



OPEN

Systematic identification of the role of gut microbiota in mental disorders: a TwinsUK cohort study

Julie Delanote¹, Alejandro Correa Rojo^{1,2}, Philippa M. Wells³, Claire J. Steves^{3,4} & Gökhan Ertaylan¹✉

Mental disorders are complex disorders influenced by multiple genetic, environmental, and biological factors. Specific microbiota imbalances seem to affect mental health status. However, the mechanisms by which microbiota disturbances impact the presence of depression, stress, anxiety, and eating disorders remain poorly understood. Currently, there are no robust biomarkers identified. We proposed a novel pyramid-layer design to accurately identify microbial/metabolomic signatures underlying mental disorders in the TwinsUK registry. Monozygotic and dizygotic twins discordant for mental disorders were screened, in a pairwise manner, for differentially abundant bacterial genera and circulating metabolites. In addition, multivariate analyses were performed, accounting for individual-level confounders. Our pyramid-layer study design allowed us to overcome the limitations of cross-sectional study designs with significant confounder effects and resulted in an association of the abundance of genus *Parabacteroides* with the diagnosis of mental disorders. Future research should explore the potential role of *Parabacteroides* as a mediator of mental health status. Our results indicate the potential role of the microbiome as a modifier in mental disorders that might contribute to the development of novel methodologies to assess personal risk and intervention strategies.

Neuropsychiatric disorders are complex conditions influenced by multiple factors, including genetic predisposition and environmental triggers^{1–5}. Several studies have shown that patients diagnosed with a mental disorder (MD), including depression, anxiety, and eating disorders, demonstrate a dysbiosis of intestinal microbiota^{6–9}. Specific microbiota imbalances affect the brain by humoral and neuronal mechanisms, with particular attention to the vagus nerve¹⁰. Gut microbiota can also have an impact through the endocrine pathway or by modification of the blood–brain barrier (BBB)¹¹. Microbial metabolites mediators in this gut–brain signaling pathway, such as short-chain fatty acids (SCFAs), secondary bile acids, and amino acids, influence behavior, memory, learning, and locomotion¹². In addition, clinical and preclinical evidence further support the existing gut–brain axis in neuropsychiatric disorders where fecal microbiota transplantation (FMT) from healthy donors has shown to be valuable in relieving depressive and anxious behaviors¹³, as well as the use of probiotics¹⁴.

Our microbiomes are inherently dynamic ecosystems that vary per individual. This complicates the identification of gut microbiota members that casually contribute to human disease. To truly identify microbial alteration associated with MDs, one must account for confounding factors, such as body mass index (BMI), dietary patterns, and antibiotic usage¹⁵. The maternal microbiota shapes the gut microbial composition, where perinatal factors are responsible for forming the adult-like microbiota¹⁶. However, there is still a wide interindividual heterogeneity in microbiota composition due to dietary traits, medication, lifestyle, and physiological variables¹⁵. Co-twin control study designs are effective in canceling off the majority of unmeasured differences. They offer a unique model for controlling confounders to interpersonal variation as they share genetic background, early-life events, and life course factors (e.g., birth delivery mode, feeding, dietary pattern, geography, etc.)¹⁷. Vilchez-Vargas et al.¹⁸ identified shared household and aging, *but not host genetics*, as factors in determining microbial similarity in

¹Sustainable Health, Flemish Institute for Technological Research (VITO), Mol, Belgium. ²Data Science Institute, Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Diepenbeek, Belgium. ³Department of Twin Research and Genetic Epidemiology, King's College London, St Thomas' Hospital, 3–4th Floor South Wing Block D, Westminster Bridge Road, London SE1 7EH, UK. ⁴Department of Ageing and Health, St Thomas' Hospital, 9th floor, North Wing, Westminster Bridge Road, London SE1 7EH, UK. ✉email: gokhan.ertaylan@vito.be

fecal specimens of twins irrespective of zygosity. These findings emphasize the predominant role of environmental factors rather than host factors in defining the twins' microbial composition¹⁸. In contrast, microbiome differences between participants, disease versus healthy, may be overpowered by interindividual variability in the microbiota in contemporary cross-sectional designs¹⁵.

Monozygotic (MZ) twins share all germline variants, whereas dizygotic (DZ) twins share, on average, half of their genetic material similar to non-twin siblings¹⁹. This study encompasses the analysis of depression, anxiety, and eating disorders as a collective group of MDs, which often co-occur and share common symptoms. The rationale behind this approach is to explore potential common impacts of the gut microbiota on these underlying disorders. Perturbations in the gut microbiome can influence various biological processes, including HPA axis dysregulation, immune system responses, mechanisms of inflammation, gut barrier permeability^{20,21}, and alterations in the production of short-chain fatty acids and neurotransmitters^{22,23}. These factors are integral to the proper functioning of neurological processes. By examining these disorders together, it is possible to uncover common patterns that contribute to the development and manifestation of these conditions. MDs were previously reported to have a considerable heritable component. Broad-sense heritability estimates for depression were reported in the range of 35–45%²⁴; anxiety disorders between 35 and 50%²⁵. For eating disorders, estimates are between 39 and 45% in binge eating disorder (BED), around 60% in bulimia nervosa (BN), and range from 28 to 74% in anorexia nervosa (AN)²⁶. Assuming no somatic mutations occur, the comparison between MZ and DZ twins can be used to obtain insight into the degree of genetic involvement of a particular trait or disease²⁷. If higher concordance rates are seen amongst MZ twins; compared to DZ twins, genetic factors might partially underlie the disease²⁷. MZ twins were found to have higher concordance rates than DZ twins in the diagnosis of AN^{28–30}, BN^{31,32}, anxiety and stress disorder^{33,34}, and depression^{35,36}. In light of this study, we posited that a desirable control is represented by the healthy co-twin of a discordant twin pair given the genetic relatedness to the MD-diagnosed twin and similar environmental exposures such as shared household as well as shared exposure from a maternal site like maternal breastmilk. By linking within-twin pair differences in mental health status to within-twin pair microbiome/metabolome differences, the discordant twin method has the unique potential to identify probable modifiable pathways between the twin's mental health and microbiome/metabolome composition^{37,38}.

Microbial alterations must be considered in light of their functional potential. It has been demonstrated that core housekeeping pathways remain stable across human microbiomes, such as ribosome and translational machinery, ATP synthesis, and glycolysis³⁹. However, the unique microbial gene content translates to a huge diversity in how microbes produce or modify metabolites⁴⁰. Seven times more associations were found between metabolites and microbial metabolic pathways than species, suggesting that functions rather than taxonomy reflect better the interplay between the microbiome and systemic environments⁴¹. Therefore, to benefit the host, modulations of the gut microbiome must target functionally related microbial communities rather than single microbes.

The gut microbiome plays an important role in mental health. A small number of research studies have identified potential microbiome-brain interactions implicated in several psychiatric conditions^{42–44}. However, to our knowledge, there has been no single study that addresses gut (microbiome)-brain (function) interaction in twins' settings. We hypothesized that the gut microbiome of healthy co-twins would differ from genetically similar twins with an MD diagnosis. Therefore, in this study, we characterized fecal microbiomes and plasma metabolomes using longitudinal data from the UK Adult Twin Registry (TwinsUK). Our aim is to discover taxa, plasma metabolites, and microbial pathways that may influence the onset/progression of MDs in adult twins, including potential differences in the gut microbiome between MZ and DZ twins. We propose a pyramid (layered) design approach for multiple comparison testing to identify differentially abundant features between sub-populations of twin types. The general flow of this study is represented in Fig. 1. The idea behind this structure is that we start from a specific population (MZ population) where twins share all of their genetic material, while the next population (DZ population) twins share, on average, 50% of their genetic material. Both populations are analyzed separately. We later combine the data of both populations to obtain more statistical confidence in possible associations between microbial genera/metabolites and MD status. To compensate for heterogeneity in the bottom layers of the pyramid, we account for potential individual confounding factors through larger sample sizes. Individuals from MD-discordant twin pairs who are not diagnosed with MD are referred to as “*healthy co-twins*” and individuals from MD-discordant or MD-concordant twin pairs who are diagnosed with MD are referred to as “*MD-twins*”.

The co-twin control study setup allows, *by design*, to control for genetic diversity and shared (childhood) environmental factors as important confounding effects. To this end, pairwise tests were performed in the population of MZ (Fig. 1, layer a), DZ (Fig. 1, layer b), and all discordant twins (Fig. 1, layer c). This setup allows us to do a direct comparison between twins who share the same genetic background and early environmental exposure and reduces the effect of potential confounders. Between MZ twins (Fig. 1, layer a), differences are more likely to be caused by other than shared genetic factors, while those identified in the DZ population (Fig. 1, layer b) could be due to both genetic and environmental effects. In layer c (Fig. 1, layer c), we combine MZ and DZ twins to increase sample size and potentially identify genera/metabolites associated with MD status. With this approach, the identification of microbial genera provides insight into the potential role of environmental factors in shaping the gut microbiome and its link with MD. These findings are then verified for non-shared environmental confounders, including BMI, dietary variables, and antibiotics usage, following multivariate analysis by using linear mixed models (Fig. 1, layers d and e). Regression analyses with linear mixed models allow us to examine the relationship between multiple variables while accounting for the correlation structure between the two twin types. By including non-shared environmental confounders as covariates, we can determine the independent associations of microbial genus/metabolite with health status. Specifically, layers (d) and (e) represent different aspects of the study design. Layer (d) focuses on studying all healthy co-twins, and are assumed to be at an

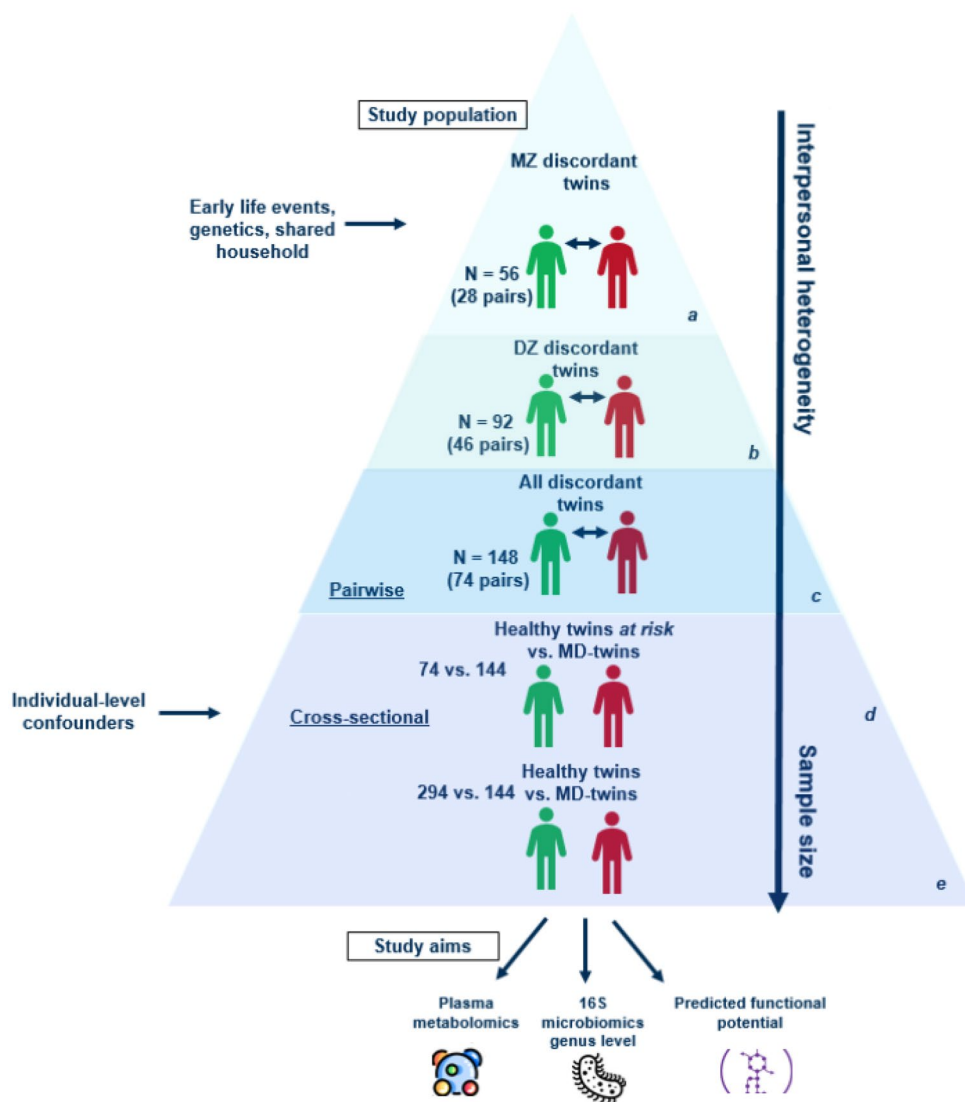


Figure 1. Visualization of the study design. The pyramid layers represent the different setups of this study. There are two main layers: (i) top layer: pairwise analyses in which we control for early life events, genetic background, and shared environment/household in which the twins grow up (ii) bottom layer: linear mixed regression analyses in which individual-level confounders were controlled for (antibiotics, diet differences, BMI, age). The top layer is divided into three parts (a) analysis of MZ twins, discordant in MD for a total of 28 twin pairs (b) analysis of DZ twins, discordant in MD for a total of 46 pairs, (c) analysis of twins, discordant for MD for a total of 74 pairs. The bottom layer studies (d) the healthy co-twins (who are assumed to be at increased risk of developing MD) with MD twins (a comparison of 74–144 twins) and (e) all healthy twins compared to all twins with MD diagnosis. The arrow indicates a higher interindividual heterogeneity and an increase in sample size the more you go toward the bottom of the pyramid. The study aims to find alterations in the microbiome and metabolome landscape in MD.

increased risk of developing MD. This layer specifically examines the co-twins of individuals who have been diagnosed with MD (discordant twin pairs) in relation to any twin who is diagnosed with MD (from both concordant and discordant pairs). Layer (e) involves a broader comparison between all healthy twins and all MD twins (both from discordant and concordant twin pairs). Thus, layer (e) allows for a more specific investigation of the unique characteristics associated with being a healthy co-twin and the potential factors that contribute to their increased risk of developing MD, whereas layer (e) encompasses the larger, broad population. This study is the first exploratory analysis to discover potential microbiome–host interactions in the role of MD from genetically identical or similar individuals hence aims to minimize the role of genetics in gut microbiome–neurological phenotype associations. We only analyzed individuals with matched and complete microbiome and metabolome datasets, comprising a total of 438 individuals (219 twin pairs). Due to the small sample size of our study, we report associations with a false discovery rate (FDR) cut-off below 0.2 and report our findings in the context of earlier studies.

Results

General description cohort

A total of 438 twins (219 twin pairs) were included in the study, amongst which 424 were female and 14 male twins. 97 twin pairs were MZ, 122 twin pairs were DZ, and all twin pairs were same-sex pairs. Descriptive cohort characteristics are found in Table 1.

In this study, there is a prevalence of affected twins in the MZ population of 37.1%, while 29.5% in the DZ population. We found proband-wise concordance rates (probability that healthy co-twin will also have a history of MD) being 61% among MZ twins and 36% in DZ twins. No significant relationship was found between zygosity and the presence of MD ($p = 0.114$), nor between zygosity and phenotype concordance ($p = 0.06562$). However, when considering all MD-diagnosed twins, a significant relationship was found between concordance of the phenotype and zygosity ($p = 0.004591$). A significant relationship was also found between MD diagnosis and zygosity when considering all concordant twins ($p = 0.00513$). These results suggest that genetic factors may partially explain the MD phenotype that runs in twins. Given these results, we hypothesize that the healthy co-twin is at an increased risk of developing an MD which raises questions on how the microbiome might act as a potential modifier of this risk. Why is the healthy co-twin still healthy? Can the microbiome composition along with its functional effects contribute to his/her health status?

Microbiome and metabolome composition are significantly more similar in MZ twins compared to DZ twins

A comparison of Bray–Curtis dissimilarity metrics revealed a statistically significant greater dissimilarity in DZ twin pairs compared to MZ twin pairs ($p = 0.028$, Student's t-test). Additionally, twins in the same twin pair demonstrated a significantly lower dissimilarity in their microbiomes, compared to twins from other pairs ($p = 3.4 \times 10^{-11}$, Student's t-test). Similar findings apply to the plasma metabolome composition, for which the Euclidean distance was used to compare metabolic profiles in MZ and DZ twin pairs. MZ twin pairs revealed a significantly lower Euclidean distance compared to DZ twin pairs ($p = 0.036$, Student's t-test), as well as individuals in the same twin pair showed a significantly lower Euclidean distance compared to twins from other pairs ($p < 2.2 \times 10^{-16}$, Student's t-test) (Fig. 2).

Pairwise analyses of twins who are discordant for mental disorders

A total of nine genera were found to be differentially abundant between MZ MD-twin and the healthy co-twin (MZ population, Fig. 1a), i.e., *Parabacteroides*, *Dorea*, *Ruminococcaceae UCG 014*, *Ruminococcaceae UCG 013*, *Family XIII AD3011 group*, *Victivallis*, *Pseudoflavonifactor*, unclassified *Ruminococcaceae*, and unclassified *Mollicutes RF39* ($p < 0.05$, paired samples Wilcoxon test) (Fig. 3a). Metabolites that were significantly differentially abundant in MD-twins compared to healthy co-twins in the MZ population ($p < 0.05$, paired samples Wilcoxon test) included mainly high-density lipoprotein subfractions that were consistently lower in the MD group compared to the healthy co-twins. We found, in order of p-value, extra-large high-density lipoprotein free cholesterol (XL-HDL-FC %), Valine (Val), high-density lipoprotein 2 cholesterol (HDL2-C), XL HDL phospholipids, XL HDL free cholesterol, HDL cholesterol (HDL-C), large HDL particles (L-HDL-P), L HDL lipids (L-HDL-L), L-HDL cholesterol (L-HDL-C), L-HDL esterified cholesterol (L-HDL-CE) (Fig. 3b). Considering predicted metabolic pathways (Kyoto Encyclopedia of Genes and Genomes, hierarchical level 3), we found 13 predicted metabolic pathways to be significantly differentially abundant between MD and healthy co-twins ($p < 0.1$) (Supplementary Fig. S1).

The same pairwise analysis was performed in the population of DZ twins discordant in MD (DZ population, Fig. 1b). Out of the 18 significant genera identified, two genera overlapped with the findings in the MZ twin pairs, i.e., *Ruminococcaceae UCG 013* and *Family XIII AD3011 group* (Supplementary Fig. S2). Moreover,

		TwinsUK cohort		
Age (years)	mean (sd)	64.89 (7.71)		
BMI (kg/m ²)	mean (sd)	26.33 (4.72)		
BMI ≥ 25	n (%)	155 (35%)		
BMI ≥ 30	n (%)	94 (21%)		
Gender (female)	n (%)	424 (96.8%)		
Gender (male)	n (%)	14 (3.2%)		
		All	MZ	DZ
Diagnosis of mental disorder(s)	n (%)	144	72	72
Healthy controls	n (%)	294	172	122
		All	MZ	DZ
Discordant pairs	# pairs	74	28	46
Concordant pairs (both MD)	# pairs	35	22	13
Concordant pairs (both healthy)	# pairs	110	47	63
		Total pairs: 219		

Table 1. Cohort characteristics.

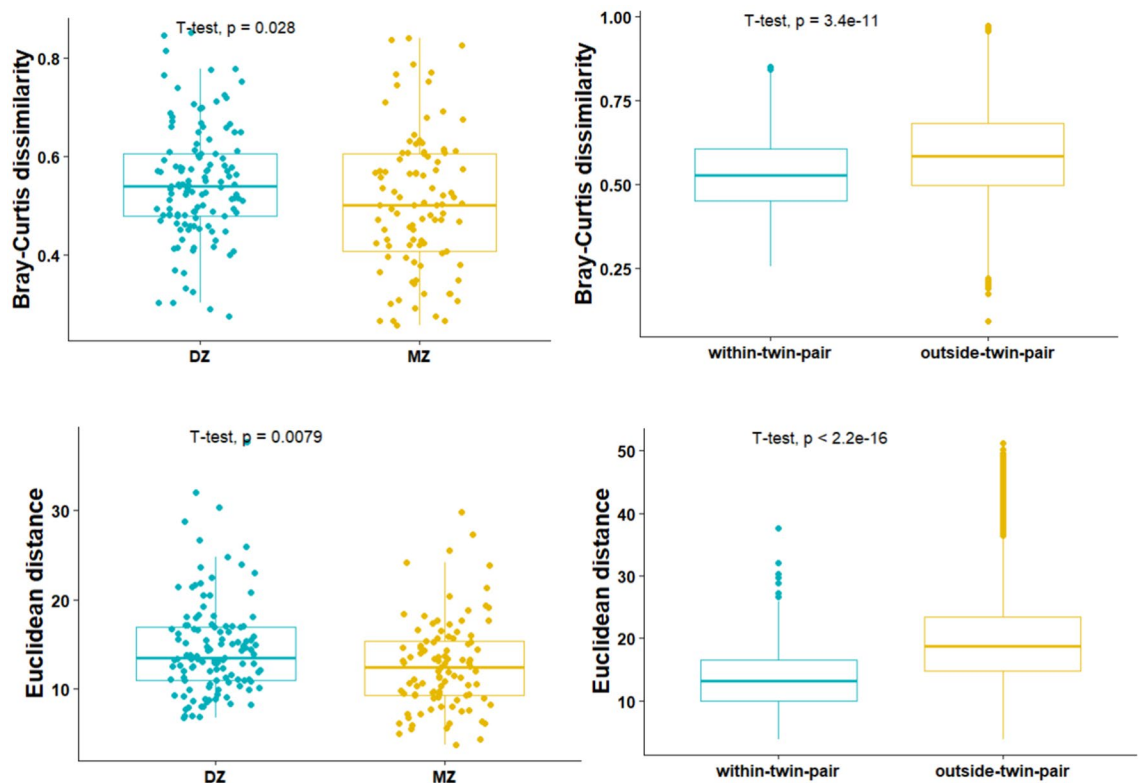


Figure 2. Microbiome and metabolome (dis)similarities between twins. Top: Bray–Curtis dissimilarity in DZ and MZ twins. within-twin pair refers to twin-twin dissimilarity in microbiome composition within a twin; outside-twin pair refers to each of the twins’ microbiome composition in comparison to each of the other twins (except his/her own co-twin). Bottom: Euclidean distances in DZ and MZ twins; and within-twin pair and between-twin analyses. Each dot represents a pair of twins (within-pair) or a pair of twins (independent of each other). Differences in dissimilarity were reported (Student’s t-test, significance level 0.05).

60 metabolites were identified to be significantly different between MD and healthy twins ($p < 0.05$), but none of the ten metabolites identified in the MZ twin pair analysis were re-identified in the DZ twin pair analysis (Supplementary Fig. S3). The predicted metabolic pathway, folate biosynthesis, was identified as significantly differentially abundant between MD and healthy co-twins ($p < 0.05$). Other metabolic pathways involving vitamin metabolism were identified at a significance threshold of 0.1: ascorbate and aldarate metabolism ($p = 0.09$), and retinol metabolism ($p = 0.09$) (Supplementary Fig. S4).

When analyzing *all* discordant twins (combined MZ-DZ population, Fig. 1c), four of the genera identified in the MZ twin pair analysis, and 17 of the genera identified in the DZ twin pair analysis remained significantly abundant ($p < 0.05$). Regarding the metabolomics analyses, ten metabolites that were identified amongst the MZ twins could not be re-identified, whereas nine metabolites of the DZ twin pair analysis were significantly re-identified. Two metabolic pathways were identified in common between the pairwise analyses in DZ and all discordant twins: folate biosynthesis and homologous recombination. A summary of the overlapping findings is shown in Fig. 4.

A distinct set of bacterial genera were found to be significantly associated with the diagnosis of mental disorder in linear mixed regression analyses

A distinct set of associated bacterial genera and metabolites, resulting from previous analyses (Fig. 4), were regressed against individual-level confounders by using linear mixed models (LMs) in two setups: healthy co-twins versus all MD-twins (risk-bearing, healthy co-twins vs MD twins, Fig. 1d), and all healthy twins versus all MD-twins (all healthy vs all MD, Fig. 1e). Confounders such as age, gender, BMI, antibiotics usage of the prior month, and vegetable and fruit intake were included. The MD diagnosis was used as a fixed effect, and random slopes were adopted to capture the twin’s zygosity. Out of the 41 bacterial genera, 14 genera were significantly associated with the MD diagnosis in a comparison of healthy co-twins (74) with MD-twins (144) ($FDR < 0.2$) (Fig. 5). Comparing their distributions, shifts can be seen in the more abundant taxa (left-skewed distributions with fewer negative CLR-transformed values), including genera Parabacteroides, Ruminococcaceae UCG 014, Ruminococcaceae UCG 013, Family XIII AD3011 group and unclassified Christensenellaceae. The existing imbalance in the microbial ecosystem coincides with shifts of more abundant taxa, along with multiple shifts in lower abundant taxa. This was not the case for the comparative setup of all healthy twins (294) with MD twins (144).

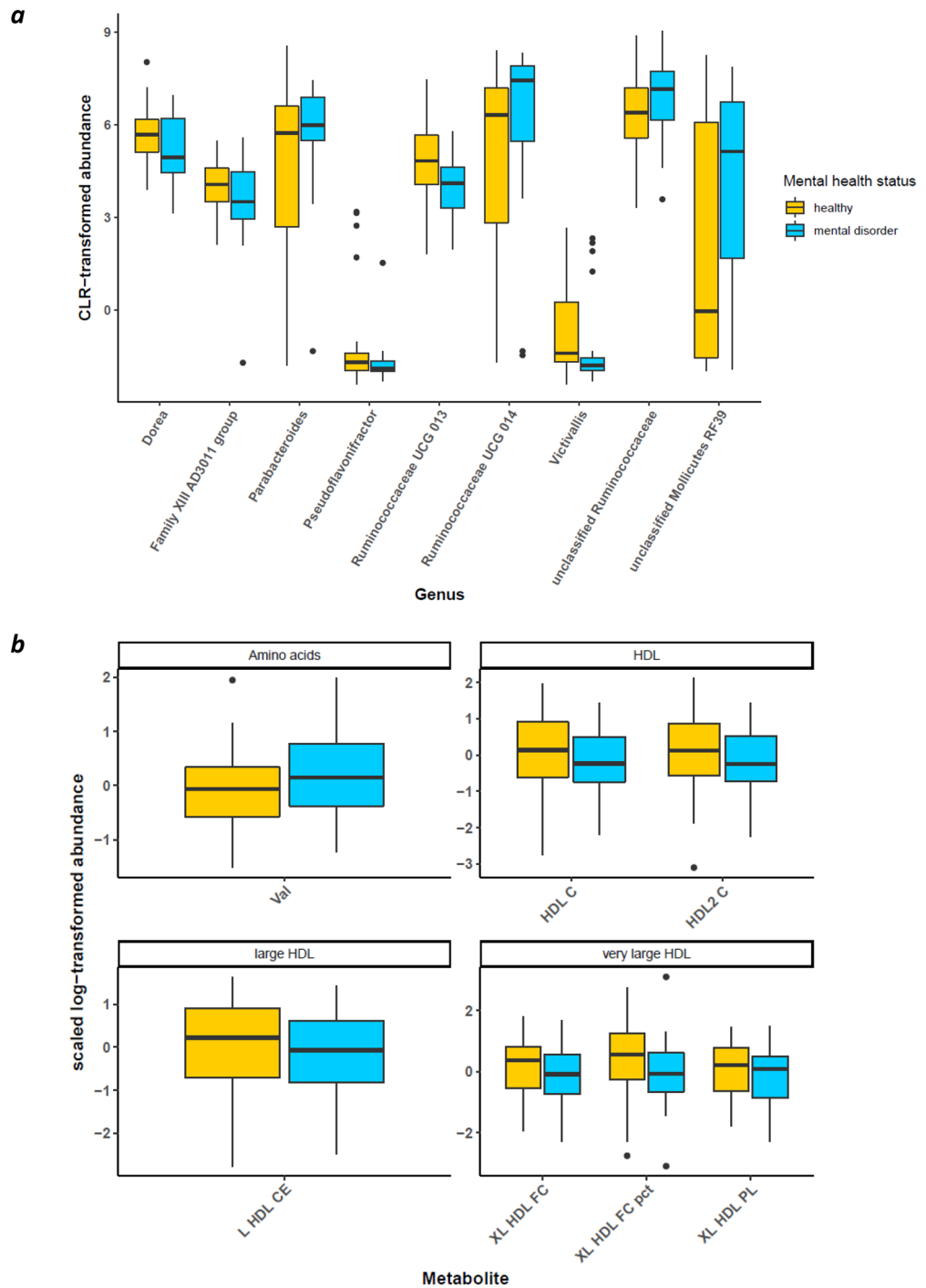


Figure 3. Pairwise analyses in the monozygotic twin population for microbiome and metabolome. **(a)** Significantly differentially abundant ($p < 0.05$, paired samples Wilcoxon test) bacterial genera identified in the MZ discordant analysis. **(b)** Significantly differentially abundant metabolites were identified in the MZ discordant analysis (same ordering as in **(a)**). Val, Valine; HDL-C, high-density lipoprotein (HDL) cholesterol; HDL2-C, HDL-2 cholesterol; L-HDL-CE, large HDL esterified cholesterol; XL-HDL-FC, very large HDL free cholesterol; XL-HDL-FC %, percentage very large HDL free cholesterol; XL-HDL-PL, very large HDL phospholipids.

Predicted metabolic pathways (KEGG Level 3) from gut microbiome resulting from previous pairwise analyses were regressed against the same set of confounders and the fixed effect used for MD-diagnosis in LMMs. Metabolic functions of the MD gut microbiome were predicted to have roles in ascorbate and aldarate metabolism,

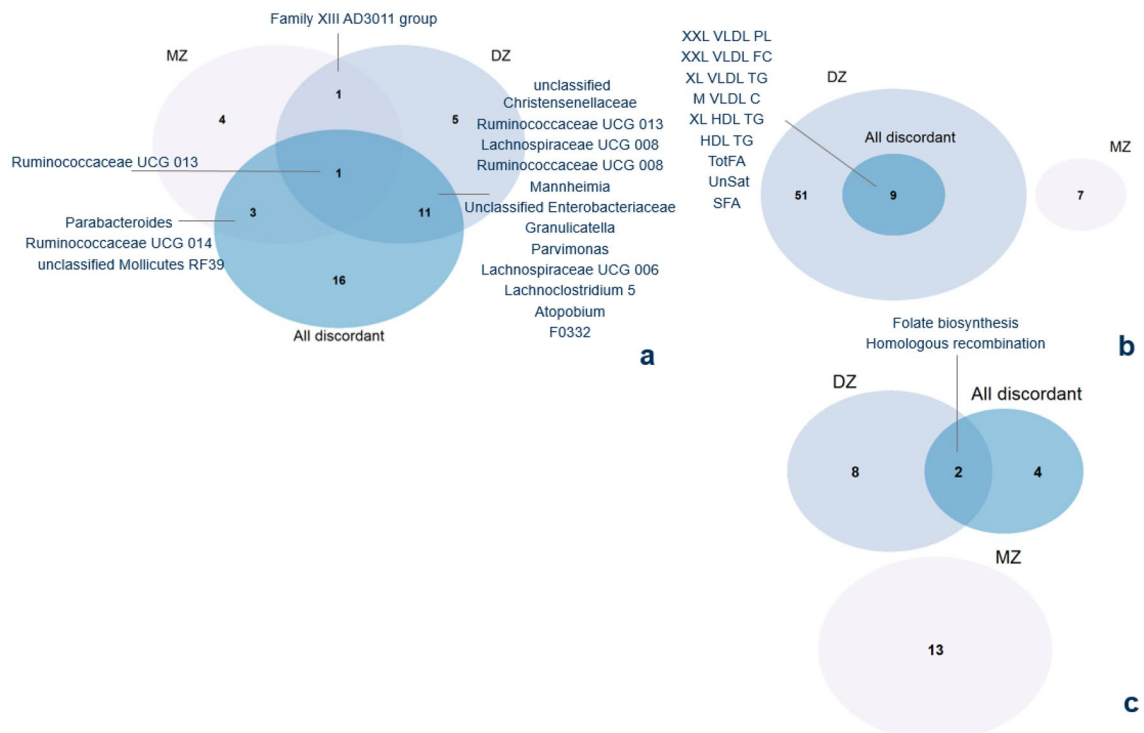


Figure 4. Venn diagrams representing overlaps between each of the pyramid layers: significantly identified microbes (a), metabolites (b), and predicted metabolic pathways (c). XXL-VLDL-PL, extremely large very-low-density lipoprotein phospholipids; FC, free cholesterol; TG, triglycerides; XL, very large; M, medium; HDL, high-density lipoprotein; TotFA, total fatty acids; UnSat, unsaturated fatty acids; SFA, saturated fatty acids; MZ, monozygotic; DZ, dizygotic.

retinol metabolism, metabolism of xenobiotics by cytochrome P450, drug metabolism by cytochrome P450, and microbial metabolism in diverse environments. On the other hand, non-MD gut microbiome metabolic functions involved pathways such as mismatch repair, carbapenem biosynthesis, glucagon signaling pathway, and homologous recombination (FDR < 0.25) (Supplementary Table S1). Regarding the metabolites significantly identified in previous analyses (Fig. 1a–c), none were significantly associated with MD diagnosis in either of the setups.

Multivariate analyses

As indicated in Fig. 1d, e, we compared healthy co-twins (n = 74) with all MD twins (n = 144) (Fig. 1d), followed by an analysis in which all healthy twins (294) were compared to all MD twins (144) considering all genera and all metabolites (Fig. 1e). Resulting from analyses in layer (d) we found, overlapping with previous analyses, genus *Parabacteroides* (increased), *Ruminococcaceae UCG 013* (decreased), *Family XIII AD3011 group* (increased), *Ruminococcaceae UCG 014* (increased), *Lachnospiraceae UCG 006* (decreased), *Lachnospiraceae UCG 008* (decreased), *Oxalobacter* (increased), *Sanguibacteroides* (decreased), *unclassified Christensenellaceae* (increased) and *Subgroup 5* to be significantly associated with MD diagnosis (p-value < 0.05), after accounting for confounders gender, age, antibiotics usage of the prior month, vegetable and fruit intake. However, after performing FDR correction, these appeared no longer significant (FDR < 0.2) (Supplementary Table S2). Metabolites XL HDL TG, total FA, and saturated FA were identified to be significantly associated with MD diagnosis (p-value < 0.05) but remained no longer significant after FDR correction (FDR < 0.20) (Supplementary Table S3). Regarding analyses from layer (e), genus *Parabacteroides*, and *Lachnospiraceae UCG 008* popped up to be significantly associated with MD diagnosis (P-value < 0.05), after accounting for confounders, though no longer significant after FDR-correction (FDR < 0.2) (Supplementary Table S4). None of the previously identified metabolites were found to be significant (Supplementary Table S5). Results of the microbiome analyses across the pyramid (Fig. 1) indicated that genus *Parabacteroides* was strongly associated with MD diagnosis (lowest p-value, highest estimate in LMM), though a stronger effect was seen in the comparison of healthy twins, at risk, in comparison with MD-twins. All bacterial genera that were significantly identified across the multiple layers of the pyramid are visualized in Fig. 6. Of the previously identified metabolic pathways, ascorbate and aldarate metabolism (increased), carbapenem biosynthesis (decreased), retinol metabolism (increased), drug metabolism by cytochrome P450 (increased), metabolism of xenobiotics by cytochrome P450 (increased) and homologous recombination (decreased) were re-identified in the setup of Fig. 1d (Supplementary Table S6). None of them were re-identified in the setup of Fig. 1e (Supplementary Table S7).

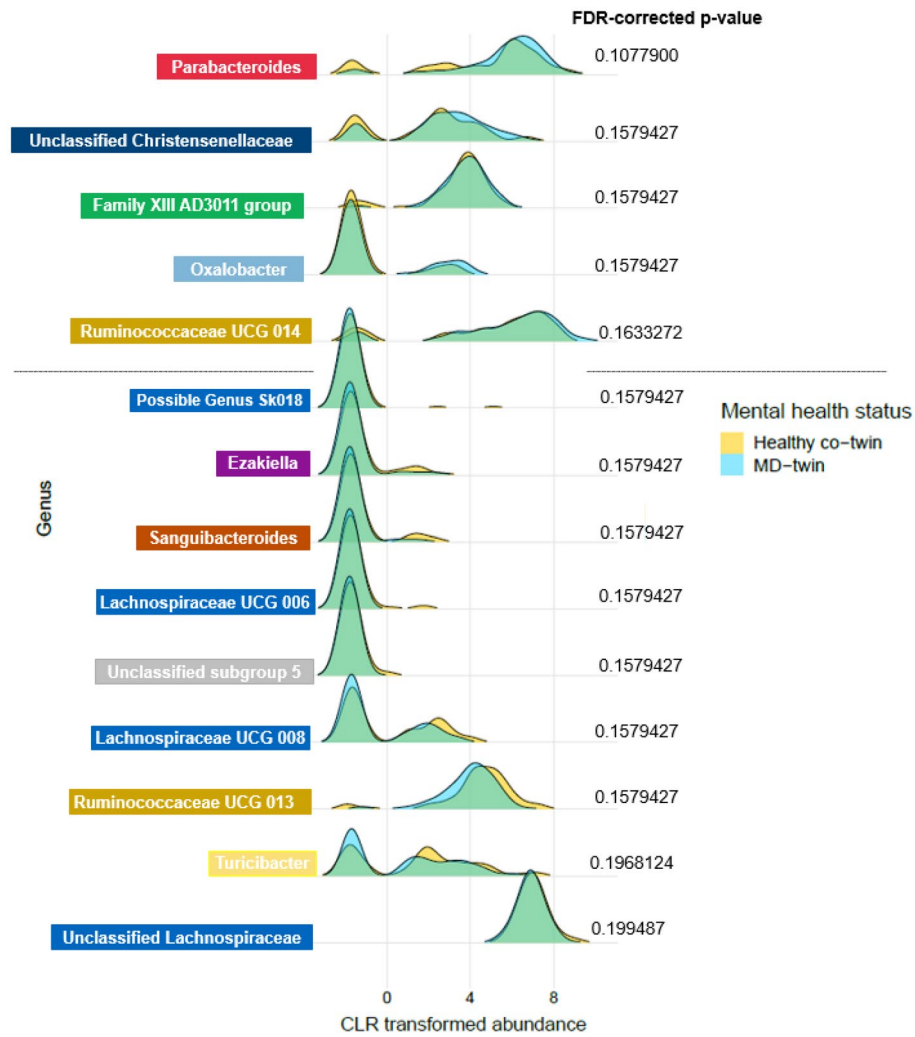


Figure 5. Distributions shown for genera significantly associated with MD-diagnosis in setup healthy co-twin (74) vs. MD-twins (144, all). Multiple testing corrections were performed using the Benjamini–Hochberg method and an FDR value < 0.2 was considered significant. Only significant genera are shown. The grey line separates those genera positively associated (top) and negatively associated with MD diagnosis (bottom). Genera in boxes with the same color are belonging to the same family.

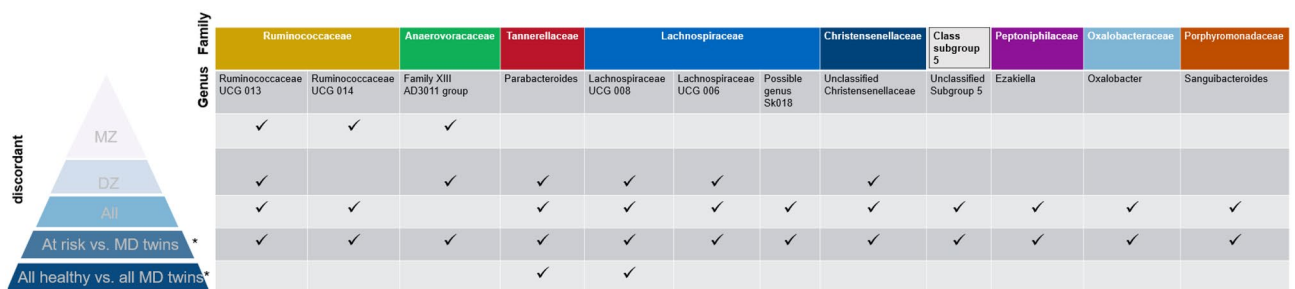


Figure 6. Pyramid layers complemented with significantly associated bacterial genera. Ticks (✓) indicate whether the bacterial genus was significantly identified in a particular layer of the pyramid. Genera are categorized per family represented by distinct colors (except for unclassified subgroup 5, which is identified up to the class level). MZ, monozygotic; DZ, dizygotic; MD, mental disorder; *: genera in these layers were significantly identified at a p-value threshold of 0.05, though no longer significantly identified after FDR correction.

Gut microbiota composition and functional profiles.

A correlation analysis using Spearman's rank correlation coefficients revealed various significant correlations between previously identified differential abundant microbial genera and differentially abundant predicted metabolic pathways (Tax4Fun analyses). The correlation analysis was performed in the population of healthy co-twins versus MD-twins (Fig. 1d). A total of 24 correlations were identified (coefficient $>|0.2|$, FDR < 0.05). The top five correlations were related to genus *Lachnospiraceae* UCG 006: “ko00332: Carbapenem biosynthesis” ($\rho = 0.54$, $p < 0.05$), “ko03440: Homologous recombination” ($\rho = 0.55$, $p < 0.05$), “ko04922: Glucagon signaling pathway” ($\rho = 0.45$, $p < 0.05$). Moreover, unclassified subgroup 5 was significantly correlated with the “ko03440: Homologous recombination” ($\rho = 0.57$, $p < 0.05$) and “ko04922: Glucagon signaling pathway” ($\rho = 0.47$, $p < 0.05$). A summary of the correlation analysis is shown in Fig. 7.

Discussion

Gut microbiome dysbiosis has been suggested to be a hallmark of mental disorders⁴⁵. In our study, we profiled the gut microbiome and plasma metabolome in both MZ and DZ twins with MD-diagnosed and unaffected twins. Since MDs share genetic underpinnings, healthy co-twins were considered to be at an increased risk of developing an MD. We observed significantly higher similarity in gut microbiome and plasma metabolome composition in MZ twins, compared to DZ twins. This suggests that both microbiome and metabolome composition is shaped by experiencing similar environments, yet hinting at a host genetic effect. These findings support the co-twin control design, in which we control for both genetic diversity and early-life factors/environment that contribute to within-twin microbiome similarity.

Studying the gut microbiome in MD-discordant and concordant twin pairs aims to elucidate the contribution of the gut microbiota in the (risk of) MD development, because of the shared (childhood) environment, i.e., maternal sites and shared household as well as genetic factors between individuals from the same twin pair. We adopted a pyramid-layer design for our analyses. First, twins with the same zygosity were compared for MD in a pairwise manner. Second, healthy co-twins (healthy twins at risk) were compared with all MD twins, as well as *all* healthy twins against all MD twins. The latter, due to its higher sample size, increased statistical power and allowed us to validate the robustness of previous findings. These results allowed us to find candidate alterations in the gut microbiome or plasma metabolome landscape, assuming a shared genetic background and (childhood) environment. However, since these candidate alterations might as well be the result of individual-level confounders such as antibiotics usage, dietary influences, BMI, and age differences, they were validated in a multivariate design. There, we made a separation between two layers: (1) healthy cotwins in comparison to all MD twins and (2) all healthy versus all MD twins. This separation helps to identify specific alterations that may be associated with the risk of developing MD on the one hand (layer d) and on the other hand, layer (e) aims to provide a broader, comprehensive assessment of healthy vs MD twins across the larger population. The bottom

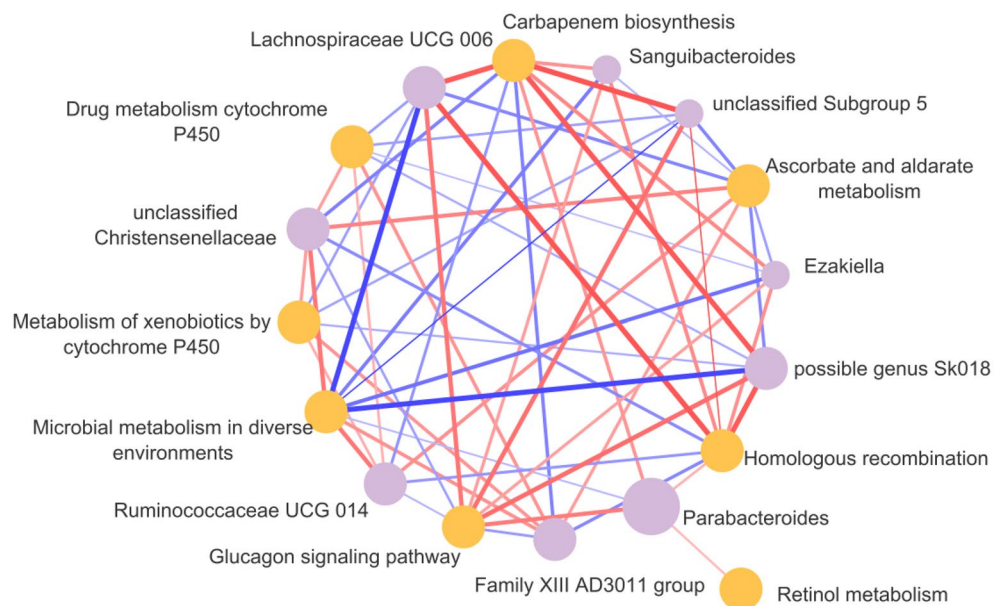


Figure 7. Correlation networks showing the relationships between differential microbial genera and the differential microbial KEGG level 3 metabolic pathways. The node sizes and colors (purple: differential microbial genera; orange: differential microbial metabolic pathways) are proportional to the number of significant associations across the layers of the pyramid. The width and color of the lines (red: positive, blue: negative) are proportional to the correlation strength. Only significant correlations (FDR-adjusted p-values with Benjamini–Hochberg procedure, < 0.05) are displayed, with a correlation strength $>|0.2|$. Permission has been obtained from Kanehisa laboratories for using KEGG pathway database⁸⁰.

layer aids in confirming the significance and association of microbial genera/metabolites with the MD diagnosis. In this way, significant *candidate* alterations in both microbiome and plasma metabolome were considered potential biomarkers when their presence was consistently noticed in all layers of the pyramid. As a result, genus *Parabacteroides* was identified in all layers, as most strongly associated with the MD diagnosis rather than with other individual-level confounders.

The increased abundance of genus *Parabacteroides* (family *Tenerellaceae*) is in concordance with recently published reports in which the genus was found to be elevated in depressed subjects compared to healthy controls^{46–48}. Since *Parabacteroides* has been considered to have health-promoting effects, i.e., the production of SCFAs, the elevation of the genus in depressed subjects has been suggested to be a compensatory mechanism⁴⁸. *Parabacteroides* species have also been linked to the production of metabolites γ -aminobutyric acid (GABA)⁴⁹ and indole⁵⁰ of which dysfunctions in their signaling pathways are related to depression and anxiety^{51,52}. *Parabacteroides* enrichment was shown to alter gene expression in pathways associated with metabolic function, neurodegenerative diseases, and dopaminergic signaling⁵³. Hence, the *Parabacteroides* genus has been associated in previous reports with both protective and pathogenic effects on human health^{54,55}. Nevertheless, the exact role played by this bacteria in depressive, stress, and anxiety-like behaviors remains to be investigated.

We found a positive correlation between the abundance of *Parabacteroides* and the glucagon signaling pathway (Fig. 7). The role of *Parabacteroides* species in the context of metabolic dysfunctions has been discussed in several earlier reports, though with inconsistent findings.

As previously suggested, we would like to argue that functionally related microbial communities rather than single microbes have an overall impact. First, genus *Parabacteroides* has been confirmed in a recent meta-analysis to be enriched in major depressive disorder in more than 20% of the included studies (24 in total). The authors highlighted *Parabacteroides*' involvement in greater utilization of glutamate and increased synthesis of GABA⁵⁶. In our research, we noticed overlapping distributions of this genus' (and other genera) abundance in healthy co-twins compared with MD-twins, though with a slightly larger abundance in MD-twins (Fig. 5). However, due to the resolution of the TwinsUK dataset, the higher abundance of this specific genus cannot be brought back to specific species. Second, diagnoses of mental disorders in our research were not limited to depressive disorders, but also include anxiety, stress, and eating disorders that might explain the less strong association compared to earlier studies. For instance, a high abundance of *Parabacteroides* could be an indication of MD symptoms. Other genera that were identified through multiple layers from our pyramid design include decreased abundances of genera *Ruminococcaceae* UCG 013, *Lachnospiraceae* UCG 008, *Lachnospiraceae* UCG 006, and increased abundances of *Ruminococcaceae* UCG 014, *Family XIII AD3011 group*, and unclassified *Christensenellaceae*. Genera of families *Lachnospiraceae* and *Ruminococcaceae* were previously associated with the maintenance of gut health⁵⁷. Members of *Lachnospiraceae* are among the main producers of SCFA, though different taxa of this family have been associated with intra- and extraintestinal diseases⁵⁸. Since *Lachnospiraceae* are strongly influenced by diet specifics, such as non-starch polysaccharides and high-fat feeding⁵⁸, the associations in both directions were not unexpected. In contrast to other reports, only a few genera were identified in significant association with MD diagnosis. Gut microbiome composition may be different in patients with an MD in partial or full remission compared to patients with a current illness phase. Since we could not account for illness states, state-related changes in the gut microbiome composition could not be established in our study design⁵⁹. We suggest combining multiple omics approaches (meta-transcriptomics, proteomics, metabolomics) to reveal a *systems-level insight* into functional differences rather than to retain a compositional focus.

In this study, we observed that metabolic functions involving vitamin metabolism, including retinol and ascorbate and aldarate metabolism were more abundant in the gut microbiome of patients with MDs compared to non-MD controls. It has been hypothesized that vitamin A levels are inversely associated with depression^{60,61}. In addition, previous work has established the role of gut microbiota in vitamin A (retinol) metabolism^{62,63}. By regulating the expression of gene *Rdh7*, gut bacteria can curb the production of vitamin A metabolite retinoic acid (RA). The increase in ascorbate metabolism, associated with MD is, in contrast, to what has been found earlier⁶³. However, the pathophysiological association of ascorbate with depression is currently not clear⁶⁴.

In the setup of healthy co-twins, assumed to be at risk of developing MD, in comparison with MD-twins (Fig. 1d), more genera were found to be significantly associated with MD than when considering the setup of all healthy twins compared with the MD-twins (discordant and concordant twins included). In addition, the size of the estimate of *Parabacteroides* in association with MD diagnosis was stronger in the former setup. Thus, the inclusion of healthy concordant twins (layer e) (no identifiable risk of MD) weakened the association of *Parabacteroides* with MD diagnosis. The number of genera identified and consequently the metabolic pathways differentially identified were different between the two setups (Supplementary Tables S2, S4, S6–7). Further studies with increased study sample sizes are expected to unravel the contributions of microbiota in individuals at increased genetic risk. Unlike other omics data, the microbiome's dynamic nature does not only suggest a role as a *modifier* to disease risk, but, not unimportantly, appears very suitable for real-life applications in terms of privacy and security reasons⁶⁵.

Compared to healthy controls, none of the metabolomic alterations identified in the pairwise analyses (both MZ and DZ) remained significant after correcting for individual-level confounders. This stands in contrast to several reports that showed plasma metabolomic markers to discriminate healthy controls from individuals with depression and anxiety^{66–69}. Plasma metabolites identified in pairwise analyses in the MZ group (Fig. 2a) varied greatly from those identified in the DZ group (Supplementary Fig. S3). First, the NMR platform mainly captures lipid metabolites rather than the full metabolome profile. Second, the variation might be explained by the BMI differences in the MZ and DZ population, between the MD-diagnosed and control group (Supplementary Fig. S5). The control group in the DZ population demonstrates a higher BMI, compared with the MZ population, along with higher levels of VLDL metabolites in the DZ control group and higher levels of HDL in the MZ control group. This is in line with previous reports on dyslipidemia and obesity^{70,71}. To cope with this

issue, future work should envision individual reference intervals, that exploit longitudinal subject-specific data rather than looking into population ranges⁷².

Our study has several limitations. First, while highly relevant covariates were included (age, antibiotics, gender, vegetable, and fruit intake), other environmental triggers might not have been comprehensively tracked. Seen the major contribution of dietary influences on the host's microbiome, more detailed dietary information on micro- and macronutrients and different food groups must be considered to yield better insights into the microbiome-metabolome-brain crosstalk. This will allow us to gain insight into what dietary patterns are protective and contribute to better mental health. Second, *all* individuals with diagnoses of MD were included, i.e., diagnoses that were ongoing at the moment of sample collection, as well as individuals previously diagnosed with an MD. When the patient is under treatment or has overcome his/her mental illness, the microbiome composition is expected to change again to restore gut dysbiosis⁷³. Third, a more balanced population sampling exploiting additional cohorts is necessary to obtain more reliable results. The TwinsUK cohort consists of mainly female individuals (96%). Therefore, the research results are mainly representative of the female population. A gender-specific version of performed analyses might be considered to see the existing sex differences in behavior and risk for neuropsychiatric disorders⁷⁴. Fourth, several diagnoses of MDs were categorized under the same heading to identify common pathways emerging in individuals with MDs, including depression, anxiety and stress, and eating disorders. Although the included diagnoses of disorders differed in their pathophysiology, common pathways are emerging in psychiatric diseases with immune dysregulation being present in a broad spectrum of neuropsychiatric disorders, increasing mental vulnerability and risk of psychiatric symptoms⁷⁵. Furthermore, these diagnoses have shown associations with increased gut permeability, impaired gut-brain communication, and disrupted neurotransmitter production^{20–23,76–78}. The combination of these factors highlights the intricate interplay between the gut microbiota, the immune system, and neurological processes in the context of MDs. Fifth, although the identified changes in the intestinal microbiota have been previously identified along with its mechanisms of altered glutamate and GABA metabolism and butyrate production⁵⁶, we would like to emphasize the need for more established protocols in microbiome studies to reveal the relationships between mental disorders and the gut microbiome (Supplementary Information, Appendix 1). Importantly, the utilization of the 16S rRNA gene sequencing method, which is inherent to the TwinsUK study design, imposes certain limitations on the extent of the microbiome analysis. The method is unable to provide species-level resolution or detailed information on the functional capacity of the microbiota. However, we applied the method with the overall idea to identify differences in the structure and composition of microbial communities. Moreover, we agree that longitudinal cohort studies, the inclusion of confounding covariates to disentangle cause/consequence as well as the exploitation of multi-omics approaches will aid to uncover the activity of our microbiomes beyond their composition.

Conclusions

In this study, we robustly identified that individuals diagnosed with MD showed a greater abundance of the genus *Parabacteroides*. This genus might offer a novel opportunity for primary prevention. However, the exact role and effect of this genus in MD initiation and progression remain to be investigated in further studies as the effects are likely species-specific. In general, we can conclude that our presented pyramid-layering design in a population of twins offers a robust manner to identify differentially abundant features in diseases with both genetic and environmental components underlying. We propose microbiome-mediated mechanisms as likely gut biomarkers and/or mediators in the context of MD as a rationale for future research.

Materials and methods

Study population

The TwinsUK cohort is a longitudinal dataset and the largest research cohort of adult twins from the United Kingdom (<http://www.twinsuk.ac.uk/>). The cohort comprises over 14,000 volunteers followed over more than two decades. Participants were predominantly female (>80%) and aged 18–103. This cohort has been extensively studied for a wide range of clinical and behavioral outcomes. All work involving human subjects was approved by the Cornell University IRB (Protocol ID 1108002388) and all methods were performed following relevant guidelines and regulations. Informed consent was obtained from all participants.

In our study, 438 twins (219 pairs) were included, who have both 16S rRNA gut microbiome data and concurrent plasma metabolomics data at the same time point. Amongst the 438 twins (219 twin pairs), 122 were DZ, and 97 were MZ twin pairs (Fig. 8). Regarding the phenotypic variable on MD, different groups were specified according to the Hospital Anxiety and Depression (HADS) questionnaire as follows. For participants that have one of the questions related to Depression (see Appendix 1) specified on “1” (yes), “Anxiety/Stress disorder” and “eating disorder”, the variable “MD” was specified on “1” (yes). Individuals who had none of the MD-related variables specified on “1”, were defined as “healthy control”, in the assumption that “healthy” means no diagnosed MDs. The final dataset consisted of 294 controls and 144 cases with one or more MDs diagnosed. Detailed description of the implementation of the *Hospital Anxiety and Depression (HADS)* questionnaire in TwinsUK cohort is available⁷⁹.

Microbiome sequences from 3335 samples, including metabolic abundances from 4830 samples, were collected from the TwinsUK cohort. First, for all samples in the microbiome dataset, twin pairs were screened based on zygosity, age, year of birth, the time point at which the microbiome sample was taken, and type of descendant. Only pairs with microbiome samples matched with their corresponding time point were considered, resulting in 1254 pairs with 2508 samples. Next, from these microbiome samples, only 524 individuals were matched with their complementary metabolome samples at the same time point. Finally, only twin pairs without missing

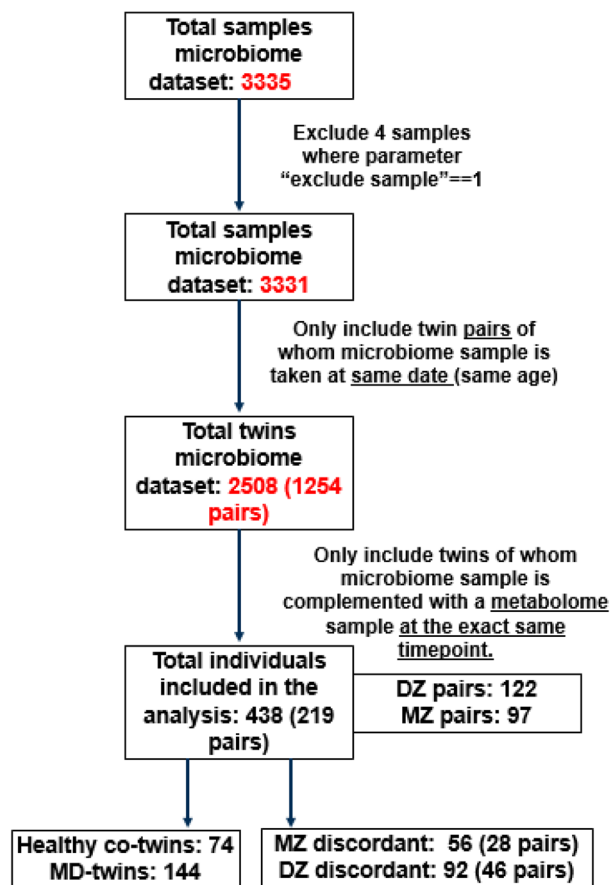


Figure 8. Overview of selected samples (microbiome/plasma metabolomics).

measurements across time points were considered in the analysis. The final dataset consisted of 438 individuals (219 twin pairs), with microbiome and metabolome measurements from 2010 to 2015.

Data preprocessing

The microbiome data from the TwinsUK cohort was delivered in a preprocessed format, i.e. amplicon sequence variants (ASVs) were generated using the DADA2 pipeline. Samples achieving a sequence depth of less than 10,000 were excluded, and sequences that were unassigned at Kingdom/Phylum level were removed. Taxonomy was assigned to the ASV sequences with the SILVA reference database⁸⁰. In the next step, the ASVs were imported, together with their associated metadata (see Metadata processing) using the *phyloseq* package (v1.40.0)⁸¹. The dataset was filtered using the following criteria: (1) taxa observed only once in the whole dataset were removed, (2) taxa must appear in more than one sample, (3) ASVs with no counts were removed (4) empty samples were removed (none in the dataset). Finally, data were agglomerated to genus level and transformed by using the centered log-ratio transformation (CLR), i.e. each of the taxon counts is divided by the geometric mean of all taxa in a sample, and subsequently logarithmically transformed. The CLR transformation was implemented in the *microbiome* R package (v1.18.0).

Metabolic measurements were available for 228 metabolic traits: 147 metabolite concentrations (comprising groups: Cholesterol, Lipoprotein subclasses, Glycerides and Phospholipids, Apolipoproteins, Total Fatty acids, Amino Acids, Ketone bodies, Fluid balance, and Inflammation), 77 lipid ratios, three lipoprotein particle sizes and a semi-quantitative measure of albumin. Metabolic profiling was conducted by Nightingale Health Ltd. (Helsinki, Finland) using a targeted NMR spectroscopy platform. These metabolic measurements were log-transformed; for all measurements before transformation, a pseudo-count of 1 was added. Next, measurements were shifted to zero mean and scaled to a standard deviation (SD) of 1⁸². Metabolites that were assigned missing values in more than 25% of the individuals were removed. Metabolites that had missing values, in less than 25% of the individuals, were replaced by their median value. For the confounder-related questionnaire data, selected features were included that are known to confound gut microbiome composition: age, gender, BMI, antibiotics, vegetable and fruit intake⁸³. Data was taken at the time point at which the microbiome/metabolome sample was collected; if not available, data were collected from the closest available time point. Missing values were replaced by the median value of the respective variable.

Statistical analyses

The statistical analyses were performed using the R statistical software version 4.1.2 (2021–11-01, <https://www.rstudio.com/products/rstudio/download/>), and different libraries included in Bioconductor (v3.13) were used, including *microbiome*, *vegan* (v2.6.2), and *phyloseq*. Microbiota beta-diversity was calculated by the Bray–Curtis dissimilarity metric, which quantifies the compositional dissimilarity between two samples or groups, considering both overall abundance per sample as well as the abundance of each taxon⁸⁴. The Euclidean distance was used for comparisons of metabolic profiles. Tests for paired data were used to address the between-group comparisons (MD vs healthy) of continuous variables. The Shapiro–Wilk’s method was used to test for normality of the distribution and accordingly, the student’s t-test (p-value > 0.05) or Wilcoxon’s test was used. The level of significance was set at the two-tailed p-value of 0.05.

For the multivariate analyses, association analyses were performed for (1) healthy co-twins vs. MD-twins and (2) healthy twins vs. twins with MD. For this, a modified version of the model proposed by New et al.⁸⁵ is proposed as follows:

$$Y_i \sim N(\alpha_{Z_{j|i}} + X\beta, \sigma^2)$$

$$\alpha_{MZ,j} \sim N(0, \sigma_{MZ}^2), \text{ for } MZ \text{ twin pair } j = 1, \dots, J_{MZ}$$

$$\alpha_{DZ,j} \sim N(0, \sigma_{DZ}^2), \text{ for } DZ \text{ twin pair } j = 1, \dots, J_{DZ}$$

where Y_i is the microbe/metabolite abundance; $\alpha_{Z_{j|i}}$ is the random effect where accounted for either MZ or DZ twins, i.e. $\alpha_{MZ,j}$ and $\alpha_{DZ,j}$ respectively; X are the predictor variable(s), β are the slope estimates explaining the contribution of the predictor variable to the dependent variable, and σ^2 is the variance. In the model, each twin was assumed to have a random slope, specifically random slopes for MZ twins and for DZ twins coming from two different distributions. The model estimates variance (relatedness) between MZ twins separately from the variance between DZ twins (relatedness) between DZ twins. CLR-transformed bacterial genera abundances and scaled, log-transformed metabolite abundances were regressed against main confounders: age, gender, BMI, antibiotics usage prior month, vegetable and fruit intake. The regression was done by using a LMM, using the *nlme* R package (v3.1.157). The MD categorical variable was included as fixed effect. We corrected for multiple testing by using the Benjamini–Hochberg approach. For multivariate analyses, FDR-thresholds were relaxed to 0.25 since we are interested in an exploration and screening of potential features (bacteria/metabolites) of interest and thus less strict on the inflated type-I error.

Differentially abundant bacterial genera along with differentially abundant KEGG pathways, that were significantly associated with MD-diagnosis in the multivariate analyses, were used in the correlation analysis based on Spearman’s rank correlation coefficients. The cutoffs for the correlation coefficients and adjusted p-values (FDR correction with Benjamini–Hochberg procedure) were determined to equal 0.2 and 0.05, respectively. Networks were built using Cytoscape (<https://cytoscape.org/>; v3.9.1) to represent the correlation network of bacterial genera and metabolic pathways.

Predicted functional potential

The Tax4Fun2 open-source R package (v1.1.5) was used to predict the functional capabilities of the microbial communities based on 16S rRNA datasets, applicable to SILVA-labeled microbial abundances⁸⁶. To predict the community function, the table with sequences of ASVs and the table with sequence abundances were required. Both tables were read by Tax4Fun2 through an external file; sequences were saved in FASTA format through the `write.fasta()` function from the `seqinr` package. In the first step, the sequences of the samples were aligned with the package reference database via BLAST, using the `runRefBlast()` function with `database_mode` option “Ref100NR” for greater precision in alignment. Next, `makeFunctionalPrediction()` function was used to obtain the prediction of the abundance of functions in the community, with a similarity cut-off set to 97%. As a result, two files were generated. The first one contained KEGG ortholog profiles, i.e., predictions of the relative abundances of each enzyme indicated for each sample, according to their composition (functional predictions). The second one contained predictions on the abundances of the metabolic pathways of each sample according to its composition. The latter was organized into three hierarchical levels, with different categories as corresponding to the metabolic pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG)⁸⁷. Tax4Fun2 output for the level 3 KEGG entries were analyzed.

Ethics declarations

Ethics approval was granted by the St. Thomas’ Hospital Research Ethics Committee. Following the restructure and merging of the research ethics committee, subsequent amendments were approved by the National Research Ethics Service (NRES) Committee London–Westminster (TwinsUK reference EC04/015); approval for the use of microbiota samples was granted by the NRES Committee London–Westminster (The Flora Twin Study reference 12/LO/0227).

Data availability

The data that support the findings of this study are available from TwinsUK but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of TwinsUK. Accession

codes: The European Bioinformatics Institute (EBI) accession numbers for the microbial DNA sequences reported in this paper is ERP015317.

Received: 11 November 2022; Accepted: 6 February 2024

Published online: 13 February 2024

References

- Koch, S. V. *et al.* Associations between parental socioeconomic-, family-, and sibling status and risk of eating disorders in offspring in a Danish national female cohort. *Int. J. Eat. Disord.* **55**, 1130–1142 (2022).
- Ding, Q., Bi, D., Zhou, Y., Bai, X. & Li, X. Genetic and environmental sources of familial resemblance in anxiety: A nuclear twin family design. *Psychol. Med.* **2021**, 1–9. <https://doi.org/10.1017/S0033291721001197> (2021).
- Baba, A., Kloiber, S. & Zai, G. Genetics of social anxiety disorder: A systematic review. *Psychiatr. Genet.* **32**, 37–66 (2022).
- Li, C. *et al.* A multi-environments-gene interaction study of anxiety, depression and self-harm in the UK Biobank cohort. *J. Psychiatr. Res.* **147**, 59–66 (2022).
- Fu, Z. *et al.* Air pollution, genetic factors and the risk of depression. *Sci. Total Env.* **850**, 158001 (2022).
- Chung, S.-Y., Kostev, K. & Tanislav, C. Dysbiosis: A potential precursor to the development of a depressive disorder. *Healthcare* **10**, 1503 (2022).
- Sonali, S. *et al.* Mechanistic insights into the link between gut dysbiosis and major depression: An extensive review. *Cells* **11**, 1362 (2022).
- Seitz, J., Trinh, S. & Herpertz-Dahlmann, B. The microbiome and eating disorders. *Psychiatr. Clin. N. Am.* **42**, 93–103 (2019).
- Queiroz, S. A. L. *et al.* The gut microbiota-brain axis: A new frontier on neuropsychiatric disorders. *Front. Psychiatry* **2022**, 13 (2022).
- Martin, C. R., Osadchiy, V., Kalani, A. & Mayer, E. A. The brain-gut-microbiome axis. *Cmgh* **6**, 133–148 (2018).
- Grochowska, M., Wojnar, M. & Radkowski, M. The gut microbiota in neuropsychiatric disorders. *Acta Neurobiol. Exp. (Wars)* **78**, 69–81 (2018).
- Rutsch, A., Kantsjö, J. B. & Ronchi, F. The gut-brain axis: How microbiota and host inflammasome influence brain physiology and pathology. *Front. Immunol.* **11**, 59. <https://doi.org/10.3389/fimmu.2020.604179> (2020).
- Settanni, C. R., Ianiro, G., Bibbò, S., Cammarota, G. & Gasbarrini, A. Gut microbiota alteration and modulation in psychiatric disorders: Current evidence on fecal microbiota transplantation. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **109**, 110258 (2021).
- Wallace, C. J. K., Foster, J. A., Soares, C. N. & Milev, R. V. The effects of probiotics on symptoms of depression: Protocol for a double-blind randomized placebo-controlled trial. *Neuropsychobiology* **79**, 108–116 (2020).
- Vujkovic-Cvijin, I. *et al.* Host variables confound gut microbiota studies of human disease. *Nature* **587**, 448–454 (2020).
- Rodríguez, J. M. *et al.* The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb. Ecol. Health Dis.* **26**, 1456 (2015).
- Sung, J., Park, H. A. & Lee, S. J. Microbiome studies and twin research. In *Twin Research for Everyone* 649–664 (Elsevier, 2022). <https://doi.org/10.1016/B978-0-12-821514-2.00010-6>.
- Vilchez-Vargas, R. *et al.* Gut microbial similarity in twins is driven by shared environment and aging. *EBioMedicine* **79**, 104011 (2022).
- Goldberg, J. & Fischer, M. Co-twin Control Methods. In *Encyclopedia of Statistics in Behavioral Science* (Wiley, 2005). <https://doi.org/10.1002/0470013192.bsa143>.
- Fan, Y. *et al.* The gut microbiota contributes to the pathogenesis of anorexia nervosa in humans and mice. *Nat. Microbiol.* **8**, 787–802 (2023).
- Huang, F. & Wu, X. Brain Neurotransmitter Modulation by Gut Microbiota in Anxiety and Depression. *Front. Cell Dev. Biol.* **9**, 649103 (2021).
- Malan-Muller, S. *et al.* The gut microbiome and mental health: Implications for anxiety- and trauma-related disorders. *OMICS* **22**(2), 90–107 (2018).
- Chami, R., Monteleone, A. M., Treasure, J. & Monteleone, M. Stress hormones and eating disorders. *Mol. Cell. Endocrinol.* **2019**, 497 (2019).
- Gonda, X. *et al.* Genetic variants in major depressive disorder: From pathophysiology to therapy. *Pharmacol. Therapeut.* **194**, 22–43. <https://doi.org/10.1016/j.pharmthera.2018.09.002> (2019).
- Penninx, B. W., Pine, D. S., Holmes, E. A. & Reif, A. Anxiety disorders. *The Lancet* **397**, 914–927. [https://doi.org/10.1016/S0140-6736\(21\)00359-7](https://doi.org/10.1016/S0140-6736(21)00359-7) (2021).
- Bulik, C. M., Blake, L. & Austin, J. Genetics of eating disorders: What the clinician needs to know. *Psychiatric Clin. N. Am.* **42**, 59–73. <https://doi.org/10.1016/j.psc.2018.10.007> (2019).
- Huang, W. Y. *et al.* Concordance among monozygotic and dizygotic twins from a population-based sample for self-reported atopic triad, syndrome x, and psychiatric conditions. *Genet. Med.* **2**, 79 (2000).
- Kipman, A., Gorwood, P., Mouren-Siméoni, M. C. & Adès, J. Genetic factors in anorexia nervosa. *Eur. Psychiatry* **14**, 189–198 (1999).
- Holland, A. J., Sicotte, N. & Treasure, J. Anorexia nervosa: Evidence for a genetic basis. *J. Psychosom. Res.* **32**, 561–571 (1988).
- Holland, A. J., Hall, A., Murray, R., Russell, G. F. M. & Crisp, A. H. Anorexia nervosa: A study of 34 twin pairs and one set of triplets. *Br. J. Psychiatry* **145**, 414–419 (1984).
- Fichter, M. M. & Noegel, R. Concordance for bulimia nervosa in twins. *Int. J. Eat. Disord.* **9**, 255–263 (1990).
- kortegeard, L. S., Hoerder, K., Joergensen, J., Gillberg, C. & Kyvik, K. O. A preliminary population-based twin study of self-reported eating disorder. *Psychol. Med.* **31**, 361–365 (2001).
- Skre, I., Onstad, S., Torgersen, S., Lygren, S. & Kringlen, E. A twin study of DSM-III-R anxiety disorders. *Acta Psychiatr. Scand.* **88**, 85–92 (1993).
- Kim, S. Y. *et al.* Comparison of the differences in state-trait anxiety inventory scores and insomnia histories between monozygotic and dizygotic twins: A cross-sectional study using KoGES HTS Data. *J. Clin. Med.* **11**, 4011 (2022).
- McGuffin, P. A hospital-based twin register of the heritability of DSM-IV unipolar depression. *Arch. Gen. Psychiatry* **53**, 129 (1996).
- Whitfield, K. E., Edwards, C. L., Brandon, D. & McDougald, C. Genetic and environmental influences on depressive symptoms by age and gender in African American twins. *Aging Ment. Health* **12**, 221–227 (2008).
- Vitaro, F., Brendgen, M. & Arseneault, L. The discordant MZ-twin method: One step closer to the holy grail of causality. *Int. J. Behav. Dev.* **33**, 376–382 (2009).
- Crush, E., Arseneault, L., Danese, A., Jaffee, S. R. & Fisher, H. L. Using discordant twin methods to investigate an environmentally mediated pathway between social support and the reduced likelihood of adolescent psychotic experiences. *Psychol. Med.* **50**, 1898–1905 (2020).
- Smith, M. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- Lee-Sarwar, K. A., Lasky-Su, J., Kelly, R. S., Litonjua, A. A. & Weiss, S. T. Metabolome-microbiome crosstalk and human disease. *Metabolites* **10**, 563. <https://doi.org/10.3390/metabo10050181> (2020).

41. Visconti, A. *et al.* Interplay between the human gut microbiome and host metabolism. *Nat. Commun.* **2019**, 10 (2019).
42. Lim, M. Y. *et al.* Association between gut microbiome and frailty in the older adult population in Korea. *J. Gerontol.: Ser. A* **76**(8), 1362–1368 (2021).
43. Chen, J. *et al.* Age-specific differential changes on gut microbiota composition in patients with major depressive disorder. *Aging (Albany N.Y.)* **10**(12), 2764–2776 (2020).
44. Caso, J. R. *et al.* Gut microbiota, innate immune pathways, and inflammatory control mechanisms in patients with major depressive disorder. *Transl. Psychiatry* **11**, 645 (2021).
45. Halverson, T. & Alagiakrishnan, K. Gut microbes in neurocognitive and mental health disorders. *Ann. Med.* **52**, 423–443 (2020).
46. Cheung, S. G. *et al.* Systematic review of gut microbiota and major depression. *Front. Psychiatry* **10**, 56. <https://doi.org/10.3389/fpsy.2019.00034> (2019).
47. Amirkhazadeh-Barandouzi, Z., Starkweather, A. R., Henderson, W. A., Gyamfi, A. & Cong, X. S. Altered composition of gut microbiota in depression: A systematic review. *Front. Psychiatry* **11**, 1–10. <https://doi.org/10.3389/fpsy.2020.00541> (2020).
48. Knuesel, T. & Mohajeri, M. H. The role of the gut microbiota in the development and progression of major depressive and bipolar disorder. *Nutrients* **14**, 256. <https://doi.org/10.3390/nu14010037> (2022).
49. Strandwitz, P. *et al.* GABA-modulating bacteria of the human gut microbiota. *Nat. Microbiol.* **4**, 396–403 (2019).
50. Averina, O. V. *et al.* Bacterial metabolites of human gut microbiota correlating with depression. *Int. J. Mol. Sci.* **21**, 1–40 (2020).
51. Kovtun, A. S. *et al.* Alterations of the composition and neurometabolic profile of human gut microbiota in major depressive disorder. *Biomedicines* **10**, 2162 (2022).
52. Hamid, H. A., Ramli, A. N. M. & Yusoff, M. M. Indole alkaloids from plants as potential leads for antidepressant drugs: A mini review. *Front. Pharmacol.* **8**, 145. <https://doi.org/10.3389/fphar.2017.00096> (2017).
53. Noble, E. E. *et al.* Gut microbial taxa elevated by dietary sugar disrupt memory function. *Transl. Psychiatry* **11**, 194 (2021).
54. Cui, Y. *et al.* Roles of intestinal Parabacteroides in human health and diseases. *FEMS Microbiol. Lett.* **369**, 1 (2022).
55. Ezeji, J. C. *et al.* Parabacteroides distansoni: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. *Gut Microbes* **13**, 1 (2021).
56. McGuinness, A. J. *et al.* A systematic review of gut microbiota composition in observational studies of major depressive disorder, bipolar disorder and schizophrenia. *Mol. Psychiatry* **27**, 1920–1935 (2022).
57. Biddle, A., Stewart, L., Blanchard, J. & Leschine, S. Untangling the genetic basis of fibrolytic specialization by lachnospiraceae and rumिनococcaceae in diverse gut communities. *Divers. (Basel)* **5**, 627–640 (2013).
58. Vacca, M. *et al.* The controversial role of human gut lachnospiraceae. *Microorganisms* **8**, 125. <https://doi.org/10.3390/microorganisms8040573> (2020).
59. Bolte, L. A. *et al.* Long-term dietary patterns are associated with pro-inflammatory and anti-inflammatory features of the gut microbiome. *Gut* **70**, 1287–1298 (2021).
60. Zhang, Y., Ding, J. & Liang, J. Associations of dietary vitamin a and beta-carotene intake with depression. A meta-analysis of observational studies. *Front. Nutr.* **2022**, 9 (2022).
61. Caldiroli, A. *et al.* Is there an association between inflammatory/anti-oxidant markers and the presence of psychotic symptoms or severity of illness in mood and psychotic disorders? A multi-centric study on a drug-free sample. *Brain Behav. Immun. Health* **22**, 100453 (2022).
62. Grizotte-Lake, M. *et al.* Commensals suppress intestinal epithelial cell retinoic acid synthesis to regulate interleukin-22 activity and prevent microbial dysbiosis. *Immunity* **49**, 1103–1115.e6 (2018).
63. Song, J. *et al.* Metabolomic signatures and microbial community profiling of depressive rat model induced by adrenocorticotrophic hormone. *J. Transl. Med.* **17**, 224 (2019).
64. Endres, K. Retinoic acid and the gut microbiota in alzheimer's disease: Fighting back-to-back?. *Curr. Alzheimer. Res.* **16**, 405–417 (2019).
65. Boonen, K. *et al.* Beyond genes: Re-identifiability of proteomic data and its implications for personalized medicine. *Genes Basel* **10**, 682 (2019).
66. Liu, Y. *et al.* Systematic metabolic characterization of mental disorders reveals age-related metabolic disturbances as potential risk factors for depression in older adults. *Med. Comm. Beijing* **2022**, 3 (2022).
67. Jiang, Y. *et al.* Identification of sex-specific plasma biomarkers using metabolomics for major depressive disorder in children and adolescents. *Front. Psychiatry* **2022**, 13 (2022).
68. Kluiver, H. *et al.* Metabolomic profiles discriminating anxiety from depression. *Acta Psychiatr. Scand.* **144**, 178–193 (2021).
69. Setoyama, D. *et al.* Plasma metabolites predict severity of depression and suicidal ideation in psychiatric patients—a multicenter pilot analysis. *PLoS One* **11**, e0165267 (2016).
70. Kuwabara, M. *et al.* Different risk for hypertension, diabetes, dyslipidemia, and hyperuricemia according to level of body mass index in japanese and american subjects. *Nutrients* **10**, 1011 (2018).
71. Yudin, R. *et al.* Risk of dyslipidemia in obese young adult subjects as measured by various obesity indices. *J. Endocrinol. Metab.* **12**, 102–106 (2022).
72. Pusparum, M., Ertaylan, G. & Thas, O. Individual reference intervals for personalised interpretation of clinical and metabolomics measurements. *J. Biomed. Inform.* **131**, 104111 (2022).
73. Carbone, E. A., D'Amato, P., Vicchio, G., de Fazio, P. & Segura-Garcia, C. A systematic review on the role of microbiota in the pathogenesis and treatment of eating disorders. *Eur. Psychiatry* **2021**, 64 (2021).
74. Bao, A.-M. & Swaab, D. F. Sex differences in the brain, behavior, and neuropsychiatric disorders. *The Neuroscientist* **16**, 550–565 (2010).
75. Pape, K., Tamouza, R., Leboyer, M. & Zipp, F. Immunoneuropsychiatry—novel perspectives on brain disorders. *Nat. Rev. Neurol.* **15**, 317–328 (2019).
76. Obrenovich, M. E. M. Leaky gut, leaky brain?. *Microorganisms* **2018**, 6 (2018).
77. Ohlsson, L. *et al.* Leaky gut biomarkers in depression and suicidal behaviour. *Acta Psychiatr. Scand.* **139**(2), 185–193 (2019).
78. Maes, M. *et al.* Depression and sickness behavior are Janus-faced responses to shared inflammatory pathways. *BMC Med.* **10**, 66 (2012).
79. Abbasian, G. *et al.* An overview of the TwinsUK cohort's anxiety and depression assessment, using the self-reported Hospital Anxiety and Depression Scale [version 1; peer review: 2 approved with reservations]. *Wellcome Open Res.* **4**, 10 (2019).
80. Glöckner, F. O. The SILVA database project: An ELIXIR core data resource for high-quality ribosomal RNA sequences. *Biodivers. Inf. Sci. Stand.* **2019**, 3 (2019).
81. McMurdie, P. J. & Holmes, S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **2013**, 8 (2013).
82. Barrios, C. *et al.* Circulating metabolic biomarkers of renal function in diabetic and non-diabetic populations. *Sci. Rep.* **2018**, 8 (2018).
83. Valles-Colomer, M. *et al.* The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat. Microbiol.* **4**, 623–632 (2019).
84. Faith, D. P., Minchin, P. R. & Belbin, L. Compositional dissimilarity as a robust measure of ecological distance. *Vegetation* **69**, 57–68 (1987).

85. New, F. N., Baer, B. R., Clark, A. G., Wells, M. T. & Brito, I. L. Collective effects of human genomic variation on microbiome function. *Sci. Rep.* **2022**, 12 (2022).
86. Wemheuer, F. *et al.* Tax4Fun2: Prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. *Env. Microbiomes* **2020**, 15 (2020).
87. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).

Acknowledgements

This work is partially funded by European Commission (DG CNECT) as part of EDITH and by The Research Foundation – Flanders (FWO) as a Junior Project (project ID: FN 701000004). TwinsUK is funded by the Wellcome Trust, Medical Research Council, Versus Arthritis, European Union Horizon 2020, Chronic Disease Research Foundation (CDRF), Zoe Ltd, the National Institute for Health and Care Research (NIHR) Clinical Research Network (CRN) and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

Author contributions

J.D., A.C.R and G.E. conceived this study. P.W. and C.S. have performed the 16S rRNA preprocessing of the TwinsUK data for downstream analysis. J.D. has performed statistical data analyses and prepared figures. A.C.R and G.E. have supervised the data analysis. J.D., A.C.R. and G.E. have written the manuscript. All authors have read and agreed on the manuscript. Corresponding author: Gökhan Ertaylan.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53929-w>.

Correspondence and requests for materials should be addressed to G.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024