Transcriptome analysis in whole blood reveals increased microbial diversity in schizophrenia

Authors: Loes M Olde Loohuis[#1], Serghei Mangul[#2], Anil PS Ori[1], Guillaume Jospin[3], David Koslicki[4], Harry Taegyun Yang[2], Timothy Wu[1], Marco P Boks[5], Catherine Lomen-Hoerth[6], Martina Wiedau-Pazos[7], Rita M Cantor[8], Willem M de Vos[9,10], René S Kahn[5], Eleazar Eskin[2], Roel A Ophoff[*,1,5,8]

Affiliations:

[1]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University California Los Angeles, Los Angeles, USA

[2]Department of Computer Science, University of California Los Angeles, Los Angeles, USA

[3]Davis Genome Center, University of California, Davis, USA

[4]Mathematics Department, Oregon State University, Corvallis, USA

[5]Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

[6]Department of Neurology, University of California San Francisco, San Francisco, USA

[7]Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, USA

[8]Department of Human Genetics, University of California Los Angeles, Los Angeles, USA

[9]Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands

[10]Department of Bacteriology and Immunology, Immunobiology Research Program, University of Helsinki, Helsinki, Finland.

* Correspondence to: Roel A. Ophoff ophoff@ucla.edu

# Equal contribution

Key words: RNA sequencing, microbiome, schizophrenia, unmapped reads, blood

Abstract:

The role of the human microbiome in health and disease is increasingly appreciated. We studied the composition of microbial communities present in blood across 192 individuals, including healthy controls and patients with three disorders affecting the brain: schizophrenia, amyotrophic lateral sclerosis and bipolar disorder. By using high quality unmapped RNA sequencing reads as candidate microbial reads, we performed profiling of microbial transcripts detected in whole blood. We were able to detect a wide range of bacterial and archaeal phyla in blood. Interestingly, we observed an increased microbial diversity in schizophrenia patients compared to the three other groups. We replicated this finding in an independent schizophrenia case-control cohort. This increased diversity is inversely correlated with estimated cell abundance of a subpopulation of $CD8^+$ memory T cells in healthy controls, supporting a link between microbial products found in blood, immunity and schizophrenia.

Introduction

Microbial communities in and on the human body represent a complex mixture of eukaryotes, bacteria, archaea and viruses. In recent years, mounting evidence has demonstrated the involvement of the microbiome in human health and disease. In particular, through the 'microbiota-gut-brain-axis' (1, 2), the microbiome has been implicated in complex psychiatric disorders, including schizophrenia and major depressive disorder(3-8), possibly via an impact on intestinal permeability(9).

High-throughput sequencing offers a powerful culture-independent approach to study the underlying diversity of microbial communities in their natural habitats across different human tissues (10) and diseases (3, 11-15). The majority of current microbiome studies use fecal samples and target 16S ribosomal RNA gene sequencing (16). With the availability of comprehensive compendia of reference microbial genomes and phylogenetic marker genes (17), it has become feasible to use non-targeted sequencing data to identify the microbial species across different human tissues and diseases in a relatively inexpensive and easy way.

Other than in cases of sepsis, we currently lack a comprehensive understanding of the human microbiome in blood, as blood has been generally considered a sterile environment lacking proliferating microbes (18). However, over the past few decades, this assumption has been challenged (19, 20), and the presence of a microbiome in the blood has received increasing attention (21-23).

To explore potential connections between the microbiome and diseases of the brain, we performed a comprehensive analysis of microbial products detected in blood in almost two

67 hundred individuals, including patients with schizophrenia, bipolar disorder and sporadic
68 amyotrophic lateral sclerosis. These three disease groups represent complex polygenic traits
69 that affect the central nervous system with largely unknown etiology. Moreover, roles for the
70 microbiome in all the diseases have been previously hypothesized (5, 24-26). We used available
71 high quality RNA sequencing (RNA-Seq) reads from whole blood that fail to map to the human
72 genome as candidate microbial reads for microbial classification. We observed an increased
73 diversity of microbial communities in schizophrenia patients, and we replicated this finding in
74 an independent dataset. Careful analyses, including the use of positive and negative control
75 datasets, suggest that these detected phyla represent true microbial communities in whole
76 blood and are not present in samples due to contaminants. With the increasing number of RNA-
77 Seq data sets, our approach may have great potential for application across different tissues
78 and disease types.

79

80

Materials and Methods

A brief description of Materials and Methods follows below; see Supplementary Methods for the full details.

Sample Description

The discovery sample consists of unaffected controls (Controls, n=49) and patients with three brain-related disorders: schizophrenia (SCZ, n=48), amyotrophic lateral sclerosis (ALS, n=47) and bipolar disorder (BPD, n=48). The replication sample includes Controls (n=88) and SCZ samples (n=91). Sample recruitment of the cohorts is described in the Supplementary Methods. All study methods were approved by the institutional review board of the University of California at Los Angeles, San Francisco or the Medical Research Ethics Committee of the University Medical Center Utrecht at The Netherlands. All participants provided written informed consent.

Sample sequencing

For the discovery sample, RNA-Seq libraries were prepared using Illumina's TruSeq RNA v2 protocol, including ribo-depletion protocol (Ribo-Zero Gold). In total, we obtained 6.8 billion 2x100bp paired-end reads for the primary study (35.3M ± 6.0 paired-end reads per sample). The replication sample was processed at the same core facility using the same standardized procedures as the discovery sample. However, the RNA-Seq libraries were prepared with poly(A) enrichment, a procedure more selective than the total RNA that was used for the discovery sample. A total of 3.8 billion reads were obtained (26.3M ± 12.0).

101 **Sequence Analysis**

102 We separated human and non-human reads, and use the latter as candidate microbial reads for
103 taxonomic profiling of microbial communities. To identify potentially microbial reads, we
104 developed the following pipeline. First, we filtered read pairs and singleton reads mapped to
105 the human genome or transcriptome. Because total number of reads may affect microbial
106 profiling, we performed normalization by sub-sampling to 100,000 reads for each sample. Next,
107 we filtered out low-quality and low-complexity reads using FASTX and SEQCLEAN (see urls).
108 Finally, the remaining reads were realigned to the human references using the Megablast
109 aligner (27) in order to exclude any potentially human reads. The remaining reads were used as
110 candidate microbial reads in subsequent analyses. Figure 1 displays an overview of our pipeline.

111 **Taxonomic profiling**

112 To access the assembly and richness of the microbiomial RNA in blood, we used phylogenetic
113 marker genes to assign the candidate microbial reads to the bacterial and archaeal taxa. We
114 used PhyloSift (v 1.0.1 with default parameters) to perform taxonomic profiling of the whole
115 blood samples (17). PhyloSift makes use of a set of protein coding genes found to be relatively
116 universal (i.e., present in nearly all bacterial and archaeal taxa) and have low variation in copy
117 number between taxa. Homologs of these genes in new sequence data (e.g., the
118 transcriptomes used here) are identified and then placed into a phylogenetic and taxonomic
119 context by comparison to references from sequenced genomes. For our replication study, we
120 used MetaPhlAn for microbial profiling v.1.7.7(28). MetaPhlAn was run in 2 stages; the first
121 stage identifies the candidate microbial reads (i.e., reads hitting a marker) and the second stage

122 profiles meta-genomes in terms of relative abundances. We used MetaPhlAn, rather than
123 PhyloSift, due to differences in library preparation (polyA enrichment versus Ribo-Zero); there
124 were an insufficient number of reads matching the database of the marker genes curated by
125 PhyloSift for adequate microbial profiling of the replication sample.
126

127 **Estimating Microbial diversity**

128 Microbial diversity, or alpha diversity, within each sample was determined using the inverse
129 Simpson index. This index simultaneously assesses both richness (corresponding to the number
130 of distinct taxa) and relative abundance of the microbial communities within each sample (29).
131 In particular, it enables effective differentiation between the microbial communities shaped by
132 the dominant taxa and the communities with many taxa with even abundances (30) (*asbio* R
133 package). To measure sample-to-sample dissimilarities between microbial communities, we use
134 Bray-Curtis beta diversity index, which accounts for both changes in the abundances of the
135 shared taxa and for taxa uniquely present in one of the samples (*vegan* R package). Higher beta
136 diversity indicates higher level of dissimilarity between microbial communities, providing a link
137 between diversity at local scales (alpha diversity) and the diversity corresponding to total
138 microbial richness of the subject group (gamma diversity (31)).
139

140 **Statistical analysis of microbiome diversity**

141 To test for differences in alpha diversity between disease groups, we fit an analysis of
142 covariance (ANCOVA) model using normalized values of alpha, including sex and age, and

technical covariates (RNA INtegrity value (RIN), batch, flow cell lane and RNA concentration) into the model. Bonferroni correction for multiple testing was used. To determine the relative effect size of alpha diversity on schizophrenia status, we fit a logistic regression model including the same covariates and measure reduction in R^2 comparing the full logistic regression model versus a reduced model with alpha removed. Analysis of beta diversity was performed analogously (see Supplementary Methods).

Reference-free microbiome analysis

We complement the reference-based taxonomic analysis with a reference independent analysis. We use EMDeBruijn (https://github.com/dkoslicki/EMDeBruijn), a reference-free approach capable of quantifying differences in microbiome composition between the samples. EMDeBruijn compresses the k-mer counts of two given samples onto de Bruijn graphs and then measures the minimal cost of transforming one of these graphs into the other. To determine overlap between the results from PhyloSift and EMdeBruin, we correlated principal components of EMdeBruin and PhyloSift by Spearman rank correlation, including all samples.

Estimation of cell proportions in whole blood

We assessed DNA methylation data from 65 controls taken from our replication sample, and we compared methylation-derived blood cell proportions estimated using Houseman's estimation method (32, 33) to alpha diversity after adjusting for age, gender, RIN and all technical parameters. We tested whether alpha diversity levels are associated with cell type abundance estimates. More details on the method, quality control pipeline of the methylation data and statistical analysis can be found in Supplementary Methods.

Results

Studying microbial RNA in blood

To study the composition of microbial RNA in blood, we determined the microbial meta-transcriptome present in the blood of unaffected controls (Controls, n=49) and patients with three brain-related disorders: schizophrenia (SCZ, n=48), amyotrophic lateral sclerosis (ALS, n=47) and bipolar disorder (BPD, n=48) (Figure 1, Table 1).

[Figure 1 about here]

[Table 1 about here]

Using our filtering pipeline, an average of 33,546 of 100,000 unmapped reads are identified as high quality, unique non-host reads and were used as candidate microbial reads in our analyses. From these, PhyloSift was able to assign an average of 1,235 reads (1.24% ± 0.41%, mean ± standard deviation) to the bacterial and archaeal gene families. A total of 1,880 taxa were assigned, with 23 taxa at the phylum level (Figure 2). Most of the taxa we observed derived from bacteria (relative genomic abundance 89.8% ± 7.4%), and a smaller portion derived from archaea (relative genomic abundance 12.28% ±6.4%).

[Figure 2 about here]

In total, we observed 23 distinct microbial phyla with on average 4.1 ± 2.0 phyla per individual. The large majority of taxa observed in our sample is not universally present in all individuals; the single exception is Proteobacteria, which dominates all samples with 73.4% ± 18.3% relative abundance (Figure 2 dark green color). Several bacterial phyla show a broad

186 prevalence across individuals and disorders (present in 1/4 of the samples of each subject
187 group). Those phyla include Proteobacteria, Firmicutes and Cyanobacteria, with relative
188 abundance 73.4% ± 18.3%, 14.9 ±10.9%, and 11.0% ± 8.9% (Table S2). This is in line with recent
189 published work on the blood microbiome using 16S targeted metagenomic sequencing
190 reporting relative abundance of 80.4-87.4% and 3.0-6.4% for Proteobacteria and Firmicutes,
191 respectively (23). The other two phyla identified in this study (Actinobacteria and
192 Bacteroidetes) were also detected in our sample in more than 25 individuals. Although
193 Proteobacteria and Firmicutes, are commonly associated with the human microbiome (34),
194 some members of these phyla might be associated with reagent and environmental
195 contaminants (35, 36).

196 To validate our pipeline and investigate the possibility of contamination introduced
197 during RNA isolation, library preparation and sequencing steps, we performed both negative
198 and positive control experiments (see Supplementary Results and Methods for details). In brief:
199 no microbiome sequences were detected in transcriptome data in lymphopblast cell lines
200 (negative control), and we only detected the Chlamydiae phylum in RNA-Seq from cells infected
201 with Chlamydiae (positive control). We examined experimental procedures and technical
202 parameters on microbial composition, and we observed no link between the presence of
203 microbial communities and possible confounders.

204 To compare the inferred microbial composition found in blood with that in other body
205 sites, we used taxonomic composition of 499 meta-genomic samples from Human Microbiome
206 Project (HMP) obtained by MetaPhlAn or five major body habitats (gut, oral, airways, and skin)
207 (10). Of the 23 phyla discovered in our sample, 15 were also found in HMP samples, of which 13

208 are confirmed by at least ten samples. Our data suggest that the predominant phyla detected in
209 blood are most closely related to the known oral and gut microbiome (Table S2). Comparing the
210 microbial composition of whole blood with the microbiome detected in atherosclerotic plaques
211 (37), we observe that the four phyla that together make up for >97% of the microbiome in
212 plaques are also identified in our sample (Firmicutes, Bacteroidetes, Proteobacteria,
213 Actinobacteria).

214 Finally, it should be noted that the sequencing technology does not allow for
215 identification of the origin of microbial RNA. That is, we can not distinguish whether the
216 observed microbial signatures in blood are originate from bacterial communities actually
217 present in the blood, or whether the RNA crossed into the blood stream from elsewhere.

218

219 **Increased microbial diversity in schizophrenia samples**

220 To evaluate potential differences in microbial profiles of individuals with the different disorders
221 (SCZ, BPD, ALS) and unaffected controls, we explored the composition and richness of the
222 microbial communities across the groups.

223 We observed increased alpha diversity in schizophrenia samples compared to all other
224 groups (*ANCOVA* $P < 0.005$ for all groups, Figure 3a, Table 2 and Table S5, *Bonferroni*
225 correction). These differences are corrected for covariates and are independent of potential
226 confounders, such as experimenter and RNA extraction run (Figure S1 and S2), and they are not
227 the consequence of a different number of reads being detected as microbial in schizophrenia
228 samples (see Supplementary Results). No significant differences were observed between the

three remaining groups (BPD, ALS, Controls). In our sample, alpha diversity was found to be a significant predictor of schizophrenia status and explained 5.0% of the variation as measured by reduction in Nagelkerke's R^2 from logistic regression. We observe no correlation between polygenic risk scores (39) and alpha diversity in our schizophrenia sample (n= 32, Kendall's tau= 0.008, P = 0.96, Supplementary Methods). We also did not observe differences in alpha diversity between sexes or across ages, nor are our results driven by the relatively younger schizophrenia cohort (Supplementary Results). Alpha diversity at other main taxonomic ranks yields a similar pattern of increased diversity in schizophrenia (Figure S3).

The increased diversity observed in schizophrenia patients may be due to specific phyla characteristic to schizophrenia, or due to a more general increased microbial diversity in people affected by the disease. To investigate this, we compared diversity across individuals within the schizophrenia group to control samples. We compared beta diversity across pairs of samples with schizophrenia and controls, resulting in three subject groups: SCZ_Controls, SCZ_SCZ and Controls_Controls. The lowest diversity was observed in the Controls_Controls group (0.43 ± 0.21), followed by SCZ_SCZ (0.50 ± 0.14), and the highest beta diversity values were observed for SCZ_Controls (0.51 ± 0.17) ($P< 0.05$ for each comparison, by ANCOVA after correcting for three tests). This result was confirmed by permanova ($P<0.001$) based on 1,000 permutations. Thus, the observed increased alpha diversity in schizophrenia is not caused by a particular microbial profile, but most likely represents a non-specific overall increased microbial burden (see also Figure S4 and Supplementary Results).

In addition to measuring individual microbial diversity (alpha), and diversity between individuals (beta), we measured the total richness of the microbiome by the total number of

distinct taxa of the microbiome community observed within an entire subject group (gamma diversity (40)). We observed that all 23 distinct phyla are observed in schizophrenia: gamma(SCZ)=23 compared to gamma(Controls)=20, gamma(ALS)=16 and gamma(BPD)=18.

We complemented reference-based methods (PhyloSift and MetaPhlAn) with EMDeBruijn, a reference-independent method. EMDeBruin distances measured between samples correlated significantly with beta diversity (Spearman rank $P < 2.2e-16$, rho = 0.37, including SCZ and Controls). Also, EMDeBruijn PCs correlated with principal components obtained from edge PCA based on the PhyloSift taxonomic classification (correlation between EMDeBruijn PC1, and PhyloSift PC1 is $P = 1.824e-09$; Spearman rank correlation is rho = -0.42; see also Figure S5). After correcting covariates, the first three EMDeBruijn PCs are significant predictors of schizophrenia status, and jointly explained 7.1% of the variance ($P< 0.05$ for each PC).

[Figure 3 about here]

[Table 2 about here]

Group differences of individual phyla

In addition to a global difference between schizophrenia and the other groups, we also investigated whether there are particular individual phyla contributing to the differences between schizophrenia and other groups. There are two phyla detected more often in schizophrenia cases versus all the other groups: Plactomycetes, observed in 20 SCZ cases compared to 3(ALS) 2(BPD) 5(Controls) ($P= 0.0002$ Fisher's exact for four groups, Bonferroni

271 corrected for 23 tests P=0.0057) and Thermotogae, observed in 20 SCZ cases compared to 6
272 ALS, 3 BPD and 6 Controls (P= 0.0006 Fisher's exact, corrected P=0.014). No outliers were
273 observed for the other groups (see Table S7).

274

275 **Replication**

276 We performed a replication experiment in an independent case-control sample: schizophrenia
277 (SCZ n=91) and healthy controls (Controls n=88) (See Table S1.D). MetaPhlAn was able to assign
278 5,174 reads (0.089% ± 0.039%, mean ± standard deviation) on average to the bacterial gene
279 families.

280 Schizophrenia samples showed increased alpha diversity on genus level (2.73 ± 0.77 for
281 cases, versus 2.32 ± 0.57 for controls, corrected P = 0.003 Figure 3b) and explained 2.5% of
282 variance as measured by reduction in Nagelkerke R^2, thus replicating our main finding of
283 increased diversity in schizophrenia. While our original analysis was performed on the phylum
284 level, in our discovery sample we observe a similar increase of diversity at the genus level (see
285 Figure S3). Similar to our discovery cohort, we observed no significant correlation between
286 alpha diversity and age or differences across gender. Beta diversity and EMDeBruijn analyses
287 also show similar, though not identical, patterns of nonspecific increased diversity in
288 schizophrenia samples (Supplementary Results).

289

290 **Cell type composition and diversity**

We hypothesized that differences in microbial diversity may be linked to whole blood cell type composition. Our analysis shows that the proportion of one cell type, $CD8^+ CD28^- CD45RA^-$ cells, is significantly negatively correlated with alpha diversity after correction for all other cell-count estimates as estimated from whole blood DNA methylation data (correlation = -0.41, P=7.3e-4, n= 65 Controls from the Replication study, Figure S6, Table S6). These cells are T cells that lack $CD8^+$ naïve cell markers CD28 and CD45RA and are thought to represent a subpopulation of $CD8^+$ memory T cells (41, 42). We observed that low alpha diversity correlates with high levels of cell abundance of this population of T cells.

Discussion

We used high throughput RNA sequencing from whole blood to perform microbiome profiling and identified an increased diversity in schizophrenia patients.

While other studies of human microbiome using RNA-Seq have been conducted (43, 44), this is the first assessing the microbiome from whole blood by using unmapped non-human reads. Despite the fact that transcripts are present at much lower fractions than human reads, we were able to detect microbial transcripts from bacteria and archaea in almost all samples. The microbes found in blood are thought to be originating from the gut as well as oral cavities (45, 46), which is in line with our finding that the microbial profiles found in our study most closely resemble the gut and oral microbiome as profiled by the HMP (10). The taxonomic profile of the cohort samples suggests the prevalence of the several phyla, Proteobacteria, Firmicutes and Cyanobacteria, across individuals and different disorders included in our study.

313 This is in line with a recent study that used 16S targeted meta-genomic sequencing, which
314 reported Proteobacteria and Firmicutes among the most abundant phyla detected in blood
315 (23).

316 Rigorous quality control is critically important for any high-throughput sequencing
317 project, especially in the context of studying the microbiome (35). To this end, we performed
318 both negative and positive quality control experiments, and we carefully evaluated possible
319 contamination effects introduced during the experiments. Our results suggest that the detected
320 phyla represent true microbial communities in whole blood and are not due to contaminants.
321 However, it should be noted that whether only the microbial products crossed into the
322 bloodstream or whether the microbes themselves are present in blood cannot be answered
323 using sequencing techniques. Future experiments, for example, using microscopy, culturing or
324 direct measures of gut permeability, may be able to shed light on this question.

325

326 The most striking finding of our study that relates to diseases affecting the central
327 nervous system is the increased microbial alpha diversity in schizophrenia patients compared to
328 controls and the other two disease groups (ALS, BPD). We replicate this finding in an
329 independent cohort of schizophrenia cases and controls. The replication experiment, while
330 based on different library preparation (Ribo-Zero versus Poly(A)), provides strong evidence for
331 an increased alpha diversity of the microbiome detected in blood in schizophrenia and explains
332 roughly 5% of disease variation. We not only observe an increased individual microbial
333 diversity, but also an increased diversity between individuals (Beta diversity) with schizophrenia
334 compared to controls, rendering it unlikely that a single phylum or microbial profile is causing

335 the disease-specific increase in diversity. Nevertheless, in our study we observed that two phyla
336 in particular, Planctomycetes and Thermotogae, were present in significantly more
337 schizophrenia samples when compared to the other groups. Interestingly, Planctomycetes is
338 group of gram-negative bacteria closely related to Verrucomicrobia and Chlamydiae; together
339 these comprise the Planctomycetes-Verrucomicrobia-Chlamydiae superphylum (47). From
340 peripheral blood, infection with Chlamydiaceae species has been reported to be increased in
341 schizophrenia (40%) compared to controls (7%) (48). Since Chlamydiae is one of the taxa of the
342 superphylum, it is possible that the increase in Planctomycetes we observe is related to the
343 observed increase in Chlamydiaceae species. As the collection of available reference genomes
344 continues to grow and improve, future studies are needed to corroborate and refine these
345 findings.

346 For the study of microbiome diversity, we employed reference-based methods
347 (PhyloSift and MethPhlAn) and the EMDebruin method, a purely reference-agnostic approach.
348 The latter showed strong correspondence to both reference-based methods, highlighting the
349 value of this unbiased sequence-based analysis for investigating microbial differences across
350 groups. However, in addition to differences in distribution of microbial transcripts, EMDebruin
351 may capture variation of other, yet unknown, origin.

352 In addition to our observation that microbial diversity is more generally increased in
353 schizophrenia, our study demonstrates the value of analyzing non-human reads present in the
354 RNA-Seq data to study the microbial composition of a tissue of interest (49, 50). The RNA-Seq
355 approach avoids biases introduced by primers in targeted 16S ribosomal RNA gene profiling. In
356 addition, since *mRNA stability is low in prokaryotes,* RNA-Seq might offer a potential advantage

357 of avoiding contamination of genomic DNA by dead cells compared to genome sequencing (51).
358 Given the many large-scale RNA-Seq datasets that are becoming available, we anticipate that
359 high-throughput meta-transcriptome-based microbiome profiling will find broad applications as
360 a hypothesis-generating tool in studies across different tissues and disease types.

361

362 The increased microbial diversity observed in schizophrenia could be part of the disease
363 etiology (i.e., causing schizophrenia) or may be a secondary effect of disease status. In our
364 sample, we observed no correlation between increased microbial diversity and genetic risk for
365 schizophrenia as measured by polygenic risk scores (52). In addition, it is remarkable that
366 bipolar disorder, which is genetically and clinically correlated to schizophrenia (53), does not
367 show a similar increased diversity. We did observe a strong inverse correlation between
368 increased diversity and estimated cell abundance of a population of T-cells in healthy controls.
369 Even though this finding is based on indirect cell count measures using DNA methylation data
370 (41), the significant correlation highlights a likely close connection between the immune system
371 and the blood microbiome, a relationship that has been documented before (54). More
372 extensive cell count measures and/or better markers of immune sensing of microbial products
373 could be used to study this relationship more directly. In the absence of a direct link with
374 genetic susceptibility and the reported correlation with the immune system, we hypothesize
375 that the observed effect in schizophrenia may be mostly a consequence of disease. This may be
376 affected by lifestyle and/or health status differences of schizophrenia patients, including
377 smoking, treatment plans, (chronic) infection, GI status, the use of probiotics, antibiotics and
378 other drug use or other environmental exposures. Future targeted and/or longitudinal studies

379 with larger sample sizes, detailed clinical phenotypes and more in-depth sequencing are
380 needed to corroborate this hypothesis. Another interesting direction for future work is to study
381 gut permeability in the context of our findings more directly. For example, how does damage to
382 the gut (such as measured using I-FABP) affect observed microbial diversity in blood? These
383 studies would likely result in an expanded understanding of the functional mechanisms
384 underlying the connection between the human immune system, microbiome, and disease
385 etiology. In particular, we hope that these future efforts will provide a useful quantitative and
386 qualitative assessment of the microbiome and its role across the gut-blood barrier in the
387 context of psychiatric disorders.

388

389

390 **Availability of Data and Materials:**

391 The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus
392 (55) and are accessible through GEO Series accession number GSE80974
393 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80974).

394

395 **Acknowledgments:**

396 This work is supported by NIH/NIMH R01 5R01MH090553 (BP samples), 5R01NS058980 (ALS
397 samples) and R01MH078075 (SCZ, Control samples), R21MH098035 (replication sample)
398 awarded to R.A.O. L.M.O.L was financially supported by the National Institute of Neurological
399 Disorders And Stroke of the National Institutes of Health under Award Number T32NS048004.
400 The authors thank Dr. Jonathan Eisen for helpful discussions and insights throughout the course
401 of this project. We thank Dr. Lana Martin for helpful edits to the manuscript. Finally, we thank
402 the study subjects for their participation.

403

404 **Financial Declarations:**

405 The authors have no conflicts of interest to declare.

406

407

References and Notes

1. Clarke G, Grenham S, Scully P, Fitzgerald P, Moloney RD, Shanahan F, et al. The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. Mol Psychiatry. 2013;18(6):666-73.
2. Cryan JF, Dinan TG. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. Nat Rev Neurosci. 2012;13(10):701-12.
3. Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, et al. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell. 2013;155(7):1451-63.
4. Erny D, Hrabe de Angelis AL, Jaitin D, Wieghofer P, Staszewski O, David E, et al. Host microbiota constantly control maturation and function of microglia in the CNS. Nat Neurosci. 2015;18(7):965-77.
5. Foster JA, McVey Neufeld KA. Gut-brain axis: how the microbiome influences anxiety and depression. Trends Neurosci. 2013;36(5):305-12.
6. Castro-Nallar E, Bendall ML, Perez-Losada M, Sabuncyan S, Severance EG, Dickerson FB, et al. Composition, taxonomy and functional diversity of the oropharynx microbiome in individuals with schizophrenia and controls. PeerJ. 2015;3:e1140.
7. Rogers GB, Keating DJ, Young RL, Wong ML, Licinio J, Wesselingh S. From gut dysbiosis to altered brain function and mental illness: mechanisms and pathways. Mol Psychiatry. 2016;21(6):738-48.
8. Zheng P, Zeng B, Zhou C, Liu M, Fang Z, Xu X, et al. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. Mol Psychiatry. 2016;21(6):786-96.
9. Kelly JR, Kennedy PJ, Cryan JF, Dinan TG, Clarke G, Hyland NP. Breaking down the barriers: the gut microbiome, intestinal permeability and stress-related psychiatric disorders. Front Cell Neurosci. 2015;9:392.
10. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-14.
11. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027-31.
12. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2009;457(7228):480-4.
13. Abu-Shanab A, Quigley EM. The role of the gut microbiota in nonalcoholic fatty liver disease. Nature reviews Gastroenterology & hepatology. 2010;7(12):691-701.
14. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nature reviews Genetics. 2012;13(4):260-70.
15. Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc Natl Acad Sci U S A. 2012;109(2):594-9.
16. de Vos WM, de Vos EA. Role of the intestinal microbiome in health and disease: from correlation to causation. Nutrition reviews. 2012;70 Suppl 1:S45-56.
17. Darling AE, Jospin G, Lowe E, Matsen FAt, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ. 2014;2:e243.
18. Drennan MR. What is “Sterile Blood”? British Medical Journal. 1942;2(4269):526-.

19. McLaughlin RW, Vali H, Lau PC, Palfree RG, De Ciccio A, Sirois M, et al. Are there naturally occurring pleomorphic bacteria in the blood of healthy humans? J Clin Microbiol. 2002;40(12):4771-5.
20. Nikkari S, McLaughlin IJ, Bi W, Dodge DE, Relman DA. Does blood of healthy subjects contain bacterial ribosomal DNA? J Clin Microbiol. 2001;39(5):1956-9.
21. Amar J, Serino M, Lange C, Chabo C, Iacovoni J, Mondot S, et al. Involvement of tissue bacteria in the onset of diabetes in humans: evidence for a concept. Diabetologia. 2011;54(12):3055-61.
22. Sato J, Kanazawa A, Ikeda F, Yoshihara T, Goto H, Abe H, et al. Gut dysbiosis and detection of "live gut bacteria" in blood of Japanese patients with type 2 diabetes. Diabetes Care. 2014;37(8):2343-50.
23. Paisse S, Valle C, Servant F, Courtney M, Burcelin R, Amar J, et al. Comprehensive description of blood microbiome from healthy donors assessed by 16S targeted metagenomic sequencing. Transfusion. 2016.
24. Wu S, Yi J, Zhang YG, Zhou J, Sun J. Leaky intestine and impaired microbiome in an amyotrophic lateral sclerosis mouse model. Physiol Rep. 2015;3(4).
25. Dinan TG, Borre YE, Cryan JF. Genomics of schizophrenia: time to consider the gut microbiome? Mol Psychiatry. 2014;19(12):1252-7.
26. Dickerson F, Severance E, Yolken R. The microbiome, immunity, and schizophrenia and bipolar disorder. Brain Behav Immun. 2016.
27. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC bioinformatics. 2009;10:421.
28. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nature methods. 2012;9(8):811-4.
29. Simpson EH. Measurement of diversity. Nature. 1949.
30. Whittaker RH. Evolution and measurement of species diversity. Taxon. 1972:213-51.
31. Koleff P, Gaston KJ, Lennon JJ. Measuring beta diversity for presence–absence data. Journal of Animal Ecology. 2003;72(3):367-82.
32. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13:86.
33. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.
34. Human Microbiome Project C. A framework for human microbiome research. Nature. 2012;486(7402):215-21.
35. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87.
36. Strong MJ, Xu G, Morici L, Splinter Bon-Durant S, Baddoo M, Lin Z, et al. Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples. PLoS Pathog. 2014;10(11):e1004437.
37. Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V, et al. Human oral, gut, and plaque microbiota in patients with atherosclerosis. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4592-8.

38. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648-60.
39. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421--7
.
40. Jost L. Partitioning diversity into independent alpha and beta components. Ecology. 2007;88(10):2427-39.
41. Horvath S, Levine AJ. HIV-1 Infection Accelerates Age According to the Epigenetic Clock. J Infect Dis. 2015.
42. Koch S, Larbi A, Derhovanessian E, Ozcelik D, Naumova E, Pawelec G. Multiparameter flow cytometric analysis of CD4 and CD8 T cell subsets in young and old people. Immun Ageing. 2008;5:6.
43. Croucher NJ, Thomson NR. Studying bacterial transcriptomes using RNA-seq. Curr Opin Microbiol. 2010;13(5):619-24.
44. McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA, et al. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res. 2013;41(14):e140.
45. Potgieter M, Bester J, Kell DB, Pretorius E. The dormant blood microbiome in chronic, inflammatory diseases. FEMS Microbiol Rev. 2015;39(4):567-91.
46. Spadoni I, Zagato E, Bertocchi A, Paolinelli R, Hot E, Di Sabatino A, et al. A gut-vascular barrier controls the systemic dissemination of bacteria. Science. 2015;350(6262):830-4.
47. Hou S, Makarova KS, Saw JH, Senin P, Ly BV, Zhou Z, et al. Complete genome sequence of the extremely acidophilic methanotroph isolate V4, Methylacidiphilum infernorum, a representative of the bacterial phylum Verrucomicrobia. Biol Direct. 2008;3:26.
48. Fellerhoff B, Laumbacher B, Mueller N, Gu S, Wank R. Associations between Chlamydophila infections, schizophrenia and risk of HLA-A10. Mol Psychiatry. 2007;12(3):264-72.
49Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. Metatranscriptomics of the human oral microbiome during health and disease. MBio. 2014;5(2):e01012-14.
50. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RG, Getz G, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. Nat Biotechnol. 2011;29(5):393-6.
51. Ben-Amor K, Heilig H, Smidt H, Vaughan EE, Abee T, de Vos WM. Genetic diversity of viable, injured, and dead fecal bacteria assessed by fluorescence-activated cell sorting and 16S rRNA gene analysis. Applied and environmental microbiology. 2005;71(8):4679-89.
52. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet. 2013;45(10):1150-9.
53. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nature genetics. 2015;47(11):1236-41.
54. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell. 2014;157(1):121-41.
55. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10.

544

545 **Figure Legends**

546

547 **Figure 1.** Microbial profiling using RNA-Seq data from whole blood. (A) We analyzed a cohort of
548 192 individuals from four subject groups, i.e. Schizophrenia (SCZ, n=48), amyotrophic lateral
549 sclerosis (ALS n=47), bipolar disorder (BPD n=48), unaffected control subjects (Controls n=49).
550 (B) Peripheral blood was collected for RNA collection. (B) RNA-Seq libraries were prepared from
551 total RNA using ribo-depletion protocol. Reads that failed to map to the human reference
552 genome and transcriptome were sub-sampled and further filtered to exclude low-quality, low
553 complexity, and remaining potentially human reads. High quality, unique, non-host reads were
554 used to determine the taxonomic composition and diversity of the detected microbiome. See
555 also Table S1.

556

557 **Figure 2.** Relative abundances of microbial taxa at phylum level. Phylogenetic classification is
558 performed using PhyloSift, which is able to assign the filtered candidate microbial reads to the
559 microbial genes from 23 distinct taxa on the phylum level.

560

561 **Figure 3.** Increased diversity of microbiome detected in blood in schizophrenia samples. (A)
562 Alpha diversity per sample for four subject groups (Controls, ALS, BPD, SCZ), measured using
563 the inverse Simpson index on the phylum level of classification. Schizophrenia samples show
564 increased diversity compared to all three other groups (ANCOVA $P < 0.005$ for all groups, after
565 adjustment of covariates, see also Methods, Table S5 and Figure S3). (B) Alpha diversity per
566 sample of schizophrenia cases and controls, measured using the inverse Simpson index on the

567 genus level of classification. Schizophrenia samples show increased within-subject diversity

568 compared to Controls (P = 0.003 after adjustment of covariates).

569