



OPEN ACCESS

Original research

# Genome-wide analysis of 944 133 individuals provides insights into the etiology of haemorrhoidal disease

Tenghao Zheng <sup>1,2</sup> David Ellinghaus <sup>3,4</sup> Simonas Juzenas <sup>3,5</sup> François Cossais,<sup>6</sup> Greta Burmeister,<sup>7</sup> Gabriele Mayr,<sup>3</sup> Isabella Friis Jørgensen,<sup>4</sup> Maris Teder-Laving,<sup>8</sup> Anne Heidi Skogholt,<sup>9</sup> Sisi Chen,<sup>10</sup> Peter R Stregé,<sup>10</sup> Go Ito,<sup>3,11</sup> Karina Banasik,<sup>4</sup> Thomas Becker,<sup>12</sup> Frank Bokelmann,<sup>13</sup> Søren Brunak,<sup>4</sup> Stephan Buch,<sup>14</sup> Hartmut Clausnitzer,<sup>15</sup> Christian Datz <sup>16</sup> DBDS Consortium, Frauke Degenhardt,<sup>3</sup> Marek Doniec,<sup>17</sup> Christian Erikstrup,<sup>18</sup> Tõnu Esko,<sup>8</sup> Michael Forster <sup>3</sup> Norbert Frey,<sup>19,20,21</sup> Lars G Fritsche,<sup>22</sup> Maiken Elvestad Gabrielsen,<sup>9</sup> Tobias Gräßle,<sup>23,24</sup> Andrea Gsur <sup>25</sup> Justus Gross,<sup>7</sup> Jochen Hampe <sup>14,26</sup> Alexander Hendricks,<sup>7</sup> Sebastian Hinz,<sup>7</sup> Kristian Hveem,<sup>9</sup> Johannes Jongen,<sup>27,28</sup> Ralf Junker,<sup>15</sup> Tom Hemming Karlsen,<sup>29</sup> Georg Hemmrich-Stanisak,<sup>3</sup> Wolfgang Kruis,<sup>30</sup> Juozas Kupcinskas,<sup>31</sup> Tilman Laubert,<sup>27,28,32</sup> Philip C Rosenstiel,<sup>3,33</sup> Christoph Röcken <sup>34</sup> Matthias Laudes,<sup>35</sup> Fabian H Leendertz,<sup>23,24</sup> Wolfgang Lieb,<sup>36</sup> Verena Limperger,<sup>15</sup> Nikolaos Margetis,<sup>37</sup> Kerstin Mätz-Rensing,<sup>38</sup> Christopher Georg Németh,<sup>12,39</sup> Eivind Ness-Jensen <sup>9,40,41</sup> Ulrike Nowak-Göttl,<sup>15</sup> Anita Pandit,<sup>22</sup> Ole Birger Pedersen,<sup>42</sup> Hans Günter Peleikis,<sup>27,28</sup> Kenneth Peuker,<sup>14,26</sup> Cristina Leal Rodriguez,<sup>4</sup> Malte Christoph Rühlemann <sup>3</sup> Bodo Schniewind,<sup>43</sup> Martin Schulzky,<sup>3</sup> Jurgita Skieceviciene,<sup>31</sup> Jürgen Tepel,<sup>44</sup> Laurent Thomas,<sup>9,45,46,47</sup> Florian Uellendahl-Werth,<sup>3</sup> Henrik Ullum,<sup>48</sup> Ilka Vogel,<sup>49</sup> Henry Volzke,<sup>50</sup> Lorenzo von Fersen,<sup>51</sup> Witigo von Schönfels,<sup>12</sup> Brett Vanderwerff,<sup>22</sup> Julia Wilking,<sup>7</sup> Michael Wittig,<sup>3</sup> Sebastian Zeissig,<sup>14,26</sup> Myrko Zobel,<sup>52</sup> Matthew Zawistowski,<sup>22</sup> Vladimir Vacic,<sup>53</sup> Olga Sazonova,<sup>53</sup> Elizabeth S Noblin,<sup>53</sup> The 23andMe Research Team, Gianrico Farrugia <sup>10</sup> Arthur Beyder,<sup>10</sup> Thilo Wedel,<sup>6</sup> Volker Kahlke,<sup>27,28,54</sup> Clemens Schafmayer,<sup>7</sup> Mauro D'Amato <sup>1,2,55,56</sup> Andre Franke <sup>3,33</sup>

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2020-323868>).

For numbered affiliations see end of article.

## Correspondence to

Prof Mauro D'Amato, Gastrointestinal Genetics Lab, CIC bioGUNE - BRTA, Bizkaia Science and Technology Park, Building 801A, 48160 Derio, Spain; [mdamato@cicbiogune.es](mailto:mdamato@cicbiogune.es) and Prof Andre Franke, Institute of Clinical Molecular Biology (IKMB) & University Hospital Schleswig-Holstein (UKSH), Christian-Albrechts-University of Kiel (CAU), Rosalind-Franklin-Str. 12, D-24105 Kiel, Germany; [a.franke@mucosa.de](mailto:a.franke@mucosa.de)

TZ, DE and SJ contributed equally.

CS, MD and AF contributed equally.

Received 18 December 2020

Revised 19 March 2021

Accepted 30 March 2021

Published Online First

22 April 2021



► <http://dx.doi.org/10.1136/gutjnl-2021-324817>



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY. Published by BMJ.

**To cite:** Zheng T, Ellinghaus D, Juzenas S, *et al.* *Gut* 2021;**70**:1538–1549.

## ABSTRACT

**Objective** Haemorrhoidal disease (HEM) affects a large and silently suffering fraction of the population but its aetiology, including suspected genetic predisposition, is poorly understood. We report the first genome-wide association study (GWAS) meta-analysis to identify genetic risk factors for HEM to date.

**Design** We conducted a GWAS meta-analysis of 218 920 patients with HEM and 725 213 controls of European ancestry. Using GWAS summary statistics, we performed multiple genetic correlation analyses between HEM and other traits as well as calculated HEM polygenic risk scores (PRS) and evaluated their translational potential in independent datasets. Using functional annotation of GWAS results, we identified HEM candidate genes, which differential expression and coexpression in HEM tissues were evaluated employing RNA-seq analyses. The localisation of expressed proteins at selected loci was investigated by immunohistochemistry.

**Results** We demonstrate modest heritability and genetic correlation of HEM with several other diseases from the GI, neuroaffective and cardiovascular domains. HEM PRS validated in 180 435 individuals from independent datasets allowed the identification of those at risk and correlated with younger age of onset and recurrent surgery. We identified 102 independent HEM risk loci harbouring genes whose expression is enriched in blood vessels and GI tissues, and in pathways associated with smooth muscles, epithelial and endothelial development and morphogenesis. Network transcriptomic analyses highlighted HEM gene coexpression modules that are relevant to the development and integrity of the musculoskeletal and epidermal systems, and the organisation of the extracellular matrix.

**Conclusion** HEM has a genetic component that predisposes to smooth muscle, epithelial and connective tissue dysfunction.

## Significance of this study

## What is already known about this subject?

- ▶ Human haemorrhoidal disease (HEM) is a prevalent anorectal pathology characterised by the symptomatic enlargement and distal displacement of anal cushions.
- ▶ As a consequence of scarce research and of being a taboo topic in society, only a few HEM non-genetic risk factors have been suggested, thus the aetiology of the disease still remains unclear.
- ▶ Genetic susceptibility to HEM has been suspected but was never systematically investigated.

## What are the new findings?

- ▶ Here, we report the first well-powered genome-wide association study (GWAS) meta-analysis with a sample size of 944 133 individuals to identify genetic risk factors for HEM.
- ▶ We describe 102 novel independent HEM risk loci that are functionally linked to pathways associated with smooth muscles, epithelial and endothelial development and morphogenesis.
- ▶ We show genetic correlations of HEM with several other diseases that are classified as GI, neuroaffective and cardiovascular disorders.
- ▶ We report significance of computed HEM polygenic risk scores, which were validated in independent population-based cohorts.
- ▶ We show significant enrichment of HEM genes in tissue coexpression modules responsible for the development of musculoskeletal and epidermal systems, and the organisation of the extracellular matrix.
- ▶ Based on our data, we outline HEM as a disorder of impaired neuromuscular motility, smooth muscle contraction and extracellular matrix organisation.

## How might it impact on clinical practice in the foreseeable future?

- ▶ The results from this GWAS provide new insights on HEM genetic predisposition to smooth muscle, epithelial and connective tissue dysfunction.
- ▶ HEM polygenic risk scores identify individuals at risk and correlate with a more severe phenotype.

## INTRODUCTION

Haemorrhoids are normal anal vascular cushions filled with blood at the junction of the rectum and the anus. It is assumed that their main role in humans is to maintain continence<sup>1</sup> but other functions such as sensing fullness, pressure and perceiving anal contents have been suggested given the sensory innervation.<sup>2</sup> Haemorrhoidal disease (hereafter referred to as HEM) occurs when haemorrhoids enlarge and become symptomatic (sometimes associated with rectal bleeding and itching/soiling) due to the deterioration or prolapse of the anchoring connective tissue, the dilation of the haemorrhoidal plexus or the formation of blood clots. Severe forms of HEM often require surgical treatment and the removal of abnormally enlarged and/or thrombosed haemorrhoids.<sup>1</sup> HEM prevalence increases with age and shows staggering figures worldwide (up to 86% prevalence in some reports),<sup>3</sup> whereby a large proportion of cases remain undetected as asymptomatic or mild enough to be self-treated with over-the-counter treatment. HEM represents a considerable medical and socioeconomic burden with an estimated annual cost of US\$800 million in the USA alone, mainly

related to the large number of haemorrhoidectomies performed every year.<sup>4</sup>

A number of HEM risk factors have been suggested, including human erect position. The tight anal sealing provided by the elaborated haemorrhoidal plexus may have developed during human evolution co-occurring with permanent bipedalism, as shown by our histology comparison of four different mammals (human, gorilla, baboon, mouse; online supplemental figure 1, online supplemental material 1). Other suggested risk factors are a sedentary lifestyle, obesity, reduced dietary fibre intake, spending excess time on the toilet, straining during defecation, strenuous lifting, constipation, diarrhoea, pelvic floor dysfunction, pregnancy and giving natural birth, with several being controversially reported. The hypothetical model shown in online supplemental figure 2 summarises the contemporary concepts regarding the pathophysiology of HEM development.<sup>5</sup> Until today, HEM aetiopathogenesis is poorly investigated, and neither the exact molecular mechanisms nor the reason(s) why only some people develop HEM are known. Genetic susceptibility may play a role in HEM development, but no large-scale, genome-wide association study (GWAS) for HEM has ever been conducted. To evaluate the contribution of genetic variation to the genetic architecture of HEM, we carried out a GWAS meta-analysis in 218 920 affected individuals and 725 213 population controls of European ancestry.

## METHODS

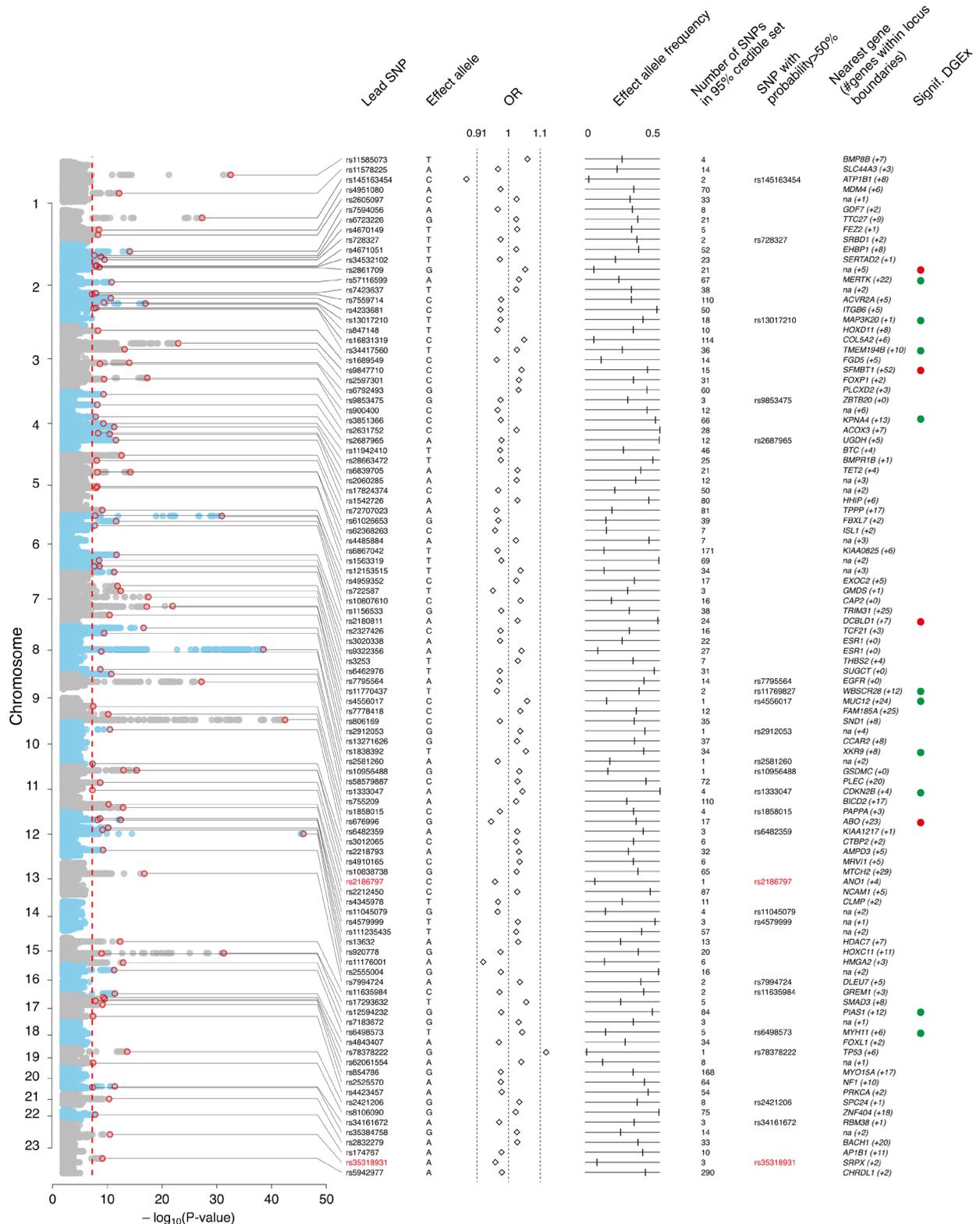
Detailed methods are provided in the online supplemental material 1 and summarised in the online supplemental figure 3.

## RESULTS

## GWAS meta-analysis and fine-mapping genomic regions

We conducted a GWAS meta-analysis in 944 133 individuals of European ancestry from five large population-based cohorts (23andMe,<sup>6</sup> UK Biobank (UKBB),<sup>7</sup> Estonian Genome Centre at the University of Tartu,<sup>8</sup> Michigan Genomics Initiative,<sup>9</sup> and Genetic Epidemiology Research on Aging (GERA)<sup>10</sup>) including 218 920 HEM cases and 725 213 controls (online supplemental table 1). HEM cases had higher body mass index (BMI), were significantly older and more often women compared with non-HEM controls; hence, age, sex, BMI (if available) and top principal components from principal component analysis (PCA) were included as covariates in individual GWAS (see 'Methods' section in online supplemental material 1, hereafter referred to as online methods).

After data harmonisation and quality control, 8 494 288 high-quality, common (minor allele frequency >0.01) single nucleotide polymorphisms (SNPs) were included in a fixed-effect inverse variance meta-analysis using the software METAL<sup>11</sup> (online methods). We identified 5480 genome-wide significant associations ( $P_{\text{Meta}} < 5 \times 10^{-8}$ ), which were mapped to 102 independent genomic regions using FUMA<sup>12</sup> (online methods). A summary of the association results for 102 lead SNPs is provided in figure 1 and online supplemental table 2. Although genomic inflation was observed ( $\lambda = 1.3$ ; online supplemental figure 4), this was likely due to polygenicity rather than population stratification, as determined via linkage disequilibrium score regression analysis (LDSC, intercept = 1.06; online methods) and based on a normalised  $\lambda_{1000} = 1.001$ . The heritability of HEM was estimated at 5% (SNP-based heritability,  $h^2_{\text{SNP}}$  computed with LDSC), whereby the newly identified 102 risk variants explained about 0.9% of the variance (online methods).



**Figure 1** Annotation of 102 haemorrhoidal disease (HEM) genome-wide association study (GWAS) risk loci. From left to right: Manhattan plot of GWAS meta-analysis results, (genome-wide significance level— $P_{Meta} < 5 \times 10^{-8}$ —indicated with vertical dotted red line); Lead single nucleotide polymorphism (SNP)—marker associated with the strongest association signal from each locus (also annotated with a red circle in the Manhattan plot); Effect allele—allele associated with reported genetic risk effects (OR), also always the minor allele; OR with respect to the effect allele; Effect allele frequency—frequency of the effect allele in the discovery dataset; Number of SNPs in 95% credible set—the minimum set of variants from Bayesian fine-mapping analysis that is >95% likely to contain the causal variant; SNP with probability >50%—single variant (if detected) with >50% probability of being causal (coding SNPs highlighted in red); Nearest gene (#genes within locus boundaries)—gene closest to the lead SNP (if within 100 kb distance, otherwise 'na') and number of additional genes positionally mapped to the locus using FUMA (online supplemental table 2 and online methods). Signif. DGEx—locus containing HEM genes differentially expressed in RNA Combo-Seq analysis of HEM affected tissue, detected at higher (green) and/or lower (red) level of expression (see online methods).

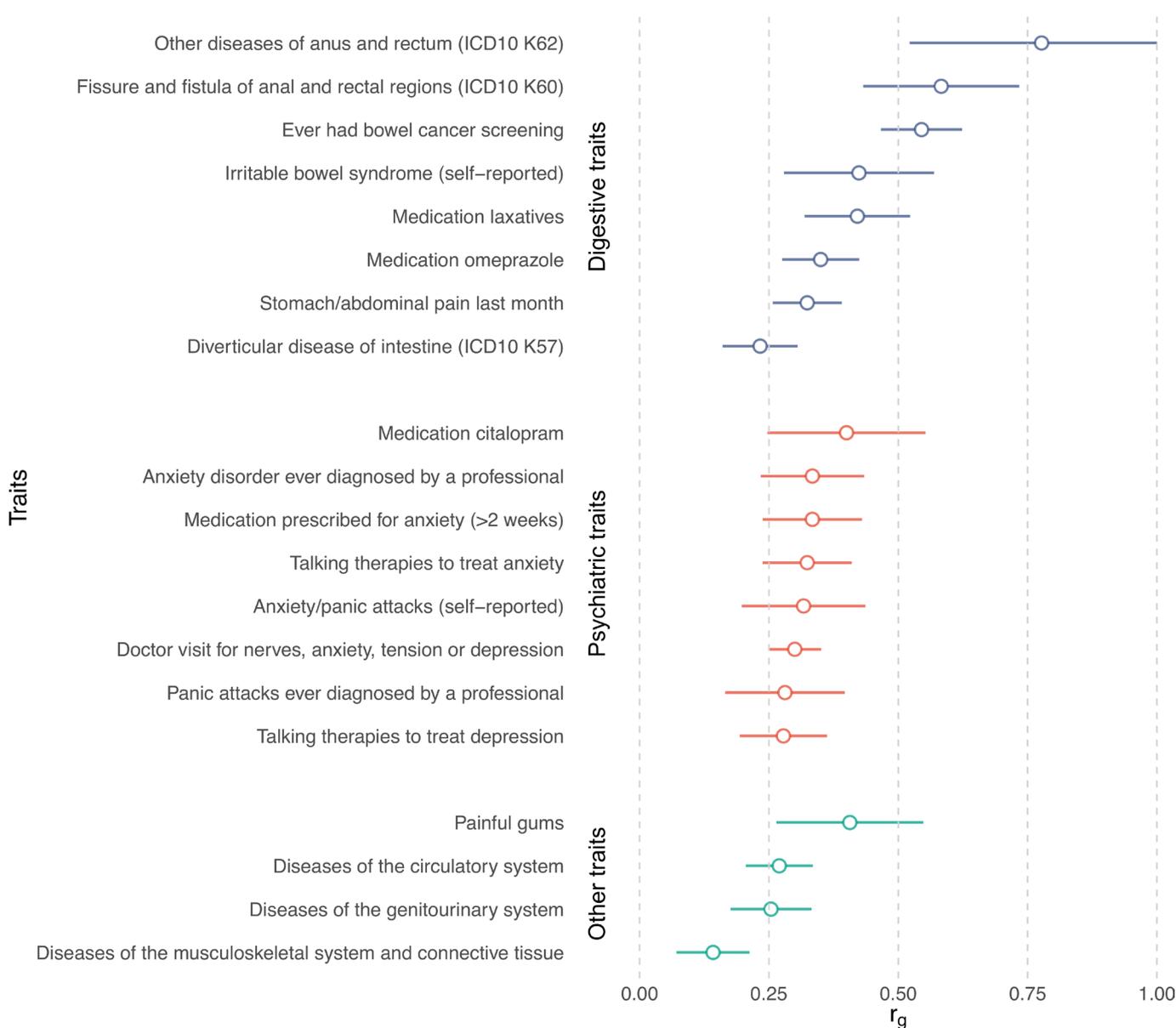
Gut: first published as 10.1136/gutjnl-2020-323868 on 22 April 2021. Downloaded from <http://gut.bmj.com/> on April 19, 2024 by guest. Protected by copyright.

Bayesian fine-mapping analysis with FINEMAP<sup>13</sup> identified a total of 3323 SNPs that belong to the 95% credible sets of variants most likely to be causal at each locus (online methods, online supplemental figure 5 and online supplemental table 3). For six loci, we pinpointed the association signal to a single causal variant with greater than 95% certainty, including two missense variants rs2186797 (*ANO1*) and rs35318931 (*SRPX*). For another 24 loci, there was evidence that the lead variant is causal with >50% certainty (figure 1).

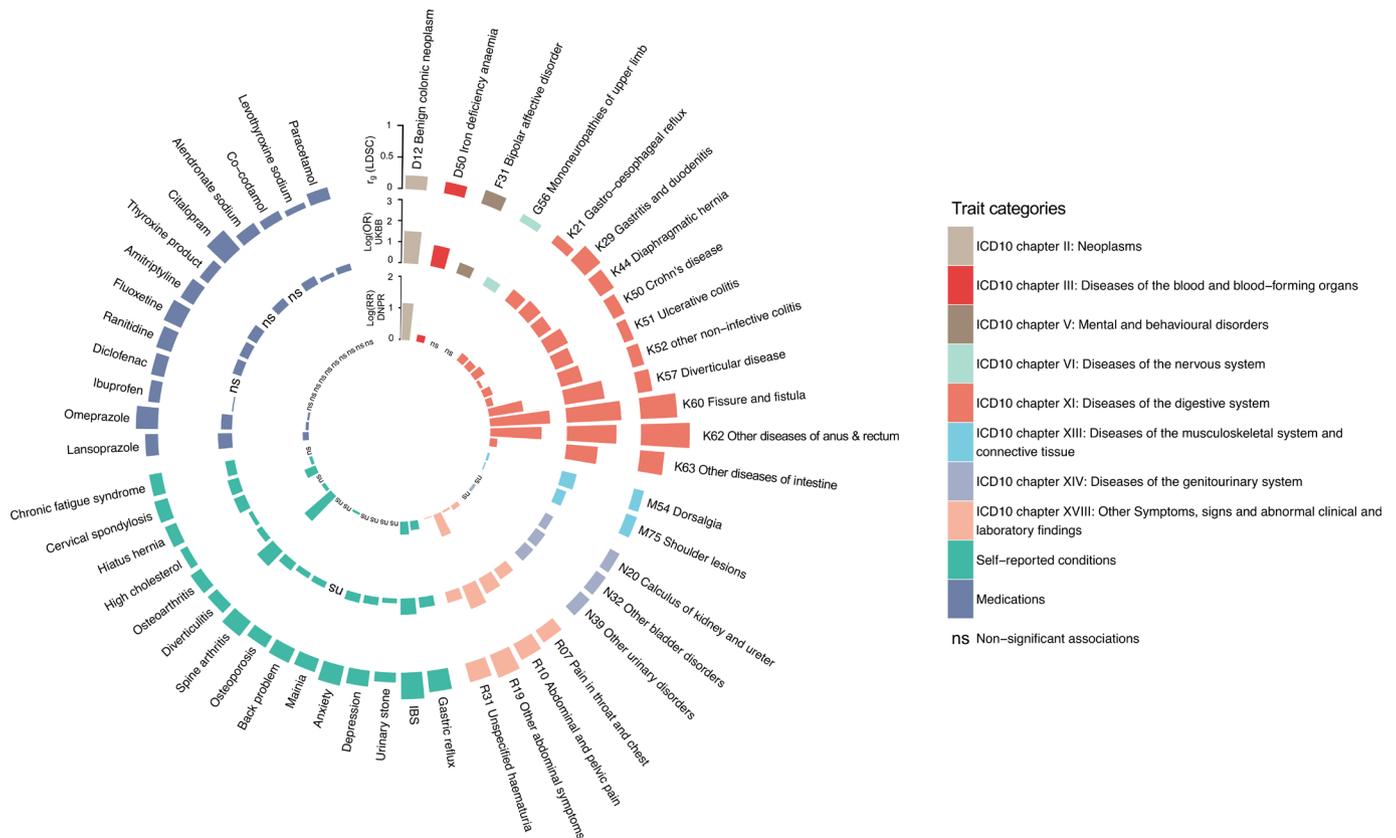
### Cross-trait analyses

A lookup of HEM association signals in previous GWAS studies retrieved from GWAS Atlas,<sup>14</sup> GWAS Catalog<sup>15</sup> and via PhenoScanner v2<sup>16</sup> (online methods) revealed that 76/102 loci had been previously associated with diseases and traits across the metabolic, cardiovascular, digestive, psychiatric, environmental and other domains (online supplemental figure 6 and online supplemental table 4).

We then investigated genetic correlations with 1387 other human traits and conditions using LDSC as implemented in CTG-VL<sup>17</sup> (online methods). The strongest correlations ( $r_g$ ) were observed with diseases and traits from the GI domain, including ‘other diseases of anus and rectum’ ( $r_g=0.78$ ,  $P_{FDR}=4.94\times 10^{-8}$ ), ‘fissure and fistula of anal and rectal regions’ ( $r_g=0.58$ ,  $P_{FDR}=2.70\times 10^{-12}$ ), ‘self-reported IBS’ ( $r_g=0.42$ ,  $P_{FDR}=1.87\times 10^{-7}$ ), use of ‘laxatives’ ( $r_g=0.42$ ,  $P_{FDR}=2.35\times 10^{-14}$ ), and ‘diverticular disease’ ( $r_g=0.23$ ,  $P_{FDR}=6.68\times 10^{-9}$ ) among others (figure 2). Given these similarities, we performed a genome-wide pleiotropy analysis for diverticular disease, IBS and HEM, revealing 44 independent genomic regions shared by at least two phenotypes from the group of diverticular disease, IBS and HEM phenotypes (online supplemental table 5). Other notable correlations were detected for psychiatric and neuro-affective disorders (anxiety, depression and neuroticism), pain-related traits (including abdominal pain and painful gums), diseases of the circulatory system, and diseases of the musculoskeletal system and connective tissue (figure 2 and online supplemental table 6).



**Figure 2** Genetic correlation between haemorrhoidal disease and other traits estimated by linkage disequilibrium score regression analysis. Genetic correlations ( $r_g$  +se) are shown for selected traits, grouped by domain. Only correlations significant after Bonferroni correction were considered (full list available in online supplemental table 6). ICD, International Classification of Diseases.



**Figure 3** Analysis of haemorrhoidal disease (HEM) genetically correlated traits in UK Biobank (UKBB) and the Danish National Patient Registry (DNPR). Traits and conditions identified in linkage disequilibrium score regression analyses of genetic correlation with HEM (outer ring in the circo plot, see also figure 2 and online supplemental table 6) were studied for their differential prevalence in UKBB and DNPR, based on data extracted from participants' healthcare records. Significant results are reported, respectively, as ORs (log(OR), UKBB, middle ring) and relative risk (log(RR), DNPR, inner ring) or 'ns' (for non-significant findings). Diseases and traits are categorised according to ICD10 diagnostic codes or self-reported conditions and use of medications from questionnaire data (see online methods). Self-reported traits in UKBB (dark blue colour) were manually mapped to ICD10-codes in DNPR.

To gain further insight into potential cross-trait overlaps and to validate the LDSC results, we assessed whether genetically correlated traits and conditions also occur more frequently in patients with HEM by analysing data on diagnoses and medications from UKBB. The results were highly consistent with those obtained via LDSC (figure 3, online supplemental table 6). Compared with controls, patients with HEM additionally suffered more often from diverticular disease, IBS and other functional GI disorders (FGIDs), abdominal pain, hypertension, ischaemic heart disease, depression, anxiety, and diseases of the musculoskeletal system and connective tissue among others. To further consolidate these findings, we analysed an independent population-scale healthcare record comprising 8 172 531 individuals from the Danish National Patient Registry (online methods), obtaining similar results (see inner circle of figure 3). Therefore, based on these largely overlapping observations at the genetic and epidemiological level, HEM appears to be strongly associated with other diseases of the digestive, neuroaffective and cardiovascular domains.

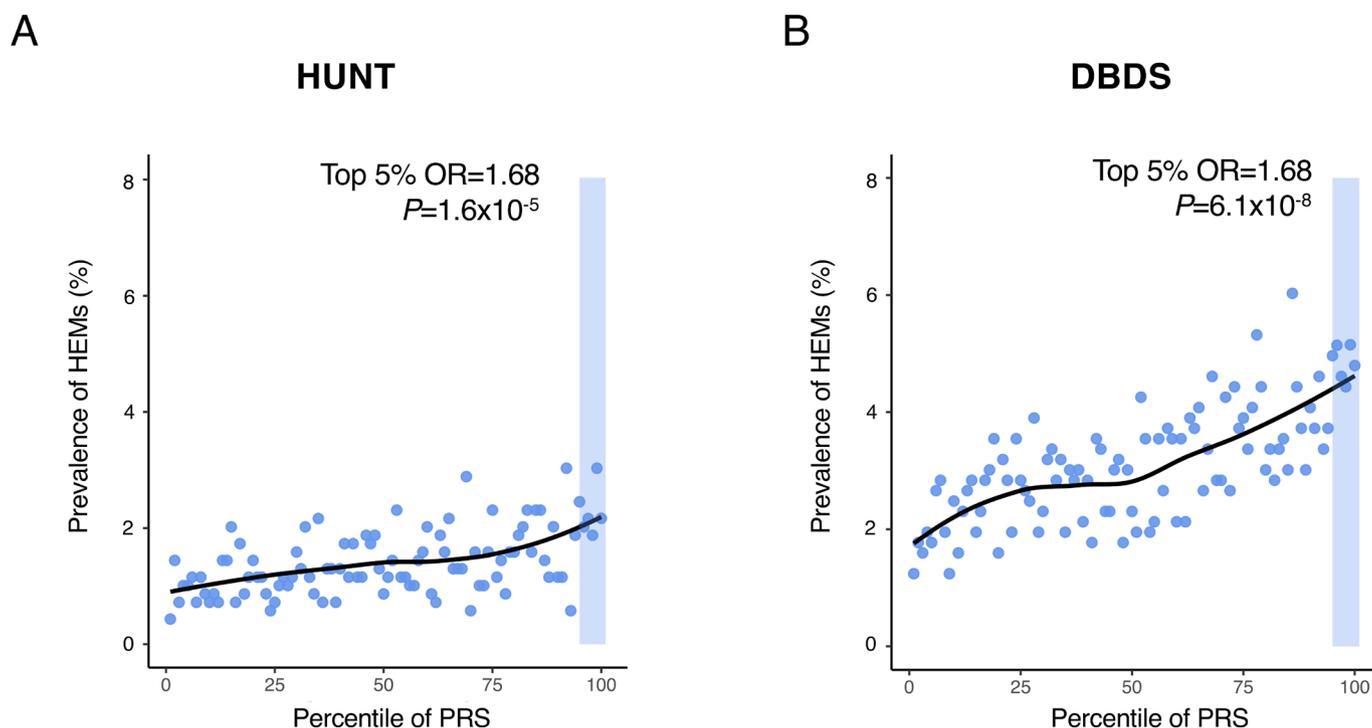
### Polygenic risk scores (PRS)

We next exploited meta-analysis summary statistics to compute HEM PRS and evaluate their relevance and translational potential in independent datasets. HEM PRS were calculated with PRSice-2<sup>18</sup> (online methods) and their performance tested in three independent datasets: (1) the Norwegian Trøndelag Health

Study (HUNT<sup>19</sup>; n=69 291; 977 cases vs 68 314 controls), (2) the Danish Blood Donor Study (DBDS<sup>20</sup>; n=56 397; 1754 cases vs 54 643 controls), and (3) 1144 HEM cases from gastroenterology clinics compared with 2740 cross-sectional population controls from Germany<sup>21</sup> (online supplemental table 1). In all three datasets, patients with HEM showed significantly higher PRS values compared with controls (OR=1.24,  $p=1.98 \times 10^{-11}$ ; OR=1.28,  $p=3.52 \times 10^{-22}$  and OR=1.36,  $p=7.64 \times 10^{-15}$ , respectively for the two population-based cohorts HUNT and DBDS, and the German case-control cohort). In HUNT and DBDS, HEM prevalence increased across PRS percentile distributions, with individuals from the top 5% tail being exposed to higher HEM risk compared with the rest of the population (OR=1.68,  $p=1.55 \times 10^{-5}$ , and OR=1.68,  $p=6.11 \times 10^{-8}$ , respectively for HUNT and DBDS, figure 4). Higher HEM PRS were also associated with a more severe phenotype as defined by the need for recurrent invasive procedures (OR=1.03,  $p=8.63 \times 10^{-3}$  in German patients) and a younger age of onset ( $p=1.90 \times 10^{-3}$  in DBDS;  $p=4.01 \times 10^{-3}$  in German patients).

### Functional annotation of GWAS loci, tissue and pathway enrichment analyses

In order to identify most likely candidate genes and relevant molecular pathways, we used independent computational pipelines for the functional annotation of GWAS results. This yielded



**Figure 4** Risk haemorrhoidal disease (HEM) prevalence across polygenic risk score (PRS) percentile distributions. PRS was derived from the results of the association meta-analysis (see online methods). HEM prevalence (%; Y-axis) is reported on a scatter plot in relation to PRS percentile distribution (X-axis) in the Norwegian Trøndelag Health Study (HUNT) (A) and the Danish Blood Donor Study (DBDS) (B) population cohorts. The top 5% of the distribution is highlighted with a shaded area in both cohorts, and the results of testing HEM prevalence in this group versus the rest of the population are also reported (p value and OR from logistic regression; online methods).

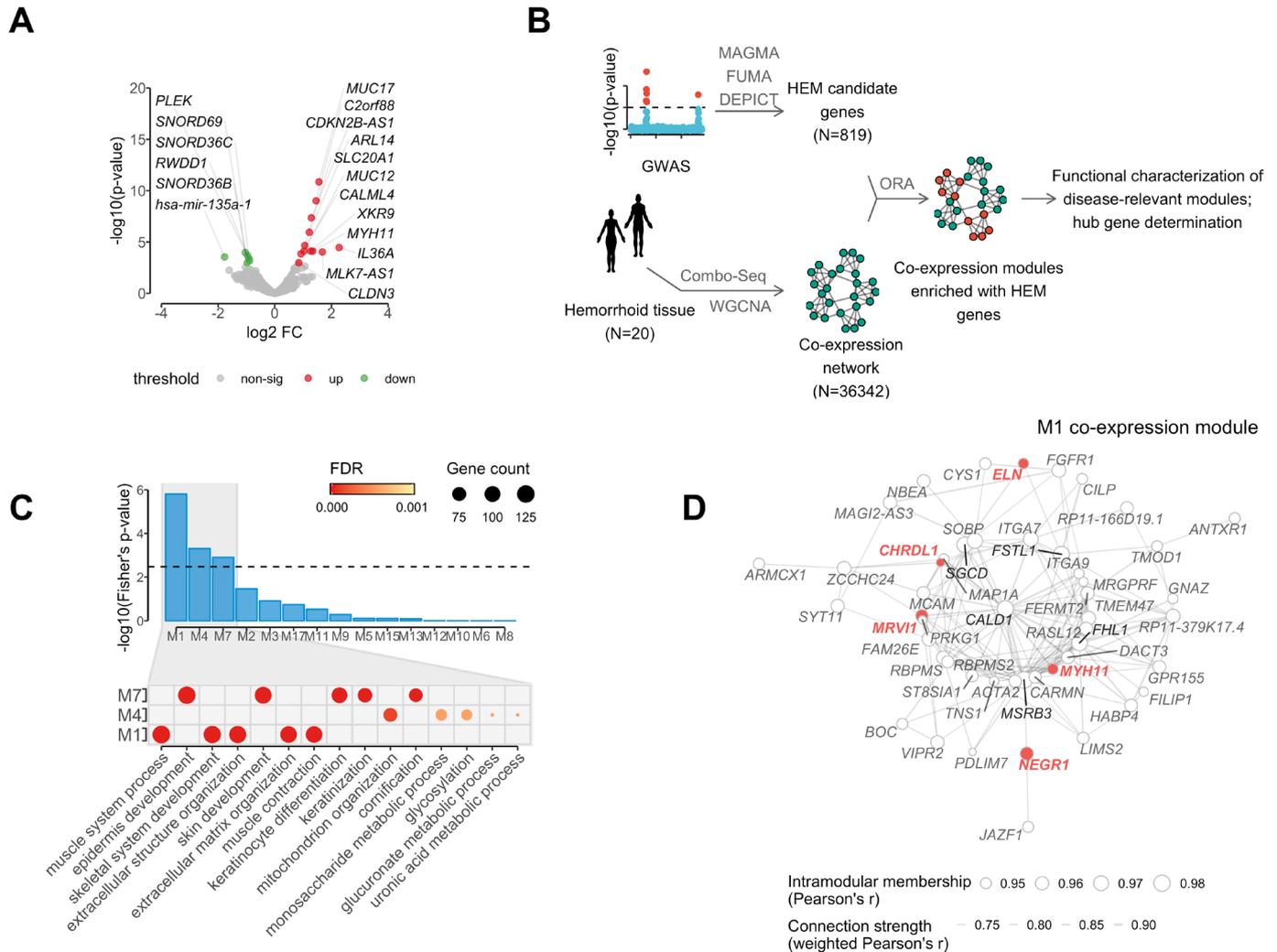
a total of 819 non-redundant HEM-associated transcripts (hereafter referred to as HEM genes; online supplemental table 7) derived from alternative positional (N=540 total from FUMA,<sup>12</sup> DEPICT<sup>22</sup> and MAGMA<sup>23</sup>) and expression quantitative trait (eQTL, N=562 from FUMA) mapping efforts (online methods). Tissue-specific enrichment analyses (TSEA) of 540 positional candidates led to very similar results for all three approaches, with enrichment of HEM gene expression in blood vessels, colon and other relevant tissues (online supplemental figure 7 and online supplemental table 8). Similarly, gene-set enrichment analyses (GSEA) highlighted common pathways important in the development of vasculature and the intestinal tissue including the gene ontology (GO) terms ‘tube morphogenesis and development’, ‘artery morphogenesis and development’, ‘epithelium morphogenesis’, ‘smooth muscle tissue morphogenesis’ and others. Additionally, DEPICT detected enrichment for a number of traits from the Mammalian Phenotype Ontology including ‘abnormal intestinal morphology’, ‘rectal prolapse’, ‘abnormal blood vessel (and artery) morphology’, and ‘abnormal smooth muscle morphology and physiology’ (online supplemental table 8). TSEA and GSEA analyses performed on all 819 transcripts including eQTL genes from FUMA gave rise to similar results although these did not reach statistical significance (not shown).

#### Gene expression in HEM tissue and gene-network analyses

The expression of HEM genes was studied in integrated mRNA and microRNA Combo-Seq analysis of enlarged haemorrhoidal tissue from 20 patients with HEM and normal specimens from 18 controls (online methods, online supplemental table 1). HEM genes were examined with regard to their expression status, differential expression between cases and controls, and their connectivity and topology in gene coexpression networks. After

normalisation for cell-type heterogeneity in different tissues (online methods, online supplemental figure 8), 720 out of 819 candidate genes were found to be expressed in haemorrhoidal tissue, with 287 (39.9%) of these being among the most strongly expressed transcripts (upper quartile) (online supplemental table 7). Compared with normal tissue from controls, 18 HEM candidate genes from 14 independent loci showed differential expression in haemorrhoidal tissue ( $P_{\text{FDR}} < 0.05$  and  $|\log_2 \text{fold change}| > 0.5$ ), with 12 genes showing increased and 6 decreased expression (figure 5A).

To obtain further biological insight from transcriptomic profiling, HEM candidate genes were further characterised for their membership in coexpressed gene modules identified via weighted gene correlation network analysis (WGCNA)<sup>24</sup> (figure 5B; online methods). The final network consisted of 36 342 genes partitioned into 41 coexpression modules, and 3 of these (M1, M4 and M7; referred to as HEM modules) were significantly enriched ( $P_{\text{FDR}} < 0.05$ ) for HEM genes (figure 5C). In total, 260 (35.7%) expressed HEM genes were members of modules M1 (N=121), M4 (N=75) or M7 (N=64) (online supplemental table 9). Functional annotation of these modules revealed M1 enrichment for ‘extracellular matrix (ECM) organisation’ ( $P_{\text{FDR}} = 5.17 \times 10^{-18}$ ) and ‘muscle contraction’ ( $P_{\text{FDR}} = 7.90 \times 10^{-17}$ ); M4 for ‘mitochondrion organisation’ ( $P_{\text{FDR}} = 2.35 \times 10^{-4}$ ) and ‘glycosylation’ ( $P_{\text{FDR}} = 5.70 \times 10^{-4}$ ); and M7 for ‘cornification’ ( $P_{\text{FDR}} = 3.97 \times 10^{-52}$ ) and ‘epidermis development’ ( $P_{\text{FDR}} = 2.07 \times 10^{-41}$ ) (figure 5C). The WGCNA analysis also allowed for the identification of the most interconnected genes, or module hub genes (ie, central nodes in the scale-free network). Of note, several HEM candidate genes, namely *NEGR1*, *MRV11*, *MYH11*, *ELN* and *CHRDL1*, were among the top 50 hub genes for module M1 (figure 5D). M1 is the module



**Figure 5** Analysis of mRNA and microRNA (Combo-Seq) data from haemorrhoidal disease (HEM) affected tissue, in relation to HEM genes coexpression networks. (A) Volcano plot reporting HEM genes differentially expressed in haemorrhoidal tissue (significantly upregulated=red, and downregulated=green); (B) schematic representation of the analytical flow for HEM genes coexpression network identification and characterisation; (C) upper panel (barplot): overrepresentation analysis of HEM genes in coexpression network modules, with significant enrichment ( $P_{\text{FDR}} < 0.05$ ) in modules M1, M4 and M7; lower panel (dotplot): top five gene ontology terms (biological process) from gene set enrichment analysis relative to M1, M4 and M7 coexpression modules (gene counts and false discovery rate (FDR)-adjusted significance level are also reported as indicated); (D) coexpression hub network of module M1. The network represents strength of connections (weighted Pearson's correlation  $> 0.7$ ) among the top 50 hub genes with highest values of intramodular membership (size of the node). HEM genes and the top 5 hub genes are highlighted in red and black, respectively.

most significantly enriched for HEM genes ( $P_{\text{FDR}} = 6.40 \times 10^{-5}$ ), and also the largest ( $n = 3975$ ) coexpression module in the haemorrhoidal tissue (online supplemental table 9), thus pointing to the importance of its associated GO terms 'ECM organisation' and muscle contraction in HEM.

### Prioritised HEM genes

In order to identify genes most likely to play a causative role in HEM, we selected candidates based on a scoring approach by prioritising those associated with one or more of the following: (1) linked to a high-confidence fine-mapped variant (posterior probability (PP)  $> 50\%$ ), (2) differentially expressed in enlarged haemorrhoidal tissue, (3) highlighted by pathway and tissue/cell-type enrichment DEPICT analysis (online methods), or (4) predicted hub of a WGCNA coexpression HEM module (M1, M4, M7). This reduced the number of candidates from 819 to 100 prioritised genes associated with 58 independent HEM loci

(online supplemental table 7). Some notable observations were made in relation to a subset of these prioritised genes, whose associated evidence and known biological function(s) make them remarkably good candidates to play a role in HEM risk.

Two genes, *ANO1* and *SRPX*, were both linked to a single coding variant fine-mapped as causal with very high confidence (rs2186797, *ANO1*, p.Phe608Ser with PP=97.0%, and rs35318931, *SRPX*, p.Ser413Phe with PP=87.3%, respectively). *ANO1* encodes the voltage-gated calcium-activated anion channel anoctamin-1 protein, which is highly expressed in the interstitial cells of Cajal (ICCs) throughout the human GI tract, where it contributes to the control of intestinal motility and peristalsis.<sup>25</sup> The *ANO1*:p.Phe608Ser (F608S) variant is predicted to destabilise local protein structure and to disturb ANO1-activating phospholipid interactions (online supplemental figure 9). Indeed, site-directed mutagenesis and electrophysiology experiments in vitro showed that the amino acid 608 Phe to

Ser change leads to an increased voltage-dependent instantaneous  $\text{Cl}^-$  current, and a slowing of activation and deactivation kinetics (especially at high  $\text{Ca}^{2+}$  concentrations) (online supplemental figure 10). The X-linked gene *SRPX* codes for a Sushi repeat-containing protein whose domain composition implies a role in ECM, and is expressed in various ECM tissues including colon and liver.<sup>26</sup> The *SRPX*:p.Ser413Phe variant (rs35318931) may potentially destabilise the C-terminal domain of unknown function (DUF4174) (online supplemental figure 11), which is conserved in various ECM proteins (online methods, section In silico variant protein analysis).

At locus 7q22.1, the signal was fine-mapped with very high confidence to rs4556017 (PP=96.2%), which exerts eQTL effects on *ACHE* and *SRRT*, both showing high levels of expression in enlarged haemorrhoidal tissue, and found in the M4 and M7 coexpression modules (figure 5), respectively. However, while *SRRT* encodes a poorly characterised capped-RNA binding protein, *ACHE* appears a much better candidate as the gene encodes an enzyme that hydrolyses the neurotransmitter acetylcholine at neuromuscular junctions, and corresponds to the Cartwright blood group antigen Yt.<sup>27</sup> One more locus was fine-mapped to single-variant resolution with high confidence, SNP rs10956488 from locus 8q24.21 (PP=0.962), which is linked to an eQTL for *GSDMC*. Gasdermin C (encoded by *GSDMC*) is a poorly characterised member of the gasdermin family of proteins expressed in epithelial cells and in enlarged haemorrhoidal tissue, though the mechanisms of its eventual HEM involvement remain elusive.

Other genes were linked via eQTL to a variant mapped with PP >50% and were prioritised based on additional experimental evidence. Among these, the fine-mapped SNP rs6498573 from locus 16p13.11 (PP=63.2%) is associated with eQTL effects on *MYH11* (encoding muscle myosin heavy chain 11; online supplemental table 10), a gene coding for a smooth muscle myosin heavy chain that shows mRNA upregulation in enlarged haemorrhoidal tissue and constitutes a hub for the M1 coexpression module associated with ECM organisation and muscle contraction. Notable prioritised candidates were also observed at loci that were not fine-mapped, including *ELN* (encoding elastin, a key component of the ECM found in the connective tissue of many organs, highly expressed in enlarged haemorrhoidal tissue and hub gene for the M1 coexpression module), *COL5A2* (encoding type V collagen protein; highly expressed in HEM tissue and belong to M1 coexpression module), *PRDM6* (encoding a putative histone methyltransferase regulating vascular smooth muscle cells contractility, expressed in enlarged haemorrhoidal tissue and hub for the M1 module), and others (online supplemental table 7).

Finally, while no candidate genes could be highlighted from the top GWAS hit region on chr12q14.3 (rs11176001), both the second and third strongest GWAS signals were detected at loci linked to genes involved in the determination of blood groups, namely *ABO* (rs676996) and the Kell Blood Group Complex Subunit-Related Family member *XKR9* (rs1838392). In addition, blood group antigens are encoded at additional loci, including *ACHE* (rs4556017) and *XKR6* (identified by *MAGMA*). Imputation of human ABO blood types from genotype data revealed the O type to be associated with increased and A and B types decreased HEM risk, both in UKBB and GERA datasets (online supplemental figure 12).

### Localisation of selected HEM gene-encoded proteins

A site-specific analysis of selected candidate proteins for their localisation in anorectal tissues underlined the complex multifactorial nature of cellular components potentially involved in the pathogenesis of HEM. Indeed, analysed candidates displayed a broad spectrum of expression in intestinal mucosal, neuromuscular, immune and anodermal tissues (figure 6, online supplemental figure 13, online supplemental tables 11 and 12). They were also directly colocalised with haemorrhoidal blood vessels, suggesting a putative role connected with the haemorrhoidal vasculature itself.

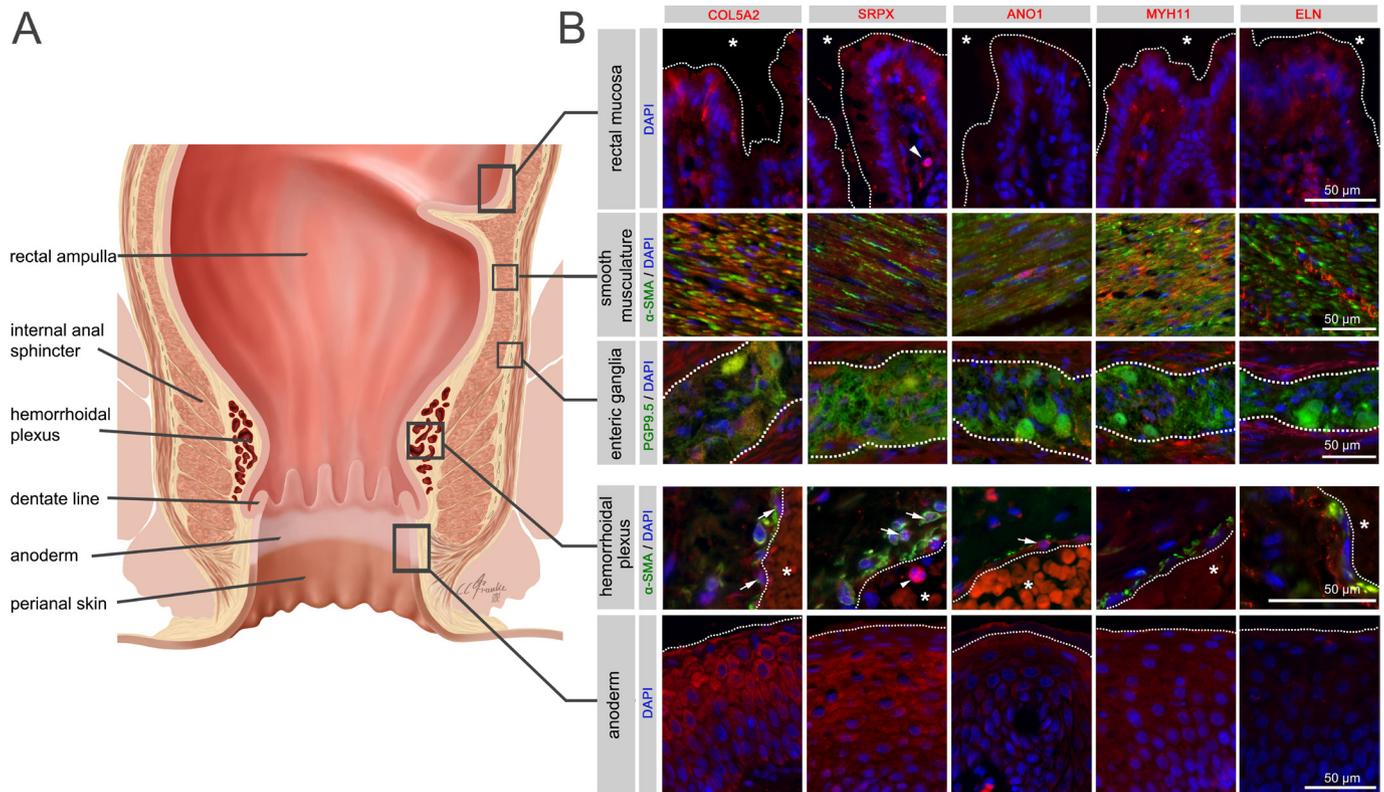
### DISCUSSION

Given the lack of large and systematic epidemiological and molecular studies, and despite its worldwide distribution, HEM still can be regarded as an understudied disease. In this study, we report the largest and most detailed genome-wide analysis of HEM, implemented via a combination of classical GWAS approaches and the use of minimal phenotyping, as recently shown to be effective in boosting sample size for increased statistical power.<sup>28</sup> We demonstrate for the first time that HEM is a partly inherited condition with a weak but detectable heritability estimated at 5% based on SNP data. We identify 102 independent risk loci, which were functionally annotated based on computational predictions and gene expression analysis of diseased and normal tissue. These loci alone explain approximately 0.9% of HEM heritability, which is of similar magnitude with respect to the genetic contribution of other common complex traits.<sup>29</sup> They provide important novel pathophysiological insight, which we discuss below in relation to individual pathways and mechanisms proposed to contribute to HEM aetiology.

### ECM, elasticity of the connective tissue and smooth muscle function

The non-vascular components of the anal cushions consist of the transitional epithelium, connective tissue (elastic and collagenous) and the submucosal anal muscle (muscle of Treitz).<sup>5</sup> Treitz's muscle tightly maintains the anal cushions in their normal position, and its deterioration is considered one of the most important pathogenetic factors in the formation of enlarged and prolapsed haemorrhoids (online supplemental figure 2). Anal cushion fixation is further facilitated by elastic and collagenous connective tissue, whose degeneration, due, for instance, to abnormalities in collagen composition, has been involved in HEM aetiology,<sup>30</sup> although without strong molecular evidence.

The coexpression module M1 identified from HEM tissue is linked to ECM organisation and muscle function and is enriched for HEM gene thus providing novel important evidence for a role of these two interconnected processes in HEM pathogenesis. *ELN* (lead SNP rs11770437) is one of the HEM prioritised genes, and also a main hub gene for the M1 module. *ELN* codes for the elastin protein, a key component of elastic fibres that comprise part of the ECM and confer elasticity to organs and tissues including blood vessels. Mutations in the *ELN* gene have been shown to cause *cutis laxa*, a disease in which dysfunctional elastin interferes with the formation of elastic fibres, thus weakening connective tissue in the skin and blood vessels. Of note, *ELN* has been recently implicated also in diverticular disease,<sup>31</sup> a condition characterised by outpouchings of the colonic wall at sites of relative weakness and/or defective elasticity of the connective and muscle layers, as well as in the common skin condition non-syndromic *striae distensae* (NSD, also known as stretch marks), whose manifestation is due to lost tissue elasticity



**Figure 6** Immunohistochemistry for selected haemorrhoidal disease (HEM) candidate proteins. Illustration of the rectum and anal canal (A) indicating the site-specific localisation of the immunohistochemical panels analysed in (B). Results of fluorescence immunohistochemistry are shown for selected HEM candidate proteins encoded by HEM genes *COL5A2* (rs16831319), *SRPX* (rs35318931), *ANO1* (rs2186797), *MYH11* (rs6498573) and *ELN* (rs11770437) (see also online supplemental table 11 and online supplemental figure 13). Antibody staining was performed on FFPE colorectal tissue specimens from control individuals. Picture layers correspond to the rectal mucosa (top row, epithelial surface delimited by a white dotted line, \*=intestinal lumen), smooth musculature (second row), enteric ganglia (third row, ganglionic boundaries delimited by a white dotted line), haemorrhoidal plexus (fourth row, endothelial surface delimited by a white dotted line, \*=vascular lumen) and the anoderm (bottom row, surface of the anoderm delimited by a white dotted line). Blue: DAPI; green:  $\alpha$ -SMA (anti-alpha smooth muscle actin antibody) for rows 2 and 4 (smooth musculature/haemorrhoidal plexus) and PGP9.5 (member of the ubiquitin hydrolase family of proteins, neuronal marker) for row 3 (enteric ganglia); red: antibody for the respective candidate protein. Arrows point to respective candidate-positive cells within the vascular wall. Arrowheads point to respective candidate-positive nucleated immune cells.

at affected skin sites.<sup>32</sup> Hence, similar mechanisms may underly HEM risk due to genetic variation in *ELN* and, notably, also the *SRPX* (lead SNP rs35318931) and *COL5A2* (lead SNP rs16831319) genes. The *SRPX* lead SNP is also strongly associated with NSD, and is a coding variant potentially impacting the function of an ECM protein. *COL5A2* codes for type V fibril-forming collagen that has regulatory roles during development and growth of type I collagen-positive tissues. Mutations in this gene are known to cause Ehlers-Danlos syndrome, a rare connective tissue disease that affects the skin, joints and blood vessels, and for which a link to HEM has already been postulated.<sup>33</sup> An additional hub gene for the coexpression module M1, *MYH11* (lead SNP rs6498573) encodes a smooth muscle myosin protein that is important for muscle contraction and relaxation, and whose dysfunction has been linked to vascular diseases<sup>34</sup> and GI dysmotility.<sup>35</sup> Functional studies in smooth muscle cells showed that overexpression of *MYH11* led to a paradoxical decrease of protein levels through increased autophagic degradation followed by disruption of contractile signalling.<sup>36</sup> Our transcriptome analysis also showed *MYH11* RNA upregulation in HEM tissue, thus suggesting possible alterations in smooth muscle action that may be relevant, for instance, to the Treitz's muscle function(s). Additional evidence for the involvement of the musculoskeletal system may come from the associations detected

for *GSDMC* (lead SNP rs10956488), an uncharacterised gene also shown to be relevant to lumbar disc herniation and back pain<sup>37</sup> and *PRDM6*, a histone methyltransferase that acts as a transcriptional repressor of smooth muscle gene expression. Finally, besides individual association signals, we detected significant genetic correlation with several other complex diseases of shared aetiology (hernia, dorsalgia), for which connective tissue and/or muscle alterations are described.

### Gut motility

Several lines of evidence link gut motility to the pathophysiology of HEM in this study. In our UKBB analyses, patients with HEM were found to suffer more often from IBS and other dysmotility syndromes than controls, as evidenced also by the increased use of medications including laxatives. These conditions also showed strongest genetic correlation with HEM among all tested traits, indicating similar genetic architecture and predisposing mechanisms. Moreover, given the important role of the gut-brain axis in IBS and other FGIDs,<sup>38</sup> it is possible that the correlations observed for anxiety, depression and other neuroaffective traits may mediate genetic risk effects at least in part via similar mechanisms also involving gut motility. Constipation and prolonged sitting and straining during defecation are associated

with delayed GI transit and reduced peristalsis and are among the proposed HEM risk factors (online supplemental figure 2). Harder stools, due for instance to infrequent defecations, can cause difficulty in bowel emptying and therefore increase pressure and mechanical friction on the haemorrhoidal cushions, leading to excessive engorgement and stretching or tissue damage. The relationship between HEM and gut motility is probably best evidenced by the association signal detected at the *ANO1* locus: its lead SNP rs2186797 corresponds to the missense variant *ANO1*:p.Phe608Ser (F608S) that was fine-mapped with very high confidence and shown to impact anoctamin-1 function. Anoctamin-1 is an ion channel expressed in the ICCs, the pacemakers of the GI tract controlling intestinal peristalsis, has already been implicated in IBS,<sup>39</sup> and is also expressed in the vasculature. Hence, it represents an ideal candidate to also affect HEM risk via genotype-driven modulation of ICC function. An additional interesting candidate is *ACHE*, which shows an eQTL for the fine-mapped lead SNP rs4556017: *ACHE* codes for an enzyme that hydrolyses the neurotransmitter acetylcholine at neuromuscular junctions and is overexpressed in Hirschsprung's disease, a condition in which gut motility is compromised due to the absence of nerve cells (aganglionosis) in the distal or entire segments of the large bowel.<sup>40</sup> Of interest, expression in the enteric ganglia next to the haemorrhoidal plexus was observed for several proteins encoded by HEM genes in our immunofluorescence experiments.

### Vasculature and circulatory system

Previous observations showed that HEM is not varicosities and accelerated blood flow velocities were observed in afferent vessels of patients with HEM.<sup>41</sup> An impaired drainage or filling of the anal cushion may contribute to cushion slippage and may thus be considered as one of many disease-causing factors, as already previously proposed.<sup>41</sup> Our genetic data support the involvement of the vasculature as an important player in HEM pathophysiology. TSEA and GSEA results downstream of HEM GWAS meta-analysis highlight blood vessels and artery morphogenesis among the HEM gene-enriched tissues and GO pathways, respectively. At the same time, moderate genetic correlation is detected for diseases of the circulatory system in the LDSC analyses. We identified a very strong association signal in correspondence of the ABO locus on chromosome 9, which determines the corresponding ABO blood group type (A, B, AB and O). In addition to red blood cells, ABO antigens are expressed on the surface of many cells and tissues, and have been strongly associated with coronary artery disease, thrombosis, haemorrhage,<sup>42</sup> GI bleeding<sup>43</sup> and other conditions related to the circulatory system. Interestingly, increased risk for coronary artery disease has been reported in patients with HEM in at least some studies<sup>44</sup> and replicated in our UKBB cross-disease analyses although amidst other hundreds of associations of similar magnitude of effects. We imputed ABO blood types from genotype data in UKBB, and detected increased HEM risk for carriers of the O type. O type has been reported to be protective for coronary artery disease in UKBB, although also predisposing to hypertension.<sup>45</sup> Hence, the potential mechanism(s) by which variation at this locus impacts HEM risk remain elusive at this stage. Blood antigens are nevertheless likely relevant, as other genes involved in the determination of specific blood groups are also among the 102 HEM GWAS hits (Kell blood group locus *XKR9*, lead SNP rs1838392).

In summary, our data provide important new insight into currently lacking evidence<sup>46</sup> about HEM pathogenesis (online

supplemental figure 14). This sets the stage for more detailed genetic and mechanistic follow-up analyses, the search for therapeutically actionable genes and pathways, and the eventual exploitation for the adoption of preventive measures based on computed individual predisposition.

### Author affiliations

- <sup>1</sup>School of Biological Sciences, Monash University, Clayton, Victoria, Australia
- <sup>2</sup>Unit of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden
- <sup>3</sup>Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
- <sup>4</sup>Novo Nordisk Foundation Center for Protein Research, Disease Systems Biology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
- <sup>5</sup>Institute of Biotechnology, Life Science Centre, Vilnius University, Vilnius, Lithuania
- <sup>6</sup>Institute of Anatomy, Christian-Albrechts-University of Kiel, Kiel, Germany
- <sup>7</sup>Department for General, Visceral, Vascular and Transplantation Surgery, University Medical Center Rostock, Rostock, Germany
- <sup>8</sup>Institute of Genomics, University of Tartu, Tartu, Estonia
- <sup>9</sup>K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway
- <sup>10</sup>Enteric NeuroScience Program, Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA
- <sup>11</sup>Institute of Advanced Research, Tokyo Medical and Dental University, Tokyo, Japan
- <sup>12</sup>Department of General-, Visceral- Transplant-, Thoracic and Pediatric Surgery, Kiel University, Kiel, Germany
- <sup>13</sup>Medical Office for Surgery Preetz, Preetz, Germany
- <sup>14</sup>Medical Department 1, University Hospital Dresden, Technische Universität Dresden (TU Dresden), Dresden, Germany
- <sup>15</sup>University Hospital Schleswig-Holstein, Institute of Clinical Chemistry, Thrombosis & Hemostasis Treatment Center, Campus Kiel & Lübeck, Kiel, Germany
- <sup>16</sup>Department of Internal Medicine, Hospital Oberndorf, Teaching Hospital of the Paracelsus Private Medical University of Salzburg, Oberndorf, Austria
- <sup>17</sup>Medical office for Colo-Proctology Kiel, Kiel, Germany
- <sup>18</sup>Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark
- <sup>19</sup>Department of Internal Medicine III, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany
- <sup>20</sup>Department of Internal Medicine III, University Hospital Heidelberg, Heidelberg, Germany
- <sup>21</sup>DZHK (German Centre for Cardiovascular Research), Partner Site Hamburg/Kiel/Lübeck, Kiel, Germany
- <sup>22</sup>Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
- <sup>23</sup>Epidemiology of highly pathogenic microorganisms, Robert Koch-Institute, Berlin, Germany
- <sup>24</sup>Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
- <sup>25</sup>Department of Medicine I, Institute of Cancer Research, Medical University Vienna, Vienna, Austria
- <sup>26</sup>Center for Regenerative Therapies Dresden (CRTD), Technische Universität (TU) Dresden, Dresden, Germany
- <sup>27</sup>Department of Proctological Surgery Park Klinik Kiel, Kiel, Germany
- <sup>28</sup>Proctological Office Kiel, Kiel, Germany
- <sup>29</sup>Research Institute for Internal Medicine, Division of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo, Oslo, Norway
- <sup>30</sup>Faculty of Medicine, University of Cologne, Cologne, Germany
- <sup>31</sup>Department of Gastroenterology, Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas, Lithuania
- <sup>32</sup>University of Lübeck, Lübeck, Germany
- <sup>33</sup>University Hospital of Schleswig-Holstein (UKSH), Kiel Campus, Kiel, Germany
- <sup>34</sup>Department of Pathology, University Medical Center Schleswig-Holstein, Kiel, Germany
- <sup>35</sup>Division of Endocrinology, Diabetes and Clinical Nutrition, Department of Medicine 1, University of Kiel, Kiel, Germany
- <sup>36</sup>Institute of Epidemiology, Christian-Albrechts-University of Kiel, Kiel, Germany
- <sup>37</sup>University of Athens, Athens Euroclinic, Athens, Greece
- <sup>38</sup>Pathology Unit, German Primate Center, Leibniz Institute for Primatology, Göttingen, Germany
- <sup>39</sup>Department of Ophthalmology, Hospital Frankfurt Hoechst, Frankfurt, Germany
- <sup>40</sup>Upper Gastrointestinal Surgery, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
- <sup>41</sup>Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway
- <sup>42</sup>Department of Clinical Immunology, Naestved Hospital, Naestved, Denmark
- <sup>43</sup>General Hospital Lüneburg, Lüneburg, Germany

<sup>44</sup>Department of General and Thoracic Surgery, Hospital Osnabrück, Osnabrück, Germany

<sup>45</sup>Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway

<sup>46</sup>BioCore – Bioinformatics Core Facility, Norwegian University of Science and Technology, Trondheim, Norway

<sup>47</sup>Clinic of Laboratory Medicine, St.Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

<sup>48</sup>Department of Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark

<sup>49</sup>Department of Surgery, Community Hospital Kiel, Kiel, Germany

<sup>50</sup>Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

<sup>51</sup>Department Research & Conservation, Zoo Nuremberg, Nuremberg, Germany

<sup>52</sup>Department of Gastroenterology, Helios Hospital Weißeritztal, Freital, Germany

<sup>53</sup>23andMe Inc, Sunnyvale, CA, USA

<sup>54</sup>Christian-Albrechts-University of Kiel, Kiel, Germany

<sup>55</sup>Gastrointestinal Genetics Lab, CIC bioGUNE - BRTA, Derio, Spain

<sup>56</sup>Ikerbasque, Basque Foundation for Science, Bilbao, Spain

**Twitter** Malte Christoph Rühlemann @mruehlemann, Arthur Beyder @beyderlab and Mauro D'Amato @damato\_mauro

**Contributors** AF, CS and MD'A were responsible for the concept and the design of the study. CS and GB coordinated the recruitment of the German clinical cohort. VK, JG, IV, BS, HGP, AH, JJ, JW, TB, MD, FB, SH, JK, JS, WK, MZo, CGN, TL, JT and WS recruited and phenotyped patients of the clinical cohorts. VK collected and phenotyped the biopsy samples that were used in the immunohistochemistry and transcriptome analyses. Animal samples for histology were taken and provided by KP, SZ, LF, TG, FHL and KM-R. Immunohistochemistry and histology analyses were performed by FC and TW with support of JW and CR. TZ, DE and FD performed genotype and phenotype data collection. TZ, DE, and FD performed GWAS data quality control and imputation. SJ performed the transcriptome (RNA-Seq) analysis with help from GI and PR. FU-W, SBu, and MF helped with bioinformatic analyses. IFJ, KB, SBr, CLR, CE, OBP, and HU conducted analyses in DBDS and DNPR. HC, VL, RJ, NF, WL, ML, and UN-G contributed German control data. SBu, CD, HV, AG and JH contributed the GWAS data on diverticular disease. Norwegian HUNT data were contributed by AHS, THK, MEG, LT, EN-J and KH. The Estonian dataset was provided by TE and MT-L. Protein models and analyses were performed by GM. The analysis of the MGI data set was performed by MZa, BV, AP and LGF. The 23andMe data set was analyzed and provided by OS, ESN, and VV. MW implemented ABO blood-group inference. SC, PRS, GF and AB performed site-directed ANO1 mutagenesis and whole-cell electrophysiology experiments. TZ, DE implemented statistical models and performed the (meta-)analysis. TZ, DE, and SJ curated and interpreted results. MS designed online supplementary figure S2 with scientific input from NM and AF. The GWAS data browser was implemented by MR and set up by GH-S. AF, MD'A and DE wrote the manuscript draft with substantial contributions from TZ and SJ. All authors reviewed, edited and approved the final manuscript.

**Funding** This project was funded by Andre Franke's and Clemens Schafmayer's DFG grant "Discovery of risk factors for hemorrhoids" (ID: FR 2821/19-1). The study received infrastructure support from the DFG Cluster of Excellence 2167 "Precision Medicine in Chronic Inflammation (PMI)" (DFG Grant: "EXC2167"). The project was supported by grants from the Swedish Research Council to MD (VR 2017-02403), the Novo Nordisk Foundation (grants NNF17OC0027594 and NNF14CC0001) and BigTempHealth (grant 5153-00002B). We are indebted to the valuable assistance by Tanja Wesse (Genotyping of the German samples) and Petra Röthgen (German DZHK control data set). We thank Clemens Franke (Institute of Anatomy, Christian-Albrechts-University of Kiel, Kiel, Germany) for graphic assistance (figure 6, online supplementary figure S1 and S13). This research has been conducted using the UK Biobank Resource under Application Number 31435. The work on cross-trait analysis for diverticular disease presented in this manuscript was supported by the German Research Council (DFG, ID: Ha3091/9-1) and the Austrian Science Fund (FWF, ID: I1542-B13). Data access to the UK biobank data for diverticular disease was granted under project numbers 22691. EGCUT work has also been supported by the European Regional Development Fund and grants SP1G118045T, No. 2014-2020.4.01.15-0012 GENTRANMED and 2014-2020.4.01.16-0125 This study was also funded by EU H2020 grant 692145, Estonian Research Council Grant PUT1660. Data analyzes with Estonian datasets were carried out in part in the High-Performance Computing Center of University of Tartu. The work on site-directed mutagenesis of ANO1 and whole-cell electrophysiology was funded by NIH DK057061. We would like to thank the research participants and employees of 23andMe for making this work possible. The following members of the 23andMe Research Team contributed to this study: Michelle Agee, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, and Vladimir Vacić.

**Competing interests** Vladimir Vacić and Olga V. Sazonova are/were employed by and hold stock or stock options in 23andMe, Inc. All other authors have nothing to declare.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository. Genome-wide summary statistics of our analyses are publicly available through our web browser (<http://hemorrhoids.online>) and have been submitted to the European Bioinformatics Institute ([www.ebi.ac.uk/gwas](http://www.ebi.ac.uk/gwas)) under the accession number GCST90014033. RNA-seq data have been deposited at NCBI Gene Expression Omnibus (GEO) under the accession number GSE154650.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iDs

Tenghao Zheng <http://orcid.org/0000-0003-1587-7481>

David Ellinghaus <http://orcid.org/0000-0002-4332-6110>

Simonas Juzenas <http://orcid.org/0000-0001-9293-0691>

Christian Datz <http://orcid.org/0000-0001-7838-4532>

Michael Forster <http://orcid.org/0000-0001-9927-5124>

Andrea Gsur <http://orcid.org/0000-0002-9795-1528>

Jochen Hampe <http://orcid.org/0000-0002-2421-6127>

Christoph Röcken <http://orcid.org/0000-0002-6989-8002>

Eivind Ness-Jensen <http://orcid.org/0000-0001-6005-0729>

Malte Christoph Rühlemann <http://orcid.org/0000-0002-0685-0052>

Gianrico Farrugia <http://orcid.org/0000-0003-3473-5235>

Mauro D'Amato <http://orcid.org/0000-0003-2743-5197>

Andre Franke <http://orcid.org/0000-0003-1530-5811>

#### REFERENCES

- Jacobs D. Clinical practice. *Hemorrhoids*. *N Engl J Med* 2014;371:944–51.
- Duthie HL, Cairns FW. Sensory nerve-endings and sensation in the anal region of man. *Br J Surg* 1960;47:585–95.
- Haas PA, Haas GP, Schmaltz S, Fox TA, Jr. The prevalence of hemorrhoids. *Dis Colon Rectum* 1983;26:435–9.
- Yang JY, Peery AF, Lund JL, et al. Burden and cost of outpatient hemorrhoids in the United States Employer-Insured population, 2014. *Am J Gastroenterol* 2019;114:798–803.
- Margetis N. Pathophysiology of internal hemorrhoids. *Ann Gastroenterol* 2019;32:264–72.
- Tung JY, Do CB, Hinds DA, et al. Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS One* 2011;6:e23473.
- Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–9.
- Leitsalu L, Haller T, Esko T, et al. Cohort profile: Estonian Biobank of the Estonian genome center, University of Tartu. *Int J Epidemiol* 2015;44:1137–47.
- Fritsche LG, Gruber SB, Wu Z, et al. Association of polygenic risk scores for multiple cancers in a Phenome-wide study: results from the Michigan genomics initiative. *Am J Hum Genet* 2018;102:1048–61.
- Kvale MN, Hesselton S, Hoffmann TJ, et al. Genotyping informatics and quality control for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics* 2015;200:1051–60.
- Willer CJ, Li Y, Abecasis GR. Metal: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–1.
- Watanabe K, Taskesen E, van Bochoven A, et al. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017;8:8.
- Benner C, Spencer CCA, Havulinna AS, et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 2016;32:1493–501.
- Watanabe K, Stringer S, Frei O, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 2019;51:1339–48.

- 15 Buniello A, MacArthur JAL, Cerezo M, *et al.* The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47:D1005–12.
- 16 Kamat MA, Blackshaw JA, Young R, *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* 2019;35:4851–3.
- 17 Cuéllar-Partida G, Lundberg M, Kho PF, *et al.* Complex-Traits genetics virtual lab: a community-driven web platform for post-GWAS analyses. *bioRxiv* 2019;518027.
- 18 Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience* 2019;8. doi:10.1093/gigascience/giz082. [Epub ahead of print: 01 07 2019].
- 19 Krokstad S, Langhammer A, Hveem K, *et al.* Cohort profile: the HUNT study, Norway. *Int J Epidemiol* 2013;42:968–77.
- 20 Hansen TF, Banasik K, Erikstrup C, *et al.* Dbds genomic cohort, a prospective and comprehensive resource for integrative and temporal analysis of genetic, environmental and lifestyle factors affecting health of blood donors. *BMI Open* 2019;9:e028401.
- 21 Krawczak M, Nikolaus S, von Eberstein H, *et al.* PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 2006;9:55–61.
- 22 Pers TH, Karjalainen JM, Chan Y, *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* 2015;6:5890.
- 23 de Leeuw CA, Mooij JM, Heskes T, *et al.* MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015;11:e1004219.
- 24 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- 25 Malysz J, Gibbons SJ, Saravanaperumal SA, *et al.* Conditional genetic deletion of *Ano1* in interstitial cells of Cajal impairs  $Ca^{2+}$  transients and slow waves in adult mouse small intestine. *Am J Physiol Gastrointest Liver Physiol* 2017;312:G228–45.
- 26 Naba A, Clauser KR, Whittaker CA, *et al.* Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. *BMC Cancer* 2014;14:518.
- 27 Bartels CF, Zelinski T, Lockridge O. Mutation at codon 322 in the human acetylcholinesterase (ACHE) gene accounts for YT blood group polymorphism. *Am J Hum Genet* 1993;52:928–36.
- 28 DeBoever C, Tanigawa Y, Aguirre M, *et al.* Assessing digital phenotyping to enhance genetic studies of human diseases. *Am J Hum Genet* 2020;106:611–22.
- 29 Okada Y, Wu D, Trynka G, *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;506:376–81.
- 30 Nasserri YY, Krott E, Van Groningen KM, *et al.* Abnormalities in collagen composition may contribute to the pathogenesis of hemorrhoids: morphometric analysis. *Tech Coloproctol* 2015;19:83–7.
- 31 Schafmayer C, Harrison JW, Buch S, *et al.* Genome-Wide association analysis of diverticular disease points towards neuromuscular, connective tissue and epithelial pathomechanisms. *Gut* 2019;68:854–65.
- 32 Tung JY, Kiefer AK, Mullins M, *et al.* Genome-Wide association analysis implicates elastic microfibrils in the development of nonsyndromic striae distensae. *J Invest Dermatol* 2013;133:2628–31.
- 33 Plackett TP, Kwon E, Gagliano RA, *et al.* Ehlers-Danlos syndrome-hypermobility type and hemorrhoids. *Case Rep Surg* 2014;2014:1–3.
- 34 Zhu L, Vranckx R, Khau Van Kien P, *et al.* Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. *Nat Genet* 2006;38:343–9.
- 35 Gilbert MA, Schultz-Rogers L, Rajagopalan R, *et al.* Protein-elongating mutations in MYH11 are implicated in a dominantly inherited smooth muscle dysmotility syndrome with severe esophageal, gastric, and intestinal disease. *Hum Mutat* 2020;41:973–82.
- 36 Kwartler CS, Chen J, Thakur D, *et al.* Overexpression of smooth muscle myosin heavy chain leads to activation of the unfolded protein response and autophagic turnover of thick filament-associated proteins in vascular smooth muscle cells. *J Biol Chem* 2014;289:14075–88.
- 37 Suri P, Palmer MR, Tsepilov YA, *et al.* Genome-Wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. *PLoS Genet* 2018;14:e1007601.
- 38 Holtmann G, Shah A, Morrison M. Pathophysiology of functional gastrointestinal disorders: a holistic overview. *Dig Dis* 2017;35 Suppl 1:5–13.
- 39 Mazzone A, Gibbons SJ, Eisenman ST, *et al.* Direct repression of anoctamin 1 (*ANO1*) gene transcription by Gli proteins. *FASEB J* 2019;33:6632–42.
- 40 Moore SW, Johnson G. Acetylcholinesterase in Hirschsprung's disease. *Pediatr Surg Int* 2005;21:255–63.
- 41 Aigner F, Gruber H, Conrad F, *et al.* Revised morphology and hemodynamics of the anorectal vascular plexus: impact on the course of hemorrhoidal disease. *Int J Colorectal Dis* 2009;24:105–13.
- 42 Franchini M, Lippi G. Relative risks of thrombosis and bleeding in different ABO blood groups. *Semin Thromb Hemost* 2016;42:112–7.
- 43 Huang ES, Khalili H, Strate LL, *et al.* Su1786 ABO blood group and the risk of gastrointestinal bleeding. *Gastroenterology* 2012;142:S-503.
- 44 Chang S-S, Sung F-C, Lin C-L, *et al.* Association between hemorrhoid and risk of coronary heart disease: a nationwide population-based cohort study. *Medicine* 2017;96:e7662.
- 45 Groot HE, Villegas Sierra LE, Said MA, *et al.* Genetically determined ABO blood group and its associations with health and disease. *Arterioscler Thromb Vasc Biol* 2020;40:830–8.
- 46 Sandler RS, Peery AF. Rethinking what we know about hemorrhoids. *Clin Gastroenterol Hepatol* 2019;17:8–15.