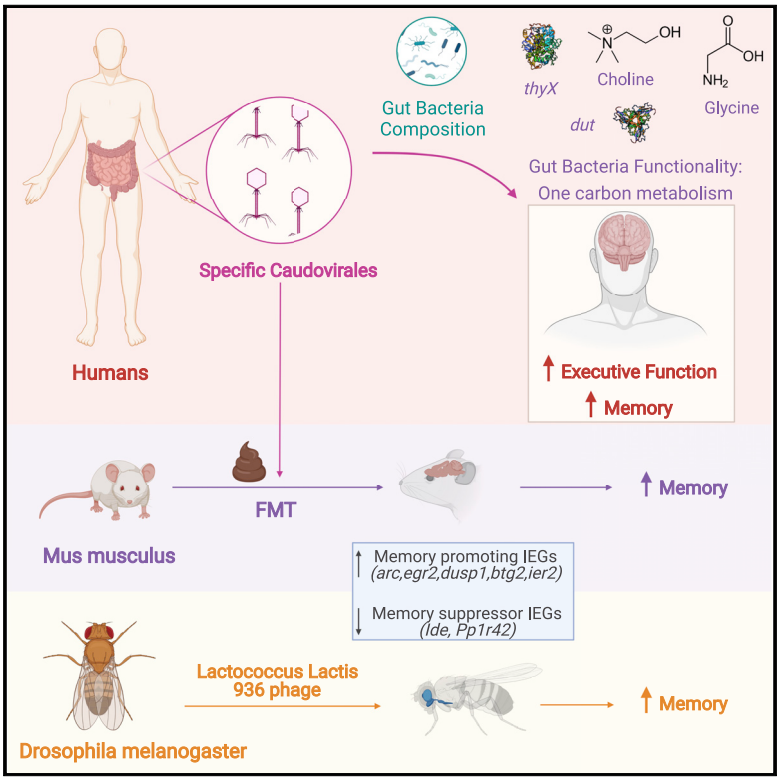
Article

Cell Host & Microbe

Caudovirales bacteriophages are associated with improved executive function and memory in flies, mice, and humans

Graphical abstract



Highlights

- Specific *Caudovirales* are linked to better executive function and memory
- *Caudovirales* are associated with specific bacterial composition and functionality
- High *Caudovirales* FMT increases memory and IEG expression in recipient mice
- *Lactococcal 936* bacteriophage supplementation increases memory in flies

Authors

Jordi Mayneris-Perxachs,
Anna Castells-Nobau,
María Arnoriaga-Rodríguez, ...,
Manuel Martínez-García,
Rafael Maldonado,
José-Manuel Fernández-Real

Correspondence

jmayneris@idibgi.org (J.M.-P.),
rafael.maldonado@upf.edu (R.M.),
jmfreal@idibgi.org (J.-M.F.-R.)

In brief

Here, Mayneris-Perxachs et al. reveal an association of the dominant bacteriophages with human host cognition in parallel to a specific bacterial composition and functionality. Microbiota transplantation and bacteriophage supplementation increased the memory of recipient mice and flies through the upregulation of memory-promoting genes. This highlights the potential of targeting bacteriophages cognitive improvement.

Mayneris-Perxachs et al., 2022, Cell Host & Microbe *30*, 340–356
March 9, 2022
https://doi.org/10.1016/j.chom.2022.01.013

CellPress

CellPress
OPEN ACCESS

Cell Host & Microbe

Article

Caudovirales bacteriophages are associated with improved executive function and memory in flies, mice, and humans

Jordi Mayneris-Perxachs,[1,2,3,21,22,*] Anna Castells-Nobau,[1,2,3,22] María Arnoriaga-Rodríguez,[1,2,3,4] Josep Garre-Olmo,[5,18] Josep Puig,[4,6,7,8] Rafael Ramos,[4,9,10] Francisco Martínez-Hernández,[11] Aurelijus Burokas,[12,19] Clàudia Coll,[13] José Maria Moreno-Navarrete,[1,2,3,4] Cristina Zapata-Tona,[1,2,3,4] Salvador Pedraza,[4,7,8] Vicente Pérez-Brocal,[14,15] Lluís Ramió-Torrentà,[4,13,20] Wifredo Ricart,[1,2,3,4] Andrés Moya,[14,15,16] Manuel Martínez-García,[11] Rafael Maldonado,[12,17,*] and José-Manuel Fernández-Real[1,2,3,4,21,23,*]

[1]Department of Diabetes, Endocrinology, and Nutrition, Dr. Josep Trueta University Hospital, Girona, Spain
[2]Nutrition, Eumetabolism, and Health Group, Girona Biomedical Research Institute (IdibGi), Girona, Spain
[3]Centro de Investigación Biomédica en Red Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Madrid, Spain
[4]Department of Medical Sciences, School of Medicine, University of Girona, Girona, Spain
[5]Research Group on Aging, Disability, and Health, Girona Biomedical Research Institute (IdibGi), Girona, Spain
[6]Institute of Diagnostic Imaging (IDI)-Research Unit (IDIR), Parc Sanitari Pere Virgili, Barcelona, Spain
[7]Medical Imaging, Girona Biomedical Research Institute (IdibGi), Girona, Spain
[8]Department of Radiology (IDI), Dr. Josep Trueta University Hospital, Girona, Spain
[9]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 Gorina-IDIAPJGol), Girona Biomedical Research Institute, (IDIBGI), Dr. Josep Trueta University Hospital, Catalonia, Spain
[10]Girona Biomedical Research Institute (IDIBGI), Dr. Josep Trueta University Hospital, Catalonia, Spain
[11]Department of Physiology, Genetics, and Microbiology, University of Alicante, Alicante, Spain
[12]Laboratory of Neuropharmacology, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain
[13]Neuroimmunology and Multiple Sclerosis Unit, Department of Neurology, Dr. Josep Trueta University Hospital, Girona, Spain
[14]Area of Genomics and Health, Foundation for the Promotion of Sanitary and Biomedical Research of Valencia Region (FISABIO-Public Health), Valencia, Spain
[15]Biomedical Research Networking Center for Epidemiology and Public Health (CIBERESP), Madrid, Spain
[16]Institute for Integrative Systems Biology (I2SysBio), University of Valencia and Spanish National Research Council (CSIC), Valencia, Spain
[17]Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain
[18]Serra-Hunter Fellow. Department of Nursing, University of Girona, Girona, Spain
[19]Department of Biological Models, Institute of Biochemistry, Life Sciences Center, Vilnius University, Vilnius, Lithuania
[20]Neurodegeneration and Neuroinflammation research group. Girona Biomedical Research Institute (IdibGi), Girona, Spain
[21]These authors contributed equally
[22]These authors contributed equally
[23]Lead contact
*Correspondence: jmayneris@idibgi.org (J.M.-P.), rafael.maldonado@upf.edu (R.M.), jmfreal@idibgi.org (J.-M.F.-R.)
https://doi.org/10.1016/j.chom.2022.01.013

SUMMARY

Growing evidence implicates the gut microbiome in cognition. Viruses, the most abundant life entities on the planet, are a commonly overlooked component of the gut virome, dominated by the *Caudovirales* and *Microviridae* bacteriophages. Here, we show in a discovery (n = 114) and a validation cohort (n = 942) that subjects with increased *Caudovirales* and *Siphoviridae* levels in the gut microbiome had better performance in executive processes and verbal memory. Conversely, increased *Microviridae* levels were linked to a greater impairment in executive abilities. Microbiota transplantation from human donors with increased specific *Caudovirales* (>90% from the *Siphoviridae* family) levels led to increased scores in the novel object recognition test in mice and up-regulated memory-promoting immediate early genes in the prefrontal cortex. Supplementation of the *Drosophila* diet with the 936 group of lactococcal *Siphoviridae* bacteriophages resulted in increased memory scores and upregulation of memory-involved brain genes. Thus, bacteriophages warrant consideration as novel actors in the microbiome-brain axis.

INTRODUCTION

Increasing clinical and pre-clinical studies implicate the gut microbiome as a key player in the regulation of neurodegenerative processes, the modulation of cognition, and neurological disorders (Cryan et al., 2020; Morais et al., 2021). A large amount of this evidence comes from studies of bacteria (Morais et al., 2021). However, emerging evidence suggests that viruses can

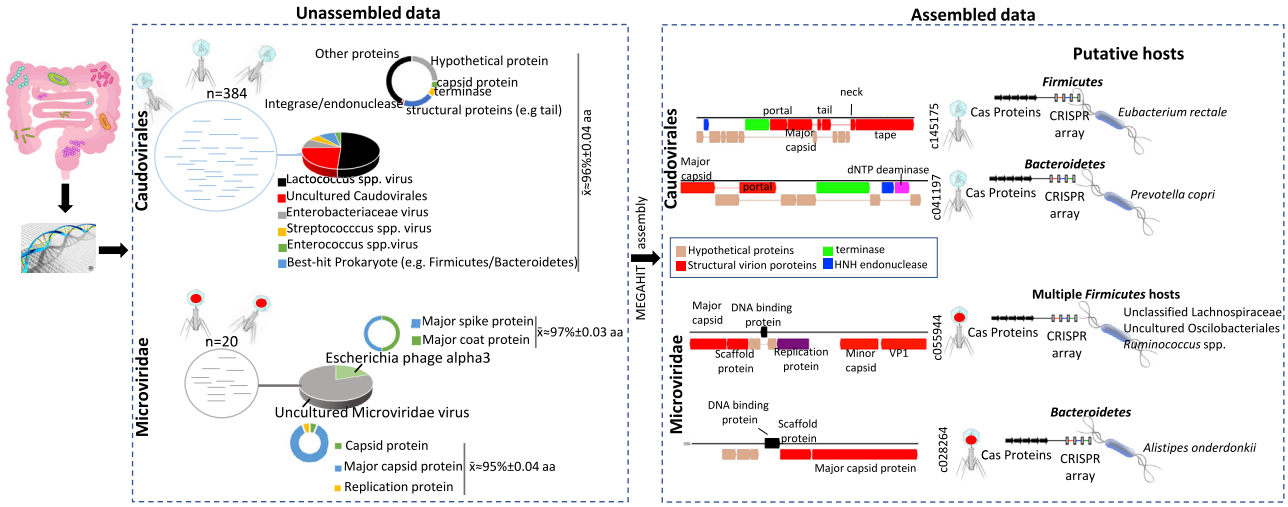


Figure 1. Gene and genomic features of the gut phageome associated with cognitive function

Left panel shows gene annotation based on the best hit score against the nr database of NCBI from unassembled viral data (n = 384 for *Caudovirales* and n = 20 for *Microviridae*) previously identified by Kaiju that correlated with cognitive function. Mean amino acid similarity is indicated along with standard deviation error. Right panel shows the genomic feature of some representative viruses that were further assembled by Megahit software. Virus-host assignment was performed based on CRISPR spacer match, best k-mer frequency signature score of corresponding virus-host pair, and gene/protein similarity against nr NCBI database. For more details on genome size, genome sequence, number of ORFs, and other genomic features, refer to Table S2.

also profoundly affect host physiology and disease (Keen and Dantas, 2018; Mirzaei and Maurice, 2017). The most numerous constituents of the human virome are bacteriophages. Temperate (lysogenic) bacteriophages can transfer genes to their bacterial hosts, thereby modulating bacterial gene expression and altering their phenotype. In fact, more than 80% of the bacterial genomes harbor prophages. Hence, bacteriophages may play an important role in shaping bacterial diversity and function and, thus, human health (Keen and Dantas, 2018; Mirzaei and Maurice, 2017).

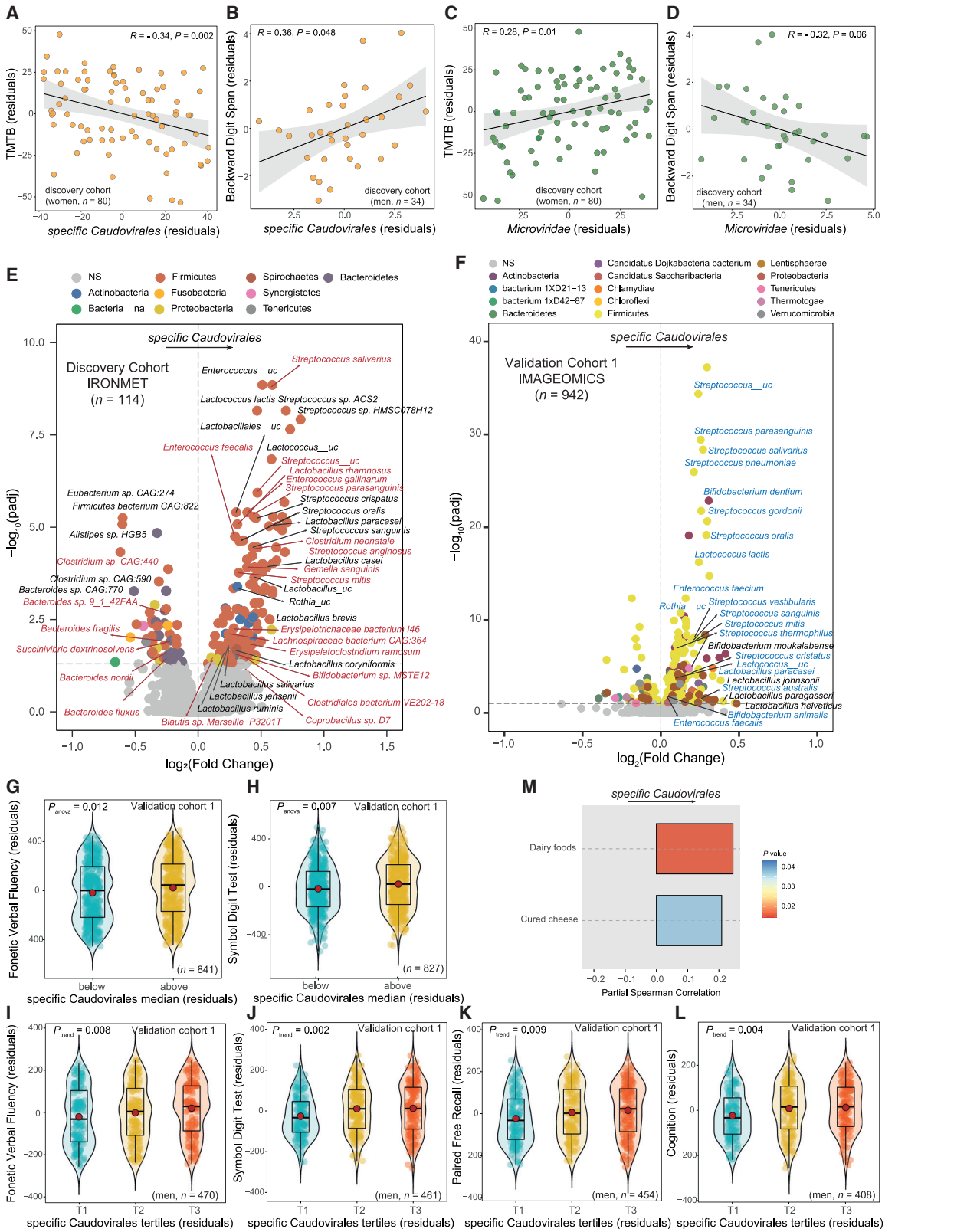
Gut bacteriophage communities display large interindividual variation and temporal stability and are dominated by temperate *Caudovirales* and virulent (lytic) *Microviridae* (Shkoporov et al., 2019). Historically, phages have been classified according to their morphology. However, the advances in the sequencing of phage genomes during this century revealed a high genomic diversity, particularly in the *Caudovirales* order. Hence, in March 2021, the International Committee on Taxonomy of Viruses (ICTV) changed the virus taxonomy by defining new genome-based families (Walker et al., 2021). The order *Caudovirales*, unifying all tailed phages, traditionally contained three families: *Myoviridae*, *Podoviridae*, and *Siphoviridae*. Within this genome-based taxonomy, the siphoviruses have been divided into three families (*Demerecviridae*, *Drexlerviridae*, and *Siphoviridae*). It is worth nothing that the old *Siphoviridae* family is the one with fewer changes: from the 1,371 species included in the old classification, 1,166 (85%) were still classified as *Siphoviridae* in the new genome-based taxonomy.

Due to the recent change in taxonomy, the evidence so far is based only on the old taxonomy. The levels of phages in the *Microviridae* family increase with age and negatively correlate with the relative abundance of phages in the *Siphoviridae* family (Gregory et al., 2020; Lim et al., 2015). In addition, diet has been shown to alter the composition of the gut virome. Specifically, transition from normal chow to high fat diet resulted in a significant decrease in bacteriophages from the *Siphoviridae* family, accompanied by gains in *Microviridae* (Schulfer et al., 2020). The existence of a healthy human gut phageome, dominated by temperate bacteriophages, has been suggested (Manrique et al., 2016; Moreno-Gallego et al., 2019), where phage dysbiosis, characterized by increases in lytic phages and/or activated prophages, has been associated with the disease (Manrique et al., 2017).

RESULTS AND DISCUSSION

Bacteriophages in the gut microbiome are associated with executive function

Despite being the most abundant biological entities on the planet and having the potential to modulate bacterial communities, bacteriophages, in general, and their relationship with cognition, in particular, remain largely unexplored. Therefore, we aimed to study the interplay among the gut phageome, the gut bacteriome, and executive function (one of the six key domains of cognition), combining fecal shotgun metagenomics and metabolomics in 4 human cohorts together with fecal microbiota transplantation (FMT) in mice and bacteriophage supplementation in *Drosophila melanogaster*. We first assessed these relationships in a longitudinal discovery cohort (IRONMET, n = 114, Table S1A) that underwent a battery of cognitive tests. In all analyses shown below, we controlled for the following confounding factors: age, BMI, sex, and years of education. To account for the compositional nature of the microbiome data, we applied a centred log-ratio transformation to the raw counts from the discovery cohort using the geometric mean of all species as the reference frame. We observed that the levels of specific *Caudovirales* (the former *Siphoviridae* family with the old taxonomy containing the new *Demerecviridae*, *Drexlerviridae*, and



(legend on next page)

Siphoviridae families) (Figure 1) were negatively associated with the trail making test B (TMTB) time (R = −0.22, p = 0.037), whereas *Microviridae* levels had a positive association (R = 0.19, p = 0.047). In the TMTB, higher times to completion indicate worse executive functioning. As sexual dimorphism has a strong impact on the gut microbiome, we also performed sex-specific analyses (Mayneris-Perxachs et al., 2020). Women with higher specific *Caudovirales* levels had lower scores in the TMTB (Figure 2A), implying a better performance in central executive processes. Although no associations with TMTB were found in men (R = 0.03, p = 0.985), specific *Caudovirales* levels were also associated with better executive function measured by the backward digit span test (Figure 2B). Higher scores in the backward digit span tests indicate better performance in executive function. No associations were observed between specific *Caudovirales* levels and the backward digit span test in women (R = 0.02, p = 0.998).

Gene and genomic analysis of these *Caudovirales* from unassembled and assembled data indicated that most of them were uncultured and uncharacterized, while others putatively infected mostly *Lactococcus* spp., and other gut bacteria belonging to *Enterobacteriacaea*, *Firmicutes* (e.g., *Eubacterium rectale*), or *Bacteroidetes* (*Prevotella copri*). Host assignment was carried out by a combination of CRISPR spacer-protospacer match search, estimation of k-mer frequencies of host and virus genome, and gene/protein search in databases (Table S2). Gene content and annotation show common gene features of these *Caudovirales* with high predominance of hypothetical proteins along with structural proteins (capsid, portal, neck, and tail) and other functional proteins of *Caudovirales* (e.g., terminases) (Figure 1). Metagenomic assembly delivered a fragmented genome assembly for some of these *Caudovirales* (3.5–42.9-kb genome size; Table S2), especially for Lactococcus viruses (<3.5-kb mean genome assembly) identified to be correlated with better performance in central executive processes. The very high similarity of predicted proteins and viral hallmark genes (mean amino acid similarity $\geq$ 95%) identified by Virsorter 2.0 from these *Caudovirales* was paramount to ascertain the identity of these viruses despite obtaining a fragmented metagenome assembly.

These specific *Caudovirales* belong to the three described families (*Siphoviridae*, *Demerecviridae*, and *Drexlerviridae*) that comprised the former *Siphoviridae* family with the old taxonomy. As in our data, the new genome-based *Siphoviridae* family still contained >86% of the genera included in the old *Siphoviridae* family, we also found significant negative associations between *Siphoviridae* levels (as per new genome-based taxonomy) and the TMTB (Figure S1). Remarkably, unlike specific *Caudovirales* levels, *Siphoviridae* levels were also significantly associated with better performance with inhibitory control (measured by the Stroop test) and both short- and long-term memory (Figures S1C–S1E), highlighting a particular role of *Siphoviridae* in cognitive function. These results were also found in women (Figures S1F–S1J), whereas in men, we only found positive associations with the backward digit span test (Figures S1K and S1L).

Conversely, some detected ssDNA *Microviridae* virus (Figure 1; Table S2) levels were associated with a greater impairment in executive function in both women and men (Figures 2C and 2D). In line with previous observations showing an increase after a high fat diet (Schulfer et al., 2020), *Microviridae* levels correlated positively with fat mass (Figure S2A). *Microviridae* hallmark genes and proteins were clearly identified in unassembled and assembled data (Figure 1). Some of them were similar to Escherichia phage alpha3 and to uncultured *Microviridae* detected previously in the gut. Remarkably, identification of putative hosts indicated that some of these *Microviridae* infect *Bacteroidetes* (likely *Alistipes onderdonkii*), and in particular, one *Microviridae* virus (contig name c055944) showed a broad host range because CRISPR spacers from *Ruminococcus* spp., Oscilobacteriales, and Lachnospiracea matched viral protospacers of this virus (Table S2).

It has been suggested that bacteriophages play an important role in host health and disease by modulating bacterial communities through transposition and induction and horizontal gene transfer (Keen and Dantas, 2018). Therefore, we then evaluated the associations of these bacteriophages with the bacterial composition and functionality. Specific *Caudovirales* levels were strongly positively associated with lactic acid bacteria (*Lactobacillales* order), particularly with *Streptococcus*, *Lactobacillus*, *Lactococcus*, and *Enterococcus* species, and negatively associated with species from the genus *Bacteroides* (Figure 2E; Table S3A). This is consistent with the fact that all known lactic acid bacteria phages belong to the *Caudovirales* order and

Figure 2. The gut phageome is associated with cognitive function and bacterial composition

(A–D) Scatter plots of the partial Spearman's rank correlations (adjusted for age, BMI, and education years) between the fecal centered log-ratio (clr) transformed bacteriophage values and executive function assessed by the trail making test B (TMTB) and the digit backward span test in men and women from the discovery cohort (IRONMET, n = 114). The ranked residuals are plotted.

(E and F) (E) Volcano plots of differential bacterial abundance associated with the fecal clr-transformed specific *Caudovirales* [*Siphoviridae* (>90%), *Demerecviridae*, and *Drexlerviridae*] in the discovery cohort and (F) the validation cohort 1 (IMAGEOMICS, n = 942) identified by DESeq2 analysis. The x axis represents the logarithm to the base 2 ($\log_2$) of the fold change associated with a unit change in the clr-transformed specific *Caudovirales* values. The y axis represents the logarithm to the base 10 ($\log_{10}$) of the p values adjusted for multiple testing (padj). Each dot represents a specific taxon. Significantly different taxa (padj < 0.05) are colored according to the phylum. Taxa highlighted in red were also associated with the TMTB scores. Taxa highlighted in blue were also associated with the clr-transformed specific *Caudovirales* levels in the discovery cohort.

(G and H) (G) Violin plots of the Fonetic verbal fluency (n = 841) and (H) symbol digit test scores (n = 827) according to the clr-transformed specific *Caudovirales* median after adjusting age, BMI, gender, education years, intake of medications with central nervous systems' side effects (ATC code N), physical activity (METS/min/week), depression scores, and glucose and cholesterol levels in the validation cohort 1 (IMAGEOMICS, n = 942). Red dots represent the mean.

(I–L) (I) Violin plots of the Fonetic verbal fluency (n = 470), (J) symbol digit test (n = 461), (K) cognition (n = 408), and (L) paired free recall (n = 454) scores according the clr-transformed specific *Caudovirales* tertiles (T1, T2, and T3) after adjusting for the previous covariates in men from the validation cohort 1 (IMAGEOMICS, n = 942). Overall significance was assessed using a Mann-Kendall trend test.

(M) Partial Spearman's rank correlation among consumption of dairy food items derived from food frequency questionnaires and the clr-transformed specific *Caudovirales* values.

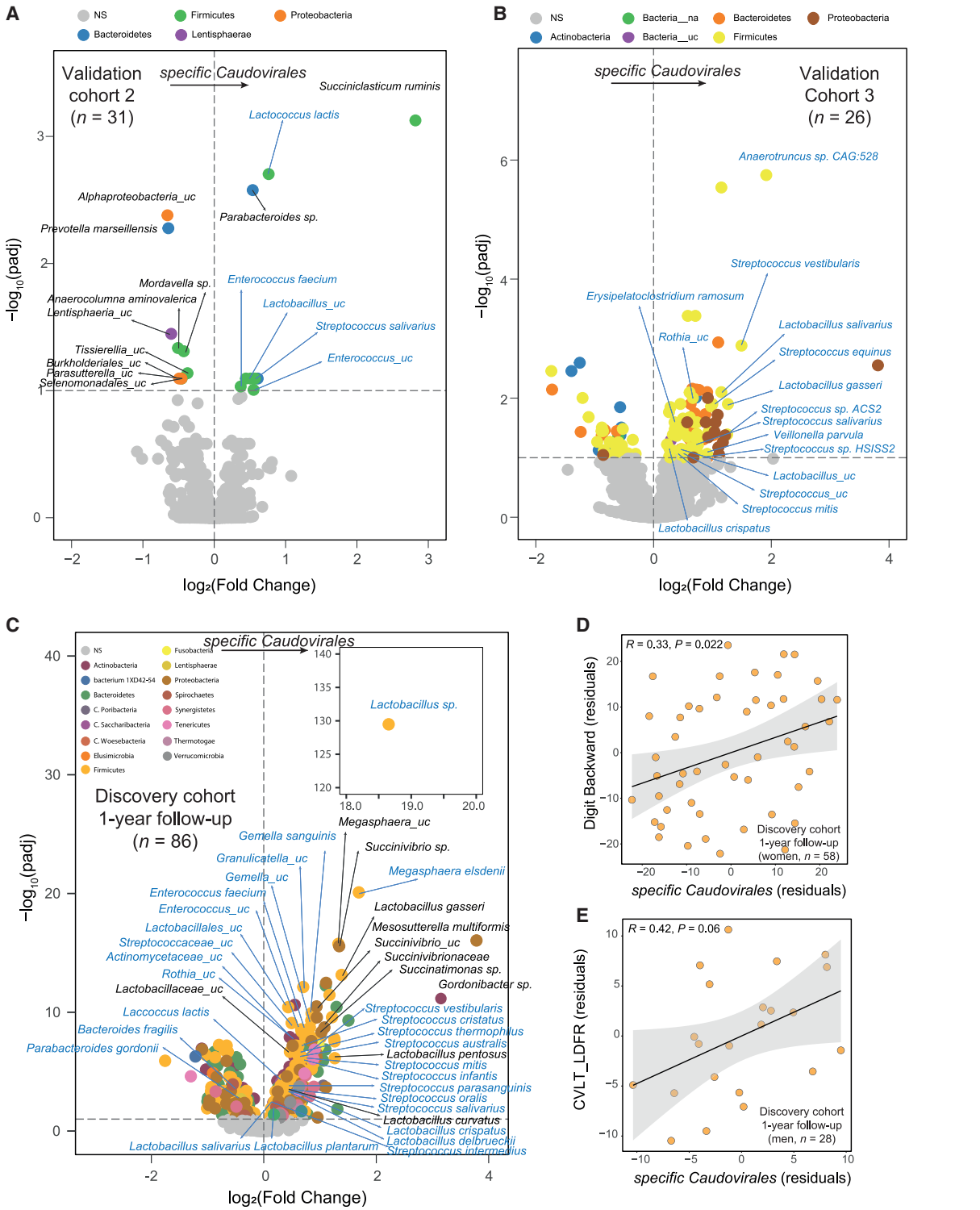


Figure 3. The gut phageome is also longitudinally associated with cognitive function

(A–C) (A) Volcano plots of differential bacterial abundance associated with the fecal clr-transformed specific *Caudovirales* [*Siphoviridae* (>86%), *Demerecviridae*, and *Drexlerviridae*] values in the validation cohort 2 (n = 31), (B) the validation cohort 3 (n = 26), and (C) 1-year follow-up in the discovery cohort (n = 86) after controlling for age, BMI, sex, and education years. The x axis represents the logarithm to the base 2 ($\log_2$) of the fold change associated with a unit change in the clr-transformed specific *Caudovirales* values. The y axis represents the logarithm to the base 10 ($\log_{10}$) of the p values adjusted for multiple testing (padj). Each dot represents a specific taxon. Significantly different taxa (padj < 0.1) are colored according to the phylum. Bacterial species highlighted in blue were also associated with the clr-transformed specific *Caudovirales* levels in the discovery cohort.

(legend continued on next page)

most of them to the *Siphoviridae* family (Murphy et al., 2017). Notably, ~40% of the species most strongly linked with specific *Caudovirales* were also significantly associated with the TMTB scores (highlighted in red in Figure 2E). As stated above, more than 86% of the specific *Caudovirales* in the discovery cohort belonged to the *Siphoviridae* family. Therefore, we found similar results when we evaluated the associations of the *Siphoviridae* family with the bacterial composition (Figure S1M). Conversely, *Microviridae* levels were negatively associated with several *Lactobacillus*, *Streptococcus*, and *Enterococcus*, while positively linked to *Bacteroides* and *Prevotella* species (Figure S2B; Table S3B). This is consistent with a recent analysis of CRISPR spacer sequence matches predicting that *Microviridae* bacteriophages infect highly predominant gut bacterial taxa, such as *Bacteroides* and *Prevotella* (Shkoporov et al., 2019).

We next sought to validate these results in an independent validation cohort (IMAGEOMICS, n = 942, Table S1B) (Puig et al., 2020). In this cohort, >94% of the specific *Caudovirales* species (*Siphoviridae*, *Demerecviridae*, and *Drexlerviridae* families) belonged to the *Siphoviridae* family. Consistent with the previous findings, specific *Caudovirales* levels were positively associated with lactic acid bacteria (Figure 2F; Table S3C). Notably, the bacterial species most associated with this bacteriophage family in the validation cohort were the same as those identified in the discovery cohort (highlighted in blue in Figure 2F). Not only that, but in line with the results in the discovery cohort, subjects with clr-transformed specific *Caudovirales* levels above the median had higher scores both in measures of frontal cortex activity such as executive function (Phonemic verbal fluency, Figure 2G) (Herrmann et al., 2017) and information processing speed (symbol digit test, Figure 2H) (Silva et al., 2019), after controlling for age, BMI, gender, education years, intake of medications with central nervous systems' side effects (ATC code N), physical activity (metabolic equivalent of tasks(METS)/min/week), depression scores (PHQ-9), and glucose and cholesterol levels in the validation cohort. In men, specific *Caudovirales* levels were not only associated with executive function and information processing speed (Figures 2I and 2J) but also with short- and long-term verbal memory (Paired, Delayed, and Total Free Recall; Figures 2K and S3A–S3C) and general cognition (Figure 2L). No significant associations were found for these cognitive tests in the case of women ($p > 0.05$). We further validated the associations among specific *Caudovirales* and lactic acid bacteria in two additional independent validation cohorts (Figures 3A and 3B; Tables S1C, S1D, S3D, and S3E).

These results were also replicated after 1-year follow-up in the discovery cohort (IRONMET, n = 86, Table S1E). Hence, once again, we found a positive strong association between *Streptococcus*, *Lactobacillus*, *Lactococcus*, and *Enterococcus* species and specific *Caudovirales* levels after 1 year (Figure 3C; Table S3F). The positive association between specific *Caudovirales* levels and improved executive functioning was also mirrored after 1-year follow-up (Figure 3D). In line with the findings in the validation cohort (IMAGEOMICS), specific *Caudovirales* levels in men from the discovery cohort also had a strong positive association with long-term verbal memory after 1-year follow-up (Figure 3E). No significant associations were found between specific *Caudovirales* levels and long-term memory (California Verbal Learning Test (CVLT) Long Delayed Free Recall: $R = -0.11$, $p = 0.460$; CVLT Long Delayed Cued Recall: $R = -0.14$, $p = 0.343$) in women after 1-year follow-up. *Microviridae* associations were also reproduced one year later (Figure S2C; Table S3G). Remarkably, in three out of four cohorts, we found a consistent positive association among specific *Caudovirales* levels and *Lactococcus lactis* and several *Lactobacillus* (*L. crispatus*, *L. plantarum*, *L. Salivarius*, and *Lactobacillus_uc*) and *Streptococcus* (*S. mitis*, *S. salivarius*, *S. vestibularis*, and *Streptococcus_uc*) species. *L. lactis* and *Lactobacillus* sp. are widely used in the fermentation of dairy products (Murphy et al., 2017), while *S. salivarius* and *S. mitis* are the predominant streptococcal species in human milk microbiota (Martín et al., 2016). Consistently, we found positive associations between specific *Caudovirales* levels and dairy product consumption (Figure 2M) and medium-chain fatty acids (Figure S3D), which are naturally occurring in dairy fat, as well as between specific *Caudovirales*-linked lactic acid bacteria and dairy products (Figure S3E). On the contrary, the *Microviridae* family had negative associations with medium-chain fatty acids (Figure S3F). Interestingly, supplementation with medium-chain fatty acids has shown to improve synaptic plasticity and cognitive function in mice and humans (Page et al., 2009; Wang and Mitchell, 2016).

Bacterial functions are linked to the gut bacteriophages

Functional analyses based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway also revealed significant associations between bacterial pathways, bacteriophages, and human host executive function. Specifically, specific *Caudovirales* levels were strongly negatively associated with folate-mediated one-carbon metabolism (Figure 4A). Folate metabolism is central to multiple physiological processes, providing 1C units required for cellular processes (Ducker and Rabinowitz, 2017). Folate-activated 1Cs are required for the nucleotide synthesis pathway, which is essential for DNA replication and repair. It is also involved in the methionine cycle, which is necessary for the synthesis of S-adenosylmethionine (SAM), the universal methyl donor to numerous methylation reactions, including DNA methylation. Through the trans-sulfuration pathway, it regulates redox defence by synthesizing antioxidants such as taurine and glutathione from cysteine. Remarkably, all these bacterial pathways were also strongly associated with the specific *Caudovirales* levels and inversely with *Microviridae* (Figures 4A and 4B). We also found negative associations among the metabolism of vitamins B2 and B6, which are essential cofactors in the folate cycle, and specific *Caudovirales*. Functional analyses at the enzyme level revealed that bacterial genes *thyX* and *dut*, both involved in folate-mediated pyrimidine biosynthesis, had the strongest associations with specific *Caudovirales* levels

(D) Scatter plot of the partial Spearman's rank correlation between the specific *Caudovirales* clr-transformed values and the digit backward span test in women from the discovery cohort after 1-year follow-up.
(E) Scatter plots of the partial Spearman's rank correlations between clr-transformed specific *Caudovirales* values and the California Verbal Learning Test-Long Delayed Free Recall (CVLT_LDFR) in men from the discovery cohort after 1-year follow-up.

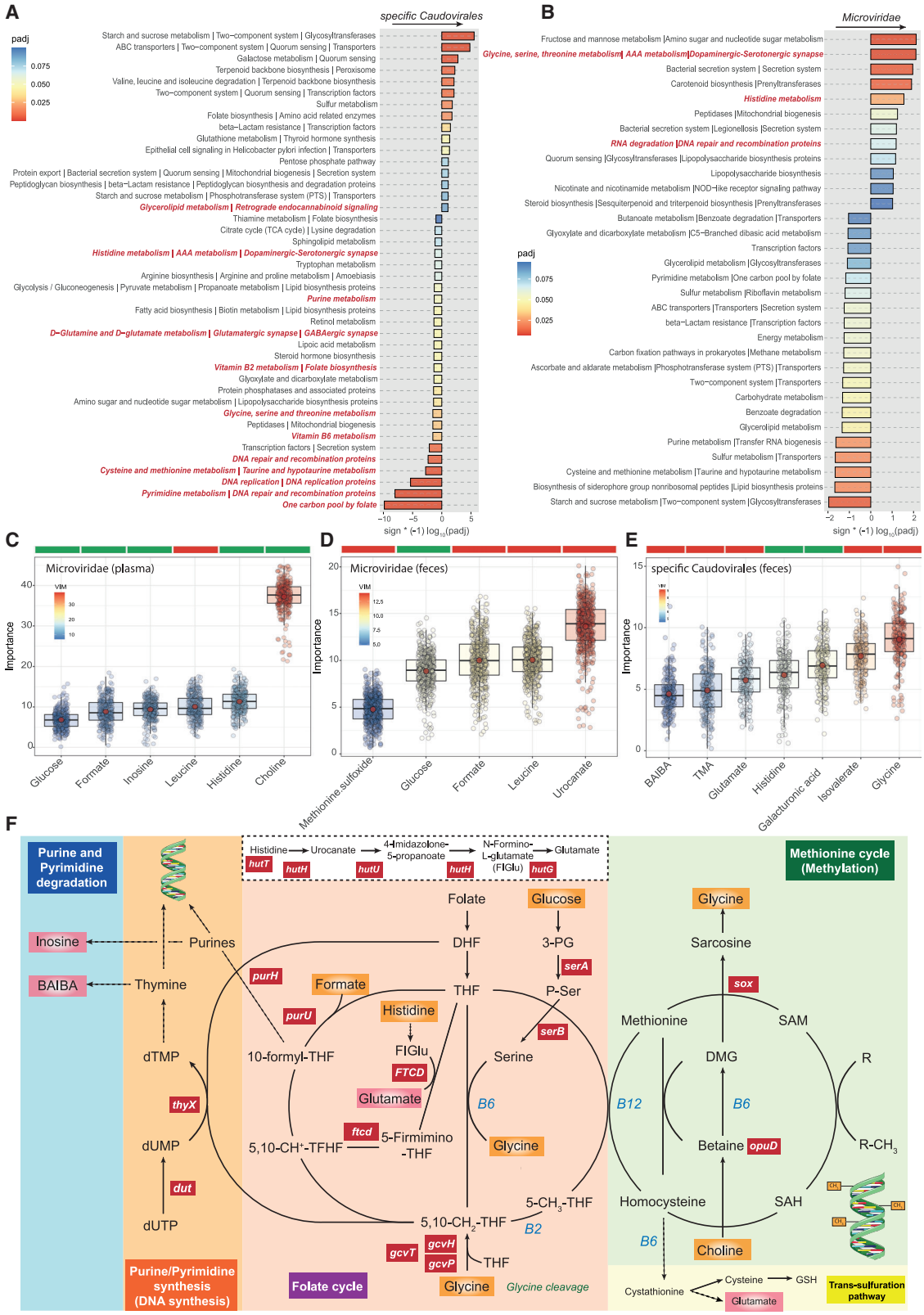


Figure 4. The gut phageome is linked to bacterial pathways involved in the folate-mediate one-carbon metabolism and related metabolites

(A and B) (A) Manhattan-like plot of significantly expressed KEGG bacterial pathways associated with the specific *Caudovirales* [*Siphoviridae* (>86%), *Demerecviridae*, and *Drexlerviridae*] and (B) *Microviridae* clr-transformed values identified by DESeq2 analysis controlling age, BMI, sex, and educations years. Bars are

(legend continued on next page)

(Figure S4A; Table S3H). Specifically, *thyX* encodes for thymidylate synthase (*TYMS* in humans). Decreased *TYMS* expression levels lead to imbalances between DNA synthesis and methylation, which is essential for neurodevelopment, synaptic plasticity, and memory (Heyward and Sweatt, 2015). In fact, impairment in folate-mediated one-carbon metabolism has been associated with neurodegenerative disease that may result from dTMP synthesis impairment and consequent uracil misincorporation into DNA (Blount et al., 1997; Ducker and Rabinowitz, 2017). Specific *Caudovirales* levels were also negatively associated with other pathways that play a key role in the central nervous system such as glutamatergic, GABAergic, dopaminergic, serotonergic synapse, and retrograde endocannabinoid signaling. Other strongly associated bacterial genes were linked to folate-mediated histidine catabolism (*FTCD*, *ftcd*) and purine biosynthesis (*purH*, *purU*) (Figure S4A).

Human plasma and fecal metabolites reflect bacterial functions modulated by the bacteriophages

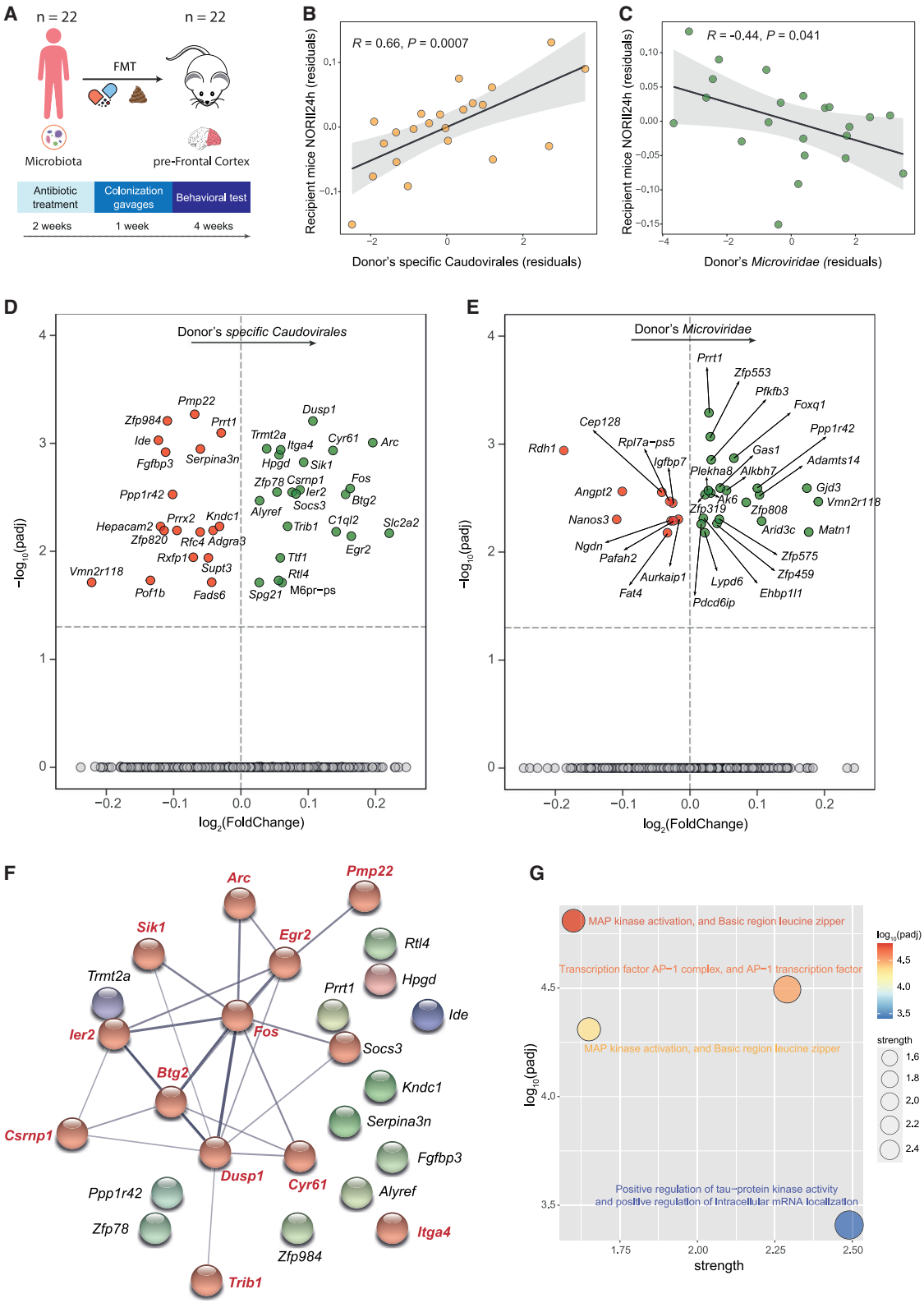
We next performed metabolic profiling of plasma and feces by ^{1}H NMR to reveal metabolic signatures associated with the bacteriophages. Applying a machine learning variable selection strategy based on random forest (Kursa and Rudnicki, 2010), we identified several metabolites linked to the *Microviridae* and specific *Caudovirales* levels (Figures 4C and 4E). Notably, most of these metabolites were directly involved in one-carbon metabolism. They included metabolites that feed 1C units to the folate pool (choline, glycine, formate, histidine, and glucose) and related catabolites (urocanate, glutamate, inosine, BAIBA, and methionine sulfoxide). Particularly, choline and glycine, which are the most important source of folate 1C units (Ducker and Rabinowitz, 2017), had the strongest associations with *Microviridae* and specific *Caudovirales* levels, respectively. In line with these results, glycine and histidine bacterial metabolic pathways were also negatively and positively associated with specific *Caudovirales* and *Microviridae* levels, respectively (Figures 4A and 4B). Specifically, we identified strong associations among bacteriophages and several genes encoding for enzymes involved in glycine and histidine metabolism and transport (Figures 4F and S4B–S4E). Glycine can be obtained from catabolism of dietary choline and serine that feed carbon units to the 1C-metabolism. Serine can also be synthesized from 3-phosphorylglycerate and intermediate in glycolysis. Glycine itself is also a 1C source through the glycine cleavage system (GCS) that produces a carbon unit for the methylation of Tetrahydrofolate (THF). Consistently, bacteriophage levels were associated with genes participating in the GCS (*gcvH*, *gcvP*, and *gcvR*), serine synthesis (*serB* and *serA*), and choline transport and catabolism (*sox* and *opuD*). Of note, GCS genes had the strongest negative association with specific *Caudovirales* levels, whereas the GCS transcriptional repressor (*gcvR*) was positively associated with the *Microviridae* family (Figures S4D and S4E). Mutations in genes encoding the GCS have shown to predispose to neural tube defects and neurological dysfunction both in mice and humans (Kure et al., 2006; Narisawa et al., 2012).

Gut bacteriophages induce changes in cognitive function and brain gene expression of recipient mice after fecal transplantation

To test whether bacteriophages and associated bacterial communities from human donors could trigger cognitive changes in recipient mice, we next performed a FMT experiment. Microbiota from 22 patients was delivered to antibiotic-treated recipient mice (Figure 5A). After FMT, we found that from the 1,994 taxa in the donors' microbiota, 1,385 (70%) were also present in recipient mice microbiota, indicating a favorable engraftment (Figure S5A). Importantly, 216 from these 1,385 taxa were present in the mice receiving FMT but not in control mice treated only with antibiotics but no FMT. In fact, a PCA score plot on the clr-transformed data revealed a clear difference between the microbiota of mice receiving FMT vs. mice not receiving FMT (Figure S5B). After 4 weeks, we observed a kind of dose-response effect according to the specific *Caudovirales* levels present in the donor's microbiome (Figure 5B): the higher the specific *Caudovirales* levels, the higher the scores in the novel object recognition (NOR) test, which is used to evaluate cognition, particularly memory. In line with this result, the recipient mice clr-transformed specific *Caudovirales* levels were also associated with a better performance in the recipient mice short-term memory (positive association with the NOR) and emotional memory (negative association with the freezing time in the cue-induced fear conditioning) (Figures S5C and S5D). On the contrary, increased *Microviridae* levels in donor's microbiome were associated with impaired cognition of recipient mice (Figure 5C). We also tested whether FMT could impact the recipient's mice brain transcriptome. An RNA sequencing of the recipient mice prefrontal cortex (PFC), which is implicated in executive functions and memory, identified 23 and 18 out of 15,547 genes up- and down-regulated associated with the donor's specific *Caudovirales* levels, respectively (Figure 5D). Donor's *Microviridae* levels were associated with 18 and 10 up- and down-regulated genes, respectively (Figure 5E). Notably, several of the most up-regulated gene transcripts with increased donor's specific *Caudovirales* levels are known memory-promoting genes (i.e., *Arc*, *Fos*, *Egr2*, and *Btg2*), whereas those down-regulated (*Ide* and *Ppp1r42*) are memory suppressors (Poon et al., 2020).

To gain better insight into bacteriophages' molecular mechanisms associated with cognition, significant genes were used to construct a gene-gene interaction network using the STRING database (Szklarczyk et al., 2019). It revealed a cluster of highly connected genes mostly up-regulated with the specific

colored according to the p value adjusted for multiple testing (padj). The x axis represents the $-\log_{10}$(padj) values multiplied by the fold-change sign to take into account the direction of the association. Those pathways related to folate-mediated one-carbon metabolism and central nervous system highlighted in bold red.
(C–E) Barplots of normalized variable importance measure (VIM) for the plasma and fecal metabolites associated with the bacteriophages clr-transformed values identified in the discovery cohort using a multiple random forest-based machine learning variable selection strategy using the Boruta algorithm. The above bar represents the sign of the correlation, with green indicating positive association and red negative association.
(F) Overview of the one-carbon metabolism, including the methionine cycle, the folate cycle, and the purine and pyrimidine synthesis. Metabolites and bacterial functions significantly associated with bacteriophages' clr-transformed values are highlighted in bold in orange and red boxes, respectively.



(legend on next page)

Caudovirales that included *Arc*, *Dusp1*, *Fos*, *Btg2*, *Egr2*, *Cyr61*, *Pmp22*, *Sik1*, *Csrnp1*, *Ier2*, and *Trib1* (Figure 5F). Genes from this cluster were involved in positive regulation of tau protein kinase activity, MAP kinase activation, and AP-1 transcription factor network (Figure 5G). The vast majority of these transcripts are encoded by immediate early genes (IEGs), crucially involved in synaptic plasticity and neuronal development (Alberini, 2009; Minatohara et al., 2015). In line with our results, both synaptic plasticity and induction of IEGs depend upon activation of AP-1 transcription factor and MAPK/ERK signaling cascade (Fowler et al., 2011). In addition, environmental enrichment in transgenic mice markedly reduced amyloid-β levels and deposition in parallel to up-regulated hippocampal expression of IEGs, including *Fos*, *Arc*, *Egr2*, *Dusp1*, *Btg2*, and *Ier2* (Lazarov et al., 2005).

We further complemented these analyses by performing over-representation analyses using the consensus pathway database (Kamburov et al., 2013). Analysis based on gene ontology highlighted cognition as the most over-represented biological process associated with donor's specific *Caudovirales* levels (Figure 6A). Other relevant biological processes included associative learning and neuron differentiation and development, neurogenesis, myelination, and nervous system development. This is in line with the IEGs playing a major role in synaptic plasticity, which is critical for key brain functions including learning and memory (Minatohara et al., 2015). Also consistent with our findings, learning and memory acquisition up-regulated the expression of the IEGs *Arc*, *Fos*, *Btg2*, *Sik1*, *Dusp1*, *Ier2*, and *Egr2* in the hippocampus and retrosplenial cortex of adult mice (Peixoto et al., 2015). DNA methylation plays a critical role in regulating gene expression and mediating synaptic plasticity by modulating the expression of memory-enhancing/-supressing genes (Heyward and Sweatt, 2015). Notably, the transcription of IEGs is highly susceptible to DNA methylation due to the large presence of CpG islands (Fowler et al., 2011; Halder et al., 2016; Kaas et al., 2013). We have shown that specific *Caudovirales* levels were mainly associated with folate-mediated 1C-metabolism. Because it plays a critical role in epigenetic regulation, including DNA methylation (Mentch et al., 2015), specific *Caudovirales* might improve cognitive performance through the regulation of the methylation status of IEGs. Finally, a pathway-based analysis revealed again over-representation of the AP-1 transcription factor network and MAPK signaling (Figures 6B and 6C). It also highlighted the serotonin, oncostatin M, brain-derived neurotrophic factor (BDNF), and IL-4 and IL-13 signaling pathways, among others. The anti-inflammatory cytokines IL-4 and IL-13 have shown to affect both learning and memory via the IL-4 receptor influencing the production of BDNF (Brombacher et al., 2020), while oncostatin M is a member of the IL-6 cytokine family that has shown neuroprotective properties (Moidunny et al., 2010). Dot plots showing a dose-response effect on relevant gene transcripts according to the specific *Caudovirales* and *Microviridae* are shown in Figures 6D and S6.

Lactococcal 936-type bacteriophages increase memory retention and alter the expression of memory-related genes in *Drosophila melanogaster*

Our results showed significant associations between bacteriophages and bacterial structure and function. However, *in vitro* studies have suggested that bacteriophages not only affect host physiology indirectly by shaping microbial communities, but they also modulate eukaryotic physiology directly (Keen and Dantas, 2018). Considering this, to gain further support for the role of *Siphoviridae* in cognition, we evaluated the impact of *Lactococcal 936-type* bacteriophages in learning and memory using the model organism *Drosophila melanogaster*. We used the *Lactococcal 936-type phages* because they were the most abundant characterized *Siphoviridae* species in the discovery cohort. In line with our results, the so-called 936-type phages are the most frequently isolated lactococcal phage species in dairy fermentations (Murphy et al., 2017). Flies were fed with diets supplemented with whey, a milk derivate known to be rich in *Lactococcal 936-type* bacteriophages (members of the *Siphoviridae* family), reaching concentrations up to 10^8 PFU/mL in this product (Ly-Chatain et al., 2011).

Since bacteriophages are thermolabile (Marcó et al., 2019), we also generated whey diets (WD) free of bacteriophages by submitting whey powder to a thermic treatment for 15 min at 121°C (WD 121°C) or 3 min at 100°C (WD 100°C). To preserve bacteriophage integrity, we also generated WD 45°C, adding whey powder to the fly food at 45°C. After performing conventional PCR, PCR products confirming the presence of *Lactococcal 936-type phages* DNA were detected in whey powder and WD 45°C samples (Figure S7A, lanes 1–4), whereas absence or weak PCR products were observed after PCR amplification of WD 100°C or WD 121°C samples (Figure S7A, lanes 5–8), in the context of probable degradation of the bacteriophage DNA due to the thermic treatment. No *lactococcal 936-species DNA* was amplified after conventional PCR from samples of diets absent of whey, standard diet (SD), or sterilized SD (SD 121°C),

Figure 5. Cognitive traits were phenocopied to recipient mice in parallel to changes in prefrontal cortex genes involved in memory formation

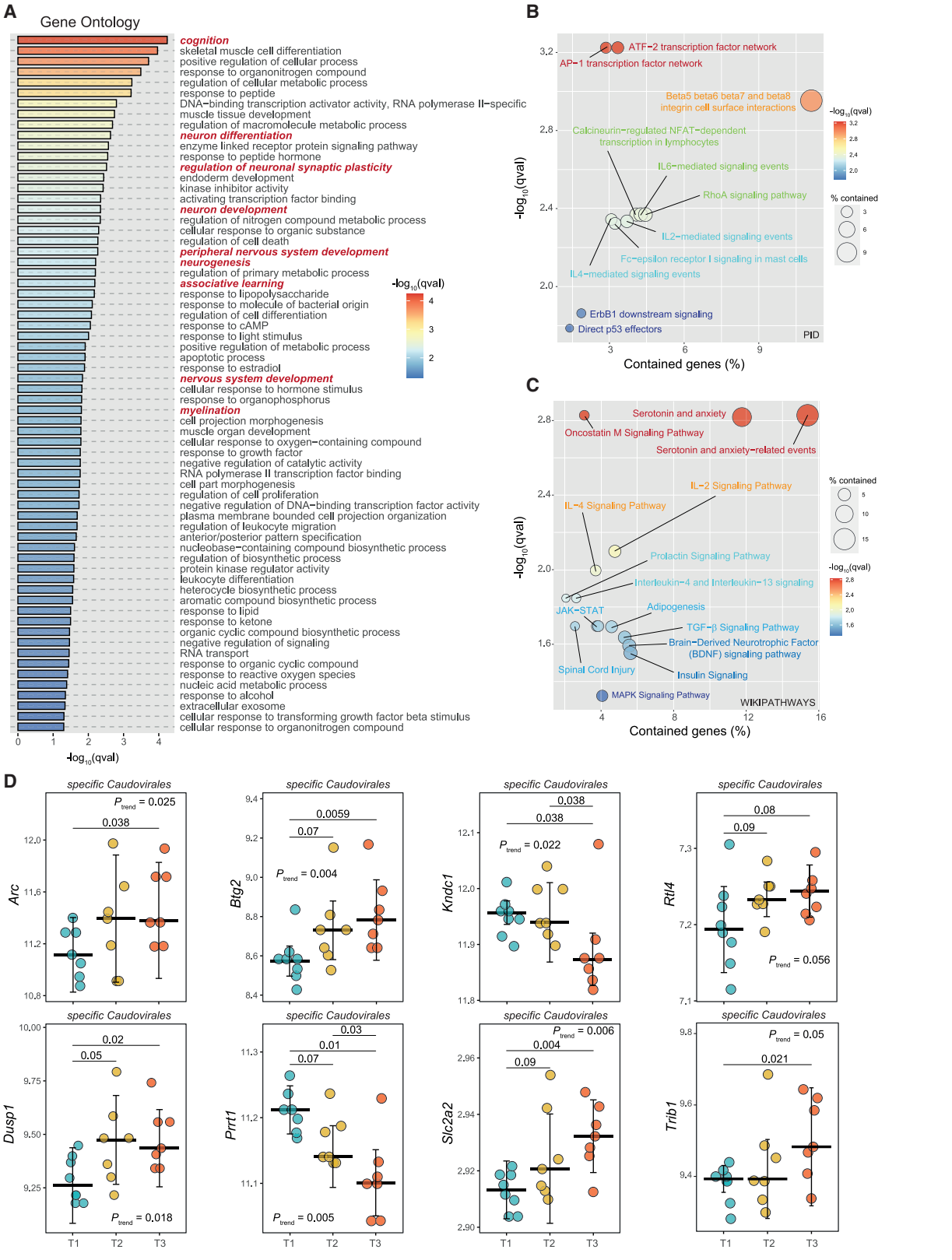
(A) Experimental design for the fecal microbiota transplantation study.

(B and C) (B) Scatter plots of the partial Spearman's rank correlations between the donor's clr-transformed specific *Caudovirales* [*Siphoviridae* (>86%), *Demerecviridae*, and *Drexlerviridae*] and (C) *Microviridae* values and the recipient's mice novel object recognition (NOR) test score, controlling for donor's age, BMI, sex, and education years.

(D and E) (D) Volcano plot of differentially expressed gene transcripts in the medial prefrontal cortex of the recipient mice prefrontal cortex associated with the donor's specific *Caudovirales* and (E) *Microviridae* clr-transformed values identified by limma-voom after adjusting p values for multiple testing (padj). The x axis represents the logarithm to the base 2 ($\log_2$) of the fold change associated with a unit change in the clr-transformed specific *Caudovirales* values. The y axis represents the logarithm to the base 10 ($\log_{10}$) of the p values adjusted for multiple testing (padj). Each dot represents a specific transcript. Significantly different transcripts (padj < 0.05) with positive fold changes are colored in green, and those with significantly negative associations are colored in red.

(F) Gene interaction network constructed using differentially expressed genes associated with the specific *Caudovirales* clr-transformed values 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. A functional cluster (highlighted in red) was detected comprising genes involved in memory formation.

(G) Local network clusters identified by STRING.



(legend on next page)

(Figure S7A, lanes 9–12). These results suggested that thermolabile components in whey powder and the presence of bacteriophages in this product might promote memory. The aversive taste memory paradigm evaluates learning and memory capabilities. *Drosophila* discriminates district taste modalities, including sweet and bitter (see STAR Methods). Flies are stimulated with sweet tasting like a sucrose solution to the tarsi (legs), inducing the proboscis extension reflex (PER). During the training period, repeated paired applications of sucrose to the tarsi and quinine to the proboscis reduce PER significantly. Due to associative memory, PER is reduced in time to the subsequent application of sucrose alone in the tarsi (Figures 7A, 7B, 7B′, and 7B″). Wild-type flies fed with WD 45°C showed longer PER suppression during the memory test compared with flies fed with thermic treated diets WD 121°C or WD 100°C, indicating that they were able to maintain aversive taste memory for longer periods (Figures S7B and S7B″). Although an increment in learning performance was observed, this was not statistically significant (Figure S7B′). When flies were fed with SD absent of whey (SD or 121°C SD), no differences in learning and memory performance were observed during the aversive taste memory paradigm (Figures S7C, S7C′, and S7C″). These results suggest that *phage 936* is by itself able to improve memory capabilities of *Drosophila melanogaster* when ingested through diet.

Subsequently, to determine whether the increment in memory retention capabilities was mediated by *Lactococcal 936-type bacteriophages*, whey and SD were treated at 121°C to eliminate any possible bacteriophage contamination. Afterwards, the conditions SD + *phage 936* and WD + *phage 936* were supplemented with 50.000 UFP of *phage-936* when the food was at room temperature. Learning and memory performance of female flies was assessed with the aversive taste memory paradigm. Flies fed with WD + *phage 936* and SD + *phage 936* showed longer PER suppression during the memory test, indicating longer memory retention than flies fed with WD and SD, respectively (Figures 7C and 7D). Flies fed with diets supplemented with the *phage 936* showed increased learning but this did not reach statistical significance (Figures 7C and 7D) but showed a significant reduction of PER 10 and 20 min after training, indicating longer memory retention (Figures 7C and 7D). These results suggest that *phage 936* is by itself able to improve memory capabilities of *Drosophila melanogaster* when ingested through diet.

To gain better insights into the molecular mechanisms mediated by *phage 936* and its impact on cognition, we then focused on the cluster of learning and memory-related genes (the corresponding *Drosophila* orthologs are shown in Table S4A) with up-regulated expression in mouse brain after specific *Caudovirales*-enriched microbiota transplantation (Figure 5D, red nodes in the gene interaction network of Figure 5F). Remarkably, the relative transcript expression in fly heads (qRT-PCR) of these genes was significantly up-regulated in fly heads fed with WD + *phage 936*, correlating with the results obtained in mouse brains. The exception was *Kay*, the Drosophila homolog of *Fos*, that was significantly down-regulated (Figure 7E). Similarly, flies fed with SD + *phage 936* showed significant increments in gene expression of *Axud*, *Trbl*, and *Socs44a*, while the expression of *Kay* was again significantly down-regulated. A trend toward increased *Arc1*, *sr*, and *Sik2* was also observed (Figure 7F).

Sr, *puc*, *kay*, *Sik2*, and *Arc1* are part of the *Drosophila* activity-regulated genes (ARGs). Similar to mammalian IEGs, these genes are initial transducers of neural activity implicated in the first steps of memory formation, their expression being transiently up-regulated upon neural activity (Chen et al., 2016). The basal expression levels of these genes appear to be differentially regulated in *Drosophila* after the administration of *phage 936* with one or both diets. Although the aversive taste memory paradigm does not measure long-term memory, ARG genes are also involved in neuronal plasticity and neurodevelopment (Kim et al., 2018). The differential expression of these genes in steady state may result in a neuronal state that enhances different forms of short-term memory.

Tribbles (*Trbl*) serine/threonine pseudokinase (the ortholog of mammalian TRIBs) was strongly up-regulated in *Drosophila* after supplementing both diets with *phage 936*. Similarly, we observed that the expression of *Trib1* was up-regulated in the mouse PFC after microbiota transplantation from human donors with increased specific *Caudovirales* levels. Mammals have three related Tribbles family members (*TRIB1-3*), whereas *Drosophila* only one. All TRIB pseudokinases show widespread expression in the nervous system in invertebrates and vertebrates, and their molecular function appears to be evolutionarily conserved. TRIBs act as transcriptional co-activators and repressors in the nucleus, as MAP kinase and AKT inhibitors and as proteasome adapters in the cytoplasm (Eyers et al., 2017). In *Drosophila*, *Trbl* is essential for proper memory formation (LaFerriere and Zars, 2017), and mutations in *Trbl* are associated with short-term memory defects (LaFerriere et al., 2008). Interestingly, *Trbl* is a known regulator of *slow boarder* (a homolog to the C/EBP family of transcription factors) by targeting it to the proteasome pathway (Masoner et al., 2013). Similarly, TRIBs bind and direct the degradation of several members of the C/EBP family to the proteasome in mammals (Keeshan et al., 2006; Naiki et al., 2007). Several members of the C/EBP superfamily of transcription factors act to constrain the conversion of short- to long-term synaptic potentiation and short- to long-term memory storage, thereby acting as memory

Figure 6. Over-representation analysis of differentially expressed gene transcripts associated with the specific *Caudovirales* [*Siphoviridae* (>86%), *Demerecviridae*, and *Drexlerviridae*] clr values

(A) Manhattan-like plot of over-represented gene ontology biological processes. Bars are colored according to the p value adjusted for multiple testing (q values). The x axis represents the $-\log_{10}$(q values) values multiplied by the fold-change sign to take into account the direction of the association. Biological processes related with cognition, learning, and nervous system development are highlighted in bold red.

(B and C) (B) Pathway over-representation analysis of significant gene transcripts based on pathway interaction database (PID) and (C) Wikipathway database. The bubble size represents the percentage of significant genes associated with the specific *Caudovirales* clr-transformed values contained in each pathway. The y axis represents the $-\log_{10}$(q value). Bubbles are colored according to the p value adjusted for multiple testing (q values).

(D) Boxplots of selected relevant genes significantly associated with the specific *Caudovirales* clr-transformed values according to the specific *Caudovirales* tertiles.

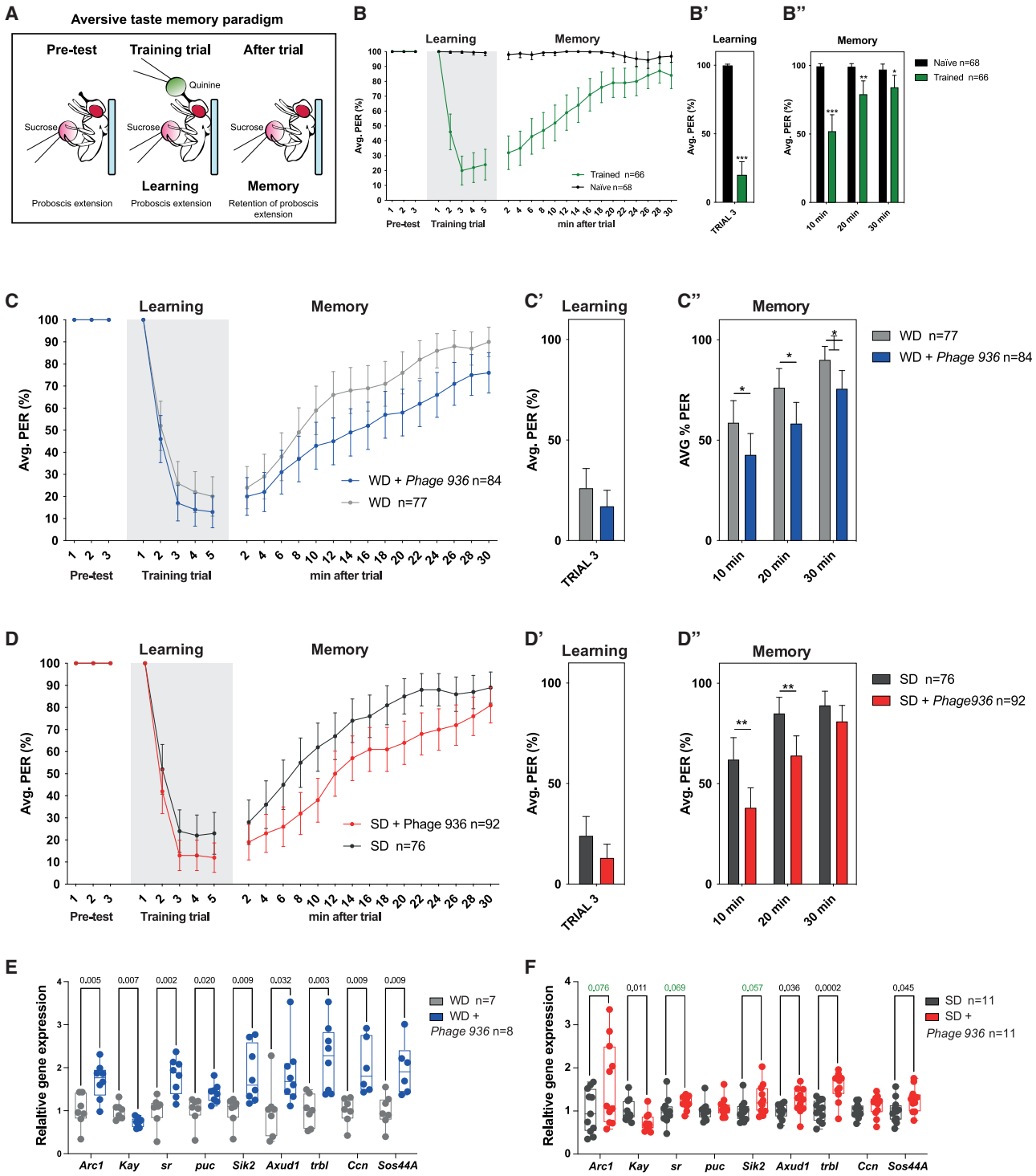


Figure 7. Dietary supplementation with *Lactococcus lactis* bacteriophage 936 promotes memory retention in *Drosophila melanogaster*

(A–D) (A) Schematic representation of the aversive taste memory paradigm. During the pre-test, flies are screened for proboscis (mouth) full extension after applying a droplet of sucrose on the tarsi. During the training trials (learning period), a droplet of sucrose is applied on the tarsi (legs), followed by the application of a droplet of quinine on the proboscis. After training, memory is assessed by observing the suppression of the PER in response to sucrose stimulation on the tarsi. Data show 10-day-old pre-mated female average PER (%) ±95% CI of all tested flies, plotted over 5 training trials (3 min inter-trial interval), followed by 15 memory trials (2 min inter-trial interval) (B–D). Bars represent the average PER (%) 95% CI at learning trial 3 (B′–D′) and at 10, 20, and 30 min after learning (B″–D″). (B) Trained wild-type flies (green) show PER suppression during the memory test, whereas naive (black) flies show a strong PER. (B′) Trained flies show a significant reduction of PER at learning trial 3 and (B″) 10 and 20 min after training. (C) Flies fed with sterile whey diet supplemented with *phage 936* (WD sterile + *phage 936*) (blue) show

(legend continued on next page)

inhibitors (Chen et al., 2003). Consequently, upregulation of TRIBs may lead to an enhancement of C/EBP degradation resulting in better memory capabilities. This is an attractive molecular mechanism to explain how the upregulation of *Tribbles* pseudokinases, in both animal models used in this study (Drosophila and mouse), are associated to an enhancement of memory capabilities.

The complex bacteriophage communities represent one of the biggest gaps in our understanding of the human microbiome. In fact, most studies have focused on the dysbiotic process only in bacterial populations. In this study, we show that two of the most prevalent bacteriophages in the gut microbiota run in parallel to human host cognition by possibly hijacking the bacterial host metabolism, resulting in a favorable/unfavorable metabolic profile depending on specific *Caudovirales/Microviridae* ratio. In addition, the direct supplementation of *936-type Siphoviridae* bacteriophages in the diet of flies and after microbiota transplantation in mice favors increased memory capabilities through upregulation of the expression of genes involved in synaptic plasticity, neuronal development, and memory formation. All these findings may have implications in the design of dietary interventions targeted at improving cognitive and memory dysfunction.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Clinical study
  - Animal studies
  - Drosophila studies
- METHODS DETAILS
  - Neuropsychological assessment in humans
  - Extraction of faecal genomic DNA and Whole-Genome Shotgun Sequencing
  - ^{1}H-NMR metabolomics analyses
  - Animal studies
  - Study of gene expression in mouse prefrontal cortex
  - Drosophila melanogaster studies
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Metagenomics analysis
  - Metabolomics analysis
  - RNA-seq analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chom.2022.01.013.

ACKNOWLEDGMENTS

This work was partially funded by the Instituto de Salud Carlos III (Madrid, Spain) through the project PI15/01934, PI18/01022, PI21/01361) to J.M.F.-R. and the project PI20/01090 (co-funded by the European Regional Development Fund. "A way to make Europe") to J.M.-P., the grants SAF2015-65878-R from the Ministry of Economy and Competitiveness, Prometeo/2018/A/133 from Generalitat Valenciana, Spain and also by the Fondo Europeo de Desarrollo Regional (FEDER) funds, European Commission (FP7, NeuroPain #2013-602891), the Catalan Government (AGAUR, #SGR2017-669, #2017 SGR-734, ICREA Academia Award 2015 to R.M. and ICREA Academia Award 2022 to J.M.F.R.), the Spanish Instituto de Salud Carlos III (RTA, #RD16/0017/0020), the European Regional Development Fund (project No. 01.2.2-LMT-K-718-02-0014) under grant agreement with the Research Council of Lithuania (LMTLT), and the Project ThinkGut (EFA345/19) 65% co-financed by the European Regional Development Fund (ERDF) through the Interreg V-A Spain-France-Andorra programme (POCTEFA 2014-2020). CIBERobn is also co-funded by the European Regional Development Fund. We also acknowledge the 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.M M.A.-R. is funded by the Instituto de Salud Carlos III, Río Hortega (CM19/00190). J.M.-P. is funded by the Miguel Servet Program from the Instituto de Salud Carlos III (ISCIII CP18/00009), co-funded by the European Social Fund "Investing in your future." A.C.-N. is funded by the Instituto de Salud Carlos III, Sara Borrell. MMG was funded by the Spanish Ministry of Science, Innovation and Universities RTI2018-094248-B-I00. The graphical abstract was created with BioRender.com.

AUTHOR CONTRIBUTIONS

J.M.-P. and M.A.-R. researched the data. A.C.-N. and C.Z.-T. performed the experiments in *Drosophila* and A.C.-N. also wrote the manuscript. A.B. and R.M. performed the mice experiment. C.C. performed the neuropsychological examination. V.P.-B. and A.M. contributed to the determination and analysis of the microbiota; L.R. and W.R. contributed to the discussion and reviewed the manuscript. 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 the 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.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 23, 2021
Revised: November 12, 2021
Accepted: January 21, 2022
Published: February 16, 2022

prolonged supersession of average PER relative to flies fed with sterile whey diet (WD sterile) (gray). (C′) No significant reduction of PER was observed at learning trial 3 and (C″) WD sterile + *phage 936* showed significant supersession of average PER 10, 20, and 30 min after training. (D) Flies fed with standard diet supplemented with *phage 936* (red) show longer supersession of average PER relative to flies fed with standard diet 121°C (dark gray). (D′) No significant reduction of PER was observed at learning trial 3 and (D″) SD sterile + *phage 936* showed significant supersession of average PER 10 and 20 min after training. Significance was calculated with Fisher's exact test (* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$).

(E and F) qRT-PCR results, bars represent relative gene expression in fly heats. Each sample contains a pool of 10–15 fly heads. Transcript levels of flies fed with (E) WD + *phage* 936 compared to WD and (F) SD + *phage* 936 compared to SD (threshold line). Error bars represent normalized SEM, p values were determined using the Brown-Forsythe and Welch ANOVA tests.

REFERENCES

Alberini, C.M. (2009). Transcription factors in long-term memory and synaptic plasticity. Physiol. Rev. *89*, 121–145.

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders (5th ed.) (Springer).

Blount, B.C., Mack, M.M., Wehr, C.M., Macgregor, J.T., Hiatt, R.A., Wang, G., Wickramasinghe, S.N., Everson, R.B., and Ames, B.N. (1997). Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. Proc. Natl. Acad. Sci. USA *94*, 3290–3295.

Brombacher, T.M., Berkiks, I., Pillay, S., Scibiorek, M., Moses, B.O., and Brombacher, F. (2020). IL-4R alpha deficiency influences hippocampal-BDNF signaling pathway to impair reference memory. Sci. Rep. *10*, 16506.

Burokas, A., Martín-García, E., Gutiérrez-Cuesta, J., Rojas, S., Herance, J.R., Gispert, J.D., Serra, M.Á., and Maldonado, R. (2014). Relationships between serotonergic and cannabinoid system in depressive-like behavior: a PET study with [11C]-DASB. J. Neurochem. *130*, 126–135.

Carvajal-Rodríguez, A., de Uña-Alvarez, J., and Rolán-Alvarez, E. (2009). A new multitest correction (SGoF) that increases its statistical power when increasing the number of tests. BMC Bioinformatics *10*, 209.

Castells-Nobau, A., Eidhof, I., Fenckova, M., Brenman-Suttner, D.B., Scheffer-De Gooyert, J.M., Christine, S., Schellevis, R.L., Van Der Laan, K., Quentin, C., Van Ninhuijs, L., et al. (2019). Conserved regulation of neurodevelopmental processes and behavior by FoxP in *Drosophila*. PLoS One *14*, e0211652.

Chen, A., Muzzio, I.A., Malleret, G., Bartsch, D., Verbitsky, M., Pavlidis, P., Yonan, A.L., Vronskaya, S., Grody, M.B., Cepeda, I., et al. (2003). Inducible enhancement of memory storage and synaptic plasticity in transgenic mice expressing an inhibitor of ATF4 (CREB-2) and C/EBP proteins. Neuron *39*, 655–669.

Chen, X., Rahman, R., Guo, F., and Rosbash, M. (2016). Genome-wide identification of neuronal activity-regulated genes in *Drosophila*. eLife *5*, e19942.

Corrigan, J.D., and Hinkeldey, N.S. (1987). Relationships between parts A and B of the trail making test. J. Clin. Psychol. *43*, 402–409.

Cryan, J.F., O'Riordan, K.J., Sandhu, K., Peterson, V., and Dinan, T.G. (2020). The gut microbiome in neurological disorders. Lancet Neurol *19*, 179–194.

Degenhardt, F., Seifert, S., and Szymczak, S. (2019). Evaluation of variable selection methods for random forests and omics data sets. Brief. Bioinform. *20*, 492–503.

Delis, D.C., Kramer, J.H., Kaplan, E., and Ober, B.A. (2000). Manual for the California Verbal Learning Test, (CVLT-II) (The Psychological Corporation).

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Ducker, G.S., and Rabinowitz, J.D. (2017). One-carbon metabolism in health and disease. Cell Metab *25*, 27–42.

Durbin, R., Eddy, S.R., Krogh, A., and Mitchison, G. (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids (Cambridge University Press).

Eyers, P.A., Keeshan, K., and Kannan, N. (2017). Tribbles in the 21st century: the evolving roles of tribbles pseudokinases in biology and disease. Trends Cell Biol *27*, 284–298.

Fernandes, A.D., Reid, J.N.S., Macklaim, J.M., McMurrough, T.A., Edgell, D.R., and Gloor, G.B. (2014). 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.

Fowler, T., Sen, R., and Roy, A.L. (2011). Regulation of primary response genes. Mol. Cell *44*, 348–360.

Galiez, C., Siebert, M., Enault, F., Vincent, J., and Söding, J. (2017). WisH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics *33*, 3113–3114.

Gregory, A.C., Zablocki, O., Zayed, A.A., Howell, A., Bolduc, B., and Sullivan, M.B. (2020). The gut virome database reveals age-dependent patterns of virome diversity in the human gut. Cell Host Microbe *28*, 724–740.e8.

Halder, R., Hennion, M., Vidal, R.O., Shomroni, O., Rahman, R.U., Rajput, A., Centeno, T.P., Van Bebber, F., Capece, V., Vizcaino, J.C.G., et al. (2016). DNA methylation changes in plasticity genes accompany the formation and maintenance of memory. Nat. Neurosci. *19*, 102–110.

Herrmann, M.J., Horst, A.K., Löble, S., Möll, M.T., Katzorke, A., and Polak, T. (2017). Relevance of dorsolateral and frontotemporal cortex on the phonemic verbal fluency—a fNIRS-study. Neuroscience *367*, 169–177.

Heyward, F.D., and Sweatt, J.D. (2015). DNA methylation in memory formation: emerging insights. Neuroscientist *21*, 475–489.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119.

Kaas, G.A., Zhong, C., Eason, D.E., Ross, D.L., Vachhani, R.V., Ming, G.L. li, King, J.R., Song, H., and Sweatt, J.D. (2013). TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. Neuron *79*, 1086–1093.

Kamburov, A., Stelzl, U., Lehrach, H., and Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res *41*, D793–D800.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27–30.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

Keen, E.C., and Dantas, G. (2018). Close encounters of three kinds: bacteriophages, commensal bacteria, and host immunity. Trends Microbiol *26*, 943–954.

Keene, A.C., and Masek, P. (2012). Optogenetic induction of aversive taste memory. Neuroscience *222*, 173–180.

Keeshan, K., He, Y., Wouters, B.J., Shestova, O., Xu, L., Sai, H., Rodriguez, C.G., Maillard, I., Tobias, J.W., Valk, P., et al. (2006). Tribbles homolog 2 inactivates C/EBPα and causes acute myelogenous leukemia. Cancer Cell *10*, 401–411.

Kelly, J.R., Borre, Y., O' Brien, C., Patterson, E., El Aidy, S., Deane, J., Kennedy, P.J., Beers, S., Scott, K., Moloney, G., et al. (2016). Transferring the blues: depression-associated gut microbiota induces neurobehavioural changes in the rat. J. Psychiatr. Res. *82*, 109–118.

Kim, S., Kim, H., and Um, J.W. (2018). Synapse development organized by neuronal activity-regulated immediate-early genes. Exp. Mol. Med. *50*, 1–7.

Kirkhart, C., and Scott, K. (2015). Gustatory learning and processing in the Drosophila mushroom bodies. J. Neurosci. *35*, 5950–5958.

Kure, S., Kato, K., Dinopoulos, A., Gail, C., DeGrauw, T.J., Christodoulou, J., Bzduch, V., Kalmanchey, R., Fekete, G., Trojovsky, A., et al. (2006). Comprehensive mutation analysis of GLDC, AMT, and GCSH in nonketotic hyperglycinemia. Hum. Mutat. *27*, 343–352.

Kursa, M.B., and Rudnicki, W.R. (2010). Feature selection with the boruta package. J. Stat. Softw. *36*, 1–13.

LaFerriere, H., Guarnieri, D.J., Sitaraman, D., Diegelmann, S., Heberlein, U., and Zars, T. (2008). Genetic dissociation of ethanol sensitivity and memory formation in *Drosophila melanogaster*. Genetics *178*, 1895–1902.

LaFerriere, H., and Zars, T. (2017). The *Drosophila melanogaster tribbles pseudokinase* is necessary for proper memory formation. Neurobiol. Learn. Mem. *144*, 68–76.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lazarov, O., Robinson, J., Tang, Y.P., Hairston, I.S., Korade-Mirnics, Z., Lee, V.M.Y., Hersh, L.B., Sapolsky, R.M., Mirnics, K., and Sisodia, S.S. (2005). Environmental enrichment reduces Aβ levels and amyloid deposition in transgenic mice. Cell *120*, 701–713.

Levis, B., Benedetti, A., and Thombs, B.D.; DEPRESsion Screening Data (DEPRESSD) Collaboration (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ *365*, l1476.

Lezak, M.D. (1984). Neuropsychological assessment in behavioral toxicology—developing techniques and interpretative issues. Scand. J. Work Environ. Health *10* (*Suppl 1*), 25–29.

Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674–1676.

Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., and Holtz, L.R. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat. Med. *21*, 1228–1234.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2-\Delta\Delta CT$ method. Methods *25*, 402–408.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Ly-Chatain, M.H., Durand, L., Rigobello, V., Vera, A., and Demarigny, Y. (2011). Direct quantitative detection and identification of lactococcal bacteriophages from milk and whey by real-time PCR: application for the detection of lactococcal bacteriophages in goat's raw milk whey in France. Int. J. Microbiol. *2011*, 594369.

Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963.

Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J. Der, De Vos, W.M., and Young, M.J. (2016). Healthy human gut phageome. Proc. Natl. Acad. Sci. USA *113*, 10400–10405.

Manrique, P., Dills, M., and Young, M.J. (2017). The human gut phage community and its implications for health and disease. Viruses *9*, 141.

Marcó, M.B., Suárez, V.B., Quiberoni, A., and Pujato, S.A. (2019). Inactivation of dairy bacteriophages by thermal and chemical treatments. Viruses *11*, 480.

Martín, V., Mediano, P., Del Campo, R., Rodríguez, J.M., and Marín, M. (2016). Streptococcal diversity of human milk and comparison of different methods for the taxonomic identification of streptococci. J. Hum. Lact. *32*, NP84–NP94.

Masek, P., and Keene, A.C. (2016). Gustatory processing and taste memory in Drosophila. J. Neurogenet. *30*, 112–121.

Masek, P., Worden, K., Aso, Y., Rubin, G.M., and Keene, A.C. (2015). A dopamine-modulated neural circuit regulating aversive taste memory in drosophila. Curr. Biol. *25*, 1535–1541.

Masoner, V., Das, R., Pence, L., Anand, G., LaFerriere, H., Zars, T., Bouyain, S., and Dobens, L.L. (2013). The kinase domain of Drosophila Tribbles is required for turnover of fly C/EBP during cellmigration. Dev. Biol. *375*, 33–44.

Mayneris-Perxachs, J., Arnoriaga-Rodríguez, M., Luque-Córdoba, D., Priego-Capote, F., Pérez-Brocal, V., Moya, A., Burokas, A., Maldonado, R., and Fernández-Real, J.M. (2020). Gut microbiota steroid sexual dimorphism and its impact on gonadal steroids: influences of obesity and menopausal status. Microbiome *8*, 136.

Mentch, S.J., Mehrmohamadi, M., Huang, L., Liu, X., Gupta, D., Mattocks, D., Gómez Padilla, P., Ables, G., Bamman, M.M., Thalacker-Mercer, A.E., et al. (2015). Histone methylation dynamics and gene regulation occur through the sensing of one-carbon metabolism. Cell Metab *22*, 861–873.

Menzel, P., Ng, K.L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. *7*, 11257.

Mills, R., Rozanov, M., Lomsadze, A., Tatusova, T., and Borodovsky, M. (2003). Improving gene annotation of complete viral genomes. Nucleic Acids Res *31*, 7041–7055.

Minatohara, K., Akiyoshi, M., and Okuno, H. (2015). Role of immediate-early genes in synaptic plasticity and neuronal ensembles underlying the memory trace. Front. Mol. Neurosci. *8*, 78.

Mirzaei, M.K., and Maurice, C.F. (2017). Ménage à trois in the human gut: interactions between host, bacteria and phages. Nat. Rev. Microbiol. *15*, 397–408.

Moidunny, S., Dias, R.B., Wesseling, E., Sekino, Y., Boddeke, H.W.G.M., Sebastião, A.M., and Biber, K. (2010). Interleukin-6-type cytokines in neuroprotection and neuromodulation: oncostatin M, but not leukemia inhibitory factor, requires neuronal adenosine A1 receptor function. J. Neurochem. *114*, 1667–1677.

Morais, L.H., Schreiber, H.L., and Mazmanian, S.K. (2021). The gut microbiota–brain axis in behaviour and brain disorders. Nat. Rev. Microbiol. *19*, 241–255.

Moreno-Gallego, J.L., Chou, S.P., Di Rienzi, S.C., Goodrich, J.K., Spector, T.D., Bell, J.T., Youngblut, N.D., Hewson, I., Reyes, A., and Ley, R.E. (2019). Virome diversity correlates with intestinal microbiome diversity in adult monozygotic twins. Cell Host Microbe *25*, 261–272.e5.

Murphy, J., Mahony, J., Fitzgerald, G.F., and van Sinderen, D. (2017). Bacteriophages infecting lactic acid bacteria. In Cheese: Chemistry, Physics and Microbiology, Fourth Edition (Elsevier Inc.)), pp. 249–272.

Naiki, T., Saijou, E., Miyaoka, Y., Sekine, K., and Miyajima, A. (2007). TRB2, a mouse tribbles ortholog, suppresses adipocyte differentiation by inhibiting AKT and C/EBPβ. J. Biol. Chem. *282*, 24075–24082.

Narisawa, A., Komatsuzaki, S., Kikuchi, A., Niihori, T., Aoki, Y., Fujiwara, K., Tanemura, M., Hata, A., Suzuki, Y., Relton, C.L., et al. (2012). Mutations in genes encoding the glycine cleavage system predispose to neural tube defects in mice and humans. Hum. Mol. Genet. *21*, 1496–1503.

Page, K.A., Williamson, A., Yu, N., McNay, E.C., Dzuira, J., McCrimmon, R.J., and Sherwin, R.S. (2009). Medium-chain fatty acids improve cognitive function in intensively treated type 1 diabetic patients and support in vitro synaptic transmission during acute hypoglycemia. Diabetes *58*, 1237–1244.

Paolo, A.M., Tröster, A.I., and Ryan, J.J. (1997). Test-retest stability of the California Verbal Learning Test in older persons. Neuropsychology *11*, 613–616.

Paxinos, G., and Franklin, K.B.J. (1997). The Mouse Brain in Stereotaxic Coordinates (Academic Press).

Peixoto, L.L., Wimmer, M.E., Poplawski, S.G., Tudor, J.C., Kenworthy, C.A., Liu, S., Mizuno, K., Garcia, B.A., Zhang, N.R., Giese, K.P., and Abel, T. (2015). Memory acquisition and retrieval impact different epigenetic processes that regulate gene expression. BMC Genomics *16* (*Suppl 5*), S5.

Poon, C.H., Chan, Y.S., Fung, M.L., and Lim, L.W. (2020). Memory and neuromodulation: a perspective of DNA methylation. Neurosci. Biobehav. Rev. *111*, 57–68.

Puig, J., Biarnes, C., Pedraza, S., Vilanova, J.C., Pamplona, R., Fernández-Real, J.M., Brugada, R., Ramos, R., Coll-de-Tuero, G., Calvo-Perxas, L., et al. (2020). The aging imageomics study: rationale, design and baseline characteristics of the study population. Mech. Ageing Dev. *189*, 111257.

Qureshi, R., and Sacan, A. (2013). A novel method for the normalization of microRNA RT-PCR data. BMC Med. Genomics *6*, S14.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015). VirSorter: mining viral signal from microbial genomic data. PeerJ *3*, e985.

Saravia, R., Ten-Blanco, M., Julià-Hernández, M., Gagliano, H., Andero, R., Armario, A., Maldonado, R., and Berrendero, F. (2019). Concomitant THC and stress adolescent exposure induces impaired fear extinction and related neurobiological changes in adulthood. Neuropharmacology *144*, 345–357.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863–864.

Schulfer, A., Santiago-Rodriguez, T.M., Ly, M., Borin, J.M., Chopyk, J., Blaser, M.J., and Pride, D.T. (2020). Fecal viral community responses to high-fat diet in mice. mSphere *5*, e00833–19.

Shkoporov, A.N., Clooney, A.G., Sutton, T.D.S., Ryan, F.J., Daly, K.M., Nolan, J.A., McDonnell, S.A., Khokhlova, E.V., Draper, L.A., Forde, A., et al. (2019). The human gut Virome is highly diverse, stable, and individual specific. Cell Host Microbe *26*, 527–541.e5.

Silva, P.H.R., Spedo, C.T., Baldassarini, C.R., Benini, C.D., Ferreira, D.A., Barreira, A.A., and Leoni, R.F. (2019). Brain functional and effective connectivity underlying the information processing speed assessed by the Symbol Digit Modalities Test. Neuroimage *184*, 761–770.

Strauss, E., Sherman, E.M.S., and Spreen, O. (2006). A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary (Oxford University Press).

Sun, Y., Gooch, H., and Sah, P. (2020). Fear conditioning and the basolateral amygdala. F1000Res. *9*, 53.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

R Core Team (2013). R development core team. RA Lang. Environ. Stat. Comput. *55*, 275–286.

Vioque, J., Navarrete-Muñoz, E.M., Gimenez-Monzó, D., García-de-la-Hera, M., Granado, F., Young, I.S., Ramón, R., Ballester, F., Murcia, M., Rebagliato, M., et al. (2013). Reproducibility and validity of a food frequency questionnaire among pregnant women in a Mediterranean area. Nutr. J. *12*, 26.

Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Adriaenssens, E.M., Alfenas-Zerbini, P., Davison, A.J., Dempsey, D.M., Dutilh, B.E., García, M.L., et al. (2021). Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). Arch. Virol. *166*, 2633–2648.

Wang, D., and Mitchell, E.S. (2016). Cognition and synaptic-plasticity related changes in aged rats supplemented with 8- and 10-carbon medium chain triglycerides. PLoS One *11*, e0160159.

Wechsler, D. (2012). WAIS-IV. Escala de inteligencia de Wechsler para adultos-IV. Manual técnico y de interpretación (NCS Pearson).

Zhang, R., Mirdita, M., Levy Karin, E., Norroy, C., Galiez, C., and Söding, J. (2021). SpacePHARER: sensitive identification of phages from CRISPR spacers in prokaryotic hosts. Bioinformatics *37*, 3364–3366.

Smyth, G.K., Michaud, J., and Scott, H.S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics. *21*, 2067–2075. https://doi.org/10.1093/bioinformatics/bti270.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Human body fluids (feces, plasma)	This paper	N/A
Mice feces	This paper	N/A
Mice prefrontal cortex	This paper	N/A
Chemicals, Peptides, and Recombinant Proteins		
3-(Trimethylsilyl)propionic-2,2,3,3-d4 acid sodium salt (TSP)	Sigma-Aldrich	Cat# 269913
Disodium hydrogen phosphate	Sigma-Aldrich	Cat#1.06586
Sodium dihydrogen phosphate	Sigma-Aldrich	Cat#1.06370
Deuterated water 99.8%	Thermo Fisher Scientific	Cat#10255880
Critical Commercial Assays		
QIAamp DNA mini stool kit	Qiagen	Cat#51504
Nextera DNA Flex Library Preparation kit	Illumina	Cat#20018705
ALLPrep DNA/RNA/miRNA Universal kit	Qiagen	Cat#80224
TrueSeq stranded mRNA library preparation kit	Illumina	Cat#20020594
Truseq RNA Single Indexes	Illumina	Cat#20020492
Truseq RNA Single Indexes	Illumina	Cat#20020493
RNA 6000 Nano chip	Agilent	Cat#5067-1511
DNA 1000 chip	Agilent	Cat#5067-1504
KAPA Library Quantification Kit	Roche	Cat#07960204001
DNA-free Ambion kit	Thermo Fisher Scientific	Cat#AM1906
High-capacity cDNA reverse transcription	Applied Biosystems	Cat#4368814
LightCycler 480 SYBR Green Master	Roche	Cat#04707516001
Deposited Data		
Metagenome Sequencing Data of Fecal Samples from Human subjects and Mice	European Nucleotide Archive (ENA)	Project number: PRJEB39631 Human samples accession numbers: ERS4859818-ERS4859933 Mice samples accession numbers: ERS4859934-ERS4859966
Experimental Models: Organisms/Strains		
Mouse C57BL/6J	Charles River	N/A
Drosophila Melanogaster	BestGene Inc.	N/A
Lactococcus Lactis phage 936	Félix d'Hérelle Reference Center for Bacterial Viruses from the Université Laval (Québec, Canada)	HER Number: 203
Software and Algorithms		
SPSS software (version 19)	IBM	https://www.ibm.com/analytics/spss-statistics-software
Rstudio (version 1.3.959)	Rstudio Team	https://rstudio.com/
R (version 3.6)	R	https://www.r-project.org/
Prinseq-lite-0.20.4	(Schmieder and Edwards, 2011)	http://prinseq.sourceforge.net/
FLASh 1.2.11	(Magoč and Salzberg, 2011)	https://ccb.jhu.edu/software/FLASH/
Bowtie2-2.3.4.3	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
MEGAHIT v1.1.2	(Li et al., 2015)	https://github.com/voutcn/megahit
Prodigal v2.6.342	(Hyatt et al., 2010)	https://github.com/hyattpd/Prodigal

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HMMER	(Durbin et al., 1998)	http://hmmer.org/
Kaiju v1.6.2	(Menzel et al., 2016)	https://github.com/bioinformatics-centre/kaiju
Geneious	(Kearse et al., 2012)	http://nebc.nerc.ac.uk/news/geneiousonbl
Virsorter 2.0	(Roux et al., 2015)	https://github.com/simroux/VirSorter
Genemark	(Mills et al., 2003)	http://opal.biology.gatech.edu/GeneMark/
WiSH	(Galiez et al., 2017)	https://github.com/soedinglab/WIsH
SPACEPharer	(Zhang et al., 2021)	https://github.com/soedinglab/spacepharer
STAR software (version 2.5.3a)	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
Subread version (1.5.1)	(Liao et al., 2014)	http://subread.sourceforge.net/
Limma (version 3.30.13)	(Smyth et al., 2005)	https://bioconductor.org/packages/release/bioc/html/limma.html
edgeR (version 3.26.8)	(Robinson et al., 2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html
DESeq2 (version 1.26.0)	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
ALDEx2 (version 1.18.0)	(Fernandes et al., 2014)	https://www.bioconductor.org/packages/release/bioc/html/ALDEx2.html
SGoF (version 2.3.2)	(Carvajal-Rodríguez et al., 2009)	https://cran.r-project.org/web/packages/sgof/index.html
STRING (version 11.0b)	(Szklarczyk et al., 2019)	https://string-db.org/
Consensus Pathway Data Base (CPDB)	(Kamburov et al., 2013)	http://cpdb.molgen.mpg.de/
Boruta (version 6.0.0)	(Kursa and Rudnicki, 2010)	https://cran.r-project.org/web/packages/Boruta/
Other		
Avance III 600 spectrometer	Bruker	N/A
5Mmm PABBO gradient probe	Bruker	N/A
Dual energy X-ray absorptiometry	GE Healthcare	N/A
Cobas 8000 c702 analyzer	Roche Diagnostics	N/A
ADAM®A1c HA-8180V	ARKRAY, Inc	N/A
Shuttle chamber LE918	Panlab	N/A
Bioanalyzer 2100	Agilent	N/A
ABI 7900HT qPCR	Applied Biosystems	N/A
HiSeq 2500	Illumina	N/A
Qubit 3.0 fluorometer	Thermo Fisher Scientific	N/A
NextSeq 500	Illumina	N/A
LightCycler 480 II	Roche	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, José Manuel Fernández-Real (jmfreal@idibgi.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The raw metagenomic sequence data derived from human samples in the IRONMET cohort and mouse samples have been deposited in the European Nucleotide Archive (ENA) an accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Clinical study

Longitudinal discovery cohort (IRONMET, baseline: n=114, 1-year follow-up: n=86)

The discovery cohort is a longitudinal study including participants with and without obesity in order to evaluate differences in cognitive function between groups and their interactions with the gut microbiome. As a pilot study, it was foreseen to recruit 120 subjects, 60 with obesity (20 men, 20 pre- and 20 postmenopausal women) and 60 without obesity (20 men, 20 pre- and 20 postmenopausal women). Finally, 180 subjects were enrolled with an increase of around 10 subjects per groups to prevent data losses. Only subjects with incomplete data were excluded from the analysis. The cross-sectional case-control study was undertaken from January 2016 to July 2018 in the Endocrinology Department of Dr. Josep Trueta University Hospital of Girona, Spain. The visit included a medical history, physical examination, dietary assessment, DEXA and MRI explorations and a neuropsychological assessment conducted by a neuropsychologist. Eligible subjects were subjects with obesity (body mass index, BMI $\geq$30 kg/m^2) and age- and sex-matched non-obese subjects (BMI 18.5-<30kg/m2). Exclusion criteria were: type 2 diabetes mellitus, chronic inflammatory systemic diseases, acute or chronic infections in the previous month; use of antibiotic, antifungal, antiviral or treatment with proton-pump inhibitors in the previous three months; severe disorders of eating behaviour or major psychiatric antecedents; neurological diseases, history of trauma or injured brain, language disorders; and excessive alcohol intake ($\geq$40 g OH/day in women or 80 g OH/day in men).The institutional medical ethical committee approved the study, and informed written consent was provided by all participants.

The longitudinal cohort included a subsample of 86 participants who accepted to participate in a longitudinal visit and met the same exclusion and inclusion criteria as the baseline visit (see above). After 1 year of follow-up, the all the explorations included in the first visit were performed. The longitudinal study was undertaken from February 2017 to April 2019 in the same facilities of the Endocrinology Department of Dr. Josep Trueta University Hospital of Girona, Spain. The institutional medical ethical committee approved the study, and informed written consent was provided by all participants.

Validation cohort 1 (IMAGEOMICS, n=942)

The Aging 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 (Puig et al., 2020). Briefly, the MESGI50 cohort included a population aged $\geq$ 50 years old, 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. The Aging Imageomics Study protocol was approved by the ethics committee of the Dr. Josep Trueta University Hospital.

Validation cohort 2 (n =31)

This is an observational prospective case-control study which included 31 subjects (24 with obesity) recruited between March 2019 and February 2020 in the Endocrinology Department of Dr. Josep Trueta University Hospital of Girona, Spain. The aim of this study was to replicate the findings of the prior study (discovery cohort), assessing the interplay between the gut microbiota and cognitive dysfunction in subjects with obesity. Therefore, subjects with obesity (body mass index, BMI $\geq$30kg/m^2) and age- and sex-matched non-obese subjects (BMI 18.5-<30kg/m2) were eligible. The initial idea was to recruit the same number of subjects as the discovery cohort. However, due to Covid-19 pandemic we were forced to suspend the recruitment. Exclusion and inclusion criteria were also described above as well as the examinations. 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.

Validation cohort 3 (n=26)

As a part of an entire project to evaluate the role of intestinal microflora in non-alcoholic fatty liver disease; a substudy was conducted to compare differences between subjects with and without obesity including as well as neuropsychological examination. Finally, n=14 participants with obesity (BMI $\geq$30 kg/m^2) and n=12 without obesity (BMI <30 kg/m2) were included. The exclusion criteria were systemic diseases, infection in the previous month, serious chronic illness, ethanol intake >20 g/day or use of antibiotic, antifungal or antivirals and PPIs in the previous three months. All control subjects were normotensive and were selected on the basis of similarity in age and sex compared with subjects with obesity and the absence of a personal history of inflammatory diseases or current drug treatment. The recruitment was conducted at the Endocrinology Service of the Dr. Josep Trueta University Hospital of Girona, Spain, from 2014-2017. The visit included a neuropsychological assessment, medical history, physical exploration and DEXA and MRI examinations. All subjects gave written informed consent, validated and approved by the ethical committee of the Dr. Josep Trueta University Hospital (Girona, Spain).

Clinical and Laboratory Parameters

Body composition was assessed using a dual energy X-ray absorptiometry (DEXA, GE lunar, Madison, Wisconsin). Fasting plasma glucose (FPG), lipids profile and high-sensitivity C-reactive protein (hsCRP) levels were measured using an analyser (Cobas®8000 c702, Roche Diagnostics, Basel, Switzerland). Glycated hemoglobin (HbA1c) was determined by performance liquid chromatography (ADAM_A1c HA-8180V, ARKRAY, Inc., Kyoto, Japan).

Dietary pattern

The dietary characteristics of the subjects were collected in a personal interview using a validated food-frequency questionnaire (Vioque et al., 2013).

Depressive symptomatology

The Patient Health Questionnaire-9 (PHQ-9) was used to measure the frequency and severity of depressive symptomatology during the past 32-weeks. The PHQ-9 is a common instrument used as depression screening tool in with a cut-off $\geq$10, which has a sensitivity and specificity of 0.88 and 0.85, respectively, to detect a major depressive episode (Levis et al., 2019). The PHQ-9 includes nine questions to quantify the frequency over the last two weeks of nine symptoms derived from diagnostic criteria for a major depressive episode of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) ((American Psychiatric Association, 2013)). The PHQ-9 score can range between 0 and 27, higher scores indicate higher severity.

Animal studies

Male C57BL/6J mice (Charles River, France), weighing 23–26 g at the beginning of the experiment were used in this study. Mice were housed individually in controlled laboratory conditions with the temperature maintained at 21 ± 1_C, humidity at 55 ± 10%, and 7 h30/19 h30 light/dark cycles. All animals were fed a standard chow diet RM1 (Irradiated Vacuum packed, Dietex International Ltd.). 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. Animal procedures were conducted in strict accordance with the guidelines of the European Communities Directive 86/609/EEC regulating animal research and were approved by the local ethical committee (CEEA-PRBB). All the experiments were performed under blinded conditions (the researcher who administered the microbiota was blinded in relation to the memory scores of the subjects who provided the faeces). Mice were given a cocktail of ampicillin and metronidazole, vancomycin (all at 500 mg/L), ciprofloxacin HCl (200 mg/L), imipenem (250 mg/L) once daily for 14 consecutive days in drinking water, as previously described (Kelly et al., 2016). Seventy-two h later, animals were colonized via daily oral gavage of donor microbiota (150 mL) for 3 days. Animals were orally gavaged with saline (n = 11) and faecal material from healthy volunteers with a wide range of clr-transformed bacteriophages levels and cognitive scores (n=22). In particular, patients were selected so that n=11 patients had low memory scores and n=11 patients with high memory scores in the in the CVLT immediate recall. In each group, patients had a wide range of scores so as to have a continuous scale of scores when both groups were considered together. Both groups were matched for age, BMI, sex, and years of education. To offset potential confounder and/or cage effects and to reinforce the donor microbiota phenotype, booster inoculations were given twice per week throughout the study. Animals were exposed to a series of behavioural testing including novel object recognition (NOR) test and fear conditioning with nociception assessed by the hot plate test to ensure specificity.

All animals were immediately sacrificed by decapitation after the last behavioural test. The cecum was removed, weighted and stored, and the faeces collected and stored at -80°C for further microbiota analysis. The mice brains were quickly removed and the medial prefrontal cortex was dissected according to the atlas of stereotaxic coordinates of mouse brain (Paxinos and Franklin, 1997). Brain tissues were then frozen by immersion in 2-methylbutane surrounded by dry ice, and stored at -80°C.

Drosophila studies

Drosophila melanogaster *stocks and maintenance*

The *Drosophila* wild-type strain was originally obtained from Bestgene. The final stock has been obtained by exchanging the w- allele of the strain that the company regularly uses to inject P-element-based transgenes by a w+ allele (Castells-Nobau et al., 2019).

Flies were raised and maintained on standard diet (SD) (1L: yeast 27,5 g, yellow cornmeal 52 g, sugar 110 g, agar 5 g, propionic acid 5 ml, 0,1% methylparaben in ethanol 5 ml) or whey diet (WD) (1L: Whey 30 g, yellow cornmeal 52 g, sugar 110 g, agar 5 g, propionic acid 5 ml, 0,1% methylparaben in ethanol 5 ml). Standard diet sterile (SD 121 °C) and whey diet sterile (WD 121 °C) were generated following the same recipe as SD and WD respectively and sterilized in the autoclave for 20 min at 121°C. To generate WD 100°C and WD 45°C, 30g/L of whey powder were added to the food mixture at 100 °C and boiled for 3 min or 45 °C respectively. To generate SD + *phage 936* and WD + *phage 936* and the corresponding control diets SD sterile or WD sterile, first SD and WD were sterilized in the autoclave for 20 min at 121°C. SD + *phage 936* and WD + *phage 936* were supplemented with a droplet of 50 µl containing 50.000 PFU of *bacteriophage 936*. The phage was added to the food surface when the food was at room temperature (21 °C). Flies were raised and maintained thought-out their entire development in the corresponding dietary condition (SD, WD, SD + *phage 936* and WD + *phage 936*), adult flies were transferred to a fresh food vial containing the corresponding diet every 4 days. Fly stocks were maintained at 25 °C, in a 12:12 hours light-dark cycle. *Lactococcus lactis phage 936* (HER Number: 203) was obtained from Félix d'Hérelle Reference Center for Bacterial Viruses from the Université Laval (Québec, Canada).

METHODS DETAILS

Neuropsychological assessment in humans

Backward digit span

The Digit Span is a subtest of the Wechsler Adult Intelligence Scale-III (WAIS-III) (Wechsler, 2012). It is based on numbers and includes the Forward and the Backward Digit Span tests. In the Backward Digit Span task, the examinee repeats in reverse order series of digits that became gradually longer. This is an executive task particularly dependent on working memory. A higher score reflects a better working memory. In a standardization sample of 394 participants (aged 16-89 years), the reliability coefficient was very high, ranging from 0.94-0.97 (Strauss et al., 2006).

Trail making tests B

The Trail Making Test (TMT) is composed by the subtests TMTA and TMTB. The TMTA (TMTA, greater focus on attention) consisted of a standardized page in which numbers 1 to 25 are scattered within the circles, and participants were asked to connect the numbers in order as quickly as possible. Before starting the test, a 6-item practice test was administered to the participants to make sure they understood both tasks. A maximum time of 300 s was allowed before suspending the test. The direct scores of TMTA were the time in seconds taken to complete each task. In the same way, TMTB (Trail B, greater focus on executive function) consisted of an alternating sequence of numbered circles and letters (Corrigan and Hinkeldey, 1987; Lezak, 1984). In both tests, shorter times to completion indicate better performance.

California Verbal Learning Test-II (CVLT)

CVLT is used to assess verbal learning and memory (Delis et al., 2000). The CVLT Long Delayed Free Recall scores assesses long-term memory. After 20 minutes, in which non-verbal tasks are carried on, the subjects are asked to repeat a list of word that the examinee has previously presented, without semantic facilitation. A higher score reflects a better memory function. About 50 min are necessary to administrate this test and its reliability ranges from 0.78-0.94 (Paolo et al., 1997).

Symbol Digit Test (SDT)

The SDT is a subtest of the Wechsler Adult Intelligence Scale-III (WAIS-III). It includes a coding key that shows nine abstract symbols, each paired with a number and below the key, series of symbols were presented. The participants were asked to write down the corresponding numbers associated with the abstract symbols as quickly as possible. This task is a measure of information processing speed. The number of correct substitutions during a 90-second interval was scored, and higher numbers indicate better performance.

Phonemic Verbal Fluency (PVF)

The PVF is a spontaneous verbal production task that consists to produce words with a specified letter (P) during one minute. This task is a measure of language ability and executive function and is influenced by processing speed. The number of correct words was scored and higher numbers indicate better performance.

General Cognition

A composite cognitive score was computed as the sum of the standardized scores of each individual for each test included in the IMAGEOMICS cohort.

Extraction of faecal genomic DNA and Whole-Genome Shotgun Sequencing

Total DNA was extracted from frozen human stool using the QIAamp DNA mini stool kit (QIAGEN, Courtaboeuf, France) following the manufacturer's instructions. 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/μl) was used for shotgun library preparation for high-throughput sequencing, using the Nextera DNA Flex Library Prep kit (Illumina, Inc., San Diego, CA, USA) according to the manufacturer's 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 (Schmieder and Edwards, 2011) and overlapping pairs were joined using FLASH-1.2.11 (Magoč and Salzberg, 2011). Fastq files, converted into fasta files, were mapped against the reference human genomes (GRCh38.p11, Dec 2013 and GRCm38.p6, Sept 2017), respectively, to remove reads from host origin, by using bowtie2-2.3.4.3 (Langmead and Salzberg, 2012) with end-to-end and very sensitive options. Next, functional analyses were carried out by assembling the host-free reads into contigs by MEGAHIT v1.1.2 (Li et al., 2015) and mapping those against the contigs with bowtie2. Reads that did not assemble were appended to the contigs. Then, prediction of codifying regions was implemented by Prodigal v2.6.342 (Hyatt et al., 2010), and subsequent functional annotation was carried out with HMMER (Durbin et al., 1998) against the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database, version 2016 (Kanehisa and Goto, 2000) to obtain the gene functional annotation. The best annotations were filtered and orf annotation were assigned to every read using the statistical package R 3.1.0 (R Core Team, 2013) which also was used to count the aligned reads, to add the category and its coverage, and finally to build abundance matrices. Taxonomic annotation was implemented with Kaiju v1.6.2 (Menzel et al., 2016) on the host-free reads, using a greedy mode. Addition of lineage information, counting of taxa and generation of an abundance matrix for all samples were performed using the package R. Finally, non-viral taxa were excluded for the downstream analyses. The average number of reads per sample were 17M.

Classified viral reads from different samples previously analysed by Kaiju that correlated with cognitive function were further co-assembled by Megahit and Geneious (Kearse et al., 2012) software assemblers using default parameters delivering very small contigs (<2.5 kb size). Then, these contigs were used as query to search for larger genome assembled viral contigs by MEGAHIT in gut assembled samples by using stand-alone BLASTn (min_perc_identity 99; evalue 0.00001). Genome annotation of those detected large assembled contigs of *Caudovirales* and *Microviridae* that correlate with cognitive function were further analysed with Virsorter version 2.0 to identify viral hallmark genes (Roux et al., 2015). ORFs were predicted with a combination of Genemark and Prodigal (Hyatt et al., 2010; Mills et al., 2003). Search of putative host was performed with WiSH program with default parameters (Galiez et al., 2017) that find the most likely assignment based on similar k-mer nucleotide frequency signatures between a virus and its host. Search of CRISPR spacers matching to viral protospacer was carried out with SPACEPharer (Zhang et al., 2021).

^{1}H-NMR metabolomics analyses

Plasma samples were thawed at room temperature. For each sample, 400 μL of plasma were combined with 200 μL of phosphate buffer ((9%w/V NaCl, 100% D_2O) that contained 10mM of 3-trimethylsilyl-1-[2,2,3,3-2H4] (TSP). Samples were mixed with the use of a vortex and centrifuged (10.000 x g) for 10 min. Then, a 550 μL aliquot was transferred into a 5 mm NMR tube prior to NMR analysis. ^{1}H spectra of low molecular weight metabolites were performed using a CPMG sequence (RD–90°–[τ–180°– τ]$_n$–ACQ-FID) with spin-echo delay of 400 μs (for a total T2 filter of 210 ms) allowing an efficient attenuation of the lipid NMR signals. The CPMG sequence generates spectra edited by T2 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 μs. 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 μL of 0.05 M PBS buffer in H_2O (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 μL of prepared faecal water was placed in appropriate 2 ml Eppendorf tube and then, 400 μL of 0.05M PBS buffer in D_2O (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}H pulse experiments were carried out using the NOESY-presaturation sequence [recycle delay (RD)-90°–τ1–90°–τm–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}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.

Animal studies

Behavioural Testing in Mice

The Novel Object Recognition (NOR) Test. The NOR was performed in a V-maze as previously published (Burokas et al., 2014). Three phases of 9-min were performed on consecutive days. Mice were first habituated to the V-maze. On the second day, 2 identical objects (chess pieces) were presented to the mice, and the time that they spent exploring each object was recorded. In the test phase (3 h later for short-term memory or 24 h later for long-term memory), 1 of the familiar objects was replaced with a novel object (a different chess piece), and the time spent exploring each object (novel and familiar) was computed. A discrimination index was calculated as the difference between the times that the animal spent exploring the novel (Tn) and familiar (Tf) object divided by the total time of object exploration: (Tn-Tf)/(Tn + Tf).

Cued-Induced Fear Conditioning. The cue-induced fear conditioning is a well-recognized model of emotional memories (Sun et al., 2020). Fear conditioning was conducted as described previously with some modifications (Saravia et al., 2019). 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 mouse prefrontal cortex

Sample preparation

For RNA preparation, each brain sample was treated individually. Total RNA isolated from the brains using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen Düsseldorf, Germany) according to the manufacturer's protocol.

RNA quality control

Quality control of the RNA was performed using the RNA 6000 Nano chip (Agilent) on an Agilent Bioanalyzer 2100 obtaining RIN values between 8.7 - 9.8.

RNA libraries

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 min. Prepared libraries were analysed 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 (Dobin et al., 2013) 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 (Liao et al., 2014).

Drosophila melanogaster studies

Aversive taste memory

Flies can detect tastants by gustatory receptors located on the mouth (proboscis) and legs (tarsi) (Masek and Keene, 2016). The detection of sweet tastes promotes feeding, inducing the proboscis extension reflex (PER), a robust innate behaviour (Kirkhart and Scott, 2015; Masek et al., 2015). On the other hand, the detection of bitter compounds inhibits feeding and bitter-tasting is a potent inducer of aversive taste memory, which enables animals to modify food choice by prior experience (Keene and Masek, 2012).

The aversive taste memory paradigm was adapted from the previously described (Kirkhart and Scott, 2015; Masek et al., 2015). Ten-day old mated female flies were collected, placed on fresh food for 24 h, and starved for 40 h in a vial on a wet Kimwipe paper. Flies were then anesthetized with CO_2 and glued to a glass slide with nail polish by their thorax, any fly with their tarsi or proboscis glued to the nail polish was removed from the study. Flies were left to recover for at least 2,5 hours in a humid chamber at 25C. Before training flies were water-satiated and were presented a stimulus with 100 mM sucrose on the tarsi three times (Pre-test). Any fly that did not show a robust proboscis extension response (PER) to each sucrose stimulation of the pre-test was removed from the study. During training, each fly was summited with 5 training trials with 3 min inter-trial interval, a droplet of 100mM sucrose was presented on their tarsi quickly followed by a stimulation of 50 mM quinine on the proboscis. Every 2 training trials, the flies were able to consume water. After training, memory was tested, flies were presented with only a droplet of 100 mM sucrose on their tarsi. Memory tests were performed with an inter-trial interval of 2 min, during a period of 30 min. Every five-memory trial, flies were able to consume water. The number of PERs of each fly were annotated during the learning trials and the memory test, a value of 1 was given for full extension of the proboscis, 0,5 for a partial extension of the proboscis, and 0 for suppression of proboscis extension. The percentage of PER (PER %) was calculated per condition summing all flies tested within a minimum of 4 independent experiments with 10-15 flies each. The average percentages of PER for the total number of flies are reported on the graphs, the average percentage of PER was calculated doing the average of PER (%) between the trial of interest, the anterior and the posterior trials. Differences in PER frequency between groups were compared using Fisher's exact test, for categorical data.

DNA extraction and PCR

To determine the presence of phage 936 family in the different diets and the whey powder, 200 mg of food were collected. DNA extraction was performed following the phenol-chloroform method. A total of 200 mg of each sample were added to 250µl of Lysis buffer (TrisHCl 0,1M (pH9.0), EDTA 0,1M and SDS 1%) and incubated for 30 min at 65°C. To precipitate DNA, 56 µl of sodium acetate 5M pH 5 where added, vortexed for 1 min and incubated in ice for 15 min. 250µl of buffer-saturated phenol/chloroform (pH 7.5-7.8) were added to each sample and centrifuged for 10 min at 6.000 rpm twice. The aqueous layer was transferred to a 1.5 ml tube containing 150 µl of isopropanol. The sample was centrifuged (12 min, 13.000 rpm), and the pellet washed with70% ethanol. The pellet was air-dried and resuspended in 50 µl of Milli-Q water by vortexing and stored at -20°C prior to use.

mRNA extraction and cDNA synthesis

Ten-day old wild-type pre-mated females were selected for qRT-PCR. A total of ten adult heads were collected per sample and transferred to RNAlater solution (Sigma). Total RNA was extracted using the Arcturus PicoPure RNA Purification kit (Thermo Fisher Scientific). To avoid amplification from genomic DNA, DNase treatment was performed using the DNA-free Ambion kit and RNA was reverse transcribed into cDNA using the High-Capacity cDNA reverse transcription (Applied Biosystems) according to manufacturers' procedures.

Quantitative real-time PCR (qRT-PCR)

Quantitative PCRs (qPCRs) were performed using the LightCycler 480 SYBR Green Master (Roche) on an LightCycler 480 II machine (Roche). Primer sequences of the analysed and the reference genes RNApol2 and 1tub23cf are provided in Table S4B. Each sample contained a pool of 10-15 fly heads. For each condition, a minimum of 7 biological replicates and two technical replicates were analysed. Differential gene expression was calculated using the $2\Delta\Delta Ct$ method (Livak and Schmittgen, 2001). The average Ct value for each sample was calculated and subtracted from the geometric mean Ct value of the reference genes RNApol2 and 1tub23cf to calculate the ΔCt value (Qureshi and Sacan, 2013). Brown-Forsythe and Welch ANOVA tests (GraphPad Prism version 9.00 for Mac) were employed for calculations of P-values.

QUANTIFICATION AND STATISTICAL ANALYSIS

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 χ^2 for categorical variables, unpaired Student's t test in normal quantitative and Mann-Whitney U test for non-normal quantitative variables. Partial Spearman's correlation analysis was used to determine the correlation between clr-transformed specific *Caudovirales* (*Siphoviridae*, *Demerecviridae*, *Drexlerviridae*), *Siphoviridae* and *Microviridae* levels and cognitive tests after controlling for covariates. As it works by obtaining the residuals of the ranked variables after removing the effect of the ranked covariates, scatter plots were generated with the ranked residuals of the model adjusting for selected covariates. Non-parametric monotonic trends according the *Siphoviridae*

tertiles for the variables of interest were assessed by the Mann-Kendall trend test. As, sexual dimorphism has shown a strong impact on the gut microbiome, we also performed sex-specific analyses (Mayneris-Perxachs et al., 2020). Theses statistical analyses were performed with SPSS, version 19 (SPSS, Inc, Chicago, IL) or R. 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, *specific Caudovirales* and *Microviridae* raw counts were applied a centred log-ratio (clr) transformation using the "ALDEx2" R package (Fernandes et al., 2014). It first uses a Dirichlet-multinomial model to inter 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 differentially associated with the clr-transformed bacteriophage levels were identified using the "DESeq2" R package (Love et al., 2014), adjusting for age, body mass index, sex, and education years. A generalized linear model (GLM) was fitted modelling read counts (K_{ij}) for each taxa (i) and sample (j) as following a negative binomial distribution. The GLMs were used with a logarithmic link: $logq_{ij} = \sum x_{jr}\beta_{ir}$, where x_{jr} represents the covariate r in sample j and β_{ir} are the coefficients (logarithmic fold change) for taxa i in covariate r. For each taxa, we controlled for age, BMI, sex, and education years. Therefore, our models were the following:

$$logq_{ij} = \beta_{age,i}age_j + \beta_{sex,i}sex_j + \beta_{BMI,i}BMI_j + \beta_{edu,i}edu_j + \beta_{Siphoviridae,i}Siphoviridae_j$$

Taxa and bacterial functions were previously filtered so that only those with more than 10 reads in at least two samples were selected. The p-values for bacterial taxa were then adjusted (p adj) for multiple comparisons using the Benjamini-Hochberg procedure for False Discovery Rate (FDR). For bacterial functions, that included a higher number of tests (5714 bacterial functions were included after filtering), p-values were adjusted for multiple comparisons using a Sequential Goodness of Fit (Carvajal-Rodríguez et al., 2009) 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 padj<0.1.

Metabolomics analysis

Metabolomics data were first normalized using a probabilistic quotient normalisation. Then 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 (Kursa and Rudnicki, 2010). It has been recently proposed as one of the two best-performing variable selection methods making use of RF for high-dimensional omics datasets (Degenhardt et al., 2019). The Boruta algorithm is a wrapper algorithm that performs feature selection based on the learning performance of the model (Kursa and Rudnicki, 2010). It performs variables selection in three steps: a) Randomization, which is based on creating a duplicate copy of the original features randomly permutate 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 permutate 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 (ntree), and a number of features randomly sampled at each split given by the rounded down number of features/3 (the mtry recommended for regression).

RNA-seq analysis

Differential expression gene analyses were performed on gene counts using the "limma" R package (Ritchie et al., 2015). First, low expressed genes were filtered, so that only gene with more than 10 reads in at least 2 samples were selected. After filtering 15,546 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 (Robinson et al., 2010). Normalized counts were then converted to $\log_2$ 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, and education years 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. P-values were adjusted for multiple comparisons using a SGoF procedure and statistical significance was set at padj<0.05.

Differentially expressed genes were 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 (Szklarczyk et al., 2019). Then, functional local clusters in the interaction network were determined using a Markov Cluster algorithm (MCL) with an 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. In addition, the functional roles of differentially expressed genes were characterized using over-representation analyses based on the Gene Ontology, Pathway Interaction Database (PID), and Wikipathways 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.