PROFESSOR MAURO D'AMATO (Orcid ID : 0000-0003-2743-5197)

Article type : Original Article

A survey of functional dyspepsia in 361,360 individuals: phenotypic and genetic cross-disease analyses

Running title: Cross-disease analyses in functional dyspepsia

Koldo Garcia-Etxebarria[1,2,3], Florencia Carbone[4], Maris Teder-Laving[5], Anita Pandit[6], Lieselot Holvoet[4], Vincent Thijs[7], Robin Lemmens[8], Luis Bujanda[1,2,9], Andre Franke[10], Sebastian Zöllner[6,11], Michael Boehnke[6], Matthew Zawistowski[6], Tonu Esko[5], Jan Tack[4],* Mauro D'Amato[1,3,12,13]*

* these authors equally contributed to the study

[1] Biodonostia Health Research Institute, Department of Gastrointestinal and Liver Diseases, San Sebastian, Spain
[2] Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Madrid, Spain
[3] Center for Molecular Medicine and Clinical Epidemiology Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden
[4] Translational Research Center for GI Disorders (TARGID), University of Leuven, Leuven, Belgium
[5] Estonian Genome Center, University of Tartu, Tartu, Estonia
[6] Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA
[7] Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg, Australia
[8] Department of Neurosciences, Leuven Brain Institute (LBI), University of Leuven, Leuven, Belgium
[9] Universidad del País Vasco (UPV/EHU), San Sebastián, Spain
[10] Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
[11] Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA
[12] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[13] Gastrointestinal Genetics Lab, CIC bioGUNE – BRTA, Derio, Spain

Correspondence
Mauro D'Amato, Prof
Gastrointestinal Genetics Lab
CIC bioGUNE, Basque Research and Technology Alliance
Parque Tecnológico de Bizkaia, 48160 Derio, Spain
mdamato@cicbiogune.es

Abbreviations
EGCUT: Estonian Genome Center of the University of Tartu
FD: Functional dyspepsia
FGID: Functional gastrointestinal disorder
GWAS: Genome-wide association study
ICD9: International Classification of Diseases 9th Revision
ICD10: International Classification of Diseases 10th Revision
MGI: Michigan Genome Initiative
UKBB: UK Biobank

ABSTRACT

Background. Functional dyspepsia (FD) is a common gastrointestinal condition of poorly understood pathophysiology. While symptoms' overlap with other conditions may indicate common pathogenetic mechanisms, genetic predisposition is suspected but has not been adequately investigated.

Methods. Using healthcare, questionnaire and genetic data from three large population-based biobanks (UK Biobank, EGCUT and MGI), we surveyed FD comorbidities, heritability and genetic correlations across a wide spectrum of conditions and traits in 10,078 cases and 351,282 non-FD controls of European ancestry.

Key Results. In UK Biobank, 281 diagnoses were detected at increased prevalence in FD, based on healthcare records. Among these, gastrointestinal conditions (OR=4.0, $P<1.0x10^{-300}$), anxiety disorders (OR=2.3, $P<1.4x10^{-27}$), ischemic heart disease (OR=2.2, $P<2.3x10^{-76}$), and infectious and parasitic diseases (OR=2.1, $P=1.5x10^{-73}$) showed strongest association with FD. Similar results were obtained in an analysis of self-reported conditions and use of medications from questionnaire data. Based on a genome-wide association meta-analysis of

genotypes across all cohorts, FD heritability was estimated close to 5% (h^2_{SNP}=0.047, P=0.014). Genetic correlations indicate FD predisposition is shared with several other diseases and traits (r_g>0.344), mostly overlapping with those also enriched in FD patients. Suggestive ($P<5.0\times10^{-6}$) association with FD risk was detected for 13 loci, with 2 showing nominal replication (P<0.05) in an independent cohort of 192 FD patients.

Conclusions & Inferences. FD has a weak heritable component that shows commonalities with multiple conditions across a wide spectrum of pathophysiological domains. This new knowledge contributes to a better understanding of FD aetiology, and may have implications for improving its treatment.

Keywords: functional dyspepsia; genetics; biobank studies; genome-wide association study; comorbidities

INTRODUCTION

Functional dyspepsia (FD) is a functional gastrointestinal disorder (FGID) estimated to affect approximately 10% of the adult population worldwide, with symptoms including postprandial fullness, early satiation, and epigastric pain or burning in the absence of detectable organic disease.[1,2] FD is often associated with symptoms overlapping with other FGIDs, like irritable bowel syndrome (IBS), and strongly impacts patients' quality of life.[1,3] This is particularly true in the important portion of FD patients (20%) also experiencing psychosocial conditions like anxiety, depression or somatization, and who are therefore often referred to advanced care because of severely impaired daily functioning.[4,5] Overall, FD is associated with major substantial direct (healthcare) and indirect (work loss) cost, thus posing a considerable burden to society.[6,7] Current treatment options for FD have shown limited efficacy, and the development of novel therapeutic approaches is hampered by the heterogeneity of symptoms, the observed large placebo response in clinical trials (30-40%), and the intrinsic difficulty in establishing the underlying pathophysiology in individual patients.[8–10]

The aetiology of FD is heterogeneous and not fully elucidated. Several candidate mechanisms have been implicated in symptom generation, such as altered visceral sensitivity and motility, mucosal perturbations (low grade duodenal inflammation, impaired mucosal duodenal integrity) and abnormal processing in the brain (visceral specific anxiety, somatisation).[11–14] Genetic predisposition to FD is suspected from the observation of familial clustering, though genetic studies have been very scarce and limited to candidate genes in small sample sets of cases and controls, hence no unequivocal FD risk gene has been proposed.[15,16]

In the study of human complex conditions, cross-disease analyses are useful approaches to unearth similar aetiology and common pathophysiological mechanisms, including those eventually accounted for by shared genetic factors. This may be particularly relevant in FD

and other FGIDs, whereby studying commonalities may improve disease understanding, and eventually inform the development of novel or alternative preventive and therapeutic strategies. Here, we leveraged data from UK Biobank (UKBB) and two smaller population-based cohorts (Michigan Genome Initiative - MGI, and the Biobank of the Estonian Genome Center, University of Tartu - EGCUT) for large-scale cross-disease analyses in order to study i) the co-occurrence of FD and other conditions, based on participants' healthcare records and questionnaire data and, ii) genetic correlation between FD and other traits and diseases, based on previous and newly generated FD GWAS data. We report the results obtained from such surveys conducted on a total of 361,360 individuals of European ancestry.

MATERIALS AND METHODS

Definition of functional dyspepsia cases

A similar protocol was adopted in all cohorts for the identification of FD cases based on respective participants' healthcare data and diagnoses classified according to the International Classification of Diseases 10th Revision (ICD10) for UKBB and EGCUT, and/or 9th revision (ICD9) for MGI. Cases were selected as those with a FD diagnosis (ICD9 536.8 or ICD10 K30) in their medical records, while the remainder of the cohorts served as control group. Exclusion criteria included *Diseases of oesophagus* (ICD9 530.*), *Gastro-oesophageal reflux disease* (ICD10 K21, except K21.9 in cases), *Other diseases of oesophagus* (ICD10 K22), *Gastric ulcer* (ICD9 531.* or ICD10 K25), *Duodenal ulcer* (ICD9 532* or ICD10 K26), *Peptic ulcer* (ICD9 533* or ICD10 K27), *Gastrojejunal ulcer* (ICD9 534* or ICD10 K28), *Inflammatory Bowel Disease* (IBD; ICD9 555.*, 556* or ICD10 K50), *IBS* (ICD9 564.1 or ICD10 K58) and *Coeliac disease* (CD; ICD9 579.0 or ICD10 K90.0).

Study cohorts

*UKBB***.** UK Biobank is a longitudinal cohort including approximately 500,000 individuals (aged 40–69 years), recruited between 2006 and 2010 in the UK.[17] Together with genotypes (Affymetrix Biobank Axiom and UK BiLEVE Axiom arrays), participants' health-related information is available and spans questionnaire data on general health and lifestyle, self-reported conditions and use of medications. In the informed consent participants gave "*permission for access to my medical and other health-related records, and for long-term storage and use of this and other information about me, for health-related research purposes (even after my incapacity or death)*". For the purpose of this study, genotypes and the following phenotypic data have been used in relation to UKBB participants self-reporting British ancestry (as from UK Biobank data field 21000): main and secondary ICD10 diagnoses (respectively, UK Biobank data fields 41202 and 41024); doctor's diagnosis of non-cancer illnesses, self-reported (UK Biobank data field 20002), and use of medications (UK Biobank data fields 6153, 6154, 6177). The study was approved by the regional Ethics Committee of the Basque Country (CEIm-E ref PI2020167), and performed with UKBB data

accessed under UKBB registration nr 30537. The demographics of UKBB individuals are reported in Table 1.

EGCUT. EGCUT is a volunteer-based sample of the Estonian adult population, which comprises more than 52,000 participants aged >18 years, and previously described eslsewhere.[18] In the informed consent the participants were made aware that "*The Gene Bank enables scientific and applied gene and health research to be carried out in order to determine genes that influence the development of diseases*". For the purpose of this study, among EGCUT participants with genotype data available (Illumina 370K, OmniExpress, CoreExome and Global Screening arrays - GSA), cases were selected as those with a K30 ICD10 record in their healthcare data, while unrelated individuals were randomly drawn from the remainder of the cohort as population controls. The study was approved by the local ethics committee at the Estonian eHealth Foundation. EGCUT demographics are reported in Table 1.

MGI. MGI is a longitudinal cohort that includes participants recruited via the Michigan Medicine healthcare system, who provided consent for linking medical records to genetic data (Illumina HumanCoreExome v12.1 arrays).[19] With their informed consent participants gave "*permission for researchers to use your samples and health information to study any disease or health condition*". We used a data freeze containing 40,000 individuals of PCA-derived European ancestry for the study, which was reviewed and approved by the Michigan University Institutional Review Board. The demographics of MGI individuals are reported in Table 1.

FD patients from TARGID. Consecutive FD patients (N=192) were recruited at the Translational Research Center for Gastrointestinal Disorder (TARGID), University Hospital Leuven, Belgium, according to Rome III criteria. Individuals were excluded if they showed abnormal findings on upper GI endoscopy, and/or if they had a history of former upper digestive surgery, diabetes, IBS, CD, IBD or other predominant symptoms typically associated with perturbed upper GI motility. FD patients' genotypes were extracted from available Illumina GSA data. Patients provided written informed consent and the study was approved by the local ethics committee (S52032). Belgian population control genotype data (Illumina HumanOmni 5M Exome) were obtained from a dataset previously described (N=442).[20] The demographics are reported in Table 1.

Cross-disease phenotype analysis

Associations between FD and other ICD10 common diagnoses (≥ 1% prevalence), self-reported conditions, and use of medications from UKBB data were tested in a logistic regression model adjusting for sex and age using base functions implemented in R (v3.6.1) A False Discovery Rate (Benjamini & Hochberg) correction for multiple tests (corrected α=0.05)

was included in the calculation of statistical significance, in order to control for type I error. Results obtained for ICD10 diagnoses were visualized with circular plots using the Circlize (v0.4.8) package (http://cran.r-project.org/web/packages/circlize/) implemented in R (v3.6.1).

Genotype quality control, imputation and individual GWAS

A common pipeline with minor modifications was applied to all cohorts for quality control (QC) of genotype data, imputation, and FD GWAS tests. Briefly, genotype QC filters were applied per sample (missing rate <95-99%; heterozygosity rate >3*SD) and per marker (call rate >95-99%; Hardy-Weinberg equilibrium $P<1\times10^{-4}$), and individuals of non-European ancestry (detected from principal component analysis [PCA]), with excessive relatedness (KING kinship coefficient >.0663), or with genotype-phenotype sex discrepancy were excluded from the analyses. Missing genotypes were imputed using the Haplotype Reference Consortium (HRC), UK10K, 1KG or Estonian reference panels, and only high quality (INFO>0.8) common (MAF>0.01) markers were included in downstream analyses. GWAS association tests were performed with a mixed linear model using SAIGE (v0.29.4.2),[21] in order to control for low case:control ratio and population stratification in UKBB and MGI, and logistic regression implemented in the Epacts software (https://genome.sph.umich.edu/wiki/EPACTS) for EGCUT. Association testing included sex, age, genotyping array and the first 4-10 principal components as covariates in the analyses.

GWAS meta-analysis

Individual GWAS results were brought forward into the meta-analysis pipeline, with absence of population stratification (inflation factor λ range 0.97-1.03). Prior to meta-analysis, individual GWAS association results were inspected with the R package EasyQC (v9.2),[22] in order to check for data integrity, remove invalid or unmapped markers, and harmonize SNP rsIDs and allele strand coding across datasets. A total of 7,347,476 high-quality SNP markers passing QC, with available summary statistics from at least 2 datasets, and absence of heterogeneity across studies (Cochran's Q > 0.05) were brought forward into a meta-analysis performed with METAL (v2011-03-25)[23] and the fixed-effect model weighted by p-value (showing no population stratification based on a Lambda inflation factor = 1.0195).

Downstream annotation of GWAS meta-analysis results

Annotation of suggestive ($P<5.0\times10^{-6}$) FD risk loci was done with Functional Annotation of GWAS (FUMA) v1.3.5 (https://fuma.ctglab.nl/).[24] The 13 association signals were identified based on SNP P-value and linkage disequilibrium between markers (LD; r2<0.4), and lead SNPs defined as those with the lowest p-values in each region. Gene content at FD loci was annotated based on positional, expression quantitative trait loci (eQTL) and 3D chromatin interactions also using FUMA with default parameters. Gene-set enrichment analysis for biological pathways, functions, or tissues was performed using FUMA, and False Discovery Rate was used for multiple testing correction.

PheWAS and genetic correlation analyses

A phenome-wide association study (PheWAS) approach was used to identify known associations ($P<5\times10^{-8}$) of other traits at FD GWAS suggestive loci using PhenoScanner2.[25,26] Partitioned FD heritability (h^2_{SNP}) on the liability scale was calculated as described (https://nealelab.github.io/UKBB_ldsc/methods.html). Genetic correlation (r_g) with other traits was computed using linkage disequilibrium score regression (LDSC)[27] implemented in the online platform CTG-VL (https://vl.genoma.io/). False discovery rate (FDR) correction was adopted in order to control for type I error.

Pilot replication in a case-control cohort

Genotypes of lead SNPs (or best LD proxies with minimum r2>0.5) from FD GWAS suggestive loci were extracted from QC'ed imputed genotype data available for TARGID patients and Belgian controls. GWAS effect alleles were tested for replication using 1-tailed P under an additive model with logistic regression adjusting for sex, age and first 10 principal components from PCA analysis in Plink (v1.9, www.cog-genomics.org/plink/1.9/).[28]

Data availability

FD GWAS meta-analysis results have been deposited in the GWAS Catalog (https://www.ebi.ac.uk/gwas/) and are publicly available under accession nr GCST90010719.

RESULTS

UK Biobank participants with FD

After excluding potentially confounding diagnoses, participants of non-European ancestry and population outliers, related individuals, and participants with poor phenotype or genotype data, a total of 300,152 UKBB participants were included in the study (Table 1). This subset of UKBB was selected for all analyses described below, in order to be able to compare phenotype and genetic cross-disease results. Among these, 5,950 had a K30 diagnosis of functional dyspepsia in their medical records, and were therefore selected as FD cases, while the remainder of the population (N=294,202) were assigned to the control group. FD cases were more likely to be female and older than controls (respectively, OR=1.6; $P=5x10^{-62}$ and OR=1.03; $P=9.99x10^{-88}$).

Cross-disease analyses I: ICD10 diagnoses

We tested common (≥ 1% occurrence in UKBB) medical conditions for their differential prevalence in FD cases and controls, based on logistic regression adjusting for sex and age. UKBB diagnoses were studied at various ICD10 code levels (Chapters to four-digit codes), for a total of 99,350,312 data points (331 common ICD10 codes in 300,152 individuals). In these analyses, we observed 284 codes differentially distributed in FD cases and controls (Figure 1 and Supplementary Table 1). Most ICD10 codes showed a significant increase in FD cases at the Chapter level, including higher risk of *Diseases of the digestive system*

(K00-K93; OR=4.0 [95% CI 3.8-4.2], P<1.0x10^{-300}), *Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified* (R00-R99; OR=3.2 [95% CI 3.0-3.4], P<1.0x10^{-300}), and *Certain infectious and parasitic diseases* (A00-B99; OR=2.1 [95% CI 2.0-2.3], P=1.5x10^{-73}). Expectedly, while reflux-related diagnoses were among the exclusion criteria (see Methods), *Diseases of oesophagus, stomach and duodenum* (K20-K31) still showed the strongest association and most pronounced risk effects (OR=8.2 [95% CI 7.7-8.7], P<1.0x10^{-300}). At the four-digit level (ie most informative, detailed diagnosis code) *Gastritis, unspecified* (K29.7) and *Hernia without obstruction or gangrene* (K44.9) accounted for the strongest risk effects (respective OR=5.4 [95% CI 5.0-5.9] and OR=5.9 [95% CI 5.5-6.4], both with P values<1x10^{-100}) within the digestive disease domain. A high association was also observed for *cholelithiasis* (K80, OR=3.1 [95% CI 2.8-3.4], P<4.5x10^{-129}) among the digestive *Disorders of gallbladder, biliary tract and pancreas* (K80-K87). Similarly, most prominent R00-R99 differences were observed for *Symptoms and signs involving the digestive system and abdomen* (R10-R19) like *Dysphagia* (R13) and *Nausea and vomiting* (R11) (respective OR=6.0 [95% CI 5.3-6.7] and OR=3.2 [95% CI 2.9-3.5], both P values<1x10^{-100}), together with a series of codes related to pain and the circulatory system like *abdominal pain* (R10.1, R10.4) and *chest pain* (R07.3 and R07.4) (ORs>2, P values<1x10^{-11}). Of interest, other notable associations (ORs>2, P values<1x10^{-2}) were detected for *Neurotic, stress-related and somatoform disorders* (F40-F48) including *anxiety disorders* (F41), as well as *Ischaemic heart diseases* (I20-I25) including *chronic ischaemic heart disease* (I25.8 and I25.9), *angina* (I20.0 and I20.9) and *myocardial infarction* (I25.2), and finally *Diseases of the musculoskeletal system and connective tissue* (M00-M99) like *osteoporosis* (M81) and *spondylosis* (M47) (Figure 1 and Supplementary Table 1).

Cross-disease analyses II: conditions and medications, self-reported

We also studied UKBB data collected via touchscreen questionnaire at UKBB participants' enrolment, in relation to common (occurrence ≥1%, N=55) self-reported non-cancer illnesses and use of medications. We detected significant FD association for 15/55 self-reported conditions (Table 2), notably often from disease domains similar to those highlighted in the ICD10 analyses. Expectedly, strongest association was observed for self-reported *gastroesophageal reflux* (OR=4.3 [95% CI 3.9-4.6], P=3.8x10^{-296}). In addition, *hiatus hernia*, *IBS*, *anxiety/panic attacks*, *angina*, *heart attack*/*myocardial infarction* and *osteoporosis*/*spondylitis* were also markedly increased, (Table 2), similar to what observed for diagnoses directly accessed from healthcare records (ICD10). As expected, use of *medication for pain relief, constipation or heartburn* was more common in FD patients (OR=0.55 [95% CI 0.5-0.6], P=2.7x10^{-113} when testing *none of the above* from questionnaire data, Table 2), with the PPI *Omeprazole* (OR=4.4 [95% CI 4.1-4.7], P<1.0x10^{-300}) and the H2

receptor blocker *Ranitidine* (OR=3.1 [95% CI 2.7-3.5], P=6.2x10^{-71}) showing strongest associations.

Genetic analyses: GWAS and follow-up

For the purpose of genetic analyses, we produced GWAS data from UKBB, MGI and EGCUT (see Methods). Harmonization and QC filtering of individual datasets allowed GWAS association results to be included in a meta-analysis spanning 7,347,476 high-quality SNP markers in 10,078 FD cases and 351,282 non-FD controls. Due to its larger size, UKB had a more pronounced weight in the meta-analysis (Effective size, N_e=11,664.10) compared to EGCUT (N_e=5,900.16); and MGI (N_e=1,458.08). We detected no genome-wide significant association (P<5x10^{-8}), though 28 markers gave rise to suggestive (P<5x10^{-6}) signals mapped to 13 independent loci using FUMA (see Methods) (Table 3). Computational annotation of suggestive FD risk loci based on FUMA positional, eQTL and chromatin interaction mapping (Methods) resulted in the identification of 80 candidate genes (Table 3). Among these were genes related to neuronal development (*NFASC*, *CNTN2* or *NXPH1*), ion channels (*KCNE4*), adherens junctions (*PTPRK*), cell apoptosis (*DIDO1*) and cell migration (*GID8*). However, tissue-specific and gene-set enrichment analyses did not disclose any significant associations with specific biological pathways, functions, or tissues after correction for multiple testing (not shown). In a pilot experiment of replication, we tested lead SNPs (or their best LD proxies, see Methods) from the 13 GWAS risk loci in a small cohort of 192 well-characterized FD patients and 442 population controls from Belgium (Supplementary Table 2). We detected nominal significance for two loci tagged by markers rs2093045 on chromosome 20, and rs2189730 (a weak proxy for rs6463848, see Methods) on chromosome 9, with concordant genetic risk effects across all tested cohorts (Figure 2).

Cross-disease analyses III: PheWAS lookup and genetic correlations

We studied FD suggestive GWAS loci for their eventual relevance to other traits and conditions, by inspecting public repositories of GWAS data using PhenoScanner2 (see Methods). This highlighted 3/13 loci to be relevant to other diseases and traits, namely mean platelet counts (locus rs2595968), impedance measures, C-reactive protein and white blood cell counts (rs6719560), and B-cell lymphoma (rs184132620). Broader evidence of genetic correlation with other conditions and traits was obtained from larger analyses of FD GWAS summary statistics using LDSC (implemented in CTG-VL, see Methods). Although weak, FD showed detectable heritability (FD h^2_{SNP}=0.047, P=0.014), and significant genetic correlation with other upper GI diseases (*diaphragmatic hernia*, *gastritis* and *duodenitis*, *use of Omeprazole*, *self-reported reflux*), several pain-related traits (*pain in throat and chest*, *stomach*, *abdominal and pelvic pain*, *back pain*), personality traits (neuroticism), mood/anxiety disorders (*anxiety*, *depression*), and joint disorders (*osteoarthritis*) (Figure 3, Supplementary Table 3). Thus, once again, these observations were similar to those made in

cross-disease phenotypic analyses of UKBB ICD10 and self-reported diagnoses, suggesting shared genetic predisposition.

DISCUSSION

FD is one of the most common functional gastrointestinal disorders, as also highlighted by a recent global epidemiological study from the Rome Foundation.[2] In spite of its large prevalence, FD has received less attention than its lower-bowel counterpart, IBS, and many aspects of its pathogenesis remain poorly understood. In this study, we have used complementary approaches to gain insight into FD pathophysiology through a large-scale investigation of cross-disease overlap at the epidemiological and genetic level. We extracted FD and other diagnoses from UK Biobank participants' healthcare records and self-reported questionnaires, and we produced and exploited GWAS results for genetic correlation analyses. Hence, with the limitation that ICD10 and self-reported diagnoses could not be formally verified by inspection of individual medical notes, we hereby report the largest population-based survey of FD to-date.

The overlap between FD, other GI conditions and, to a lesser extent, personality traits and mood/anxiety disorders have been studied in a few surveys, while the analysis of other conditions has been scarce. The most comprehensive survey, where several disease domains were investigated, included 1,669 FD cases and 83,450 matched controls, with clinical data retrieved from paid health insurance claims.[29] In the current report we studied >300,000 UK Biobank participants including 5,950 FD cases ascertained via their healthcare records, and interrogated 335 diagnostic codes at four ICD10 classification levels, making it the largest and most comprehensive survey of FD comorbidities.

A first important observation from these analyses is the large number of conditions (N=281) whose prevalence is significantly higher in FD patients compared to controls. This is consistent with previous reports on insurance claims,[29] as well as studies looking at FD patients' self-evaluation of their health status.[30] Overall, therefore, there appears to be compelling evidence that FD patients generally experience worse health conditions than the rest of the population, a notion that may have implication for the clinical management of this common disorder and the pressure it poses on the healthcare system. Also noteworthy, cross-disease analyses gave rise to remarkably consistent results across ICD10 diagnoses identified via linkage to healthcare records, self-reported conditions from questionnaire data, and genetic correlation using GWAS summary statistics. This suggests that FD and comorbid conditions co-manifest because of shared genetic architecture and the contribution of genes and pathways with multiple (pleiotropic) disease risk effects on a spectrum of ailments.

A series of conditions from the digestive disease domain were expectedly found to occur more often in FD patients than controls, either as ICD10 codes from healthcare records or

self-reported traits. For instance, while excluded in the ICD10 analyses (see Methods), IBS and reflux (including its prescriptions drugs omeprazole and ranitidine) were among the conditions most often self-reported by FD patients, and are notoriously associated with dyspepsia.[31–34] Similarly, gastritis, hiatus hernia and eventually chest pain and angina, which were all linked to FD in UK Biobank (as ICD10 and self-reported diagnoses as well as related medications), may represent proxies for reflux due to heartburn and/or other overlapping symptoms. At the same time, upper abdominal pain is a key feature of FD patients with epigastric pain syndrome,[35] but can also occur postprandially in the postprandial distress syndrome group.[36] Finally among digestive diseases, gallstone prevalence was much higher in FD patients, who in fact represent 6.3% of cholelithiasis cases among the selected UK Biobank participants, an observation that may be relevant to studying the source of symptoms and avoiding the risk of ineffective cholecystectomies.[37] We also detected increased prevalence of personality traits and mood/anxiety disorders among FD patients (both as ICD10 and self-reported conditions), which is consistent with previous reports and the fact that FD individuals tend to show a higher degree of somatization.[38,39] In particular, the prevalence of anxiety and depression has been shown to correlate with the number and severity of comorbid FGIDs,[40] again an observation made here in relation to the high number of co-manifestations detected in FD cases compared to the rest of UK Biobank. Finally, several other associations were detected that represent novel observations, including the associations with osteoarthropathies and, to a lesser extent, diseases of the circulatory system. Association with osteoporosis was detected in the analysis of ICD10 and self-reported diagnoses but not among the significant genetic correlations, which could be interpreted as a possible consequence of long-term PPI use and its known effect on bone density.[41] Spondylosis, spondylitis and other diseases of the bone and joints were linked to FD both in the survey of ICD10 and self-reported diagnoses and in the genetic correlations with other conditions and traits, suggesting a possible common pathogenetic denominator. Increased prevalence of rheumatological disorders has been previously reported in FGIDs,[42,43] while spondylosis has not been studied in FD. Of interest, cervical spondylosis has been recently proposed to induce FGIDs in an experimental rat model.[44] Heart disease has been studied in relation to upper gastrointestinal symptoms, and FD patients appear to more often suffer from circulatory system disease, myocardial infarction, angina, and chest pain.[45] While some risk factors (smoking, obesity, high blood pressure) are known to be common to FD and heart disease, the explanation for the observed comorbidity has often been sought in the side effects of respective pharmacological treatments. However, the LDSC results reported here, where significant genetic correlation is detected for FD and other heart conditions, suggest commonalities exist in the genetic architecture of these traits, which may thus arise from similar, partially overlapping pathophysiological mechanisms.

We also report here the first GWAS and meta-analysis of FD, performed exploiting genotype data and information on ICD10 diagnoses from healthcare records in three large population-based cohorts. Although more than 360,000 individuals were included in the analyses, the total number of FD cases was limited to 10,078, and likely not enough to identify robust association signals at the genome-wide level of significance ($P<5x10^{-8}$). This is not surprising given that FD shows detectable but only weak heritability (5%) and, similar to other common FGIDs like IBS, this represents a challenge for gene hunting efforts requiring massive sample size. However, genetic correlation analyses (which do not solely rely on genome-wide significant signals) suggest that the genetic factors predisposing to FD also contribute to risk of several other conditions, indeed many of which often co-manifesting in FD, as reported in this and previous surveys. Hence, genetic liability in FD may extend across a much wider spectrum of conditions and traits, with potential implications for a better understanding of its pathophysiology and the identification of therapeutically actionable pathways from the study of genetically overlapping diseases.

Finally, while the thirteen GWAS-suggestive signals identified here may represent best candidates for investigation in future studies, we report preliminary replication for 2 loci in a pilot case/control analysis of 192 FD patients defined according to gold-standard Rome III Criteria. This is for markers rs2189730 and rs2093045, respectively tagging FD suggestive risk loci on chromosomes 7 and 20. Among four genes mapped via chromatin interactions to the risk locus on chromosome 7, *NXPH1* likely represents the best candidate to play a role in FD, since polymorphisms in this gene have been previously reported to also affect risk of IBS.[46] *NXPH1* codes for neurexophilin-1, a glycoprotein ligand of the α-neurexin receptors involved in synaptic neurotransmission and plasticity in the brain.[47] Neurexophilin-1 is primarily expressed in inhibitory neurons where it modulates γ-aminobutyric acid (GABA) receptor-mediated signalling, thus possibly affecting FD-relevant motor functions and GI motility similar to what has been shown for other members of the neurexophilin family.[48] Several genes map to the chromosome 20 risk locus, including *GID8*, *DIDO1*, *SLC17A9* and *TCFL5,* whose expression is detected in human oesophagus and is affected by eQTLs for the rs2093045 lead SNP. Among these, *GID8* may be of particular interest as it codes for TWA1, a nuclear protein that is an important member of the CTLH complex regulating cell migration,[49] hence potentially relevant to upper GI development and integrity.

In summary, we report here a large-scale survey of FD comorbidity and genetic predisposition, which highlights considerable clinical and genetic overlap with several other conditions. This novel knowledge contributes to a better understanding of FD pathophysiology, and may be relevant to the identification of actionable pathways from the study of common disease mechanisms.

ACKNOWLEDGMENTS

This research has been conducted using the UK Biobank Resource under Application Number 30537. The authors acknowledge the University of Michigan Precision Health Initiative and Medical School Central Biorepository for providing biospecimen storage, management, processing and distribution services and the Center for Statistical Genetics in the Department of Biostatistics at the School of Public Health for genotype data curation, imputation, and management in support of this research.

FUNDING

Supported by grants from the Spanish Ministry of Economy and Competitiveness (FIS PI17/00308) and the Swedish Research Council (VR 2017-02403) to MDA. The Florey Institute of Neuroscience and Mental Health acknowledges the strong support from the Victorian Government and in particular the funding from the Operational Infrastructure Support Grant. Funding was provided by a Methusalem grant from Leuven University to JT.

CONFLICT OF INTEREST

The authors have no competing interests.

AUTHORS CONTRIBUTION

JT and MD: study concept and design; FC, MTL, AP, LH, VT, RL, AF, SZ, MB, MZ, TE, JT, MDA: cohorts, patients characterization, data collection; KGE, MTL, AP, MZ: statistical analyses; KGE, FC, MTL, LB, MZ, TE, JT, MD: data analysis and interpretation; MD: obtained funding, administrative and technical support, study supervision; KGE, FC, JT and MD: drafted the manuscript, with input and critical revision from all other authors.

REFERENCES

1. Stanghellini V, Chan FKL, Hasler WL, et al. Gastroduodenal disorders. *Gastroenterology*. 2016;150(6):1380-1392.
2. Sperber AD, Bangdiwala SI, Drossman DA, et al. Worldwide Prevalence and Burden of Functional Gastrointestinal Disorders, Results of Rome Foundation Global Study. *Gastroenterology*. May 2020.
3. Drossman DA. Functional gastrointestinal disorders: History, pathophysiology, clinical features, and Rome IV. *Gastroenterology*. 2016;150(6):1262–1279.e2.
4. Piessevaux H, De Winter B, Louis E, et al. Dyspeptic symptoms in the general population: A factor and cluster analysis of symptom groupings. *Neurogastroenterol Motil*. 2009;21(4):378-388.

5. Camilleri M, Dubois D, Coulie B, et al. Prevalence and socioeconomic impact of upper gastrointestinal disorders in the United States: Results of the US upper gastrointestinal study. *Clin Gastroenterol Hepatol*. 2005;3(6):543-552.
6. Brook RA, Kleinman NL, Choung RS, Melkonian AK, Smeeding JE, Talley NJ. Functional Dyspepsia Impacts Absenteeism and Direct and Indirect Costs. *Clin Gastroenterol Hepatol*. 2010;8(6):498-503.
7. Ford AC, Forman D, Bailey AG, Axon ATR, Moayyedi P. Effect of dyspepsia on survival: A longitudinal 10-year follow-up study. *Am J Gastroenterol*. 2012;107(6):912-921.
8. Camilleri M, Stanghellini V. Current management strategies and emerging treatments for functional dyspepsia. *Nat Rev Gastroenterol Hepatol*. 2013;10(3):187-194.
9. Masuy I, Van Oudenhove L, Tack J. Review article: treatment options for functional dyspepsia. *Aliment Pharmacol Ther*. 2019;49(9):1134-1172. doi:10.1111/apt.15191
10. Tack J. Prokinetics and fundic relaxants in upper functional GI disorders. *Curr Opin Pharmacol*. 2008;8(6):690-696.
11. B Biomed GB, Carroll G, Mathe A, et al. Evidence for Local and Systemic Immune Activation in Functional Dyspepsia and the Irritable Bowel Syndrome: A Systematic Review. *Am J Gastroenterol*. 2019;114(3):429-436.
12. Cirillo C, Bessissow T, Desmet AS, Vanheel H, Tack J, Berghe P Vanden. Evidence for neuronal and structural changes in submucous ganglia of patients with functional dyspepsia. *Am J Gastroenterol*. 2015;110(8):1205-1215.
13. Carbone F, Tack J. Gastroduodenal mechanisms underlying functional gastric disorders. *Dig Dis*. 2014;32(3):222-229.
14. Lee KJ, Tack J. Duodenal Implications in the Pathophysiology of Functional Dyspepsia. *J Neurogastroenterol Motil*. 2010;16(3):251-257.
15. Gathaiya N, Locke GR, Camilleri M, Schleck CD, Zinsmeister AR, Talley NJ. Novel associations with dyspepsia: A community-based study of familial aggregation, sleep dysfunction and somatization. *Neurogastroenterol Motil*. 2009;21(9):922-e69.
16. Triantafyllou K, Kourikou A, Gazouli M, Karamanolis GP, Dimitriadis GD. Functional dyspepsia susceptibility is related to CD14, GNB3, MIF, and TRPV1 gene polymorphisms in the Greek population. *Neurogastroenterol Motil*. 2017;29(1):1-8.
17. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.
18. Leitsalu L, Haller T, Esko T, et al. Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *Int J Epidemiol*. 2015;44(4):1137-1147.
19. Fritsche LG, Gruber SB, Wu Z, et al. Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative.

Am J Hum Genet. 2018;102(6):1048-1061.

20. Meschia JF, Arnett DK, Ay H, et al. Stroke genetics network (SiGN) study design and rationale for a genome-wide association study of ischemic stroke subtypes. *Stroke*. 2013;44(10):2694-2702.
21. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50(9):1335-1341.
22. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192-1212.
23. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191.
24. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826.
25. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: A database of human genotype-phenotype associations. *Bioinformatics*. 2016;32(20):3207-3209.
26. Kamat MA, Blackshaw JA, Young R, et al. PhenoScanner V2: An expanded tool for searching human genotype-phenotype associations. *Bioinformatics*. 2019;35(22):4851-4853.
27. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241.
28. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
29. Brook RA, Kleinman NL, Choung RS, Smeeding JE, Talley NJ. Excess comorbidity prevalence and cost associated with functional dyspepsia in an employed population. *Dig Dis Sci*. 2012;57(1):109-118.
30. Jarbøl DE, Rasmussen S, Balasubramaniam K, Elnegaard S, Haastrup PF. Self-rated health and functional capacity in individuals reporting overlapping symptoms of gastroesophageal reflux disease, functional dyspepsia and irritable bowel syndrome - a population based study. *BMC Gastroenterol*. 2017;17(1):1-9.
31. De Bortoli N, Tolone S, Frazzoni M, et al. Gastroesophageal reflux disease, functional dyspepsia and irritable bowel syndrome: Common overlapping gastrointestinal disorders. *Ann Gastroenterol*. 2018;31(6):639-648.
32. Aziz I, Palsson OS, Törnblom H, Sperber AD, Whitehead WE, Simrén M. Epidemiology, clinical characteristics, and associations for symptom-based Rome IV functional dyspepsia in adults in the USA, Canada, and the UK: a cross-sectional population-based study. *Lancet Gastroenterol Hepatol*. 2018;3(4):252-262.

33. von Wulffen M, Talley NJ, Hammer J, et al. Overlap of Irritable Bowel Syndrome and Functional Dyspepsia in the Clinical Setting: Prevalence and Risk Factors. *Dig Dis Sci*. 2019;64(2):480-486.
34. Geeraerts A, Van Houtte B, Clevers E, et al. Gastroesophageal Reflux Disease—Functional Dyspepsia Overlap: Do Birds of a Feather Flock Together? *Am J Gastroenterol*. 2020;115(8):1167-1182.
35. Bharucha AE, Chakraborty S, Sletten CD. Common Functional Gastroenterological Disorders Associated With Abdominal Pain. *Mayo Clin Proc*. 2016;91(8):1118-1132.
36. Carbone F, Vanuytsel T, Tack J. Analysis of Postprandial Symptom Patterns in Subgroups of Patients With Rome III or Rome IV Functional Dyspepsia. *Clin Gastroenterol Hepatol*. 2020;18(4):838-846.e3.
37. Latenstein CSS, de Jong JJ, Eppink JJ, et al. Prevalence of dyspepsia in patients with cholecystolithiasis: a systematic review and meta-analysis. *Eur J Gastroenterol Hepatol*. 2019;31(8):928-934.
38. Aro P, Talley NJ, Ronkainen J, et al. Anxiety Is Associated With Uninvestigated and Functional Dyspepsia (Rome III Criteria) in a Swedish Population-Based Study. *Gastroenterology*. 2009;137(1):94-100.
39. Clauwaert N, Jones MP, Holvoet L, et al. Associations between gastric sensorimotor function, depression, somatization, and symptom-based subgroups in functional gastroduodenal disorders: are all symptoms equal? *Neurogastroenterol Motil*. 2012;24(12):1088-e565.
40. Pinto-Sanchez MI, Ford AC, Avila CA, et al. Anxiety and depression increase in a stepwise manner in parallel with multiple FGIDs and symptom severity and frequency. *Am J Gastroenterol*. 2015;110(7):1038-1048.
41. Andersen BN, Johansen PB, Abrahamsen B. Proton pump inhibitors and osteoporosis. *Curr Opin Rheumatol*. 2016;28(4)..
42. Koloski N, Jones M, Walker MM, et al. Population based study: atopy and autoimmune diseases are associated with functional dyspepsia and irritable bowel syndrome, independent of psychological distress. *Aliment Pharmacol Ther*. 2019;49(5):546-555.
43. Ford AC, Talley NJ, Walker MM, Jones MP. Increased prevalence of autoimmune diseases in functional gastrointestinal disorders: case–control study of 23 471 primary care patients. *Aliment Pharmacol Ther*. 2014;40(7):827-834.
44. Song X-H, Xu X-X, Ding L-W, Cao L, Sadel A, Wen H. A preliminary study of neck-stomach syndrome. *World J Gastroenterol*. 2007;13(18):2575-2580.
45. Coleman CI, Limone BL, Schein JR, et al. Upper Gastrointestinal Symptoms and Cardiovascular Disease. In: Curley M, Shaffer E, eds. *Dyspepsia - Advances in*

Understanding and Management. London: IntechOpen Limited; 2013:135-168.

46. Wouters MM, Lambrechts D, Knapp M, et al. Genetic variants in CDC42 and NXPH1 as susceptibility factors for constipation and diarrhoea predominant irritable bowel syndrome. *Gut*. 2014;63:1103-1111.

47. Born G, Breuer D, Wang S, et al. Modulation of synaptic function through the α-neurexin–specific ligand neurexophilin-1. *Proc Natl Acad Sci*. 2014;111(13):E1274-E1283.

48. Beglopoulos V, Montag-Sallaz M, Rohlmann A, et al. Neurexophilin 3 Is Highly Localized in Cortical and Cerebellar Regions and Is Functionally Important for Sensorimotor Gating and Motor Coordination. *Mol Cell Biol*. 2005;25(16):7278-7288.

49. Salemi LM, Maitland MER, Yefet ER, Schild-Poulter C. Inhibition of HDAC6 activity through interaction with RanBPM and its associated CTLH complex. *BMC Cancer*. 2017;17(1):1-16.

SUPPORTING INFORMATION

Supplementary Table S1: Co-morbidities in FD cases and controls based on ICD10 diagnoses from UKBB Biobank

Supplementary Table S2: Replication of GWAS findings in an independent case-control cohort.

Supplementary Table S3: Genetic correlation of functional dyspepsia with other traits.

608 **Table 1.** Demographics of cohorts included in the study.

	FD cases	CTRLS
UKB	5,950	294,202
Age (SD)	58.6 (7.4)	56.6 (8.0)
% Female	64.4	53.6
EGCUT	3,383	23,054
Age (SD)	42.3 (18.2)	42.3 (17.2)
% Female	72.3	64.2
MGI	745	34,026
Age (SD)	57.6 (14.8)	56.0 (16.2)
% Female	61.7	52.1
TARGID	192	442
Age (SD)	43.0 (14.2)	55.8 (14.3)
% Female	81.2	53.4

610 **Table 2.** Prevalence of self-reported conditions, diagnoses, traits and medications in UKB
611 patients and controls.

	FD %	CTRL %	P_{FDR}	OR (95%CI)
Self-reported conditions, diagnoses or traits				
gastro-oesophageal reflux (gord) / gastric reflux	16.3	4.3	3.8E-296	4.3 (3.9 - 4.6)
hiatus hernia	6.3	2.0	6.3E-73	3.0 (2.6 - 3.3)
irritable bowel syndrome	5.6	2.7	4.1E-29	2.0 (1.8 - 2.3)
angina	6.6	3.7	5.2E-26	1.9 (1.7 - 2.1)
cholelithiasis/gall stones	4.4	2.0	2.1E-20	1.9 (1.7 - 2.2)
heart attack/myocardial infarction	3.9	2.8	2.1E-08	1.6 (1.3 - 1.8)
anxiety/panic attacks	3.0	1.9	2.6E-08	1.6 (1.4 - 1.9)
back problem	3.3	2.3	4.6E-06	1.5 (1.3 - 1.7)
osteoporosis	3.7	2.0	6.2E-06	1.5 (1.3 - 1.7)
osteoarthritis	14.8	11.1	1.4E-05	1.2 (1.1 - 1.3)
spine arthritis/spondylitis	1.8	1.1	2.7E-04	1.5 (1.2 - 1.9)
asthma	16.5	15.0	3.1E-04	1.2 (1.1 - 1.3)
migraine	5.5	4.2	4.0E-04	1.3 (1.1 - 1.4)
emphysema/chronic bronchitis	2.3	1.6	1.4E-03	1.4 (1.2 - 1.7)
diverticular disease/diverticulitis	2.2	1.4	2.7E-03	1.4 (1.1 - 1.7)
prolapsed disc/slipped disc	2.9	2.3	3.7E-03	1.3 (1.1 - 1.6)
depression	8.8	7.7	5.9E-03	1.2 (1.1 - 1.3)
ovarian cyst or cysts	1.8	1.1	5.9E-03	1.4 (1.1 - 1.7)
enlarged prostate	2.1	2.0	8.3E-03	1.4 (1.1 - 1.6)
pulmonary embolism +/- dvt	1.5	1.0	2.0E-02	1.4 (1.1 - 1.7)
Self-reported medications				
Omeprazole (e.g. Zanprol)	18.2	4.4	<1.0E-300	4.4 (4.1-4.7)
None of the above (Medication for pain relief, constipation, heartburn)	41.9	58.1	2.7E-113	0.5 (0.5-0.6)
Ranitidine (e.g. Zantac)	4.8	16	6.2E-71	3.1 (2.7-3.5)
Paracetamol	29.6	21.2	9.9E-50	1.5 (1.5-1.6)
Laxatives (e.g. Dulcolax, Senokot)	5.5	2.7	1.5E-23	1.8 (1.6-2.0)
None of the above (Medication for cholesterol, blood pressure, diabetes, or exogenous hormones) (female)	62.2	70.4	9.3E-14	0.8 (0.7-0.8)
Hormone replacement therapy (female)	10.1	7.3	4.4E-10	1.4 (1.3-1.6)
Cholesterol lowering medication (female)	17.5	12.0	8.6E-09	1.3 (1.2-1.4)
None of the above (Medication for cholesterol, blood pressure or diabetes) (male)	59.1	67.6	1.6E-07	0.8 (0.7-0.8)
Cholesterol lowering medication (male)	30.0	22.8	4.6E-07	1.3 (1.2-1.4)
Blood pressure medication (female)	22.0	16.8	1.8E-04	1.2 (1.1-1.3)
Aspirin	15.4	13.7	2.3E-02	1.1 (1.0-1.2)
Blood pressure medication (male)	29.2	24.4	2.5E-02	1.1 (1.0-1.3)

612
613

614 **Table 3**. GWAS meta-analysis association results and annotation of suggestive loci.

Lead SNP	Chr	EA	OA	EAF	Z-Score	P	Nearest gene	Mapped genes
rs2595968	1q32.1	A	G	0.283	4.8	2.0E-06	NFASC	NFASC[p,c], CNTN2[e,c], TMEM81[e,c], CDK18[e], MDM4[c], LRRN2[c], DSTYK[c], TMCC2[c], KLHDC8A[c]
rs6719560	2q36.1	T	C	0.108	4.9	8.4E-07	KCNE4	KCNE4[p,e,c], FARSB[e,c], PAX3[c], CCDC140[c], MOGAT1[c], SCG2[c], AP1S3[c], WDFY1[c], MRPL44[c], FAM124B[c], CUL3[c]
rs144885331	3p24.3	C	T	0.027	4.9	8.2E-07	ZNF385D	ZNF385D[p]
rs9868674	3q22.1	T	C	0.410	4.7	2.5E-06	CPNE4	CPNE4[p,e], ASTE1[c], NEK11[c], MRPL3[c]
rs80062354	3q26.31	T	C	0.194	-4.8	1.7E-06	SPATA16	SPATA16[p,c], NLGN1[c], NAALADL2[c]
rs961136	5p15.33	T	C	0.014	4.6	4.7E-06	IRX2	ADAMTS16[c]
rs67375755	5q34	T	C	0.065	-4.9	1.0E-06		HMMR[c], MAT2B[c], TENM2[c]
rs17245411	6q22.33	C	T	0.040	-4.6	3.3E-06	PTPRK	PTPRK[p,e],THEMIS[c], LAMA2[c], TMEM244[c], L3MBTL3[c]
rs6463848	7p21.3	T	C	0.404	-5	5.2E-07		NXPH1[c], NDUFA4[c], PHF14[c], THSD7A[c]
rs9696092	9p21.3	C	T	0.459	-4.6	4.7E-06	CDKN2B	MTAP[c], RP11-145E5.5[c], C9orf53[c], CDKN2A[c], CDKN2B[c]
rs184132620	15q21.2	C	A	0.014	-4.6	3.5E-06	ARPP19	ARPP19[p,c], FAM214A[p,c], LYSMD2[e], SCG3[c], TMOD3[c], LEO1[c], MAPK6[c], BCL2L10[c], GNB5[c], MYO5C[c], UNC13C[c], RSL24D1[c], RAB27A[c], PIGB[c], DYX1C1[c], UQCRFS1[c]
rs17597505	19q13.11	T	C	0.051	4.6	3.8E-06	-	VSTM2B[c], POP4[c], C19orf12[c], CCNE1[c], URI1[c], ZNF536[c], TSHZ3[c], DPY19L3[c], PDCD[c], SLC7A9[c]
rs2093045	20q13.33	C	T	0.323	-4.8	1.8E-06	SLC17A9	DIDO1[p,e,c], GID8[p,e,c], SLC17A9[p,e,c], COL9A3[e], TCFL5[e,c], BHLHE23[c]

615 EA, effect allele; OA, other allele; EAF, effect allele frequency; Nearest gene, nearest gene within 100kb from lead SNP (if any). Mapped genes include
616 protein coding genes mapped by FUMA, [p]genes physically mapped; [e]genes mapped by eQTLs; [c]genes mapped by chromatin interactions.

FIGURE LEGENDS

Figure 1. Results of cross-disease analysis of selected ICD10 diagnoses in UK Biobank. The results are schematically represented with a circos plot, where each circle represents a level of ICD10 coding (chapter, block and three-digits levels, respectively from inner to outer circles). Odds-ratios (OR) are reported on the Y axis at each level, while statistical significance is expressed across a color gradient of p values.

Figure 2. GWAS and replication results for selected markers. The Forest plots shows association results for SNPs rs6463848 and rs2093045, reported for each with respective odds-ratio (OR) and 95% confidence intervals, according to sample size (size of the symbol). (*) For the TARGID case/ctrl analysis, data for the rs6463848 locus correspond to marker rs2189730, which is the best LD proxy (r2=0.5) with genotype data available in this cohort.

Figure 3. Results of LDSC analyses and genetic overlap with selected traits. Genetic overlap is reported as regression score (r_g) with standard error, for selected conditions, diagnoses or traits from different disease domains. ICD10: conditions, diagnoses or traits based on healthcare records; self-reported: conditions, diagnoses or traits based on questionnaires data.