Màster Universitari
Anàlisi de Dades Òmiques / Omics Data Analysis
FACULTAT DE CIÈNCIES I TECNOLOGIA
UVIC | UVIC·UCC

Master of Science in Omics Data Analysis

Master Thesis

Investigating the Genetic Basis of the Co-occurrence of Irritable Bowel Syndrome and Anxiety

by

Dana Fakhreddine

Supervisor: Silvia Alemany, Group of Psychiatry, Mental Health and Addiction
Vall d'Hebron Research Institute (VHIR)

Co-supervisor: Marta Ribasés, Group of Psychiatry, Mental Health and Addiction
Vall d'Hebron Research Institute (VHIR)

Academic tutor: Mireia Olivella, Bioinformatics and Medical Statistics Group,
University of Vic – Central University of Catalonia

Biosciences Department

University of Vic – Central University of Catalonia

September 13, 2021

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

Investigating the Genetic Basis of the Co-occurrence of Irritable Bowel Syndrome and Anxiety

Fakhreddine, Dana

1 ABSTRACT

Irritable bowel syndrome (IBS) is a chronic gastrointestinal disorder representing a serious burden to the healthcare system. IBS research is extremely challenging due to the multifactorial etiology of the disease and the heterogeneity of patients that present high comorbidity rates with mental disorders including anxiety and depression. Such highly comorbid disorders show substantial heritability and are partly determined by a genetic component. In the present study, we used data available from large pre-existing genome-wide association studies on IBS (n=455,321) and anxiety (n=117,751) to assess the genetic overlap and causal relationship between these comorbid disorders and found strong and positive genetic correlation between them (rg=0.713, se= 0.076, P= 3.6e-20). The multi-trait analysis of GWAS (MTAG) highlighted three new genome-wide significant loci for IBS located in, or nearby, genes related to synaptic transmission, nervous system development, neuroticism, and epigenetic modification of chromatin. We also used Mendelian randomization with a range of sensitivity analyses to clarify the causal relationship between these disorders and found consistent evidence for a causal effect of the genetic liability of IBS on anxiety (P=7.6e-05) but no evidence of causal effect was detected in the opposite direction. Our results are consistent with a shared common genetic background between IBS and anxiety, highlight the importance of common genetic factors in the risk of these comorbid disorders, and add insight into the relationship between gastrointestinal and mental illnesses.

2 INTRODUCTION

Irritable bowel syndrome (IBS) is a chronic functional gastrointestinal disorder characterized by recurrent abdominal pain and altered bowel habits (1). With a global prevalence of 11%, it is associated with a substantial economic burden on healthcare systems worldwide. IBS is considered a complex genetic disorder whose manifestation is attributed to a combination of multiple genetic variants and an interplay among the central nervous system, immune activation of the intestinal mucosa, previous infections, the gut microbiome, and dietary habits (2–4). Evidence supporting the genetic contribution to IBS stems from family studies reporting that having a first-degree relative with IBS predicts IBS over other gastrointestinal disorders (5). Results from twin studies also support the genetic component of IBS given the higher concordance rate in monozygotic twins than dizygotic twins (5). On the other hand, the non-genetic contribution of environmental factors is expressed through an increased risk of IBS among non-related spouses who share a common household with similar lifestyle and dietary habits (6).

Genetic studies have been conducted to further understand the genetic basis of IBS and identify potential genetic risk factors. At least 60 candidate genes have been investigated in candidate gene case-control studies most of which are involved in serotonin synthesis reuptake, mucosal immune activation, neuropeptide signaling, inflammation, and intestinal secretion (7). However, such hypothesis-driven studies remain limited in design and sample size due to the challenges imposed by the clinical heterogeneity of IBS, its poorly characterized molecular pathway, and the lack of established biomarkers (5,7). To overcome these limitations, large-scale hypothesis-free analyses such as genome-wide association studies (GWAS) have presented a solid opportunity into a discovery-driven approach in uncovering genetic variants and genes associated with IBS (7). Despite the scarcity of clinical cohorts with available genetic data on this disorder, a number of genome-wide association studies have been performed on this disorder. A genetic association meta-analysis performed on five population-based cohorts detected seven genomic regions harboring 64 candidate genes affecting IBS (8). A recently published large-scale GWAS performed an array of gastrointestinal disorders revealed 2 other independent loci associated with IBS (2).

The burden of IBS is demonstrated through its association with poor quality of life, an impairment in social function and work abilities, and with psychological – psychiatric conditions such as anxiety and depression. A recent systematic review performed on 73 published articles revealed that the prevalence of anxiety and depression symptoms among IBS patients is 39.1% and 28.8%, respectively, compared to healthy controls (9). The growing body of evidence that links IBS with psychological – psychiatric conditions has led to the development of a biopsychosocial model for IBS that aims at understanding the bi-directional relationship between this gastrointestinal illness and the mind to help clinicians better understand the determinants of this disease, manage treatment, and improve clinical care (10). In fact, the close interaction between the physiological abnormalities of IBS with psychological stress, GI intestinal motor functioning, and the central nervous system activity is thought to be an important factor in the development of IBS symptoms through a communication system between the brain and the gut termed the brain-gut axis (10). This has been reflected in the clinic where IBS patients are often prescribed psychopharmacological such as tricyclic anti-depressants or psychological treatments such as cognitive behavioral therapy (10,11). However, the genetic underpinnings that explain the high rates of comorbidity between IBS and psychiatric symptoms remain largely unclear.

In this research, we aimed to further investigate the genetic basis of IBS and to examine the genetic overlap between IBS and anxiety. First, we jointly analyzed publicly available GWAS summary statistics of

IBS and anxiety to identify new IBS genetic variants and to explore the shared genetics between both disorders. Second, we conduct Mendelian randomization analyses to investigate the existence of a uni- or bi-directional causal relationship between them.

3 MATERIALS AND METHODS

3.1 SUMMARY STATISTICS

Publicly available SNP-level GWAS summary statistics for IBS and anxiety were used in this research. The IBS summary statistics data was obtained from the published article Wu et al. (2). This data was originally obtained from the UK Biobank (UKB data field: 131639) and was later curated by the authors by removing patients who were also diagnosed with inflammatory bowel disease (2). Thus, the sample size for the IBS summary statistics used is 455,321 with 28,518 cases and 426,803 controls, and the total number of SNPs analyzed is 8,546,066 (Table 1) (2).

For the anxiety trait, summary statistics from the binary phenotype *"Mental health problems ever diagnosed by a professional: Anxiety, nerves or generalized anxiety disorder*" (UKB data field: 20544_15) available from the UK Biobank summary statistics database was used. This dataset harbored information on a total of 117,751 individuals with 16,730 cases and 101,021 controls and a total of 13,791,467 SNPs (Table 1).

The UK Biobank is a prospective cohort study with intricate genetic and phenotypic data on approximately 500,000 individuals aged between 40 and 69 at the time of recruitment (between 2006 and 2010) recruited across 22 centers in the United Kingdom (12). All analysis done in this research is based on publicly available SNP-level GWAS summary statistics for each trait. No individual-level data was accessed or used in this research.

Table 1: Summary of the IBS and anxiety summary statistics information.

Trait	Number of SNP markers	cases	controls	total	sample prevalence	population prevalence
Irritable Bowel Syndrome (IBS)	8,546,066	28,518	426,803	455,321	0.063	0.11*
Anxiety	13,791,467	16,730	101,021	117,751	0.142	0.142**

Sample prevalence = cases/total for each trait.
**Obtained from Wu et al.* (2)
***Assumed equal to sample prevalence (see Methods section 3.2)*

3.2 SNP-Based Heritability and Genetic Correlation

SNP-based heritability (i.e. proportion of variance in trait liability attributable to genome-wide common SNPs) and genome-wide genetic correlation between the IBS and anxiety traits were estimated from the SNP-level information in the GWAS summary statistics of each trait using the single-trait and the cross-trait the linkage disequilibrium (LD) score regression methods, respectively (13,14). These methods rely on the fact that the GWAS effect size estimate of a given SNP incorporates the effects of all SNPs that are in linkage disequilibrium with this SNP (13).

The strength of these methods relies on their ability to distinguish between inflation in GWAS test-statistics caused by polygenicity (multiple small genetic effects) and confounding bias (population stratification and cryptic relatedness) due to the relationship of these factors with LD (14). Polygenicity that is affected by genetic variation is correlated with LD which is not the case with cryptic relatedness within cohorts or population stratification that are caused by genetic drift rather than LD.

Liability-scale LD score regression was performed to account for the population and sample prevalence of each trait. As such, the sample prevalence was calculated as the proportion of cases in each sample which was 0.063 for IBS and 0.142 for the studied anxiety-related phenotype. The population prevalence of the IBS trait was 0.11 as reported by Wu et al. (2), and due to the absence of population prevalence information on the UK Biobank anxiety phenotype selected for this study, its population prevalence was assumed to be equal to its sample prevalence; 0.142 (Table 1).

The LD score regression model software used in this research was developed by Neale's lab (LDSC v1.0.1) (https://github.com/bulik/ldsc), and analysis was conducted on European LD scores' reference panel available within the software.

3.3 Joint Analysis of IBS and Anxiety GWAS Summary Statistics

GWAS summary statistics of IBS and anxiety were jointly analyzed using multi-trait analysis of GWAS (MTAG) developed by Turley et al. (15). MTAG exploits information contained in the GWAS estimates of a trait to improve the effect estimates of other traits (15). MTAG, which is a generalization of inverse-variance-weighted meta-analysis, takes as input summary statistics from single-trait GWAS to output trait specific association studies (15). The key assumption of MTAG is that SNPs share the same variance-covariance matrix of effect sizes across traits. By default, and before running joint analysis of GWAS, MTAG automatically runs a set of default filters on the involved SNPs. First, SNPs with a minor allele frequency less than 1% are dropped (15). Second, for each trait, the variation in SNP sample size is restricted by

calculating the 90th percentile of the SNP sample size distribution and removing any SNP with a sample size smaller than 75% of this value (15). Finally, SNPs within regions that contain other SNPs with outlier effect sizes are removed (15). In our analysis, the sample size in our data was constant for all SNPs for each trait, so the second filter was not applied. In addition, strand ambiguous SNPs were not included.

Since summary statistics of both disorders used in this study were obtained from the UK Biobank, the possibility of a sample overlap was elevated. Yet, the use of MTAG was advantageous due to its use of cross-trait linkage disequilibrium that accounts for sample overlap (15).

Upon running MTAG on IBS and anxiety, two files containing updated SNP effect sizes and P-values were generated. SNPs were considered genome-wide significant if they achieved a P-value below 5×10^{-8}. The gain in statistical power for MTAG relative to GWAS characterized by the increase in the mean chi-squared (χ^2) statistic and the GWAS equivalent sample size needed to attain such statistical power was calculated based on the formula: $N_{GWAS-equiv,j} = N_{GWAS,j} \frac{\overline{\chi^2_{MTAG}} - 1}{\overline{\chi^2_{GWAS}} - 1}$

Where $N_{GWAS,j}$ is the sample size in the single-trait GWAS study for a given SNP j and χ^2_{GWAS} and χ^2_{MTAG} are the mean χ^2 statistics for the GWAS and MTAG studies, respectively.

For our MTAG analysis, we used the python code for MTAG available at https://github.com/omeed-maghzian/mtag via Python 2.7.16.

3.4 Identifying Hits

The PLINK 1.9 clumping algorithm was used to reveal independent lead SNP identified in the multi-trait analysis of GWAS (16). First, the SNP with the lowest P-value was selected as the index SNP, then all SNPs within $\pm 250kb$ and $r^2 > 0.5$ (squared correlation between the index SNP and the evaluated SNP) with respect to the lead SNP were included in this clump. After that, the SNP with the second lowest P-value outside the first clump was deemed the index SNP in the next clump that accepts SNPs with the same parameters. Only SNPs not yet included within a clump were evaluated by the algorithm for inclusion in the next clump. In this study, the P-value threshold selected for the index SNP was 5×10^{-8} while no P-value restriction was made to the clumped SNPs (selected P-value threshold was 1).

To identify whether MTAG led to the identification of new hits, overlap between the MTAG hits for IBS with the original GWAS hits for IBS was assessed within a $\pm$ 500,000kb distance.

3.5 Bayesian Credible Set Analysis

Bayesian statistics analysis developed by Maller et. al (17) was used as a refinement approach in the identification of causal variants. The posterior probability of each SNP was calculated for each region identified by the clumping algorithm described above (located $\pm 250kb$ and with $r^2 > 0.5$ with respect to index SNP (17). Credible sets were then identified by the set of SNPs with a $P < 0.05$ that account for 99% of the posterior probability within each region. Credible sets with a small number of SNPs were deemed informative in narrowing down the SNPs that could be causal.

Regional plots were then developed using LocusZoom.js v0.13.3 using the *GRCH37* genome assembly and the European population for LD score classification.

3.6 Biological Annotation & Gene Analysis

The annotation feature in MAGMA (Multi-marker Analysis of GenoMic Annotation) developed by de Leeuw et al. (18) was used to map SNPs to genes based on genomic location based on the NCBI 37.7 gene definitions. For this purpose, a window of 10kb upstream and 5kb downstream was used to extend the annotation region at the 5' and 3' ends of each gene, respectively, to account for SNPs near genes but outside their transcription regions. This method was used to identify the biological annotation of SNPs that scored a P-value $< 5 \times 10^{-8}$ within our data. In addition, the UCSC Variant Annotation Integrator with the GRCH37 genome assembly was used to define genomic regions captured by the identified loci.

This annotation method was also used as a pre-processing step for genome-wide gene-based analysis of the MTAG results for IBS where 18,234 genes were assessed for association with the studied traits with false-discovery rate (FDR) correction. The output of gene-based analysis was used to perform gene-set analysis using MSigDB gene sets denoting canonical pathways that include Kyoto Encyclopedia of Genes and Genomes (KEGG), BioCarta, Reactome, and Gene Ontology with FDR multiple-testing correction where statistical significance level was attained with an adjusted P-value < 0.05.

3.7 Mendelian Randomization

Two-sample Mendelian randomization (MR) analyses were conducted to assess whether causal relationships exist between IBS and anxiety. These analyses were performed in both directions, (i) using IBS as exposure and anxiety as outcome and (ii) using anxiety as exposure and IBS as outcome. The SNPs used as instruments in this analysis were obtained by PLINK 1.9 (16) clumping using the thresholds $r^2 = 0.5, p2 = 1, kb = 250$ as clumping parameters. Two index SNP P-values (p1) were used to run this analysis, $p1 < 5 \times 10^{-7}$ & $p1 < 5 \times 10^{-6}$. After removal of SNPs that overlapped between the studied

traits within a distance of 500,000kb and substitution of palindromic SNPs that have an intermediate allele frequency detected during analysis with their best proxies having an r^2 of at least 0.7, nine variants were included for IBS as exposure versus two for anxiety at $p1 < 5 \times 10^{-7}$. On the other hand, 32 variants were included for IBS as exposure versus 17 for anxiety at $p1 < 5 \times 10^{-6}$. Several sensitivity analyses were used to clarify the causal relationship between the disorders and to test whether the data presented fulfills Mendelian randomization's key assumptions that the genetic variants used as instruments:

(i) Are strongly associated with the exposure,

(ii) Are independent from confounding factors,

(iii) And only affect the outcome through the exposure and not through an independent pathway (19).

The main method used to assess the overall estimate of a causal effect was the inverse-variance weighted (IVW) method. When all assumptions are fulfilled, this method provides efficient estimates of causal relationship (19). Three sensitivity analyses were used to test the robustness of the MR results depicted by the IVW method. The first method, MR-Egger, loosens the third assumption of the genetic instruments only and measures the pleiotropic effect across the variants included in the MR analysis to test if they are affecting the exposure and the outcome independently. In that case, MR-Egger provides an estimate of the causal effect accounting for the pleiotropic effects (20).The second method used was the weighted median method that provides consistent results when up to 50% of the included variants are invalid instruments violating the second and third assumptions (21). The third method used was the weighted mode method, that gives consistent estimates when the largest number of causal effect estimates come from valid instruments despite the majority of instruments being invalid (22). Mendelian randomization analyses were carried out in R version 3.6.1 using the TwoSampleMR package.

4 RESULTS

4.1 SNP-BASED HERITABILITY AND GENETIC CORRELATION

The estimated liability-scale SNP-based heritability was 6.91% for IBS and 8.61% for anxiety (table 2). The cross-disorder LD score regression revealed a strong genetic correlation between the two phenotypes (rg=0.713, se= 0.078, P= 3.6e-20).

Table 2: SNP-based heritability and genetic correlation of IBS and anxiety*. For each disorder, N, LD score regression heritability (h2) with standard error (SE), intercept with standard error (Intercept (SE)), and ratio obtained as (intercept-1)/(mean(chi^2)-1) with standard error (Ratio (SE)) are presented. Genetic covariance with standard error (SE), genetic correlation with standard error (SE) and the P-value (P) are also displayed.*

Trait	N	h2(%)	SE	Intercept	Ratio (SE)	Genetic covariance (SE)	genetic correlation (SE)	P
IBS	455,321	0.0691	0.0063	1.0013 (0.0068)	0.0094 (0.0485)	0.0165 (0.0019)	0.7129 (0.0775)	3.66E-20
Anxiety	117751	0.0861	0.0106	0.9954 (0.0073)	< 0			

4.2 JOINT ANALYSIS OF IBS AND ANXIETY GWAS SUMMARY STATISTICS

After applying MTAG SNP filters, 6,594,820 SNPs were included in the multi-trait analysis of GWAS summary statistics of both disorders. This analysis revealed four genome-wide significant independent loci in chromosomes 3, 6, 9, and 11 for IBS and three new genome-wide significant independent loci in chromosomes 3, 7, and 9 for anxiety. Such results represent an improvement from the two genome-wide significant loci identified in each of the original GWAS for IBS (reported by Wu et al. (2)) and anxiety (the UK biobank original GWAS) as can be observed in Figure 1. Figure A1 in the Appendix displays the Quantile-Quantile plots of the expected and observed P-values for each disorder. The calculated slope (lambda λ) for IBS was 1.13 and 1.11 for anxiety.

Three of the four lead SNPs identified by MTAG in IBS were new, and the one in chromosome 11, rs10891490, overlapped with a previously identified SNP in the same chromosome that was originally reported by Wu et al. (2). These two SNPs were in LD with an r^2 of 0.992. In addition, two of them were statistically significant in both disorders: rs1872552 in chromosome 3 and rs10156602 in chromosome 9. No overlap was observed between the MTAG-identified hits for IBS and the original GWAS hits for anxiety. The known and novel SNPs are displayed in Table 3.

The gain in power attained by applying MTAG on the GWAS summary statistics of IBS and anxiety can be illustrated by the GWAS-equivalent sample size reported by MTAG which is 602,390 for IBS and 202,243 for anxiety, indicating a gain in power with MTAG equivalent to an increase in sample size relative to the original GWAS by 71.8% and 32.3%, respectively.

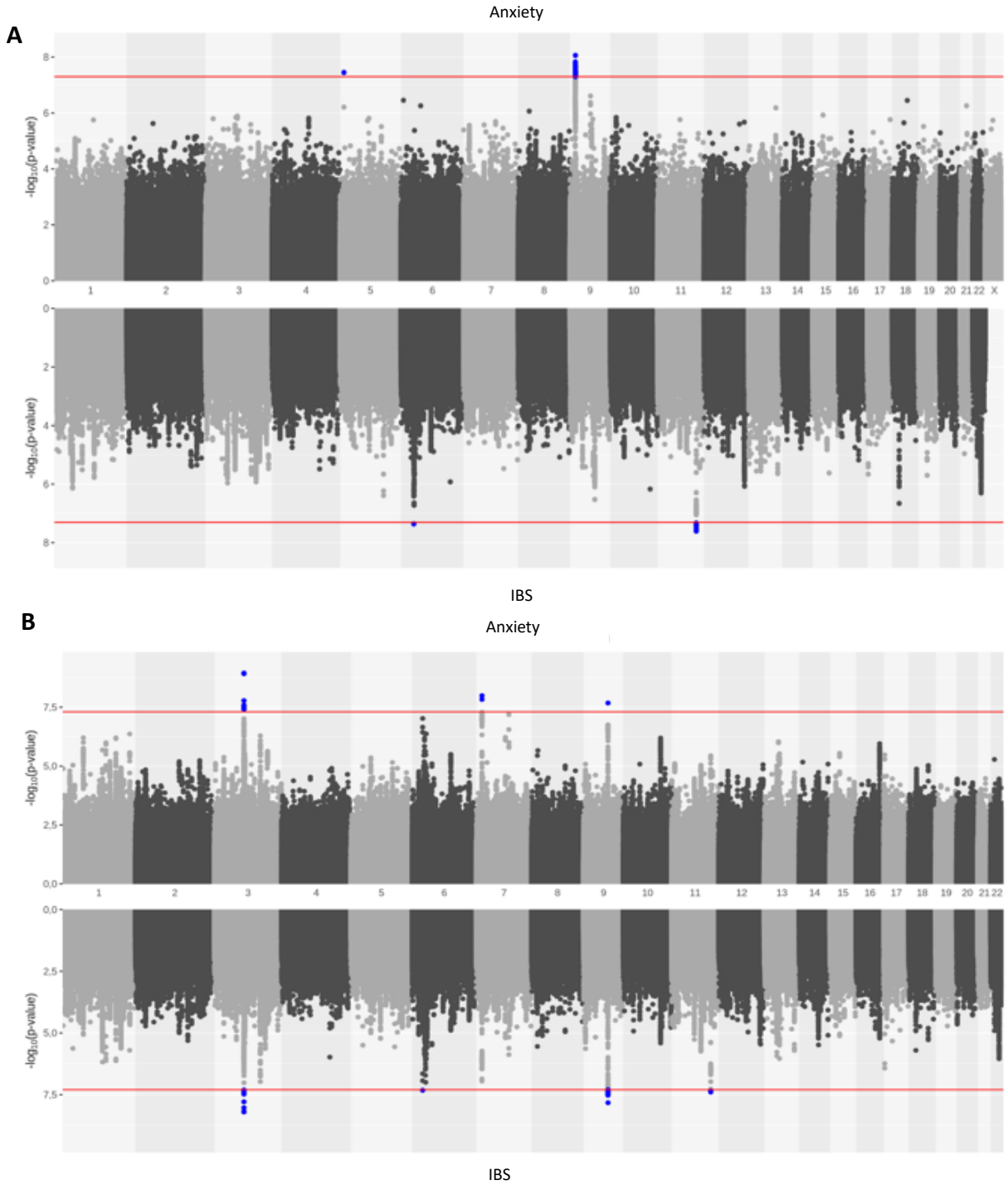

Figure 1: Miami plots of anxiety and IBS. *(A) represents the original GWAS of Anxiety (above) and IBS (bottom). (B) represents the Miami plots of both disorders after joint analysis using MTAG. Loci above the red line (P<5e-8) are shown in blue.*

Table 3: The main results of the loci identified for IBS by MTAG*. chr,=chromosome; Pos,=genomic position based on the GRCH37 genome assembly; the first/second alleles =the effect and alternative alleles. FRQ allele frequency of effect allele; Beta =regression coefficient; SE =standard error of beta; Genes =genes identified through MAGMA annotation and UCSC Variant Annotation Integrator.*

						MTAG			GWAS				
	SNP	**chr**	**pos**	**Alleles**	**FRQ**	**beta**	**SE**	**P**	**beta**	**SE**	**P**	**note**	**genes**
known	rs10891490	11	112885527	T/C	0.406	0.0102	0.00186	4.00E-08	0.0487	0.0087	2.60E-08		*NCAM1*
novel	rs1872552	3	85049088	G/A	0.601	0.0108	0.00186	6.23E-09	0.0366	0.0088	3.50E-05	significant SNP in MTAG for anxiety	*CADM2*
	rs117552817	6	26609105	C/T	0.986	0.0418	0.00765	4.74E-08	0.1514	0.0339	8.10E-06		*intergenic*
	rs10156602	9	96345328	A/G	0.64	0.0108	0.00190	1.48E-08	0.0414	0.0091	6.00E-06	significant SNP in MTAG for anxiety	*PHF2, FAM120A, FAM120AOS*

4.3 Bayesian Credible Set Analysis

Our results indicated that all lead SNPs at each locus showed the maximum posterior probability within their credible set. For IBS, the identified locus in chromosome 6 harbored a credible set of 3 SNPs, and the posterior probability of the lead SNP (rs117552817) within this locus was 60.8%. While MAGMA annotation did not reveal any protein coding genes containing this locus, the Variant Annotation Integrator feature within UCSC Genome Browser identified this SNP as being intergenic. The rest of the MTAG – identified loci for IBS contained much larger credible sets with lower posterior probabilities of their respective lead SNPs with the credible set in chromosome 3 containing 66 SNP and a posterior probability of the lead SNP rs1872552 being 22%, the credible set in chromosome 9 with 84 SNPs and a posterior probability of the lead SNP rs10156602 being 3.6%, and the credible set in chromosome 11 containing 75 SNPs with a posterior probability of the lead SNP rs10891490 being 7.5%. The locus identified in chromosome 3 was found within the gene *CADM2* and the locus in chromosome 9 was found within the *PHF2* and in close proximity to the *FAM120A* and *FAM120AOS* genes. Additionally, the UCSC Variant Annotation Integrator identified the lead SNP in chromosome 9 as falling within a regulatory region of the *PHF2* gene. As reported in Wu et al. (2), the locus identified in chromosome 11 locus was within the *NCAM1* gene.

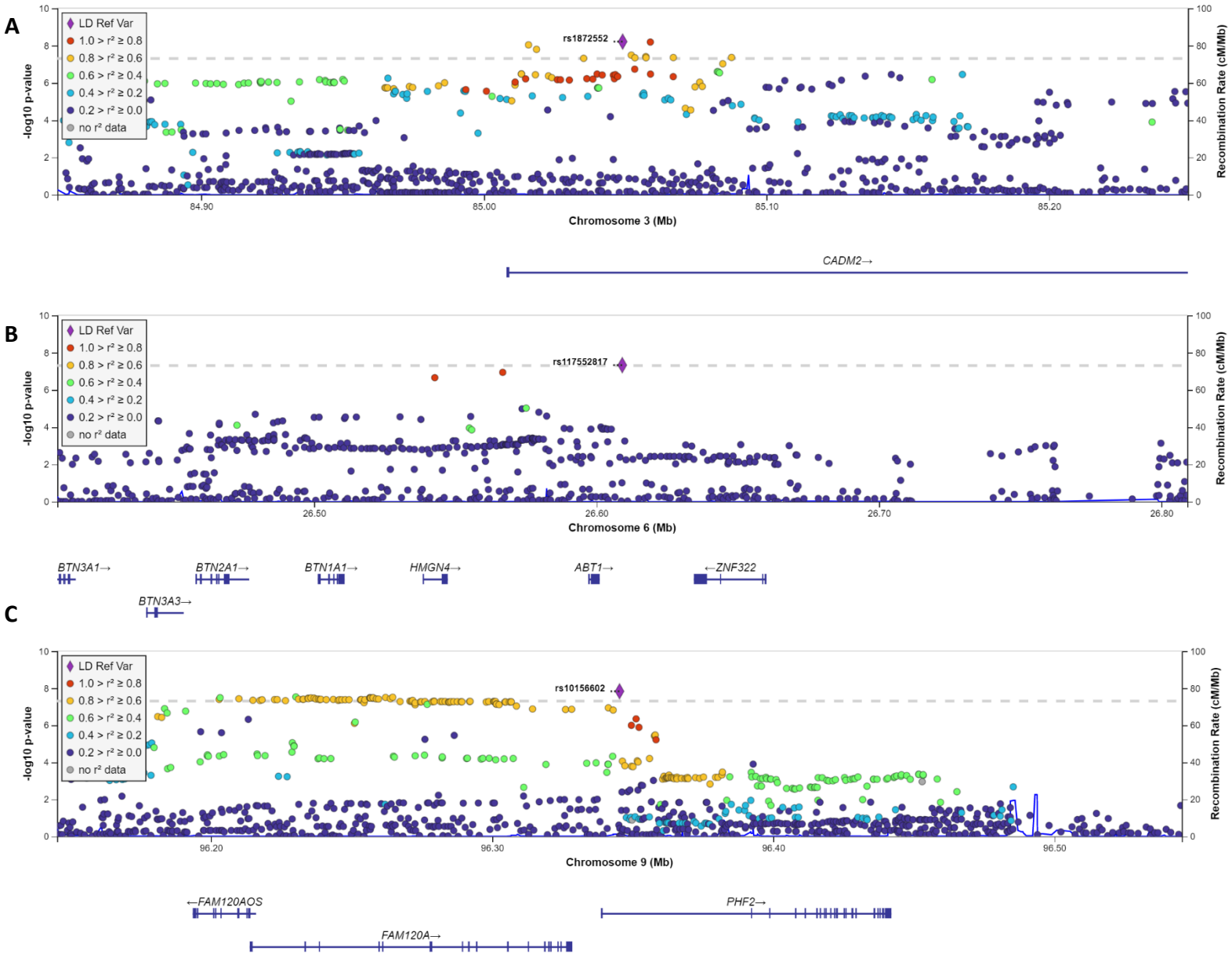

***Figure 2**: Regional plots of novel SNPs identified in IBS by MTAG. (A), (B), and (C) show the regional plots in the novel loci in chromosomes 3, 6, and 9, respectively. The dashed horizontal line represents the p-value threshold (5e-8) and the LD score (r^2) is based on the European population. Genes overlapping with hits are found beneath each plot (based on GRCH37 genome assembly) Plots were generated using my.locuszoom.org*

4.4 GENOME-WIDE GENE-BASED AND GENE-SET ANALYSIS

An array of genes appeared to be genome-wide significant after gene-based analysis of MTAG-generated summary statistics performed in MAGMA (v1.06). For IBS, MTAG revealed 258 genes displaying genome-wide significance after FDR correction (adjusted P<0.05) (Table 4) which represents an improvement from the 179 genes that were identified after applying the same annotation and gene-based analysis parameters on the original GWAS for IBS.

Table 4: Top 20 genome-wide significant SNPs in IBS after MTAG with anxiety. *MAGMA annotation was performed to elucidate the gene symbols. CHR=chromosome, Start & stop= the annotation boundaries of the gene including window around gene applied in annotation; NSNPS= the number of SNPs annotated to that gene; P= unadjusted P-value of gene; FDR-P=adjusted P-value after false discovery rate adjustment.*

Gene Symbol	CHR	START	STOP	NSNPS	P	FDR-P
CADM2	3	84998133	86128579	2376	2.43E-11	4.43E-07
FAM120A	9	96204173	96333397	304	4.65E-10	4.24E-06
NCAM1	11	112821969	113154158	778	1.49E-09	7.62E-06
SORCS3	10	106390859	107029993	1551	1.96E-09	7.62E-06
SYT14	1	210101519	210342636	326	2.09E-09	7.62E-06
PITPNM2	12	123463027	123645376	173	4.15E-09	1.26E-05
HLA-B	6	31316649	31334989	229	6.82E-09	1.78E-05
TMEM106B	7	12240848	12281890	188	9.11E-09	2.08E-05
S100A1	1	153590873	153609513	38	3.94E-08	7.23E-05
S100A13	1	153586275	153616582	60	3.97E-08	7.23E-05
S100A16	1	153574362	153595514	37	4.48E-08	7.39E-05
RAB13	1	153949093	153968853	21	4.86E-08	7.39E-05
IL20RB	3	136666707	136734927	45	6.35E-08	8.90E-05
ARL6IP4	12	123454607	123472460	13	1.14E-07	1.49E-04
S100A14	1	153581731	153598808	30	1.48E-07	1.80E-04
NUP210L	1	153960166	154137592	216	2.03E-07	2.31E-04
NCOR2	12	124803957	125062079	641	2.33E-07	2.49E-04
CREB3L4	1	153930315	153951840	26	2.65E-07	2.69E-04
TNXB	6	32003932	32087151	106	2.81E-07	2.70E-04
SLC39A1	1	153926575	153950660	27	4.18E-07	3.74E-04

MAGMA gene-set analysis was then applied to the identified group of genes using the C2 curated gene sets – canonical pathways from MSigDB. One gene set, the "*Glypican 1"* pathway, passed the FDR multiple testing correction for MTAG results of IBS (adjusted P=0.044). Gene set analyses with Gene Ontology terms returned no statistically significant results. Additionally, gene set analyses using the same parameters performed on the original GWAS results of IBS did not return any FDR-significant gene sets.

4.5 MENDELIAN RANDOMIZATION

The main analyses with the strictest threshold used for association with IBS ($P < 5 \times 10^{-7}$, 9 variants) showed evidence of causal effect of the genetic liability of IBS on anxiety with the inverse weighted variance (IVW) method (beta = 0.042, se = 0.011, P = 7.58E-05). Sensitivity analyses were overall consistent with this result. Specifically, analyses with the weighted median analysis method were also significant (beta = 0.042, se = 0.014, P = 0.002), although the association was weaker when using the weighted mode method (beta = 0.045, se = 0.021, P = 0.071)(Figure 3A). Leave-one-out analyses provided evidence that such findings were not driven by a single genetic instrument (Figure 3B). There was no indication of pleiotropy (MR pleiotropy test P = 0.513) or heterogeneity in the IWV method (P=0.933). When loosening the P-value threshold to $P < 5 \times 10^{-6}$, a total of 32 genetic instruments were included. MR analyses based on these 32 variants as instruments yielded weaker yet statistically significant results that continued to support the causal effect of IBS on anxiety (IVW: beta=0.019, se= 0.019, P= 0.003; Weighted median: beta=0.019, se=0.008, P=0.022) (Figure 3C). Although results with the weighted mode method were no longer significant (beta=0.017, se=0.018, P= 0.347)(Figure 3C), leave-one-out analysis showed concordance with the previous analysis with neither of the variants expressing outlying contribution to the MR result (Figure 3D). Finally, the pleiotropy test supported the absence of pleiotropic effects (P=0.053) and the heterogeneity test on IWV confirmed the absence of heterogeneity (P = 0.424).

MR analysis in the opposite direction did not identify a causal effect of the genetic liability of anxiety on IBS. With a P-value threshold of $P < 5 \times 10^{-7}$, only two SNPs were considered valid instruments for MR analysis which was not enough to conduct any robust analysis. Lowering the P-value threshold to $P < 5 \times 10^{-6}$ resulted in 17 SNPs being valid genetic instruments but returned a non-significant IWV result (P=0.322). There was no evidence of pleiotropy with the pleiotropy test showing a P-value of 0.92, but heterogeneity was observed in the IWV result (P=0.017).

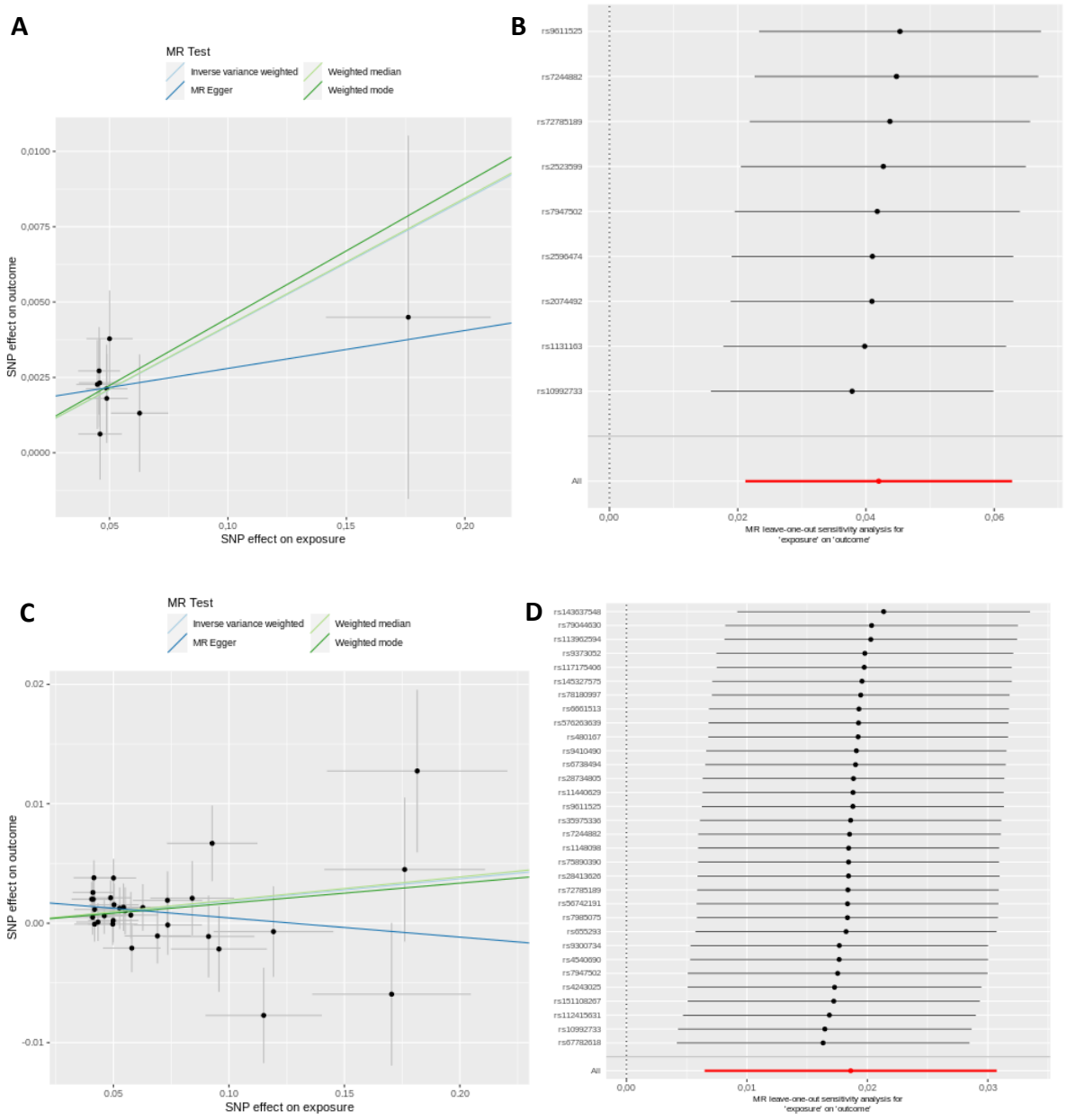

Figure 3: Mendelian randomization study of the causal effect estimate of IBS on anxiety. *(Top) Plots for a P-value threshold of 5 × 10-7 and (bottom) plots for a P-value threshold of 5x10-6. (A & C) Scatter plot of SNP effect estimates on IBS vs. effect estimates on anxiety use with error bars. Lines are drawn for each Mendelian randomization method used, with the slope of each line corresponding to the estimated causal effect. (B & D) Leave-one-out plot. Causal effect sizes of the genetic liability of IBS on anxiety with 95% confidence interval for the full IVW results (in red) and the IVW results excluding one SNP at a time*.

5 DISCUSSION

In this study, we investigated the genetic basis of the co-occurrence of IBS and anxiety by using publicly available SNP-level summary statistics of two large genome-wide association studies representing IBS and anxiety from the UK Biobank. We found evidence of a strong and positive genetic correlation between IBS

and anxiety, identified three new loci associated with IBS and provided evidence for a causal relationship of the genetic liability of IBS on anxiety.

In line with previous reports of phenotypic overlap or comorbidity between IBS and psychiatric disorders such as anxiety (9), the cross-trait linkage disequilibrium regression analysis conducted in this study revealed statistically significant genetic correlation between both disorders (rg=0.713, P= 3.6e-20). These results suggest that there is a common genetic background shared by these phenotypes.

The strong genetic correlation that was identified between IBS and anxiety motivated the exploitation of the available SNP-level GWAS summary statistics to increase the understanding of the genetic underpinnings of IBS and its relationship with anxiety. Multi-trait analysis of GWAS (MTAG) was used as an integrative tool to perform joint association analysis of the summary statistics of both disorders. The ability of this method to increase the statistical power of GWAS allowed the identification of three independent loci for IBS in chromosomes 3, 6, and 9 not previously identified in the original GWAS (2). With one locus in chromosome 11 identified in the original GWAS for this disorder and replicated in our analysis and another independent locus in chromosome 6 identified by the original GWAS only, the number of independent loci associated with this disorder increased from two to five.

The top hit identified in our analysis, rs1872552 in chromosome 3, is an intronic variant in the *CADM2* gene. This gene encodes the synaptic cell adhesion molecule 2, a protein predominantly expressed in the brain. Interestingly, this gene has been associated with psychological and psychiatric phenotypes of relevance for IBS such as neuroticism, mood instability, and risk-taking behavior (23). The second most significant hit in this analysis, rs10156602, is an intronic variant and lies within a transcription factor binding site of the gene *PHF2* which encodes the plant homeodomain (PHD) finger 2 protein that is involved in histone modification, chromatin rearrangement, and epigenetic activation (24). This variant has also shown close proximity with the *FAM120A* gene that encodes the protein involved in the signal transduction pathway of interlukin-13 (IL13) for the development of colon cancer (25). The identification such genes after joint-association analysis of IBS with anxiety provides evidence of the genetic overlap between IBS and anxiety and gives insight into the epigenetic factors that can bridge the knowledge on the genetic and environmental contributors in IBS and its relationship to anxiety.

Genome-wide gene level analysis after MTAG on IBS revealed 258 FDR-significant genes. Among the most significant genes, and in addition to the genes discussed above, we found the *NCAM1 gene* that was also identified by Wu et al. and encodes the protein involved in the development of the nervous system (2). Other genes include *SORCS3* that encodes sortilin-related receptor CNS expressed 3, a sorting receptor

involved in neuronal signaling (26), *SYT14* encoding synaptotagmin XIV and *TMEM106B* encoding transmembrane protein 106B, proteins involved in neurodegenerative disorders (27,28), and the *PITPNM2* gene that encodes a Phosphatidylinositol transfer protein membrane associated protein (29). To our best knowledge, most of these genes have not been reported to be associated with IBS in the literature before, thus replication of this data and validation of these genes is required to move forward in understanding the genes that may play a role in the pathophysiology of IBS.

Gene set analyses identified the *Glypican 1 pathway* to be statistically significant in IBS after FDR correction (P=0.0044). *Glypican 1* belongs to the glypican subfamily of heparan sulfate proteoglycans, is anchored to the cell membrane via a glycosylphosphatidylinositol (GPI) anchor and is involved in the regulation of cellular signaling with emerging evidence of involvement in neural connectivity and synaptic signaling (30). Such results further point at the genetic overlap between IBS and anxiety and the possible common genetic pathways that govern their co-occurrence.

Mendelian randomization analyses performed on the original GWAS summary statistics of IBS and anxiety found evidence for a causal effect of IBS on anxiety but not vice-versa. With two P-value thresholds for the IBS genetic instruments, a causal relationship was sustained for the genetic liability of IBS on anxiety. Statistically significant results for the IWV method, weighted median sensitivity method, and the MR-Egger pleiotropy test were recorded. Although the weighted mode method showed no significant results at either threshold, the direction and magnitude of its effect size remained consistent with the other methods which is characteristic of the robustness of our results (22). There was no evidence of a causal relationship in the opposite direction. These findings are in partial alignment with the brain-gut axis hypothesis that describes a bi-directional communication between the brain and the gut. Larger datasets and other MR analyses methods are warranted to strengthen our evidence on the causal relationship between these disorders.

The present study should be considered in the context of some limitations. While the UK Biobank provides access to well-sized genetic datasets, it does not ensure well-characterized phenotypes that are often ambiguous or self-reported. In our study, the anxiety phenotype "*Mental health problems ever diagnosed by a professional: Anxiety, nerves or generalized anxiety disorder*" was defined in a broad manner which may result in misclassification of anxiety cases. Moreover, IBS is considered a highly heterogenous disorder with pathophysiological differences observed among its subgroups, between genders, and across age groups and geographic locations (31). Accounting for such factors may contribute to better characterization of the disorder but may limit the sample size of future GWAS on the disorder. In addition,

we could not control for comorbid conditions in our analysis meaning that the presence of anxiety cases within our IBS dataset and vice-versa may bias our results. Another limitation imposed by the use of UK Biobank datasets is the expected sample overlap between both traits. While this does not affect the joint association analysis that accounts for sample overlap, it could cause bias in MR results (32). Finally, the sample size for the anxiety phenotype (117,751) was much smaller than that for IBS (455,321) which could explain the low-powered MR analysis for the causal relationship of the genetic liability of anxiety on IBS.

Our study provides solid evidence of positive genetic correlation between IBS and anxiety along with a potential causal relationship of the genetic liability of IBS on anxiety. Joint association analysis of both disorders revealed new genetic variants and genes that may be involved in the pathophysiology of both disorders. Future goals in this research could aim to investigate the identified SNPs within eQTL databases to understand their role in regulating gene expression and the validation of the identified genes in targeted hypothesis-driven analyses with well-characterized samples. Furthermore, replication of our findings in larger datasets is warranted to increase the credibility of our results.

6 REFERENCES

1. Lacy BE, Mearin F, Chang L, Chey WD, Lembo AJ, Simren M, et al. Bowel Disorders. Gastroenterology [Internet]. 2016 May [cited 2021 Aug 3];150(6):1393-1407.e5. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0016508516002225

2. Wu Y, Murray GK, Byrne EM, Sidorenko J, Visscher PM, Wray NR. GWAS of peptic ulcer disease implicates Helicobacter pylori infection, other gastrointestinal disorders and depression. Nat Commun [Internet]. 2021 Dec [cited 2021 Aug 3];12(1):1146. Available from: http://www.nature.com/articles/s41467-021-21280-7

3. Buono JL, Mathur K, Averitt AJ, Andrae DA. Economic Burden of Irritable Bowel Syndrome with Diarrhea: Retrospective Analysis of a U.S. Commercially Insured Population. JMCP [Internet]. 2017 Apr [cited 2021 Aug 4];23(4):453–60. Available from: http://www.jmcp.org/doi/10.18553/jmcp.2016.16138

4. Tack J, Stanghellini V, Mearin F, Yiannakou Y, Layer P, Coffin B, et al. Economic burden of moderate to severe irritable bowel syndrome with constipation in six European countries. BMC Gastroenterol [Internet]. 2019 Dec [cited 2021 Aug 4];19(1):69. Available from: https://bmcgastroenterol.biomedcentral.com/articles/10.1186/s12876-019-0985-1

5. Saito YA. The Role of Genetics in IBS. Gastroenterology Clinics of North America [Internet]. 2011 Mar [cited 2021 Aug 6];40(1):45–67. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0889855310001354

6. Waehrens R, Ohlsson H, Sundquist J, Sundquist K, Zöller B. Risk of irritable bowel syndrome in first-degree, second-degree and thirddegree relatives of affected individuals: a nationwide family study in Sweden. Gut [Internet]. 2015 Feb [cited 2021 Aug 6];64(2):215–21. Available from: https://gut.bmj.com/lookup/doi/10.1136/gutjnl-2013-305705

7. Henström M, D'Amato M. Genetics of irritable bowel syndrome. Mol Cell Pediatr [Internet]. 2016 Dec [cited 2021 Sep 5];3(1):7. Available from: http://www.molcellped.com/content/3/1/7

8. Bonfiglio F, Henström M, Nag A, Hadizadeh F, Zheng T, Cenit MC, et al. A GWAS meta-analysis from 5 population-based cohorts implicates ion channel genes in the pathogenesis of irritable bowel syndrome. Neurogastroenterology & Motility. 2018 Apr 19;30(9):11.

9. Zamani M, Alizadeh-Tabari S, Zamani V. Systematic review with meta-analysis: the prevalence of anxiety and depression in patients with irritable bowel syndrome. Aliment Pharmacol Ther [Internet]. 2019 Jul [cited 2021 Aug 3];50(2):132–43. Available from: https://onlinelibrary.wiley.com/doi/10.1111/apt.15325

10. Tanaka Y, Kanazawa M, Fukudo S, Drossman DA. Biopsychosocial Model of Irritable Bowel Syndrome. J Neurogastroenterol Motil [Internet]. 2011 Apr 30 [cited 2021 Aug 4];17(2):131–9. Available from: http://www.jnmjournal.org/journal/view.html?doi=10.5056/jnm.2011.17.2.131

11. Hetterich L, Stengel A. Psychotherapeutic Interventions in Irritable Bowel Syndrome. Front Psychiatry [Internet]. 2020 Apr 30 [cited 2021 Aug 5];11:286. Available from: https://www.frontiersin.org/article/10.3389/fpsyt.2020.00286/full

12. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature [Internet]. 2018 Oct [cited 2021 Aug 9];562(7726):203–9. Available from: http://www.nature.com/articles/s41586-018-0579-z

13. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet [Internet]. 2015 Nov [cited 2021 Aug 10];47(11):1236–41. Available from: http://www.nature.com/articles/ng.3406

14. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet [Internet]. 2015 Mar [cited 2021 Aug 10];47(3):291–5. Available from: http://www.nature.com/articles/ng.3211

15. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet [Internet]. 2018 Feb [cited 2021 Aug 13];50(2):229–37. Available from: http://www.nature.com/articles/s41588-017-0009-4

16. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci [Internet]. 2015 Dec [cited 2021 Aug 19];4(1):7. Available from: https://academic.oup.com/gigascience/article-lookup/doi/10.1186/s13742-015-0047-8

17. Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, Su Z, et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet [Internet]. 2012 Dec [cited 2021 Aug 17];44(12):1294–301. Available from: http://www.nature.com/articles/ng.2435

18. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. Tang H, editor. PLoS Comput Biol [Internet]. 2015 Apr 17 [cited 2021 Aug 18];11(4):e1004219. Available from: https://dx.plos.org/10.1371/journal.pcbi.1004219

19. Treur JL, Demontis D, Smith GD, Sallis H, Richardson TG, Wiers RW, et al. Investigating causality between liability to ADHD and substance use, and liability to substance use and ADHD risk, using Mendelian randomization. Addiction Biology [Internet]. 2021 Jan [cited 2021 Aug 19];26(1). Available from: https://onlinelibrary.wiley.com/doi/10.1111/adb.12849

20. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology [Internet]. 2015 Apr 1 [cited 2021 Aug 19];44(2):512–25. Available from: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyv080

21. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol [Internet]. 2016 May [cited 2021 Aug 19];40(4):304–14. Available from: https://onlinelibrary.wiley.com/doi/10.1002/gepi.21965

22. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International Journal of Epidemiology [Internet]. 2017 Dec 1 [cited 2021 Aug 19];46(6):1985–98. Available from: https://academic.oup.com/ije/article/46/6/1985/3957932

23. Morris J, Bailey MES, Baldassarre D, Cullen B, de Faire U, Ferguson A, et al. Genetic variation in CADM2 as a link between psychological traits and obesity. Sci Rep [Internet]. 2019 Dec [cited 2021 Sep 7];9(1):7339. Available from: http://www.nature.com/articles/s41598-019-43861-9

24. Kim H, Hur SW, Park JB, Seo J, Shin JJ, Kim S, et al. Histone demethylase PHF2 activates CREB and promotes memory consolidation. EMBO Rep [Internet]. 2019 Sep [cited 2021 Sep 7];20(9). Available from: https://onlinelibrary.wiley.com/doi/10.15252/embr.201845907

25. Bartolomé RA, García-Palmero I, Torres S, López-Lucendo M, Balyasnikova IV, Casal JI. IL13 Receptor α2 Signaling Requires a Scaffold Protein, FAM120A, to Activate the FAK and PI3K Pathways in Colon Cancer Metastasis. Cancer Res [Internet]. 2015 Jun 15 [cited 2021 Sep 7];75(12):2434–44. Available from: http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-14-3650

26. Christiansen GB, Andersen KH, Riis S, Nykjaer A, Bolcho U, Jensen MS, et al. The sorting receptor SorCS3 is a stronger regulator of glutamate receptor functions compared to GABAergic mechanisms in the hippocampus: SORCS3 DEFICIENCY RESULTS IN PRONOUNCED GLUTAMATERGIC CHANGES. Hippocampus [Internet]. 2017 Mar [cited 2021 Sep 8];27(3):235–48. Available from: https://onlinelibrary.wiley.com/doi/10.1002/hipo.22689

27. Doi H, Yoshida K, Yasuda T, Fukuda M, Fukuda Y, Morita H, et al. Exome Sequencing Reveals a Homozygous SYT14 Mutation in Adult-Onset, Autosomal-Recessive Spinocerebellar Ataxia with Psychomotor Retardation. The American Journal of Human Genetics [Internet]. 2011 Aug [cited 2021 Sep 8];89(2):320–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0002929711003065

28. Ren Y, van Blitterswijk M, Allen M, Carrasquillo MM, Reddy JS, Wang X, et al. TMEM106B haplotypes have distinct gene expression patterns in aged brain. Mol Neurodegeneration [Internet]. 2018 Dec [cited 2021 Sep 8];13(1):35. Available from: https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0268-2

29. Ocaka L, Spalluto C, Wilson DI, Hunt DM, Halford S. Chromosomal localization, genomic organization and evolution of the genes encoding human phosphatidylinositol transfer protein membrane-associated (PITPNM) 1, 2 and 3. Cytogenet Genome Res [Internet]. 2005 [cited 2021 Sep 8];108(4):293–302. Available from: https://www.karger.com/Article/FullText/81519

30. Condomitti G, de Wit J. Heparan Sulfate Proteoglycans as Emerging Players in Synaptic Specificity. Front Mol Neurosci [Internet]. 2018 Jan 26 [cited 2021 Sep 8];11:14. Available from: http://journal.frontiersin.org/article/10.3389/fnmol.2018.00014/full

31. Card T, Canavan C, West J. The epidemiology of irritable bowel syndrome. CLEP [Internet]. 2014 Feb [cited 2021 Sep 9];71. Available from: http://www.dovepress.com/the-epidemiology-of-irritable-bowel-syndrome-peer-reviewed-article-CLEP

32. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genetic Epidemiology. 2016 Sep 14;40(7):12.