Single-cell omics for complex microbial communities

Xihong Lan[1,2,3†], Qiaoxing Liang[1,2,3†], Jiayi Wu[1], Jinhua He[4], Xiaoying Zhang[4], Fei Li[1,2,3], Guoping Zhao[1,2], Ruidong Guo[1], Huijue Jia[1,2,3]*

[1]Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China.

[2]School of Life Sciences, Fudan University, Shanghai, China.

[3]State Key Laboratory of Genetic Engineering, Fudan University, Shanghai, China.

[4]Central Laboratory of Panyu Central Hospital, The Affiliated Panyu Central Hospital of Guangzhou Medical University, Guangzhou, China.

†These authors contributed equally to this work

*Corresponding author. Email: huijue_jia@fudan.edu.cn

Abstract

Despite over a decade of effort, microbial single-cell genomics and transcriptomics remain a challenge, especially for complex communities such as the human microbiome. Here we report a solution based on the Laser-Induced Forward Transfer (LIFT) technology, which circumvent the need for droplet microfluidics and complex barcoding, and allow selective amplification of the genome and the transcriptome of commensal bacteria from complex samples. We capture the diverse oral microbiome, and elucidate the single-cell transcriptome in sporulation. This single-cell method is both scalable and precise, and enables investigation into tissue samples. The approach shows great promise for elucidating the single-cell fate of microbial populations and the native environment of the host-associated microbiome.

Introduction

Single-cell transcriptome analyses for mammalian cells are reaching their prime, further boosted by spatial omics technology (*1*). Single-cell omics for bacterial cells, however, have remained a daunting task, due to the orders of magnitude lower concentration of nucleic acids, the lack of a poly(A) tail, and the Poisson distribution of droplet microfluidics (*2*). Marine cyanobacteria were among the first bacteria to be analyzed using single-cell genomes (*3*), due to their dominance in the global ocean. Single-cell genomes of extreme environments have revealed genomic features of uncultured bacteria and archaea *(4)*. Other

than low-diversity communities, microbial single-cell studies have mostly been illustrated for model organisms such as *E. coli*, and the genome or transcriptome analyzed, often using targeted methods, has remained far from complete *(5-8)*.

Laser-Induced Forward Transfer (LIFT) is a straight-forward way of manipulating microparticles *(9)*. With a laser beam pushing on the matrix, LIFT is ideally suited for picking particles in the size range of bacteria. It can potentially tolerate food particles, cell debris, and other objects commonly present in the human microbiome that could undermine analyses using droplet microfluidics.

Metagenomic analyses of low-biomass environment such as tumors and placenta has been costly and controversial *(10)*. Thus, a direct approach to characterize the microbes in tissues with over 99.9% host sequences would also be very much welcomed.

Here we present microbial single-cell genomes and transcriptomes using a commercially available optics equipment, and demonstrate their superb utility in a number of scenarios.

Results

Draft genome quality assembly for microbial single cells

To develop microbial single-cell technology for complex samples, we started with saliva samples which are even more diverse than feces. A fresh saliva

45 sample was centrifuged and spotted onto the slide (Fig.S1), and bacteria cells
46 were visualized before LIFT (Fig1.A). Single cells were ejected by LIFT without
47 affecting neighboring cells (Fig1.A and B, no burnt circle with the low power
48 <75nJ) and Multiple Displacement Amplification (MDA) was performed in a
49 PCR tube. Without the need for sophisticated algorithms to pool assembly from
50 fragmental genomes, as was reported recently *(5)*, here we used state-of-the-
51 art algorithms developed for metagenomes *(11)*, and obtained a bacterial
52 genome of 99.99% completeness and low contamination 0.33% (Fig1.C,
53 Fig.S1). This genome was identified as *Megamonas funiformis*. Gene
54 annotation through Bakta revealed the detection of coding sequences (CDSs),
55 tRNAs, rRNAs, ncRNAs, and CRISPRs (Fig1.D, Table S1). Calculating the
56 coding density showed that the genome sequence is >90% coding (Fig.S1),
57 which is within the range for commensal bacteria (on average 87% for human
58 gut *(12)*). Thus, LIFT from complex samples could obtain high-quality single-
59 cell bacteria genomes (single-cell assembled genomes, SAGs) that far exceeds
60 typical single-cell genomes (*5*) or metagenomically assembled genomes *(13)*,
61 and are on-par with genomes from cultured isolates.

62 To increase the throughput and eliminate individual differences in operation,
63 we made use of an automated pipetting platform to take over the molecular
64 biology work after automated single-cell ejection (Fig.S1). Individual salivary
65 microbes of three fields from a single donor were marked for collection (0.1s
66 per cell), which consisted of diverse morphological characters (Fig2.A). After de

67 novo assembly, 4 of the 39 SAGs have completeness >90 and contamination
68 <5, which are regarded as high quality *(14)*. 27 of the 39 SAGs have
69 completeness >50 and contamination <10, which are medium quality (Table S2).
70 The species were identified according to GTDB-Tk with ANI >95%. With this
71 relatively small-scale attempt, we already obtained oral bacteria from all major
72 phyla, including *Actinobacteriota*, *Bacteroidota*, *Campylobacterota*, *Firmicutes*,
73 *Fusobacteriota*, *Patescibacteria* and *Proteobacteria* (Fig.2B). Species from the
74 oral cavity and related body sites were seen, such as *TM7x*, *Haemophilus sp.*,
75 *Neisserria sp.*, *Porphyromonas gingivalis*, *Filifactor alocis*, *Fusobacterium*
76 *peridonticum*, *Bifidobacterium vaginale* (renamed from *Gardnerella vaginalis*)
77 and *Parabacteroides merdae* (Fig.2B). These results indicate that the single-
78 cell genomes could be reliably obtained with medium to high quality in an
79 automated and selective manner to reveal the microbial landscape in a complex
80 sample.
81

82 **Cell-fate commitment revealed by single-cell transcriptome**

83 Sporulation is a strategy for bacteria to survive harsh environments (*15, 16*),
84 and to potentially distribute to other hosts. We cultured a strain of *Bacillus*
85 *licheniformis* isolated from human saliva, and LIFT-picked single cells to
86 analyze their transcriptome following citric acid treatment. Thousands of
87 transcripts were detected from cells stimulated by citric acid (Fig.3A), which is
88 consistent with widespread yet incomplete expression of the bacterial genome

89 per cell cycle, known for model organisms such as *E. coli (17)*. Besides protein
90 coding genes, tRNAs and regulatory sRNAs were also detected (Fig.3B),
91 favoring the robustness of the experimental protocol.

92 In addition to scRNA analysis according to the bacteria reference genome
93 (Fig.S2), the high-quality reads obtained could also be de novo assembled.
94 Gene counts obtained by these two analytical methods were similar
95 (significance $p < 0.05$) and the expression profile was highly correlated (Fig.3C).
96 Thus, both known and unknown bacteria could be investigated using this
97 scRNA approach.

98 Oligo(A) tails target bacterial mRNAs for degradation *(18)*; this coupling to
99 exonucleases in bacteria is evolutionarily similar to exosome degradation of
100 transcripts in the eukaryotic nucleus. We detected stretches of three or more
101 adenylates in a number of transcripts (Fig.S3, Table S2). TctA (Tripartite
102 tricarboxylate transporter TctA family), a large transmembrane protein involved
103 in citrate uptake, for example, showed long stretches of poly(A) spanning the
104 entire gene, which led to three contigs in de novo assembly (Fig3.D and Fig.S3).
105 The shorter form was more abundant in five of the five cells analyzed (Fig.3E),
106 consistent with degradation from 3' to 5'. These results highlight the
107 fundamental questions that can be addressed using high-quality microbial
108 single-cell transcriptome.

109 We next explored the dynamic transcription changes during the
110 commitment to sporulation in *B. licheniformis* following citric acid treatment for

111 nine days. In addition to the upright laser for LIFT, the commercially available
112 equipment we used also harbor an inverted laser for single-cell Raman
113 spectrometry (Fig1.B). Both the cell morphology and Raman spectra could be
114 rapidly obtained from a sample. During the citric acid-induced sporulation, the
115 day 0, day 3 and day 9, *B. licheniformis* cells can be visually and spectrally
116 divided into three groups, from day 0 rod (I), day 3 intermediate state (II) to day
117 9 oval spore (III) (Fig4.A and B). The mean spectrum of cells I, II, III were
118 different, and spectra from the individual microbial cells formed three clusters
119 (Fig4.C). Consistent with the decrease in growth, while ribosomal RNAs,
120 ribosomal proteins, and nucleoid associated proteins dominated the day 0
121 transcriptome, the reads proportion of other genes notably increased as the
122 cells undergo morphological changes towards sporulation (Fig.S4).
123 Comparison of the gene expression of cells I, II, III allowed us to identify 194
124 differentially expressed genes, with these transcripts mainly clustered into three
125 groups according to variations in their reads counts (Fig4.D). Gene ontology
126 (GO) analysis of Biological Process (BP), Molecular Function (MF), Cell
127 Component (CC) showed that the genes preferentially expressed in the
128 intermediate state of group II are linked to the pathways such as cellular
129 response to stress, sporulation, hydrolase activity and intrinsic component of
130 membrane (Fig.S4). Consistently, electron microscopy confirmed that cellular
131 component assembly occurred in cells II, which was also observed before
132 Raman microscopy and LIFT (Fig.S4). Comparing cells I and cells II, we found

133 that germination gene *gerD* co-present with sporulation related genes such as
134 spore coat protein and *YjcZ* family sporulation protein were upregulated after
135 citric acid treatment at day 3 (Fig4.D). When cells I transformed into spore,
136 *spoIIIAC* (stage III sporulation protein AC), *YtvI* (sporulation integral membrane
137 protein) and spore-specific gene *splB* (spore photoproduct lyase) were highly
138 expressed (Fig4.E).

139 Further looking into the day 3 samples led to five types of cells with distinct
140 cell morphology, and the scRNA also formed five groups (Fig5.A). Dimensional
141 reduction of gene expression profiles revealed that cells of type I, type IV and
142 type V were relatively clustered independently, while cells of type II and type III
143 were relatively dispersed (Fig5.B). Comparing gene counts in each type, we
144 found that cells of type II in cell division state were active in gene transcription
145 (Fig5.C). Genes whose expression were enriched in the different cell types
146 included enzymes, transcription regulator and biosynthesis proteins (Fig5.D).
147 Genes highly expressed in type I are FdhF/YdeP family oxidoreductase, manA
148 (mannose-6-phosphate isomerase, class I), PTS mannose/fructose/sorbose
149 transporter subunit IIC, phosphoribosylaminoimidazolesuccinocarboxamide
150 synthase, *ComGE* and 2-dehydropantoate 2-reductase, which enable type I
151 cells resistance to low PH caused by citric acid, transporting sugar molecules,
152 pantothenate biosynthesis, DNA uptake and purine biosynthesis (*19-23*).
153 Genes highly expressed in type II cells included DEAD/DEAH box helicase,
154 primosomal protein DnaI, DNA polymerase/3'-5' exonuclease PolX, DNA

155 helicase RecQ, RNA-binding protein and sigma-54-dependent Fis family
156 transcriptional regulator, which indicated DNA replication and gene transcription
157 process in cell division (*24, 25*). Genes observed in type II were also expressed
158 in type III, but relatively lower. When cells changed into type IV, GTP associated
159 genes MnmE, elongation factor Tu and GTP 3',8-cyclase MoaA, 50S proteins
160 and GntR family transcriptional regulator were highly expressed, which
161 regulated biological process such as gene translation and cell motility (*26, 27*).
162 Genes in type V were mainly transporters, Na+/H+ antiporter NhaC, permease
163 and bacteriocin, which functioned in stress resistance, substance exchange
164 and inhibition of spore germination (*28-30*). Comparing mean spectrum of five
165 type cells, we observed Raman signal at band 750nm-785nm contributed by
166 Cytochrome c and metabolites of purine degradation decreased (Fig5.E) (*31,*
167 *32*), which is consistent with variation of metabolic and gene transcriptional
168 activity in five type cells. In addition, transcription of sporulation and spore
169 related genes also varied with cell state (Fig5.F). These results illustrate the
170 discovery power of high-quality scRNA, and indicate that visual and/or label-
171 free Raman spectra-guided selection of microbial single cells could greatly
172 facilitate the investigation of heterogeneous microbial populations. All cells
173 could be rapidly scanned in seconds, before more costly amplification and
174 analyses.
175
176

177 **Single-cell microbial genome and transcriptome from tissue-resident**
178 **microbiota *in situ***

179 Having demonstrated single-cell draft genomes from complex samples and
180 high-quality precision single-cell transcriptome from heterogenous populations,
181 we went on to directly study microbial single-cells in tissue samples, which has
182 been impossible with other methods. Resected tumor tissues from a patient
183 diagnosed as colorectal cancer were cut into tissue sections and loaded onto
184 the slide for microbial single-cell omics (Fig6.A, Fig.S5). We confirmed cancer
185 infiltrating and staging by immunohistochemical staining of marker genes such
186 as P53, BrafV600E, Ki67 (Fig.S5). Using antibodies against bacterial
187 lipopolysaccharide (LPS) and vancomycin to detect Gram-negative and Gram-
188 positive bacteria respectively, we observed microbes with diverse
189 morphological features including rod shape and spherical shape, distributed in
190 colorectal cancer (Fig.S5). Without special treatment, we searched for microbes
191 in the bright field. The microbes have distinct Raman spectral fingerprints
192 compared to histiocytic Raman spectrum (Fig.S5), which means this technology
193 can very effectively avoid sequencing of host cells or collection of non-microbial
194 particles. For this colorectal cancer sample, a 1.33μm long and rod-shaped
195 single cell was spotted and separated for genomic sequencing combined with
196 Raman spectrum acquisition (Fig6.B and C). We obtained a medium-quality
197 SAG identified as *Bacteroides intestinalis* based on Mash analysis (Fig6.D).
198 Phylogenetic analysis of all *B. intestinalis* genomes from metadata-gut of Mash

database and our SAG revealed that SAG was similar to MGYG000028343, which is also from China (Fig6.E).

Moreover, a rod-shaped single cell adhered to intestinal mucosal was separated for RNA sequencing (Fig7.A). After removal of host gene reads, we obtained de novo transcriptome assembly identified as *Bacteroides xylanisolvens* by GTDB-Tk, in which 70% contigs and 75% transcripts belonged to *Bacteroides* (Fig7.B). Of note, we observed dust particles in the lungs of this colorectal cancer patient with 50 years of smoking when cancer metastasized to the lungs (Fig.S6), which is consistent with the finding that *Bacteroides xylanisolvens* is involved in nicotine degradation in smokers (*33*). Membrane proteins SusC/RagA family TonB-linked outer membrane protein, ABC transporter, sialic acid-specific acetylesterase and porin (Fig7.C), which are crucial for nutrients uptake and thrive of *Bacteroidetes* in the gut (*34-36*), were also detected. Gene ontology analysis further verified that these genes were enriched in pathways of passive transmembrane transporter activity, sialic acid transmembrane transporter activity and porin activity (Fig7.D). These results suggest the application potential of this method for tissue microbe detection *in situ* and single cell spatial microbiome.

Discussion

In summary, here we present an optics-based strategy for precision single-cell microbial genomics, and transcriptomes that could fully capture the genomic

221 landscape of host-associated microbiome, and allow nucleotide-resolution
222 understanding of microbial gene transcription and turnover in individual cells in
223 their native environment.

224 Knowing the complete single-cell genome and transcriptome would be very
225 useful to understanding fundamental questions of the microbiome beyond
226 model organisms. Both the single-cell genome and the transcriptome could be
227 either reference-based or de novo assembled. Rare populations could be
228 enriched before sequencing. Visually, the existing software could already be set
229 to select for cell size, shape, separation, etc. Further standardization and
230 algorithm development for the single-cell Raman spectra would allow more
231 precision in sub-species analyses.

232 We have previously reported multiple *Bacteroides* species in an Austrian
233 cohort of colorectal adenoma and carcinoma (*37, 38*) and *B. fragilis* were
234 typically studied in mice models of colitis (*39*). Further application of our
235 microbial single-cell omics method to clinical samples would facilitate patient
236 stratification and treatment.

Materials and Methods

Study participants and sample collection

All the adult volunteers were not known to have chronic diseases, and did not take antibiotics within the past three months. All healthy volunteers and the patient had written informed consent. All the samples were self-collected into a sterile tube, and handed in for further processing on the same day. The study was approved by the Institutional Review Board of Fudan University. Tissue collection was approved by the Ethics Committee at The Affiliated Panyu Central Hospital of Guangzhou Medical University and was conducted following guidelines established by this committee.

Laser Induced Forward Transfer (LIFT) of bacterial cells

The equipment, PRECI SCS-R300 (Hooke Instruments, Ltd., China), was turned on and adjusted according to standards following the manufacturer's handbook *(40)*. The filtered and washed samples were spotted onto a quartz slide (HSC24, Hooke Instruments, Ltd.), and air-dried in the hood. Visually selected bacterial cells (the manufacturer's software) were LIFT-transferred into collectors (HSR04, Hooke Instruments, Ltd.) using a laser power below 75mJ, and then flipped into individual PCR tubes with brief spinning at 2000g (Yooning Mini-6K).

Whole genome amplification and sequencing from single bacterial cells

Following cell lysis at 65℃ for 10 min, Multiple displacement amplification (MDA) was performed using the REPLI-g scDNA Polymerase, according to kit

instructions (REPLI-g single cell kit, #150345, QIAGEN). Sequencing libraries (PE150) were constructed without PCR, and sequenced on a DNBSEQ-T7 high-throughput sequencer (MGI).

Genome assembly and analyses

Quality control on paired-end sequencing reads of raw data was performed by fastp (0.23.2) tool. The filtered reads were assembled by metaSPAdes (3.13.0) in single cell mode. Based on binning information provided by Kraken2 (2.0.7), the obtained assembled bins were processed with bin refinement using BinSpreader. The quality of genome bins was analysed by Checkm2 and QUAST. GTDB-Tk (1.0.2) was used to classify and taxonomically assign genome bins. Phylogenetic trees were constructed by PhyloPhlAn (3.0.67) tool.

Whole transcriptome amplification and sequencing from single bacterial cells

In addition to phosphate buffered saline (PBS) and lysis buffer according to the QIAGEN kit (#150065), cell lysis was performed at 24℃ for 5 min, in the presence of murine RNase Inhibitor (Vazyme), and MetaPolyzyme (MAC4L, Invitrogen). Total RNA Reverse-transcription, ligation and MDA were performed using the REPLI-g WTA single cell kit (#150065, QIAGEN), with extended reaction time for bacteria instead of mammalian cells: 2 hours for ligation, 4 hours for MDA. Sequencing libraries (PE150) were constructed without PCR, and sequenced on a DNBSEQ-T7 high-throughput sequencer (MGI).

Single-cell transcriptome analyses

Filtered reads of scRNA sequencing sample was assembled using rnaSPAdes (3.15.5) *(41)*. Bakta (1.8.1) conducts a comprehensive annotation of assembled transcriptome. Protein sequences generated by Bakta (version 1.8.1; AMRFinderPlus for AMR gene annotations; a generalized protein sequence expert system for identity, distinct coverage and identity and priority values for each sequence) were further annotated by eggnog-mapper (5.0.2) for gene function. KofamScan is used to annotate protein sequences with KEGG Orthology (KO). DeepFRI is used to predict Enzyme Commission (EC) numbers and Gene Ontology (GO) terms for Bakta annotated genes. Bowtie2 is used to align reads to transcripts and calculate gene FPKM values. For genome-based analysis, reads are referred against *Bacillus licheniformis* genome and genome annotation GTF file downloaded from NCBI database. For analyzing the distribution and characteristics of oligoA, bioawk tool is used to analyze oligoA positions in gene/contig level, and then process the results to obtain matrix containing gene/contig lengths, oligoA positions, gene start and gene end positions. All transcripts in de novo transcriptome were evenly divided into ten percentile according to gene length and the number of occurrences of oligoA positions within each percentile range of the gene lengths was recorded. Standard deviation of oligoA distribution was calculated according to the fraction of counts in each percentile range of the gene length. The proportion of oligoA occurrence was normalized by R package “Mfuzz”. The analysis of

poly(A) with nine adenylate residues (9A) or more followed these steps. Genes expression with fold change ≥1.5 and pvalue < 0.05 were determined as significant change using DESeq2 (version 1.20.0). Gene enrichment of Biological Process (BP), Molecular Function (MF), Cell Component (CC) was analysed by R package Clusterprofiler with pvalueCutoff = 0.05. Heatmaps of gene expression were using R package “pheatmap”. The bar plots, bubble plots of GO analysis, scatterplots, and boxplots were visualized using on-line tool imageGP (http://www.ehbio.com/ImageGP/index.php/Home/Index/index.html).

Bacillus licheniformis culture and transmission electron microscopy

The stored *Bacillus licheniformis* strain was inoculated into the Columbia Broth medium overnight at 37°C. 100μl *Bacillus licheniformis* suspension was transferred to 2ml medium containing 10% Columbia Broth, in which PH was adjusted to 4.0 by adding citric acid. *Bacillus licheniformis* were cultured in citric acid-amended 10% Columbia Broth for day 3 and day 9.

For transmission electron microscopy, citric acid treated *Bacillus licheniformis* were fixed with 4% PFA and stored at 4℃. The copper mesh was incubated in 10μl fixed *Bacillus licheniformis* solution for 2 minutes. After removal of excessive solution using filter paper, copper mesh was incubated with 1% phosphotungstic acid (PTA) for 2 minutes and observed by electron microscope (FEI Tecnai G2 spirit).

Tissue pretreatment for LIFT cell sorting and Hematoxylin-Eosin staining, Immunofluorescence staining, Gomoris staining

Several pieces of 2mm thickness tissues (20×20×2mm) separated by laparoscopic colorectal cancer surgery were placed on filter paper and dry gauze pad. After removal of excessive water, tissues were transferred to supporter of freezing microtome (Leica CM1860 UV) and embedded with OCT (optimal cutting temperature compound). In order to ensure the operation of the tissue section, OCT embedded tissue was placed on cryo-stage until it's a solid ice cube. Processed tissue was sectioned into 7μm thickness for Raman spectrometry and LIFT on PRECI SCS-R300 (Hooke Instruments, Ltd., China). Dehydrated and paraffin embedded tissues were cut into 5μm thickness and were stained with hematoxylin-eosin after de-paraffinization. For immunofluorescence staining, tissue sections were treated with antigen retrieval buffers, and then permeabilized and blocked by mixed solution of 0.3% Triton X-100 and 5% BSA for 1 hour. Primary antibodies were used to incubate tissue sections overnight at 4°C. Phosphate Buffer Saline containing 0.1% Tween-20 (PBST) was used to wash tissue sections five times. Well-washed sections were carefully incubated with secondary antibodies for 1 hour at room temperature. DAPI (4',6-diamidino-2-phenylindole) diluted at PBS with concentration 1:1000 was used to cell nuclei staining. Immunofluorescence images were captured by Leica sp8 system. Antibodies and their working concentration are listed as followed: anti-CK20 (ZSGB-BIO, #ZA-0574, 1:200),

CDX2 (ZSGB-BIO, #ZA-0520, 1:200), Villin (ZSGB-BIO, #ZA-0575, 1:300), MSH2 (ZSGB-BIO, #ZA-0622, 1:250), MSH6 (ZSGB-BIO, #ZM-0367, 1:250), MLH1 (ZSGB-BIO, #ZM-0152, 1:300), PMS2 (ZSGB-BIO, #ZA-0542, 1:200), P53 (ZSGB-BIO, #ZM-0408, 1:200), BrafV600E (Roche, #760-5095, 1:300), and Ki67 (ZSGB-BIO, #ZM-0166, 1:300), anti-LPS (Cloud-Clone, #MAB526Ge22, 1:150), anti-Vancomycin (Invitrogen, #V34850, 1:150).

Gomoris staining method followed the manufacturer's protocol (Baso diagnostics. Inc. Zhuhai. China, #BA4375A). Briefly, 5μm thickness paraffin tissue sections were brought to water via xylene and descending grades of ethanol for deparaffinization. Following nuclear staining by Weigert's Iron Hematoxylin, sections were placed in Solution A for 15 minutes. After rinsing off the stain with distilled water, sections were stained with Solution B. After dehydration with ethanol and clearing with xylene, sections were mounted with a resinous medium.

References and Notes

1. P. L. Stahl *et al.*, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82 (2016).
2. A. Kuchina *et al.*, Microbial single-cell RNA sequencing by split-pool barcoding. *Science* **371**, (2021).
3. N. Kashtan *et al.*, Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. *Science* **344**, 416-420 (2014).
4. C. Rinke *et al.*, Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431-437 (2013).
5. W. Zheng *et al.*, High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science* **376**, eabm1483 (2022).
6. D. Dar, N. Dar, L. Cai, D. K. Newman, Spatial transcriptomics of planktonic and sessile

bacterial populations at single-cell resolution. *Science* **373**, (2021).
7. A. W. Pountain *et al.*, Transcription-replication interactions reveal bacterial genome regulation. *Nature* **626**, 661-669 (2024).
8. F. Lan *et al.*, Massively parallel single-cell sequencing of diverse microbial populations. *Nat Methods* **21**, 228-235 (2024).
9. C. Florian, P. Serra, Printing via Laser-Induced Forward Transfer and the Future of Digital Manufacturing. *Materials (Basel)* **16**, (2023).
10. H. Jia, Investigating Human Diseases with the Microbiome (Elsevier, ed. 1st, 2022, https://www.elsevier.com/books/investigating-human-diseases-with-the-microbiome/jia/978-0-323-91369-0). (2022).
11. F. Meyer *et al.*, Critical Assessment of Metagenome Interpretation: the second round of challenges. *Nat Methods* **19**, 429-440 (2022).
12. J. Li *et al.*, An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* **32**, 834-841 (2014).
13. J. Zhu *et al.*, Over 50,000 Metagenomically Assembled Draft Genomes for the Human Oral Microbiome Reveal New Taxa. *Genomics Proteomics Bioinformatics* **20**, 246-259 (2022).
14. R. M. Bowers *et al.*, Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725-731 (2017).
15. R. A. Greene, R. A. Slepecky, Minimal requirements for commitment to sporulation in Bacillus megaterium. *J Bacteriol* **111**, 557-565 (1972).
16. D. Schultz, P. G. Wolynes, E. Ben Jacob, J. N. Onuchic, Deciding fate in adverse times: sporulation and competence in Bacillus subtilis. *Proc Natl Acad Sci U S A* **106**, 21027-21034 (2009).
17. D. N. Macklin *et al.*, Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation. *Science* **369**, (2020).
18. K. J. Bandyra, B. F. Luisi, Licensing and due process in the turnover of bacterial RNA. *RNA Biol* **10**, 627-635 (2013).
19. C. Iobbi-Nivol, S. Leimkuhler, Molybdenum enzymes, their maturation and molybdenum cofactor biosynthesis in Escherichia coli. *Biochim Biophys Acta* **1827**, 1086-1101 (2013).
20. J. G. McCoy, E. J. Levin, M. Zhou, Structural insight into the PTS sugar transporter EIIC. *Biochim Biophys Acta* **1850**, 577-585 (2015).
21. D. R. Price, A. C. Wilson, A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol* **12**, 110 (2014).
22. J. Mom, I. Chouikha, O. Valette, L. Pieulle, V. Pelicic, Systematic functional analysis of the Com pilus in Streptococcus sanguinis: a minimalistic type 4 filament dedicated to DNA uptake in monoderm bacteria. *mBio* **15**, e0266723 (2024).
23. V. M. Levdikov *et al.*, The structure of SAICAR synthase: an enzyme in the de novo pathway of purine nucleotide biosynthesis. *Structure* **6**, 363-376 (1998).
24. C. Bruand, M. Farache, S. McGovern, S. D. Ehrlich, P. Polard, DnaB, DnaD and DnaI proteins are components of the Bacillus subtilis replication restart primosome. *Mol Microbiol* **42**, 245-255 (2001).
25. V. Shingler, Signal sensing by sigma 54-dependent regulators: derepression as a control mechanism. *Mol Microbiol* **19**, 409-416 (1996).

26. V. Ramakrishnan, Ribosome structure and the mechanism of translation. *Cell* **108**, 557-572 (2002).
27. G. F. Liu, X. X. Wang, H. Z. Su, G. T. Lu, Progress on the GntR family transcription regulators in bacteria. *Yi Chuan* **43**, 66-73 (2021).
28. D. L. Jack, N. M. Yang, M. H. Saier, Jr., The drug/metabolite transporter superfamily. *Eur J Biochem* **268**, 3620-3639 (2001).
29. R. Zhang *et al.*, A Novel MFS-MDR Transporter, MdrP, Employs D223 as a Key Determinant in the Na(+) Translocation Coupled to Norfloxacin Efflux. *Front Microbiol* **11**, 955 (2020).
30. K. Egan *et al.*, Bacteriocins: Novel Solutions to Age Old Spore-Related Problems? *Front Microbiol* **7**, 461 (2016).
31. W. R. Premasiri *et al.*, The biochemical origins of the surface-enhanced Raman spectra of bacteria: a metabolomics profiling by SERS. *Anal Bioanal Chem* **408**, 4631-4647 (2016).
32. V. R. I. Kaila, M. Wikstrom, Architecture of bacterial respiratory chains. *Nat Rev Microbiol* **19**, 319-330 (2021).
33. B. Chen *et al.*, Gut bacteria alleviate smoking-related NASH by degrading gut nicotine. *Nature* **610**, 562-568 (2022).
34. L. S. Robinson, W. G. Lewis, A. L. Lewis, The sialate O-acetylesterase EstA from gut Bacteroidetes species enables sialidase-mediated cross-species foraging of 9-O-acetylated sialoglycans. *J Biol Chem* **292**, 11861-11872 (2017).
35. D. Papamichail, N. Delihas, Outer membrane protein genes and their small non-coding RNA regulator genes in Photorhabdus luminescens. *Biol Direct* **1**, 12 (2006).
36. L. M. Hall *et al.*, Sequence diversity and antigenic variation at the rag locus of Porphyromonas gingivalis. *Infect Immun* **73**, 4253-4262 (2005).
37. Q. Feng *et al.*, Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun* **6**, 6528 (2015).
38. J. Wang, H. Jia, Metagenome-wide association studies: fine-mining the microbiome. *Nat Rev Microbiol* **14**, 508-522 (2016).
39. C. C. Wong, J. Yu, Gut microbiota in colorectal cancer development and therapy. *Nat Rev Clin Oncol* **20**, 429-452 (2023).
40. H. Instruments, PRECI SCS-R300 Raman Single Cell Sorter. (2022). (available at https://en.hooke-instruments.com/ProductCenter/unicellular/2022-02-18/41.html).
41. E. Bushmanova, D. Antipov, A. Lapidus, A. D. Prjibelski, rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* **8**, (2019).

Acknowledgements

The authors were grateful to Zhong Yang, Lili Li, Jiong Ma, Xu Zhang, Hang Li, Meiya Lin for equipment support, Ping Sun, Miaoyi Kuang for sequencing support.

Author contributions

Xihong Lan and Huijue Jia conceived and designed the methodology; Xihong Lan developed and performed the experiments with assistance from Jiayi Wu; Xihong Lan and Jiayi Wu optimized sequencing library preparation; Jinhua He and Xiaoying Zhang prepared tissue sections; Xihong Lan, Qiaoxing Liang and Ruidong Guo performed the sequencing data analysis; Xihong Lan wrote the original manuscript; Huijue Jia and Guoping Zhao revised the manuscript; Huijue Jia supervised the study.

Competing interests

Authors declare that they have no competing interests.

Figure captions

Fig.1

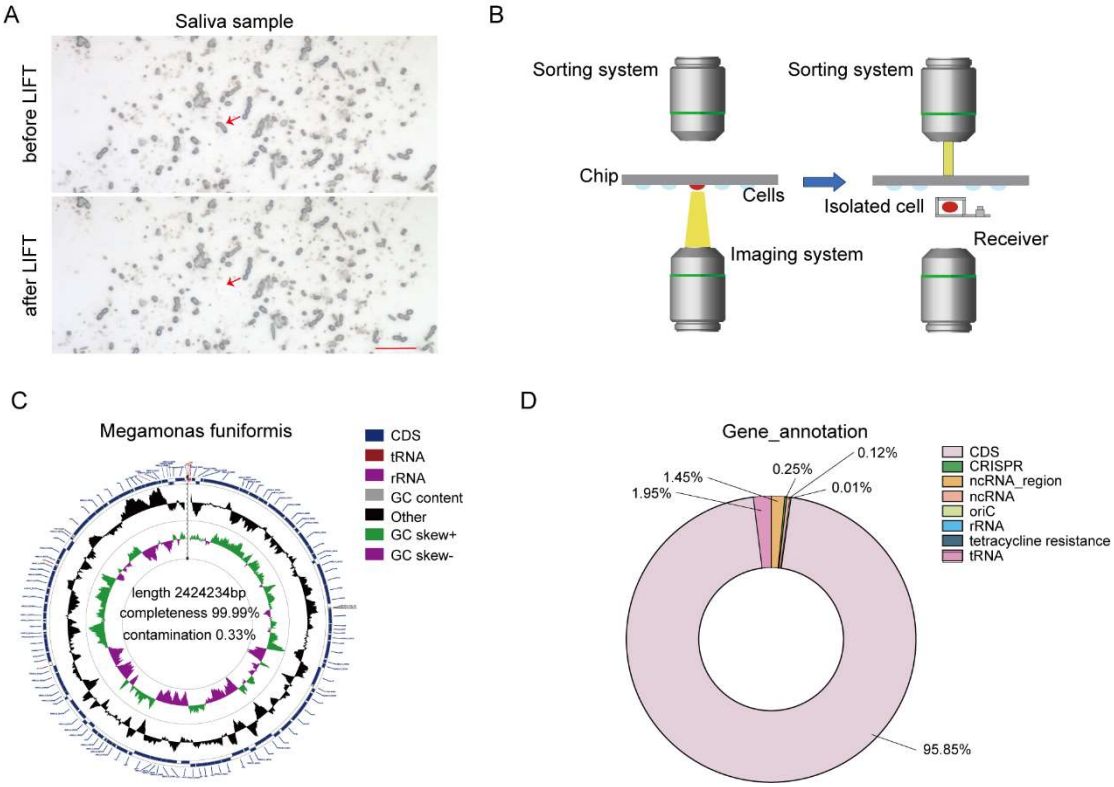

Figure.1 High-quality microbial single-cell draft genome obtained from saliva. (A) A view of the sample before and after ejection of the selected microbial cell. A single cell as indicated was selected for genomic sequencing. Scale bar 10μm. **(B)** Diagram illustrating the imaging by the inverted microscope and LIFT by the upright microscope, on the commercially available PRECI-SCS equipment. The collectors are placed on a wheel, whose rotation is controlled by the manufacturer's software (Multiple cells could also be transferred into the same collector, depending on the experimental design). **(C)** Visualization of single-cell assembled genome by CGView tool displaying GC skew, feature labels, GC content, divider rings, and tick density. **(D)** The pie chart of single-cell assembled genome annotation display different categories including coding sequence (CDS), CRISPR, ncRNA region, ncRNA, replication origin (oriC), rRNA, tetracycline resistance, tRNA with their respective proportions.

Fig.2

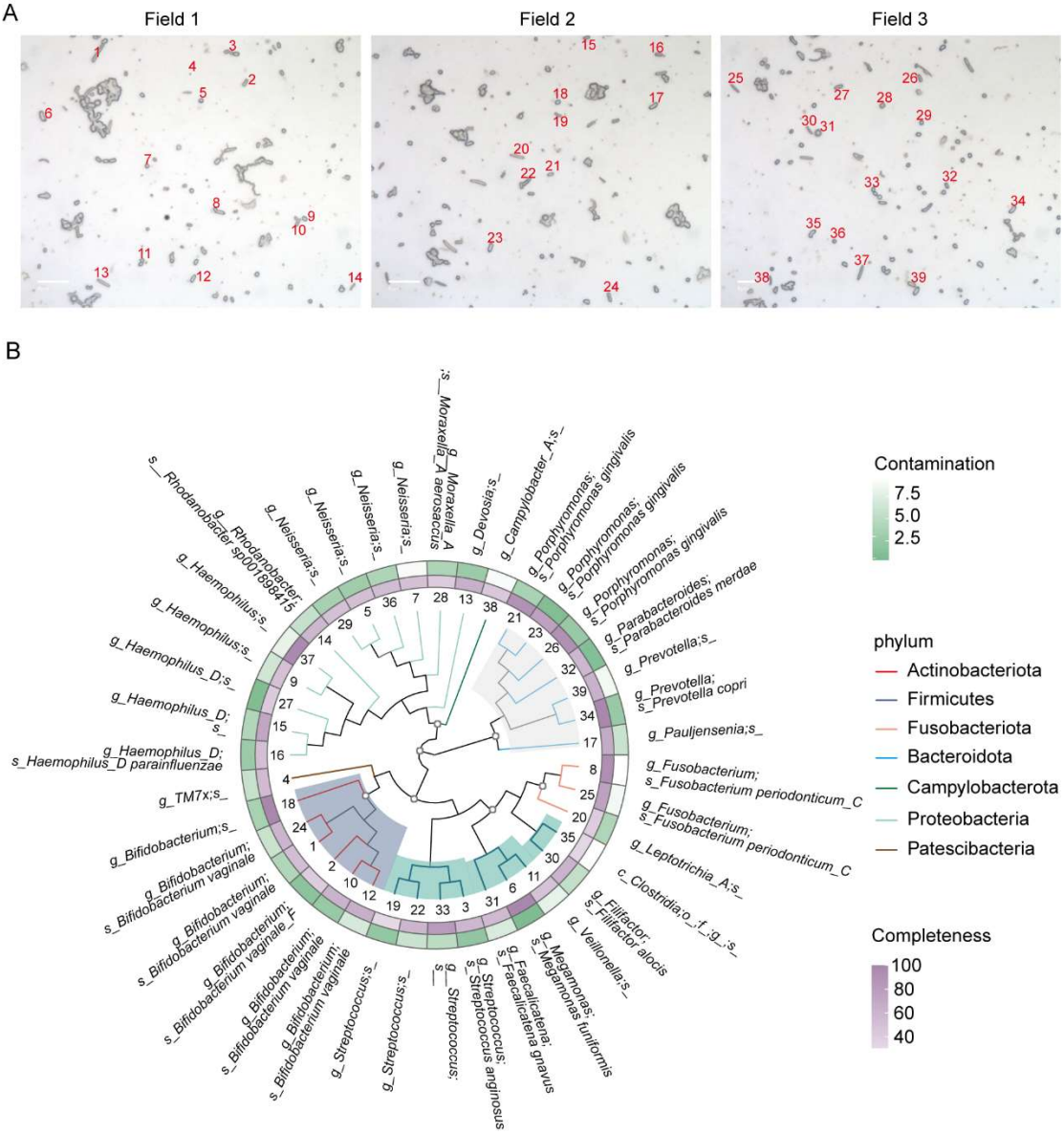

Figure.2 Automated microbial single-cell genomics. (A) Individual microbial cells of saliva distributed on the chip were observed before LIFT. The selected microbes numerically marked with variable morphology are ejected for single cell sequencing. **(B)** A phylogeny constructed from 39 obtained SAGs genome sequence is represented by the inner dendrogram. The phylum of each SAGs is indicated by the branch color. Numbers for the tip nodes of the dendrogram are corresponding to numbers labeled in panel **A**. Completeness and contamination of each SAGs are represented by two heatmaps overlay on the phylogenetic tree. The species name of each SAGs is listed in the outermost ring.

Fig.3

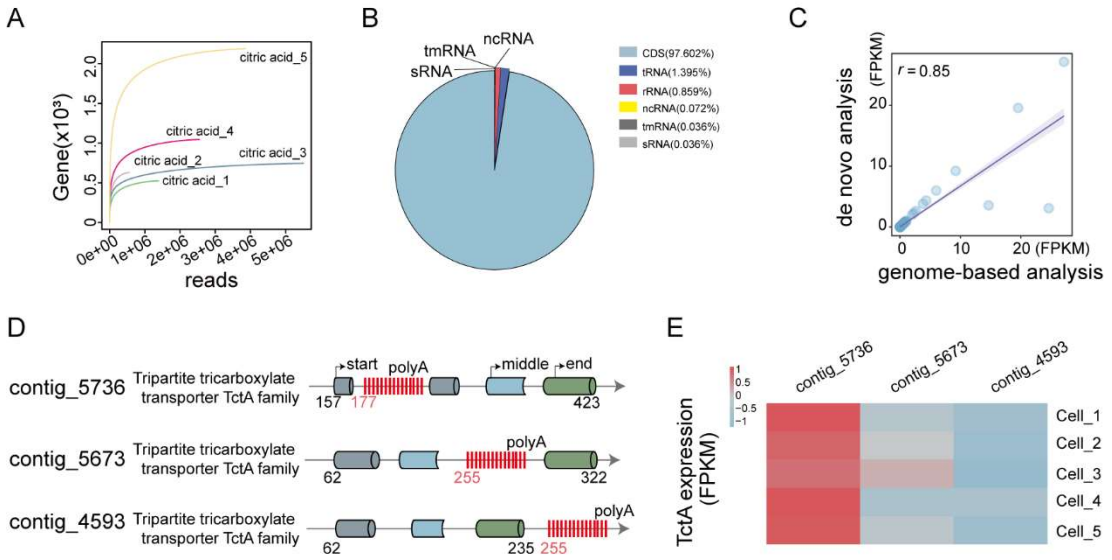

Figure.3 Single-cell transcriptome of sporulating *Bacillus licheniformis*. **(A)** Rarefaction curve of samples treated with (citric acid_1/2/3/4/5) citric acid indicated sufficient sequencing depth. The mapped genes selected for rarefaction curve were protein coding genes against reference genome, which varied from less to more in each sample. **(B)** Gene type annotation of a representative sample in pie chart. **(C)** Scatterplot of gene expression correlation in one of three representative samples by two methods: de novo analysis and genome-based analysis. Pearson's correlation in all samples r > 0.5. **(D)** Illustration for oligoA detection: Polyadenylation (15 adenylate residues) in TctA gene located in three contigs from de novo assembly genome. Numbers colored black are gene start and end positions. TctA gene in each contig is split evenly into three parts according to gene total length: start part, middle part and end part. Numbers colored red indicate start positions of the polyA tracts (15 adenylate residues) according to the de novo assembly. **(E)** Gene expression heatmap (FPKM) of TctA with different polyA site in five scRNA samples treated with citric acid.

Fig4

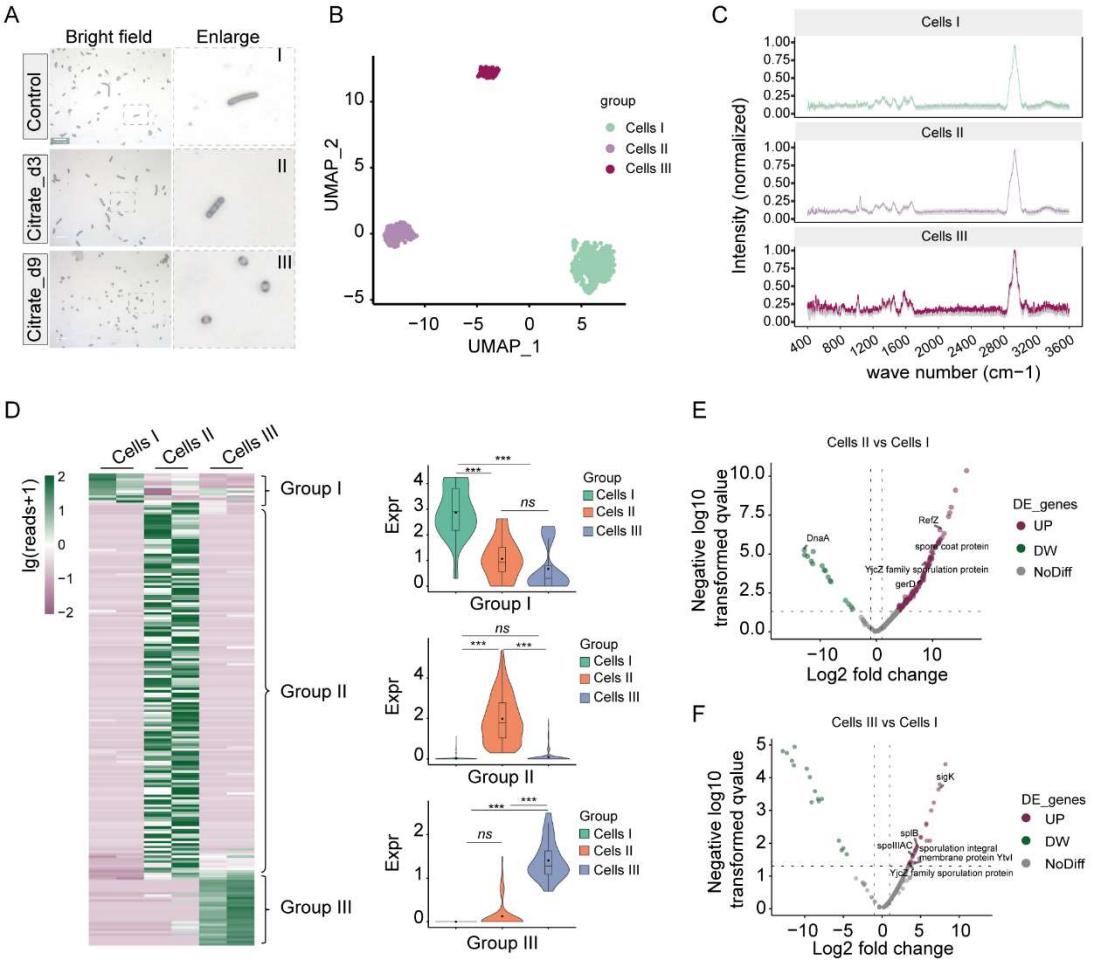

Figure.4 Coherent grouping of *B. licheniformis* cells visually, spectrally and with scRNA transcriptome. (A) Individual *B. licheniformis* cells distributed on the chip were observed in bright field before Raman spectrometry and LIFT. Cells I from *B. licheniformis* treated without citric acid (control), Cells II from *B. licheniformis* treated with citric acid for 3 days and Cells III from *B. licheniformis* treated with citric acid for 9 days are ejected for scRNA sequencing. **(B)** More than 50 cells from each group (Cells I, Cells II and Cells III) are selected for Raman spectra acquisition. Dimensionality reduction is performed on the obtained Raman spectra by UMAP. **(C)** The means and standard deviations of Raman spectra were calculated for each group. **(D)** Heatmap of differentially expressed transcripts in Cells I, Cells II and Cells III. The candidate transcripts are mainly clustered into three groups (left panel). Violin plot of normalized transcripts expression in Group I, Group II and Group III (right panel). Data are represented as mean ± SE (ns: $P > 0.05$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). **(E)** Volcano plot of differentially expressed genes of Cells II versus Cells I. Gene counts normalized and compared by DeSeq2. **(F)** Volcano plot of differentially expressed genes of Cells III versus Cells I. Gene counts normalized and compared by DeSeq2.

Fig.5

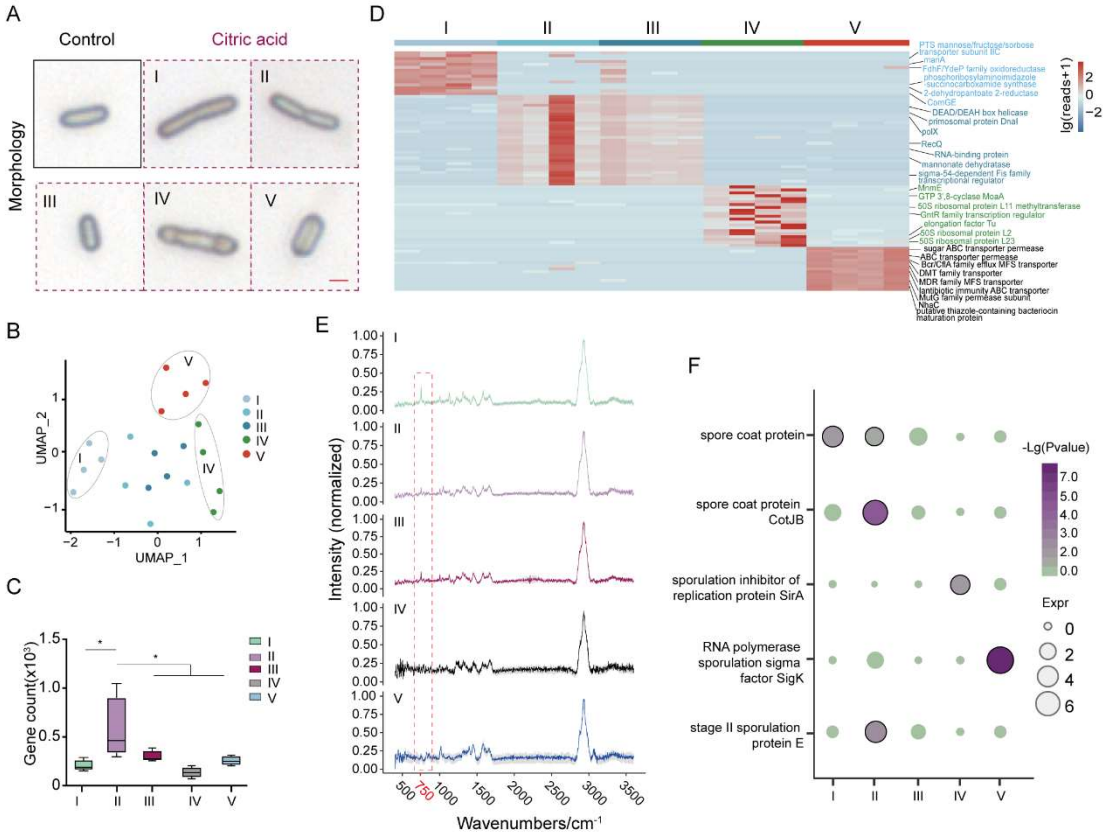

Figure.5 Single-cell transcriptome reveal heterogeneity in sporulation. (**A**) Image of five type *B. licheniformis* cells in sample treated with citric acid for 3 days and cells without citric acid treatment as control. Scale bar, 1μm. (**B**) UMAP two-dimensional representation of gene expression profile from the five cell types. **(C)** Calculation of gene counts in each cell type. Data are represented as mean ± SE (**P* < 0.05, ***P* < 0.01, ****P* < 0.001). **(D)** Heatmap of representative gene expression (z-score of log-transformed values) from 20 individual *B. licheniformis* cells grown in day 3 citric acid-treated media. (**E**) Raman spectra of five type *B. licheniformis* cells. The spectra are averages of ten individual spectra and normalized to the intensity of the strongest feature in each spectrum. **(F)** Dot plot of genes involved in *B. licheniformis* sporulation. Gene expression and pvalue were calculated between the selected cell type and the remaining cell type.

Fig.6

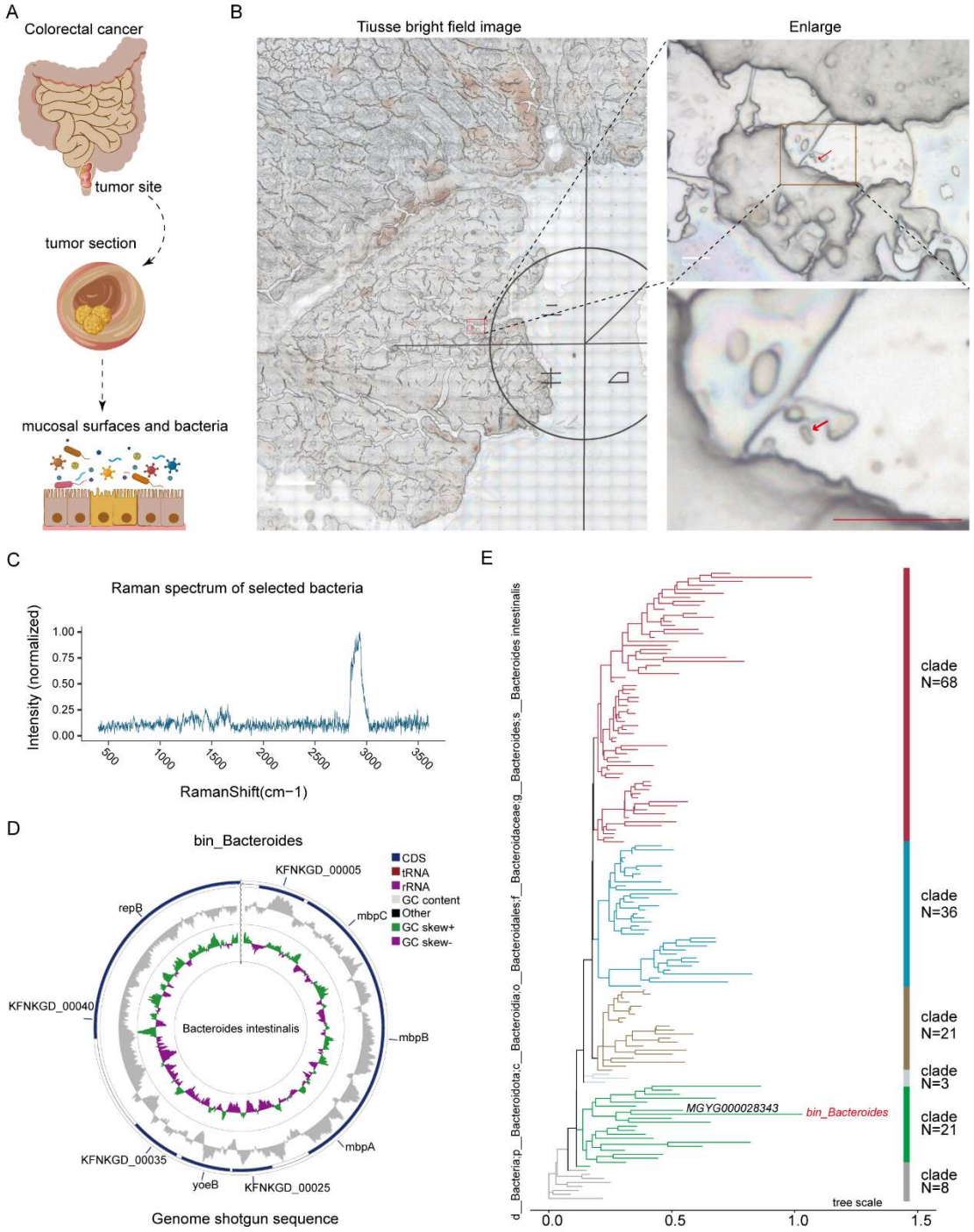

Figure.6 Genomic sequencing for single microbe in tissue. (A) Diagram for the colorectal cancer sample. **(B)** Tissue section containing the microbial single-cell ejected and genome sequenced. Scale bar, 10μm. **(C)** Raman spectrum of the indicated single cell. **(D)** Visualization of obtained SAG (bin Bacteroides) by CGView tool displaying GC skew, feature labels, GC content, divider rings, and tick density. **(E)** Phylogenetic tree was constructed by obtained SAG (bin Bacteroides) and all Bacteroides intestinalis genomes

downloaded from metadata-gut of Mash database. Among Bacteroides intestinalis genomes from Mash database, genome *MGYG000028343* was isolated from China.

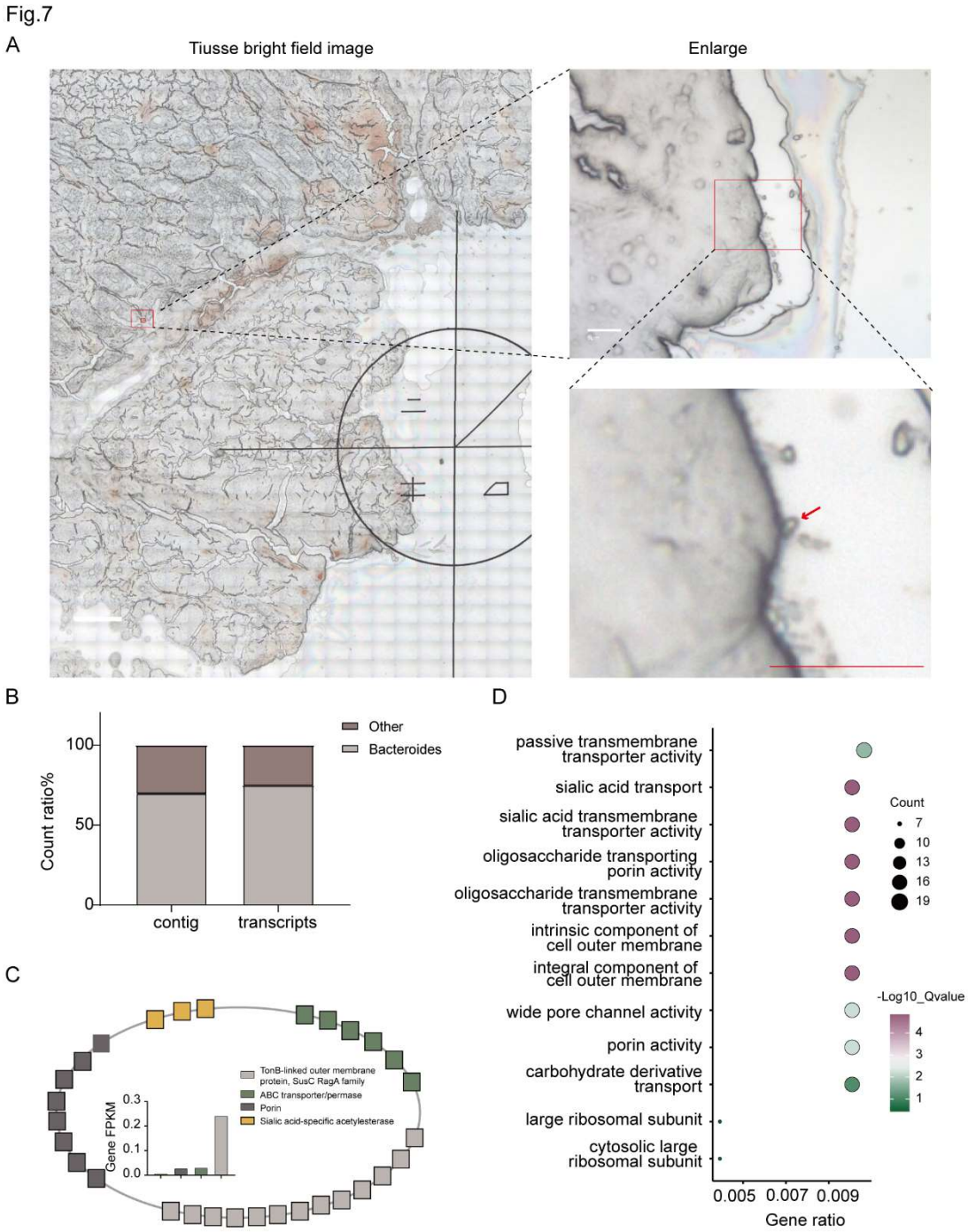

Figure.7 Transcriptome sequencing for single microbes from tissue. (A) Tissue section as in Figure 6, with inlets indicating the position of the bacteria for scRNA analyses. Scale bar, 10μm. **(B)** Count ratio of contigs and transcripts

in single cell assembled transcriptome. Contig classification was annotated by Kraken2. Transcript annotation was referred against de novo transcriptome annotated by eggnog-mapper software. **(C)** The detection of proteins located on cell membrane. Gene FPKM was calculated by Bakta. **(D)** Gene ontology analysis of transcripts classified to Bacteroidetes.