OPEN ACCESS

University of BRISTOL

Bonfiglio, F., Liu, X., Smillie, C., Pandit, A., Timpson, N. J., Kurilshikov, A., Bacigalupe, R., Zheng, T., Nim, H., Garcia-Etzebarria, K., Bujanda, L., Andreasson, A., Agreus, L., Walter, S., Abecasis, G. R., Eijsbouts, C., Jostins, L., Parkes, M., Hughes, D. A., ... D'Amato, M. (2021). GWAS of stool frequency provides insights into gastrointestinal motility and irritable bowel syndrome. *Cell Genomics*, *1*(3), [100069]. https://doi.org/10.1016/j.xgen.2021.100069

Publisher's PDF, also known as Version of record

License (if available):
CC BY-NC-ND

Link to published version (if available):
10.1016/j.xgen.2021.100069

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). It first appeared online via Elsevier at https://doi.org/10.1016/j.xgen.2021.100069.Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

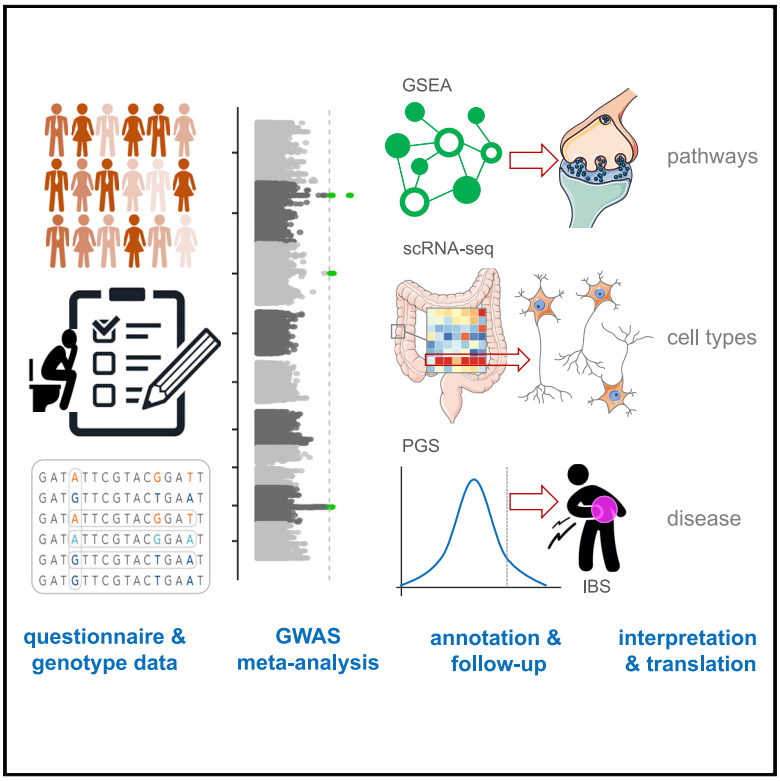
This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

Cell Genomics

Short article

GWAS of stool frequency provides insights into gastrointestinal motility and irritable bowel syndrome

Graphical abstract



Highlights

- Genetics of gut motility via data on stool frequency in 167,875 individuals
- GWAS identifies 14 loci associated with stool frequency
- Candidate genes enriched in neurotransmitter signaling and enteric motor neurons
- Polygenic scores predict increased risk of irritable bowel syndrome

Authors

Ferdinando Bonfiglio, Xingrong Liu, Christopher Smillie, ..., Magnus Simren, Michael Camilleri, Mauro D'Amato

Correspondence

mdamato@cicbiogune.es

In brief

Bonfiglio et al. investigate the genetics of human gut motility using genotype and questionnaire data on stool frequency in 167,875 individuals from the UK Biobank and four other cohorts. They identify 14 loci associated with stool frequency as well as candidate genes enriched in neurotransmitter signaling and preferentially expressed in enteric neurons controlling peristalsis. The genetic architecture of stool frequency correlates with that of irritable bowel syndrome (IBS), and stool frequency polygenic scores were predictive of IBS risk.

Bonfiglio et al., 2021, Cell Genomics *1*, 100069
December 8, 2021 © 2021 The Author(s).
https://doi.org/10.1016/j.xgen.2021.100069

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Short article

GWAS of stool frequency provides insights into gastrointestinal motility and irritable bowel syndrome

Ferdinando Bonfiglio,[1,2] Xingrong Liu,[2,3,30] Christopher Smillie,[4,30] Anita Pandit,[5,30] Alexander Kurilshikov,[6,30] Rodrigo Bacigalupe,[7,8,30] Tenghao Zheng,[1,2,3] Hieu Nim,[1] Koldo Garcia-Etxebarria,[9] Luis Bujanda,[9,10,11] Anna Andreasson,[12] Lars Agreus,[13] Susanna Walter,[14] Gonçalo Abecasis,[5] Chris Eijsbouts,[15,16]

(Author list continued on next page)

[1]School of Biological Sciences, Monash University, Clayton, VIC, Australia
[2]Unit of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden
[3]Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden
[4]Klarman Cell Observatory, Broad Institute, Cambridge, MA, USA
[5]Department of Biostatistics, University of Michigan, School of Public Health, Ann Arbor, MI, USA
[6]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
[7]Department of Microbiology and Immunology, Rega Instituut, KU Leuven, Leuven, Belgium
[8]Center for Microbiology, VIB, Leuven 3000, Belgium
[9]Department of Gastrointestinal and Liver Diseases, Biodonostia HRI, San Sebastian, Spain
[10]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Madrid, Spain

(Affiliations continued on next page)

SUMMARY

Gut dysmotility is associated with constipation, diarrhea, and functional gastrointestinal disorders like irritable bowel syndrome (IBS), although its molecular underpinnings are poorly characterized. We studied stool frequency (defined by the number of bowel movements per day, based on questionnaire data) as a proxy for gut motility in a GWAS meta-analysis including 167,875 individuals from UK Biobank and four smaller population-based cohorts. We identify 14 loci associated with stool frequency ($p \leq 5.0 \times 10^{-8}$). Gene set and pathway analyses detected enrichment for genes involved in neurotransmitter/neuropeptide signaling and preferentially expressed in enteric motor neurons controlling peristalsis. PheWAS identified pleiotropic associations with dysmotility syndromes and the response to their pharmacological treatment. The genetic architecture of stool frequency correlates with that of IBS, and UK Biobank participants from the top 1% of stool frequency polygenic score distribution were associated with 5× higher risk of IBS with diarrhea. These findings pave the way for the identification of actionable pathological mechanisms in IBS and the dysmotility syndromes.

INTRODUCTION

Gastrointestinal (GI) motility is essential to digestion, nutrients absorption, and overall human health, including bi-directional host-microbiome interactions.[1,2] Gut dysmotility and altered peristalsis are observed in constipation, diarrhea, and common functional GI disorders like irritable bowel syndrome (IBS).[3,4] There is only incomplete understanding of the physiological mechanisms regulating intestinal motility and their perturbation in the dysmotility syndromes, and therapeutic options mostly rely on targeting specific symptoms rather than (currently unknown) underlying mechanisms. There is evidence of heritability for gut motility from previous studies in relation to colonic transit time measured with detectable tracers,[5–7] suggesting that genetic association studies may be useful in identifying biological pathways for therapeutic exploitation.

Direct assessment of GI motility in humans requires clinical procedures that are exclusively performed to support patient diagnosis and therapeutic management and are therefore not suitable for large-scale population-wide genetic surveys.[8] However, stool consistency and, to a lesser extent, stool frequency (defined by the number of bowel movements over a period of time) are valuable indicators of bowel function that correlate with colonic transit time.[9,10] These represent practical surrogate tools that can be adopted and scaled for studying GI motility at the population level, thanks to their ease of data collection via questionnaires and survey-based approaches.

Here, we leverage data from UK Biobank and four smaller population-based cohorts for a stool frequency genome-wide association study (GWAS) meta-analysis in a total of 167,875 individuals of European descent. We show modest but detectable heritability for this trait, identify 14 loci harboring genes

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Luke Jostins,[17,18] Miles Parkes,[19] David A. Hughes,[20,21] Nicholas Timpson,[20,21] Jeroen Raes,[7,8] Andre Franke,[22] Nicholas A. Kennedy,[23,31] Aviv Regev,[4,28,31] Alexandra Zhernakova,[6,31] Magnus Simren,[24,31] Michael Camilleri,[25,31] and Mauro D'Amato[1,2,9,26,27,29,*]

[11]Universidad del País Vasco (UPV/EHU), San Sebastian, Spain
[12]Division of Clinical Medicine, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden
[13]Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
[14]Division of Neuro and Inflammation Science, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden
[15]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK
[16]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK
[17]Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK
[18]Christ Church, University of Oxford, Oxford, UK
[19]Division of Gastroenterology, Department of Medicine, University of Cambridge, Cambridge, UK
[20]MRC Integrative Epidemiology Unit at University of Bristol, Bristol, UK
[21]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[22]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
[23]IBD Pharmacogenetics, College of Medicine and Health, University of Exeter, Exeter, UK
[24]Dept of Internal Medicine & Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[25]Clinical Enteric Neuroscience Translational and Epidemiological Research (CENTER) and Division of Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic, Rochester, MN, USA
[26]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[27]Gastrointestinal Genetics Lab, CIC bioGUNE - BRTA, Derio, Spain
[28]Present address: Genentech, South San Francisco, CA, USA
[29]Lead contact
[30]These authors contributed equally
[31]These authors contributed equally
*Correspondence: mdamato@cicbiogune.es
https://doi.org/10.1016/j.xgen.2021.100069

associated with pathways and cell types plausibly involved in the control of GI motility in humans, and provide compelling evidence of the relevance of these findings to IBS. The identification of genetic factors predisposing to altered gut motility may eventually allow early identification of individuals at higher risk of functional GI disorders and therapeutically actionable pathways that may be targeted for the delineation of alternative treatment options.

RESULTS

Stool frequency GWAS meta-analysis

We studied genotype and stool frequency questionnaire data in 167,875 individuals of European descent from five population-based cohorts. These include UK Biobank (UKBB; n = 163,616), LifeLines-DEEP (LLD; n = 942), the Genes for Good study (GFG; n = 1,069), the Flemish Gut Flora Project (FGFP; n = 2,001), and the Population-based Colonoscopy study (PopCol; n = 247) (Table S1; STAR Methods). Our GWAS meta-analysis for stool frequency identified 3,751 genome-wide-significant associations ($p \leq 5.0 \times 10^{-8}$) at 14 independent loci (Figures 1 and S1; Tables 1 and S2). The strongest signal was detected for marker rs12273363 on chromosome 11 ($p = 4.8 \times 10^{-21}$), in proximity to the brain-derived neurotrophic factor (*BDNF*) gene.

Gene set and pathway enrichment analyses

A classical gene set enrichment analysis[11] highlighted the REACTOME pathways "Class B/2 secretin family receptors" and "transmission across chemical synapses," with involvement of genes from multiple loci (Table S3). GeneNetwork[12] co-expression analysis identified significant enrichment for relevant KEGG pathways, including "olfactory transduction" and "neuroactive ligand receptor interaction" and Gene Ontology terms "detection of chemical stimulus involved in sensory perception," "olfactory receptor activity," and "neuropeptide signaling pathway" (Figure S2; Tables S4–S6). Similarly, PASCAL[13] pathway-level analysis returned "neurotransmitter receptor binding and downstream transmission in the postsynaptic cell" and "serotonin receptors" as the top-enriched REACTOME pathways (Table S7).

Functional annotation at the single-cell level

Using single-cell transcriptomic data available from human colonic mucosa and muscularis propria,[14,15] we inspected the expression of candidate genes at the associated loci in relation to immune, epithelial, stromal, and glial cells, muscle cells, and enteric neuron subtype (Figures 2A and S3; STAR Methods). They were strongly enriched in enteric neurons ($p = 1.1 \times 10^{-4}$) and more so in putative motor neurons (excitatory and/or inhibitory subtypes, $p = 7.3 \times 10^{-8}$) reportedly involved in the control of peristalsis,[15] with contribution from several loci (Figure 2B).

Prioritization of candidate genes

FINEMAP analysis of candidate causative SNPs from each locus (STAR Methods) mapped 5 out of 14 signals at single-marker resolution with >50% probability (Table 1). Variants rs4556017 and rs13162291 were mapped with highest confidence (respective probabilities 95.1% and 83.5%) and are both associated with expression quantitative trait loci (eQTLs) in multiple tissues (Figure S4). In particular, rs4556017 shows eQTLs for the acetylcholinesterase *ACHE*, and rs13162291 shows eQTLs for the fatty

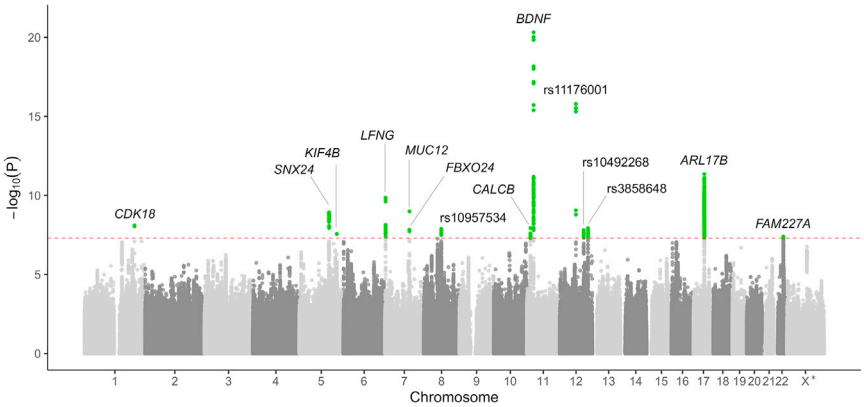


Figure 1. Manhattan plot of stool frequency GWAS meta-analysis results

GWAS association signals ($-\log_{10}$ P) are reported for SNP markers across all chromosomes shown in alternate gray colors. Significance level corresponding to the genome-wide significant ($p = 5.0 \times 10^{-8}$) threshold is indicated with a dashed red horizontal line. For each independent association signal, the nearest gene (within 100 kb, otherwise the lead SNP) is reported. Genome-wide significant markers are highlighted in green. *GWAS results for the X chromosome are only available for UKBB.

acid hydroxylase *FAXDC2* (Figure S5), two genes expressed in enteric and motor neurons (Figure 2A). The SNP rs11240503 is associated with a colon-specific eQTL for *CDK18* (Figure S5), a protein kinase expressed in colonic M cells and BEST4+ enterocytes (Figure 2A), while rs62482222 is associated with eQTLs for 21 genes in multiple tissues (Table S2; Figure S4). Finally, the rs12273363 marker is associated with eQTLs for a long non-coding antisense RNA (*BDNF-AS*, Figure S5) modulating the expression of BDNF.[16] Of note, rs12273363 emerged as top GWAS signal in our meta-analysis ($p = 4.8 \times 10^{-21}$; Table 1) and also showed consistent genetic effects in pilot follow-up analyses of stool consistency and colonic transit time (Tables S8 and S9): in these, the T allele linked to more frequent stools was significantly associated with (1) softer/looser stools in 2,338 individuals from GFG, LLD, and PopCol and (2) faster transit in a small cohort of 160 IBS patients. Other loci harbored candidate genes of known relevance to GI motility and dysmotility syndromes, like neuropeptides/neurotransmitters and their receptors (*CALCA/CALCB* and *CRHR1*), ion channels (*KCNJ4*), tight junction proteins (*CLDN15*), and others (Table S2).

Correlations with other diseases and traits

Phenome-wide association study (PheWAS) of the 14 stool frequency-associated signals in publicly available GWAS data (using lead SNPs and their linkage disequilibrium [LD] proxies; STAR Methods) suggested some of these loci to be associated also to other traits and diseases across multiple domains (Figure S6A). The loci tagged by markers rs12273363 and rs2732706 showed the largest number of associations, mostly with anthropometric and psychiatric traits (Table S10). Because of the known effects of gut microbiota composition on GI motility and vice versa,[2] we also tested the 14 loci for association in recent GWAS data from the MiBioGen consortium (same PheWAS protocol; STAR Methods), in relation to host genetic variation influencing the composition of the human fecal microbiome.[17] Inspection of GWAS data for 211 different taxa (STAR Methods) identified the *BDNF* locus to be associated also with fecal abundance (reduced for the rs12273363 T allele) of the genus Ruminococcaceae UCG-005 (Figure S7).

LD score regression (LDSC) analysis estimated SNP-based stool frequency heritability around 7% (SNP heritability, h^2_{SNP} = 0.073) and identified strong genetic correlation with IBS ($r_g = 0.42$, $p = 5.1 \times 10^{-5}$), while additional significant findings were obtained for other GI (diverticular disease, use of proton pump inhibitors) and psychiatric (anxiety, depression) traits (Figure S6B; Table S11). Medications did not attenuate genetic effects or the strength of the association signals (Table S12).

Sex-specific analyses

Three GWAS signals were detected upon stratification into male and female groups, respectively, on chromosome 20 for man (rs6123818) and chromosomes 6 (rs1268068) and 19 for women (rs10407548) (Table S13; Figure S8). The locus on chromosome 19 (Figure S9) harbors genes coding for free fatty acid receptors (*FFAR1* and *FFRA3*) that are known to bind short-chain fatty acids whose luminal concentration can affect gut motility (i.e., butyrate).[18] The GWAS signal was fine-mapped with high confidence (probability 98.9%) to the lead SNP rs10407548, linked to *FFAR3* gene expression via eQTLs detected in multiple tissues for its LD proxies rs756904, rs756905, and rs756906 (based on data from FUMA; STAR Methods). Medications did not attenuate genetic effects or the strength of sex-specific association signals (Table S14).

Stool frequency polygenic scores and IBS

We further explored the relevance of stool frequency GWAS findings to IBS by computing polygenic scores (PGSs) based on stool frequency GWAS summary statistics obtained in this study. This was done using a two-step testing and validation approach on two independent, non-overlapping subsets of UKBB data: participants with Digestive Health Questionnaire data available (DHQ+ individuals, n = 163,616, the same cohort used for GWAS analyses based on stool frequency data from the DHQ survey) and those without DHQ data (DHQ− individuals; n = 291,496) (Table S15; STAR Methods). Based on Rome III criteria for IBS from DHQ data (in DHQ+ participants), the distribution of PGSs was significantly different in IBS versus asymptomatic individuals, most pronouncedly for the IBS diarrhea subtype (IBS-D; $p < 1 \times 10^{-300}$) (Figure 3). Risk of IBS-D markedly increased toward the upper tail of PGS distribution (odds ratio [OR] = 5.5; $p = 1.3 \times 10^{-155}$ for the top 1% and OR = 4.3; $p < 1 \times 10^{-300}$ for the top 5% of the distribution; Figure 3; Table S16). Although the diarrhea subtype could not be tested (IBS Rome III data not available for DHQ− individuals), PGSs were validated in the

Table 1. Stool frequency GWAS meta-analysis and fine mapping results

Chr	Lead SNP	Start–end (bp)	EA	OA	EAF	Beta (SE)[a]	p	p Het	Nearest gene (other genes)[b]	Most likely causal SNP (%probability)[d]
1	rs11240503	205473772–205485290	A	G	0.300	0.018 (0.003)	7.80E−09	0.66	*CDK18 (5)*	rs11240503 (0.588)[e]
5	rs39819	122069447–122402133	A	G	0.671	0.018 (0.003)	1.20E−09	0.87	*SNX24 (4)*	–
5	rs13162291	154369987–154447535	A	G	0.191	0.020 (0.004)	2.70E−08	0.08	*KIF4B (4)*	rs13162291 (0.835)[e]
7	rs12700026	2556512–2605424	A	C	0.890	−0.029 (0.005)	1.40E−10	0.38	*LFNG (4)*	rs12700026, rs12700027 (0.350)
7	rs62482222	100122391–100197866	A	G	0.118	0.023 (0.004)	1.40E−08	0.08	*FBXO24 (29)*[c]	rs62482222 (0.770)[e]
7	rs4556017	100618993–100632790	T	C	0.853	0.024 (0.004)	1.00E−09	0.30	*MUC12 (10)*[c]	rs4556017 (0.951)[e]
8	rs10957534	71502376–72012331	C	G	0.367	−0.016 (0.003)	1.30E−08	0.11	*(5)*	–
11	rs6486216	14999189–15120775	T	C	0.276	0.018 (0.003)	1.10E−08	0.53	*CALCB (4)*	–
11	rs12273363	27477864–27748493	T	C	0.795	0.032 (0.003)	4.80E−21	0.09	*BDNF (3)*	rs12273363 (0.525)[e]
12	rs11176001	66393756–66410673	A	C	0.132	0.034 (0.004)	1.60E−16	0.73	*(1)*	rs11176001 (0.392)
12	rs10492268	98344454–98385439	T	C	0.552	0.016 (0.003)	1.60E−08	0.81	–	rs10492268 (0.187)
12	rs3858648	115873190–115940482	A	C	0.508	−0.016 (0.003)	1.20E−08	0.60	–	rs3858648 (0.077)
17	rs2732706	43463493–44865603	T	C	0.221	0.024 (0.003)	4.40E−12	0.15	*ARL17B (109)*	–
22	rs5757162	38869463–39152412	T	C	0.286	0.017 (0.003)	4.00E−08	0.25	*FAM227A (15)*	–

Chr, chromosome; EA, effect allele; OA, other allele; EAF, effect allele frequency

[a]Positive beta = higher stool frequency

[b]Nearest gene (within 100 kb from lead SNP) and other genes in the region, based on FUMA positional and eQTL mapping

[c]These loci share four mapped genes

[d]Only causal SNPs identified with >5% probability are reported

[e]Causal SNPs identified with >50% probability

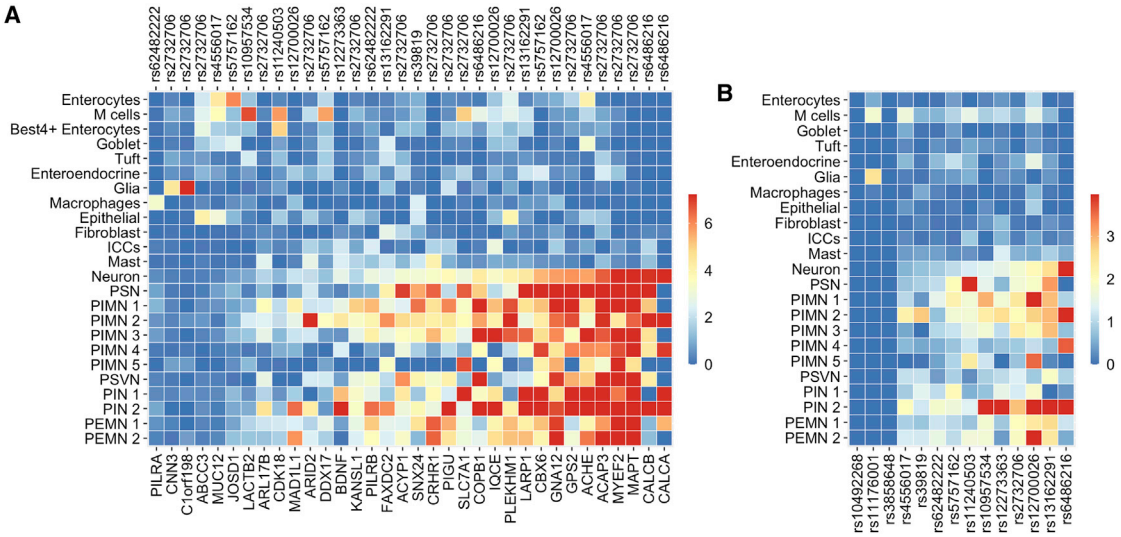


Figure 2. Heatmap of colonic stool frequency gene expression

(A) An illustrative selection of stool frequency genes (annotated with locus membership at the top) is reported for their expression in relevant cell types from colonic mucosa and colonic muscularis, ordered according to increasing expression.

(B) The expression of all stool frequency genes (see Figure S3 for an extended heatmap of the full dataset) is reported collapsed by locus (median values), demonstrating that multiple loci are enriched for neuronal expression.

The heatmaps display log2(TP10K+1) transformed data, and the expression of each gene is scaled across all cells and shown in color scale ranging from 0 to the 99th data quantile (to avoid high/low expressed genes dominating the heatmap). ICCs, interstitial cells of Cajal; PSN, putative sensory neuron; PEMN, putative excitatory motor neuron; PIMN, putative inhibitory motor neuron; PIN, putative interneuron; PSVN, secretomotor/vasodilator neuron. Cell types, neuron types, and subtypes are classified as defined previously in Drokhlyansky et al.[15]

independent DHQ− dataset in which the prevalence of IBS diagnoses (from medical records or self-reported) increased across PGS percentiles (OR = 1.3; p = 7.5 × 10^{-3}, for the top 1% of the distribution; Table S16), and their values were significantly higher in cases than in controls (p = 1.9 × 10^{-8}). These results suggest that stool frequency PGSs may contribute to the identification of individuals at increased risk of IBS, particularly IBS-D.

DISCUSSION

We report the results of a GWAS for stool frequency based on the meta-analysis of genetic and health-related data in 167,875 individuals of European descent. Our approach aimed at identifying relevant physiological pathways and mechanisms via indirectly measuring GI function based on suitable questionnaire data on bowel habits. A similar strategy was adopted in a previous study; however, no significant results were obtained, likely due to the small size of the cohorts analyzed (total n = 1,281 from LLD and PopCol, also included here after re-analysis using a GWAS pipeline common to all cohorts).[19] We report modest but detectable heritability for stool frequency, confirming previous evidence coming from studies of defecation in invertebrates (*C. elegans*) and experimental rat models.[20,21] At the same time, we identify genome-wide significant association at 14 independent loci, which harbor genes and variants implicating pathways, cell types, and mechanisms plausibly affecting human gut motility in health and disease.

GWAS-downstream analyses suggest an enrichment for candidate genes involved in neuropeptide and neurotransmitter signaling, sensory perception, and control of motor function in the gut, pathways notoriously central to the enteric nervous system.[22] Our cell-type-specific analyses show that stool frequency candidate genes are strongly enriched for their expression in enteric neurons, a specific pattern otherwise undetected at the whole-tissue level. This appears to be more pronounced in putative excitatory and inhibitory motor neurons that have been associated with peristalsis and mechanosensation of gut distention (PEMN and PIMN subtypes expressing the mechanosensitive ion channel *PIEZO2*).[15] Further investigation of such expression patterns may contribute important insight into the exact mechanisms underlying neurogenic motor control in the gut and eventually aid the development of future therapeutic strategies to modify GI function and motility.

Individual candidate genes that seem most likely to play an important role in the control of stool frequency also suggest the involvement of neuropeptide/neurotransmitter signaling pathways. This is best exemplified by the strongest GWAS association we detected for the *BDNF* locus on chromosome 11. The signal is mapped with relatively high confidence (>50% probability) to the rs12273363 marker, which is linked to multiple functional effects on *BDNF* expression: it has eQTL effects on an antisense transcript (*BDNF-AS*) that induces *BDNF* mRNA degradation[16] and lies in a regulatory region previously shown to impart allele-specific, direct repression of *BDNF* promoter activity (with rs12273363 T a less active repressor).[23] *BDNF* is a neurotrophin expressed in the central and peripheral nervous systems, with neurotransmitter modulatory properties and a crucial role in neuronal growth, differentiation, survival, and

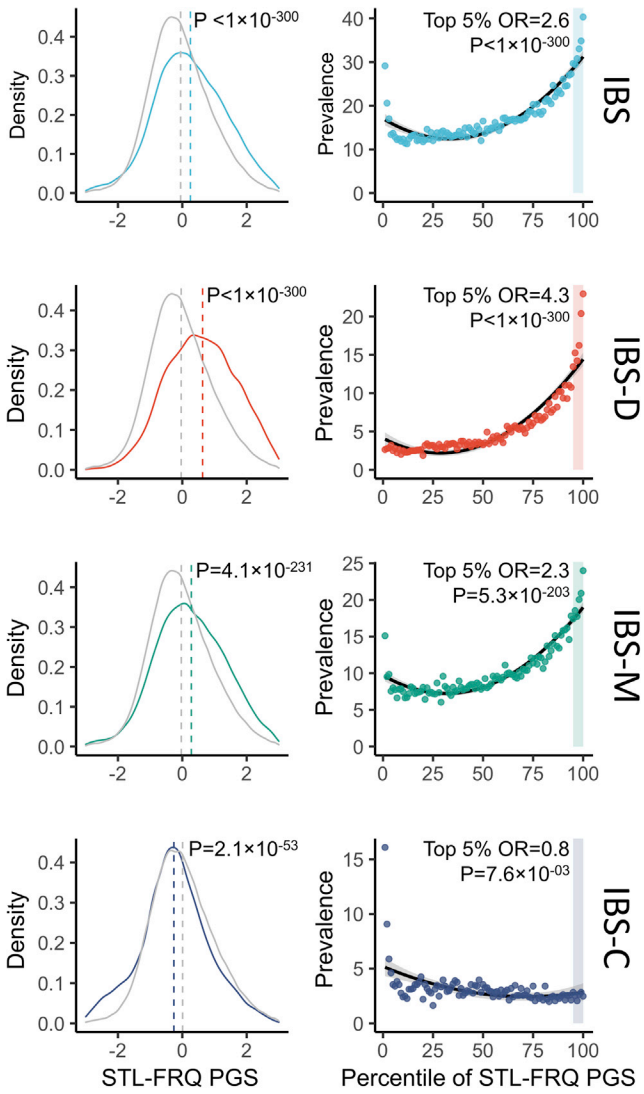


Figure 3. Stool frequency polygenic scores and IBS in UK Biobank

Results are reported (including statistical significance) in relation to PGS distribution in IBS cases versus controls (left panels; p values from t test) and in relation to the prevalence of IBS across PGS percentiles in the entire cohort (right panels; with top 5% of the distribution highlighted with shaded area, p values versus the rest of the cohort from logistic regression). IBS and subtypes defined according to Rome III Criteria based on DHQ questionnaire data (see STAR Methods and Table S15).

plasticity.[24] It has also been implicated in several diseases including major depression, bipolar disorder, and other psychiatric conditions.[25] *BDNF* is recognized to influence many important gut functions, including sensation, motility, epithelial barrier, neuroprotection, and neuroplasticity.[26] Multiple lines of evidence indicate that *BDNF* has prokinetic effects on gut motility,[27–29] including accelerated GI and colonic transit in healthy individuals administered recombinant *BDNF* (r-metHuBDNF).[30] Hence, our findings are in line with these observations in that the rs12273363 T allele associated with more frequent stools, shorter colonic transit time, and reduced stool consistency has also been shown to associate with stronger *BDNF* expression.[23] Altogether, this suggests a bona fide role for *BDNF* in the genetically determined modulation of human gut motility and warrants new analyses of recombinant *BDNF* trials based on genotype stratification, eventually also in relation to some of its reported side effects.[30]

Our results also point to interesting candidate genes from other stool frequency loci where the association signal has been refined: *ACHE*, *FAXDC2*, *CDK18*, and the female-specific signal for *FFAR3* all show eQTL associations with individual variants that have been fine mapped with >50% probability. *ACHE* codes for an enzyme that hydrolyzes the neurotransmitter acetylcholine at neuromuscular junctions and is overexpressed in Hirschsprung's disease,[31] while *FAXDC2* is a hydroxylase of fatty acids whose luminal concentrations are known to affect gut motility;[32,33] they are both expressed in enteric and motor neurons and therefore represent ideal functional candidates. *CDK18* encodes a protein kinase expressed in colonic M cells and BEST4+ enterocytes specialized in electrolyte and pH sensing;[15,34] hence, its associated colon-specific eQTL may be relevant to colonic osmolarity and, consequently, transit. Finally, *FFAR3* is expressed on enterochromaffin cells and enteric neurons, where it can sense short-chain fatty acids derived from bacterial fermentation and induce suppression of serotonin-mediated gut motility.[35,36]

Other strong functional candidates with a well-known role in GI motility include alpha and beta calcitonin gene-related peptides (*CALCA* and *CALCB* genes) from the rs6486216 locus[37] and the corticotropin-releasing hormone receptor (*CRHR1*) from the rs2732706 locus.[38] In particular, *CALCA* and *CALCB* code for a series of small hormones with a known role in gut physiology (due to tissue-specific alternative RNA splicing and post-translational processing).[39] These include alpha (*CALCA*) and beta (*CALCB*) versions of calcitonin, calcitonin gene-related peptide, and katacalcin, resulting in the potential involvement of up to a dozen or more signaling proteins in the mechanism(s) linking genetic variation to the modulation of stool frequency.

Our PheWAS analysis finds that 10 out of 14 signals associated with stool frequency in the current study have been previously associated with lifestyle, anthropometric, and disease-related traits (psychiatric conditions in particular). Also of interest is the association with fecal abundance of Ruminococcaceae UCG-005 detected for the *BDNF* locus, where the rs12273363 T allele linked to more frequent stools was also associated with decreased abundance of this genus. This is in line with previous observations linking increased fecal levels of Ruminococcaceae to harder (usually less frequent) stools.[40]

Broader evidence of genetic overlap with other conditions came from our LDSC analyses, which further highlighted shared genetic architecture with gastrointestinal diseases and comorbid neuroaffective traits,[41] independent of risk factors associated with use of specific medications. We further explored this in relation to IBS by computing PGSs based on our stool frequency GWAS meta-analysis and testing them in two independent subsets of UKBB. Stool frequency PGSs were significantly higher in IBS cases versus asymptomatic controls defined according to Rome III criteria (from questionnaire) or a doctor's

diagnosis (self-reported or from medical records). Individuals from the upper tail of the PGS distribution were more likely affected by IBS and exposed to up to >5× higher risk of IBS-D compared to the rest of the population (in the top 1% of the distribution). Of note, at least in UKBB, the heritability of stool frequency (h^2_{SNP} = 0.073) appears to be higher than that of IBS (h^2_{SNP} = 0.037 on the liability scale, based on previous GWAS data on self-reported IBS).[42] This suggests that, once necessarily refined and further validated in independent cohorts, PGSs derived from the simple stool frequency trait may ultimately contribute to an early identification, and eventual preventive treatment, of individuals at higher risk of developing IBS and other complex dysmotility syndromes.

Limitations of the study

Our study has some limitations based on the study design. First, the current analyses could not take into account likely contributing environmental factors, including diet, as this information was unavailable in most of the datasets. Second, cell types and neuronal subtypes relevant to stool frequency candidate genes have been classified here based on single-cell gene expression data in a small number of tissue types. Additional expression analyses in other tissues and cell types, as well as functional characterizations, are required in order to examine specific mechanisms involved in the control of motility. Finally, additional genetic and mechanistic studies are needed to resolve causative gene(s) and/or variant(s) at the associated loci. These are important directions for follow-up investigation and future studies.

In conclusion, we identify 14 loci associated with stool frequency. Our gene and pathway analyses identify an enrichment of genes with a plausible role in GI motility, possibly acting via neurotransmission and similar pathways in subsets of enteric neurons. The demonstrated relevance of these findings to IBS warrants further study for the identification of actionable pathological mechanisms in the dysmotility syndromes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2021.100069.

ACKNOWLEDGMENTS

This research has been conducted using the UK Biobank Resource under application number 17435. This study was supported by grants from the Swedish Research Council (VR 2017-02403), the Health Department of the Basque Government (2015111133), and the Spanish Ministry of Economy and Competitiveness (FIS PI17/00308) to M.D.; the research leading to these results has received funding from the EU FP7 under grant nr. 313010 (BBMRI-LPC); the FGFP project received support from the Flemish government (IWT130359), the Research Fund–Flanders (FWO) Odysseus program (G.0924.09), the King Baudouin Foundation (2012-J80000-004), FP7 METACARDIS HEALTH-F4-2012-305312, VIB, the Rega Institute for Medical Research, and KU Leuven. R.B. is funded by the Research Fund–Flanders (FWO) through a postdoctoral fellowship (1221620N). A.Z. is supported by the ERC Starting Grant 715772, Netherlands Organization for Scientific Research NWO-VIDI grant 016.178.056, the Netherlands Heart Foundation CVON grant 2018-27, and the NWO Gravitation grant ExposomeNL 024.004.017.

AUTHOR CONTRIBUTIONS

Conceptualization, M.D. and A.Z.; investigation and data curation, A.A., L.A., S.W., G.A., M.P., D.A.H., N.T., J.R., A.F., N.A.K., A.R., A.Z., M.S., M.C., and M.D.; formal analysis, T.Z., H.N., and K.G.-E.; formal analysis and investigation, F.B., X.L., C.S., A.P., A.K., R.B., L.B., C.E., L.J., M.P., N.T., J.R., A.F., N.A.K., A.R., A.Z., M.S., M.C., and M.D.; visualization and validation, F.B.; funding acquisition, project administration and supervision, M.D.; writing original draft, F.B. and M.D. with input and critical revision from all other authors.

DECLARATION OF INTERESTS

A.R. is a founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas Therapeutics, and, until August 31, 2020, was a scientific advisory board member of Syros Pharmaceuticals, Neogene Therapeutics, Asimov, and Thermo Fisher Scientific. A.R. is a member of the journal advisory board. From August 1, 2020, A.R. is an employee of Genentech.

Received: August 13, 2020
Revised: July 27, 2021
Accepted: October 11, 2021
Published: December 8, 2021

REFERENCES

1. Scratcherd, T., and Grundy, D. (1984). The physiology of intestinal motility and secretion. Br. J. Anaesth. *56*, 3–18.
2. Quigley, E.M.M. (2011). Microflora modulation of motility. J. Neurogastroenterol. Motil. *17*, 140–147.
3. Keller, J., Bassotti, G., Clarke, J., Dinning, P., Fox, M., Grover, M., Hellström, P.M., Ke, M., Layer, P., Malagelada, C., et al.; International Working Group for Disorders of Gastrointestinal Motility and Function (2018). Expert consensus document: Advances in the diagnosis and classification of gastric and intestinal motility disorders. Nat. Rev. Gastroenterol. Hepatol. *15*, 291–308.
4. Simrén, M., and Tack, J. (2018). New treatments and therapeutic targets for IBS and other functional bowel disorders. Nat. Rev. Gastroenterol. Hepatol. *15*, 589–605.

5. Camilleri, M., Shin, A., Busciglio, I., Carlson, P., Acosta, A., Bharucha, A.E., Burton, D., Lamsam, J., Lueke, A., Donato, L.J., and Zinsmeister, A.R. (2014). Genetic variation in GPBAR1 predisposes to quantitative changes in colonic transit and bile acid excretion. Am. J. Physiol. Gastrointest. Liver Physiol. *307*, G508–G516.

6. Camilleri, M., Carlson, P., Zinsmeister, A.R., McKinzie, S., Busciglio, I., Burton, D., Zucchelli, M., and D'Amato, M. (2010). Neuropeptide S receptor induces neuropeptide expression and associates with intermediate phenotypes of functional gastrointestinal disorders. Gastroenterology *138*, 98–107.e4.

7. Wong, B.S., Camilleri, M., Carlson, P.J., Guicciardi, M.E., Burton, D., McKinzie, S., Rao, A.S., Zinsmeister, A.R., and Gores, G.J. (2011). A Klothoβ variant mediates protein stability and associates with colon transit in irritable bowel syndrome with diarrhea. Gastroenterology *140*, 1934–1942.

8. Fox, M.R., Kahrilas, P.J., Roman, S., Gyawali, C.P., Scott, S.M., Rao, S.S., Keller, J., and Camilleri, M.; International Working Group for Disorders of Gastrointestinal Motility and Function (2018). Clinical measurement of gastrointestinal motility and function: who, when and which test? Nat. Rev. Gastroenterol. Hepatol. *15*, 568–579.

9. Jaruvongvanich, V., Patcharatrakul, T., and Gonlachanvit, S. (2017). Prediction of Delayed Colonic Transit Using Bristol Stool Form and Stool Frequency in Eastern Constipated Patients: A Difference From the West. J. Neurogastroenterol. Motil. *23*, 561–568.

10. Saad, R.J., Rao, S.S.C., Koch, K.L., Kuo, B., Parkman, H.P., McCallum, R.W., Sitrin, M.D., Wilding, G.E., Semler, J.R., and Chey, W.D. (2010). Do stool form and frequency correlate with whole-gut and colonic transit? Results from a multicenter study in constipated individuals and healthy controls. Am. J. Gastroenterol. *105*, 403–411.

11. Fang, H., Knezevic, B., Burnham, K.L., and Knight, J.C. (2016). XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. Genome Med. *8*, 129.

12. Deelen, P., van Dam, S., Herkert, J.C., Karjalainen, J.M., Brugge, H., Abbott, K.M., van Diemen, C.C., van der Zwaag, P.A., Gerkes, E.H., Zonneveld-Huijssoon, E., et al. (2019). Improving the diagnostic yield of exome- sequencing by predicting gene-phenotype associations using large-scale gene expression analysis. Nat. Commun. *10*, 2837.

13. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z., and Bergmann, S. (2016). Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. PLoS Comput. Biol. *12*, e1004714.

14. Smillie, C.S., Biton, M., Ordovas-Montanes, J., Sullivan, K.M., Burgin, G., Graham, D.B., Herbst, R.H., Rogel, N., Slyper, M., Waldman, J., et al. (2019). Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell *178*, 714–730.e22.

15. Drokhlyansky, E., Smillie, C.S., Van Wittenberghe, N., Ericsson, M., Griffin, G.K., Eraslan, G., Dionne, D., Cuoco, M.S., Goder-Reiser, M.N., Sharova, T., et al. (2020). The Human and Mouse Enteric Nervous System at Single-Cell Resolution. Cell *182*, 1606–1622.e23.

16. Modarresi, F., Faghihi, M.A., Lopez-Toledano, M.A., Fatemi, R.P., Magistri, M., Brothers, S.P., van der Brug, M.P., and Wahlestedt, C. (2012). Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. Nat. Biotechnol. *30*, 453–459.

17. Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., Le Roy, C.I., Raygoza Garay, J.A., Finnicum, C.T., Liu, X., et al. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat. Genet. *53*, 156–165.

18. Soret, R., Chevalier, J., De Coppet, P., Poupeau, G., Derkinderen, P., Segain, J.P., and Neunlist, M. (2010). Short-chain fatty acids regulate the enteric neurons and control gastrointestinal motility in rats. Gastroenterology *138*, 1772–1782.

19. Jankipersadsing, S.A., Hadizadeh, F., Bonder, M.J., Tigchelaar, E.F., Deelen, P., Fu, J., Andreasson, A., Agreus, L., Walter, S., Wijmenga, C., et al. (2017). A GWAS meta-analysis suggests roles for xenobiotic metabolism and ion channel activity in the biology of stool frequency. Gut *66*, 756–758.

20. Blizard, D.A., and Adams, N. (2002). The Maudsley Reactive and Nonreactive strains: a new perspective. Behav. Genet. *32*, 277–299.

21. Branicky, R., and Hekimi, S. (2006). What keeps C. elegans regular: the genetics of defecation. Trends Genet. *22*, 571–579.

22. Tack, J., and Vanden Berghe, P. (2000). Neuropeptides and colonic motility: it's all in the little brain. Gastroenterology *119*, 257–260.

23. Hing, B., Davidson, S., Lear, M., Breen, G., Quinn, J., McGuffin, P., and MacKenzie, A. (2012). A polymorphism associated with depressive disorders differentially regulates brain derived neurotrophic factor promoter IV activity. Biol. Psychiatry *71*, 618–626.

24. Maisonpierre, P.C., Belluscio, L., Friedman, B., Alderson, R.F., Wiegand, S.J., Furth, M.E., Lindsay, R.M., and Yancopoulos, G.D. (1990). NT-3, BDNF, and NGF in the developing rat nervous system: parallel as well as reciprocal patterns of expression. Neuron *5*, 501–509.

25. Di Carlo, P., Punzi, G., and Ursini, G. (2019). Brain-derived neurotrophic factor and schizophrenia. Psychiatr. Genet. *29*, 200–210.

26. Liu, S. (2018). Neurotrophic factors in enteric physiology and pathophysiology. Neurogastroenterol. Motil. *30*, e13446.

27. Grider, J.R., Piland, B.E., Gulick, M.A., and Qiao, L.Y. (2006). Brain-derived neurotrophic factor augments peristalsis by augmenting 5-HT and calcitonin gene-related peptide release. Gastroenterology *130*, 771–780.

28. Chai, N.-L., Dong, L., Li, Z.-F., Du, K.-X., Wang, J.-H., Yan, L.-K., and Dong, X.-L. (2003). Effects of neurotrophins on gastrointestinal myoelectric activities of rats. World J. Gastroenterol. *9*, 1874–1877.

29. Chen, F., Yu, Y., Wang, P., Dong, Y., Wang, T., Zuo, X., and Li, Y. (2014). Brain-derived neurotrophic factor accelerates gut motility in slow-transit constipation. Acta Physiol. (Oxf.) *212*, 226–238.

30. Coulie, B., Szarka, L.A., Camilleri, M., Burton, D.D., McKinzie, S., Stambler, N., and Cedarbaum, J.M. (2000). Recombinant human neurotrophic factors accelerate colonic transit and relieve constipation in humans. Gastroenterology *119*, 41–50.

31. Moore, S.W., and Johnson, G. (2005). Acetylcholinesterase in Hirschsprung's disease. Pediatr. Surg. Int. *21*, 255–263.

32. Jin, Q., Ren, Y., Wang, M., Suraneni, P.K., Li, D., Crispino, J.D., Fan, J., and Huang, Z. (2016). Novel function of FAXDC2 in megakaryopoiesis. Blood Cancer J. *6*, e478, e478.

33. Neunlist, M., and Schemann, M. (2014). Nutrient-induced changes in the phenotype and function of the enteric nervous system. J. Physiol. *592*, 2959–2965.

34. Malumbres, M., Harlow, E., Hunt, T., Hunter, T., Lahti, J.M., Manning, G., Morgan, D.O., Tsai, L.-H., and Wolgemuth, D.J. (2009). Cyclin-dependent kinases: a family portrait. Nat. Cell Biol. *11*, 1275–1276.

35. Kaji, I., Akiba, Y., Furuyama, T., Adelson, D.W., Iwamoto, K., Watanabe, M., Kuwahara, A., and Kaunitz, J.D. (2018). Free fatty acid receptor 3 activation suppresses neurogenic motility in rat proximal colon. Neurogastroenterol. Motil. *30*, 30.

36. Nøhr, M.K., Pedersen, M.H., Gille, A., Egerod, K.L., Engelstoft, M.S., Husted, A.S., Sichlau, R.M., Grunddal, K.V., Poulsen, S.S., Han, S., et al. (2013). GPR41/FFAR3 and GPR43/FFAR2 as cosensors for short-chain fatty acids in enteroendocrine cells vs FFAR3 in enteric neurons and FFAR2 in enteric leukocytes. Endocrinology *154*, 3552–3564.

37. Taché, Y., Garrick, T., and Raybould, H. (1990). Central nervous system action of peptides to influence gastrointestinal motor function. Gastroenterology *98*, 517–528.

38. Stengel, A., and Taché, Y. (2009). Neuroendocrine control of the gut during stress: corticotropin-releasing factor signaling pathways in the spotlight. Annu. Rev. Physiol. *71*, 219–239.

39. Findlay, D.M., and Sexton, P. (2003). Calcitonin Helen Henry and Anthony Norman (Academic Press).

40. Vandeputte, D., Falony, G., Vieira-Silva, S., Tito, R.Y., Joossens, M., and Raes, J. (2016). Stool consistency is strongly associated with gut

microbiota richness and composition, enterotypes and bacterial growth rates. Gut *65*, 57–62.

41. Fond, G., Loundou, A., Hamdani, N., Boukouaci, W., Dargel, A., Oliveira, J., Roger, M., Tamouza, R., Leboyer, M., and Boyer, L. (2014). Anxiety and depression comorbidities in irritable bowel syndrome (IBS): a systematic review and meta-analysis. Eur. Arch. Psychiatry Clin. Neurosci. *264*, 651–660.
42. Bonfiglio, F., Zheng, T., Garcia-Etxebarria, K., Hadizadeh, F., Bujanda, L., Bresso, F., Agreus, L., Andreasson, A., Dlugosz, A., Lindberg, G., et al. (2018). Female-Specific Association Between Variants on Chromosome 9 and Self-Reported Diagnosis of Irritable Bowel Syndrome. Gastroenterology *155*, 168–179.
43. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47* (*D1*), D1005–D1012.
44. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A.P., and Price, A.L. (2018). Mixed-model association for biobank-scale datasets. Nat. Genet. *50*, 906–908.
45. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat. Genet. *50*, 1335–1341.
46. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
47. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826.
48. Benner, C., Spencer, C.C.A., Havulinna, A.S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics *32*, 1493–1501.
49. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M.; Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.
50. Choi, S.W., and O'Reilly, P.F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience *8*, 8.
51. Kamat, M.A., Blackshaw, J.A., Young, R., Surendran, P., Burgess, S., Danesh, J., Butterworth, A.S., and Staley, J.R. (2019). PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics *35*, 4851–4853.
52. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J., et al.; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. *44*, 369–375, S1–S3.
53. Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M.R., McGrath, J.J., Visscher, P.M., Wray, N.R., and Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat. Commun. *9*, 224.
54. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.
55. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.
56. Tigchelaar, E.F., Zhernakova, A., Dekens, J.A.M., Hermes, G., Baranska, A., Mujagic, Z., Swertz, M.A., Muñoz, A.M., Deelen, P., Cénit, M.C., et al. (2015). Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. BMJ Open *5*, e006772.
57. Brieger, K., Zajac, G.J.M., Pandit, A., Foerster, J.R., Li, K.W., Annis, A.C., Schmidt, E.M., Clark, C.P., McMorrow, K., Zhou, W., et al. (2019). Genes for Good: Engaging the Public in Genetics Research via Social Media. Am. J. Hum. Genet. *105*, 441–442.
58. Vandeputte, D., Tito, R.Y., Vanleeuwen, R., Falony, G., and Raes, J. (2017). Practical considerations for large-scale gut microbiome studies. FEMS Microbiol. Rev. *41* (*Supp_1*), S154–S167.
59. Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., et al. (2016). Population-level analysis of gut microbiome variation. Science *352*, 560–564.
60. Kjellström, L., Molinder, H., Agréus, L., Nyhlin, H., Talley, N.J., and Andreasson, A. (2014). A randomly selected population sample undergoing colonoscopy: prevalence of the irritable bowel syndrome and the impact of selection factors. Eur. J. Gastroenterol. Hepatol. *26*, 268–275.
61. Walter, S.A., Kjellström, L., Nyhlin, H., Talley, N.J., and Agréus, L. (2010). Assessment of normal bowel habits in the general adult population: the Popcol study. Scand. J. Gastroenterol. *45*, 556–566.
62. Le Nevé, B., Brazeilles, R., Derrien, M., Tap, J., Guyonnet, D., Ohman, L., Törnblom, H., and Simrén, M. (2016). Lactulose Challenge Determines Visceral Sensitivity and Severity of Symptoms in Patients With Irritable Bowel Syndrome. Clin. Gastroenterol. Hepatol. *14*, 226, 33.e1, 3.
63. Tap, J., Derrien, M., Törnblom, H., Brazeilles, R., Cools-Portier, S., Doré, J., Störsrud, S., Le Nevé, B., Öhman, L., and Simrén, M. (2017). Identification of an Intestinal Microbiota Signature Associated With Severity of Irritable Bowel Syndrome. Gastroenterology *152*, 111–123.e8.
64. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B.S., et al.; Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics *33*, 272–279.
65. Battle, A., Brown, C.D., Engelhardt, B.E., and Montgomery, S.B.GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.
66. Törnblom, H., Van Oudenhove, L., Sadik, R., Abrahamsson, H., Tack, J., and Simrén, M. (2012). Colonic transit time and IBS symptoms: what's the link? Am. J. Gastroenterol. *107*, 754–760.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Stool frequency meta-analysis summary statistics	This paper	GWAS Catalog accession ID GCST90002250
Stool frequency PGS	This paper	https://www.pgscatalog.org
Human colon scRNA-Seq - processed data	Smillie et al., 2019[14]	Single cell portal accession SCP259 (https://singlecell.broadinstitute.org/single_cell)
Human enteric neurons 10X – processed data	Drokhlyansky et al., 2020[15]	Single cell portal accession SCP1038 (https://singlecell.broadinstitute.org/single_cell)
Mibiogen GWAS summary statistics	Kurilshikov et al., 2021[17]	https://mibiogen.gcc.rug.nl/
GWAS Catalog	Buniello et al., 2019[43]	https://www.ebi.ac.uk/gwas/
Software and algorithms		
BOLT-LMM v2.3.4	Loh et al., 2018[44]	https://alkesgroup.broadinstitute.org/BOLT-LMM/downloads/
SAIGE v0.38	Zhou et al., 2018[45]	https://github.com/weizhouUMICH/SAIGE/
METAL v2011-03-25	Willer et al., 2010[46]	https://genome.sph.umich.edu/wiki/METAL_Documentation
FUMA v1.3.5	Watanabe et al., 2017[47]	https://fuma.ctglab.nl
FINEMAP v1.3	Benner et al., 2016[48]	http://www.christianbenner.com/#
LDSC v1.0.1	Bulik-Sullivan et al., 2015[49]	https://github.com/bulik/ldsc
PRSice-2 v2.2.11	Choi and O'Reilly, 2019[50]	https://www.prsice.info
GeneNetwork v2.0	Deelen et al., 2019[12]	https://www.genenetwork.nl
PASCAL v1.0	Lamparter et al., 2016[13]	https://www2.unil.ch/cbg/index.php?title=Pascal
XGR v1.1.8	Fang et al., 2016[11]	http://galahad.well.ox.ac.uk:3030/
PhenoScanner v2	Kamat et al., 2019[51]	http://www.phenoscanner.medschl.cam.ac.uk
COJO 1.93.2beta	Yang et al., 2012[52]	https://yanglab.westlake.edu.cn/software/gcta/#COJO
mtCOJO 1.93.2beta	Zhu et al., 2018[53]	https://yanglab.westlake.edu.cn/software/gcta/#mtCOJO
PLINK v2.0	Chang et al., 2015[54]	https://www.cog-genomics.org/plink/2.0/
Other		
UK Biobank	Bycroft et al., 2018[55]	https://www.ukbiobank.ac.uk
LifeLines Deep	Tigchelaar et al., 2015[56]	https://www.lifelines.nl
Genes for Good Project	Brieger et al., 2019[57]	https://genesforgood.sph.umich.edu/
Flemish Gut Flora Project	Vandeputte et al., 2017;[58] Falony et al., 2016[59]	N/A
PopCol	Kjellström et al., 2014;[60] Walter et al., 2010[61]	N/A
IBS cohort with transit time data	Le Nevé et al., 2016;[62] Tap et al., 2017[63]	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Prof. Mauro D'Amato (mdamato@cicbiogune.es).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The stool frequency GWAS meta-analysis summary statistics have been deposited in the GWAS Catalog (https://www.ebi.ac.uk/gwas/) with the study accession ID GCST90002250. Stool frequency PGS is deposited in the PGS Catalog (https://www.pgscatalog.org/), under an ID associated with the publication. Publicly available analysis software and code were used as described in the method details.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study cohorts

We included in this study analyses of a total of 459,531 individuals with available phenotype and genotype data across five population-based cohorts, and a case cohort for transit studies (Table S1) Health-related information was derived from questionnaires and participants' electronic medical records, with stool frequency defined as the number of stool passes per day, after data harmonization (see below for cohort-specific definitions of stool frequency). Selected genotype data was also studied in relation to colonic transit time in a pilot follow-up analysis of a small cohort of 160 IBS patients from Sweden. A detailed description of all cohorts follows below (see also Tables S1 and S15 for summary information).

UK Biobank

UK Biobank (UKBB, https://www.ukbiobank.ac.uk) is a population-based cohort consisting of approximately 500,000 individuals (aged 40–69 years) genotyped on Affymetrix UK BiLEVE Axiom and Affymetrix UKBB Axiom arrays and recruited between 2006 and 2010 in the United Kingdom.[55] Each participant underwent cognitive and physical assessment at enrollment and, together with genotypes, extensive health-related information has been collected over time, including data from their healthcare medical records. Available genotype and a series of phenotypes derived from UKBB data have been used for the purpose of this study in relation to individuals self-reporting *white-British*, *Irish*, *white* ethnic background (UKBB data field 21000), as follows: 1) stool frequency was defined as a quantitative trait (mean = 1.42 ± 0.002SE) based on data extracted from the Digestive health web-based questionnaire (DHQ, https://biobank.ctsu.ox.ac.uk/showcase/showcase/docs/digestive_health.pdf) in relation to the question "What is the average number of times you open your bowels per day"? (UKBB data field 21044); 2) irritable bowel syndrome (IBS) and its diarrhea (IBS-D), constipation (IBS-C) and mixed (IBS-M) subtypes according to consensus Rome III Criteria were also derived from specific items in the DHQ (UKBB data fields 21025-21034), after excluding individuals reporting celiac disease or gluten sensitivity (UKBB data field 21068); 3) IBS diagnoses for a non-overlapping set of UKBB participants (those who did not fill the DHQ) were obtained based on the combination of non-cancer illness codes, self-reported (UKBB data field 20002) and K58 (IBS) codes according to the International Classifications of Diseases version-10 (ICD10) from hospital inpatient records (UKBB data fields 41202 and 41024) after excluding individuals self-reporting or ever diagnosed with celiac disease (UKBB data fields 20002, 41202 and 41024). UKBB received ethical approval from the competent Research Ethics Committee (REC reference for UKBB is 11/NW/0382). The demographics of UKBB individuals included in this study are reported in Tables S1 and S15, together with the studied phenotypes and the specific tools used to derive these.

LifeLines-DEEP

The LifeLines-DEEP cohort (LLD) is a subcohort of the prospective population-based LifeLines cohort (https://www.lifelines.nl) from the northern provinces of the Netherlands (Groningen, Drenthe and Friesland) and includes participants of Dutch ancestry whose molecular profiles (genome, epigenome, transcriptome, microbiome, metabolome and other biological parameters) and health-related information were collected for integrative analyses and the identification of disease biomarkers.[56] Data on stool frequency and consistency were extracted from gastrointestinal health questionnaires including daily records of defecation patterns (including number of bowel movements per day [mean = 1.37 ± 0.19SE], and Bristol Scale values [mean = 3.85 ± 0.03SE]), available for 942 individuals included in this study genotyped on customized Illumina Infinium array (HumanCytoSNP-12 BeadChip15 and the ImmunoChip). The Ethics Committee of the University Medical Centre Groningen approved the study. The demographics of LLD are reported in Table S1.

Genes for Good

The Genes for Good study (GFG, https://genesforgood.sph.umich.edu/) is a population-based cohort initiated by the University of Michigan in 2015, which includes participants across the US who provided a saliva sample for DNA genotyping on Illumina Inðnium CoreExome array.[57] Health history and daily tracking surveys (including bowel habits from a questionnaire on Gastrointestinal Conditions) were collected via social media (Facebook application). At the time of analysis, information on average weekly stool frequency (mean = 7.84 ± 0.12SE) and genetic data were available for 1,069 participants who answered the question "On average, how many times per week do you empty the bowel"? (harmonized to daily frequency dividing values by 7, with resulting mean = 1.12 ± 0.02SE). Stool consistency (mean = 3.79 ± 0.04SE) was derived from answers to the question "Looking at the picture provided (showing the Bristol Stool chart), what faeces type best represents your typical stool passing"?. The study was approved by the University of Michigan Institutional Review Board. The demographics of GFG are reported in Table S1.

Flemish Gut Flora Project

The Flemish Gut Flora Project (FGFP) is a population-based cohort that comprises individuals from the Flanders region of Belgium.[58,59] Primary aim of FGFP is the implementation of large-scale microbiome analyses in relation to health and lifestyle factors, and a 2018 release containing genotype (Illumina Human CoreExome array) and bowel data for 2,027 individuals was included in this study. Volunteers filled in a health and lifestyle questionnaire in which they reported frequency of defecation in two separate questions. In the first question, participants provided information on how many days they defecated the previous week. In the following question, they were asked about how many bowel movements they had, on average, on those days they had defecated. The stool frequency per day (mean = 1.37 ± 0.02SE) was obtained combining these two questions, that is multiplying the number of defecation days by the number of bowel movements on those days. The FGFP study was approved by the medical ethics committee of the University of Brussels - Brussels University Hospital (approval 143201215505, 5/12/2012), and the demographics of participants included in this study are reported in Table S1.

PopCol

The Population-based Colonoscopy (PopCol) study is a cohort of Swedish-born individuals from Stockholm, Sweden, which includes an omics-rich set of participants with available data from bowel symptoms questionnaires, gastroenterology visits, and biospecimens.[60,61] For the purpose of this study, we included data from 247 participants with available genotype (Illumina HumanOmniExpressExome 8v1 array) and stool frequency and consistency data (number of bowel movements per day, mean = 1.42 ± 0.044SE; Bristol Scale values mean = 3.96 ± 0.06SE), extracted from records of defecation patterns kept in daily diaries over a 1-2 weeks period. The study protocol was approved by Karolinska Institutet's Research Ethics Committee, and PopCol demographics of PopCol are reported in Table S1.

IBS cohort with transit time data

Swedish adult IBS patients (N = 160) were recruited at Sahlgrenska Hospital, University of Gothenburg, Sweden, and have been previously described elsewhere.[62,63] The diagnosis of IBS was based on clinical presentation and according to the Rome III Criteria. Exclusion criteria were other GI diseases including celiac disease, inflammatory bowel disease or microscopic colitis, medication history of probiotics or antibiotics within 1 month before the recruitment, severe psychiatric disorders, and history of drug or alcohol abuse. Selected genotypes (extracted from available Illumina Inðnium CoreExome24 array data) were studied in relation to colonic transit time (also called oro-anal transit time). Consent was obtained from all participants and the study was approved by the local ethics committee. The demographics of this cohort are reported in Table S1.

METHOD DETAILS

Genotype quality control, imputation, and individual GWAS

A common pipeline with minor modifications was applied to all cohorts for quality control (QC) of genotype data, imputation, and stool frequency GWAS tests. Briefly, Genotype QC filters were applied per sample (missing rate < 95%–99%; heterozygosity rate > 3SD) and per marker (call rate > 95%–99%; Hardy-Weinberg equilibrium $p < 1.0 \times 10^{-04}$), and individuals of non-European ancestry (detected from principal component analysis of European reference populations 1000 genomes or Human Genome Diversity Project), or with genotype-phenotype sex discrepancy were excluded from the analyses. Missing genotypes were imputed using the Haplotype Reference Consortium (HRC), UK10K, or 1KG as reference panels, and association tests were performed with mixed linear models to control for population stratification using BOLT-LMM v2.3.4[44] or SAIGE v0.38[45] on high quality (INFO > 0.8) common (MAF > 0.01) markers including sex, age, the first 10 PCs and genotyping array (if relevant) as covariates in the GWAS analyses. Sex-specific analyses in UKBB were carried out using the same GWAS pipeline, after stratifying participants' data into male and female groups. In order to counteract departures from normality, stool frequency values were rank-based inverse normal transformed before testing.

GWAS meta-analysis

Individual stool frequency GWAS results were brought forward into the pipeline for the purpose of the meta-analysis, with no GWAS showing genomic inflation (λ<1.1). Post-imputation QC on GWAS summary statistics was first implemented using EasyQC v9.2 (https://www.genepi-regensburg.de/easyqc), to check for data integrity and harmonize SNP IDs and allele coding across datasets. Markers with allele frequency deviating > 20% from the HRC reference panel were excluded, together with indels and multi-allelic markers. A fixed-effect meta-analysis based on the inverse-variance weighted method was performed with METAL v2011-03-25,[46] on a total of 167,875 individuals and 8,817,117 markers showing no heterogeneity of effects (Cochran's Q-test p value > 0.05) and present in at least 2 cohorts. The Manhattan plot was generated using the R package qqman v0.1.6 (https://github.com/stephenturner/qqman), annotating each locus with the nearest gene (if within 100kb from the tag SNP). Regional plots were produced with Locuszoom (http://locuszoom.org). The calculated linkage disequilibrium (LD) score regression intercept was equal to 1.02 (compared to a lambda inflation of 1.14), suggesting polygenicity mostly accounts for statistical inflation.[64]

Functional annotation of stool frequency loci

Locus definition and content

Annotation of stool frequency loci was done with functional mapping and annotation of genetic associations (FUMA) v1.3.5,[47] based on GWAS meta-analysis summary statistics. Independent association signals were identified based on SNP P value ($\leq 5.0 \times 10^{-8}$) and LD between markers (r2 < 0.6). Association signals were merged into a single locus for LD blocks closer than 250kb apart. Gene

content at stool frequency loci was annotated based on positional and eQTL mapping, also with FUMA using default parameters and FDR $p < 0.05$. The COnditional and JOint (COJO) function from the GCTA 1.93.2beta software package[52] was used to run conditional analyses at each locus, and identify independent association signals based on stool frequency GWAS meta-analysis summary statistics (COJO performs secondary association scans conditioning on lead SNPs).

Fine mapping

Fine-mapping was performed for the 14 genome-wide significant loci using FINEMAP v1.3,[48] with z-scores from the stool frequency GWAS meta-analysis and LD matrices derived from the genotype probabilities (.bgen files) of UKBB data. Specific eQTL traits associated with fine mapped SNPs were identified based on data from GTEx v8.[65]

Enrichment analyses

Gene-set and pathway enrichment analyses

Functional enrichment of stool frequency associated genes (as from positional and eQTL mapping with FUMA) was evaluated using multiple alternative computational approaches. Classical gene-set enrichment analyses were performed with the R implementation of XGR v1.1.8 (http://galahad.well.ox.ac.uk:3030/),[11] using the "xEnricherGenes" function to screen with hypergeometric test the REACTOME pathways with a minimum overlap of 5 genes. We then used GeneNetwork v2.0,[12] in relation to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) terms, exploiting gene co-regulation matrices to predict pathway membership by integrating 31,499 public RNA-seq samples. The algorithm uses a method based on principal component analysis of co-expression matrices, and subsequent non-parametric testing (Mann-Whitney U-test) between components of known gene sets and the genes of interest. Enrichment of molecular pathways from the REACTOME libraries was tested using PASCAL with stool frequency GWAS meta-analysis summary statistics.[13] PASCAL uses a pathway scoring based on a modified Fisher method with a better control on type I errors. The algorithm uses GWAS meta-analysis summary statistics as input and allows to aggregate SNP P values across pathway genes, while controlling for genes in LD that cannot be treated independently.

Cell-type enrichment analyses

Look up of stool frequency gene expression was done on previously reported scRNA-seq and RAISIN RNA-seq data from human colonic mucosa and muscularis propria,[14,15] in relation to 76 cell types including immune, stromal and enteric neurons, among others. As described,[15] enteric neurons were partitioned into 5 classes based on the expression of major neurotransmitters/neuropeptides (*CHAT*, *SLC5A7*, *NOS1*, *VIP*) and other known markers: putative sensory neurons (PSN), interneurons (PIN subsets 1 and 2), secretomotor/vasodilator neurons (PSVN), and excitatory (PEMN subsets 1 and 2) and inhibitory motor neurons (PIMN subsets 1-5). Heatmaps were produced using the ggplot2 R package (https://ggplot2.tidyverse.org) on log2(TP10K+1) transformed data, and the expression of each gene scaled across all cells. A heatmap showing gene expression collapsed by locus was created by calculating and reporting locus-specific median values of gene expression for genes at corresponding loci. Enrichment tests, comparing the expression of stool frequency genes versus background genes in enteric and motor neurons, were conducted using Fisher's Exact test, controlling for type I error by FDR adjustment.

Cross-trait analyses

PheWAS for stool frequency GWAS signals in other traits

The GWAS Catalog[43] and PhenoScanner v2[51] were screened with 14 stool frequency lead SNPs and their high LD proxies ($r2 > 0.8$) in order to highlight associations ($P \leq 5.0 \times 10^{-8}$) with other traits. Associations were plotted with Circlize (cran.r-project.org/web/packages/circlize). GWAS meta-analyses of human genome-microbiome association studies (N = 211 taxa) were downloaded from the MiBioGen consortium (https://mibiogen.gcc.rug.nl) and screened with stool frequency lead SNPs and their high LD proxies ($r2 > 0.8$). For each risk locus, we considered the best P value obtained for the lead SNP or its LD proxies, after Bonferroni correction for multiple comparisons ($p < 1.7 \times 10^{-5}$ significance threshold, considering adjustment for 211 MiBioGen traits × 14 stool frequency loci).

Genetic correlations

Stool frequency h^2_{SNP} and the r_g between stool frequency and other complex traits were estimated using LDSC v1.0.1,[49] implemented in the CTG-VL platform (vl.genoma.io), which integrates public summary statistics of 1,387 traits from multiple repositories, filtered based on significant ($p < 0.05$) heritability with h^2_{SNP} z-score^3 2. Tests for statistical significance were FDR adjusted to control for type I errors.

Multi trait conditional analyses

Multi-trait COnditional and JOint analysis 1.93.2beta (mtCOJO),[53] was adopted to generate conditioned GWAS summary statistics for stool frequency (outcome) after correcting for the SNP effects of medication traits (exposures). mtCOJO estimates the effect of the exposure on the outcome either by generalized summary-data-based mendelian randomization or from genetic correlation (r_g) analysis when there are not enough genome-wide significant SNPs for the exposure trait (at least 10 required by default). mtCOJO analysis was carried out on stool frequency GWAS meta-analysis data versus medication traits (derived from UKBB fields 20003 and 6154) that showed significant r_g from LDSC analysis.

Polygenic score analyses

PGS based on a pruning and thresholding approach were built using PRSice-2 v2.2.11.[50] Effect estimates and corresponding standard errors from the stool frequency GWAS meta-analysis were studied for their relevance to IBS in a 2-step approach including testing and validation. First, we used them to generate the best PGS model for IBS according to gold-standard Rome III Criteria

(based on DHQ questionnaire data available for N = 163,616 individuals, see Table S15), using PRSice-2 default settings and UKBB v3 imputed data. Specifically, the best model was derived for IBS and subtypes (IBS-C, IBS-D and IBS-M) based on the largest Nagelkerke's R2 value obtained by testing varying numbers of SNPs selected from a large range of stool frequency GWAS P value thresholds (from $p = 5.0 \times 10^{-8}$ to p = 1) according to the standard PRSice-2 pipeline. A total of 263,279 LD pruned variants (clumping parameters clump-kb = 250, clump-p = 1, clump-r2 = 0.1 as default) entered the analysis, and the best models included 263,279 variants for Rome III IBS, IBS-D and IBS-M (P threshold = 1), and 146,077 variants for Rome III IBS-C (P threshold = 0.332). Then, PGS were tested again versus IBS for validation (using the 263,279 best-model markers with PLINK[54] v2.0 –score function) in the independent DHQ- remainder UKBB subset (N = 291,496), based on IBS defined according to ICD10 and self-reported diagnoses from participants' healthcare records. The PGS for each individual in the target IBS trait were computed as the sum of effect sizes of all the SNPs included in the respective model. To account for differences in the numbers of variants per cohort and facilitate interpretation, a normalized polygenic score standardized to a normal distribution (mean = 0, SD = 1) was created per cohort. Student's t test was employed to determine the significance of the difference between the mean PGSs in IBS and controls. PGSs were binned into percentiles and the subset of IBS patients within a given magnitude of increased stool frequency PGS (top percentiles) was compared to the reminder of the population in a logistic regression adjusting for sex, age, the first 10 PCs and genotyping array.

Colonic transit time measurement

Patients were asked to ingest 10 radiopaque rings every morning for 6 consecutive days. On the morning of the seventh day, the radiopaque rings retained in the bowel were measured using fluoroscopy (Exposcop 7000 Compact; Ziehm GmbH, Nüremberg, Germany). Colonic transit time per day was calculated as the number of retained radiopaque rings divided by the daily dose number, that is by 10.[66] Colonic transit time values were scrutinized with the R package "bestNormalize" (https://github.com/petersonR/bestNormalize) and an ordered quantile transformation applied to obtain best normality statistics for the genetic analyses.

Look-up of GWAS signals in relation to stool consistency and colonic transit time

Pilot follow-up analyses of stool frequency GWAS signals were carried out in relation to stool consistency and colonic transit time. Summary statistics for the 14 stool frequency associated loci were extracted from GWAS data produced for 2338 individuals with Bristol stool score values from LLD, FGF and PopCol (see Study Cohorts section), and meta-analyzed using the same pipeline adopted for the stool frequency GWAS. The genotypes of lead SNPs (or r2 > 0.9 proxies) from the 14 stool frequency associated loci were extracted from imputed genotype data available for a small cohort of 160 Swedish IBS patients with measured colonic transit time (see Study Cohorts section). Following outliers' removal and genotype QC, GWAS effect alleles were tested for association in a linear regression model, adjusting for age, sex and first 10 PCs from PCA analysis of genotype data.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details on statistical tests, significance cut-offs and sample sizes can be found in the tables and figure legends. P values are reported based on FDR correction for type I errors, unless specified differently. When relevant, further details are found in the method details for the specific measurement.