Oral microbiome profiles relate to periodontal disease and brain health - a microbiome network analysis in the Hamburg City Health Study

Marvin Petersen[1*], Carolin Walther[2*], Katrin Borof[2], Guido Heydecke[3], Thomas Beikler[2], Malik Alawi[4], Christian Müller[4], Felix L. Nägele[1], Birgit-Christiane Zyriax[5], Jens Fiehler[6], Jürgen Gallinat[7], Simone Kühn[7,8], Raphael Twerenbold[9,10,11,12], Corinna Bang[13], Götz Thomalla[1], Bastian Cheng[1#], Ghazal Aarabi[2#]

[1]*Department of Neurology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany.*

[2]*Department of Periodontics, Preventive and Restorative Dentistry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[3]*Department of Prosthetic Dentistry, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[4]*Bioinformatics Core, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[5]*Midwifery Science-Health Services Research and Prevention, Institute for Health Services Research in Dermatology and Nursing (IVDP), University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[6]*Department of Neuroradiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[7]*Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[8]*Lise Meitner Group for Environmental Neuroscience, Max Planck Institute for Human Development, Berlin, Germany*

[9]*Department of Cardiology, University Heart and Vascular Center, Hamburg, Germany*

[10]*Epidemiological Study Center, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

[11]*German Center for Cardiovascular Research (DZHK), partner site Hamburg/Kiel/Luebeck, Hamburg, Germany*

[12]*University Center of Cardiovascular Science, University Heart and Vascular Center, Hamburg, Germany*

[13]*Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany*

**shared first authorship*

[#]shared last authorship

Correspondence to:

Marvin Petersen

Department of Neurology

University Medical Center Hamburg-Eppendorf

Email: mar.petersen@uke.de

ORCID: 0000-0001-6426-7167

and

Ghazal Aarabi

Department of Periodontics, Preventive and Restorative Dentistry

University Medical Center Hamburg-Eppendorf

Email: g.aarabi@uke.de

ORCID: 0000-0001-5484-2594

Keywords: Oral microbiome; Periodontitis; Brain health; Cognition; Brain structure; Neuroimaging; Mental health; Inflammation; Nutrition

Abstract

The relationship between oral microbiome composition and brain health in the general population remains poorly understood. In this study, we inferred a microbiome similarity network based on 16S rRNA sequencing data of crevicular fluid collected from 1,026 participants in the Hamburg City Health Study, which revealed a continuous disease gradient mirroring the microbial pathogenicity spectrum of periodontitis. Leveraging this network, we systematically examined associations between periodontal microbiome profiles and 37 brain health-related phenotypes, including cognitive function, brain structure, mental health, inflammatory biomarkers, diet, vascular risk factors, and demographics. We found that higher abundance of periodontitis-related microbial taxa was linked to poorer cognitive performance, lower cortical thickness and subcortical volumes, and elevated leukocyte counts, after adjusting for covariates. Notably, we identified both, previously known as well as novel microbial associations with brain health phenotypes. These findings advance the understanding of the oral microbiome-brain axis, suggesting mechanistic links between periodontal health and brain function with potential implications for future causal and interventional studies.

Introduction

The human oral microbiome harbors a diverse community of microorganisms that influences the health and well-being of their hosts.[1] In recent years, periodontitis, which is the inflammatory disruption in the host–microbial homeostasis of the periodontal pocket, has gained increasing attention as a key factor impacting brain health. Studies indicate that periodontitis and the linked bacterial communities are associated with the incidence of cognitive decline and Alzheimer's dementia.[2] Given its estimated prevalence between 10% and 50% in elderly people, this renders periodontitis a relevant public health concern.[3] As therapeutic interventions can alter the progression of periodontitis, comprehending its impact on the brain is vital for effective prevention and management of cognitive sequelae.

Among bacterial species of the subgingival biofilm, several are known to be associated with periodontitis including: *Porphyromonas gingivalis*, *Tannerella forsythia* and *Treponema denticola*.[4,5] Mechanistic models have been proposed to explain the connection between these bacterial communities and cognitive decline. The connection is considered to arise from bacterial species promoting systemic inflammation, neurodegenerative processes, blood-brain barrier disruption, and endothelial dysfunction.[6]

Despite existing research efforts, our understanding of the association between oral microbiome composition and brain health remains limited. Recent studies predominantly focus on clinical Alzheimer's disease and mild cognitive impairment cohorts, with a marked gap in research exploring this link within the general population.[2,7–10] Additionally, many investigations are constrained by small sample sizes and a lack of multi-modal imaging and biomarker data potentially leading to

inconsistent findings.[2] These issues are further complicated by the inherent complexity of microbiome data.

Tapping into these research needs, we aim to advance the understanding of the oral microbiome-brain axis, applying advanced analysis techniques capable of integrating large-scale, multi-domain data that are crucial for achieving robust and interpretable insights. We analyze large-scale population-based data of n=1,026 individuals from the Hamburg City Health Study.[11] Our “big data” approach focuses on a topology-based analysis that integrates oral microbiome abundance data derived from 16S rRNA sequencing of subgingival samples with in-depth clinical and lifestyle data including oral health assessments, cognitive test scores, mental health scores, neuroimaging, circulating inflammatory markers, dietary patterns, and vascular risk measures.

Results

Our methodology, which integrates microbiome, clinical and lifestyle data into a unified analysis framework, is illustrated in *figure 1*. Quality assessment of oral microbiome abundance and brain MR imaging data resulted in a final analysis sample of 1,026 individuals ([mean ± SD] age 63.72 ± 8.2 years, 42.7% female; for details see *table 1*).

A similarity network of subgingival bacterial communities

We applied an unsupervised, topology-based technique to abundance data of 85 different bacterial genera to infer a low-dimensional network representation of oral microbiome similarity. The resulting microbiome similarity network consisted of 528 nodes – representing participant groups with highly similar oral microbiome compositions – and 4,061 edges – connecting nodes that share at least one participant. For a visualization of the network see *figure 2a*.

In an enrichment analysis, we examined whether the microbiome similarity network captures inter-individual differences in participant traits. Therefore, we computed enrichment scores, which are node-level indices that indicate whether participant traits are significantly higher or lower than expected by chance in specific regions, i.e., participant clusters, of the network. Enrichment scores were obtained for genus-level abundance data as well as all non-microbiome phenotypes. Furthermore, we calculated the enrichment ratio (= number of significantly enriched nodes / total number of nodes) based on the enrichment scores for each phenotype quantifying the overall amount of enrichment on the microbiome similarity network. This measure indicates how strongly the network-based organization of participants reflects inter-individual variance in a respective phenotype.

Refer to *figure 2b and 2c* for an exemplary display of *Porphyromonas* abundance and the corresponding enrichment map on the microbiome similarity network.

Enrichment analysis of genus-level abundance

Microbiome abundance showed significant enrichment across all detected genera ranging from 43% to 87%. *Figure 3a* displays the enrichment ratios alongside mean relative abundance and phylogenetic associations. We performed a dominance analysis by identifying for each network node the bacterial genus with the highest enrichment scores (*figure 3b*). Of the 85 investigated genera, 17 showed the highest enrichment score for at least 5 nodes. Distribution of dominant genera in the network representation of oral microbiome similarity followed a horizontal pathogenicity gradient: Bacterial genera with strongest enrichments at the *left* end of the microbiome similarity network were periodontitis-associated taxa including *Dialister* (n_{nodes} = 6), *Fusobacterium* (134), *Parvimonas* (19), *Porphyromonas* (6), and *Saccharibacteria* (TM7) [G-5] (48). In the center of the network, *Arachnia* (14), *Gemella* (5), *Neisseria* (25), *Ruminococcaceae* [G-1] (26), *Selenomonas* (8), and *Weeksellaceae* [G-1] (10) exhibited highest enrichment scores. At the right end, genera with strongest enrichments were of low periodontal pathogenicity or related to other dental diseases including *Enterococcus* (42), *Haemophilus* (28), *Lactococcus* (7), *Rothia* (20), *Streptococcus* (48) and *Veillonella* (50).

Enrichment analysis of non-microbiome phenotypes

All non-microbiome phenotypes showed significant enrichment on the microbiome similarity network including measures of oral health status, cognition, mental health, brain structure, circulating inflammatory markers, diet, vascular risk factors and demographics (*figure 4*). The top 10 ranking non-microbiome phenotypes were

plaque index (78.8%), leukocytes (72.0%), general cognitive ability (68.8%), clinical attachment loss (68.2%), gray matter free-water (66.1%), smoking behavior (64.2%), information processing speed (62.7%), executive function (61.1%), Mini Mental Health Exam (59.5%), and decayed/missing/filled teeth (DMFT) index (59.2%) (*figure 4a*). Complementary microbiome covariate identification using a linear approach (*envfit*) confirmed these results. Among the phenotypes with the top 10 highest enrichment ratios, all but information processing speed showed a significant linear association ($p_{FDR} < 0.05$) with genus-level principal coordinate analysis (PCoA) ordination of the Bray-Curtis dissimilarity (*supplementary figure S1)*.

Based on the enrichment maps, we assessed inter-participant differences of phenotypes. Phenotype enrichment maps are presented in *figure 4b*. Most participant traits varied along the microbiome similarity network in a linear left-right trajectory aligning with the pathogenicity gradient identified via dominance analysis.

Put differently, there was a gradient of enrichment from left to right that reflected the transition of participant characteristics. Participants at the left end of this gradient exhibited higher severity of clinical periodontitis, elevated circulating inflammatory markers, older age, a higher percentage of smokers (indicated by significant positive enrichment), as well as lower cognitive function, lower brain structural integrity, a less healthy diet, a lower percentage of females, and lower education levels (indicated by significant negative enrichment). Conversely, the right end of the gradient featured participants with opposite traits: lower severity of clinical periodontitis, lower circulating inflammatory markers, younger age, fewer smokers, as well as higher cognitive performance, higher brain structural integrity, a healthier diet, a higher percentage of females, and higher education levels. Mental health

scores and vascular risk factors beyond smoking displayed a non-linear enrichment pattern not aligned with the pathogenicity gradient.

Co-enrichment analysis

Principal component analysis of enrichment scores revealed that enrichment patterns of phenotypes differed along two dominant axes of inter-phenotype variation explaining 60.67% (principal component 1, PC1) and 11.87% (principal component 2, PC2) of variance, respectively. Phenotypes with similar enrichment patterns co-localized in the principal component space formed by these axes (*figure 5a, supplementary figure S2*). Phenotypes with enrichment patterns indicating increasing values from left to right were localized on the right extreme of the principal component space including periodontitis-associated bacterial genera, clinical oral health measurements, circulating inflammatory markers, smoking behavior and gray matter free-water measured by diffusion-weighted MRI. Phenotypes increasing right to left were located on the left including health-associated bacterial genera, cognitive performance measurements, diet scores, cortical thickness and subcortical volume as well as tissue fractional anisotropy as measured by MRI. Phenotypes showing a non-linear enrichment pattern or transition pattern in the left-right right-left orientation were localized in the middle including mental health scores and vascular risk measurements apart from smoking. For a heatmap depicting the Spearman correlation of enrichment scores for all non-microbiome phenotypes see *figure 5b*. For the same co-enrichment heatmap indicating genus-genus correlation see *supplementary figure S3*.

Microbiome similarity groups differ in periodontal disease, inflammatory markers, cognition and brain structure

Applying k-means clustering on the microbiome similarity network topology resulted in an unsupervised separation of the sample into two participant groups ("A" and "B") along the pathogenicity gradient. The following statistics to characterize the identified participant groups were adjusted for potential confounders including age, sex, education and vascular risk factors. The procedure is illustrated in *figure 6a*. Participants assigned to nodes in both groups were excluded from the analysis (n=63).

We compared genus-level abundance to identify bacterial genera that showed significant differences between groups (*figure 6b* and *6c*). Of the 85 tested genera, 6 showed no significant group differences (*supplementary figure S4*). Top 15 bacterial genera which were significantly higher in the "A" group were *Fusobacterium* (β_{std} = -1.23)., *Treponema* (β_{std} = -1.11), *Fretibacterium* (β_{std} = 1.07), *Dialister* (β_{std} = -0.97), *Porphyromonas* (β_{std} = -0.95), *Tannerella* (β_{std} = -0.90), *Mogibacterium* (β_{std} = -0.89), *Parvimonas* (β_{std} = -0.87), *Peptostreptococcaceae* [G-9] (β_{std} = -0.87), *Peptostreptococcaceae* [G-4] (β_{std} = -0.83), *Prevotella* (β_{std} = -0.82), *Veillonellaceae* [G-1] (β_{std} = -0.81), *Campylobacter* (β_{std} = -0.80), *Peptostreptoccocaceae* [G-2] (β_{std} = -0.77), and *Desulfobulbus* (β_{std} = -0.76) (all $p_{FDR} < 0.001$). Top 15 bacterial genera which were significantly higher in the "B" group were *Streptococcus* (β_{std} = 1.16), *Haemophilus* (β_{std} = 1.07), *Rothia* (β_{std} = 0.90), *Granulicatella* (β_{std} = 0.77), *Gemella* (β_{std} = 0.71), *Riemerella* (β_{std} = 0.66), *Veillonella* (β_{std} = 0.63), *Lautropia* (β_{std} = 0.58), *Actinomyces* (β_{std} = 0.57), *Neisseria* (β_{std} = 0.56), *Enterococcus* (β_{std} = 0.53), *Kingella* (β_{std} = 0.51), *Abiotrophia* (β_{std} = 0.47), *Paracoccus* (β_{std} = 0.38), *Roseomonas* (β_{std} = 0.38) (all $p_{FDR} < 0.001$).

In addition, the "A" group showed significantly higher clinical severity of periodontitis, higher circulating inflammatory markers, lower cognitive performance and lower

brain structural integrity compared to the “B” group (*figure 5b*): a significantly higher clinical attachment loss (β_{std} = -0.50, p_{FDR} < 0.001), DMFT index (β_{std} = -0.17, p_{FDR} = 0.029), plaque index (β_{std} = -0.32, p_{FDR} < 0.001), bleeding on probing index (β_{std} = -0.49, p_{FDR} < 0.001), leukocytes (β_{std} = -0.25, p_{FDR} < 0.001) as well as a significantly lower general cognitive ability (β_{std} = 0.22, p_{FDR} = 0.001), executive function (β_{std} = 0.15, p_{FDR} = 0.027), verbal fluency (β_{std} = 0.18, p_{FDR} = 0.013), Mini Mental State Exam score (β_{std} = 0.19, p_{FDR} = 0.011) as well as cortical thickness and subcortical volume (β_{std} = 0.15, p_{FDR} = 0.029). The remaining cognitive scores, brain structural indices and hsCRP showed no significant differences (*supplementary figure S5*). Moreover, mental health scores and diet scores also showed no significant differences.

Sensitivity analysis

To assess the robustness of our result we performed two-fold sensitivity analyses.

First, we conducted repeated enrichment analyses on randomly selected subsamples, reducing the sample size from 100% to 10% in 1% decrements, with 100 random samples at each step (*supplementary figure S6*). Throughout this process, the Spearman correlation for enrichment ratios (covering both non-microbiome and microbiome phenotypes) remained high (Spearman ρ > 0.8), demonstrating high robustness to different sample compositions until the sample size was reduced to approximately 50% (n = 513). Additionally, the Adjusted Rand Index (ARI) remained high (ARI > 0.8), indicating strong consistency in microbiome similarity network-based group assignments from k-means clustering, until the sample size was reduced to approximately 50%.

Second, to evaluate the potential influence of pipeline design choices on our findings, we reanalyzed the data using 15 different pipeline configurations, varying the components and parameters of the topological data analysis (*supplementary table S7*). The results demonstrated high robustness, with Spearman correlations between enrichment ratios from the original and alternative configurations averaging 0.88 ± 0.06. Furthermore, the ARI for group assignments derived from k-means clustering via the original pipeline, compared to those from different configurations, remained consistently high at 0.85 ± 0.14.

Discussion

The oral cavity hosts the second most diverse microbiome in the human body, following the gut.[12] Interactions within these oral microbial communities, along with host factors, impacting both oral and systemic health.[13] In this study, we present a comprehensive population-level analysis examining the association between the oral microbiome and multi-domain brain health phenotypes ranging from cognitive functions, brain structure, mental health, inflammatory blood biomarkers, to dietary behavior and vascular risk factors. Using 16S rRNA sequencing and topological data analysis, we inferred a microbiome similarity network that revealed a continuous pathogenicity gradient, mapping individuals based on periodontal microbiome profiles. Leveraging this network, we identified associations of periodontitis-related microbial communities with multiple brain health phenotypes: Individuals with higher abundance of periodontitis-related taxa exhibited significantly lower performance in multiple cognitive domains, lower cortical thickness and subcortical volume as well as increased leukocyte counts adjusting for demographics and cardiovascular risk. These findings exhibited strong robustness, as confirmed through sensitivity analyses involving diverse random subsamples and varying analytical configurations. Our findings highlight several key insights into the oral microbiome-brain axis, potential mechanistic pathways and clinical implications.

Topological data analysis reveals a latent axis of periodontitis-related microbial composition

We performed a dominance analysis to determine which bacterial genera are particularly abundant within specific participant clusters of the microbiome similarity network. Based on this analysis, we show that the microbiome similarity network

captured interindividual variations and a gradient of microbial compositions from periodontitis-associated taxa to health-associated taxa (*Figure 3b*): Periodontitis-associated genera such as *Porphyromonas*, *Fusobacterium*, *Treponema, Saccharibacteria (TM7)* and *Parvimonas* were enriched at the left part of the network, consistent with their established roles in biofilm formation, immune modulation, and periodontal tissue destruction.[4,14–16] Also positioned at this pathogenic end were bacteria of the genus *Dialister*, which have only recently been associated with periodontitis and could be a key periodontal pathogen, warranting further research into its mechanisms.[17] In contrast, genera such as *Enterococcus*, *Lactococcus, Streptococcus*, *Haemophilus*, and *Rothia* were enriched at the right part, reflecting their protective roles maintaining periodontal health or their association with oral conditions different from periodontitis, such as dental caries.[18,19] Enrichment ratios for bacterial genera were overall high, affirming that the observed interindividual differences were relevantly captured by the microbiome similarity network.

These findings indicate that periodontitis emerged as the most parsimonious explanation for the topology of the microbiome similarity network, with participants harboring higher abundances of periodontitis-related taxa clustering at one end, and those with higher abundances of other genera clustering at the other. Put differently, the identified network revealed a latent axis and continuum of periodontitis (referred to as pathogenicity gradient in the following) mirroring the pathogenicity spectrum of periodontitis and suggesting that the disease constitutes a key driver of the observed interindividual variance of microbiome compositions. This evidence demonstrates that a topological data analysis-based approach can uncover subtle yet biologically meaningful patterns of disease severity in large-scale, highly complex datasets.

Microbial compositions are linked to non-microbiome phenotypes

The microbiome similarity network offers a means to integrate complex, high-dimensional microbiome data with comprehensive health information by mapping non-microbial phenotypes on the network and statistically testing their association with the network structure. We observed that the investigated non-microbiome phenotypes differentially enriched on the network, implying varying degrees of association with microbial configuration. These enrichment patterns ranged from linear enrichments correlated or anti-correlated with the pathogenicity gradient (e.g., plaque index) – where variance in these variables coincided with changes in the abundance of periodontitis-related taxa – to non-linear patterns indicating weaker or more complex associations with the pathogenicity gradient (e.g., PHQ-9) (*Figure 4b*). Co-enrichment analyses further detailed these observations, showing that microbial and non-microbial phenotypes with visually overlapping enrichment patterns occupied similar positions in principal component space and exhibited strong correlations in their enrichment scores (*Figure 5* and *supplementary figures S2-S3*). Among the phenotypes most closely aligned with the pathogenicity gradient were the plaque index, clinical attachment loss, the DMFT index and bleeding on probing index, reinforcing periodontitis as the primary pathology captured by the network.[20,21] These findings highlight not only the complex covariance structure of oral microbiome composition but also the potential of genus-level abundance data as a promising biomarker source, reflecting a broad range of clinical and lifestyle factors.

Periodontal dysbiosis and brain health

The mechanisms behind the relationship between periodontitis and brain health are multifaceted and remain to be fully elucidated. Our analysis sought to address these

complexities by integrating data on multiple brain health phenotypes into a single analysis framework.

Notably, multiple markers of cognitive performance, brain structural integrity (gray matter free-water, cortical thickness, subcortical volume), and systemic inflammation (leukocyte counts) showed the strongest enrichment ratios among non-oral phenotypes (*Figure 4a*) and varied in tandem with the pathogenicity gradient from one network end to the other (*Figure 4b*). Specifically, participants with higher abundance of periodontitis-associated taxa and poorer oral health exhibited significantly lower cognitive performance, reduced cortical and subcortical volume, and elevated systemic inflammation after adjusting for covariates (*Figure 6c* and *6d*). Importantly, these findings build on prior reports linking shifts in oral microbiota to cognitive changes and brain structure among dementia patients and demonstrate that such associations also manifest in a healthy population of non-demented, middle-aged individuals.[7,22,23] Our findings align with earlier analyses indicating relationships between clinical markers of periodontitis, cognitive health, and brain structural integrity.[24,25]

Our analysis highlights associations between brain health and the abundance of various bacterial genera indicating pathomechanistic links. For example, many of the highlighted genera have already been implicated in Alzheimer's disease in previous studies including *Porphyromonas*, *Treponema*, *Fusobacterium* and *Prevotella*. Prominently, experimental evidence from animal models demonstrated that *Porphyromonas gingivalis* contributes to Alzheimer's pathology by contributing to the formation of amyloid-β, neurofibrillary tangles and neuroinflammation.[6,26] Furthermore, postmortem studies report a higher abundance of *Porphyromonas gingivalis* in the brain samples of AD patients compared to non-demented controls,

with intracerebral presence of the bacterium being related to a six-fold increased risk of AD.[26,27] *Treponema denticola* has been demonstrated to induce tau hyperphosphorylation and neuroinflammatory processes in mice.[28] The serum levels of antibodies against *Prevotella intermedia* and *Fusobacterium nucleatum* are significantly increased in Alzheimer's disease patients.[8] Additionally, elevated *Prevotella intermedia* abundance is associated with *APOE4*-carrier status, and *Fusobacterium* species have been linked to cerebrovascular lesions indicative of small vessel pathology.[10,29] Our findings align with these results, indicating that the prevalence of these pathogenic bacteria in the subgingival biofilm corresponds to reduced brain health in cognitively normal individuals. At the same time, we found genera not yet reported in relation to brain health – *Fretibacterium*, *Tannerella*, *Dialister*, *Mogibacterium*, and *Parvimonas* – warranting further experimental research to determine whether they independently drive cognitive deficits or merely coincide with other pathogenic species.

Our findings emphasize the role of periodontitis-related taxa in systemic inflammation and possible connections to neurodegenerative pathways. We observed that individuals positioned at the pathogenic end of the oral microbiome similarity network exhibited significantly higher leukocyte levels after covariate adjustment (*Figure 6d*), indicating presence of systemic inflammation. This aligns with prior studies linking clinically diagnosed periodontitis to elevated systemic inflammatory markers.[30] Pathomechanistically, such responses can stem from dense leukocytic infiltration during gingival inflammation, the stimulation of bone marrow for sustained inflammatory cell production, or the systemic response to dissemination of pathogens from ulcerated gingival tissues.[30] Importantly, we present novel direct evidence linking periodontitis-associated microbial compositions to systemic

inflammation. Notably, this systemic inflammation was accompanied by reduced cognitive performance and diminished brain structural integrity, supporting the hypothesis that chronic oral infections contribute to both systemic inflammation and neurodegenerative processes. Although our findings cannot prove causality due to the cross-sectional nature of the study design, they align with the inflammatory hypothesis of Alzheimer's disease, suggesting that immune responses triggered by oral pathogens may disrupt the blood-brain barrier and activate microglia, thereby driving neurodegeneration.[31]

In contrast to cognitive and brain structural measures, mental health scores displayed weaker associations with the pathogenicity gradient, suggesting more nuanced or less direct links with periodontitis-related microbiome composition. These associations did not persist after adjusting for demographic and cardiovascular risk factors. We interpret these results as pointing to a more specific impact of periodontitis on cognitive rather than mental health. Nevertheless, prior studies have reported connections between anxiety, depression, and clinical periodontitis or oral microbiome composition.[32] Longitudinal research will be essential to elucidate how these factors collectively influence oral and brain health.

Influence of diet and vascular risk factors on oral and brain health

We investigated the role of nutritional behavior as a key factor influencing oral microbiome composition.[33] Enrichment analysis revealed that adherence to cognitively beneficial diets, such as the MEDAS, MIND, and DASH diets,[34–36] aligned with the pathogenicity gradient: participants with greater abundance of periodontitis-related taxa tended to adhere less to these diets. Prior findings from the HCHS showed a significant association between higher adherence to DASH and

Mediterranean diets and lower odds of periodontal disease.[37] Although speculative, these results suggest that such diets may exert favorable effects on cognition through modulation of the oral microbiome. However, after adjusting for covariates, including vascular risk factors, no significant differences in diet scores were observed among individuals with higher abundance of periodontitis-related taxa. This raises the alternative hypothesis that the cognitive benefits of these diets may be mediated primarily through cardiovascular pathways rather than direct effects on the oral microbiome. Interventional studies are needed to clarify the causal relationships between specific dietary patterns, the oral microbiome, and cognitive health.

Among demographic and vascular risk factors, smoking behavior was the only factor clearly associated with the pathogenicity gradient. Smoking has been documented to promote the colonization of periodontal pathogens, facilitate biofilm formation, and compromise the immune response, thereby exacerbating periodontal disease and systemic inflammation.[38] Our findings highlight the role of smoking on promoting dysbiosis and inflammation and positions it as a direct contributor to the pathogenicity gradient and brain health phenotypes.

Clinical implications

Our findings not only shed light on the intricate systemic pathophysiology of periodontal dysbiosis and brain health, but they also hint at potential avenues of clinical utilization. Our study reveals an association between oral microbiome signatures and variance in cognitive performance and brain structure that could enhance early screening measures for risk of cognitive decline and improve targeted recruitment of individuals at critical early stages of cognitive impairment and dementia. Particularly, identifying microbial community changes before emergence

of cognitive symptoms could enhance diagnostics by providing microbiome-based markers as a convenient complement to existing dementia biomarkers. Regarding future therapeutic interventions, microbiome signatures of early periodontal dysbiosis may guide the development of oral microbiome-directed therapies to slow or prevent dementia progression.[9] However, the definitive role of microbiome biomarkers in cognitive disorders is yet to be determined and large-scale longitudinal as well as interventional studies are required for moving in this direction.

Strengths and limitations

Strengths of this work lie in its considerable sample size; state-of-the-art oral microbiome profiling; in-depth phenotyping of clinical and lifestyle data; as well as a novel, robust data analysis pipeline that unifies multiple complex data domains into a single analysis framework. However, our study also exhibits limitations. Due to the cross-sectional nature of the performed microbiome-phenotype association analyses, our findings cannot firmly establish causal links and should be regarded as hypothesis-generating. Moreover, even after adjusting for confounders in our statistical models, we cannot fully exclude their potential influence on the relationship between oral microbiome composition and non-microbiome phenotypes. Longitudinal and experimental studies are needed that expand on our findings to further discern effects of the oral microbiome on brain health. Lastly, high-resolution sequencing methods, such as shotgun sequencing, could reveal strain-specific microbial variations (e.g., virulence factors) linked to periodontitis and cognitive decline, while also identifying potential non-bacterial microbiome members involved in these processes.

Conclusion

Drawing on a comprehensive investigation integrating multi-domain data with cutting-edge analysis techniques we characterized latent associations of periodontitis-related oral microbial composition with cognitive function, mental health, brain structural integrity, systemic inflammation, diet and vascular risk in predominantly healthy individuals. Notably, periodontitis-related oral dysbiosis was associated with lower cognitive performance, lower cortical and subcortical volume, and higher leukocytes. As this research field progresses, oral microbiome profiling could contribute to improved dementia risk stratification and guide preventive interventions.

Materials and methods

Study design

Our methodology is illustrated in *figure 1*. This analysis integrates different data domains related to oral and brain health into a unified analysis framework: oral microbiome composition, oral health status, cognitive function, brain structure, mental health status, circulating inflammatory markers, dietary patterns, vascular risk and demographics. In brief, we employed topological data analysis, specifically the Mapper algorithm, to create a low-dimensional network representation of oral microbiome abundance data. This network representation positions individuals based on similarities in their oral microbiome composition. Subsequently, we conducted an enrichment analysis using Spatial Analysis of Functional Enrichment (SAFE). This analysis statistically assesses regions of the network derived via Mapper to identify where specific phenotypes are significantly higher or lower than expected by chance. This approach allowed us to examine regional differences in phenotypes, revealing whether the network representation captures variance in specific participant traits related to oral and brain health. Finally, we performed a post-hoc group comparison between participants of two distinct participant clusters within the microbiome network, enabling us to assess the link of oral microbiome composition and clinical and lifestyle phenotypes while adjusting for relevant covariates.

Study population

PAROMIND is a cross-sectional study nested within the Hamburg City Health Study (HCHS). The HCHS is a prospective, single-center, population-based cohort study investigating adults aged 45-75 to enhance the detection of major chronic disease

risks through extensive clinical and imaging phenotyping.[11] The study protocol includes assessments of oral health (including collection of gingival crevicular fluids for microbiome analyses), cognition, mental health, diet, vascular risk, brain MRI, and blood sampling.

Ethics statement

PAROMIND and the HCHS were approved by the local ethics committee of the Landesärztekammer Hamburg (State of Hamburg Chamber of Medical Practitioners, PV5131). The conduct of PAROMIND is governed by ethical guidelines of Good Clinical Practice (GCP), Good Epidemiological Practice (GEP) and the Declaration of Helsinki.[39] Written informed consent was obtained from all participants investigated in this work.

Molecular analysis and phenotyping of the oral microbiome

Oral microbiome phenotyping followed a standardized procedure implementing latest developments. During the dental examination, gingival crevicular fluid samples were collected with sterile paper points. Two samples per participant were obtained: (1) from a single deep periodontal pocket and (2) one pooled sample (4 paper points) from the deepest periodontal pockets per quadrant. Each paper point remained in situ for 15 seconds. Subsequently, samples were placed in a sterile 2ml-Eppendorf tube and stored at -80°C in the HCHS biobank. For the following molecular analysis, only pooled samples were further processed.

Bacterial composition in the oral microbiome was determined via 16S rRNA Illumina sequencing. DNA Isolation was performed using the DNA extraction kit from Innuprep Analytics (Analytik Jena AG, Überlingen, Germany). For the initial lysis with lysozyme and mutanolysin (3 mg lysozyme, 100 U mutanolysin, in 200 µl Tris EDTA

buffer), all samples were incubated at 37°C for 10 min and further isolation was performed according to the manual with elution volumes of 100 µl. For 16S rDNA sequencing of isolated DNA, variable regions V3 and V4 of the 16S rRNA gene were amplified using the primer pair 341F (5‘-CCTACGGGAGGCAGCAG-3‘) and 806R (5‘-GGACTACHVGGGTWTCTAAT-3’) 27F-338R in a dual barcoding approach.[40,41] Agarose gel electrophoresis was used to verify resulting PCR products before normalization using the SequalPrep Normalization Plate Kit (Thermo Fischer Scientific, Waltham, MA, USA), pooling and sequencing on the Illumina MiSeq with v3 2x300bp chemistry (Illumina Inc., San Diego, CA, USA). Demultiplexing after sequencing was based on 0 mismatches in the barcode sequences.

Data processing was performed using the DADA2 workflow for big datasets (v. 1.10.42, https://benjjneb.github.io/dada2/bigdata.html), resulting in abundance tables of amplicon sequence variants (ASVs). For this, all sequencing runs were handled separately and finally collected in a single abundance table per dataset, which underwent chimera filtering. ASVs underwent taxonomic annotation using the Bayesian classifier provided in DADA2 and using the expanded Human Oral Microbiome Project (eHOMD) version V15.23. Samples with less than 10,000 sequences were not considered for further analysis. Genera that were observed with a frequency of less than 0.1% of all genera detected in a sample were discarded.

Oral health assessment

A certified study nurse assessed probing depth and gingival recession at six sites of the tooth (mesio-buccal, buccal, disto-buccal, disto-palatinal, palatinal and mesio-palatinal). The clinical attachment loss (CAL) was calculated (CAL = probing depth + gingival recession).[42] The bleeding on probing (BOP) index was determined by probing 2 sites per tooth (vestibular, oral) and expressed as a percentage of

bleeding sites. The number of bleeding sites on probing is a reliable and consistent measure to assess the degree of gingival inflammation. Measurements were completed with a standard periodontal probe (PCP 15, Hu-Friedy, Chicago, IL, USA). Additionally, the DMFT index was calculated. The oral health assessment followed a standardized protocol for the reporting of the prevalence and severity of periodontal diseases.[43]

Cognitive and mental health assessments

Cognitive testing was conducted using the extended version of the Consortium to Establish a Registry for Alzheimer's Disease Neuropsychological Assessment Battery (CERAD-NP/Plus).[44] A trained study nurse administered all tests. For this analysis, we considered cognitive scores measuring executive function (Trail Making Test B), information processing speed (Trail Making Test A), memory (Word List Recall Test), reasoning (Multiple Choice Vocabulary Intelligence Test B), verbal fluency (Animal Naming Test), and the Mini Mental State Exam. To ensure higher scores indicated better cognitive performance across all tests, we inverted the results of Trail Making Test A and B. Subsequently, we performed a principal component analysis (PCA) on all individual test scores. Following previous procedures, the first principal component, which accounted for the greatest variance (40.5%), was defined as a measure of general cognitive ability (g) (for details see *supplementary figure S8*).[45] According to the principal component loadings, higher values of this measure corresponded to lower cognitive performance; thus, it was also inverted. Furthermore, participants underwent mental health assessments via established questionnaires for depression (PHQ-9, Geriatric Depression Scale), somatic symptom severity (PHQ-15), and anxiety (GAD-7).[46–48]

Neuroimaging of brain macro- and microstructure

Neuroimaging markers representing different aspects of macro- and microstructural brain integrity were computed following previous procedures.[49] After cortical surface reconstruction and subcortical segmentation based on T1-weighted images with FreeSurfer (v. 6.0.1), cortical thickness and subcortical volume were estimated representing morphometric measures of neurodegenerative processes.[50,51] The mean cortical thickness and mean subcortical volume were z-scored. Subsequently, the two resulting z scores were averaged to obtain a single summary measure reflecting cortical thickness and subcortical volume. Following preprocessing of the diffusion-weighted images, free-water imaging was employed to compute free-water which quantifies the amount of extracellular water as well as tissue fractional anisotropy reflecting neurite architecture and integrity.[52,53] The free-water and tissue fractional anisotropy values were then averaged across cortical and subcortical gray matter voxels, as well as white matter voxels, to obtain global measures of gray and white matter, respectively. For a detailed account on the acquisition protocol, quality assessment, preprocessing and computation procedures on the different imaging measures see *supplementary text S9*.

Diet

The dietary behavior of all participants was assessed using a validated food frequency questionnaire with 102 items, developed for the European Prospective Investigation into Cancer and Nutrition Study (EPIC).[54] The adherence to different dietary patterns was measured based on the food frequency questionnaire scores, including the Mediterranean diet (MEDAS diet), the Dietary Approaches to Stop Hypertension (DASH diet), and the Mediterranean-DASH Intervention for

Neurodegenerative Delay (MIND diet).[34–36] Adherence to the Mediterranean diet was determined using the German version of the Mediterranean Diet Adherence Screener, which assigns a score of 0 or 1 to 14 specific food items, producing a total adherence score ranging from 0 (no adherence) to 14 (maximum adherence).[34] The DASH diet was evaluated using a previously established scoring method that assigns a score of 0, 0.5, or 1 to each of 10 items, resulting in an adherence score between 0 (no adherence) and 10 (maximum adherence).[36] Finally, adherence to the MIND diet was calculated according to standard procedures: scores of 0, 0.5, or 1 to are assigned 10 healthy and 5 unhealthy food items, culminating in a total adherence score ranging from 0 (no adherence) to 15 (maximum adherence).[35]

Topological data analysis

We implemented a novel approach based on topological data analysis that integrates two key components: (1) the Mapper algorithm, which performs an unsupervised reconstruction of a topological network based on genus-level abundance data, capturing microbiome composition similarity, and (2) SAFE, which conducts statistical tests to examine the relationship between the network's structure and different phenotypes.[55,56] This method effectively combines dimensionality reduction with topological insights, offering a powerful tool for understanding the intrinsic geometry of high-dimensional microbiome data and its relationships with other data domains. The analysis was performed in *python* v3.8.1 based on the packages *NetworkX* v2.2 (https://github.com/networkx/networkx), *safepy* (https://github.com/baryshnikova-lab/safepy), *scikit-learn* v1.5.1 (https://github.com/scikit-learn/scikit-learn) and *tmap* v1.2 (https://github.com/GPZ-Bioinfo/tmap) as well as *R* v4.4.0 based on the package *vegan v2.6-6.1*.[55,56] Data visualization was based on *plotly* v5.22 (https://github.com/plotly/plotly.py) and iTOL

v6 (https://itol.embl.de/). HTML versions of the presented plots can be found on OSF allowing interactive data exploration (https://osf.io/9ykv5/).

Mapper: Reconstruction of the microbiome similarity network

Genus-level oral microbiome abundance data served as input to the Mapper algorithm, a topological data analysis technique that simplifies complex high-dimensional data by constructing a topological network capturing essential relationships and patterns in the data. Importantly, the resulting network can represent non-linear associations that conventional linear techniques might miss. The applied Mapper pipeline comprises multiple analysis steps to reconstruct the topological network (for an illustration see *figure 1b*): filtering, covering, clustering and network reconstruction.[57] First, as the filtering step, we conducted a principal coordinate analysis of the Bray-Curtis dissimilarity of genus-level abundance ($n_{participants}$ x n_{genera}). By that, we obtained two components – also called lenses – capturing the major axes of variation in the microbial community composition across the participants. Based on these axes, overlapping covers were defined (overlap = 75%, resolution = 30) to segment the data into overlapping bins, each representing a local region of inter-individual variation. Unsupervised clustering of datapoints within each bin was performed using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN, epsilon threshold = 0.95).[58] Lastly, the network reconstruction was accomplished by connecting clusters sharing common participants. Of note, not all datapoints are retained by Mapper resulting in the omission of some participants ($n_{not\ retained}$ = 106).

SAFE: Enrichment analysis

SAFE is an annotation technique for biological networks that enables the computation and statistical assessment of local enrichment for specific phenotypes.[56] In our work, we performed SAFE on the microbiome similarity network derived using the Mapper algorithm to identify regions within the network that are significantly enriched for specific participant traits. Specifically, we investigated the enrichment of demographics (age, sex, education), genus-level microbiome abundance, oral health measures (clinical attachment loss, plaque index, bleeding on probing index, DMFT index), cognitive scores (general cognitive ability, Animal Naming Test, Mini Mental State Exam, Multiple Choice Vocabulary Intelligence Test B, Trail Making Tests A and B, Word List Recall), mental health scores (PHQ-9, PHQ-15, Geriatric Depression Scale, GAD-7), imaging measures (cortical thickness and subcortical volume, gray matter free-water, white matter free-water, gray matter tissue fractional anisotropy, white matter tissue fractional anisotropy), circulating inflammatory markers (hsCRP, leukocytes), diet scores (MEDAS score, DASH score, MIND score), and vascular risk factors (systolic and diastolic blood pressure, body mass index, smoking behavior, blood triglycerides, cholesterol, low density lipoprotein, high density lipoprotein, HbA1c).

Following previous analyses leveraging SAFE,[56,59] the microbiome similarity network was spring-embedded, i.e., nodes in the network were positioned so that those with interactions are placed closer together, while repulsive forces push non-connected nodes apart, resulting in a visually balanced and interpretable representation of the network's structure.[60] Next, we computed node attributes for all phenotypes by averaging the values of each variable across all participants within a node. Subsequently, enrichment scores were calculated for each node following a four-step process. First, we defined the local neighborhood by identifying all nodes within

a maximum distance threshold of 0.75 from the central node. The distance was measured using the map-weighted shortest path length (MSPL).[56] Second, we calculated a neighborhood score by summing the attribute values of neighboring nodes. Third, we computed a p-value by comparing the empirical neighborhood score against a distribution derived from 5000 permutations. Permutations were performed by randomly reassigning attributes to nodes while preserving the network topology.[56] The resulting p-value was corrected for multiple comparisons across all phenotypes. Finally, we assigned an enrichment score to the neighborhood center by applying a -log10 transformation to the corrected p-value. Given the 5000 permutations the maximal enrichment score is -log10(1/5000) = 3.70 with -log10(0.05) = 1.30 indicating significance. Positive enrichment scores indicate that observed values are higher than the permuted distribution, negative enrichment scores indicate that they are lower than the permuted distribution. This procedure is repeated for each node, resulting in an enrichment map indicating where attributes are higher or lower than expected by chance.

To measure how strongly the microbiome similarity network reflects a specific phenotype we computed the enrichment ratio as the number of significantly enriched nodes (corrected $p < 0.05$) divided by the total number of nodes. A higher enrichment ratio indicates that more nodes are significantly enriched, suggesting that the network's topology captures a greater extent of a phenotype's variance. As a complementary approach, we used *envfit* ($n_{permutations}$ = 10,000, false discovery rate correction of p-values) from the *vegan* package to test the linear association between non-microbiome phenotypes and microbiome configuration as used for covariate identification in previous studies.[61] This approach assesses phenotype associations with genus-level community ordination achieved via principal coordinate

analysis based on Bray-Curtis dissimilarity of genus-level microbiome abundance. The *envfit* analysis was only performed for non-microbiome phenotypes and not for abundance information of individual genera.

To determine whether specific phenotypes co-enrich, i.e., exhibit similar enrichment patterns, we performed an ordination of the enrichment scores using principal component analysis retaining the first two principal components and assessed the pairwise Spearman correlation between the enrichment scores.

Group analysis

To further examine the relationship between the microbiome similarity network and clinical as well as lifestyle phenotypes, we performed a post-hoc group analysis which allowed us to adjust for relevant covariates. Therefore, nodes of the microbiome similarity network were clustered in two non-overlapping groups using k-Means clustering of the node positions. We chose k-Means as it is widely used and arguably represents the simplest unsupervised clustering technique.[62] k-Means requires to predefine the number of clusters to assign datapoints to. Given that enrichment analysis indicated that most participant traits varied along the microbiome similarity network in a linear left-right trajectory, i.e., participants on the left end differed from those on the right end, two was predefined as the number clusters post-hoc.

After the clustering, participants were categorized based on the resulting groupings, with individuals present in both groups – due to being assigned to nodes in both groups – excluded from the analysis. Given that not all datapoints are retained by the Mapper algorithm during the network reconstruction step, participants that were not represented in the microbiome similarity network were not considered for this

analysis. The groups were statistically compared for the phenotypes using multiple linear regression and age, sex and education and vascular risk factors (systolic and diastolic blood pressure, body mass index, smoking behavior, triglycerides, cholesterol, LDL, HDL, HbA1c) were included as covariates:

$$Phenotype \sim Group + Age + Sex + Education + Vascular\ risk\ factors$$

Sensitivity analyses

To verify the robustness of our results against variations in the examined sample as well as the analysis pipeline, we conducted a comprehensive sensitivity analysis.

This involved repeating the entire analysis across random subsamples of different sizes. Specifically, we varied the sample sizes from 100% to 10% of the total dataset, in 1% decrements. For each decrement, we randomly sampled subsets 100 times, resulting in a total of 9,000 iterations (90 different sample sizes × 100 random samples per steps). For each iteration, we assessed the robustness of our findings by comparing the results from the subsamples to the original results. The stability of the enrichment analysis results was evaluated by calculating the Spearman correlation of enrichment ratios for the non-microbiome phenotypes and genus-level abundance. In addition to this, we measured the agreement of the participant-group assignments resulting from k-Means clustering used for group analysis employing the Adjusted Rand Index (ARI) which ranges from 0 (no agreement) to 1 (full agreement).

Our analysis pipeline involved several design choices, which are based on best practices or are the default settings of the software tools used. To ensure that our results were not biased by these design choices, we performed a sensitivity analysis

with different pipeline configurations. We systematically explored variations in parameter values and pipeline components, altering the Mapper cover overlap from the original 0.75 to 0.5 and 0.99; the Mapper cover resolution from 30 to 10, 20, 40, and 50; the Mapper epsilon threshold from 0.95 to 0.99, 0.90, and 0.85; and the Mapper filter function from principal coordinate analysis of the Bray-Curtis dissimilarity to multidimensional scaling and Uniform Manifold Approximation and Projection (UMAP) on the raw data. Additionally, we adjusted the SAFE distance threshold from 0.75 to 0.5 and 0.99, and the SAFE neighborhood radius from 0.1 to 0.05 and 0.15. We adjusted one parameter from the original pipeline per iteration, resulting in a total of 15 iterations. The robustness of our results was evaluated by comparing the findings from the original configuration to those from alternative setups. Matching the approach for the previous sensitivity analysis, we calculated the Spearman correlation of enrichment ratios for clinical and genus abundance phenotypes and the ARI for group assignments from k-Means clustering to assess stability.

Data availability

HCHS data can be obtained by qualified researchers on reasonable request to the study's steering committee. The analysis code for this work is publicly available on GitHub (https://github.com/csi-hamburg/oral_microbiome_brain_health). Interactive versions of the plots can be found on OSF (https://osf.io/9ykv5).

Acknowledgments

The authors would like to acknowledge all participants, all cooperation partners, patrons and the Deanery from the University Medical Center Hamburg-Eppendorf for supporting the HCHS. Special thanks are due to the staff at the Population Health

Research Department for conducting the study. The publication of its results has been approved by the Steering Board of the HCHS. We also thank the staff of the microbiome and sequencing laboratories of the Institute of Clinical Molecular Biology Kiel for their excellent support.

Competing interests

JG has received speaker fees from Lundbeck, Janssen-Cilag, Lilly, Otsuka and Boehringer outside the submitted work. JF reported receiving personal fees from Acandis, Cerenovus, Microvention, Medtronic, Phenox, and Penumbra; receiving grants from Stryker and Route 92; being managing director of eppdata; and owning shares in Tegus and Vastrax; all outside the submitted work. RT is a co-inventor of an international patent on the use of a computing device to estimate the probability of myocardial infarction (PCT/EP2021/073193, International Publication Number WO2022043229A1). RT is shareholder of the company ART-EMIS GmbH Hamburg. GT has received fees as consultant or lecturer from Acandis, Alexion, Amarin, Bayer, Boehringer Ingelheim, BristolMyersSquibb/Pfizer, Daichi Sankyo, Portola, and Stryker outside the submitted work. The remaining authors declare no conflicts of interest.

Author Contributions

Each author has made a significant contribution to the manuscript and all authors read and approved its final version. We describe contributions to the paper using the CRediT contributor role taxonomy. M.P.: Conceptualization, Data curation, Formal

analysis, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing—original draft, Writing—review & editing; C.W.: Conceptualization, Data curation, Investigation, Methodology, Resources, Writing—original draft, Writing—review & editing; K.B.: Data curation, Methodology, Investigation, Resources, Software, Writing—review & editing; G.H.: Resources, Writing—review & editing; T.B.: Resources, Writing—review & editing; M.A.: Resources, Writing—review & editing; C.M.: Resources, Writing—review & editing; F.L.N.: Data curation, Resources, Writing—review & editing; B.Z.: Resources, Writing—review & editing; J.F.: Resources, Writing—review & editing; J.G.: Resources, Writing—review & editing; S.K.: Resources, Writing—review & editing; R.T.: Resources, Writing—review & editing; C.B.: Methodology, Writing—review & editing; G.T.: Supervision, Funding, Writing—review & editing; B.C.: Conceptualization, Funding, Project administration, Resources, Supervision, Writing—original draft, Writing—review & editing; G.A.: Conceptualization, Funding, Project administration, Resources, Supervision, Writing—original draft, Writing—review & editing.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): PAROMIND – 514762487 – GZ AA 93/9-1 and GZ CH 2631/4-1 (B.C., G.A.); Sonderforschungsbereich 936 – 178316478 – C2 (G.T., B.C.); and Schwerpunktprogramm 2041 – 454012190 (G.T.). Microbiome sequencing received infrastructure support from the DFG Research Unit 5042 „miTarget" and the DFG Excellence Cluster 2167 "Precision Medicine in Chronic Inflammation" (PMI).

References

1. Baker, J. L., Mark Welch, J. L., Kauffman, K. M., McLean, J. S. & He, X. The oral microbiome: diversity, biogeography and human health. *Nat Rev Microbiol* **22**, 89–104 (2024).
2. Dioguardi, M. *et al.* The Role of Periodontitis and Periodontal Bacteria in the Onset and Progression of Alzheimer's Disease: A Systematic Review. *Journal of Clinical Medicine* **9**, 495 (2020).
3. Demmer, R. T. & Papapanou, P. N. Epidemiologic patterns of chronic and aggressive periodontitis. *Periodontol 2000* **53**, 28–44 (2010).
4. Socransky, S. s., Haffajee, A. d., Cugini, M. a., Smith, C. & Kent Jr., R. L. Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology* **25**, 134–144 (1998).
5. Belstrøm, D. *et al.* Microbial profile comparisons of saliva, pooled and site-specific subgingival samples in periodontitis patients. *PLoS One* **12**, e0182992 (2017).
6. Hajishengallis, G. & Chavakis, T. Local and systemic mechanisms linking periodontal disease and inflammatory comorbidities. *Nat Rev Immunol* **21**, 426–440 (2021).
7. Rubinstein, T. *et al.* Periodontitis and brain magnetic resonance imaging markers of Alzheimer's disease and cognitive aging. *Alzheimer's & Dementia* **20**, 2191–2208 (2024).
8. Sparks Stein, P. *et al.* Serum antibodies to periodontal pathogens are a risk factor for Alzheimer's disease. *Alzheimers Dement* **8**, 196–203 (2012).
9. Schwahn, C. *et al.* Effect of periodontal treatment on preclinical Alzheimer's disease—Results of a trial emulation approach. *Alzheimer's & Dementia* alz.12378 (2021) doi:10.1002/alz.12378.
10. L'Heureux, J. E. *et al.* Oral microbiome and nitric oxide biomarkers in older people with mild cognitive impairment and APOE4 genotype. *PNAS Nexus* **4**, pgae543 (2025).
11. Jagodzinski, A., Koch-gromus, U., Adam, G., Anders, S. & Augustin, M. Rationale and Design of the Hamburg City Health Study. *European Journal of Epidemiology* (2019) doi:10.1007/s10654-019-00577-4.
12. Escapa, I. F. *et al.* New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems* **3**, e00187-18 (2018).

13. Curtis, M. A., Diaz, P. I. & Van Dyke, T. E. The role of the microbiota in periodontal disease. *Periodontol 2000* **83**, 14–25 (2020).
14. Sedghi, L. M., Bacino, M. & Kapila, Y. L. Periodontal Disease: The Good, The Bad, and The Unknown. *Front. Cell. Infect. Microbiol.* **11**, (2021).
15. Horiuchi, A., Kokubu, E., Warita, T. & Ishihara, K. Synergistic biofilm formation by *Parvimonas micra* and *Fusobacterium nucleatum*. *Anaerobe* **62**, 102100 (2020).
16. Hajishengallis, G., Darveau, R. P. & Curtis, M. A. The Keystone Pathogen Hypothesis. *Nat Rev Microbiol* **10**, 717–725 (2012).
17. Antezack, A., Etchecopar-Etchart, D., La Scola, B. & Monnet-Corti, V. New putative periodontopathogens and periodontal health-associated species: A systematic review and meta-analysis. *Journal of Periodontal Research* **58**, 893–906 (2023).
18. Costalonga, M. & Herzberg, M. C. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunology Letters* **162**, 22–38 (2014).
19. Li, X., Liu, Y., Yang, X., Li, C. & Song, Z. The Oral Microbiota: Community Composition, Influencing Factors, Pathogenesis, and Interventions. *Front. Microbiol.* **13**, (2022).
20. Hajishengallis, G. Periodontitis: from microbial immune subversion to systemic inflammation. *Nat Rev Immunol* **15**, 30–44 (2015).
21. Hajishengallis, G. Immunomicrobial pathogenesis of periodontitis: keystones, pathobionts, and host response. *Trends Immunol* **35**, 3–11 (2014).
22. Troci, A. *et al.* Disease- and stage-specific alterations of the oral and fecal microbiota in Alzheimer's disease. *PNAS Nexus* **3**, pgad427 (2024).
23. Fogelholm, N. *et al.* Subgingival microbiome at different levels of cognition. *Journal of Oral Microbiology* **15**, 2178765 (2023).
24. Mayer, C. *et al.* Association between periodontal disease and microstructural brain alterations in the Hamburg City Health Study. *Journal of Clinical Periodontology* **n/a**, (2023).
25. Said-Sadier, N. *et al.* Association between Periodontal Disease and Cognitive Impairment in Adults. *Int J Environ Res Public Health* **20**, 4707 (2023).
26. Dominy, S. S. *et al.* Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors. *Science Advances* **5**, eaau3333 (2019).

27. Sixin, L., Stuart G, D. & Rui, Z. Association Between Oral Bacteria and Alzheimer's Disease: A Systematic Review and Meta-Analysis. *Journal of Alzheimer's disease□: JAD* **91**, (2023).
28. Tang, Z. *et al.* Treponema denticola Induces Alzheimer-Like Tau Hyperphosphorylation by Activating Hippocampal Neuroinflammation in Mice. *J Dent Res* **101**, 992–1001 (2022).
29. Kato, Y. *et al.* Fusobacterium in oral bacterial flora relates with asymptomatic brain lesions. *Heliyon* **10**, (2024).
30. Botelho, J. *et al.* Periodontitis and circulating blood cell profiles: a systematic review and meta-analysis. *Experimental Hematology* **93**, 1–13 (2021).
31. Kamer, A. R. *et al.* Inflammation and Alzheimer's disease: Possible role of periodontal diseases. *Alzheimer's & Dementia* **4**, 242–250 (2008).
32. Malan-Müller, S. *et al.* Probing the oral-brain connection: oral microbiome patterns in a large community cohort with anxiety, depression, and trauma symptoms, and periodontal outcomes. *Transl Psychiatry* **14**, 1–16 (2024).
33. Santonocito, S. *et al.* A Cross-Talk between Diet and the Oral Microbiome: Balance of Nutrition on Inflammation and Immune System's Response during Periodontitis. *Nutrients* **14**, 2426 (2022).
34. Hebestreit, K. *et al.* Validation of the German version of the Mediterranean Diet Adherence Screener (MEDAS) questionnaire. *BMC Cancer* **17**, 341 (2017).
35. Morris, M. C. *et al.* MIND diet slows cognitive decline with aging. *Alzheimer's & Dementia* **11**, 1015–1022 (2015).
36. Folsom, A. R., Parker, E. D. & Harnack, L. J. Degree of Concordance With DASH Diet Guidelines and Incidence of Hypertension and Fatal Cardiovascular Disease*: *American Journal of Hypertension* **20**, 225–232 (2007).
37. Altun, E. *et al.* Association between Dietary Pattern and Periodontitis—A Cross-Sectional Study. *Nutrients* **13**, 4167 (2021).
38. Jiang, Y., Zhou, X., Cheng, L. & Li, M. The Impact of Smoking on Subgingival Microflora: From Periodontal Health to Disease. *Front. Microbiol.* **11**, (2020).
39. Petersen, M. *et al.* Network Localisation of White Matter Damage in Cerebral Small Vessel Disease. *Scientific Reports* **10**, 9210 (2020).

40. Muyzer, G., de Waal, E. C. & Uitterlinden, A. G. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**, 695–700 (1993).
41. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621–1624 (2012).
42. Eke, P. I., Page, R. C., Wei, L., Thornton-Evans, G. & Genco, R. J. Update of the case definitions for population-based surveillance of periodontitis. *J Periodontol* **83**, 1449–1454 (2012).
43. Holtfreter, B. *et al.* Standards for reporting chronic periodontitis prevalence and severity in epidemiologic studies: Proposed standards from the Joint EU/USA Periodontal Epidemiology Working Group. *J Clin Periodontol* **42**, 407–412 (2015).
44. Moms, J. C. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assesment of Alzheimer's disease. *Neurology* **39**, 1159–1159 (1989).
45. Fawns-Ritchie, C. & Deary, I. J. Reliability and validity of the UK Biobank cognitive tests. *PLoS One* **15**, e0231627 (2020).
46. Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med* **16**, 606–613 (2001).
47. Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ-15: Validity of a New Measure for Evaluating the Severity of Somatic Symptoms: *Psychosomatic Medicine* **64**, 258–266 (2002).
48. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Arch Intern Med* **166**, 1092 (2006).
49. Petersen, M. *et al.* Brain imaging and neuropsychological assessment of individuals recovered from a mild to moderate SARS-CoV-2 infection. *Proceedings of the National Academy of Sciences* **120**, e2217232120 (2023).
50. Fischl, B. & Dale, A. M. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences* **97**, 11050–11055 (2000).
51. Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
52. Reveley, C. *et al.* Diffusion MRI anisotropy in the cerebral cortex is determined by unmyelinated tissue features. *Nat Commun* **13**, 6702 (2022).

53. Pasternak, O., Sochen, N., Gur, Y., Intrator, N. & Assaf, Y. Free water elimination and mapping from diffusion MRI. *Magnetic Resonance in Medicine* **62**, 717–730 (2009).

54. Boeing, H., Wahrendorf, J. & Becker, N. EPIC-Germany--A source for studies into diet and risk of chronic diseases. European Investigation into Cancer and Nutrition. *Ann Nutr Metab* **43**, 195–204 (1999).

55. Liao, T., Wei, Y., Luo, M., Zhao, G.-P. & Zhou, H. tmap: an integrative framework based on topological data analysis for population-scale microbiome stratification and association studies. *Genome Biol* **20**, 293 (2019).

56. Baryshnikova, A. Systematic Functional Annotation and Visualization of Biological Networks. *Cell Systems* **2**, 412–421 (2016).

57. Singh, G., Memoli, F. & Carlsson, G. *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*. (The Eurographics Association, 2007).

58. McInnes, L., Healy, J. & Astels, S. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software* **2**, 205 (2017).

59. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016).

60. Fruchterman, T. M. J. & Reingold, E. M. Graph drawing by force-directed placement. *Software: Practice and Experience* **21**, 1129–1164 (1991).

61. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).

62. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. & Heming, J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences* **622**, 178–210 (2023).

63. How, K. Y., Song, K. P. & Chan, K. G. Porphyromonas gingivalis: An Overview of Periodontopathic Pathogen below the Gum Line. *Front Microbiol* **7**, 53 (2016).

64. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **52**, W78–W82 (2024).

Figures

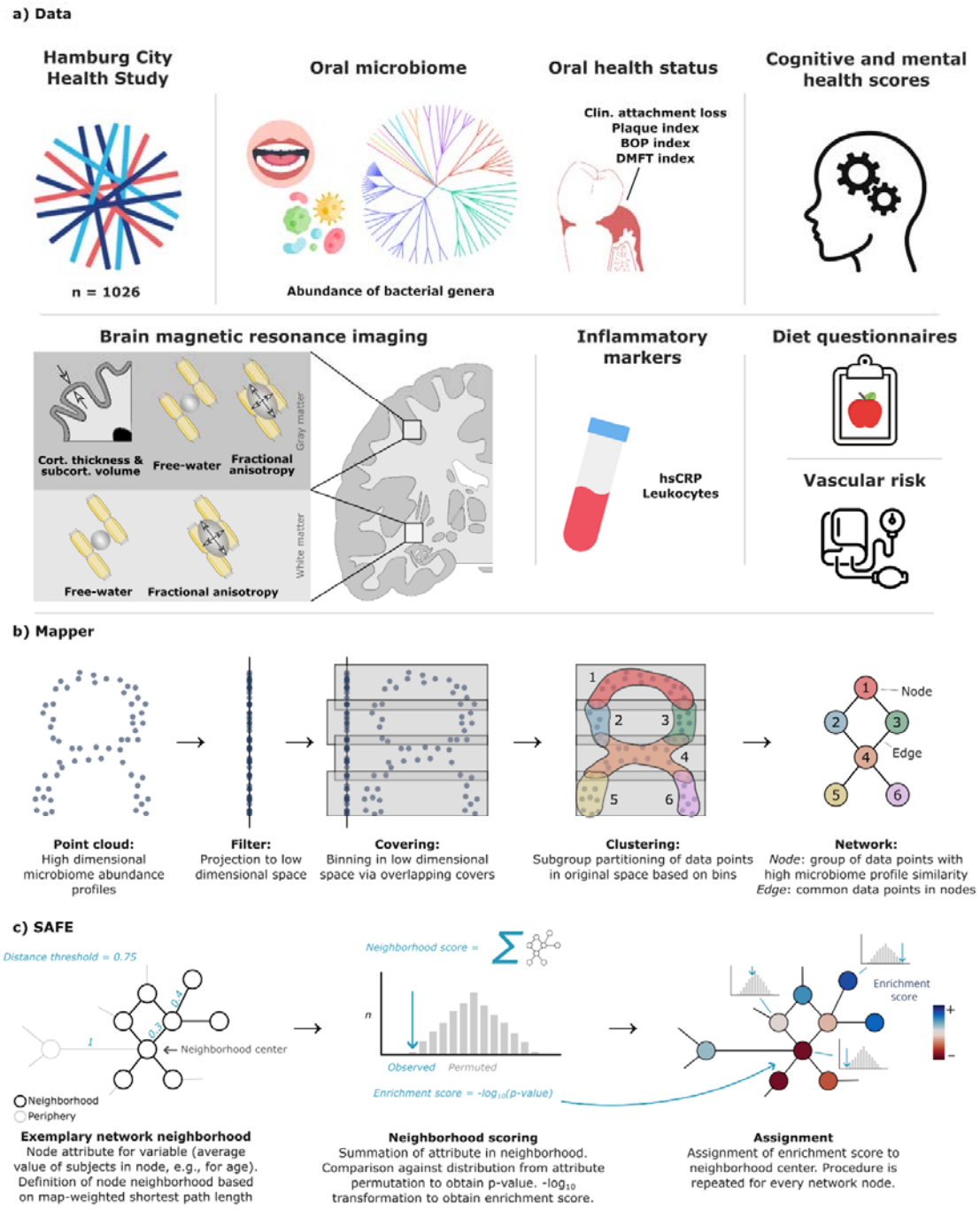
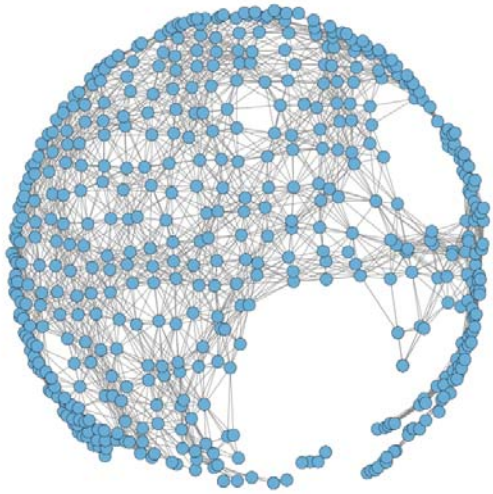


Figure 1. Methodology. a) Population-based data from the Hamburg City Health Study were used including oral microbiome abundance, assessments of oral health

status, cognitive function and mental health, multimodal brain MRI, circulating inflammatory markers, diet questionnaires, and vascular risk measures. b) We applied the Mapper algorithm to transform the high-dimensional genus-level oral microbiome profiles into an interpretable low-dimensional network representation. Mapper consists of multiple steps: (1) Inputting participant-level genus abundance; (2) Projecting data points to a low-dimensional space using a filter function for dimensionality reduction; (3) Dividing the low-dimensional space into overlapping covers, each containing a subset of data points; (4) Clustering data points within each cover based on their distances in the original high-dimensional space; (5) Constructing a network from the clustering results, where each node represents a cluster of participants, and links between nodes indicate shared participants between clusters. c) Spatial Analysis of Functional Enrichment (SAFE) was used to annotate the network derived via Mapper, identifying regions significantly enriched with specific attributes (e.g., systolic blood pressure). SAFE involves the following steps: (1) Computing node attributes by averaging variables of interest across participants within each node; (2) Defining the local neighborhood by identifying all nodes within a maximum distance threshold from the center node, with distance measured by map-weighted shortest path length (MSPL); (3) Calculating a neighborhood score by summing the attribute values of neighboring nodes; (4) Computing a p-value by comparing the empirical neighborhood score against a randomly permuted distribution, achieved through random node-to-attribute reassignments while preserving network topology; (5) Assigning an enrichment score to the neighborhood center using -log10 transformation of the multiple testing-corrected p-value. This procedure is repeated for each node, resulting in an enrichment map indicating where attributes are higher or lower than expected by chance. *Abbreviations*: BOP

index = bleeding on probing index; clin. = clinical; cort. = cortical; DMFT index = decayed/missing/filled teeth index; hsCRP = high-sensitivity c-reactive protein; subcort. = subcortical.

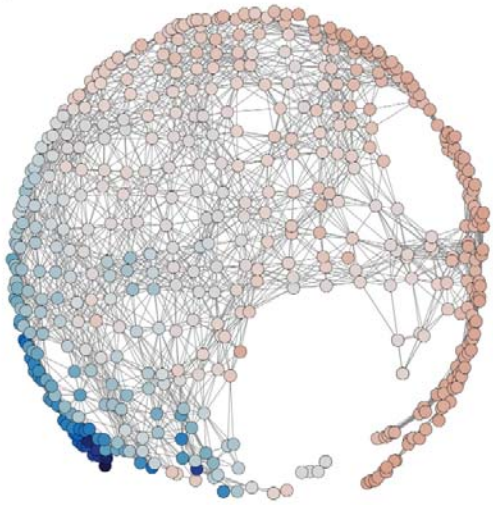
a) Microbiome similarity network



Node: group of subjects with similar microbiome profiles

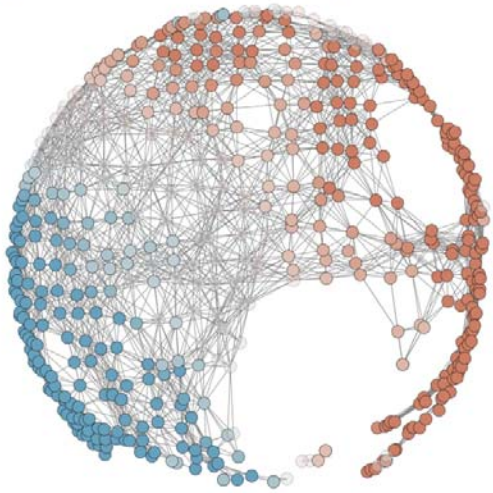
Edge: connected nodes have common subjects

b) Abundance - Porphyromonas



Average across subjects within node

c) Enrichment map - Porphyromonas



Significant positive enrichment

Not significant

Significant negative enrichment

$$\text{Enrichment ratio} = \frac{n_{\bullet} + n_{\bullet}}{n_{\bullet} + n_{\bullet} + n_{\circ}}$$

Enrichment score

Figure 2. Microbiome similarity network based on genus-level abundance. a) Microbiome similarity network obtained by applying the Mapper algorithm to the abundances of oral microbiome genera consisting of 528 nodes and 4061 edges. Nodes represent groups of participants with similar microbiome profiles. Edges connect nodes that share common participants. Within the network, proximity signifies microbiome profile similarity among participants. b) Exemplary network distribution of z-scored *Porphyromonas* abundance. *Porphyromonas* was selected given its high relevance in periodontal disease.[63] c) Enrichment map of *Porphyromonas* abundance. Nodes with positive enrichment scores are colored blue indicating a higher *Porphyromonas* abundance of the node and its neighborhood than expected by random permutation. Nodes with negative enrichment scores are colored red indicating lower *Porphyromonas* abundance than expected by permutation. Non-significant nodes are shown in a lighter shade. The enrichment ratio represents a measure of how strongly a phenotype is enriched on a network. It is computed by dividing the count of significantly enriched nodes by all nodes.

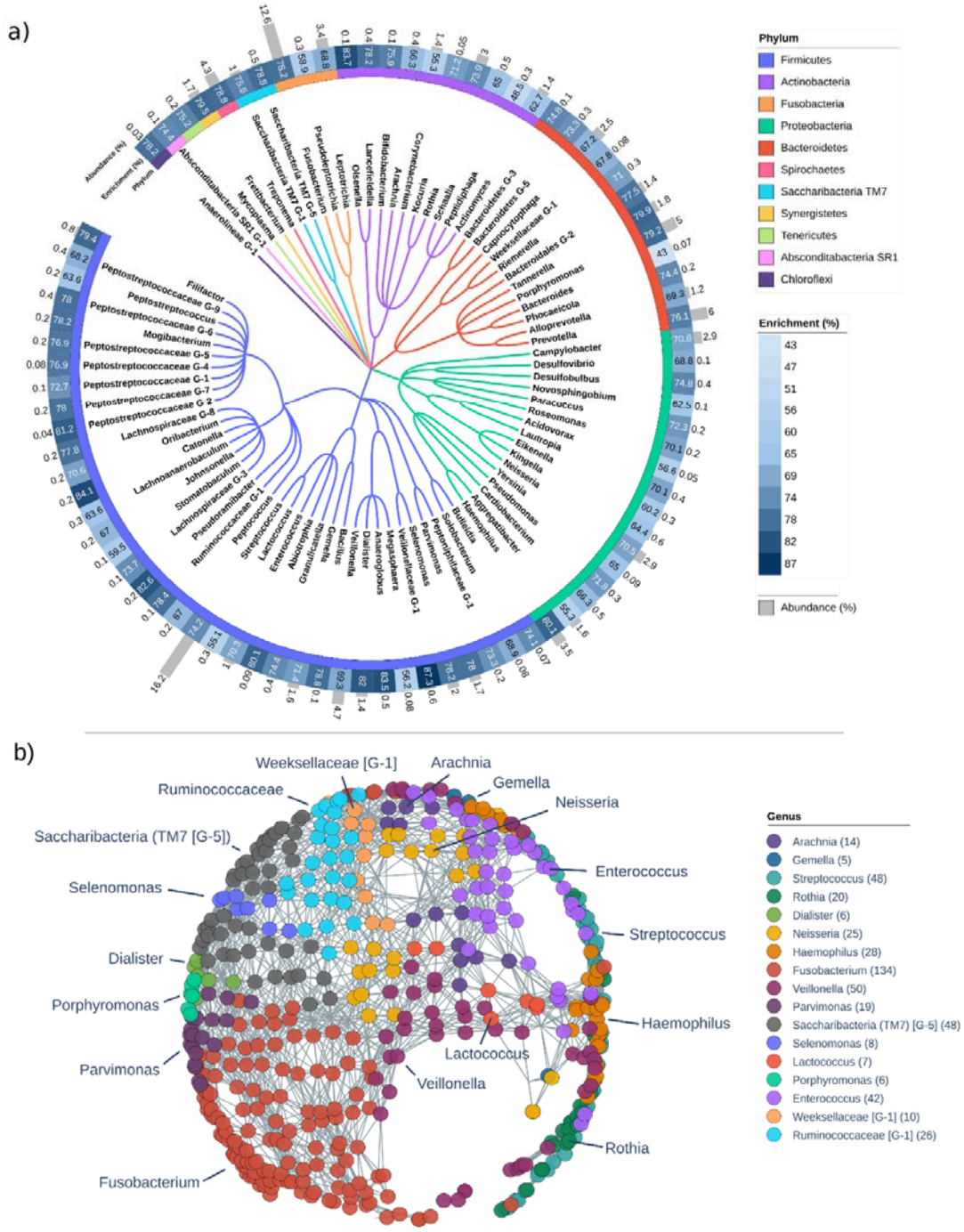


Figure 3. Enrichment analysis of microbiome phenotypes. a) Circular phylogenetic tree at the level of oral microbiome genera. The inner band shows phyla with corresponding coloring of the tree, the mid band displays the enrichment ratio and the outer band shows overall relative abundance. The phylogenetic tree

was visualized using iTOL (v6).[64] b) Dominance analysis: Nodes of the microbiome similarity network are colored to represent the bacterial genus with the highest enrichment score. Genera with dominance in five or more nodes are shown. *Abbreviations*: GM = gray matter, WM = white matter.

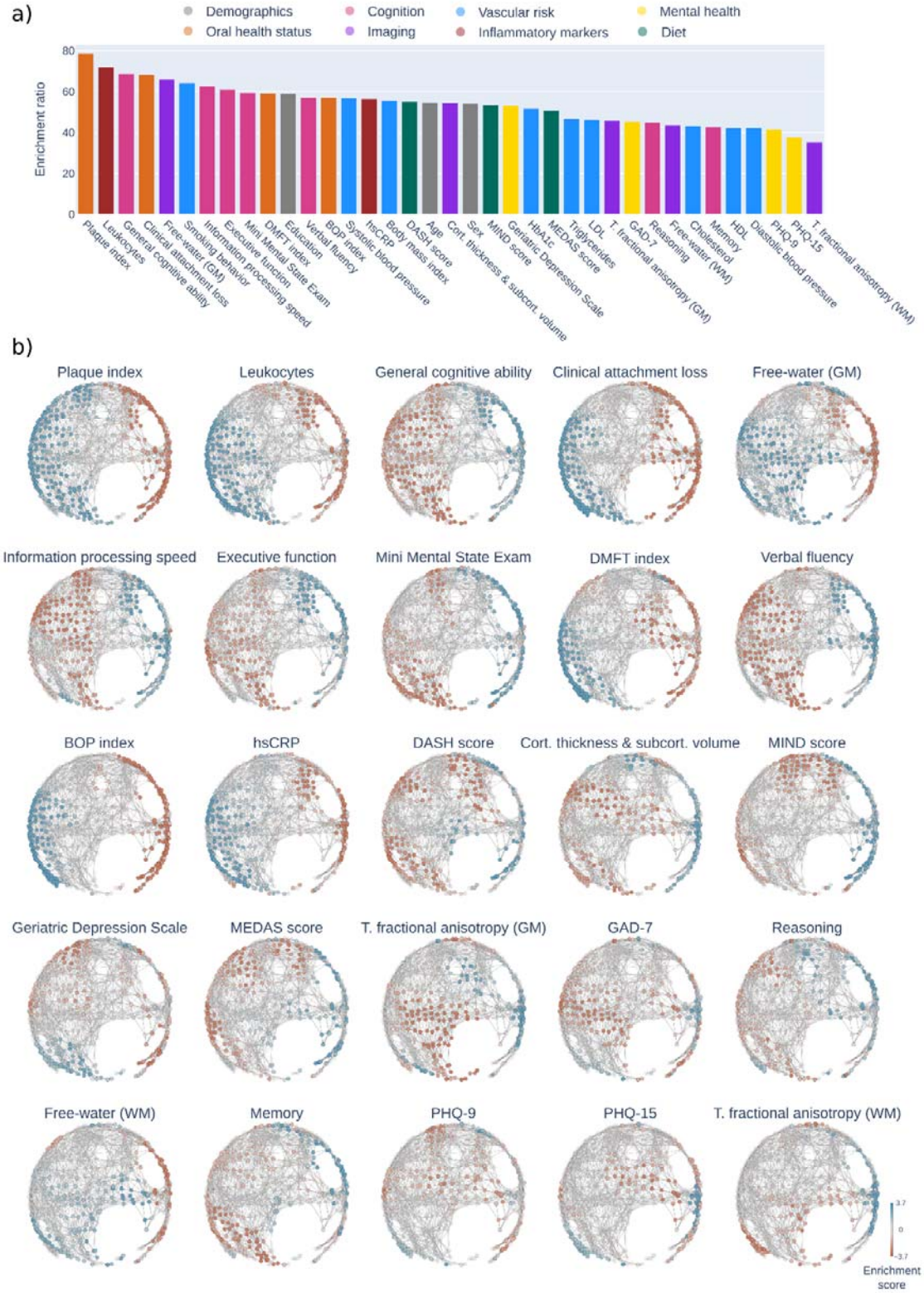


Figure 4. Enrichment analysis of non-microbiome phenotypes. a) The bar plot

displays the enrichment ratio (= significantly enriched network nodes / all network nodes) of non-microbiome phenotypes. The enrichment ratio reflects how strongly a phenotype's variance is captured by the topology of the microbiome network obtained via Mapper. The bars of the plot are colored by variable category. b) Enrichment maps: Enrichment scores (-$\log_{10}$-transformed p-values) are mapped on the microbiome similarity network. Nodes with positive enrichment scores, indicating a higher neighborhood score than expected by chance, are colored blue, while nodes with negative enrichment scores, indicating a lower neighborhood score than expected by chance, are colored red. Non-significant nodes are shown in a lighter shade. Non-microbiome phenotypes are ordered by enrichment ratio. Enrichment maps for demographics and vascular risk factors are not displayed and can be found in the *supplementary materials.* For interactive versions of these plots see: https://osf.io/9ykv5. *Abbreviations*: BOP index = bleeding on probing index; DMFT index = decayed/missing/filled teeth index; GM = gray matter, hsCRP = high sensitivity c-reactive peptide; LDL = low density lipoprotein, T. fractional anisotropy = tissue fractional anisotropy; WM = white matter.

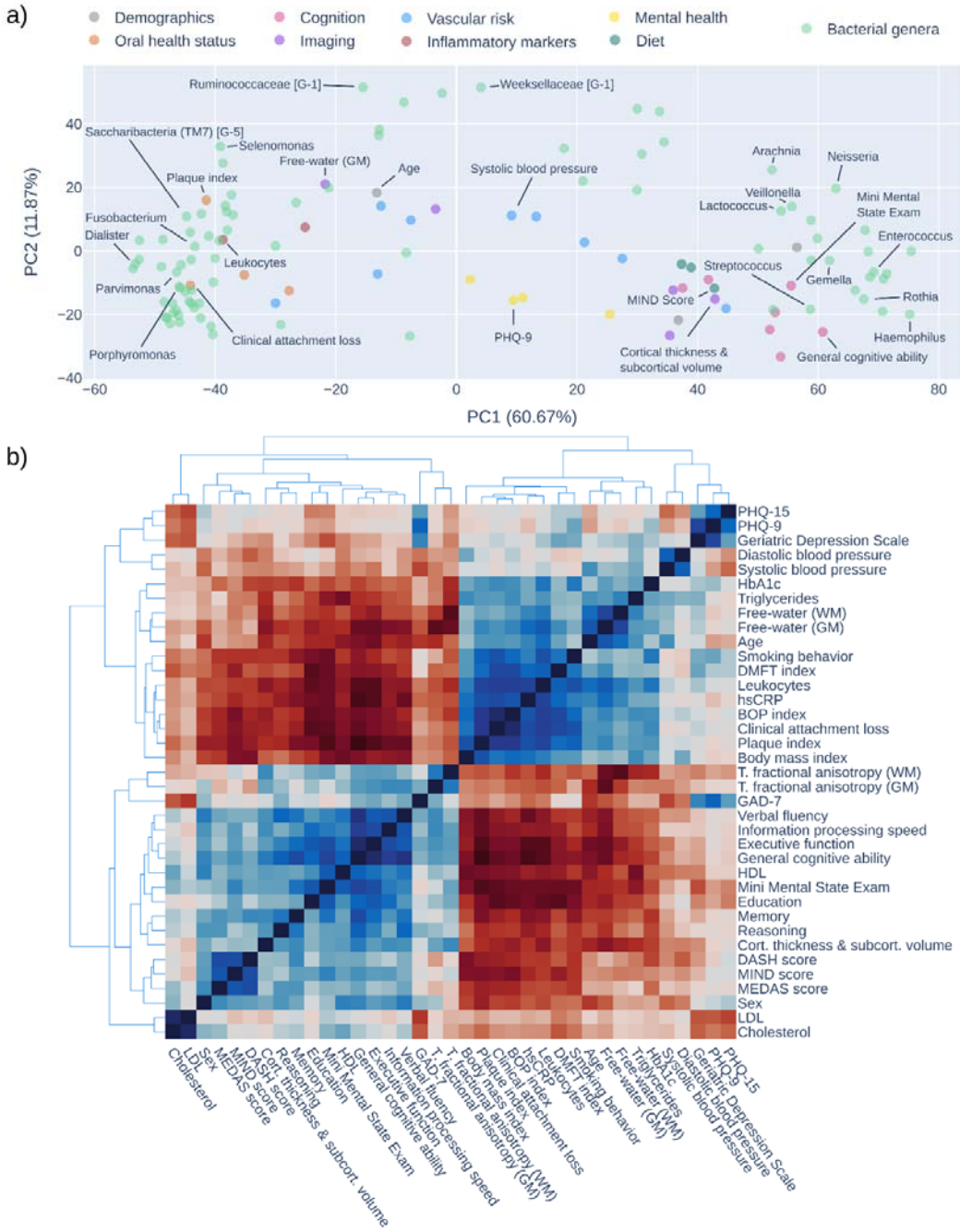


Figure 5. Co-enrichment analysis. a) Principal component analysis of enrichment scores: The plot illustrates the phenotypes' location in principal component space, where proximity suggests similarity in enrichment patterns. Points are color-coded by phenotype category. Dominant genera and selected non-microbiome measures from each phenotype group were highlighted with annotations. For a fully annotated

scatter plot, refer to the *supplementary materials*. b) Co-enrichment heatmap: The heatmap presents the Spearman correlation of enrichment scores for non-microbiome phenotypes. The order is determined by hierarchical clustering of the enrichment scores. For interactive versions of these plots see: https://osf.io/9ykv5. *Abbreviations*: BOP index = bleeding on probing index, DMFT index = decayed/missing/filled teeth index, GM = gray matter, LDL = low density lipoprotein, T. fractional anisotropy = tissue fractional anisotropy, WM = white matter.

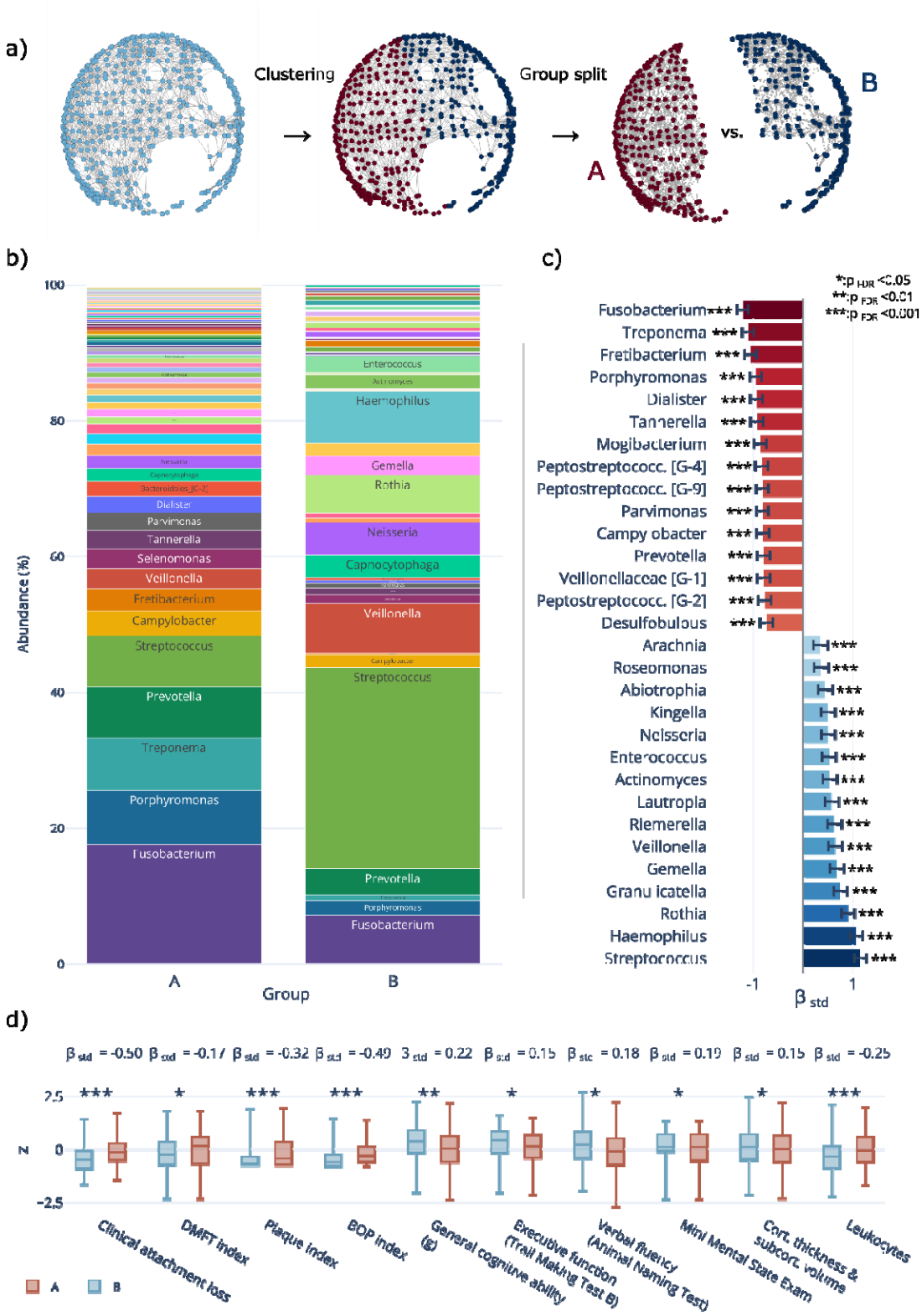

a)
Clustering
Group split
A
vs.
B
b)
Abundance (%)
Group
A
B
Fusobacterium
Porphyromonas
Treponema
Prevotella
Streptococcus
Campylobacter
Fretibacterium
Veillonella
Selenomonas
Tannerella
Parvimonas
Dialister
Capnocytophaga
Neisseria
Haemophilus
Gemella
Rothia
Enterococcus
c)
*:p FDR <0.05
**:p FDR <0.01
***:p FDR <0.001
Fusobacterium
Treponema
Fretibacterium
Porphyromonas
Dialister
Tannerella
Mogibacterium
Peptostreptococc. [G-4]
Peptostreptococc. [G-9]
Parvimonas
Campy obacter
Prevotella
Veillonellaceae [G-1]
Peptostreptococc. [G-2]
Desulfobulous
Arachnia
Roseomonas
Abiotrophia
Kingella
Neisseria
Enterococcus
Actinomyces
Lautropia
Riemerella
Veillonella
Gemella
Granu icatella
Rothia
Haemophilus
Streptococcus
-1
1
β std
d)
β std = -0.50 β std = -0.17 β std = -0.32 β std = -0.49 β std = 0.22 β std = 0.15 β std = 0.18 β std = 0.19 β std = 0.15 β std = -0.25
z
Clinical attachment loss
DMFT index
Plaque index
BOP index
General cognitive ability (g)
Executive function (Trail Making Test B)
Verbal fluency (Animal Naming Test)
Mini Mental State Exam
Cort. thickness & subcort. volume
Leukocytes
A
B

Figure 6. Group analysis. a) Approach: The nodes of the microbiome similarity network were divided into two distinct groups ("A" and "B") using k-means clustering. Participants were categorized based on these groupings, with individuals present in both groups being excluded from the analysis (n = 63). The two participant groups were then statistically compared using linear regression analysis, adjusting for age, sex, education and vascular risk factors. b) Relative abundance of bacterial genera within groups. c) Top 15 positive and negative differences in microbiome genera: Negative regression coefficients (higher in the "A" group) are shown in red, while positive coefficients (higher in the "B" group) are shown in blue. Significance is indicated by asterisks. For a bar plot displaying coefficients of all investigated genera see *supplementary materials*. d) Shown are box plots for significant associations. Box plot colors correspond to the groups. Statistical significance level is indicated by asterisks. For a box plots of remaining non-microbiome phenotypes see *supplementary materials*. For interactive versions of these plots see: https://osf.io/9ykv5. *Abbreviations*: β_{std} = standardized beta coefficient for the group variable. BOP index = bleeding on probing index, DMFT index = decayed/missing/filled teeth index, GM = gray matter; WM = white matter.

Tables

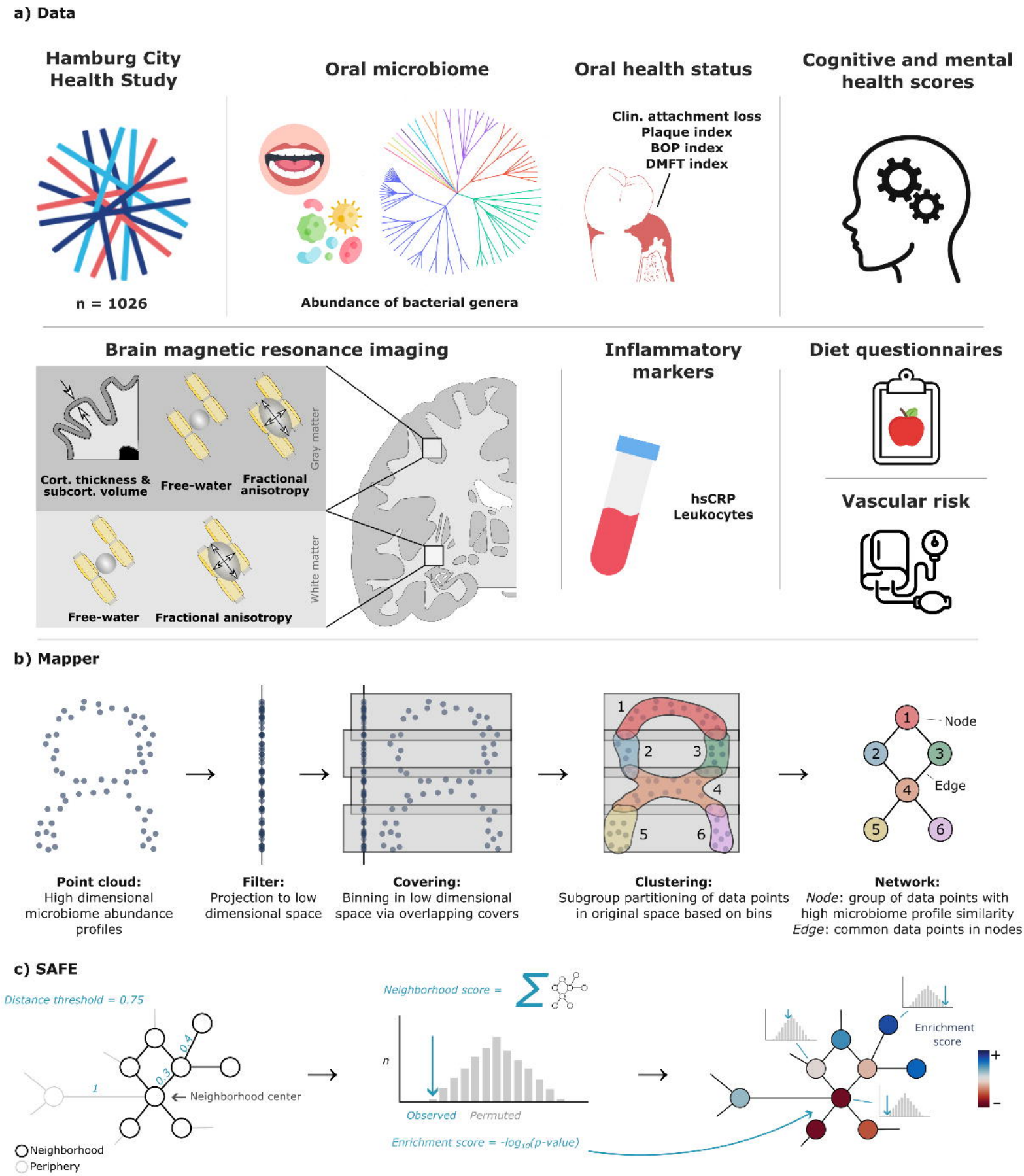
Table 1. Sample characteristics

Metric	Value[a]
Age, years	63.72 ± 8.17 (1,026)
Female, n (% female)	438 (42.7)
Education, ISCED	2.42 ± 0.58 (999)
Oral health measures	
Clinical attachment loss	2.60 ± 0.83 (842)
Plaque index	19.22 ± 26.30 (830)
Bleeding on probing index	15.38 ± 19.91 (1,011)
DMFT index	19.27 ± 4.90 (842)
Cognitive scores	
Animal Naming Test	24.66 ± 6.67 (977)
Mini Mental State Exam	27.85 ± 1.68 (973)
Multiple Choice Vocabulary Intelligence Test	31.42 ± 3.43 (838)
Trail Making Test A, sec	39.82 ± 14.07 (925)
Trail Making Test B, sec	89.44 ± 37.66 (917)
Word List Recall	7.70 ± 1.88 (949)
Mental health scores	
PHQ-9	3.21 ± 3.61 (1,025)
PHQ-15	4.27 ± 3.90 (1,025)
Geriatric Depression Scale	1.83 ± 2.20 (914)
GAD-7	2.41 ± 2.90 (1,025)
Global neuroimaging features	
Free-water (gray matter)	0.50 ± 0.05 (909)
Free-water (white matter)	0.31 ± 0.03 (909)
Tissue fractional anisotropy (gray matter)	0.19 ± 0.01 (909)
Tissue fractional anisotropy (white matter)	0.41 ± 0.01 (909)
Mean cortical thickness, mm	2.63 ± 0.11 (986)
Mean subcortical volume, ml	3188.31 ± 313.66 (986)
Inflammatory markers	
hsCRP	0.25 ± 0.45 (970)
Leukocytes	6.13 ± 1.75 (1,006)
Diet	
MEDAS score	4.46 ± 1.89 (970)
DASH score	4.46 ± 1.06 (970)
MIND score	6.30 ± 1.78 (970)
Vascular risk measurements	
Systolic blood pressure	142.12 ± 20.06 (997)
Diastolic blood pressure	83.46 ± 10.55 (997)
Body mass index	26.60 ± 4.32 (995)
Smoking, %	173 (16.9)
Triglycerides	117.67 ± 75.68 (1,007)
Cholesterol	208.28 ± 40.83 (1,008)

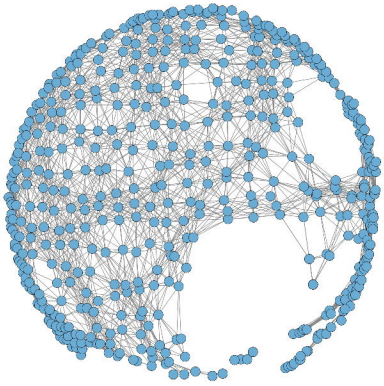
LDL	121.56 ± 36.53 (997)
HDL	63.51 ± 19.00 (1,008)
HbA1c	5.63 ± 0.54 (1,007)

Abbreviations: DMFT index = decayed/missing/filled teeth index, HDL = high density lipoprotein, hsCRP = high sensitivity c-reactive peptide, LDL = low density lipoprotein, ISCED = International Standard Classification of Education, mm = millimeters, sec = seconds

[a]Presented as median [IQR] (N)


a) Data
Hamburg City Health Study
n = 1026
Oral microbiome
Abundance of bacterial genera
Oral health status
Clin. attachment loss
Plaque index
BOP index
DMFT index
Cognitive and mental health scores
Brain magnetic resonance imaging
Cort. thickness & subcort. volume
Free-water
Fractional anisotropy
Gray matter
Free-water
Fractional anisotropy
White matter
Inflammatory markers
hsCRP
Leukocytes
Diet questionnaires
Vascular risk
b) Mapper
Point cloud:
High dimensional microbiome abundance profiles
Filter:
Projection to low dimensional space
Covering:
Binning in low dimensional space via overlapping covers
Clustering:
Subgroup partitioning of data points in original space based on bins
Network:
Node: group of data points with high microbiome profile similarity
Edge: common data points in nodes
Node
Edge
c) SAFE
Distance threshold = 0.75
Neighborhood center
Neighborhood
Periphery
Exemplary network neighborhood
Node attribute for variable (average value of subjects in node, e.g., for age). Definition of node neighborhood based on map-weighted shortest path length
Neighborhood score =
Observed
Permuted
Enrichment score = -log10(p-value)
Neighborhood scoring
Summation of attribute in neighborhood. Comparison against distribution from attribute permutation to obtain p-value. -log10 transformation to obtain enrichment score.
Enrichment score
Assignment
Assignment of enrichment score to neighborhood center. Procedure is repeated for every network node.

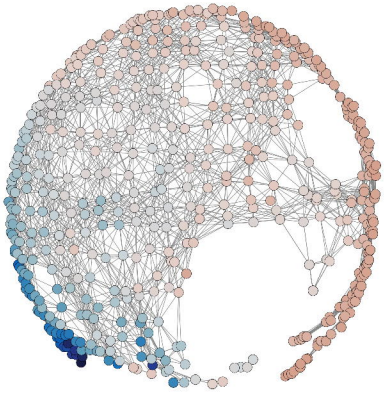
a) Microbiome similarity network



Node: group of subjects with similar microbiome profiles

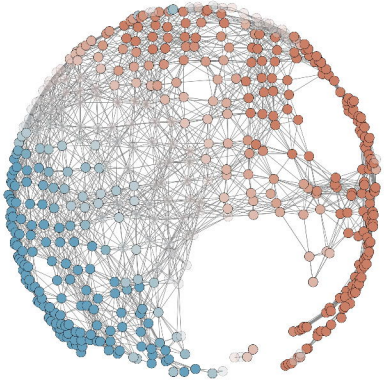
Edge: connected nodes have common subjects

b) Abundance - Porphyromonas



Average across subjects within node

c) Enrichment map - Porphyromonas



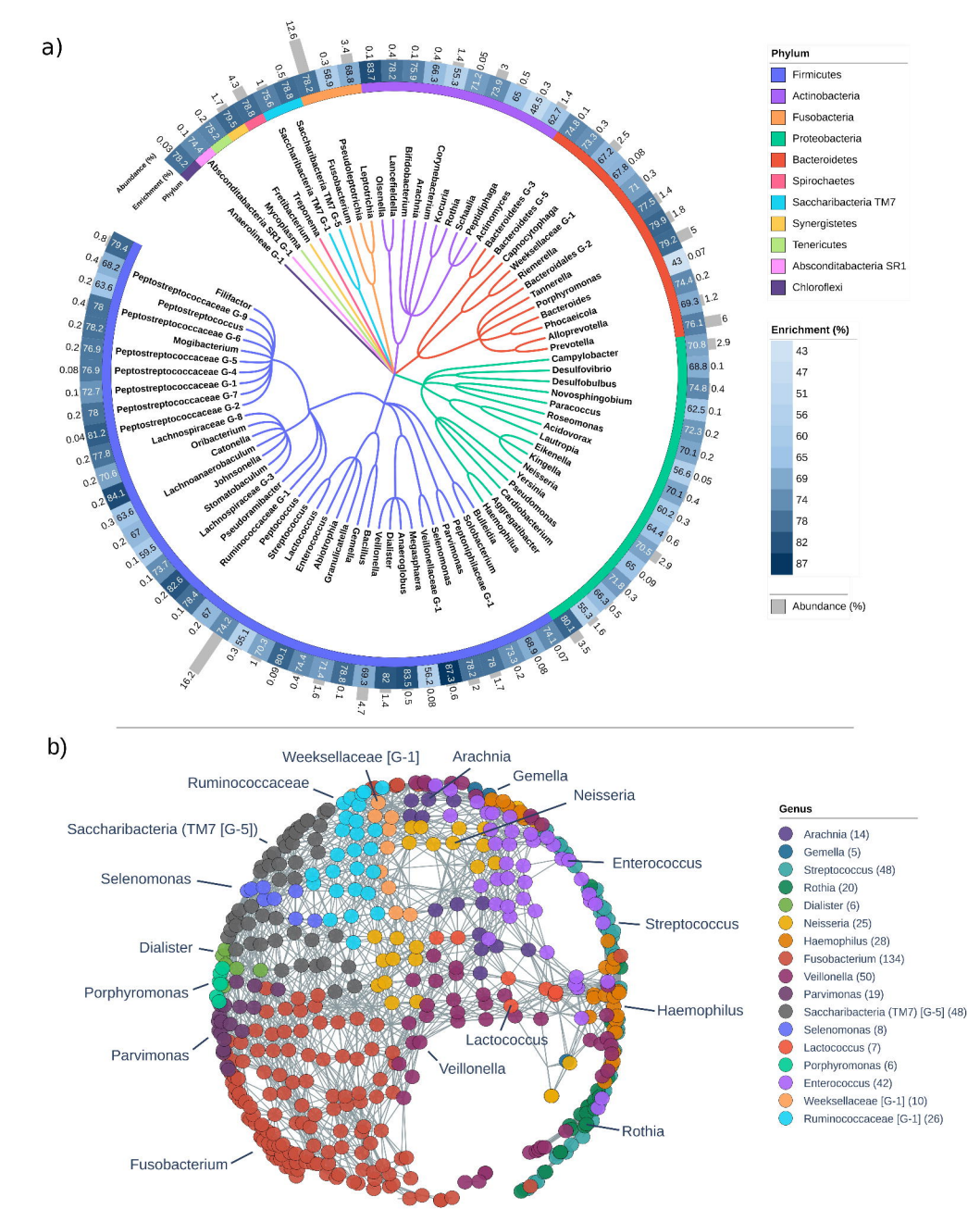
Significant positive enrichment

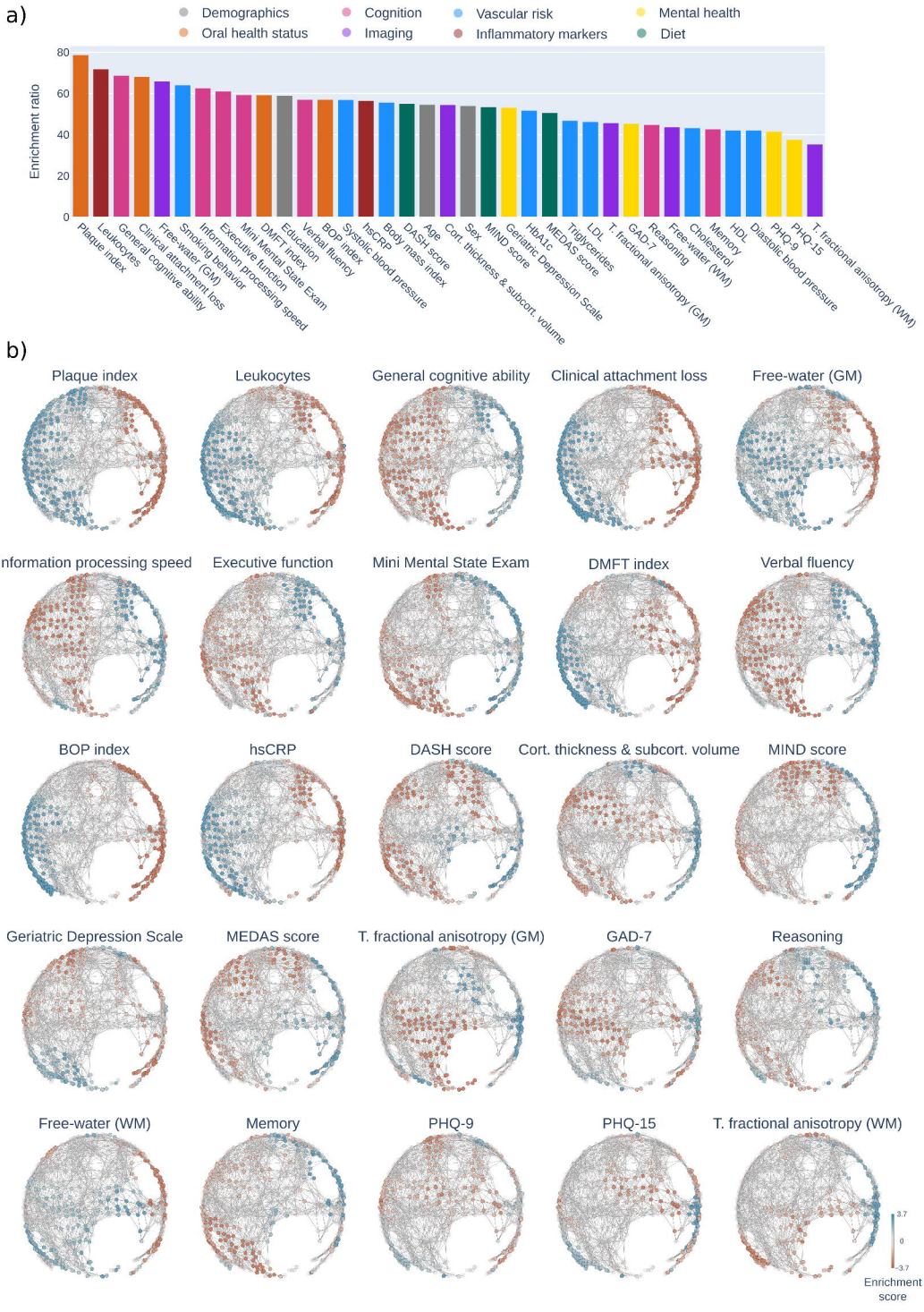
Not significant

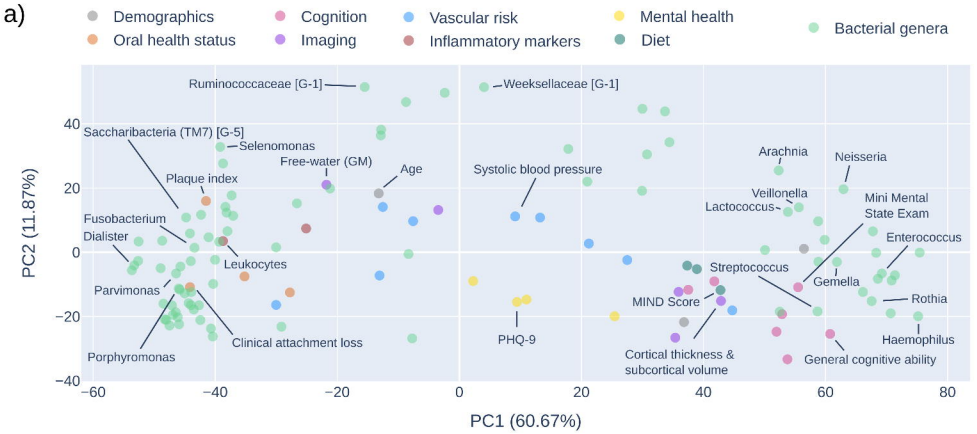
Significant negative enrichment

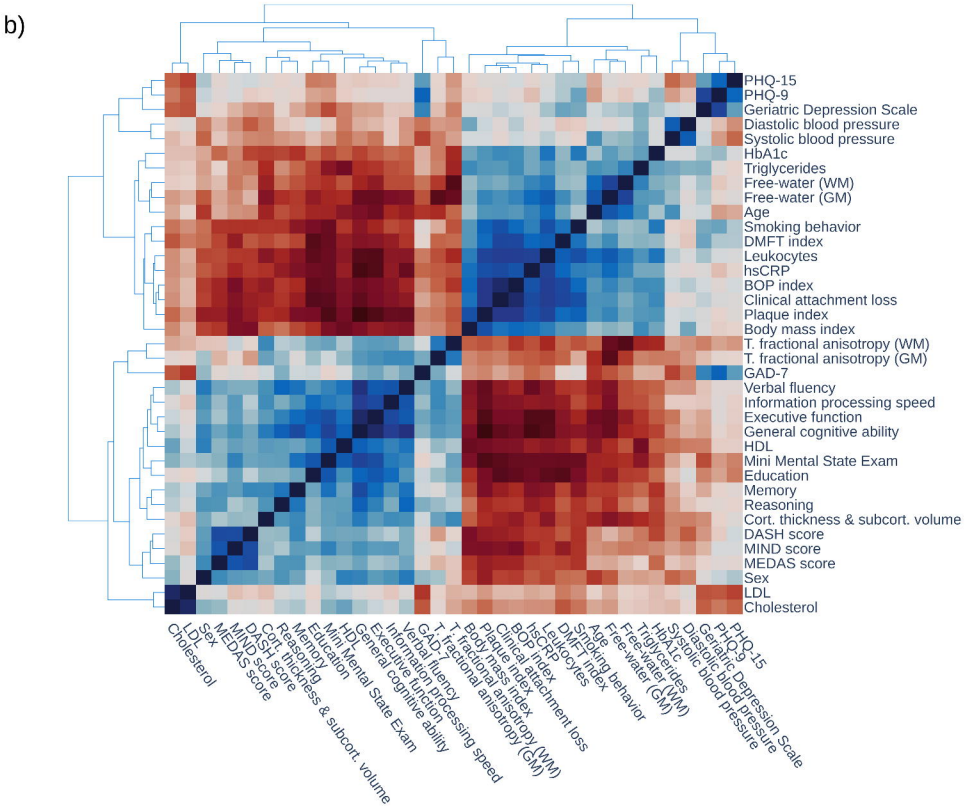
$$\text{Enrichment ratio} = \frac{n_{\bullet} + n_{\bullet}}{n_{\bullet} + n_{\bullet} + n_{\circ}}$$

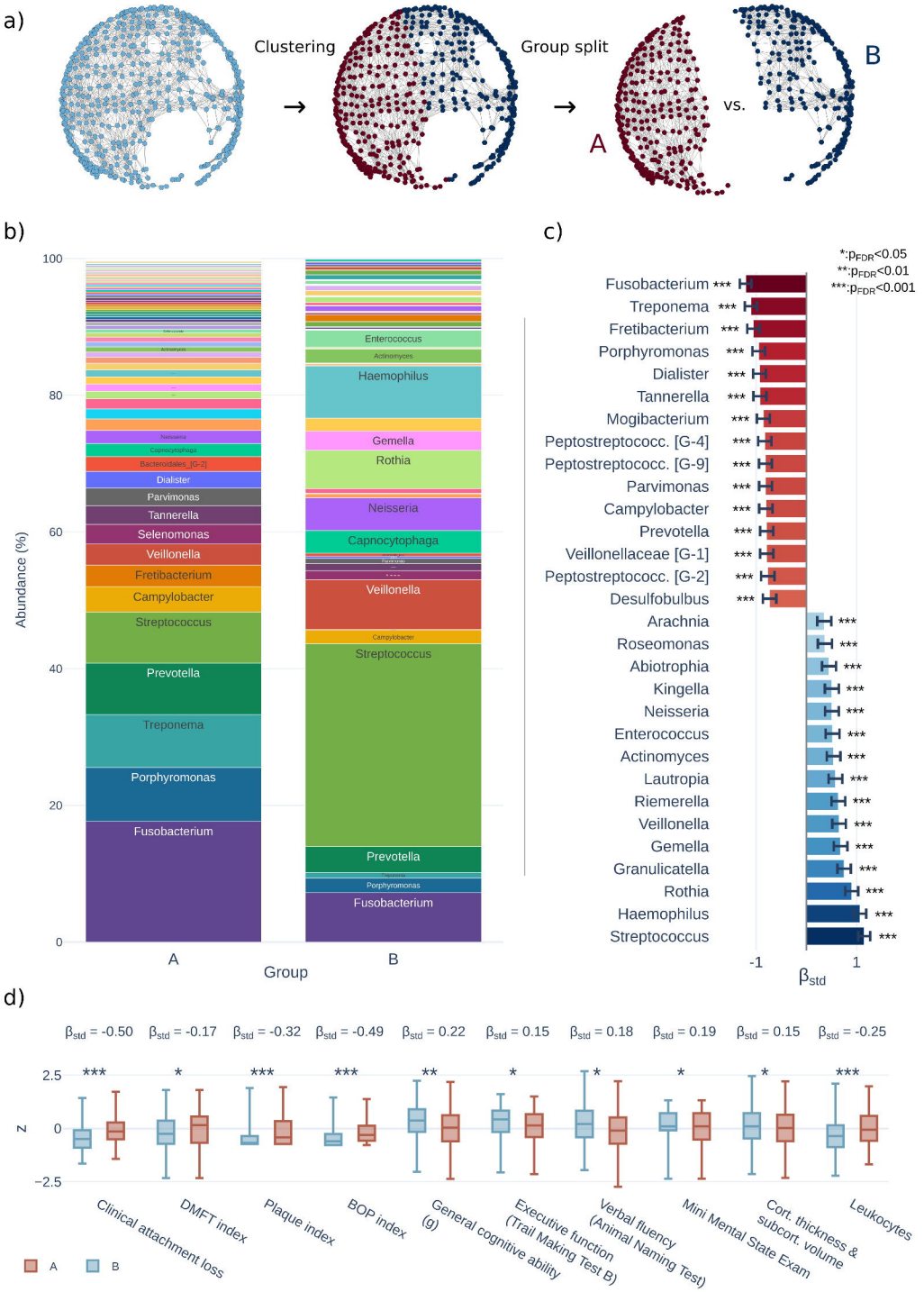
Enrichment score


a)
Phylum
Firmicutes
Actinobacteria
Fusobacteria
Proteobacteria
Bacteroidetes
Spirochaetes
Saccharibacteria TM7
Synergistetes
Tenericutes
Absconditabacteria SR1
Chloroflexi
Enrichment (%)
43
47
51
56
60
65
69
74
78
82
87
Abundance (%)
Abundance (%)
Enrichment (%)
Phylum
b)
Weeksellaceae [G-1]
Arachnia
Gemella
Neisseria
Ruminococcaceae
Saccharibacteria (TM7 [G-5])
Enterococcus
Selenomonas
Streptococcus
Dialister
Porphyromonas
Haemophilus
Lactococcus
Parvimonas
Veillonella
Rothia
Fusobacterium
Genus
Arachnia (14)
Gemella (5)
Streptococcus (48)
Rothia (20)
Dialister (6)
Neisseria (25)
Haemophilus (28)
Fusobacterium (134)
Veillonella (50)
Parvimonas (19)
Saccharibacteria (TM7) [G-5] (48)
Selenomonas (8)
Lactococcus (7)
Porphyromonas (6)
Enterococcus (42)
Weeksellaceae [G-1] (10)
Ruminococcaceae [G-1] (26)


a)
Demographics
Oral health status
Cognition
Imaging
Vascular risk
Inflammatory markers
Mental health
Diet
Enrichment ratio
0
20
40
60
80
Plaque index
Leukocytes
General cognitive ability
Clinical attachment loss
Free-water (GM)
Smoking behavior
Information processing speed
Executive function
Mini Mental State Exam
DMFT index
Education
Verbal fluency
BOP index
Systolic blood pressure
hsCRP
Body mass index
DASH score
Age
Cort. thickness & subcort. volume
Sex
MIND score
Geriatric Depression Scale
HbA1c
MEDAS score
Triglycerides
LDL
T. fractional anisotropy (GM)
GAD-7
Reasoning
Free-water (WM)
Cholesterol
Memory
HDL
Diastolic blood pressure
PHQ-9
PHQ-15
T. fractional anisotropy (WM)
b)
Plaque index
Leukocytes
General cognitive ability
Clinical attachment loss
Free-water (GM)
Information processing speed
Executive function
Mini Mental State Exam
DMFT index
Verbal fluency
BOP index
hsCRP
DASH score
Cort. thickness & subcort. volume
MIND score
Geriatric Depression Scale
MEDAS score
T. fractional anisotropy (GM)
GAD-7
Reasoning
Free-water (WM)
Memory
PHQ-9
PHQ-15
T. fractional anisotropy (WM)
3.7
0
-3.7
Enrichment score


a)
Demographics
Oral health status
Cognition
Imaging
Vascular risk
Inflammatory markers
Mental health
Diet
Bacterial genera
Ruminococcaceae [G-1]
Weeksellaceae [G-1]
Saccharibacteria (TM7) [G-5]
Selenomonas
Free-water (GM)
Age
Plaque index
Systolic blood pressure
Arachnia
Neisseria
Veillonella
Lactococcus
Mini Mental State Exam
Fusobacterium
Dialister
Leukocytes
Enterococcus
Streptococcus
Gemella
Parvimonas
MIND Score
Rothia
Clinical attachment loss
PHQ-9
Haemophilus
Porphyromonas
Cortical thickness & subcortical volume
General cognitive ability
PC2 (11.87%)
PC1 (60.67%)


b)
PHQ-15
PHQ-9
Geriatric Depression Scale
Diastolic blood pressure
Systolic blood pressure
HbA1c
Triglycerides
Free-water (WM)
Free-water (GM)
Age
Smoking behavior
DMFT index
Leukocytes
hsCRP
BOP index
Clinical attachment loss
Plaque index
Body mass index
T. fractional anisotropy (WM)
T. fractional anisotropy (GM)
GAD-7
Verbal fluency
Information processing speed
Executive function
General cognitive ability
HDL
Mini Mental State Exam
Education
Memory
Reasoning
Cort. thickness & subcort. volume
DASH score
MIND score
MEDAS score
Sex
LDL
Cholesterol


a)
Clustering
Group split
A
vs.
B
b)
Abundance (%)
100
80
60
40
20
0
Fusobacterium
Porphyromonas
Treponema
Prevotella
Streptococcus
Campylobacter
Veillonella
Fretibacterium
Selenomonas
Tannerella
Parvimonas
Dialister
Bacteroidales_[G-2]
Capnocytophaga
Neisseria
Gemella
Actinomyces
Enterococcus
Haemophilus
Rothia
A
B
Group
c)
*:pFDR<0.05
**:pFDR<0.01
***:pFDR<0.001
Fusobacterium ***
Treponema ***
Fretibacterium ***
Porphyromonas ***
Dialister ***
Tannerella ***
Mogibacterium ***
Peptostreptococc. [G-4] ***
Peptostreptococc. [G-9] ***
Parvimonas ***
Campylobacter ***
Prevotella ***
Veillonellaceae [G-1] ***
Peptostreptococc. [G-2] ***
Desulfobulbus ***
Arachnia ***
Roseomonas ***
Abiotrophia ***
Kingella ***
Neisseria ***
Enterococcus ***
Actinomyces ***
Lautropia ***
Riemerella ***
Veillonella ***
Gemella ***
Granulicatella ***
Rothia ***
Haemophilus ***
Streptococcus ***
-1
1
βstd
d)
βstd = -0.50
βstd = -0.17
βstd = -0.32
βstd = -0.49
βstd = 0.22
βstd = 0.15
βstd = 0.18
βstd = 0.19
βstd = 0.15
βstd = -0.25
2.5
0
−2.5
z
Clinical attachment loss
DMFT index
Plaque index
BOP index
General cognitive ability (g)
Executive function (Trail Making Test B)
Verbal fluency (Animal Naming Test)
Mini Mental State Exam
Cort. thickness & subcort. volume
Leukocytes
A
B