RNA-Seq analysis in giant pandas reveals the differential expression of multiple genes involved in cataract formation

Yuyan You1†‡, Chao Bai1†, Xuefeng Liu2, Yan Lu1, Ting Jia2, Maohua Xia2, Yanqiang Yin3, Wei Wang1, Yucun Chen4, Chenglin Zhang1, Yan Liu1, Liqin Wang5, Tianchun Pu2, Tao Ma2, Yanhui Liu2, Jun Zhou3, Lili Niu5, Suhui Xu4, Yanxia Ni2, Xin Hu1 and Zengshuai Zhang2

Abstract

Background: The giant panda (Ailuropoda melanoleuca) is an endangered mammalian species native to China. Fewer than 2500 giant pandas are known to exist, many of which are bred in captivity as a means to preserve and repopulate the species. Like other captive mammals, giant pandas acquire age-related cataracts, reducing their quality of life. Recent comparative genome-wide methylation analysis revealed 110 differentially methylated genes associated with cataract formation including six also associated with the formation of age-related cataracts in humans.

Results: To investigate the pathological pathway in greater detail, here we used RNA-Seq analysis to investigate the differential expression profiles of genes in three giant pandas with cataracts and three healthy controls. We identified more than 700 differentially expressed genes, 29 of which were selected for further analysis based on their low q-value. We found that many of the genes encoded regulatory and signaling proteins associated with the control of cell growth, migration, differentiation and apoptosis, supporting previous research indicating a key role for apoptosis in cataract formation.

Conclusion: The identification of genes involved in the formation of age-related cataracts could facilitate the development of predictive markers, preventative measures and even new therapies to improve the life of captive animals.

Keywords: Giant panda, endangered mammals, Cataracts, RNA-Seq

Background

The housing of mammals in zoos and reservations is an efficient strategy to protect endangered species and encourage repopulation, but captive mammals tend to live longer than their wild counterparts and thus experience diseases of ageing, which are uncommon in the wild. One example is the development of age-related cataracts, which are associated with ageing mammals due to the accumulation of oxidative damage in the lens [1]. Cataracts are the main cause of blindness in ageing humans and other primates, as well as companion animals such as dogs and cats [2–4]. They are also prevalent in captive giant pandas (Ailuropoda melanoleuca), which live to 15–20 years in the wild but 25–30 years in captivity [5, 6]. Giant pandas 18 or more years old are described as aged because they have reached an equivalent human age of ~ 75. The prevalence of cataracts in the current population of aged giant pandas is ~ 20%. Age-related cataracts are heritable with significant environmental triggers, including oxidative stress and the resulting accumulation of DNA damage [7–10].
genes most strongly associated with cataracts therefore include those related to oxidative stress responses, the production of antioxidant enzymes and metabolites, and various DNA repair pathways [11–13]. Previous studies identified changes in DNA methylation associated with cataract formation in several mammals [14–16] and we recently reported that 110 genes with functions relevant to cataract pathogenesis are differentially methylated in giant pandas, including six genes known to be associated with age-related cataracts in humans [17].

Epigenetic modifications such as DNA methylation affect gene expression and therefore control the availability of the corresponding gene products. To gain more insight into the role of differential gene expression in the pathogenesis of age-related cataracts in giant pandas, we took blood samples from three aged giant pandas with cataracts and three healthy controls for RNA-Seq analysis. Following the alignment of reads with the reference genome, we identified expression profiles representing genes expressed exclusively or preferentially in the cataractogenic or healthy samples, and determined the corresponding functional annotations. The identification of genes that are overexpressed or suppressed during the formation of cataracts could lead to the development of new diagnostics, preventative treatments and therapeutic approaches to improve the quality of life for captive giant pandas and other mammals.

Results

Samples
Peripheral blood samples were collected from six giant pandas (five females and one male) ranging in age from 19 to 37, with three of the females affected by cataracts and the other specimens defined as healthy based on regular physical examinations (Table 1). The blood samples were assigned to three sample bands: A (af- fected females), B (unaffected male) and C (unaffected females). These sample bands were subsequently used for comparative transcriptomic analysis.

RNA-Seq data processing and quality analysis
Blood samples from all six giant pandas were used to prepare a de novo RNA-Seq dataset with an average of 52.05 ± 5.63 million reads per specimen and an average length of 144.35 ± 1.86 bp after quality control and trimming. For each specimen, ~ 87% of the reads mapped to unique sequences in the reference genome (Supplementary Table S1). Homogeneity distribution analysis and the gene coverage ratio indicated the anticipated distribution of reads within each gene and across different genes (Supplementary Fig. S1a,b) and also confirmed that ~ 44.72% of the reads mapped to exonic regions of the reference sequence (Supplementary Fig. S1c).

Analysis of gene structure characteristics
Analysis of the RNA-Seq dataset for gene coverage and chromosome distribution revealed differences in mapping density across different chromosomes, with chromosomes GL192341.1 and GL192355.1 showing a particularly low read density and chromosomes GL192348.1 showing the opposite phenomenon (Fig. 1a). The mapped reads were screened for single-nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms (indels) revealing 5.9–6.7 × 10^4 SNPs but < 1 × 10^4 indels in each sample (Fig. 1b). Among the SNPs, the number of transitions vastly exceeded the number of transversions, and the transitions A > G, C > T, G > A and T > C were particularly abundant (Supplementary Fig. S2). We detected more than 2.5 × 10^4 alternative mRNA processing events, the most abundant of which were alternative transcriptional start sites (8 × 10^3) followed by alternative transcriptional stop sites (7 × 10^3) (Fig. 2). We also identified at least eight new transcript regions representing previously unknown isoforms of known genes, with seven of these new regions mapping to four genes: GNAO1, OGFOD1, HERPUD1 and NLRC5 (Supplementary Table S2).

Analysis of gene expression
The number of reads mapping to each gene in the reference genome represents the abundance of the corresponding mRNA when corrected for factors such as gene length, sequencing depth and saturation. We also ensured the accuracy of our results by testing for correlation between samples (Supplementary Table S3). We then determined the number of genes that were co-expressed in our samples, revealing a core of 2711

Table 1 Description of giant panda sample donors. Sample band A features the affected females, whereas bands B and C feature the healthy males and females, respectively

| Name | Sample band | Spectrum number | Birth year (age) | Status | Sex | Remarks |
|------|-------------|-----------------|------------------|--------|-----|---------|
| BD   | C3          | 520             | 2000 (21)        | Healthy| Male|         |
| YY   | C1          | 362             | 1990 (31)        | Healthy| Female|         |
| YE   | C2          | 493             | 1999 (22)        | Healthy| Female|         |
| JN   | A3          | 403             | 1993 (28)        | Age-related cataracts | Female| Died in 2021 |
| LL   | A2          | 320             | 1986 (32)        | Age-related cataracts | Female| Died in 2018 |
| XX   | A1          | 253             | 1982 (39)        | Age-related cataracts | Female|         |
common genes that were universally expressed and a minimum of 172 and a maximum of 263 genes in each animal that were not part of this common set (Fig. 3). PCA indicated that 36.68% of the variation could be explained by PC1, a further 33.06% by PC2, and 13.65% by PC3, primarily separating the affected and unaffected animals but to a lesser extent separating the unaffected male from the unaffected females. Overall, this resulted in the clustering of the affected females together and the dispersion of the other three samples, suggesting the affected females had more in common than any of the unaffected individuals had in common with each other (Fig. 4a). This
outcome was broadly supported by PoCA, which revealed that 50.02% of the variation could be explained by PoC1, a further 13.65% by PoC2 and 5.62% by PoC3, again leading to the formation of an affected cluster with the other samples dispersed (Fig. 4b). The individual two-dimensional plots showing the PCA and PoCA data in more detail are provided in Fig. 5.

Identification and analysis of differentially expressed genes

We observed a much greater number of differentially expressed genes when comparing female pandas with cataracts to the healthy male (A vs B, 705 genes) than when comparing female pandas with cataracts to healthy females (A vs C, 33 genes). However, most of these genes were not differentially expressed when comparing healthy male and female pandas (B vs C, 116 genes), suggesting that the A vs B profile cannot be wholly explained by sex-specific differences in gene expression (Fig. 6). Given that very few genes were differentially expressed when comparing affected and healthy females (A vs C), cataract formation appears to influence a larger number of genes in male than female pandas. Among the 705 differentially expressed genes in the A vs B comparison, 533 were upregulated and 172 downregulated. When visualized as a scatter plot (Fig. 7), it is clear that more genes are upregulated in the healthy male than the affected females and are more likely to show a statistically significant change in expression (Fig. 7b) whereas the 23 upregulated and 10 downregulated genes in the comparison A vs C show limited statistical significance (Fig. 7a). The results are emphasized in the corresponding heat maps (Fig. 8). The comparison of affected females (A) with unaffected pandas of either sex (B + C) revealed 29 genes satisfying q < 0.05 and |log2 fold change (FC)| > 1 (Table 2).

Functional annotation of differentially expressed genes

The biological functions and related pathways of the differentially expressed genes were investigated by screening the sequences against the GO and KEGG databases, as well as the Clusters of Orthologous Groups of proteins (COG) and euKaryotic Ortholog Groups (KOG) maintained by the NCBI. GO annotations revealed that the differentially expressed genes represented a wide range of biological processes (111 genes), cellular components
**Fig. 4** (a) Principal component analysis. The three-dimensional plot separates the individual pandas along three principal components. The affected females (red, band A) form one group whereas the unaffected male (blue, band B) and females (green, band C) are scattered. (b) Principal coordinates analysis. The three-dimensional plot separates the individual pandas along three principal coordinates. The affected females (red) form one group whereas the unaffected male (blue) and females (green) are scattered. The corresponding two-dimensional plots are shown in Fig. 5.

**Fig. 5** Principal component analysis (a) and principal coordinates analysis (b). The two-dimensional plots provide more detailed context for the three-dimensional plots shown in Fig. 4a (PCA) and Fig. 4b (PoCA).
(73 genes) and molecular functions (5 genes). Strongly represented biological process categories (accounting for > 10% of the genes) included cellular process, metabolic process, biological regulation/regulation of a biological process (particularly positive regulation), and response to stimulus. With few exceptions, the functional profile of the differentially expressed genes mirrored that of the total gene catalog, although the differentially expressed genes were overrepresented in the categories immune system process and multi-organism process but underrepresented in the categories biological phase and cell aggregation (Fig. 9). There was little difference in terms of cellular

![Venn diagram](image)

**Fig. 6** Venn diagram showing the differentially expressed genes when comparing the three bands in three pairwise comparisons. **Fig. 7** Scatter plots showing the differentially expressed genes when comparing the three sample bands in three pairwise comparisons. The horizontal and vertical axes represent the log2(TPM) value of two group samples. Each point represents a gene, and the closer each point is to the origin, the lower the expression level. Red represents upregulated genes, green represents downregulated genes, and black represents genes with no difference in expression.

![Scatter plots](image)

**Fig. 7** Scatter plots showing the differentially expressed genes when comparing the three sample bands in three pairwise comparisons. The horizontal and vertical axes represent the log2(TPM) value of two group samples. Each point represents a gene, and the closer each point is to the origin, the lower the expression level. Red represents upregulated genes, green represents downregulated genes, and black represents genes with no difference in expression.
component categories, with the exception of the differentially expressed genes being unrepresented in the category *nucleoid*. However, one of the most abundant categories among the differentially expressed and complete gene catalog was *protein-containing complex*, and preliminary protein network analysis revealed that the genes most strongly upregulated in the A vs B comparison also tended to be more likely to interact with other proteins and also tended to have more connections (data not shown). In terms of molecular functions, there was again little difference between the differentially expressed genes and complete gene catalog, with the exception of the differentially expressed genes being unrepresented in the categories involving *morphogen*, *metallochaperone* and *chemoattractant/chemorepellant* activity (Fig. 9). Interestingly, several genes encoding enzymes involved in histidine metabolism were included among the differentially expressed pathways revealed when screening KEGG (Fig. 10). The analysis of GO categories that were enriched among the differentially expressed genes revealed strong hits for immunity and defense-related functions, which showed the highest Rich factors (Table 3, Fig. 11). Visualization of the biological, cellular and molecular functions by means of directed acyclic graphs revealed that 45
| No. | GeneName | result | Transcript id | GeneID | MeanTPM (A) | MeanTPM (B) | log2FoldChange |
|-----|----------|--------|---------------|--------|-------------|-------------|----------------|
| 1   | GALNS    | up     | ENSAMET0000000379 | ENSAMEG00000000312 | 41.27       | 18.8        | 1.134360776    |
| 2   | NCF1     | down   | ENSAMET00000016724 | ENSAMEG00000015198 | 1.534666667 | 268.47      | -7.448182824   |
| 3   | GLG1     | down   | ENSAMET0000001501 | ENSAMEG0000001334  | 0.283333333 | 14.76       | -5.703048571   |
| 4   | C12orf56 | up      | ENSAMET00000019892 | ENSAMEG00000018106 | 14.698666667 | 2.5         | 2.554069994    |
| 5   | AZN1     | down   | ENSAMET00000016178 | ENSAMEG00000014703 | 25.47       | 61.03       | -1.260719631   |
| 6   | RCAN1    | down   | ENSAMET00000005825 | ENSAMEG00000005295 | 0.0001      | 82.1        | -16.3250946    |
| 7   | RASD1    | up      | ENSAMET00000002517 | ENSAMEG00000002311 | 15.45533333 | 0.76        | 4.345774837    |
| 8   | CYSLTR1  | up      | ENSAMET00000002176 | ENSAMEG000000019893 | 19.67666667 | 80.1        | 1.296611993    |
| 9   | EGR1     | up      | ENSAMET000000018184 | ENSAMEG000000016544 | 63.94333333 | 48.1        | 3.7326585154   |
| 10  | RANBP9   | up      | ENSAMET00000000514 | ENSAMEG00000000415 | 8.37        | 1.31        | 2.675660811    |
| 11  | UCHL1    | down   | ENSAMET00000018946 | ENSAMEG00000017217 | 2.423333333 | 84.8        | -1.807071401   |
| 12  | TNF SF12 | up      | ENSAMET00000001886 | ENSAMEG000000016729 | 8.51        | 0.0001      | 16.37687151    |
| 13  | SERPINB10| up      | ENSAMET000000016540 | ENSAMEG000000015156 | 58.52       | 13.83       | 2.081128613    |
| 14  | SPP1L2B  | up      | ENSAMET000000006059 | ENSAMEG000000005486 | 65.44666667 | 26.61       | 1.294351204    |
| 15  | PTP4A3   | up      | ENSAMET00000001697 | ENSAMEG000000015150 | 38.59866667 | 5.66        | 2.76928463     |
| 16  | FOS      | up      | ENSAMET000000020143 | ENSAMEG000000018335 | 16.93143333 | 71.19       | 4.52512604     |
| 17  | ARHGAP21 | up      | ENSAMET000000007161 | ENSAMEG000000006453 | 5.335333333 | 2.03        | 1.446669108    |
| 18  | FOSB     | up      | ENSAMET000000013608 | ENSAMEG000000012405 | 204.76      | 68.3        | 4.905904523    |
| 19  | IFI27L2  | up      | ENSAMET000000006889 | ENSAMEG000000006288 | 79.83       | 337.07      | 1.24804284     |
| 20  | ANXA3    | down   | ENSAMET000000005999 | ENSAMEG000000005375 | 13.09966667 | 44.12       | -1.75233124    |
| 21  | CLIC4    | up      | ENSAMET000000011061 | ENSAMEG000000010080 | 14.94833333 | 6.36        | 2.132403328    |
| 22  | COL2A1   | up      | ENSAMET000000014144 | ENSAMEG000000012885 | 15.23       | 6.57        | 2.121950666    |
| 23  | DUSP22   | up      | ENSAMET000000006552 | ENSAMEG000000005967 | 41.65333333 | 15.57       | 1.41966308     |
| 24  | EPCAM    | down   | ENSAMET000000006851 | ENSAMEG000000006223 | 2.763333333 | 59.1        | -1.106748529   |
| 25  | PDE4A    | up      | ENSAMET000000012636 | ENSAMEG000000011455 | 6.78        | 2.24        | 1.597786541    |
| 26  | MYBL1    | down   | ENSAMET000000013645 | ENSAMEG000000012401 | 4.073333333 | 15.41       | -1.391958077   |
| 27  | STEAP3   | up      | ENSAMET000000009325 | ENSAMEG000000008497 | 18.19       | 69.5        | 1.388006056    |
| 28  | UNC119   | up      | ENSAMET000000005520 | ENSAMEG0000000049930 | 111.16      | 50.33       | 1.143147237    |
| 29  | ENG      | up      | ENSAMET000000016023 | ENSAMEG000000015119 | 30.99666667 | 8.75        | 2.08004037     |
| No. | pValue         | qValue          | GeneDescription | GO                                      | KEGG                                      | KOG                                    |
|-----|----------------|-----------------|-----------------|-----------------------------------------|-------------------------------------------|----------------------------------------|
| 1   | 0.0000000000000000000000 | 0.000000000000002714 | galactosamine (N-acetyl-6-sulfatase [Source:HGNC Symbol:AcchGNC:4122]) | GO:0007066 GO:0008484 GO:0008152; GO:0003824 GO:0003890 | –                          | KOG3867-Sulfatase [R]                   |
| 2   | 0.0000000000000000000000 | 0.0000000000000000000000 | neutrophil cytosolic factor 1 [Source:HGNC Symbol:Acc:HGNC:7660] | GO:0016175 GO:0035091 GO:0055114; GO:0005735 GO:0004332; GO:0006611 GO:0042754 GO:0005829; GO:0019816 GO:0003902 GO:0004168; GO:0004631 GO:0043574 GO:0004589; GO:0004635 GO:0071276 GO:1900745 GO:0005886 | –                          | –                                      |
| 3   | 0.0000000000000000000000 | 0.0000000000000000000000 | golgi glycoprotein 1 [Source:HGNC Symbol:Acc:HGNC:4316] | GO:0007066 GO:0016020 GO:0005794; GO:0000139 GO:0016021 | –                          | KOG3648-Golgi apparatus protein (cysteine-rich fibroblast growth factor receptor)[U] |
| 4   | 0.0000000000000000000000 | 0.0000000000000000000000 | chromosome 12 open reading frame 56 [Source:HGNC Symbol:Acc:HGNC:26907] | –                          | –                          | –                                      |
| 5   | 0.0000000000000000000000 | 0.000000000000005027 | antizyme inhibitor 1 [Source: RefSeq peptide:Acc:NP_001278387] | GO:0006596 GO:0042978 GO:0043085; GO:0003824 GO:0042177 GO:1902269 GO:0005634 | –                          | KOG0622-Ornithine decarboxylase [E]     |
| 6   | 0.0000000000000000000000 | 0.000000000000003632 | regulator of calcineurin 1 [Source:HGNC Symbol:Acc:HGNC:3040] | GO:0033173 GO:0019722 GO:0003676; GO:005622 | –                          | KOG4019-Calcineurin-mediated signaling pathway inhibitor DSCR1[TR] |
| 7   | 0.0000000000000000000000 | 0.000000000000003903 | ras related dexamethasone induced 1 [Source:HGNC Symbol:Acc:HGNC:15882] | GO:0016529 GO:0003924 GO:0005525; GO:0001016 GO:007165 GO:0016020 | –                          | KOG0395-Ras-related GTPase [R]         |
| 8   | 0.0000000000000000000000 | 0.000000000000001126 | cytoxyl leukotriene receptor 1 [Source:HGNC Symbol:Acc:HGNC:17451] | GO:0004974 GO:0061737 GO:0007186; GO:0004930 GO:0016021 GO:0016020; GO:000243 lg:0006816 GO:0006935; GO:0007166 GO:0003887 | –                          | KOG0322-Leukotriene receptor[C]         |
| 9   | 0.0000000000000000000000 | 0.000000000000001781 | early growth response 1 [Source:HGNC Symbol:Acc:HGNC:3238] | GO:0005735 GO:0005654 GO:0005634; GO:2000182 GO:0098759 GO:0090909; GO:0071480 GO:0071310 GO:0071498; GO:0061418 GO:0030725 GO:0004886; GO:0045946 GO:0045893 GO:0045080; GO:0044849 GO:004281 GO:0035914; GO:003323 G0:003268 GO:0035050; GO:0030217 GO:0039749 GO:0006836; GO:0006535 GO:0002931 GO:0001666; GO:0000122 GO:1900841 GO:0044212; GO:0044729 GO:0043565 GO:0035035; GO:0010386 GO:0010270 GO:0003570; GO:0003677 GO:0001077 GO:0000982; GO:000977 GO:000976 GO:0006351; GO:0003676 GO:0046872 | –                          | KOG1721-FOG: Zn-finger [R]              |
|   | q-value | log2 fold change | Gene Name | Source | GO Terms | KOG Annotation |
|---|--------|-----------------|-----------|--------|----------|----------------|
| 10 | 0.0000000000223677291954 | 0.0000000029087367841100 | RAN binding protein 9 | HGNC Symbol: HGNC:13727 | GO:0005737;GO:0005634;GO:0070373; GO:0019899 | – |
| 11 | 0.00000000003728989939510 | 0.0000000029087367841100 | ubiquitin C-terminal hydrolase L1 | HGNC Symbol: HGNC:12513 | GO:0006511;GO:1904115;GO:0016874; GO:0006508; GO:0006824;GO:0008234;GO:0008235; GO:0043417;GO:0019899 | – |
| 12 | 0.000000000046303228987080 | 0.0000000029087367841100 | TNF superfamily member 12 | HGNC Symbol: HGNC:11927 | GO:0005164;GO:0006955;GO:0016020; GO:0045732;GO:2001238;GO:0005576;GO:0048471 | – |
| 13 | 0.000000000059926447639400 | 0.0000000029087367841100 | serpin family B member 10 | HGNC Symbol: HGNC:8942 | GO:0005615;GO:0005634;GO:0005829 | – |
| 14 | 0.000000000042149703973800 | 0.0000000029087367841100 | signal peptide peptidase like 2B | HGNC Symbol: HGNC:30627 | GO:1900746;GO:0043542;GO:0043117; GO:0007219;GO:0006470;GO:0016311;GO:0016791;GO:0008138;GO:0006355;GO:1901224;GO:1904951;GO:0005634;GO:0005737;GO:0031668 | – |
| 15 | 0.000000000059926447639400 | 0.0000000029087367841100 | protein tyrosine phosphatase type IVA, member 3 | HGNC Symbol: HGNC:9636 | GO:1900746;GO:0043542;GO:0043117; GO:0007219;GO:0006470;GO:0016311;GO:0016791;GO:0008138;GO:0006355;GO:1901224;GO:1904951;GO:0005634;GO:0005737;GO:0190830 | – |
| 16 | 0.000000000010152143300000 | 0.0000000029087367841100 | Fox proto-oncogene, AP-1 transcription factor subunit | HGNC Symbol: HGNC:5794 | GO:0005654;GO:0005667;GO:0005634; GO:0043542;GO:0043117;GO:0007219;GO:0006470;GO:0016311;GO:0016791;GO:0008138;GO:0006355;GO:1901224;GO:1904951;GO:0005634 | K04379-k04010 MAPK signaling pathway; k04668 TNF signaling pathway; k04024 MAPK signaling pathway; k04210 Apoptosis; k04620 Toll-like receptor signaling pathway; k04600 T cell receptor signaling pathway; k04602 B cell receptor signaling pathway; k04915 Estrogen signaling pathway; k04917 ProtecIn signaling pathway; k04923 Ocytocin signaling pathway; k04725 Cholinergic synapse; k04728 Dopaminergic synapse; k04800 Osteoclast differentiation; k04713 Caidian entertain; k05020 Pathways in cancer; k05233 Choline metabolism in cancer; k05210 Colorectal cancer; k05224 Breast cancer; k05323 Rheumatoid arthritis; k05031 | – |
| Gene ID   | FDR    | Log2 FC | Function/Condition                                                                 |
|----------|--------|---------|-----------------------------------------------------------------------------------|
| 17       | 0.000197666 | 0.00050966G | GO:0005096;GO:0043547;GO:0045177;GO:0005829;GO:0005886;GO:0005975;GO:0005902;GO:0005902;GO:0005902 | Amphetamine addiction; HTLV-I infection; Hepatitis B; Herpes simplex infection; Leprosy; Amphetamine addiction; Chagas disease (American trypanosomiasis); Endocrine resistance |
| 18       | 0.000266574 | 0.00056544G | GO:0005654;GO:0005654;GO:0005654;GO:0005654;GO:0005654 | Oxygen metabolism; FosB proto-oncogene, AP-1 transcription factor subunit |
| 19       | 0.000292568 | 0.00160212 | GO:0016021;GO:0016021;GO:0016021;GO:0016021;GO:0016021 | Interferon-alpha inducible protein 27 like 2; KOG0819-Annexin |
| 20       | 0.001808525 | 0.00395670 | GO:0039567;GO:0039567;GO:0039567;GO:0039567;GO:0039567 | Annexin A3; KOG0819-Annexin |
| 21       | 0.002467927 | 0.00494711 | GO:0049471;GO:0049471;GO:0049471;GO:0049471;GO:0049471 | Chloride intracellular channel 4; Collagens (type IV and type XIII), and related proteins |
| 22       | 0.002568066 | 0.00469824 | GO:0046982;GO:0046982;GO:0046982;GO:0046982;GO:0046982 | Collagen type II alpha 1 chain; KOG3544-Collagens (type IV and type XIII), and related proteins |
| No. | FDR | P-value | Symbol/GO Terms | Description |
|-----|-----|---------|----------------|-------------|
| 23  | 0.00020624 | 0.00368372 | dual specificity phosphatase 22 [Source: HGNC Symbol; Acc: HGNC:16077] | dual specificity phosphatase [V] |
| 24  | 0.000258111 | 0.00446881 | epithelial cell adhesion molecule [Source: HGNC Symbol; Acc: HGNC:11528] | epithelial cell adhesion molecule [T] |
| 25  | 0.000451773 | 0.006951735 | phosphodiesterase 4A [Source: HGNC Symbol; Acc: HGNC:8781] | phosphodiesterase 4A [T] |
| 26  | 0.000777204 | 0.010437403 | MYB proto-oncogene like 1 [Source: HGNC Symbol; Acc: HGNC:7547] | MYB proto-oncogene like 1 [K] |
| 27  | 0.000805245 | 0.01071711 | STEAP metalloenductase [Source: HGNC Symbol; Acc: HGNC:26592] | STEAP metalloenductase [K] |
| 28  | 0.002244887 | 0.022926346 | unc-119 lipid binding chaperone [Source: HGNC Symbol; Acc: HGNC:12565] | unc-119 lipid binding chaperone [K] |
| 29  | 0.003054818 | 0.028698842 | endoglin [Source: HGNC Symbol; Acc: HGNC:3349] | endoglin [K] |
genes differed between the healthy and unhealthy pandas (Supplementary Fig. S3).

Discussion
Ageing mammals in captivity often develop cataracts due to the accumulation of oxidative damage [1]. This has been observed in captive giant pandas, which typically live up to 10 years longer than their wild counterparts [5, 6]. The prevalence of cataracts in the current population of aged giant pandas is ~ 20% and this is associated with a declining quality of life, as the animals find it difficult to feed and negotiate their surroundings. Although genetic factors have been identified that promote age-related cataracts, the most important triggers are environmental, particularly oxidative stress and DNA damage [7–10]. We previously compared the DNA methylation status of giant pandas with and without cataracts in an attempt to identify epigenetic effects that might influence the expression of genes associated with cataract formation [17]. We identified multiple differentially methylated genes with potential roles in cataract-related pathways, including base excision repair, apoptosis and p53 signaling. Certain genes also showed abnormal methylation profiles specifically in the pandas with cataracts, including the cysteine-aspartate protease gene CASP3, a pro-apoptotic mediator already linked to cataracts in rats [18, 19], and the glutathione S-transferase gene GSTM3, which is expressed in the lens tissues of human patients with age-related cataracts [20].

DNA methylation is an epigenetic mechanism that regulates gene expression by modifying the structure of chromatin, usually leading to the suppression of gene expression. As a logical extension of our previous study, we were therefore interested in the analysis of differential gene expression between giant pandas with cataracts and controls with healthy eyes. Using the same animals as in our previous study, we carried out RNA-Seq analysis to identify panels of genes either upregulated or downregulated in pandas with cataracts. Following the alignment of reads with the reference genome, we identified expression profiles representing genes expressed exclusively or preferentially in the cataractogenic or healthy samples, and identified the corresponding functional annotations by screening the KEGG database and GO categories. Our RNA-Seq data also revealed abundant alternative splicing and novel transcripts within the dataset, as previously reported during murine lens development [21]. Two genes that were shown in our DNA methylation study [17] to be methylated specifically in affected pandas were also shown to be downregulated in the affected pandas by RNA-Seq analysis. CASP3, encoding the apoptotic cysteine-aspartate protease was downregulated with a log2FC of −1.03, whereas GSTM3, encoding the glutathione S-transferase, was downregulated with a log2FC of −1.28.

A link between the glutathione (GSH) pathway and cataract formation was identified in a transcriptomic study in mice [22]. The authors considered the
consequences of GSH deficiency in naïve and buthionine sulfoximine-treated C57Bl/6 LEGSKO (lens GSH-synthesis knockout) mice compared to wild-type controls, thus providing a more relevant model of oxidative damage. Among 24,415 mapped reads, 441 genes showed significantly modulated expression, including genes involved in epithelial-mesenchymal transition (EMT) signaling, the visual cycle, and lipid metabolism. Several detoxification genes were upregulated, including the aldehyde dehydrogenases Aldh1a1 and Aldh3a1, the metallothioneins Mt1 and Mt2, the carboxylesterase Ces1g, and the urea transporter Slc14a1, whereas genes encoding lens crystallins and other vision-related genes were downregulated [22]. The authors concluded that GSH deficiency in the lens leads to the expression of detoxifying genes and the activation of EMT signaling (providing evidence of adaption to the loss of antioxidant capacity) and also revealing a pathogenetic mechanism of cataract formation.

Similarly, our RNA-Seq analysis revealed the differential expression of several genes related to oxidative damage, the visual cycle, developmental functions, and lipid metabolism, suggesting that ageing pandas develop cataracts as the natural capacity for oxidative stress responses begins to diminish. For example, we observed a 7.45-fold down-regulation of NCF1, encoding a membrane-bound subunit of NADPH oxidase which is involved in superoxide production and the induction of apoptosis [23, 24]. Examples of differentially regulated genes involved in visual perception included COL2A1, encoding a collagen component of the extracellular matrix (1.2-fold induction) and UNC119, encoding a G-protein-binding factor that is required for protein trafficking in photoreceptor cells (1.1-fold induction) [25]. The differentially expressed genes related to developmental functions included EPCAM, encoding the epithelial cell adhesion molecule involved in cell–cell adhesion, cell signaling, migration, proliferation, and differentiation (downregulated 1.1-fold) [26, 27] as well as a chloride channel that interacts with the cytoskeleton and is known to regulate vascular morphogenesis in the eye (CLIC4, 1.2-fold induction) [28, 29] and

![Fig. 10 Differential representation of metabolic pathways. The rectangular nodes represent gene products (enzymes or regulatory factors) and the circular nodes represent metabolites. The rounded white boxes indicate linked pathways. Red indicates upregulated genes and green indicates downregulated genes, with deeper coloring indicating a greater fold change in expression.](image-url)
| GO ID       | Term Description          | GO ID | Term | Ontology | Significant | Annotated | P-value | Q-value | Signi_id | Signi_symbol |
|------------|---------------------------|-------|------|----------|-------------|-----------|---------|---------|---------|-------------|
| GO:0031982 | vesicle cellular component |       |      |          | 84/287      | 2493/14310 | 0.00000041 | 0.0004  |         | PDZK1P1, NUDT14, GALNS, LPL, MINDY1, SLC1A4, ZNF61, ECM1, KCNN1, ANPEP, HNMT, SCPEP1, PLOD3, CRYZ, EPS8, CMBl, KCTD12, SLPI, GAS6, NAN5, SPP1L28, Rab32, RIHO, ARAS, TMEBRA, TRIP10, SMPDL3A, SLC7A8, OAF, SLC1A5, STEAP3, F5, FBP1, SEPT5, COMT, SERPINE2, HAX1, LY2, APOH, ANGPTL6, FM1688, BVR, PLD3, CTSZ, SOG1, GIP, LAMB2, TUBB6, C3orf58, TIMP1, FES, A2M, RRA5, IGF8, NAA1, ITSN1, GAA, SCSH, MICL2, PRAD1, HFE, ACTG1, PEDP1, DPG4, EFHD1, TNFSF13, TMEI75, ARRB1, APP, ABCA3, PTGR1, CDC42BPB, ENDO1, KNE3, ATP6, |

You et al. BMC Genomic Data (2021) 22:44
Page 15 of 56
Table 3: Functional enrichment analysis based on the representation of GO terms (Continued)

| GO.ID   | Term                      | Ontology              | Significant | Annotated | Pvalue   | Qvalue | Signi_id | Signi_symbol |
|---------|---------------------------|-----------------------|-------------|-----------|----------|--------|----------|--------------|
| GO:0005576  | extracellular region    | cellular component    | 91/287      | 2809/14310 | 0.00000064 | 0.0004 | ENSAMEG00000000015, ENSAMEG00000000093, ENSAMEG000000000312, ENSAMEG000000000640, ENSAMEG000000000539, ENSAMEG000000000609, ENSAMEG0000000005291, ENSAMEG0000000002718, ENSAMEG0000000002732, ENSAMEG0000000003068, ENSAMEG0000000003442, ENSAMEG0000000003627, ENSAMEG0000000003820, ENSAMEG0000000004013, ENSAMEG0000000004196, ENSAMEG0000000004399, ENSAMEG0000000004441, ENSAMEG0000000004680, ENSAMEG0000000004836, ENSAMEG0000000006082, ENSAMEG0000000006368, ENSAMEG0000000007046, ENSAMEG0000000007327, ENSAMEG0000000007340, ENSAMEG0000000007342, ENSAMEG0000000007468, ENSAMEG0000000007631, ENSAMEG0000000008008, ENSAMEG0000000008907, ENSAMEG0000000008620, ENSAMEG0000000008780, ENSAMEG0000000009105, ENSAMEG0000000009193, ENSAMEG0000000009998, ENSAMEG0000000010219, ENSAMEG0000000010536, ENSAMEG0000000010602, ENSAMEG0000000010841, ENSAMEG0000000011403, ENSAMEG0000000011494, ENSAMEG0000000011617, ENSAMEG0000000011687, ENSAMEG0000000011820, ENSAMEG0000000011842, ENSAMEG0000000012016, ENSAMEG0000000012043, ENSAMEG0000000012457, ENSAMEG0000000012482, ENSAMEG0000000012483, ENSAMEG0000000012511, ENSAMEG0000000012590, ENSAMEG0000000012708, ENSAMEG0000000012743, ENSAMEG0000000012804, ENSAMEG0000000012869, PDZK1IP1, NUDT14, GALNS, HSPB6, LPL, MINDY1, SLC1A4, ECM1, DKK2, CRTP, ANPEP, HNMT, SCPEP1, PLOD3, MMP19, CRYZ, EPS8, CMBL, KCTD12, SLPI, PFPBP2, GAS6, NANS, TIMP4, RHOQ, MRRAS, TMEMBA, TRBP10, ZPI, SDC2, SMD3L3A, SLCA8, OAF, SLC1A5, CYBRD1, ENSAMEG0000000008780, AHCY, DMXL2, EMLN2, GNPDA1, VIM, PLED1, HSPB1, F5, FBP1, COMT, SERPINE2, LYZ, APOH, ANGPTL6, FAM168B, BLVRB, PLD3, CTSZ, SOGA1, GXP1, LAMB2, TUBB6, C3orf58, CFp, TIMP1, ECM2, A2M, RRA5, GSF8, ENSAMEG0000000014335, NAAA, CCL14, ENSAMEG0000000014908, SERPINE1, DAA, SGSH, ELN, ENSAMEG0000000015548, PRADC1, HFE, ACTG1, PEBP1, CD36, EFHD1, TNF5F12, TNF5F13, IL25, APP, F13A1, ABCA3, PTGR1, CDC180, CDC42BPB, ENDO1D1, ATP6 |
Table 3 Functional enrichment analysis based on the representation of GO terms (Continued)

| GO.ID      | Term                        | Ontology                  | Significant | Annotated | Pvalue | Qvalue | Signi_id | Signi_symbol |
|------------|-----------------------------|---------------------------|-------------|-----------|--------|--------|----------|--------------|
| ENSAMEG000000012943, ENSAMEG000000013536, ENSAMEG000000013869, ENSAMEG000000013931, ENSAMEG000000014335, ENSAMEG000000014605, ENSAMEG000000014906, ENSAMEG000000014908, ENSAMEG000000015156, ENSAMEG000000015157, ENSAMEG000000015221, ENSAMEG000000015429, ENSAMEG000000015548, ENSAMEG000000015564, ENSAMEG000000015938, ENSAMEG000000016015, ENSAMEG000000016143, ENSAMEG000000016224, ENSAMEG000000016446, ENSAMEG000000016729, ENSAMEG000000016736, ENSAMEG000000016820, ENSAMEG000000017184, ENSAMEG000000017422, ENSAMEG000000017862, ENSAMEG000000017964, ENSAMEG000000018103, ENSAMEG000000018199, ENSAMEG000000019907, ENSAMEG000000023442 | GO:0044421 extracellular region part                           | cellular component     | 84/287      |          | 2524/14310 | 0.000000070 | 0.0004 | PDZK1IP1, NUDT14, GALNS, LPL, MINDY1, SLC1A4, ECM1, DKK2, CRTAP, ANPEP, HNMT, SCPBP1, PLOD3, MMP19, CRYZ, EPS8, CMBl, KCTD12, SLPI, PFPBP2, GA56, NANS, RHOCQ, MRAS, TMEBBA, TRIP10, ZP1, SDC2, SMDP1A, SLC7A8, DAF, SLC1A5, CYBRD1, ENSAMEG000000008780, AHCY, DMXL2, EMILIN2, GNPD1A, VIM, PLBD1, HSFB1, F5, FBP1, COMT, SERPINE2, LY2, APOH, ANGPTL6, FAM168B, BLVRB, PLD3, CTS2, SOGA1, GPX1, LAMB2, TUBB6, C3orf58, CFP, TIMP1, ECM2, A2M, MRAS, ICSF8, ENSAMEG000000014335, NAAA, SERPINB10, GAA, SGSH, ELN, PRADCl, HFE, ACTG1, PEPB1, CD36, EFHD1, TNFSF13, APP, F13A1, ABCA3, PTGR1, Ccdc80, CDC40B8, ENDOD1, ATP6 |
| GO.ID | Term | Ontology | Significant | Annotated | Pvalue | Qvalue | Signi_id | Signi_symbol |
|-------|------|----------|-------------|-----------|--------|--------|----------|--------------|
| ENSAMEG000000011403 | ENSAMEG000000011494 | ENSAMEG000000011617 | ENSAMEG000000011687 | ENSAMEG000000011820 | ENSAMEG000000011842 | ENSAMEG000000012016 | ENSAMEG000000012043 | ENSAMEG000000012457 | ENSAMEG000000012482 | ENSAMEG000000012483 | ENSAMEG000000012511 | ENSAMEG000000012590 | ENSAMEG000000012708 | ENSAMEG000000012743 | ENSAMEG000000012804 | ENSAMEG000000012869 | ENSAMEG000000012943 | ENSAMEG000000013931 | ENSAMEG000000013980 | ENSAMEG000000014335 | ENSAMEG000000014605 | ENSAMEG000000015156 | ENSAMEG000000015157 | ENSAMEG000000015221 | ENSAMEG000000015429 | ENSAMEG000000015564 | ENSAMEG000000015938 | ENSAMEG000000016015 | ENSAMEG000000016143 | ENSAMEG000000016224 | ENSAMEG000000016446 | ENSAMEG000000016736 | ENSAMEG000000017184 | ENSAMEG000000017422 | ENSAMEG000000017862 | ENSAMEG000000017964 | ENSAMEG000000018103 | ENSAMEG000000018199 | ENSAMEG000000019007 | ENSAMEG000000023442 |
| GO:0005615 | extracellular space cellular component | 79/287 | 0.00000072 | 0.0005 | PDZK1IP1, NUDT14, GALNS, LPL, MINDY1, SLC1A4, ECM1, DKK2, CRTAP, ANPEP, HNMT, SCPEP1, PLOD3, MMP19, CRYP, EPSP, CMBL, XCTD12, SLP, PPIFBP2, GAS6, NANS, RHOD, MRAS, TMEM8A, TRIP10, SMPDL3A, SLC7A8, OAF, SLC1A5, CYBRD1, ENSAMEG00000008780, ACHE, DMD2, EMLIN3, GNPD1A1, VM, PLBD1, HSPB1, F5, FBIP1, COMT, SRPRINE2, LYZ, APOH, ANGPTL6, FAM168B, BLVRB, PLD3, CTSZ, SOGA1, GPX1, LAMB2, TUBB6, C3orf58, TIMP1, A2M, Rras, IGF1, ENSAMEG000000014335, NAA, SERPINB10, GAA, SGSH, PRAD1, HFE, ACTG1, PEBP1, CD36, EFHD1, TNFSF13, APP, F13A1, ABCA3, PTGR1, CDC48B, CDC42BPB, ENDOD1, ATP6 |
### Table 3: Functional enrichment analysis based on the representation of GO terms (Continued)

| GO ID     | Term Description         | Ontology        | Significant | Annotated | P-value       | Q-value       | Signi_id | Signi_symbol |
|-----------|--------------------------|-----------------|-------------|-----------|---------------|---------------|----------|--------------|
| ENSAMEG000000000015 | extracellular organelle | Cellular component | 63/287     | 1929/14310 | 0.00004900    | 0.013         |          |              |
| ENSAMEG000000000015 | extracellular organelle | Cellular component | 63/287     | 1929/14310 | 0.00004900    | 0.013         |          |              |

You et al. BMC Genomic Data (2021) 22:44 Page 19 of 56
Table 3 Functional enrichment analysis based on the representation of GO terms (Continued)

| GO.ID     | Term                  | Ontology        | Significant | Annotated | Pvalue  | Qvalue  | Signi_id | Signi_symbol |
|-----------|-----------------------|-----------------|-------------|-----------|---------|---------|----------|--------------|
|          |                       |                 |             |           |         |         |          |              |
| ENSAMEG00000007468, ENSAMEG00000007631, ENSAMEG00000008008, ENSAMEG00000008407, ENSAMEG00000008620, ENSAMEG00000009105, ENSAMEG00000010219, ENSAMEG00000010536, ENSAMEG00000010841, ENSAMEG00000011403, ENSAMEG00000011494, ENSAMEG00000011617, ENSAMEG00000011687, ENSAMEG00000011820, ENSAMEG00000011842, ENSAMEG00000012016, ENSAMEG00000012043, ENSAMEG00000012457, ENSAMEG00000012482, ENSAMEG00000012483, ENSAMEG00000012511, ENSAMEG00000012590, ENSAMEG00000012708, ENSAMEG00000012743, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000013931, ENSAMEG00000013980, ENSAMEG00000014605, ENSAMEG00000015157, ENSAMEG00000015221, ENSAMEG00000015564, ENSAMEG00000016015, ENSAMEG00000016143, ENSAMEG00000016446, ENSAMEG00000016736, ENSAMEG00000017184, ENSAMEG00000017964, ENSAMEG00000018199, ENSAMEG00000019907, ENSAMEG00000023442 | GO: 1903561 extracellular vesicle cellular component | 63/287 | 1929/14310 | 0.00004900 | 0.013 |          |          |
|          |                       |                 | 63/287      | 1929/14310 | 0.00004900 | 0.013 |          |              |
| PDZK1IP1, NUDT14, GALNS, LPL, MINDY1, SLCA14, ECM1, ANPEP, HNMT, SCPEP1, POLD3, CRYZ, EPS8, CMBL, KCTD12, SLPI, GAS5, NANS, RHOQ, MRAS, TMEM8A, TRIP10, SMFDL3A, SLCA78, DAF, SLCA9S, CYB5D1, AHCY, GNPD, A1, VM, HSPB1, F5, FBP1, COMT, SERPINE2, LYZ, APOH, ANGPTL6, FAM168B, BLVRB, PLD3, CTSZ, SOGA1, GPX1, LAMB2, TUBB6, TIMP1, A2M, RRAS, KGFB3, NAAA, GAA, SGSH, PRAD1C1, ACTG1, PEBP1, EFHD1, TNFSF13, APP, PTGR1, CDC42BPB, ENDOD1, ATP6 |
Table 3 Functional enrichment analysis based on the representation of GO terms *(Continued)*

| GO.ID          | Term                  | Ontology  | Significant | Annotated | Pvalue     | Qvalue     | Signi_id | Signi_symbol |
|----------------|-----------------------|-----------|-------------|-----------|------------|------------|----------|--------------|
| ENSAMEG00000011403, ENSAMEG00000011494, ENSAMEG00000011617, ENSAMEG00000011687, ENSAMEG00000011842, ENSAMEG00000012016, ENSAMEG00000012043, ENSAMEG00000012457, ENSAMEG00000012482, ENSAMEG00000012483, ENSAMEG00000012511, ENSAMEG00000012590, ENSAMEG00000012708, ENSAMEG00000012743, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000013931, ENSAMEG00000013980, ENSAMEG00000014605, ENSAMEG00000015157, ENSAMEG00000015221, ENSAMEG00000015564, ENSAMEG00000016015, ENSAMEG00000016143, ENSAMEG00000016446, ENSAMEG00000016736, ENSAMEG00000017184, ENSAMEG00000017964, ENSAMEG00000018199, ENSAMEG00000019907, ENSAMEG00000023442 | GO:0070062 extracellular exosome cellular component | 61/287 1918/14310 | 0.00014000 0.0319 | 1918/14310 | 0.00014000 0.0319 | PDZK1IP1, NUDT14, GALNS, LPL, MINDY1, SLC1A4, ECM1, ANPEP, HNMT, SCPEP1, PLOD3, CRYZ, EPS58, CMBL, KCTD12, SLP1, GAS6, NANS, RHOQ, MRA5, TMEM8A, TRIP10, SNAPC3, SLC7A8, DAF, SLC1A5, CYRD1, AHCY, GPN2, A1, VM, HSPB1, FBP1, COMT, LYZ, APOH, ANGPTL6, FAM168B, BLVRB, PLOD3, CTSZ, SOGA1, GPX1, LAMR2, TUBB6, TIMP1, A2M, RAS, IGFBF8, NAAA, GAA, SGSH, PRAD1, ACTG1, PEBP1, EFHD1, TNFSF13, APP, PTGR1, CDC42BPI, ENDO1, ATP6 |

You et al. BMC Genomic Data (2021) 22:44 Page 21 of 56
| GO.ID   | Term                          | Ontology    | Significant | Annotated  | Pvalue  | Qvalue | Signi_id   | Signi_symbol                                                                 |
|--------|-------------------------------|-------------|-------------|------------|---------|--------|------------|----------------------------------------------------------------------------|
| ENSAMEG000000012511, ENSAMEG000000012590, ENSAMEG000000012708, ENSAMEG000000012743, ENSAMEG000000012943, ENSAMEG000000013869, ENSAMEG000000013931, ENSAMEG000000013980, ENSAMEG000000014605, ENSAMEG000000015157, ENSAMEG000000015221, ENSAMEG000000015564, ENSAMEG000000016015, ENSAMEG000000016143, ENSAMEG000000016446, ENSAMEG000000016736, ENSAMEG000000017184, ENSAMEG000000017964, ENSAMEG000000019007, ENSAMEG0000000223442 | GO:0031012, extracellular matrix, cellular component | 17/287 | 342/14310 | 0.00056000 | 0.1117 | ENSAMEG00000001502, ENSAMEG00000003068, ENSAMEG00000004196, ENSAMEG00000007340, ENSAMEG00000007342, ENSAMEG00000009998, ENSAMEG00000010536, ENSAMEG00000010841, ENSAMEG00000011687, ENSAMEG00000011842, ENSAMEG00000012708, ENSAMEG00000012869, ENSAMEG00000012943, ENSAMEG00000013536, ENSAMEG00000015429, ENSAMEG00000016015, ENSAMEG00000018103 | ECM1, MMP19, SLPI, ZP1, SDC2, EMILIN2, VM, HSPB1, SERPINE2, APOH, LAMB2, CFP, TIMP1, ECM2, ELN, ACTG1, CCDC80 |
| ENSAMEG00000001502, ENSAMEG00000003068, ENSAMEG00000004196, ENSAMEG00000004441, ENSAMEG00000004774, ENSAMEG00000004836, ENSAMEG00000006630, ENSAMEG00000010076, ENSAMEG00000010287, ENSAMEG00000010991, ENSAMEG00000011687, ENSAMEG00000012590, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017990 | GO:0052547, regulation of peptidase activity, biological process | 17/269 | 260/13500 | 0.00001800 | 0.127 | ENSAMEG00000001502, ENSAMEG00000004196, ENSAMEG00000004441, ENSAMEG00000004774, ENSAMEG00000004836, ENSAMEG00000006630, ENSAMEG00000010076, ENSAMEG00000010287, ENSAMEG00000010991, ENSAMEG00000011687, ENSAMEG00000012590, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017990 | ECM1, SLPI, GAS6, PPARG, TIMP4, NOD1, ENSAMEG000000010076, KIAA0141, HIP1, SERPINE2, GPX1, TIMP1, A2M, NLRP3, ARRB1, APP, BAD |
| ENSAMEG00000001502, ENSAMEG00000004196, ENSAMEG00000004441, ENSAMEG00000004774, ENSAMEG00000004836, ENSAMEG00000006630, ENSAMEG00000010076, ENSAMEG00000010287, ENSAMEG00000010991, ENSAMEG00000011687, ENSAMEG00000012590, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017990 | GO:0052548, regulation of endopeptidase activity, biological process | 16/269 | 234/13500 | 0.00001900 | 0.127 | ENSAMEG00000001502, ENSAMEG00000004196, ENSAMEG00000004441, ENSAMEG00000004774, ENSAMEG00000004836, ENSAMEG00000006630, ENSAMEG00000010076, ENSAMEG00000010287, ENSAMEG00000010991, ENSAMEG00000011687, ENSAMEG00000012590, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017990 | SLPI, GAS6, PPARG, TIMP4, NOD1, ENSAMEG000000010076, KIAA0141, HIP1, SERPINE2, GPX1, TIMP1, A2M, NLRP3, ARRB1, APP, BAD |
### Table 3 Functional enrichment analysis based on the representation of GO terms (Continued)

| GO.ID  | Term                                      | Ontology                  | Significant | Annotated | Pvalue   | Qvalue | Signi_id                                                                 | Signi_symbol                  |
|--------|-------------------------------------------|----------------------------|-------------|-----------|----------|--------|---------------------------------------------------------------------------|--------------------------------|
| GO:0019897 | extrinsic component of plasma membrane    | cellular component        | 5/287       | 63/14310  | 0.00161  | 0.2699 | ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000011687, ENSAMEG000000013803, ENSAMEG00000016396, ENSAMEG00000017737 | CYTH3, ENSAMEG000000008528, SEPRINE2, FES, DLG4, GNG11 |
| GO:0031234 | extrinsic component of cytoplasmic side of plasma membrane | cellular component        | 5/287       | 44/14310  | 0.00179  | 0.2699 | ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000013803, ENSAMEG00000016396, ENSAMEG00000017737 | CYTH3, ENSAMEG000000008528, FES, DLG4, GNG11 |
| GO:0098802 | plasma membrane receptor complex          | cellular component        | 8/287       | 112/14310 | 0.00186  | 0.2699 | ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000013803, ENSAMEG00000016396, ENSAMEG00000017737 | EP58, ENSAMEG000000004337, ITGAX, PORCN, ACVR1, HFE, GABBR1, DLG4 |
| GO:0005774 | vacuolar membrane                        | cellular component        | 10/287      | 174/14310 | 0.00271  | 0.3487 | ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000013803, ENSAMEG00000016396, ENSAMEG00000017737 | TCIRG1, ANPEP, SFB2, SPPL2B, TMEM8A, CLCN7, GAA, CD68, TMEM175, SLC29A3 |
| GO:0043235 | receptor complex                         | cellular component        | 11/287      | 206/14310 | 0.003    | 0.3487 | ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000013803, ENSAMEG00000016396, ENSAMEG00000017737 | EP58, ENSAMEG000000004337, ITGAX, ENSAMEG000000010559, PORCN, ACVR1, HFE, GABBR1, CD36, DLG4, APP |
| GO:0044437 | vacuolar part                            | cellular component        | 10/287      | 180/14310 | 0.00345  | 0.3487 | ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000013803, ENSAMEG00000016396, ENSAMEG00000017737 | TCIRG1, ANPEP, SFB2, SPPL2B, TMEM8A, CLCN7, GAA, CD68, TMEM175, SLC29A3 |
| GO:0031966 | mitochondrial membrane                   | cellular component        | 15/287      | 338/14310 | 0.00348  | 0.3487 | ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000013803, ENSAMEG00000016396, ENSAMEG00000017737 | HADH, ENSAMEG000000002314, CHDH, RAB32, RNF144B, TMEM173, HAX1, IKBKE, EHD1, BAD, ND1, ND2, COX1, ATP6, CYTB |
| GO:0008805 | whole membrane                           | cellular component        | 22/287      | 592/14310 | 0.00409  | 0.3487 | ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000013803, ENSAMEG00000016396, ENSAMEG00000017737 | TCIRG1, KCNQ1, ANPEP, SFB2, SPPL2B, RAB32, RHQO, TMEM8A, CLCN7, HIP1, TMEM173, HAX1, |
| GO.ID   | Term                          | Ontology     | Significant | Annotated | Pvalue  | Qvalue  | Signi_id | Signi_symbol |
|---------|-------------------------------|--------------|-------------|-----------|---------|---------|----------|--------------|
| GO: 0005764 | lysosome cellular component   | 13/287       | 279/14310   | 0.00424   | 0.3487 | ENSAMEG0000001230, ENSAMEG0000001230, ENSAMEG0000002069, ENSAMEG0000002535, ENSAMEG0000005486, ENSAMEG0000007046, ENSAMEG0000007997, ENSAMEG0000012483, ENSAMEG0000014605, ENSAMEG0000019739, ENSAMEG0000017739, ENSAMEG0000017739 | C3orf58, GAA, CD36, CD68, TMEM175, APP, SLC29A3, ABCA3, BAD, KCNE3 |
| GO: 0005765 | lysosomal membrane cellular component | 9/287       | 156/14310   | 0.00424   | 0.3487 | ENSAMEG0000001230, ENSAMEG0000002535, ENSAMEG0000005486, ENSAMEG0000007046, ENSAMEG0000007997, ENSAMEG0000012483, ENSAMEG0000014605, ENSAMEG0000019739, ENSAMEG0000017739, ENSAMEG0000017739 | TCIRG1, ANPEP, SPP2L8, TMEM8A, CLCN7, CTSZ, NAAA, GAA, SGSH, CD68, TMEM175, SLC29A3 |
| GO: 0098852 | lytic vacuole membrane cellular component | 9/287       | 156/14310   | 0.00424   | 0.3487 | ENSAMEG0000001230, ENSAMEG0000002535, ENSAMEG0000005486, ENSAMEG0000007046, ENSAMEG0000007997, ENSAMEG0000012483, ENSAMEG0000014605, ENSAMEG0000019739, ENSAMEG0000017739, ENSAMEG0000017739 | TCIRG1, ANPEP, SPP2L8, TMEM8A, CLCN7, GAA, CD68, TMEM175, SLC29A3 |
| GO: 0000323 | lytic vacuole cellular component | 13/287       | 280/14310   | 0.00437   | 0.3487 | ENSAMEG0000001230, ENSAMEG0000002535, ENSAMEG0000005486, ENSAMEG0000007046, ENSAMEG0000007997, ENSAMEG0000012483, ENSAMEG0000014605, ENSAMEG0000019739, ENSAMEG0000017739, ENSAMEG0000017739 | TCIRG1, ANPEP, SPP2L8, TMEM8A, CLCN7, GAA, CD68, TMEM175, SLC29A3 |
| GO: 0048866 | endopeptidase inhibitor activity molecular function | 9/273       | 102/13119   | 0.00027   | 0.3916 | ENSAMEG0000004196, ENSAMEG0000004441, ENSAMEG0000004836, ENSAMEG0000010076, ENSAMEG0000011687, ENSAMEG0000012943, ENSAMEG0000012943, ENSAMEG0000012943, ENSAMEG0000012943, ENSAMEG0000012943 | SLPI, GAS6, TIMP4, ENSAMEG00000010076, SERPINE2, TIMP1, A2M, ARRB1, APP |
| GO.ID  | Term                              | Ontology     | Significant | Annotated | Pvalue  | Qvalue  | Signi_id                                                                 | Signi_symbol                  |
|--------|-----------------------------------|--------------|-------------|-----------|---------|---------|----------------------------------------------------------------------------|------------------------------|
| GO:0061135 | endopeptidase regulator activity | molecular function | 9/273       | 105/13119 | 0.00034 | 0.3916 | ENSAMEG00000013869, ENSAMEG00000017062, ENSAMEG00000017184 | SLPI, GAS6, TIMP4, ENSAMEG00000010076, SERPINE2, TIMP1, A2M, ARRB1, APP |
| GO:0008484 | sulfuric ester hydrolase activity | molecular function | 4/273       | 17/13119  | 0.00035 | 0.3916 | ENSAMEG00000000312, ENSAMEG00000013098, ENSAMEG00000014912, ENSAMEG00000015221 | GALNS, ENSAMEG00000013098, ENSAMEG00000014912, SGSH |
| GO:0061134 | peptidase regulator activity     | molecular function | 10/273      | 132/13119 | 0.00043 | 0.3916 | ENSAMEG00000004196, ENSAMEG00000004441, ENSAMEG00000004836, ENSAMEG00000010076, ENSAMEG00000011687, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000017062, ENSAMEG00000017184 | SLPI, GAS6, TIMP4, ENSAMEG00000010076, SERPINE2, TIMP1, A2M, ARRB1, APP, BAD |
| GO:0045275 | respiratory chain complex III     | cellular component | 2/287       | 6/14310   | 0.0057  | 0.4233 | ENSAMEG000000023436, ENSAMEG000000023455 | COX1, CYTB |
| GO:0008328 | ionotropic glutamate receptor complex | cellular component | 3/287       | 19/14310  | 0.0061  | 0.4233 | ENSAMEG000000003627, ENSAMEG000000012689, ENSAMEG000000016396 | EPS8, PORCN, DLG4 |
| GO:0098878 | neurotransmitter receptor complex | cellular component | 3/287       | 19/14310  | 0.0061  | 0.4233 | ENSAMEG000000003627, ENSAMEG000000012689, ENSAMEG000000016396 | EPS8, PORCN, DLG4 |
| GO:0098552 | side of membrane                  | cellular component | 10/287      | 198/14310 | 0.00671 | 0.425  | ENSAMEG000000005486, ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000011687, ENSAMEG000000013414, ENSAMEG000000013803, ENSAMEG000000015938, ENSAMEG000000016224, ENSAMEG000000016396, ENSAMEG000000017737 | SPPL2B, CYTH3, ENSAMEG000000008528, SERPINE2, CD163, FES, HFE, CD36, DLG4, GNG11 |
| GO:0005773 | vacuole                           | cellular component | 14/287      | 329/14310 | 0.00673 | 0.425  | ENSAMEG000000001230, ENSAMEG000000002069, ENSAMEG000000002535, ENSAMEG000000003625, ENSAMEG000000005486, ENSAMEG000000007046, ENSAMEG000000007997, ENSAMEG000000012483, ENSAMEG000000014605, ENSAMEG000000015157, ENSAMEG000000015221, ENSAMEG000000016765, ENSAMEG000000016795, ENSAMEG000000017737 | TCIRG1, KCNQ1, ANPEP, SFB2, SPPL2B, TMEM8A, CLCN7, CTSZ, NAAA, GAA, SGSH, CD68, TMEM175, SLC29A3 |
| GO:0005740 | mitochondrial envelope            | cellular component | 15/287      | 365/14310 | 0.00699 | 0.425  | ENSAMEG000000001814, ENSAMEG000000002314, ENSAMEG000000002398, ENSAMEG000000006014, ENSAMEG000000008424, ENSAMEG000000011225, ENSAMEG000000011750, ENSAMEG000000011831, ENSAMEG000000016446, ENSAMEG000000018369, ENSAMEG000000019717, ENSAMEG00000002069, ENSAMEG00000002535, ENSAMEG00000003625, ENSAMEG00000005486, ENSAMEG00000007046, ENSAMEG00000007997, ENSAMEG00000012483, ENSAMEG00000014605, ENSAMEG00000015157, ENSAMEG00000015221, ENSAMEG00000016765, ENSAMEG00000016795, ENSAMEG00000017737 | HADH, ENSAMEG00000002314, CHDH, RAB32, RNF144B, TMEM173, HAX1, KPBK, EFHD1, BAD, ND1, ND2, COX1, ATP6, CYTB |
| GO ID   | Term                               | Ontology            | Significant | Annotated | P value | Q value | Signi_id | Signi_symbol |
|---------|------------------------------------|---------------------|-------------|-----------|---------|---------|----------|--------------|
| GO: 0098562 | cytoplasmic side of membrane       | cellular component  | 6/287       | 85/14310  | 0.00719 | 0.425   | ENSAMEG0000017990, ENSAMEG0000023426, ENSAMEG0000023430, ENSAMEG0000023436, ENSAMEG0000023442, ENSAMEG0000023455 | SPPL2B, CYTH3, ENSAMEG0000008528, FES, DLG4, GNG11 |
| GO: 0005614 | interstitial matrix                | cellular component  | 2/287       | 7/14310   | 0.00788 | 0.4492  | ENSAMEG0000013536, ENSAMEG0000018103 | ECM2, CCDC80 |
| GO: 0030414 | peptidase inhibitor activity       | molecular function  | 9/273       | 114/13119 | 0.0062  | 0.4517  | ENSAMEG00000013536, ENSAMEG0000018103 | SLPI, GAS6, TIMP4, ENSAMEG00000010076, SERPINE2, TIMP1, A2M, ARRB1, APP |
| GO: 005912  | adherens junction                  | cellular component  | 14/287      | 337/14310 | 0.00824 | 0.4535  | ENSAMEG0000003894, ENSAMEG0000006768, ENSAMEG0000010536, ENSAMEG0000010841, ENSAMEG0000013263, ENSAMEG0000013297, ENSAMEG0000013803, ENSAMEG0000013931, ENSAMEG0000014689, ENSAMEG0000015431, ENSAMEG0000015837, ENSAMEG0000016015, ENSAMEG0000016087, ENSAMEG0000019485 | MRC2, CYTH3, VIM, Hspb1, REXO2, FHL2, FES, RRAS, TNS5, LIMK1, STXB5, ACTG1, POLIM1, CDC42EP4 |
| GO: 0031410 | cytoplasmic vesicle                | cellular component  | 25/287      | 755/14310 | 0.00965 | 0.4966  | ENSAMEG0000001276, ENSAMEG0000002069, ENSAMEG0000005486, ENSAMEG0000006014, ENSAMEG0000008800, ENSAMEG0000009193, ENSAMEG0000010113, ENSAMEG0000010991, ENSAMEG0000011403, ENSAMEG0000011509, ENSAMEG0000011687, ENSAMEG0000011750, ENSAMEG0000012016, ENSAMEG0000012483, ENSAMEG0000012804, ENSAMEG0000013803, ENSAMEG0000013931, ENSAMEG0000014724, ENSAMEG0000015326, ENSAMEG0000015938, ENSAMEG0000016396, ENSAMEG0000016795, ENSAMEG0000017062, ENSAMEG0000017184, ENSAMEG0000017862 | ZNRF1, KCNQ1, SPPL2B, RAB32, STEAP3, DGKH, DMXL2, SLA2, HIP1, F5, SEPT5, SERPINE2, HAX1, ANGPTL6, CTSZ, C3orf58, FES, ITSNI, MICALL2, HFE, DLG4, TMEM175, ARRB1, APP, ABCA3 |
| GO: 0007708 | intracellular vesicle              | cellular component  | 25/287      | 756/14310 | 0.0098  | 0.4966  | ENSAMEG0000001276, ENSAMEG0000002069, ENSAMEG0000005486, ENSAMEG0000006014, ENSAMEG0000008800, ENSAMEG0000009193, ENSAMEG0000010113, ENSAMEG0000010991, ENSAMEG0000011403, ENSAMEG0000011509, ENSAMEG0000011687, ENSAMEG0000011750, ENSAMEG0000012016, ENSAMEG0000012483, ENSAMEG0000012804, ENSAMEG0000013803, ENSAMEG0000013931, ENSAMEG0000014724, ENSAMEG0000015326, ENSAMEG0000015938, ENSAMEG0000016396, ENSAMEG0000016795, ENSAMEG0000017062, ENSAMEG0000017184, ENSAMEG0000017862 | ZNRF1, KCNQ1, SPPL2B, RAB32, STEAP3, DGKH, DMXL2, SLA2, HIP1, F5, SEPT5, SERPINE2, HAX1, ANGPTL6, CTSZ, C3orf58, FES, ITSNI, MICALL2, HFE, DLG4, TMEM175, ARRB1, APP, ABCA3 |
| GO.ID  | Term                        | Ontology          | Significant | Annotated | Pvalue  | Qvalue | Signi_id | Signi_symbol |
|--------|-----------------------------|-------------------|-------------|-----------|---------|--------|----------|-------------|
| GO:0070161 | anchoring junction           | cellular component | 14/287      | 345/14310 | 0.01    | 0.4966 | ENSAMEG0000003894, ENSAMEG0000006768, ENSAMEG00000010536, ENSAMEG00000010841, ENSAMEG00000013263, ENSAMEG00000013479, ENSAMEG00000013803, ENSAMEG00000013931, ENSAMEG00000014689, ENSAMEG00000015431, ENSAMEG00000015837, ENSAMEG00000016015, ENSAMEG00000016087, ENSAMEG00000019485, MRC2, CYTH3, VIM, HSPB1, REXO2, FHL2, FES, RAS, TNS3, LIMK1, STXB6, ACTG1, PDLIM1, CDC42EP4 |
| GO:0042627 | chylomicron                 | cellular component | 2/287       | 8/14310   | 0.01036 | 0.4966 | ENSAMEG0000000943, ENSAMEG00000011842, LPL, APOH |
| GO:0042641 | actomyosin                  | cellular component | 4/287       | 43/14310  | 0.01058 | 0.4966 | ENSAMEG00000015326, ENSAMEG00000017327, ENSAMEG00000017670, ENSAMEG00000018199, MICALL2, FHOD1, PDLIM4, CDC42BPB |
| GO:0030027 | lamellipodium               | cellular component | 6/287       | 96/14310  | 0.01272 | 0.5715 | ENSAMEG00000004387, ENSAMEG00000007278, ENSAMEG00000011750, ENSAMEG00000014724, ENSAMEG00000015431, ENSAMEG00000016087, ENSAMEG00000017670, PLXND1, ARHGAP31, HAX1, ITSN1, LIMK1, PDLIM4 |
| GO:0032281 | AMPA glutamate receptor complex | cellular component | 2/287       | 9/14310   | 0.01315 | 0.5715 | ENSAMEG00000012689, ENSAMEG00000016396, PORCN, DLG4 |
| GO:0031252 | cell leading edge           | cellular component | 9/287       | 187/14310 | 0.01325 | 0.5715 | ENSAMEG00000003627, ENSAMEG00000004387, ENSAMEG00000007278, ENSAMEG00000010536, ENSAMEG00000011750, ENSAMEG00000014724, ENSAMEG00000015431, ENSAMEG00000016087, ENSAMEG00000017670, EP58, PLXND1, CYTH3, ARHGAP31, VIM, HAX1, ITSN1, LIMK1, PDLIM4 |
| GO:0051336 | regulation of hydrolase activity | biological process | 28/269      | 690/13500 | 0.00028 | 0.5728 | ENSAMEG0000001502, ENSAMEG0000003625, ENSAMEG0000004106, ENSAMEG0000004196, ENSAMEG0000004441, ENSAMEG0000004474, ENSAMEG0000004836, ENSAMEG0000005271, ENSAMEG0000005566, ENSAMEG0000006453, ENSAMEG0000006630, ENSAMEG0000007278, ENSAMEG0000008477, ENSAMEG00000010076, ECM1, SBF2, SLPI, GAS6, PPARG, TIMP4, ENSAMEG0000005271, ADAP2, ARHGAP21, NOD1, ARHGAP31, HSCB, ENSAMEG00000010076, KIAA0141, HIP1, SDR14E2, APOH, GPx1, TIMP1, A2M, PPI1R1S, ENSAMEG00000015030, NLRP3, ARRB1, APP, BAD, CDC42EP2 |
| GO.ID   | Term                                      | Ontology                      | Significant | Annotated | Pvalue   | Qvalue   | Signi_id | Signi_symbol |
|---------|-------------------------------------------|-------------------------------|-------------|-----------|----------|----------|----------|--------------|
| GO:     | negative regulation of peptidase activity | biological process           | 11/269      | 157/13500 | 0.00031  | 0.5728   |          |              |
| 0010466 |                                          |                               |             |           |          |          |          |              |
| GO:     | immune system process                     | biological process           | 43/269      | 1258/13500| 0.00031  | 0.5728   |          |              |
| 002376  |                                          |                               |             |           |          |          |          |              |

**Table 3** Functional enrichment analysis based on the representation of GO terms (Continued)
### Table 3: Functional enrichment analysis based on the representation of GO terms (Continued)

| GO ID | Term                                               | Ontology                  | Significant | Annotated | P-value   | Q-value | Signi_id                     | Signi_symbol                  |
|-------|----------------------------------------------------|---------------------------|-------------|-----------|-----------|---------|------------------------------|------------------------------|
| GO:0055095 | lipoprotein particle mediated signaling          | biological process       | 2/269      | 1/13500   | 0.0004    | 0.5728  | ENSAMEG00000000943, ENSAMEG00000016224 | LPL, CD36                    |
| GO:0055096 | low-density lipoprotein particle mediated signaling | biological process       | 2/269      | 1/13500   | 0.0004    | 0.5728  | ENSAMEG00000000943, ENSAMEG00000016224 | LPL, CD36                    |
| GO:0010885 | regulation of cholesterol storage               | biological process       | 3/269      | 8/13500   | 0.00041   | 0.5728  | ENSAMEG00000000943, ENSAMEG00000004774, ENSAMEG00000016224 | LPL, PPARG, CD36             |
| GO:0045861 | negative regulation of proteolysis             | biological process       | 13/269     | 218/13500 | 0.00043   | 0.5728  | ENSAMEG00000001502, ENSAMEG00000004196, ENSAMEG00000004441, ENSAMEG00000004836, ENSAMEG0000010076, ENSAMEG0000011687, ENSAMEG0000012483, ENSAMEG0000012590, ENSAMEG0000012943, ENSAMEG0000013869, ENSAMEG0000015938, ENSAMEG0000017062, ENSAMEG0000017184 | ECM1, SLPI, GAS6, TIMP4, ENSAMEG00000010076, SERPINE2, CTSZ, GPX1, TIMP1, A2M, HFE, ARRB1, APP |
| GO:0030162 | regulation of proteolysis                        | biological process       | 20/269     | 436/13500 | 0.00047   | 0.5728  | ENSAMEG00000000272, ENSAMEG00000001502, ENSAMEG00000004196, ENSAMEG00000004441, ENSAMEG00000004836, ENSAMEG00000006630, ENSAMEG00000010076, ENSAMEG00000010287, ENSAMEG00000010991, ENSAMEG00000011687, ENSAMEG00000012483, ENSAMEG00000012590, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000015938, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017990 | CBFA2T3, ECM1, SLPI, GAS6, PPARG, TIMP4, NOD1, ENSAMEG00000010076, KIAA0141, HIP1, SERPINE2, CTSZ, GPX1, TIMP1, A2M, HFE, NLRP3, ARRB1, APP, BAD |
| GO:0006955 | immune response                                 | biological process       | 27/269     | 686/13500 | 0.00058   | 0.5728  | ENSAMEG00000001502, ENSAMEG00000003641, ENSAMEG00000004196, ENSAMEG00000004441, ENSAMEG00000004663, ENSAMEG00000004774, ENSAMEG00000004836, ENSAMEG00000006630, ENSAMEG00000010076, ENSAMEG00000010287, ENSAMEG00000010991, ENSAMEG00000011687, ENSAMEG00000012483, ENSAMEG00000012590, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000015938, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017990, ENSAMEG00000018841 | ECM1, CARDS, SLPI, GAS6, TRIM14, PPARG, POLJ2F2, SPPL2B, CYBB, TMEM173, IRBKE, SDHAF4, ERC1, GPX1, FES, A2M, HFE, NLRP3, ARRB1, APP, CEBPG, CYSLTR1 |
Table 3 Functional enrichment analysis based on the representation of GO terms (Continued)

| GO.ID | Term                          | Ontology         | Significant | Annotated | Pvalue   | Qvalue   | Signi_id                  | Signi_symbol                        |
|-------|-------------------------------|-------------------|-------------|-----------|----------|----------|---------------------------|-------------------------------------|
| GO: 0030851 | granulocyte differentiation    | biological process | 4/269       | 20/13500  | 0.00058  | 0.5728   | ENSAMEG00000019893         | CBFA2T3, HAX1, GATA2, IL2S          |
| GO: 0010951 | negative regulation of endopeptidase activity | biological process | 10/269      | 143/13500 | 0.00058  | 0.5728   | ENSAMEG00000000272, ENSAMEG00000011750, ENSAMEG00000015180, ENSAMEG00000016820 | SLPI, GAS6, TIMP4, ENSAMEG00000010076, SERPINE2, GPX1, TIMP1, A2M, ARRB1, APP |
| GO: 0010878 | cholesterol storage           | biological process | 3/269       | 9/13500   | 0.0006   | 0.5728   | ENSAMEG00000000943, ENSAMEG00000004774, ENSAMEG00000016224 | LPL, PPARC, CD36                     |
| GO: 0098797 | plasma membrane protein complex | cellular component | 14/287      | 360/14310 | 0.01408  | 0.5914   | ENSAMEG0000002069, ENSAMEG00000036327, ENSAMEG0000004337, ENSAMEG0000006806, ENSAMEG0000007284, ENSAMEG0000008528, ENSAMEG0000010991, ENSAMEG0000012689, ENSAMEG0000014318, ENSAMEG0000015938, ENSAMEG0000016101, ENSAMEG0000016396, ENSAMEG0000017737, ENSAMEG0000020287 | KCNQ1, EPS8, ENSAMEG0000004337, ITGAX, CYBB, ENSAMEG0000008528, HIP1, PORCN, ACVR1, HFE, GABBR1, DLG4, GNG11, KCNE3 |
| GO: 1990204 | oxidoreductase complex        | cellular component | 5/287       | 72/14310  | 0.01471  | 0.602    | ENSAMEG00000007284, ENSAMEG00000023426, ENSAMEG00000023430, ENSAMEG00000023436, ENSAMEG00000023455 | CYBB, ND1, ND2, COX1, CYTB           |
| GO: 0098796 | membrane protein complex      | cellular component | 24/287      | 749/14310 | 0.01602  | 0.6392   | ENSAMEG0000001230, ENSAMEG0000002069, ENSAMEG00000036327, ENSAMEG0000004337, ENSAMEG0000006806, ENSAMEG0000007284, ENSAMEG0000008528, ENSAMEG0000009860, ENSAMEG0000010991, ENSAMEG0000012689, ENSAMEG0000012804, ENSAMEG0000014318, ENSAMEG0000015938, ENSAMEG0000016101, ENSAMEG0000016396, ENSAMEG0000017737, ENSAMEG0000020287, ENSAMEG0000023426, ENSAMEG0000023430, ENSAMEG0000023436, ENSAMEG0000023442, ENSAMEG0000023455 | TCIRG1, KCNQ1, EPS8, ENSAMEG0000004337, AP2A2, ITGAX, CYBB, ENSAMEG0000008528, ORMDL3, HIP1, PORCN, C3orf58, ACVR1, HFE, GABBR1, DLG4, GNG11, OSTM4, KCNE3, ND1, ND2, COX1, ATP6, CYTB |
| GO: 0009898 | cytoplasmic side of plasma membrane | cellular component | 5/287       | 75/14310  | 0.01729  | 0.6534   | ENSAMEG0000006768, ENSAMEG0000008528, ENSAMEG0000013803, ENSAMEG0000016396, ENSAMEG0000017737 | CYTH3, ENSAMEG0000008528, FES, DLG4, GNG11 |
Table 3 Functional enrichment analysis based on the representation of GO terms (Continued)

| GO ID     | Term          | Ontology     | Significant Annotated | Pvalue  | Qvalue | Signi id                | Signi_symbol       |
|-----------|---------------|--------------|-----------------------|---------|--------|-------------------------|--------------------|
| GO:0044304| main axon     | cellular     | 3/287                 | 28/14310| 0.01806| ENSAMEG00000012095,     | NAV1, DLG4, DAGLA   |
|           | cellular      | component    |                       |         |        | ENSAMEG0000016396,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000016637       |                    |
| GO:0016020| membrane      | cellular     | 165/287               | 7329/14310| 0.0182 | PDZK1IP1, MARVELD1,     |                    |
|           | cellular      | component    |                       |         |        | ENSAMEG00000000015,     |                    |
|           |               |              |                       |         |        | ENSAMEG0000000095,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000000381,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000000590,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000000602,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000000751,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000000943,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000001092,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000001230,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000001267,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000001276,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000001318,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000001514,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000001784,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000001814,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000001879,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000002069,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000002314,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000002398,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000002415,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000002508,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000002535,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003020,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003086,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003089,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003400,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003591,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003625,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003627,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003733,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003778,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003894,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000003970,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000004337,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000004387,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000004469,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000004481,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000004482,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000004516,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000004600,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000004900,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000005113,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000005412,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000005486,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000005566,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000005794,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000005927,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006014,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006082,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006083,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006288,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006368,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006453,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006630,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006677,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006768,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006806,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000006896,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000007046,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000007123,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000007284,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000007302,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000007340,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000007342,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000007375,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000007631,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000007997,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000008407,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000008416,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000008424,      |                    |
|           |               |              |                       |         |        | ENSAMEG0000008433,      |                    |
| GO.ID          | Term                      | Ontology | Significant | Annotated | Pvalue | Qvalue | Signi_id | Signi_symbol |
|---------------|---------------------------|----------|-------------|-----------|--------|--------|----------|--------------|
| ENSAMEG00000008497, ENSAMEG00000008528, ENSAMEG00000008575, ENSAMEG00000008620, ENSAMEG00000009198, ENSAMEG00000009367, ENSAMEG00000009654, ENSAMEG00000009860, ENSAMEG00000010113, ENSAMEG00000010536, ENSAMEG00000010841, ENSAMEG00000010857, ENSAMEG00000010898, ENSAMEG00000010991, ENSAMEG00000011225, ENSAMEG00000011403, ENSAMEG00000011408, ENSAMEG00000011455, ENSAMEG00000011509, ENSAMEG00000011564, ENSAMEG00000011617, ENSAMEG00000011687, ENSAMEG00000011750, ENSAMEG00000011831, ENSAMEG00000011849, ENSAMEG00000011977, ENSAMEG00000012457, ENSAMEG00000012482, ENSAMEG00000012576, ENSAMEG00000012804, ENSAMEG00000013414, ENSAMEG00000013445, ENSAMEG00000013447, ENSAMEG00000013560, ENSAMEG00000013803, ENSAMEG00000013851, ENSAMEG00000013931, ENSAMEG00000013980, ENSAMEG00000014079, ENSAMEG00000014229, ENSAMEG00000014318, ENSAMEG00000014354, ENSAMEG00000014490, ENSAMEG00000014678, ENSAMEG00000014705, ENSAMEG00000014724, ENSAMEG00000015157, ENSAMEG00000015252, ENSAMEG00000015326, ENSAMEG00000015431, ENSAMEG00000015837, ENSAMEG00000015938, ENSAMEG00000015964, ENSAMEG00000016015, ENSAMEG00000016101, ENSAMEG00000016224, ENSAMEG00000016396, ENSAMEG00000016446, ENSAMEG00000016637, ENSAMEG00000016729, ENSAMEG00000016736, ENSAMEG00000016739, ENSAMEG00000016765, ENSAMEG00000016795, ENSAMEG00000016820, ENSAMEG00000017107, ENSAMEG00000017184, ENSAMEG00000017327, ENSAMEG00000017329, ENSAMEG00000017737, ENSAMEG00000017739, |
| GO.ID       | Term                                         | Ontology               | Significant | Annotated | Pvalue | Qvalue | Signi_id   | Signi_symbol |
|-------------|----------------------------------------------|------------------------|-------------|-----------|--------|--------|------------|--------------|
| GO:0005608  | laminin-3 complex                            | cellular component     | 1/287       | 1/14310   | 0.0206 | 0.6534 | ENSAMEG0000012708 | LAMB2        |
| GO:0038039  | G-protein coupled receptor heterodimeric complex | cellular component     | 1/287       | 1/14310   | 0.0206 | 0.6534 | ENSAMEG0000016101 | GABBR1       |
| GO:0043196  | varicosity                                    | cellular component     | 1/287       | 1/14310   | 0.0206 | 0.6534 | ENSAMEG0000016637 | DAGLA        |
| GO:0097232  | lamellar body membrane                        | cellular component     | 1/287       | 1/14310   | 0.0206 | 0.6534 | ENSAMEG0000017862 | ABCA3        |
| GO:0097233  | alveolar lamellar body membrane               | cellular component     | 1/287       | 1/14310   | 0.0206 | 0.6534 | ENSAMEG0000017862 | ABCA3        |
| GO:0098839  | postsynaptic density membrane                 | cellular component     | 1/287       | 1/14310   | 0.0206 | 0.6534 | ENSAMEG0000016396 | DLG4         |
| GO:0098803  | respiratory chain complex                     | cellular component     | 4/287       | 54/14310  | 0.02287| 0.73   | ENSAMEG00000223426, ENSAMEG00000223430, ENSAMEG00000223436, ENSAMEG00000223455 | ND1, ND2, COX1, CYTB |
| GO:0043281  | regulation of cysteine-type endopeptidase activity involved in apoptotic process | biological process | 9/269       | 125/13500 | 0.000388| 0.7841 | ENSAMEG00000000004441, ENSAMEG00000000004774, ENSAMEG00000000006630, ENSAMEG00000000010287, ENSAMEG00000000010991, ENSAMEG00000000012590, ENSAMEG00000000016947, ENSAMEG00000000017062, ENSAMEG00000000017990 | GAS6, PPARG, NOD1, KIAA0141, HIP1, GPX1, NLRP3, ARRB1, BAD |
| GO:0005925  | focal adhesion                                | cellular component     | 11/287      | 281/14310 | 0.02663| 0.8278 | ENSAMEG0000000003894, ENSAMEG0000000010536, ENSAMEG0000000010841, ENSAMEG0000000013263, ENSAMEG0000000013297, ENSAMEG0000000013803, ENSAMEG0000000013931, ENSAMEG0000000014689, ENSAMEG0000000015431, ENSAMEG0000000016015, ENSAMEG0000000016087 | MRC2, VIM, HSPB1, REXO2, FHL2, FES, RRAS, TNS3, LIMK1, ACTG1, PDUM1 |
| GO:0070069  | cytochrome complex                            | cellular component     | 2/287       | 13/14310  | 0.02702| 0.8278 | ENSAMEG00000000223436, ENSAMEG00000000223455 | COX1, CYTB |

You et al. BMC Genomic Data (2021) 22:44 Page 33 of 56
### Table 3  Functional enrichment analysis based on the representation of GO terms (Continued)

| GO.ID | Term                        | Ontology          | Significant | Annotated   | pvalue  | Qvalue  | Signi_id                          | Signi_symbol                          |
|-------|-----------------------------|--------------------|-------------|-------------|---------|---------|----------------------------------|----------------------------------------|
| GO:0019898 | extrinsic component of membrane | cellular component | 6/287       | 115/14310   | 0.02828 | 0.8278 | ENSAMEG000000006768, ENSAMEG000000008528, ENSAMEG000000011687, ENSAMEG000000013803, ENSAMEG000000016396, ENSAMEG000000017737 | CYTH3, ENSAMEG000000008528, SERPINE2, FES, DLG4, GNB1 |
| GO:0005924 | cell-substrate adherens junction | cellular component | 11/287      | 284/14310   | 0.02848 | 0.8278 | ENSAMEG00000003894, ENSAMEG00000010536, ENSAMEG00000010841, ENSAMEG00000013263, ENSAMEG00000013297, ENSAMEG00000013803, ENSAMEG00000013931, ENSAMEG00000014689, ENSAMEG00000015431, ENSAMEG00000016015, ENSAMEG00000016087 | MRC2, VIM, HSPB1, REXO2, FHL2, FES, RAR5, TNF5, LIMP1, ACTG1, PDUM1 |
| GO:0120025 | plasma membrane bounded cell projection | cellular component | 24/287      | 793/14310   | 0.02923 | 0.8278 | ENSAMEG00000003086, ENSAMEG00000003627, ENSAMEG00000003733, ENSAMEG000000004387, ENSAMEG00000006768, ENSAMEG00000007278, ENSAMEG00000008620, ENSAMEG00000010536, ENSAMEG00000010841, ENSAMEG00000011509, ENSAMEG00000011750, ENSAMEG00000012095, ENSAMEG00000014724, ENSAMEG00000015252, ENSAMEG00000015431, ENSAMEG00000015964, ENSAMEG00000016224, ENSAMEG00000016396, ENSAMEG00000016637, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017670, ENSAMEG00000018345, ENSAMEG00000020287 | NRP2, EPS8, INPP5E, PLXND1, CYTH3, ARHGAP31, CYBRD1, VIM, HSPB1, SEPT5, HAX1, NAV1, ITSN1, RILPL1, LIMP1, ADCA4, CD36, DLG4, DAGLA, ARRB1, APP, PDUM1, TLL5, KCNE3 |
| GO:0031090 | organelle membrane | cellular component | 31/287      | 1086/14310  | 0.02933 | 0.8278 | ENSAMEG00000001230, ENSAMEG00000001814, ENSAMEG00000002314, ENSAMEG00000002398, ENSAMEG00000002535, ENSAMEG00000003625, ENSAMEG00000003733, ENSAMEG00000005486, ENSAMEG00000006014, ENSAMEG00000007046, ENSAMEG00000007997, ENSAMEG00000008424, ENSAMEG00000010991, ENSAMEG00000011225, ENSAMEG00000011750, ENSAMEG00000011831, ENSAMEG00000012804, ENSAMEG00000015157, ENSAMEG00000016446, ENSAMEG00000016765, ENSAMEG00000017095, ENSAMEG00000017739, ENSAMEG00000017862, ENSAMEG00000017990, ENSAMEG00000019719, ENSAMEG00000020281, ENSAMEG00000023426, ENSAMEG00000023430, ENSAMEG00000023436, ENSAMEG00000023442, TCRG1, HADH, ENSAMEG00000002314, CDH1, ANPEP, SBF2, INPP5E, SPPL2B, RAB32, TMEM18A, CLCN7, RNF1-448, HIP1, TMEM173, HAX1, JKBK, Cxcr5, GAA, EFHD1, CD68, TMEM175, SLC29A3, ABCA3, BAD, B3GNT9, B3GALT2, ND1, ND2, COX1, ATP6, CYTB |
| GO.ID  | Term                                      | Ontology                  | Significant | Annotated | Pvalue  | Qvalue  | Signi_id | Signi_symbol                      |
|--------|-------------------------------------------|---------------------------|-------------|-----------|---------|---------|----------|-----------------------------------|
| GO:0030055 | cell-substrate junction cellular component     |                           | 11/287      | 287/14310 | 0.03042 | 0.8278 | ENSAMEG000000023455 ENSAMEG00000003894, ENSAMEG000000010536, ENSAMEG00000001841, ENSAMEG000000013263, ENSAMEG000000013297, ENSAMEG000000013803, ENSAMEG000000013931, ENSAMEG000000014689, ENSAMEG000000015431, ENSAMEG000000016015, ENSAMEG000000016087 | MRC2, VIM, HSPB1, REXO2, FHL2, FES, RAR5, TNS5, LIMK1, ACTG1, PDUM1 |
| GO:0031091 | platelet alpha granule cellular component |                           | 2/287       | 14/14310  | 0.03112 | 0.8278 | ENSAMEG00000011403, ENSAMEG00000011687 | F5, SERPINE2 |
| GO:0034361 | very-low-density lipoprotein particle cellular component |                           | 2/287       | 14/14310  | 0.03112 | 0.8278 | ENSAMEG00000000943, ENSAMEG00000011842 | LPL, APOH |
| GO:0034385 | triglyceride-rich plasma lipoprotein particle cellular component |                           | 2/287       | 14/14310  | 0.03112 | 0.8278 | ENSAMEG00000000943, ENSAMEG00000011842 | LPL, APOH |
| GO:0061077 | chaperone-mediated protein folding biological process |                           | 4/269       | 23/13500  | 0.00101 | 0.8293 | ENSAMEG00000000640, ENSAMEG00000002444, ENSAMEG000000010841, ENSAMEG000000014637 | HSPB6, CRTAP, HSPB1, UNC45B |
| GO:0030222 | eosinophil differentiation biological process |                           | 2/269       | 3/13500   | 0.00117 | 0.8293 | ENSAMEG00000015180, ENSAMEG00000016820 | GATA2, IL25 |
| GO:0031532 | actin cytoskeleton reorganization biological process |                           | 6/269       | 60/13500  | 0.0012  | 0.8293 | ENSAMEG00000003627, ENSAMEG00000011750, ENSAMEG00000013803, ENSAMEG00000015226, ENSAMEG00000017670, ENSAMEG00000019659 | EP58, HAX1, FES, MICALL2, PDL1 M4, S1PR2 |
| GO:0030029 | actin filament-based process biological process |                           | 19/269      | 442/13500 | 0.0014  | 0.8293 | ENSAMEG00000000602, ENSAMEG00000002069, ENSAMEG00000003627, ENSAMEG00000005610, ENSAMEG00000006802, ENSAMEG00000007327, ENSAMEG00000011750, ENSAMEG00000012992, ENSAMEG00000013803, ENSAMEG00000015226, ENSAMEG00000015431, ENSAMEG00000017062, ENSAMEG00000017147, ENSAMEG00000017327, ENSAMEG00000017670, ENSAMEG00000019659, ENSAMEG00000019280, ENSAMEG00000019659, ENSAMEG0000002087 | EPB41L3, KCNQ1, EP58, INPPL1, RHQO, TRIP10, HAX1, SNTA1, FES, MICALL2, LIMK1, ARR81, FMR1L2, FHOD1, PDLIM4, CDC42BPB, CDC42EP2, S1PR2, KCNE3 |
| GO:0042133 | neurotransmitter metabolic process biological process |                           | 6/269       | 62/13500  | 0.0014  | 0.8293 | ENSAMEG00000002398, ENSAMEG00000002581, ENSAMEG000000011617, ENSAMEG000000016224, ENSAMEG000000016637, ENSAMEG000000016662 | CHDH, HNMT, COMT, CD36, DAGLA, DPYD |
| GO:0032634 | interleukin-5 production biological process |                           | 3/269       | 12/13500  | 0.00151 | 0.8293 | ENSAMEG00000002415, ENSAMEG00000014490, ENSAMEG00000016820 | IL17RB, ILSRA, IL25 |
| GO:0060307 | regulation of ventricular cardiac muscle cell membrane repolarization biological process |                           | 3/269       | 12/13500  | 0.00151 | 0.8293 | ENSAMEG00000002069, ENSAMEG00000012992, ENSAMEG0000002087 | KCNQ1, SNTA1, KCNE3 |
| GO.ID    | Term                                      | Ontology          | Significant | Annotated | Pvalue  | Qvalue | Signi_id                  | Signi_symbol                  |
|----------|-------------------------------------------|-------------------|-------------|-----------|---------|--------|---------------------------|-----------------------------|
| GO:0097164 | ammonium ion metabolic process            | biological process | 7/269       | 85/13500  | 0.00152 | 0.8293 | ENSAMEG00000001272, ENSAMEG00000002314, ENSAMEG00000002398, ENSAMEG00000002581, ENSAMEG00000003089, ENSAMEG00000007468, ENSAMEG00000011617 | CHKA, ENSAMEG00000002314, CHDH, HNMT, PNKD, SMPDL3A, COMT |
| GO:0002521 | leukocyte differentiation                 | biological process | 14/269      | 281/13500 | 0.00154 | 0.8293 | ENSAMEG00000000272, ENSAMEG00000004441, ENSAMEG00000004774, ENSAMEG00000004811, ENSAMEG00000006928, ENSAMEG00000011750, ENSAMEG00000012594, ENSAMEG00000013931, ENSAMEG00000015180, ENSAMEG00000016820, ENSAMEG00000017184, ENSAMEG00000017745, ENSAMEG00000017990, ENSAMEG00000018841 | CBFA2T3, GAS6, PPARG, POL2F2, DUSP10, HAX1, LYL1, BRAS, GATA2, IL25, APP, BATF2, BAD, CEBPG |
| GO:0010883 | regulation of lipid storage               | biological process | 4/269       | 26/13500  | 0.00163 | 0.8293 | ENSAMEG00000000943, ENSAMEG00000004774, ENSAMEG00000011831, ENSAMEG00000016224 | LPL, PPARG, IKBKE, CD36 |
| GO:2000116 | regulation of cysteine-type endopeptidase activity | biological process | 9/269       | 138/13500 | 0.00177 | 0.8293 | ENSAMEG00000004441, ENSAMEG00000004774, ENSAMEG00000006630, ENSAMEG00000010287, ENSAMEG00000010991, ENSAMEG00000012590, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017990 | GAS6, PPARG, NOD1, KIAA0141, HIP1, GPX1, NLRP3, ARRB1, BAD |
| GO:0007596 | blood coagulation                         | biological process | 8/269       | 114/13500 | 0.002    | 0.8293 | ENSAMEG00000004441, ENSAMEG00000010841, ENSAMEG00000011403, ENSAMEG00000011687, ENSAMEG00000011842, ENSAMEG00000016015, ENSAMEG00000016224, ENSAMEG00000017422 | GAS6, HSPB1, F5, SERPINE2, APOH, ACTG1, CD36, F13A1 |
| GO:0006952 | defense response                          | biological process | 24/269      | 638/13500 | 0.00213 | 0.8293 | ENSAMEG00000011502, ENSAMEG0000002415, ENSAMEG00000035641, ENSAMEG0000004196, ENSAMEG0000004663, ENSAMEG0000004774, ENSAMEG0000006630, ENSAMEG0000007284, ENSAMEG0000008499, ENSAMEG0000008575, ENSAMEG0000011225, ENSAMEG0000011820, ENSAMEG0000011831, ENSAMEG0000012283, ENSAMEG0000012590, ENSAMEG0000013869, ENSAMEG0000014335, ENSAMEG0000014490, ENSAMEG0000016224, ENSAMEG0000016820, ENSAMEG0000016947, ENSAMEG0000017745, ENSAMEG0000018841, ENSAMEG0000019893 | ECM1, IL17RB, CARD9, SLPL, TRIM14, PPARG, NOD1, CYBB, ZFP36, STAB1, TMEM173, LYZ, IKBKE, SEHAF4, GPX1, A2M, ENSAMEG0000014335, IL5RA, CD36, IL25, NLRP3, BATF2, CEBPG, CYSLTR1 |
| GO:0039611 | response to wounding                      | biological process | 13/269      | 260/13500 | 0.00218 | 0.8293 | ENSAMEG00000004441, ENSAMEG00000006577, ENSAMEG0000008499 | GAS6, NINJ1, ZFP36, ENSAMEG00000010280, HSPB1, F5, SERPINE2, APOH, GPX1 |
| GO.ID | Term                              | Ontology                  | Significant | Annotated | Pvalue  | Qvalue | Signi_id                                      | Signi_symbol                  |
|-------|-----------------------------------|---------------------------|-------------|-----------|---------|--------|----------------------------------------------|--------------------------------|
| GO: 0007599 | hemostasis                        | biological process        | 8/269       | 0.00223   | 0.8293 | ENSAMEG00000010280, ENSAMEG00000010841, ENSAMEG00000011403, ENSAMEG00000011687, ENSAMEG00000011842, ENSAMEG00000012590, ENSAMEG00000012708, ENSAMEG00000016015, ENSAMEG00000016224, ENSAMEG00000017422 | LAMB2, ACTG1, CD36, F13A1 |
| GO: 0042089 | cytokine biosynthetic process     | biological process        | 6/269       | 0.0023    | 0.8293 | ENSAMEG00000003641, ENSAMEG00000006630, ENSAMEG00000008409, ENSAMEG00000010841, ENSAMEG00000013744, ENSAMEG00000018841 | GAS6, HSPB1, F5, SERPINE2, APOH, ACTG1, CD36, F13A1 |
| GO: 0010757 | negative regulation of plasminogen activation | biological process        | 2/269       | 0.00231   | 0.8293 | ENSAMEG00000011687, ENSAMEG00000012483 | SERPINE2, CTSZ |
| GO: 0010886 | positive regulation of cholesterol storage | biological process        | 2/269       | 0.00231   | 0.8293 | ENSAMEG00000000943, ENSAMEG00000016224 | LPL, CD36 |
| GO: 0106049 | regulation of cellular response to osmotic stress | biological process        | 2/269       | 0.00231   | 0.8293 | ENSAMEG00000009393, ENSAMEG00000017990 | YBX3, BAD |
| GO: 1902218 | regulation of intrinsic apoptotic signaling pathway in response to osmotic stress | biological process        | 2/269       | 0.00231   | 0.8293 | ENSAMEG00000009393, ENSAMEG00000017990 | YBX3, BAD |
| GO: 0050817 | coagulation                       | biological process        | 8/269       | 0.00235   | 0.8293 | ENSAMEG00000004441, ENSAMEG00000010841, ENSAMEG00000011403, ENSAMEG00000011687, ENSAMEG00000011842, ENSAMEG00000016015, ENSAMEG00000016224, ENSAMEG00000017422 | GAS6, HSPB1, F5, SERPINE2, APOH, ACTG1, CD36, F13A1 |
| GO: 0009141 | nucleoside triphosphate metabolic process | biological process        | 10/269      | 0.00236   | 0.8293 | ENSAMEG00000000272, ENSAMEG00000003519, ENSAMEG00000006082, ENSAMEG00000007468, ENSAMEG00000011494, ENSAMEG00000017990, ENSAMEG00000023430, ENSAMEG00000023436, ENSAMEG00000023442, ENSAMEG00000023455 | CBFA2T3, PFKFB2, RHOQ, SMPD |
| GO: 0099623 | regulation of cardiac muscle cell membrane repolarization | biological process        | 3/269       | 0.00242   | 0.8293 | ENSAMEG00000012992, ENSAMEG00000020287 | KCNQ1, SNTA1, KCNE3 |
| GO: 0099625 | ventricular cardiac muscle cell membrane repolarization | biological process        | 3/269       | 0.00242   | 0.8293 | ENSAMEG00000012992, ENSAMEG00000020287 | KCNQ1, SNTA1, KCNE3 |
| GO:0055086 | nucleobase-containing small molecule metabolic process | biological process | 18/269 | 431/13500 | 0.00254 | 0.8487 | ENSAMEG00000000272, ENSAMEG00000003519, ENSAMEG00000006082, ENSAMEG00000007468, ENSAMEG00000007497, ENSAMEG00000008034, ENSAMEG00000009105, ENSAMEG00000010857, ENSAMEG00000011455, ENSAMEG00000011494, ENSAMEG00000014118, ENSAMEG00000015964, ENSAMEG00000016662, ENSAMEG00000017990, ENSAMEG00000023430, ENSAMEG00000023436, ENSAMEG00000023442, ENSAMEG00000023455, CBFA2T3, PFKFB2, RHOQ, SMPD3, NADSYN1, ENSAMEG00000011418, ADCCY4, DPYD, BAD, ND2, COX1, ATP6, CYTB |
| GO:0030054 | cell junction | cellular component | 20/287 | 640/14310 | 0.03342 | 0.8682 | ENSAMEG00000000602, ENSAMEG00000003894, ENSAMEG00000006453, ENSAMEG00000006768, ENSAMEG00000008034, ENSAMEG00000010536, ENSAMEG00000010841, ENSAMEG00000013263, ENSAMEG00000013297, ENSAMEG00000013803, ENSAMEG00000013931, ENSAMEG00000014689, ENSAMEG00000015326, ENSAMEG00000015431, ENSAMEG00000015837, ENSAMEG00000016015, ENSAMEG00000016087, ENSAMEG00000016396, ENSAMEG00000017184, ENSAMEG00000017327, ENSAMEG00000019485, EPB41L3, MRCC2, ARHGAP21, CYTH3, VIM, HSPB1, REXO2, FHL2, FES, RRAS, TNS3, MICALL2, LIMK1, STXB5, ACTG1, PDLIM1, DLG4, APP, RHOD1, CDC42EP4 |
| GO:0042995 | cell projection | cellular component | 24/287 | 808/14310 | 0.03526 | 0.8682 | ENSAMEG00000003086, ENSAMEG00000003627, ENSAMEG00000003733, ENSAMEG00000004387, ENSAMEG00000006768, ENSAMEG00000007278, ENSAMEG00000008620, ENSAMEG00000010536, ENSAMEG00000010841, ENSAMEG00000011509, ENSAMEG00000011750, ENSAMEG00000012095, ENSAMEG00000014724, ENSAMEG00000015252, ENSAMEG00000015431, ENSAMEG00000015964, ENSAMEG00000016224, ENSAMEG00000016396, ENSAMEG00000016637, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017670, ENSAMEG00000018345, ENSAMEG00000020287, NRP2, EPS8, INPP5E, PLXND1, CYTH3, ARHGAP31, CYBRD1, VIM, HSPB1, SEPT5, HAX1, NAVI1, ITSN1, RILPL1, LIMK1, ADCY4, CD36, DLG4, DAGLA, ARRB1, APP, PDLIM4, TTL5, KCNE3 |
| GO:0015629 | actin cytoskeleton | cellular component | 11/287 | 294/14310 | 0.03531 | 0.8682 | ENSAMEG000000005486, ENSAMEG00000006082, ENSAMEG00000006453, ENSAMEG00000011202, ENSAMEG00000011750, ENSAMEG00000015326, ENSAMEG00000016015, ENSAMEG00000017327, ENSAMEG00000017670, SPPL2B, RHOQ, ARHGAP21, ZNF74, HAX1, MICALL2, ACTG1, RHOD1, PDLIM4, CDC42BPB, CDC42EP4 |
| GO:ID   | Term                          | Ontology                  | Significant| Annotated | Pvalue  | Qvalue  | Signi_id          | Signi_symbol   |
|---------|-------------------------------|---------------------------|------------|-----------|---------|---------|-------------------|----------------|
| GO:0070469 | respiratory chain cellular component | 4/287                     | 63/14310   | 0.03751   | 0.8682  |         | ENSAMEG00000018199, ENSAMEG00000019485 | ND1, ND2, COX1, CYTB |
| GO:0001725 | stress fiber cellular component | 3/287                     | 37/14310   | 0.03757   | 0.8682  |         | ENSAMEG00000023426, ENSAMEG00000023430, ENSAMEG00000023436, ENSAMEG00000023455 | NDI2, FHOD1, PDIM4 |
| GO:0097517 | contractile actin filament bundle cellular component | 3/287                     | 37/14310   | 0.03757   | 0.8682  |         | ENSAMEG00000023426, ENSAMEG00000023430, ENSAMEG00000023436, ENSAMEG00000023442, ENSAMEG00000023455 | NDI2, FHOD1, PDIM4 |
| GO:0044429 | mitochondrial part cellular component | 16/287                   | 491/14310  | 0.03918   | 0.8682  |         | ENSAMEG00000023426, ENSAMEG00000023430, ENSAMEG00000023436, ENSAMEG00000023442, ENSAMEG00000023455 | HADH, ENSAMEG00000002314, CHDH, RAB32, GLR05, RNF1448, TME173, HAX1, IQK6E, EFHD1, BAD, ND1, ND2, COX1, ATPI6, CYTB |
| GO:0005737 | cytoplasm cellular component | 140/287                   | 6226/14310 | 0.03957   | 0.8682  |         | ENSAMEG00000000640, ENSAMEG00000012330, ENSAMEG0000001276, ENSAMEG00000017894, ENSAMEG0000001814, ENSAMEG0000002069, ENSAMEG0000002314, ENSAMEG00000023998, ENSAMEG00000025315, ENSAMEG0000002581, ENSAMEG00000027332, ENSAMEG00000030200, ENSAMEG0000003089, ENSAMEG0000003442, ENSAMEG00000036227, ENSAMEG00000036481, ENSAMEG00000037333, ENSAMEG0000004195, ENSAMEG00000043448, ENSAMEG00000044441, ENSAMEG00000045116, ENSAMEG00000045600, ENSAMEG00000046800, ENSAMEG00000047744, ENSAMEG00000055113, ENSAMEG00000055458, ENSAMEG0000005566, ENSAMEG0000005610, ENSAMEG00000056798, ENSAMEG00000058747, ENSAMEG00000059227, ENSAMEG00000060014, ENSAMEG00000060893, ENSAMEG00000064553, ENSAMEG00000064540, ENSAMEG00000066300, ENSAMEG00000068768, ENSAMEG00000069282, ENSAMEG0000007046, | HSPB6, TCIRG1, ZNF1, TCT7B, HADH, KCNQ1, ENSAMEG00000002314, CHDH, IL17B8, CRATP, ANPEP, HNMT, PLOD3, HACD4, PNKD, CRYZ, SBF2, EPS8, CARD9, INPP5E, SLPI, ENSAMEG00000004348, GASS6, CD1D, NANS, PPARG, SCDS, SPPL28, ADAP2, INPPL1, JAZ2F1, TCC2, ATP10A, RAB32, AP2P2, ARHGA21, GLR5X, NOD1, CYTH3, DUSP10, TME173, NAD5, YN1, CLCN7, RNFL448B, HSCB, STEAP3, ZF936, DGK4, AIKCY, DMYXL2, ZERB, XXTL17, YBX3, CPT2, S0C52, ORMDL3, TRIM47, SLA2, GNPDA1, KIAA0141, VIM, HSPB1, HIF1, PGEP1, TME173, F5, PDE4A, FBP1, SEPT5, CCDC50, COMT, SERPINE2, HAX1, ENSAMEG000000011792, IKBKE, ANGPT16, SDHAF4, ERCC1, RNF7, BLVRB, CTS2, DNAJ4A4, GPX1, PORCN, Clorh58, ENSAMEG00000013098, ACS53, RX02, SNA11, FES, NAAA, UNC45B, PPP1R15A, TNS3, ITSN1, SERPINB10, GAA, SGSH, RLP1L1, MICALL2, CLP2, LIMK1, NUAK1, HIF, ADFOCY, ACTG1, CD36, TRIM3, DLG4, EFHD1, DPDY, TNFSF12, TNSFSF13, CD68, TME175, NLRP3, ARRB1, RN1, APP, FHOD1, ENSAMEG00000017441, PDIM4, CASB, SLC29A3, ABCA3, PTO1G1, BAD, TNFAIP2, TITL5, CDC42E2, CDC42E4, B3GNT9, OST4, B3GALT2, KNE3, ND1, ND2, COX1, ATP6, CYTB |
Table 3 Functional enrichment analysis based on the representation of GO terms (Continued)

| GO.ID | Term | Ontology | Significant | Annotated | Pvalue | Qvalue | Signi_id | Signi_symbol |
|-------|------|----------|-------------|-----------|--------|--------|----------|--------------|
| ENSAMEG00000007997, ENSAMEG00000007997, ENSAMEG00000008424, ENSAMEG00000008477, ENSAMEG00000008497, ENSAMEG00000008949, ENSAMEG00000008800, ENSAMEG00000009105, ENSAMEG00000009193, ENSAMEG00000009337, ENSAMEG00000009367, ENSAMEG00000009393, ENSAMEG00000009555, ENSAMEG00000009824, ENSAMEG00000009860, ENSAMEG00000009928, ENSAMEG00000010113, ENSAMEG00000010219, ENSAMEG00000010287, ENSAMEG00000010536, ENSAMEG00000010841, ENSAMEG00000010991, ENSAMEG00000011170, ENSAMEG00000011225, ENSAMEG00000011403, ENSAMEG00000011455, ENSAMEG00000011494, ENSAMEG00000011509, ENSAMEG00000011582, ENSAMEG00000011617, ENSAMEG00000011687, ENSAMEG00000011750, ENSAMEG00000011792, ENSAMEG00000011831, ENSAMEG00000012016, ENSAMEG00000012283, ENSAMEG00000012418, ENSAMEG00000012425, ENSAMEG00000012457, ENSAMEG00000012463, ENSAMEG00000012576, ENSAMEG00000012590, ENSAMEG00000012689, ENSAMEG00000012804, ENSAMEG00000013098, ENSAMEG00000013162, ENSAMEG00000013263, ENSAMEG00000013478, ENSAMEG00000013803, ENSAMEG00000014605, ENSAMEG00000014637, ENSAMEG00000014678, ENSAMEG00000014689, ENSAMEG00000014724, ENSAMEG00000015156, ENSAMEG00000015157, ENSAMEG00000015221, ENSAMEG00000015252, ENSAMEG0000001526, ENSAMEG00000015346, ENSAMEG00000015431, ENSAMEG00000015604, ENSAMEG00000015938, ENSAMEG00000015964, ENSAMEG00000016015, ENSAMEG00000016224, ENSAMEG00000016339, ENSAMEG00000016396, ENSAMEG00000016446, ENSAMEG00000016662, ENSAMEG00000016729, ENSAMEG00000016736, |
| GO:ID  | Term                              | Ontology                  | Significant | Annotated | Pvalue  | Qvalue  | Signi_id     | Signi_symbol |
|--------|-----------------------------------|---------------------------|-------------|-----------|---------|---------|--------------|--------------|
| 0000110| nucleotide-excision repair factor 1 complex | cellular component        | 1/287       | 2/14310   | 0.03971 | 0.8682  | ENSAMEG00000012418 | ERCC1        |
| 0031232| extrinsic component of external side of plasma membrane | cellular component        | 1/287       | 2/14310   | 0.03971 | 0.8682  | ENSAMEG00000011687 | SERPINE2     |
| 0035339| SPOTS complex                     | cellular component        | 1/287       | 2/14310   | 0.03971 | 0.8682  | ENSAMEG00000009860 | ORMDL3       |
| 0043083| synaptic cleft                    | cellular component        | 1/287       | 2/14310   | 0.03971 | 0.8682  | ENSAMEG00000012708 | LAMB2        |
| 0072559| NLRP3 inflammasome complex        | cellular component        | 1/287       | 2/14310   | 0.03971 | 0.8682  | ENSAMEG00000016947 | NLRP3        |
| 0042107| cytokine metabolic process        | biological process        | 6/269       | 70/13500  | 0.00267 | 0.8704  | ENSAMEG00000003641 | CARD9, NOD1, ZFP36, HSPB1, MAST2, CEBPG |
| 0043393| regulation of protein binding     | biological process        | 9/269       | 148/13500 | 0.00286 | 0.882   | ENSAMEG00000004196 | SLPI, PLXND1, TMEM173, CTSZ, HFE, CD36, CAMK1, ARRB1, APP |
| 0032640| tumor necrosis factor production  | biological process        | 6/269       | 71/13500  | 0.00287 | 0.882   | ENSAMEG00000003641 | CARD9, GAS6, NOD1, HSPB1, CD36, APP |
| 0010884| positive regulation of lipid storage | biological process        | 3/269       | 15/13500  | 0.00298 | 0.882   | ENSAMEG00000011831 | LPL, IKBKE, CD36 |
### Table 3 Functional enrichment analysis based on the representation of GO terms (Continued)

| GO.ID   | Term                                      | Ontology                   | Significant | Annotated | Pvalue  | Qvalue  | Signi_id              | Signi_symbol                        |
|---------|-------------------------------------------|----------------------------|-------------|-----------|---------|---------|-----------------------|-------------------------------------|
| GO:0051239 | regulation of multicellular organismal process | biological process        | 47/269      | 158/1350  | 0.00314 | 0.882  | ENSAMEG00000000528, ENSAMEG00000000640, ENSAMEG00000000943, ENSAMEG00000001502, ENSAMEG0000002069, ENSAMEG00000002415, ENSAMEG00000003641, ENSAMEG00000004387, ENSAMEG00000004441, ENSAMEG00000004774, ENSAMEG00000006630, ENSAMEG00000006928, ENSAMEG00000008416, ENSAMEG00000008499, ENSAMEG00000008575, ENSAMEG00000009824, ENSAMEG0000010280, ENSAMEG0000010536, ENSAMEG0000010841, ENSAMEG0000011225, ENSAMEG0000011687, ENSAMEG0000011750, ENSAMEG0000011842, ENSAMEG0000012590, ENSAMEG0000012943, ENSAMEG0000012992, ENSAMEG0000013478, ENSAMEG0000013744, ENSAMEG0000013803, ENSAMEG0000013931, ENSAMEG0000014318, ENSAMEG0000014490, ENSAMEG0000015157, ENSAMEG0000015180, ENSAMEG0000015431, ENSAMEG0000015938, ENSAMEG0000016224, ENSAMEG0000016268, ENSAMEG0000016625, ENSAMEG0000016736, ENSAMEG0000016947, ENSAMEG0000017062, ENSAMEG0000017184, ENSAMEG0000017990, ENSAMEG0000018841, ENSAMEG0000019059, ENSAMEG0000020287 | NCKIPSD, HSPB6, LPL, ECM1, KCNQ1, IL17RB, CARD9, PLXND1, GAS6, PPARG, NOD1, DUSP10, KREMEN1, ZFP36, STAB1, SOCS2, ENSAMEG00000010280, VIM, HSPB1, TMEM173, SERPINE2, HAX1, APOH, GPX1, TIMP1, SNTA1, SNAI1, MAST2, FES, RRAS, ACVR1, ILSRA, GAA, GATA2, LINIK1, HFE, CD36, CAMK1, PBX3, TNFSF13, NRF2, ARRB1, APP, BAD, CEBPB, S1PR2, KCNE3 |
| GO:0007585 | respiratory gaseous exchange | biological process        | 4/269       | 31/1350   | 0.00318 | 0.882  | ENSAMEG00000004600, ENSAMEG00000010280, ENSAMEG00000015157, ENSAMEG00000016625 | NDS1, ENSAMEG00000010280, GAA, PBX3 |
| GO:0050830 | defense response to Gram-positive bacterium | biological process        | 4/269       | 31/1350   | 0.00318 | 0.882  | ENSAMEG00000003641, ENSAMEG00000009630, ENSAMEG00000011820, ENSAMEG00000016224 | CARD9, NOD1, LYZ, CD36 |
| GO:0071706 | tumor necrosis factor superfamily cytokine production | biological process        | 6/269       | 74/1350   | 0.00353 | 0.882  | ENSAMEG00000003641, ENSAMEG00000009441, ENSAMEG00000009630, ENSAMEG00000010841, ENSAMEG00000016224, ENSAMEG00000017184 | CARD9, GAS6, NOD1, HSPB1, CD36, APP |
| GO:0009205 | purine ribonucleoside triphosphate metabolic process | biological process        | 9/269       | 153/1350  | 0.00357 | 0.882  | ENSAMEG00000002072, ENSAMEG00000003519, ENSAMEG00000006302, ENSAMEG00000011494, ENSAMEG00000017990, ENSAMEG00000023430, ENSAMEG00000023436, ENSAMEG00000023442, ENSAMEG00000023455 | CFB24T3, PFKFB2, RHOD, FBP1, BAD, ND2, COX1, ATP6, CYTB |
| GO:0031638 zymogen activation | biological process | 4/269 | 32/13500 | 0.00357 | 0.882 | ENSAMEG00000011687, ENSAMEG00000011842, ENSAMEG00000012483, ENSAMEG00000017990 | SERPINE2, APOH, CTSZ, BAD | LPL, IL17R, CARD9, GA56, NOD1, ZFP36, HSPB1, TIMM173, MAST2, IL5RA, HFE, CD36, IL25, NLRP3, ARRB1, APP, CEBPG |
| GO:0001816 cytokine production | biological process | 17/269 | 411/13500 | 0.00366 | 0.882 | ENSAMEG00000000943, ENSAMEG00000002415, ENSAMEG00000003641, ENSAMEG00000004441, ENSAMEG00000006630, ENSAMEG00000008499, ENSAMEG00000010841, ENSAMEG00000011225, ENSAMEG00000013764, ENSAMEG00000014400, ENSAMEG00000015938, ENSAMEG00000016224, ENSAMEG00000016820, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000018841 | LPL, IL17R, CARD9, GA56, NOD1, ZFP36, HSPB1, TIMM173, MAST2, IL5RA, HFE, CD36, IL25, NLRP3, ARRB1, APP, CEBPG |
| GO:0035754 B cell chemotaxis | biological process | 2/269 | 5/13500 | 0.0038 | 0.882 | ENSAMEG00000004441, ENSAMEG00000007512 | GAS6, HSD3B7 |
| GO:0090240 positive regulation of histone H4 acetylation | biological process | 2/269 | 5/13500 | 0.0038 | 0.882 | ENSAMEG00000000640, ENSAMEG00000001502, ENSAMEG00000004387, ENSAMEG00000004774, ENSAMEG00000008575, ENSAMEG00000010841, ENSAMEG00000011842, ENSAMEG00000013931, ENSAMEG00000015180 | HSPB6, ECM1, PLXND1, PPARG, STAB1, HSPB1, APOH, RRAS, GATA2 |
| GO:0045765 regulation of angiogenesis | biological process | 9/269 | 155/13500 | 0.00389 | 0.882 | ENSAMEG00000000640, ENSAMEG00000001502, ENSAMEG00000002069, ENSAMEG00000002415, ENSAMEG00000003641, ENSAMEG00000004441, ENSAMEG00000006630, ENSAMEG00000006928, ENSAMEG00000009824, ENSAMEG00000010841, ENSAMEG00000011225, ENSAMEG00000011687, ENSAMEG00000011750, ENSAMEG00000011842, ENSAMEG00000013478, ENSAMEG00000013803, ENSAMEG00000013931, ENSAMEG00000014318, ENSAMEG00000015180, ENSAMEG00000015431, ENSAMEG00000016224, ENSAMEG00000016268, ENSAMEG00000016736, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017990, ENSAMEG00000018841 | NCKIPSD, HSPB6, LPL, ECM1, KCNQ1, IL17RB, CARD9, GA56, NOD1, DUJSP10, SOC2, HSPB1, TIMM173, SERPINE2, HAX1, APOH, SNAI1, FES, RRAS, ACVR1, GATA2, LIMK1, CD36, CAMK1, TNFSF13, NLRP3, APP, BAD, CEBPG |
| GO:0051240 positive regulation of multicellular organismal process | biological process | 29/269 | 863/13500 | 0.004 | 0.882 | ENSAMEG00000000528, ENSAMEG00000000540, ENSAMEG00000002640, ENSAMEG00000002943, ENSAMEG000000051502, ENSAMEG00000005269, ENSAMEG00000005415, ENSAMEG00000005641, ENSAMEG00000005641, ENSAMEG00000007577, ENSAMEG00000010841, ENSAMEG00000011225, ENSAMEG00000011687, ENSAMEG00000011750, ENSAMEG00000011842, ENSAMEG00000013478, ENSAMEG00000013803, ENSAMEG00000013931, ENSAMEG00000014318, ENSAMEG00000015180, ENSAMEG00000015431, ENSAMEG00000016224, ENSAMEG00000016268, ENSAMEG00000016736, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017990, ENSAMEG00000018841 | ENSAMEG00000000528, ENSAMEG00000000540, ENSAMEG00000002640, ENSAMEG00000002943, ENSAMEG000000051502, ENSAMEG00000005269, ENSAMEG00000005415, ENSAMEG00000005641, ENSAMEG00000005641, ENSAMEG00000007577, ENSAMEG00000010841, ENSAMEG00000011225, ENSAMEG00000011687, ENSAMEG00000011750, ENSAMEG00000011842, ENSAMEG00000013478, ENSAMEG00000013803, ENSAMEG00000013931, ENSAMEG00000014318, ENSAMEG00000015180, ENSAMEG00000015431, ENSAMEG00000016224, ENSAMEG00000016268, ENSAMEG00000016736, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000017990, ENSAMEG00000018841 | NCKIPSD, HSPB6, LPL, ECM1, KCNQ1, IL17RB, CARD9, GA56, NOD1, DUJSP10, SOC2, HSPB1, TIMM173, SERPINE2, HAX1, APOH, SNAI1, FES, RRAS, ACVR1, GATA2, LIMK1, CD36, CAMK1, TNFSF13, NLRP3, APP, BAD, CEBPG |
| GO:0009199 ribonucleoside triphosphate metabolic process | biological process | 9/269 | 157/13500 | 0.00423 | 0.882 | ENSAMEG0000000011783, ENSAMEG00000000272, ENSAMEG000000003519, ENSAMEG000000006082, ENSAMEG000000011494, ENSAMEG000000017990, ENSAMEG000000018841 | CBFA2T3, PFKFB2, RHOQ, FBPI, BAD, ND2, COX1, ATP6, CYTB |
| GO.ID   | Term                                      | Ontology                  | Significant | Annotated | Pvalue  | Qvalue | Signi_id                  | Signi_symbol                        |
|---------|-------------------------------------------|---------------------------|-------------|-----------|---------|--------|---------------------------|-------------------------------------|
| GO: 0042060 | wound healing biological process          |                           | 11/269      | 218/13500 | 0.00436 | 0.882  | ENSAMEG000000023430,     | GAS6, NINJ1, ENSAMEG000000010280, HSPB1, F5, SERPINE2, APOH, GPX1, ACTG1, CD36, F15A1 |
| GO: 0009144 | purine nucleoside triphosphate metabolic process |                           | 9/269       | 158/13500 | 0.00441 | 0.882  | ENSAMEG00000002722,     | CBF2A3T3, PFKFB2, RHOQ, FBP1, BAD, ND2, COX1, ATP6, CYTB |
| GO: 0006753 | nucleoside phosphate metabolic process     |                           | 16/269      | 388/13500 | 0.0049  | 0.882  | ENSAMEG00000002722,     | CBF2A3T3, PFKFB2, RHOQ, SMPD L3A, NADSYN1, ENSAMEG00000008034, ADCY19, PDE4A, FBP1, ENSAMEG00000014118, ADCY4, BAD, ND2, COX1, ATP6, CYTB |
| GO: 0032612 | interleukin-1 production biological process |                           | 4/269       | 35/13500  | 0.00497 | 0.882  | ENSAMEG00000004441,     | GAS6, HSPB1, CD36, NLRP3 |
| GO: 0051346 | negative regulation of hydrolase activity biological process |                           | 12/269      | 254/13500 | 0.00501 | 0.882  | ENSAMEG00000001502,     | ECM1, SLPI, GAS6, TIMP4, ENSAMEG00000010076, SERPINE2, GPX1, TIMP1, A2M, PPP1R5A, ARRB1, APP |
| GO: 0010955 | negative regulation of protein processing biological process |                           | 3/269       | 18/13500  | 0.00511 | 0.882  | ENSAMEG000000011687,     | SERPINE2, CTSZ, A2M |
| GO: 0014002 | astrocyte development biological process   |                           | 3/269       | 18/13500  | 0.00511 | 0.882  | ENSAMEG00000001503,     | VIM, LAMB2, APP |
| GO: 0099622 | cardiac muscle cell membrane repolarization biological process |                           | 3/269       | 18/13500  | 0.00511 | 0.882  | ENSAMEG00000002069,     | KCNQ1, SNTA1, KCNE3 |
Table 3 Functional enrichment analysis based on the representation of GO terms (Continued)

| GO.ID | Term                                                                 | Ontology                                    | Significant | Annotated | Pvalue  | Qvalue  | Signi_id          | Signi_symbol |
|-------|----------------------------------------------------------------------|---------------------------------------------|-------------|-----------|---------|---------|--------------------|--------------|
| GO:1903318 | negative regulation of protein maturation                           | biological process                          | 3/269       | 1/3/1500  | 0.00511 | 0.882   | ENSAMEG00000011687, ENSAMEG00000012483, ENSAMEG00000013869 | SERPINE2, CTSZ, A2M |
| GO:0008627 | intrinsic apoptotic signaling pathway in response to osmotic stress | biological process                          | 2/269       | 6/3/1500  | 0.00563 | 0.882   | ENSAMEG00000009393, ENSAMEG00000017990 | YBX3, BAD |
| GO:0015858 | nucleoside transport                                                 | biological process                          | 2/269       | 6/3/1500  | 0.00563 | 0.882   | ENSAMEG00000009393, ENSAMEG00000017990 | YBX3, BAD |
| GO:0019359 | nicotinamide nucleotide biosynthetic process                         | biological process                          | 5/269       | 58/3/1500 | 0.00587 | 0.882   | ENSAMEG00000000272, ENSAMEG00000003519, ENSAMEG00000007497, ENSAMEG0000000834, ENSAMEG00000011494 | CBFA2T3, PFKFB2, NADSYN1, ENSAMEG00000008034, FBPI |
| GO:0019363 | pyridine nucleotide biosynthetic process                             | biological process                          | 5/269       | 58/3/1500 | 0.00587 | 0.882   | ENSAMEG00000000272, ENSAMEG00000003519, ENSAMEG00000007497, ENSAMEG0000000834, ENSAMEG00000011494 | CBFA2T3, PFKFB2, NADSYN1, ENSAMEG00000008034, FBPI |
| GO:0046034 | ATP metabolic process                                                 | biological process                          | 8/269       | 136/3/1500| 0.00588 | 0.882   | ENSAMEG00000000272, ENSAMEG00000003519, ENSAMEG00000007497, ENSAMEG0000000834, ENSAMEG00000011494 | CBFA2T3, PFKFB2, FBPI, BAD, ND2, COX1, ATP6, CYTB |
| GO:0030036 | actin cytoskeleton organization                                       | biological process                          | 16/269      | 396/3/1500| 0.00593 | 0.882   | ENSAMEG00000000602, ENSAMEG00000003627, ENSAMEG00000005610, ENSAMEG00000006082, ENSAMEG00000007327, ENSAMEG00000007803, ENSAMEG00000011526, ENSAMEG00000014531, ENSAMEG00000017062, ENSAMEG00000017147, ENSAMEG00000017327, ENSAMEG00000017670, ENSAMEG00000018199, ENSAMEG00000019280, ENSAMEG00000019659 | EPB4L13, EPS8, INPPL1, RHOQ, TRIP10, HAX1, FES, MKC2L2, LIN1K1, ARRB1, FMNL2, PHD1, PDUM4, CDC42BPB, CDC42EP2, S1PR2 |
| GO:0031639 | plasminogen activation                                               | biological process                          | 3/269       | 19/3/1500 | 0.00599 | 0.882   | ENSAMEG000000011687, ENSAMEG000000011842, ENSAMEG000000012483 | SERPINE2, APOH, CTSZ |

You et al. BMC Genomic Data (2021) 22:44 Page 45 of 56
| GO:        | Term                                      | Ontology          | Significant | Annotated | Pvalue       | Qvalue       | Signi_id                        | Signi_symbol                    |
|------------|-------------------------------------------|-------------------|-------------|-----------|--------------|--------------|---------------------------------|---------------------------------|
| 0086005    | ventricular cardiac muscle cell action potential | biological process | 3/269       | 19/13500  | 0.00599     | 0.882        | ENSAMEG00000002069, ENSAMEG00000012902, ENSAMEG00000020287 | KCNQ1, SNTA1, KCNE3                |
| 0072521    | purine-containing compound metabolic process | biological process | 14/269      | 327/13500 | 0.00601     | 0.882        | ENSAMEG00000000272, ENSAMEG00000003519, ENSAMEG00000006082, ENSAMEG00000009105, ENSAMEG00000010857, ENSAMEG00000011455, ENSAMEG00000011494, ENSAMEG00000015964, ENSAMEG00000016662, ENSAMEG00000017990, ENSAMEG00000023430, ENSAMEG00000023436, ENSAMEG00000023442, ENSAMEG00000023455 | CBFA2T3, PPPFB2, RHOQ, AHCY, ADCY9, PDE4A, FBPI, ADCY4, DPDY, BAD, ND2, COX1, ATP6, CYTB |
| 0044092    | negative regulation of molecular function | biological process | 23/269      | 655/13500 | 0.00605     | 0.882        | ENSAMEG00000001502, ENSAMEG00000004196, ENSAMEG00000004441, ENSAMEG00000004836, ENSAMEG00000005412, ENSAMEG00000006928, ENSAMEG00000008499, ENSAMEG00000010076, ENSAMEG00000010841, ENSAMEG00000011684, ENSAMEG00000012483, ENSAMEG00000012590, ENSAMEG00000012943, ENSAMEG00000013869, ENSAMEG00000014678, ENSAMEG00000015431, ENSAMEG00000015938, ENSAMEG00000016268, ENSAMEG00000016947, ENSAMEG00000017062, ENSAMEG00000017184, ENSAMEG00000018841, ENSAMEG00000020287 | ECM1, SLPI, GAS6, TIMP4, ADORA3, DUSP10, ZFP36, ENSAMEG00000010076, HSPB1, SERPINE2, CTSZ, GPX1, TIMP1, A2M, PPP1R15A, LIMK1, HFE, CAK1, NLRP3, ARRB1, APP, CEBPG, KCNE3 |
| 0045597    | positive regulation of cell differentiation | biological process | 18/269      | 469/13500 | 0.00612     | 0.882        | ENSAMEG00000000528, ENSAMEG00000006630, ENSAMEG00000010841, ENSAMEG00000016224 | NCKIPSD, GAS6, DUSP10, ZFP36, SOCS2, SERPINE2, HAX1, SNAI1, FES, ACVR1, GATA2, LIMK1, CD36, CAK1, APP, HSF4, BAD, S1PR2 |
| 1901342    | regulation of vasculature development      | biological process | 9/269       | 167/13500 | 0.0063      | 0.882        | ENSAMEG00000000640, ENSAMEG0000001502, ENSAMEG0000004387, ENSAMEG0000004774, ENSAMEG0000008575, ENSAMEG0000010841, ENSAMEG0000011842, ENSAMEG0000012184, ENSAMEG0000012742, ENSAMEG0000017990, ENSAMEG0000019659 | HSPB6, ECM1, PLXND1, PPARG, STAB1, HSPB1, APOH, RRAS, GATA2 |
Table 3  Functional enrichment analysis based on the representation of GO terms (Continued)

| GO ID | Term                                           | Ontology                        | Significant | Annotated | Pvalue  | Qvalue  | Signi_id                     | Signi_symbol                |
|-------|-----------------------------------------------|---------------------------------|-------------|-----------|---------|---------|------------------------------|-----------------------------|
| GO:  | regulation of cytokine biosynthetic process   | biological process              | 5/269       | 59/13500  | 0.00631 | 0.882   | ENSAMEG000000013931, ENSAMEG000000015180 | CARD9, ZFP36, HSPB1, MAST2, CEBPG |
| GO:  | positive regulation of tumor necrosis factor superfamily cytokine production | biological process              | 4/269       | 38/13500  | 0.00668 | 0.882   | ENSAMEG000000013931, ENSAMEG000000015180 | CARD9, NOD1, HSPB1, CD36 |
| GO:  | aromatic compound catabolic process           | biological process              | 12/269      | 264/13500 | 0.00676 | 0.882   | ENSAMEG000000013931, ENSAMEG000000015180 | CBFA2T3, HNMT, PFKFB2, LSM7, SMPL3A, ZFP36, AHCY, PD4E4A, FBP1, COMT, BLVRB, DPYD |
| GO:  | respiratory system process                    | biological process              | 3/269       | 20/13500  | 0.0694  | 0.882   | ENSAMEG000000013931, ENSAMEG000000015180 | ENSAMEG000000010280, GAA, PBX3 |
| GO:  | tissue regeneration                           | biological process              | 3/269       | 20/13500  | 0.0694  | 0.882   | ENSAMEG000000013931, ENSAMEG000000015180 | NIN1, ENSAMEG000000010280, GPX1 |
| GO:  | regulation of membrane repolarization         | biological process              | 3/269       | 20/13500  | 0.0694  | 0.882   | ENSAMEG000000013931, ENSAMEG000000015180 | KCNQ1, SNTA1, KCNE3 |
| GO:  | pyridine-containing compound biosynthetic process | biological process              | 5/269       | 61/13500  | 0.0726  | 0.882   | ENSAMEG000000013931, ENSAMEG000000015180 | CBFA2T3, PFKFB2, NADSYN1, ENSAMEG00000008034, FBP1 |
| GO:  | organophosphate metabolic process             | biological process              | 22/269      | 627/13500 | 0.0726  | 0.882   | ENSAMEG000000013931, ENSAMEG000000015180 | CBFA2T3, CHKA, TTC7B, ENSAMEG00000002034, PFKFB2, INPP5E, INPPL1, ROH2, SMPL3A, NADSYN1, ENSAMEG00000008034, DGH, ADCY9, PD4E4A, FBP1, ENSAMEG00000014118, ADCY4, BAD, ND2, COX1, ATP6, CYT6 |
| GO:  | glycerolipid metabolic process                | biological process              | 10/269      | 202/13500 | 0.0733  | 0.882   | ENSAMEG000000013931, ENSAMEG000000015180 | LPL, CHKA, TTC7B, ENSAMEG00000002834, INPP5E, INPPL1, DGH, APOH, GPX1, DAGLA |
| GO: ID | Term Description | Ontology | Significant | Annotated | Pvalue | Qvalue | Signi_id | Signi_symbol |
|--------|-----------------|----------|--------------|-----------|--------|--------|----------|--------------|
| GO: 1901135 | carbohydrate derivative metabolic process | biological process | 23/269 | 667/13500 | 0.00747 | 0.882 | ENSAMEG000000011842, ENSAMEG000000012590, ENSAMEG000000016637 | CBFA2T3, PLOD3, PFKFB2, NDST1, RHOQ, AHCY, XXYLT1, GNPD1, ADCY9, PEDE4A, FBP1, PORCN, SGSH, ADCY4, DPYD, BAD, B3GNT9, GST4, B3GALT2, ND2, COX1, ATP6, CYTB |
| GO: 0030812 | negative regulation of nucleotide catabolic process | biological process | 2/269 | 7/13500 | 0.00778 | 0.882 | ENSAMEG000000012590, ENSAMEG000000011842 | CBFA2T3, FBP1 |
| GO: 0031269 | pseudopodium assembly | biological process | 2/269 | 7/13500 | 0.00778 | 0.882 | ENSAMEG000000011842, ENSAMEG000000012590 | CDC42EP2, CDC42EP4 |
| GO: 0031272 | regulation of pseudopodium assembly | biological process | 2/269 | 7/13500 | 0.00778 | 0.882 | ENSAMEG000000011842, ENSAMEG000000012590 | CDC42EP2, CDC42EP4 |
| GO: 0031274 | positive regulation of pseudopodium assembly | biological process | 2/269 | 7/13500 | 0.00778 | 0.882 | ENSAMEG000000011842, ENSAMEG000000012590 | CDC42EP2, CDC42EP4 |
| GO: 0001817 | regulation of cytokine production | biological process | 15/269 | 373/13500 | 0.00793 | 0.882 | ENSAMEG00000000943, ENSAMEG00000002415, ENSAMEG00000003641, ENSAMEG00000009343 | LPL, IL17RA, CARD9, GAS6, NOD1, ZFP36, HSPB1, TMEM173, MAST2, IL5RA, HFE, CD36, NLRP3, ARRB1,
another gene involved in vascular remodeling (ENG, 2.1-fold induction) [30].

Previous transcriptomic studies of cataract formation have focused on the analysis of gene expression in models of human congenital cataracts rather than age-related cataract formation, but it is possible that some of the pathway elements are conserved. For example, mutations in the major intrinsic protein (MIP, also known as aquaporin 0) promote cataract formation in human infants, and several strains of rats and mice with loss-of-function mip mutations also develop fully penetrant cataracts [31–35]. More recently, we reported a novel mutation in the panda Mip gene also associated with cataract formation [36]. Transcriptomic analysis in juvenile mip−/− knockout mice identified 29 genes with > 2-fold changes in expression, including the mitochondrial translocase
(Timmdc1), a matrix metallopeptidase (Mmp2), a Rho GTPase-interacting protein (Ubxn11) and a transcription factor (Twist2) which were strongly upregulated, and a proteasome subunit (Psmd8), a ribonuclease (Pop4), and a heat-shock protein (Hspb1) that were strongly downregulated [37].

The discovery of differentially expressed proteasome subunits and heat shock factors indicates that the regulation of protein turnover may contribute to cataract formation, and likewise we identified several differentially expressed genes with similar functions. These included GLG1, which encodes a negative regulator of protein processing (downregulated 5.7-fold in affected pandas), AZINI, which encodes a regulator of protein turnover (downregulated 1.2-fold in affected pandas), and UCHL1, which encodes a thiol-dependent ubiquitin-specific protease (downregulated 1.8-fold in affected pandas). Furthermore, we detected the strong (16.4-fold) upregulation of TNFSF12, encoding a regulator of protein turnover linked to angiogenesis and apoptosis [38], and a 2.1-fold induction of SERPINB10, another regulator of proteases associated with apoptosis [39]. The link

![Significant enrichment functional scatter plot. The vertical axis represents the functional annotation, and the horizontal axis represents the Rich factor of that function. The q-value is represented by the color of the dots, and the number of differentially expressed genes representing each function is shown by the size of the dots. Only the 30 highest enrichments are shown.](image-url)
with apoptosis was also supported by the 3.7-fold induction of EGR1, which mediates p53-independent apoptosis induced by c-Myc [40]. We also observed the 1.5-fold induction of ARHGAP21, a target of p53 that promotes the degradation of MDM2 to maintain TP53 stability [41], and the 1.4-fold induction of STEAP3, another target of p53 that regulates cell cycle progression and apoptosis [42]. The IFIT2L2 gene, encoding an interferon-induced mitochondrial membrane protein that also promotes apoptosis, was induced 1.24-fold [43]. Recent studies have shown that lens epithelial cells undergo apoptosis as a common cytological basis for all kinds of cataract except congenital lesions.

Several of the most strongly modulated genes we identified are involved in signaling, indicating the activation of cell–cell signaling as a response to lens deterioration. The top-ranking downregulated gene was RCAN1 (16-fold repression), which encodes a calcineurin-binding regulator of CNS development. The relevance of this gene in cataracts is unclear, but may be linked to the vascular remodeling discussed above given its role in the suppression of angiogenesis [44]. Other genes we identified appear to be involved in signaling pathways involving G-protein-coupled receptors (GPCRs). For example, CYSLTR1 (induced 1.3-fold) encodes a leukotriene-specific GPCR typically involved in bronchoconstriction, but it also regulates vascular permeability, cell migration and collagen deposition, which may be the more relevant functions here [45]. The UV irradiation of lens epithelial cells was previously shown to increase the expression of LET-7B, the ligand for another GPCR (LGRT), resulting in the induction of apoptosis [46]. A targeted deletion in LGR4 reduced the resistance of rat lens epithelial cells to oxidative stress and accelerated the development of age-related cataracts [47]. RASD1 (induced 4.3-fold) encodes a member of the Ras family of G-protein regulators acting downstream of GPCRs, and its upregulation in human cell lines has been shown to suppress cell growth and induce apoptosis [48, 49]. Similarly RANBP9 (induced 2.7-fold) encodes a Ras protein that influences cytoskeletal organization and interacts with several regulators of cell growth [50]. PDE4A (induced 1.6-fold) encodes a cAMP-specific 3',5'-cyclic phosphodiesterase that regulates second messenger signaling downstream of GPCRs by cleaving cAMP [51]. Other modulated genes encoding signaling proteins potentially involved in the regulation of cell growth included PTP4A3, encoding a membrane-associated protein tyrosine phosphatase (induced 2.8-fold) [52], ANXA3, encoding annexin 3 (downregulated 1.75-fold) [53], and DUISP22, encoding a dual-specificity protein kinase (induced 1.42-fold) [54]. We also observed the 1.92-fold downregulation of the MYBL1 gene, whose principal function is the regulation of meiosis, so its relevance in the context of cataract formation is unclear.

Mutations in the HSF4 gene have been shown to cause cataracts in humans and other mammals [55], including the recent discovery of a novel HSF4 mutation in pandas [56]. HSF4 is a transcriptional repressor and we proposed that the novel mutation we discovered in this gene is likely to affect its interactions with upstream signaling components, thus disrupting the genetic control of lens functions. Although HSF4 was not among the differentially expressed genes we detected in this study, we found that both FOS and FOSB were strongly upregulated (4.5-fold and 4.9-fold, respectively). These genes encode two leucine zipper proteins that dimerize with members of the JUN family to form transcriptional regulators [57]. Importantly, the FOS family of transcription factors is involved in the regulation of many of the processes discussed above revealed by the functional analysis of other differentially expressed genes, including cell proliferation, differentiation and survival, oxidative stress and angiogenesis [58]. It is therefore possible that FOS/FOSB play a key role in the regulation of the other genes discussed above, or that the FOS/FOSB genes are targets of the modulated signaling pathways we have identified.

The analysis of GO enrichment provided results that were broadly consistent with the functional annotation of the differentially expressed genes, with cellular process, metabolic process, biological regulation/regulation of a biological process as the most overrepresented categories in the differentially expressed gene catalog. The categories immune system process and multi-organism process were also represented, which may reflect the modulation of signaling proteins that are also known to participate in the regulation of immunity responses, including cytokine release and immune cell recruitment. Another enriched category was protein-containing complex, which agreed with our preliminary protein network analysis and is consistent with the tendency of signaling proteins to form complexes. Finally, functional annotation by screening the differentially expressed genes against the KEGG database generated several hits in the histidine metabolic pathway, which is particularly interesting given that dietary histidine and/or carnitine are known to prevent cataract development in salmon, presumably by countering the effect of oxidative stress [59, 60]. Taken together, our results confirm that age-related cataracts in pandas bear many of the same hallmarks as cataracts in other animals, including the modulation of stress response genes, metabolic adaptations, and signaling pathways regulating cell growth and apoptosis. The identification of genes that are overexpressed or suppressed during the formation of cataracts could lead to the development of markers for early diagnosis,
preventative strategies and therapies that improve the quality of life for captive giant pandas and other mammals.

Conclusions
Blood samples from six giant pandas with and without cataracts were used for de novo RNA-Seq analysis. This revealed the differential expression of several genes related to oxidative damage, the visual cycle, developmental functions, and lipid metabolism, suggesting that ageing pandas may develop cataracts as the natural capacity for oxidative stress responses begins to diminish. We found that many of the genes encoded regulatory and signaling proteins associated with the control of cell growth, migration, differentiation and apoptosis, supporting previous research indicating a key role for apoptosis in cataract formation. The identification of genes potentially involved in the formation of age-related cataracts could facilitate the development of predictive markers, preventative measures and even new therapies to improve the life of captive animals.

Methods
Clinical findings
Routine physical examinations were carried out every month on the living captive animals, including eye, mouth and nose and general physical appearance, abdominal palpation, and general clinical signs. Blood was collected once a month for the analysis of physiological and biochemical indicators in order to exclude risk factors such as injury, diabetes or other diseases that can promote cataract formation [36].

RNA isolation
Peripheral blood samples (2 ml) were collected from all six giant panda specimens (Table 1) because markers for many diseases can be detected in blood using transcriptomics methods [61] including eye diseases [62–64]. JN and LL were from Beijing Zoo, XX and YE were from Chongqing Zoo, and BD and YY were from Strait (Fuzhou) Giant Panda Research and Exchange Center. Three of the females were diagnosed with age-related cataracts and the other three donors were healthy controls. Blood samples were stored at 4 °C and total RNA was extracted within 2 days using the Trizol Total RNA Extraction kit (Sangon Biotech, Shanghai, China) according to the manufacturer’s protocol. The samples were treated with RNase-free DNase I to remove genomic DNA. RNA integrity was evaluated by 1.0% agarose gel electrophoresis, and RNA quality and quantity were assessed using a NanoPhotometer (Implen, Westlake Village, CA, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

Library preparation and sequencing
High-quality RNA samples were sent to Sangon Biotech for library preparation and sequencing. The libraries were generated from 2 μg RNA using the VAHTS mRNA-seq V2 Library Prep Kit for Illumina (Vazyme Biotech, Nanjing, China) and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations at 94 °C in VAHTS 5× First Strand Synthesis Reaction Buffer, and first-strand cDNA was synthesized using DNA polymerase I and RNase H, which also created blunt ends for 3′ polyadenylation. Following adapter ligation, cDNA fragments in the size range 150–200 bp were selected using the AMPure XP system (Beckman Coulter, Brea, CA, USA). The size-selected and adaptor-ligated cDNA was incubated with 3 μL USER enzyme mix (New England Biolabs, Ipswich, MA, USA) for 15 min at 37 °C then 5 min at 95 °C. We then carried out PCR using Phusion high-fidelity DNA polymerase, universal PCR primers i7 (5′-CAA GCA GAA GAC GGC ATA CGA GAT-index primer-GTG ACT GGA GTT CAG ACG TGT GCT-3′) and i5 (5′-A*A TGA TAC GGC GAC CAC CGA GAT-CTA CAC-index primer-ACA CTC TTT CCC TAC AAC ACC TCA ACG TGC TCC CGA T*T*C-3′) and sample-specific index primers (Table 4). The PCR products were purified (AMPure XP system) and library quality was assessed using the Agilent Bioanalyzer 2100 system. The libraries were then quantified and pooled. Paired-end sequencing was carried out using a HiSeq XTen device (Illumina, San Diego, CA, USA).

Data assessment and quality control
The quality of the sequence data was determined using FastQC v0.11.2. Raw reads were filtered using Trimomatic v0.36 in five steps. First, the adaptor sequences were removed. Next, low-quality bases (Q < 20) were removed from reads in the 3′ to 5′ direction and then in the 5′ to 3′ direction. Then we used a sliding window

| Name     | Index primer (5′ → 3′) | Name     | Index primer (5′ → 3′) |
|----------|------------------------|----------|------------------------|
| Bing-Dian| ACTCGGTAA              | Le-Le    | CTGCTGTC               |
| Ya-Ya    | CGTAAAGCC              | Ji-Ni    | GACTGTTCA              |
| Ya-Er    | CTCATGTC               | Le-Le    | CTGCCTGTT              |
| Ji-Ni    | GACTGTTCA              | Xin-Xing | TCAGTCAC               |

Table 4 Sample-specific index primers for sequencing
method to remove low-quality bases from the read tails (window size = 5 bp). Finally, any reads smaller than 35 nt were removed along with the corresponding paired reads.

**Alignment with reference genome**

Sequences randomly selected from clean data were used as blastn queries against the NCBI nucleotide database (http://ncbi.nlm.nih.gov/). Hits with an E-value cutoff of ≤1×10⁻¹⁰, similarity > 90% and coverage > 80% of the results were used to calculate the species distribution and pollution detection. The remaining clean reads were mapped to the reference genome using HISAT2 v2.0 with default parameters. RSeQC v2.6.1 was used for the statistical analysis of the alignment results.

**Gene structure analysis**

The sequence of each chromosome was compared and mapped statistically to show the distribution of sequences using BEDtools v2.26.0. After comparing the reads to the reference genome, the proportion of each gene structure in the reads was counted, including exons, introns and intergenic DNA, using Qualimap v2.2.1. Only 30 chromosomes were displayed. BCFtools v1.5 was used to find SNPs and SnpEff v2.36 was used to determine their effects. The sequences were filtered according to the mass value (> 20) and coverage (> 8). The sequences were assembled using StringTie v1.3.3b, and compared with existing genomic data in the STRING database (http://string-db.org/) using GUCompare v0.10.1 to find new transcription regions. The predicted transcripts were used as gene models to identify variable splicing. ASprofile v1.0.4 is used to classify variable splicing events according to the predicted gene model of each sample.

**Expression analysis**

StringTie v1.3.3b was applied to the transcriptionally assembled reference genome to determine gene expression levels by measuring transcript abundances as transcripts per million reads (TPM) for both the protein-coding genes and lncRNAs in each sample. After quality control, the sequence was compared with the reference genome using HISAT2 v2.1.0 and the results were statistically compared using RSeQC v2.6.1, with the feature count used to homogenize the read count matrix (gene length and sequence depth) after each gene count to determine the TPM. Principal component analysis (PCA) and principal co-ordinates analysis (PCoA) were used to determine the distance and difference between samples in vegan v2.0.10. Data were presented as TPM avoid the influence of gene length or sequencing discrepancies during sample comparison. We used DESeq2 v1.12.4 to identify genes that were differentially expressed between two samples. Differential expression was considered significant if the false discovery rate (q-value) was ≤0.05 and the log2 fold change (|FC| value) was ≥2. If the normalized expression of a gene between two samples was zero, its expression value was adjusted to 0.01 (because 0 cannot be plotted on a log plot). If the normalized expression of a gene in two libraries was < 1, further differential expression analysis was conducted without this gene.

**Functional analysis of differentially expressed genes**

Given the absence of biological replicate samples, we first standardized all the data with TMM. This removed all the unexpressed genes, identified a sample with average data trends from many samples as the reference sample, calculated the total number of reads for all samples, and divided each sample by its own total number of reads to get the modified number of reads. The Q3 value (the third quartile) of each sample’s modified read number was calculated, averaged, and the sample with the smallest difference from the average Q3 was used as the reference sample. To find the representative gene sets in each sample and calculate the standardization factor of the sample, we referred to the fold change of these representative gene sets. Differentially expressed genes were identified by standardizing the read count data in DESeq2 v1.12.4 as described above. TopGO v2.24.0 was then used for GO enrichment analysis and to prepare the significant GO-directed acyclic graph. ClusterProfiler v3.0.5 was used to analyze KEGG pathways and KOG taxonomic enrichment analysis, allowing the construction of a network diagram. Functional enrichment analysis was carried out to identify differentially expressed genes significantly enriched in GO terms (biological, cellular and molecular functions) or KEGG metabolic pathways. Genes were mapped to the GO database (http://www.geneontology.org) and KEGG database (http://www.kegg.jp) [65–67], the number of genes representing each term or pathway was calculated, and hypergeometric tests were performed to identify significantly enriched GO terms or KEGG pathways in the gene list. GO terms and KEGG pathways were considered significant if the q-value was ≤0.05.

The results of differential gene expression analysis were visualized using DESeq2 v1.12.4 and mapped to the STRING protein interaction network database (http://string-db.org/) to construct the protein interaction network. Based on these results, Venn diagrams were prepared with VennDiagram v1.6.17 in the R package. The correlation of gene expression levels between
samples was determined in R using the cor.test package to ensure data reliability and rational sample selection. The sample-to-sample and group-to-group distance heat map were prepared in R using the gplots v2.17.0 package, which showed the distance relationship between the samples or the groups directly. Scatter plots were used to show the degree of difference between groups of genes and were prepared in R v3.2.

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s12863-021-00996-x.

**Funding**

This study was supported by the National Natural Science Foundation of China (NSFC 31872257). The project funding agency only provided the research funds.

**Availability of data and materials**

The RNA sequencing data are available at NCBI GenBank (https://dataview.ncbi.nlm.nih.gov/object/PRJNA720280?reviewer=m87rldcpcp4b1f1lq490483re). All data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

**Declarations**

**Ethics approval and consent to participate**

All samples were authorized by the source institution (Beijing Zoo, Chongqing Zoo, or Strait (Fuzhou) Giant Panda Research and Exchange Center). All three source institutions conform to the guidelines laid down by the Beijing Zoo Academic and Ethics Committee. The study is reported in accordance with ARRIVE guidelines. All blood samples were collected in accordance with the Wildlife Protection Law of the People’s Republic of China (President of the People’s Republic of China No. 16), and the experimental approach and samples were approved by the Beijing Zoo Academic and Ethics Committee.

**Consent for publication**

Not applicable.

**Competing interests**

The author declare no conflict of interest.

**Author details**

1. Beijing Key Laboratory of Captive Wildlife Technologies, Beijing Zoo, Beijing, China. 2. Beijin Zoo, Beijing, China. 3. Chongqing Zoo, Chongqing, China. 4. Strait (Fuzhou) Giant Panda Research and Exchange Centers, Fuzhou, China. 5. Chengdu Zoo, Chengdu, China.

Received: 28 April 2021 Accepted: 9 September 2021

Published online: 27 October 2021

**References**

1. Triscott RWJEBR. Age-related nuclear cataract-oxidation is the key. Exp Eye Res. 2005;80(5):709–25. https://doi.org/10.1016/j.exere.2004.12.007.

2. Uno HJA. Age-related pathology and biosenescent markers in captive rhesus macaques. Age. 1997;20(1):1–13. https://doi.org/10.1007/BF01357-997-0001-5.

3. Asbell PA, Dulanl I, Mindel JS, Brooks D, Ahmad M, Epstein SPJT. Age-related cataract. Lancet. 2005;365(9459):599–609. https://doi.org/10.1016/S0140-6736(05)67802-0.

4. Urfer SR, Gerek K, Wolf NS. Age-related cataract in dogs: a biomarker for life span and its relation to body size. Age. 2011;33(3):451–60. https://doi.org/10.1007/s11357-010-9158-4.

5. Jin Y, Lin W, Huang S, Zhang C, Pu T, Ma W, et al. Dental abnormalities in eight captive giant pandas (Ailuropoda melanoleuca) in China. J Comp Pathol. 2012;146(4):357–64. https://doi.org/10.1016/j.jcpa.2011.08.001.

6. Jin Y, Chen S, Chao Y, Pu T, Xu H, Liu X, et al. Dental abnormalities of eight wild Qinling giant pandas (Ailuropoda melanoleuca qinlingensis), Shaanxi Province. China J Wildlife Dis. 2015;51(4):849–59. https://doi.org/10.17589/2014-12-289.

7. Hammond CJ, Duncan DD, Snieder H, De Lange M, West SK, Spector TD, et al. The heritability of age-related cortical cataract: the twin eye study. Invest Ophthalmo Vis Sci. 2001;42(3):601–5.

8. Ottonello S, Foroni C, Carta A, Petrucco S, Maraini GJO. Oxidative stress and age-related cataract. Ophthalmologica. 2000;214(1):78–85. https://doi.org/10.1111/j.0029-8419.200027474.x.

9. Ho M, Peng Y, Chen S, Chiu SJ. Senile cataracts and oxidative stress. J Clin Gerontol Geriatr. 2010;1(1):17–21. https://doi.org/10.1016/j.jcgg.2010.10.006.

10. Tinaztepe OE, Ay M, Eser E. CERK: Nuclear and mitochondrial DNA of age-related cataract patients are susceptible to oxidative damage. Curr Eye Res. 2017;42(4):1–6. https://doi.org/10.1080/02713683.2016.1200100.

11. Billingsley G, Santhiya ST, Paterson AD, Ogata K, Wodak SJ, Hosseini SM, et al. CRYPBA, a novel human cataract gene, is also involved in
microphthalmia. Am J Hum Genet. 2006;79(4):702–9. https://doi.org/10.1086/507712.

12. Hasanov N, Kubo E, Kumamoto Y, Takamura Y, Y/JboO A. Age-related cataracts and Pdrx6 correlation between severity of lens opacity, age and the level of Pdrx6 expression. Br J Ophthalmol. 2009;93(1):1081–4.

13. Zhang Y, Zhang L, Sun D, Li Z, Wang L, Liu PJ. Genetic polymorphisms of superoxide dismutases, catalase, and glutathione peroxidase in age-related cataract. Mol Vis. 2011;17:2325–32.

14. Liu X, Luo Y, Zhou P, Lu YAO, Science V. DNA methylation mediated and oxidative stress related genes CRYAA and GJA3 in nuclear age-related cataract (ARC) and its mechanism. Invest Ophthalmol Vis Sci. 2015;56:5877.

15. Zhou P, Luo Y, Liu F, Fan L, Lu Y1FJ. Down-regulation and Cbp island hypermethylation of CRYAA in age-related nuclear cataract. FASEB J. 2012; 26(2):4897–902. https://doi.org/10.1096/fj.11-213702.

16. Wang Y, Li F, Zhang G, Kang L, Qin B, Guan HICER. Altered DNA methylation and expression profiles of 8-oxoguanine DNA glycosylase 1 in lens tissue from age-related cataract patients. Curr Eye Res. 2015;40(8):815–21. https://doi.org/10.1007/s00109-015-2473-2.

17. You Y, Bai C, Liu X, Xia M, Ji A, Li X, et al. Genome-wide analysis of gene expression changes in lens tissue from age-related cataract patients. Curr Eye Res. 2015;40(8):815–21. https://doi.org/10.1016/j.exer.2015.09.004.

18. Doshna CW, Fortner JH, Pfohl JC, Aleo TMW, MEJIO V. Science V. DNA methylation mediated and oxidative stress related genes CRYAA and GJA3 in nuclear age-related cataract (ARC) and its mechanism. Invest Ophthalmol Vis Sci. 2015;56:5877.

19. Salichan K, Sedlund J, Soderberg PG. Kinetics of GADD45 expression in the lens. Invest Ophthalmol Vis Sci. 2002;43:2377.

20. Shiels A, Mackay D, Bassnett S, al-Ghouli K, Kusaka J. Disruption of lens fiber cell architecture in mice expressing a chimeric AQP9-LTR protein. FASEB J. 2000;14(14):2207–12. https://doi.org/10.1096/faseb.99-1071com.

21. Sidjanin DJ, Parker-Williams DM, Neuberger-Klaus A, et al. A 7-bp deletion in the Mip gene causes autosomal dominant cataract in Hf mice. Genomics. 2001;74(3):313–9. https://doi.org/10.1006/geno.2001.6509.

22. Okamura T, Miyoshi I, Takahashi K, Mototani Y, Ishigaki S, Kan Y, et al. Bilateral congenital cataracts result from a gain-of-function mutation in the gene for aquaporin-0 in mice. Genomics. 2003;81(4):361–8. https://doi.org/10.1016/S0888-7543(03)00299-6.

23. Watanabe K, Wada K, Ohashi T, Okubo S, Takekuma K, Hashizume R, et al. A 5-bp insertion in Mip causes recessive congenital cataract in KF54/Kyots rats. Plos One. 2012;7(11):e50737. https://doi.org/10.1371/journal.pone.0050737.

24. Bai C, You YY, Liu XF, Xi MH, Wang W, Ji A, et al. A novel misense mutation in the gene encoding major intrinsic protein (MIP) in a Giant panda with unilateral cataract formation. BMC Genomics. 2021;22(1):100. https://doi.org/10.1186/s12864-021-07386-8.

25. Bennett TM, Zhou Y, Shiels A. Lens transcriptome profile during cataract development in Mip-null mice. Biochem Biophys Res Commun. 2016;472(2):998–93. https://doi.org/10.1016/j.bbrc.2016.08.068.

26. Burkly LC. TWEAK/Fn14 axis: the current paradigm of tissue injury-inducible function in the midst of complexities. Seminars in Immunology. The TNF family - challenges ahead. 2014;26:229–36.

27. Schleef RR, Chuang TL. Protease inhibitor 10 inhibits tumor necrosis factor alpha-induced cell death. Evidence for the formation of intracellular high-M(r) protease inhibitor 10-containing complexes. J Biol Chem. 2000;275(34):36385–9. https://doi.org/10.1074/jbc.C000389200.

28. Boone DN, Qi Y, Li Z, Hans SR. EgR1 mediates p53-independent c-Myc-induced apoptosis via a noncanonical ARF-dependent transcriptional mechanism. Proc Natl Acad Sci U S A. 2011;108(26):10599–604. https://doi.org/10.1073/pnas.1102689108.

29. Rosa LRO, Soares GM, Silveira LR, Boscheri AC, Barbosa-Sampaio HCL. ARHGAP21 as a master regulator of multiple cellular processes. J Cell Physiol. 2018;233(11):8477–81. https://doi.org/10.1002/jcp.26289.

30. Pasier B, Nancy-Portebois V, Arnzallag N, Prieur S, Cans C, Roborel de Climens A, et al. The p53-inducible TAP66 gene product regulates apoptosis and the cell cycle and interacts with nix and the Myt1 kinase. Proc Natl Acad Sci U S A. 2003;100(5):2284–9. https://doi.org/10.1073/pnas.0101733100.

31. Gütz H, Hansen MF, Skovbjerg S, Krietschen AC, Harlyck S, Jensen MB, et al. Apoptotic properties of the type 1 interferon induced family of human mitochondrial membrane 1G12 proteins. BioCell. 2017;19(2):94–112. https://doi.org/10.1111/boc.201600034.

32. Baek KH, Zaslavsky A, Lynch RC, Britt C, Okada Y, Siarey RJ, et al. Down's syndrome suppression of tumour growth and the role of the calcineurin inhibitor Nsflkin. Nature. 2004;99(7250):1126–30. https://doi.org/10.1038/nature06862.

33. Liu M, Yokomizo T. The role of leutokinotors in allergic diseases. Allergol Int. 2015;64(1):17–26. https://doi.org/10.1016/j.jallint.2014.09.001.

34. Dong Y, Zheng Y, Xiao J, Zhu C, Zhao M. MicroRNA let-7b induces lens epithelial cell apoptosis by targeting leucinerich repeat containing G protein-coupled receptor 4 (Lgr4) in age-related cataract. Exp Eye Res. 2016;147:98–104. https://doi.org/10.1016/j.exer.2016.04.018.

35. Zhu J, Hou Q, Dong XD, Wang Z, Chen X, Zheng D, et al. Targeted deletion of the murine Lgr4 gene decreases lens epithelial cell resistance to oxidative stress and induces age-related cataract formation. Plos one. 2015;10(3):e0119599.

36. Vaidyanathan G, Cismowski MJ, Wang G, Vincent TS, Brown KD, Lanier SM. The Ras-related protein AGS1/RASD1 suppresses cell proliferation. Oncogene. 2004;23(4):5848–55. https://doi.org/10.1038/sj.onc.1207865.

37. Jhih Jh Jhih Jhih Jhih Jhih. Chloride intracellular channels 1 and 4 function in distinct branches of SNIP signaling to regulate endothelial cell behavior and vascular development. PhD thesis, Columbia University, 2017.

38. Ulmasov B, Brun B, Gordon N, Hartnett ME, Edwards JC. Chloride intracellular channel protein-4 functions in angiogenesis by supporting acidification of vacuoles along the intracellular tubulogenic pathway. Am J Pathol. 2009;174(4):1084–96. https://doi.org/10.2353/ajpath.2009.080625.

39. Sano-Sanchez S, Guinovart-Maroto A, Botella LM, Vanvill D, Gary CP, Bernabeu C. Endolgin regulates cytoskeletal organization through binding to ZAP-1, a member of the Lim family of proteins. J Biol Chem. 2004; 279(31):32888–6. https://doi.org/10.1074/jbc.M400843200.

40. Shiel A, Bassnett S. Mutations in the founder of the MIP gene family underlie cataract development in the mouse. Nat Genet. 1996;16(2):212–5. https://doi.org/10.1038/ng0296-212.
phosphodiesterase gene. Mol Pharmacol. 2005;67(6):1920–34. https://doi.org/10.1124/mol.104.009423.

52. Zeng Q, Dong JM, Guo K, Li J, Tan HK, Koh V, et al. PRL-3 and PRL-1 promote cell migration, invasion, and metastasis. Cancer Res. 2003;63(11):2716–22.

53. Park JE, Lee DH, Lee JA, Park SG, Kim NS, Park BC, et al. Annexin A3 is a potential angiogenic mediator. Biochem Biophys Res Commun. 2005;337(4):1283–7. https://doi.org/10.1016/j.bbrc.2005.10.004.

54. Sekine Y, Ikeda O, Hayakawa Y, Tsuji S, Imoto S, Aoki N, et al. DUSP22/JMM-DSP2 regulates estrogen receptor-alpha-mediated signaling through dephosphorylation of Ser-118. Oncogene. 2007;26(41):6038–49. https://doi.org/10.1038/sj.onc.1210426.

55. Bu L, Jin Y, Shi Y, Chu R, Ban A, Elberg H, et al. Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. Nature Genet. 2002;31(3):276–8. https://doi.org/10.1038/ng921.

56. You YY, Bai C, Liu XF, Xia MH, Yin YQ, Chen YC, et al. A novel missense mutation in the HSF4 gene of giant pandas with senile congenital cataracts. Sci Rep. 2021;11:5411.

57. Halazonetis TD, Georgopoulous K, Greenberg ME, Leder P. C-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. Cell. 1988;55(5):917–24. https://doi.org/10.1016/0092-8674(88)90147-X.

58. Tulchinsky E. Fos family members: regulation, structure and role in oncogenic transformation. Histol Histopathol. 2000;15(3):921–8. https://doi.org/10.14670/HH-15.921.

59. Waagbar R, Tröse C, Koppe W, Fontanillas R, Breck O. Dietary histidine supplementation prevents cataract development in adult Atlantic salmon, *Salmo salar* L., in seawater. Br J Nutr. 2010;104(10):1460–70. https://doi.org/10.1017/S0007114510002485.

60. Sambraus F, Fjelldal PG, Rema SC, Hevray EM, Nilsen TO, Thorsen A, et al. Water temperature and dietary histidine affect cataract formation in Atlantic salmon (*Salmo salar* L.) diploid and triploid yearling smolt. J Fish Dis. 2017;40(9):1195–212. https://doi.org/10.1111/jfd.12594.

61. Mohr S, Liew CC. The peripheral-blood transcriptome: new insights into disease and risk assessment. Trends Mol Med. 2007;13(10):422–32. https://doi.org/10.1016/j.molmed.2007.08.003.

62. Jeoung JW, Ko JH, Kim YJ, Kim YW, Park KH, Oh JY. Microarray-based analysis of gene expression profiles in peripheral blood of patients with acute primary angle closure. Ophthalmic Genet. 2017;38(6):520–6. https://doi.org/10.1080/13816810.2017.1300922.

63. Rosenbaum JT, Harrington CA, Searles RP, Fei SS, Zaki A, Arepalli S, et al. Revising the diagnosis of idiopathic uveitis by peripheral blood transcriptomics. Am J Ophthalmol. 2021;222:15–23. https://doi.org/10.1016/j.ajo.2020.09.012.

64. Ye HH, Zhang JM, Lu YF, Qian YY. Expression of nitric oxide synthase in peripheral blood of cataract patients and its clinical significance. Chin J Ophthalmol. 2015;54:351–5.

65. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545–51. https://doi.org/10.1093/nar/gkaa970.

66. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30. https://doi.org/10.1093/nar/28.1.27.

67. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 1947-1951;2019:28.

**Publisher’s Note**

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