

1 **GWAS of stool frequency reveals genes, pathways, and cell types**  
2 **relevant to human gastrointestinal motility and irritable bowel**  
3 **syndrome**

4  
5 Ferdinando Bonfiglio<sup>1,2</sup>, Xingrong Liu<sup>2,3,\*</sup>, Christopher Smillie<sup>4,\*</sup>, Anita Pandit<sup>5,\*</sup>, Alexander  
6 Kurilshikov<sup>6,\*</sup>, Rodrigo Bacigalupe<sup>7,8,\*</sup>, Tenghao Zheng<sup>1,2,3</sup>, Hieu Nim<sup>1</sup>, Koldo Garcia-  
7 Etxebarria<sup>9</sup>, Luis Bujanda<sup>9,10,11</sup>, Anna Andreasson<sup>12</sup>, Lars Agreus<sup>13</sup>, Susanna Walter<sup>14</sup>,  
8 Goncalo Abecasis<sup>5</sup>, Chris Eijsbouts<sup>15,16</sup>, Luke Jostins<sup>17,18</sup>, Miles Parkes<sup>19</sup>, David A  
9 Hughes<sup>20,21</sup>, Nicholas Timpson<sup>20,21</sup>, Jeroen Raes<sup>7,8</sup>, Andre Franke<sup>22</sup>, Nicholas A Kennedy<sup>23, #</sup>,  
10 Aviv Regev<sup>4, #</sup>, Alexandra Zhernakova<sup>6, #</sup>, Magnus Simren<sup>24, #</sup>, Michael Camilleri<sup>25, #</sup>, Mauro  
11 D'Amato<sup>1,2,9,26,§</sup>

12  
13 \*equal contribution

14 #equal contribution

15 §corresponding author

16  
17 <sup>1</sup>School of Biological Sciences, Monash University, Clayton VIC, Australia; <sup>2</sup>Unit of Clinical  
18 Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; <sup>3</sup>Center for  
19 Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; <sup>4</sup>Klarman Cell Observatory, Broad  
20 Institute, Cambridge, MA, USA; <sup>5</sup>Department of Biostatistics, University of Michigan, School of Public  
21 Health, Ann Arbor, MI, USA; <sup>6</sup>Department of Genetics, University of Groningen, University Medical  
22 Center Groningen, Groningen, the Netherlands; <sup>7</sup>Department of Microbiology and Immunology, Rega  
23 Instituut, KU Leuven, Leuven, Belgium; <sup>8</sup>Center for Microbiology, VIB, Leuven, 3000, Belgium.;  
24 <sup>9</sup>Department of Gastrointestinal and Liver Diseases, Biodonostia HRI, San Sebastián, Spain; <sup>10</sup>Centro  
25 de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd);  
26 <sup>11</sup>Universidad del País Vasco (UPV/EHU), San Sebastian, Spain; <sup>12</sup>Division of Clinical Medicine,  
27 Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; <sup>13</sup>Division of Family  
28 Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska  
29 Institutet, Stockholm, Sweden; <sup>14</sup>Division of Neuro and Inflammation Science, Department of Clinical  
30 and Experimental Medicine, Linköping University, Linköping, Sweden; <sup>15</sup>Wellcome Centre for Human  
31 Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK; <sup>16</sup>Big Data Institute, Li  
32 Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; <sup>17</sup>Kennedy  
33 Institute of Rheumatology, University of Oxford, Oxford, UK; <sup>18</sup>Christ Church, University of Oxford,  
34 Oxford, UK; <sup>19</sup>Division of Gastroenterology, Department of Medicine, University of Cambridge, UK;  
35 <sup>20</sup>MRC Integrative Epidemiology Unit at University of Bristol, Bristol, UK; <sup>21</sup>Population Health Sciences,  
36 Bristol Medical School, University of Bristol, Bristol, UK; <sup>22</sup>Institute of Clinical Molecular Biology,  
37 Christian-Albrechts-University of Kiel, Kiel, Germany; <sup>23</sup>IBD Pharmacogenetics, College of Medicine  
38 and Health, University of Exeter, Exeter, UK; <sup>24</sup>Dept of Internal Medicine & Clinical Nutrition, Institute  
39 of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; <sup>25</sup>Clinical Enteric  
40 Neuroscience Translational and Epidemiological Research (CENTER), and Division of  
41 Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic, Rochester, MN, USA;  
42 <sup>26</sup>IKERBASQUE, Basque Foundation for Science, Bilbao, Spain.

43  
44 **Correspondence:**

45 Mauro D'Amato

46 School of Biological Sciences, Monash University

47 25 Rainforest Walk, 3800 Clayton VIC, Australia

48 +61 3 99053751; [mauro.damato@monash.edu](mailto:mauro.damato@monash.edu)

49

50 **Word count:** 3998

51 **ABSTRACT**

52 **Objective.** Gut dysmotility is associated with constipation, diarrhea, and functional  
53 gastrointestinal disorders (FGID) like irritable bowel syndrome (IBS). Its molecular  
54 underpinnings, and their anomalies in FGID disorders are poorly characterized, hence we  
55 sought to gain mechanistic insight through a large-scale genetic investigation.

56 **Design.** We used stool frequency (STL-FRQ) as a (surrogate) quantitative trait to study the  
57 genetics of gut motility, exploiting questionnaire and genotype data from UK Biobank and  
58 four smaller population-based cohorts (LifeLines-Deep, Genes for Good, Flemish Gut Flora  
59 Project and PopCol), in a GWAS meta-analysis spanning 8,817,117 high-quality SNP  
60 markers and 167,875 individuals of European descent.

61 **Results.** We identify 13 genome-wide significant loci ( $P \leq 5.0 \times 10^{-8}$ ) harboring prioritized  
62 genes that are: i) involved in sensory perception and neurotransmitter/neuropeptide  
63 signaling; ii) enriched for their expression in enteric motor neurons associated with the  
64 control of peristalsis ( $P = 7.0 \times 10^{-8}$ ) iii) previously linked to other traits and conditions, including  
65 GI motility and dysmotility syndromes, and the response to their pharmacological treatment.  
66 The genetic architecture of STL-FRQ most strongly correlates with that of IBS ( $r_g = 0.42$ ;  
67  $P = 1.1 \times 10^{-3}$ ). In UK Biobank, the risk of IBS with diarrhea was 4x higher in individuals from  
68 the top 1% of the distribution of polygenic scores (PGS) computed based on STL-FRQ  
69 GWAS summary statistics (ORs=4.14;  $P = 1.2 \times 10^{-97}$ ).

70 **Conclusion.** We identify loci harboring genes with a plausible role in GI motility, possibly  
71 acting via neurotransmission and similar pathways in specialized enteric neurons. The  
72 demonstrated relevance of these findings to IBS warrants further study for the identification  
73 of actionable pathomechanisms in the dysmotility syndromes.

74

75 **Keywords.** Intestinal motility, genetics, irritable bowel syndrome

## 76 INTRODUCTION

77

78 Gastrointestinal (GI) motility is essential to digestion, nutrients absorption and overall  
79 human health, including bi-directional host-microbiome interactions.[1,2] Gut  
80 dysmotility and altered peristalsis are observed in constipation, diarrhea, and  
81 common functional GI disorders (FGIDs) like irritable bowel syndrome (IBS), which  
82 affect a large portion of the population and pose a remarkable socio-economic and  
83 healthcare burden.[3,4] While showing considerable symptoms overlap, most FGIDs  
84 are associated with some degree of GI motor dysfunction, possibly best exemplified  
85 by the observation that colonic transit time is generally delayed in patients with  
86 constipation-predominant IBS (IBS-C), and accelerated in patients with diarrhea-  
87 predominant IBS (IBS-D).[5,6] Dysmotility plausibly represents a key pathogenetic  
88 mechanism contributing to several GI conditions across a continuum ranging from  
89 mild symptoms, through functional disorders, and extreme cases of severely  
90 impaired peristalsis as observed in chronic idiopathic intestinal pseudo-obstruction.[7]  
91 There is only incomplete understanding of the physiological mechanisms regulating  
92 intestinal motility, and their perturbation in the dysmotility syndromes. While some  
93 medications (opioids, antidepressants, laxative and others) are known to influence GI  
94 motility, intrinsic triggers are generally believed to come from disturbed  
95 communication along the gut-brain axis, inflammatory or degenerative processes,  
96 and overstimulation of visceral sensory pathways that ultimately affect the gut  
97 musculature and local GI motor function via the enteric nervous system (ENS).[8]  
98 Therapeutic options are limited in the dysmotility syndromes, and rely on targeting  
99 specific symptoms rather than (currently unknown) underlying mechanisms. Genetic  
100 research may reveal biological pathways amenable to therapeutic exploitation, and

101 some evidence of heritability can be derived for gut motility from previous studies in  
102 relation to colonic transit time measured with detectable tracers.[9–11] These  
103 studies, however, were largely underpowered to capture reliable genetic effects, as  
104 they lacked replication and focused on few DNA variants from individual candidate  
105 genes in small cohorts of IBS patients, hence results are not conclusive or  
106 transferable to the general population. As of today, no large-scale study has been  
107 performed to identify the genetic determinants of gut motility.

108 Direct assessment of GI motility in humans requires demanding clinical procedures  
109 (like transit time scintigraphy or the radiopaque markers method), which are  
110 exclusively performed to support patient diagnosis and therapeutic management, and  
111 are therefore not suitable for large-scale population-wide genetic surveys.[12]  
112 However, stool consistency and, to a lesser extent, stool frequency (STL-FRQ which  
113 refers to the number of bowel movements over a period of time) are valuable  
114 indicators of bowel function that correlate with colonic transit time, and can be  
115 practically recorded based on questionnaire- or diary survey-based  
116 approaches.[13,14] These represent therefore practical surrogate tools that can be  
117 adopted and scaled for studying GI motility at the population level thanks to their  
118 ease of data collection. Precedent for this approach comes, for instance, from a  
119 similar strategy recently applied in a genome-wide association study (GWAS) of IBS  
120 as self-reported condition from questionnaire data.[15]

121 Here, we leverage data from UK Biobank and four smaller population-based cohorts  
122 (LifeLines-Deep, Genes for Good, Flemish Gut Flora Project and PopCol) for a STL-  
123 FRQ GWAS meta-analysis across 8,817,117 high-quality single nucleotide  
124 polymorphism (SNP) markers in a total of 167,875 individuals of European descent.  
125 We identify 13 loci that harbor genes associated with pathways and cell types

126 plausibly involved in the control of GI motility in humans, and provide compelling  
127 evidence of the relevance of these findings to IBS. The identification of genetic  
128 factors predisposing to altered gut motility may allow early identification of individuals  
129 at higher risk of FGID and, more importantly, therapeutically actionable pathways that  
130 may be targeted for the delineation of alternative treatment options.

## 131 **MATERIALS AND METHODS**

132

### 133 **STUDY COHORTS**

134 We studied phenotype and genotype data in 460,734 individuals from 5 population-  
135 based cohorts: UK Biobank (UKBB), LifeLines-DEEP cohort (LLD), The Genes for  
136 Good study (GFG), the Flemish Gut Flora Project (FGFP) and the Population-based  
137 Colonoscopy (PopCol) study. Health-related information was derived from  
138 questionnaires and participants' electronic medical records, with STL-FRQ defined as  
139 the number of stool passes per day, after data harmonization. Selected genotype  
140 data was also studied in relation to colonic transit time (CTT) in a small cohort of 160  
141 IBS patients from Sweden. A detailed description of all cohorts is reported in  
142 Supplementary Table 1 and the Supplementary Methods.

143

### 144 **GWAS META-ANALYSIS**

145 A common GWAS pipeline was applied to individual cohorts based on mixed linear  
146 models and high quality (INFO>0.8) common (MAF>0.01) markers. A fixed-effect  
147 meta-analysis based on the inverse-variance weighted method was performed on a  
148 total of 167,875 individuals and 8,817,117 markers. In the analysis of CTT, STL-FRQ  
149 GWAS effect alleles were tested for association with linear regression adjusting for  
150 age, sex and first 10 principal components. A summary of relevant data, quality  
151 control measures, and analytical procedures is reported in the Supplementary  
152 Methods.

153

### 154 **FUNCTIONAL ANNOTATION OF STL-FRQ LOCI**

155 **Locus definition and content.** Annotation of *STL-FRQ* loci was done with FUMA  
156 v1.3.5 (<https://fuma.ctglab.nl/>), based on GWAS meta-analysis summary statistics.  
157 Independent association signals were identified based on SNP P-value ( $\leq 5.0 \times 10^{-8}$ )  
158 and linkage disequilibrium (LD) between markers ( $r^2 < 0.4$ ). Association signals were  
159 merged into a single locus for LD blocks closer than 250kb apart. Gene content at  
160 *STL-FRQ* loci was annotated based on positional and expression quantitative trait  
161 loci (eQTL) mapping, also with FUMA using default parameters and false discovery  
162 rate (FDR)  $P < 0.05$ .

163 **Fine mapping.** Fine-mapping was performed for the 13 genome-wide significant loci  
164 using FINEMAP v1.3,[16] with z-scores from the *STL-FRQ* GWAS meta-analysis and  
165 LD matrices derived from the genotype probabilities (.bgen files) of UK Biobank data.  
166 Specific eQTL traits associated with fine mapped SNPs were identified based on  
167 data from GTEx v8.[17]

168

## 169 **ENRICHMENT ANALYSES**

170 **Gene-set and pathway enrichment analyses.** Functional enrichment of *STL-FRQ*  
171 associated genes (as from positional and eQTL mapping with FUMA) was evaluated  
172 using GeneNetwork v2.0 ([www.genenetwork.nl](http://www.genenetwork.nl)), in relation to KEGG pathways and  
173 Gene Ontology (GO) terms, using pre-computed co-regulation Z-scores and a Mann-  
174 Whitney U-test with FDR correction. Enrichment of molecular pathways from the  
175 REACTOME libraries was tested using PASCAL with *STL-FRQ* GWAS meta-  
176 analysis summary statistics, and default parameters including type I error control.[18]

177 **Cell-type enrichment analyses.** Look up of *STL-FRQ* gene expression was done on  
178 previously reported single cell RNA-seq (scRNA-seq) and Ribosomes And Intact  
179 Single Nucleus isolation (RAISIN) RNA-seq data from human colonic mucosa and

180 muscularis propria,[19,20] in relation to 76 cell types including immune, stromal and  
181 enteric neurons, among others. As described,[20] enteric neurons were partitioned  
182 into 5 classes based on the expression of major neurotransmitters/neuropeptides  
183 (*CHAT*, *SLC5A7*, *NOS1*, *VIP*) and other known markers: putative sensory neurons  
184 (PSN), interneurons (PIN subsets 1 and 2), secretomotor/vasodilator neurons  
185 (PSVN), and excitatory (PEMN subsets 1 and 2) and inhibitory motor neurons (PIMN  
186 subsets 1-5). Enrichment tests, comparing the expression of STL-FRQ genes versus  
187 background genes in enteric and motor neurons, were conducted using Fisher's  
188 Exact test, controlling for type 1 error by FDR adjustment.

189

## 190 **CROSS-TRAIT ANALYSES**

191 **Lookup of STL-FRQ GWAS signals in other traits.** The GWAS catalog[21] and  
192 PhenoScanner[22] were screened with STL-FRQ lead SNPs and their high LD  
193 proxies ( $r^2 > 0.8$ ) in order to highlight associations ( $P \leq 5.0 \times 10^{-8}$ ) with other traits.  
194 Associations were plotted with Circlize ([cran.r-project.org/web/packages/circlize](http://cran.r-project.org/web/packages/circlize)).

195 **Genetic correlations.** STL-FRQ SNP heritability ( $h^2_{\text{SNP}}$ ) and genetic correlation ( $r_g$ )  
196 between STL-FRQ and other complex traits were estimated using LD score  
197 regression (LDSC v1.0.1),[23] implemented in the CTG-VL platform ([vl.genoma.io](http://vl.genoma.io)),  
198 which integrates public summary statistics of 1,387 traits from multiple repositories.  
199 Tests for statistical significance were FDR adjusted to control for type I errors.

200

## 201 **POLYGENIC SCORE ANALYSES**

202 Polygenic scores (PGS) based on a pruning and thresholding approach were built  
203 using PRSice-2.[24] Effect estimates and corresponding standard errors from the  
204 STL-FRQ GWAS meta-analysis were used as the base dataset to generate weights,

205 and then applied to IBS traits from UK Biobank (including subtypes) to derive PGS  
206 using PRSice-2 default settings. To account for the differences in the numbers of  
207 variants per cohort, a normalized polygenic score (mean=0, SD=1) was created per  
208 cohort. Student's t-test was employed to determine the significance of the difference  
209 between the mean PGSs in IBS and controls. PGSs were binned into percentiles and  
210 the subset of IBS patients within a given magnitude of increased STL-FRQ PGS (top  
211 percentiles) was compared to the remainder of the population in a logistic regression  
212 adjusting for sex, age, the first 10 PCs and genotyping array.

## 213 **RESULTS**

214

### 215 **STL-FRQ GWAS meta-analysis**

216 The distribution of STL-FRQ (harmonized to the number of stool passes per day) was  
217 similar in the studied cohorts, with average ranging from 1.12 in GFG to 1.42 in  
218 UKBB and PopCol. Independent GWAS were carried out in individual cohorts using a  
219 common pipeline (Supplementary Methods), and later included in a meta-analysis  
220 encompassing 167,875 participants and 8,817,117 high-quality SNP markers. The  
221 STL-FRQ GWAS meta-analysis showed no population stratification (LDSC intercept  
222 = 1.02, see Methods), and identified 3751 genome-wide-significant associations  
223 ( $P \leq 5.0 \times 10^{-8}$ ) from 13 independent loci (Table 1, Figure 1 and Supplementary Figure  
224 S1). The strongest signal was detected for marker rs12273363 on chromosome 11  
225 ( $P = 4.8 \times 10^{-21}$ ), in proximity of the BDNF gene.

226

### 227 **Gene-set and pathway enrichment analyses**

228 In order to obtain biological insight from the observed associations, we analysed  
229 STL-FRQ GWAS data with a computational pipeline for the functional annotation of  
230 associated loci. FUMA was used to define STL-FRQ GWAS loci, their boundaries  
231 and respective gene content, based on positional and eQTL mapping (see Methods).  
232 Several relevant genes were located at the associated loci (Supplementary Table 2),  
233 therefore we proceeded to perform gene-set and pathway enrichment analysis.  
234 GeneNetwork analysis revealed significant (FDR  $P < 0.01$ ) enrichment for relevant  
235 KEGG pathways including “*neuroactive ligand receptor interaction*”, and GO terms  
236 “*detection of chemical stimulus involved in sensory perception*” and “*neuropeptide  
237 signaling pathway*” (Figure 2 and Supplementary Table 3). Similarly, PASCAL

238 pathway-level analysis highlighted “*neurotransmitter receptor binding and*  
239 *downstream transmission in the postsynaptic cell*” and “*serotonin receptors*” as the  
240 top enriched REACTOME pathways (Supplementary Table 4).

241

#### 242 **Functional annotation at the single cell level**

243 STL-FRQ genes did not show any preferential tissue expression based on FUMA or  
244 DEPICT analyses (not shown), hence we turned to study cell-type specific  
245 expression using single cell transcriptomic data available from human colonic  
246 mucosa and muscularis propria.[19,20] Look up of STL-FRQ gene expression was  
247 carried out using scRNA-seq and RAISINs RNA-seq data (see Methods), in relation  
248 to immune, epithelial, stromal and glial cells, muscle cells, and 11 subtypes of enteric  
249 neurons grouped into 5 major classes based on the relative expression of major  
250 neurotransmitters neuropeptides and other markers (Figure 3 and Supplementary  
251 Figure S2). Of note, the expression of STL-FRQ associated genes was strongly  
252 enriched in enteric neurons (FDR  $P=1.4\times 10^{-3}$ ), and more so in specific putative  
253 excitatory and inhibitory motor neurons (PEMN and PIMN subtypes, FDR  $P=7.0\times 10^{-8}$ )  
254 reportedly involved in the control of peristalsis.[20] Hence, functional annotation of  
255 STL-FRQ genes at the single cell level points to a potential role in controlling human  
256 gut motility through the involvement of specialized neuronal populations.

257

#### 258 **Prioritization of causative genes**

259 FINEMAP analysis of candidate causative SNPs from STL-FRQ loci mapped 4/13  
260 signals at single-marker resolution with >50% probability (Table 1). Variants  
261 rs4556017 and rs13162291 were mapped with highest confidence (respective  
262 probabilities 95.1% and 83.5%) and are both associated with eQTLs in multiple

263 tissues (Supplementary Figure S3). In particular, rs4556017 shows eQTLs for the  
264 acetylcholinesterase *ACHE*, and rs13162291 eQTLs for the fatty acid hydroxylase  
265 *FAXDC2*, two genes expressed in enteric and motor neurons (Figure 3). The SNP  
266 rs11240503 is associated with a colon-specific eQTL for *CDK18* (Supplementary  
267 Figure S3), a protein kinase expressed in colonic M cells and BEST4+ enterocytes  
268 (Figure 3). Finally, the rs12273363 marker is associated with eQTLs for a long non  
269 coding antisense RNA (*BDNF-AS*, Supplementary Figure S3) modulating the  
270 expression of the brain-derived neurotrophic factor BDNF.[25] Of note, rs12273363  
271 emerged as top GWAS signal in our meta-analysis ( $P=4.8 \times 10^{-21}$ , Table 1), and was  
272 also associated with consistent genetic effects on CTT measured by the radiopaque  
273 method in a small set of 160 IBS individuals ( $P=0.036$ , with the T allele associated  
274 both with more frequent stools and faster transit; Table 1 and Supplementary Table  
275 5). Although fine mapping did not highlight most likely causative variants at other loci,  
276 their gene content includes candidate genes of known relevance to GI motility and  
277 dysmotility syndromes, like neuropeptides/neurotransmitters and their receptors  
278 (*CALCA/CALCB*, *CRHR1*), ion channels (*KCNJ4*), tight junction proteins (*CLDN15*)  
279 and others (Supplementary Table 2). Genetic variation at these loci also appears to  
280 affect gene expression across several tissues, as evidenced by eQTL analysis  
281 (Supplementary Figure S4).

282

### 283 **Correlations with other disease and traits**

284 A lookup of STL-FRQ association signals in publicly available GWAS data suggested  
285 some of the 13 loci to be relevant to other traits and diseases across multiple  
286 domains, including health outcomes, lifestyle factors and anthropometric traits  
287 (Figure 4A). In particular, the loci tagged by markers rs12273363 and rs2732706

288 showed the largest number of associations, mostly with anthropometric and  
289 psychiatric traits, respectively (Supplementary Table 6).

290 Evidence of genetic correlation with other conditions and traits was obtained from  
291 broader analyses of STL-FRQ GWAS summary statistics using LDSC analysis,  
292 which estimated SNP-based STL-FRQ heritability around 7% ( $h^2_{\text{SNP}}=0.073$ ). When  
293 screening publicly available GWAS data (see Methods), strongest correlation was  
294 observed for IBS ( $r_g=0.42$ , FDR  $P=5.1 \times 10^{-5}$ ), while additional significant findings were  
295 obtained for other gastrointestinal (diverticular disease, use of proton pump inhibitors  
296 - Omeprazole) and psychiatric (anxiety, depression) traits, as well as a number of  
297 traits associated with pain and fatigue (Figure 4B and Supplementary Table 7),  
298 comorbidities often seen in dysmotility syndromes like IBS. Typifying inverse  
299 correlations were also detected for dietary fibers (bran cereals) and laxatives  
300 (dulcolax), which are usually consumed to avoid or relieve constipation, indeed a trait  
301 at or near the root of the STL-FRQ distribution tested here.

302

### 303 **STL-FRQ polygenic scores and irritable bowel syndrome**

304 We further explored the relevance of STL-FRQ GWAS findings to IBS by computing  
305 PGS with PRSice-2 (see Methods) using STL-FRQ GWAS summary statistics. We  
306 studied STL-FRQ PGS in relation to IBS and its subtypes (constipation, IBS-C;  
307 diarrhea, IBS-D and mixed, IBS-M) defined according to gold-standard consensus  
308 Rome III criteria,[26] using data available for a subset of 164,979 UK Biobank  
309 participants who filled a digestive health questionnaire (the same used for the  
310 derivation of STL-FRQ in UK Biobank, Supplementary Methods). PGS distribution  
311 was significantly different in IBS vs asymptomatic individuals, most pronouncedly for  
312 IBS-D (mean PGS 0.456 vs -0.022 in cases and controls;  $P < 1 \times 10^{-300}$ ) (Figure 5).

313 IBS-D prevalence increased from 2.9% in the lowest to 16.4% in the highest PGS  
314 percentile, with individuals showing markedly increased IBS-D risk towards the tail of  
315 PGS distribution (ORs 4.14;  $P=1.2\times 10^{-97}$  for the top 1% and ORs 2.88;  $P=3.1\times 10^{-209}$   
316 for the top 5% of the distribution, respectively; Figure 5 and Supplementary Table 8).  
317 However, although focused on a different trait (IBS), these analyses were performed  
318 on UK Biobank participants also included in the STL-FRQ GWAS (97% overlap),  
319 hence we further tested STL-FRQ PGS in the independent remainder of UK Biobank  
320 (N=291,496), in relation to combined IBS diagnoses available from touchscreen  
321 questionnaire (self-reported) and electronic medical records (ICD10 codes) (see  
322 Supplementary Methods). Although the prevalence of IBS defined by this approach  
323 was much lower (3.3%), and the diarrhea or other IBS subtypes could not be tested,  
324 similar results were obtained, thus replicating Rome-III findings: PGS values were  
325 significantly higher in cases than controls (respective means 0.042 and -0.001;  
326  $P=2.9\times 10^{-5}$ ), IBS prevalence increased across PGS percentiles (3.6-4.2% bottom-top  
327 percentile range) and the risk of IBS was highest in the top 1% of the PGS  
328 distribution (1.29 OR;  $P=6.7\times 10^{-3}$ ; Supplementary Table 8).

## 329 **DISCUSSION**

330

331 We report the results of a powered STL-FRQ GWAS, based on the meta-analysis of  
332 genetic and health-related data from five population-based cohorts. We undertook  
333 this study because of the known correlation between stool frequency and gut motility,  
334 whereas the latter cannot be feasibly studied in humans in numbers large enough for  
335 meaningful genetic investigations. Our approach therefore aimed at revealing  
336 relevant physiological pathways and mechanisms via indirectly measuring GI function  
337 based on suitable questionnaire data on bowel habits. A similar strategy was adopted  
338 in a previous study, however no significant results were obtained, likely due to the  
339 small size of the cohorts analyzed (total N=1281).[27] Studying almost 170,000  
340 individuals, here we identify thousands of associations from 13 independent genome-  
341 wide significant loci, which harbor genes and DNA variants implicating pathways, cell  
342 types and mechanisms plausibly affecting human gut motility in health and disease.  
343 Functional annotation and GWAS-downstream analyses suggest that genes from  
344 STL-FRQ loci are largely involved in neuropeptide and neurotransmitter signaling,  
345 sensory perception and control of motor function in the gut, which provides further  
346 evidence of the validity of our approach. These pathways are notoriously central to  
347 the ENS and its effects on GI motility, enabling bidirectional communication along the  
348 gut-brain axis.[28] Neuropeptides and neurotransmitters regulate gut behavior by  
349 propagating neuronal signals to the mucosal, immune and muscle systems, with  
350 excitatory and inhibitory effects on muscle contraction and peristalsis, among other  
351 functions. Our gene expression and cell-type enrichment analyses provide additional  
352 insight by harnessing the power of single-cell transcriptomics: exploiting RAISIN data  
353 from human colonic muscularis propria, we reveal how STL-FRQ genes are strongly

354 enriched for their expression in enteric neurons, a specific pattern otherwise  
355 undetected at the whole tissue level. In particular, the enrichment appears to be more  
356 pronounced in putative excitatory and inhibitory motor neurons that have been  
357 associated with peristalsis and mechanosensation of gut distention (PEMN and PIMN  
358 subtypes expressing the mechanosensitive ion channel *PIEZO2*).[20] Further  
359 investigation of such expression patterns, coupled with functional characterization of  
360 STL-FRQ genes in specialized cell types and neuronal subtypes, may therefore  
361 contribute important insight into the exact mechanisms underlying neurogenic motor  
362 control in the gut, including dissecting specific classes of ENS neurons into their  
363 respective functional roles. This may aid the development of future therapeutic  
364 strategies to modify GI function and motility.

365 Individual genes most likely candidate to play an important role in the control of STL-  
366 FRQ also clearly point to the involvement of neuropeptide/neurotransmitter signaling  
367 pathways. This is best exemplified by the strongest association we detected in our  
368 GWAS meta-analysis at the *BDNF* locus on chromosome 13, which also replicated in  
369 CTT analyses of IBS patients. The association signal is mapped with relatively high  
370 confidence (>50% probability) to the rs12273363 marker, which is linked to multiple  
371 functional effects on *BDNF* expression: it has eQTL effects on an antisense  
372 transcript (*BDNF-AS*) that induces *BDNF* mRNA degradation,[25] and lies in a  
373 regulatory region previously shown to impart allele-specific, direct repression of  
374 *BDNF* promoter activity (with rs12273363 T a less active repressor).[29] *BDNF* is a  
375 neurotrophin expressed in the central and peripheral nervous systems, with  
376 neurotransmitter modulatory properties and a crucial role in neuronal growth,  
377 differentiation, survival and plasticity.[30] It has also been implicated in several  
378 diseases including major depression, bipolar disorder and other psychiatric

379 conditions.[31] *BDNF* is recognized to influence many important gut functions,  
380 including sensation, motility, epithelial barrier, neuroprotection, and  
381 neuroplasticity.[32] Multiple lines of evidence indicate *BDNF* has prokinetic effects on  
382 gut motility, as shown by impaired peristalsis and delayed GI transit in *BDNF*+/-  
383 mice,[33] increased colonic myoelectric activity in *BDNF*-treated rats,[34] reduced  
384 *BDNF* colonic levels in patients with slow-transit constipation[35] and, notably,  
385 accelerated GI and colonic transit in individuals administered recombinant *BDNF* (r-  
386 metHuBDNF).[36] Hence, our findings are in line with these observations, in that the  
387 rs12273363 T allele associated with more frequent stools and shorter CTT has also  
388 been shown to induce stronger *BDNF* expression (weaker repressor).[29] Altogether,  
389 this suggests a *bona fide* role for *BDNF* in the genetically-determined modulation of  
390 human gut motility, and warrants new analyses of recombinant *BDNF* trials based on  
391 genotype stratification.

392 Our results also point to interesting candidate genes from other STL-FRQ loci where  
393 the association signal has been refined: *ACHE*, *FAXDC2* and *CDK18* all show eQTLs  
394 association with individual variants that have been fine mapped with >50% probability  
395 (respectively, rs4556017 on chromosome 7, rs13162291 on chromosome 5 and  
396 rs11240503 on chromosome 1). *ACHE* codes for an enzyme that hydrolyzes the  
397 neurotransmitter acetylcholine at neuromuscular junctions and is overexpressed in  
398 Hirschsprung's disease,[37] while *FAXDC2* is a hydroxylase of fatty acids whose  
399 luminal concentrations are known to affect gut motility;[38,39] they are both  
400 expressed in enteric and motor neurons and therefore represent ideal functional  
401 candidates. *CDK18* encodes a protein kinase expressed in colonic M cells and  
402 BEST4+ enterocytes specialized in electrolyte and pH sensing,[20,40] hence its

403 associated colon-specific eQTL may be relevant to colonic osmolarity and,  
404 consequently, transit.

405 Finally, strong functional candidates with a well-known role in GI motility map to  
406 additional loci where the association could not be attributed to specific variants.  
407 These involve additional neuropeptides/neurotransmitter systems, including alpha  
408 and beta calcitonin-gene related peptides (*CALCA* and *CALCB* genes) from the  
409 rs6486216 locus,[41] and the corticotropin-releasing hormone receptor (*CRHR1*)  
410 from the rs2732706 locus.[42] Altogether, these and previous observations made for  
411 STL-FRQ candidate causative genes are particularly interesting, in that they may  
412 provide rationale for future translational opportunities in the dysmotility syndromes.

413 Several STL-FRQ associations were previously detected also in other health and  
414 disease-related traits, as from our cross-trait approach to interrogating publicly  
415 available GWAS data. Ten out of thirteen GWAS loci were already linked to lifestyle,  
416 anthropometric and disease-related traits (psychiatric conditions in particular).  
417 Broader evidence of genetic overlap with these conditions came from our LDSC  
418 analyses, which further highlighted shared genetic architecture with gastrointestinal  
419 diseases and often co-morbid neuroaffective traits,[43] among others. This likely  
420 reflects the recognized importance of the gut-brain axis, and suggests our results  
421 may be exploited to gain disease insight in addition to their relevance to better  
422 understanding the physiology of human gut motility. We explored this in relation to  
423 IBS, the most common FGID and the archetype of dysmotility syndromes, which also  
424 showed strongest correlation with STL-FRQ among all traits tested in the LDSC  
425 analysis.

426 Polygenic scores (PGS; calculated by summing multiple alleles weighted by their  
427 effect sizes, usually derived from GWAS studies), are an attractive way to capture an

428 individual's predisposition to develop a specific trait or disease, and hold strong  
429 potential for clinical translation and patient stratification. We computed PGS based on  
430 our STL-FRQ GWAS meta-analysis, and tested them in relation to IBS in the large  
431 UK Biobank cohort. STL-FRQ PGS were significantly higher in IBS cases vs  
432 asymptomatic controls defined according to Rome III criteria available for  
433 approximately 165,000 individuals, as well as in individuals with a doctor's diagnosis  
434 of IBS (self-reported or in their medical records) compared to all other participants in  
435 the remainder of UK Biobank (almost 300,000 people). Individuals from the upper tail  
436 of the PGS distribution were more likely affected by IBS, and exposed to up to >4x  
437 higher risk of IBS-D compared to the rest of the population (in the top 1% of the  
438 distribution). Of note, at least in UK Biobank, the heritability of STL-FRQ  
439 ( $h^2_{\text{SNP}}=0.073$ ) appears to be higher than that of IBS ( $h^2_{\text{SNP}}=0.037$  on the liability  
440 scale, based on previous GWAS data on self-reported IBS).[15] This suggests that,  
441 once refined and further validated in independent cohorts, PGS derived from the  
442 simple STL-FRQ trait may ultimately contribute to an early identification, and  
443 eventual preventive treatment, of individuals at higher risk of developing IBS and  
444 other complex dysmotility syndromes.

445 Finally, our study has a number of limitations: i) stool frequency defined according to  
446 questionnaire data only equates to human gut motility to a certain extent, as its  
447 correlation with GI transit time has been shown to be weaker than, for instance, stool  
448 consistency; ii) current analyses could not take into account likely contributing  
449 environmental factors like diet, medications and others (whose related information  
450 was unavailable in most datasets); iii) relevant cell types and neuronal species have  
451 been identified and further classified here based on gene expression data, hence  
452 functional characterizations may be necessary to confirm specific mechanisms

453 involved in the control of STL-FRQ and motility, as proposed; and finally iv) most  
454 STL-FRQ loci still require conclusive identification of the individual causative gene  
455 and variant(s). These issues can be addressed in future studies, and should  
456 therefore stimulate further investigation as follow-up to the novel findings reported  
457 here.

458 In conclusion, we identify loci harboring prioritized genes with a plausible role in GI  
459 motility, possibly acting via neurotransmission and similar pathways in specialized  
460 enteric neurons. The demonstrated relevance of these findings to IBS warrants  
461 further study for the identification of actionable pathomechanisms in the dysmotility  
462 syndromes.

463 **ACKNOWLEDGEMENTS.** This research has been conducted using the UK Biobank  
464 Resource under Application Number 17435.

465 **AUTHOR CONTRIBUTORSHIP.** MD and AZ: study concept and design; AA, LA,  
466 SW, GA, MP, DAH, NT, JR, AF, NAK, AR, AZ, MS, MC, MD: cohorts, patients  
467 characterization, data collection; FB, XL, CS, AP, AK, RB, TZ, HN, KGE: statistical  
468 analyses; FB, XL, CS, AP, AK, RB, LB, CE, LJ, MP, NT, JR, AF, NAK, AR, AZ, MS,  
469 MC, MD: data analysis and interpretation; MD: obtained funding, administrative and  
470 technical support, study supervision; FB and MD: drafted the manuscript, with input  
471 and critical revision from all other authors.

472 **COMPETING INTERESTS.** None declared.

473 **FUNDING.** Supported by grants from the Swedish Research Council (VR 2017-  
474 02403), the Health Department of the Basque Government (2015111133), and the  
475 Spanish Ministry of Economy and Competitiveness (FIS PI17/00308) to MDA; the  
476 research leading to these results has received funding from the EU FP7 under grant  
477 nr. 313010 (BBMRI-LPC); the FGFP project received support from the Flemish  
478 government (IWT130359), the Research Fund–Flanders (FWO) Odysseus program  
479 (G.0924.09), the King Baudouin Foundation (2012-J80000-004), FP7 METACARDIS  
480 HEALTH-F4-2012-305312, VIB, the Rega Institute for Medical Research, and KU  
481 Leuven. RB is funded by the Research Fund–Flanders (FWO) through a Postdoctoral  
482 Fellowship (1221620N). AZ is supported by the ERC Starting Grant 715772,  
483 Netherlands Organization for Scientific Research NWO-VIDI grant 016.178.056, the  
484 Netherlands Heart Foundation CVON grant 2018-27, and the NWO Gravitation grant  
485 ExposomeNL 024.004.017.

## 486 REFERENCES

487

- 488 1 Scratcherd T, Grundy D. The physiology of intestinal motility and secretion. *Br J Anaesth*  
489 1984;**56**:3–18. doi:10.1093/bja/56.1.3
- 490 2 Quigley EMM. Microflora modulation of motility. *J Neurogastroenterol Motil* 2011;**17**:140–  
491 7. doi:10.5056/jnm.2011.17.2.140
- 492 3 Keller J, Bassotti G, Clarke J, *et al.* Expert consensus document: Advances in the  
493 diagnosis and classification of gastric and intestinal motility disorders. *Nat Rev*  
494 *Gastroenterol Hepatol* 2018;**15**:291–308. doi:10.1038/nrgastro.2018.7
- 495 4 Simrén M, Tack J. New treatments and therapeutic targets for IBS and other functional  
496 bowel disorders. *Nat Rev Gastroenterol Hepatol* 2018;**15**:589–605. doi:10.1038/s41575-  
497 018-0034-5
- 498 5 Manabe N, Wong BS, Camilleri M, *et al.* Lower functional gastrointestinal disorders:  
499 evidence of abnormal colonic transit in a 287 patient cohort. *Neurogastroenterol Motil*  
500 2010;**22**:293–e82. doi:10.1111/j.1365-2982.2009.01442.x
- 501 6 Törnblom H, Van Oudenhove L, Sadik R, *et al.* Colonic transit time and IBS symptoms:  
502 what's the link? *Am J Gastroenterol* 2012;**107**:754–60. doi:10.1038/ajg.2012.5
- 503 7 Knowles CH, Lindberg G, Panza E, *et al.* New perspectives in the diagnosis and  
504 management of enteric neuropathies. *Nat Rev Gastroenterol Hepatol* 2013;**10**:206–18.  
505 doi:10.1038/nrgastro.2013.18
- 506 8 Boeckxstaens G, Camilleri M, Sifrim D, *et al.* Fundamentals of Neurogastroenterology:  
507 Physiology/Motility – Sensation. *Gastroenterology* 2016;**150**:1292–1304.e2.  
508 doi:10.1053/j.gastro.2016.02.030
- 509 9 Camilleri M, Shin A, Busciglio I, *et al.* Genetic variation in GPBAR1 predisposes to  
510 quantitative changes in colonic transit and bile acid excretion. *Am J Physiol Gastrointest*  
511 *Liver Physiol* 2014;**307**:G508–516. doi:10.1152/ajpgi.00178.2014
- 512 10 Camilleri M, Carlson P, Zinsmeister AR, *et al.* Neuropeptide S receptor induces  
513 neuropeptide expression and associates with intermediate phenotypes of functional  
514 gastrointestinal disorders. *Gastroenterology* 2010;**138**:98–107.e4.  
515 doi:10.1053/j.gastro.2009.08.051
- 516 11 Wong BS, Camilleri M, Carlson PJ, *et al.* A Klotho $\beta$  variant mediates protein stability and  
517 associates with colon transit in irritable bowel syndrome with diarrhea. *Gastroenterology*  
518 2011;**140**:1934–42. doi:10.1053/j.gastro.2011.02.063
- 519 12 Fox MR, Kahrilas PJ, Roman S, *et al.* Clinical measurement of gastrointestinal motility  
520 and function: who, when and which test? *Nat Rev Gastroenterol Hepatol* 2018;**15**:568–  
521 79. doi:10.1038/s41575-018-0030-9
- 522 13 Jaruvongvanich V, Patcharatrakul T, Gonlachanvit S. Prediction of Delayed Colonic  
523 Transit Using Bristol Stool Form and Stool Frequency in Eastern Constipated Patients: A  
524 Difference From the West. *J Neurogastroenterol Motil* 2017;**23**:561–8.  
525 doi:10.5056/jnm17022

- 526 14 Saad RJ, Rao SSC, Koch KL, *et al.* Do stool form and frequency correlate with whole-gut  
527 and colonic transit? Results from a multicenter study in constipated individuals and  
528 healthy controls. *Am J Gastroenterol* 2010;**105**:403–11. doi:10.1038/ajg.2009.612
- 529 15 Bonfiglio F, Zheng T, Garcia-Etxebarria K, *et al.* Female-Specific Association Between  
530 Variants on Chromosome 9 and Self-Reported Diagnosis of Irritable Bowel Syndrome.  
531 *Gastroenterology* 2018;**155**:168–79. doi:10.1053/j.gastro.2018.03.064
- 532 16 Benner C, Spencer CCA, Havulinna AS, *et al.* FINEMAP: efficient variable selection  
533 using summary data from genome-wide association studies. *Bioinformatics*  
534 2016;**32**:1493–501. doi:10.1093/bioinformatics/btw018
- 535 17 GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis  
536 Working Group, Statistical Methods groups—Analysis Working Group, *et al.* Genetic  
537 effects on gene expression across human tissues. *Nature* 2017;**550**:204–13.  
538 doi:10.1038/nature24277
- 539 18 Lamparter D, Marbach D, Rueedi R, *et al.* Fast and Rigorous Computation of Gene and  
540 Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput Biol*  
541 2016;**12**:e1004714. doi:10.1371/journal.pcbi.1004714
- 542 19 Smillie CS, Biton M, Ordovas-Montanes J, *et al.* Intra- and Inter-cellular Rewiring of the  
543 Human Colon during Ulcerative Colitis. *Cell* 2019;**178**:714-730.e22.  
544 doi:10.1016/j.cell.2019.06.029
- 545 20 Drokhlyansky E, Smillie CS, Wittenberghe NV, *et al.* The enteric nervous system of the  
546 human and mouse colon at a single-cell resolution. *bioRxiv* 2019;:746743.  
547 doi:10.1101/746743
- 548 21 Buniello A, MacArthur JAL, Cerezo M, *et al.* The NHGRI-EBI GWAS Catalog of published  
549 genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic  
550 Acids Res* 2019;**47**:D1005–12. doi:10.1093/nar/gky1120
- 551 22 Kamat MA, Blackshaw JA, Young R, *et al.* PhenoScanner V2: an expanded tool for  
552 searching human genotype-phenotype associations. *Bioinformatics* 2019;**35**:4851–3.  
553 doi:10.1093/bioinformatics/btz469
- 554 23 Bulik-Sullivan BK, Loh P-R, Finucane HK, *et al.* LD Score regression distinguishes  
555 confounding from polygenicity in genome-wide association studies. *Nature Genetics*  
556 2015;**47**:291–5. doi:10.1038/ng.3211
- 557 24 Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data.  
558 *Gigascience* 2019;**8**. doi:10.1093/gigascience/giz082
- 559 25 Modarresi F, Faghihi MA, Lopez-Toledano MA, *et al.* Inhibition of natural antisense  
560 transcripts in vivo results in gene-specific transcriptional upregulation. *Nat Biotechnol*  
561 2012;**30**:453–9. doi:10.1038/nbt.2158
- 562 26 Drossman DA. The functional gastrointestinal disorders and the Rome III process.  
563 *Gastroenterology* 2006;**130**:1377–90. doi:10.1053/j.gastro.2006.03.008
- 564 27 Jankipersadsing SA, Hadizadeh F, Bonder MJ, *et al.* A GWAS meta-analysis suggests  
565 roles for xenobiotic metabolism and ion channel activity in the biology of stool frequency.  
566 *Gut* 2017;**66**:756–8. doi:10.1136/gutjnl-2016-312398

- 567 28 Tack J, Berghe PV. Neuropeptides and colonic motility: It's all in the little brain.  
568 *Gastroenterology* 2000;**119**:257–60. doi:10.1053/gast.2000.9115
- 569 29 Hing B, Davidson S, Lear M, *et al.* A polymorphism associated with depressive disorders  
570 differentially regulates brain derived neurotrophic factor promoter IV activity. *Biol*  
571 *Psychiatry* 2012;**71**:618–26. doi:10.1016/j.biopsych.2011.11.030
- 572 30 Maisonpierre PC, Belluscio L, Friedman B, *et al.* NT-3, BDNF, and NGF in the  
573 developing rat nervous system: Parallel as well as reciprocal patterns of expression.  
574 *Neuron* 1990;**5**:501–9. doi:10.1016/0896-6273(90)90089-X
- 575 31 Di Carlo P, Punzi G, Ursini G. Brain-derived neurotrophic factor and schizophrenia.  
576 *Psychiatr Genet* 2019;**29**:200–10. doi:10.1097/YPG.0000000000000237
- 577 32 Liu S. Neurotrophic factors in enteric physiology and pathophysiology.  
578 *Neurogastroenterology & Motility* 2018;**30**:e13446. doi:10.1111/nmo.13446
- 579 33 Grider JR, Piland BE, Gulick MA, *et al.* Brain-derived neurotrophic factor augments  
580 peristalsis by augmenting 5-HT and calcitonin gene-related peptide release.  
581 *Gastroenterology* 2006;**130**:771–80. doi:10.1053/j.gastro.2005.12.026
- 582 34 Chai N-L, Dong L, Li Z-F, *et al.* Effects of neurotrophins on gastrointestinal myoelectric  
583 activities of rats. *World J Gastroenterol* 2003;**9**:1874–7. doi:10.3748/wjg.v9.i8.1874
- 584 35 Chen F, Yu Y, Wang P, *et al.* Brain-derived neurotrophic factor accelerates gut motility in  
585 slow-transit constipation. *Acta Physiol (Oxf)* 2014;**212**:226–38. doi:10.1111/apha.12374
- 586 36 Coulie B, Szarka LA, Camilleri M, *et al.* Recombinant human neurotrophic factors  
587 accelerate colonic transit and relieve constipation in humans. *Gastroenterology*  
588 2000;**119**:41–50. doi:10.1053/gast.2000.8553
- 589 37 Moore SW, Johnson G. Acetylcholinesterase in Hirschsprung's disease. *Pediatr Surg Int*  
590 2005;**21**:255–63. doi:10.1007/s00383-005-1383-z
- 591 38 Jin Q, Ren Y, Wang M, *et al.* Novel function of FAXDC2 in megakaryopoiesis. *Blood*  
592 *Cancer Journal* 2016;**6**:e478–e478. doi:10.1038/bcj.2016.87
- 593 39 Neunlist M, Schemann M. Nutrient-induced changes in the phenotype and function of the  
594 enteric nervous system. *J Physiol (Lond)* 2014;**592**:2959–65.  
595 doi:10.1113/jphysiol.2014.272948
- 596 40 Malumbres M, Harlow E, Hunt T, *et al.* Cyclin-dependent kinases: a family portrait.  
597 *Nature Cell Biology* 2009;**11**:1275–6. doi:10.1038/ncb1109-1275
- 598 41 Taché Y, Garrick T, Raybould H. Central nervous system action of peptides to influence  
599 gastrointestinal motor function. *Gastroenterology* 1990;**98**:517–28. doi:10.1016/0016-  
600 5085(90)90849-v
- 601 42 Stengel A, Taché Y. Neuroendocrine control of the gut during stress: corticotropin-  
602 releasing factor signaling pathways in the spotlight. *Annu Rev Physiol* 2009;**71**:219–39.  
603 doi:10.1146/annurev.physiol.010908.163221
- 604 43 Fond G, Loundou A, Hamdani N, *et al.* Anxiety and depression comorbidities in irritable  
605 bowel syndrome (IBS): a systematic review and meta-analysis. *Eur Arch Psychiatry Clin*  
606 *Neurosci* 2014;**264**:651–60. doi:10.1007/s00406-014-0502-z

## 607 **FIGURE LEGENDS**

608

609 **Figure 1. Manhattan plot of STL-FRQ GWAS meta-analysis results.** GWAS  
610 association signals ( $-\log_{10} P$ ) are reported for SNP markers across all chromosomes  
611 shown in alternate gray colors. Significance level corresponding to genome wide  
612 ( $P=5.0 \times 10^{-8}$ ) threshold is indicated with a dashed red horizontal line. For each  
613 independent association signal, the nearest gene (within 100kb, otherwise the lead  
614 SNP) is reported. Genome-wide significant markers are highlighted in green.

615 **Figure 2. Gene set enrichment analysis results.** Top significant findings from  
616 GeneNetwork analysis ranked by  $-\log_{10}$  FDR adjusted P and color-coded according  
617 to KEGG pathways, Gene Ontology biological process (GOBP) and molecular  
618 function (GOMF) categories.

619 **Figure 3. Heatmap of STL-FRQ gene expression in colonic cells.** A selection of  
620 *STL-FRQ* genes is reported for their expression in relevant cell types from colonic  
621 mucosa and colonic muscularis, ordered according to increasing expression. The  
622 heatmap displays  $\log_2(TP10K+1)$  transformed data, and the expression of each gene  
623 is scaled across all cells and shown in color scale ranging from 0 to the 99th data  
624 quantile (to avoid high/low expressed genes dominating the heatmap). ICCs:  
625 interstitial cells of Cajal; PSN: putative sensory neurons; PEMN: putative excitatory  
626 motor neurons; PIMN: putative inhibitory motor neurons; PIN: putative interneurons;  
627 PSVN: secretomotor/vasodilator neurons. Cell types, neuron types and subtypes are  
628 classified as defined previously in Drokhyansky et al. (Ref [20]).

629 **Figure 4. Cross-trait analysis of STL-FRQ GWAS results.** A) Circus plot showing  
630 previously reported genome-wide significant associations ( $P=5.0 \times 10^{-8}$ ) at the 13  
631 *STL-FRQ* loci (lead SNPs or  $r^2 > 0.8$  LD proxies, see Methods). Associations are

632 grouped by category, and the ribbon size is proportional to the number of associated  
633 traits in that category. In order to avoid redundancy, for multiple markers (lead or  
634 proxy SNP) linked to the same trait only the one with the lowest P is reported. No  
635 association were detected for 3 loci. B) Results obtained in the LDSC analyses of  
636 genetic correlation, in relation to a selection of most relevant traits (full results  
637 reported in Supplementary Table 7).

638 **Figure 5. STL-FRQ polygenic scores (PGS) and IBS in UK Biobank.** Results are  
639 reported (including statistical significance) in relation to PGS distribution in IBS cases  
640 vs controls (left panels, P-values from t-test), and in relation to the prevalence of IBS  
641 across PGS percentiles in the entire cohort (right panels; with top 5% of the  
642 distribution highlighted with shaded area, P-values vs the rest of the cohort from  
643 logistic regression). IBS and subtypes defined according to Rome III Criteria based  
644 on DHQ questionnaire data (see Methods).

**Table 1. STL-FRQ GWAS meta-analysis and fine mapping results**

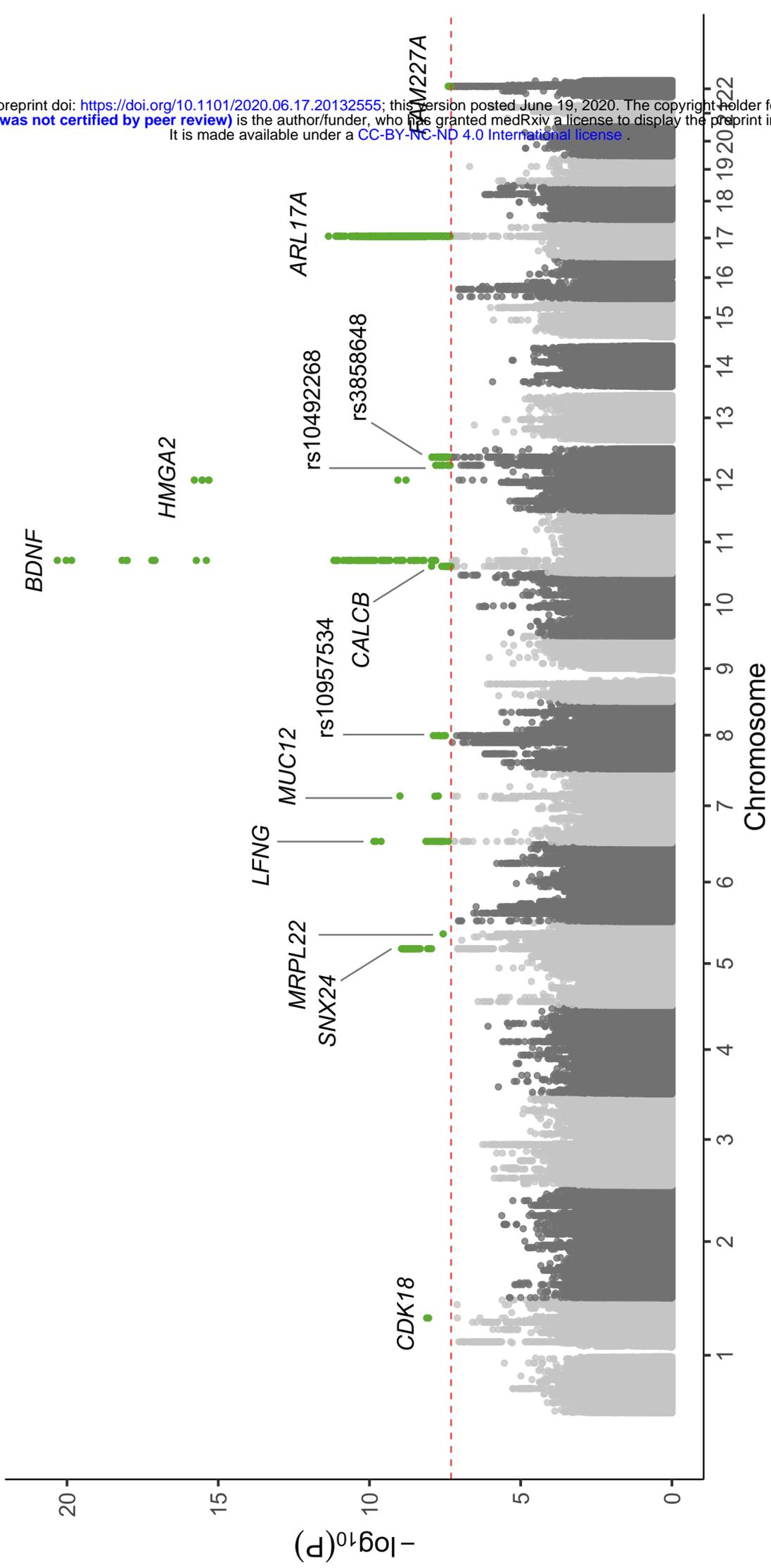
Chr	Lead SNP	Start-end (bp)	EA	OA	EAF	Beta (SE) #	P	Nearest gene (other genes) *	Most likely causal SNP (% probability) ^
1	rs11240503	205469956-205485290	A	G	0.300	0.018 (0.003)	7.8E-09	CDK18 (5)	<u>rs11240503</u> (0.588)
5	rs39819	122032544-122636855	A	G	0.671	0.018 (0.003)	1.2E-09	SNX24 (3)	-
5	rs13162291	154119471-154448827	A	G	0.191	0.020 (0.004)	2.7E-08	MRPL22 (5)	<u>rs13162291</u> (0.835)
7	rs12700026	2554037-2605424	A	C	0.890	-0.029 (0.005)	1.4E-10	LFNG (5)	rs12700026, rs12700027 (0.350)
7	rs4556017	99919517-100678086	T	C	0.853	0.024 (0.004)	1.0E-09	MUC12 (46)	<u>rs4556017</u> (0.951)
8	rs10957534	71482998-72012331	C	G	0.367	-0.016 (0.003)	1.3E-08	(5)	-
11	rs6486216	14980848-15127148	T	C	0.276	0.018 (0.003)	1.1E-08	CALCB (5)	-
11	rs12273363	27455582-27748493	T	C	0.795	0.032 (0.003)	4.8E-21	BDNF (3)	<u>rs12273363</u> (0.525)
12	rs11176001	66317487-66410673	A	C	0.132	0.034 (0.004)	1.6E-16	HMGA2 (1)	rs11176001 (0.392)
12	rs10492268	98298807-98506148	T	C	0.552	0.016 (0.003)	1.6E-08	(1)	rs10492268 (0.187)
12	rs3858648	115861753-115950227	A	C	0.508	-0.016 (0.003)	1.2E-08		rs3858648 (0.077)
17	rs2732706	43460181-44865603	T	C	0.221	0.024 (0.003)	4.4E-12	ARL17A (109)	-
22	rs5757162	38869463-39152412	T	C	0.286	0.017 (0.003)	4.0E-08	FAM227A (21)	-

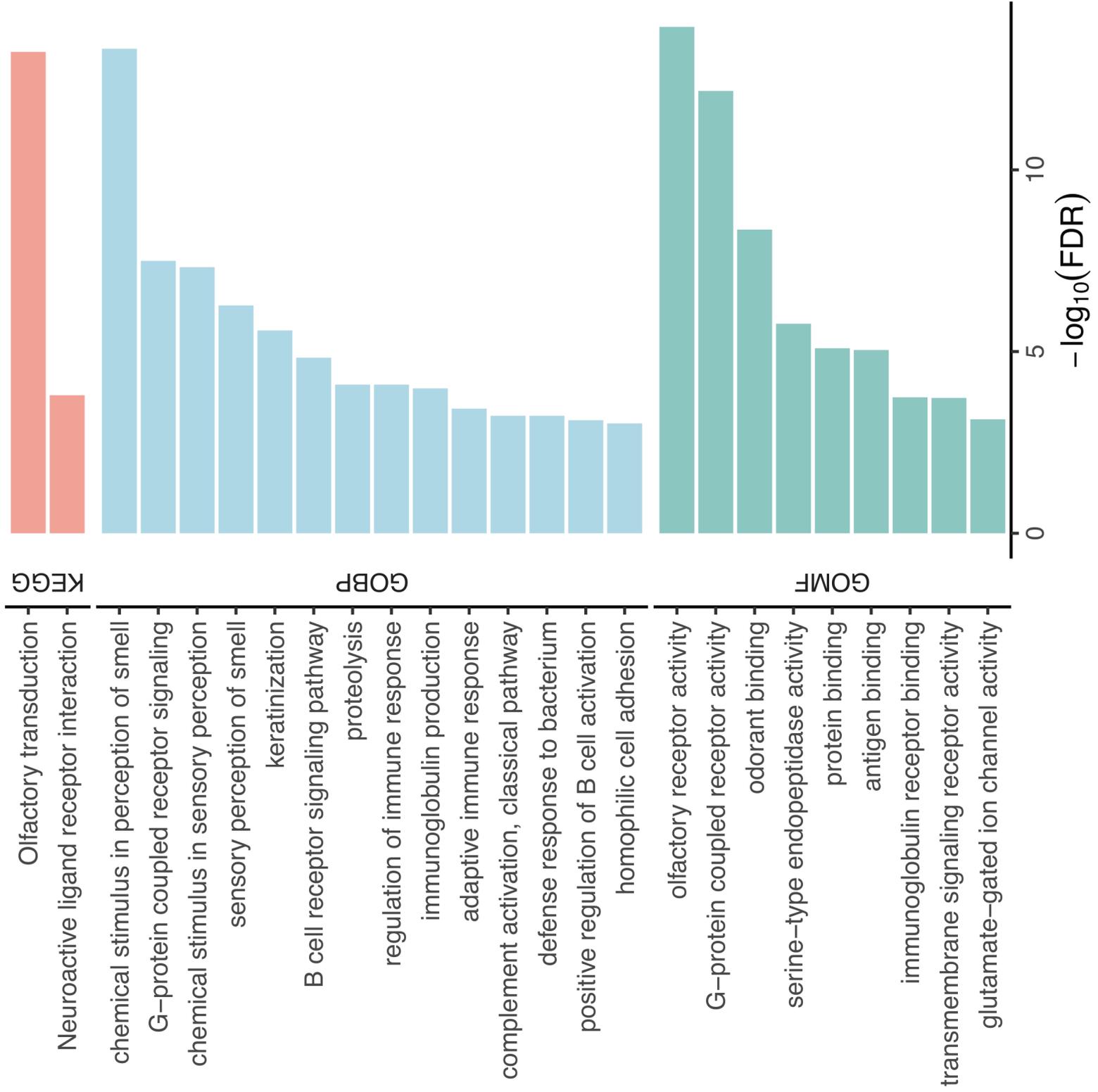
Chr: chromosome, EA: effect allele; OA: other allele; EAF: effect allele frequency

# positive beta = higher stool frequency

\* nearest gene (within 100kb from lead SNP) and other genes in the region, based on FUMA positional and eQTL mapping

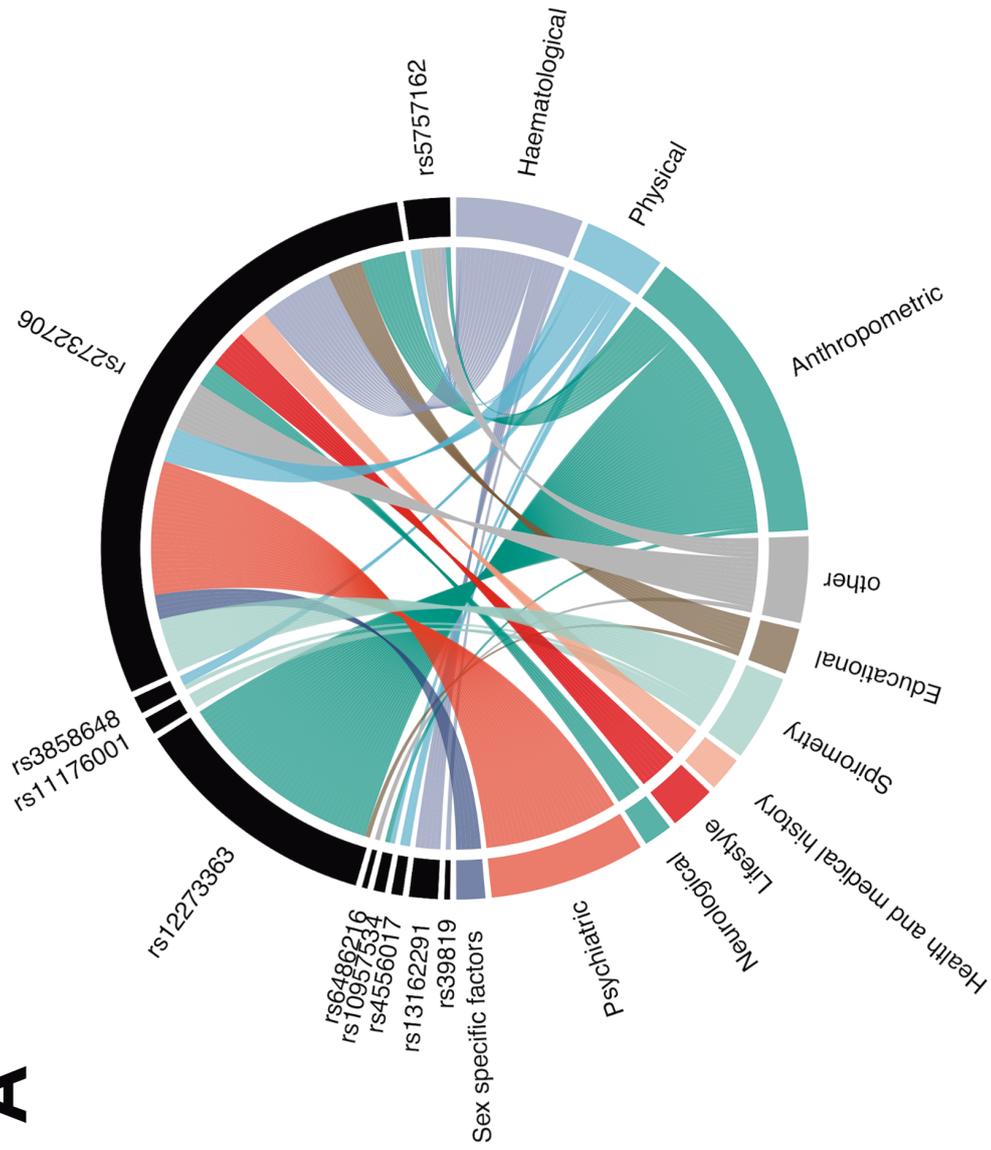
^ only causal SNPs identified with >5% probability are reported (>50% underlined)







**A**



**B**

