SUPPLEMENTARY MATERIALS AND METHODS

Phenotype definition in the UK Biobank

An individual was classified as a diverticular disease case if they matched hospital-bases ICD9 or ICD10 coding (562 and K57 respectively) for primary (n = 16,560), secondary (n = 13,375) and self-reported diverticular disease diagnosis (n = 1,982; information collected and placed within the coding tree during the verbal interview at the assessment clinic by a trained nurse) in the UK Biobank dataset. Control individuals where classified on the basis of absence of a diverticular disease diagnosis (n = 419,135). We did not have complete information on the proportion of diverticulitis patients in the UK Biobank as depth of ICD coding was insufficient to differentiate disease subtype diverticulosis (i.e. diverticular disease without inflammation) from diverticulitis. The majority of cases had diagnosis K57.3 coding for both disease subtypes diverticulosis without perforation or abscess (K57.30, K57.31) and diverticulitis without perforation or abscess (K57.32, K57.33) not allowing this differentiation at available coding depth K57.3.

Recruitment and phenotyping of the German replication samples

The Northern and Western German samples were phenotyped as follows: Controls and cases with uncomplicated diverticulosis were defined by manual review of colonoscopy (complete colonoscopy required) and patient records at participating hospitals and gastroenterology outpatient services. In addition, the patient questionnaire was reviewed for absence of diverticulitis (including a hospital or outpatient diagnosis of diverticulitis or episodes of lower left quadrant paint and fever in diverticulosis patients). Cases with diverticulitis were defined either as patients with colonoscopy proven diverticular disease, that required antibiotics due to diverticulitis or as patients diagnosed by abdominal ultrasound or CT scan. To obtain patient information on additional cases with diverticulitis, hospital or medical office information systems were screened for ICD- Code K57.x and OPS Code 5-455.7 for resection of the sigma and patient records where then further reviewed as described above. All patients were of self-reported Caucasian ancestry. Patients with carcinoma or inflammatory bowel

disease (IBD) were excluded from all groups. Patients from the Germany / North cohort were recruited through the popgen biobank as described previously [1,2]. Patients from the Germany / West cohort were recruited at the Department of Medicine II, Saarland University Medical Center, Homburg between 2012 and 2017. All German participants provided written informed consent. The study protocol was approved by the Research Ethics Committees of the Saarland University (approval 63/11), the Medical faculty of Christian-Albrechts-University Kiel, Germany (A 156/03) and the Medical Faculty of the Technische Universität Dresden (EK470122013).

Austrian replication samples

For the Austria / Vienna cohort, the ongoing molecular epidemiology colorectal cancer study of Austria (CORSA) was used. More than 16,000 Caucasian participants were recruited since May 2003 through the province-wide screening program "Burgenland Prevention Trial of Colorectal Disease with Immunological Testing" (B-PREDICT). All habitants of Burgenland aged between 40 and 80 years are invited to take part in the screening program. Participants with a positive faecal occult blood test receive further diagnostic workup including colonoscopy. Results of colonoscopies are collected in a central database and standardized documentation guidelines are followed. Demographic and anthropometric factors, dietary and smoking habits are assessed by questionnaire. All subjects gave written informed consent and the study was approved by the institutional review board "Ethikkommission Burgenland (EK33/2010)". Ascertainment of the diverticulosis / diverticulitis phenotype in the CORSA study population was based on database review of colonoscopy and clinical data for the years 2003 to 2009. Individuals with colon cancer or IBD were excluded. Genotypes for replication analysis were extracted from available QC'ed Axiom Genome-Wide CEU1 array (Affymetrix, Santa Clara, CA) data which was imputed to 1000 Genomes Project Phase 3 reference (using IMPUTE2). For the Austria / Oberndorf samples, a systematic, prospective recruitment of all patients undergoing screening colonoscopy at the Krankenhaus Oberndorf is being performed since 2007. Patients were recruited on site according to the same phenotypic criteria as for the German

cohorts. The study was approved by the local ethics committee (Ethikkommission des Landes Salzburg, approval no. 415-E/ 1262/2-2010) and informed consent was obtained from all participants.

Lithuanian replication cohort

Patients were recruited at the Department of Gastroenterology at the Lithuanian University of Health Sciences, Kaunas in Lithuania between 2012 and 2017 from patients referred for colonoscopy according to the same criteria as the Germany / West samples as described previously [3]. All patients have signed an informed consent form to participate in the study. The study protocol was approved by the Regional Kaunas Ethics Committee (BE-10-2). The study was performed according to the Declaration of Helsinki.

Swedish replication samples

The population-based colonoscopy study (PopCol) was performed at Ersta hospital in Stockholm, Sweden from 2002 to 2006, where 3356 randomly selected adults from the general population were sent an Abdominal Symptom Questionnaire and 2293 responders were contacted for further investigations. Of the 745 individuals (426 women and 319 men) who underwent an ileo-colonoscopy as part of the study, 130 individuals (17.4 %) had diverticulosis. No individual presented with diverticulitis. Illumina OmniExpressExome-8 v1 genotypes were extracted from available QC'ed and imputed data, which have already been used and described in previous publications [4,5]. Study approval was obtained from the local Ethics committee (No 394/01, Karolinska Institutet Huddinge Regional Ethics Board, Sweden) and written informed consent was obtained from all participants. The PopCol study is described in detail in Kjellström *et al.* Eur J Gastroenterol Hepatol. 2014 Mar; 26(3):268-75.

GWAS analysis

Discovery GWAS analysis was performed on UK Biobank on Version 3 imputed genotypes. Genome-wide association tests were performed using BOLT-LMM v2.34, which applies a linear mixed model to adjust for the effects of population structure and individual relatedness[6]. This enabled the inclusion of all related individuals in our white European subset allowing a sample size of up 451,099 individuals, as opposed to a maximal set of 379,767 unrelated individuals. We limited our analysis to 11,977,111 genetic variants centrally imputed using both the Haplotype Reference Consortium imputation reference panel and a combined UK10K and 1000 Genomes reference panel with a minimum minor allele frequency (MAF) >0.1% and imputation quality score (INFO) >0.4.

Data availability

The GWAS summary statistics is publicly available on our group websites:

http://gengastro.med.tu-dresden.de/suppl/diverticular_disease/

http://www.t2diabetesgenes.org/data/.

The data from UK Biobank reported in this paper are available via application directly to the UK Biobank (http://www.ukbiobank.ac.uk/).

Validation genotyping

The 51 loci were validated in a combined European sample of 3,893 cases and 2,829 diverticula-free controls based on colonoscopy (Table 1) using the most significant discovery variant. When direct genotyping of a lead variant was not technical feasible, appropriate proxies were selected instead, defined as the variant with the next-lowest P-value within 250 kb of the index SNP (Table 2,3 Supplementary Table 5). Genotyping of SNPs was performed using the Agena® iPLEX Gold chemistry MassARRAY platform and TaqMan® technology from Life Technologies on an automated platform as described previously [2]. The choice of genotyping technology per variant was based on technical considerations of assay design feasibility and is indicated in Table 2 and 3. Genomic DNA was amplified with the GenomiPhi (Amersham) whole-genome amplification kit and fragmented at 99 °C for 3 min. All data were logged and managed with a database-driven laboratory information

management system (LIMS) [7]. Individual samples with >5% missing data were excluded from further analyses. SNPs that had >5% missing data or deviated from Hardy-Weinberg equilibrium (exact $P < 10^{-6}$ in controls) were excluded and replaced. Logistic regression analyses under additive model of inheritance were performed with PLINK [8] adjusting for gender and age for Taqman and Plex genotyped cohorts (i.e. German data sets (Germany North and West), the Austrian data set (Austria Oberndorf) and the Lithuanian data set). For cohorts Austria (Vienna) and Sweden, 1000 genomes imputed GWAS data was available, here association tests were performed using SNPTEST (v2.5) [9] adjusting for gender and age. Study-specific β effect estimates from all European replication/differentiation cohorts were then combined by fixed-effect meta-analysis using an inverse variance–weighted method implemented in META (1.6.0) [10]. For replication a nominal significance level of $P<0.05$ and consistency in odds ratio direction between the discovery and replication stage was applied.

To increase statistical power to replicate risk loci with lower observed odds ratios the sample size of the replication cohort was increased by including replication data for diverticulitis presented by Maguire et al. from European samples (N=29,367) from the Michigan genome initiative (MGI) into a meta-analysis of all European replication cohorts (N=36,089 samples). Study-specific z-scores for each allele were combined across samples in a weighted sum, with weights proportional to the square-root of the sample size for each study implemented in METAL (http://csg.sph.umich.edu/abecasis/Metal/).

Sudy-specific effective sample sizes were calculated as 4 / (1/[# of cases] + 1/[# of controls]).

mRNA expression analysis

Colonic tissue samples were obtained during surgical resection for the controls and non-inflamed diverticulosis during partial colectomy for nonobstructive colorectal carcinoma with tissue obtained distal from the tumor at a distance >5cm from any additional pathology or the tumour. Anorectal evacuation and colonic motility disorders were excluded. Diverticulitis samples were obtained from patients operated after two or more attacks of diverticulitis during elective surgery. Full-thickness

specimens were harvested from sites adjacent to colonic diverticula. All specimens were immediately transferred from the operating room to the laboratory for tissue processing in PBS (phosphate-buffered saline, pH 7.2). The study of human tissue received approval from the Local Ethics Committee of the Faculty of Medicine, Kiel University, Germany (B299/07). Patient characteristics are provided in Supplementary Table 7.

For dissection of colonic layers, full-thickness rectangular tissue blocks (30 mm x 10 mm) were pinned out flat on a cork plate by fine needles as described previously [11]. Mucosa, submucosa, and muscularis propria were separated using microsurgical scissors and immediately frozen in isopentane after resection and stored at −70°C until further processing. RNA was extracted using the NucleoSpin® RNA Kit (Macherey-Nagel, Düren, Germany). Reverse transcription was carried out using the M-MLV Reverse Transcriptase RNase (H-) (Promega, Mannheim, Germany) according to the manufacturer's protocol. Duplicate real-time quantitative PCR reactions were performed with qPCR Master Mix Plus (Eurogentec, Seraing, Belgium) using an ABI Prism 7500 fast Real-Time PCR cycler (Life Technologies). The housekeeping gene *HPRT* was used for normalization. The primer/probe sets (Eurogentec) or TaqMan assays (Thermo Fisher Scientific, Waltham, MA, USA) used are listed in Supplementary Table 8. Quality filtering: samples showing deviations at duplicate Ct values of >1.5 units were excluded from analysis. Potential outliers were not excluded from non-parametric group comparison analysis.

Statistical analysis of mRNA expression data and creation of box plots with jittered points were performed using ggpubr: 'ggplot2' package in r (http://www.sthda.com/english/rpkgs/ggpubr/). For the comparison of two groups a non-parametric Mann-Whitney U test (i.e. unpaired two-samples Wilcoxon test) was used. For the comparison of three groups a non-parametric Kruskal-Wallis test was used. Differences were considered significant after correction for multiple testing at $P<0.0025$ (0.05/20 tests). All results are expressed as medians with interquartile ranges.

Immunohistochemistry

Fluorescence immunohistochemistry was performed as previously described [12]. Briefly, colonic specimens were fixed in 4% paraformaldehyde for 24 h. Paraffin-embedded tissue sections were pre-treated with citrate buffer and incubated overnight with following primary antibodies: mouse anti-Col6a1 (B-4, Santa Cruz Biotechnology, Santa Cruz, USA), rabbit anti-PHGR1 (gift of Oddmund Nordgård, Stavanger University Hospital, Norway), rabbit anti-GPR158 (ABIN 2890856, Antibodies-online.com, Aachen, Germany), mouse anti-EFEMP1 (mAB3-5, Santa Cruz Biotechnology), rabbit anti-CRISPLD2 (NBP1-85143, Novusbio, Littleton, USA), rabbit anti-Elastin (Ab21610, Abcam, Cambridge, U.K.), rabbit anti-a-SMA (Ab5694, Abcam), mouse anti-a-SMA (M0851, Dakocytomation, Glostrup, Denmark), rabbit anti-PGP9.5 (RA95101, UltraClone, Isle of Wight, U.K.) and mouse anti-PGP9.5 (BM699, Acris, Herford, Germany). Anti-rabbit AlexaFluor488, anti-rabbit Alexafluor555, anti-mouse AlexaFluor488 and anti-mouse Alexafluor555 (Life Technologies, Karlsruhe, Germany) were used as secondary antibodies. All antibodies were diluted in antibody diluent (Life Technologies). Nuclei were counterstained with DAPI (Roche, Mannheim, Germany). Image acquisition was performed on a fluorescence inverted microscope (Axiovert 200 M, Zeiss, Gottingen, Germany) coupled to an AxioCam MR3 camera (Zeiss) using Axiovision software (version 4.7, Zeiss).

Loci Discovery and Functional Annotation (FUMA)

Genomic risk loci and lead variants were derived from FUnctional Mapping and Annotation of genetic associations (FUMAv1.3.1, http://fuma.ctglab.nl) [13] based on GWAS summary statistics obtained from BOLT-LMM. Independent significant SNPs were identified using the SNP2GENE function and were defined as SNPs with a P-value of $<5\times10^{-8}$ and independence to other genome wide significant SNPs at $r^2 < 0.6$ based on reference panel 1000 Genomes phase 3. Unique genomic risk loci were identified as LD blocks of independent significant SNPs that are >250kb apart, closer blocks were merged into a single locus. For each genomic risk locus one or more lead SNPs were identified among the independent significant SNPs and were defined as those that were independent from each other at

r^2<0.1. The independent significant variant with the lowest p-value at each risk locus was classified as the top lead variant for that respective locus and was followed up in replication genotyping (N=51 top lead variants) (Supplementary Table 5). Based on these independent significant SNPs, candidate SNPs used in subsequent functional annotations were identified as all SNPs that had a P-value of $<1\times10^{-5}$, MAF > 0.01 and were in LD of $r^2\geq0.6$ with at least one of the independent significant SNPs (Supplementary Table 1).

In order to identify candidate gene(s) at the respective genomic risk locus we followed i) **a manually curated selection process** based on local LD structure and supporting evidence from regulatory elements (eQTL and chromatin interaction) as detailed below and in Supplementary Table 3 and ii) we performed hypothesis-free functional and gene annotations based on the genomic positions of risk loci using **FUMA** [21] as manually curated selection process of candidate genes might not capture the full biology of the risk architecture. In the manually curated selection process, except for locus #25 (*PHGR1* and *DISP2*), a single curated candidate gene was assigned to each locus. The FUMA based approach resulted in 1080 unique mapped candidate genes (Supplementary Figure 6 & 7 and Supplementary Table 2).

Manually curated selection process (Supplementary Table 3): For lead variants located intronic, exonic, in the 3´or 5´UTR to a single annotated gene, the respective gene was identified as the candidate gene (30 out of 51 loci). For lead variants located upstream or downstream to a single annotated gene, this gene was assigned as the candidate gene, if the respective gene contains a variant with at least an r^2>0.5 to the lead variant at the locus (12 out of 51 loci). For loci, where variants in more than one neighbouring or overlapping transcript showed significant LD (r^2>0.5) to the lead variant, the transcript with higher expression in the tissue of interest was selected (3 out of 51 loci). For loci, where variants in more than one neighbouring or overlapping transcript showed significant LD (r^2>0.5) to the lead variant and no clear differences in expression were evident, the curated candidate gene was selected if additional regulatory evidence was present, i.e. an eQTL in a tissue of interest according to GTEx_v7 or a chromatin interaction pointed to a particular gene (1 out of 51 loci). If the lead variant was not located in a gene region and did not show significant LD to a variant in a neighbouring transcript, the impact of the lead variant and the variants in LD to the lead variant on

regulatory elements (eQTLs or chromatin interaction) was evaluated. If such elements were identified, the closest respective gene was annotated as the curated candidate gene (Criterion 5: 4 out of 51 loci). For the remaining loci, the closest transcript to the lead variant was annotated as the curated candidate gene if the distance to the variant was less than 1MB (Criterion 6: 1 out of 51 loci).

Using **FUMA**, all candidate SNPs were by default mapped to Ensembl genes (build 85) using ANNOVAR. The maximum physical distance to map SNPs to genes was 10kb. Intergenic SNPs were mapped to the two closest up and down stream genes thus with possible assignment to multiple genes. Candidate SNP and gene positions are referring to the human reference assembly (GRCh37/hg19) and are provided in Supplementary Table 1. Functional consequences of candidate SNPs were assessed using ANNOVAR, a tissue-specific *cis*-eQTL dataset (GTExV7, https://gtexportal.org) and 15-core chromatin states (ENCODE Project Consortium, 2012) [16]. Candidate genes with eQTL variants affecting gene expression in sigmoid colon at FDR<0.05 or at nominal p_eQTL <0.05 are shown in Table 2 and Table 3 and Supplementary Table 16. Enhancer and promoter regions were obtained from Roadmap Epigenomics Projects for 111 epigenomes [16]. Those regions were predicted using DNase peaks and core 15-state chromatin state model.

Gene set and pathway analysis

We used two gene set and pathway analysis approaches (MSigDB [14] and VEGAS2pathway [15]) to determine if the polygenic signal measured in the diverticular disease associated genes clustered in specific biological pathways. First, positional candidate genes from genomic risk loci showing consistency in effect direction between both discovery and replication stage, as outlined in Supplementary Table 5, were tested for overrepresentation with gene sets from the C5 collection: GO Biological Processes and gene sets from C2 sub-collection CP: Canonical pathways, curated in the Molecular Signatures Database (MSigDB 6.1; http://software.broadinstitute.org/gsea/msigdb/). A hypergeometric over-representation p-value: (k, K, n, N) was calculated for each gene set from K (the number of genes in the set), k (the number of genes in the intersection of the query set), n (the number of genes in comparison) and N (the number of all known human gene symbols). To control the false

positive error rates a FDR ($p < 0.05$) threshold was applied for significance. Results are provided in Supplementary Table 9 and 11. Secondly, we used VEGAS2pathway (https://vegas2.qimrberghofer.edu.au/) an extension of the VEGAS2 approach (VErsatile Gene-based Association Study, VEGAS2v02) to test hypothesis-free for pathway and gene set enrichment using the GWAS summary statistics obtained from BOLT-LMM v2.34. VEGAS2Pathway is a two-step pathway analysis strategy. Firstly, we calculated the gene-based test statistics for all genes using VEGAS2, which accounts for the LD between the SNPs within a gene through simulation. Variants lying within 50 kb on either side of a gene's transcription site (hg19 annotated RefSeq genes from UCSC table browser) were assigned to the respective gene to compute its association p value. This selection criterion was used to balance between inclusions of possible cis-regulatory variants and maintaining specificity of a gene. Secondly, for each of a set of pre-specified gene-sets, the relevant gene-based results were carried forward to compute a pathway-based test for gene-sets from the Gene Ontology, curated gene-sets from MSigDB; containing canonical pathways and gene-sets from BIOCARTA, REACTOME, KEGG databases, PANTHER, and pathway commons databases. Gene-sets were filtered to include only those with size between 10 and 1,000 genes. Overall, there were 6,212 gene-sets, including 18,399 genes with 511,336 annotations. Results are provided in Supplementary Table 11 and 12.

Enrichment analyses in cell lines and primary tissues.

GARFIELD (**G**WAS **A**nalysis of **R**egulatory or **F**unctional **I**nformation **E**nrichment with **LD** correction). The GARFIELD (http://europepmc.org/preprints/ppr7035) approach is independent from FUMA annotated genes and VEGAS2pathway results. GARFIELD used the whole number of GWAS SNPs as input then performs greedy pruning of GWAS SNPs (LD $r2 > 0.1$) and then annotates them based on functional information overlap. Functional enrichment analysis of diverticular disease variants in DNaseI Hypersensitive sites from ENCODE and Roadmap Epigenomics data are provided in Supplementary Table 13. GARFIELD allows for parallel enrichment analyses at multiple p-value sub-thresholds, which improves power to define statistically significant enrichment patterns by

increasing the number of variants tested. GARFIELD uses a nonparametric approach to weight GWAS findings with regulatory or functional annotations to find features relevant to a phenotype of interest. GARFIELD accounts for LD, minor allele frequency, matched genotyping variants and local gene density with the application of permutations to derive statistical significance. GARFIELD quantifies enrichment using odds ratios (OR) at various GWAS p-value cutoffs and assesses their significance by employing generalized linear model testing, while accounting for minor allele frequency, distance to nearest transcription start site and number of LD proxies (r2 > 0.8). The fold enrichment at various GWAS p-value cutoffs is indicated by color coding as described in the figure legend of Supplementary Figure 11. Fold enrichment values are shown in black and blue, for the GWAS P-value thresholds <1×10-8 and <1×10-5, respectively. The innermost and outermost dots along the inside edge denote significant enrichment for the cell type at <1×10-5 and < 1×10-8, respectively.

URLs:

ANNOVAR: http://annovar.openbioinformatics.org/en/latest/
BOLT-LMM: https://data.broadinstitute.org/alkesgroup/BOLT-LMM/
FUMA: http://fuma.ctglab.nl/
GARFIELD: http://europepmc.org/preprints/ppr7035
GSEA/MSigDB 6.1: http://software.broadinstitute.org/gsea/msigdb/
GWAS Catalog: https://www.ebi.ac.uk/gwas/
LOCUSZOOM: http://locuszoom.org/
META: https://mathgen.stats.ox.ac.uk/genetics_software/meta/meta.html
METAL: https://genome.sph.umich.edu/wiki/METAL
OMIM: https://omim.org/
PLINK: http://zzz.bwh.harvard.edu/plink/
R, ‘ggplot2’ package: http://www.sthda.com/english/rpkgs/ggpubr/
SNPTEST: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html
UCSC: https://genome.ucsc.edu/
VEGAS2: https://vegas2.qimrberghofer.edu.au/

ACKNOWLEDGEMENTS

The authors wish to thank all study participants, researchers, clinicians and administrative staff who contributed to this study. We thank Julia Wilking, Inka Geurink, Katrin Neblung-Masuhr, Sanaz Sedghpour Sabet, Tanja Wesse, Anja Tanck, Frank Lichte and Karin Stengel for expert technical assistance and Oddmund Nordgård (Stavanger University Hospital, Norway) for kindly providing the anti-PHGR1 antibody.

REFERENCES

1 Buch S, Schafmayer C, Völzke H, *et al.* A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat Genet* 2007;**39**:995–9. doi:10.1038/ng2101

2 Buch S, Stickel F, Trépo E, *et al.* A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat Genet* 2015;**47**:1443–8. doi:10.1038/ng.3417

3 Reichert MC, Kupcinskas J, Krawczyk M, *et al.* A Variant of COL3A1 (rs3134646) Is Associated With Risk of Developing Diverticulosis in White Men. *Dis Colon Rectum* 2018;**61**:604–11. doi:10.1097/DCR.0000000000001001

4 Bonfiglio F, Zheng T, Garcia-Etxebarria K, *et al.* Female-Specific Association Between Variants on Chromosome 9 and Self-Reported Diagnosis of Irritable Bowel Syndrome. *Gastroenterology* 2018;**155**:168–79. doi:10.1053/j.gastro.2018.03.064

5 Henström M, Diekmann L, Bonfiglio F, *et al.* Functional variants in the sucrase–isomaltase gene associate with increased risk of irritable bowel syndrome. *Gut* 2018;**67**:263–70. doi:10.1136/gutjnl-2016-312456

6 Loh P-R, Tucker G, Bulik-Sullivan BK, *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;**47**:284–90. doi:10.1038/ng.3190

7 Hampe J, Wollstein A, Lu T, *et al.* An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics* 2001;**17**:654–5.

8 Chang CC, Chow CC, Tellier LC, *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;**4**:7. doi:10.1186/s13742-015-0047-8

9 Johnson EO, Hancock DB, Levy JL, *et al.* Imputation across genotyping arrays for genome-wide association studies: assessment of bias and a correction strategy. *Hum Genet* 2013;**132**:509–22. doi:10.1007/s00439-013-1266-7

10 Liu JZ, Tozzi F, Waterworth DM, *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 2010;**42**:436–40. doi:10.1038/ng.572

11 Böttner M, Barrenschee M, Hellwig I, *et al.* The enteric serotonergic system is altered in patients with diverticular disease. *Gut* 2012.

12 Cossais F, Leuschner S, Barrenschee M, *et al.* Persistent Increased Enteric Glial Expression of S100β is Associated With Low-grade Inflammation in Patients With Diverticular Disease. *J Clin Gastroenterol* Published Online First: March 2018. doi:10.1097/MCG.0000000000001011

13 Watanabe K, Taskesen E, van Bochoven A, *et al.* Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017;**8**:1826. doi:10.1038/s41467-017-01261-5

14 Subramanian A, Tamayo P, Mootha VK, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545–50. doi:10.1073/pnas.0506580102

15 Mishra A, MacGregor S. A Novel Approach for Pathway Analysis of GWAS Data Highlights Role of BMP Signaling and Muscle Cell Differentiation in Colorectal Cancer Susceptibility. *Twin Res Hum Genet* 2017;**20**:1–9. doi:10.1017/thg.2016.100

16 Schmitt AD, Hu M, Jung I, Xu Z et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep* 2016 Nov 15;17(8)