**Results**

**Differentially expressed genes (DEGs) among hcq, pcq, phpma samples**

DEGs were considered as genes which had foldchange > 2 or < 0.5 and False Discovery Rate

(FDR) q-vaule < 0.05. The distribution of all the genes in hcq vs con, pcq vs con and phpma vs con were presented in the volcano plots (Fig. 1). Pcq vs con has the most amount of 190 DEGs (143 downregulated and 47 upregulated) among the three groups (Fig. 1a). Then, the next is hcq vs con which has 11 DEGs (2 downregulated and 9 upregulated) (Fig 1b). Phpma vs con do not has any DEGs (Fig 1c). Details were in Table1.

**Common and unique of the downregulated and upregulated DEGs in three groups**

Overlaps in DEGs were assessed by using Venn diagram. There are 2 common downregulated and 6 upregulated DEGs in hcq vs con and pcq vs con. (Fig. 2a, b).

**Pathway analysis of using Gene Set Enrichment Analysis (GSEA)**

To extrapolate the regulation mechanism of genes to cellular mechanisms, R package (fgsea) was used to calculate the enrichment of the dysregulated genes in GSEA. Kyoto Encyclopedia of Genes and Genomes (KEGG) systemic lupus erythematosus pathway is enriched in hcq vs con group. This pathway is also enriched in pcq vs con group. KEGG leishmania infection pathway is another enriched pathway in pcq vs con group.

**Methods**

**RNA-seq data processing** *FastQC* was utilized to perform quality control for the raw fastq files [1]. The *STAR* software package executes mapping of large sets of high-throughput sequencing reads to a mouse reference genome (GRCm39) with high levels of accuracy and speed [2]. PCR replicates mapped in the human genome were removed with *picard MarkDuplicates* program (v2.22.7) [3]. *repair* was used for repair paired-ends reads [4]. *featureCounts* was applied to quantify the reads [4].

**Differentially Expressed Genes (DEGs)** Read counts finding out the differentially expressed genes by using DESeq2 [5]. Genes with foldchange > 2 or < 0.5 and FDR q-vaule < 0.05 were considered as statistically significantly DEGs [6]. All of the genes were shown in a volcano plot generated using *R* (4.0.4) software. Green color dots are indicated DEGs. Venn diagrams were drawn by using <http://bioinformatics.psb.ugent.be/webtools/Venn/>

**Gene Set Enrichment Analysis (GSEA)** Pre-ranked Gene Set Enrichment Analyses were performed using the *fgsea* (Version 1.16.0) package with a default setting, and no restrictions imposed on the size of gene sets that could be included. The *fgsea* takes two objects as input: a list of query gene sets and an array of the gene statistic values [7]. Kyoto Encyclopedia of Genes and Genomes (KEGG) subset of canonical pathways (c2.cp.kegg.v7.2.symbols.gmt) were used as database [8]. Function *fgsea* was utilited to generate the pathway, and the *plotEnrichment* was used to plot the enrichment pathway.

**Availability of data and materials**

All analyses were performed in R version 4.0.4 and Bioconductor version 3.12 (R Core Team, 2021; [http://www.R-proj](http://www.R-project.org/)ect.org/). All codes are available at https://github.com/yueli8/unmc.

**Figures**

**Fig. 1** The variation of gene expressions could be visualized in the volcano plot. **a,** hcq vs con. **b,** pcq vs con. **c,** phpma vs con. Volcano plots are constructed by using fold-change values and False Discovery Rate (FDR) q-vaule < 0.05, so they could be visualized by the relationship between fold change (magnitude of change) and statistical significance (which considers both magnitudes of change and variability into consideration). Based on those values, subsets of genes were also allowed to be isolated. 2.0-fold chang of up and down were corresponding to the vertical lines, and a FDR q-vaule of 0.05 was represented to the horizontal line. Thus, the differentially expressed genes with statistical significance were represented in green points.

**Fig. 2** The Venn diagrams illustrating the unique and the common of the paired-wise overlap between the differentially expressed genes among hcq vs con and pcq vs con **a,** downregulated. **b,** upregulated. The colors of circles in the Venn diagrams are as follows: red, hcq vs con; green, pcq vs con.

**Fig. 3** Gene set enrichment analysis (GSEA) of signaling pathways activated in comparison of data sets of **a,** hcq vs con of Kyoto Encyclopedia of Genes and Genomes (KEGG) systemic lupus erythematosus signaling pathway. **b,** pcq vs con of KEGG leishmania infection signaling pathway. **c,** pcq vs con of KEGG systemic lupus erythematosus signaling pathway. (Genes sets with *P-value* < 0.05, FDR q-value < 0.25 are considered significance) NES, normalized enrichment score; FDR, false discovery rate.

**Reference**

1. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17:1–19.

2. Dobin A, Gingeras TR, Spring C, Flores R, Sampson J, Knight R, et al. Mapping RNA-seq with STAR. Curr Protoc Bioinforma. 2016;51:586–97. doi:10.1002/0471250953.bi1114s51.Mapping.

3. Ebbert MTW, Wadsworth ME, Staley LA, Hoyt KL, Pickett B, Miller J, et al. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. BMC Bioinformatics. 2016;17 Suppl 7. doi:10.1186/s12859-016-1097-3.

4. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019;47.

5. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

6. Heng S, Wang L, Yang X, Huang H, Chen G, Cui M, et al. Genetic and Comparative Transcriptome Analysis Revealed DEGs Involved in the Purple Leaf Formation in Brassica juncea. Front Genet. 2020;11 April:1–10.

7. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov M, Sergushichev A. Fast gene set enrichment analysis. 2016.

8. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008;36 SUPPL. 1:480–4.