

# Microbiota alterations in proline metabolism impact on depression through GABA and ECM homeostasis

**Jordi Mayneris-Perxachs**

Institut d'Investigació Biomèdica de Girona (IdibGi) <https://orcid.org/0000-0003-3788-3815>

**María Arnoriaga-Rodríguez**

Institut d'Investigació Biomèdica de Girona (IdibGi)

**Miquel Martín**

Pompeu Fabra University

**Aurelijus Burokas**

Universitat Pompeu Fabra <https://orcid.org/0000-0002-0364-3496>

**Gerard Blasco**

Institut d'Investigació Biomèdica de Girona (IdibGi)

**Clàudia Coll**

Neuroimmunology and Multiple Sclerosis Unit, Department of Neurology, Dr. Josep Trueta University Hospital

**Anira Escrichs**

Computational Neuroscience Group, Center for Brain and Cognition, Department of Information and Communication Technologies, Universitat Pompeu Fabra

**Carles Biarnés**

Institut d'Investigació Biomèdica de Girona (IdibGi)

**José María Moreno-Navarrete**

Nutrition, Eumetabolism and Health Group, Girona Biomedical Research Institute (IdibGi)

**Josep Puig**

Department of Medical Sciences, School of Medicine, University of Girona

**Josep Garre**

University of Girona

**Rafael Ramos**

University of Girona

**Salvador Pedraza**

Dr. Josep Trueta University Hospital

**Ramón Brugada**

Girona Biomedical Research Institute (IDIBGI), Dr. Josep Trueta University Hospital

**Joan C Vilanova**

Dr. Josep Trueta University Hospital <https://orcid.org/0000-0003-2148-6751>

**Joaquín Serena**

Institut d'Investigació Biomèdica de Girona (IdibGi)

**Jordi Gich**

University of Girona

**Lluís Ramió-Torrentà**

Dr. Josep Trueta University Hospital

**Vicente Pérez-Brocal**

Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO-Public Health)

**Andrés Moya**

Universitat de València <https://orcid.org/0000-0002-2867-1119>

**Reinald Pamplona**

Universitat de Lleida-IRBLLEIDA

**Joaquim Sol**

Lleida Biochemical Research Institute

**Mariona Jové**

Universitat de Lleida-IRBLLEIDA

**Wifredo Ricart**

Nutrition, Eumetabolism and Health Group, Girona Biomedical Research Institute (IdibGi)

**Manuel Portero-Otin**

Universitat de Lleida-IRBLLEIDA

**Gustavo Deco**

Universitat Pompeu Fabra <https://orcid.org/0000-0002-8995-7583>

**Rafael Maldonado**

Pompeu Fabra University <https://orcid.org/0000-0002-4359-8773>

**Jose-Manuel Fernandez-Real (✉ [jmfreal@idibgi.org](mailto:jmfreal@idibgi.org))**

Biomedical Research Networking Center for Physiopathology of Obesity and Nutrition (CIBEROBN), Madrid <https://orcid.org/0000-0002-7442-9323>

---

**Article**

**Keywords:** Microbiota-gut-brain Axis, Multi-omics Approach, Mild Depression, Whole-brain Dynamics, Club Network Disruptions

**Posted Date:** March 1st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-251090/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Microbiota alterations in proline metabolism impact on depression through GABA and ECM homeostasis

Jordi Mayneris-Perxachs<sup>1,2,3\*</sup>, María Arnoriaga-Rodríguez<sup>1,2,3,4</sup>, Miquel Martín<sup>5</sup>, Aurelijus Burokas<sup>5,6</sup>, Gerard Blasco<sup>7,8</sup>, Clàudia Coll<sup>9</sup>, Anira Escrichs<sup>10</sup>, Carles Biarnés<sup>7,8,11</sup>, José Maria Moreno-Navarrete<sup>1,2,3,4</sup>, Josep Puig<sup>4,7,8,11</sup>, Josep Garre-Olmo<sup>4,12,13</sup>, Rafel Ramos<sup>4,14,15</sup>, Salvador Pedraza<sup>4,8,11</sup>, Ramón Brugada<sup>15,16</sup>, Joan Carles Vilanova<sup>11,15</sup>, Joaquín Serena<sup>15,17</sup>, Jordi Gich<sup>4,17</sup>, Lluís Ramió-Torrentà<sup>4,11,17</sup>, Vicente Pérez-Brocal<sup>18,19</sup>, Andrés Moya<sup>18,19,20</sup>, Reinald Pamplona<sup>21</sup>, Joaquim Sol<sup>21,22,23</sup>, Marionna Jové<sup>21</sup>, Wifredo Ricart<sup>1,2,3,4</sup>, Manuel Portero-Otin<sup>21</sup>, Gustavo Deco<sup>10,24,25,26</sup>, Rafael Maldonado<sup>5,27\*</sup>, José Manuel Fernández-Real<sup>1,2,3,4\*</sup>

## Affiliations:

<sup>1</sup>Department of Diabetes, Endocrinology and Nutrition, Dr. Josep Trueta University Hospital, Girona, Spain.

<sup>2</sup>Nutrition, Eumetabolism and Health Group, Girona Biomedical Research Institute (IdibGi), Girona, Spain.

<sup>3</sup>Biomedical Research Networking Center for Physiopathology of Obesity and Nutrition (CIBEROBN), Madrid, Spain.

<sup>4</sup>Department of Medical Sciences, School of Medicine, University of Girona, Girona, Spain.

<sup>5</sup>Laboratory of Neuropharmacology, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain.

<sup>6</sup>Present address: Institute of Biochemistry, Life Sciences Center, Vilnius University, Vilnius, Lithuania.

<sup>7</sup>Institute of Diagnostic Imaging (IDI)-Research Unit (IDIR), Parc Sanitari Pere Virgili, Barcelona, Spain.

<sup>8</sup>Medical Imaging, Girona Biomedical Research Institute (IdibGi), Girona, Spain.

<sup>9</sup>Neuroimmunology and Multiple Sclerosis Unit, Department of Neurology, Dr. Josep Trueta University Hospital, Girona, Spain.

<sup>10</sup>Computational Neuroscience Group, Center for Brain and Cognition, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain.

<sup>11</sup>Department of Radiology (IDI), Dr. Josep Trueta University Hospital, Girona, Spain.

<sup>12</sup>Research Group on Aging, Disability and Health, Girona Biomedical Research Institute (IdibGi), Girona, Spain.

<sup>13</sup>Institut d'Assistència Sanitària, Girona, Spain.

<sup>14</sup>Vascular Health Research Group of Girona (ISV-Girona). Jordi Gol Institute for Primary Care Research (*Institut Universitari per a la Recerca en Atenció Primària Jordi Gol i Gurina* -IDIAPJGol), Girona Biomedical Research Institute (IDIBGI), Dr. Josep Trueta University Hospital, Catalonia, Spain.

<sup>15</sup>Girona Biomedical Research Institute (IDIBGI), Dr. Josep Trueta University Hospital, Catalonia, Spain.

<sup>16</sup>Biomedical Research Networking Center for Cardiovascular Diseases (CIBER).

<sup>17</sup>Girona Neurodegeneration and Neuroinflammation Group, Girona Biomedical Research Institute (IdibGi), Girona, Spain.

<sup>18</sup>Department of Genomics and Health, Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO-Public Health), Valencia, Spain.

<sup>19</sup>Biomedical Research Networking Center for Epidemiology and Public Health (CIBERESP), Madrid, Spain.

<sup>20</sup>Institute for Integrative Systems Biology (I2SysBio), University of Valencia, Spanish National Research Council (CSIC-UVEG), Valencia, Spain.

<sup>21</sup>Metabolic Physiopathology Research Group, Experimental Medicine Department, Lleida University-Lleida Biochemical Research Institute (UdL-IRBLleida), Lleida, Spain.

<sup>22</sup>Institut Català de la Salut, Atenció Primària, Lleida, Spain.

<sup>23</sup>Research Support Unit Lleida, Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Lleida, Spain.

<sup>24</sup>Institució Catalana de la Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

<sup>25</sup>Department of Neuropsychology, Max Planck Institute for human Cognitive and Brain Sciences, Leipzig, Germany.

<sup>26</sup>Turner Institute for Brain and Mental Health, Monash University, Melbourne, Victoria, Australia.

<sup>27</sup>Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain.

*Corresponding authors:*

\* Jordi Mayneris-Perxachs, M.D., PhD. e-mail: [jmayneris@idibgi.org](mailto:jmayneris@idibgi.org)  
Department of Diabetes, Endocrinology and Nutrition, Dr. Josep Trueta University Hospital.  
Nutrition, Eumetabolism and Health Group, Girona Biomedical Research Institute (IdibGi).  
Avinguda de França s/n, 17007, Girona, Spain.  
Phone: +34 972940200 / Fax: +34 97294027.

\* Rafael Maldonado, PhD., e-mail: [rafael.maldonado@upf.edu](mailto:rafael.maldonado@upf.edu)  
Laboratory of Neuropharmacology, Department of Experimental and Health Sciences,  
Universitat Pompeu Fabra, Barcelona.  
Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain.  
Carrer del Doctor Aiguader, 88, 08003 Barcelona, Spain.  
Phone: +34 933160866.

\* José-Manuel Fernández-Real, M.D., PhD. e-mail: [jmfreal@idibgi.org](mailto:jmfreal@idibgi.org)  
Department of Diabetes, Endocrinology and Nutrition, Dr. Josep Trueta University Hospital.  
Department of Medical Sciences, Faculty of Medicine, Girona University, Girona, Spain.  
Nutrition, Eumetabolism and Health Group, Girona Biomedical Research Institute (IdibGi).  
Avinguda de França s/n, 17007, Girona, Spain.  
Phone: +34 972940200 / Fax: +34 97294027.

## **Abstract**

The microbiota-gut-brain axis has emerged as a novel target in depression, a disorder with low treatment efficacy. However, the field is dominated by underpowered studies focusing on major depression not addressing microbiome functionality, compositional nature, or confounding factors. We applied a multi-omics approach combining pre-clinical models with three human cohorts including mild-depressed patients. Microbial functions and metabolites converging into glutamate/GABA metabolism, particularly proline, were linked to depression. Whole-brain dynamics revealed rich club network disruptions associated with depression and circulating proline. Proline supplementation in mice exacerbated depression along with microbial translocation. Human microbiota transplantation induced an emotional-impaired phenotype in mice and alterations in GABA-, proline-, and extracellular matrix-related pre-frontal cortex genes. Targeting the microbiome and dietary proline may open new windows for an efficient depression treatment.

## Main

Depression affects more than 300 million people worldwide and is well known to constitute one of the main causes of disability.<sup>1,2</sup> Despite this, the underlying mechanisms of depression still remains a crucial unresolved research topic, which is evidenced by the lack of an appropriate treatment, with an overall efficacy below 50% and a relapse rate of 40% in responders.<sup>3,4</sup> In addition, benefits become clinically relevant only in the small minority of patient populations with severe major depression. Therefore, there is an urgent need for new insights into the pathophysiology of depression.

A promising novel area of investigation involves the microbiota-gut-brain axis,<sup>5</sup> which has recently shown to control cognitive function<sup>6</sup> and inhibitory behavior (Gut 2021, *in press*). The microbiome influences the gut-brain communication through neural, endocrine, immune and neuroactive pathways. The latter includes microbial-derived neurotransmitters (*e.g.*, GABA, catecholamines) and metabolites (SCFA, bile acids), and brain-derived neurotrophic factors (*e.g.*, BDNF).<sup>7</sup> However, a recent systemic review that identified 19 studies analyzing the gut microbiota of clinical groups with depression revealed large inconsistent findings.<sup>8</sup> Remarkably, this body of research is dominated by underpowered cross-sectional case-control studies comparing healthy controls to patients with major depressive disorders (MDD) focusing solely on taxonomic composition based on 16S rRNA sequencing. Therefore, it is mandatory to develop powerful longitudinal studies under controlled conditions to clarify the specific role of microbiota on depression. Subjects with mild depression should also be investigated in order to evaluate all the spectrum of depressive symptomatology and potentially provide preventive measures. In addition, different microbial signatures may result in the same phenotype due to functional redundancy. Going beyond taxonomic composition to include analyses of microbial functionality, such as metabolomics and shotgun metagenomics sequencing, is thus vital to accurately capture the host-microbiome interplay. Finally, with just two exceptions, all these studies did not consider all essential confounding variables known to influence depression or the gut microbiota composition, particularly diet<sup>9</sup> and psychotropic medication; and only one study applied

appropriate statistical methods to take into account the compositional nature of the microbiome datasets, a key point to obtain reliable results.

## Results

### Depression scores are associated with a specific microbial ecosystem

To overcome all these issues, we applied a unique integrative longitudinal, multi-cohort and multi-omics approach to reveal molecular mechanisms underlying the microbiome-gut-brain axis interplay in depression combining fecal shotgun metagenomics, plasma and fecal metabolomics, and whole-brain dynamic functional magnetic resonance imaging (fMRI) in three human cohorts with complementary mice experiments. We first assessed the relationships of bacterial composition and functionality with depression, diagnosed using the Patient Health Questionnaire 9 (PHQ-9), in a longitudinal discovery cohort (IRONMET,  $n = 116$ , **Table S1**) comprising non-depressed ( $n = 44$ , PHQ-9: 0-4), mild depressed ( $n = 47$ , PHQ-9: 5-9), and major depressed subjects ( $n = 25$ , PHQ-9 >10).<sup>10</sup>

With few exceptions,<sup>11,12</sup> previous studies analyzing the associations between the gut microbiome and depression have not used appropriate compositional data analysis, thereby rendering potential misleading conclusions. In order to take into account the compositional nature of the microbiome data<sup>13</sup> we modelled read counts using a Dirichlet distribution to deal with 0 count values and then applied a centered log-ratio (clr) transformation as implemented in the ALDEx2 R package.<sup>14</sup> Previous studies have reported inconsistent alpha diversity findings, with a limited number reporting lower indices in subject with MDD compared to controls.<sup>8</sup> In line with this, we did not observe significant differences in the alpha diversity measures such as species richness or the Shannon index among groups (**Fig. 1a,b**). However, non-depressed individuals had higher Fisher's alpha diversity indices than depressed subjects, while no differences were found between mild and major depression (**Fig. 1c**). The compositional alternative to the principal coordinate analysis plots of  $\beta$ -diversity is the principal component analysis (PCA). Therefore, we applied a PCA analysis to the clr-transformed data to reveal global variance patterns in the microbial

profiles and identify outliers. This initial unsupervised exploratory analysis revealed significant differences in the microbiome composition between non-depressed and depressed subjects (**Fig. 1b**).

We next sought to identify those bacterial species contributing the alterations in the gut microbiome associated with depression. Importantly, the vast majority of previous studies were based on 16S rRNA gene sequencing and, consequently, they do not provide enough taxonomic resolution to report results at the species level. For each taxa, we fitted a robust linear regression model between the PHQ-9 scores and the clr-transformed data controlling for age, gender, BMI, education years, and antidepressant and anxiety medication. Using this approach, we were able to identify thirty bacterial species significantly associated with depression ( $p_{adj} < 0.1$ ) (**Fig. 1d, Table S2**). Patients with higher PHQ-9 scores had higher levels of *Parabacteroides spp.* and *Acidaminococcus spp.* but lower levels of *Bifidobacterium pseudolongum* and species from the butyrate-producing *Lachnospiraceae family*, including *Roseburia spp.* All previous case-control studies were also cross-sectional in nature. Longitudinal research is required to elucidate causality in the associations. When we evaluated baseline bacterial taxa predictive of the PHQ-9 score one-year later, we found that lower levels of several *Bifidobacterium spp.* and *Lachnospiraceae* species and higher levels of *Prevotella* and *Enterobacter* species were associated with higher PHQ-9 scores one year later (**Fig. 1e, Table S3**).

### **Microbial functions involved in glutamate family of amino acids metabolism/transport and neurotransmitter transporters are linked to depression**

Due to functional redundancy, significantly different microbial compositions may produce the same functional outcome.<sup>15</sup> Different phylotypes can cover identical functions and produce the same metabolic outputs under the same environmental conditions. It is therefore imperative to include functional analyses to disentangle the complex host-microbiome interaction in health and disease. Unlike most previous studies based on 16S rRNA sequencing, the use of a shotgun metagenomics sequencing approach allowed us to study the functional role of the microbiome in depression. Reads originating from microbial genes were mapped to the Kyoto Encyclopedia of

Genes and Genomes (KEGG) orthologs. The resulting KEGG counts were then used to identify differentially abundant microbial gene functions in the metagenome associated with depression. Fitted generalized linear models to the clr-transformed KEGG counts partialling out the effects of age, gender, BMI, education years, and antidepressants and anxiolytics medication, revealed several pathways (KEGG Orthology level 3) associated with the PHQ-9 score ( $p_{adj} < 0.1$ ) (**Fig. 1f**). Of note, bacterial pathway involving arginine, proline and histidine metabolism were negatively associated with depression. Along with glutamine, these amino acids comprise the glutamate family of amino acids, as their catabolism converges into glutamate, which serves as a carbon source for fueling the TCA cycle and subsequent gluconeogenic reactions as well as the synthesis of  $\gamma$ -aminobutyric acid (GABA). Consistently, we also found significant associations of bacterial glutamate metabolism, glutamatergic synapse and GABAergic synapse with the host PHQ-9 scores. Deeper analysis of the functional terms (KEGG Orthology level 4) identified 327 out of 5714 bacterial functions significantly associated with depression scores (**Table S4**). Notably, some of the bacterial functions with strongest associations were involved in TCA cycle; arginine, proline, histidine, and glutamate metabolism; as well as the transport of these amino acids (**Fig. 1g**). Remarkably, the solute carrier family 6 (SLC6) also had a negative association with PHQ-9 scores. SLC6 genes mainly encode transporters for neurotransmitters (*e.g.*, GABA, monoamines, glycine and proline), but also proteinogenic amino acids, betaine, taurine and creatine,<sup>16</sup> which is consistent with alterations in the GABAergic synapse.

### **Metabolites from the histidine, arginine and proline degradation pathways converging into glutamate and GABA shunt are associated with depression**

To further explore the microbiome functionally, we next performed a metabolic profiling (<sup>1</sup>H-NMR and HPLC-ESI-MS/MS) of plasma and fecal samples. Unlike shotgun metagenomics, which only provides information about microbial genes and thus predicts the microbiome potential functionality, metabolic phenotyping provides a readout of the actual microbiota functional activity.<sup>15</sup> Applying a machine learning variable selection strategy based on multiple random-forest,<sup>17</sup> we were able to identify several metabolites linked to the PHQ-9 scores (**Fig.**

**2a-f).** Consistent with our previous findings, several of these metabolites were involved in the TCA cycle (succinate, fumarate), histidine metabolism (urocanate) and proline and glutamate metabolism. We then aimed to validate these findings by performing an HPLC-MS/MS-based metabolic profiling of plasma samples from in an independent large validation cohort (IMAGEOMICS,  $n = 919$ , **Table S5**) consisting also of non-depressed, mild and major depressed subjects. The most consistent finding was the strong positive association of circulating proline with the depression scores (**Fig. 2g**). Other consistent findings implicated alterations in retinol metabolism (**Fig. 2g**) and TCA cycle intermediates (citric acid, **Fig. 2h**). To facilitate analysis and interpretation of these metabolomic results, we performed a pathway over-representation analysis mapping metabolites that were significantly associated with the PHQ-9 in the discovery cohort to the KEGG, Reactome, INOH and HumanCyc databases included in the ConsensusPathDB.<sup>18</sup> This enrichment analysis further highlighted a significant ( $qval < 0.1$ ) over-representation of pathways associated with the TCA cycle and oxidative phosphorylation, glutamate metabolism, and arginine, proline and histidine catabolism (**Fig. 2i**). In agreement with metagenomic findings, it also stressed a significant over-representation of SLC and amino acid transporters as well as the GABA synthesis/degradation pathway and its link with the TCA cycle through the GABA shunt, which is predominantly associated with neurotransmission in the mammalian brain. A summary of the main metabolites and bacterial functions associated with depression is shown in **Fig. 2j**.

### **High proline consumption is associated with increased depression scores and small intestine genes participating in glutamatergic and GABAergic synapse and extracellular matrix homeostasis**

Diet has a strong impact in modulating the composition and metabolic activity of the gut microbiome.<sup>19</sup> Therefore, we next evaluated the associations of macronutrients, vitamins, minerals, amino acids and fatty acids derived from food frequency questionnaires with the PHQ-9 scores in the IRONMET cohort. Strikingly, partial spearman's rank correlation analysis revealed proline as the dietary factor with the strongest impact on depression (**Fig. 3a**). When subjects were categorized according the median levels of plasma proline and dietary proline

consumption, those with both low circulating and dietary proline levels had the lowest PHQ-9 scores, while individuals with both high dietary proline consumption and circulating proline levels had the highest depression scores (**Fig. 3b**). As dietary proline is extensively metabolized by enterocytes in the small intestine (**Fig. 3c**), we performed an RNA sequencing of jejunum samples from a second independent cohort (INTESTINE cohort,  $n = 28$ , **Table S6**) to identify those transcripts associated with proline consumption. Differential gene expression analysis was performed following TMM normalization using the limma pipeline with the *voom* transformation, linear modelling and empirical Bayes moderation.<sup>20</sup> We identified 1,547 out of 15,144 significant gene transcripts associated with dietary proline (**Fig. 3d and Table S7**). To facilitate functional interpretation of differentially expressed genes, we performed over-representation analyses mapping those genes to both Reactome (**Fig. 3 e,f and Table S8**) and KEGG (**Fig. S1a,b and Table S9**) pathways. As pathway information is inherently redundant, with genes often participating in multiple pathways, we collapsed redundant pathways into a single biological theme using EnrichmentMap<sup>21</sup> to overcome redundancy and further simplify interpretation (**Fig. 3f and Fig. S1b**). Notably, Reactome-based analyses identified pathways involved in GABA receptor activation, synaptic interactions and axon guidance. It also revealed pathways participating in extracellular matrix (ECM), muscle contraction, MAPK signaling and GPCR signaling. In line with these results, KEGG-based analyses highlighted several pathways involved in neuron synapse (**Fig. S1a,b**), in particular GABAergic (**Fig. 3g**) and glutamatergic synapse (**Fig. S1c**), which is consistent with metabolomics results. Our analyses in the jejunum also identified pathways participating in ECM and muscle contraction. We also mapped significant transcripts associated with dietary proline to the DisGeNET disease-based database, which contains one of the largest collections of genes associated with human diseases, using both expert curated and text mining data. Remarkably, enrichment analysis highlighted a significant over-representation of diseases associated with cognitive, neurodegenerative and CNS disorders, including schizophrenia and major depressive disorders (**Fig. S2a-c and Table S10**).

### **Proline supplementation exacerbates depressive-like behavior in mice in association with microbial translocation**

To further investigate the direct effect of dietary proline on depression, we performed a supplementation study in mice (**Fig. 3h**). A total of  $n=40$  C57BL/6J mice were fed either a standard diet (SD) or a crude fiber rich diet (CFD) and supplemented with either water or proline (36 g/L). Throughout the experiment, mice were chronically exposed to unpredictable mild stressors (UCMS model), an animal model with high face, constant and predictive validity, to develop and evaluate depressive-like behavior. After 6 weeks of UCMS exposure, mice supplemented with SD + proline had higher immobility times in the forced swim test (**Fig. 3i**), a well validated model of the despair behavior, and reduced sucrose intake (**Fig. 3j**), a behavioral model of anhedonia, a core symptom of depression, compared to stress-free control mice that only received water. Additionally, compared to control animals, SD + proline mice had higher levels of lipopolysaccharide binding protein (LBP) (**Fig. 3k**), which is indicative of microbial translocation and in agreement with our microbiome results. Interestingly, these behavioral and biochemical effects of proline did not appear in mice fed a CFD, suggesting a potential protective effect of this diet.

### **Brain iron deposition and whole-brain functional dynamics reveal rich-club network disruptions associated to both depression and circulating proline**

Metagenomics, metabolomics and RNA-seq analyses were consistent in identifying alterations in glutamatergic and GABAergic systems. Since iron plays a crucial role in glutamate and GABA homeostasis, we assessed iron deposition in the brain of the IRONMET patients using magnetic resonance imaging (MRI) based on T2\* and R2\* relaxometry. In fact, iron overload in certain brain areas has been implicated in neurodegenerative disorders<sup>22</sup> and alterations in anxiety-like behaviour and mood,<sup>23</sup> but only few studies have characterized the impact of iron on depression. Partial spearman's rank correlation analysis among relaxometry parameters in brain regions (AAL Atlas) and PHQ-9 scores highlighted low iron deposition (mean T2\* values) of regions along the cingulum and frontal lobe positively associated with depression (**Fig. 4a**). These results

were validated using a machine learning variable selection strategy (**Fig. 4b**). Moreover, the human brain is a complex network of structurally and functionally connected regions. Functional communication between these regions is thought to play a vital role in complex processes such as depression. However, most studies have focused on static functional connectivity, but functional connectivity among brain networks is not static over time. Studying the dynamics of resting-state brain activity across the whole-brain functional network might facilitate interpretation of brain functioning and provide better insights into the pathophysiology of disease.<sup>24</sup> Therefore, we analyzed resting-state fMRI data (**Fig. 4c**) in a subset of patients from the IMAGEOMICS cohort ( $n=591$ ) and studied whole-brain functional dynamics applying a novel intrinsic-ignition framework across the whole-brain functional network (214 brain areas),<sup>25</sup> to assess the effect of spontaneous local activation events on local-global integration (**Fig. 4d**), followed by machine learning algorithms to identify those nodes predictive of the PHQ-9 scores. Using the Boruta algorithm we identified 72 intrinsic-ignition nodes associated with depression (**Fig. 4e**). Similarly, it revealed 68 nodes linked to the circulating proline levels (**Fig. 4f**), 30 of which were also associated with the PHQ-9 scores (**Fig. 4g and Table S11**). Notably, these shared intrinsic-ignition areas, including the superior frontal cortex, the precuneus, insula and subcortical areas such as the caudate, putamen, and hippocampus, mostly belong to the so-called “rich-club” (**Fig. 4h**), a set of high-degree nodes that tend to be more closely connected among themselves than with peripheral regions, *i.e.* lower degree nodes.<sup>26,27</sup>

### **An emotional impairment is transferred to mice through the gut microbiota. Direct impact on the expression of mPFC genes participating in proline and GABA transport and ECM and collagen metabolism in recipient mice**

Finally, to evaluate a potential causal role of the microbiota in the development of emotional disorders, we transplanted microbiota from 20 human donors with different PHQ-9 scores into 20 antibiotic-treated mice (**Fig. 5a**). We then assessed whether an emotional impaired phenotype emerged in mice receiving microbiota from donors with higher depression scores using a fear conditioning induced freezing test, a well-recognized model of maladaptive response to stress. Notably, donor’s PHQ-9 scores were significantly correlated with the freezing time in recipient

mice (**Fig. 5b**). One of the most consistent findings in MDD include decreased frontal lobe function, mainly involving the medial prefrontal cortex (mPFC).<sup>28</sup> In addition, results from a recent meta-analysis suggested that decreased levels of glutamatergic metabolites in the mPFC are linked with the pathophysiology of depression.<sup>29</sup> Considering our findings highlighting the impact of the microbiome on glutamatergic and GABAergic systems, we performed an RNA-seq of the mPFC of recipient mice. We identified 59 out of 15,537 gene transcripts in recipient's mice mPFC significantly associated with donor's PHQ-9 scores (**Fig 5.c and Table S12**). To gain a better insight into the potential mechanisms underlying the microbial effects on depression, we built gene-gene interaction networks using the STRING database<sup>30</sup> (**Fig. 5d,e**). Furthermore, we mapped significant genes to KEGG (**Fig. 5f**) and Reactome (**Fig. 5g**) databases. Half of the transcripts negatively associated with the donor's PHQ-9 score clustered together and were involved in oxidative phosphorylation and neurodegenerative disease. Notably, we identified a cluster comprising gene transcripts encoding for transporters of GABA (*slc6a12* and *slc6a13*) and Proline (*slc6a20*) (**Fig. 5h**), which is in agreement with our functional analyses. Consequently, we found an over-representation of Reactome pathways associated with GABA neurotransmission, which is also in consonance with the results from the jejunal RNA-seq analysis. These results are in line with metagenomics and metabolomics and highlight again the importance of the GABA shunt interconnecting the TCA cycle with GABA and glutamate metabolism. Not only that, but we also found another cluster of gene transcripts participating in ECM and collagen homeostasis and muscle contraction. Importantly, proline is essential for collagen biosynthesis, constituting 10% of its amino acid content. Therefore, these results validate and further highlight the importance of the proline-glutamate-GABA-microbiome interplay identified in our metabolomics and metagenomics analyses.

## Discussion

The gut microbiota has emerged as a novel actor in the pathophysiology of depression. However, recent meta-analyses have revealed a strong inconsistency among studies in terms of gut microbiome signatures associated with depression.<sup>8</sup> These inconsistencies mostly arise from underpowered studies, methodological heterogeneity, varying depression diagnostic criteria, lack of consideration of confounding variables, inappropriate statistical analysis, and functional redundancy of the microbiota. Furthermore, a limitation of the vast majority previous studies includes the use of 16S-RNA sequencing, which does not have enough taxonomic resolution to report results at the species levels does not provide information about microbial functionality. Here, we tackled these limitations by applying a multi-omics approach in three human cohorts and pre-clinical studies.

In concordance with some previous findings,<sup>8</sup> we found decreased levels of SCFA-producing bacteria such as species from the *Lachnospiraceae* family (including *Roseburia spp.*) and *Bifidobacterium spp.*<sup>31</sup> Notably, *Bifidobacterium* strains are amongst the most efficient GABA-producers.<sup>32</sup> The current functional metabolomics and metagenomics analyses highlighted several bacterial functions and metabolites involved in proline, histidine and arginine pathways that converge into glutamate and GABA metabolism linked to depression scores. In line with this, we found that patients with higher PHQ-9 scores had higher levels of *Acidaminococcus spp.*, which use glutamate as the only carbon source and have shown to grow only in culture media containing arginine, glutamate or histidine.<sup>33</sup> Dysregulation of glutamate and GABA neurotransmission and increased circulating levels of glutamate and GABA have been reported in participants with MDD.<sup>34-36</sup> Therefore, the current findings point towards a potential role of the microbiome in depression through glutamate/GABA metabolism, compatible with the glutamate hypothesis of depression.<sup>36</sup> So far, only two studies have performed functional metagenomics analysis.<sup>11,12</sup> Although not significant, a recent study performing a targeted analysis pointed towards alterations in GABA shunt and glutamate degradation pathways in subjects with depression<sup>11</sup>. Similarly, in

a small study ( $n=40$ ) the GABA degradation pathway was prominent in the microbiome of individuals with MDD.<sup>12</sup>

The most consistent finding was the positive association of circulating proline with the depression scores. In a recent meta-analysis of peripheral blood metabolites in major depressive disorders, a subgroup analysis revealed that antidepressant-free MDD patients had higher levels of L-proline.<sup>37</sup> It is worth noting that our analyses were controlled for antidepressant medication. Hyperprolinemia has also been linked to epilepsy, schizophrenia, seizures and impaired cognitive function. Importantly, the proline degradative pathway can eventually generate glutamate and GABA. Hence, proline accumulation has shown to disrupt GABA production, glutamate release and impair synaptic transmission,<sup>38,39</sup> while inactivation of proline transporter altered glutamatergic synapse and perturbed behaviours in mice.<sup>40</sup> In line with previous findings, we found that dietary proline was strongly associated with several pathways involving GABAergic and glutamatergic synapse. We also demonstrated a potential causal role of this amino acid in depression by supplementing mice with proline. Notably, circulating proline levels were consistently associated with nodes of intrinsic brain networks linked to depression. Alterations in the circulating proline levels were also linked to depression cores through the so-called “rich-club” of highly interconnected nodes. The rich-club of the brain network plays a critical role in global integration of neural information and is essential for efficient communication across multiple segregated and distant brain regions.<sup>26,27</sup> Therefore, network disruptions within the rich club have a direct impact on various behavioral and cognitive tasks. Connectivity-based neuroimaging studies have identified alterations in rich-club network in several neuropsychiatric disorders.<sup>26</sup> Recently, disruptions in the rich-club network organization has also been implicated in MDD ( $n=32$ ).<sup>41</sup> Here, we extend these findings to mild-moderate depression using a much larger cohort ( $n=591$ ).

Dietary proline was also strongly linked to the ECM homeostasis not only in the human intestine but also in the mouse brain. Collagen makes up 80% of ECM, with proline and hydroxyproline accounting for 25% of collagen amino acid content. Therefore, proline is stored in the ECM as

collagen, which can serve both as a reservoir or dump for proline.<sup>42</sup> Of note, components of the ECM such as matrix metalloproteinases, which degrade collagen and release proline, and perineuronal nets, have been recently suggested as new key players in the development of psychiatric disorders.<sup>43</sup> Strikingly, microbiota transfer from humans to mice highlighted a strong up-regulation of pathways associated with the ECM and collagen homeostasis as well as GABA and proline-dependent neurotransmission in the mPFC of mice that received microbiota from human donors with high depression scores. These results further highlight a potential causative role of the microbiome in depression by disrupting of glutamatergic and GABAergic homeostasis through the alteration of proline metabolism. All classical depression treatments target directly the brain with limited effectiveness. Understanding how the microbiome links proline to glutamate and GABA metabolism is thus crucial to design new effective therapies for the treatment of depression. We propose that diets with reduced proline content may have a strong impact in ameliorating depressive symptomatology.

## Methods

**IRONMET cohort (n=116).** This is a cross-sectional case-control study setting at the Endocrinology Department of Dr. Josep Trueta University Hospital. The recruitment of subjects started in January 2016 and finished in October 2017. Consecutive middle-aged subjects, 27.2–66.6 years, were included. Patients with obesity (body mass index (BMI)  $\geq 30$  kg/m<sup>2</sup>) and age-matched and sex-matched subjects without obesity (BMI 18.5– $<30$  kg/m<sup>2</sup>) were eligible. Exclusion criteria were type 2 diabetes mellitus, chronic inflammatory systemic diseases, acute or chronic infections in the previous month; use of antibiotic, antifungal, antiviral or treatment with proton pump inhibitors; severe disorders of eating behaviour or major psychiatric antecedents; neurological diseases, history of trauma or injured brain, language disorders and excessive alcohol intake ( $\geq 40$  g OH/day in women or 80 g OH/day in men). The Institutional review board - Ethics Committee and the Committee for Clinical Research (CEIC) of Dr. Josep Trueta University Hospital (Girona, Spain) approved the study protocol and informed written consent was obtained from all participants.

**IRONMET longitudinal cohort (n=70).** after 1-year of follow-up individuals were re-evaluated and cognitive function, brain structure and fecal and plasma samples were collected.

**IMAGEOMICS cohort (n=919).** The Ageing Imageomics Study is an observational study including participants from two independent cohort studies (MESGI50 and MARK). Detailed description of the cohorts can be found elsewhere.<sup>27</sup> Briefly, the MESGI50 cohort included a population aged  $\geq 50$  years, while the MARK cohort included a random sample of patients aged 35–74 years with intermediate cardiovascular risk. Eligibility criteria included age  $\geq 50$  years, dwelling in the community, no history of infection during the last 15 days, no contraindications for MRI and consent to be informed of potential incidental findings.

**INTESTINE cohort (n=28).** This is a pilot project that included morbidly obese (BMI  $>35$  kg/m<sup>2</sup>) subjects recruited at the Endocrinology Department of Dr. Josep Trueta University Hospital. All subjects were of Caucasian origin and reported a body weight stable for at least three

months before the study. Subjects were studied in the post-absorptive state. The following exclusion criteria were considered: i) no systemic disease other than obesity; ii) free of any infections in the previous month before the study; iii) no liver diseases (specifically tumour disease and infections) and thyroid dysfunction, which will be specifically excluded by biochemical work-up. This protocol was revised, validated and approved by the Ethics committee of the Hospital Dr Josep Trueta. The purpose of the study was explained to participants and they signed written informed consent before being enrolled in the study.

### **Clinical and laboratory parameters**

Completed medical history and anthropometric variables were collected from all participants. In fasting conditions, a blood sample was provided. Fasting plasma glucose (FPG) and lipid profiles were measured by standard laboratory methods using an analyzer (CobasR 8000 c702, Roche Diagnostics, Basel, Switzerland). High-sensitivity C-reactive protein (hsCRP) levels were determined by immunoturbidimetric method (CobasR 8000 c702, Roche Diagnostics, Basel, Switzerland). Glycated hemoglobin (HbA1c) was determined by high performance liquid chromatography (ADAMRA1c HA-8180V, ARKRAY, Inc., Kyoto, Japan).

### **Body composition**

Fat mass (FM), fat free mass (FFM) and their distribution was measured by a dual energy xray absorptiometry (DEXA, GE lunar, Madison, Wisconsin).

### **Dietary pattern**

The dietary characteristics of the subjects were collected in a personal interview using a validated food- frequency questionnaire.<sup>44</sup>

### **Cognitive assessment**

The Patient Health Questionnaire-9 (PHQ-9) is the depression module of the PRIME-MD diagnostic instrument for mental disorders.<sup>45</sup> It is a self-administered questionnaire and consists of 9 items of depression symptoms plus a question about functional impairment. The PHQ-9 can be scored either as a depression severity rating (range 0-27 points) or with an algorithm based on

the DSM-IV criteria (major and minor episode). It can also be interpreted using a cut-off point applied the symptom severity score. Scores of 10 to 14 represent a moderate symptom severity level, 15 to 19 represent moderately severe symptoms, and 20 to 27 severe depressive symptoms. Scores of 10 or more have an 88% sensitivity and specificity.

### **MRI acquisition and image pre-processing**

All subjects were studied on a 1.5T Ingenia (Philips Healthcare, Best, The Netherlands) with eight channel head coils. As a part of a larger study protocol a multislice fluid attenuation inversion recovery (T2-FLAIR) with TR/TE/TI=6500/120/2200ms, flip angle 90°, in-plane resolution 0.78x0.78mm, slice thickness 5mm without gap and 20 axial slices was used to exclude pre-existing brain lesions. MRI relaxometry was assessed by using a multi-echo gradient echo sequence with TR/1stTE/ $\Delta$ TE=800/2.2/5ms, flip angle 80°, in-plane resolution 2x2mm, slice thickness 5mm without gap and 20 axial slices. After acquisition, T2\* and R2\* maps were computed using Olea Sphere 3.0 (Olea Medical, La Ciotat, France) with Bayesian analysis algorithm. T2\* maps were calculated by fitting the signal decay curve of the respective magnitude multiecho data and R2\* maps were calculated as  $R2^* = 1/T2^*$ . In addition, a brain extraction tool was used to delete all non-brain tissues of calculated T2\* and R2\* maps. R2\* were measured in  $s^{-1}$ .

### **Intrinsic-Ignition Framework**

We applied the Intrinsic-Ignition Framework<sup>25</sup> to obtain the effect of naturally occurring activation events that reflect the capability of a given brain area to propagate activity to other brain areas. In brief, we transformed the BOLD time series to phase space by filtering the signals within the narrowband (0.04-0.07 Hz) and computed the Hilbert transform to obtain the phases of the signal between each pair of brain areas at each time point. Fig. 4c shows the representation of the Hilbert BOLD phase for a brain area over time in the complex plane. Fig. 4d shows the algorithm used to obtain the ignition value of each brain area evoked by an event within a set time window. The binary events were defined by transforming the time series into

z-scores,  $z_i(t)$ , and fixing a threshold,  $\theta$ . Then, a phase lock matrix  $P_{jk}(t)$ , was computed which describes the state of phase synchronization between brain areas  $j$  and  $k$  at time  $t$  as:

$$P_{jk}(t) = e^{-3|\varphi_j(t) - \varphi_k(t)|}$$

where  $\varphi_j(t)$  and  $\varphi_k(t)$  correspond to the phases of the BOLD time series for brain areas  $j$  and  $k$  at time  $t$ . Then, the integration was defined by measuring the length of largest connected component in the phase lock matrix  $P_{jk}(t)$  and the value of integration was computed as the length of the connected component considered as an adjacent graph (*i.e.*, the largest subcomponent). Finally, for each brain area, we averaged across the events the integration evoked at each time  $t$  within the set time window. A complete description of the method can be consulted in <sup>25</sup>.

### **Extraction of faecal genomic DNA and whole-genome shotgun sequencing**

Total DNA was extracted from frozen human stools using the QIAamp DNA mini stool kit (Qiagen, Courtaboeuf, France). Quantification of DNA was performed with a Qubit 3.0 fluorometer (Thermo Fisher Scientific, Carlsbad, CA, USA), and 1 ng of each sample (0.2 ng/ $\mu$ l) was used for shot gun library preparation for high-throughput sequencing, using the Nextera DNA Flex Library Prep kit (Illumina, Inc., San Diego, CA, USA) according to the manufacturers' protocol.

Sequencing was carried out on a NextSeq 500 sequencing system (Illumina) with 2 X 150-bp paired-end chemistry, at the facilities of the Sequencing and Bioinformatic Service of the FISABIO (Valencia, Spain). The obtained input fastq files were decompressed, filtered and 3·ends-trimmed by quality, using prinseq-lite-0.20.4 program<sup>46</sup> and overlapping pairs were joined using FLASH-1.2.11.<sup>47</sup> Fastq files were then converted into fast files, and human and mouse host reads were removed by mapping the reads against the GRCh38.p11, reference human genome (Dec 2013), and GRCm38.p6, reference mouse genome (Sept 2017), respectively, by using bowtie2-2.3.4.3 <sup>48</sup> with end-to-end and very sensitive options. Next, functional analyses were

carried out by assembling the non-host reads into contigs by MEGAHIT v1.1.2<sup>49</sup> and mapping those reads against the contigs with bowtie2. Reads that did not assemble were appended to the contigs. Next, the program Prodigal v2.6.342<sup>50</sup> was used for predicting coding regions. Functional annotation was carried out with HMMER<sup>51</sup> against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, version 2016<sup>52</sup> to obtain the functional subcategory, route and annotation of the genes. The filtering of the best annotations and the assignment of the orf annotation to every read were carried out using the statistical package R 3.1.0<sup>53</sup> which also was used to count the aligned reads and to add the category and its coverage, and finally to build abundance matrices. Taxonomic annotation, was implemented with Kaiju v1.6.2<sup>54</sup> on the human and mouse-free reads. Addition of lineage information was added, counting of taxa and generation of an abundance matrix for all samples were performed using the package R.

### **Jejunal stranded RNA sequencing**

Intestinal epithelium from jejunum was collected during gastric by-pass surgery in RNAlater (Thermo Fisher Scientific), to preserve RNA integrity. Then, samples were immediately transported to the laboratory. The handling of tissue was carried out under strictly aseptic conditions and stored at -80°C. RNA purification was performed using RNeasy-Tissue Mini-Kit (QIAGEN). Total RNA was quantified by Qubit® RNA BR Assay kit (Thermo Fisher Scientific) and the integrity was checked by using the RNA Kit (15NT) on 5300 Fragment Analyzer System (Agilent). The RNASeq libraries were prepared with Illumina® TruSeq® Stranded Total RNA Sample Preparation kit following the manufacturer's recommendations with some modifications. Briefly, in function of availability 100-500ng of total RNA was rRNA depleted using the RiboZero Magnetic Gold Kit and fragmented by divalent cations. The strand specificity was achieved during the second strand synthesis performed in the presence of dUTP. The cDNA was adenylated and ligated to Illumina platform compatible IDT adaptors with unique dual indexes with unique molecular identifiers (Integrated DNA Technologies), for paired end sequencing. The ligation products were enriched with 15 PCR cycles and the final library was validated on an Agilent 2100 Bioanalyzer with the DNA 7500 assay (Agilent). The libraries were sequenced on

NovaSeq 6000 (Illumina) in a fraction of sequencing flow cell with a read length of 2x101bp following the manufacturer's protocol for dual indexing. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA v3.4.4) and followed by generation of FASTQ sequence files. RNA-seq reads were mapped against human reference genome (GRCh38) using STAR software version 2.5.3a<sup>55</sup> with ENCODE parameters. Genes were quantified using RSEM version 1.3.0<sup>56</sup> with default parameters and using the annotation file from GENCODE version 29.

### **Metabolomics analyses**

*HPCL-ESI-MS/MS metabolomics analyses (IRONMET, IMAGEOMICS).* For non-targeted metabolomics analysis, metabolites were extracted from plasma and faecal samples with methanol (containing phenylalanine-C13 as an internal standard) according to previously described methods.<sup>57</sup> Briefly, for plasma samples 30µl of cold methanol were added to 10 µl of each sample, vortexed for 1 minute and incubated for one hour at -20 °C. For faecal samples, the content of a 1.2 ml tube of Lysing Matrix E (MP biomedical) and 600 µl of cold methanol were added to 10mg of sample. Samples were homogenized using FastPrep-24™ (MP biomedical) and were incubated overnight in a rocker at 4°C. Then, all samples were centrifuged for three minutes at 12.000g, the supernatant was recovered and filtered with a 0.2 µm Eppendorf filter. Two µL of the extracted sample were applied onto a reversed-phase column (Zorbax SB-Aq 1.8 µm 2.1 x 50 mm; Agilent Technologies) equipped with a precolumn (Zorbax-SB-C8 Rapid Resolution Cartridge 2.1 x 30 mm 3.5 µm; Agilent Technologies) with a column temperature of 60°C. The flow rate was 0.6 mL/min. Solvent A was composed of water containing 0.2% acetic acid and solvent B was composed of methanol 0.2% acetic acid. The gradient started at 2% B and increased to 98% B in 13 min and held at 98% B for 6 min. Post-time was established in 5 min.

Data were collected in positive and negative electrospray modes time of flight operated in full-scan mode at 50–3000 m/z in an extended dynamic range (2 GHz), using N2 as the nebulizer gas (5 L/min, 350°C). The capillary voltage was 3500 V with a scan rate of 1 scan/s. The ESI source used a separate nebulizer for the continuous, low-level (10 L/min) introduction of reference mass

compounds 121.050873 and 922.009798, which were used for continuous, online mass calibration. MassHunter Data Analysis Software (Agilent Technologies, Barcelona, Spain) was used to collect the results, and MassHunter Qualitative Analysis Software (Agilent Technologies, Barcelona, Spain) to obtain the molecular features of the samples, representing different, co-migrating ionic species of a given molecular entity using the Molecular Feature Extractor algorithm (Agilent Technologies, Barcelona, Spain). We selected samples with a minimum of 2 ions. Multiple charge states were forbidden. Compounds from different samples were aligned using a retention time window of  $0.1\% \pm 0.25$  minutes and a mass window of  $20.0 \text{ ppm} \pm 2.0 \text{ mDa}$ . We selected only those present in at least 50% of the samples of one group and corrected for individual bias.

*<sup>1</sup>H-NMR Metabolomics analyses (IRONMET)*. Plasma samples were thawed at room temperature. For each sample, 400  $\mu\text{L}$  of plasma were combined with 200  $\mu\text{L}$  of phosphate buffer ((9%w/V NaCl, 100% D<sub>2</sub>O) that contained 10mM of 3-trimethylsilyl-1-[2,2,3,3-2H<sub>4</sub>] (TSP). Samples were mixed with the use of a vortex and centrifuged (10.000 x g) for 10 min. Then, a 550  $\mu\text{L}$  aliquot was transferred into a 5 mm NMR tube prior to NMR analysis. <sup>1</sup>H spectra of low molecular weight metabolites were performed using a CPMG sequence (RD-90°-[t-180°- t]<sub>n</sub>-ACQ-FID) with spin-echo delay of 400  $\mu\text{s}$  (for a total T<sub>2</sub> filter of 210 ms) allowing an efficient attenuation of the lipid NMR signals. The CPMG sequence generates spectra edited by T<sub>2</sub> relaxation times, reducing broad resonances from high molecular weight compounds facilitating the observation of low molecular weight metabolites. The total acquisition time was 2.73 s with a RD of 2s and the 90° pulse length was automatically calibrated for each sample at around 11.1  $\mu\text{s}$ . For each sample, 8 dummy scans were followed by 256 scans and collected in 64-K points over a spectral width of 20 ppm. TSP was used a general reference for NMR samples because it does not introduce any additional signals apart from the sharp methylsilyl resonance at 0 ppm. In addition, a high concentration of TSP was used to release low-molecular weight metabolites with high affinity for serum proteins by binding competition with TSP.

For faecal samples, 15-20 mg of dried faecal matter was placed in a 2 ml Eppendorf tube. Then, 500  $\mu\text{L}$  of 0.05 M PBS buffer in  $\text{H}_2\text{O}$  (pH=7.3) was added and vortexed vigorously, frozen and thawed twice and centrifuged (21000 g, 15 min, 4°C) to obtain a clear faecal water over the precipitated stool. From the upper layer, 200  $\mu\text{L}$  of prepared faecal water was placed in appropriate 2 ml Eppendorf tube and then, 400  $\mu\text{L}$  of 0.05M PBS buffer in  $\text{D}_2\text{O}$  (pH=7.2, TSP 0.7mM) was added. The sample was vigorously vortexed and sonicated until complete homogenization and the mixture (clear dispersion), if necessary, was centrifuged again (14000 rpm around 14000 g, 5 min, 4°C). For NMR measurement the clear upper phase was placed into a 5mm o.d. NMR tube. One dimensional  $^1\text{H}$  pulse experiments were carried out using the NOESY-presaturation sequence [recycle delay (RD)-90°- $t_1$ -90°- $t_m$ -90° acquire (ACQ) free induction decay (FID)] to suppress the residual water peak. For each sample, 8 dummy scans were followed by 256 scans and collected in 64-K points over a spectral width of 20 ppm.

All  $^1\text{H}$ -NMR spectra were recorded at 300 K on an Avance III 600 spectrometer (Bruker®, Germany) operating at a proton frequency of 600.20 MHz using a 5 mm PABBO gradient probe and automatic sample changer with a cooling rack at 4°C.

### **Proline supplementation mice experiment**

*Animals.* Male CD-1 mice (8 weeks old at the beginning of the experiment) were used. Animals were housed in reverse light-dark cycle (lights on from 20:00 to 8:00), standard temperature (21° +/- 1° C) and with food and water available ad libitum. Mice exposed to the Unpredictable Chronic Mild Stress (UCMS) procedure were singly housed in one room. Control mice, not exposed to UCMS, were housed 2-4 per cage in a different room. Animals were divided into the following experimental groups based on the type of diet and water they were exposed to. Thus, 10 control mice were exposed to standard diet and water whilst another 10 control mice were exposed to an especial diet rich in crude fiber (RCFD; Altromin C1014, 20% crude fiber content) and to l-proline (Sigma-Aldrich P0380) diluted in the drinking water (36 g/l) for the entire experimental procedure. Similar, mice exposed to the UCMS protocol were divided in the following 4 groups of 10 mice each: i) standard diet and water, ii) standard diet and l-proline (36 g/l) in drinking

water, iii) RCFD and water, and iv) RCFD and l-proline (36 g/l) in the drinking water. L-proline was freshly prepared and replaced every 3-4 days.

*Unpredictable Chronic Mild Stress procedure (UCMS).* Three days after the beginning of the exposure to the different diets and to l-proline the UCMS protocol started (similar to the one previously published by Farley et al., 2010). During the dark part of the cycle stressors were applied twice a day for a maximum of 3 h/stressor (and a minimum of 1 h resting in between stressors) in a randomized order. Mice were exposed to tilted cages (45°), wet bedding (150 ml of water/100 g sawdust bedding), paired housing, crowding, home cage bedding with the smell of an unknown mouse, unpredictable illumination, housed in a cage with no bedding and with 2 cm depth of water, forced swim test (only once) and tail suspension test (only once). Stressors during the light part of the cycle (12 h/stressor) included tilted cages, wet bedding and bedding with the smell of unknown mouse. Exposure to the UCMS protocol lasted for 6 weeks. Body weight and fur conditions were regularly checked. Control mice were left undisturbed in a separate room, with the exception of the weekly measurements of body weight and the sucrose intake (individualized), tail suspension and forced swim tests evaluations (performed on the same days as the UCMS mice). Mice were sacrificed 12 h after the exposure to the last stressor, 43 days after the beginning of the UCMS protocol. Plasma samples, cecum and fecal pellets were obtained.

*Sucrose consumption test.* To evaluate the development of an anhedonic-like behavior. Previous habituation to sucrose consumption, sucrose intake (4%, 2 h) measurement were performed on both, control and UCMS mice, 41 days after the beginning of the UCMS protocol by using calibrated pipettes.

*Tail suspension test.* The tail suspension test was performed as previously reported (Mato et al., 2007). Animals were individually suspended by the tail using adhesive tape (distance from tip of tail was 2 cm). Mice were suspended for 6 min. Immobility time (depressive-like behavior) was evaluated on the last 4 min of the test. Mice were exposed to the tail suspension test 40 days after

the beginning of the UCMS protocol. Both, control and UCMS mice, were exposed to the test the same day.

*LBP measurement.* Serum LBP (HK205-02, LBP mouse ELISA kit, Hycult Biotech Inc., PA, USA) were measured using commercial kits according to manufacturer's instructions with intra- and interassay coefficients of variation <8%.

### **Animals faecal microbiome transplantation (FMT) experiment.**

*Animals.* Male C57BL/6J mice (Charles River, France), weighing 23–26 g at the beginning of the experiment were used. Upon arrival to the animal facilities, animals were let to adapt during 5 days to housing conditions (12 hours reversed light/dark cycle, 08:00 AM lights off). Mice were housed individually in controlled laboratory conditions with temperature maintained at  $21 \pm 1$  °C and humidity at  $55 \pm 10\%$ . All animals were fed a standard chow diet RM1 (Irradiated Vacuum packed, Dietex International Ltd.). Food and water were available *ad libitum* during all the experiment. The health status of each mouse included in the experimental schedule was checked every day before the experimental sessions and recorded in the experimenter protocol notebook. Health status checks included body weight, physical aspect, behaviour, and clinical signs. No abnormalities were recorded in the animals included in this study. All animal procedures were performed in accordance with the guidelines of the European Communities Council Directive 2010/63/EU regulating animal research and were approved by the local ethical committee (Comitè Ètic d'Experimentació Animal-Parc de Recerca Biomèdica de Barcelona, CEEA-PRBB).

*Experimental design.* Mice were given *ad libitum* cocktail of antibiotics during 14 days in drinking water to deplete gut microbiota. Antibiotic cocktail consisted of ampicillin (1 g/L), metronidazole (1 g/L), vancomycin (400 mg/L), ciprofloxacin HCl (250 mg/L) and imipenem (250 mg/L). After 14 days of antibiotic intake animals were subjected to a 72 hours wash out and then colonized via daily oral gavage of donor microbiota (200 µL) for 3 days. Donor microbiota was acquired from fecal samples patients with a range of PHQ-9 scores. Booster inoculations

were given twice weekly to throughout the study to reinforce donor microbiota phenotype. Animals were exposed to a fear conditioning test with nociception assessed by the hot plate test to ensure specificity. At the end of the study, animals were consecutively sacrificed and the brains were quickly removed and the medial prefrontal cortex was dissected according to the atlas of stereotaxic coordinates of mouse brain.<sup>58</sup> Brain tissues were then frozen by immersion in 2-methylbutane surrounded by dry ice, and stored at -80°C.

*Emotional testing in mice.* Fear conditioning was conducted as described previously with some modifications.<sup>59</sup> Mice were individually placed in a shuttle chamber (LE918, Panlab, Barcelona) surrounded by a sound-attenuating cabinet. The chamber floor was formed by parallel stainless-steel bars connected to a scrambled shock generator. On the training day, mice were habituated to the chamber during 180 s before the exposure to an acute beeping 30 s sound (80 dB). Each animal received an unconditioned stimulus (US) (0.6 mA footshock during 2 s) paired with the end of the sound (conditioned stimulus, CS). After the shock, the animal remained for 60 s in the shuttle chamber. To evaluate cued fear conditioning, mice were re-exposed to the CS in a novel environment (a wide white cylinder in the chamber) 24 h after the conditioning session. Mice were allowed to adapt for 180 s to the new environment which was followed by 30 s of the sound used in the training day. After the last sound trial, mice remained in the cylinder for 60 s. Fear memory was assessed as the percentage of time that mice spent freezing during the session. Freezing response, a rodent's natural response to fear, was evaluated by direct observation and defined as complete lack of movement, except for respiration for more than 1 s. The procedure was performed between 8.00 and 12.00 h in an experimental room different to the housing room.

### **Study of gene expression in the mice prefrontal cortex**

RNA quality control performed using the RNA 6000 Nano chip (Agilent) on an Agilent Bioanalyzer 2100 obtaining RIN values between 8.7 - 9.8. Libraries were prepared from 500 ng of total RNA using the TruSeq stranded mRNA library preparation kit (Illumina, #20020594) with TruSeq RNA Single Indexes (Illumina, #20020492 and #20020493) according to the manufacturer's instruction reducing the RNA fragmentation time to 4.5 minutes. Prepared libraries were analyzed

on a DNA 1000 chip on the Bioanalyzer and quantified using the KAPA Library Quantification Kit (Roche, #07960204001) on an ABI 7900HT qPCR instrument (Applied Biosystems). Sequencing was performed with 2x50 bp paired-end reads on a HiSeq 2500 (Illumina) using HiSeq v4 sequencing chemistry. Raw sequencing reads in the fastq files were mapped with STAR version 2.5.3a<sup>55</sup> to the Gencode release 17 based on the GRCm38.p6 reference genome and the corresponding GTF file. The table of counts was obtained with FeatureCounts function in the package subread, version 1.5.1.<sup>60</sup> Genes having less than 10 counts in at least 5 samples were excluded from the analysis.

### **Statistical analysis**

*Clinical variables.* First, normal distribution and homogeneity of variances were tested. Results are expressed as number and frequencies for categorical variables, mean and standard deviation (SD) for normal distributed continuous variables and median and interquartile range [IQ] for non-normal distributed continuous variables. To determine differences between study groups, we used  $\chi^2$  for categorical variables, unpaired Student's t test in normal quantitative and Mann-Whitney U test for non-normal quantitative variables. Spearman or Pearson analysis was used to determine the correlation between quantitative variables. These statistical analyses were performed with SPSS, version 19 (SPSS, Inc, Chicago, IL). Statistics can be found in the figures and legends.

*Metagenomics analysis.* To take into account the compositional structure of the microbiome data and rule out possible spurious associations, raw counts were transformed using a centered log-ratio (clr) transformation as implemented in the "ALDEx2" R package.<sup>14</sup> It first uses a Dirichlet-multinomial model to infer abundance from read counts and then applies a clr transformation to each instance. We used 128 Dirichlet Monte Carlo instances in the `aldex.clr` function. Bacterial species and functions associated with the PHQ-9 scores were identified using robust linear regression models as implemented in the Limma R package,<sup>20</sup> adjusting for age, body mass index, sex, education years, antidepressant and anxiety medication. Taxa and bacterial functions were previously filtered so that only those with more than 10 reads in at least five samples were selected. The *p*-values were adjusted for multiple comparisons using a Sequential Goodness of

Fit<sup>61</sup> as implemented in the “SGoF” R package. Unlike FDR methods, which decrease their statistical power as the number of test increases, SGoF methods increase their power with increasing number of tests. SGoF has proven to behave particularly better than FDR methods with high number of tests and low sample size, which is the case of omics large datasets. Statistical significance was set at  $p_{adj} < 0.1$ .

*Metabolomics and whole-brain functional dynamics analysis.* Metabolomics data were first normalized using a probabilistic quotient normalisation. Metabolomics and whole-brain functional dynamics data were analysed using machine learning (ML) methods. In particular, we adopted an all-relevant ML variable selection strategy applying a multiple random forest (RF)-based method as implemented in the Boruta algorithm.<sup>17</sup> It has been recently proposed as one of the two best-performing variable selection methods making use of RF for high-dimensional omics datasets.<sup>62</sup> The Boruta algorithm is a wrapper algorithm that performs feature selection based on the learning performance of the model.<sup>17</sup> It performs variables selection in three steps: a) Randomization, which is based on creating a duplicate copy of the original features randomly permuted across the observations; b) Model building, based on RF with the extended data set to compute the normalized permutation variable importance (VIM) scores; c) Statistical testing, to find those relevant features with a VIM higher than the best randomly permuted variable using a Bonferroni corrected two-tailed binomial test; and d) Iteration, until the status of all features is decided. We run the Boruta algorithm with 500 iterations, a confidence level cut-off of 0.005 for the Bonferroni adjusted  $p$ -values, 5000 trees to grow the forest ( $n_{tree}$ ), and a number of features randomly sampled at each split given by the rounded down number of features/3 (the  $m_{try}$  recommended for regression). Pathway over-representation analysis was performed mapping metabolites that were significantly associated with the PHQ-9 in the discovery cohort to the KEGG, Reactome, INOH and HumanCyc databases included in the ConsensusPathDB.<sup>18</sup>

*RNA-seq analysis.* Differential expression gene analyses were performed on gene counts using the “limma” R package.<sup>20</sup> First, low expressed genes were filtered, so that only gene with more than 10 reads in at least 2 samples were selected. After filtering, 15,144 genes out of 22,204 were

retained for subsequent analyses. RNA-seq data were then normalized for RNA composition using the trimmed mean of M-value (TMM) as implemented in edgeR package.<sup>63</sup> Normalized counts were then converted to log2 count per million (logCPM) with associated precision weights to account for variations in precision between different observations using the “voom” function with donor’s age, BMI, sex, education years, antidepressant and anxiety medication, and kcal intake as covariates. A robust linear regression model adjusted the previous covariates was then fitted to the data using the “lmFit” function with the option method = “robust”, to limit the influence of outlying samples. Finally, an empirical Bayes method was applied to borrow information between genes with the “eBayes” function. Over-representation analyses were performed by mapping differentially expressed genes into the Reactome, KEGG, and DisGeNET databases. Pathway significance was assessed using a hypergeometric test and a Storey procedure ( $q$ -values) was applied for multiple testing correction. Statistical significance was set at  $qval < 0.1$ . For the FMT study, differentially expressed genes were also mapped to the Search Tool for Retrieval of Interacting Proteins/Genes (STRING) database (which integrates known and predicted protein/gene interactions) to predict functional gene-gene interaction networks.<sup>30</sup> Then, functional local clusters in the interaction network were determined using a Markov Cluster algorithm (MCL) with a inflation parameter = 3. Active interacting sources including text mining, experiments, databases, co-expression, and co-occurrence and an interaction score  $> 0.4$  were used to construct the interaction networks.

## **Acknowledgments**

This work was partially supported by research grants FIS (PI15/01934) from the Instituto de Salud Carlos III from Spain, SAF2015-65878-R from Ministry of Economy and Competitiveness, Prometeo/2018/A/133 from Generalitat Valenciana, Spain and also by Fondo Europeo de Desarrollo Regional (FEDER) funds, European Commission (FP7, NeuroPain #2013-602891), the Catalan Government (AGAUR, #SGR2017-669, ICREA Academia Award 2015), the Spanish Instituto de Salud Carlos III (RTA, #RD16/0017/0020) and the European Regional Development Fund (No. 01.2.2-LMT-K-718-02-0014). We also acknowledge funding from the Spanish Ministry of Science, Innovation and Universities (RTI2018-099200-B-I00), and the Generalitat of Catalonia (Agency for Management of University and Research Grants (2017SGR696) and Department of Health (SLT002/16/00250)) to R.P. María Arnoriaga-Rodríguez is funded by Instituto de Salud Carlos III, Río Hortega (CM19/00190). Jordi Mayneris-Perxachs is funded by a Miguel Servet contract (CP18/00009) from the Instituto de Salud Carlos III. Joaquin Sol is funded by a predoctoral PERIS contract (SLT002/16/00250) from the Catalan Government. Mariona Jové is a professor under “Serra Hunter” program (Generalitat de Catalunya).

## **Author contributions**

J.M-P. researched the data, performed the statistical analysis and wrote the manuscript; M.A-R. researched the data. A.B., M.M., and R.M. performed or analyzed the experiments in mice; G.B., C.B., A.E., and G.D. researched the MRI data; C.C. performed the neuropsychological examination. J.M.M-N., performed the RNA sequencing of jejunal samples. V.P-B. and A.M. contributed with the determination and analysis of the microbiota; J.P., J.G-O., R.R., S.P., R.B., J.C.V., J.S., J.G., L.R-T. and W. R. contributed to the discussion and reviewed the manuscript. M.J., R.P., J.S. and M.P-O. performed the metabolomics analyses. J.M-P. and J.M.F.-R. carried out the conception and coordination of the study, performed the statistical analysis and wrote the manuscript. All authors participated in final approval of the version to be published. J.M.F.-R. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data.

## **Competing interests statement**

The authors declare no competing interests.

## **Data and materials availability**

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

## References

1. James, S. L. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1789–1858 (2018).
2. WHO. World Health Organization. Depression 2020. <https://www.who.int/news-room/fact-sheets/detail/depression>.
3. Hansen, R. *et al.* Meta-analysis of Major Depressive Disorder Relapse and Recurrence With Second-Generation Antidepressants. *Psychiatr. Serv.* **59**, 1121–1130 (2008).
4. Zhou, X. *et al.* Comparative efficacy and acceptability of antidepressants, psychotherapies, and their combination for acute treatment of children and adolescents with depressive disorder: a systematic review and network meta-analysis. *The Lancet Psychiatry* **7**, 581–601 (2020).
5. Liang, S., Wu, X., Hu, X., Wang, T. & Jin, F. Recognizing depression from the microbiota–gut–brain axis. *International Journal of Molecular Sciences* vol. 19 (2018).
6. Arnoriaga-Rodríguez, M. *et al.* Obesity Impairs Short-Term and Working Memory through Gut Microbial Metabolism of Aromatic Amino Acids. *Cell Metab.* **32**, 548–560.e7 (2020).
7. Cryan, J. F. *et al.* The microbiota-gut-brain axis. *Physiol. Rev.* **99**, 1877–2013 (2019).
8. Simpson, C. A. *et al.* The gut microbiota in anxiety and depression – A systematic review. *Clin. Psychol. Rev.* **83**, 101943 (2021).
9. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science (80-. )*. **334**, 105–108 (2011).

10. Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613 (2001).
11. 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).
12. Stevens, B. R. *et al.* Depression phenotype identified by using single nucleotide exact amplicon sequence variants of the human gut microbiome. *Mol. Psychiatry* (2020) doi:10.1038/s41380-020-0652-5.
13. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology* vol. 8 (2017).
14. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**, 15 (2014).
15. Mayneris-Perxachs, J. & Fernández-Real, J.-M. Exploration of the microbiota and metabolites within body fluids could pinpoint novel disease mechanisms. *FEBS J.* **287**, 856–865 (2020).
16. Bröer, S. & Gether, U. The solute carrier 6 family of transporters. *British Journal of Pharmacology* vol. 167 256–278 (2012).
17. Kursa, M. B. & Rudnicki, W. R. Feature selection with the boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).
18. Kamburov, A., Stelzl, U., Lehrach, H. & Herwig, R. The ConsensusPathDB interaction database: 2013 Update. *Nucleic Acids Res.* **41**, D793–D800 (2013).
19. Sonnenburg, J. L. & Bäckhed, F. Diet–microbiota interactions as moderators of human metabolism. *Nature* **535**, 56–64 (2016).

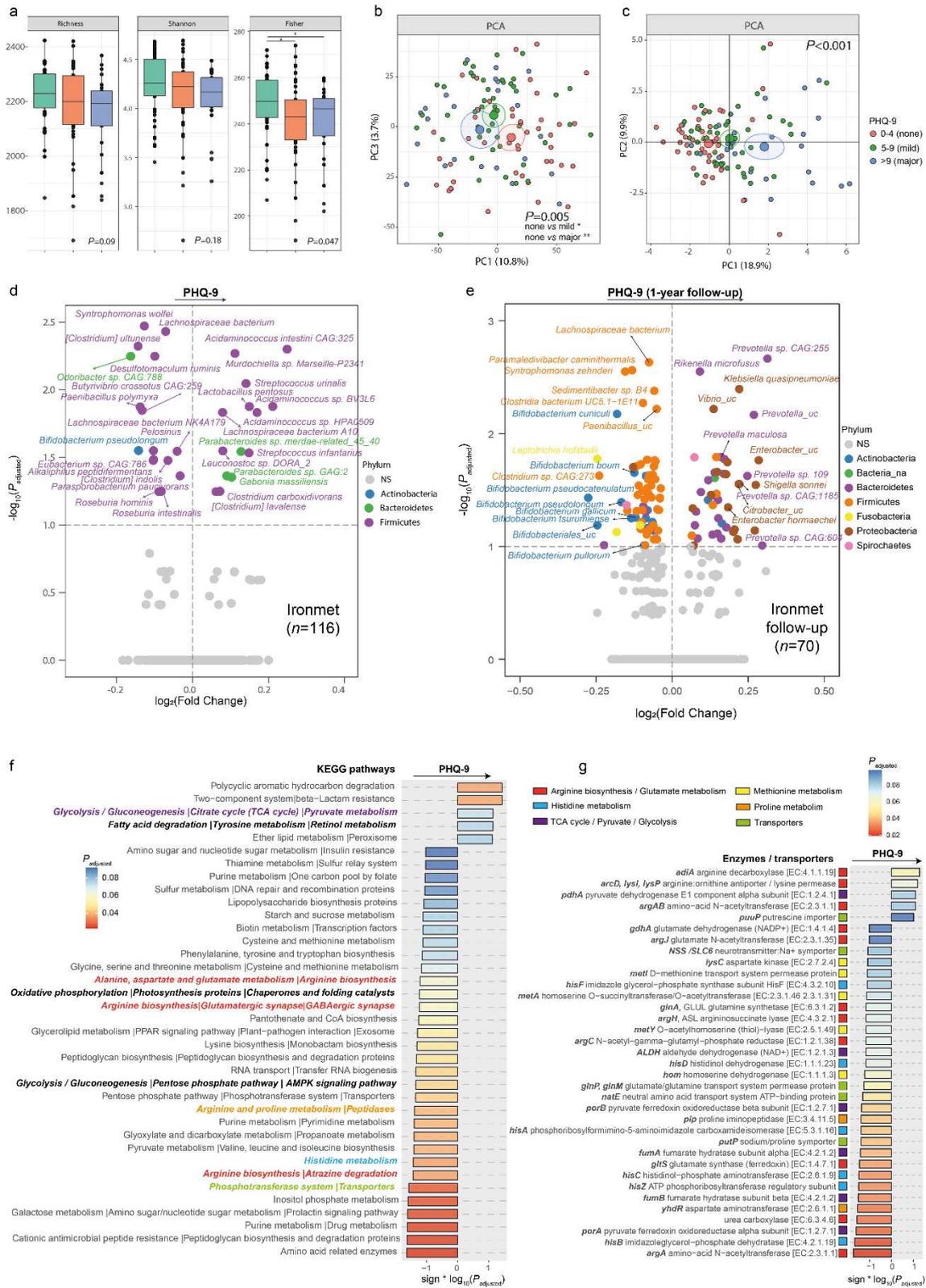
20. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
21. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, (2010).
22. Zecca, L., Youdim, M. B. H., Riederer, P., Connor, J. R. & Crichton, R. R. Iron, brain ageing and neurodegenerative disorders. *Nature Reviews Neuroscience* vol. 5 863–873 (2004).
23. Kim, J. & Wessling-Resnick, M. Iron and mechanisms of emotional behavior. *Journal of Nutritional Biochemistry* vol. 25 1101–1107 (2014).
24. Escrichs, A. *et al.* Whole-Brain Dynamics in Aging: Disruptions in Functional Connectivity and the Role of the Rich Club. *Cereb. Cortex* (2020)  
doi:10.1093/cercor/bhaa367.
25. Deco, G. & Kringelbach, M. L. Hierarchy of Information Processing in the Brain: A Novel ‘Intrinsic Ignition’ Framework. *Neuron* vol. 94 961–968 (2017).
26. Kim, D. J. & Min, B. K. Rich-club in the brain’s macrostructure: Insights from graph theoretical analysis. *Computational and Structural Biotechnology Journal* vol. 18 1761–1773 (2020).
27. Puig, J. *et al.* The aging imageomics study: rationale, design and baseline characteristics of the study population. *Mech. Ageing Dev.* 111257 (2020)  
doi:10.1016/j.mad.2020.111257.
28. Wang, L., Hermens, D. F., Hickie, I. B. & Lagopoulos, J. A systematic review of resting-state functional-MRI studies in major depression. *Journal of Affective Disorders* vol. 142 6–12 (2012).

29. Moriguchi, S. *et al.* Glutamatergic neurometabolite levels in major depressive disorder: a systematic review and meta-analysis of proton magnetic resonance spectroscopy studies. *Molecular Psychiatry* vol. 24 952–964 (2019).
30. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
31. Louis, P. & Flint, H. J. Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology* vol. 19 29–41 (2017).
32. Yunes, R. A. *et al.* GABA production and structure of gadB/gadC genes in *Lactobacillus* and *Bifidobacterium* strains from human microbiota. *Anaerobe* **42**, 197–204 (2016).
33. Dai, Z. L., Zhang, J., Wu, G. & Zhu, W. Y. Utilization of amino acids by bacteria from the pig small intestine. *Amino Acids* **39**, 1201–1215 (2010).
34. Inoshita, M. *et al.* Elevated peripheral blood glutamate levels in major depressive disorder. *Neuropsychiatr. Dis. Treat.* **14**, 945–953 (2018).
35. Pan, J. X. *et al.* Diagnosis of major depressive disorder based on changes in multiple plasma neurotransmitters: A targeted metabolomics study. *Transl. Psychiatry* **8**, 130 (2018).
36. Duman, R. S., Sanacora, G. & Krystal, J. H. Altered Connectivity in Depression: GABA and Glutamate Neurotransmitter Deficits and Reversal by Novel Treatments. *Neuron* vol. 102 75–90 (2019).
37. Pu, J. *et al.* An integrated meta-analysis of peripheral blood metabolites and biological functions in major depressive disorder. *Mol. Psychiatry* 1–12 (2020)  
doi:10.1038/s41380-020-0645-4.
38. Crabtree, G. W., Park, A. J., Gordon, J. A. & Gogos, J. A. Cytosolic Accumulation of L-

- Proline Disrupts GABA-Ergic Transmission through GAD Blockade. *Cell Rep.* **17**, 570–582 (2016).
39. Wyse, A. T. S. & Netto, C. A. Behavioral and neurochemical effects of proline. *Metabolic Brain Disease* vol. 26 159–172 (2011).
  40. Schulz, D., Morschel, J., Schuster, S., Eulenburg, V. & Gomeza, J. Inactivation of the Mouse L-Proline Transporter PROT Alters Glutamatergic Synapse Biochemistry and Perturbs Behaviors Required to Respond to Environmental Changes. *Front. Mol. Neurosci.* **11**, (2018).
  41. Liu, X. *et al.* Disrupted rich-club network organization and individualized identification of patients with major depressive disorder. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* (2020) doi:10.1016/j.pnpbp.2020.110074.
  42. Phang, J. M., Liu, W. & Zabirnyk, O. Proline Metabolism and Microenvironmental Stress. *Annu. Rev. Nutr.* **30**, 441–463 (2010).
  43. Bach, D. R., Brown, S. A., Kleim, B. & Tyagarajan, S. K. Extracellular matrix: A new player in memory maintenance and psychiatric disorders. *Swiss Medical Weekly* vol. 149 (2019).
  44. Vioque, J. *et al.* Reproducibility and validity of a food frequency questionnaire among pregnant women in a Mediterranean area. *Nutr. J.* **12**, 26 (2013).
  45. Spitzer, R. L., Kroenke, K. & Williams, J. B. W. Validation and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. *J. Am. Med. Assoc.* **282**, 1737–1744 (1999).
  46. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
  47. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve

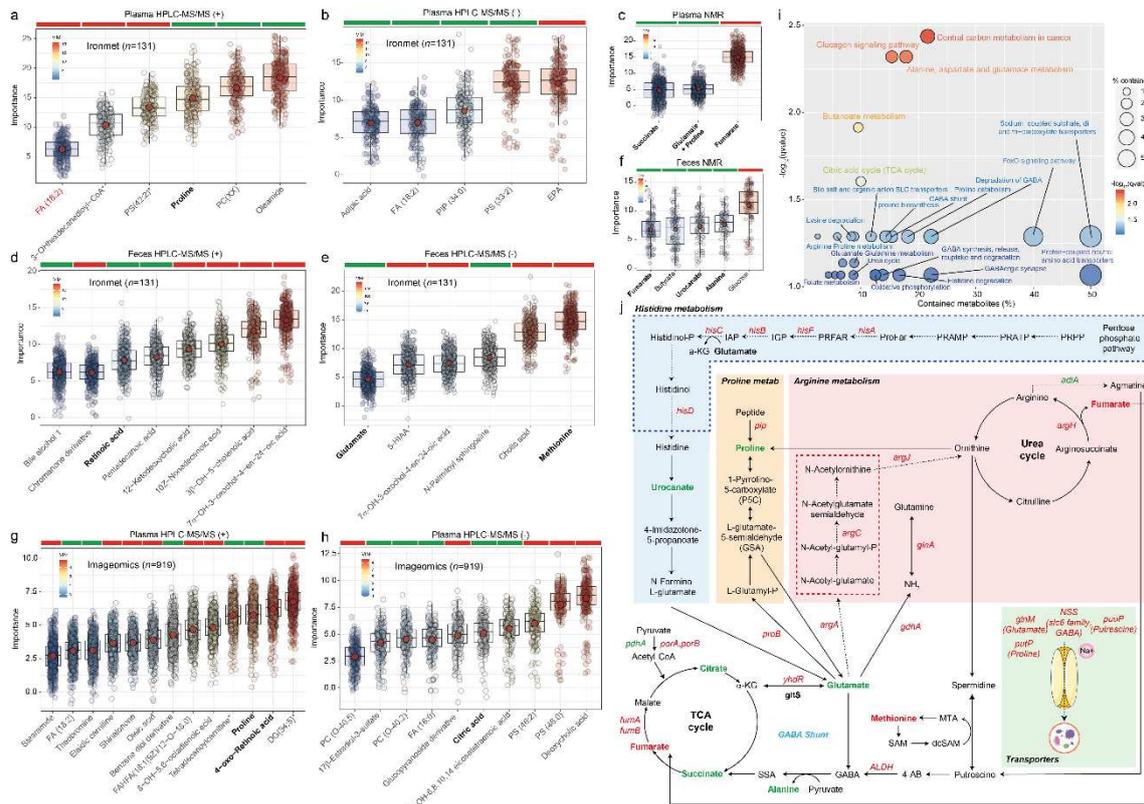
- genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9 (2012).
  49. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
  50. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
  51. Durbin, R., Eddy, S. R., Krogh, A. & Mitchison, G. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. (1998).
  52. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
  53. Team, R. R Development Core Team. *R A Lang. Environ. Stat. Comput.* (2013).
  54. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
  55. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
  56. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, (2011).
  57. Wikoff, W. R., Pendyala, G., Siuzdak, G. & Fox, H. S. Metabolomic analysis of the cerebrospinal fluid reveals changes in phospholipase expression in the CNS of SIV-infected macaques. *J. Clin. Invest.* **118**, 2661–9 (2008).
  58. Paxinos, G. & Franklin, K. B. J. *The mouse brain in stereotaxic coordinates*. (Academic

- Press, 1997).
59. Saravia, R. *et al.* Concomitant THC and stress adolescent exposure induces impaired fear extinction and related neurobiological changes in adulthood. *Neuropharmacology* **144**, 345–357 (2019).
  60. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
  61. Carvajal-Rodríguez, A., de Uña-Alvarez, J. & Rolán-Alvarez, E. A new multitest correction (SGoF) that increases its statistical power when increasing the number of tests. *BMC Bioinformatics* **10**, 209 (2009).
  62. Degenhardt, F., Seifert, S. & Szymczak, S. Evaluation of variable selection methods for random forests and omics data sets. *Brief. Bioinform.* **20**, 492–503 (2019).
  63. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

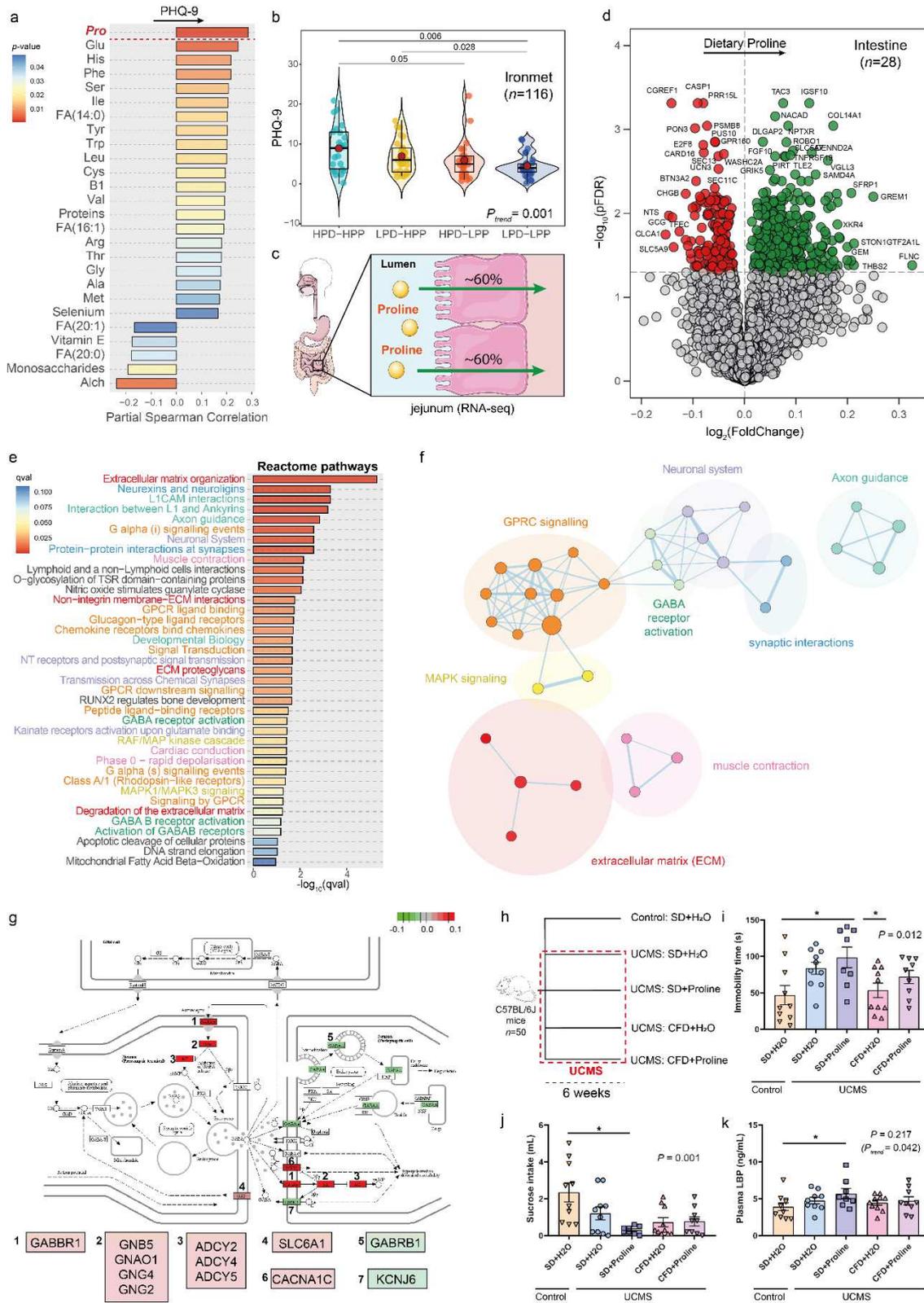


**Fig. 1. Associations of the gut microbiota composition and functionality with depression cores.** **a)** Alpha diversity indices (IRONMET,  $n=116$ , Kruskal-Wallis). **b)** Principal component analysis scores plot based on clr-transformed shotgun sequencing metagenomic microbial taxonomy data colored according to depression status. Overall differences in the microbiome composition were assessed by PERMANOVA using 1000 permutations and Euclidean distances. Pairwise differences between groups were assessed using the *pairwise.adonis* function adjusted for Bonferroni correction. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ . **c)** Principal component analysis scores plot

based only on significant microbial species associated with the PHQ-9 scores. **d)** Volcano plots of differential bacteria associated with the baseline and **e)** 1-year follow-up PHQ-9 scores identified in the IRONMET cohort after fitting a robust linear regression model to the clr-transformed data controlling for age, gender, BMI, education years, and antidepressant and anxiety medication. The  $\log_2$  fold change associated with a unit change in the clr-transformed values and the  $\log_{10}$   $p$ -values adjusted for multiple testing are plotted for each taxon. Significantly different taxa are colored according to phylum. **f)** Manhattan-like plot of significantly expressed KEGG bacterial pathways and **g)** functional associated with the PHQ-9 scores identified from the clr-transformed shotgun metagenomics data in the IRONMET cohort controlling for age, gender, BMI, education years, and antidepressant and anxiety medication. Bars are colored according to the  $p$ -value adjusted for multiple testing. The  $-\log_{10}(P_{\text{adjusted}})$  values are multiplied by the fold change sign to take into account the direction of the association. Pathways and functions involved in arginine, histidine, proline, methionine, glutamate and GABA and related transporters are highlighted in color.

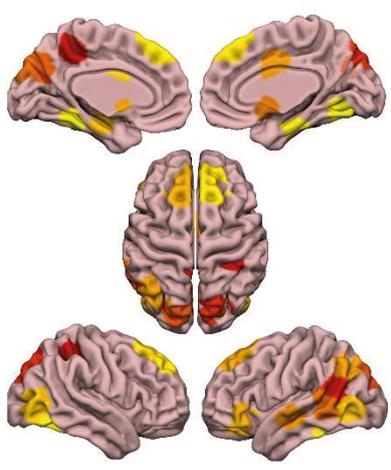
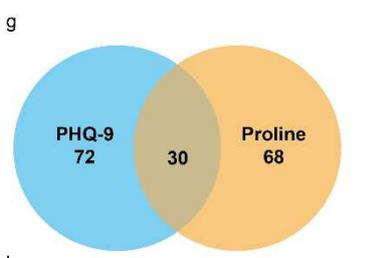
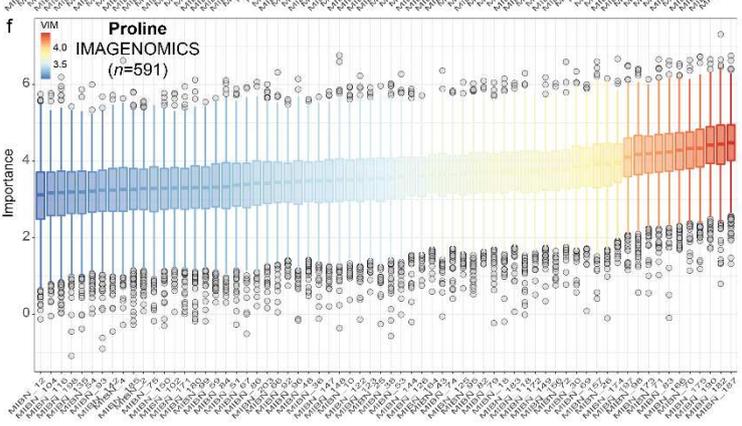
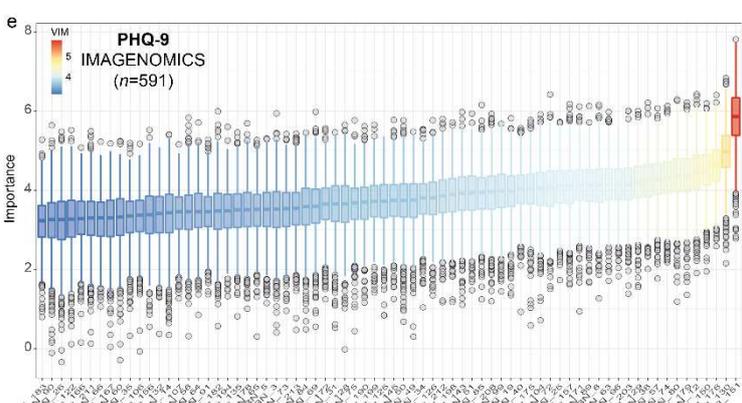
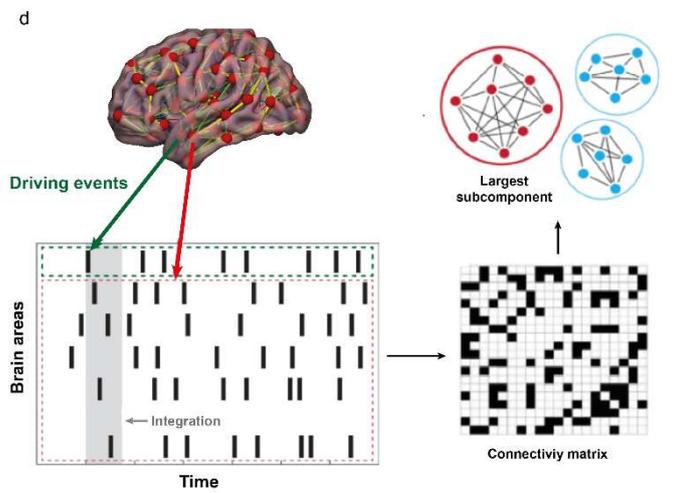
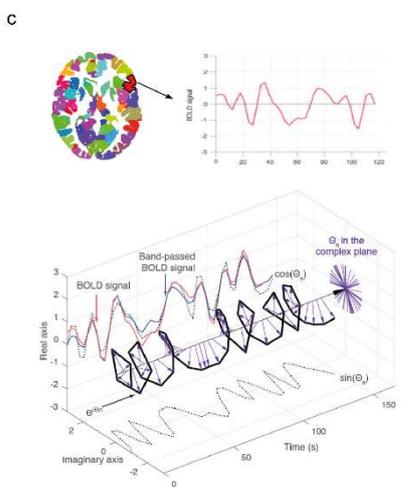
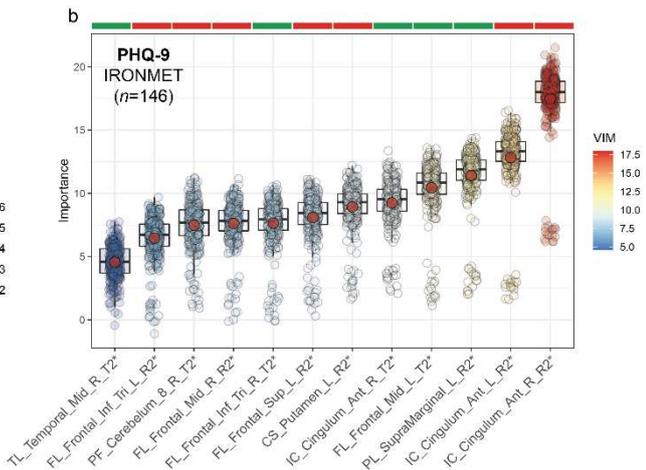
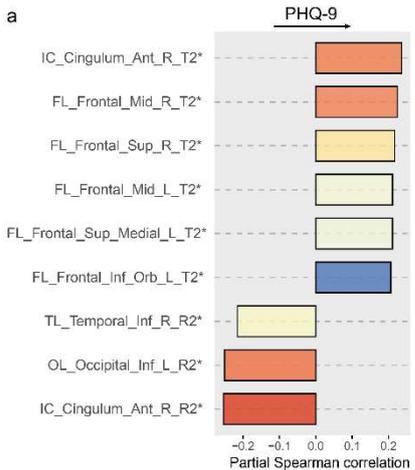


**Fig. 2. Plasma and fecal metabolites associated with the PHQ-9 depression scores in the IRONMET and IMAGEOMICS cohorts identified by machine learning.** Barplots of the normalized variable importance measure (VIM) for the metabolites associated with the PHQ-9 scores in **a**) plasma (HPLC-MS/MS in positive mode), **b**) plasma (HPLC-MS/MS in negative model), **c**) plasma (1H-NMR), **d**) feces (HPLC-MS/MS in positive mode), **e**) feces (HPLC-MS/MS in negative model), and **f**) feces (1H-NMR), in the IRONMET ( $n=131$ ) cohort. **g,h**) Barplots of VIM for the metabolites associated with the PHQ-9 score plasma samples from the IMAGEOMICS cohort ( $n=919$ ) identified by HPLC-ESI-MS/MS in positive and negative modes, respectively. The color bar above each plot indicates the sign of the association among the metabolites the PHQ-9 scores, with red indicating negative correlation and green positive correlation. Significant metabolites were identified using a multiple random forest-based machine learning variable selection strategy as implemented in the Boruta algorithm with 5000 trees and 500 iterations. All metabolites were identified based on exact mass, retention time and MS/MS spectrum, except those with (\*) that were only identified based on exact mass and retention time. **i**) Pathway over-representation analysis of metabolites significantly associated with the PHQ-9 scores based on KEGG, Reactome, INOH, and HumanCyc databases. The bubble size represents the percentage of significant genes contained in each pathway. Bubbles are colored according to the p-value adjusted for multiple testing (q-values). **j**) Representation of the connections of histidine (blue), proline (orange), and arginine (red) metabolic pathways with the glutamate and GABA metabolism. It also shows how the GABA shunt interconnects the TCA cycle with GABA and glutamate. Dashed lines indicate pathways only occurring in bacteria. Metabolites and bacterial functions positively and negatively associated with the PHQ-9 scores are highlighted in green and red, respectively. 5-HIAA, 5-hydroxyindole acetic acid; Benzene diol derivative, 3-methoxy-5-[7-methoxy-3-methyl-5-(prop-1-en-1-yl)-2,3-dihydro-1-benzofuran-2-yl]benzene-1,2-diol; Bile alcohol 1, b-cholestane-3 a,7a,12a,23,25-pentol; chromanone derivative; 3-(3,7-dimethylocta-2,6-dien-1-yl)-5-hydroxy-8,8-dimethyl-2-(2,4,5-trihydroxyphenyl)-4H,8H-pyrano[3,2-g]chromen-4-one; EPA, eicosapentanoic acid, fatty acid; Glycopyranoside derivative, 1,2,10-Trihydroxydihydro-trans-linalyl oxide 7-O-b-D-glucopyranoside; PC, phosphatidylcholine; PIP, phosphatidylinositol phosphate; PS, phosphatidylserine.

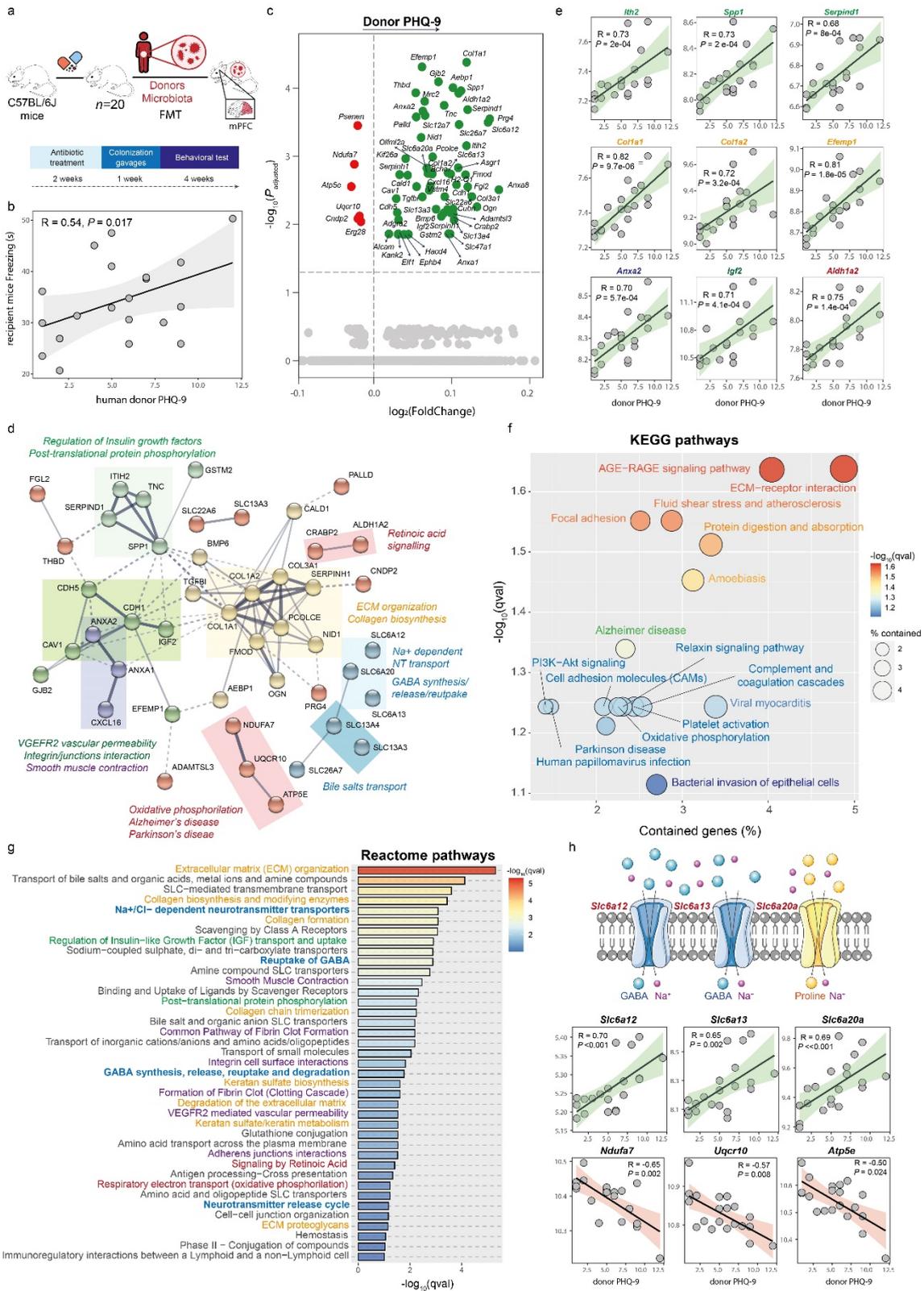


**Fig. 3. Associations of dietary proline with PHQ-9 scores and jejunal RNA sequencing.** **a)** Partial Spearman's rank correlations of the depression scores with macronutrients, vitamins, minerals, amino acids and fatty acids derived from food frequency questionnaires controlling for age, BMI, sex, education years, kcal intake and antidepressant and anxiety medication in the IRONMET cohort ( $n=116$ ). Items above the red dashed line are significant after adjusting for multiple comparisons (FDR). **b)** Violin plots of the PHQ-9 scores in the IRONMET cohort

according to high or low proline levels in plasma (HPP or LPP, respectively) and high or low proline from the diet (HPD or LPD, respectively). Overall significance was assessed using a Mann-Kendall trend test and between groups significance using a Wilcoxon test. **c)** Dietary proline is extensively metabolized by enterocytes in the small intestine (~40%). **d)** Volcano plot of differentially expressed jejunal gene transcripts associated with the consumption of proline identified by limma-voom controlling for age, BMI, sex, education years, kcal intake, and antidepressant medication ( $pFDR < 0.05$ ) in the INTESTINE cohort ( $n=28$ ). **e)** Pathway over-representation analysis of the jejunal genes significantly associated with the dietary proline based on Reactome database ( $qval < 0.1$ ). Bars are colored according to the  $qval$  and pathway names are colored based on the results of functional enrichment map. **f)** Over-representation results were mapped as a functional network of pathways using Cytoscape and enrichment map. Nodes represent over-represented pathways. Node size reflects the total number of genes in each pathway. Edge thickness represents the degree of overlapping genes between pathways. Groups of functionally related pathways are circled and labelled. Pathways (nodes) are colored according to the functional group they belong. Nodes with not overlapping are not represented. **g)** KEGG map of regulation of the GABAergic synapse pathway, with each block representing a group of genes. A white block indicates that no gene in that block was significantly associated with dietary proline, whereas green/red backgrounds indicate significantly down-/up-regulated genes, respectively. Node color reflects  $\log_2$  fold change range. Over-expressed genes in each block are highlighted in the boxes below. Data were integrated and visualized using Pathview. **h)** Scheme of the proline supplementation study design. CFD, crude fiber rich diet; SD, standard diet; UCMS, unpredictable chronic mild stress. **i)** Immobility time (measured by tail suspension), **j)** sucrose intake, and **l)** plasma lipopolysaccharide binding protein (LBP) in the experimental groups. Differences were assessed by one-way ANOVA followed by Tukey or Dunnett's post-hoc test in case of equal or different variances, respectively. A Mann-Kendal trend test was also used to analyze the linear trend in the plasma LBP measure in the SD groups.

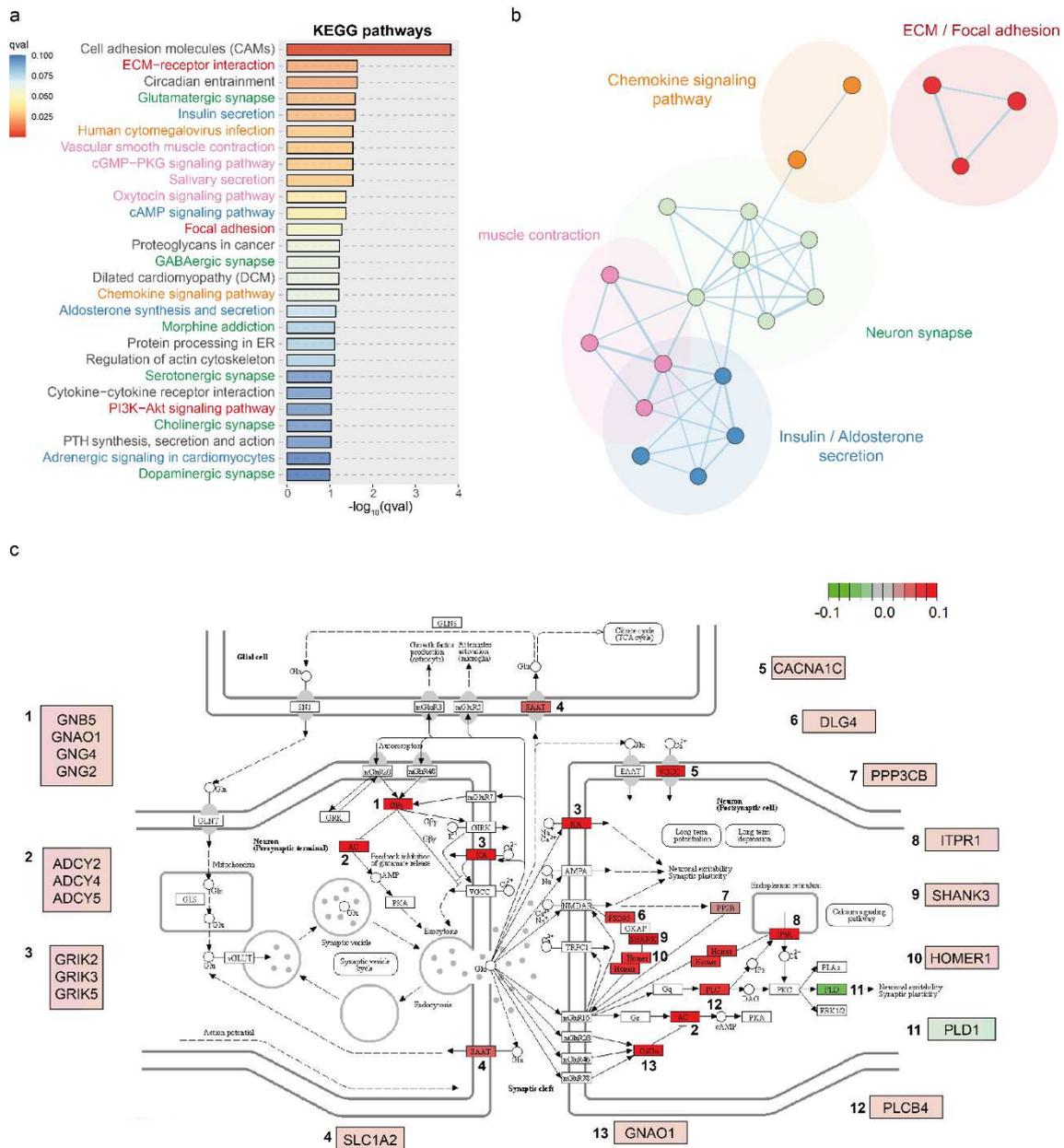


**Fig. 4. Associations of brain iron deposition and intrinsic brain networks with depression scores and circulating proline identified using machine learning.** **a)** Partial Spearman's rank correlations of the PHQ-9 scores with relaxometry parameters (mean R2\* and T2\* values) of brain regions from the Anatomical Automatic Labelling brain atlas controlling for age, BMI, sex, education years, antidepressant and anxiety medications in the IRONMET cohort ( $n=146$ ). **b)** Barplots of the normalized variable importance measure (VIM) for the relaxometry measures of brain regions associated with the PHQ-9 scores. **c)** We extracted the BOLD time series using a resting-state atlas and computed the phase space of the BOLD signal for each brain area by using the Hilbert transform. The phase dynamics can be represented in the complex plane as  $e^{i\phi}$  (black bold line), the real part as  $\cos(\phi)$  (black dotted line), and the imaginary part as  $\sin(\phi)$  (black dotted line). The purple arrows represent the Hilbert phases at each TR. *Adapted from* <sup>27</sup>. **d)** Intrinsic Ignition Framework. Events were obtained by applying a threshold method, and for each driving event, the activity in the rest of the network (see red stippled area) was obtained in the set 4TR time window (gray area). A binarized phase matrix was extracted from the time window. From this phase matrix, we obtained the integration by calculating the largest subcomponent and computed the mean ignition for each brain area across time. *Adapted from* <sup>27</sup>. **e)** Barplots of VIM for the mean ignition of brain areas associated with the PHQ-9 scores and **f)** circulating proline in the IMAGENOMICS cohort ( $n = 591$ ). Significant variable in the machine learning analyses were identified using a multiple random forest-based variable selection strategy as implemented in the Boruta algorithm with 5000 trees and 500 iterations. **g)** Venn diagram for the associations of the PHQ-9 and circulating proline. **h)** Some of these brain areas common to the circulating proline and PHQ-9 scores such as the superior frontal cortex, the precuneus, insula, and subcortical (*i.e.*, the caudate, putamen, and hippocampus) belong to the so-called rich club.

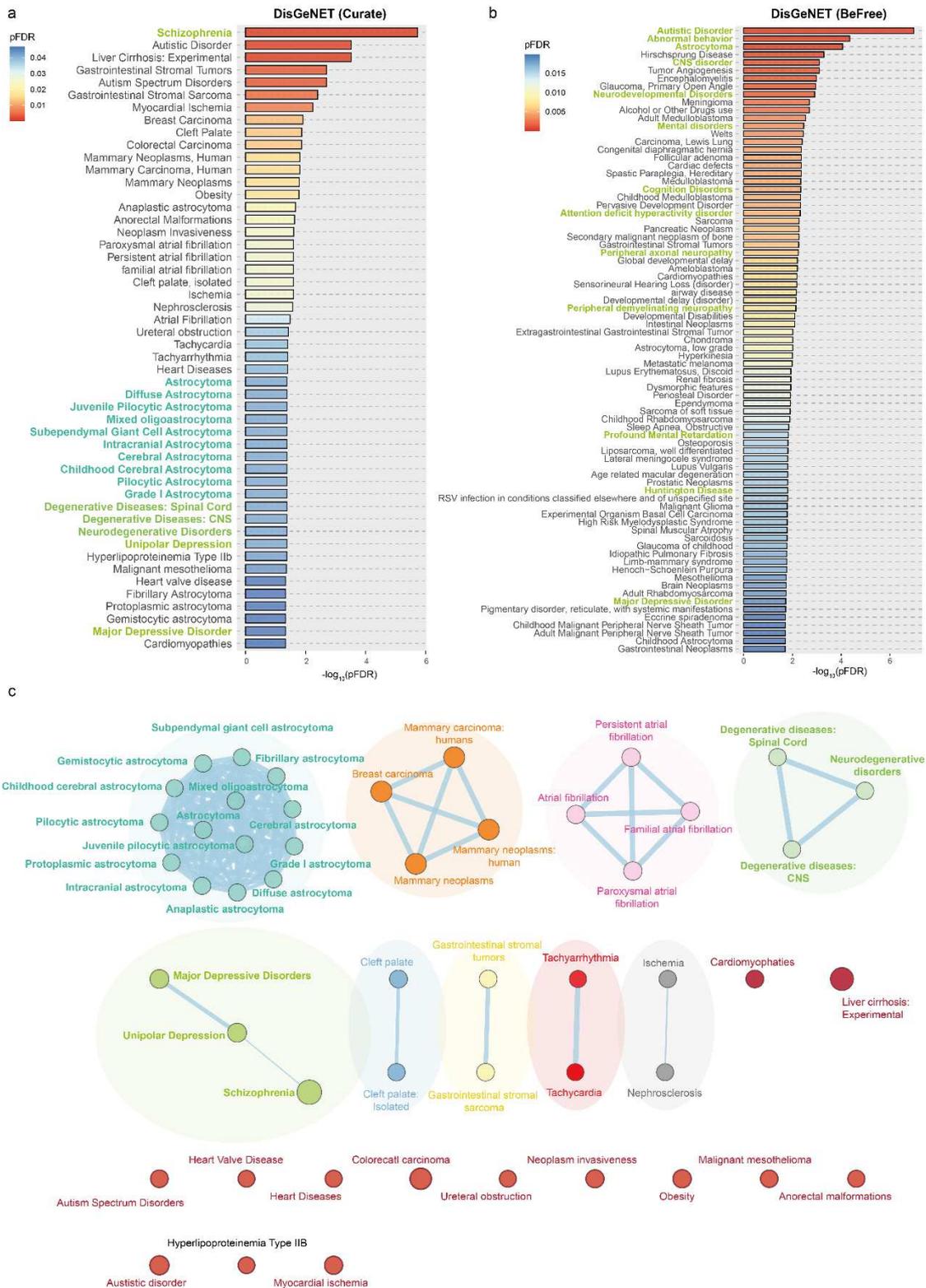


**Fig. 5. An emotional-impaired phenotype was phenocopied to recipient mice in parallel to changes in the expression of mPFC gene transcripts.** a) Experimental design for the FMT study. Microbiota from  $n = 20$  human donors with different PHQ-9 scores was delivered to  $n = 20$  recipient mice pre-treated with antibiotics for 14 days. A freezing test was performed after 4 weeks. b) Scatter plot and correlation between the donor's PHQ-9 scores and the recipient's mice

freezing test scores controlling for donor's age, BMI, sex, education years, and antidepressant and anxiety medication. **c)** Volcano plot of differentially expressed gene transcripts in the medial prefrontal cortex of the recipient mice associated with the donor's PHQ-9 scores identified by both limma-voom and DESeq2 after adjusting  $p$ -values for multiple testing ( $p_{adj}$ ). **d)** Gene interaction network constructed using differentially expressed mPFC gene transcripts associated with donor's PHQ-9 via the Search Tool for the Retrieval of Interacting Proteins/Genes (STRING) database. The network nodes are genes and the edges represent the predicted functional interactions. The thickness indicates the degree of confidence prediction of the interaction. Functional gene clusters are coloured and annotated based on pathway over-representation analysis results. **e)** Scatter plots and correlations of selected relevant genes from the functional clusters. **f)** KEGG-based pathway over-representation analysis of recipient's mice genes associated with the donor's PHQ-9 scores. The bubble size represents the percentage of significant genes contained in each pathway. Bubbles are colored according to the  $p$ -value adjusted for multiple testing ( $q$ -values). **g)** Reactome-based pathway over-representation analysis of the recipient's mice genes significantly associated with the donor's PHQ-9 scores ( $q_{val} < 0.1$ ). Bars are colored according to the  $q_{val}$  and pathway names are colored based on the functional clusters identified with STRING. **h)** Scatter plots and correlations of gene transcripts encoding for GABA and Proline transporters as well as those involved in oxidative phosphorylation.



**Fig. S1. Over-representation analysis of significant jejunal gene transcripts associated with the dietary proline consumption.** a) Reactome-based pathway over-representation analysis ( $qval < 0.1$ ). Bars are colored according to the  $qval$  and pathway names are colored based on the results of functional enrichment map. b) Reactome over-represented pathways were mapped as a functional network of pathways using Cytoscape and enrichment map. Nodes represent over-represented pathways. Node size reflects the total number of genes in each pathway. Edge thickness represents the degree of overlapping genes between pathways. Groups of functionally related pathways are circled and labelled. Pathways (nodes) are colored according to the functional group they belong. Nodes with not overlapping are not represented. c) KEGG map of regulation of the glutamatergic synapse pathway, with each block representing a group of genes. A white block indicates that no gene in that block was significantly associated with dietary proline, whereas green/red backgrounds indicate significantly down-/up-regulated genes, respectively. Node color reflects  $\log_2$  fold change range. Over-expressed genes in each block are highlighted in the boxes below. Data were integrated and visualized using Pathview.



**Fig. S2. Over-representation analysis of gene-disease associations.** Pathway over-representation analysis of the jejunal gene transcripts significantly associated with the dietary proline based on DisGeNET **a)** expert curated and **b)** text mining-based BeFree databases ( $q_{val} < 0.1$ ). **c)** Over-representation results were mapped as a functional network of diseases using Cytoscape and enrichment map. Nodes represent over-represented diseases. Node size reflects the total number of genes involved in each disease. Edge thickness represents the degree of overlapping genes between diseases. Groups of functionally related pathways are circled and labelled. Diseases (nodes) are colored according to the functional group they belong.



depression status. Overall differences in the microbiome composition were assessed by PERMANOVA using 1000 permutations and Euclidean distances. Pairwise differences between groups were assessed using the pairwise.adonis function adjusted for Bonferroni correction. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ . c) Principal component analysis scores plot based only on significant microbial species associated with the PHQ-9 scores. d) Volcano plots of differential bacterial bacteria associated with the baseline and e) 1-year follow-up PHQ-9 scores identified in the IRONMET cohort after fitting a robust linear regression model to the clr-transformed data controlling for age, gender, BMI, education years, and antidepressant and anxiety medication. The log2 fold change associated with a unit change in the clr-transformed values and the log10 p-values adjusted for multiple testing are plotted for each taxon. Significantly different taxa are colored according to phylum. f) Manhattan-like plot of significantly expressed KEGG bacterial pathways and g) functional associated with the PHQ-9 scores identified from the clr-transformed shotgun metagenomics data in the IRONMET cohort controlling for age, gender, BMI, education years, and antidepressant and anxiety medication. Bars are colored according to the p-value adjusted for multiple testing. The  $-\log_{10}(P_{\text{adjusted}})$  values are multiplied by the fold change sign to take into account the direction of the association. Pathways and functions involved in arginine, histidine, proline, methionine, glutamate and GABA and related transporters are highlighted in color.

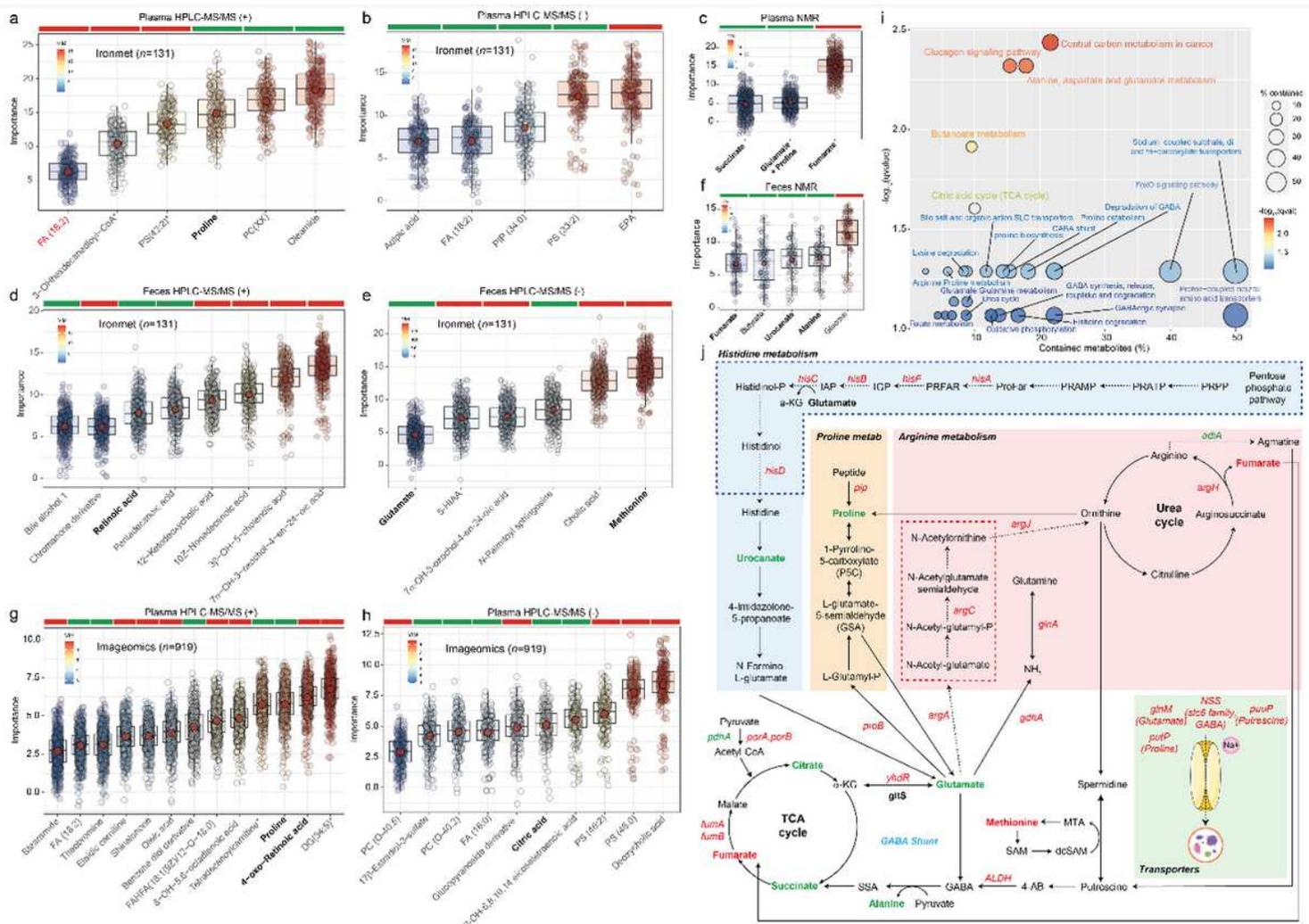
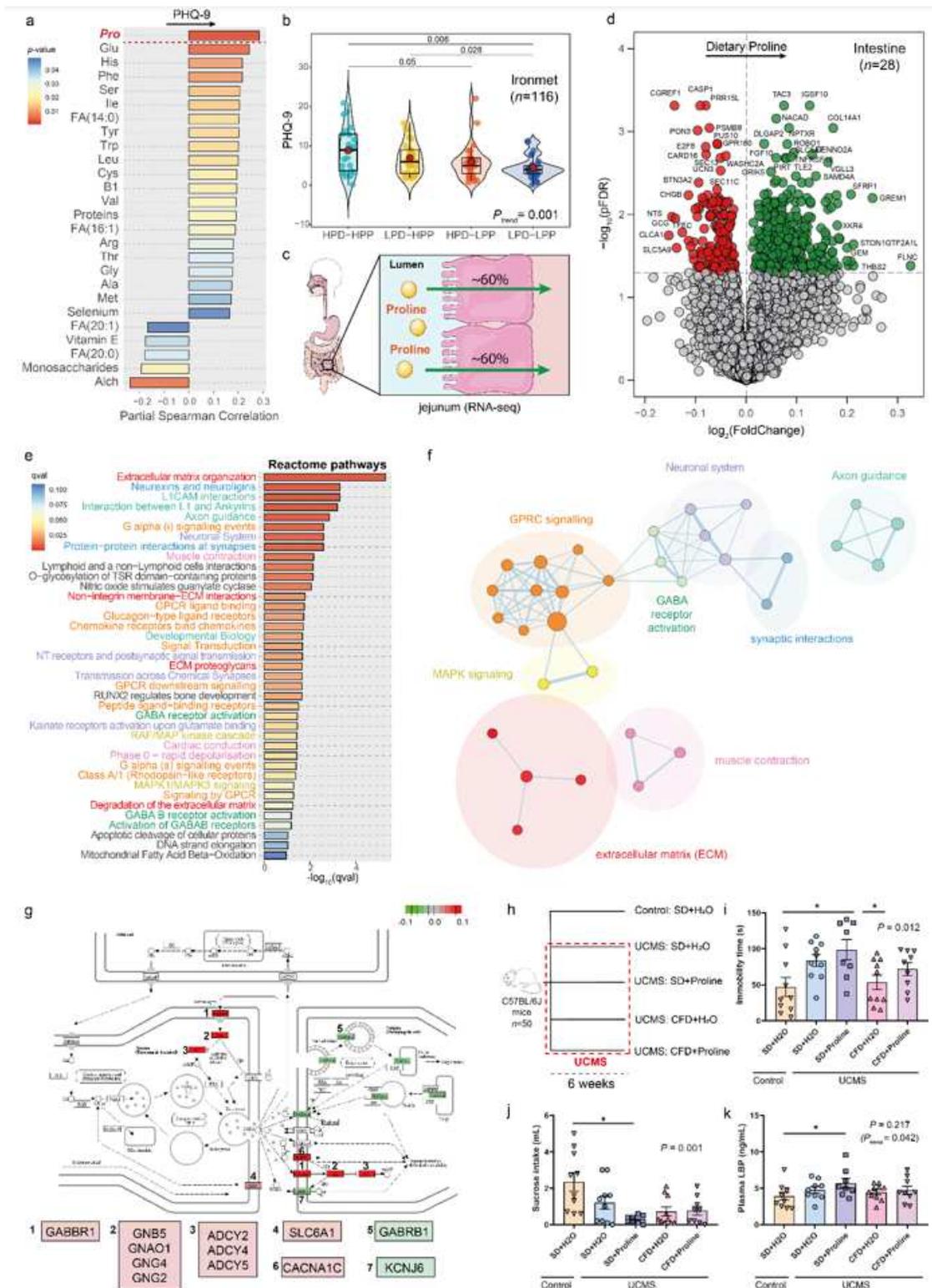


Figure 2

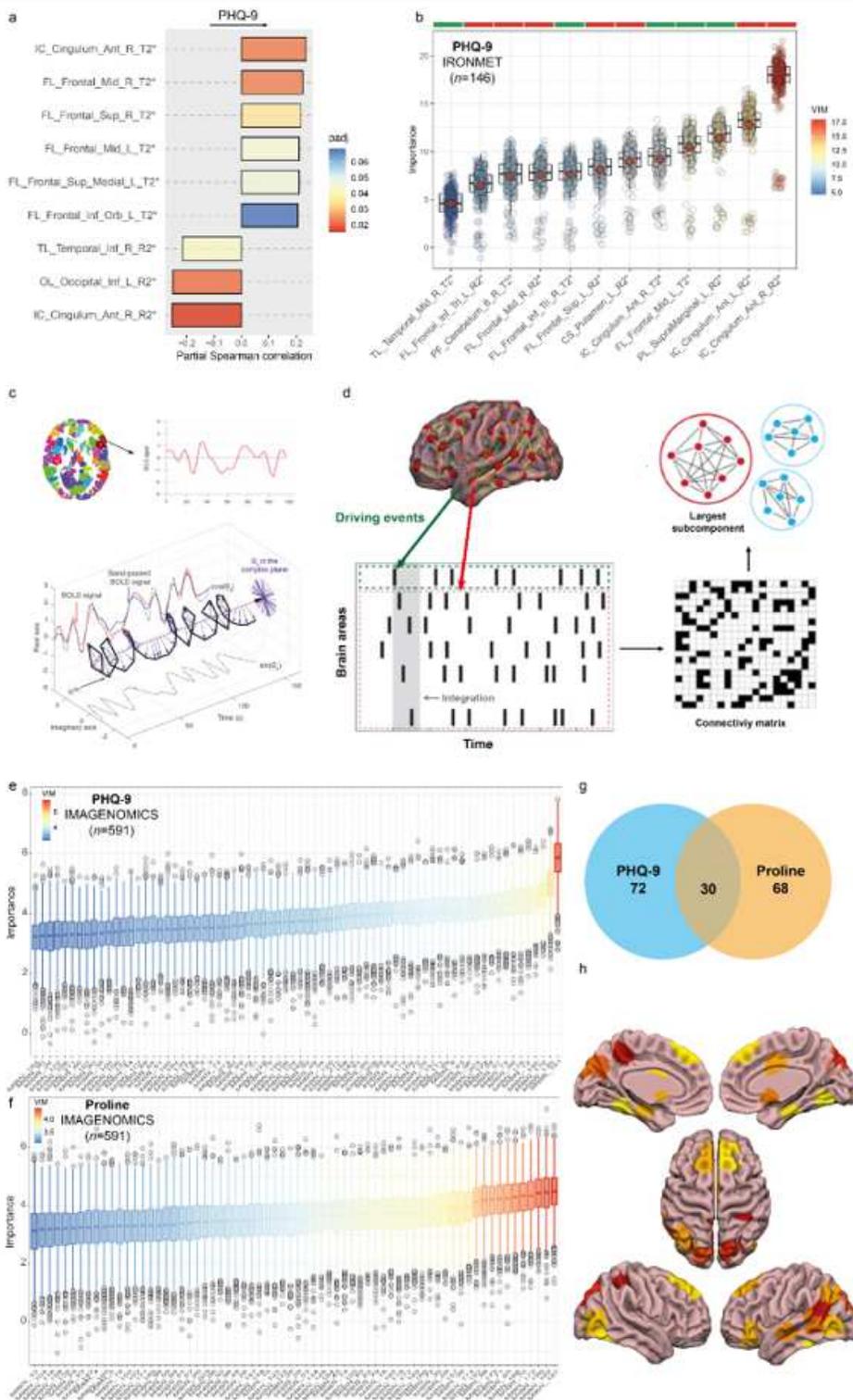
Plasma and fecal metabolites associated with the PHQ-9 depression scores in the IRONMET and IMAGEOMICS cohorts identified by machine learning. Barplots of the normalized variable importance measure (VIM) for the metabolites associated with the PHQ-9 scores in a) plasma (HPLC-MS/MS in positive mode), b) plasma (HPLC-MS/MS in negative model), c) plasma (1H-NMR), d) feces (HPLC-MS/MS in positive mode), e) feces (HPLC-MS/MS in negative model), and f) feces (1H-NMR), in the IRONMET (n=131) cohort. g,h) Barplots of VIM for the metabolites associated with the PHQ-9 score plasma samples from the IMAGEOMICS cohort (n=919) identified by HPCL-ESI-MS/MS in positive and negative modes, respectively. The color bar above each plot indicates the sign of the association among the metabolites the PHQ-9 scores, with red indicating negative correlation and green positive correlation. Significant metabolites were identified using a multiple random forest-based machine learning variable selection strategy as implemented in the Boruta algorithm with 5000 trees and 500 iterations. All metabolites were identified based on exact mass, retention time and MS/MS spectrum, except those with (\*) that were only identified based on exact mass and retention time. i) Pathway over-representation analysis of metabolites significantly associated with the PHQ-9 scores based on KEGG, Reactome, INOH, and HumanCyc databases. The bubble size represents the percentage of significant genes contained in each pathway. Bubbles are colored according to the p-value adjusted for multiple testing (q-values). j) Representation of the connections of histidine (blue), proline (orange), and arginine (red) metabolic pathways with the glutamate and GABA metabolism. It also shows how the GABA shunt interconnects the TCA cycle with GABA and glutamate. Dashed lines indicate pathways only occurring in bacteria. Metabolites and bacterial functions positively and negatively associated with the PHQ-9 scores are highlighted in green and red, respectively. 5-HIAA, 5-hydroxyindole acetic acid; Benzene diol derivative, 3-methoxy-5-[7-methoxy-3-methyl-5-(prop-1-en-1-yl)-2,3-dihydro-1-benzofuran-2-yl]benzene-1,2-diol; Bile alcohol 1, b-cholestane-3 a,7a,12a,23,25-pentol; chromanone derivative; 3-(3,7-dimethylocta-2,6-dien-1-yl)-5-hydroxy-8,8-dimethyl-2-(2,4,5-trihydroxyphenyl)-4H,8H-pyrano[3,2-g]chromen-4-one; EPA, eicosapentanoic acid, fatty acid; Glycopyranoside derivative, 1,2,10-Trihydroxydihydro-trans-linalyl oxide 7-O-b-D-glucopyranoside; PC, phosphatidylcholine; PIP, phosphatidylinositol phosphate; PS, phosphatidylserine.



**Figure 3**

Associations of dietary proline with PHQ-9 scores and jejunal RNA sequencing. a) Partial Spearman's rank correlations of the depression scores with macronutrients, vitamins, minerals, amino acids and fatty acids derived from food frequency questionnaires controlling for age, BMI, sex, education years, kcal intake and antidepressant and anxiety medication in the IRONMET cohort (n=116). Items above the red dashed line are significant after adjusting for multiple comparisons (FDR). b) Violin plots of the PHQ-9

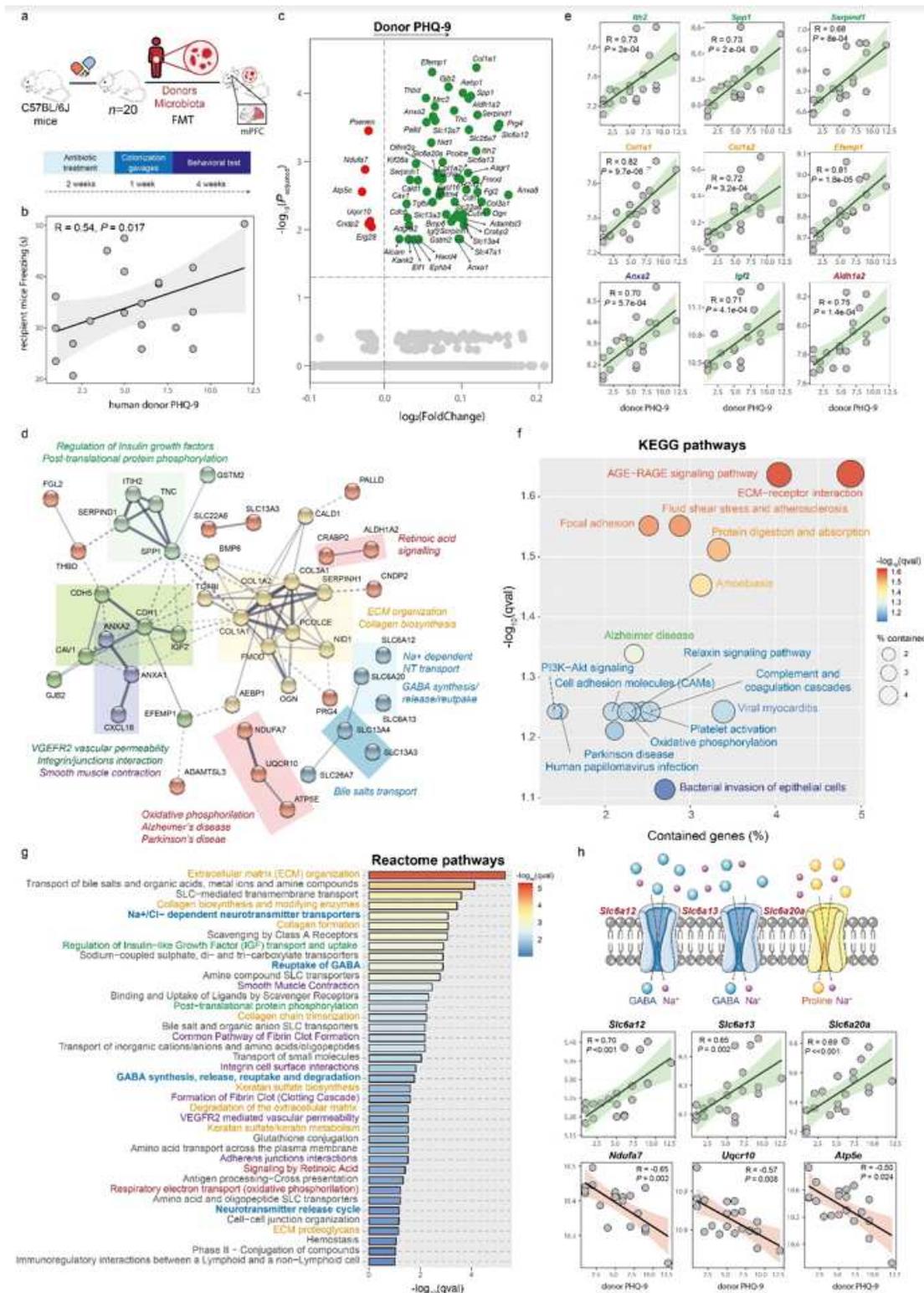
scores in the IRONMET cohort according to high or low proline levels in plasma (HPP or LPP, respectively) and high or low proline from the diet (HPD or LPD, respectively). Overall significance was assessed using a Mann-Kendall trend test and between groups significance using a Wilcoxon test. c) Dietary proline is extensively metabolized by enterocytes in the small intestine (~40%). d) Volcano plot of differentially expressed jejunal gene transcripts associated with the consumption of proline identified by limma-voom controlling for age, BMI, sex, education years, kcal intake, and antidepressant medication (pFDR<0.05) in the INTESTINE cohort (n=28). e) Pathway over-representation analysis of the jejunal genes significantly associated with the dietary proline based on Reactome database (qval<0.1). Bars are colored according to the qval and pathway names are colored based on the results of functional enrichment map. f) Over-representation results were mapped as a functional network of pathways using Cytoscape and enrichment map. Nodes represent over-represented pathways. Node size reflects the total number of genes in each pathway. Edge thickness represents the degree of overlapping genes between pathways. Groups of functionally related pathways are circled and labelled. Pathways (nodes) are colored according to the functional group they belong. Nodes with not overlapping are not represented. g) KEGG map of regulation of the GABAergic synapse pathway, with each block representing a group of genes. A white block indicates that no gene in that block was significantly associated with dietary proline, whereas green/red backgrounds indicate significantly down-/up-regulated genes, respectively. Node color reflects log<sub>2</sub> fold change range. Over-expressed genes in each block are highlighted in the boxes below. Data were integrated and visualized using Pathview. h) Scheme of the proline supplementation study design. CFD, crude fiber rich diet; SD, standard diet; UCMS, unpredictable chronic mild stress. i) Immobility time (measured by tail suspension), j) sucrose intake, and l) plasma lipopolysaccharide binding protein (LBP) in the experimental groups. Differences were assessed by one-way ANOVA followed by Tukey or Dunnett's post-hoc test in case of equal or different variances, respectively. A Mann-Kendal trend test was also used to analyze the linear trend in the plasma LBP measure in the SD groups.



**Figure 4**

Associations of brain iron deposition and intrinsic brain networks with depression scores and circulating proline identified using machine learning. a) Partial Spearman's rank correlations of the PHQ-9 scores with relaxometry parameters (mean R2\* and T2\* values) of brain regions from the Anatomical Automatic Labelling brain atlas controlling for age, BMI, sex, education years, antidepressant and anxiety medications in the IRONMET cohort (n=146). b) Barplots of the normalized variable importance measure

(VIM) for the relaxometry measures of brain regions associated with the PHQ-9 scores. c) We extracted the BOLD time series using a resting-state atlas and computed the phase space of the BOLD signal for each brain area by using the Hilbert transform. The phase dynamics can be represented in the complex plane as  $e^{i\varphi}$  (black bold line), the real part as  $\cos(\varphi)$  (black dotted line), and the imaginary part as  $\sin(\varphi)$  (black dotted line). The purple arrows represent the Hilbert phases at each TR. Adapted from 27. d) Intrinsic Ignition Framework. Events were obtained by applying a threshold method, and for each driving event, the activity in the rest of the network (see red stippled area) was obtained in the set 4TR time window (gray area). A binarized phase matrix was extracted from the time window. From this phase matrix, we obtained the integration by calculating the largest subcomponent and computed the mean ignition for each brain area across time. Adapted from 27. e) Barplots of VIM for the mean ignition of brain areas associated with the PHQ-9 scores and f) circulating proline in the IMAGENOMICS cohort (n = 591). Significant variable in the machine learning analyses were identified using a multiple random forest-based variable selection strategy as implemented in the Boruta algorithm with 5000 trees and 500 iterations. g) Venn diagram for the associations of the PHQ-9 and circulating proline. h) Some of these brain areas common to the circulating proline and PHQ-9 scores such as the superior frontal cortex, the precuneus, insula, and subcortical (i.e., the caudate, putamen, and hippocampus) belong to the so-called rich club.



**Figure 5**

An emotional-impaired phenotype was phenocopied to recipient mice in parallel to changes in the expression of mPFC gene transcripts. a) Experimental design for the FMT study. Microbiota from  $n = 20$  human donors with different PHQ-9 scores was delivered to  $n = 20$  recipient mice pre-treated with antibiotics for 14 days. A freezing test was performed after 4 weeks. b) Scatter plot and correlation between the donor's PHQ-9 scores and the recipient's mice freezing test scores controlling for donor's age,

BMI, sex, education years, and antidepressant and anxiety medication. c) Volcano plot of differentially expressed gene transcripts in the medial pre-frontal cortex of the recipient mice associated with the donor's PHQ-9 scores identified by both limma-voom and DESeq2 after adjusting p-values for multiple testing (padj). d) Gene interaction network constructed using differentially expressed mPFC gene transcripts associated with donor's PHQ-9 via the Search Tool for the Retrieval of Interacting Proteins/Genes (STRING) database. The network nodes are genes and the edges represent the predicted functional interactions. The thickness indicates the degree of confidence prediction of the interaction. Functional gene clusters are coloured and annotated based on pathway over-representation analysis results. e) Scatter plots and correlations of selected relevant genes from the functional clusters. f) KEGG-based pathway over-representation analysis of recipient's mice genes associated with the donor's PHQ-9 scores. The bubble size represents the percentage of significant genes contained in each pathway. Bubbles are colored according to the p-value adjusted for multiple testing (q-values). g) Reactome-based pathway over-representation analysis of the recipient's mice genes significantly associated with the donor's PHQ-9 scores ( $q_{val} < 0.1$ ). Bars are colored according to the  $q_{val}$  and pathway names are colored based on the functional clusters identified with STRING. h) Scatter plots and correlations of gene transcripts encoding for GABA and Proline transporters as well as those involved in oxidative phosphorylation.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarytables.xlsx](#)