

ARTICLE

Received 23 Dec 2016 | Accepted 24 Apr 2017 | Published 6 Jun 2017

DOI: 10.1038/ncomms15789

OPEN

Sequence variants in *ARHGAP15*, *COLQ* and *FAM155A* associate with diverticular disease and diverticulitis

Snaevar Sigurdsson¹, Kristjan F. Alexandersson¹, Patrick Sulem¹, Bjarke Feenstra², Steinunn Gudmundsdottir¹, Gisli H. Halldorsson¹, Sigurgeir Olafsson¹, Asgeir Sigurdsson¹, Thorunn Rafnar¹, Thorgeir Thorgeirsson¹, Erik Sørensen³, Andreas Nordholm-Carstensen⁴, Jakob Burcharth⁵, Jens Andersen⁶, Henrik Stig Jørgensen⁷, Emma Possfelt-Møller⁸, Henrik Ullum³, Gudmar Thorleifsson¹, Gisli Masson¹, Unnur Thorsteinsdottir^{1,9}, Mads Melbye^{2,10,11}, Daniel F. Gudbjartsson^{1,12}, Tryggvi Stefansson¹³, Ingileif Jonsdottir^{1,9,14} & Kari Stefansson^{1,9}

Diverticular disease is characterized by pouches (that is, diverticulae) due to weakness in the bowel wall, which can become infected and inflamed causing diverticulitis, with potentially severe complications. Here, we test 32.4 million sequence variants identified through whole-genome sequencing (WGS) of 15,220 Icelanders for association with diverticular disease (5,426 cases) and its more severe form diverticulitis (2,764 cases). Subsequently, 16 sequence variants are followed up in a diverticular disease sample from Denmark (5,970 cases, 3,020 controls). In the combined Icelandic and Danish data sets we observe significant association of intronic variants in *ARHGAP15* (Rho GTPase-activating protein 15; rs4662344-T: $P = 1.9 \times 10^{-18}$, odds ratio (OR) = 1.23) and *COLQ* (collagen-like tail subunit of asymmetric acetylcholinesterase; rs7609897-T: $P = 1.5 \times 10^{-10}$, OR = 0.87) with diverticular disease and in *FAM155A* (family with sequence similarity 155A; rs67153654-A: $P = 3.0 \times 10^{-11}$, OR = 0.82) with diverticulitis. These are the first loci shown to associate with diverticular disease in a genome-wide study.

¹deCODE Genetics/Amgen Inc., Sturlugata 8, 101 Reykjavik, Iceland. ²Department of Epidemiology Research, Statens Serum Institut, 2300 Copenhagen, Denmark. ³Department of Clinical Immunology, Copenhagen University Hospital/Rigshospitalet, 2100 Copenhagen, Denmark. ⁴Digestive Disease Center, Bispebjerg Hospital, University of Copenhagen, Copenhagen, 2400 Nevada, Denmark. ⁵Department of Surgery, Herlev Hospital, University of Copenhagen, 2730 Herlev, Denmark. ⁶Hvidovre university Hospital, Department of Surgery, Gastroenterology, 2650 Hvidovre, Denmark. ⁷Department of Surgery, Northern Sealand Hospital, 3400 Hillerød, Denmark. ⁸Department of Surgical Gastroenterology, Copenhagen University Hospital/Rigshospitalet, 2100 Copenhagen, Denmark. ⁹Faculty of Medicine, School of Health Sciences, University of Iceland, 101 Reykjavik, Iceland. ¹⁰Department of Clinical Medicine, University of Copenhagen, 2100 Copenhagen, Denmark. ¹¹Department of Medicine, Stanford University School of Medicine, Stanford, California 94305-5475, USA. ¹²School of Engineering and Natural Sciences, University of Iceland, 101 Reykjavik, Iceland. ¹³Department of Surgery, Landspítali, the National University Hospital of Iceland, 101 Reykjavik, Iceland. ¹⁴Department of Immunology, Landspítali, the National University Hospital of Iceland, 101 Reykjavik, Iceland. Correspondence and requests for materials should be addressed to I.J. (email: ingileif.jonsdottir@decode.is) or to K.S. (email: kstefans@decode.is).

Diverticular disease is thought to be due to complex interactions between diet, lifestyle, colonic motility, structural changes in the gut, enteric neuropathy and smoking^{1–3}. The intestinal microflora and low-grade inflammation may contribute to diverticular disease and acute diverticulitis³. Diverticulae are commonly found during routine colonoscopy with increased prevalence from age 50–59 years (32.6%) to age ≥ 80 years (71.4%)^{4,5}. Up to 20% will experience complications of the disease⁶ but only 1–4% of individuals with diverticula develop acute diverticulitis, with a recurrence risk of 20% within 5 years⁷.

Relative risk of siblings of diverticular disease cases is 2.9 (ref. 8) and the heritability in twin studies estimated to be 40–50% (refs 8,9). This indicates that there is a strong genetic component to the risk. No sequence variants associating with risk of diverticular disease have been found and no genome-wide association studies (GWAS) have been published.

We performed GWAS to search for sequence variants that affect the risk of diverticular disease and diverticulitis in Iceland with a follow up in Danish sample of diverticular disease. We find association of three intronic variants in the genes *ARHGAP15* (Rho GTPase-activating protein 15) and *COLQ* (collagen-like tail subunit of asymmetric acetylcholinesterase) with diverticular disease and in *FAM155A* (family with sequence similarity 155A) with diverticulitis. These are the first sequence variants found to show genome-wide significant association with diverticular disease.

Results

Association of three loci with diverticular disease. We imputed 32.4 million sequence variants identified through WGS of 15,220 Icelanders into 151,677 chip typed Icelanders and their first- and second-degree relatives^{10,11} and performed two GWAS to search for sequence variants that affect the risk of diverticular disease (5,426 cases) and diverticulitis (2,764 cases) using the same 245,951 controls (Supplementary Table 1). We applied weighted thresholds for genome-wide significance that depend on the functional class of each variant, based on its prior probability of affecting gene function¹² (Supplementary Table 2).

We chose 16 variants for follow-up in a Danish diverticular disease sample set that were within two orders of magnitude from genome-wide significance threshold in a variant class for either diverticular disease or diverticulitis in Iceland (Table 1 and Supplementary Table 3a–c). We do not have information on diverticulitis in the Danish cohort, but chose to follow-up the Icelandic diverticulitis findings based on the assumption that the Danish cohort includes diverticulitis, although the proportion is unknown. With these data sets we identified three loci that are of genome wide-significance in the combined analysis of the Icelandic and Danish samples; intronic variants at the *ARHGAP15* and *COLQ* loci associate significantly with diverticular disease and at *FAM155A* locus with diverticulitis ($P < 2.3 \times 10^{-9}$, the threshold for intronic variants within a DNase hypersensitivity site¹²) (Table 1 and Supplementary Table 3). No significant heterogeneity was observed between the study groups and the three single-nucleotide polymorphisms (SNPs) are nominally significant in the Danish follow-up.

Potential causal genes at the diverticular disease loci. The strongest diverticular disease association in Iceland was with 45 correlated ($r^2 > 0.97$) sequence variants (minor allele frequency (MAF) = 17.6–17.8%) in 88 kb region spanning introns 9 and 10 of *ARHGAP15* (Rho GTPase-activating protein 15) (Fig. 1a). The variants are represented by rs4662344-T that associates at genome-wide significance in Iceland (chr2:143,591,289,

odds ratio (OR) = 1.23, $P = 4.9 \times 10^{-13}$) (Table 1). In Iceland rs4662344-T confers similar risk of diverticulitis (OR = 1.26, $P = 4.5 \times 10^{-9}$) and uncomplicated diverticular disease (OR = 1.20, $P = 2.6 \times 10^{-6}$) ($P_{\text{het}} = 0.36$). No association signal remains at the locus after conditional analysis using rs4662344-T as a covariate (Fig. 2a), indicating that one of the 45 intronic variants is likely to mediate the signal at this locus. The association replicates in the Danish samples (OR = 1.22 and $P = 7.3 \times 10^{-7}$) for a combined P value of 1.9×10^{-18} and OR of 1.23 for the Icelandic and Danish samples. None of the three missense variants in *ARHGAP15* (15 exons 475 amino acids) associate with diverticular disease ($P > 0.44$). No other gene is within 100 kb of any of the SNPs in linkage disequilibrium (LD) ($R^2 > 0.8$) with rs4662344-T.

rs4662344-T did not associate with expression of *ARHGAP15* or any other gene in the region (± 500 kb) in any of the tissues in GTExV6 database (including whole blood and small intestine, the most relevant tissues) nor in RNA sequencing data at deCODE from whole blood ($n = 2,246$) and adipose tissue ($n = 708$) (Supplementary Fig. 1, shown for each transcript and each exon).

ARHGAP15 encodes Rho GTPase-activating protein 15, a member of the Rac-specific GTPase-activating protein (GAP). Rac is a small GTPase, important for cell proliferation, apoptosis, attachment and motility¹³. *ARHGAP15*s activation of Rac affects the actin cytoskeleton and cell morphogenesis¹⁴ and overexpression of *ARHGAP15* causes increase in actin stress fibres and cell contraction¹⁵. Neutrophils of mice that are knockout for *ArhGAP15* show increased migration, phagocytosis, reactive oxygen species (ROS) production and bacterial killing, and reduced inflammation¹³. We therefore tested the effect of rs4662344-T on ROS production by neutrophils stimulated with *E. coli*, phorbol 12-myristate 13-acetate or *N*-formyl-MetLeuPhe, but found no effect of rs4662344-T carrier status on neutrophil ROS production for any of the stimulants (Supplementary Fig. 2).

The second strongest signal that associates with diverticular disease in the Icelandic samples is captured by a single intronic SNP in *COLQ*, rs7609897-T (chr3:15,461,174, MAF = 24.7%) (Fig. 2b), that associates with diverticular disease (OR = 0.85, $P = 1.6 \times 10^{-9}$) in Iceland; this association replicates nominally in the Danish samples with a consistent direction of the effect (OR = 0.91, $P = 0.010$) for a combined OR of 0.87 and P value of 1.5×10^{-10} for the Icelandic and Danish samples. In the Icelandic and in 1,000G European data sets, rs7609897-T is weakly correlated with other markers ($r^2 < 0.26$ and $r^2 < 0.48$, respectively). We performed conditional analysis to look for additional signals at the locus (Supplementary Table 5). We found one rare missense variant rs146687198-G (p.Gly246Ala, MAF = 0.22%) in *COLQ* with large effect on diverticular disease in Iceland (OR = 2.06, 95% confidence interval (CI): 1.4, 3.0, $P = 3.5 \times 10^{-4}$). This rare missense variant is not correlated with the intronic rs7609897-T ($r^2 < 0.001$). Follow-up genotyping in the Danish samples showed a weaker and not significant effect (OR = 1.15, 95% CI: 0.63, 2.10, $P = 0.65$, MAF = 0.26%). However, the effect is consistent in the two populations ($P_{\text{het}} = 0.11$, for Iceland and Danish samples). Although the association of this rare missense variant with diverticular disease is not of genome-wide significance, the prior probability established by the association of rs7609897-T suggests that the association of rs146687198-G may be real and points to *COLQ* as the causative gene at this locus.

COLQ has 18 exons that span a 71 kb region (Supplementary Table 5). *COLQ* is expressed in most tissues (GTEx V6)¹⁶. rs7609897-T did not associate with RNA expression of *COLQ* or any of the 11 genes within 500 kb of rs7609897-T in any of the tissues in the GTExV6 database or in the deCODE in blood or adipocyte RNA sequencing data (Supplementary Fig. 3).

Table 1 | Icelandic GWAS results, follow-up in a Danish diverticular disease sample set and association in the Icelandic and Danish sample sets combined.

Nearest gene	SNP	Amin/ Amaj	Icelandic diverticular disease			Danish diverticular disease*			Combined Icelandic and Danish diverticular disease sample sets		
			5,426 cases; 245,951 controls			5,970 cases; 3,020 controls			11,396 cases; 248,971 controls		
			MAF % cases- controls	OR (95% CI)	P value	MAF % cases- controls	OR (95% CI)	P value	OR (95% CI)	P value	P _{het}
ARHGAP15	rs4662344	T/C	20.9/17.7	1.23 (1.16, 1.31)	4.9 × 10 ⁻¹³	21.7/18.5	1.22 (1.13, 1.32)	7.0 × 10 ⁻⁷	1.23 (1.17, 1.29)	1.9 × 10 ⁻¹⁸	0.86
COLQ	rs7609897	T/G	22.1/24.7	0.85 (0.80, 0.89)	1.6 × 10 ⁻⁹	21.6/23.3	0.91 (0.84, 0.98)	1.0 × 10 ⁻²	0.87 (0.83, 0.91)	1.5 × 10 ⁻¹⁰	0.17
FAM155A	rs67153654	A/T	17.0/18.6	0.89 (0.84, 0.94)	8.7 × 10 ⁻⁵	17.4/20.1	0.84 (0.78, 0.91)	2.2 × 10 ⁻⁵	0.87 (0.83, 0.91)	1.3 × 10 ⁻⁸	0.30

Nearest gene	SNP	Amin/ Amaj	Icelandic diverticulitis			Danish diverticular disease*			Combined Icelandic diverticulitis and Danish diverticular disease sample sets		
			2,764 cases; 245,951 controls			5,970 cases; 3,020 controls			8,734 cases; 248,971 controls		
			MAF % cases- controls	OR (95% CI)	P value	MAF % cases- controls	OR (95% CI)	P value	OR (95% CI)	P value	P _{het}
ARHGAP15	rs4662344	T/C	21.3/17.7	1.26 (1.16, 1.36)	4.5 × 10 ⁻⁹	21.7/18.5	1.22 (1.13, 1.32)	7.0 × 10 ⁻⁷	1.24 (1.17, 1.31)	1.8 × 10 ⁻¹⁴	0.62
COLQ	rs7609897	T/G	21.0/24.7	0.8 (0.74, 0.86)	1.9 × 10 ⁻⁹	21.6/23.3	0.91 (0.84, 0.98)	1.0 × 10 ⁻²	0.85 (0.80, 0.89)	1.0 × 10 ⁻⁹	0.02
FAM155A	rs67153654	A/T	15.6/18.6	0.8 (0.74, 0.87)	2.3 × 10 ⁻⁷	17.4/20.1	0.84 (0.78, 0.91)	2.2 × 10 ⁻⁵	0.82 (0.78, 0.87)	3.0 × 10 ⁻¹¹	0.43

Nearest gene	SNP	Amin/ Amaj	Icelandic uncomplicated diverticular disease			Danish diverticular disease*			Combined uncomplicated diverticular disease and Danish diverticular disease sample sets		
			2,662 cases; 245,951 controls			5,970 cases; 3,020 controls			8,632 cases; 248,971 controls		
			MAF % cases/ controls	OR (95% CI)	P value	MAF % cases/ controls	OR (95% CI)	P value	OR (95% CI)	P value	P _{het}
ARHGAP15	rs4662344	T/C	20.4/17.7	1.2 (1.11, 1.30)	2.6 × 10 ⁻⁶	21.7/18.5	1.22 (1.13, 1.32)	7.0 × 10 ⁻⁷	1.21 (1.15, 1.28)	8.6 × 10 ⁻¹²	0.77
COLQ	rs7609897	T/G	23.7/24.7	0.91 (0.84, 0.97)	6.1 × 10 ⁻³	21.6/23.3	0.91 (0.84, 0.98)	0.01	0.91 (0.86, 0.95)	1.7 × 10 ⁻⁴	1.0
FAM155A	rs67153654	A/T	18.87/18.6	0.99 (0.91, 1.06)	0.72	17.4/20.1	0.84 (0.78, 0.91)	2.2 × 10 ⁻⁵	0.91 (0.87, 0.97)	1.4 × 10 ⁻³	4.9 × 10 ⁻³

Amaj, major allele; Amin, minor allele; 95% CI, 95% confidence interval; MAF, minor allele frequency; OR, odds ratio of the minor allele; P_{het}, P value for the heterogeneity between cohorts; SNP, single-nucleotide variant polymorphism.
MAF is calculated on the chip-typed samples, excluding family imputed genotypes. The three variants rs4662344, rs7609897 and rs67153654 are annotated as intronic variant within a DNase hypersensitivity site, giving the class-specific Bonferroni threshold for genome-wide significance as $P < 2.3 \times 10^{-9}$ (ref.12).
*Note that the Danish diverticular disease results are the same in all three parts of the table, since diverticulitis diagnosis was not available.

COLQ encodes a subunit of a collagen-like molecule (ColQ) associated with acetylcholinesterase (AChE), whose catalytic subunits are anchored in the basal lamina of neuromuscular junction through ColQ. Homozygote mutations (or compound heterozygotes) in *COLQ* can reduce AChE availability resulting in prolonged nerve to muscle signalling, that can cause muscle weakness and congenital myasthenic syndromes¹⁷.

The third locus harbours sequence variants in the first intron of *FAM155A*, marked by rs67153654-A, showing a suggestive association with diverticulitis in Iceland (Fig. 1b and Fig. 2c) (chr13:107,572,636, MAF = 18.6%, OR = 0.80, 95% CI: 0.74, 0.87, $P = 2.3 \times 10^{-7}$). Although diagnosis of diverticulitis is not available for the Danish samples, we genotyped the *FAM155A* variant in the Danish diverticular disease samples. We replicated the association with rs67153654-A (Danish diverticular disease: OR = 0.84, $P = 2.2 \times 10^{-5}$) despite the lack of information on the proportion of diverticulitis among the Danish diverticular disease samples. In the combined analysis of the Icelandic diverticulitis and Danish total diverticular disease sample set, rs67153654-A reached genome-wide significance ($P = 3.0 \times 10^{-11}$, OR = 0.82) (Table 1). The association of rs67153654-A is driven by those in the diverticular disease sample who have developed diverticulitis, with no association with the subset of uncomplicated diverticular disease in Iceland (OR = 0.99, 95% CI: 0.92, 1.07, $P = 0.78$), suggesting that this variant is not likely to influence the integrity of the wall of the colon, but rather protection from infection or inflammation.

FAM155A spans 703 kb with only three exons and no other protein coding genes lie within 500 kb of rs67153654-A. None of the 16 missense variants found in *FAM155A* associates with diverticular disease or diverticulitis (Supplementary Table 5). *FAM155A* is mainly expressed in the hypothalamus and pituitary gland^{18,19} with low expression in the colon and blood

(GTEx V6)¹⁶. We found no effect of rs67153654-A on the expression of *FAM155A* or nearby genes in GTExV6 or in blood or adipocytes using RNA sequencing (Supplementary Fig. 4). Little is known about the function of *FAM155A* but close SNPs ($r^2 < 0.01$ with rs67153654-A) have been associated with increased fat mass in children²⁰ and anorexia nervosa²¹. We tested the association of diverticulitis versus uncomplicated diverticular disease for variants at the *ARHGAP15*, *COLQ* and *FAM155A* loci. (Supplementary Table 4). Only the *FAM155A* variant is significantly less frequent in diverticulitis (OR = 0.84, 95% CI: 0.74, 0.94, $P = 3.8 \times 10^{-3}$).

The 13 variants at the other loci selected for validation in the Danish samples lack evidence for association with either diverticular disease or diverticulitis (Supplementary Table 3a-c and Supplementary Note).

Inflammation contributes to the development and recurrence of diverticulitis²². Therefore, we tested the effects of the diverticular disease variants at the *ARHGAP15*, *COLQ* and *FAM155A* loci on other inflammatory diseases of intestine and colon, namely ulcerative colitis (UC) and Crohn's disease (CD) (inflammatory bowel disease (IBD)) and found no association with a $P < 1 \times 10^{-3}$. Furthermore, none of the 184 IBD/UC/CD variants (from the GWAS catalogue)²³⁻²⁹ associate with diverticulitis or diverticular disease ($0.05/184 = P < 2.7 \times 10^{-4}$) in Iceland. Neither did polygenic risk scores (PRS) for IBD, UC and CD capture risk of diverticular disease or diverticulitis (Supplementary Table 6).

Few sequence variants have previously been reported to associate with diverticular disease, diverticulitis or diverticular disease-related diseases in small candidate gene studies³⁰⁻³³. We show no evidence for association of these variants with the disease in the Icelandic data (Supplementary Table 7).

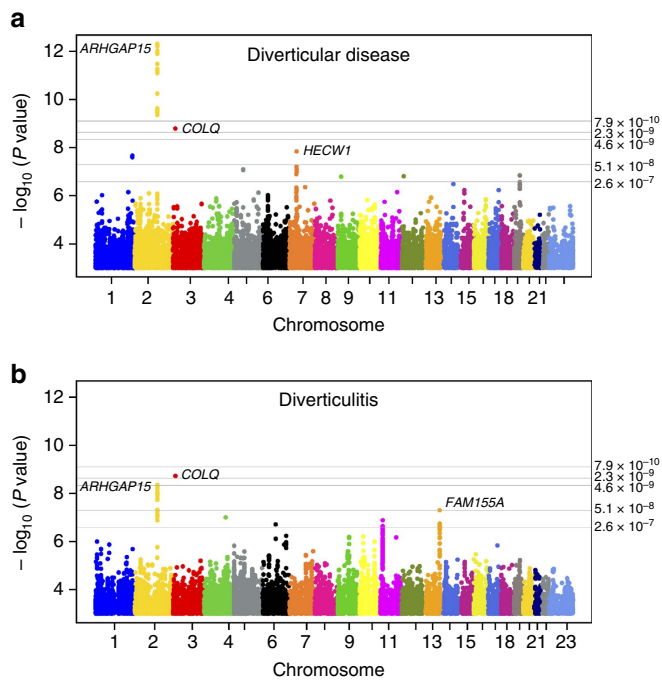


Figure 1 | Manhattan plot for genome-wide association results. The P values ($-\log_{10}$) are plotted against their respective positions on each chromosome. P value thresholds for the different annotation classes are indicated with gray lines. For intronic/intergenic variants outside DNase hypersensitivity site: $P = 1.1 \times 10^{-9}$, intronic/intergenic variants within DNase hypersensitivity site $P = 2.3 \times 10^{-9}$, low-impact variants: $P = 5.3 \times 10^{-9}$, medium-impact variants: $P = 7.4 \times 10^{-8}$ and high-impact variants: $P = 3.7 \times 10^{-7}$. The plots were created using qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots⁵⁴. **(a)** GWAS results for diverticular disease (cases $n = 5,292$; controls $n = 245,951$). *ARHGAP15* $P = 5.1 \times 10^{-12}$, *COLQ* $P = 2.8 \times 10^{-9}$ and *FAM155A* $P = 8.0 \times 10^{-5}$. **(b)** GWAS results for diverticulitis (Cases $n = 2,764$ cases; controls $n = 245,951$) (excluding diverticular disease). *ARHGAP15* $P = 6.0 \times 10^{-9}$, *COLQ* $P = 3.0 \times 10^{-9}$ and *FAM155A* $P = 1.7 \times 10^{-7}$. The three variants rs4662344, rs7609897 and rs67153654 are annotated as intronic variant within a DNase hypersensitivity site, giving the class-specific Bonferroni threshold for genome-wide significance as $P < 2.3 \times 10^{-9}$ (ref. 12).

Discussion

We have found common sequence variants in introns of the *ARHGAP15*, *COLQ* and *FAM155A* that associate with risk of diverticular disease or diverticulitis. These sequence variants do not overlap with known GWAS signals in other diseases or traits, including established risk loci for immune-mediated and inflammatory diseases³⁴. Diverticulitis occurs when the mucosa of diverticula becomes inflamed. Often the flat colon mucosa between the orifices of the diverticula is inflamed, with changes indistinguishable from those of UC or CD³⁵. We found no genetic overlap between diverticular disease and UC and CD. We found no association of the diverticular disease variants reported here with these diseases and well-established risk variants, and PRS for IBD, UC and CD do not associate with diverticular disease or diverticulitis. This indicates that the pathogenic mechanisms differ from those of autoimmune diseases of the colon and intestine. This is further supported by the complete lack of association with the HLA region that associates strongly with IBD and UC²³.

The stronger association of the *FAM155A* variants with diverticulitis than diverticular disease in general may reflect

effects on disease progression, such as inflammation or infection. Various inflammatory components have been suggested as biomarkers of diverticular disease and diverticulitis, including C-reactive protein, white blood cell count, erythrocyte sedimentation rate and faecal calprotectin³⁶. Still we found that the *FAM155A* variants have no effect on C-reactive protein levels, white blood cell count or neutrophil count, erythrocyte sedimentation rate ($P < 10^{-3}$) (Supplementary Methods), neither do the *ARHGAP15* and *COLQ* variants. Despite the role of Rho GTPase-activating protein 15, encoded by *ARHGAP15*, on phagocyte function and inflammation the *ARHGAP15* variant associating with diverticular disease does not affect ROS production by neutrophils. Whether the diverticular disease variants mediate their effects by modulating inflammation is thus unclear. None of the diverticular disease variants affect the expression of *ARGHAP15*, *COLQ* or *FAM155A* or of nearby genes, neither in deCODE's RNAseq data on blood and adipocytes nor in data from the various tissues of the GTEx database. Thus, the mechanism by which they affect the risk of diverticular disease remains to be elucidated.

Smoking is a risk factor for symptomatic diverticular disease in both men and women increasing risk of developing complicated diverticular disease¹ and hospital admission for acute colonic diverticulitis³⁷. We found that heavy smokers ($N = 26,113$, > 10 pack-years)^{38,39} have higher risk of developing diverticular disease than never smokers ($N = 22,815$) (Supplementary Methods), with relative risk = 1.35; 95% CI: 1.21–1.51, $P = 1.11 \times 10^{-7}$ (adjusted for sex and age). However, smoking showed no interaction with the effect of any of the three diverticular disease variants.

This first genome-wide association scan may pave the way for studies on the mechanism underlying the development of diverticular disease and diverticulitis.

Methods

Discovery cohort. We have collected phenotype data on the Icelandic population from everyone diagnosed with diverticular disease (ICD 9: 562.1–2 and ICD 10 K57.2–9) at the National University Hospital and from Akureyri Hospital in northern Iceland during the years 1985–2014. Phenotype data was available for 5,777 individuals with diverticular disease, including 2,923 with the primary diagnosis of diverticulitis and 2,854 with uncomplicated diverticular disease. Patients who came to the hospital primary for diverticulitis complications or if the diagnosis was coupled to a resection of the left colon or sigmoid colon were classified as diverticulitis. Genotype information was available for 94% of the diverticular disease cases or 2,764 of the 2,923 with diverticulitis and 2,662 of 2,854 uncomplicated diverticular disease. The diverticular disease patients were 60% female, with the mean age of 67.7 years (s.d. = 4.7) and mean body mass index of 27.8 (s.d. = 5.2). Information on the patients is summarized in Supplementary Table 1.

The study was approved by the National Bioethics Committee (ref. VSN 12-121) and the Data Protection Authority (2013030423PS/-) in Iceland. All participating subjects who donated blood signed informed consent. Personal identities of the participant's data and biological samples were encrypted by a third-party system (Identity Protection System), approved and monitored by the Data Protection Authority.

Follow-up cohort. Statens Serum Institut (SSI) hosts the Danish National Biobank and one of the associated biobanks under the Danish National Biobank umbrella is the Copenhagen Hospital Biobank, which stores EDTA whole blood from patient samples submitted for blood typing at hospitals in the Capital Region of Denmark (the Greater Copenhagen Area). SSI identified 6,500 individuals with diverticular disease diagnosis (ICD 9: 562.1–2 and ICD 10 K57.2–9), with EDTA whole-blood samples in the Copenhagen Hospital Biobank and further 3,000 control individuals who have no records of DD. Information on the proportion of diverticulitis was not available. The study was approved by the Scientific Ethics Committee of the Capital Region of Denmark (H-15000405) and the Danish Data Protection Agency. The Scientific Ethics Committee granted exemption from obtaining informed consent from participants as the study was based on the biobank material. The DNA extraction from whole blood and marker genotyping was performed at deCODE genetics.

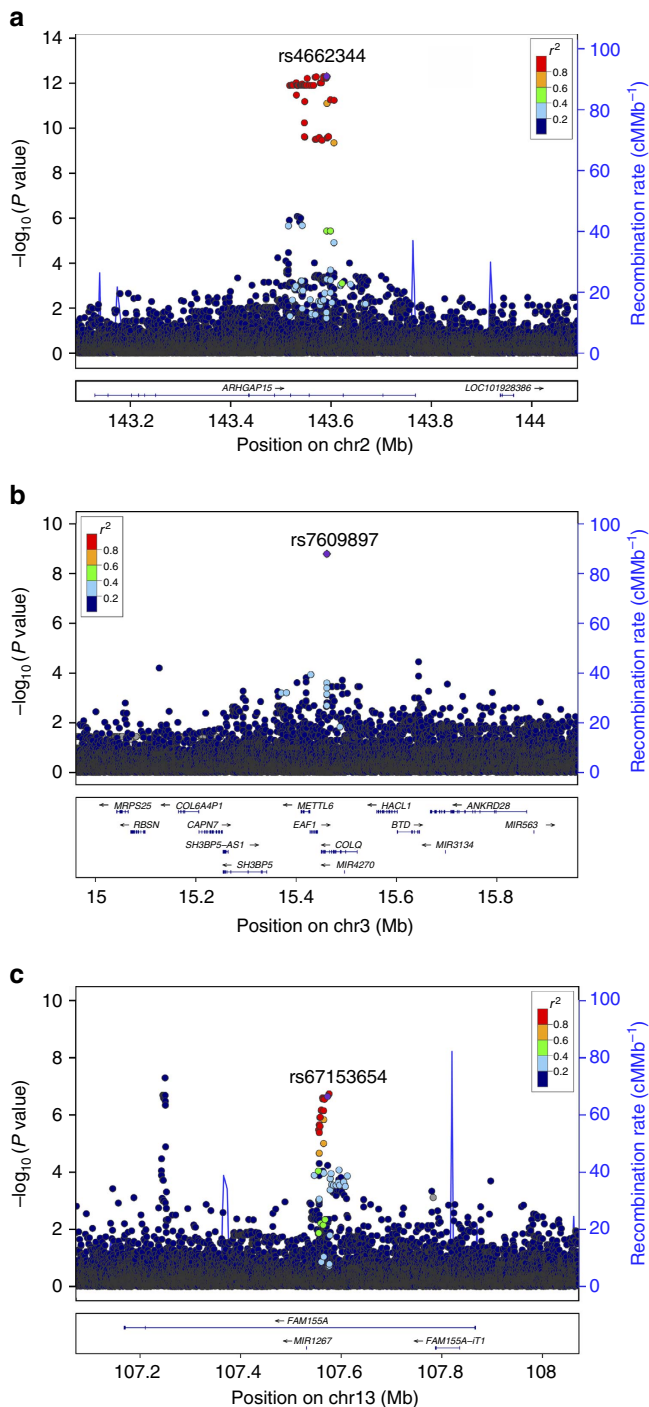


Figure 2 | Regional association plot for the three associated loci. P values ($-\log_{10}$) for the marker associations are plotted against the chromosomal location (human genome build 38) at each locus. The colour of the genomic variants reflects the linkage disequilibrium (r^2 LD) with the lead SNP in the Icelandic dataset. The blue line indicates recombination rates from the Icelandic recombination map for males and females⁵⁵. Known genes and exons are shown below using data from the UCSC genes track. The plot was created with a stand-alone version of the LocusZoom Software⁵⁶. **(a)** Locus plot for the marker rs4662344-T (chr2:143,591,289) at the *ARHGAP15* locus. P values plotted are for association with diverticular disease in Iceland. **(b)** Locus plot for the marker rs7609897-T (chr3:15,461,174) in intron of the *COLQ* gene. P values plotted are for the association with diverticular disease in Iceland. **(c)** Locus plot for the marker rs67153654-A (chr13:107,572,636) at the *FAM155A* locus. P values plotted are for the association with diverticulitis in Iceland.

Genotyping and association. Genotyping and imputation methods and the association analysis in the Icelandic samples was performed as follows: In brief, we sequenced the whole genome of 15,220 Icelanders using Illumina sequencers to a mean depth of at least $\times 10$ (mean 30, median $\times 32$)¹¹, using three different library preparation methods from Illumina: (a) the standard TruSeq DNA library preparation method; Illumina GAIIX and/or HiSeq 2000 sequencers; (b) the TruSeq DNA PCR-free library preparation method; Illumina HiSeq 2500 sequencers; and (c) the TruSeq Nano DNA library preparation method; Illumina HiSeq X sequencers (see Supplementary Methods for a detailed description of the sequencing methods)¹¹. Genotypes of SNPs and indels were called using joint calling with the Genome Analysis Toolkit HaplotypeCaller (GATK version 3.3.0)⁴⁰. Using information about haplotype sharing genotype calls were improved, taking advantage of chip-typing and long-range phasing of all the sequenced individuals. In total, 32,463,443 genetic variants were called ($\text{info} > 0.8$ and $\text{MAF} > 0.01\%$). SNPs and indels that met the quality criteria were imputed into the 151,677 chip-typed Icelanders with the help of extensive genealogical information and long-range phased haplotypes¹¹. The sequence variants were imputed into 294,212 untyped relatives of the chip-typed individuals to further increase the sample size for association analysis and increase the power to detect associations^{10,11,41}.

We used the variant effect predictor⁴² to predict the maximal consequence of each sequence variant on all neighbouring RefSeq genes⁴³. There is a substantial variation in the enrichment of phenotype-associating sequence variants based on their annotations¹². On the basis of these enrichments, it is possible to group sequence variants into categories, in order of decreasing impact on biological function. We used the enrichment of variant classes to correct the threshold for genome-wide significance with a weighted Bonferroni adjustment¹². With 32,463,443 sequence variants tested, the weights given in Sveinbjornsson *et al.*¹² were rescaled to control the family-wise error rate (Supplementary Table 2). This yielded significance thresholds of $P < 2.6 \times 10^{-7}$ for high-impact variants ($N = 8,474$, including stop gained, frameshift, splice acceptor or donor), $P < 5.1 \times 10^{-8}$ for moderate-impact variants ($N = 149,983$, including missense, splice-region variants and in-frame INDELs), $P < 4.6 \times 10^{-9}$ for low-impact variants ($N = 2,283,889$, including synonymous variants 3'- and 5'-untranslated region variants), $P < 2.3 \times 10^{-9}$ for other variants overlapping DNase hypersensitivity sites ($N = 3,913,058$) and $P < 7.9 \times 10^{-10}$ for other non-DNase hypersensitivity site variants, intergenic and deep intronic ($N = 26,108,039$)¹² (Supplementary Table 2). For association testing in the case-control analysis, we used logistic regression; disease status was treated as the response and genotype counts were used as covariates. We also included in the model as nuisance variables the following available individual characteristics that correlate with disease status; county of birth, sex, current age or age of death (first- and second-order terms included), availability of blood sample for the individual and an indicator function for the overlap of the timespan of phenotype collection with lifetime of the individual^{11,44,45}. We applied LD score regression to estimate a correction factor to distinguish polygenicity from population stratification in the GWAS results⁴⁶. To correct for the relatedness of the Icelandic individuals included in this study, we applied the method of genomic control⁴⁷ where the inflation in the χ^2 values was estimated on the basis of a subset of about 300,000 common variants, and P values were adjusted by dividing the corresponding χ^2 values by this factor. For the diverticular disease, this factor was 1.18 and for diverticulitis 1.17.

A total of 5,426 individuals with diverticular disease were included in the association analysis; 3,368 of these were genotyped using various Illumina chips and imputed using long-range phased haplotypes and the remaining 1,958 were imputed on the basis of genotypes of first- and second-degree relatives¹¹. The same population controls, 245,951 individuals recruited through different deCODE projects, were used for association analysis of the three diverticular disease phenotypes: 124,228 genotyped and 121,723 imputed on the basis of genotypes of first- and second-degree relatives. All individuals with diverticular disease were excluded from the control list.

Single SNP genotyping in the replication cohort was performed at deCODE genetics with the Centaurus (Nanogen) platform⁴⁸. The rs761545809 indel was typed using a PCR-based method using NED-labelled (yellow fluorescent dye-labelled primer, Applied Biosystems) primers. An internal size standard was added to the resulting PCR products and the fragments were separated and detected on an Applied Biosystems Model 3730 Sequencer, using in-house Allele Caller Software. Test for association in the Danish replication samples was done using logistic regression implemented in the NEMO Software⁴⁹. The results from the replication were combined with the discovery results using a Mantel-Haenszel model⁵⁰.

Expression analysis. RNA sequencing. Preparation of Poly-A cDNA sequencing libraries. Isolated total RNA samples were assessed for quality and quantity using the Total RNA 6000 Nano Chip for the Agilent 2100 Bioanalyser. We generated cDNA libraries derived from Poly-A mRNA using Illumina's TruSeq RNA Sample Prep Kit. Briefly, using hybridization to Poly-T beads we isolated Poly-A mRNA from total RNA samples (1–4 μg input). The Poly-A mRNA was fragmented at 94 °C, and first-strand cDNA prepared using SuperScript II Reverse Transcriptase (Invitrogen) and random hexamers, followed by second-strand cDNA synthesis, end repair, addition of a single A base, adaptor ligation, AMPure bead purification and PCR amplification. The resulting cDNA was measured using the DNA 1000 Lab Chip on a Bioanalyser.

Sequencing. We used using Illumina's cBot and the TruSeq PE Cluster Kits v2 to cluster the samples on to flow cells. Then, we performed paired-end sequencing with either HiSeq 2000 Instruments using TruSeq v3 Flow Cells/SBS Kits or GAIIx Instruments using the TruSeq SBS Kits v5 from Illumina. Read lengths were 2×125 cycles.

Read alignment. We aligned the RNA sequencing reads to Homo Sapiens (Build 38) with TopHat version 2.0.12 with a set of known transcripts supplied in GTF format (RefSeq hg38; Homo sapiens, NCBI, build 38). TopHat was configured in such a way to first attempt to align reads to the provided transcriptome, following, for reads that do not map fully to the transcriptome, an attempt to map them onto the genome.

RNA-seq quality control. RNA libraries were excluded if the number of mapped reads were $< 10^7$ or number of mapped read pairs were $< 10^6$ or if the mapping rate of the first or second read end fell below 80% relative to the mapping of the other read end. Genotype concordance was determined by comparing imputed genotypes to those derived from RNA-seq. Samples surpassing exclusion had median 103 million mapped reads.

RNA transcript expression. Transcript abundance was estimated with kallisto⁵¹ version 0.43 using Ensembl r87 transcriptome with subset to transcripts annotated as GENECODE Basic or Transcript support level 1. Transcripts with minimum five counts in each sample for at least 47% of the samples were included in the downstream analysis. Association between sequence variants and log-transformed transcripts abundances (transcripts per million) was tested on samples from the whole blood ($n = 2,947$) and adipose tissue ($n = 766$) using linear regression model with sequencing covariates listed in RNA exon expression analysis.

RNA exon expression. Fragments basepair aligning to exons were counted and scaled in terms of exon length and sequenced library size. Association between sequence variants and normalized expression was tested on samples from whole blood ($n = 2,246$) and adipose tissue ($n = 708$) using linear regression model with sequencing covariate terms: (1) fragment length mean, (2) exonic mapping rate, (3) number of genes detected. For blood samples, covariate terms: (4) sample preparation method and (5) read length was included. For association with adipose base libraries, the covariate terms: (4) number of alternative alignments, (5) number of mapped pairs and (6) percentage of coding bases were included.

ROS production test. Phagocytosis assay was performed with the Phagoburst Kit (Glycotape-Biotechnology) using concentrations of reagents and incubation times according to the manufacturer's protocol. Whole blood of 12 heterozygote carriers of the sequence variant, 14 homozygotes and 14 non-carriers as controls was stimulated with *E. coli* (particulate stimulus), phorbol 12-myristate 13-acetate (high stimulus), *N*-formyl-MetLeuPhe (low physiological stimulus) or without stimulus to serve as a negative background control. To quantify reactive oxygen metabolites, the fluorogenic substrate dihydrorhodamine 123 was used and the reaction stopped by the addition of a lysing solution that removes erythrocytes and results in partial fixation of leukocytes. The fluorescent signal was measured with a BD FACSCalibur Flow Cytometer.

Screening for overlap with regulatory regions. To identify which associated variants might have regulatory effects, we selected the lead variant in each locus and searched with Haploreg 4.1⁵² for SNPs in LD $r^2 > 0.6$ that overlapped with predicted enhancers, DNase I hypersensitivity clusters and H3K4me1, H3K27ac and H3K9ac chromatin state assignments.

PRS for IBD in diverticular disease. PRS were calculated using publicly available summary statistics of IBD in Europeans from an Immunochip GWAS study on IBD⁵³. Summary statistics of the European subcohort, available at <https://www.ibdgenetics.org/downloads.html> (downloaded on 25 October 2016), were used to assign weights to SNPs. PLINK 1.9 was used to prune SNPs in a sliding window of 500 kb, retaining the SNP that showed the strongest evidence of association with the phenotype in the training data and removed SNPs having $r^2 > 0.1$ with that SNP. We excluded the extended MHC region (chr6:25,000,000–35,000,000) from the PRS calculations due to the complex linkage disequilibrium in the region. A set of 960 whole-genome sequenced Icelanders, unrelated at six meioses served as LD reference. We calculated a polygenic score for each individual, in the target data at two different *P* value inclusion thresholds, $P < 5 \times 10^{-8}$ and $P < 0.05$. Each PRS was then tested for association with each disease using generalized additive regression with smoothed age, sex and the first five principal components as covariates. *P* values were adjusted for population stratification estimated by calculating association statistics from 10,000 randomly chosen SNPs with MAF $> 5\%$ and variance explained was estimated using Nagelkerke's pseudo- R^2 (Supplementary Table 6).

Data availability. The lead variants for all association signals with *P* values less than two orders of magnitude above the relevant class-specific Bonferroni threshold are given in Supplementary Table 3a–c. Other relevant data are available from the corresponding authors on reasonable request.

References

- Humes, D. J., Ludvigsson, J. F. & Jarvholm, B. Smoking and the risk of hospitalization for symptomatic diverticular disease: a Population-Based Cohort Study from Sweden. *Dis. Colon Rectum* **59**, 110–114 (2016).
- Pfützer, R. H. & Kruijs, W. Management of diverticular disease. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 629–638 (2015).
- Tursi, A. Diverticulosis today: unfashionable and still under-researched. *Ther. Adv. Gastroenterol.* **9**, 213–228 (2016).
- Everhart, J. E. & Ruhl, C. E. Burden of digestive diseases in the United States part II: lower gastrointestinal diseases. *Gastroenterology* **136**, 741–754 (2009).
- Matrana, M. R. & Margolin, D. A. Epidemiology and pathophysiology of diverticular disease. *Clin. Colon Rectal Surg.* **22**, 141–146 (2009).
- Sheth, A. A., Longo, W. & Floch, M. H. Diverticular disease and diverticulitis. *Am. J. Gastroenterol.* **103**, 1550–1556 (2008).
- Peery, A. F. Recent advances in diverticular disease. *Curr. Gastroenterol. Rep.* **18**, 37 (2016).
- Strate, L. L. *et al.* Heritability and familial aggregation of diverticular disease: a population-based study of twins and siblings. *Gastroenterology* **144**, 736–742 e731 quiz e714 (2013).
- Granolund, J. *et al.* The genetic influence on diverticular disease—a twin study. *Aliment Pharmacol. Ther.* **35**, 1103–1107 (2012).
- Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
- Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
- Sveinbjornsson, G. *et al.* Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat. Genet.* **48**, 314–317 (2016).
- Costa, C. *et al.* The RacGAP ArhGAP15 is a master negative regulator of neutrophil functions. *Blood* **118**, 1099–1108 (2011).
- Radu, M. *et al.* ArhGAP15, a Rac-specific GTPase-activating protein, plays a dual role in inhibiting small GTPase signaling. *J. Biol. Chem.* **288**, 21117–21125 (2013).
- Seoh, M. L., Ng, C. H., Yong, J., Lim, L. & Leung, T. ArhGAP15, a novel human RacGAP protein with GTPase binding property. *FEBS Lett.* **539**, 131–137 (2003).
- Consortium G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
- Engel, A. G., Shen, X. M., Selcen, D. & Sine, S. M. Congenital myasthenic syndromes: pathogenesis, diagnosis, and treatment. *Lancet Neurol.* **14**, 420–434 (2015).
- Ge, X. *et al.* Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* **86**, 127–141 (2005).
- Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA* **101**, 6062–6067 (2004).
- Comuzzie, A. G. *et al.* Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS ONE* **7**, e51954 (2012).
- Wang, K. *et al.* A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. *Mol. Psychiatry* **16**, 949–959 (2011).
- Tursi, A. *et al.* Detection of endoscopic and histological inflammation after an attack of colonic diverticulitis is associated with higher diverticulitis recurrence. *J. Gastrointest. Liver Dis.* **22**, 13–19 (2013).
- Jostins, L. *et al.* Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
- Barrett, J. C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
- Parkes, M. *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* **39**, 830–832 (2007).
- Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42**, 1118–1125 (2010).
- Yang, S. K. *et al.* Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. *Gut* **63**, 80–87 (2014).
- Julia, A. *et al.* A genome-wide association study identifies a novel locus at 6q22.1 associated with ulcerative colitis. *Hum. Mol. Genet.* **23**, 6927–6934 (2014).
- Kenny, E. E. *et al.* A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS Genet.* **8**, e1002559 (2012).
- Connelly, T. M. *et al.* The TNFSF15 gene single nucleotide polymorphism rs7848647 is associated with surgical diverticulitis. *Ann. Surg.* **259**, 1132–1137 (2014).
- Beasley, W. D., Beynon, J., Jenkins, G. J. & Parry, J. M. Reprimo 824 G>C and p53R2 4696 C>G single nucleotide polymorphisms and colorectal cancer: a case-control disease association study. *Int. J. Colorectal Dis.* **23**, 375–381 (2008).

32. Asling, B. *et al.* Collagen type III alpha I is a gastro-oesophageal reflux disease susceptibility gene and a male risk factor for hiatus hernia. *Gut* **58**, 1063–1069 (2009).
33. Kluijvers, K. B. *et al.* COL3A1 2209G>A is a predictor of pelvic organ prolapse. *Int. Urogynecol. J. Pelvic Floor Dysfunct.* **20**, 1113–1118 (2009).
34. Parkes, M., Cortes, A., van Heel, D. A. & Brown, M. A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **14**, 661–673 (2013).
35. West, A. B. Ndsq. The pathology of diverticulitis. *J. Clin. Gastroenterol.* **42**, 1137–1138 (2008).
36. Tursi, A. Biomarkers in diverticular diseases of the colon. *Dig. Dis.* **30**, 12–18 (2012).
37. Jamal Talabani, A., Lydersen, S., Ness-Jensen, E., Endreseth, B. H. & Edna, T. H. Risk factors of admission for acute colonic diverticulitis in a population-based cohort study: the North Trondelag Health Study, Norway. *World J. Gastroenterol.* **22**, 10663–10672 (2016).
38. Thorgeirsson, T. E. *et al.* Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat. Genet.* **42**, 448–453 (2010).
39. Thorgeirsson, T. E. *et al.* A rare missense mutation in CHRNA4 associates with smoking behavior and its consequences. *Mol. Psychiatry* **21**, 594–600 (2016).
40. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
41. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868–874 (2009).
42. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
43. Eilbeck, K. *et al.* The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.* **6**, R44 (2005).
44. Steinhorsdottir, V. *et al.* Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat. Genet.* **46**, 294–298 (2014).
45. Sveinbjornsson, G. *et al.* Rare mutations associating with serum creatinine and chronic kidney disease. *Hum. Mol. Genet.* **23**, 6935–6943 (2014).
46. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
47. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
48. Kutuyavin, I. V. *et al.* A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res.* **34**, e128 (2006).
49. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.* **35**, 131–138 (2003).
50. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer. Inst.* **22**, 719–748 (1959).
51. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
52. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, D877–D881 (2016).
53. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
54. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Preprint at <http://dxdoi.org/10.1101/005165> (2014).
55. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103 (2010).
56. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).

Acknowledgements

We thank all study participants who provided data for this study and our valued colleagues who contributed to data collection and phenotypic characterization of clinical samples, genotyping and analysis of genome sequences data. This study was funded in part by the National Institute on Drug Abuse (NIDA) (R01-DA017932).

Author contributions

S.S., U.T., D.F.G., T.S., I.J. and K.S. designed the study, coordinated the project and interpreted the results. S.S., B.F., E.S., A.N.-C., J.B., J.A., H.S.J., E.P.-M., H.U., M.M., T.R., T.T., I.J. and T.S. coordinated and managed collection of samples and ascertainment of phenotype data. S.G. and A.S. performed experiments and analysed results. S.S., K.F.A., P.S., G.H.H., S.O., G.Th., G.M. and D.F.G. performed statistical and bioinformatic analysis. S.S., D.F.G., I.J. and K.S. drafted the manuscript. All authors contributed to the final version of manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing interests: S.S., K.F.A., P.S., S.G., G.H.H., S.O., A.S., T.R., T.T., G.Th., G.M., U.T., D.F.G., I.J. and K.S. are employees of deCODE Genetics/Amgen Inc. The remaining authors declares no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Sigurdsson, S. *et al.* Sequence variants in *ARHGAP15*, *COLQ* and *FAM155A* associate with diverticular disease and diverticulitis. *Nat. Commun.* **8**, 15789 doi: 10.1038/ncomms15789 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017