

ARTICLE

Open Access

Transcriptome analysis in whole blood reveals increased microbial diversity in schizophrenia

Loes M. Olde Loohuis¹, Serghei Mangul^{2,11}, Anil P. S. Ori¹, Guillaume Jospin³, David Koslicki⁴, Harry Taegyun Yang², Timothy Wu¹, Marco P. Boks⁵, Catherine Lomen-Hoerth⁶, Martina Wiedau-Pazos⁷, Rita M. Cantor⁸, Willem M. de Vos^{9,10}, René S. Kahn^{5,12}, Eleazar Eskin² and Roel A. Ophoff^{1,5,8}

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 (SCZ) and major depressive disorder^{3–8}, possibly via an impact on intestinal permeability⁹.

High-throughput sequencing offers a powerful culture-independent approach to study the underlying diversity of microbial communities in their natural habitats across

different human tissues¹⁰ and diseases^{3,11–15}. The majority of current microbiome studies use fecal samples and target 16S ribosomal RNA gene sequencing¹⁶. With the availability of comprehensive compendia of reference microbial genomes and phylogenetic marker genes¹⁷, 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 microbes¹⁸. 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

Correspondence: Roel A. Ophoff (ophoff@ucla.edu)

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

²Department of Computer, Science University of California Los Angeles, Los Angeles, CA, USA

Full list of author information is available at the end of the article

These authors contributed equally: Loes M. Olde Loohuis, Serghei Mangul.

© The Author(s) 2018



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

in blood in almost 200 individuals, including patients with SCZ, bipolar disorder (BPD) and sporadic amyotrophic lateral sclerosis (ALS). These three disease groups represent complex polygenic traits that affect the central nervous system with largely unknown etiology. Moreover, roles for the microbiome in all the diseases have been previously hypothesized^{5,24–26}. We used available high quality RNA sequencing (RNA-Seq) reads from whole blood that fail to map to the human genome as candidate microbial reads for microbial classification. We observed an increased diversity of microbial communities in SCZ patients, and we replicated this finding in an independent dataset. Careful analyses, including the use of positive and negative control data sets, suggest that these detected phyla represent true microbial communities in whole blood and are not present in samples due to contaminants. With the increasing number of RNA-Seq data sets, our approach may have great potential for application across different tissues and disease types.

Materials and methods

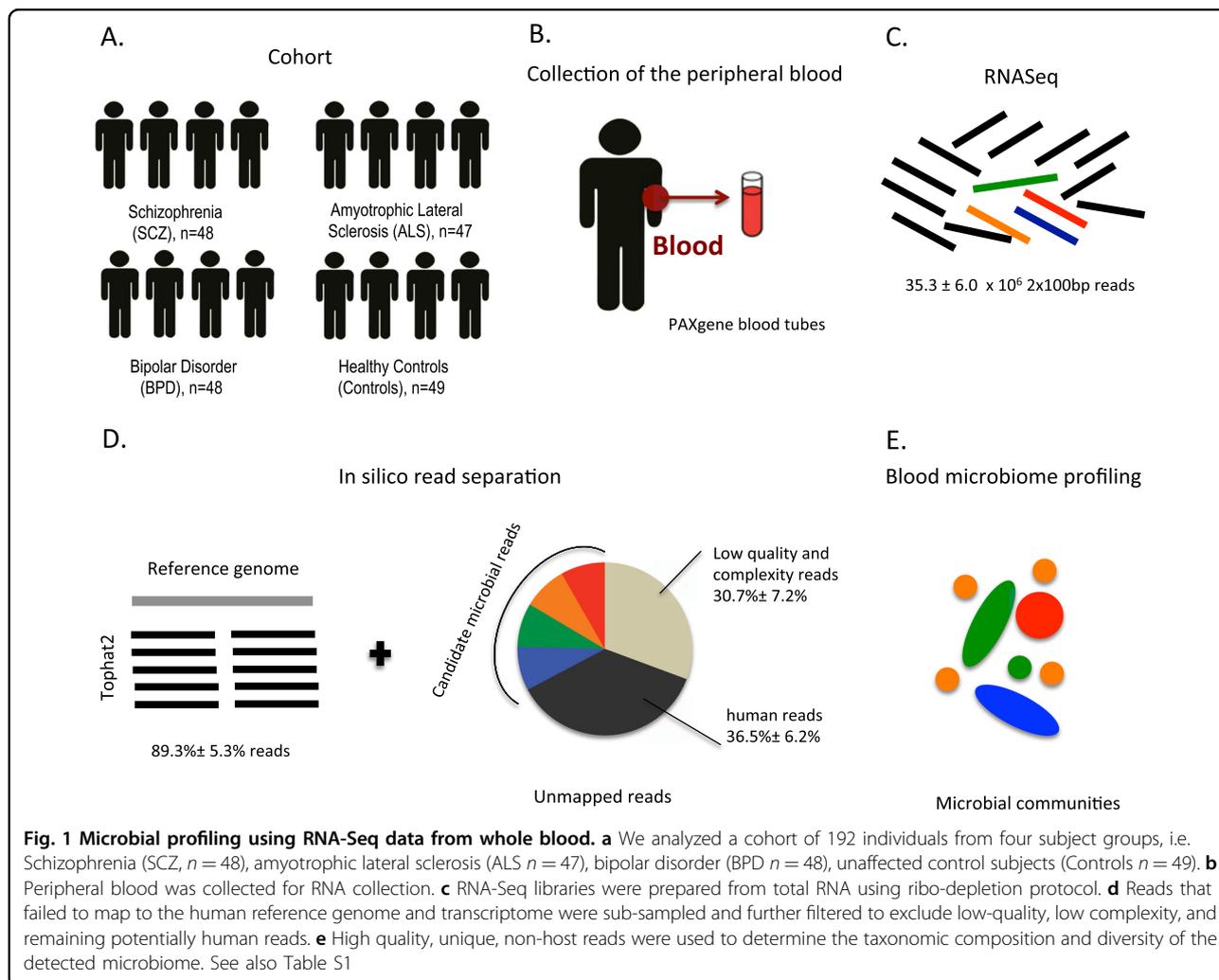
A brief description of Materials and methods follows; see Supplementary Methods for the full details

Sample description

The discovery sample consists of unaffected controls (Controls, $n = 49$) and patients with three brain-related disorders: SCZ ($n = 48$), ALS ($n = 47$), and 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 2×100 bp paired-end reads for the primary study ($35.3\text{M} \pm 6.0$ 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 ($26.3\text{M} \pm 12.0$).

Sequence analysis

We separated human and non-human reads, and use the latter as candidate microbial reads for taxonomic profiling of microbial communities. To identify potentially microbial reads, we developed the following pipeline. First, we filtered read pairs and singleton reads mapped to the human genome or transcriptome. Because total number of reads may affect microbial profiling, we performed normalization by sub-sampling to 100,000 reads for each sample. Next, we filtered out low-quality and low-complexity reads using FASTX and SEQCLEAN (see urls). Finally, the remaining reads were realigned to the human references using the Megablast aligner²⁷ in order to exclude any potentially human reads. The remaining reads were used as candidate microbial reads in subsequent analyses. Fig. 1 displays an overview of our pipeline.

Taxonomic profiling

To access the assembly and richness of the microbial RNA in blood, we used phylogenetic marker genes to assign the candidate microbial reads to the bacterial and archaeal taxa. We used PhyloSift (v 1.0.1 with default parameters) to perform taxonomic profiling of the whole blood samples¹⁷. PhyloSift makes use of a set of protein coding genes found to be relatively universal (i.e., present in nearly all bacterial and archaeal taxa) and have low variation in copy number between taxa. Homologs of these genes in new sequence data (e.g., the transcriptomes used here) are identified and then placed into a phylogenetic and taxonomic context by comparison to references from sequenced genomes. For our replication study, we used MetaPhlAn for microbial profiling v.1.7.7²⁸. MetaPhlAn was run in two stages; the first stage identifies the candidate microbial reads (i.e., reads hitting a marker) and the second stage profiles metagenomes in terms of relative abundances. We used MetaPhlAn, rather than PhyloSift, due to differences in library preparation (polyA enrichment versus Ribo-Zero); there were an insufficient number of reads matching the database of the marker genes curated by PhyloSift for adequate microbial profiling of the replication sample.

Estimating microbial diversity

Microbial diversity, or alpha diversity, within each sample was determined using the inverse Simpson index. This index simultaneously assesses both richness (corresponding to the number of distinct taxa) and relative abundance of the microbial communities within each sample²⁹. In particular, it enables effective differentiation between the microbial communities shaped by the dominant taxa and the communities with many taxa with even abundances³⁰ (*asbio* R package). To measure sample-to-sample dissimilarities between microbial communities, we use Bray-Curtis beta diversity index, which accounts for both changes in the abundances of the shared taxa and for taxa uniquely present in one of the samples (*vegan* R package). Higher beta diversity indicates higher level of dissimilarity between microbial communities, providing a link between diversity at local scales (alpha diversity) and the diversity corresponding to total microbial richness of the subject group (gamma diversity³¹).

Statistical analysis of microbiome diversity

To test for differences in alpha diversity between disease groups, we fit an analysis of 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 SCZ 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 EMDeBruijn, we correlated principal components of EMDeBruijn 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: SCZ ($n = 48$), ALS ($n = 47$), and bipolar disorder (BPD, $n = 48$) (Fig. 1, Table 1).

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 1235 reads ($1.24 \pm 0.41\%$, mean \pm standard deviation) to the bacterial and archaeal gene families. A total of 1880 taxa were assigned, with 23 taxa at the phylum level (Fig. 2). Most of the taxa we observed derived from bacteria (relative genomic abundance $89.8 \pm 7.4\%$), and a smaller portion derived from archaea (relative genomic abundance $12.28 \pm 6.4\%$).

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 $73.4 \pm 18.3\%$ relative abundance (Fig. 2, dark green color). Several bacterial phyla show a broad prevalence across individuals and disorders (present in 1/4 of the samples of each subject group). Those phyla include Proteobacteria, Firmicutes, and Cyanobacteria, with relative abundance $73.4 \pm 18.3\%$, $14.9 \pm 10.9\%$, and $11.0 \pm 8.9\%$ (Table S2). This is in line with recent published work on the blood microbiome using 16S targeted metagenomic sequencing reporting relative abundance of 80.4–87.4 and 3.0–6.4% for Proteobacteria and Firmicutes, respectively²³. The other two phyla identified in this study (Actinobacteria and Bacteroidetes) were also detected in our sample in more than 25

individuals. Although Proteobacteria and Firmicutes are commonly associated with the human microbiome³⁴, some members of these phyla might be associated with reagent and environmental contaminants^{35,36}.

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

To compare the inferred microbial composition found in blood with that in other body sites, we used taxonomic composition of 499 metagenomic samples from Human Microbiome Project (HMP) obtained by MetaPhlan or five major body habitats (gut, oral, airways, and skin)¹⁰. Of the 23 phyla discovered in our sample, 15 were also found in HMP samples, of which 13 are confirmed by at least ten samples. Our data suggest that the predominant phyla detected in blood are most closely related to the known oral and gut microbiome (Table S2). Comparing the microbial composition of whole blood with the microbiome detected in atherosclerotic plaques³⁷, we observe that the four phyla that together make up for >97% of the microbiome in plaques are also identified in our sample (Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria).

Finally, it should be noted that the sequencing technology does not allow for identification of the origin of microbial RNA. That is, we cannot distinguish whether the observed microbial signatures in blood are originated from bacterial communities actually present in the blood, or whether the RNA crossed into the bloodstream from elsewhere.

Increased microbial diversity in SCZ samples

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

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

Table 1 Sample description

Disease status	Control	SCZ	BPD	ALS
<i>N</i>	49	48	48	47
Age mean (SD)	41.1 (10.7)	29.9 (5.8)	46.5 (9.9)	56.4 (10.3)
Age range	[21–60]	[22–46]	[26–71]	[35–76]
Male/Female	38/11	39/9	20/28	29/18

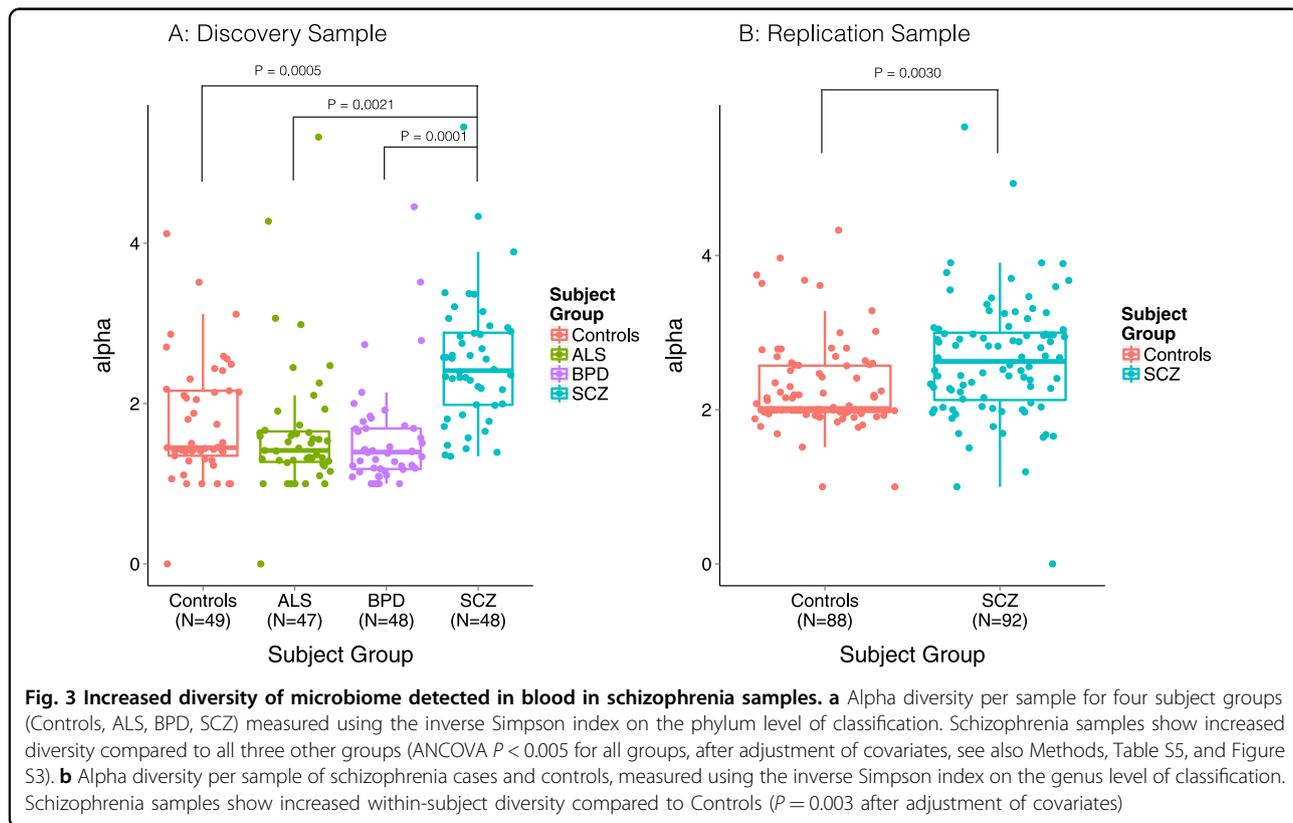
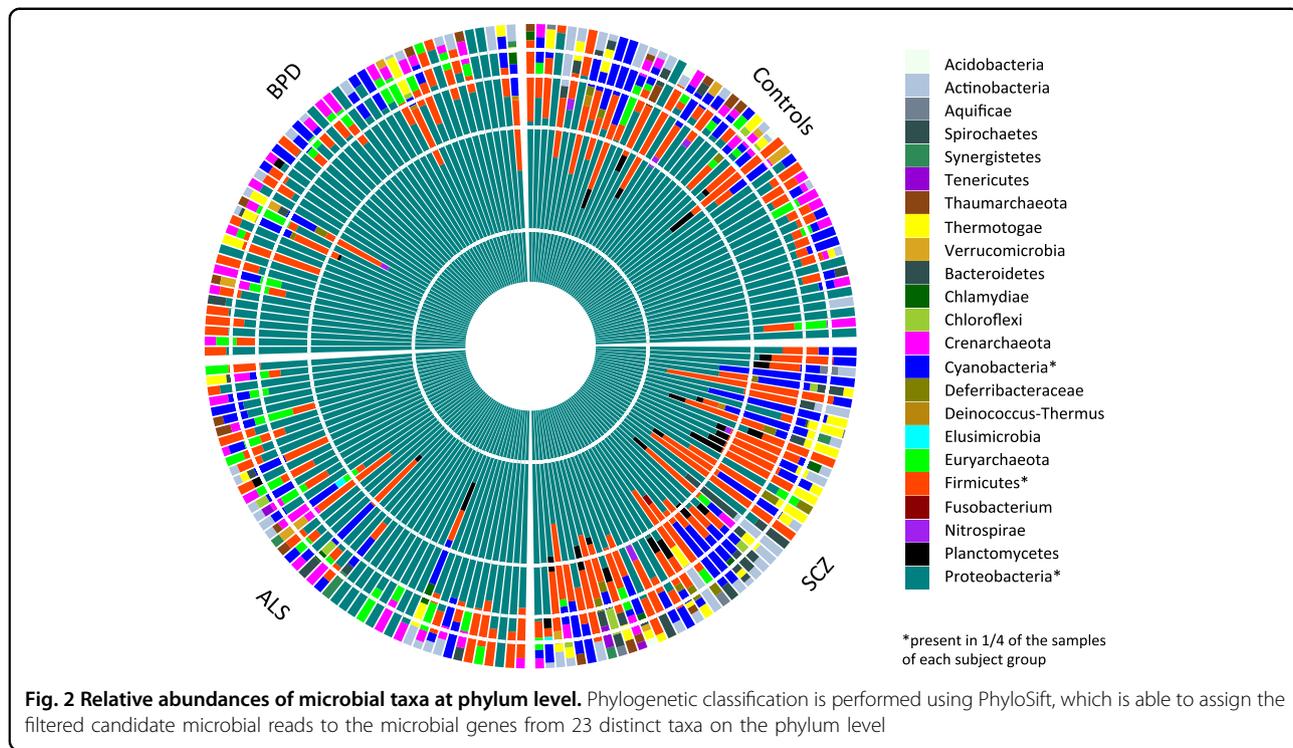


Table 2 Microbial diversity measures

Disease status	Control	SCZ	BPD	ALS
N	49	48	48	47
Alpha diversity, mean (SD)	1.77 (0.74)	2.50 (0.79)	1.55 (0.66)	1.65 (0.86)
Beta diversity, mean (SD)	0.43 (0.21)	0.50 (0.14)	0.31 (0.17)	0.38 (0.22)
Gamma diversity (per group)	20	23	18	16

observed between the three remaining groups (BPD, ALS, Controls). In our sample, alpha diversity was found to be a significant predictor of SCZ 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 scores³⁸ and alpha diversity in our SCZ 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 SCZ cohort (Supplementary Results). Alpha diversity at other main taxonomic ranks yields a similar pattern of increased diversity in SCZ (Figure S3).

The increased diversity observed in SCZ patients may be due to specific phyla characteristic to SCZ, 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 SCZ group to control samples. We compared beta diversity across pairs of samples with SCZ 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 1000 permutations. Thus, the observed increased alpha diversity in SCZ is not caused by a particular microbial profile, but most likely represents a nonspecific 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 diversity³⁹). We observed that all 23 distinct phyla are observed in SCZ: 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. EMDeBruijn 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 SCZ status, and jointly explained 7.1% of the variance ($P < 0.05$ for each PC).

Group differences of individual phyla

In addition to a global difference between SCZ and the other groups, we also investigated whether there are particular individual phyla contributing to the differences between SCZ and other groups. There are two phyla detected more often in SCZ 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 corrected for 23 tests $P = 0.0057$), and Thermotogae, observed in 20 SCZ cases compared to 6 ALS, 3 BPD and 6 Controls ($P = 0.0006$ Fisher's exact, corrected $P = 0.014$). No outliers were observed for the other groups (see Table S7).

Replication

We performed a replication experiment in an independent case-control sample: SCZ ($n = 91$) and healthy controls (Controls $n = 88$) (see Table S1.D). MetaPhlAn was able to assign 5174 reads ($0.089\% \pm 0.039\%$, mean \pm standard deviation) on average to the bacterial gene families.

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

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 sub-population of CD8⁺ memory T cells^{40,41}. 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 SCZ patients.

While other studies of human microbiome using RNA-Seq have been conducted^{42,43}, 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^{44,45}, 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 HMP¹⁰. 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. This is in line with a recent study that used 16S targeted metagenomic sequencing, which reported Proteobacteria and Firmicutes among the most abundant phyla detected in blood²³.

Rigorous quality control is critically important for any high-throughput sequencing project, especially in the context of studying the microbiome³⁵. To this end, we performed both negative and positive quality control experiments, and we carefully evaluated possible contamination effects introduced during the experiments. Our results suggest that the detected phyla represent true microbial communities in whole blood and are not due to contaminants. However, it should be noted that whether only the microbial products crossed into the bloodstream or whether the microbes themselves are present in blood cannot be answered using sequencing techniques. Future experiments, for example, using microscopy, culturing, or direct measures of gut permeability, may be able to shed light on this question.

The most striking finding of our study that relates to diseases affecting the central nervous system is the increased microbial alpha diversity in SCZ patients compared to controls and the other two disease groups (A.L.

S., B.P.D.). We replicate this finding in an independent cohort of SCZ cases and controls. The replication experiment, while based on different library preparation (Ribo-Zero versus Poly(A)), provides strong evidence for an increased alpha diversity of the microbiome detected in blood in SCZ and explains roughly 5% of disease variation. We not only observe an increased individual microbial diversity but also an increased diversity between individuals (Beta diversity) with SCZ compared to controls, rendering it unlikely that a single phylum or microbial profile is causing the disease-specific increase in diversity. Nevertheless, in our study we observed that two phyla in particular, Planctomycetes and Thermotogae, were present in significantly more schizophrenia samples when compared to the other groups. Interestingly, Planctomycetes is group of Gram-negative bacteria closely related to Verrucomicrobia and Chlamydiae; together these comprise the Planctomycetes–Verrucomicrobia–Chlamydiae superphylum⁴⁶. From peripheral blood, infection with Chlamydiaceae species has been reported to be increased in SCZ (40%) compared to controls (7%)⁴⁷. Since Chlamydiae is one of the taxa of the superphylum, it is possible that the increase in Planctomycetes we observe is related to the observed increase in Chlamydiaceae species. As the collection of available reference genomes continues to grow and improve, future studies are needed to corroborate and refine these findings.

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

In addition to our observation that microbial diversity is more generally increased in SCZ, our study demonstrates the value of analyzing non-human reads present in the RNA-Seq data to study the microbial composition of a tissue of interest^{48,49}. The RNA-Seq approach avoids biases introduced by primers in targeted 16S ribosomal RNA gene profiling. In addition, since *mRNA stability is low in prokaryotes*, RNA-Seq might offer a potential advantage of avoiding contamination of genomic DNA by dead cells compared to genome sequencing⁵⁰. Given the many large-scale RNA-Seq data sets that are becoming available, we anticipate that high-throughput metatranscriptome-based microbiome profiling will find broad applications as a hypothesis-generating tool in studies across different tissues and disease types.

The increased microbial diversity observed in SCZ could be part of the disease etiology (i.e., causing SCZ) or

may be a secondary effect of disease status. In our sample, we observed no correlation between increased microbial diversity and genetic risk for SCZ as measured by polygenic risk scores⁵¹. In addition, it is remarkable that bipolar disorder, which is genetically and clinically correlated to SCZ⁵², does not show a similar increased diversity. We did observe a strong inverse correlation between increased diversity and estimated cell abundance of a population of T cells in healthy controls. Even though this finding is based on indirect cell-count measures using DNA methylation data⁴⁰, the significant correlation highlights a likely close connection between the immune system and the blood microbiome, a relationship that has been documented before⁵³. More extensive cell-count measures and/or better markers of immune sensing of microbial products could be used to study this relationship more directly. In the absence of a direct link with genetic susceptibility and the reported correlation with the immune system, we hypothesize that the observed effect in SCZ may be mostly a consequence of disease. This may be affected by lifestyle and/or health status differences of SCZ patients, including smoking, treatment plans, (chronic) infection, GI status, the use of probiotics, antibiotics and other drug use or other environmental exposures. Future targeted and/or longitudinal studies with larger sample sizes, detailed clinical phenotypes, and more in-depth sequencing are needed to corroborate this hypothesis. Another interesting direction for future work is to study gut permeability in the context of our findings more directly. For example, how does damage to the gut (such as measured using I-FABP) affect observed microbial diversity in blood? These studies would likely result in an expanded understanding of the functional mechanisms underlying the connection between the human immune system, microbiome, and disease etiology. In particular, we hope that these future efforts will provide a useful quantitative and qualitative assessment of the microbiome and its role across the gut–blood barrier in the context of psychiatric disorders.

Availability of Data and Materials

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus⁵⁴ and are accessible through GEO Series accession number GSE80974 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80974>).

Acknowledgements

This work is supported by NIH/NIMH R01 5R01MH090553 (BP samples), 5R01NS058980 (ALS samples) and R01MH078075 (SCZ, Control samples), R21MH098035 (replication sample) awarded to R.A.O. L.M.O.L. was financially supported by the National Institute of Neurological Disorders And Stroke of the National Institutes of Health under Award Number T32NS048004. SM acknowledges support from a QCB Collaboratory Postdoctoral Fellowship, and the QCB Collaboratory community directed by Matteo Pellegrini. SM and EE are supported by National Science Foundation grants 0513612, 0731455,

0729049, 0916676, 1065276, 1302448, 1320589 and 1331176, and National Institutes of Health grants K25-HL080079, U01-DA024417, P01-HL30568, P01-HL28481, R01-GM083198, R01-ES021801, R01-MH101782 and R01-ES022282. The authors thank Dr. Jonathan Eisen for helpful discussions and insights throughout the course of this project. We thank Dr. Lana Martin for helpful edits to the manuscript. Finally, we thank the study subjects for their participation.

Author details

¹Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University California Los Angeles, Los Angeles, CA, USA. ²Department of Computer, Science University of California Los Angeles, Los Angeles, CA, USA. ³Davis Genome Center, University of California, Davis, CA, USA. ⁴Mathematics Department, Oregon State University, Corvallis, OR, USA. ⁵Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands. ⁶Department of Neurology, University of California San Francisco, San Francisco, CA, USA. ⁷Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ⁸Department of Human Genetics, University of California Los Angeles, Los Angeles, CA, USA. ⁹Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands. ¹⁰Department of Bacteriology and Immunology, Immunobiology Research Program, University of Helsinki, Helsinki, Finland. ¹¹Institute for Quantitative and Computational Biosciences, University of California Los Angeles, Los Angeles, CA, USA. ¹²Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

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

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41398-018-0107-9>).

Received: 24 January 2018 Accepted: 31 January 2018

Published online: 10 May 2018

References

- Clarke, G. et al. The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. *Mol. Psychiatry* **18**, 666–673 (2013).
- Cryan, J. F. & Dinan, T. G. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat. Rev. Neurosci.* **13**, 701–712 (2012).
- Hsiao, E. Y. et al. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155**, 1451–1463 (2013).
- Erny, D. et al. Host microbiota constantly control maturation and function of microglia in the CNS. *Nat. Neurosci.* **18**, 965–977 (2015).
- Foster, J. A. & McVey Neufeld, K. A. Gut-brain axis: how the microbiome influences anxiety and depression. *Trends Neurosci.* **36**, 305–312 (2013).
- Castro-Nallar, E. et al. Composition, taxonomy and functional diversity of the oropharynx microbiome in individuals with schizophrenia and controls. *PeerJ* **3**, e1140 (2015).
- Rogers, G. B. et al. From gut dysbiosis to altered brain function and mental illness: mechanisms and pathways. *Mol. Psychiatry* **21**, 738–748 (2016).
- Zheng, P. et al. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. *Mol. Psychiatry* **21**, 786–796 (2016).
- Kelly, J. R. et al. Breaking down the barriers: the gut microbiome, intestinal permeability and stress-related psychiatric disorders. *Front. Cell Neurosci.* **9**, 392 (2015).
- Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).

12. Turnbaugh, P. J. et al. A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
13. Abu-Shanab, A. & Quigley, E. M. The role of the gut microbiota in nonalcoholic fatty liver disease. *Nat. Rev. Gastroenterol. Hepatol.* **7**, 691–701 (2010).
14. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **13**, 260–270 (2012).
15. Greenblum, S., Turnbaugh, P. J. & Borenstein, E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc. Natl. Acad. Sci. USA* **109**, 594–599 (2012).
16. de Vos, W. M. & de Vos, E. A. Role of the intestinal microbiome in health and disease: from correlation to causation. *Nutr. Rev.* **70**(Suppl 1), S45–S56 (2012).
17. Darling, A. E., Jospin, G. & Lowe, E. Matsen FAT, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2**, e243 (2014).
18. Drennan, M. R. What is "Sterile Blood". *Br. Med. J.* **2**, 526 (1942).
19. McLaughlin, R. W. et al. Are there naturally occurring pleomorphic bacteria in the blood of healthy humans? *J. Clin. Microbiol.* **40**, 4771–4775 (2002).
20. Nikkari, S., McLaughlin, I. J., Bi, W., Dodge, D. E. & Relman, D. A. Does blood of healthy subjects contain bacterial ribosomal DNA? *J. Clin. Microbiol.* **39**, 1956–1959 (2001).
21. Amar, J. et al. Involvement of tissue bacteria in the onset of diabetes in humans: evidence for a concept. *Diabetologia* **54**, 3055–3061 (2011).
22. Sato, J. et al. Gut dysbiosis and detection of "live gut bacteria" in blood of Japanese patients with type 2 diabetes. *Diabetes Care* **37**, 2343–2350 (2014).
23. Paise, S. et al. Comprehensive description of blood microbiome from healthy donors assessed by 16S targeted metagenomic sequencing. *Transfusion* **56**, 1138–47 (2016).
24. Wu S., Yi J., Zhang Y. G., Zhou J. & Sun J. Leaky intestine and impaired microbiome in an amyotrophic lateral sclerosis mouse model. *Physiol. Rep.* 2015;3. See also: <http://physreports.physiology.org/content/3/4/e12356>.
25. Dinan, T. G., Borre, Y. E. & Cryan, J. F. Genomics of schizophrenia: time to consider the gut microbiome? *Mol. Psychiatry* **19**, 1252–1257 (2014).
26. Dickerson, F., Severance, E. & Yolken, R. The microbiome, immunity, and schizophrenia and bipolar disorder. *Brain Behav. Immun.* **62**, 46–52 (2017).
27. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
28. Segata, N. et al. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814 (2012).
29. Simpson E. H. Measurement of diversity. *Nature* **163**, 688 (1949). See also: <https://www.nature.com/articles/163688a0>
30. Whittaker, R. H. Evolution and measurement of species diversity. *Taxon* **21**, 213–251 (1972).
31. Koleff, P., Gaston, K. J. & Lennon, J. J. Measuring beta diversity for presence-absence data. *J. Anim. Ecol.* **72**, 367–382 (2003).
32. Houseman, E. A. et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).
33. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
34. Human Microbiome Project C. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
35. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
36. Strong, M. J. et al. Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples. *PLoS Pathog.* **10**, e1004437 (2014).
37. Koren, O. et al. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc. Natl. Acad. Sci. USA* **108**(Suppl 1), 4592–4598 (2011).
38. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
39. Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427–2439 (2007).
40. Horvath, S. & Levine, A. J. HIV-1 infection accelerates age according to the epigenetic clock. *J. Infect. Dis.* **212**, 1563–1573 (2015).
41. Koch, S. et al. Multiparameter flow cytometric analysis of CD4 and CD8 T cell subsets in young and old people. *Immun. Ageing* **5**, 6 (2008).
42. Croucher, N. J. & Thomson, N. R. Studying bacterial transcriptomes using RNA-seq. *Curr. Opin. Microbiol.* **13**, 619–624 (2010).
43. McClure, R. et al. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* **41**, e140 (2013).
44. Potgieter, M., Bester, J., Kell, D. B. & Pretorius, E. The dormant blood microbiome in chronic, inflammatory diseases. *FEMS Microbiol. Rev.* **39**, 567–591 (2015).
45. Spadoni, I. et al. A gut-vascular barrier controls the systemic dissemination of bacteria. *Science* **350**, 830–834 (2015).
46. Hou, S. et al. Complete genome sequence of the extremely acidophilic methanotroph isolate V4, *Methylophilum inferorum*, a representative of the bacterial phylum Verrucomicrobia. *Biol. Direct* **3**, 26 (2008).
47. Fellerhoff, B., Laumbacher, B., Mueller, N., Gu, S. & Wank, R. Associations between *Chlamydomonas* infections, schizophrenia and risk of HLA-A10. *Mol. Psychiatry* **12**, 264–272 (2007).
48. Jorth, P. et al. Metatranscriptomics of the human oral microbiome during health and disease. *MBio* **5**, e01012–e01014 (2014).
49. Kotic, A. D. et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat. Biotechnol.* **29**, 393–396 (2011).
50. Ben-Amor, K. et al. Genetic diversity of viable, injured, and dead fecal bacteria assessed by fluorescence-activated cell sorting and 16S rRNA gene analysis. *Appl. Environ. Microbiol.* **71**, 4679–4689 (2005).
51. Ripke, S. et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).
52. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
53. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141 (2014).
54. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).