Check for updates

OPEN

Genome-wide analysis of 53,400 people with irritable bowel syndrome highlights shared genetic pathways with mood and anxiety disorders

Chris Eijsbouts [1,2], Tenghao Zheng [3,4,86], Nicholas A. Kennedy [5,86], Ferdinando Bonfiglio [3,4], Carl A. Anderson [6], Loukas Moutsianas [6,7], Joanne Holliday [8], Jingchunzi Shi [9], Suyash Shringarpure [9], 23andMe Research Team*, Alexandru-Ioan Voda [10,11], The Bellygenes Initiative*, Gianrico Farrugia [12], Andre Franke [13], Matthias Hübenthal [13,14], Gonçalo Abecasis [15], Matthew Zawistowski [15], Anne Heidi Skogholt [16,17], Eivind Ness-Jensen [17,18,19], Kristian Hveem [17], Tõnu Esko [20], Maris Teder-Laving [20], Alexandra Zhernakova [21], Michael Camilleri [22], Guy Boeckxstaens [23], Peter J. Whorwell [24], Robin Spiller [25], Gil McVean [1], Mauro D'Amato [3,4,13,26,27,28,87 ✉], Luke Jostins [1,10,29 ✉] and Miles Parkes [30,31 ✉]

Irritable bowel syndrome (IBS) results from disordered brain–gut interactions. Identifying susceptibility genes could highlight the underlying pathophysiological mechanisms. We designed a digestive health questionnaire for UK Biobank and combined identified cases with IBS with independent cohorts. We conducted a genome-wide association study with 53,400 cases and 433,201 controls and replicated significant associations in a 23andMe panel (205,252 cases and 1,384,055 controls). Our study identified and confirmed six genetic susceptibility loci for IBS. Implicated genes included *NCAM1*, *CADM2*, *PHF2/ FAM120A*, *DOCK9*, *CKAP2/TPTE2P3* and *BAG6*. The first four are associated with mood and anxiety disorders, expressed in the nervous system, or both. Mirroring this, we also found strong genome-wide correlation between the risk of IBS and anxiety, neuroticism and depression ($r_g > 0.5$). Additional analyses suggested this arises due to shared pathogenic pathways rather than, for example, anxiety causing abdominal symptoms. Implicated mechanisms require further exploration to help understand the altered brain–gut interactions underlying IBS.

BS is common worldwide and typically presents in early adulthood with symptoms including abdominal pain, bloating and bowel dysfunction[1]. Symptom intensity varies over time and between individuals but IBS has been reported, in severe cases, to affect quality of life as much as renal impairment or diabetes[2].

IBS accounts for approximately half of all referrals to gastroenterology clinics[3]. Although of doubtful clinical value, many patients undergo multiple investigations, including colonoscopies, to exclude conditions such as Crohn's disease or cancer. To the frustration of patients and clinicians alike, all tests are characteristically normal. The healthcare costs, combined with indirect employment costs, of IBS amount to at least €13 billion (£11.7 billion) annually in Europe[4].

Attempts have been made to identify positive clinical diagnostic features and reduce investigations. The widely used and recently revised Rome III criteria define IBS as unexplained abdominal pain or discomfort eased by defecation, with altered stool form or frequency, for more than six months[5]. Three main subtypes are recognized: constipation-predominant (IBS-C), diarrhea-predominant (IBS-D) or 'mixed'/alternating constipation and diarrhea (IBS-M). Individuals with functional constipation and functional diarrhea share the disordered bowel pattern with IBS but do not suffer from pain.

The commonly used IBS treatments, ranging from dietary exclusion to psychoactive medications, are relatively ineffective and their variety reflects the uncertain etiology[6]. Behavioral therapies, while

more effective (number needed to treat = 4), are not widely available[7]. Given the high frequency, impact and cost of IBS, there is a pressing need for improved pathophysiological understanding to enable better therapeutic approaches.

The relative importance of peripheral and gut versus central and psychological factors to IBS etiology is uncertain. The consensus view is that IBS results from abnormal brain–gut interactions. Recent epidemiological data suggested that, in individuals developing both IBS and psychological features, the former preceded the latter in two thirds of cases and the latter preceded the former in one third[8]. IBS is associated with abnormalities of central pain processing but also increased gut permeability, mast cell activation, disordered motility and dysbiosis[9]. Up to one in ten cases are triggered after infection, so-called postinfectious IBS (PI-IBS)[10].

IBS aggregates in families, with individuals being two to three times more likely to develop IBS if they have an affected relative. Estimates of heritability from twin studies range widely from 0 up to 57%[11]. Twin studies have indirectly investigated whether IBS and mental health conditions share a genetic basis but have proved inconclusive[12]. Although genetic association studies have provided pathogenic and therapeutic insights for many conditions[13], only one variant (rs10512344) has previously been identified at genome-wide significance in IBS, with only modest replication[14]. Larger datasets are clearly required to characterize the recognized heritable

A full list of affiliations appears at the end of the paper.

component of IBS; the use of large population-based cohorts had been proposed[11,15]. Broad meta-analysis of cases from different cohorts and case definitions is also a proven method of increasing power[16]. Between 2006 and 2010, UK Biobank (UKB) recruited half a million people aged 40–69 years. Participants underwent baseline assessment and consented to long-term follow-up, including questionnaires, and linkage to routinely collected health data. All participants underwent genome-wide SNP genotyping. Therefore, UKB provides a powerful epidemiological resource for exploring risk factors for health outcomes[17,18].

The main aim of the current study was to identify genetic risk factors for IBS through an analysis involving over 250,000 affected individuals. We report robustly validated genetic susceptibility loci for IBS and provide evidence of its shared genetic etiology with mood and anxiety disorders.

Results

Epidemiology. We designed a digestive health questionnaire (DHQ) for the UKB website, with a link e-mailed to 332,793 UKB participants with valid e-mail addresses. A total of 171,061 (51.4%) responses were received and analyzed (Supplementary Fig. 1). The DHQ included validated instruments for IBS diagnosis (Rome III symptom criteria), the IBS symptom severity score (IBS-SSS, measured using the IBS Severity Scoring System)[19] and the Patient Health Questionnaire 12 Somatic Symptom score (PHQ-12)[20]. It also asked about previous IBS diagnosis, environmental exposures and associated conditions (including anxiety or depression, based on treatment sought or offered).

After sample exclusions and quality control, we identified a total of 40,548 UKB participants of European ancestry (Fig. 1a) who met the diagnostic criteria for IBS (based on DHQ Rome III symptom data, self-report of previous medical IBS diagnosis or electronic medical records; see Methods and Supplementary Tables 1 and 2). The demographics of the DHQ respondents are presented in Table 1. Females were affected by IBS more commonly than males (72.1%). IBS-M, with hard and loose stools present at least 'sometimes' (alternating), was the most common subtype in patients defined according to the Rome III criteria (55.0%).

A total of 24,845 respondents reported current abdominal symptoms meeting standard diagnostic criteria (DHQ Rome III, Fig. 1) at the time of the survey, providing a point prevalence of IBS of 14.5%. Of these, only 8,836 (35.6%) had a hospital-documented IBS diagnosis, via an IBS ICD-10 code (International Statistical Classification of Diseases and Related Health Problems, 10th revision), or reported having been diagnosed with IBS by a doctor (DHQ self-report and/or unprompted self-report). They reported greater gastrointestinal symptom severity (quantified via IBS-SSS) than those not medically diagnosed (odds ratio (OR) and 95% confidence interval (CI) = 1.07 (1.07–1.08) per IBS-SSS unit); Fig. 1b and Supplementary Table 3).

Risk factors and associated conditions. As reported previously, a family history of IBS was more common in cases than controls (24.0 versus 9.5%, OR and 95% CI = 3.73 (3.60–3.88)). However, birth by Cesarean section was not (2.6 versus 2.5%, OR and 95% CI = 1.02 (0.94–1.11); Table 1). A significantly higher proportion of cases with IBS recalled receiving long-term or recurrent antibiotics during childhood compared with controls (20.0 versus 9.6%, OR and 95% CI = 2.22 (2.13–2.30)). The severity of current IBS symptoms correlated positively with recalled childhood antibiotic exposure (OR and 95% CI = 1.04 (1.04–1.04) per IBS-SSS unit) and family history (OR and 95% CI = 1.05 (1.05–1.06)). Participants with anxiety also reported increased antibiotics use in childhood (18.4%, OR and 95% CI = 1.64 (1.59–1.70); Supplementary Table 4).

Regarding comorbidities, as documented previously, the rates of appendicectomy, cholecystectomy and hysterectomy were all increased in IBS (Supplementary Table 5), as were the rates of atopic disease (Table 1). Anxiety and depression were each approximately twice as common (Table 1 and Supplementary Table 4); 34.3% of cases reported treatment for anxiety compared with 16.1% of controls. This effect was more prominent in individuals medically diagnosed with IBS.

The median PHQ-12 score was 4 (interquartile range (IQR) = 2–6) in controls and 6 (IQR = 4–9) in pooled cases (7 in all four constituent subgroups, Supplementary Table 6; see also 'Median scores among pooled and individual diagnoses' in the Supplementary Note). The PHQ-12 score correlated with IBS symptom severity (Pearson's correlation = 0.40 (95% CI = 0.39–0.41) among 31,402 IBS cases completing all PHQ-12 and IBS-SSS questions), with back pain, limb pain and tiredness driving this association (Fig. 1c and Supplementary Figs. 2 and 3). Among UKB participants with previous generalized anxiety disorder-7 (GAD-7) scores (n = 79,430; Supplementary Table 7) or Patient Health Questionnaire-9 (PHQ-9) depression scores (n = 79,087, Supplementary Table 8; see Supplementary Note for definitions), these scores were consistently higher in cases with IBS than controls (OR and 95% CI = 1.14 (1.14–1.15) per GAD-7 unit and 1.15 (1.15–1.16) per PHQ-9 unit) and correlated with IBS-SSS (Pearson's correlations and 95% CIs among cases = 0.24 (0.22–0.25) and 0.27 (0.25–0.28), respectively; Fig. 1d and Supplementary Fig. 4).

The respective prevalence for functional constipation and functional diarrhea (that is, bowel disturbance without abdominal pain or discomfort; see 'Definitions of IBS cases' in the Supplementary Note) were 6.4% and 11.7% (Table 1). Somatic symptoms (PHQ-12) and treatment for anxiety or depression were less strongly associated with functional diarrhea than with IBS-D (excess OR and 95% CI = 1.24 (1.22–1.26) per PHQ-12 unit and 1.58 (1.46–1.72), respectively), with similar effects for functional constipation and IBS-C (Table 1).

Genetics. We identified six independent IBS susceptibility loci at genome-wide significance (P < 5 × 10⁻⁸) in a discovery cohort totaling 53,400 cases and 433,201 controls (Fig. 2a and Supplementary Fig. 5). This resulted from pooling IBS cases across all case definitions to maximize power, in a meta-analysis of data from UKB (40,548 cases and 293,220 controls; Supplementary Tables 1 and 2) and the international collaborative Bellygenes initiative (12,852 cases and 139,981 controls; Methods and Supplementary Table 9). Using data from an independent panel from 23andMe (Supplementary Note), all six loci were replicated at Bonferroni significance (P < 0.0083) with the same direction of effect (Table 2). All were found on autosomal chromosomes (none on the X chromosome) and conferred modest ORs < 1.05. Three out of six loci also had reported associations with mood and anxiety disorders and related phenotypes[21–25].

We undertook genetic fine-mapping to establish plausible causal variants (Supplementary Fig. 5) and used several techniques to identify candidate causal genes within IBS risk loci (Supplementary Table 10; see 'Gene mapping' in the Supplementary Note). Among the genes implicated (Table 2) were two encoding neural adhesion molecules: neural cell adhesion molecule 1 (*NCAM1*) and cell adhesion molecule 2 (*CADM2*). Ranking tissues according to enrichment for risk gene expression (Supplementary Fig. 6), the brain came top of the list (LDSC applied to specifically expressed genes[26] coefficient = 8.32 × 10⁻¹⁰, s.e.m. = 4.5 × 10⁻¹⁰, P = 0.03). However, this result was not statistically significant after correcting for multiple testing, which may in part reflect lack of power due to low SNP heritability. Using expression colocalization analysis as a separate method to implicate specific gene–tissue combinations, we found evidence that the six IBS-associated variants regulate gene expression across a number of tissues, with many genes particularly expressed in the brain (Fig. 2b).
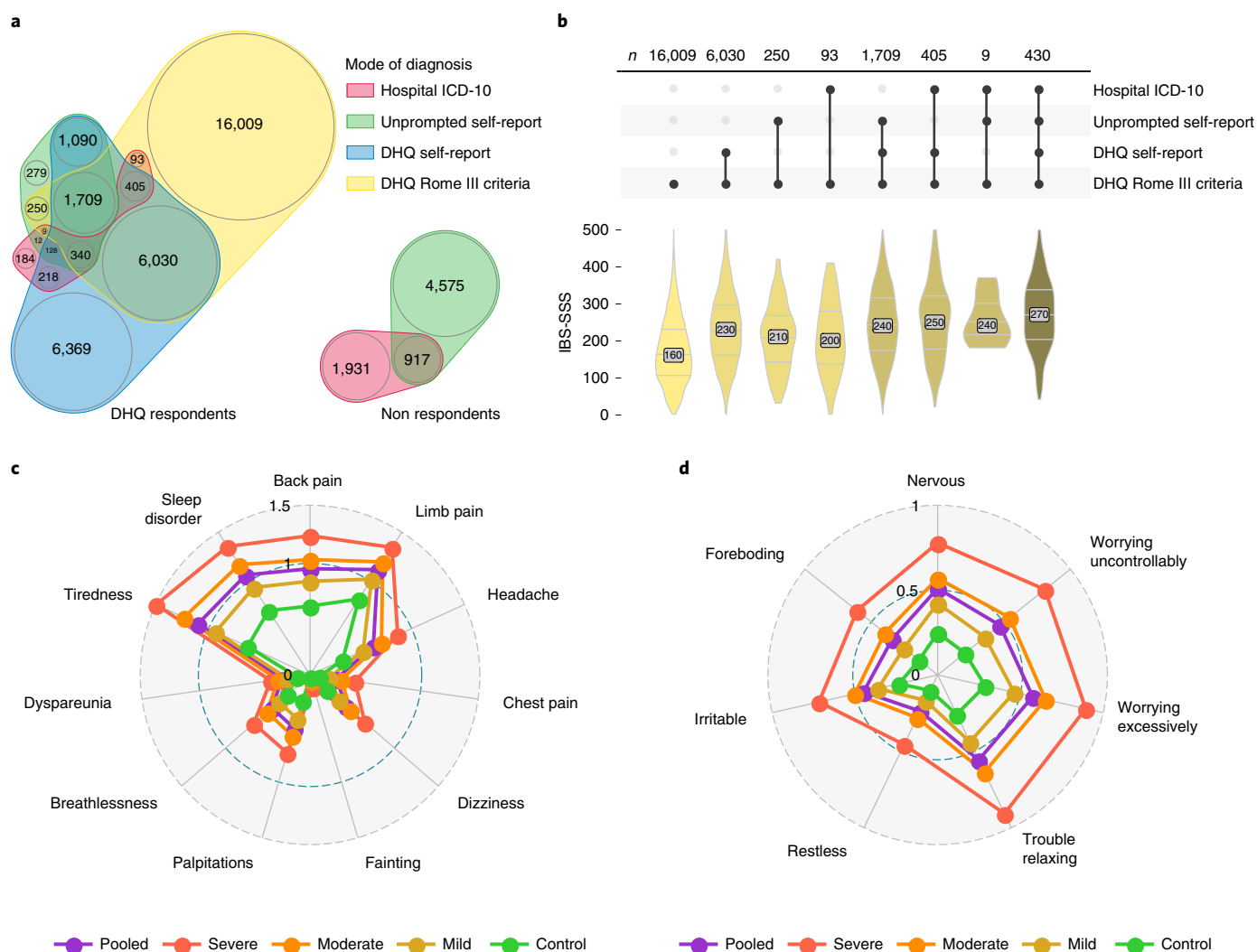
Fig. 1 | Diagnostic modalities and comorbidities of IBS. a, Venn diagram of overlap between UKB IBS cases by different diagnostic modality, split by DHQ respondents and nonrespondents. The areas and numbers indicate the sample size. Most participants with current symptoms (DHQ Rome III, yellow) did not report being diagnosed with IBS either when listing medical conditions unprompted at UKB enrollment (unprompted self-report, green) or when asked specifically about a previous IBS diagnosis when completing the DHQ (DHQ self-report, blue). Conversely, many participants previously diagnosed with IBS, even those formally recorded during a hospital admission (hospital ICD-10, red), did not have symptoms sufficient for Rome III criteria IBS diagnosis at the time of their DHQ response. **b**, Among individuals experiencing IBS symptoms (DHQ Rome III positive), those previously diagnosed by a clinician had greater symptom severity, with an increase in the number of IBS diagnostic modalities (connected dots, middle; top: sample size is shown) being associated with an increase in symptom severity score (IBS-SSS, bottom). Distributions are colored by the number of diagnoses and the groups shown are mutually exclusive. For post-hoc statistics, see Supplementary Table 3. **c**, Severity of different somatic symptoms in the past three months among digestively healthy controls and IBS cases (classified as mild, moderate and severe based on IBS-SSS). Mean scores for PHQ-12 items ranked from 0 (not bothered at all) to 2 (bothered a lot) are shown. Pooled refers to all UKB cases in the discovery cohort. **d**, As above, for symptoms of anxiety in the last two weeks, measured using average scores for GAD-7 items ranked from 0 (never bothered) to 3 (bothered nearly every day).

One association mapped to the major histocompatibility complex (MHC) class 3 region close to BAG cochaperone 6 (*BAG6*). The signal is not driven by human leukocyte antigen (HLA) alleles and is independent of known MHC associations with ulcerative colitis, celiac disease or microscopic colitis (Supplementary Fig. 7 and Supplementary Tables 11 and 12) (refs. [27–30]). It is also independent of lead variants for neuroticism at this locus (highest $r^2 = 0.51$)[23].

Eight additional loci showed genome-wide significant association with various IBS definitions (Methods) but not the whole discovery cohort, of which five were replicated in the 23andMe data (Supplementary Fig. 8 and Supplementary Tables 13 and 14). These require further study. The female-specific signal identified previously[14] for unprompted self-reported IBS in the UKB was also observed in our female-specific analysis of unprompted

self-reported data but was not detected in female-specific analyses of any other case definitions from UKB or Bellygenes initiative, nor replicated in the 23andMe unstratified analyses of both sexes (Supplementary Table 15), possibly suggesting survey-specific factors playing a role. Specific candidate gene associations previously reviewed in the literature[31,32] also did not show significant evidence of association after multiple testing correction (all $P > 0.015$).

LDSC estimated a modest but significant genome-wide SNP heritability for IBS of 5.77% (s.e.m. = 0.35%) in the discovery cohort, with no evidence of population stratification (LDSC intercept = 0.9951, s.e.m. = 0.007). This was consistent across case definitions within UKB (h^2 range of 5.42–7.71%), with similar values seen in the Bellygenes ($h^2 = 3.14\%$, s.e.m. = 0.74%) and 23andMe cohorts ($h^2 = 5.39\%$, s.e.m. = 0.02%).

Table 1 | Demographics, symptom severity, family history and associated conditions for IBS patients diagnosed via different modalities and with different IBS subtypes

Group	n	Male (%)	Female (%)	Age (years)	Age (s.d.)	Mean IBS-SSS	Mean PHQ-12	Mean GAD-7	Family history of IBS	Childhood antibiotic exposure	Born by Cesarean section	Treatment for anxiety offered or sought	Treatment for depression offered or sought	Pooled asthma/ hay fever/ eczema
						Scores among DHQ participants completing the relevant sections			**Among DHQ participants (%)**					**Atopy (%)**
Controls (DHQ respondents)	72,356	50.3	49.7	65.3	7.5	33	4.0	1.5	9.5	9.6	2.5	16.1	18.0	18.1
Hospital ICD-10	4,237	23.4	76.6	65.9	7.9	202*	7.4*	3.6*	31.8*	24.0*	2.5	43.0*	43.6*	29.3*
Unprompted self-report	9,309	25.7	74.3	65.2	8.0	196*	7.1*	3.3*	30.7*	19.8*	2.3	41.1*	40.0*	29.6*
DHQ, Rome III criteria	24,845	27.9	72.1	63.9	7.7	194*	7.1*	3.3*	24.3*	20.4*	2.6	33.7*	35.1*	24.0*
DHQ, self-report	16,289	26.2	73.8	64.0	7.7	177*	6.9*	3.1*	29.0*	21.6*	2.7	39.7*	39.3*	25.9*
Pooled (any of the four definitions above)	40,548	27.9	72.1	64.3	7.8	173*	6.8*	3.1*	24.0*	20.0*	2.6	34.3*	35.2*	24.8*
DHQ, Rome III criteria, type C	3,989	16.9	83.1	64.5	7.8	190*	6.8*	3.2*	22.2*	19.2*	2.2	32.3*	34.1*	21.4*
DHQ, Rome III criteria, type D	6,506	33.2	66.8	63.8	7.7	185*	6.5*	3.0*	23.8*	19.1*	2.8	32.1*	33.0*	24.0*
DHQ, Rome III criteria, type M	13,666	28.5	71.5	63.8	7.8	204*	7.5*	3.6*	25.6*	21.5*	2.6	35.2*	36.9*	24.8*
DHQ, Rome III criteria, type U	672	31.0	69.0	65.6	7.5	124*	5.6*	2.4*	16.7*	17.7*	2.8	27.5*	24.9*	23.8*
DHQ, postinfectious	860	24.2	75.8	62.2	7.8	196*	7.4*	3.6*	30.3*	27.7*	3.5	42.0*	42.4*	26.6*
DHQ, functional constipation	3,502	33.9	66.1	67.1	7.4	64*	4.7*	1.8*	12.5*	11.0	2.5	23.1*	25.4*	19.0
DHQ, functional diarrhea	5,386	61.0	39.0	65.1	7.5	49*	4.3*	1.8*	11.0*	10.7*	2.6	20.3*	22.0*	20.5*

Digestively healthy controls and functional constipation as well as diarrhea groups are shown for reference. Gastrointestinal symptoms are captured by the IBS-SSS (range 0–500), while somatic symptoms are captured by the (modified) PHQ-12 (range 0–22). The GAD-7 score captures symptoms of anxiety (range 0–21). The single asterisk marks significant differences from the control group after adjusting for age, sex, DHQ participation and (Bonferroni) multiple testing at $P < 0.05/108$ (two-sided logistic regression test). Age and sex differences were not tested.

IBS-C showed weak genetic correlation with functional constipation, as did IBS-D with functional diarrhea (Supplementary Fig. 9). IBS-C and IBS-D correlated with each other but there were no cross-correlations, that is, IBS-C did not correlate with functional diarrhea. Heritability for the IBS subtypes was comparable with IBS overall; IBS subtypes showed similar genetic correlation with mental health and personality traits (Supplementary Table 16).

We compared the overlap between susceptibility with IBS and 751 other traits and diseases listed in the LD Hub[33]. The strongest correlations in genome-wide risk were with mood and anxiety disorders and related phenotypes, including anxiety ($r_g = 0.58$, s.e.m. $= 0.10$), neuroticism ($r_g = 0.54$, s.e.m. $= 0.04$), depression ($r_g = 0.53$, s.e.m. $= 0.05$) and insomnia ($r_g = 0.42$, s.e.m. $= 0.05$)[33]. Across the genome, the same alleles that predisposed to IBS also predisposed to mood and anxiety disorders. The correlations were consistent regardless of the mode of diagnosis of anxiety or depression (Supplementary Fig. 10) (refs. [34,35]). We calculated phenotypic correlations for these traits on a comparable liability scale (Fig. 3 and Supplementary Table 16). Mostly, the phenotypic and genotypic correlations mirrored each other, although genetic correlations were often larger. Notably, other digestive diseases presenting with similar symptoms, including celiac disease ($r_g = 0.03$, s.e.m. $= 0.08$,

$P = 0.69$) and Crohn's disease ($r_g = 0.08$, s.e.m. $= 0.04$, $P = 0.06$), were not genetically correlated with IBS.

We also ran higher-specificity (IBS cases meeting at least 2 of the 4 UKB case definitions, 11,201 cases and 293,220 controls) and high-severity (IBS-SSS > 300, 4,296 cases and 72,356 controls) analyses in UKB. The former produced no new associations. The latter, while being more heritable (liability scale $h^2 = 0.42$, s.e.m. $= 0.05$, Cochran's $Q = 51.7$, $P = 6.31 \times 10^{-13}$ compared with the discovery cohort IBS), produced one association (rs9947289, $P = 2.80 \times 10^{-8}$) that did not replicate ($P = 0.57$ in the 23andMe data; Supplementary Table 13). Both of these phenotypes recapitulated the same genetic correlation with mood and anxiety disorders as found in the discovery cohort (Supplementary Fig. 11).

To explore the role of shared genetic risk versus direct phenotypic overlap, we compared the genome-wide association study (GWAS) results for IBS having removed participants with anxiety to the GWAS results for anxiety having removed participants with IBS (for anxiety definitions, see Supplementary Tables 17 and 18). The genetic correlation between IBS and anxiety attenuated but remained strong ($r_g = 0.31$, s.e.m. $= 0.06$; Supplementary Fig. 12). We next used bidirectional Mendelian randomization[36] with an independent anxiety GWAS[37], as well as genome-wide latent vari-
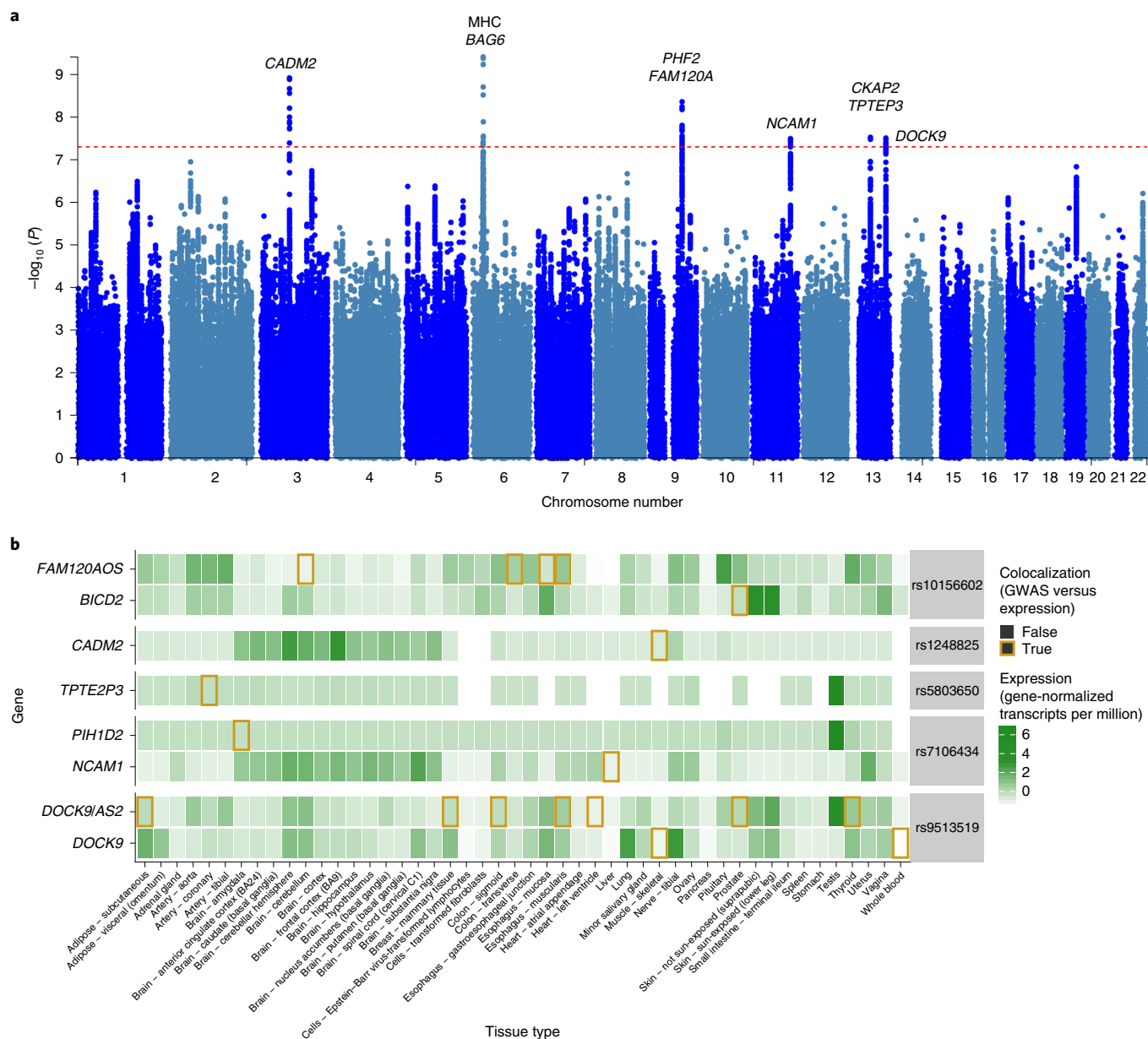
Fig. 2 | Genome-wide association results for IBS. a, Manhattan plot showing the distribution of IBS-associated SNPs across the genome. The dashed line indicates the genome-wide significance threshold at $P = 5 \times 10^{-8}$. *P* values are from a two-sided test, after inverse variance-weighted fixed-effects meta-analysis. See Supplementary Fig. 5 for zoomed-in visualizations with linkage disequilibrium data, posterior probabilities of causality and transcript annotations. **b**, Expression of genes implicated in IBS etiology through causal effects of associated variants on gene expression across a range of tissues. Darker green indicates higher gene expression and the golden outlines indicate tissue–gene combinations where the IBS-associated variant was known to influence gene expression in this tissue (colocalization posterior > 0.5). Genes with at least one colocalization event across tissues with expression quantitative trait locus data from the GTEx dataset are shown.

able Mendelian randomization[38], to explore directionality. Multiple models could explain our data (Supplementary Table 19) but they were best explained by shared genetic risk pathways rather than causal effects between the two traits. Similar complex causal relationships were evident between IBS and mental health and personality traits other than anxiety (Supplementary Table 19).

Discussion

The importance of this study lies in its scale and therefore the robustness of its genetic results. We have identified replicable genetic associations for IBS, providing new biological insights, while demonstrating that overall its heritability is modest. Two

observations are particularly striking: the genetic overlap between IBS and mood and anxiety disorders and the lack of signals implicating genes expressed specifically in the gut or overlapping other intestinal disorders. Our findings suggest that, with respect to the genetically determined risk for IBS, neuronal pathways play a dominant role.

Increasing abdominal symptom severity correlated with increasing PHQ-12 somatic symptom scores, particularly for the domains of tiredness, back pain, limb pain and headache (Fig. 1). Multifocal pain suggests either poor coping skills, perhaps relating to psychological comorbidity, or visceral hypersensitivity from aberrant antinociceptive mechanisms[39]. By contrast, the painless bowel disorders,

Table 2 | Variants associated with IBS, their effect measured in the discovery cohort and *P* values for association in the discovery cohort, the replication cohort and the meta-analysis of these two

	Variant				Effect (discovery)		*P* values			Annotation	
SNP	Chromosome number	Position	Alleles	Frequency	OR	95% CI	Discovery	Replication	Meta-analysis	Mapped gene	Previously implicated in
rs1248825	3	84,993,411	C/A	0.33	1.05	1.03–1.07	1.20×10^{-9}	4.90×10^{-8}	7.48×10^{-15}	*CADM2*	Personality traits (risk-taking, neuroticism, anxiety)[21], cannabis use[22]
rs2736155	6	31,605,199	G/C	0.48	1.05	1.02–1.07	3.88×10^{-10}	8.28×10^{-6}	3.19×10^{-12}	*BAG6*	
rs10156602	9	96,345,328	G/A	0.63	1.04	1.02–1.06	4.36×10^{-9}	1.18×10^{-8}	3.04×10^{-15}	*PHF2, FAM120AOS*	Neuroticism[23], depression[23], autism[24]
rs7106434	11	112,860,579	C/T	0.41	1.04	1.02–1.06	3.19×10^{-8}	2.27×10^{-5}	9.17×10^{-11}	*NCAM1*	Neuroticism[23], depression[25], cannabis use[22], anorexia nervosa[40]
rs5803650	13	53,939,598	CT/C	0.48	1.05	1.03–1.07	2.97×10^{-8}	2.25×10^{-8}	6.31×10^{-14}	*CKAP2, TPTE2P3*	
rs9513519	13	99,610,146	G/A	0.62	1.04	1.02–1.06	3.09×10^{-8}	4.20×10^{-5}	2.31×10^{-10}	*DOCK9*	

The reported frequencies and effects are those of the second allele. The second allele is defined such that it increases IBS risk. Allele frequencies are taken from UKB. Previous associations were obtained from the literature and GWAS Catalog (Supplementary Note).

functional constipation and functional diarrhea, were less strongly associated with raised PHQ-12 scores or psychological comorbidity.

IBS showed the strongest genome-wide overlap with psychological traits: anxiety, neuroticism, depression and schizophrenia (Fig. 3). GAD-7 anxiety scores correlated with IBS severity (Fig. 1) and 34.3% of cases with IBS had sought or had been treated for anxiety versus 16.1% of controls (Table 1). Although the phenotypic correlation was strong, the genetic correlation appeared quantitatively even greater (Fig. 3). Furthermore, this genetic correlation between IBS and anxiety persisted even after eliminating data from individuals with phenotypic overlap (that is, between GWAS for 'IBS excluding anxiety' and 'anxiety excluding IBS'; Supplementary Fig. 12). Thus, their co-occurrence probably reflects shared etiologic pathways between IBS and anxiety rather than one condition simply causing the other. This conclusion was supported by the Mendelian randomization analysis.

Four out of six of the confirmed IBS loci implicated genes influencing mood or anxiety disorders, genes expressed in the nervous system or both. These include *NCAM1* (also associated with neuroticism, anxiety, mood disorders and anorexia nervosa)[23,25,40], *CADM2* (also associated with neuroticism, anxiety and cannabis use)[21,22], PHD finger protein 2 (*PHF2*)/family with sequence similarity 120A (*FAM120A*) (also associated with neuroticism, depression and autism)[23,24] and dedicator of cytokinesis 9 (*DOCK9*). Brain expression of *NCAM1* and *CADM2* was implicated in our colocalization analysis (Fig. 2b and Supplementary Table 10): both regulate neural circuit formation and influence changes in white matter microstructure found in both mood disorders and IBS[25,41,42]. *PHF2* and *DOCK9* also play key roles in brain development[43,44]. Of note, *NCAM1*, *PHF2* and *DOCK9* are also expressed in the rich network of nerve fibers and ganglia of the gut, while *CADM2* is not[45]. Predominant brain expression, combined with the coassociation of IBS with several psychological traits, perhaps most strongly implicates the central nervous system as the site where these gene variants exert their action. However, the genetic variants may also be acting peripherally for the subset expressed in the enteric nervous system, which shares many neurotransmitters, signaling pathways and anatomical properties as well as rich communication with the brain.

The MHC signal is independent of known HLA associations with ulcerative colitis and celiac disease; in fact, it localizes to *BAG6* (Supplementary Fig. 7 and Supplementary Tables 10–12). BAG6 is known to chaperone misfolded proteins, regulate membrane protein dynamics and affect diverse processes from apoptosis to antigen presentation[46,47]. Functional exploration of *BAG6* may yield new IBS pathophysiological insights unconnected to the nervous system.

IBS genome-wide SNP heritability was just 5.8% (s.e.m. < 0.01) in the European ancestry population in this study and the effect sizes of our susceptibility loci were modest (OR < 1.05). Earlier genetic studies of IBS were underpowered to detect such small effects. By comparison, SNP heritability estimates for Crohn's disease, ulcerative colitis and anxiety are 41%, 23% and 26%, respectively[48,49]. Previous IBS heritability estimates, from family and twin studies, varied widely at 0–57% (ref. [11]). Our results indicate that the genetic contribution to IBS heritability is modest and imply that additional environmental factors, including dysbiosis, diet, stress and learned behaviors, all potentially shared within families, play a more prominent role.

Regarding dysbiosis, we noted increased childhood exposure to antibiotics among IBS cases (20.0%) versus controls (9.6%). While there are clearly biases inherent to recall of events from childhood, this result is corroborated by previous studies specifically set up to address this question[50]. Interestingly, we saw the same association with anxiety (18.4%). Among possible explanations, childhood antibiotics might increase the risk of IBS (and perhaps anxiety) by embedding a dysbiotic gut flora and disturbing the balance of short-chain fatty acid metabolites known to influence microglial development and mood[50,51]. Equally, anxiety in late adulthood might influence recall of childhood antibiotic exposure, and familial anxiety might lead parents to take their offspring to the doctor repeatedly for minor ailments, resulting in recurrent antibiotic exposure. While enteric infection can alter the baseline gut microbiota and trigger PI-IBS, in the UKB PI-IBS closely mirrored 'conventional' IBS in terms of symptom severity, frequency of family history and association with psychological traits, suggesting that the infectious 'seed' falls on fertile ground to trigger IBS in predisposed individuals.

One question is whether the neuronal emphasis of our results derives from our strategy of combining multiple IBS definitions
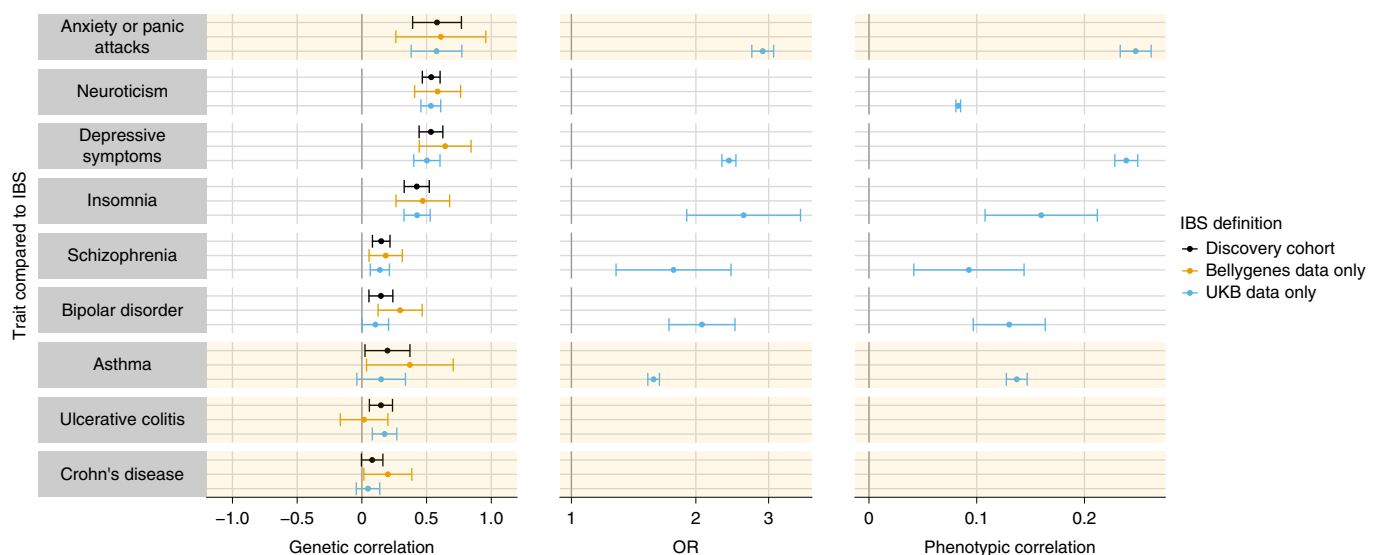
Fig. 3 | Genetic and phenotypic correlations between IBS and other traits. Correlations of genetic risk (coheritability estimates, left) and phenotype (ORs, middle, and liability-scale correlation, right) between IBS and other traits. The subset of all LD Hub traits (excluding rapid GWAS results) with significant genetic correlation (two-tailed coheritability test as implemented in the LDSC with unadjusted $P < 0.05$) in both UKB data (light blue, 40,548 cases and 293,220 controls) and the independent Bellygenes cohort (orange, 12,852 cases and 139,981 controls) is shown, as well as the meta-analysis used for discovery (black, 53,400 cases and 433,201 controls). IBS had a similar genetic risk profile to traits such as neuroticism, depression and insomnia (for which P values were significant in all datasets after multiple testing correction for the number of traits tested). Traits highlighted in yellow were added manually given their clinical relevance. Anxiety was not initially included since it is only available in the LD Hub as a rapid GWAS result. Among UKB participants, we present ORs (middle) and phenotypic correlations values (right) for these traits. Sample sizes (restricted to UKB participants who were either cases or controls in the discovery cohort; $n = 333,768$) were as follows: anxiety or panic attacks, 8,714; neuroticism, 271,423 (scores); depressive symptoms, 24,311; insomnia, 326; schizophrenia, 574; bipolar disorder, 1,207; asthma, 41,178.

to increase statistical power, including pooling 'opposite' subtypes (for example, IBS-C and IBS-D), that is, whether gut-specific effects might be lost in the pooling such that the brain remains the common link between these. However, the heritability of IBS subtypes is comparable with IBS overall; IBS-C and IBS-D share approximately 50% of their genetic susceptibility and each of the subgroups also individually genetically correlates with mental health and personality traits (Supplementary Table 16). Furthermore, subtype GWAS identified only one significant signal in IBS-C and none in IBS-D, suggesting an absence of strong subtype-specific, possibly gut-focused genetic effects.

Aside from the pooling strategy, justified by our LDSC analysis (Methods), other potential weaknesses include the use of Rome III criteria instead of the more restricted Rome IV criteria, since the former were the standard at the time of study design[52], the fact the IBS diagnosis was made based on Rome III symptoms reported via the DHQ rather than by medical review for nearly half of cases in the UKB cohort and the limited age range and ancestry of UKB. However, we believe that the fact that all of the loci identified at genome-wide significance thresholds in the discovery panel replicated in the independent 23andMe panel validates both the findings and the approach taken.

Our GWAS and the results of our polygenic analyses provide important new insights. Individual loci identified by the GWAS implicate new target genes within previously under-researched pathways (for example, neuronal adhesion). Mendelian randomization and genome-wide correlation analyses demonstrate shared genetic risk pathways between anxiety and IBS that are independent of the comorbidity between these two traits. This may point toward a mechanistic rationale for the efficacy of psychoactive medications and behavioral therapies and suggest that more attention should be paid to identifying new therapeutics that target neuronal function. We anticipate that future research will build on our discoveries,

both by investigating the target genes identified and exploring the shared genetic risk across traits to improve our understanding of the disordered brain–gut interactions that characterize IBS.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00950-8.

References

1. Ford, A. C. Commentary: estimating the prevalence of IBS globally––past, present and future. *Aliment. Pharmacol. Ther.* **51**, 198–199 (2020).
2. Gralnek, I. M., Hays, R. D., Kilbourne, A., Naliboff, B. & Mayer, E. A. The impact of irritable bowel syndrome on health-related quality of life. *Gastroenterology* **119**, 654–660 (2000).
3. Soubieres, A., Wilson, P., Poullis, A., Wilkins, J. & Rance, M. Burden of irritable bowel syndrome in an increasingly cost-aware National Health Service. *Frontline Gastroenterol.* **6**, 246–251 (2015).
4. Hillilä, M. T., Färkkilä, N. J. & Färkkilä, M. A. Societal costs for irritable bowel syndrome—a population based study. *Scand. J. Gastroenterol.* **45**, 582–591 (2010).
5. Longstreth, G. F. et al. Functional bowel disorders. *Gastroenterology* **130**, 1480–1491 (2006).
6. Pacheco, R. L. et al. What do Cochrane systematic reviews say about the management of irritable bowel syndrome? *Sao Paulo Med. J.* **137**, 82–91 (2019).
7. Windgassen, S. et al. The journey between brain and gut: a systematic review of psychological mechanisms of treatment effect in irritable bowel syndrome. *Br. J. Health Psychol.* **22**, 701–736 (2017).
8. Koloski, N. A., Jones, M. & Talley, N. J. Evidence that independent gut-to-brain and brain-to-gut pathways operate in the irritable bowel

syndrome and functional dyspepsia: a 1-year population-based prospective study. *Aliment. Pharmacol. Ther.* **44**, 592–600 (2016).

9. Enck, P. et al. Irritable bowel syndrome. *Nat. Rev. Dis. Primers* **2**, 16014 (2016).

10. Jalanka, J., Salonen, A., Fuentes, S. & de Vos, W. M. Microbial signatures in post-infectious irritable bowel syndrome––toward patient stratification for improved diagnostics and treatment. *Gut Microbes* **6**, 364–369 (2015).

11. Saito, Y. A. The role of genetics in IBS. *Gastroenterol. Clin. North Am.* **40**, 45–67 (2011).

12. Bengtson, M.-B., Aamodt, G., Vatn, M. H. & Harris, J. R. Co-occurrence of IBS and symptoms of anxiety or depression, among Norwegian twins, is influenced by both heredity and intrauterine growth. *BMC Gastroenterol.* **15**, 9 (2015).

13. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).

14. Bonfiglio, F. et al. Female-specific association between variants on chromosome 9 and self-reported diagnosis of irritable bowel syndrome. *Gastroenterology* **155**, 168–179 (2018).

15. D'Amato, M. Genes and functional GI disorders: from casual to causal relationship. *Neurogastroenterol. Motil.* **25**, 638–649 (2013).

16. DeBoever, C. et al. Assessing digital phenotyping to enhance genetic studies of human diseases. *Am. J. Hum. Genet.* **106**, 611–622 (2020).

17. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

18. Manolio, T. A., Bailey-Wilson, J. E. & Collins, F. S. Genes, environment and the value of prospective cohort studies. *Nat. Rev. Genet.* **7**, 812–820 (2006).

19. Francis, C. Y., Morris, J. & Whorwell, P. J. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. *Aliment. Pharmacol. Ther.* **11**, 395–402 (1997).

20. Spiller, R. C. et al. The Patient Health Questionnaire 12 Somatic Symptom scale as a predictor of symptom severity and consulting behaviour in patients with irritable bowel syndrome and symptomatic diverticular disease. *Aliment. Pharmacol. Ther.* **32**, 811–820 (2010).

21. Boutwell, B. et al. Replication and characterization of *CADM2* and *MSRA* genes on human behavior. *Heliyon* **3**, e00349 (2017).

22. Pasman, J. A. et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.* **21**, 1161–1170 (2018).

23. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & van der Sluis, S. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat. Commun.* **9**, 905 (2018).

24. Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

25. Petrovska, J. et al. The NCAM1 gene set is linked to depressive symptoms and their brain structural correlates in healthy individuals. *J. Psychiatr. Res.* **91**, 116–123 (2017).

26. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

27. Goyette, P. et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat. Genet.* **47**, 172–179 (2015).

28. Monsuur, A. J. et al. Effective detection of human leukocyte antigen risk alleles in celiac disease using tag single nucleotide polymorphisms. *PLoS ONE* **3**, e2270 (2008).

29. Green, H. D. et al. Genome-wide association study of microscopic colitis in the UK Biobank confirms immune-related pathogenesis. *J. Crohns Colitis* **13**, 1578–1582 (2019).

30. Westerlind, H. et al. Dense genotyping of immune-related loci identifies HLA variants associated with increased risk of collagenous colitis. *Gut* **66**, 421–428 (2017).

31. Camilleri, M. & Katzka, D. A. Irritable bowel syndrome: methods, mechanisms, and pathophysiology. Genetic epidemiology and pharmacogenetics in irritable bowel syndrome. *Am. J. Physiol. Gastrointest. Liver Physiol.* **302**, G1075–G1084 (2012).

32. Camilleri, M. et al. Neuropeptide S receptor induces neuropeptide expression and associates with intermediate phenotypes of functional gastrointestinal disorders. *Gastroenterology* **138**, 98–107.e4 (2010).

33. Zheng, J. et al. 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 (2017).

34. Ripke, S. et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).

35. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

36. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).

37. Levey, D. F. et al. Reproducible genetic risk loci for anxiety: results from ~200,000 participants in the Million Veteran Program. *Am. J. Psychiatry* **177**, 223–232 (2020).

38. Darrous, L., Mounier, N. & Kutalik, Z. Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. Preprint at *medRxiv* https://www.medrxiv.org/content/10.1101/2020.01.27.20018929v3 (2020).

39. Whitehead, W. E., Palsson, O. & Jones, K. R. Systematic review of the comorbidity of irritable bowel syndrome with other disorders: what are the causes and implications? *Gastroenterology* **122**, 1140–1156 (2002).

40. Watson, H. J. et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).

41. Kolkova, K., Novitskaya, V., Pedersen, N., Berezin, V. & Bock, E. Neural cell adhesion molecule-stimulated neurite outgrowth depends on activation of protein kinase C and the Ras-mitogen-activated protein kinase pathway. *J. Neurosci.* **20**, 2238–2246 (2000).

42. Frei, J. A., Andermatt, I., Gesemann, M. & Stoeckli, E. T. The SynCAM synaptic cell adhesion molecules are involved in sensory axon pathfinding by regulating axon–axon contacts. *J. Cell Sci.* **127**, 5288–5302 (2014).

43. Pappa, S. et al. PHF2 histone demethylase prevents DNA damage and genome instability by controlling cell cycle progression of neural progenitors. *Proc. Natl Acad. Sci. USA* **116**, 19464–19473 (2019).

44. Shi, L. Dock protein family in brain development and neurological disease. *Commun. Integr. Biol.* **6**, e26839 (2013).

45. Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

46. Kawahara, H., Minami, R. & Yokota, N. BAG6/BAT3: emerging roles in quality control for nascent polypeptides. *J. Biochem.* **153**, 147–160 (2013).

47. Mock, J.-Y. et al. Bag6 complex contains a minimal tail-anchor-targeting module and a mock BAG domain. *Proc. Natl Acad. Sci. USA* **112**, 106–111 (2015).

48. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

49. Purves, K. L. et al. A major role for common genetic variation in anxiety disorders. *Mol. Psychiatry* **25**, 3292–3303 (2020).

50. Krogsgaard, L. R., Engsbro, A. L. & Bytzer, P. Antibiotics: a risk factor for irritable bowel syndrome in a population-based cohort. *Scand. J. Gastroenterol.* **53**, 1027–1030 (2018).

51. Silva, Y. P., Bernardi, A. & Frozza, R. L. The role of short-chain fatty acids from gut microbiota in gut-brain communication. *Front. Endocrinol. (Lausanne)* **11**, 25 (2020).

52. Schmulson, M. J. & Drossman, D. A. What is new in Rome IV. *J. Neurogastroenterol. Motil.* **23**, 151–163 (2017).

¹Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. ²Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. ³Center for Molecular Medicine & Clinical Epidemiology Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. ⁴School of Biological Sciences, Monash University, Clayton, Victoria, Australia. ⁵IBD Pharmacogenetics, College of Medicine and Health, University of Exeter, Exeter, UK. ⁶Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. ⁷William Harvey Research Institute, Barts & The London School of Medicine & Dentistry, Queen Mary University of London, London, UK. ⁸Nuffield Department of Population Health, University

of Oxford, Oxford, UK. [9]23andMe, Inc., Sunnyvale, CA, USA. [10]Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK. [11]Saint Edmund Hall, University of Oxford, Oxford, UK. [12]Enteric NeuroScience Program, Division of Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic, Rochester, MN, USA. [13]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany. [14]Department of Dermatology, Quincke Research Center, University Hospital Schleswig-Holstein, Kiel, Germany. [15]Department of Biostatistics, University of Michigan, School of Public Health, Ann Arbor, MI, USA. [16]Department of Laboratory Medicine, Children's and Women's Health, Norwegian University of Science and Technology, Trondheim, Norway. [17]Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway. [18]Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway. [19]Department of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden. [20]Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. [21]Department of Genetics, University Medical Center Groningen, Groningen, the Netherlands. [22]Clinical Enteric Neuroscience Translational and Epidemiological Research and Division of Gastroenterology and Hepatology, Department of Medicine, Mayo Clinic, Rochester, MN, USA. [23]David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [24]Neurogastroenterology Unit, Wythenshawe Hospital, Centre for Gastrointestinal Sciences, University of Manchester, Manchester, UK. [25]Nottingham Digestive Diseases Centre, National Institute for Health Research Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK. [26]Biodonostia Health Research Institute, San Sebastian, Spain. [27]Gastrointestinal Genetics Lab, CIC bioGUNE – Basque Research and Technology Alliance, Derio, Spain. [28]IKERBASQUE, The Basque Science Foundation, Bilbao, Spain. [29]Christ Church, University of Oxford, Oxford, UK. [30]Division of Gastroenterology and Hepatology, Department of Medicine, University of Cambridge, Cambridge, UK. [31]Department of Gastroenterology, Cambridge University Hospital, Cambridge, UK. [86]These authors contributed equally: Tenghao Zheng, Nicholas A. Kennedy. [87]These authors jointly supervised this work: Mauro D'Amato, Luke Jostins, Miles Parkes. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: mdamato@cicbiogune.es; luke.jostins@kennedy.ox.ac.uk; miles.parkes@addenbrookes.nhs.uk

23andMe Research Team

Michelle Agee[9], Stella Aslibekyan[9], Adam Auton[9], Robert K. Bell[9], Katarzyna Bryc[9], Sarah K. Clark[9], Sarah L. Elson[9], Kipper Fletez-Brant[9], Pierre Fontanillas[9], Nicholas A. Furlotte[9], Pooja M. Gandhi[9], Karl Heilbron[9], Barry Hicks[9], David A. Hinds[9], Karen E. Huber[9], Ethan M. Jewett[9], Yunxuan Jiang[9], Aaron Kleinman[9], Keng-Han Lin[9], Nadia K. Litterman[9], Marie K. Luff[9], Jey C. McCreight[9], Matthew H. McIntyre[9], Kimberly F. McManus[9], Joanna L. Mountain[9], Sahar V. Mozaffari[9], Priyanka Nandakumar[9], Elizabeth S. Noblin[9], Carrie A. M. Northover[9], Jared O'Connell[9], Aaron A. Petrakovitz[9], Steven J. Pitts[9], G. David Poznik[9], J. Fah Sathirapongsasuti[9], Anjali J. Shastri[9], Janie F. Shelton[9], Jingchunzi Shi[9], Suyash Shringarpure[9], Chao Tian[9], Joyce Y. Tung[9], Robert J. Tunney[9], Vladimir Vacic[9], Xin Wang[9] and Amir S. Zare[9]

The Bellygenes Initiative

Mauro D'Amato[3,4,13,26,27,28,87], Tenghao Zheng[3,4,86], Ferdinando Bonfiglio[3,4], Michael Camilleri[22], Gianrico Farrugia[12], Purna Kashyap[12], Lin Chang[32], Emeran Mayer[23], Margaret Heitkemper[33], Gregory S. Sayuk[34], Tamar Ringel-Kulka[35], Yehuda Ringel[36,37], William D. Chey[38], Shanti Eswaran[38], Juanita L. Merchant[39], Robert J. Shulman[40], Luis Bujanda[41,42,43], Koldo Garcia-Etxebarria[41,43], Aldona Dlugosz[44], Greger Lindberg[44], Peter T. Schmidt[44], Pontus Karling[45], Bodil Ohlsson[46], Susanna Walter[47], Åshild O. Faresjö[48], Magnus Simren[49], Jonas Halfvarson[50], Piero Portincasa[51], Giovanni Barbara[52], Paolo Usai-Satta[53], Matteo Neri[54], Gerardo Nardone[55], Rosario Cuomo[56], Francesca Galeazzi[57], Massimo Bellini[58], Anna Latiano[59], Lesley Houghton[60,61,62], Daisy Jonkers[63], Alexandra Zhernakova[21], Alexander Kurilshikov[21], Rinse K. Weersma[64], Mihai Netea[65], Andre Franke[13], Matthias Hübenthal[13,14], Jonas Tesarz[66], Annika Gauss[67], Miriam Goebel-Stengel[68,69], Viola Andresen[70], Thomas Frieling[71], Christian Pehl[72], Rainer Schaefert[73,74], Beate Niesler[75,76], Wolfgang Lieb[77], Kurt Hanevik[78], Nina Langeland[78], Knut-Arne Wensaas[79], Sverre Litleskare[80], Anne Heidi Skogholt[16,17], Maiken E. Gabrielsen[81], Laurent Thomas[82], Kristian Hveem[17], Eivind Ness-Jensen[17,18,19], Tõnu Esko[20], Maris Teder-Laving[20], Vincent Thijs[83], Robin Lemmens[84], Lukas Van Oudenhove[85], Mira Wouters[85] and Guy Boeckxstaens[23]

[32]Division of Digestive Diseases/Gastroenterology, David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, CA, USA. [33]Department of Biobehavioral Nursing and Health Informatics, University of Washington, Seattle, WA, USA. [34]Division of Gastroenterology, Washington

University School of Medicine, St. Louis, MO, USA. [35]Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [36]Department of Medicine, Division of Gastroenterology and Hepatology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [37]Division of Gastroenterology and Hepatology, Meir Medical Center, affiliated to Tel-Aviv University, Kfar Saba, Israel. [38]Division of Gastroenterology, Michigan Medicine, Ann Arbor, MI, USA. [39]Division of Gastroenterology, University of Arizona College of Medicine, Tucson, AZ, USA. [40]Children's Nutrition Research Center, Baylor College of Medicine, Texas Children's Hospital, Houston, TX, USA. [41]Universidad del País Vasco, San Sebastian, Spain. [42]Department of Gastrointestinal and Liver Diseases, Biodonostia Health Research Institute, Sebastian, Spain. [43]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Instituto Carlos III, Madrid, Spain. [44]Department of Medicine Solna, Karolinska Institutet, Center for Digestive Diseases, Karolinska University Hospital, Stockholm, Sweden. [45]Division of Medicine, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden. [46]Department of Internal Medicine, Lund University, Skåne University Hospital, Lund, Sweden. [47]Division of Neuro and Inflammation Science, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden. [48]Department of Health, Medicine and Caring Science/Society and Health/Public health, Linköpings University, Linköping, Sweden. [49]Department of Internal Medicine & Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [50]Department of Gastroenterology, School of Medical Sciences, Örebro University, Örebro, Sweden. [51]Department of Biomedical Sciences and Human Oncology, Clinica Medica 'A. Murri', University of Bari Medical School, Bari, Italy. [52]Department of Medical and Surgical Sciences, University of Bologna, St. Orsola – Malpighi Hospital, Bologna, Italy. [53]S.C. Gastroenterologia, Azienda Ospedaliera G. Brotzu, Cagliari, Italy. [54]Department of Medicine and Aging Sciences and CESI, G. D'Annunzio University & Foundation, Chieti, Italy. [55]Gastroenterology Unit, Department of Clinical Medicine and Surgery, University Federico II, Naples, Italy. [56]Gastroenterology Unit, Department of Medical Science, 'Sant'Anna e San Sebastiano' Hospital, Caserta, Italy. [57]Unità Operativa Complessa Gastroenterologia, Padova University Hospital, Padova, Italy. [58]Gastroenterology Unit, Department of Gastroenterology, University of Pisa, Pisa, Italy. [59]Division of Gastroenterology, Istituto di Ricovero e Cura a Carattere Scientifico 'Casa Sollievo della Sofferenza' Hospital, San Giovanni Rotondo, Italy. [60]Division of Gastroenterology and Surgical Sciences, Leeds Institute of Medical Research at St. James's, University of Leeds, Leeds, UK. [61]Department of Gastroenterology, Mayo Clinic, Jacksonville, FL, USA. [62]GI Sciences, Division of Diabetes, Endocrinology & Gastroenterology, University of Manchester, Manchester, UK. [63]Department of Internal Medicine, School for Nutrition and Translational Research in Metabolism, Maastricht University Medical Center+, Maastricht, the Netherlands. [64]Department of Gastroenterology and Hepatology, University Medical Center Groningen, Groningen, the Netherlands. [65]Department of Internal Medicine and Radboud Center of Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands. [66]Department of General Internal Medicine and Psychosomatics, University Hospital Heidelberg, Heidelberg, Germany. [67]Department of Gastroenterology, Infectious Diseases and Intoxications, University of Heidelberg, Heidelberg, Germany. [68]Department of Psychosomatic Medicine, University Hospital Tübingen, Tübingen, Germany. [69]Department of Internal Medicine and Gastroenterology, HELIOS Clinic Rottweil, Rottweil, Germany. [70]Israelitisches Krankenhaus, Hamburg, Germany. [71]Helios Klinik Krefeld, Krefeld, Germany. [72]Krankenhaus Vilsbiburg, Vilsbiburg, Germany. [73]Department of Psychosomatic Medicine, Division of Internal Medicine, University Hospital Basel, Basel, Switzerland. [74]Faculty of Medicine, University of Basel, Basel, Switzerland. [75]Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany. [76]Interdisciplinary Center for Neurosciences, Heidelberg University, Heidelberg, Germany. [77]Institute of Epidemiology, Christian-Albrechts-University Kiel, Kiel, Germany. [78]Department of Clinical Science, Faculty of Medicine, University of Bergen, Bergen, Norway. [79]Research Unit for General Practice, NORCE Norwegian Research Centre, Bergen, Norway. [80]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway. [81]KG Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. [82]KG Jebsen Center for Genetic Epidemiology, Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. [83]Florey Institute of Neuroscience and Mental Health, Heidelberg, Victoria, Australia. [84]Department of Neurosciences, Katholieke Universiteit Leuven, Leuven, Belgium. [85]Translational Research Center for Gastrointestinal Disorders, Katholieke Universiteit Leuven, Leuven, Belgium.

Methods

For details of the cohorts, descriptive statistics, association analysis methodologies and functional interpretation of associations, see the Supplementary Note.

Ethics oversight. The UKB DHQ was approved as a substantial amendment to the UKB protocol by the North West – Haydock Research Ethics Committee, reference no. 11/NW/038e. The Bellygenes initiative study received ethical approval from the Stockholm Ethics Examination Authority (EPN ID 2016/1620-31/2) and Monash University Human Research Ethics Committee (MUHREC ID 20326). Written informed consent was obtained from all participants. This study did not award compensation to any participant.

Data and study participants. Our discovery cohort combined cases of IBS identified in UKB with cases from the Bellygenes initiative. Replication was sought in an independent panel from 23andMe. Cases ascertained in UKB met at least one of the following four conditions (the DHQ is viewable online – UKB resource 595): (1) DHQ Rome III: met Rome III symptom criteria for IBS diagnosis without other diagnostic explanations for these symptoms; (2) DHQ 'self-report': answered 'yes' to the question 'Have you ever been diagnosed with IBS?'; (3) Unprompted 'self-report': self-reported IBS diagnosis in response to question 'Has a doctor ever told you that you have any … serious medical conditions?'; (4) hospital ICD-10: linked hospital episode statistics indicating inpatient or day-case admission with clinician diagnosis of IBS entered as main or secondary ICD-10 diagnosis.

Participants with conditions such as celiac disease, inflammatory bowel disease or previous intestinal resectional surgery that could result in IBS-like symptoms were excluded from both cases and digestively healthy controls to avoid signal contamination. For detailed case and control inclusion and exclusion criteria, see Supplementary Tables 1 and 2, respectively. To maximize sample size, cases from the 4 UKB groups were pooled ($n = 40{,}548$). This approach was supported by demonstrating high genetic correlations between them using LDSC[48] following a separate GWAS on each (minimum pairwise $r_g = 0.70$, s.e.m. $= 0.06$; Supplementary Fig. 13) and by previous literature on the consistency of genetic results obtained from different diagnostic definitions in UKB[16].

We then meta-analyzed IBS GWAS data from UKB (40,548 cases) and Bellygenes initiative (12,852 cases and 139,981 controls; Supplementary Table 9), an international collaboration studying IBS genetics based on electronic medical records, specialist diagnoses form tertiary clinics and questionnaire data (including Rome III criteria) across multiple cohorts, having again demonstrated high genetic correlation between them ($r_g = 0.998$, s.e.m. $= 0.129$). This produced a total discovery cohort of 53,400 cases and 433,201 controls. Evidence of replication was sought in a large 23andMe dataset (Supplementary Note). 23andMe cases ($n = 205{,}252$) self-reported being diagnosed or treated for IBS while controls ($n = 1{,}384{,}055$) did not.

Analyses of IBS subtypes were conducted solely using UKB DHQ data based on standard definitions of IBS-C, IBS-D, IBS-M and IBS-U according to the frequency of hard or lumpy stools versus loose, mushy or watery stools. Functional constipation and functional diarrhea cases were identified similarly, and with the same exclusions per IBS cases, but (in contrast to the Rome III definition of IBS) needed to have responded 'never' when asked about the frequency of abdominal pain in the last three months. Likewise, analyses of IBS severity (using the IBS-SSS) and associated somatic symptoms (using the PHQ-12) were restricted to DHQ respondents. Anxiety and depression were identified among UKB participants based on previously surveyed responses to GAD-7 anxiety and PHQ-9 depression questionnaires, self-report of diagnosis with depression or anxiety/panic attack, diagnostic codes for major depression and phobic or generalized anxiety disorder in electronic healthcare records or reporting of treatment being sought or offered for these conditions in our DHQ (Supplementary Note).

Statistical analysis. Association between IBS and nongenetic risk factors, including risk factors assayed by recall from the DHQ, was tested using logistic regression conditioning on age and sex (Supplementary Note, 'Nongenetic associations').

Standard genetic quality control was carried out to remove samples with poor genotype quality and variants with poor genotyping or imputation performance. Only participants of European ancestry were included in the discovery dataset due to the limited number of non-European ancestry participants. GWAS were conducted using a linear mixed model (BOLT-LMM v.2.3.2)[53] to control for population stratification and relatedness. Meta-analysis of GWAS summary statistics was carried out using METAL (March 2011 release)[54]. The UKB GWAS was stratified into DHQ respondents and nonrespondents, with results meta-analyzed to avoid genetic confounding with questionnaire response (Supplementary Fig. 14).

We assigned loci to candidate genes using annotations from FUMA v.1.3.4[55], as well as from a colocalization analysis using Coloc v.3.2-1[56] on multi-tissue expression data from the Genotype-Tissue Expression (GTEx) consortium[56,57]. We calculated SNP heritability and coheritability (r_g, genetic correlation) using univariate and bivariate LDSC[48] against a range of traits via the LD Hub website[33]. Other statistical analyses were carried out using R v.3.6.1; any P values were obtained from two-sided tests unless otherwise specified.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Genome-wide summary statistics have been deposited to the European Bioinformatics Institute GWAS Catalog (https://www.ebi.ac.uk/gwas/) under accession no. GCST90016564. Individual-level data on the DHQ responses, along with matching genotype, electronic health record and survey data, are available via an application to the UK Biobank Access Management System (https://bbams.ndph.ox.ac.uk/ams/). Individual-level data for 23andMe were not shared as part of this project to protect the privacy of 23andMe participants. Data used for the analyses pertinent to the Bellygenes initiative include both individual-level and aggregate data. Individual-level data from the following sources can be obtained via applications to the respective biobanks and cohorts: TWINGENE (https://ki.se/en/research/swedish-twin-registry-for-researchers); HUNT (https://www.ntnu.edu/hunt/data); Michigan Genomics Initiative (https://precisionhealth.umich.edu/our-research/michigangenomics/); Estonian Genome Center of the University of Tartu (https://genomics.ut.ee/en/biobank.ee/data-access); Lifelines (https://www.lifelines.nl/researcher/how-to-apply); Gene-Environment and Gene-Gene Interaction Research Application (dbGaP Study accession no. phs000674.v2.p2, now superseded by phs000674.v3.p3). Data from IBS patients from tertiary centers can be requested from Mauro D'Amato at mdamato@cicbiogune.es and may be made available depending on specific material and data transfer agreements with principal investigators at respective collaborating institutions.

Code availability

All software used is publicly available at the URLs or references cited. The R code for additional analyses is available at https://doi.org/10.5281/zenodo.5048820.

References

53. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
54. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
55. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
56. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
57. Lonsdale, J. et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

Author contributions

Competing interests

Additional information

nature research

Corresponding author(s): Dr. Luke Jostins

Last updated by author(s): Jul 1, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No relevant software was used to collect data used in the study.
Data analysis	BOLT-LMM was used to conduct the GWAS in the UK Biobank sample. Results were meta-analyzed with Bellygenes data using METAL (March 2011 version). We assigned loci to candidate genes using annotations from FUMA (v1.3.4), as well as from a colocalization analysis using Coloc (v3.2-1). To perform conditional analyses, we extracted SNPs associated with gastrointestinal disorders other than IBS from UK Biobank bgen data in expected dosage format using qctool 2.0.5 (https://www.well.ox.ac.uk/~gav/qctool_v2/). For clumping, we used PLINK version 2.00a2 and PLINK version 1.90b6.7. Use of qctool and PLINK is detailed in the Supplementary Note, as these tools were used for file format conversions rather than key analyses. We calculated SNP heritability and co-heritability (rg, genetic correlation) using univariate and bivariate LDSC (v1.0.0) against a range of traits via the LD Hub (v1.9.0) website. Further statistical analyses were carried out using R version 3.6.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Genome-wide summary statistics have been deposited to the EBI GWAS catalogue (https://www.ebi.ac.uk/gwas/), under accession GCST90016564. Individual-level data on the DHQ responses, along with matching genotype, electronic health record and survey data, are available via an application to the UK Biobank Access

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study did not recruit participants up until a target sample size was reached. Instead, this study uses the pre-defined UK Biobank sample, allowing participants to supply additional data on digestive health.
Data exclusions	UK Biobank participants with conditions such as coeliac disease, inflammatory bowel disease or previous intestinal resectional surgery which could result in IBS-like symptoms were excluded from both cases and digestively healthy controls to avoid signal contamination. These exclusion criteria were pre-established.
Replication	All genetic associations passing the genome-wide significance threshold in the discovery cohort were successfully replicated in an independent 23andMe dataset. The discovery analysis was performed once, and its results were replicated once.
Randomization	Participants were not randomly allocated into experiment groups. Instead, case and control groups for the genetic analyses were defined to have IBS or be digestively healthy. Differences between cases and controls were accounted for by including covariates such as age, sex, and transformations thereof in the genetic and phenotypic association analyses. Survey response bias was accounted for by analyzing respondents and non-respondents separately, then meta-analyzing the results.
Blinding	Data collectors were blinded to the genotype status of the participants, and participants were not made aware of their genotype status by UK Biobank. Unblinded genotype data was available to data analysts, but standard quality control and analysis software was used which makes blinding not relevant for this study.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if*

| Randomization | *allocation was not random, describe how covariates were controlled.* |

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	UK Biobank participants were between ages 40-69 upon recruitment, and predominantly of European ancestry. A subset responded to the digestive health questionnaire (DHQ), one of the means used to identify IBS cases. Respondents have lower rates of IBS as measured via hospital ICD-10 codes (1.16% vs 1.40% among 171,061 respondents and 317,234 non-respondents, respectively), but not unprompted self-reporting (2.85% vs 2.30%). They also have lower rates of mental health disorders based on hospital ICD-10 codes (schizophrenia: 0.04% vs 0.21%, depression: 1.77% vs 3.40%) and unprompted self-reporting (schizophrenia: 0.05% vs 0.16%, depression: 5.83% vs 6.20%). Respondents also had a lower mean age than non-respondents (64.8 vs 65.8 years when the DHQ data were collected), and were more often female (56.7% vs 52.9%). Response rates also varied by ethnicity, e.g. 15.8% (1205 of 7645) among participants who report a Black or Black British background compared to 36.0% (165243 of 459256) among participants reporting a White ethnic background. Ultimately, the GWAS was conducted in a sample of European ancestry (this includes the Bellygenes population described in Table S9).

Recruitment	UK Biobank recruitment has been documented elsewhere, but in brief, it is known that its participants are disproportionately older, healthier and wealthier than the UK population as a whole. This, combined with the fact that the individuals analyzed were of European ancestry, means results may not generalize to the entire population, or populations from other ancestries.
	This study expanded the UK Biobank dataset via a digestive health questionnaire (DHQ). The final digestive health questionnaire (available online as UK Biobank resource 595) was incorporated into the UK Biobank Questionnaire platform and advertised via email as an online questionnaire entitled 'digestive health'. Participants who had agreed to email contact were sent a hyperlink to their questionnaire, enabling linkage of results to other data in the UK Biobank dataset. The DHQ was also available on the participant area of the UK Biobank website and participants without email addresses received a flyer with their annual postal newsletter encouraging them to login to the participants' area of the UK Biobank website, and complete the questionnaire. Given that this resulted in a sample with a predominantly white British background (as is true for UK Biobank generally), genetic associations with IBS identified in this study may not generalize to non-European populations. Additionally, the fact that the sample consists of older individuals means that genetic association specific to IBS as it presents in younger individuals may be missed.
	Recruitment of the Bellygenes cohorts is as described in the supplement. Briefly, the Bellygenes initiative is a large international collaboration that includes data from multiple biobanks, population-based cohorts, and patients from tertiary gastroenterology clinics worldwide. Three independent datasets were included in this study, based on different definitions of IBS: Rome Criteria for population-based cohorts HUNT, TWINGENE and LifeLines; ICD10 diagnoses from healthcare records for Mayo Genome Consortium, Estonia Genome Center at the University of Tartu, Michigan Genomics Initiative, and Genetic Epidemiology Research on Ageing from the Kasier Permanente biobank; gastroenterologist diagnosis for tertiary center patients. Except for patients from tertiary centers (who were recruited into IBS-specific studies at respective sites) the recruitment of participants into other cohorts was not focused on IBS or any other disease, as these are population-based cohorts (whose implementation has been described previously, see Supplementary Methods). As in the case of UK Biobank (whose "healthy volunteer" selection bias has been described) self-selection bias cannot be excluded for individuals from the Bellygenes initiative (comprising multiple population-based cohorts). Given the observed genetic link between mood/anxiety disorders and IBS, an overestimation of genetic risk effects may not be ruled out if increased propensity to volunteering data for research were assumed as common denominator for mood/anxiety disorders and IBS.
Ethics oversight	This study meta-analyzes UK Biobank and Bellygenes data. The UK Biobank digestive health questionnaire was approved as a substantial amendment to the UK Biobank protocol by the North West - Haydock REC, reference 11/NW/038e. The Bellygenes initiative study received ethical approval from Stockholm Ethics Examination Authority (EPN ID 2016/1620-31/2) and Monash University Human Research Ethics Committee (MUHREC ID 20326). Informed consent was obtained from all human participants. This study did not award compensation to any participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.
Study protocol	Note where the full trial protocol can be accessed OR if not available, explain why.
Data collection	Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.
Outcomes	Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No | Yes
- ☐ ☐ Public health
- ☐ ☐ National security
- ☐ ☐ Crops and/or livestock
- ☐ ☐ Ecosystems
- ☐ ☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session
(e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument

Identify the instrument used for data collection, specifying make and model number.

Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI   ☐ Used   ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain   ☐ ROI-based   ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.