



OPEN ACCESS

Original research

Multimomics analysis reveals the presence of a microbiome in the gut of fetal lambs

Yanliang Bi ,¹ Yan Tu,¹ Naifeng Zhang,¹ Shiqing Wang,¹ Fan Zhang,² Garret Suen,³ Dafu Shao,⁴ Shengli Li,⁵ Qiyu Diao¹

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2020-320951>).

¹Feed Research Institute, Chinese Academy of Agricultural Sciences, National Engineering Research Center of Biological Feed, Beijing, China

²State Key Laboratory of Animal Nutrition, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China

³Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin, USA

⁴Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China

⁵State Key Laboratory of Animal Nutrition, College of Animal Science and Technology, China Agricultural University, Beijing, China

Correspondence to

Professor Qiyu Diao, Feed Research Institute, Chinese Academy of Agricultural Sciences, Beijing, China; diaoqiyu@caas.cn

YT, SL and QD contributed equally.

Received 24 February 2020

Revised 8 January 2021

Accepted 13 January 2021

Published Online First

15 February 2021

ABSTRACT

Objective Microbial exposure is critical to neonatal and infant development, growth and immunity. However, whether a microbiome is present in the fetal gut prior to birth remains debated. In this study, lambs delivered by aseptic hysterectomy at full term were used as an animal model to investigate the presence of a microbiome in the prenatal gut using a multimomics approach.

Design Lambs were euthanised immediately after aseptic caesarean section and their cecal content and umbilical cord blood samples were aseptically acquired. Cecal content samples were assessed using metagenomic and metatranscriptomic sequencing to characterise any existing microbiome. Both sample types were analysed using metabolomics in order to detect microbial metabolites.

Results We detected a low-diversity and low-biomass microbiome in the prenatal fetal gut, which was mainly composed of bacteria belonging to the phyla Proteobacteria, Actinobacteria and Firmicutes. *Escherichia coli* was the most abundant species in the prenatal fetal gut. We also detected multiple microbial metabolites including short chain fatty acids, deoxynojirimycin, mitomycin and tobramycin, further indicating the presence of metabolically active microbiota. Additionally, bacteriophage phiX174 and Orf virus, as well as antibiotic resistance genes, were detected in the fetal gut, suggesting that bacteriophage, viruses and bacteria carrying antibiotic resistance genes can be transmitted from the mother to the fetus during the gestation period.

Conclusions This study provides strong evidence that the prenatal gut harbours a microbiome and that microbial colonisation of the fetal gut commences in utero.

INTRODUCTION

The gut microbial community plays critical roles in host health by providing metabolic products, maintaining metabolic function, developing the immune system and defending against pathogens.^{1–3} Mounting evidence suggests that dysfunction of the early life gut microbiota can result in immune and metabolic disorders, contributing to the development of non-communicable diseases and obesity.⁴ However, establishing effective means to reduce disease risk by manipulating early life host–microbe interactions is hindered by considerable gaps in knowledge on the colonisation time and origin of intestinal microbiota.⁵

Significance of this study

What is already known on this subject?

► It has long been assumed that the uterus is sterile and that establishment of the fetal microbiota commences at birth. Although microbes have been detected in the placenta, amniotic fluid, fetal membranes and meconium, there is no consensus as to the presence of microbes in the fetal gut prior to delivery.

What are the new findings?

► We found that the prenatal gut in lambs harbours a low-diversity and low-biomass microbiome, suggesting that microbial colonisation of the fetal gut commences in utero. We also found multiple microbial metabolites in the fetal gut, suggesting that the prenatal gut microbes are metabolically active. In addition, bacteriophage phiX174, Orf virus and expression of multiple genes involved in antibiotic resistance, were detected in the fetal gut.

How might it impact on clinical practice in the foreseeable future?

► Understanding the community composition and function of the fetal gut microbiome may help in the development of strategies to mitigate several diseases and guide the formation of health-promoting microbiota that could be beneficial for the development of healthy offspring.

Whether there exists colonising microbes in the fetal gut prior to delivery is still controversial. It has long been assumed that the fetus is sterile and that establishment of the fetal microbiota commences during the birthing process.⁶ Recently, this dogma has been challenged by a growing body of new data. Numerous studies have reported the presence of diverse microbes in samples of human infant meconium,^{7 8} placenta,^{9–11} amniotic fluid^{7 12} and umbilical cord blood¹³ obtained from healthy normal pregnancies using sequencing-based methods. This suggests that microbial seeding of the fetal gut may initiate in utero before delivery. However, the interpretation of these findings has raised concerns regarding the validity of these conclusions and has primarily focused on methodological challenges, contradictory results and our



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Bi Y, Tu Y, Zhang N, et al. *Gut* 2021;**70**:853–864.

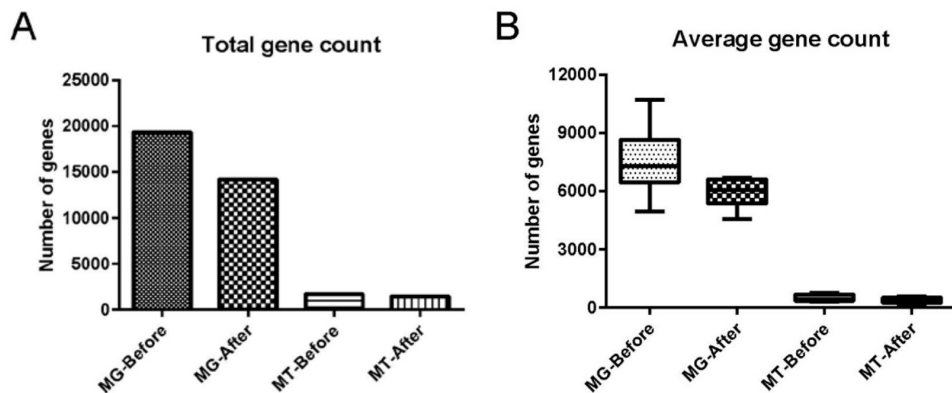


Figure 1 Effect of data decontamination on gut microbial metagenomic and metatranscriptomic gene counts. (A) Total gene count for microbial metagenome (MG) and metatranscriptome (MT) before and after data decontamination in samples of fetal cecal content. 'Before' indicates gene counts before data decontamination; 'after' indicates gene counts after data decontamination. (B) Average gene count in microbial MG and MT per cecal content sample before and after data decontamination. In (B), the boxes represent IQRs between the first and third quartiles, and the horizontal line inside the box indicates the median; whiskers represent the minima or maxima within 1.5× IQR from the first or third quartiles.

current immunological understanding.¹⁴ As a result, the concept of a 'sterile womb paradigm' has persisted, with many scientists contending that microbes detected in utero may be due to underlying contamination issues especially when working with low biomass samples.^{15–17} For example, several studies assert that microbes detected in the placenta were due to DNA present in laboratory reagents and equipment contaminants.^{18–20} Thus, it is essential to conclusively determine whether the prenatal gut harbours a microbiome, or whether the purported fetal gut microbiome is merely the result of contamination.

The aim of this study is to provide direct evidences in support of the existence of a microbiome in the prenatal gut. To accomplish this, we used lambs delivered via aseptic hysterectomy as an animal model. Lambs were euthanised immediately after birth via aseptic caesarean section and their cecal content and umbilical cord blood samples were collected. Cecal content samples were assessed using metagenomic and metatranscriptomic sequencing to investigate microbial composition and function, after which both sample types were analysed using metabolomics to detect the presence of microbial metabolites.

RESULTS

Removal of potential contaminating genes and transcripts from the metagenomic and metatranscriptomic datasets

To assess the microbiome of the six in utero lambs used in our study, we performed both metagenomic and metatranscriptomic sequencing of the cecal contents from our study animals. Given concerns from other studies of contaminating DNA/RNA in the reagents used to process samples we also included negative control where nucleic acid-free water without sample DNA was used. From the negative control, we performed a data decontamination step, as described in Methods and online supplemental figure S1.

Comparison of the metagenomic dataset (containing 19 320 non-redundant (nr) genes) with the negative control showed that cecal content samples possessed 14 116 unique genes, and that 5204 genes were shared between cecal content samples and negative control samples (online supplemental tables S1 and S2). Of these 5204 shared genes, 83 passed our filtering criteria and remained in our dataset whereas the 5121 remaining genes were discarded as potential contaminants. Comparison of the metatranscriptomic dataset (containing 1691 nr genes) with the negative control indicated that cecal content samples contained 1446

unique genes. Our filtering step identified 245 genes shared between cecal content samples and negative control (online supplemental tables S3 and S4), and of those, 10 genes passed our filtering criteria and thus remained in our dataset. After final quality assessment and data decontamination, the gut microbial metagenomic and metatranscriptomic datasets contained a total of 14 199 genes (5935 ± 711 genes per sample, expressed as mean \pm SD) and 1456 genes (358 ± 134 genes per sample) for the six cecal content samples, respectively (figure 1, online supplemental tables S5 and S6).

Taxonomic composition identified by metagenomic analysis

To identify microbial community structures of the fetal gut at a metagenomic level, we performed a metagenomic sequencing of cecal content samples obtained from six lambs delivered via aseptic hysterectomy. We aligned the metagenomic genes against the nr nucleotide database of National Center of Biotechnology Information (NCBI) for taxonomic assignment. Microbial taxa were defined as being present in at least three out of six samples. We found that the fetal gut of lambs showed low alpha diversity (online supplemental figure S2), indicating that bacterial richness in the fetal gut was low. Four major phyla were identified in cecal content samples: Proteobacteria ($95.30\% \pm 2.19\%$), Firmicutes ($4.85\% \pm 1.71\%$), Actinobacteria ($0.53\% \pm 0.22\%$) and Thaumarchaeota ($0.02\% \pm 0.01\%$) (figure 2A, online supplemental table S7). At the genus and species levels, 33 genera and 50 species were detected in cecal content samples. *Escherichia* ($88.76\% \pm 2.04\%$) and *Catelliboccus* ($4.19\% \pm 1.65\%$) were the main bacterial genera (figure 2B, online supplemental table S7), while *Escherichia coli* ($86.89\% \pm 2.21\%$) and *Catelliboccus marimammalium* ($4.11\% \pm 1.61\%$) were the most abundant species (figure 2C, online supplemental table S7). We also detected high levels of bacteriophage phiX174 ($52.29\% \pm 5.55\%$ of all mapped reads) and Orf virus ($0.03\% \pm 0.01\%$ of all mapped reads) in all cecal contents samples (online supplemental figure S3A) and 22 antibiotic resistance genes in the fetal gut (online supplemental table S8).

Taxonomic composition identified by metatranscriptomic analysis

To examine the potential activity of the detected microbes, we performed a metatranscriptomic analysis of RNA extracted

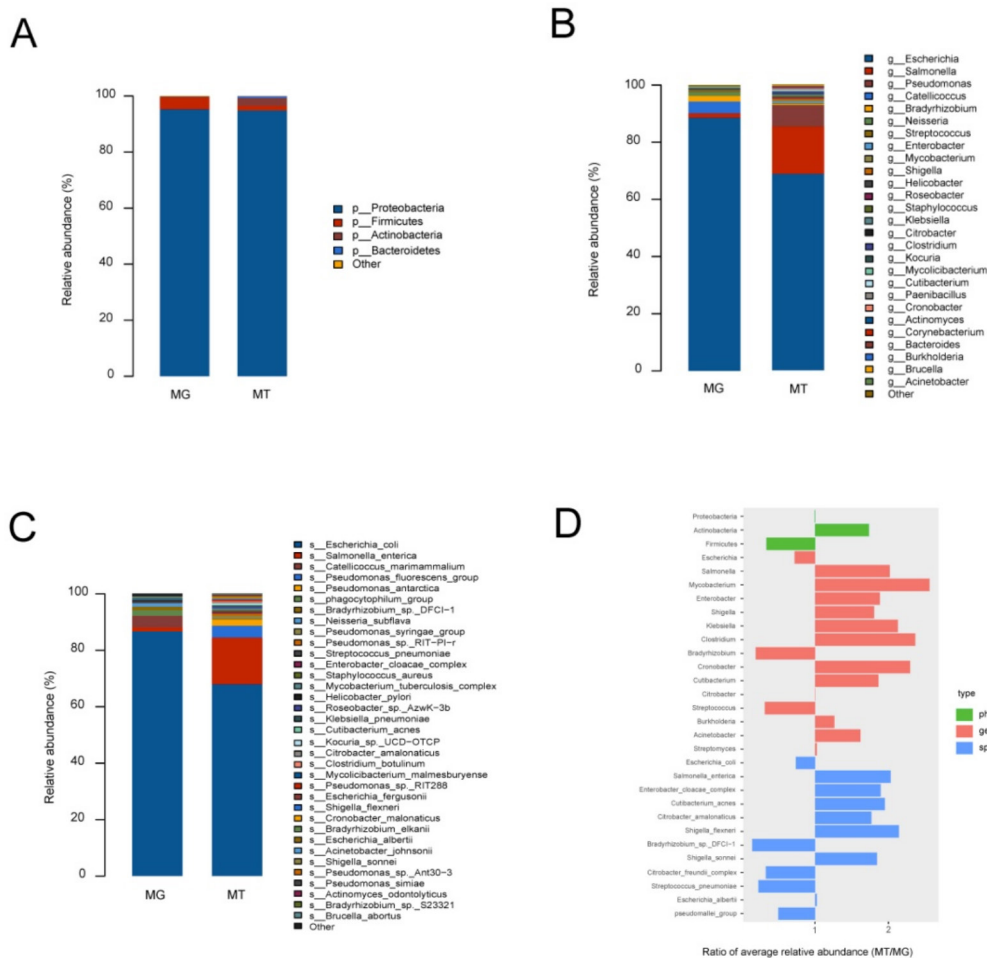


Figure 2 Microbiome composition of metagenome (MG) and metatranscriptome (MT) in fetal gut. (A–C) Represent phylum-level, genus-level and species-level bacterial composition of the fetal gut microbiome (relative abundance $\geq 0.1\%$), respectively. Microbiome composition shows average relative abundance of bacterial presence in all samples. $n=6$. (D) Ratios of average relative microorganism abundance in MT relative to the MG (MT/MG).

from cecal content samples. A taxonomic classification of the metatranscriptomic data revealed a total of bacterial phyla, 26 bacterial genera and 32 bacterial species. The four bacteria phyla included the Proteobacteria ($94.73\% \pm 3.66\%$), Actinobacteria ($2.75\% \pm 2.52\%$), Firmicutes ($1.69\% \pm 1.13\%$) and Bacteroidetes ($0.67\% \pm 0.62\%$) (figure 2A, online supplemental table S9). At the genus and species levels, *Escherichia* ($64.23\% \pm 19.30\%$) and *Salmonella* ($15.51\% \pm 6.84\%$) were the main bacterial genera (figure 2B, online supplemental table S9), while *E. coli* ($64.14\% \pm 19.17\%$) and *Salmonella enterica* ($15.60\% \pm 6.87\%$) were the most abundant species (figure 2C, online supplemental table S9). Metatranscriptomic analysis of the cecal content samples also indicated high levels of bacteriophage phiX174 ($58.88\% \pm 14.73\%$ of all mapped reads) (online supplemental figure S3B).

To infer relative microbial activity at the time of sample collection, we analysed microbial gene expression in the samples based on the ratio of microbial functional gene transcripts to genes.²¹ MT/MG ratios were defined as ratios of average relative microbe abundance in the metatranscriptomes, relative to those in the corresponding metagenomes.²¹ Members of the Actinobacteria, such as the genera *Mycobacterium* and *Cutibacterium*, showed increased MT/MG ratios (figure 2D). Members of the Firmicutes, such as the genus *Streptococcus*, exhibited decreased MT/MG ratios, but the MT/MG ratios of *Clostridium* were increased. The

MT/MG ratios of the Proteobacteria were near 1, but members belonging to this phylum varied greatly. For example, members of the genera *Escherichia* and *Citrobacter* complex showed decreased MT/MG ratios, whereas other microbes belonging to the Proteobacteria showed increased MT/MG ratios.

Quantification of microbes in samples of cecal content and in the controls

We performed absolute quantitative real-time PCR (qPCR) to quantify the copy numbers of total bacteria and five selected microbes (figure 3) showing a high relative abundance in our metagenomic data. Copy numbers per gram of total bacteria in cecal content samples (expressed as mean \pm SD) and negative controls were $4.6 \times 10^7 \pm 3.4 \times 10^7$ and $1.6 \times 10^7 \pm 1.1 \times 10^6$, respectively. Cecal content samples showed significantly higher ($p < 0.05$) copy numbers per gram of total bacteria than did negative controls. Copy numbers per gram of *E. coli*, *C. marimammillum*, *Streptococcus pneumoniae* and *Saccharomyces cerevisiae* were significantly higher ($p < 0.05$) in cecal content samples than those in the negative controls, whereas copy numbers per gram of *Bradyrhizobium* did not differ significantly between these samples.

Functional characterisation of the cecal content microbiome

To functionally characterise the cecal microbiome, we aligned nr gene catalogues of the metagenomes and metatranscriptomes

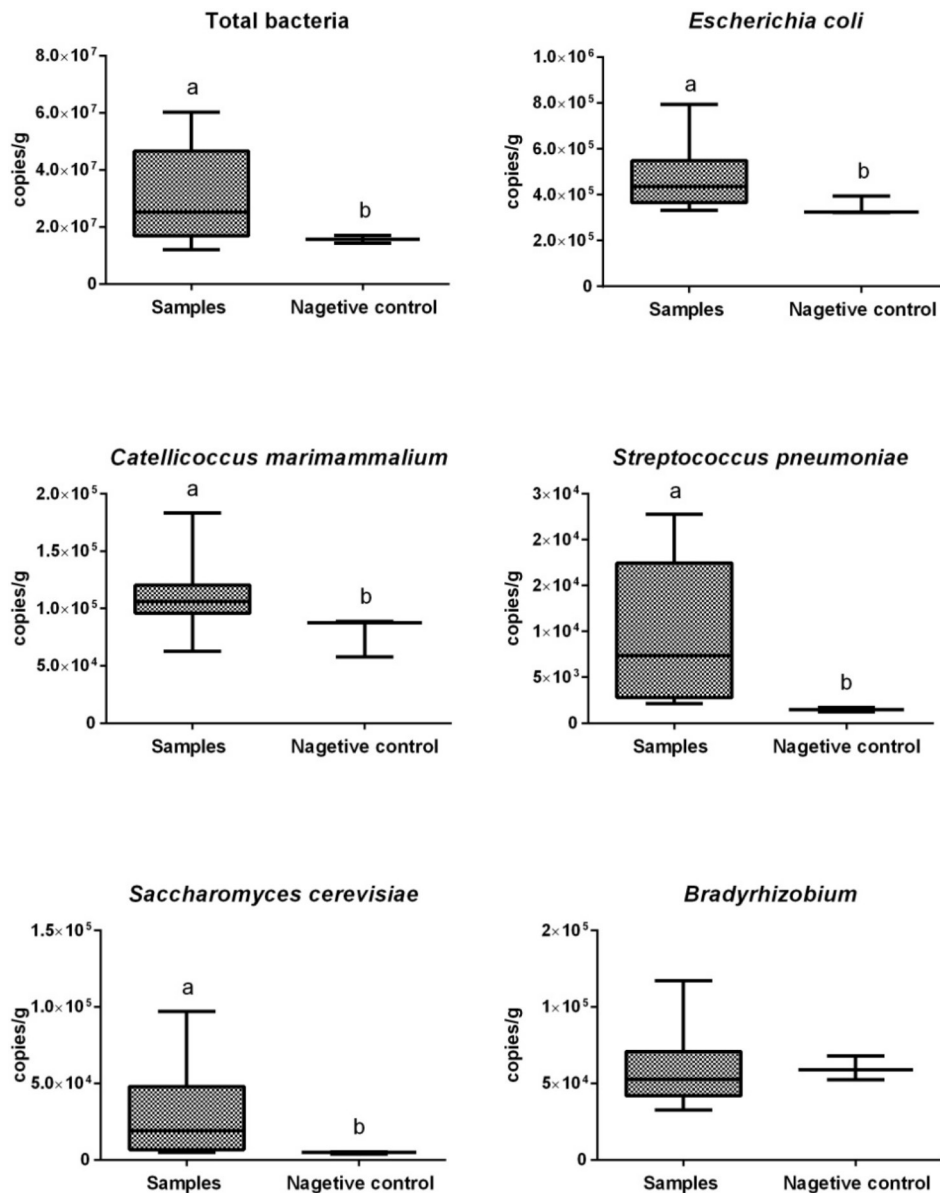


Figure 3 Copy numbers of total bacteria and selected microbes per gram cecal content. The boxes represent IQRs between the first and third quartiles, and the horizontal line inside the box indicates the median; whiskers represent the minima or maxima within $1.5 \times$ IQR from the first or third quartiles. Boxes with different letters above their whiskers are significantly different at $p < 0.05$.

against the Kyoto Encyclopaedia of Genes and Genomes (KEGG) and evolutionary genealogy of genes: non-supervised orthologous groups (eggNOG) databases. Of the metagenomic genes, 5458 and 8749 genes were annotated using KEGG and eggNOG, respectively. KEGG pathways belonging to 'Signal transduction', 'Carbohydrate metabolism', 'Amino acid metabolism' and 'Energy metabolism' were highly enriched in the microbiome of fetal gut (figure 4A). Similarly, microbial genes involved in 'Amino acid transport and metabolism', 'Carbohydrate transport and metabolism', 'Energy production and conversion' and 'Signal transduction mechanisms' pathways were highly enriched in eggNOG functional categories. Furthermore, pathways involved in 'Transcription', 'Post-translational modification, protein turnover, chaperones' and 'Intracellular trafficking, secretion and vesicular transport' were also highly enriched in eggNOG functional categories (figure 4C).

For the metatranscriptomic data, 912 and 1168 genes were annotated using KEGG and eggNOG, respectively. KEGG pathways involving the 'Carbohydrate metabolism', 'Energy metabolism', 'Amino acid metabolism', 'Membrane transport' and 'Signal transduction' were highly enriched in the microbiome of fetal gut (figure 4B). The most active eggNOG functional categories in the fetal gut were 'energy production and conversion', 'amino acid transport and metabolism' and 'carbohydrate transport and metabolism', followed by 'translation, ribosomal structure and biogenesis' and 'cell wall/membrane/envelope biogenesis' (figure 4D).

Metabolomic profiling of cecal content and umbilical cord blood

To further analyse microbial activity in the fetal gut, we conducted a metabolomic analysis to detect microbial metabolites in our samples. The concentrations of short chain fatty

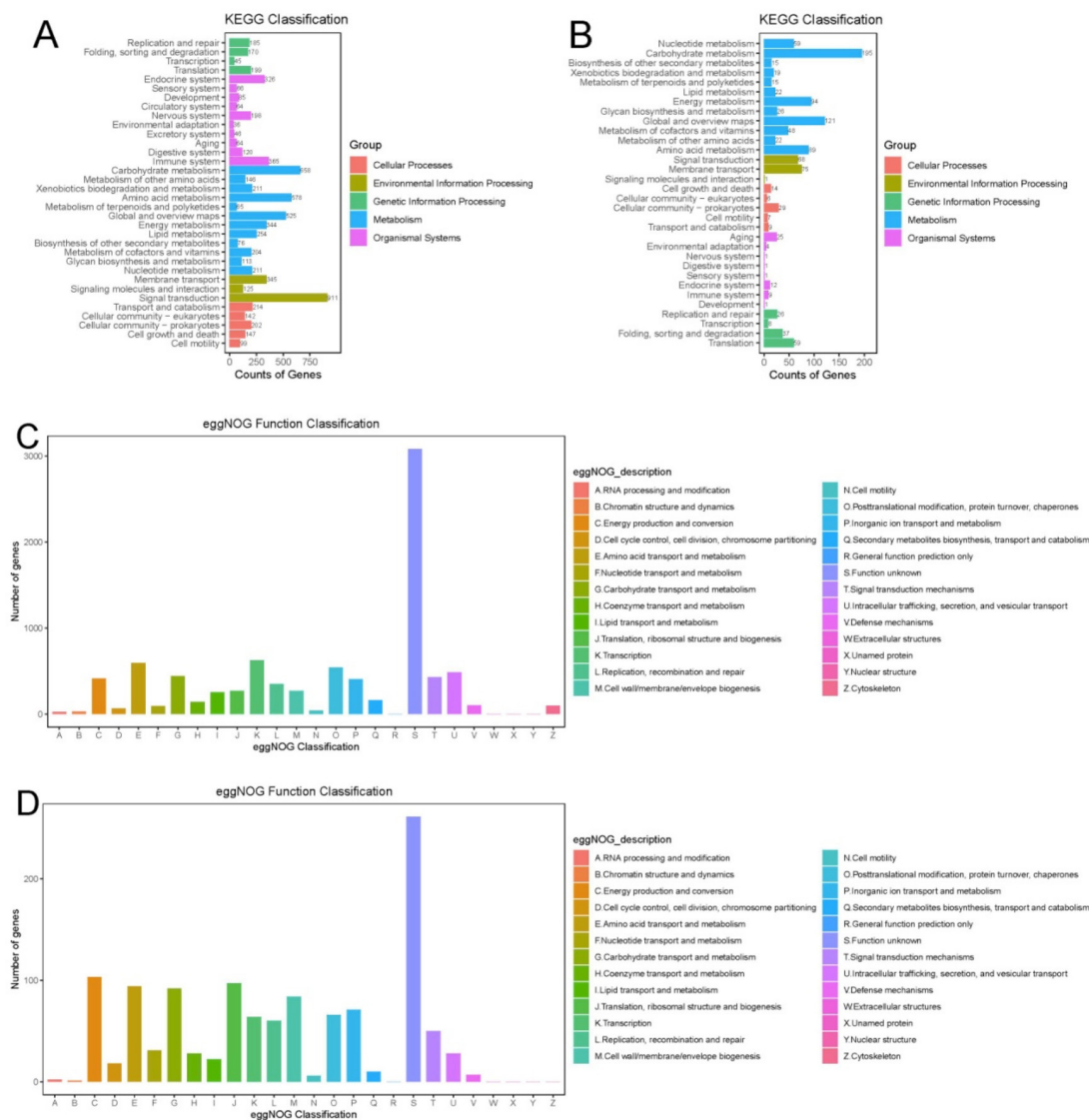


Figure 4 KEGG and eggNOG classifications of the fetal gut metagenome and metatranscriptome. (A, B) KEGG classifications for the fetal gut metagenome and metatranscriptome, respectively. (C, D) eggNOG functional classifications for the fetal gut metagenome and metatranscriptome, respectively. eggNOG, non-supervised orthologous groups; KEGG, Kyoto Encyclopaedia of Genes and Genomes.

acids (SCFAs) were measured and used as a proxy for microbial metabolic activity in the gut. For this study, we used a detection limit of 5×10^{-5} $\mu\text{g}/\text{mg}$ for acetic, propionic and isobutyric acids and a detection limit of 5×10^{-6} $\mu\text{g}/\text{mg}$ for butyric, isovaleric, valeric and hexanoic acids. This range allowed for the detection of all SCFAs in our cecal content samples, except for propionic, isobutyric and isovaleric acids. The concentrations of acetic, butyric, valeric and hexanoic acids were $1.88 \times 10^{-3} \pm 4.05 \times 10^{-4}$, $1.01 \times 10^{-4} \pm 2.43 \times 10^{-5}$, $2.71 \times 10^{-5} \pm 4.56 \times 10^{-6}$ and $1.02 \times 10^{-4} \pm 3.01 \times 10^{-5}$ $\mu\text{g}/\text{mg}$, respectively. To trace the possible sources of these SCFAs in the gut, we examined SCFAs in umbilical cord blood. The concentrations of acetic, propionic, butyric, isobutyric, valeric, isovaleric and hexanoic acids in umbilical cord blood were 6.23 ± 2.94 , 1.69 ± 0.55 , 0.72 ± 0.24 , 0.56 ± 0.19 , 0.24 ± 0.09 , 0.04 ± 0.03 and 0.41 ± 0.24 $\mu\text{g}/\text{mL}$, respectively (online supplemental table S10).

We also performed non-targeted metabolomic profiling using high-performance liquid chromatography-quadrupole time-of-flight-mass spectrometry (UHPLC-QTOF-MS) to detect other metabolites in our samples. In total, 2808 peaks were obtained

from cecal content samples, including 1271 and 1537 peaks in positive ion mode (ESI+) and negative ion mode (ESI-), respectively. These peaks were aligned using Human Metabolome Database (HMDB), PubChem and the KEGG databases and we were able to annotate 1990 total peaks (937 peaks in ESI+ and 1053 peaks in ESI-) (online supplemental tables S11 and S12). Of these 1990 metabolites, 151 were annotated to 180 KEGG pathways, whereas the other 1839 metabolites were not assigned to a specific pathway. These KEGG pathways were then divided into host-associated (ie, *Ovis aries*) pathways and non-host associated pathways (ie, all KEGG pathways not shared with *Ovis aries* KEGG pathways). Of these 151 total metabolites, 143 were involved in 127 host-related KEGG pathways (online supplemental table S13), and 77 were involved in 53 non-host microbially related KEGG pathways (online supplemental table S14). The 53 non-host microbially related metabolic pathways contained eight unique metabolites including deoxyojirimycin, mitomycin, tobramycin, p-benzoquinone, cyclohexylsulfamate, daidzein, chlorogenate and hesperetin.

We note that all 8 of these metabolites are associated with known microbial metabolic pathways. Specifically, deoxynojirimycin, mitomycin and tobramycin are involved in ‘biosynthesis of antibiotics’, tobramycin is involved in ‘neomycin, kanamycin and gentamicin biosynthesis’, p-benzoquinone and cyclohexylsulfamate are involved in ‘microbial metabolism in diverse environments’, daidzein and chlorogenate are involved in ‘Biosynthesis of secondary metabolites’ and hesperetin is involved in ‘flavonoid biosynthesis’ (online supplemental table S14). To trace the possible sources of these microbial metabolites in the gut, we performed non-targeted metabolomic profiling to detect metabolites in umbilical cord blood, but did not find the presence of deoxynojirimycin, mitomycin or tobramycin (online supplemental tables S15 and S16).

Associations between cecal microbiome and cecal content metabolites

To identify the correlations between cecal microbiome and cecal content metabolites, we performed co-occurrence network and heat map analyses using Spearman’s rank correlation coefficient. In the co-occurrence network and heat map, butyric and hexanoic acids concentrations showed positive correlations with the relative abundances of *S. pneumoniae* and *Kocuria* spp. UCD.OTCP (Spearman’s correlation value >0.6 , $p < 0.05$; figure 5A,B). Furthermore, hexanoic acid concentration was positively associated with the relative abundance of *Klebsiella pneumoniae*. Norethindrone concentration was also found to be positively correlated with the relative abundances of *Klebsiella pneumoniae* and *S. pneumoniae*, but negatively correlated with that of *Salmonella enterica*. In addition to the correlations identified between cecal microbiome and cecal content metabolites, we also found that the cecal content metagenome, macrotranscriptome and metabolome shared numerous KEGG pathways. Of the 376 and 198 KEGG pathways in our metagenomic and metatranscriptomic data, respectively, 152 and 95 KEGG pathways, respectively, were shared with the metabolome (online supplemental tables S17 and S18).

DISCUSSION

In this study, we used lambs delivered by aseptic hysterectomy and multi-omics based metagenomic, metatranscriptomic and metabolomic analyses to investigate the presence of a microbiome in the prenatal fetal gut. In order to ensure the accuracy of our results, we undertook many measures to eliminate the risk of contamination from environmental sources and handling. First, the lambs used in our animal model were delivered via aseptic hysterectomy, thereby preventing microbial contamination from the maternal vaginas and faeces. Second, the euthanasia, open surgery and collection of cecal content samples from these lambs, as well as total DNA and RNA extractions, were performed in strictly controlled, decontaminated and sterilised biosafety cabinets to avoid potential environmental contamination. Importantly, the negative control was performed and treated identically during sample collection, extraction of genetic material, sequencing and data analysis to eliminate (via data filtration) potential contamination from DNA and RNA found in extraction kits and other laboratory reagents and equipment.^{15 22}

After potentially-contaminating microbial genes were removed from our metagenomic and metatranscriptomic sequencing datasets, a diverse and metabolically active microbiota was detected in the prenatal fetal gut. The fetal gut microbiome was predominantly composed of Proteobacteria, Actinobacteria



Figure 5 Associations of cecal microbiome with cecal content metabolites. (A) Analysis of co-occurrence networks between cecal content metabolites and microbial species in the fetal gut metatranscriptome. Nodes represent metabolites and microbial species. An increase in the number of lines through a node increases the size of a node. Red edges indicate positive correlations between metabolites and microbial species; Spearman’s rank correlation coefficient >0.6 , $p < 0.05$. Blue edges indicate negative correlations between metabolites and microbial species; Spearman’s rank correlation coefficient < -0.6 , $p < 0.05$. (B) Heat map of Spearman’s rank correlation coefficients between cecal content metabolites and microbial species in the fetal gut metatranscriptome. Spearman’s rank correlation coefficient >0.6 or < -0.6 , $*p < 0.05$, $**p < 0.01$. Red and blue colours denote positive and negative correlations, respectively. Colour intensity is proportional to Spearman’s rank correlation values.

and Firmicutes, with Proteobacteria showing the highest abundance and accounting for more than 90% of all sequences in the different samples. *E. coli* showed the highest relative abundance in all the samples, accounting for more than 86% and 64% of the total sequences in our metagenomic and metatranscriptomic sequencing datasets, respectively. Additionally, other species, such as *C. marimammalium*, *Salmonella enterica*, *S. pneumoniae*, *Enterobacter cloacae*, *Escherichia albertii* and *Shigella sonnei*, were detected in the fetal gut. These bacterial taxa are similar to previous reports from humans, particularly in term and preterm placentas without intraamniotic infection.^{9 11} These observations suggest that bacterial colonisation is consistent across species.

Numerous studies have also identified microbes using non-sequencing methods in utero in the absence of immunopathology.

For example, intact placental microbes were visualised and localised by 16S rRNA in situ hybridisation and traditional histological techniques in term and preterm placentas without intraamniotic infection¹¹; gram-positive and—negative intracellular bacteria were identified in the basal plate of human placenta without clinical or histologic evidence of chorioamnionitis based on morphological evidence²³; probiotic bacterial strains administered to mothers before delivery were detected using qPCR in all placenta of full-term infants delivered by caesarean section.²⁴ Our observations are consistent with these studies, suggesting that seeding of the microbiome likely commences in utero prior to delivery.

Importantly, our qPCR results showed that the copy numbers per gram of total bacteria, *E. coli*, *C. marimammalium*, *S. pneumoniae* and *S. cerevisiae* were significantly higher in the cecal content samples than in our negative control, further providing evidence for the presence of a microbiome in the prenatal fetal gut. However, the copy numbers per gram of these microbes in the cecal content were low, with an order of magnitude of 10^3 – 10^5 . Compared with our previous reports on the gut microbiota of a 3-day-old lambs, the fetal gut of in utero lambs contained significantly lower bacterial diversity and biomass,²⁵ indicating that bacterial richness and biomass were low in the prenatal fetal gut. Our findings further support several previously published reports,^{5 9 11} which used metagenomic sequencing to demonstrate the existence of low-abundance, low-biomass and sparse microbial communities in utero.

In addition to our transcriptomics analysis, which provides support that the microbiome present in the prenatal fetal gut are active, we also conducted a metabolomics analysis. We detected several microbial metabolites including SCFAs, deoxynojirimycin, mitomycin and tobramycin in fetal cecal content samples. SCFAs are the major end products of microbial fermentation in the gut and act as a proxy for inferring microbial metabolic activity.²⁶ Previous studies have found SCFAs, such as acetic, propionic and butyric acids, in the infant first-pass meconium,^{7 27} which is consistent with our findings. Nojirimycin, a precursor to deoxynojirimycin, is known to inhibit various microbial glucosidases²⁸ and is produced by several strains of *Streptomyces*. This secondary compound has robust biological activity against several drug-resistant bacterial strains including *Sarcina lutea*, *Shigella flexneri* and *Xanthomonas oryzae*.^{29 30} Mitomycins, which are important antibiotics also produced by members of the *Streptomyces*, show reduced toxicity and broad spectrum activity against tumours and thus are widely used in the clinic.³¹ Tobramycin is an aminoglycoside antibiotic derived from nebramycin, a broad-spectrum antibiotic also produced by members of the *Streptomyces*.³² It shows excellent antibacterial activity against *Pseudomonas*, *Staphylococcus aureus* and most of the *Enterobacteriaceae*.³³ Our metabolite analysis using KEGG revealed that deoxynojirimycin, mitomycin and tobramycin are involved in the pathway ‘biosynthesis of antibiotics’. Tobramycin is also involved in ‘neomycin, kanamycin and gentamicin biosynthesis’. In this study, we detected sequences belonging to the *Streptomyces* in both metagenomic and metatranscriptomic data from fetal cecal contents and it is possible that the presence of the *Streptomyces* and the metabolites deoxynojirimycin, mitomycin and tobramycin is not coincidental. One hypothesis is that members of the *Streptomyces* present in the fetal microbiome are producing these secondary compounds as a way to outcompete other microbes in the community.

In order to trace the possible sources of these microbial metabolites in the fetal gut, we also investigated metabolites present in umbilical cord blood. The developing fetus is supported by

the umbilical cord in mammals, through which almost all nutrients, excretion products, respiratory gases and xenobiotics are passed.³⁴ In this study, we detected the presence of SCFAs in umbilical cord blood, but it is unknown if SCFAs in the fetal gut are from the fetal gut microbiome or from the ewe. It is possible that these SCFAs are the product of the gut microbiota of the ewe that are transferred from the gut, across the placenta and into the fetus.³⁴ It is important to note that we did not detect the presence of deoxynojirimycin, mitomycin and tobramycin in the umbilical cord blood, which were present in the fetal gut. Although with detected antibiotics, they could be from the bacteria locally or transmitted previously from mothers. These data further demonstrated that the prenatal gut not only harbours the presence of a microbiome, but that it may also be metabolically active.

The available evidence suggests that microbial colonisation of the fetal gut is likely established by ascension from the vagina or through the hematogenous spread of the maternal gut or oral microbiome. It is well established that certain vaginal microbes can ascend into the uterus and invade the amniotic cavity during the pregnancy.^{35–37} The amniotic fluid contains microbes that are ingested by the fetus during pregnancy, which ultimately colonise the fetal gut.^{38 39} Microbial DNA orally ingested by pregnant mice was discovered in various organs of fetuses, and a transplacental pathway was proposed as the route of transmission.⁴⁰ Genetically labelled bacteria introduced into the gut of pregnant mice were detected using culture-based and PCR-based methods in the placenta and meconium of healthy neonates born by sterile caesarean section.^{13 41} The possible mechanism of microbial transport from the maternal gut to the developing fetus is bacterial translocation from the maternal gut epithelium into the bloodstream followed by colonisation in the placenta.⁴² Several species of microbiota identified in the placenta with DNA-based technology have been found in the maternal oral cavity, but not in the maternal gut and vagina.^{43–45} Moreover, significant similarities between human oral and placenta microbial communities were previously reported,⁹ in which the placental microbiota was more similar to the oral microbiota than that in any other body site (including the vagina and gut). Some of these oral microbes such as *Fusobacterium nucleatum* can bind to the vascular endothelium and increase permeability, allowing other microbial species such as *E. coli* to cross endothelial barriers, facilitating hematogenous transmission during the placentation.⁴⁶

Prenatal seeding of a metabolically active microbiome may be clinically important for the developing fetus. First, several members of the microbiota or microbial metabolites may drive immune system development and intestinal epithelium maturation in the fetus. For example, the probiotics *Bifidobacterium lactis* and *Lactobacillus rhamnosus* GG, when supplemented to pregnant mothers, were detected in the placenta and amniotic fluid, and the expression of Toll-like receptor-related genes in the fetal gut were significantly modulated, suggesting that microbial contact in utero was associated with changes in fetal intestinal innate immune development.²⁴ The inflammatory cytokines levels of the amniotic fluid were associated with the microbiome presence within the amniotic fluid, which may influence the developing fetal immune system.⁷ In an experimental model of transient gestational colonisation, germ-free pregnant female mice were treated with a transiently colonising live *E. coli* strain that could not replicate in vivo.³ Compared with the pups born to the control germ-free mothers, germ-free pups born to transiently colonised mothers harboured increased numbers of intestinal group 3 innate lymphoid cells and F4/80⁺CD11c⁺ mononuclear cells. Maternal colonisation of *E. coli* also reshaped

the intestinal transcriptional profiles in the pups, including the increased expression of antibacterial peptides and increased cell division and differentiation of epithelial cells.³ In this study, we detected high levels of *E. coli* in the fetal gut, suggesting that fetuses were exposed to bacterial-derived immune stimulation prior to birth. Thus, the *E. coli* identified in our study may be involved in supporting fetal immune system development in utero. Microbial metabolites SCFAs serve as important energy sources and signalling molecules and are involved in the augmentation of energy metabolism, maintenance of the immune system, induction of reactive oxygen species, modulation of chemotaxis and phagocytosis, inhibition of histone deacetylases, mediation of cell proliferation and differentiation, and enhancement of intestinal barrier integrity.^{26 47 48} SCFAs contribute to host immune maturation by activating G protein-coupled receptors on the immune cell surface or by inhibiting lysine deacetylases.¹⁴ In the current study, we identified various SCFAs in the fetal gut, although the sources of the SCFAs were uncertain; however, the presence of these SCFAs might serve to modify fetal immune function and development.²⁶ Second, the microbial biomass must be maintained at extremely low levels in the intrauterine environment and excessive colonisation by microbes or potential pathogens in utero may result in intrauterine infection, causing adverse pregnancy outcomes such as miscarriage, preterm delivery or stillbirth.³⁵ For example, transmission of *Streptococcus agalactiae* from mothers to fetuses during pregnancy can lead to neonatal sepsis²⁰; oral *Fusobacterium nucleatum* colonises the mouse placenta, proliferates quickly to reach 10⁷ colony forming unit per gram tissue in 72 hours, and spreads to the fetus and amniotic fluid, causing term stillbirth^{45 49}; some bacterial species identified in the placenta, including *Campylobacter*, *E. coli*, *Leptotrichia*, *Neisseria*, *Peptostreptococcus* and *Streptococcus*, are associated with preterm birth, fetal death and neonatal sepsis.⁴⁵ Furthermore, perinatal transmission of bacteria carrying antibiotic resistance genes to fetuses may make newborns resistant to certain antibiotics after birth. Screening for certain bacteria carrying antibiotic resistance genes or potential pathogens during pregnancy and targeted use of antibiotics can prevent adverse pregnancy outcomes and reduce neonatal morbidity and mortality.^{35 50}

To our knowledge, this study is the first to provide direct evidence in support of the existence of a microbiome in the prenatal gut. However, several issues remain to be resolved. First, using the metatranscriptomic sequencing and metabolomics analysis techniques, we demonstrated the presence of metabolically active microbes in the prenatal gut, but this is not equivalent to capturing live microbes. Culture-based methods are required in the future to confirm the presence of live microbes. Second, we identified numerous microbial species and metabolites in the fetal gut. Since the knowledge regarding the functional role of these microbial species and metabolites remains limited, the effects of these on the fetal development needs to be explored. Third, our association analysis can not directly infer that these metabolites are produced by the microbes we detected, but merely suggests the relatedness between cecal content metabolites and the microbiome. For instance, we found that the butyric and hexanoic acids concentrations showed positive correlation with the relative abundance of *S. pneumoniae*. However, *S. pneumoniae* is not known to produce butyric and hexanoic acids.⁵¹ Furthermore, our study did not investigate the timing and origins of fetal gut microbiome colonisation. The fetal microbiome may begin triggering and educating the fetal immune system far earlier than previously thought. Finally, owing to the intrinsic differences between humans and animal

models, the composition and function of the gut microbiome in fetal lambs likely differ from that in humans. Clearly, such experiments would be infeasible in humans, but this study does provide some insight into the process of microbial colonisation in the fetal gut.

In conclusion, using lambs delivered by aseptic hysterectomy as an animal model and multiomics based analyses, our study provides strong evidence that the prenatal gut harbours a low-diversity and low-biomass metabolically active microbiome, and that the fetal gut microbiome is seeded antenatally. These findings advance our understanding of the fetal gut microbiome, which is helpful in designing clinical therapies aimed at improving human health and treating non-communicable diseases associated with deregulation of the gut microbiota. However, further studies are required for gaining an in-depth understanding of the origin, composition, function, dynamics and colonisation time of the fetal gut microbiome, and for elucidating the effects of the fetal gut microbiome and their metabolites on fetal development and health throughout the early stages of life.

METHODS

Establishment of animal model and study design

In this study, we used healthy newborn lambs delivered by aseptic hysterectomy as an animal model to investigate the presence of a microbiome in the prenatal fetal gut. Healthy nulliparous pregnant ewes approaching delivery were used to obtain healthy lambs via aseptic hysterectomy. Newborn lambs were euthanised using a penetrative captive bolt and their cecal content and umbilical cord blood samples were collected aseptically. Cecal content samples were analysed using metagenomic and metatranscriptomic sequencing to determine the composition and function of microbes. Cecal content and umbilical cord blood samples were assessed using metabolomics analysis to detect microbial metabolites.

Procedures involving animals and sample collection

All surgeries were performed under sodium pentobarbital anaesthesia and all efforts were made to minimise the suffering of animals.

Healthy nulliparous pregnant ewes with the same expected date of parturition were selected from the same herd. These ewes were fed the same diet without administration of antibiotics or probiotics during pregnancy. Offspring was obtained from the ewes via aseptic hysterectomy performed when the ewes exhibited obvious signs of parturition, such as swollen vulva and udders. Ewes showing the onset of labour, rupture of membranes or signs of maternal infection were excluded from the study. This resulted in a total of six ewes that were finally selected for use in this study.

Preoperative preparation of the ewes and surgical facility, as well as aseptic hysterectomy were performed as described previously⁵² using a standardised procedure by a team of trained technicians. Briefly, the ewes were narcotised when parturition was imminent and the disinfected surgical area of donor females was adhered to the sterile chamber of the surgical isolator. Hysterectomy was performed in a sterile surgical isolator using sterile disposable surgical instruments. After opening the abdominal cavity, the intact pregnant uterus and its entire contents were harvested, clamped, introduced into a germicidal bath, and transferred into a sterile disposable transfer isolator. As soon as the intact pregnant uterus was introduced into the transfer isolator, the connection between the transfer isolator and the surgery isolator was sealed to avoid contamination.

Samples were collected as previously described,¹¹ by pathology-trained personnel wearing surgical masks and sterile gloves and using sterile disposable surgical instruments. The uterus was opened without affecting the large vessels. Umbilical cord blood was sampled using a sterile needle and syringe and centrifuged at 4°C to obtain serum samples. The lamb was removed to a chemical and ultraviolet sterilised, class II B2 biosafety cabinet and placed into an autoclaved and blaze-disinfected salver. The lamb was then euthanised, and the abdominal cavity was opened with a sterile scalpel and forceps. We chose the cecum as the proxy of the intestinal tract. The cecum is the main fermentation organ in the intestine and is rich in microorganisms and microbial metabolites. Moreover, the saclike structure of the cecum has a large amount of enteral content, which is conducive to sample detection. Cecal content samples were collected with new sterile disposable scalpel and forceps. To avoid potential cross-contamination among lambs, cecal content samples, and instruments, none of the salvers or instruments were reused for any collection or dissection. All samples were collected within 1 hour of delivery under clean and sterile conditions. The serum and cecal content samples were collected into sterile closed vials, snap-frozen in liquid nitrogen, and stored at -80°C until further analysis. A negative control (nucleic acid-free water rather than samples) was subjected to the sample collection procedure.

DNA extraction and metagenomic sequencing

To reduce the risk of environmental contamination of samples, all experimental procedures (total DNA and RNA isolation, and qPCR) were performed in a stringent clean and sterile class II B2 biosafety cabinet. Total genomic DNA was extracted from cecal content samples using a DNeasy PowerSoil Kit (Cat. No.12888, Qiagen, Hilden, Germany) according to the manufacturer's instructions. The negative control during sample collection and an extraction control composed of *E. coli* (positive control) were subjected to the same protocols for DNA extraction, sequencing, and data analysis as the lamb-derived samples. DNA integrity was evaluated using agarose gel electrophoresis. DNA quality was assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and DNA concentration was measured using a Qubit 2.0 Fluorimeter (Invitrogen, Carlsbad, California, USA). Construction of the metagenomic library was performed using a TruSeq DNA PCR-Free Library Preparation Kit (Illumina, San Diego, California, USA). The quantity of each metagenomic library was evaluated with a Qubit V2.0 Fluorimeter. Metagenomic sequencing was conducted using an Illumina HiSeq 4000 platform (paired-end; insert size, 300 bp; read length, 150 bp).

RNA extraction and metatranscriptomic sequencing

Total RNA was extracted from cecal content samples using a RNeasy PowerSoil Total RNA Kit (Cat. No.12866, Qiagen) according to the manufacturer's protocols. Negative and positive controls were the same as those described for DNA extraction and metagenomic sequencing and were processed alongside cecal content samples throughout RNA extraction and metatranscriptomic sequencing. RNA concentration was measured using a Qubit V2.0 Fluorimeter and RNA quality was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). rRNA in each sample was removed using Ribo-Zero rRNA Removal Kits (Bacteria) and Ribo-Zero Magnetic Gold Kit (Yeast) (Epicentre, San Diego, California, USA) as per the manufacturer's instructions. Enriched mRNAs were used to construct an mRNA-enriched

metatranscriptomic library using a NEBNext Ultra RNA Library Prep Kit (New England Biolabs, Ipswich, Massachusetts, USA). The metatranscriptomic library was sequenced using an Illumina HiSeq 4000 (paired-end; insert size, 300 bp; read length, 150 bp).

Quality control of metagenomic and metatranscriptomic datasets

Metagenomic and metatranscriptomic sequencing generated 113.60 Gbp (18.93±2.24 Gbp per sample; data are expressed as mean±SD) and 64.86 Gbp (10.81±8.52 Gbp per sample) raw reads for the six cecal content samples, respectively. Quality control of the metagenomic and metatranscriptomic raw datasets was performed using Trimmomatic (V.0.35)⁵³ to trim adapters, filter out reads with >1% of uncertain bases, and remove low-quality bases (quality scores <20) and short reads (<150 bp). After quality control of the sequencing datasets, the remaining reads were mapped to the sheep (*Ovis aries*) genome using bowtie2⁵⁴ to eliminate potential host DNA and RNA contamination. A total of 10 544 549 and 56 746 269 high-quality pair-end clean reads for the six cecal content samples were acquired from the metagenome (1757, 425±5 62 944 clean reads per sample; 0.53±0.17 Gbp per sample) and metatranscriptome (9457, 711±5 982 821 clean reads per sample; 2.84±1.79 Gbp per sample), respectively.

Construction of a reference gene catalog for microbial metagenome and metatranscriptome analysis

High-quality metagenomic and metatranscriptomic reads were de novo assembled into contigs using MEGAHIT (V.1.0.6)⁵⁵ and Trinity,⁵⁶ respectively. This generated a total of 64 220 contigs with an average length of 864 bp (maximum 415 711 bp) and N50 length of 794 bp for the metagenomics data, and a total of 24 858 contigs with an average length of 398 bp (maximum 21 634 bp) and N50 length of 356 bp for the metatranscriptomics data. MetaGeneMark (V.2.7) was used to predict open reading frames (ORFs). All predicted ORFs were clustered via CD-HIT⁵⁷ using the following criteria: identity >95% and coverage >90%. Gene sequences with the greatest number of hits in each cluster were selected to represent the cluster, while other genes in the cluster were regarded as redundant and were not considered in downstream analyses. Metagenomic and metatranscriptomic gene catalogues containing 19 320 and 1691 nr genes, respectively, were constructed. Of the 19 320 genes identified in the metagenome, 5204 genes were shared between the cecal content samples and negative control, and 14 116 genes were unique to the samples (online supplemental tables S1 and S2). Of the 1691 genes identified in the metatranscriptome, 245 genes were shared between the cecal content samples and negative control, and 1446 genes were unique to the samples (online supplemental tables S3 and S4).

Filtering of potential contaminating genes from metagenomic and metatranscriptomic reference gene catalogs

Before phylogenetic and functional annotation of the metagenome and metatranscriptome, potential contaminant genes (originating from reagents or instruments) were removed from all sample sequencing datasets. Filtering was done by comparing the abundance of shared nr genes between each sample and the negative control. First, all shared nr genes in each sample and the negative control were selected; second, high-quality reads from each sample and the negative control were remapped to the shared genes using the RSEM software, as described previously.⁵⁸

The abundance of shared genes in cecal content samples and in the negative control were calculated using the fragments per kilobase of transcript per million fragments method⁵⁹ based on the number of uniquely mapped reads. DESeq2⁶⁰ was used to identify differentially expressed genes between each sample and the negative control. A $\log_2\text{FoldChange} > 2$ was used as the threshold for identifying contaminant genes in cecal content samples.⁶¹ After data decontamination, the gut microbial metagenome and metatranscriptome gene catalogues contained 14 199 and 1456 nr genes, respectively (online supplemental tables S5 and S6).

Taxonomic assignment and functional annotation of the metagenome and metatranscriptome

To estimate the taxonomic profiles of the gut microbial metagenome and metatranscriptome, the predicted nr genes were aligned against NCBI's nr database using DIAMOND blastx⁶² (V.0.8.23, default parameter except that $\text{evalue} = 1e-5$). Bacterial compositional profiles were summarised at the phylum, genus and species levels. To identify metagenomic and metatranscriptomic functional categories, the predicted nr genes were aligned against protein sequences in the KEGG and eggNOG databases using DIAMOND blastx⁶² (V.0.8.23, default parameter except that $\text{evalue} = 1e-5$, $\text{top hit} = 1$, $\text{max-target-seqs/k} = 1$). Antibiotic-resistance genes were annotated by alignment against the CARD database via BLASTN (V.2.9.0, default parameter except that $\text{evalue} = 1e-10$).

α -diversity analysis

Only microbial taxa detected in at least three out of six samples were considered as being present, and were used for α -diversity analysis. Within-sample diversity (α -diversity) was used to estimate the diversity and evenness of microbial taxa contained in a sample according to the Shannon index and based on the microbial taxa profile of the sample.

qPCR analysis

qPCR analysis was performed to detect the copy numbers of total bacteria, and those of five selected microbes showing a high relative abundance in our metagenomic analysis; the primers used in qPCR are shown in online supplemental table S19. A standard curve was constructed for total bacteria and for each individual strain. qPCR was performed in triplicate using 20 μL reaction mixtures containing 10 μL 2 \times qPCR Master Mix (SinoGene, Beijing, China), 0.5 μL each primer (10 μM), 8.0 μL ddH₂O and 1 μL 10 ng DNA templates. PCR cycling parameters were as follows: initial denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 20 s, annealing at 60°C for 30 s, extension at 72°C for 30 s and a final extension at 72°C for 10 min. Copy numbers of total bacteria and those of five selected microbes, per gram of cecal content were then calculated.

SCFAs profiling of cecal content and umbilical cord blood samples

Approximately 100 mg cecal content was vortexed with 1 mL HPLC-grade water. The mixture was then homogenised in a ball mill for 3 min at 45 Hz and centrifuged at 6000 \times g and 4°C for 10 min. The supernatant fluid (0.3 mL) or serum sample of umbilical cord blood (0.3 mL) was combined with 0.4 mL internal standard (25 $\mu\text{g}/\text{mL}$ 2-methylvaleric acid) and 0.1 mL 50% H₂SO₄, vortexed for 30 s, and ultrasonicated in ice water for 5 min. After centrifuging at 13,000 \times g and 4°C for 10 min,

the supernatant (60 μL) was transferred into a fresh 2 mL glass vial for gas chromatography-mass spectrometry (GC-MS) analysis. Standard solutions containing acetic, propionic, isobutyric, butyric, isovaleric, valeric and hexanoic acids (Sigma, St. Louis, Missouri, USA) were used for calibration. The concentration of SCFAs was measured using an Agilent 7890 GC system (Agilent Technologies, Santa Clara, California, USA) equipped with an HP-FFAP capillary column (30 m \times 250 μm \times 0.25 μm , Agilent Technologies) coupled with an Agilent 7000D mass spectrometer. Sample aliquots (1 μL) were injected using a split mode (5:1). Helium gas was used as a carrier at a flow rate of 1 mL/min. The oven temperature was initially maintained at 100°C for 1 min, increased to 150°C at a rate of 5°C/min and maintained for 0 min, further increased to 240°C at a rate of 20°C/min, and finally maintained at 240°C for 5 min. Temperatures used for injection, transfer line, quad and ion source were 260°C, 280°C, 150°C and 230°C, respectively. MS data were acquired in multiple reaction monitoring mode with a solvent delay of 3 min. SCFAs quantities are expressed as micrograms per milligram cecal content.

Metabolomics profiling of cecal content and umbilical cord blood samples

Sample preparation

Approximately 50 mg cecal content or 100 μL serum was placed in an Eppendorf tube, to which 1 mL ice-cold extraction solution (V methanol: V acetonitrile: V water = 2:2:1) containing an internal standard was added; the resulting mixture was vortexed for 30 s. The mixture was then homogenised in a ball mill for 4 min at 45 Hz and ultrasonicated in an ice water bath three times for 5 min each time (This step was not performed for serum samples). The mixture was incubated at -20°C for 1 hour. After centrifuging at 13,000 \times g and 4°C for 15 min, 750 μL of the supernatant was transferred into an Eppendorf tube and dried in a vacuum concentrator without heating. The residue was reconstituted in 200 μL extraction solution (V acetonitrile: V water = 1:1), vortexed for 30 s, and ultrasonicated in an ice water bath for 10 min. After centrifuging at 13,000 \times g and 4°C for 15 min, 60 μL of the supernatant was transferred into a fresh 2 mL liquid chromatography-MS glass vial for UHPLC-QTOF-MS analysis. The quality control sample was prepared by mixing an equal aliquot of the supernatants from all samples.

UHPLC-QTOF-MS analysis

UHPLC-QTOF-MS metabolomics analysis was performed using an Agilent 1290 UHPLC system coupled with a QTOF mass spectrometer (Triple TOF 5600, AB Sciex, Framingham, Massachusetts, USA). A Waters ACQUITY UPLC BEH Amide column (1.7 μm \times 2.1 mm \times 100 mm, Milford, Massachusetts, USA) was used for chromatographic separation and the column temperature was maintained at 25°C. The mobile phase consisted of 25 mM NH₄OAc and 25 mM NH₄OH in deionised water (pH 9.75; A) and acetonitrile (B); the elution gradient was as follows: 0–0.5 min, 95% B; 0.5–7 min, 95% B; 7–8 min, 65% B; 8–9 min, 40% B; 9–9.1 min, 40% B; 9.1–12 min, 95% B, delivered at 0.5 mL/min. The sample injection volume was 4 μL . Sample analysis was performed using electrosprayed electrospray ionisation (ESI) source in positive and negative ion modes. The TOF mass scanning range was m/z 50–1200 Da. MS spectra were acquired using Analyst TF 1.7 (AB Sciex) in information-dependent basis. In each cycle, precursor ions with an intensity greater than 100 were selected for fragmentation at a collision energy of 30 eV (15 MS/MS events with product ion accumulation time of 50

ms each). ESI source parameters were as follows: ion source gas 1, 60 psi; ion source gas 2, 60 psi; curtain gas, 35 psi; source temperature, 650°C; ion spray voltage floating, 5000 V in positive mode, and -4000 V in negative mode.

Data analysis and metabolite identification

MS raw data files were converted into mzXML format using ProteoWizard, and then processed using R package XCMS (V.3.2). Pretreatment of the acquired MS data included: correction of retention time (RT); peak identification, selection and grouping; integration of the peak area; and peak alignment and peak normalisation. Preprocessing results were used to generate a data matrix containing RT, mass-to-charge ratio (m/z) values and peak intensity. R package CAMERA was used for peak annotation after XCMS data processing. An in-house MS2 database was used for metabolite identification. Metabolites were identified and validated by aligning the molecular mass data (m/z) using the online HMDB (<http://www.hmdb.ca>), PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) and KEGG (<http://www.genome.jp/kegg/>) databases.

Correlations between the cecal microbiome and cecal content metabolites

Correlations between cecal bacterial species and cecal content metabolites were calculated using Spearman's rank correlation coefficient and visualised by co-occurrence networks and heat map. Cecal content metabolites were considered to associate with cecal bacterial species if $|\text{Spearman's rank correlation coefficient}| > 0.6$ and $p < 0.05$.

Statistical analyses

The results of qPCR analysis of total bacteria and five selected microorganisms in the cecal content samples and negative control were analysed by one-way analysis of variance using SAS V.9.4 (SAS Institute). Statistical significance was set to $p < 0.05$.

Acknowledgements The authors thank Beijing Allwegene Co., Ltd. for performing metagenomic and metatranscriptomic sequencing. The authors thank Professor Leluo Guan in Department of Agricultural, Food and Nutritional Science of University of Alberta for technical assistance.

Contributors The study was conceived by QD and designed by SL and YT. The experiments were performed by YB, SW, FZ and NZ. Data analysis was performed by YB, DS and GS. Paper writing was conducted by YB. All authors read and approved the final manuscript.

Funding This study was supported by the National Key Research and Development Programme of China (2017YFD0500500), National Natural Science Foundation of China (31702136), earmarked fund for the China Agriculture Research System (CARS-38) and Fundamental Research Funds for Central Non-profit Scientific Institution (16110382021008, Y2019CG08).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Animal care and use for all experiments were approved by and performed following the guidelines of the Animal Ethics Committee of Feed Research Institute, Chinese Academy of Agricultural Sciences, Beijing, China.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Metagenomic and metatranscriptomic sequencing data for all samples have been deposited in the NCBI Sequence Read Archive database under accession numbers PRJNA601636 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA601636>) and PRJNA598075 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA598075>).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability

of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Yanliang Bi <http://orcid.org/0000-0001-8183-8208>

REFERENCES

- 1 Flint HJ, Scott KP, Louis P, *et al.* The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* 2012;9:577–89.
- 2 Maranduba CMdaC, De Castro SBR, de Souza GT, *et al.* Intestinal microbiota as modulators of the immune system and neuroimmune system: impact on the host health and homeostasis. *J Immunol Res* 2015;2015:1–14.
- 3 Gomez de Agüero M, Ganai-Vonarburg SC, Fuhrer T, *et al.* The maternal microbiota drives early postnatal innate immune development. *Science* 2016;351:1296–302.
- 4 Rautava S, Luoto R, Salminen S, *et al.* Microbial contact during pregnancy, intestinal colonization and human disease. *Nat Rev Gastroenterol Hepatol* 2012;9:565–76.
- 5 Collado MC, Rautava S, Aakko J, *et al.* Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Sci Rep* 2016;6:23129.
- 6 Mackie RI, Sghir A, Gaskins HR. Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr* 1999;69:1035s–45.
- 7 Stinson LF, Boyce MC, Payne MS, *et al.* The not-so-sterile womb: evidence that the human fetus is exposed to bacteria prior to birth. *Front Microbiol* 2019;10:1124.
- 8 Chu DM, Ma J, Prince AL, *et al.* Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat Med* 2017;23:314–26.
- 9 Aagaard K, Ma J, Antony KM, *et al.* The placenta harbors a unique microbiome. *Sci Transl Med* 2014;6:237ra65.
- 10 Antony KM, Ma J, Mitchell KB, *et al.* The preterm placental microbiome varies in association with excess maternal gestational weight gain. *Am J Obstet Gynecol* 2015;212:653.e1–653.e16.
- 11 Seferovic MD, Pace RM, Carroll M, *et al.* Visualization of microbes by 16S in situ hybridization in term and preterm placentas without intraamniotic infection. *Am J Obstet Gynecol* 2019;221:146.e1–146.e23.
- 12 DiGiulio DB. Diversity of microbes in amniotic fluid. *Semin Fetal Neonatal Med* 2012;17:2–11.
- 13 Jiménez E, Fernández L, Marín ML, *et al.* Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Curr Microbiol* 2005;51:270–4.
- 14 Ganai-Vonarburg SC, Hornef MW, Macpherson AJ. Microbial-host molecular exchange and its functional consequences in early mammalian life. *Science* 2020;368:604–7.
- 15 Kim D, Hofstaedter CE, Zhao C, *et al.* Optimizing methods and dodging pitfalls in microbiome research. *Microbiome* 2017;5:52.
- 16 de Goffau MC, Lager S, Salter SJ, *et al.* Recognizing the reagent microbiome. *Nat Microbiol* 2018;3:851–3.
- 17 Perez-Muñoz ME, Arrieta M-C, Ramer-Tait AE, *et al.* A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. *Microbiome* 2017;5:48.
- 18 Leiby JS, McCormick K, Sherrill-Mix S, *et al.* Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome* 2018;6:196.
- 19 Theis KR, Romero R, Winters AD, *et al.* Does the human placenta delivered at term have a microbiota? results of cultivation, quantitative real-time PCR, 16S rRNA gene sequencing, and metagenomics. *Am J Obstet Gynecol* 2019;220:267.e1–267.e39.
- 20 de Goffau MC, Lager S, Sovio U, *et al.* Human placenta has no microbiome but can contain potential pathogens. *Nature* 2019;572:329–34.
- 21 Hultman J, Waldrop MP, Mackelprang R, *et al.* Multi-Omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 2015;521:208–12.
- 22 Salter SJ, Cox MJ, Turek EM, *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014;12:87.
- 23 Stout MJ, Conlon B, Landeau M, *et al.* Identification of intracellular bacteria in the basal plate of the human placenta in term and preterm gestations. *Am J Obstet Gynecol* 2013;208:226.e1–226.e7.
- 24 Rautava S, Collado MC, Salminen S, *et al.* Probiotics modulate host-microbe interaction in the placenta and fetal gut: a randomized, double-blind, placebo-controlled trial. *Neonatology* 2012;102:178–84.
- 25 Bi Y, Cox MS, Zhang F, *et al.* Feeding modes shape the acquisition and structure of the initial gut microbiota in newborn lambs. *Environ Microbiol* 2019;21:2333–46.
- 26 Koh A, De Vadder F, Kovatcheva-Datchary P, *et al.* From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell* 2016;165:1332–45.

- 27 Koenig JE, Spor A, Scalfone N, *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* 2011;108 Suppl 1:4578–85.
- 28 Iida H, Yamazaki N, Kibayashi C. Total synthesis of (+)-nojirimycin and (+)-1-deoxynojirimycin. *J Org Chem* 1987;52:3337–42.
- 29 Ishida N, Kumagai K, Niida T, *et al.* Nojirimycin, a new antibiotic. II. Isolation, characterization and biological activity. *J Antibiot* 1967;20:66–71.
- 30 Inouye S, Tsuruoka T, Ito T, *et al.* Structure and synthesis of nojirimycin. *Tetrahedron* 1968;24:2125–44.
- 31 Galm U, Hager MH, Van Lanen SG, *et al.* Antitumor antibiotics: bleomycin, enediynes, and mitomycin. *Chem Rev* 2005;105:739–58.
- 32 Higgins CE, Kastner RE. Nebramycin, a new broad-spectrum antibiotic complex. II. description of *Streptomyces tenebrarius*. *Antimicrob Agents Chemother* 1967;7:324–31.
- 33 Dienstag J, Neu HC. In vitro studies of tobramycin, an aminoglycoside antibiotic. *Antimicrob Agents Chemother* 1972;1:41–5.
- 34 Macpherson AJ, de Agüero MG, Ganai-Vonarburg SC. How nutrition and the maternal microbiota shape the neonatal immune system. *Nat Rev Immunol* 2017;17:508–17.
- 35 Goldenberg RL, Hauth JC, Andrews WW. Mechanisms of disease - Intrauterine infection and preterm delivery. *New Engl J Med* 2000;342:1500–7.
- 36 Kim MJ, Romero R, Gervasi MT, *et al.* Widespread microbial invasion of the chorioamniotic membranes is a consequence and not a cause of intra-amniotic infection. *Lab Invest* 2009;89:924–36.
- 37 Gonçalves LF, Chaiworapongsa T, Romero R. Intrauterine infection and prematurity. *Ment Retard Dev Disabil Res Rev* 2002;8:3–13.
- 38 Ardisson AN, de la Cruz DM, Davis-Richardson AG, *et al.* Meconium microbiome analysis identifies bacteria correlated with premature birth. *PLoS One* 2014;9:e90784.
- 39 Gilbert WM, Brace RA. Amniotic fluid volume and normal flows to and from the amniotic cavity. *Semin Perinatol* 1993;17:150–7.
- 40 Schubert R, Hohlweg U, Renz D, *et al.* On the fate of orally ingested foreign DNA in mice: chromosomal association and placental transmission to the fetus. *Mol Gen Genet* 1998;259:569–76.
- 41 Jiménez E, Marín ML, Martín R, *et al.* Is meconium from healthy newborns actually sterile? *Res Microbiol* 2008;159:187–93.
- 42 Stinson LF, Payne MS, Keelan JA. Planting the seed: origins, composition, and postnatal health significance of the fetal gastrointestinal microbiota. *Crit Rev Microbiol* 2017;43:352–69.
- 43 Han YW, Ikegami A, Bissada NF, *et al.* Transmission of an uncultivated *Bergeyella* strain from the oral cavity to amniotic fluid in a case of preterm birth. *J Clin Microbiol* 2006;44:1475–83.
- 44 Han YW, Fardini Y, Chen C, *et al.* Term stillbirth caused by oral *Fusobacterium nucleatum*. *Obstet Gynecol* 2010;115:442–5.
- 45 Fardini Y, Chung P, Dumm R, *et al.* Transmission of diverse oral bacteria to murine placenta: evidence for the oral microbiome as a potential source of intrauterine infection. *Infect Immun* 2010;78:1789–96.
- 46 Fardini Y, Wang X, Témoins S, *et al.* *Fusobacterium nucleatum* adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity. *Mol Microbiol* 2011;82:1468–80.
- 47 Furusawa Y, Obata Y, Fukuda S, *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* 2013;504:446–50.
- 48 Hamer HM, Jonkers D, Venema K, *et al.* Review article: the role of butyrate on colonic function. *Aliment Pharmacol Ther* 2008;27:104–19.
- 49 Han YW, Redline RW, Li M, *et al.* *Fusobacterium nucleatum* induces premature and term stillbirths in pregnant mice: implication of oral bacteria in preterm birth. *Infect Immun* 2004;72:2272–9.
- 50 Schrag SJ, Zywicki S, Farley MM, *et al.* Group B streptococcal disease in the era of intrapartum antibiotic prophylaxis. *N Engl J Med* 2000;342:15–20.
- 51 Vos P, Garrity G, Jones D. *Bergey's Manual of Systematic Bacteriology: The Firmicutes*. Vol 3. New York: Springer, 2009.
- 52 Riou Y, Gouet P, Dubourguier HC, *et al.* Techniques for obtaining, fistulization and rearing of axenic or gnotoxenic lambs, kids and calves. *Ann Rech Vet* 1977;8:13–24.
- 53 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 2014;30:2114–20.
- 54 Langmead B, Wilks C, Antonescu V, *et al.* Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 2019;35:421–32.
- 55 Li D, Liu C-M, Luo R, *et al.* MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;31:1674–6.
- 56 Grabherr MG, Haas BJ, Yassour M, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;29:644–52.
- 57 Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;22:1658–9.
- 58 Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;12.
- 59 Trapnell C, Williams BA, Pertea G, *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010;28:511–5.
- 60 Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;11.
- 61 Alipour MJ, Jalanka J, Pessa-Morikawa T, *et al.* The composition of the perinatal intestinal microbiota in cattle. *Sci Rep* 2018;8:1–14.
- 62 Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using diamond. *Nat Methods* 2015;12:59–60.