Cerebrospinal fluids from healthy pregnant women does not harbor a detectable microbial community

Yongyong Kang[1,3,11#], Xinchao Ji[2#], Li Guo[3,4,5#], Han Xia[3,6#], Xiaofei Yang[1,4,7], Zhen Xie[8], Xiaodan Shi[9], Rui Wu[8], Dongyun Feng[9], Chen Wang[10], Min Chen[10], Wenliang Zhang[4], Hong Wei[4], Yuanlin Guan[6], Kai Ye[1,3,4,5,11*], Gang Zhao[8,9*]

[1]Genome Institute, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

[2]Department of Neurology, Xi'an No. 3 Hospital, The Affiliated Hospital of Northwest University, Xi'an 710061, 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

[9]Department of Neurology, Xijing Hospital, The Fourth Military Medical University, Xi'an 710032, China

[10]Department of Anesthesiology, Xijing Hospital, The Fourth Military Medical University, Xi'an, 710032, China

[11]Center for Mathematical Medical, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

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

# These authors contributed equally to this work

ABSTRACT Cerebrospinal fluids circulating human central nervous system 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 reports that microbes were identified in human brains, raising the question whether a microbial community is present in cerebrospinal fluids of healthy individuals without neurological diseases. Here, we collected and analyzed metagenomic and metatranscriptomic sequencing data of cerebrospinal fluid specimens from a cohort of 23 pregnant women aged between 23 and 40 and one-to-one matched contamination controls. From data analysis of 116 specimens of eight different types, we detected 619 nonredundant microbial taxa which were dominated by bacteria (75%) and viruses (24%). In cerebrospinal fluids metagenomic samples, a total of 76 redundant species were detected including four (one nonredundant) eukaryota taxa, eleven (four nonredundant) bacteria, and 61 (21 nonredundant) viruses that were mostly bacteriophages. Metagenomic data analysis found no significant difference between cerebrospinal fluid specimens and negative controls in terms of microbial species diversity. In addition, no active or viable microbiome were present in the cerebrospinal fluid samples after subtracting microbes detected in contamination controls. In conclusion, we found no strong evidence that colonized microbial community exist in the cerebrospinal fluids of healthy individuals.

IMPORTANCE Microbiome are prevalent throughout human bodies with profound health implications. However, it remains unclear whether a microbiome is present and active in human cerebrospinal fluids that are long considered aseptic given the blood-

brain barrier. Here, we applied unbiased metagenomic and metatranscriptomic sequencing to detect microbiome in cerebrospinal fluids collected from a cohort of 23 pregnant women with matched controls. By analyzing 116 specimens of eight types, no strong evidence was found to support a presence of colonized microbiome in the cerebrospinal fluids. Our findings have profound implications to human immunity against neurological infections and disorders, providing a guide for disease diagnostics, prevention and therapeutics in clinical settings.

Keywords: cerebrospinal fluids, mNGS, metagenomics, pathogen, microbiome

62 **INTRODUCTION**

63 First defined by Joshua Lederberg in 2001(1), human microbiome has since been
64 discovered at almost every part of human bodies such as gut, oral, skin, bladder, vagina,
65 lungs(2-8). It has profound impact on human health, being associated with a broad
66 range of human diseases including cancers, diabetes, schizophrenia and autoimmune
67 diseases etc.(9-12). However, due to difficulties in the identification and traceability of
68 contaminations, it remains controversial whether there are colonized microbial
69 community at some sites such as placenta, blood and amniotic fluid (13-17), although
70 recently both experimental and analytical methods have improved in sensitivity and
71 accuracy regarding microbiome discovery.

72 Cerebrospinal fluid (CSF) circulating the human central nervous system (CNS) has long
73 been considered sterile given that the blood-brain barrier is thought to effectively
74 protect against microbial invasions. However, this traditional knowledge of microbe-
75 free CSF has been challenged in recent years with several reports of microbes detected
76 in human brains and CSF. For example, a bacterial pathogen *Porphyromonas gingivalis*
77 was identified in brain regions including cerebral cortex and hippocampus in patients
78 with Alzheimer's disease(18). In addition, a number of DNA viruses in CSF were
79 identified from a subjects of virome(19). It remains elusive whether these reports are
80 evidence of a common microbiome in human CSF and CNS, or simply sporadic and
81 accidental events.

82 Given the debate over the existence of any microbial community in CSF and the
83 importance of understanding microbial infection in human central nervous systems, we

84 performed microbiome analysis to characterize bacteria, archaea, eukaryota and viruses
85 of CSF from a cohort of 23 donors without neurological disorders, as well as one-to-
86 one matched positive controls (oral and skin) and negative controls (normal saline).
87 DNA/RNA extraction buffers and sterile swab were also collected as controls. In total,
88 116 specimens of eight types were used in this study. Considering the limitations of 16s
89 ribosomal RNA based approach to achieve a consistent result in species or strain
90 level(20-23), we choose unbiased metagenomic next-generation sequencing (mNGS)
91 for rapidly detecting all genetic materials of microbiome at species resolution and use
92 metatranscriptomic sequencing to assess the physiological states of microbial
93 communities in CSF(24, 25). mNGS as a promising approach, its clinical diagnostic
94 performance in infectious diseases has been widely adopted in the medical community
95 by multi-center studies(26-28).
96 In our data analysis, we found no significant difference between CSF specimens and
97 negative controls in microbial species diversity. In all CSF samples, no active or viable
98 microbiome was present after subtracting microbial taxa detected in CSF by those
99 detected in contamination controls. Taken together, no strong evidence was found in
100 our study supporting that colonized microbiome exists in the cerebrospinal fluids.

101 # RESULTS

102 **Metagenomic sequencing of cerebrospinal fluids in healthy individuals**

103 To investigate whether there is microbiome in CSF, we collected and analyzed
104 microbiome of CSF samples from 23 pregnant women aged 23–40 years who
105 underwent intraspinal anesthesia before the caesarean section via lumbar puncture,

106 coupled with normal saline collected with syringe as negative controls. For each subject,
107 oral and skin microbiomes were also collected and analyzed as positive controls (Figure
108 1a). All samples were then subjected to DNA extraction and metagenomic shotgun
109 sequencing and analysis. Finally, to validate whether the microbiome, if any detected
110 in CSF, is physiologically active, metatranscriptome profilings for 12 of the pregnant
111 women CSF samples were performed. After quality control (QC) with KneadData(29)
112 (v0.7.2) for metagenomic and metatranscriptomic sequencing data, MetaPhlAn(30)
113 (latest version 3), a state-of-the-art taxonomic classification tool based on unique clade-
114 specific marker genes, was used to detect microbes in each sample.
115 In total, we detected 619 nonredundant microbial taxa in 116 specimens using
116 metagenomic and metatranscriptomic sequencing and analysis (Supplementary Table
117 S1). These microbes detected in all samples were dominated by bacteria (75%) and
118 viruses (24%). Overall, skin, oral and swab samples had the most abundant microbiome
119 of all samples with 393, 199 and 137 nonredundant microbial taxa, respectively. This
120 came as no surprise because skins and orals are well known to harbor a plethora of
121 microbes. By contrast, the number of microbial taxa detected in CSF_DNA (26),
122 negative controls (49) and extraction buffers (27) were relatively fewer (Figure 1b).
123 We then compared the taxa detected in different specimen types, finding that there was
124 little overlap among all samples. Skin, oral and sterile swab had a large amount of
125 unique microbial taxa among all sample types, with 243, 129 and 36 taxa found only in
126 these samples, respectively (Figure 1b). Although a large variation in the number of
127 microbes detected was observed for skins, orals and swabs, a much smaller variation

128 was found for CSF_DNA, CSF metatranscriptomic (CSF_RNA), negative controls and
129 extraction buffers (DNA/RNA buffer) (Figure 1b). The oral samples were rich in
130 *Streptococcus*, *Veillonella*, *Neisseria, Rothia* and *Prevotella,* while the skin samples
131 were rich in C*utibacterium*, *Staphylococcus*, *Micrococcus*, *Malassezia,* consistent with
132 many previous studies(4, 6, 31) (Supplementary Figure 1). Our successful detection of
133 known microbiome for orals and skins provided a proof-of-concept of NGS-based
134 metagenomic sequencing method, laying a solid foundation for our exploration of CSF
135 microbiome using such a method.
136 We next focused on examining the microbial species detected for each CSF_DNA
137 specimen. In CSF samples, a total of 76 redundant species including 11 (4 nonredundant)
138 bacteria, 61 (21nonredundant) viruses and four (one nonredundant) eukaryota taxa were
139 detected (Figure 1c). Most of these viruses were bacteriophages. The relative
140 abundance of microbes suggested the species "*Cyprinid_herpesvirus_3*" are the
141 predominant species in 19 of 23 CSF_DNA samples (Figure 1d). *Cutibacterium_acnes*
142 in species level appeared in 5 specimens. Additionaly, 100%, 26%, 22%, and 22% of
143 all CSF_DNA samples contain *Cyprinid herpesvirus 3*, *Human alphaherpesvirus 2*,
144 *Enterobacteria phage mEp460* and *Dasheen mosaic virus*, respectively. However,
145 "*Cyprinid herpesvirus 3*" detected in all CSF_DNA samples were also found in all
146 negative and skin samples, suggesting a likely external source of these microbes during
147 the CSF sampling procedure.
148 **The microbiome signature of cerebrospinal fluids and negative controls is similar**

149 Since microbial species were identified in both CSF_DNA and negative controls, it is
150 likely that microbial cells and/or DNA present in negative controls may have been
151 introduced into CSF during the sampling process. Similarly, the possibility of skin
152 microbiome being introduced into CSF during lumbar puncture could not be ruled out,
153 despite the application of skin surface sterilization. Therefore, we asked how similar in
154 general the microbiome signature is for different sample types by comparing the
155 microbial species detected in these samples. We first performed Non-metric
156 Multidimensional Scaling (NMDS) analysis and principal coordinates analysis (PCoA),
157 and then characterized the beta diversity of CSF and other specimen types using Bray-
158 Curtis distances, a metric commonly used to evaluate microbiome difference among
159 samples supported by Wilcoxon statistical significance. NMDS, PCoA (Supplementary
160 Figure 2), and beta-diversity analysis revealed an overall clear separation of microbial
161 communities for each sample type, except that microbiome in CSF_DNA specimens
162 overlapped partially with negative controls (Figure 2a). Statistical analysis suggested
163 beta-diversity between CSF_DNA and other sample types was significantly different
164 from CSF_DNA self-comparison. However, there was no significant difference
165 between CSF_DNA self beta-diversity and CSF_DNA-negative beta diversity (Wilcox
166 test: p=0.59) (Figure 2b). In addition, the low diversity suggested the microbial
167 communities in CSF_DNA and negative controls were highly similar. In fact, shared
168 microbial taxa between CSF_DNA and negative control accounted for 42% and 22%
169 of CSF_DNA and negative control, respectively. By contrast, 58% microbial taxa in
170 CSF-DNA were detected in skin samples, whereas only 4% of skin microbes were

171 found in CSF-DNA specimens. On one hand, these results indicated the microbial cells
172 or DNA detected in CSF samples may partly have come from negative controls during
173 sample collection. On the other hand, the high beta-diversity between skin and CSF
174 specimens implied that the skin surface sterilization before lumbar punctures effectively
175 prevented the contamination of CSF samples with most, if not all, skin microbes.

176 **No microbiome is present in the CSF after subtracting microbes from controls**

177 With the detected microbiome in CSF samples, we questioned whether these microbes
178 were truly CSF inhabitants or simply brought in from external sources such as skins,
179 negative controls and DNA extraction buffer. To verify whether the CSF contains de
180 facto colonized microbial communities, we substracted the microbes collectively
181 detected in negative control samples and DNA extraction buffer samples from microbes
182 of each CSF_DNA sample, a method commonly used and previously described by
183 human microbiome study(17). After substraction, 12 CSF samples contained no
184 microbe, whereas the other 11 CSF samples contain a total of 14 microbial taxa
185 including 11 viruses, 2 bacteria and 1 eukaryota. Since an introduction of microbes from
186 skins could not be completely ruled out, we further checked whether the 14 taxa were
187 present in skins as well and found that 6 of the 14 taxa were also found in the matching
188 skin microbiome. This left eight microbial taxa after substraction as potential CSF
189 inhabitating microbes, including five viruses ("*Bovine alphaherpesvirus 1*",
190 "*Escherichia virus V5*", "*Klebsiella virus KP27*", "*Macaca mulatta polyomavirus 1*",
191 "*Trichoplusia_ni_single_nucleopolyhedrovirus*"), two bacteria (*Hydrogenophilus*
192 *thermoluteolus*, *Tepidimonas fonticaldi*), and one eukaryota (*Aspergillus turcosus*).

193 The detection of microbes using the metagenomic approach offers a glimpse of
194 microorganisms present in certain niches. However, it remains uncertain whether these
195 microbes are live or dead, as DNA from dead cells are also detectable by mNGS.
196 Therefore, we further evaluated the physiological activities of the potential CSF-
197 inhabitating microbes using metatranscriptomic sequencing, because the number of
198 microbes detected by both in metagenomic and metatranscriptome would indicate
199 active microbes may be present in CSF samples. CSF transcriptomics revealed
200 transcripts for several microbial taxa including "*Equine infectious anemia virus*" and
201 "*Cyprinid herpesvirus 3*" appearing in all samples, and *Escherichia_coli* and *Dasheen*
202 *mosaic virus* appearing in eleven and nine samples, respectively. We then asked, for the
203 eleven CSF-DNA samples with microbes left after substraction by negative controls
204 and DNA extraction buffers, whether these microbes have detected in
205 metatranscriptomic data. The result showed that only "*human alphaherpesvirus 1*" had
206 signals from both CSF genomics and transcriptomics. However, "*human*
207 *alphaherpesvirus 1*" also appeared in the skin, suggesting no active microbiome was
208 detectable in CSF after removing this species potentially originated from skins.
209 Although metagenomic analysis detected the one *Aspergillus turcosus* species from four
210 individuals (Figure 3), no transcripts of *Aspergillus turcosus* were detected in
211 metatranscriptomic, suggesting a lack of living cell activity. *Aspergillus turcosus* is well
212 known as opportunistic human pathogens and can cause infections in individuals of
213 compromised immune systems(32). How this fungal species (cells or DNA) reaches the
214 CSF of the five healthy individuals is unknown, but it shows CSF, though without an

active microbiome, might not be entirely free of opportunistic fungi which could potentially cause infections in central nervous systems when host immune system is compromised. Taken together, our study found no strong evidence supporting actively transcribed microbiome in the CSF.

DISCUSSION

Hereby, CSF samples from a cohort of 23 healthy individuals without neurological disease with a matched set of controls were collected for microbiome detection using culture-independent approach by a whole genome 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 by comparing CSF and 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.

Compared with bacteria, more viruses were detected in CSF specimens. These viruses are mainly bacteriophages, most of which are also present in negative and skin samples. Although bacteriophages in the CSF have been reported, clear evidence for regular colonization of CSF by these viruses is lacking (19). Whether these viruses appear accidentally or colonized in CSF needs further investigation.

In our results, four CSF samples contained *Aspergillus* DNA, but no *Aspergillus* nucleic acid was detected in the corresponding RNA samples, suggesting that these *Aspergillus* DNA fragments may have come from contaminations. Due to the high sensitivity of

237 mNGS, it can also detect trace amount of nucleic acid fragments released from dead
238 microorganisms present in human periphery blood or tissues, experiment reagents and
239 consumables. Furthermore, when using puncture to collect CSF specimen, tissues such
240 as skin, muscle, blood vessels are potential sources of contamination. Except for strict
241 disinfection measures before operation, constructing a database of colonizing
242 microorganisms of these healthy tissues will enable subsequent bioinformatics analysis
243 to filter out noise signals and reduce the false positive rate.
244 Highly sensitive mNGS represents a powerful tool for detecting microbiome at species
245 resolution, especially for microbiome studies in specimens of low-abundance biomass,
246 such as CSF. Currently, mNGS has become an important auxiliary method for clinical
247 pathogenic diagnosis and treatment of infectious diseases. Main challenges of studying
248 this issue have been an overall lack of CSF samples from healthy human subjects and
249 the technically sound sampling as well as data analysis methods based on different
250 reference databases and taxonomic strategies.
251 We focused on determining whether a CSF microbiome is present in healthy individuals
252 without neurological disorders, a long-disputed issue in scientific and clinical research
253 field. Our data analysis demonstrated that the microbiome of CSF was indistinguishable
254 from contamination controls. It is intriguing but remains unclear whether a microbiome
255 is present in CSF of patients diagnosed with diseases such as Alzheimer's disease,
256 multiple sclerosis, Parkinson’s disease and what roles the CSF microbiome plays in the
257 development of these disorders.
258 In conclusion, using metagenomic combined with metatranscriptomic deep sequencing,

259 we found microbiome profile in CSF samples was indistinguishable from that in
260 contamination controls. Our data indicated that by current approaches there was no
261 evidence to support the existence of a CSF microbiome in the populations without
262 known neurological disorders. Such findings shall have great implications to human
263 health especially neurological disorders and infections, providing a guide for disease
264 diagnostics, prevention and therapeutics in clinical settings.

265 MATERIALS AND METHODS

266 Subjects

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

277 Sample collection

278 Lumbar puncture was performed in the 23 subjects enrolled in this study and the CSF
279 samples were collected in a 4ml centrifugal tube with syringe and then stored in a
280 −80 °C freezer for metagenomics analysis. Twelve CSF samples were randomly

281 selected from 23 pregnant women for metatranscriptome studies and RNA protection
282 reagent was added to the CSF immediately after collection. Then, the samples were
283 centrifuged and the pellets stored at −80 °C for metatranscriptomic sequencing.
284 Meanwhile, normal saline was collected with syringes for environmental controls
285 (negative control). Furthermore, oral and skin samples were selected from 23 enrolled
286 subjects as one-to-one matched positive control. For skin positive controls: The back
287 skin of 5×5 cm^2 areas around the puncture site (L3-L4 intervertebral space) were
288 swabbed with a sterile cotton swab before skin clean with povidone iodine. To
289 maximize microbial load, no bathing was permitted within 24 hours of sample
290 collection. For oral positive controls, all subjects were forbidden to eat and drink six
291 hours before operation. The surfaces of tongue, buccal fold, hard palate, soft palate,
292 tooth, gingiva and saliva were swabbed with sterile swab. Unused sterile swabs were
293 collected for negative controls ("sterile swab"). Details of Matching information
294 between samples are described in Supplementary Table S2.

295 **DNA extraction and purification**

296 DNA was isolated using the QIAamp DNA Mini Kit (Qiagen 51304) according to the
297 manufacturer's instructions. 1) DNA extraction from swabs: Swab tips were cut and
298 placed in a 2 ml microcentrifuge tube and then 400 μl PBS were added. Next, 20 μl of
299 proteinase K and 400 μl of buffer AL were added, vortexed for 10 s, and the solution
300 was incubated for 15 min at 56 °C. And then added 400 μl ethanol (100%) and mixed
301 again by vortexing. Lastly, DNA purification was performed with buffer AW1 and AW2
302 using QIAamp Mini spin column, followed by elution with 35 μl of buffer EB. 2) DNA

303 extraction from CSF and normal saline controls: 200 μl sample was added into the
304 microcentrifuge tube, and then added 20 μl of proteinase K and 200 μl of buffer AL
305 respectively, vortexed for 10 s, and the solution was incubated for 15 min at 56 °C. Next,
306 added 200 μl ethanol (100%) and mixed again by vortexing. DNA purification was
307 performed as described above.

308 **Metagenomics library construction**

309 For preparation of metagenomics libraries, the QIAseq FX DNA Library Kit (Qiagen;
310 180715) was used. The construction involved five steps: 1) Fragmentation and End-
311 repair: to generate 200–300 bp fragments, 32.5μl purified DNA were fragmented by
312 incubation with FX buffer 5μl, FX enhancer 2.5μl and 10 μl FX enzyme mix at cycling
313 program: 4 ºC 1minute→32 º C 12minutes→65 º C 30minutes→4 º C hold. 2)
314 Adapter ligation: 5 μl of adaptor, 20 μl of ligation buffer, 10 μl of DNA ligase and 15
315 μl of nuclease-free water were added and incubated for 15 minutes at 20 °C to initiate
316 adapter ligation. Adapter ligation cleanup was performed immediately, 3) Adapter
317 ligation cleanup: 80 μl of resuspended AMPure® XP beads (0.8×) were added to each
318 ligated sample and mix well by pipetting. Next, the mixture was incubated for 5 minutes
319 at room temperature and then the beads were pelleted on a magnetic stand (Invitrogen)
320 for 2 minutes. The supernatant was discarded and the pallet was washed twice with 200
321 μl of 80% ethanol, then the beads were eluted with 52.5 μl of buffer EB. Subsequently,
322 50 μl of supernatant was transfered into a new 1.5 ml microcentrifuge tube and a second
323 purification was performed with 50 μl (1×) AMPure® XP beads. The final, 23.5 μl of
324 purified DNA sample was obtained. 4) Amplification of library DNA: 25 μl of HiFi

325 PCR Master Mix, 1.5 µl of Primer Mix and 23.5 µl of library DNA were added in PCR
326 tube. PCR enrichment was performed under the cycle conditions: 2 minutes at 94 ºC,
327 12 × (20 s at 98 ºC, 30 s at 60 ºC, 30 s at 72 ºC), and 1 minute at 72 ºC. The final, to
328 obtain libraries, the PCR products were purified with AMPure XP beads as described
329 above.

330 **RNA extraction and purification**

331 Total RNA was extracted using the RNeasy Mini Kit (Qiagen; 74104) according to the
332 manufacturer's instructions. The pellet of each sample which has been treated with
333 RNA protection reagent as described above, was resuspended in 100µl TE buffer
334 containing lysozyme, and Proteinase K was added into the mixture, then incubated for
335 10 minutes at room temperature. 350µl of buffer RLT was added and vortexed
336 vigorously. The final, RNA isolation and purification was performed with buffer AW1
337 and RPE respectively using RNeasy Mini spin column, followed by elution with
338 RNase-free water.

339 **RNA library preparation for metatranscriptomics sequencing**

340 For construction of RNA libraries, the QIAseq FX Single Cell RNA Library Kit (Qiagen;
341 180733) was used. The construction involved five steps: 1) Genomic DNA (gDNA)
342 removal: 8 µl of purified RNA and 3 µl of NA denaturation buffer were added into a
343 sterile PCR tube and incubated for 3 minutes at 95 °C. To remove genomic DNA, 2 µl
344 of gDNA wipeout buffer was added and incubated for 10 minutes at 42 °C. 2) Reverse
345 transcription: 4 µl of RT/Polymerase buffer, 1 µl of random primer, 1 µl of Oligo dT
346 primer and 1 µl of Quantiscript RT enzyme mix were added in each sample and reverse

347 transcription was carried out for 60 minutes at 42 °C. 3) Ligation: 8 μl of ligase buffer
348 and 2 μl of ligase mix were added into the RT reaction and incubated at 24ºC for 30
349 minutes. 4) Whole transcriptome amplification: 1 μl of REPLI-g SensiPhi DNA
350 Polymerase and 29μl of reaction buffer were used for Multiple Displacement
351 Amplification (MDA), then incubate at 30ºC for 2 h. The final, an approximate length
352 of 2000–70,000 bp amplified cDNA was produced. 5) Enzymatic Fragmentation: The
353 amplified cDNA was diluted 1:3 in H_2O sc, 10 μl of the diluted DNA and FX Enzyme
354 Mix were used to obtain 300 bp library fragment with reaction conditions: 4° C
355 1minute→32° C 15minutes→65° C 30minutes→4° C hold. 6) Adapter ligation: 5 μ
356 l of adapter and 45 μ l of ligation master mix were added into each sample and incubated
357 at 20° C for 15 minutes. Subsequently, the adapter ligation cleanup was performed with
358 AMPure XP beads as described above. The final, purified libraries were obtained ready
359 for sequencing without further PCR amplification.

360 **Next generation equencing**

361 Shotgun sequencing was performed on Illumina Hiseq platform for all samples (paired
362 end library of 150-bp and 150-bp read length). Approximately, 25 Gb and 5 Gb of raw
363 paired-end reads were obtained per sample in the CSF genomics samples and negative
364 samples, respectively.

365 **Data quality control**

366 To reduce the impact of host reads on the results, we need to remove human reads
367 involved in the raw sequencing data before bioinformatics analysis. KneadData(29)
368 (v0.7.2), a widely used tool, is designed to perform quality control on metagenomic and

369 metatranscriptomic sequencing data, especially for microbiome experiments. All reads
370 were filtered using KneadData with the following trimmomatic options:
371 ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10:8:true, SLIDINGWINDOW:4:20,
372 MINLEN:50 and bowtie2 options: --very-sensitive, --dovetail. The proportion of
373 human reads in CSF genomics samples is up to 92%.

374 **Detecting potential microbiome**

375 MetaPhlAn(30) (v 3.0.1) is a computational tool for profiling the composition of
376 microbial communities (bacteria, archaea, viruses and eukaryotes) from shotgun
377 sequencing data. Based on ~1.1M unique clade-specific marker genes identified from
378 ~100,000 reference genomes, MetaPhlAn can profile unambiguous taxonomic
379 assignments and accurate estimation of organismal relative abundance in species-level
380 resolution. Classifying the reads to marker genes database, MetaPhlAn outputs a file
381 containing detected microbes and relative abundance in different level. MetaPhlAn ran
382 with custom parameters: --add_viruses --input_type fastq --read_min_len 50. It's worth
383 noting that MetaPhlAn (previous version 2) was the only bioinformatics tool with 0%
384 false positive relative abundance and the best diversity estimate(33).

385 **β-diversity and phylogenetic analysis**

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

Supplementary files

Supplementary Table S1. Microbes detected in different samples.

Supplementary Figure S1. Top 10 genus in oral and skin samples, respectively. Microbial 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.

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

Supplementary Table S2. Samples label and matching information.

Availability of data and materials

The clean sequence data reported in this paper have been deposited in the Genome Sequence Archive in BIG Data Center(34, 35), Chinese Academy of Sciences, under accession number ~~PRJCA004977~~XXXXX that are publicly accessible at https://~~/bigd.big.ac.cn/bioproject~~XXX.

Acknowledgements

We thank Peng Jia, Tingjie Wang, Ningxin Dang, Honghui Shen and Tun Xu for helpful discussions regarding data analysis and Jing Hai for administrative and technical support. We thank the High-Performance Computing Cluster of the First Affiliated Hospital of Xi'an Jiaotong University for data processing.

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).

Ethics Statement

This study was approved by the Ethics Committee of the Xijing Hospital of the Fourth 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.

421 **References**

422 1. Lederberg J, McCray A. 2001. `Ome Sweet `Omics--A Genealogical Treasury of Words. The
423 Scientist 15:8-8.
424 2. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. 2018. Current
425 understanding of the human microbiome. Nature Medicine 24:392-400.
426 3. Clemente JC, Ursell LK, Parfrey LW, Knight R. 2012. The impact of the gut microbiota on
427 human health: an integrative view. Cell 148:1258-1270.
428 4. Nasidze I, Li J, Quinque D, Tang K, Stoneking M. 2009. Global diversity in the human salivary
429 microbiome. Genome research 19:636-643.
430 5. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J,
431 Tacket CO. 2011. Vaginal microbiome of reproductive-age women. Proceedings of the
432 National Academy of Sciences 108:4680-4687.
433 6. Grice EA, Segre JA. 2011. The skin microbiome. Nature reviews microbiology 9:244-253.
434 7. Beck JM, Young VB, Huffnagle GB. 2012. The microbiome of the lung. Translational Research
435 160:258-266.
436 8. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl
437 AM, FitzGerald MG, Fulton RS. 2012. Structure, function and diversity of the healthy human
438 microbiome. nature 486:207.
439 9. Iida N, Dzutsev A, Stewart CA, Smith L, Bouladoux N, Weingarten RA, Molina DA, Salcedo R,
440 Back T, Cramer S. 2013. Commensal bacteria control cancer response to therapy by
441 modulating the tumor microenvironment. Science 342:967-970.
442 10. Moreno-Indias I, Cardona F, Tinahones FJ, Queipo-Ortuño MI. 2014. Impact of the gut
443 microbiota on the development of obesity and type 2 diabetes mellitus. Frontiers in
444 microbiology 5:190.
445 11. Zhu F, Ju Y, Wang W, Wang Q, Guo R, Ma Q, Sun Q, Fan Y, Xie Y, Yang Z, Jie Z, Zhao B, Xiao L,
446 Yang L, Zhang T, Feng J, Guo L, He X, Chen Y, Chen C, Gao C, Xu X, Yang H, Wang J, Dang Y,
447 Madsen L, Brix S, Kristiansen K, Jia H, Ma X. 2020. Metagenome-wide association of gut
448 microbiome features for schizophrenia. Nature Communications 11:1612.
449 12. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N,
450 Snapper SB. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease
451 and treatment. Genome biology 13:R79.
452 13. de Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, Parkhill J, Charnock-Jones DS,
453 Smith GCS. 2019. Human placenta has no microbiome but can contain potential pathogens.
454 Nature 572:329-334.
455 14. Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, Leite R, Elovitz MA,
456 Parry S, Bushman FD. 2016. Comparison of placenta samples with contamination controls
457 does not provide evidence for a distinct placenta microbiota. Microbiome 4:29.
458 15. Castillo DJ, Rifkin RF, Cowan DA, Potgieter M. 2019. The Healthy Human Blood Microbiome:
459 Fact or Fiction? Frontiers in Cellular and Infection Microbiology 9.
460 16. Whittle E, Leonard MO, Harrison R, Gant TW, Tonge DP. 2019. Multi-Method Characterization
461 of the Human Circulating Microbiome. Frontiers in Microbiology 9.
462 17. Lim ES, Rodriguez C, Holtz LR. 2018. Amniotic fluid from healthy term pregnancies does not
463 harbor a detectable microbial community. Microbiome 6:1-8.
464 18. Dominy SS, Lynch C, Ermini F, Benedyk M, Marczyk A, Konradi A, Nguyen M, Haditsch U, Raha

465 D, Griffin C. 2019. Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for
466 disease causation and treatment with small-molecule inhibitors. Science advances
467 5:eaau3333.
468 19. Ghose C, Ly M, Schwanemann LK, Shin JH, Atabakhsh K, Barr JJ, Little M, Schooley RT, Pride
469 DT. 2019. The virome of cerebrospinal fluid: viruses where we once thought there were none.
470 Frontiers in microbiology 10:2061.
471 20. Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P.
472 2010. Experimental factors affecting PCR-based estimates of microbial species richness and
473 evenness. The ISME journal 4:642-647.
474 21. Klappenbach JA, Saxman PR, Cole JR, Schmidt TM. 2001. rrndb: the Ribosomal RNA Operon
475 Copy Number Database. Nucleic acids research 29:181-184.
476 22. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O'Toole PW. 2010.
477 Comparison of two next-generation sequencing technologies for resolving highly complex
478 microbiota composition using tandem variable 16S rRNA gene regions. Nucleic acids research
479 38:e200-e200.
480 23. Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y,
481 Vandamme P, Thompson FL. 2005. Re-evaluating prokaryotic species. Nature Reviews
482 Microbiology 3:733.
483 24. Bikel S, Valdez-Lara A, Cornejo-Granados F, Rico K, Canizales-Quinteros S, Soberón X, Del
484 Pozo-Yauner L, Ochoa-Leyva A. 2015. Combining metagenomics, metatranscriptomics and
485 viromics to explore novel microbial interactions: towards a systems-level understanding of
486 human microbiome. Computational and Structural Biotechnology Journal 13:390-401.
487 25. Shakya M, Lo C-C, Chain PSG. 2019. Advances and Challenges in Metatranscriptomic Analysis.
488 Frontiers in Genetics 10.
489 26. Ramachandran PS, Wilson MR. 2020. Metagenomics for neurological infections — expanding
490 our imagination. Nature Reviews Neurology 16:547-556.
491 27. Chiu CY, Miller SA. 2019. Clinical metagenomics. Nature Reviews Genetics 20:341-355.
492 28. Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, Federman S, Stryke D, Briggs B,
493 Langelier C, Berger A, Douglas V, Josephson SA, Chow FC, Fulton BD, DeRisi JL, Gelfand JM,
494 Naccache SN, Bender J, Dien Bard J, Murkey J, Carlson M, Vespa PM, Vijayan T, Allyn PR,
495 Campeau S, Humphries RM, Klausner JD, Ganzon CD, Memar F, Ocampo NA, Zimmermann LL,
496 Cohen SH, Polage CR, DeBiasi RL, Haller B, Dallas R, Maron G, Hayden R, Messacar K,
497 Dominguez SR, Miller S, Chiu CY. 2019. Clinical Metagenomic Sequencing for Diagnosis of
498 Meningitis and Encephalitis. New England Journal of Medicine 380:2327-2340.
499 29. Anonymous. KneadData software: http://huttenhower.sph.harvard.edu/kneaddata.
500 30. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012.
501 Metagenomic microbial community profiling using unique clade-specific marker genes.
502 Nature Methods 9:811-814.
503 31. Chen YE, Fischbach MA, Belkaid Y. 2018. Skin microbiota–host interactions. Nature 553:427-
504 436.
505 32. Wilson MR, O'Donovan BD, Gelfand JM, Sample HA, Chow FC, Betjemann JP, Shah MP, Richie
506 MB, Gorman MP, Hajj-Ali RA, Calabrese LH, Zorn KC, Chow ED, Greenlee JE, Blum JH, Green G,
507 Khan LM, Banerji D, Langelier C, Bryson-Cahn C, Harrington W, Lingappa JR, Shanbhag NM,
508 Green AJ, Brew BJ, Soldatos A, Strnad L, Doernberg SB, Jay CA, Douglas V, Josephson SA,

509 DeRisi JL. 2018. Chronic Meningitis Investigated via Metagenomic Next-Generation
510 Sequencing. JAMA Neurology 75:947-955.
511 33. Amos GCA, Logan A, Anwar S, Fritzsche M, Mate R, Bleazard T, Rijpkema S. 2020. Developing
512 standards for the microbiome field. Microbiome 8:98.
513 34. Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, Tang B, Dong L, Ding N, Zhang Q, Bai Z, Dong X,
514 Chen H, Sun M, Zhai S, Sun Y, Yu L, Lan L, Xiao J, Fang X, Lei H, Zhang Z, Zhao W. 2017. GSA:
515 Genome Sequence Archive*. Genomics, Proteomics & Bioinformatics 15:14-18.
516 35. Members NGDC, Partners. 2019. Database Resources of the National Genomics Data Center
517 in 2020. Nucleic Acids Research 48:D24-D33.
518

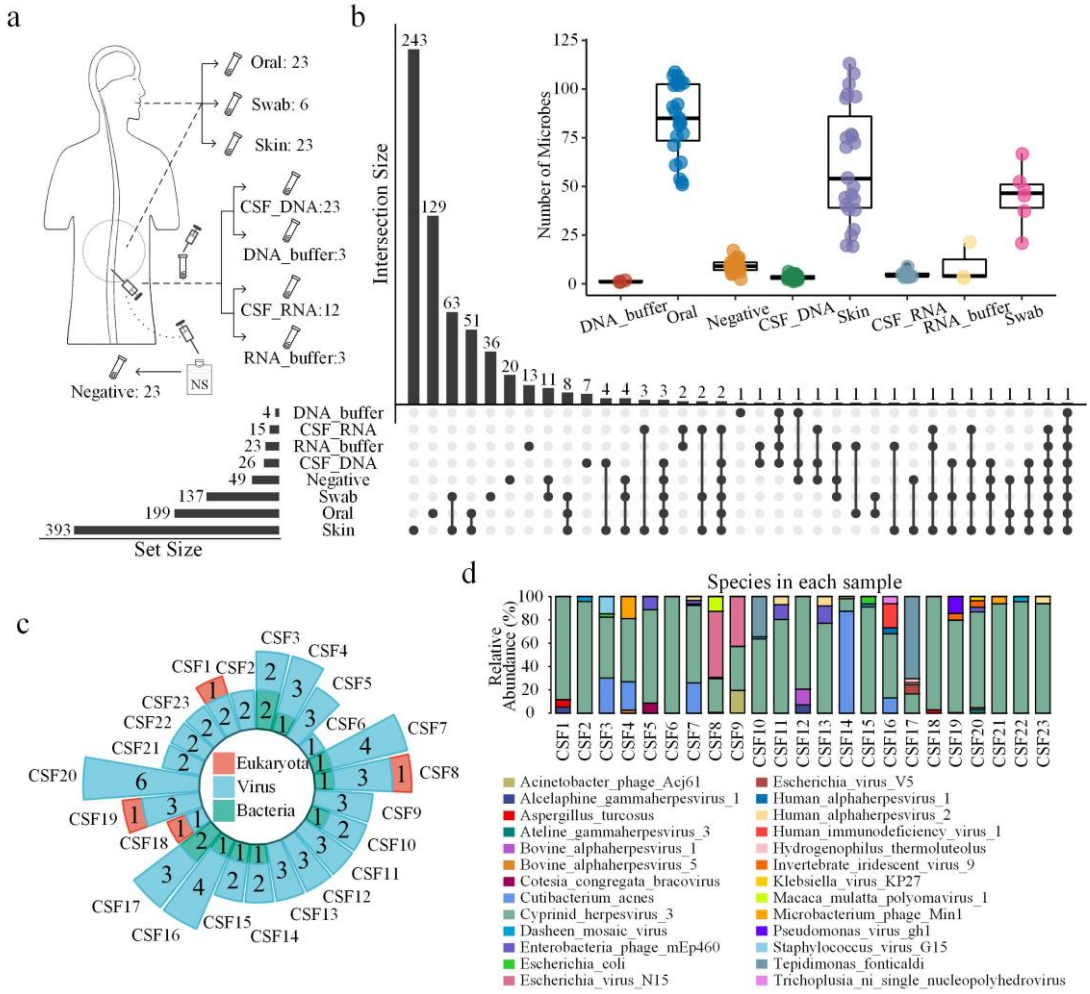


Figure 1: Microbial community structure in human CSF of 23 healthy individuals. (a) Metagenomic experimental design in this study: CSF and matched control samples (positive controls: oral and skin; negative controls: saline solution) collected from 23 pregnant women along with DNA and RNA extraction 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.

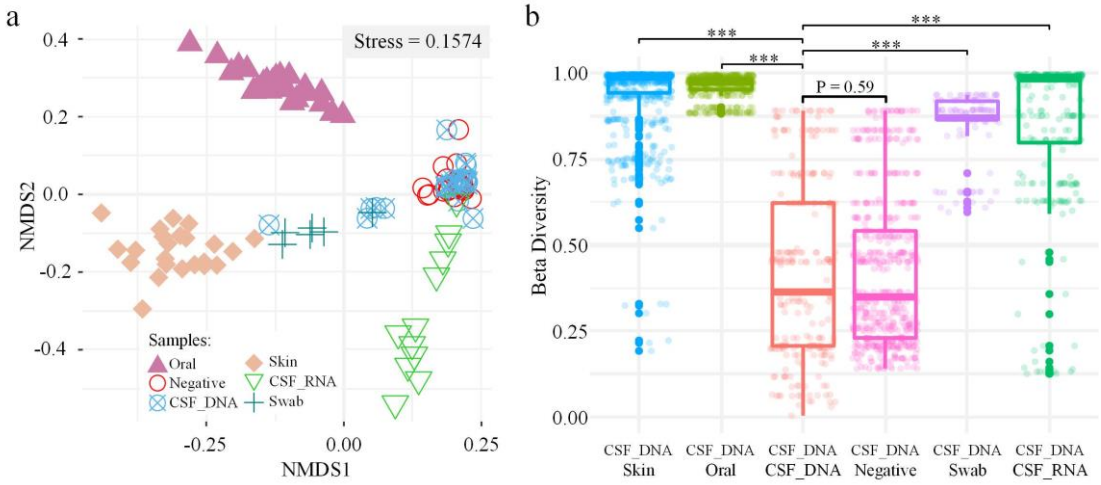


Figure 2. Microbiome similarity among sample types. (a) NMDS (Non-metric Multidimensional Scaling) analysis of microbial species detected from different sample types. Shapes and colors represent sample types. (b) Boxplot summarizing the beta diversity within CSF_DNA and between CSF_DNA and other specimens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

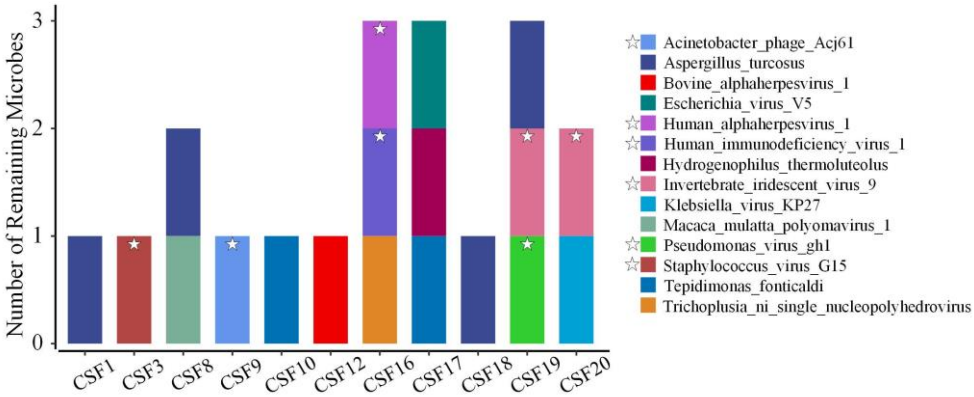
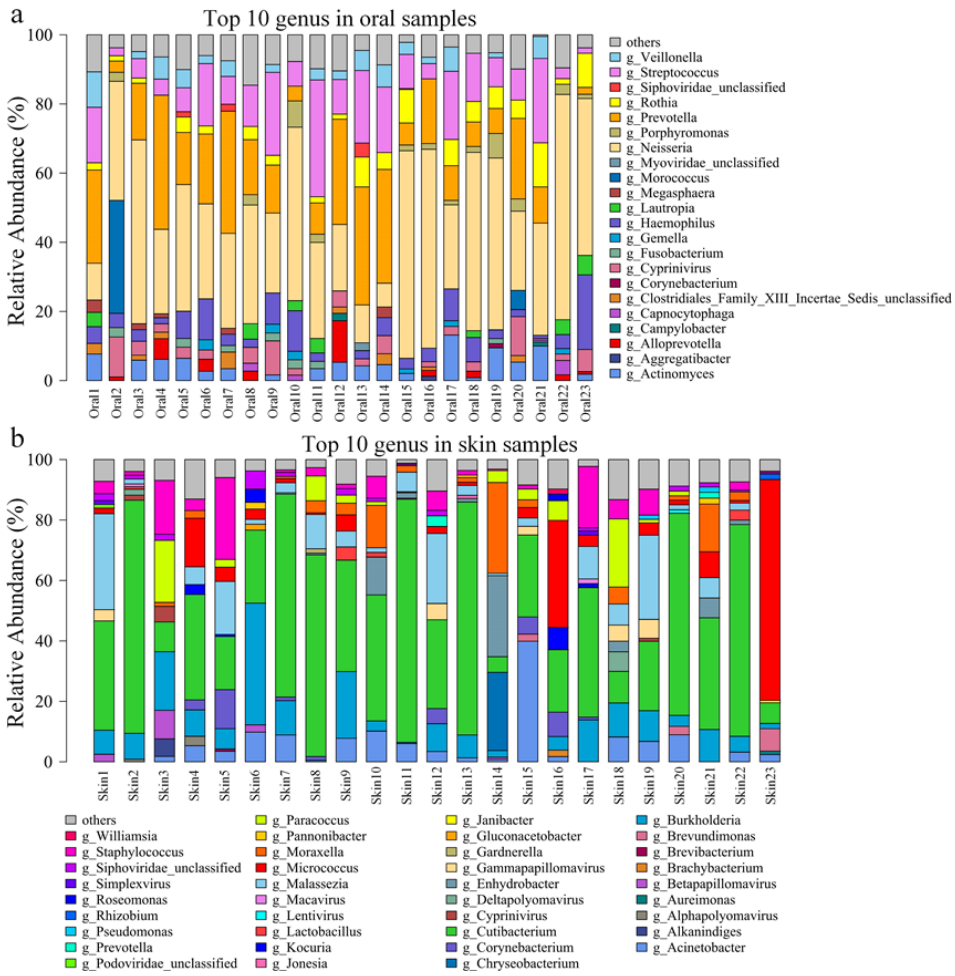


Figure 3. Microbes remained in the cerebrospinal fluids. Subtracting the microbes appeared in the negative control and DNA extraction buffer, 14 species (6 species labeled with star appeared in skin samples) remined in CSF genomic samples.

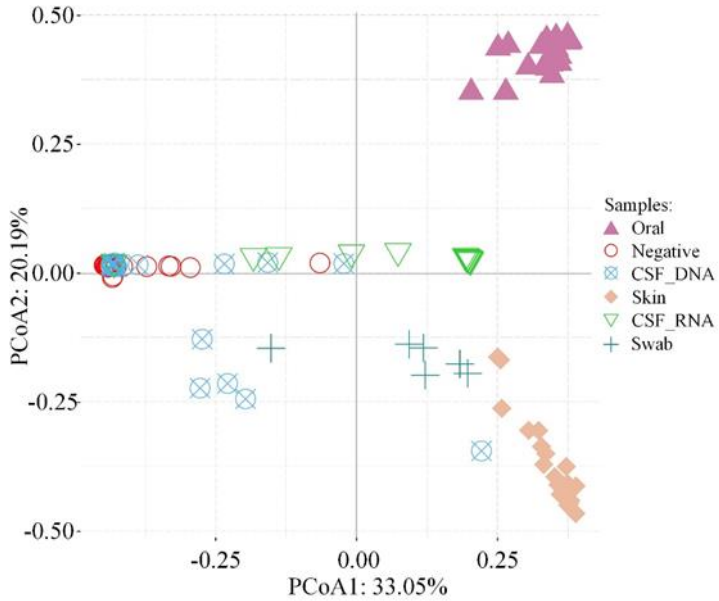
Supplementary Figure S1: Top 10 genus in oral and skin samples, respectively.



Microbial 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.

550 **Supplementary Figure S2**: Principal Coordinates Analysis (PCoA) analysis of

551 microbial species detected from different sample types.



552

553 PCoA (Principal Coordinates Analysis) analysis of microbial species detected from different sample

554 types. Shapes and colors represent sample types.

555

Supplementary Table S2: Sample labels and matching information.

Table S1: Samples label and matching information							
CSF_DNA	Negative	Skin	Oral	CSF_RNA	Swab	RNA_buffer	DNA_buffer
1	1	1	1	1	1	1	1
2	2	2	2		2	2	2
3	3	3	3		3	3	3
4	4	4	4	4	4		
5	5	5	5		5		
6	6	6	6		6		
7	7	7	7				
8	8	8	8	8			
9	9	9	9	9			
10	10	10	10	10			
11	11	11	11				
12	12	12	12				
13	13	13	13				
14	14	14	14	14			
15	15	15	15	15			
16	16	16	16	16			
17	17	17	17				
18	18	18	18	18			
19	19	19	19	19			
20	20	20	20				
21	21	21	21	21			
22	22	22	22				
23	23	23	23	23			
CSF_DNA: Cerebrospinal fluids metagenomic							
CSF_RNA: Cerebrospinal fluids metatranscriptomic							