bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Human cerebrospinal fluids have no microbiome but can contain potential pathogens

Yongyong Kang[1#], Xinchao Ji[2#], Li Guo[3,4,5#], Han Xia[3,6#], Xiaofei Yang[1,4,7], Wenliang Zhang[4], Hong Wei[4], Yuanlin Guan[6], Kai Ye[1,3,4,5*], Gang Zhao[2,8*]

[1]Genome Institute, the First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China
[2]Department of Neurology, Xijing Hospital, Air Force Military Medical University, Xi'an 710032, China
[3]School of Automation Science and Engineering, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[4]MOE Key Laboratory for Intelligent Networks & Network Security, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[5]School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049 China
[6]Hugobiotech Co., Ltd., Beijing 100000, China
[7]School of Computer Science and Technology, Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[8]School of Medicine, Northwest University, Xi'an 710069, China

* Corresponding author: kaiye@xjtu.edu.cn; zhaogang@nwu.edu.cn

# These authors contributed equally to this work

Abstract

Cerebrospinal fluids (CSF) circulating human central nervous system (CNS) have long been considered aseptic in healthy individuals, because normally the blood-brain barrier protects against microbial invasions. However, this dogma has been questioned by several recent reports that colonized microbes were identified in human brains. To investigate whether CSF from healthy individuals without neurological diseases is colonized by a microbiome, we collected and analyzed a cohort of 23 CSF specimens from pregnant women with one-to-one matched contamination controls using metagenomic and metatranscriptomic next-generation sequencing.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Metagenomic data analysis found no significant difference between CSF specimens and negative controls in terms of microbial species diversity. In addition, no active or viable microbiome were present in the CSF samples after being subtracted by microbes in negative controls and DNA extraction buffer. In conclusion, we found no strong evidence that colonized microbiome exist in the cerebrospinal fluids, but may harbor potential pathogens in some healthy individuals.

Introduction

First defined by Joshua Lederberg in 2001[1], human microbiome has since been discovered at almost every part of human bodies such as gut, oral, skin, bladder, vagina, lungs[2-8]. They have profound impact on human health, being associated with a broad range of human diseases including cancers, diabetes, schizophrenia and autoimmunue diseases etc.[9-12]. However, there are a few exceptions such as placenta, amniotic fluids and cerebrospinal fluids (CSF) where no microbiome is found (amniotic fluids and placenta)[13, 14] or known (CSF). CSF circulating the human central nervous system (CNS) has long been considered sterile given that the blood-brain barrier is thought to effectively protect against microbial invasions. However, this traditional knowledge of microbe-free CSF has been challenged in recent years with several reports of microbes detected in human brains and CSF. For example, a bacterial pathogen *Porphyromonas gingivalis* was identified in brains including cerebral cortex, hippocampus as well as CSF of Alzheimer's disease patients[15]. In addition, a community of DNA viruses of CSF was identified from a cohort of mostly healthy human subjects[16]. It remains elusive whether these case reports are evidence of a common microbiome in human CSF and CNS, or simply sporadic and accidental events. Main challenges of studying this issue have been an overall lack of CSF samples from healthy human subjects and the technically sound sampling as well as data analysis methods. Nevertheless, the answer to this question will have great implications to human health especially neurological disorders and infections, providing a guide for disease diagnostics, prevention and therapeutic measures in clinical settings.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Metagenomic sequencing in microbiome studies have traditionally relied on 16s rRNA based approaches[17-19]. Briefly, using hypervariable segment of 16S rRNA, a reliable survey of bacterial community and diversity can be obtained in genus level. However, due to amplified bias, hypervariable regions and primer collected, it is unable to achieve a high-resolution for microbial species identification at species or strain level using 16s rRNA based methods. By contrast, culture-independent, unbiased metagenomic next-generation sequencing (mNGS) represents an alternative and powerful method for rapidly detecting all genetic materials of microbiota at species resolution, an ideal choice for microbiome studies in specimens of low-abundance biomass, such as CSF. Furthermore, mNGS as a promising approach, its clinical diagnostic performance in infectious diseases has been widely adopted in the medical community by multi-center studies[20-22]. Although mNGS comprehensively identifies the presence and diversity of microbiome, it is unable to determine whether the signals come from living microbial cells or simply nucleic acids from dead and broken cells. Therefore, metatranscriptomics is widely adopted to assess the physiological states of microbial communities , providing deeper insights into the actual active microbiota under various environmental conditions[23].

Given the debate over the existence of any microbial community in CSF and the importance of understanding microbial infection in human central nervous systems, we performed microbiome analysis to characterize bacteria, archaea, eukaryota and viruses of CSF from a cohort of 23 pregnant women without neurological disorders, as well as a set of positive and negative controls. Basically, a comprehensive metagenomic and metatranscriptomic study of CSF specimens was conducted to identify potential microbiomes and investigate the potential pathogens, one-to-one matched positive controls (oral and skin) collected from 23 pregnant women along with DNA extraction buffers, as well as matched negative controls (normal saline). To distinguish whether microbes are active members, transcriptomes from twelve CSF samples along with RNA extraction buffers was sequenced using short paired-end sequencing. Data analysis found no significant difference between CSF specimens

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and negative controls in microbial species diversity. In all CSF samples, no active or viable microbiome was present after subtracting microbial taxa detected in CSF by those detected in negative controls and extraction buffers. Taken together, no strong evidence was found in our study supporting that colonized microbiome exists in the cerebrospinal fluids. Such findings shall provide guidance to future research and clinical diagnostics regarding microbial infections inhuman nervous systems.

Results

Metagenomic sequencing of cerebrospinal fluids in healthy pregnant women

Whether a microbiome is present in human CSF remains a disputed issue so far. It's worth studying because it can shed light on the immunology of human central nervous system against neurological diseases, and perhaps more. To investigate whether there is microbiome in CSF, we collected and analyzed microbiome of CSF samples from 23 pregnant women aged 23–40 years who underwent intraspinal anesthesia before the caesarean section via lumbar puncture, coupled with normal saline collected with syringe as negative controls. For each subject, oral and skin microbiomes were also collected and analyzed as positive controls (Figure 1a). All samples were then subjected to DNA extraction and shotgun sequencing for metagenomic analysis. Finally, to validate whether the microbiome, if any detected in CSF, is physiological active, metatranscriptomic for 12 of the 23 CSF samples were sequenced and analyzed. After quality control (QC) with Kneaddata (v0.7.2)[24] for raw sequencing data, MetaPhlAn (v3.0.1)[25], a state-of-the-art taxonomic classification tool based on unique clade-specific marker genes, was used to detect potential microbes in each sample.

In total, we detected 619 nonredundant microbial taxa in 116 samples of eight types from a cohort of 23 pregnant women using metagenomic and metatranscriptomic sequencing and analysis. These microbes detected in all samples are dominated by bacteria (75%) and viruses (24%), whereas a small amount (1%) of eukaryota were also detected, mainly in the skin (82%), swab (11%) and CSF metagenomic

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(CSF_DNA) (7%) samples. Overall, skin, oral and swab samples have the most abundant microbiome of all samples with 393, 199 and 137 nonredundant microbial taxa respectively. This came as no surprise because skins and orals are well known to harbor a plethora of microbes[4, 6]. By contrast, the microbial taxa detected in CSF_DNA (26), negative controls (49) and extraction buffers (27) were relatively fewer (Figure 1b). We then compared the taxa detected in different specimen types, finding that there was very little overlap among all samples. Skin, oral and swab have high number of unique microbial taxa among all sample types, with 243, 129 and 36 taxa found only in these samples, respectively (Figure 1b). Although a large variation in the number of microbes detected was observed for skins, orals and swabs, a much smaller variation was found for CSF_DNA, CSF metatranscriptomic (CSF_RNA), negative controls and extraction buffers (Figure 1b). The oral samples were rich in *Streptococcus*, *Veillonella*, *Neisseria, Rothia* and *Prevotella,* while the skin samples were rich in C*utibacterium*, *Staphylococcu*, *Micrococcus*, *Malassezia,* consistent with many previous studies[4, 6, 26] (additional figure 1). Our successful detection of known microbiota for orals and skins provided a proof-of-concept of NGS-based metagenomic sequencing method, laying a solid foundation for our exploration of CSF microbiome using such a method.

We next focused on examining the microbial species detected for each CSF_DNA specimen. In CSF samples, a total of 76 redundant species inclued 11 (4 nonredundant) bacteria, 61 (21 nonredundant) viruses and 4 (1 nonredundant) eukaryota taxa were detected (Figure 1c). The relative abundance of microbes suggested the species “Cyprinid_herpesvirus_3” are the predominant species in 19 of 23 CSF_DNA samples (Figure 1d). *Cutibacterium_acnes* in species level appeared in 5 specimens. Additionaly, 100%, 26%, 22%, and 22% of all CSF_DNA samples contain Cyprinid herpesvirus 3, Human alphaherpesvirus 2, Enterobacteria phage mEp460 and Dasheen mosaic virus, respectively. However, “Cyprinid herpesvirus 3” detected in all CSF_DNA samples were also found in all negative and skin samples, suggesting a likely external source of these microbes during the CSF sampling procedure.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

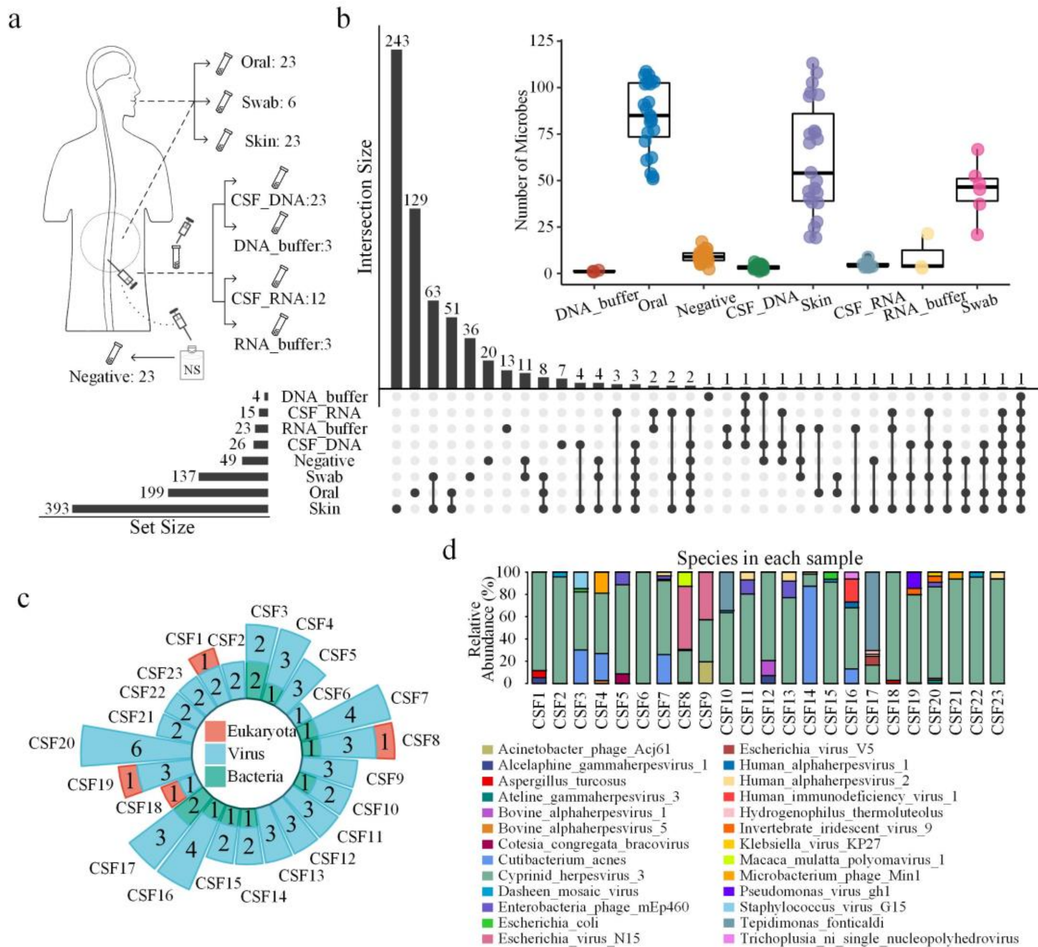

Figure 1: Microbial community structure in human CSF of 23 healthy individuals. (a) Metagenomic experimental design in this study: CSF and matched control samples (positive controls: oral and skin; negative controls: saline solution) collected from 23 pregnant women along with DNA and RNA extraction buffers (number indicates replicates) and were sequenced for metagenomic and metatranscriptomic analysis (see Methods). (b) An overview of microbes detected in each sample type. The number of microbes detected in each sample, and shared species between different samples were shown in the upset plot, with the dots representing intersections among sample types, and the bars representing the number of microbes for each sample type (horizontal bars) and ones shared for each intersection type (vertical bars). The inlet shows a box plot summarizing distributions of the number of species detected for different sample types. (c) Circle barplot summarizing the the number of microbial species in each CSF_DNA sample, categorized into three major types: eukaryota, virus and bacteria. (d) Microbial community structures of 23 CSF_DNA samples shown in a stacked barplot that summarizes the relative abundance of different species of microbes detected for each CSF_DNA sample.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The microbiome signature of cerebrospinal fluid and negative controls is similar

Since microbial species were identified in both CSF_DNA and negative controls, it is likely that microbial cells and/or DNA present in negative controls may have been introduced into CSF during the sampling process. Similarly, the possibility of skin microbiome being introduced into CSF during lumbar puncture could not be ruled out, despite the application of skin surface sterilization. Therefore, we asked how similar in general the microbiome signature is for different sample types by comparing the microbial species detected in these samples. We first performed Non-metric Multidimensional Scaling (NMDS) analysis and principal coordinates analysis (PCoA), and then characterized the beta diversity of CSF and other specimen types using Bray-Curtis distances, a metric commonly used to evaluate microbiome difference among samples supported by Wilcox statistical significance. NMDS, PCoA (Additional Figure 2), and beta-diversity analysis revealed an overall clear seperation of microbial communities for each sample type, except that microbiome in CSF_DNA specimens overlapped partially with negative controls (Figure 2a). Statistical analysis suggested beta-diversity between CSF_DNA and other sample types is significantly different from CSF_DNA self-comparision. However, there is no significant difference between CSF_DNA self beta-diversity and CSF_DNA -negative control beta diversity (Wilcox test: $p = 0.59$) (Figure 2b), suggesting the microbial communities in CSF_DNA and negative controls have a high similarity. In fact, shared microbial taxa between CSF_DNA and negative control accounted for 42% and 22% of CSF_DNA and negative control, respectively. By contrast, 58% microbial taxa in CSF-DNA were detected in skin samples, whereas only 4% of skin microbes were found in CSF-DNA specimens. On one hand, these results indicated the microbial cells or DNA detected in CSF samples may partly have come from negative controls during sample collection. On the other hand, the high beta-diversity between skin and CSF specimens implied that the skin surface sterilization before lumbar punctures effectively prevented the contamination of CSF samples with most, if not all, skin microbes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

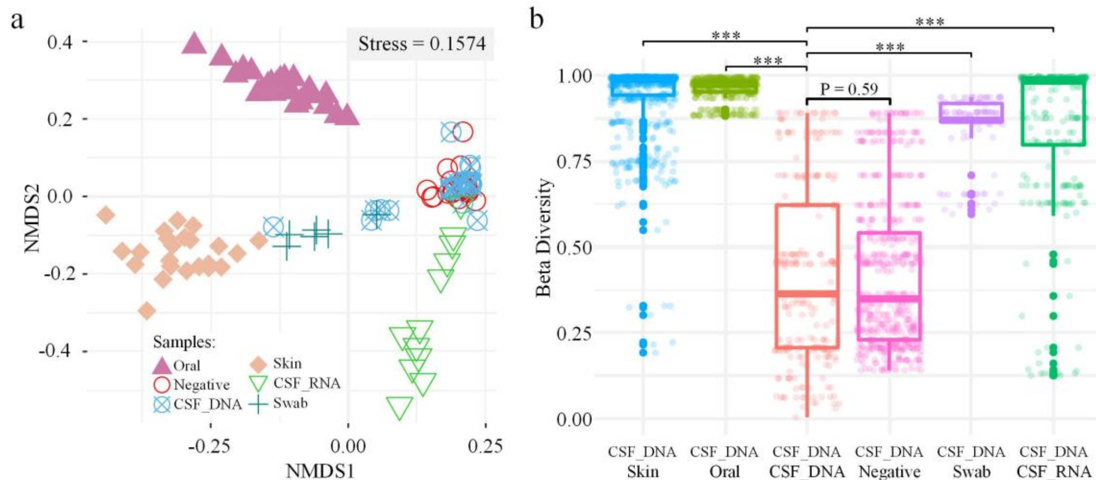

Figure 2. Microbiome similarity among sample types. (a) NMDS (Non-metric Multidimensional Scaling) analysis of microbial species detected from different sample types. Shapes and colors represent sample types. (b) Boxplot summarizing the beta diversity within CSF_DNA and between CSF_DNA and other specimens using Bray-Cruits dissimilarity. Statistical significance was assessed by Wilcoxon test whose significance level is indicated with asterisks (***: P<0.001).

No microbiome is present in the cerebrospinal fluids substrated by contamination controls

With the detected microbiome in CSF samples, we questioned whether these microbes were truly CSF inhabitants or simply brought in from external sources such as skins, negative controls and DNA extraction buffer. To verify whether the CSF contains *de facto* colonized microbial communities, we substracted the microbes collectively detected in negative control samples and DNA extraction buffer samples from microbes of each CSF_DNA sample, a method commonly used and previously described by human microbiome study [14]. After substraction, 12 CSF samples contained no microbe, whereas the other 11 CSF samples contain a total of 14 microbial taxa including 11 viruses, 2 bacteria and 1 eukaryota. Since an introduction of microbes from skins could not be completely ruled out, we further checked whether the 14 taxa were present in skins as well and found that 6 of the 14 taxa were also found in the matching skin microbiome. This left eight microbial taxa after

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

substraction as potential CSF inhabitating microbes, including five viruses ("Bovine alphaherpesvirus 1", "Escherichia virus V5", "Klebsiella virus KP27", "Macaca mulatta polyomavirus 1", "Trichoplusia_ni_single_nucleopolyhedrovirus"), two bacteria (*Hydrogenophilus thermoluteolus*, *Tepidimonas fonticaldi*), and one eukaryota (*Aspergillus turcosus*).

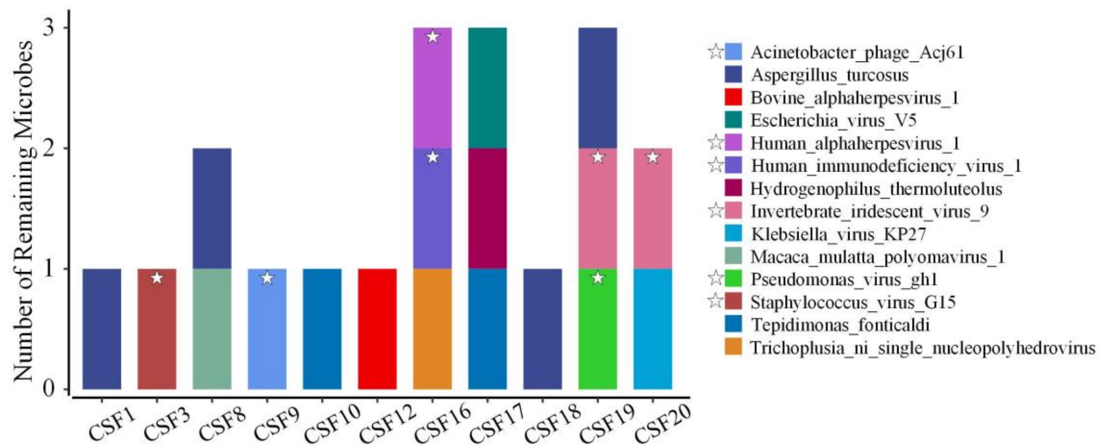

Figure 3. Microbes remained in the cerebrospinal fluids. Subtracting the microbes appeared in the negative control and DNA extraction buffer, 14 species (6 species labeled with star appeared in skin samples) remined in CSF genomic samples.

The detection of microbes using the metagenomic approach offers a glimpse of microorganisms present in certain niches. However, it remains uncertain whether these microbes are live or dead, as DNA from dead cells are also detectable by mNGS. Therefore, we further evaluated the physiological activities of the potential CSF-inhabitating microbes using metatranscriptomic sequencing, because the number of microbes detected by both in metagenomic and metatranscriptome would indicate active microbes may be present in CSF samples. CSF transcriptomics revealed transcripts for several microbial taxa including "Equine infectious anemia virus" and "Cyprinid herpesvirus 3" appearing in all samples, and *Escherichia_coli* and Dasheen mosaic virus appearing in eleven and nine samples, respectively. We then asked, for the eleven CSF-DNA samples with microbes left after substraction by negative controls and DNA extraction buffers, whether these microbes have detected in metatranscriptomic data. The result showed that only "human alphaherpesvirus 1" had

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

signals from both CSF genomics and transcriptomics. However, "human alphaherpesvirus 1" also appeared in the skin, suggesting no active microbiome was detectable in CSF after removing this species potentially originated from skins. Although metagenomic analysis detected the one *Aspergillus turcosus* species from four individuals (Figure 3), no transcripts of *Aspergillus turcosus* were detected in metatranscriptomic, suggesting a lack of living cell activity. *Aspergillus turcosus* is well known as opportunistic human pathogens and can cause infections in individuals of compromised immune systems. How this fungal species (cells or DNA) reach the CSF of the four healthy individuals is unknown, but it shows CSF, though without an active microbiome, might not be entirely free of these opportunistic fungi which could potentially cause infections in central nervous systems when factors such as the strength of host immune systems do alter. Taken together, our study found no strong evidence supporting a colonized microbiome in the cerebrospinal fluids which, however, may occasionally contain opportunistic pathogens.

Discussion

Hereby, a cohort of 23 CSF samples of the pregnant woman without neurological disease with a matched set of controls were collected for microbiome detection using culture-independent approach by a direct shotgun sequencing. The metagenomic data analysis indicated that there was no significant difference between CSF specimens and negative controls in beta diversity of detected microbes. In addition, no clear signal of active microbiome in the CSF samples was found compared to contamination controls. Except *Aspergillus turcosus* appeared in four samples, no microbiome was present in more than two CSF samples after being subtracted by microbes in negative controls and DNA extraction buffer. Collectively, our studies showed that although CSF is free of an active microbiome, the sporadic presence of the remaining microbes after sample substraction indicated potential opportunistic pathogens might be found occasionally in healthy human populations.

In this study, the specimen DNA was only identified by metagenomic analysis, not

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

combined with 16S rRNA gene sequencing, which is based on distinguish the hypervariable regions of the 16S rRNA gene of bacteria and archaea. Our focus was only on the issue of CSF microbiome in populations without neurological disorders, whether CSF contains a microbiome in disease states such as Alzheimer's disease, multiple sclerosis, parkinson's disease ect, were not investigated. Our data can only illustrate that the microbiome of CSF was indistinguishable from contamination controls. We could not rule out the existence of CSF microbiome limited by the sensitivity of current research approaches.

In conclusion, we analyzed 23 CSF specimens and an extensive set of control samples using metagenomic analysis combined with metatranscriptomic deep sequencing in the present study. Although both of these methods have detected distinctive microbial signals in all samples, the detected microbe in CSF samples was indistinguishable from contamination controls. Our data indicated that there was no evidence to support the existence of a CSF microbiome in the populations without neurological disorders by current approaches.

Methods

Subjects

Twenty-three donors were recruited from the Xi Jing Hospital of the Air Force Military Medical University. All subjects were enrolled from obstetrics department in which the pregnant woman aged 23–40 years need intraspinal anesthesia before the caesarean section. Subjects who have suffered from central nervous system infection disease (eg, meningitis, encephalitis) or any systemic infection disease and autoimmune disease (eg, hepatitis, tuberculosis, systemic lupus erythematosus, rheumatism) and have received antibiotics treatment in the past six months prior to sample collection were excluded. We also excluded subjects with a history of hypertension, diabetes, heart disease, cancer and neurological disease (eg, Alzheimer’s disease, Parkinson’s disease, multiple sclerosis, epilepsy).

Sample Collection

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Lumbar puncture was performed in the 23 subjects enrolled in this study and the CSF samples were collected in a 4ml centrifugal tube with syringe and then stored in a −80 °C freezer for metagenomics analysis. Twelve CSF samples were selected from 23 subjects at random for metatranscriptome studies and RNA protection reagent was added to the CSF immediately after collection. Then, the samples were centrifuged and the pellets stored at −80 °C for metatranscriptomic sequencing. Meanwhile, normal saline was collected with syringe for environmental controls (negative control). Furthermore, oral and skin samples were selected from 23 enrolled subjects as one-to-one matched positive control. For skin positive controls: The back skin of 5×5 cm^2 areas around the puncture site (L3-L4 intervertebral space) were swabbed with a sterile cotton swab before skin clean with povidone iodine. To maximize microbial load, no bathing was permitted within 24 hours of sample collection. For oral positive controls, all subjects were forbidden to eat and drink six hours before operation. The surfaces of tongue, buccal fold, hard palate, soft palate, tooth, gingiva and saliva were swabbed with sterile swab. Unused sterile swabs were collected for negative controls ("sterile swab"). Details of Matching information between samples are described in Additional File 1.

DNA Extraction and Purification

DNA was isolated using the QIAamp DNA Mini Kit (Qiagen 51304) according to the manufacturer's instructions. 1) DNA extraction from swabs: Swab tips were cut and placed in a 2 ml microcentrifuge tube and then 400 μl PBS were added. Next, 20 μl of proteinase K and 400 μl of buffer AL were added, vortexed for 10 s, and the solution was incubated for 15 min at 56 °C. And then added 400 μl ethanol (100%) and mixed again by vortexing. The final, DNA purification was performed with buffer AW1 and AW2 using QIAamp Mini spin column, followed by elution with 35 μl of buffer EB. 2) DNA extraction from CSF and normal saline controls: 200 μl sample was added into the microcentrifuge tube, and then added 20 μl of proteinase K and 200 μl of buffer AL respectively, vortexed for 10 s, and the solution was incubated for 15 min at 56 °C. Next, added 200 μl ethanol (100%) and mixed again by vortexing. The final,DNA purification was performed as described above. As a reagent control, ultrapure water

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

was used.

Metagenomics Library Construction

For preparation of metagenomics libraries, the QIAseq FX DNA Library Kit (Qiagen; 180715) was used. The construction involved five steps: 1) Fragmentation and End-repair: to generate 200–300 bp fragments, 32.5µl purified DNA were fragmented by incubation with FX buffer 5µl, FX enhancer 2.5µl and 10 µl FX enzyme mix at cycling program: 4 ºC 1min→32 ºC 12min→65 ºC 30min→4 ºC hold. 2) Adapter ligation: 5 µl of adaptor, 20 µl of ligation buffer, 10 µl of DNA ligase and 15 µl of nuclease-free water were added and incubated for 15 min at 20 °C to initiate adapter ligation. Adapter ligation cleanup was performed immediately, 3) Adapter ligation cleanup: 80 µl of resuspended AMPure® XP beads (0.8×) were added to each ligated sample and mix well by pipetting. Next, the mixture was incubated for 5 min at room temperature and then the beads were pelleted on a magnetic stand (Invitrogen) for 2 min. The supernatant was discarded and the pallet was washed twice with 200 µl of 80% ethanol, then the beads were eluted with 52.5 µl of buffer EB. Subsequently, 50 µl of supernatant was transfered into a new 1.5 ml microcentrifuge tube and a second purification was performed with 50 µl (1×) AMPure® XP beads. The final, 23.5 µl of purified DNA sample was obtained. 4) Amplification of library DNA: 25 µl of HiFi PCR Master Mix, 1.5 µl of Primer Mix and 23.5 µl of library DNA were added in PCR tube. PCR enrichment was performed under the cycle conditions: 2 min at 94 ºC, 12 × (20 s at 98 ºC, 30 s at 60 ºC, 30 s at 72 ºC), and 1 min at 72 ºC. The final, to obtain libraries, the PCR products were purified with AMPure XP beads as described above.

RNA Extraction and Purification

Total RNA was extracted using the RNeasy Mini Kit (Qiagen; 74104) according to the manufacturer's instructions. The pellet of each sample which has been treated with RNA protection reagent as described above, was resuspended in 100µl TE buffer containing lysozyme, and Proteinase K was added into the mixture, then incubated for 10 min at room temperature. 350µl of buffer RLT was added and vortexed vigorously. The final, RNA isolation and purification was performed with buffer AW1 and RPE

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

respectively using RNeasy Mini spin column, followed by elution with RNase-free water.

RNA Library Preparation for Metatranscriptomics Sequencing

For construction of RNA libraries, the QIAseq FX Single Cell RNA Library Kit (Qiagen; 180733) was used. The construction involved five steps: 1) Genomic DNA (gDNA) removal: 8 µl of purified RNA and 3 µl of NA denaturation buffer were added into a sterile PCR tube and incubated for 3 min at 95 °C. To remove genomic DNA, 2 µl of gDNA wipeout buffer was added and incubated for 10 min at 42 °C. 2) Reverse transcription: 4 µl of RT/Polymerase buffer, 1 µl of random primer, 1 µl of Oligo dT primer and 1 µl of Quantiscript RT enzyme mix were added in each sample and reverse transcription was carried out for 60 minutes at 42 °C. 3) Ligation: 8 µl of ligase buffer and 2 µl of ligase mix were added into the RT reaction and incubated at 24ºC for 30 min. 4) Whole transcriptome amplification: 1 µl of REPLI-g SensiPhi DNA Polymerase and 29µl of reaction buffer were used for Multiple Displacement Amplification (MDA), then incubate at 30ºC for 2 h. The final, an approximate length of 2000–70,000 bp amplified cDNA was produced. 5) Enzymatic Fragmentation: The amplified cDNA was diluted 1:3 in H_2O sc, 10 µl of the diluted DNA and FX Enzyme Mix were used to obtain 300 bp library fragment with reaction conditions: 4° C 1min →32° C 15min→65° C 30min→4° C hold. 6) Adapter ligation: 5 µ l of adapter and 45 µ l of ligation master mix were added into each sample and incubated at 20° C for 15 min. Subsequently, the adapter ligation cleanup was performed with AMPure XP beads as described above. The final, purified libraries were obtained ready for sequencing without further PCR amplification.

Sequencing

Metagenomic shotgun sequencing was performed on Illumina Hiseq platform for all samples (paired end library of 150-bp and 150-bp read length). Adaptor and low quality reads were discarded from the raw reads, and the remaining reads were filtered in order to eliminate host DNA based on the human reference genome as described below.

Data quality control

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To reduce the impact of host reads on the results, we need to remove human reads involved in the raw sequencing data before bioinformatics analysis. KneadData [24], a widely used tool, is designed to perform quality control on metagenomic sequencing data, especially for microbiome experiments. All reads were filtered using KneadData with the following trimmomatic options: ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10:8:true, SLIDINGWINDOW:4:20, MINLEN:50 and bowtie2 options: --very-sensitive, --dovetail. Approximately, 25 Gb and 5 Gb of raw paired-end reads were obtained per sample in the CSF genomics samples and negative samples, respectively. All samples data size and size after quality control can be found in Additional Figure 3a and 3b. The proportion of human reads in CSF genomics samples is up to 92%. The approximate proportion of human source data for each sample and the per base sequences quality obtained by fastQC (v.0.11.8) [27] for reads after quality control can be found in Additional Figure 3c and 3d, respectively.

Detect potential microbiome

MetaPhlAn (version 3.0.1) [25] is a computational tool for profiling the composition of microbial communities (bacteria, archaea, viruses and eukaryotes) from metagenomic shotgun sequencing data. Based on ~1.1M unique clade-specific marker genes identified from ~100,000 reference genomes, MetaPhlAn can profile unambiguous taxonomic assignments and accurate estimation of organismal relative abundance in species-level resolution. Classifying the reads to marker genes database, MetaPhlAn outputs a file containing detected microbes and relative abundance in different level. MetaPhlAn ran with custom parameters: --add_viruses --input_type fastq --read_min_len 50.

β-diversity and phylogenetic analysis

Using R (version 3.6.3) with R studio environment, β-diversity (between-sample diversity) was estimated by Bray-Curtis dissimilarity in vegan package. All pictures are ploted using R.

Additional files

Additional file 1: Figure S1. Top 10 genus in oral and skin samples, respectively. Microbial

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

community structures of 23 Oral (figure S1a) and Skin (figure S1b) samples shown in a stacked barplot that summarizes the relative abundance of different genus detected.

Additional file2: Figure S2. Principal Coordinates Analysis (PCoA) analysis of microbial species detected from different sample types.

Additional file3: Figure S3. Data information. (a) Approximately raw data size for different samples. (b) Data size for different samples after quality control. (c) The proportion of Human Data in raw data. (d) Per base sequences quality after data quality control for different sample types.

Additional file4: table S1. Samples label and matching information.

Abbreviations

CSF: Cerebrospinal fluids; CNS: central nervous system; mNGS: metagenomic next-generation sequencing; QC: quality control; CSF_DNA: CSF metagenomic; CSF_RNA: CSF metatranscriptomic; NMDS: Non-metric Multidimensional Scaling; PCoA: principal coordinates analysis; MDA: Multiple Displacement Amplification.

Funding

This study was supported by the National Key R&D Program of China (Grant Nos. 2018YFC0910400, 2017YFC0907500, and 2016 YFC0904501), the National Natural Science Foundation of China (Grant Nos. 31671372, 61702406, 31701739, and 31970317), the National Science and Technology Major Project of China (Grant No. 2018ZX10302205), as well as the General Financial Grant from the China Postdoctoral Science Foundation (Grant Nos. 2017M623178 and 2017M623188).

Contributions

KY and GZ supervised this study. XCJ, KY and GZ designed experiment and collected the specimens. YYK, WLZ, HW, XFY and YLG analyzed the data. YYK, XCJ, LG, HX, KY wrote and revised the manuscript. All authors read and approved the final manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

We thank Peng Jia, Tingjie Wang, Ningxin Dang, and Tun Xu for helpful discussions regarding data analysis and Jing Hai for administrative and technical support.

Ethics Statement

This study was approved by the Ethics Committee of the Xi Jing Hospital of the Air Force Military Medical University. All procedures were conducted in accordance with the approved guidelines. All patients read and signed the consent form before sample collection.

Competing interests

The authors declare that they have no competing interests.

References

1. Lederberg J, Mccray A: The Scientist:'Ome Sweet'Omics--A Genealogical Treasury of Words. *Scientist* 2001, 15:22-27.
2. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R: Current understanding of the human microbiome. *Nature Medicine* 2018, 24:392-400.
3. Jose C., Clemente, and, Luke K., Ursell, and, Laura Wegener, Parfrey, and, Rob: The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell* 2012.
4. Nasidze I, Li J, Quinque D, Tang K, Stoneking M: Global diversity in the human salivary microbiome. *Genome Research* 2009, 19:636.
5. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al: Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences* 2011, 108:4680-4687.
6. Grice EA, Segre JA: The skin microbiome. *Nature Reviews Microbiology* 2011, 9:244.
7. Beck JM, Young VB, Huffnagle GB: The microbiome of the lung. *Translational research : the journal of laboratory and clinical medicine* 2012, 160:258-266.
8. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, et al: Structure, function and diversity of the healthy human microbiome. *Nature* 2012, 486:207-214.
9. Iida N, Dzutsev A, Stewart CA, Smith L, Bouladoux N, Weingarten RA, Molina DA, Salcedo R, Back T, Cramer S: Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* 2013, 342:967-970.
10. Moreno-Indias I, Cardona F, Tinahones FJ, Queipo-Ortu?±o M-aI: Impact of the gut microbiota on the development of obesity and type 2 diabetes mellitus. *Frontiers in Microbiology* 2014, 5:190.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.16.299065; this version posted September 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

11. Zhu F, Ju Y, Wang W, Wang Q, Ma X: Metagenome-wide association of gut microbiome features for schizophrenia. *Nature Communications* 2020, 11:1612.
12. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB: Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome biology* 2012, 13:R79.
13. Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, Leite R, Elovitz MA, Parry S, Bushman FD: Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* 2016, 4:29.
14. Lim ES, Cynthia R, Holtz LR: Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. *Microbiome* 2018, 6:87.
15. Dominy SS, Lynch C, Ermini F, Benedyk M, Potempa J: Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors. *ence Advances* 2019, 5:eaau3333.
16. Atab K: The Virome of Cerebrospinal Fluid: Viruses Where We Once Thought There Were None. *Frontiers in Microbiology* 2019, 10:2061-.
17. Richa B, Grimm DG: Current challenges and best-practice protocols for microbiome analysis. *Briefings in Bioinformatics* 2019.
18. Hamady M, Knight R: Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome research* 2009, 19:1141-1152.
19. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N: Shotgun metagenomics, from sampling to analysis. *Nature biotechnology* 2017, 35:833-844.
20. Ramachandran PS, Wilson MR: Metagenomics for neurological infections—expanding our imagination. *Nature Reviews Neurology* 2020:1-10.
21. Chiu CY, Miller SA: Clinical metagenomics. *Nature Reviews Genetics* 2019.
22. Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, Federman S, Stryke D, Briggs B, Langelier C, et al: Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis. *New England Journal of Medicine* 2019, 380:2327-2340.
23. Bikel S, Valdez-Lara A, Cornejo-Granados F, Rico K, Canizales-Quinteros S, Soberón X, Del Pozo-Yauner L, Ochoa-Leyva A: Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Computational and Structural Biotechnology Journal* 2015, 13:390-401.
24. *KneadData software: http://huttenhower.sph.harvard.edu/kneaddata.*
25. 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, 9:811-814.
26. Chen YE, Fischbach MA, Belkaid Y: Skin microbiota-host interactions. *Nature* 2018, 553:427-436.
27. *FastQC software: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.*