HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2019 March 03.

Published in final edited form as:

Nat Genet. 2018 October ; 50(10): 1359–1365. doi:10.1038/s41588-018-0203-z.

Genome-wide association analyses identify 39 new susceptibility loci for diverticular disease

Lillias H. Maguire[1,*], **Samuel K. Handelman**[2], **Xiaomeng Du**[2], **Yanhua Chen**[2], **Tune H. Pers**[3,4], and **Elizabeth K. Speliotes**[2,5]

[1]Department of Surgery, Division of Colorectal Surgery, University of Michigan, Ann Arbor, Michigan 48109, USA

[2]Department of Internal Medicine, Division of Gastroenterology, Ann Arbor, Michigan 48109, USA

[3]The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2100, Copenhagen, Denmark

[4]Department of Epidemiology Research, Statens Serum Institut, 2300, Copenhagen, Denmark

[5]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA

Abstract

Diverticular disease is common and morbid. Treatments are limited due to poor understanding of its pathophysiology. To elucidate its etiology, we performed a genome-wide association study of diverticular disease (27,444 cases; 382,284 controls) in the UK Biobank and tested for replication

[*]Lillias H. Maguire, 2922 Taubman Center, 1500 E. Medical Center Dr., Ann Arbor, MI 48109, USA, Phone: (734) 936-7248, maguirel@med.umich.edu.

LHM: Study conception and design, data analysis and interpretation, drafting of manuscript, critical revision of manuscript, final review of submission

SKH: Data analysis and interpretation, drafting of manuscript, critical revision of manuscript, final review of submission

XD: Data acquisition, analysis and interpretation, critical revision of manuscript, final review of submission

YC: Data acquisition, analysis and interpretation, critical revision of manuscript, final review of submission

TP: Data acquisition, analysis and interpretation, critical revision of manuscript, final review of submission

EKS: Study conception and design, data acquisition, analysis and interpretation, critical revision of manuscript, final review of submission

1. Lillias H. Maguire, 2922 Taubman Center, 1500 E. Medical Center Dr., Ann Arbor, MI 48109, USA, Phone: (734) 936-7248, maguirel@med.umich.edu

2. Samuel K. Handelman, 4742A Med Sci II, 1137 Catherine Street, Ann Arbor, MI 48109, USA, Phone: (734) 647-8042, samuelkh@med.umich.edu

3. Xiaomeng Du, 4742A Med Sci II, 1137 Catherine Street, Ann Arbor, MI 48109, USA, Phone: (734) 647-8042, xmdu@med.umich.edu

4. Yanhua Chen, 4742A Med Sci II, 1137 Catherine Street, Ann Arbor, MI 48109, USA, Phone: (734) 647-8042, chenyanh@med.umich.edu

5. Tune H. Pers, Blegdamsvej 3B, 2200 København N, Denmark, Phone: (+45) 3533-5755, tune.pers@sund.ku.dk

6. Elizabeth K. Speliotes, Medical Center Drive, 6520 MSRB I, Ann Arbor, MI 48109, Phone: (734) 936-9364; Fax: (734) 763-2535; espeliot@med.umich.edu

LHM is supported by the University of Michigan Department of Surgery.

EKS, SKH, XD, and YC are supported by RO1 DK106621, RO1 DK107904, The University of Michigan Biological Sciences Scholars Program, and The University of Michigan Department of Internal Medicine. All grants made to EKS.

THP is supported by Lundbeck Foundation and Benzon Foundation.

Conflict of Interest Statement: No, I declare that the authors have no competing financial or non-financial interests as defined by Nature Research.

in the Michigan Genomics Initiative (2,572 cases; 28,649 controls). We identified 42 loci associated with diverticular disease, 39 of them novel. Using DEPICT, we show that genes in these associated regions are significantly enriched for expression in mesenchymal stem cells and multiple connective tissue cell types and are co-expressed with genes that play a role in vascular and mesenchymal biology. Genes in these associated loci play roles in immunity, extracellular matrix biology, cell adhesion, membrane transport, and intestinal motility. Phenome-wide association analysis of the 42 variants shows a common etiology of diverticular disease with obesity and hernia. These analyses establish the genomic landscape of diverticular disease.

Keywords

diverticular disease; diverticulosis; diverticulitis; genomics; genome-wide association study

Diverticulosis is an outpouching of the gastrointestinal tract present in the majority of older adults in Western countries[1]. Most patients are asymptomatic, but hundreds of thousands develop diverticular disease. Diverticulosis, the precursor lesion, is highly prevalent in the United States (US), Europe, and Canada; >50% of adults over age 60 have diverticulosis, and 10–25% will become symptomatic[2]. It is less common in other populations and demonstrates anatomic variation; diverticulosis is predominantly (90%) in the sigmoid colon in Western populations, but for unknown reasons is right-sided (70%) in Asia[3]. The low fiber Western diet has traditionally been implicated in diverticulosis, but this correlation has been questioned[4–6]. Diverticulitis (inflammation and infection of diverticula) causes >200,000 hospital admissions in the US annually[7]. The incidence is increasing; US inpatient admissions increased 26% from 1998–2005, most rapidly among patients <45 years old[8]. Complications including fistula, stricture, abscess, and intestinal perforation necessitate tens of thousands of surgical interventions annually[9]. Inpatient mortality is 1.5–3.0%[10]. Progression from diverticulosis to diverticular disease is poorly understood. Observational studies have correlated age, obesity, decreased physical activity, ultraviolet radiation, and diet with diverticular diseaes[4,11–15]. Incidence is higher in males <50, but females predominate in older ages[16]. Diverticular disease is associated with connective tissue disorders: Ehlers-Danlos Syndrome (collagen mutations)[17], Williams Syndrome (elastin mutations)[18], and polycystic kidney disease[19]. In the general population, twin studies estimate heritability at 40–53%[20,21] indicating a strong genetic component. To date, only one genome-wide association study (GWAS) has been performed, identifying three associated loci[22].

Here we report the to-date largest GWAS of diverticular disease. We examined associations of ~30 million single nucleotide polymorphisms (SNPs) with diverticular disease (ICD-10 code K57; N=27,244) in the European component of the United Kingdom Biobank (UKBB) population vs 382,284 control individuals[23]. K57 is a root code including diverticulitis and diverticular hemorrhage (Supplementary Table 1). It has been validated as a diagnostic code for diverticular disease, with a positive predictive value of 0.98[24]. Analyses were adjusted for age, sex, and principal components and relatedness using mixed linear modeling[25]. We tested the top 154 independently-associated SNPs ($p< 1\times10^{-5}$) in 31,221 unrelated European-ancestry individuals enrolled in the Michigan Genomics Initiative (MGI)[26]

adjusted for gender, age, and principal components[25]. Cases of diverticular disease in MGI were identified using ICD-9 billing codes for diverticulosis (code 562.10, 562.12, N=1,854) or diverticulitis (code 562.11, 562.13 N=718).

We identified 40 independent loci with genome-wide significant ($p < 5 \times 10–8$) associations for diverticular disease and 112 more loci with suggestive associations ($p < 1 \times 10–5$, Supplementary Table 2, Supplementary Figures 1 and 2) in UKBB. In MGI, 8/154 variants replicated with a consistent direction of effect at an MGI false discovery rate (FDR) < 10%. All loci associated with UKBB-genome-wide significant SNPs (N=40) and two MGI-replicated/UKBB-suggestive-SNPs were carried forward for analysis (Figure 1). Of these 42 loci of interest, 39 represent novel associations (Table 1). Supplementary Table 2 is a full list of associated variants. The 42 loci mapped to 99 genes within a distance of 500kb and R^2 >0.5 (Supplementary Table 3). Regional association plots better defined the associated signal (Supplementary Figure 3).

Tissue expression and pathway enrichment analyses were performed using Data-Driven Expression Prioritized Integration for Complex Traits (DEPICT) [27]. Mesenchymal stem cells and four related connective tissue cell types were enriched (FDR < 0.20). Digestive, connective, and urogenital tissues (Figure 2AB, Supplementary Table 4) were enriched (FDR < 0.20). 95 of 14,462 independent reconstituted DEPICT gene sets (FDR <0.20 and kappa of 0.5) were enriched for the 99 genes, including pathways involved in vascular biology, mesenchymal development/derivatives, and embryogenesis (Supplementary Table 5).

Of the 42 SNPs carried forward for analysis, 7 were expression quantitative trait loci (eQTLs) in sigmoid colon, and 6 were eQTLs in transverse colon, according to GTEx[28] (Table 1). The most significant eQTL-SNP was rs7086249 (NM_020752.2:c.1405–28470T>C) regulating *GPR158*, **which encodes** orphan G-protein coupled receptor, in the sigmoid colon. Mechanistic studies in fresh tissues are needed. Power to detect eQTLs is limited; post-mortem interval strongly influences colonic RNA quality[29]. 31/42 SNPs, including the 8 confirmed variants, were intronic; the remainder were intergenic.

We performed Phenome-wide association study (PheWAS) analysis for the 42 loci of interest. PheWAS can be used to agnostically assess whether phenotypes are associated with a genetic variant. Here 42 SNPs were tested for association with 780 UKBB traits[30] (Supplementary Table 6). Traits were hierarchically clustered before filtering those without significant association. Twenty-three loci correlated with morphometric traits (*ABO, BDNF, CALCA/CALCB, COL6A1, CRISPLD2, CWC27, DISP2, EFEMP1, ENSG00000224849, ENSG00000251283, FADD/ANO1, FAM185A, GTPBP1, HLX, LYPLAL1, NOV, NT5C1B, RBKS, PCSK5, S100A10, TRPS1, UBTF,* and *ZBTB4*). Fourteen loci associated with hematologic variables (*ABO, ARHGAP15, BDNF, CRISPLD2, DISP2, ENSG00000224849, GTPBP1, HLX, PPP1R14A/SPINT2, RBKS, SLC25A28, TRPS1, UBTF*, and *ZBTB4*). *LYPLAL1, GTPBP1, ELN, EFEMP1,* and *CRISPLD2* associated with hernias. *EFEMP1* and *CRISPLD2* also associated with female genital prolapse *EFEMP1* has been previously associated with hernia[31]. *SHFM1, UBTF, HLX, ABO,* and *UNC50* associated with connective tissue traits, such as osteoarthritis and soft tissue inflammation.

Eight loci (*ABO, CACNB2, ENSG00000224849, FADD/ANO1, NOV, NT5C1B, RBKS,* and *ZBTB4*) associated with vascular traits including venous thrombosis, pulmonary embolism, hypertension, and heart failure. Nineteen loci (*ABO, ARHGAP15, COLQ/METTL6, CRISPLD2, EFEMP1, ELN, ENSG00000224849, ENSG00000251283, FADD/ANO1, FAM155A, FAM185A, GPR158, GTPBP1, P2RY12/P2RY14, PPP1R14A/SPINT2, RBKS, SLC25A28, SLC35F3,* and *UNC50*) associated with gastrointestinal disease, but not with the common bowel conditions inflammatory bowel disease, polyps, and cancer. An edited heatmap (Figure 3) highlights these effects.

One prior GWAS identified risk-loci near *ARHGAP15, FAM155A* and *COLQ*[22]. These associations were confirmed, supporting the validity of our approach. *ARHGAP15* encodes a GTPase-activating protein acting on Rac[32] and negatively regulates neutrophils[33]. The function of the gene product of *FAM155A* is unknown. *COLQ* encodes a critical protein for acetylcholine-mediated signaling[34]. *CALCB*, identified in our study, was identified but not validated in the prior GWAS. *TNFSF15* has been associated with diverticular disease[35], but this was not found in our study. Despite clinical association[17], Ehlers-Danlos genes were not identified.

The 8 replicated loci were associated with 21 genes (Supplementary Table 3). Some contribute to cytoskeletal and extracellular matrix (ECM) dynamics (*ELN, SHFM1, BMPR1B, LIMK1,* and *CLSTN2*). *BMPR1B* and *SHFM1* are implicated in bone and cartilage synthesis[36,37]. *LIMK1* stabilizes the cytoskeleton by inhibiting actin de-polymerization[38]. *CLSTN2* encodes an atypical cadherin involved in cell adhesion[39]. *ELN* encodes elastin, which is altered in diverticular colons[40]. Diverticular disease is common in Williams Syndrome, a congenital disorder caused by *ELN* hemizygosity[18]. *ANO1* encodes a chloride channel in the intestinal pacemaker cells of Cajal[41]. These cells are reduced in diverticular disease[42]. *ANO1* is critical for intestinal contractility[43]. Altered intestinal motility is implicated in diverticular disease[44]. Diverticular colons demonstrate abnormal smooth muscle morphology[45] and altered contractile force[46]. *ARHGAP15, GPR158,* and *GTPBP1* are involved in GTP-signaling. Many identified genes have unknown function or unclear functional link to diverticular disease. Functional characterization should be prioritized to confirm these gene-variant associations. In the absence of strong molecular evidence to the contrary, systematic studies indicate that the closest gene is the best candidate for SNP effect[47]. All replicated SNPs were located in introns, supporting a molecular mechanism at the RNA-expression level in the surrounding gene. Therefore, expression levels of these genes is the most plausible avenue for further molecular phenotyping[48].

Among our other 99 identified genes, many have roles in the ECM, motility, and membrane transport (Figure 4), *COL6A1, CRISPLD2, EFEMP1*, *HAS2, NOV*, and *TCHH* have known roles in the ECM[49–53]. Enrichment in mesenchymal stem cells, connective tissues, and mesenchymal development pathways, suggest a role for connective tissue biology. *PPP1R14A* and *CHRNB1* effect smooth muscle motility[54,55]. Others are involved in transport of copper (*CUTC*), sodium (*SPINT2*) and calcium (*CALCA, CALCB, CACNB2*). *SPINT2* mutations result in congenital sodium diarrhea[57]. Altered absorption or motility could produce constipation, which is clinically associated with diverticular disease. Vascular

biology identified by pathway analysis/PheWAS may be relevant as diverticula tend to occur adjacent to penetrating arteries.

This study is limited in that it detects diverticular disease via inpatient coding, and does not identify asymptomatic diverticulosis. Given the epidemiology of diverticulosis, the majority of participants likely harbor the precursor lesion and the variants identified only associate with diverticular disease. However, this is the clinically relevant outcome. Given the high reliability of diverticular disease codes[24] and the derivation of cases from inpatient hospital admissions, it is likely that most patients suffered severe diverticular disease. However, patients might be erroneously identified if diverticulosis was noted incidentally. Conversely, patients with mild diverticular disease treated as outpatients may be falsely identified as controls. The de-identified nature of the data precludes coding confirmation. Another limitation is ICD9 versus ICD10 coding between populations. We chose grouped, rather than individual codes to mitigate this difference. Additionally, the UKBB entry age of 40–69 prohibits comparison of older/younger patients. Finally, some conditions in our PheWas could be a consequence of diverticular disease rather than sharing a common etiology.

In summary, the biologic basis for both the development of colonic diverticula in the majority of older adults and the triggers that produce diverticulitis in some patients are unknown. We report the largest GWAS thus far for diverticular disease and identify 39 novel loci as contributing to the pathophysiology of these diseases. This work defines the landscape for future functional studies and identifies possible targets for therapeutic development.

Online Methods

UK Biobank

The UK Biobank (UKBB) contains genotypes, clinical and demographic data on over 400,000 individuals aged 40–69 at time of study recruitment. The UKBB protocols were approved by the National Research Ethics Service Committee. Participants signed written informed consent, specifically applicable to health-related research. All ethical regulations were followed. The analyses used in this paper were carried out by Canela-Xandri et al. under UK Biobank Resource project 788[30]. Diverticular disease was recorded under the International Classification of Disease (ICD) 10 code K57 (N=27,244). Participant genotyping, data collection, and quality control has been described in detail[23]. In brief, participants were genotyped on one of two purpose-designed arrays (UK BiLEVE Axiom Array (N=50,520) and UK Biobank Axiom Array (N=438,692)) with 95% maker overlap. The Haplotype Reference Consortium was used as a reference panel to phase and impute the data. Following quality control, over 30 million variants in 408,455 white British individuals (http://geneatlas.roslin.ed.ac.uk/) were tested for association with K57 controlling for age, gender, principal components and relatedness using mixed linear modeling.

Michigan Genomics Initiative

The Michigan Genomics Initiative (MGI) is an institutional repository of DNA linked to participants' medical profile via the electronic medical record at the University of Michigan.

The MGI is approved for this research use by the Institutional Review Board of the University of Michigan. All relevant ethical regulations were followed. Diverticular disease was derived from ICD-9 codes (562.11 and 562.13 for diverticulitis and 562.10 and 562.12 for diverticulosis). Following informed consent, individuals (N=35,888) were genotyped using the Illumina HumanCoreExome Array.

Genotype analysis was performed with Illumina GenomeStudio (module 1.9.4, algorithm GenTrain 2.0). After initial clustering, we re-defined variant cluster boundaries using individuals with call rate >99% and genotyped the remaining samples. Samples were excluded if they met any of the following criteria: (1) call rate <99%, (2) estimated contamination >2.5% (BAF Regress)[57], (3) large chromosomal copy number variants, (4) lower call rate of a technical duplicate pair and twins, or (5) whose inferred sex contradicted the reported sex.

Variant-quality control was performed in the following manner: we excluded variants if: (1) their probes could not be perfectly mapped or mapped perfectly to multiple positions (2) they showed deviations from Hardy Weinberg equilibrium (p-value< 0.0001), (3) had a call rate < 99%, or (4) another variant with higher call rate assayed the same variant (PLINK (v1.90)[58]).

Phasing was carried out using SHAPEIT2 (v2. r837)[59] on autosomal chromosomes. Genotypes of the Haplotype Reference Consortium (chromosome 1–22: HRC release 1; chromosome X: HRC release 1.1) were imputed into the phased MGI data using Minimac3 (v1.0.13)[60]. Excluding variants with low imputation quality (R2 <0.3) resulted in dense mapping at 39,127,678 million quality-imputed genetic markers.

We estimated pairwise kinship using the software KING (v1.4.2)[61]. We excluded any 1st- or 2nd-degree relative pairs within the cohort. In addition, we used principal component analysis to identify ethnically homogeneous groups using individuals from the Human Genome Diversity Project[64]. We included only European samples.

Locus Identification

197 independent loci were identified for all imputed UKBB variants associated with diverticular disease at p <1×10–5 using criteria of R^2 < 0.1 within a distance of 500kb using PLINK version 1.90b4.6[62] within the DEPICT program[27]. DEPICT then assigned each SNP to genes in the specified region or genes containing variants in linkage disequilibrium with the SNP. SNPs were then queried for replication in MGI using a nominal one sided FDR of 10% by Benjamini-Hochberg[63].

SNP Annotation

Effect allele and allele frequencies were annotated using ANNOVAR chromosome 1–22 imputation data, build 37[64].

Tissue and Pathway Analysis

Tissue and pathway enrichment was carried out against 14,462 reconstituted gene sets in DEPICT[27] (version 1, release 194) for the 192 loci associated with diverticular disease at a

nominal p-value below 1×10–5. Pathways were culled using a kappa statistic of 0.5[65]. Tissue and cell type enrichment was similarly determined in DEPICT by analyzing gene expression enrichment of genes at our 192 loci of interest in 209 MeSH-defined tissue and cell types. FDR of <0.20 was set as a threshold for significance for both pathway and tissue analysis.

Colon eQTLs

Lists of eQTLs in sigmoid or transverse colon were obtained from GTEx version 7. The GTEx project has been described in detail elsewhere[28]. Briefly it is a gene expression resource created from RNA Sequencing (RNA-Seq) results obtained from post-mortem donors. Gene expression levels and individual variants were correlated to enable discovery of 697,430 gene-variant associations in sigmoid colon, and 832,983 gene-variant associations in transverse colon. In both cases, a false discovery rate below 5% was used.

PheWAS

Phenome-wide association study was carried out for all lead SNPs in our loci of interest. SNPs were queried against 778 traits ascertained for UKBB participants and reported in the Roslin Gene Atlas, including morphometric data, hematologic lab values, ICD-10 clinical diagnoses, and self-reported conditions. First, traits were hierarchically clustered using inverse-absolute Pearson correlation among the Z-scores as a distance metric. The resultant hierarchical clustering/tree was pruned at a height corresponding h=0.2, leaving a total of 97 largely independent traits. Then, the pruned matrix of trait-genotype associations was filtered at an FDR of 0.05 by Benjamini-Hochberg[63]. This filtered association matrix was used in further analysis and reporting.

Statistics

All p-values described in the manuscript are two-sided. Multiple comparison corrections were made using the method of Benjamini-Hochberg[63] at multiple points during the study, as detailed above.

Data Availability Statement

The UK BioBank genomic and phenotypic data supporting this publication are publicly available from the Roslin Institute, University of Edinburgh (http://geneatlas.roslin.ed.ac.uk/). The Michigan Genomics Initiative (MGI) genomic and phenotypic data are not publicly available due to restrictions on participant privacy. MGI data can be made available on reasonable request to the corresponding author with permission of the University of Michigan Institutional Review Board. Detailed information on software, study design, and data availability can be found in the Life Sciences Reporting Summary associated with this manuscript.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

The authors acknowledge the University of Michigan Medical School Central Biorepository/Michigan Genomics Initiative for providing biospecimen storage, management, and distribution services in support of the research reported in this publication.Graphics in Figure 4 obtained courtesy of Servier Medical Art, Les Laboratoires Servier (https://smart.servier.com). Used under Creative Commons License.

References

1. Painter NS & Burkitt DP Diverticular disease the colon, a 20th century problem. Clin Gastroenterol 4, 3–21 (1975). [PubMed: 1109818]
2. Weizman AV & Nguyen GC Diverticular disease: epidemiology and management. Can J Gastroenterol. 25, 385–389 (2011). [PubMed: 21876861]
3. Sugihara K et al. Diverticular disease of the colon in Japan: A review of 615 cases. Dis Colon Rectum 27, 531–7 (1984). [PubMed: 6468190]
4. Pan G, et al. Diverticular disease of the colon in China. Chin Med J 97, 391–4. (1984). [PubMed: 6437755]
5. Alatise OI et al. Spectrum of colonoscopy findings in Ile-Ife Nigeria. Niger Postgrad Med J. 19, 219–24 (2012). [PubMed: 23385677]
6. Peery AF et al. A high-fiber diet does not protect against asymptomatic diverticulosis. Gastroenterology 142, 266–72 (2012). [PubMed: 22062360]
7. Peery AF et al. Burden of gastrointestinal disease in the United States: 2012 update. Gastroenterology 143, 1179–1187 (2012). [PubMed: 22885331]
8. Etzioni DA, Mack TM, Beart RW, Jr, & Kaiser AM Diverticulitis in the United States: 1998–2005: changing patterns of disease and treatment. Ann Surg. 249, 210–7 (2009). [PubMed: 19212172]
9. Ricciardi R,et al. Is the decline in the surgical treatment for diverticulitis associated with an increase in complicated diverticulitis? Dis Colon Rectum 52, 1558–63 (2009). [PubMed: 19690482]
10. Delvaux M Diverticular disease of the colon in Europe: epidemiology, impact on citizen health and prevention. Aliment Pharmacol Ther. 18, 71–74. (2003). [PubMed: 14531745]
11. Strate LL, Liu YL, Aldoori WH & Giovannucci EL Physical activity decreases diverticular complications. Am J Gastroenterol 104, 1221 (2009). [PubMed: 19367267]
12. Strate LL, Liu YL, Aldoori WH, Syngal S, & Giovannucci EL Obesity increases the risks of diverticulitis and diverticular bleeding. Gastroenterology 136, 115–22 (2009). [PubMed: 18996378]
13. Maguire LH, Song M, Strate LL, Giovannucci EL & Chan AT Higher serum levels of vitamin D are associated with a reduced risk of diverticulitis. Clin Gastroenterol Hepatol. 11, 1631–5 (2013). [PubMed: 23954650]
14. Maguire LH, Song M, Strate LL, Giovannucci EL & Chan AT Association of geographic and seasonal variation with diverticulitis admissions. JAMA Surg 1, 74–77 (2015).
15. Strate LL, Liu YL, Syngal S, Aldoori WH, & Giovannucci EL Nut, corn, and popcorn consumption and the incidence of diverticular disease. JAMA 300, 907–14 (2009).
16. Warner E, et al. Fourteen-year study of hospital admissions for diverticular disease in Ontario. Can J Gastroenterol. 21: 97–9 (2007). [PubMed: 17299613]
17. Leganger J, et al. Association between diverticular disease and Ehlers-Danlos syndrome: a 13-year nationwide population-based cohort study. Int J Colorectal Dis. 12, 1863–1867 (2016).
18. Cherniske EM et al. 2004. Multisystem study of 20 older adults with Williams syndrome. Am J Med Genet 131A, 255–264 (2004).
19. Lederman ED, McCoy G, Conti DJ, & Lee EC Diverticulitis and polycystic kidney disease. Am Surg. 66, 200–203 (2000). [PubMed: 10695753]
20. Granlund J, et al. The genetic influence on diverticular disease–a twin study. Aliment Pharm Ther 35, 1103–7 (2012).
21. Strate LL, et al. Heritability and familial aggregation of diverticular disease: a population-based study of twins and siblings. Gastroenterology 144, 736–42 (2013). [PubMed: 23313967]

22. Sigurdsson S, et al. Sequence variants in ARHGAP15, COLQ and FAM155A associate with diverticular disease and diverticulitis. Nat Commun 8, 15789 (2017). [PubMed: 28585551]
23. Sudlow C, 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 12: e1001779 (2015). [PubMed: 25826379]
24. Erichsen R, Strate L, Sørensen HT & Baron JA Positive predictive values of the International Classification of Disease, 10th edition diagnoses codes for diverticular disease in the Danish National Registry of Patients. Clin Exp Gastroenterol 3, 139–42 (2010). [PubMed: 21694857]
25. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904–9 (2006). [PubMed: 16862161]
26. Dey R, Schmidt EM, Abecasis GR, & Lee S A Fast and Accurate Algorithm to Test for Binary Phenotypes and Its Application to PheWAS. Am J Hum Genet 101, 37–49. (2017). [PubMed: 28602423]
27. Pers TH, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun 6, 5890 (2015). [PubMed: 25597830]
28. Consortium GTEx. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017) [PubMed: 29022597]
29. Ferreira PG et al. The effects of death and post-mortem cold ischemia on human tissue transcriptomes. Nat Commun. 9, 490 (2018). [PubMed: 29440659]
30. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. bioRxiv 176834; doi: 10.1101/176834
31. Jorgenson E et al. A genome-wide association study identifies four novel susceptibility loci underlying inguinal hernia. Nat Commun 6, 10130 (2015). [PubMed: 26686553]
32. Seoh ML, Ng CH, Yong J, Lim L & Leung T ArhGAP15, a novel human RacGAP protein with GTPase binding property. FEBS letters 539, 131–7 (2003). [PubMed: 12650940]
33. Costa C, et al. The RacGAP ArhGAP15 is a master negative regulator of neutrophil functions. Blood 118, 1099–108 (2011). [PubMed: 21551229]
34. Arredondo et al. COOH-terminal collagen Q (COLQ) mutants causing human deficiency of endplate acetylcholinesterase impair the interaction of ColQ with proteins of the basal lamina. Hum Genet 133, 599–616 (2014). [PubMed: 24281389]
35. Connelly TM, Berg AS, Hegarty JP, Deiling S, Brinton D, Poritz LS, Koltun WA.The TNFSF15 gene single nucleotide polymorphism rs7848647 is associated with surgical diverticulitis. Ann Surg. 259, 1132–7 (2014) [PubMed: 24814505]
36. Racacho L et al. Two novel disease-causing variants in BMPR1B are associated with brachydactyly type A1. Eur J Hum Genet 23, 1640–5 (2015). [PubMed: 25758993]
37. Rasmussen MB, et al. Phenotypic subregions within the split-hand/foot malformation 1 locus. Hum Genet 135, 345–57 (2016) [PubMed: 26839112]
38. Mashiach-Farkash E et al. Computer-based identification of a novel LIMK1/2 inhibitor that synergizes with salirasib to destabilize the actin cytoskeleton. Oncotarget 6, 629–39 (2012).
39. Ortiz-Medina H, Emond MR, & Jontes JD Zebrafish calsyntenins mediate homophilic adhesion through their amino-terminal cadherin repeats. Neuroscience. 286 87–96 (2014). [PubMed: 25463516]
40. Whiteway J & Morson BC Elastosis in diverticular disease of the sigmoid colon. Gut 26, 258–66 (1985). [PubMed: 3972272]
41. Gomez-Pinilla PJ, et al. Ano1 is a selective marker of interstitial cells of Cajal in the human and mouse gastrointestinal tract. Am J Physiol Gastrointest Liver Physiol 296, G1370–81 (2009) [PubMed: 19372102]
42. Bassotti G, et al. Interstitial cells of Cajal, enteric nerves, and glial cells in colonic diverticular disease. J Clin Pathol 58, 973–7 (2005). [PubMed: 16126881]
43. Cobine CA, et al. ANO1 in intramuscular interstitial cells of Cajal plays a key role in the generation of slow waves and tone in the internal anal sphincter. J Physiol 595, 2021–2041 (2017). [PubMed: 28054347]

44. Jeyarajah S & Papagrigoriadis S Review article: the pathogenesis of diverticular disease--current perspectives on motility and neurotransmitters. Aliment Pharmacol Ther 33, 789–800 (2011). [PubMed: 21306406]
45. Hughes LE Postmortem survey of diverticular disease of the colon. II. The muscular abnormality of the sigmoid colon. Gut 10, 344–51 (1969). [PubMed: 5771666]
46. Maselli MA, et al. Colonic smooth muscle responses in patients with diverticular disease of the colon: effect of the NK2 receptor antagonist SR48968. Dig Liver Dis 36, 348–54 (2004). [PubMed: 15191205]
47. Stacey D et al. ProGeM: A framework for the prioritisation of candidate causal genes at molecular quantitative trait loci. bioRxiv 230094. doi: 10.1101/230094
48. Richards S et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 17, 405–424. (2015) [PubMed: 25741868]
49. Himes BE et al. RNA-Seq transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells. PLoS One 6, e99625 (2014).
50. Klenotic PA, Munier FL, Marmorstein LY & Anand-Apte B Tissue inhibitor of metalloproteinases-3 (TIMP-3) is a binding partner of epithelial growth factor-containing fibulin-like extracellular matrix protein 1 (EFEMP1). Implications for macular degenerations. J Biol Chem 29, 30469–73 (2004).
51. Qu C, et al. Extensive CD44-dependent hyaluronan coats on human bone marrow-derived mesenchymal stem cells produced by hyaluronan synthases HAS1, HAS2 and HAS3. Int J Biochem Cell Biol 48, 45–54 (2014). [PubMed: 24406795]
52. Yeger H & Perbal B CCN family of proteins: critical modulators of the tumor cell microenvironment. J Cell Commun Signal 10, 229–240 (2016). [PubMed: 27517291]
53. Lettmann S, et al. Col6a1 null mice as a model to study skin phenotypes in patients with collagen VI related myopathies: expression of classical and novel collagen VI variants during wound healing. PLoS One 8, e105686 (2014).
54. Mori D, et al. Synchronous phosphorylation of CPI-17 and MYPT1 is essential for inducing Ca(2+) sensitization in intestinal smooth muscle. Neurogastroenterol Motil 23, 1111–22 (2011). [PubMed: 22004286]
55. Akk G, et al. Energetic contributions to channel gating of residues in the muscle nicotinic receptor β1 subunit. PLoS One 8, e78539 (2013). [PubMed: 24194945]
56. Janecke AR, Heinz-Erian P & Müller T Congenital Sodium Diarrhea: A Form of Intractable Diarrhea, With a Link to Inflammatory Bowel Disease. J Pediatr Gastroenterol Nutr 63, 170–6 (2016). [PubMed: 26835907]
57. Jun G et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet 91, 839–848, (2012). [PubMed: 23103226]
58. Chang CC et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015). [PubMed: 25722852]
59. O'Connell J et al. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet 10, e1004234 (2014). [PubMed: 24743097]
60. Das S et al. Next-generation genotype imputation service and methods. Nat Genet 48, 1284–1287, (2016). [PubMed: 27571263]
61. Manichaikul A et al. Robust relationship inference in genome-wide association studies. Bioinformatics 26, 2867–2873 (2010). [PubMed: 20926424]
62. Purcell S, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics 81, 559–571 (2007). [PubMed: 17701901]
63. Benjamini Y & Hochberg Y Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 1, 289–300 (1995).
64. Wang K, Li M & Hakonarson H ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data. Nucleic Acids Research 38, e164 (2010). [PubMed: 20601685]
65. Huang DW, Sherman BT & Lempicki RA Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc 4, 44–57 (2009). [PubMed: 19131956]

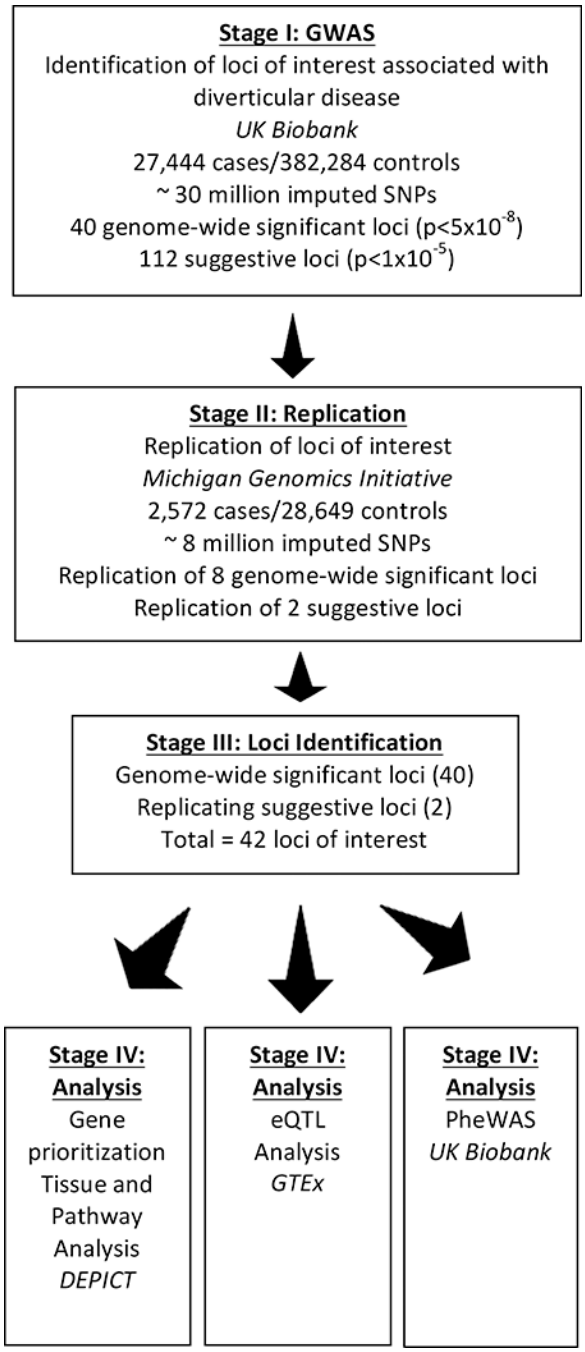
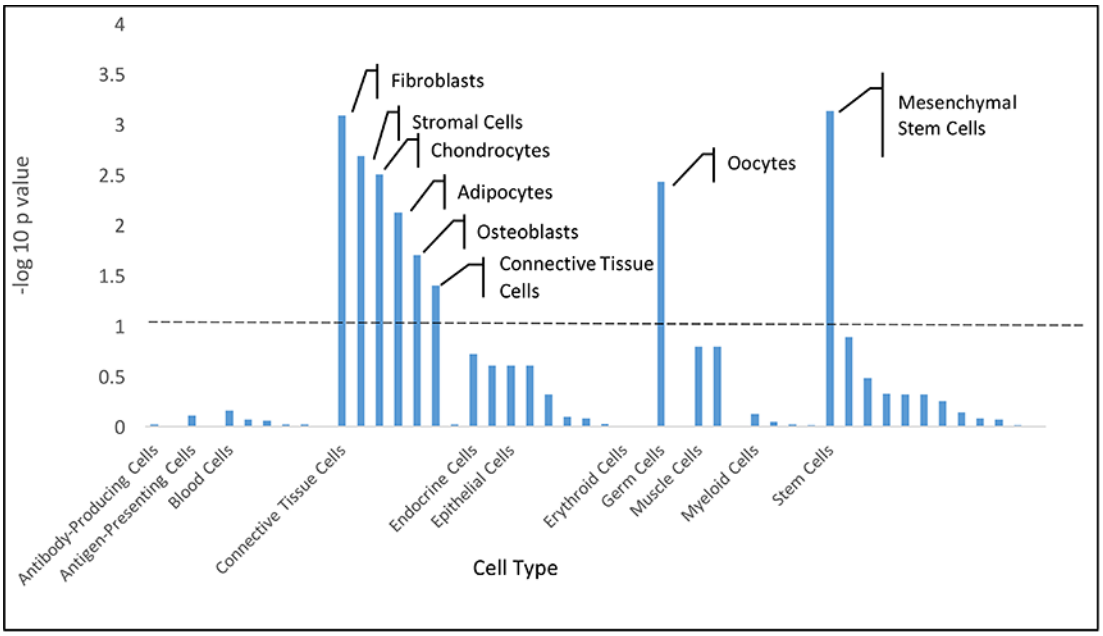


Figure 1:

Study Design. Graphic representation of study design. GWAS: Genome-wide association study. SNPs: single nucleotide polymorphisms. PheWAS: Phenome-wide association study. GTEx: Genotype-Tissue Expression project. DEPICT: Data-driven Expression-Prioritized Integration for Complex Traits. eQTL: Expression Quantitative Trait Loci



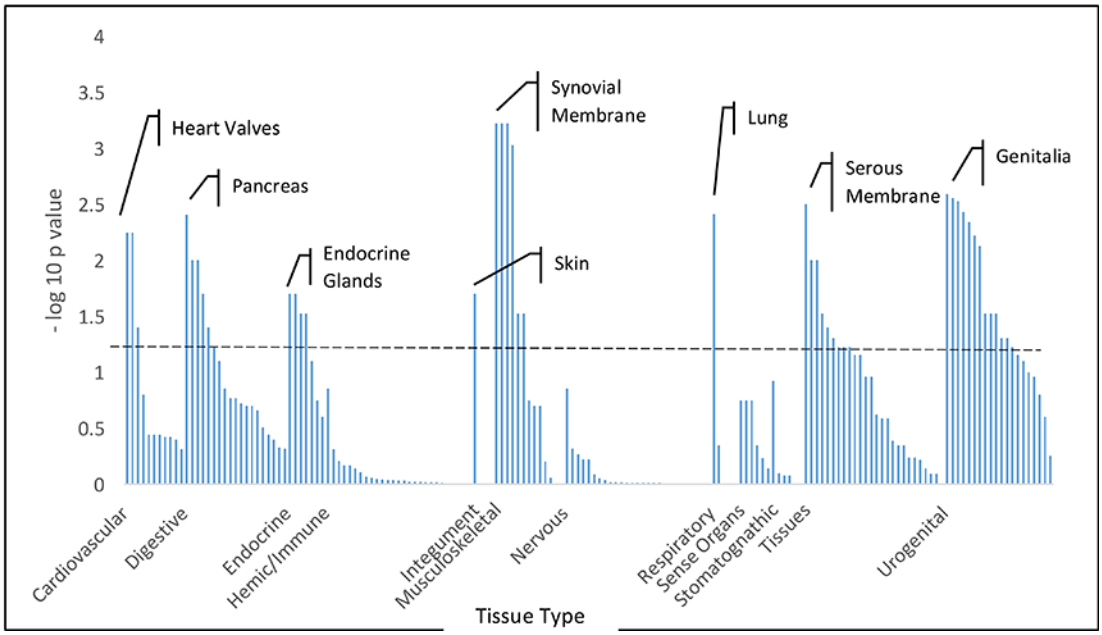


Figure 2A/B:

Tissue and cell type enrichment analysis. Plots showing the enrichment of loci associated with diverticular disease ($p < 1 \times 10{-}5$ in the UKBB; N=27,444 cases/382,284 controls) in specific cell types (A) and tissues (B). Enrichments are grouped according to system or cell type and significance; annotations above the dashed line have FDR ≤ 0.20. Data corrected for multiple comparisons using Benjamini-Hochberg method. Top tissue in each category labelled.

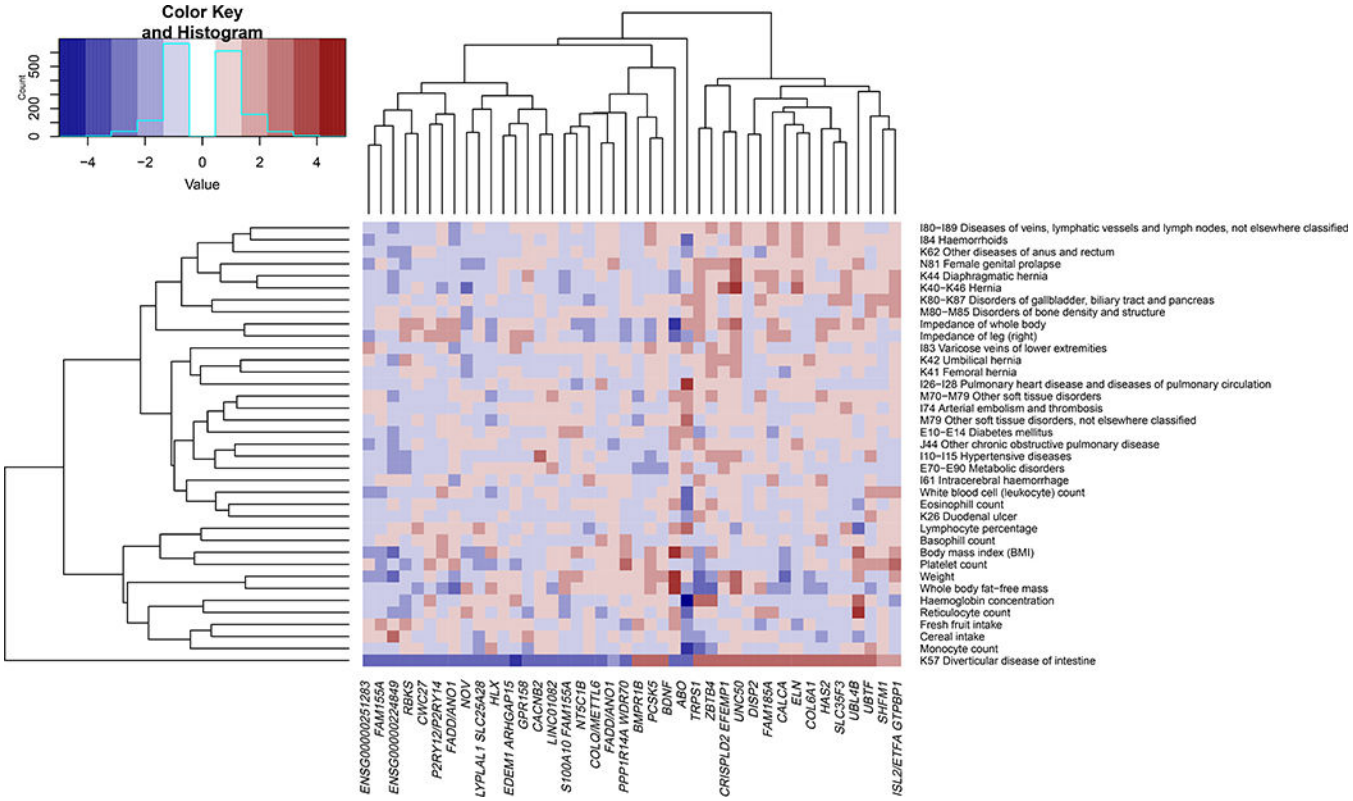


Figure 3:

Phenome-wide association matrix. Filtered association matrix highlighting vascular, gastrointestinal, connective tissue, hematologic, morphometric, and dietary traits associated with loci of interest in the UKBB (27,444 cases/382,284 controls) Data controlled for multiple comparisons using Benjamini-Hochberg method, filtered at FDR<5%, and clustered at h=0.2.

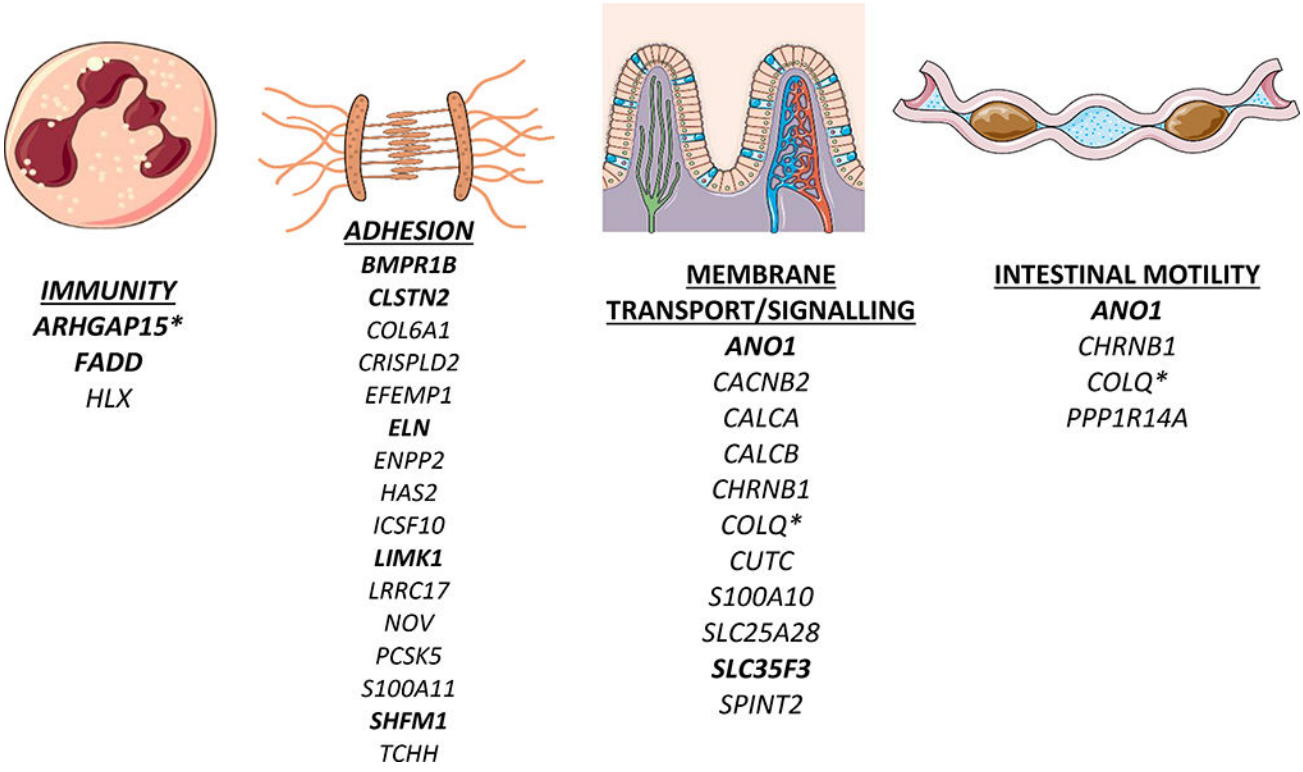


Figure 4:

Plausible biological pathways underlying risk loci associated with diverticular disease. Bold gene symbols indicate replication in MGI. * indicates prior identification in GWAS

Table One:

Loci of interest including genome-wide significant variants ($p<5\times10^{-8}$) from UKBB and highly significant SNPs ($p<5\times10^{-5}$) with replication in MGI. Bold gene symbols indicate replication in MGI at FDR <0.1 following Benjamin-Hochberg correction. X–chromosome, DD – diverticular disease, FDR – false discovery rate, GWAS – genome wide association study, EAF – estimated allele frequency. At each locus, a superscript1 or 2 indicates an eQTL and eGene for GTEx v7 sigmoid colon or transverse colon, respectively.

UK Biobank (27,444 cases/382,284 controls)									Michigan Genomics Initiative (2,572 cases/28,649 controls)	
Locus	**SNP**	**Chr**	**Position**	**Nearest Gene**	**Other Nearby Genes**	**EA**	**EAF**	**DD p- value**	**p- diverticulosis**	**p- diverticulitis**
1	**rs6734367**	2	143556678	***ARHGAP15****		**G**	**0.82**	**4.29E-44**	**0.0038**	0.07
2	**rs4333882**	**1**	**234217153**	***SLC35F3***		**G**	**0.19**	**4.44E-22**	**0.0007**	0.14
3	rs7609897	3	15461174	*COLQ*/METTL6*	*EAF1*	T	0.22	2.72E-18	0.10	0.0096
4	**rs7086249[1]**	10	25522506	***GPR158[1]***		**C**	**0.46**	**5.37E-16**	0.014	**0.0004**
5	rs1802575	2	55866069	*EFEMP1*		C	0.13	7.71E-16	0.08	0.07
6	rs11667256	19	38245164	*PPP1R14A*	*SPINT2*	T	0.52	1.24E-14	0.025	0.062
7	rs962369	11	27712873	*BDNF*	*BDNF-AS1, KIF18A, METTL15, ENSG00000255496*	C	0.31	2.16E-14	0.18	0.74
8	rs6949391[2]	7	102806416	*FAM185A, PMPCB, LRRC17*	*NAPEPLD, S100A11P1, UPK3BL1[2], DNACJ2, DPY19L2P2, ARMC10, FBXL13, ENSG00000230257*	T	0.34	3.74E-14	0.17	0.45
9	rs61823192	1	219121228	*LYPLAL1*	*ENSG00000223842*	T	0.03	1.15E-13	0.16	0.021
10	rs9520344	13	107250422	***FAM155A****		A	0.24	5.23E-12	0.08	0.34
	rs11619840		**107566610**			**A**	**0.19**	**1.70E-09**	**0.0042**	0.021
11	rs7098322	10	99631412	*SLC25A28*	*ENTPD7, CUTC, COX15*	T	0.87	9.94E-12	0.67	0.33
12	rs10472291	5	37772678	*WDR70*		A	0.33	1.01E-11	0.21	0.89
13	**rs582094[1,2]**	**9**	**136145484**	***ABO[1,2]***		**T**	**0.32**	**1.55E-11**	0.12	**0.0008**
14	rs75434097	21	45999606	*COL6A1*	*PCBP3*	A	0.15	4.90E-11	0.42	0.98
15	rs2280028	16	86199807	*LINC01082*		A	0.14	7.05E-11	0.73	0.10
16	rs9856118[2]	3	151360428	*P2RY12[2], P2RY14*	*GPR87, IGSF10, P2RY13, GPR171, MED12L*	G	0.17	8.80E-11	0.15	0.82
17	rs71472433[1]	15	40357408	*DISP2[1]*	*IVD, C15orf23*	C	0.17	8.90E-11	0.74	0.89

UK Biobank (27,444 cases/382,284 controls)									Michigan Genomics Initiative (2,572 cases/28,649 controls)	
Locus	**SNP**	**Chr**	**Position**	**Nearest Gene**	**Other Nearby Genes**	**EA**	**EAF**	**DD p- value**	**p- diverticulosis**	**p- diverticulitis**
18	rs2131755	16	84823772	*CRISPLD2*		G	0.41	1.50E-10	0.06	0.59
19	rs4839715	6	97917413	*ENSG00000224849*		A	0.37	1.62E-10	0.87	0.26
20	rs148376933	2	98612512	*UNC50*				1.88E-10	0.91	0.49
21	rs1381335[1,2]	8	119415408	*NOV*[1,2]	*ENPP2*	T	0.24	1.91E-10	0.69	0.64
22	rs61814883[1]	1	151998153	*S100A10*[1]	*S100A11, TCHH, THEM4*	A	0.30	2.05E-10	0.46	0.60
23	rs8074740	17	44235410	*UBTF*	*ASB16, TMUB2, ATXN7L3, ASB16-AS1, C17orf53*	A	0.32	2.34E-10	0.88	0.08
24	**rs3113037**	7	96449252	***SHFM1***	***CLSTN2***	**T**	**0.23**	**2.52E-10**	**0.0039**	0.0033
25	rs12293535	11	14993308	*CALCA*	*CALCB*	A	0.28	6.20E-10	0.37	0.47
26	**rs875107**	**11**	**70159268**	***FADD***	***ANO1***	**A**	**0.52**	**2.33E-09**	**0.0004**	0.10
	rs72945112		**70247466**			**T**	**0.15**	**6.30E-06**	**0.0016**	0.80
27	**rs3823878**	**7**	**74028915**	***ELN***	***LIMK1***	**A**	**0.06**	**2.63E-09**	**0.0018**	0.78
28	rs10471645	5	64999536	*CWC27*		C	0.83	3.03E-09	0.037	0.0091
29	rs1888693	10	18151515	*CACNB2*		A	0.34	3.58E-09	0.10	0.50
30	rs4871180	8	121246834	*HAS2*		T	0.25	4.15E-09	0.77	0.77
31	rs2049865	8	115576319	*TRPS1*		A	0.58	5.54E-09	0.59	0.52
32	**rs1544387**	**4**	**94852434**	***BMPR1B***		**T**	**0.58**	**5.74E-09**	0.023	**0.0005**
33	rs11934833	4	156636431	*ENSG00000251283*	*ENSG00000249479 ENS00000251511*	G	0.30	6.21E-09	0.0053	0.30
34	rs2784255	1	220893031	*HLX*		C	0.48	1.06E-08	0.99	0.61
35	rs10120333	9	76125437	*PCSK5*		G	0.53	1.54E-08	0.92	0.39
36	rs12942267[1,2]	17	7469318	*ZBTB4*	*CHRNB1*[1,2], *POLR2A*	T	0.64	2.55E-08	0.042	0.73
37	rs62126581	2	18806974	*NT5C1B*		A	0.17	3.77E-08	0.62	0.23
38	rs115490395	1	110120397	*UBL4B*		A	0.01	4.39E-08	0.42	0.81
39	rs2470653	3	5804815	*EDEM1*		A	0.23	4.51E-08	1.00	0.20
40	rs10173528	2	28065525	*RBKS*	*GPN1, BRE, SUPT7L, MRPL33*	T	0.61	4.73E-08	0.82	0.94
47	rs2056544	15	76533662	***ISL2, ETFA***	***RCN2, TMEM266, SCAPER***	A	0.57	1.01E-07	0.99	0.08

UK Biobank (27,444 cases/382,284 controls)								Michigan Genomics Initiative (2,572 cases/28,649 controls)	
Locus	SNP	Chr	Position	Nearest Gene	Other Nearby Genes	EA	EAF	DD p- value	p- diverticulosis	p- diverticulitis
	rs10519134		**76286749**			**A**	**0.11**	**7.12E-06**	**0.0036**	0.66
68	**rs138699**[1,2]	**22**	**38733703**	***GTPBP1***	***AL021707.2***[1,2], ***CBY1***[1], ***FAM227A***[2]	**A**	**0.25**	**8.02E-07**	**0.0039**	0.25