

RESEARCH

Open Access



# Application of multi-omics technology in pathogen identification and resistance gene screening of sheep pneumonia

Kai Huang<sup>1</sup>, Lvfen Yuan<sup>2</sup>, Jia Liu<sup>3</sup>, Xiaolong Li<sup>1</sup>, Dan Xu<sup>1</sup>, Xiaoxue Zhang<sup>3</sup>, Jie Peng<sup>3</sup>, Huibin Tian<sup>1</sup>, Fadi Li<sup>1</sup> and Weimin Wang<sup>1\*</sup>

## Abstract

**Background** Pneumonia constitutes a major health challenge in sheep, severely compromising growth rates and overall productivity, and resulting in considerable economic losses to the sheep industry. To address this issue, the development of disease-resistant breeding programs based on the identification of genetic markers associated with pneumonia susceptibility is of critical importance. This study investigated a sheep population on a farm where pneumonia was endemic. The purpose was to use multi-omics methods to rapidly identify the principal pathogens responsible for pneumonia outbreaks, and to screen for genetic loci and key genes related to pneumonia resistance, thereby providing a scientific basis for the implementation of targeted breeding strategies for pneumonia resistance.

**Results** Here, we assessed the impact of pneumonia on sheep growth by evaluating the pneumonia phenotypes of 912 sheep. High-throughput transcriptome sequencing of 40 lungs was conducted to obtain exogenous RNA fragments for microbial sequence alignment. Additionally, 16S rRNA sequencing was performed on lung tissues from 10 healthy and 10 diseased sheep to identify biomarkers associated with phenotypic differences. *Mycoplasma ovipneumoniae* was identified as the primary pneumonia pathogen, and its presence was further validated by load quantification and immunohistochemical analysis. Integration of genome-wide association study (GWAS) data from 266 lung pathological scores with transcriptome-based differentially expressed genes analysis enabled the identification of five single nucleotide polymorphisms (SNPs) and three potential candidate genes associated with *Mycoplasma* pneumonia. Subsequent genotyping and phenotype association analyses confirmed the significance of two SNPs and established a strong association between the *FOXF1* gene and resistance to *Mycoplasma* pneumonia.

**Conclusions** High-throughput sequencing technologies have enabled the rapid and accurate identification of the causative pathogen of sheep pneumonia. By integrating multi-omics data, two genomic loci significantly associated with *Mycoplasma* pneumonia were screened, as well as an anti-*Mycoplasma* pneumonia key gene, *FOXF1*.

**Keywords** Sheep, Multi-omics, Pathogen identification, *Mycoplasma ovipneumoniae*, Genome-wide association study, Disease-resistant breeding

\*Correspondence:

Weimin Wang  
wangweimin@lzu.edu.cn

<sup>1</sup>State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems; Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs; Engineering Research Center of

Grassland Industry, Ministry of Education; College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730020, China

<sup>2</sup>Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences (CAAS), Lanzhou 730046, China

<sup>3</sup>College of Animal Science and Technology, Gansu Agricultural University, Lanzhou 730070, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Animal diseases are one of the main threats to modern livestock and poultry production. According to reports, in developed countries, economic losses caused by diseases account for about 17% of the total output value of animal husbandry, while in developing countries, this proportion reaches 35–50% [1]. Sheep are economically important livestock species globally. Pneumonia is a respiratory disease that is widely prevalent in sheep farms. The incidence rate of pneumonia in ruminant animals ranges from 10 to 40%, with higher percentages in childhood [2]. Pneumonia can cause growth retardation and increase the risk of secondary pleurisy in sheep, resulting in major production losses [3]. Bacteria-induced infectious pneumonia is most common, and it has been confirmed that pathogens such as *Mycoplasma ovipneumoniae*, *Mycoplasma filamentosa*, *Pasteurella multocida*, *Cryptobacterium septicum*, *Mannheimia haemolytica*, *Klebsiella pneumoniae*, *Escherichia coli*, *Streptococcus*, and other pathogens can all cause pneumonia individually or in combination [4–6].

Next generation sequencing (NGS) technology is a promising method for pathogen identification, including rare and newly identified viruses [7]. NGS is independent of cultivation and can efficiently and unbiasedly detect pathogenic microorganisms in clinical samples, providing comprehensive analyses [8, 9]. Previous studies have generally focused on pulmonary microbiota composition in healthy sheep [10, 11]; while the data on the pulmonary microbiota related to diseases in sheep remains relatively scarce [6]. In microbiome research, comparing 16S rRNA gene sequences has become a reliable approach for identifying microorganisms associated with pathogenicity and infection [12]. Compared to pathogen isolation and identification in veterinary medicine, comprehensive analysis of multi-omics can help to quickly and accurately identify pathogens and discover pathogens that are difficult to cultivate.

*Mycoplasma ovipneumoniae* (*M. ovipneumoniae*) colonization has been detected in approximately 90% of pneumonia-associated clinical surgery cases [13]. Moreover, it has been identified as the main pathogen causing severe epidemic pneumonia in bighorn sheep [14]. *M. ovipneumoniae* has also caused severe respiratory disease outbreaks in goats across multiple regions [15, 16]. With the increasing proportion of large-scale breeding in recent years, *Mycoplasma pneumonia* has become the highest incidence rate infectious disease in the sheep industry. The main manifestations of *M. ovipneumoniae* infections include wheezing, coughing, fever, weight loss, and pulmonary interstitial hyperplasia inflammation [17]. It is characterized by an unusually wide host range, high phenotype, strong infectivity and high incidence rate, and is prone to secondary or mixed infection

of other bacteria [18]. Currently, antibiotics remain the main means of treating *M. ovipneumoniae* infections. However, the long-term use of antibiotics can affect meat quality, produce antibiotic residues, and even lead to the development of drug resistance. Therefore, starting from the genetic basis and screening for resistance genes of *Mycoplasma pneumoniae* for disease-resistant breeding is an important way to fundamentally address this problem.

In farm animal breeding programs, disease resistance and susceptibility—the natural, unique mechanisms by which a host responds to infectious pathogens—are often used as main references. Disease resistance depends on the host's biological and immune responses, so livestock's ability to resist multiple diseases varies among individuals [19, 20]. Studies on sheep and cattle have reported potential genetic components involved in the differences in pneumonia susceptibility observed among animals [3, 21]. Heritable variations may aid in selecting animals that can resist and recover from infections of multiple pathogenic factors [22]. In addition, omics techniques, bioinformatics, and Genome-wide association studies (GWAS) can identify genomic regions associated with potential causal mutations that affect susceptibility to infectious diseases [23–25], and the discovery of these regions may also lead to the establishment of new disease diagnostic tools and alternative therapies. Therefore, using molecular biology and molecular genetics techniques to search for disease-resistance genes for disease-resistance breeding and improve genetic disease resistance in livestock has become an effective supplement to conventional disease prevention and control methods. The screening and cultivation of disease-resistant animal strains have also become a major research direction.

In the current study, we: (1) showed the severe impact of pneumonia on the growth and productivity of sheep, (2) integrated multi-omics technologies to quickly and accurately identify the pathogen of sheep pneumonia, *M. ovipneumoniae*, and (3) identified genomic regions associated with *Mycoplasma pneumonia*. In particular, by combining GWAS, transcriptomics, and association analysis with pneumonia, two effective SNPs related to *Mycoplasma pneumonia* and a candidate *FOXF1* gene for anti-*Mycoplasma pneumonia* were screened and validated.

## Methods

### Animal parameters and sample collection

All the sheep in this study were from the Minqin experimental farm of Lanzhou University (N38°43'41'', E103°013'), and we first evaluated their respiratory disease status. At the age of 180 days and after fasting for 12 h, whole blood for measurement of blood physiological indicators, DNA extraction and whole genome

sequencing (WGS) was obtained from each sheep via the jugular vein. Subsequently, on the premise of ensuring animal welfare and reducing animal suffering, each animal was slaughtered according to standard commercial procedures following the requirements of the China Council on Animal Care. This study received approval from the Lanzhou University Research Ethics Committee (approval number: 2021-02). During two rounds of slaughter, the lungs of 912 Hu sheep (905 male and 7 female) from this studied sheep farm were examined grossly and photographed. After identifying the lung phenotype of each Hu sheep, lung tissue block samples were collected in triplicate from normal areas of healthy lungs and diseased areas of inflammatory diseased lungs, each for subsequent sequencing, slice staining, and molecular experiments.

All included sheep maintain similar birth ages in each slaughter (i.e. each batch, as detailed in Table 1). All batches of sheep were raised in single pens in identical environments, with ad libitum access to food and water and no differences in management methods or nutritional levels.

Clinical examination, histopathological observation, and phenotype data collection

The lungs health or pathological conditions, including the lung lesions and tactile perception of lung elasticity, were comprehensively evaluated. Next, we unfolded the ventral sides of the lungs horizontally and photographed them. The lung tissue blocks used for slice staining were fixed in 4% formaldehyde, embedded in paraffin according to standard procedures, then sectioned, and stained with hematoxylin and eosin for histopathological examination under an optical microscope. Moreover, we assessed the pathological changes in lung tissue and scored them on a 5-point system, ranging from 1 to 5, indicating increasing lesion severity. Additional file 1 summarizes the general histopathological observations and scores of the lung slices.

Combining clinical dissection and histopathological observations, we assessed the lung lesion phenotype and collected pathological score data. Based on these data, we divided 912 Hu sheep into healthy and diseased groups for subsequent analyses.

Table 1 Incidence of pulmonary lesions caused by *Mycoplasma pneumoniae* in sheep population

Breeds	Batch	Slaughtering time (year/month)	Sample size	Number of cases	Incidence rate
Hu sheep	1	2022/8	491	157	31.98%
Hu sheep	2	2023/1	421	168	39.91%
Total			912	325	35.64%

Collection and analysis of growth traits and blood parameters

The growth performance measurement of Hu sheep starts from 80 days and ends at 180 days, during which the body weight (BW) and feed intake (FI) are measured every 20 days. The BW of lambs at 80 days is recorded as the initial weight, and the BW at 180 days is recorded as the final weight. At the same time, the live weight at slaughter and carcass weight were recorded. In addition, calculate the Total FI of each Hu sheep throughout the entire experimental period (80 d-180 days) and based on the corresponding data, calculate the average daily feed intake (ADFI), average daily weight gain (ADG), and feed conversion rate (FCR): ADFI (kg/day) = Total FI (kg) / 100 (day); ADG (kg/day) = (final weight (kg) - initial weight (kg)) / 100 (day); FCR = ADFI (kg/day) / ADG (kg/day). Sheep whole blood samples collected through the jugular vein before slaughtered were measured on the fully automatic five-classification animal blood cell analyzer Aldex ProCyt Dx\* for their blood physiological indicators.

Based on the phenotype grouping of the lungs mentioned above, the growth performance and blood parameter data of all sheep were combined and analyzed for inter group differences to evaluate the impact of pneumonia related respiratory diseases on sheep growth performance and blood parameters.

Pathogen identification of sheep pneumonia by high-throughput transcriptome sequencing

Among the 912 Hu sheep with lung phenotype data, 20 Hu sheep in the healthy group (10 each from batches 1 and 2) and 20 Hu sheep in the diseased group (10 each from batches 1 and 2) were random selected to collect samples from normal or typical lesion areas of the lungs (n=40) for total RNA extraction. After RNA quality testing, the kit was used to remove ribosomal RNA (rRNA) and retain mRNA and other non-coding RNA. cDNA was synthesized by fragmenting and reverse transcription of RNA, and then a series of steps such as end repair and linker ligation were performed to construct the lncRNA sequencing library. Finally, the library was sequenced on the sequencing platform Illumina HiSeq (Illumina, San Diego, CA, USA) to obtain 150 bp paired-end (PE) reads.

We used Fastp to perform quality control of the offline data based on our lncRNA sequencing results and Picard to convert the results into BAM format files. For pathogen analysis, Genomic Analysis Toolkit (GATK) PathSeq was used according to the gatk-best-practices process. The tables of 40 individual pathogen analyses were merged, and the sum of the unambiguous sequences was obtained. We then screened microbial species with a high number of unambiguous sequences for between-group difference analysis. Bowtie2 was used to compare sequences with the reference genome

of the aforementioned microbial species, with a filtering condition of  $m \geq 30$ , and the mapped sequences were counted. Next, we calculated reads mapped per million reads (RPM) for each microorganism and analyzed the between-group differences. Linear regression analysis was performed on the sequence reads of pathogenic microorganisms and tissue pathological scores. Coverage statistics were performed using Bamdst for the sequenced sequence comparison on the reference genome of *M. ovipneumoniae*. Finally, the *M. ovipneumoniae* lactate dehydrogenase (*LDH*) sequence was extracted and subjected to multiple sequence alignment using MAFFT. Then, a phylogenetic tree was constructed using MEGA.

### 16S rRNA sequencing analysis for microbial community diversity in lung tissues

From 40 lung tissue samples used for transcriptome sequencing, 20 lung tissue samples (10 each from the healthy and diseased groups) were random selected for 16S rRNA sequencing to determine the colonized microbial community.

Total DNA was extracted using the OMEGA Soil DNA Kit (M5635-02; Omega Bio-Tek, Norcross, GA, USA). Next, the V3-V4 region of the standard bacterial 16S rRNA gene was amplified through polymerase chain reaction (PCR) with the following primers: forward, 5'-ACTCCTACGGGAGGCAGCA-3', and reverse, 5'-GGACTACHVGGGTWTCTAAT-3'. PCR reactions, containing 5  $\mu$ L 5Xreaction buffer, 5  $\mu$ L 5XGC buffer, 2  $\mu$ L dNTP (2.5mM), 1  $\mu$ L each primer (10 $\mu$ M), 2  $\mu$ L DNA template, 8.75  $\mu$ L ddH<sub>2</sub>O and 0.25  $\mu$ L Q5 DNA Polymerase (NEB) in a volume of 25  $\mu$ L, were amplified by thermocycling: 2 min at 98 °C for initial denaturation; 30 cycles of 15 s denaturation at 98 °C, 30 s annealing at 55 °C, and 30 s extension at 72 °C; followed by 5 min final elongation at 72 °C. The amplification products were purified and recovered to prepare a sequencing library. Next, we performed dual-end sequencing of the library on Illumina NovaSeq. The PE reads obtained through sequencing were first spliced according to their overlap relationships, using FastQC for sequence quality control. The filtered data were further processed using the DADA2 method [26] on QIIME2 (<https://qiime2.org>) to generate amplicon sequence variants (ASVs). Finally, we obtained the abundance tables for these ASVs (i.e., the ASV feature tables). After completing the denoising of all libraries, we merged all ASV feature tables and removed singleton ASVs. The Greengenes (<http://greengenes.secondgenome.com/>) database was selected for ASVs taxonomic annotation. After distinguishing and grouping the samples, we conducted species composition analysis at the phylum and genus levels, and performed a significant difference analysis of the abundance of the *Mycoplasma* and *Pasteurella* between groups at the genus

level.  $\alpha$ -Diversity metrics, including the Chao1, Simpson, Faith's PD, Pielou's evenness, and Good's coverage indexes, were calculated based on the ASV feature tables by using the diversity function in the R package vegan (<https://rdrr.io/cran/vegan/>).  $\beta$ -Diversity was estimated using the Bray–Curtis distance algorithm and visualized using Principal coordinate analysis (PCoA) and Nonmetric multidimensional scaling (NMDS). Finally, we used the Python software package Linear discriminant analysis (LDA) effect size (LEfSe) to analyze robust biomarkers with LDA scores  $> 4$  and  $p < 0.05$  in both groups.

### Verification of *M. ovipneumoniae* load in experimental sheep lung tissue

Due to the high coverage on the reference genome of *M. ovipneumoniae*, the expression of *LDH* gene was used to characterize the load of *M. ovipneumoniae*. The fluorescence quantitative detection of *LDH* gene was designed and the detailed information of the relevant primers was listed in Additional file 2. Using lung DNA as a template, fluorescence quantitative PCR (qPCR) was performed on *LDH* and standardized with  $\beta$ -tubulin to characterize the *M. ovipneumoniae* load in each sheep lung. qPCR reaction system (10  $\mu$ L): 0.5  $\mu$ L each primer (10  $\mu$ M), 1  $\mu$ L DNA template, 5  $\mu$ L TB Green Premix Ex Tag II (Takara), 3  $\mu$ L ddH<sub>2</sub>O. Quantitative program: pre-denaturation at 95 °C for 3 min, denaturation at 95 °C for 10 s, annealing at 60 °C for 10 s, extension at 72 °C for 15 s, 40 cycles. All quantitative measurements were obtained in triplicate, and the qPCR reactions were performed in a Bio-Rad CFX384 sequence detector.

### Immunohistochemistry of *M. ovipneumoniae*

In order to further characterize *M. ovipneumoniae* in lungs, the lung tissue paraffin sections prepared were sequentially dewaxed, rehydrated, antigen repaired, and quenched with endogenous peroxidase. Then, sections were blocked with 5% BSA for 40 min and incubated with primary antibody (self-made rabbit immune serum, dilution ratio 1:500) and goat anti-rabbit HRP labeled secondary antibody. Simultaneously set up a negative control that was not incubated with the antibody. DAB staining and hematoxylin counterstaining of sections were performed, and finally dehydrated and sealed. The images were examined under an optical microscope and collected.

### WGS and data processing

Genomic DNA was extracted from Hu sheep blood samples according to the instructions of the Easy Pure Blood Genomic DNA Kit (TransGen Biotech) and sent to Molbreeding Biotech (Shijiazhuang, China). A PE sequencing library with an insert size of 350 bp was constructed using a Truseq Nano DNA HT Sample Preparation Kit



(Illumina), according to the manufacturer's instructions. The libraries that have passed quality inspection were subjected to PE sequencing with 150-bp reads (PE150) on Illumina HiSeq. Raw sequencing data were converted to fastq files [27]. Here, FastQC (version 0.10.1) was used to filter the raw data; in particular, linker sequences, sequences containing the uncalled base N, and <60-bp sequences were discarded.

The obtained clean reads were mapped to the sheep reference genome Oar\_v1.0 using Burrows-Wheeler Aligner (version 0.7.8) [28]. SAMtools (version 1.12) [29] was used to convert the mapping results to the BAM format, remove duplicate reads, and retain pairs with the highest mapping quality. Then, GATK (version 3.4.0) was used to call SNPs and generate files in the VCF format. Next, we used VCFtools (version 0.1.14) software to correct the GATK results and screen out high-quality SNPs [30]. The screening conditions were as follows: (1) number of supported SNP reads > 5, (2) deletion rate < 20%, (3) minimum allele frequency > 5%, and (4) root mean squared mapping quality > 20. Beagle (version 5.0), which self-fills genotype data, was used to eliminate Not Available (NA) value present in the genotype data. Finally, 23,480,428 SNPs and 912 sheep, which passed the filtering and quality control steps, were used for subsequent analysis.

#### GWAS and significant SNPs

A GWAS was performed on lung pathological scores using the General Linear Model in the R package rMVP [31]. In addition, the influence of population stratification was corrected through the addition of the first five PCs derived from whole-genome SNPs. Lamb birthplace and BW variables were included as covariates in the analysis model. To avoid potential false positives, the genome-wide significance threshold was adjusted, and the significance threshold of GWAS results for sheep lung pathological scores traits was set to  $p < 1 \times 10^{-6}$ . The R package CMplot (<https://github.com/YinLiLin/R-CMplot>) was used to draw quantile–quantile (Q-Q) and Manhattan plots, set the parameter 'threshold =  $10^{-6}$ ', which means the gray threshold line in the graph is 6. Before gene annotation, we used PopLDdecay to assess for the attenuation of linkage imbalance. Ensembl and NCBI were used to determine the location of the significant SNPs within the sheep reference genome Oar\_v1.0 and identify annotated genes within 200 kb upstream and downstream of significant SNPs.

#### Conventional analysis of RNA-Seq data and screening of candidate key genes

The high-quality sequences (i.e., clean data) obtained by filtering the raw data of transcriptome sequencing were mapped to the sheep reference genome Oar\_v1.0, and the mapped reads were assembled and spliced to restore

the transcriptional sequence. We then used Cufflinks [32] to calculate each gene's expression level, standardized to fragments per kilobase of exon per million fragments mapped (FPKM). Next, we used the R package DESeq to perform principal component analysis (PCA) on each sample based on gene expression levels. The stat-test function in the Ballgown package was used to search for mRNAs differentially expressed between the healthy and diseased groups, based on the following  $|\log_2(\text{fold change})| > 1$  and adjusted  $p < 0.05$ . Next, further expression clustering and enrichment analyses, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses, were conducted on the identified differentially expressed genes (DEGs). To validate our RNA-Seq results, total RNA of sheep lung tissues extracted from those performed RNA-Seq were subjected to cDNA synthesis using the RevertAid First Strand cDNA synthesis kit (Thermo Scientific Co., Ltd, China). Next, we performed qRT-PCR to measure the expression levels of 10 randomly selected DEGs (*PROM1*, *PAK5*, *SLIT1*, *PGAM1*, *UBTD1*, *ADGRB3*, *FOXF1*, *PRDX6*, *PTGS1* and *SMAD7*) by using a TB Green Kit (Takara) and lung tissue cDNA. Additional file 2 presents the details of our qRT-PCR primers. The expression level of each gene was normalized by  $\beta$ -actin [33]. The reaction system and program of qRT-PCR are the same as those of qPCR mentioned above. All quantitative measurements were obtained in triplicate, and all qRT-PCR assays were performed on the Bio-Rad CFX384 sequence detector. DEGs obtained from transcriptome analysis were cross-analyzed with annotated genes from the above GWAS results to screen for potential candidate functional genes and their associated significant SNPs.

#### Genotyping and association analysis of significant SNP loci

We randomly selected 854 blood DNA samples from a population of 912 sheep, and used competitive allele-specific-based polymerase chain reaction (KASPar) (LGC Genomics, Hoddesdon, UK) to genotype the significant SNPs, according to a previously published method [34]. The primer pairs designed for our KASPar genotyping are listed in Additional file 2.

To verify the relationship between significant SNPs and screened candidate key genes, as well as their correlation with the anti-*Mycoplasma* pneumonia mechanism. We grouped the expanded sheep population ( $n = 203/197/194/189$ ) based on the different genotypes of each significant SNP locus, and then detected the differences in the expression of corresponding candidate key genes, as well as in the lung phenotypes (including MO load and lung pathological score) between groups. The expression level detection of candidate key genes was evaluated by qRT-PCR, as for DEGs. The detection method for *M. ovipneumoniae* load was qPCR.

Furthermore, we studied the association between significant SNP loci and inflammation related blood physiological indicators based on genotype grouping.

### Statistical analysis

qRT-PCR and qPCR data were processed using the  $2^{-\Delta\Delta Ct}$  method [35]. Internal reference genes were used to standardize the expression of each gene. Differences between two groups (healthy and diseased groups) were evaluated using Student's *t* test, whereas those between three groups (three genotypes groups) were assessed one-way analysis of variance followed by Dunnett's test for multiple comparisons. These tests were performed using SPSS (version 20.0; IBM, Chicago, IL, USA). Graphics were drawn using R (version 4.2.3) and GraphPad Prism (version 5.0; GraphPad Software, San Diego, CA, USA). All results are expressed as means  $\pm$  standard errors of the means. Finally,  $p < 0.05$  and  $p < 0.01$  were considered to indicate significance and extreme significance, respectively.

## Results

### Phenotypic identification of pneumonia lesions and their impact on sheep growth performance

During an investigation in early July 2022, we found that approximately 20% of sheep experienced coughing, runny nose, and elevated body temperature at our study sheep farm in Minqin County, Wuwei City, Gansu Province, China. In some of these sheep, the clinical respiratory symptoms worsened over time. Two batches of sheep fed in this farm were slaughtered. In clinical autopsy, we observed both healthy lungs and diseased lungs (Fig. 1a). Substantial lesions appeared in the pointed lobe of the diseased lung, with bleeding points on the surface of the lung. The trachea contained a lot of foam-like mucus, and hilar lymph nodes were swollen. Histopathological analysis of the lungs and trachea showed proliferation of parabronchial lymphoid follicles or lymphocyte infiltration in diseased lungs. The bronchi were filled with pus, mixed with monocytes, neutrophils, and lymphocytes. The alveoli collapsed to varying degrees, and small abscesses were often seen in the collapsed tissue area. Congestion and bleeding of small and medium-sized arteries in the pulmonary interstitium, as well as congestion of capillaries were visible (Fig. 1b). A 5-point scoring system was established on the basis of the severity of each lesion in lung tissue slices; Fig. 1c presents a standard example. These scores data were collected as phenotypic traits for subsequent research.

Based on the epidemiological, clinical anatomical, and histopathological results mentioned above, we grouped all 912 observed sheep into healthy lung (hereinafter, healthy group) and diseased lung (hereinafter, diseased group) groups, and compiled a pneumonia phenotype

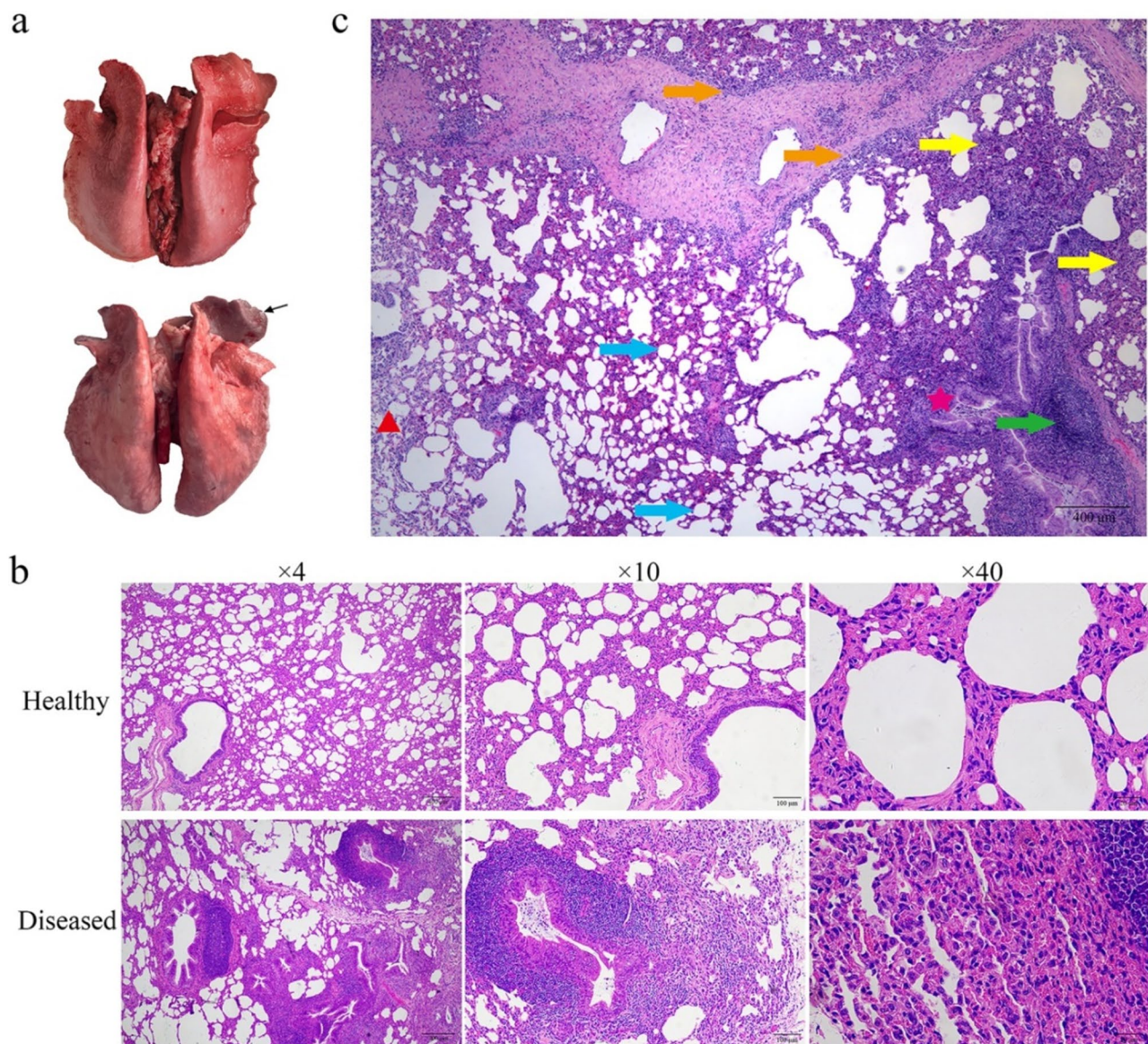
statistical table (Table 1). It can be seen that the average incidence of pulmonary lesions in sheep caused by respiratory diseases was about 35%. In addition, from July 2022 to January 2023, pneumonia in sheep farms was gradually spreading.

According to the phenotype grouping, we statistically analyzed the growth traits (Table 2) and blood physiological indicators (Table 3) of 912 sheep for differences between groups. We measured each individual's BW, FI, live weight at slaughter, and carcass weight, and calculated the corresponding ADFI, BWG, ADG, and FCR. No between-group BW differences were noted in the 80-day-old Hu sheep; however, in the 180-day-old Hu sheep, the diseased group demonstrated a extremely significant decrease in BW compared with the healthy group ( $p < 0.01$ ); the live weight before slaughter and carcass weight were also significantly lower than those in the healthy group (both  $p < 0.01$ ). Moreover, throughout the feeding trial period, the ADFI and ADG of the diseased Hu sheep remained significantly lower than those of the healthy group (all  $p < 0.01$ ). In contrast, the FCR in the diseased group was higher than in the healthy group; however, this difference was nonsignificant. We also measured 11 blood physiological indicators related to inflammation. Although most indicators such as lymphocyte (LYMPH), monocyte (MONO), eosinophil (EO), basophil (BASO), EO% and BASO% did not show statistically significant differences between groups, the total white blood cell count, neutrophil count, and proportion in the diseased group were significantly higher than those in the healthy group (all  $p < 0.01$ ).

### Pathogen identification method based on high-throughput sequencing analysis

High-throughput transcriptome sequencing of two batches of sheep lung tissue yielded average sequencing reads of approximately 82.8 M and 106.9 M, respectively (Additional file 3). Sheep genome-related sequences accounted for 74.7% and 71.3%, respectively (Additional file 4). After excluding the host genome sequence, several major microorganisms that colonize the lungs were identified by comparing exogenous RNA fragments and calculating the RPM of the aligned microorganisms (Additional file 5). The inter group difference analysis of these microorganisms related sequence quantities emphasized the importance of *M. ovipneumoniae* ( $p < 0.01$  in both batches) and *Pasteurella multocida* (*P. multocida*,  $p < 0.01$  in batch 2) in diseased groups of all batches. In addition, except for *Mycobacterium abscesses* (*M. abscessus*,  $p < 0.01$  in batch 1) which showed higher sequence quantity in the healthy group, the sequence quantity of other microorganisms was very low and there was no significant difference between the groups (Fig. 2a). Therefore, *M. ovipneumoniae* and *P. multocida*





**Fig. 1** Phenotype identification and histopathological observation of sheep pneumonia lesions. **(a)** Representative gross photographs of lungs from Hu sheep in the healthy (up) and diseased (down) lung phenotype groups. The lungs from the healthy group were elastic and presented a homogeneous pink appearance, whereas those from the diseased group demonstrated severe inflammatory lesions, including extensive red and gray hepatization (indicated using black arrows). **(b)** Representative lung tissue hematoxylin–eosin staining images (magnification: 4×, 10×, and 40×) of healthy and diseased groups. **(c)** Example of pathological scoring of lung tissue (hematoxylin–eosin; magnification: 4×). Lesions are scored as follows: +1, minor; +2, mild; +3, moderate; +4, severe; and +5, extremely severe. In the picture, red triangle denotes partial alveoli retain their original structure; purple red pentagram denotes partial bronchial lumens contain large numbers of neutrophils, lymphocytes (+1); green arrow denotes partial peribronchial lymphoid follicular hyperplasia (+1); blue arrow denotes partial alveolar compensatory dilation (+2); orange arrow denotes inflammatory cell infiltration at the junction of lung interstitium and lung parenchyma (+2); yellow arrow denotes partial collapse and substantial lesions of lung tissue, mainly filled with monocytes (+2)

were selected as candidate pathogenic microorganisms. The linear regression between sequencing results and histopathological scores showed a good positive correlation between the sequencing quantity of *M. ovipneumoniae* and histopathological scores ( $R^2=0.6788$ ,  $p<0.01$ ; Fig. 2b); while such correlation was not noted for *P. multocida* ( $R^2=0.1556$ ,  $p=0.0118$ ; Fig. 2c). Moreover, the sequencing results showed good coverage of

the reference genome of *M. ovipneumoniae* (Fig. 2d); and the genes with high expression levels of *Mycoplasma* were sequenced. Considering the strong conservation of 16S rRNA, which is not conducive to genetic analysis of closely related strains, this study used *LDH* to construct a phylogenetic tree. In the phylogenetic tree, the pathogenic strain detected in this study is *M. ovipneumoniae*, and it has high homology with the isolated strain

**Table 2** Statistical analysis of differences in growth data between healthy and diseased lung groups of sheep

Growth performance <sup>1</sup>	Feeding period (day age)	Healthy group <sup>2</sup> (n = 587)	Diseased group <sup>2</sup> (n = 325)	t-test (p-value)
BW (kg)	80	17.87 ± 4.23	17.44 ± 3.96	0.136954
BW (kg)	180	43.17 ± 7.78	41.09 ± 8.19	0.000149
ADFI (kg/day)	80–180	1.64 ± 0.34	1.54 ± 0.35	2.75E-05
ADG (kg/day)	80–180	0.25 ± 0.05	0.24 ± 0.06	7.33E-06
FCR (kg of FI/kg of BWG)	80–180	6.52 ± 0.86	6.60 ± 1.01	0.203815
Live weight before slaughter (kg)		44.83 ± 7.19	42.64 ± 7.70	4.83E-05
Carcass weight (kg)		24.12 ± 4.44	22.79 ± 4.63	5.41E-05

<sup>1</sup>BW, body weight; ADFI, average daily feed intake; ADG, average daily gain; FCR, feed conversion ratio; BWG, body weight gain

<sup>2</sup>Results are expressed as mean ± standard deviation

**Table 3** Statistical analysis of differences in blood indicators between healthy and diseased lung groups of sheep

Blood parameters <sup>1</sup>	Healthy group <sup>2</sup> (n = 587)	Diseased group <sup>2</sup> (n = 324)	t-test (p-value)
WBC (K/uL)	12.44 ± 2.93	13.05 ± 3.37	0.005
NEUT (K/uL)	5.17 ± 2.16	5.81 ± 2.68	9.24E-05
LYMPH (K/uL)	5.54 ± 1.33	5.52 ± 1.46	0.89
MONO (K/uL)	1.51 ± 0.69	1.48 ± 0.71	0.52
EO (K/uL)	0.13 ± 0.13	0.14 ± 0.14	0.25
BASO (K/uL)	0.09 ± 0.05	0.1 ± 0.06	0.7
NEUT% (%)	40.88 ± 8.57	43.5 ± 10.33	4.4E-05
LYMPH% (%)	45.14 ± 8.28	43.18 ± 9.13	0.001
MONO% (%)	12.16 ± 4.83	11.49 ± 5.13	0.049
EO% (%)	1.06 ± 0.95	1.09 ± 0.9	0.66
BASO% (%)	0.76 ± 0.37	0.74 ± 0.41	0.49

<sup>1</sup>WBC, total white blood count; NEUT, neutrophil; LYMPH, lymphocyte; MONO, monocyte; EO, eosinophil; BASO, basophil

<sup>2</sup>Results are expressed as mean ± standard deviation

(NXNK2203) from Ningxia, which is geographically close (Fig. 2e).

Meanwhile, after further 16S rRNA sequencing analysis of 20 sheep from the 2nd batch mentioned above, we obtained 1,743,066 clean reads (Additional file 6), and then counted ASVs at various taxonomic levels (Additional file 7) along with a preliminary ASVs abundance tables of 10,336 ASVs in lung samples. The average sequencing depth was 169, which helps to obtain reliable microbial community analysis results. These count data could be annotated to 530 genera. These data were then normalized to relative abundance by using the total-sum scaling method. The evaluation indicators of species richness, diversity, and evolutionary diversity, namely the Chao1, Simpson, and Faith's PD indexes, as well as the Pielou's evenness and Good's coverage indexes, revealed no significant differences in lung tissue between groups (Fig. 3a). PCoA based on Bray–Curtis distances revealed differential separation of microbial communities between

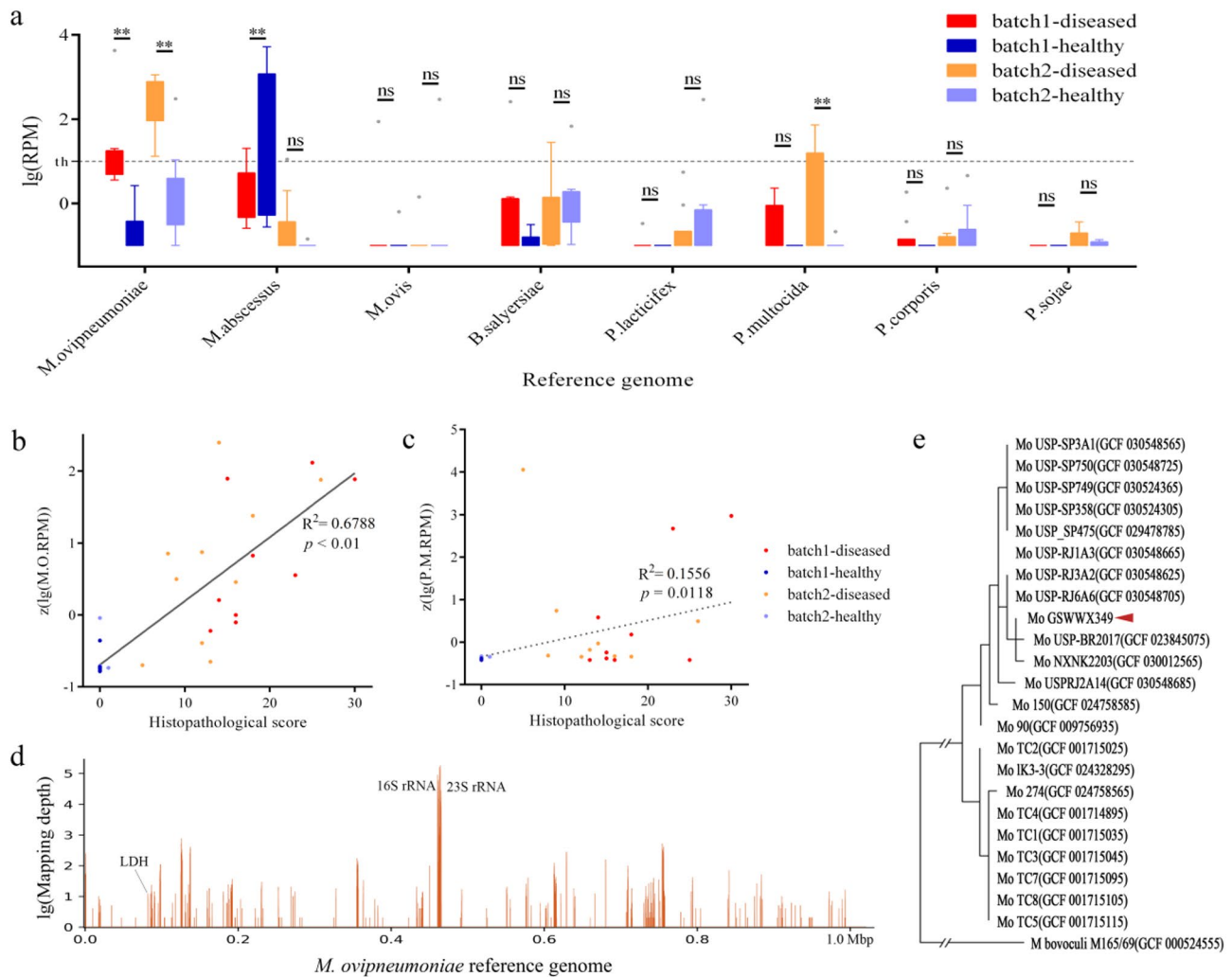
the healthy and diseased groups (Fig. 3b). NMDS analysis also demonstrated between-group separation, with a stress value of 0.131 (Fig. 3c). We also evaluated pulmonary microbiota composition at different taxonomic levels. The most predominant phylum in both healthy and diseased lungs was Bacteroidetes (41.97% and 30.75%, respectively), followed by Firmicutes (34.5% and 24.68%, respectively), Proteobacteria (11.43% and 17.56%, respectively), and Tenericutes (1.19% and 16.51%, respectively; Fig. 3d). Moreover, both healthy and diseased lungs demonstrated a predominance of the genus *Prevotella* (34.36% and 17.73%, respectively); in contrast, compared with diseased lungs, healthy lungs demonstrated significantly lower predominance of *Mycoplasma* (0.82% vs. 16.26%,  $p = 0.057$ ) and *Pasteurella* (0.15% vs. 9.03%,  $p < 0.05$ ; Fig. 3e). A heatmap displayed the distribution and variability of the top 10 bacterial genera in both groups (Fig. 3f). LEfSe analyze, with an LDA threshold set to 4, revealed that the phyla Prevotellaceae and Bacteroidetes were enriched in the healthy group. The genus *Mycoplasma*, the family Mycoplasmataceae, and the order Mycoplasmatales were highly abundant in the diseased group (Fig. 3g).

The correctness and accuracy of our high-throughput sequencing analysis in identifying *M. ovipneumoniae* as a pathogen have been further supported by the laboratory isolation of *M. ovipneumoniae* (strain number GSWWX349), but this result is related to another unpublished study and therefore not shown here.

#### Verification of *M. ovipneumoniae* load in experimental sheep lung tissue

To verify the accuracy of high-throughput sequencing analysis in identifying pathogenic bacteria causing pneumonia, we extracted lung DNA from an expanded sheep population ( $n = 270$ ) and then performed fluorescence quantitative detection of *LDH*, which characterizes the load of *M. ovipneumoniae*. Inter group analysis showed that the *M. ovipneumoniae* load in the diseased group was higher than that in the healthy group (Fig. 4a), and reached an extremely significant level ( $p < 0.01$ ), which is consistent with the results of identifying pathogens using high-throughput sequencing. The immunohistochemical results of *M. ovipneumoniae* in lung tissue showed that the negative control group had substantial lesions and a decrease in the number of alveoli, but the entire lung tissue section did not show positive staining for *M. ovipneumoniae*. After incubating with primary and secondary antibodies, the sheep lung slices showed varying degrees of positive staining (yellow/orange/brown), and positive staining was mainly observed in the cytoplasm of type I and type II alveolar cells in the alveolar septum, indicating the widespread presence of *M. ovipneumoniae* in the alveolar epithelial cells of diseased lungs (Fig. 4b).





**Fig. 2** High-throughput sequencing analysis-based identification of pathogenic microorganisms involved in sheep pneumonia. **(a)** Differential analysis of lung microbial expression between the healthy and diseased groups. **(b, c)** Linear regression between expression of candidate pathogenic microorganisms *M. ovipneumoniae* (M.O., **b**) and *P. multocida* (P.M., **c**) and pulmonary pathological scores. **(d)** Mapping of sequencing results on the *M. ovipneumoniae* reference genome. **(e)** Phylogenetic tree constructed using LDH of *M. ovipneumoniae* obtained through sequencing. Red triangle indicates *M. ovipneumoniae* strain in this study. \*\*  $p < 0.01$ , ns indicates no significant difference between groups, unpaired Student's *t* test

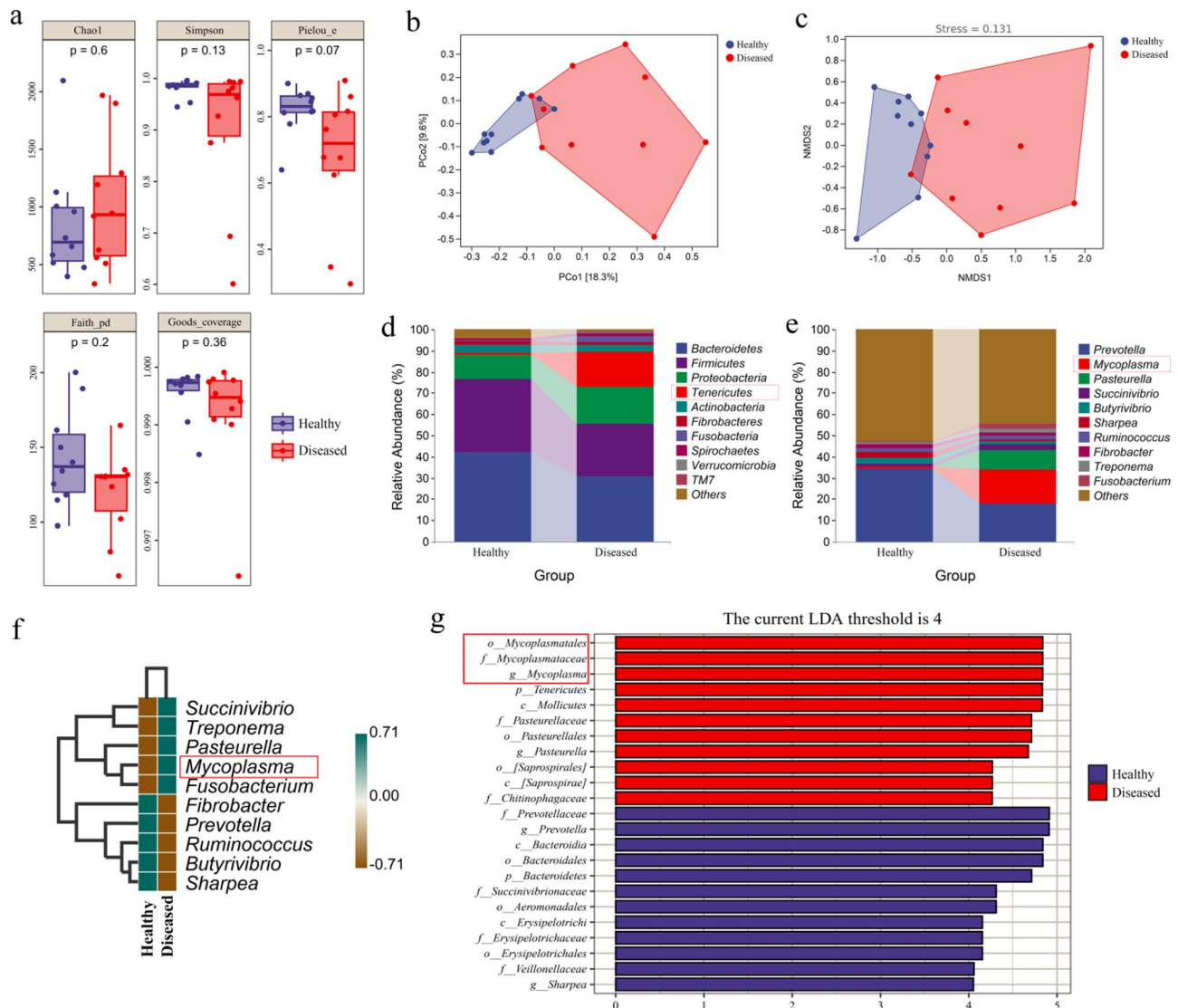
### GWAS of pulmonary pathological scores

We analyzed the relationship between 266 pathological scores of sheep lung and 23,480,428 filtered high-quality SNP markers (Additional file 8). The Q-Q plot (Fig. 5b) demonstrated that some SNPs departed from the expected probability, possibly because of the corresponding changes in the severity scores of pneumonia lesions. In the Manhattan plot, SNPs with  $p$  values below the significance threshold were considered significantly correlated with the phenotypic traits. The GWAS results demonstrated that 59 SNPs ( $p < 10^{-6}$ ) were significantly correlated with pneumonia lesion scores at the genomic level, located on chromosomes 1, 3, 6, 9, 11, 12, 13, 14, 17, 18, 20, and 21 (Fig. 5a). Table 4 presents the annotation information of these 59 significantly correlated SNPs; of them, 10 SNPs were located in the intron or exon regions

of the gene, whereas the remaining SNPs were located in noncoding RNA or intergenic regions. Genome-wide linkage disequilibrium (LD) decay analysis of 912 Hu sheep (Additional file 9) showed that  $r^2$  began to exhibit a trend toward stability when the distance reached 200 kb, suggesting that the genes within 200 kb of the significant SNPs were potential candidates. Finally, except for no genes annotated near the significant SNPs on chromosomes 1 and 18, a total of 30 genes were annotated in the 400-kb region near other significant SNPs in the genome.

### Transcriptomic analysis for screening candidate key genes associated with *M. ovipneumoniae* infection

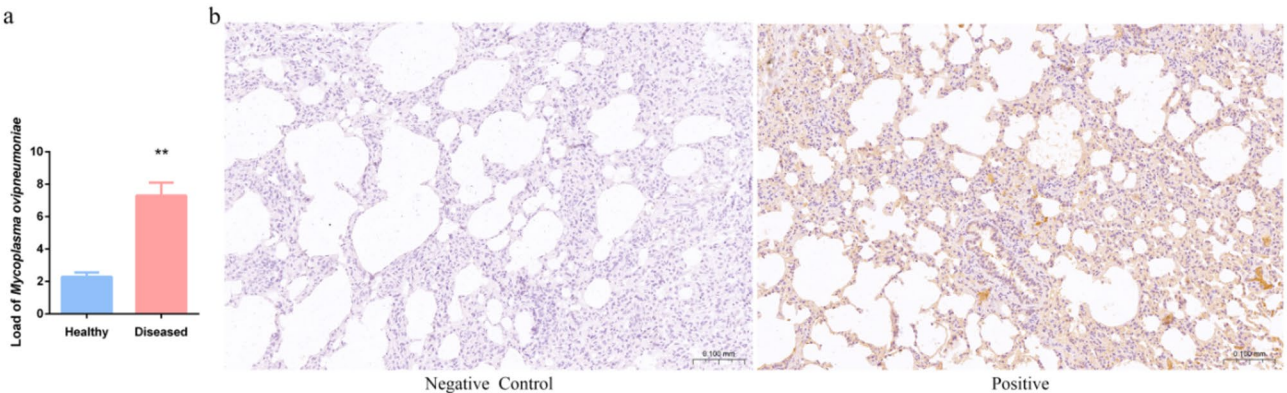
To further investigate the mechanisms underlying resistance to naturally prevalent *Mycoplasma pneumoniae* in our sheep, we performed RNA-Seq on lung tissues from



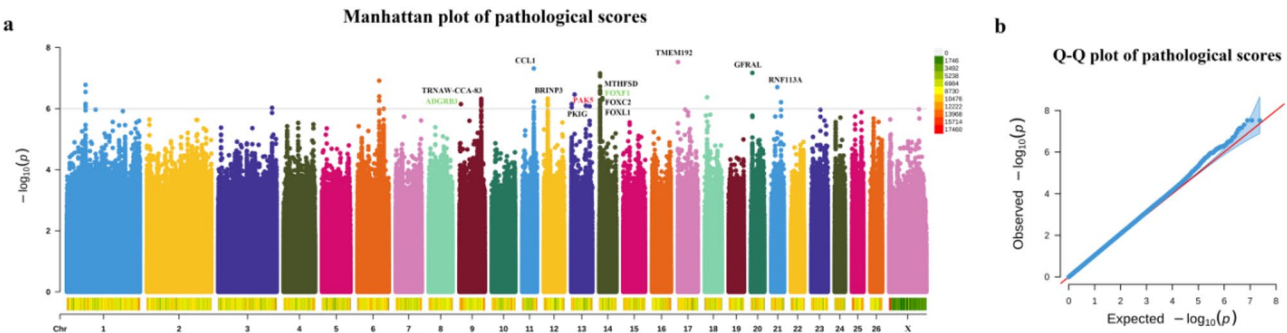
**Fig. 3** 16S rRNA sequencing confirms *M. ovipneumoniae* as a biomarker between healthy and diseased lung groups. **(a)** Microbial community α-diversity between the healthy and diseased groups, including the Chao1, Simpson, Pielou's evenness, Faith's PD, and Good's coverage indexes. **(b, c)** PCoA **(b)** and NMDS **(c)** of the ASVs level based on Bray-Curtis distances between the healthy and diseased groups. **(d)** Bar chart of microbial composition analysis at the phylum level (relative abundance > 1%). The top 10 microbial taxa in terms of abundance are displayed by different color scales. Abundance in a group is the average sample abundance in the group. **(e)** Bar chart of microbial composition analysis at the genus level (relative abundance > 1%). The top 10 microbial taxa in terms of abundance are displayed by different color scales. Abundance in a group is the average sample abundance in the group. **(f)** Heatmap of the distribution and variability of bacteria in the lungs. **(g)** LefSe analysis. The significance threshold is set to an LDA score of 4

both healthy and diseased group sheep. After removing some samples to appropriately cluster them and perform more accurate DEGs analysis between the groups, we retained 19 and 16 samples from the healthy and diseased groups, respectively. We statistically analyzed the distribution of clean reads aligned to the genome (Additional file 10) and calculated gene expression levels. Based on PCA of gene expression, 35 lung samples formed good clusters according to their respective lung phenotype states (Fig. 6a). In addition, by analyzing the genes expressed in the healthy and diseased groups, we identified 2,727 DEGs in the diseased group, and the volcano

plot (Fig. 6b) showed the distribution and expression differences of 1,606 upregulated DEGs and 1,121 downregulated DEGs. Bidirectional clustering analysis achieved clustering of samples with the same phenotype and clustering of DEGs (Fig. 6c). We then subjected the DEGs to GO and KEGG pathway enrichment analyses. GO enrichment analysis results (Fig. 6d) demonstrated that the DEGs were highly enriched in GO terms related to various cellular components. Moreover, they were mainly enriched in GO terms related to molecular functions, such as signaling receptor activity, molecular transducer activity, and transmembrane signaling receptor activity.



**Fig. 4** Verification of *M. ovipneumoniae* load in experimental sheep lung tissue. **(a)** Detection of *M. ovipneumoniae* load between groups in an expanded sheep population (healthy=86, diseased=184). *LDH* gene fluorescence quantitative detection was used to characterize the load of *M. ovipneumoniae*, with  $\beta$ -tubulin gene as an endogenous control. **(b)** Immunohistochemical detection of *M. ovipneumoniae* in the lungs of diseased group. A negative control represents the absence of incubation of *M. ovipneumoniae* primary antibody, while a positive indicates normal incubation of *M. ovipneumoniae* primary antibody. The yellow color in the field of view of the positive picture indicates positive staining of *M. ovipneumoniae*, and the darker the staining, the stronger the positive intensity. The data are presented as means  $\pm$  standard errors of the means ( $n=3$ ). \*\*  $p < 0.01$ , unpaired Student's  $t$  test. *LDH*: lactate dehydrogenase



**Fig. 5** GWAS of pulmonary pathological scores. **(a)** Manhattan plot of the GWAS results, with the threshold line (gray line) set at 6. The genes annotated in the figure are those annotated within 200 kb upstream and downstream of the significant loci on the threshold line. The genes denoted in red and green are those significantly upregulated and downregulated in the diseased lung group in differentially expressed gene analysis, respectively. **(b)** Q-Q plot of the GWAS result. The red line represents the expected statistical distribution, the blue dot represents the actual observed statistical distribution, and the light blue area represents the confidence interval

Finally, they were enriched in GO terms related to biological processes, including stimulus response, cell adhesion, and multicellular organismal process regulation. In the KEGG pathway enrichment analysis (Fig. 6e), we noted that the DEGs were highly enriched in crucial pathways including viral carcinogenesis, epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor resistance, axon guidance, fluid shear stress and atherosclerosis, and nucleocytoplasmic transport. The inter group expression of 10 randomly selected DEGs (*PROM1*, *PAK5*, *SLIT1*, *PGAM1*, *UBTD1*, *ADGRB3*, *FOXF1*, *PRDX6*, *PTGS1*, and *SMAD7*) showed consistent differential trends with our RNA-Seq results (Fig. 6f), indicating the reliability of our RNA-Seq data.

Finally, through overlap analysis, we identified three annotated DEGs (*ADGRB3*, *PAK5*, and *FOXF1*) from GWAS that had 5 significant SNPs. Compared with the healthy group, the expression of *PAK5* was increased in

diseased lung tissue, while the expression of *ADGRB3* and *FOXF1* was decreased.

#### Different genotypes of important SNPs affect *FOXF1* gene expression and individual pneumonia phenotype

Five SNPs were genotyped through KASPar, and 2 significant SNPs (chr14: g. 13215150 A>G, chr14: g. 13215157 A>G) were successfully validated (Additional file 11). Both SNPs are annotated to the same differentially expressed gene *FOXF1*. According to the different genotypes of each important SNP locus, the expanded sheep were divided into three groups, and the expression of the candidate key gene *FOXF1* and the sheep lung phenotype (including *M. ovipneumoniae* load and lung pathological score) were detected in each group.

The association analysis between genotype and *M. ovipneumoniae* load showed that at the mutation site g. 13,215,150 A>G on chromosome 14, the expression level



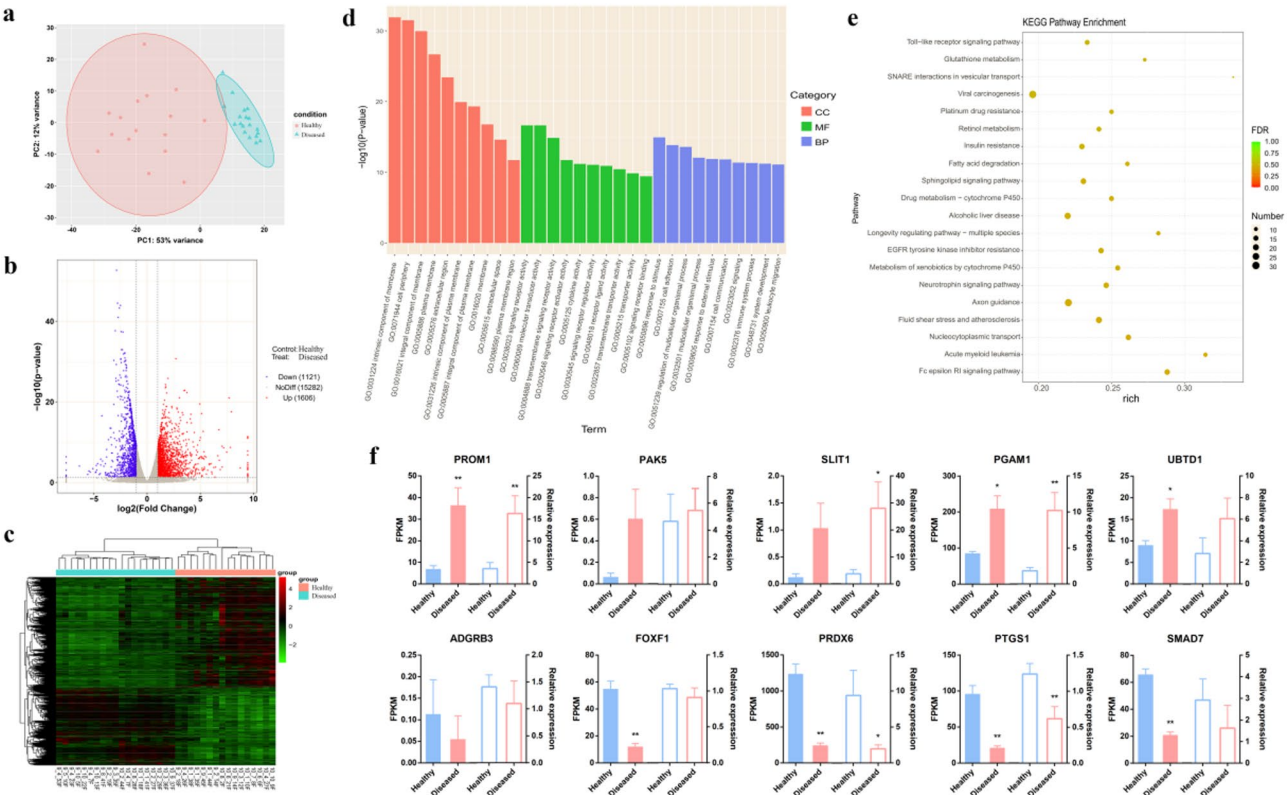
**Table 4** Information list of SNPs associated with the pathological score of sheep lung by GWAS

Chr	SNP Name	Position	p-Value	Site annotation	Adjacent genes <sup>1</sup> (± 200 kb)
1	1_77188321	77,188,321	8.80E-07	intergenic	
1	1_77198094	77,198,094	2.85E-07	intergenic	
1	1_77199179	77,199,179	1.66E-07	intergenic	
1	1_77199507	77,199,507	6.98E-07	intergenic	
1	1_77248496	77,248,496	8.14E-07	intergenic	
3	3_221919077	221,919,077	9.44E-07	ncRNA_intronic	LOC114113822
6	6_90057882	90,057,882	5.51E-07	intergenic	LOC101120245
6	6_90064693	90,064,693	1.22E-07	intergenic	
6	6_90086998	90,086,998	4.00E-07	intergenic	
9	9_6068850	6,068,850	7.12E-07	intergenic	<b>ADGRB3</b>
9	9_89666126	89,666,126	5.99E-07	intergenic	TRNAW-CCA-83
9	9_89666200	89,666,200	5.22E-07	intergenic	
9	9_89666393	89,666,393	7.87E-07	intergenic	
9	9_89666456	89,666,456	4.77E-07	intergenic	
9	9_89667773	89,667,773	6.97E-07	intergenic	
9	9_89667849	89,667,849	6.19E-07	intergenic	
9	9_89667936	89,667,936	7.00E-07	intergenic	
11	11_47969453	47,969,453	4.84E-08	intergenic	LOC114117035
11	11_47971750	47,971,750	8.99E-07	intergenic	LOC101120093
11	11_47996946	47,996,946	5.96E-07	intergenic	LOC101119832
					LOC101119572
					CCL1
					LOC101114373
12	12_16258615	16,258,615	5.41E-07	intronic	BRINP3
12	12_16264379	16,264,379	8.06E-07	intronic	BRINP3
12	12_16265274	16,265,274	6.87E-07	intronic	BRINP3
12	12_16268553	16,268,553	4.68E-07	intronic	BRINP3
12	12_16270305	16,270,305	7.68E-07	intronic	BRINP3
13	13_2458671	2,458,671	7.12E-07	intronic	<b>PAK5</b>
13	13_2458683	2,458,683	9.16E-07	intronic	<b>PAK5</b>
13	13_14289597	14,289,597	3.44E-07	intergenic	LOC114117710
13	13_60723049	60,723,049	8.07E-07	intergenic	C13H20orf85
					PMEPA1
					LOC105606212
					ZBP1
13	13_76130160	76,130,160	8.40E-07	intronic	PKIG
14	14_3800439	3,800,439	3.11E-07	intergenic	LOC114117782
14	14_3801288	3,801,288	6.85E-07	intergenic	
14	14_3801501	3,801,501	5.51E-07	intergenic	
14	14_3801715	3,801,715	5.25E-07	intergenic	
14	14_3801758	3,801,758	8.72E-08	intergenic	
14	14_3801759	3,801,759	7.05E-08	intergenic	
14	14_3802541	3,802,541	5.35E-07	intergenic	
14	14_3802553	3,802,553	5.49E-07	intergenic	
14	14_3802601	3,802,601	3.24E-07	intergenic	
14	14_3802602	3,802,602	3.24E-07	intergenic	
14	14_3802645	3,802,645	6.75E-07	intergenic	
14	14_3803552	3,803,552	8.57E-07	intergenic	
14	14_3804157	3,804,157	9.44E-07	intergenic	
14	14_3805404	3,805,404	1.89E-07	intergenic	
14	14_3805630	3,805,630	2.86E-07	intergenic	
14	14_3805762	3,805,762	5.28E-07	intergenic	
14	14_3805955	3,805,955	2.52E-07	intergenic	

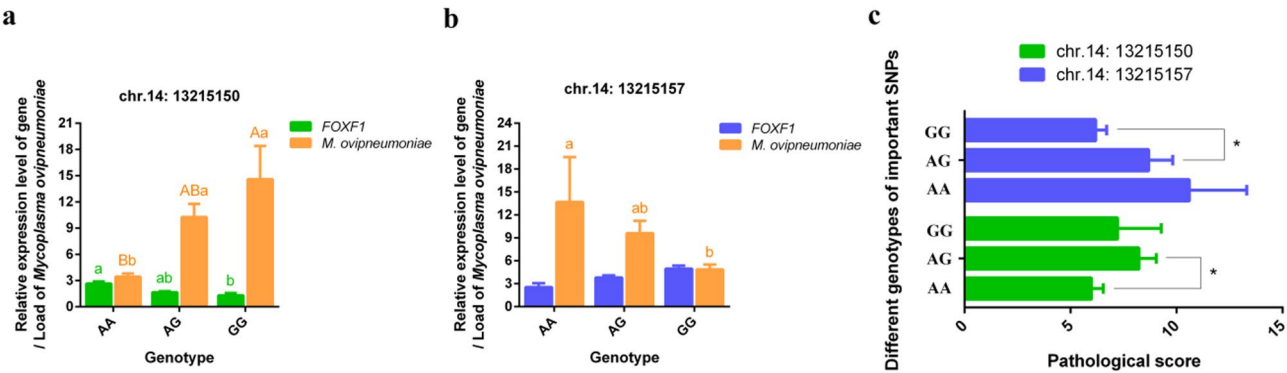
Table 4 (continued)

Chr	SNP Name	Position	p-Value	Site annotation	Adjacent genes <sup>1</sup> (± 200 kb)
14	14_13215150	13,215,150	4.53E-07	intergenic	<b><i>FOXF1</i></b>
14	14_13215157	13,215,157	5.09E-07	intergenic	<i>MTHFSD</i> <i>FOXC2</i> <i>FOXL1</i>
17	17_433379	433,379	2.98E-08	intergenic	LOC114108875
17	17_433393	433,393	2.98E-08	intergenic	LOC114108868
17	17_433402	433,402	2.98E-08	intergenic	LOC105605790 <i>TMEM192</i>
18	18_9841714	9,841,714	4.28E-07	intergenic	
20	20_6437379	6,437,379	6.76E-08	intronic	<i>GFRAL</i>
21	21_25024242	25,024,242	1.98E-07	exonic	<i>RNF113A</i>
21	21_40390544	40,390,544	6.17E-07	ncRNA_intronic	LOC114110086
21	21_40390562	40,390,562	6.17E-07	ncRNA_intronic	
21	21_40390571	40,390,571	6.17E-07	ncRNA_intronic	
21	21_40390576	40,390,576	6.17E-07	ncRNA_intronic	

<sup>1</sup> The genes highlighted in bold and underlined indicate genes that are downregulated or upregulated in the diseased lung group of RNA-Seq differentially expressed genes analysis



**Fig. 6** Transcriptomic analysis for candidate key genes associated with *M. ovipneumoniae* infection. **(a)** PCA of RNA-Seq data in the healthy and diseased groups. **(b)** A volcano plot presenting DEGs between the healthy and diseased groups. **(c)** Heatmap showing bidirectional clustering of DEGs and samples. **(d)** GO enrichment analysis of DEGs. GO classification is based on cellular components (CC), molecular functions (MF), and biological processes (BP). The top 10 GO term entries with the smallest *p* value, i.e., the most significant enrichment, in each GO classification were selected for display. **(e)** KEGG pathway enrichment analysis of DEGs. The larger the bubble, the greater the degree of enrichment. Closer FDR values to zero indicate more significant enrichment. The top 20 KEGG pathways with the smallest FDR values, i.e., the most significant enrichment, were selected for display. **(f)** Verification of RNA-Seq results through qRT-PCR. Relative expression levels calculated from standard curves were normalized to the  $\beta$ -actin gene, as the endogenous control. The data are presented as means  $\pm$  standard errors of the means ( $n=3$ ). \*  $p<0.05$ , \*\*  $p<0.01$ , unpaired Student's *t* test. *PROM1*: prominin 1; *PAK5*: p21 (RAC1) activated kinase 5; *SLIT1*: slit guidance ligand 1; *PGAM1*: phosphoglycerate mutase 1; *UBTD1*: ubiquitin domain containing 1; *ADGRB3*: adhesion G protein-coupled receptor B3; *FOXF1*: forkhead box F1; *PRDX6*: peroxiredoxin 6; *PTGS1*: prostaglandin-endoperoxide synthase 1; *SMAD7*: SMAD family member 7



**Fig. 7** Different genotypes of important SNPs affect *FOXF1* gene expression and individual pneumonia phenotype. **(a)** The expression level of candidate *FOXF1* gene and *M. ovipneumoniae* load in the lungs of individuals with different genotypes (*FOXF1*: AA=116, AG=78, GG=9; *LDH*: AA=114, AG=74, GG=9) of SNP loci (chr.14: 13215150). **(b)** The expression level of candidate *FOXF1* gene and *M. ovipneumoniae* load in the lungs of individuals with different genotypes (*FOXF1*: AA=5, AG=51, GG=138; *LDH*: AA=5, AG=50, GG=134) of SNP loci (chr.14: 13215157). **(c)** The impact of different genotypes of important SNPs (chr.14: 13215150, AA=116, AG=78, GG=9; chr.14: 13215157, AA=5, AG=51, GG=138) on their pulmonary pathological scores. Relative expression levels of *FOXF1* calculated from standard curves were normalized to the  $\beta$ -actin gene, as the endogenous control. *LDH* gene fluorescence quantitative detection was used to characterize the load of *M. ovipneumoniae*, with  $\beta$ -tubulin gene as an endogenous control. The data are presented as means  $\pm$  standard errors of the means ( $n=3$ ). Different superscript lowercase and capital letters respectively showed significant difference ( $P<0.05$ ) and extremely significant difference ( $P<0.01$ ), one-way ANOVA and Dunnett's test. \*  $p<0.05$ , unpaired Student's  $t$  test. *FOXF1*: forkhead box F1; *LDH*: lactate dehydrogenase

**Table 5** Significant SNP and their association analysis with blood parameters

Blood parameters <sup>1</sup>	chr14 g. 13,215,150 A>G		
	AA <sup>2</sup> (n=500)	AG <sup>2</sup> (n=281)	GG <sup>2</sup> (n=42)
WBC (K/uL)	12.59 $\pm$ 3.13	12.73 $\pm$ 3.22	12.92 $\pm$ 2.32
NEUT (K/uL)	5.3 $\pm$ 2.33	5.5 $\pm$ 2.6	5.37 $\pm$ 1.67
LYMPH (K/uL)	5.53 $\pm$ 1.37	5.57 $\pm$ 1.42	5.63 $\pm$ 1.34
MONO (K/uL)	1.53 $\pm$ 0.71 <sup>ab</sup>	1.45 $\pm$ 0.7 <sup>b</sup>	1.67 $\pm$ 0.63 <sup>a</sup>
EO (K/uL)	0.14 $\pm$ 0.15	0.12 $\pm$ 0.09	0.15 $\pm$ 0.12
BASO (K/uL)	0.091 $\pm$ 0.05	0.098 $\pm$ 0.05	0.103 $\pm$ 0.05
NEUT% (%)	41.33 $\pm$ 9.15	42.19 $\pm$ 9.73	41.28 $\pm$ 8.67
LYMPH% (%)	44.67 $\pm$ 8.56	44.57 $\pm$ 9.23	43.77 $\pm$ 7.47
MONO% (%)	12.14 $\pm$ 4.82 <sup>ab</sup>	11.48 $\pm$ 5.12 <sup>b</sup>	13.03 $\pm$ 4.27 <sup>a</sup>
EO% (%)	1.12 $\pm$ 1.12	0.97 $\pm$ 0.6	1.11 $\pm$ 0.83
BASO% (%)	0.74 $\pm$ 0.38	0.79 $\pm$ 0.4	0.81 $\pm$ 0.41

<sup>1</sup>WBC, total white blood count; NEUT, neutrophil; LYMPH, lymphocyte; MONO, monocyte; EO, eosinophil; BASO, basophil

<sup>2</sup>Results are expressed as mean  $\pm$  standard deviation. Different superscript lowercase and capital letters respectively showed significant difference ( $P<0.05$ ) and extremely significant difference ( $P<0.01$ )

of *FOXF1* gradually decreased in individuals with AA, AG, and GG genotypes, while the *M. ovipneumoniae* load gradually increased (Fig. 7a). At another mutation site g. 13,215,157 A>G on chromosome 14, *FOXF1* expression gradually increased in individuals with AA, AG, and GG genotypes, while the *M. ovipneumoniae* load gradually decreased (Fig. 7b). According to the genotyping results, sheep individuals exhibit AA genotype at chr.14: 13,215,150 site, while most of them also exhibit GG genotype at chr.14: 13,215,157 site. At this time, *FOXF1* expression is the highest compared to other genotype individuals, while the *M. ovipneumoniae* load

is the lowest. This is consistent with the high expression of *FOXF1* in healthy lung tissue observed in RNA-Seq analysis. This correlation trend was also validated in the association analysis between genotype and another lung phenotype trait, pathological score (Fig. 7c). What's more, the analysis of blood physiological indicators for sheep expanding population showed that individuals with GG genotype at the mutation site g. 13,215,150 A>G generally had higher inflammation-related indices (Table 5). In summary, we speculate that *FOXF1* may serve as a resistance gene for *Mycoplasma pneumonia* in sheep.

Discussion

Recently, sheep pneumonia incidence has increased in China [36]. Hu sheep is a local breed of sheep raised in China, known for its early maturity and high reproductive efficiency [37]. From August 2022 to January 2023, Hu sheep reared at a study sheep farm developed pneumonia with an increased incidence of inflammatory lesions in the lungs, but no deaths. Such a situation urgently requires a rational assessment of the epidemiologic etiology of pneumonia in this sheep farm.

When describing chronic or acute sheep pneumonia, it is necessary to consider mortality and subclinical effects, as chronic pneumonia may gradually recover, while acute pneumonia has severe symptoms and may even lead to death [14, 38]. The detection of chronic nonprogressive pneumonia usually requires autopsy or postslaughter lung examination. Compared with experimental infection, natural pneumonia occurrence more effectively reflects the disease characteristics under commercial sheep-breeding conditions. Our study strictly



starts from the macroscopic of clinical autopsy and histopathological microscopic aspects to classify the phenotype of sheep lungs to distinguish between healthy or inflammatory lesions. We observed pathological changes in sheep lungs, such as lung consolidation, inflammatory cell infiltration, and vascular lesions—similar to the *M. ovipneumoniae* infection. We confirmed the negative impact of pneumonia on sheep growth and development based on phenotype grouping analysis. This is consistent with reports that some sheep infected with pneumonia have lower ADG, FCR, live weight, and carcass yield [39]. Blood physiological indicators can reflect the health status of organisms from an immunological perspective and are important indicators for disease screening and diagnosis [40]. We found that the white blood cell, neutrophil and its proportion in diseased sheep increased, which is consistent with previous reports related to pneumonia [41, 42]. The cost of pneumonia has prompted the search for effective solutions to combat the disease, and research aims to explore potential resistance genes to cultivate advantageous livestock and poultry breeds with considerable resistance to the disease. Identifying the pathogen of pneumonia is the first step in disease resistance research. But the determination of the pathogen cannot rely solely on pathological examination.

Traditional veterinary pathogen identification first involves the isolation, purification, and cultivation of pathogens, and then relies on biochemical testing, Sanger sequencing, serological testing, and qRT-PCR to identify pathogens [38, 43]. This undoubtedly requires longer cycles and specific culture media, as well as consideration for small and difficult to culture pathogens such as *Mycoplasma*. It seems promising to achieve high-resolution rapid detection of clinical samples. The rapid development of high-throughput sequencing technology has led to the widespread use of WGS to characterize microbial genomes, especially for clinical pathogen identification and bioinformatics analysis [44, 45]. Respiratory infectious diseases are caused by pathogens invading the respiratory system [46], and although their colonization ability varies by tissue, nasal and lung isolates represent similar biological populations [47]. Direct sequencing of lung tissue not only provides information about the lung microbiome [48], but also obtains a large amount of host genome sequences. It is crucial to use high-throughput sequencing technology to design pathogen identification schemes suitable for different studies when dealing with clinical samples of unknown pathogens. In this study, considering the high transcription levels of active pathogens in diseased tissues, we anticipate that the proportion of microbial sequences in RNA sequencing results will be high. Therefore, we performed high-throughput sequencing of sheep lungs following the lncRNA analysis process. After filtering out the host genome, the

obtained exogenous RNA fragments were compared with the microbial genome database, and it was determined that the pathogen of farm epidemic pneumonia was *M. ovipneumoniae*. The physiological and pathological status of the lungs varies depending on the richness and abundance of the microbiota [49, 50]. The increase in the number of dominant pathogenic bacteria in the lungs leads to the occurrence and deterioration of pneumonia [51]. 16S rRNA sequencing confirmed significant enrichment of *Prevotella* (34.36%, 17.73%) in the lungs of healthy sheep [52]. The species composition and LEfSe analysis showed significant differences in the abundance and relative contribution of *Mycoplasma* between the healthy and diseased groups, indicating that it is a biomarker for pneumonia lesions. We wanted to verify this conclusion through experiments. Considering that *M. ovipneumoniae* is difficult to count using traditional agar plates, a fluorescent quantitative assay protocol for the *LDH* gene was specifically designed to characterize the load of *M. ovipneumoniae* based on the genome sequencing sequence of the pathogenic strain in this study. The load detection and immunohistochemical localization of *M. ovipneumoniae* confirmed the accuracy of our high-throughput sequencing analysis in identifying the pathogen. High-throughput sequencing characterizes all RNA or DNA present in the sample, enabling analysis of the entire microbiome, transcriptome, or host genome in diseased animal samples. Andriyan et al. [53], sequenced the DNA of poultry samples infected with West Nile Virus (WNV) collected in the United States between 2006 and 2007, and identified a new genetic variant strain containing 13 nucleotide deletions. Chen et al. [54], accidentally discovered a new coronavirus during a survey of domestic poultry using RNA-Seq. Sequencing technology has successfully realized qualitative and quantitative analysis of pathogens, especially diagnosis and treatment of difficult and critical infectious diseases and identification of new unknown pathogens, with its advantages of high flux, high accuracy, etc. It can also be applied to drug resistance gene monitoring, epidemiological tracking and investigation, and has incomparable advantages over traditional pathogenic identification technology [8].

In recent years, about 90% of lambs with chronic bronchopneumonia have been noted to be positive for pulmonary *Mycoplasma* after slaughter [55]. In 2001, *M. ovipneumoniae* was detected in 88% of 453 sheep farms in the United States [56]. Over 2007–2019, 16.4% of *Mycoplasma* strains isolated from small ruminants were identified to be *M. ovipneumoniae* in France [57]. In 2021, the nasal swabs from >40% of sheep in Xinjiang, a major sheep-breeding area in China, tested positive for *M. ovipneumoniae* [58]. *M. ovipneumoniae* is prevalent worldwide, mainly transmitted through close and repeated respiratory contact, and can persist in the

environment and is difficult to eliminate [59]. All our studied sheep shared an identical environment (including fence, water, and feed), which suggests that the differences in pathological changes in sheep lungs are not due to differences in environmental microorganisms, possibly indicating that *M. ovipneumoniae* infection resistance or recovery abilities can differ among sheep populations fed in identical environments.

The resistance and resilience of animals to infectious diseases is a complex trait with varying genetic patterns. Effective research strategies such as GWAS are needed to analyze the genetic basis of complex traits [60], which can analyze the associations of genetic markers representing phenotypes with variations in a set of DNA samples [61]. The GWAS tool has identified polymorphic loci or genes associated with important economic traits in the sheep genome, such as reproductive quality [62], wool [63], milk [64], and meat yield [65]. However, GWAS-based research related to disease resilience in breeding programs has been insufficient thus far [66]. Bacterial load cannot determine the severity of lesions, and uneven bacterial colonization also makes it difficult to obtain accurate load data. We believe that an intuitive judgment of the severity of lesions can better understand the sheep's response and tolerance to *M. ovipneumoniae* infection. The GWAS results of our pathological scoring identified genomic regions and loci associated with *Mycoplasma* pneumonia lesions, and annotated key genes. To screen for the most promising candidate genes for disease resistance and conduct targeted functional studies, we chose a multi-omics research approach. The integrated analysis of multi-omics such as whole genome/transcriptome has successfully identified some disease resistance genes in the past [67, 68]. Our comparative transcriptome analysis of healthy and diseased lung in study sheep population validated population segregation based on lung phenotypes. The GO enrichment results of the 2727 identified DEGs include response to stimulus, cell adhesion, and signaling receptor activity, which is consistent with the pathogenic mechanisms of *Mycoplasma*, as adhesion is the first step of *Mycoplasma* invasion into the host, thereby creating conditions for subsequent colonization and infection by *Mycoplasma* [69]. In addition, respiratory epithelial cells, as the first barrier for defense mechanisms in animals, have the functions of synthesizing and secreting mucin, degrading/inhibiting inflammatory mediators, regulating local immune responses, exerting physiological or pathological regulatory effects, and signaling across membranes [70]. When *Mycoplasma* crosses respiratory epithelial cells, it triggers innate immune responses by producing various inflammatory mediators, thereby mediating lung inflammation [71]. This is also consistent with the results showing that DEGs are enriched in the immune system

and leukocyte migration. DEGs are enriched in KEGG pathways such as Fc epsilon RI signaling pathway, Axon guidance, Toll-like receptor signaling pathway, etc. In particular, there have been many reports on the inflammatory response triggered by *Mycoplasma* invasion into host cells through the Toll-like receptor signaling pathway [17, 72]. The enriched pathways identified in our study are also worthy of further investigation. Based on GWAS and transcriptome, we further screened three overlapping DEGs: *ADRB3*, *PAK5*, and *FOXF1*, as well as five significant SNPs of these annotated genes. Two SNPs with mutations were successfully identified through genotyping. Further analysis was conducted on the different genotypes of these two SNPs and their annotated *FOXF1* gene expression, as well as their relationship with individual pneumonia phenotypes. We found that the *M. ovipneumoniae* load was lowest and pulmonary pathological score was lowest in genotypes with high *FOXF1* expression, which is fully consistent with the results of transcriptome analysis and further indicates the dominant genotype associated with resistance to *M. ovipneumoniae*. In summary, our study identified the key gene *FOXF1* and 2 important SNPs (chr.14: g. 13215150 A>G, chr.14: g. 13215157 A>G) associated with resistance to *M. ovipneumonia* lesions.

In recent years, the genetic mechanisms and candidate genes related to resistance traits of *Mycoplasma* pneumonia in sheep have received increasing attention from scholars both domestically and internationally. Cao et al. [73] have conducted an in-depth analysis of genomic data from 3,938 sheep samples and found that the introgressed alleles in a specific region of *PADI2* (chr2: 248,302,667–248,306,614) are associated with anti-pneumonia. Kathryn et al. [23] discovered five SNPs linked to pneumonic lesions through GWAS, with 37 more showing possible significance. Four SNPs were tied to pleurisy, and 11 others were of potential significance. These SNPs are located near genes that play roles in DNA repair and immune defense, some of which have known connections to respiratory diseases. Our findings contribute new SNPs and candidate resistance genes to the study of resistance to *Mycoplasma* pneumonia in sheep. It is speculated that these newly discovered SNPs may affect sheep's resistance to *Mycoplasma* pneumonia and their individual blood immune status by regulating the expression of *FOXF1* gene. *FOXF1*, a transcription factor forkhead box (FOX) family member, is a transcription factor with a role in organ development and acute injury. *FOXF1* expression is significantly higher in pulmonary endothelial cells than in endothelial cells in other organs, such as the brain, heart, pancreas, liver, and kidneys [74, 75]; however, lung injury can reduce *FOXF1* expression in pulmonary endothelial cells [76]. *FOXF1* regulates key genes involved in inflammation, extracellular matrix

remodeling, and endothelial barrier function, thereby stimulating lung repair and regeneration [77]. *FOXF1*-mediated BMP9/ACVRL1 signaling activates lung epithelial progenitor cells, promoting neonatal pulmonary angiogenesis and alveolarization [78]. It has also been proven that *FOXF1*, as an antifibrotic factor, inhibits pulmonary fibrosis by preventing the conversion of CDH2 to CDH11 cadherin in myofibroblasts [79]. Numerous studies on *FOXF1* in the lungs have emphasized its regulatory function in lung diseases, which also supports our conclusion of screening *FOXF1* as a candidate gene for resistance to *Mycoplasma pneumonia* in sheep through multi-omics methods. However, our work also has limitations. The functional validation of the *FOXF1* gene in anti-*Mycoplasma pneumonia* at the cellular level and its resistance mechanism and regulatory pathway against *M. ovipneumonia* infection are still unknown, which deserves further research.

The research strategy based on multi-omics sequencing for pathogen identification has enriched and rapidly achieved direct and accurate detection of pathogens. Integrating multi-omics data to mine key genes can be used for future disease resistant breeding, but different omics methods and experimental designs make it difficult to integrate and extract meaningful information from the data, and strict requirements are placed on the collection of phenotype data. We believe that the standardization of such research strategies will promote the development of disease resistant breeding in the livestock and poultry industry.

Conclusions

Our study identified *M. ovipneumoniae* as the primary pathogen responsible for epidemic pneumonia in sheep farms through high-throughput sequencing, pathogen load quantification in tissues, and immunohistochemical assays. We identified genomic mutation sites associated with *Mycoplasma pneumonia* through GWAS of pulmonary pathological scores. Especially by integrating multi-omics analysis and association analysis of pneumonia phenotype, two important SNPs related to *Mycoplasma pneumonia* and a candidate gene *FOXF1* for anti-*Mycoplasma pneumonia* were screened and validated.

Abbreviations

SNPs	Single nucleotide polymorphisms
NGS	Next generation sequencing
<i>M. ovipneumoniae</i>	<i>Mycoplasma ovipneumoniae</i>
GWAS	Genome-wide association studies
WGS	Whole genome sequencing
BW	Body weight
FI	Feed intake
ADFI	Average daily FI
BWG	BW gain
ADG	Average daily BW gain
FCR	Feed conversion rate
GATK	Genomic analysis toolkit

LDH	Lactate dehydrogenase
qRT-PCR	Real-time reverse transcription polymerase chain reaction
ASVs	Amplicon sequence variants
PCoA	Principal coordinate analysis
NMDS	Nonmetric multidimensional scaling
LEfSe	Linear discriminant analysis effect size
qPCR	Fluorescence quantitative PCR
Q-Q plots	Quantile–quantile plots
FPKM	Fragments per kilobase of exon per million fragments mapped
PCA	Principal component analysis
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
DEGs	Differentially expressed genes
KASPar	Competitive allele-specific-based polymerase chain reaction
LYMPH	Lymphocyte
MONO	Monocyte
EO	Eosinophil

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11699-3>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7
Supplementary Material 8
Supplementary Material 9
Supplementary Material 10
Supplementary Material 11

Acknowledgements

We would like to thank the native English speaking scientists of Elixigen Company (Huntington Beach, California) for editing our manuscript.

Author contributions

W.W.M., P.J. and H.K. designed the experiments. H.K., L.J., and X.D. performed the experiments; Y.L.F. and L.X.L. analysed the data. H.K., P.J., Z.X.X., T.H.B. and L.F.D. drafted and revised the manuscript and H.K. was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Innovation 2030-Major Project of Agricultural Biotech Breeding (grant number 2022ZD04013).

Data availability

The raw sequence data (GSA: CRA022717) reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, Chinese Academy of Sciences (<https://ngdc.cncb.ac.cn/gsa>). Further data requests and information can be obtained by contacting the corresponding author Weimin Wang at wangweimin@lzu.edu.cn.

Declarations

Ethics approval and consent to participate

All animal experiments and procedures were conducted in accordance with Chinese laws and institutional guidelines, and have been approved



by the Ethics Committee of Lanzhou University (approval number: 2021-02). In addition, all samples were collected specifically for this study following standard procedures with the informed consent of the the Minqin experimental farm of Lanzhou University who owned the animals.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Received: 25 January 2025 / Accepted: 9 May 2025

Published online: 20 May 2025

#### References

- Lewis CR, Ait-Ali T, Clapperton M, Archibald AL, Bishop S. Genetic perspectives on host responses to porcine reproductive and respiratory syndrome (PRRS). *Viral Immunol*. 2007;20(3):343–58.
- Jesse FFA, Amira NA, Isa KM, Maqbool A, Ali NM, Chung ELT, et al. Association between Mannheimia haemolytica infection with reproductive physiology and performance in small ruminants: a review. *Veterinary World*. 2019;12(7):978–83.
- McRae KM, Baird HJ, Dodds KG, Bixley MJ, Clarke SM. Incidence and heritability of ovine pneumonia, and the relationship with production traits in New Zealand sheep. *Small Ruminant Res*. 2016;145:136–41.
- Butler CJ, Edwards WH, Paterson JT, Proffitt KM, Jennings-Gaines JE, Killion HJ, et al. Respiratory pathogens and their association with population performance in Montana and Wyoming bighorn sheep populations. *PLoS ONE*. 2018;13(11):e0207780.
- Dassanayake RP, Shanthalingam S, Subramaniam R, Herndon CN, Bavananthasivam J, Haldorson GJ, et al. Role of Bibersteinia trehalosi, respiratory syncytial virus, and parainfluenza-3 virus in bighorn sheep pneumonia. *Vet Microbiol*. 2013;162(1):166–72.
- Miao Y, Zhao X, Lei J, Ding J, Feng H, Wu K et al. Characterization of lung microbiomes in pneumonic Hu sheep using culture technique and 16S rRNA gene sequencing. *Animals: an open access journal from MDPI*. 2023;13(17).
- Han J, Kang Z, Xie Y, Li H, Yan H, Song X. Acute diffuse edematous-hemorrhagic Epstein-Barr virus meningoencephalitis: a case report. *Medicine*. 2019;98(51):e18070.
- Brown JR, Bharucha T, Breuer J. Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases. *J Infect*. 2018;76(3):225–40.
- Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res*. 2014;24(7):1180–92.
- Glendinning L, Collie D, Wright S, Rutherford KMD, McLachlan G. Comparing microbiotas in the upper aerodigestive and lower respiratory tracts of lambs. *Microbiome*. 2017;5(1):145.
- Glendinning L, Wright S, Pollock J, Tennant P, Collie D, McLachlan G. Variability of the sheep lung microbiota. *Appl Environ Microbiol*. 2016;82(11):3225–38.
- Rathay R, Paul S, Ponesakki V, Kanniah P, Muthu SP, Arumugaperumal A et al. Data on genome sequencing, analysis and annotation of a pathogenic *Bacillus cereus* 062011msu. *Data in brief*. 2018; 17:15–23.
- Besser TE, Levy J, Ackerman M, Nelson D, Manlove K, Potter KA, et al. A pilot study of the effects of *Mycoplasma ovipneumoniae* exposure on domestic lamb growth and performance. *PLoS ONE*. 2019;14(2):e0207420.
- Besser TE, Frances Cassirer E, Highland MA, Wolff P, Justice-Allen A, Mansfield K, et al. Bighorn sheep pneumonia: sorting out the cause of a polymicrobial disease. *Prev Vet Med*. 2013;108(2–3):85–93.
- Gonçalves R, Mariano I, Núñez A, Branco S, Fairfoul G, Nicholas R. Atypical non-progressive pneumonia in goats. *Veterinary J (London England)*. 1997. 2010;183(2):219–21.
- Rifatbegovic M, Maksimovic Z, Hulaj B. *Mycoplasma ovipneumoniae* associated with severe respiratory disease in goats. *Vet Rec*. 2011;168(21):565.
- Xue D, Ma Y, Li M, Li Y, Luo H, Liu X, et al. *Mycoplasma ovipneumoniae* induces inflammatory response in sheep airway epithelial cells via a MyD88-dependent TLR signaling pathway. *Vet Immunol Immunopathol*. 2015;163(1–2):57–66.
- Maksimović Z, Rifatbegović M, Loria GR, Nicholas RAJ. *Mycoplasma ovipneumoniae*: a most variable pathogen. *Pathogens (Basel, Switzerland)*. 2022;11(12).
- Bishop SC, Morris CA. Genetics of disease resistance in sheep and goats. *Small Ruminant Res*. 2007;70(1):48–59.
- Davies G, Genini S, Bishop SC, Giuffra E. An assessment of opportunities to dissect host genetic variation in resistance to infectious diseases in livestock. *Animal: Int J Anim Bioscience*. 2009;3(3):415–36.
- Snowder G. Genetics, environment and bovine respiratory disease. *Anim Health Res Reviews*. 2009;10(2):117–9.
- Goddard MJAPS. Uses of genomics in livestock agriculture. *Anim Prod Sci*. 2012;52:73–77.
- McRae KM, Rowe SJ, Baird HJ, Bixley MJ, Clarke SM. Genome-wide association study of lung lesions and pleurisy in New Zealand lambs. *J Anim Sci*. 2018;96(11):4512–20.
- Ruiz-Narváez EA. What is a functional locus? Understanding the genetic basis of complex phenotypic traits. *Med Hypotheses*. 2011;76(5):638–42.
- Zhang J, Jiang K, Lv L, Wang H, Shen Z, Gao Z, et al. Use of genome-wide association studies for cancer research and drug repositioning. *PLoS ONE*. 2015;10(3):e0116477.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from illumina amplicon data. *Nat Methods*. 2016;13(7):581–3.
- Kodama Y, Shumway M, Leinonen R. The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res*. 2012;40(Database issue):D54–56.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinf (Oxford England)*. 2009;25(14):1754–60.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. *Bioinf (Oxford England)*. 2009;25(16):2078–9.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
- Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom Proteom Bioinform*. 2021;19(4):619–28.
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with tophat and cufflinks. *Nat Protoc*. 2012;7(3):562–78.
- Zhang D, Zhang X, Li F, Li C, La Y, Mo F, et al. Transcriptome analysis identifies candidate genes and pathways associated with feed efficiency in Hu sheep. *Front Genet*. 2019;10:1183.
- Smith SM, Maughan PJ. SNP genotyping using KASPar assays. *Methods in molecular biology*. (Clifton NJ). 2015;1245:243–56.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods (San Diego Calif)*. 2001;25(4):402–8.
- Miao Y, Zhao X, Adam FEA, Xie Q, Feng H, Ding J et al. Isolation and identification of *Aeromonas veronii* in sheep with fatal infection in China: a case report. *Microorganisms* 2023;11(2).
- Zhang H, Sun L-w, Wang Z-y, Ma T-w, Deng M-t, Wang F, et al. Energy and protein requirements for maintenance of Hu sheep during pregnancy. *J Integr Agric*. 2018;17(1):173–83.
- Papatsiros VG, Stylianaki I, Tsokana CN, Papakonstantinou G, Christophroru M, Papaioannou N, et al. Histopathological lesions of *Actinobacillus pleuropneumoniae* serotype 8 in infected pigs. *Veterinary Res Forum: Int Q J*. 2023;14(7):401–4.
- Goodwin-Ray KA, Stevenson M, Heuer C. Flock-level case-control study of slaughter-lamb pneumonia in New Zealand. *Prev Vet Med*. 2008;85(1–2):136–49.
- Li M, Zi X, Lv R, Zhang L, Ou W, Chen S et al. Cassava foliage effects on antioxidant capacity, growth, immunity, and ruminal microbial metabolism in Hainan black goats. *Microorganisms* 2023;11(9).
- He J, Yuan R, Cui X, Cui Y, Han S, Wang QQ, et al. Anemoside B4 protects against *Klebsiella pneumoniae*- and influenza virus FM1-induced pneumonia via the TLR4/Myd88 signaling pathway in mice. *Chin Med*. 2020;15:68.
- Liu J, Pang Z, Wang G, Guan X, Fang K, Wang Z et al. Advanced role of neutrophils in common respiratory diseases. *Journal of immunology research*. 2017;2017:6710278.

43. Hatrongjit R, Fittipaldi N, Gottschalk M, Kersdin A. Genomic epidemiology in *Streptococcus Suis*: moving beyond traditional typing techniques. *Heliyon*. 2024;10(6):e27818.
44. Ma C, Li M, Peng H, Lan M, Tao L, Li C, et al. Mesomycoplasma ovipneumoniae from goats with respiratory infection: pathogenic characteristics, population structure, and genomic features. *BMC Microbiol*. 2023;23(1):220.
45. Sousa T, Branco R, Piedade AP, Morais PV. Hyper accumulation of arsenic in mutants of *Ochrobactrum tritici* silenced for arsenite efflux pumps. *PLoS ONE*. 2015;10(7):e0131317.
46. Wang B, Yang D, Chang Z, Zhang R, Dai J, Fang Y. Wearable bioelectronic masks for wireless detection of respiratory infectious diseases by gaseous media. *Matter*. 2022;5(12):4347–62.
47. Li J, Chen C, Gao L, Wang L, Wang W, Zhang J, et al. Analysis of histopathology and changes of major cytokines in the lesions caused by *Mycoplasma ovipneumoniae* infection. *BMC Vet Res*. 2023;19(1):273.
48. Carney SM, Clemente JC, Cox MJ, Dickson RP, Huang YJ, Kitsios GD, et al. Methods in lung Microbiome research. *Am J Respir Cell Mol Biol*. 2020;62(3):283–99.
49. Fan J, Li X, Gao Y, Zhou J, Wang S, Huang B, et al. The lung tissue microbiota features of 20 deceased patients with COVID-19. *J Infect*. 2020;81(3):e64–7.
50. Lee SH, Sung JY, Yong D, Chun J, Kim SY, Song JH, et al. Characterization of Microbiome in Bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. *Lung cancer (Amsterdam Netherlands)*. 2016;102:89–95.
51. Das S, Bernasconi E, Koutsokera A, Wurlod DA, Tripathi V, Bonilla-Rosso G, et al. A prevalent and culturable microbiota links ecological balance to clinical stability of the human lung after transplantation. *Nat Commun*. 2021;12(1):2126.
52. Baker JM, Hinkle KJ, McDonald RA, Brown CA, Falkowski NR, Huffnagle GB, et al. Whole lung tissue is the preferred sampling method for amplicon-based characterization of murine lung microbiota. *Microbiome*. 2021;9(1):99.
53. Grinev A, Chancey C, Añez G, Ball C, Winkelman V, Williamson P, et al. Genetic analysis of west nile virus isolates from an outbreak in Idaho, United States, 2006–2007. *Int J Environ Res Public Health*. 2013;10(9):4486–506.
54. Chen GQ, Zhuang QY, Wang KC, Liu S, Shao JZ, Jiang WM, et al. Identification and survey of a novel avian coronavirus in ducks. *PLoS ONE*. 2013;8(8):e72918.
55. Lindström L, Tauni FA, Vargmar K. Bronchopneumonia in Swedish lambs: a study of pathological changes and bacteriological agents. *Acta Vet Scand*. 2018;60(1):54.
56. Manlove K, Branan M, Baker K, Bradway D, Cassirer EF, Marshall KL, et al. Risk factors and productivity losses associated with *Mycoplasma ovipneumoniae* infection in United States domestic sheep operations. *Prev Vet Med*. 2019;168:30–8.
57. Jaý M, Ambroset C, Tricot A, Colin A, Tardy F. Population structure and antimicrobial susceptibility of *Mycoplasma ovipneumoniae* isolates in France. *Vet Microbiol*. 2020;248:108828.
58. Zhao JY, Du YZ, Song YP, Zhou P, Chu YF, Wu JY. Investigation of the prevalence of *Mycoplasma ovipneumoniae* in Southern Xinjiang, China. *J Veterinary Res*. 2021;65(2):155–60.
59. Amores J, Corrales JC, Martín AG, Sánchez A, Contreras A, de la Fe C. Comparison of culture and PCR to detect *Mycoplasma agalactiae* and *Mycoplasma mycoides* subsp. *Capri* in ear swabs taken from goats. *Vet Microbiol*. 2010;140(1–2):105–8.
60. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol*. 2015;33(4):408–14.
61. Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. *Nat Rev Genet*. 2019;20(3):135–56.
62. Abdoli R, Mirhoseini SZ, Ghavi Hosseini-Zadeh N, Zamani P, Moradi MH, Ferdosi MH, et al. Genome-wide association study of first lambing age and lambing interval in sheep. *Small Ruminant Res*. 2019;178:43–5.
63. Wang Z, Zhang H, Yang H, Wang S, Rong E, Pei W, et al. Genome-wide association study for wool production traits in a Chinese Merino sheep population. *PLoS ONE*. 2014;9(9):e107101.
64. García-Gómez E, Gutiérrez-Gil B, Sahana G, Sánchez JP, Bayón Y, Arranz JJ. GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the LALBA gene. *PLoS ONE*. 2012;7(10):e47782.
65. Rovadoski GA, Pertile SFN, Alvarenga AB, Cesar ASM, Pértile F, Petrini J, et al. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. *BMC Genomics*. 2018;19(1):375.
66. Bai X, Plastow GS. Breeding for disease resilience: opportunities to manage polymicrobial challenge and improve commercial performance in the pig industry. *CABI Agric Bioscience*. 2022;3(1):6.
67. Lee M, Pinto NA, Kim CY, Yang S, D'Souza R, Yong D et al. Network integrative genomic and transcriptomic analysis of Carbapenem-Resistant *Klebsiella pneumoniae* strains identifies genes for antibiotic resistance and virulence. *mSystems* 2019;4(4).
68. Brajnik Z, Ogorevc J. Candidate genes for mastitis resistance in dairy cattle: a data integration approach. *J Anim Sci Biotechnol*. 2023;14(1):10.
69. Waites KB, Talkington DF. *Mycoplasma pneumoniae* and its role as a human pathogen. *Clin Microbiol Rev*. 2004;17(4):697–728. table of contents.
70. Liu X, Luo M, Zhang L, Ding W, Yan Z, Engelhardt JF. Bioelectric properties of chloride channels in human, pig, ferret, and mouse airway epithelia. *Am J Respir Cell Mol Biol*. 2007;36(3):313–23.
71. Daxboeck F, Krause R, Wenisch C. Laboratory diagnosis of *Mycoplasma pneumoniae* infection. *Clin Microbiol Infect: Official Publication Eur Soc Clin Microbiol Infect Dis*. 2003;9(4):263–73.
72. Shimizu T, Kida Y, Kuwano K. A triacylated lipoprotein from *Mycoplasma genitalium* activates NF-kappaB through toll-like receptor 1 (TLR1) and TLR2. *Infect Immun*. 2008;76(8):3672–8.
73. Cao YH, Xu SS, Shen M, Chen ZH, Gao L, Lv FH, et al. Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. *Mol Biol Evol*. 2021;38(3):838–55.
74. Kalucka J, de Rooij L, Goveia J, Rohlenova K, Dumas SJ, Meta E, et al. Single-Cell transcriptome atlas of murine endothelial cells. *Cell*. 2020;180(4):764–e779720.
75. Paik DT, Tian L, Williams IM, Rhee S, Zhang H, Liu C, et al. Single-Cell RNA sequencing unveils unique transcriptomic signatures of organ-specific endothelial cells. *Circulation*. 2020;142(19):1848–62.
76. Acharya A, Bian F, Gomez-Arroyo J, Wagner KA, Kalinichenko VV, Kalin TV. Hypoxia represses FOXF1 in lung endothelial cells through HIF-1α. *Front Physiol*. 2023;14:1309155.
77. Cai Y, Bolte C, Le T, Goda C, Xu Y, Kalin TV, et al. FOXF1 maintains endothelial barrier function and prevents edema after lung injury. *Sci Signal*. 2016;9(424):ra40.
78. Wang G, Wen B, Deng Z, Zhang Y, Kolesnichenko OA, Ustiyan V, et al. Endothelial progenitor cells stimulate neonatal lung angiogenesis through FOXF1-mediated activation of BMP9/ACVRL1 signaling. *Nat Commun*. 2022;13(1):2080.
79. Black M, Milewski D, Le T, Ren X, Xu Y, Kalinichenko VV, et al. FOXF1 inhibits pulmonary fibrosis by preventing CDH2-CDH11 Cadherin switch in myofibroblasts. *Cell Rep*. 2018;23(2):442–58.

## Publisher's note

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