

Transcriptome Assembly of Asian Buffalo Leech *Hirudinaria Manillensis* and An Accurate Identification of Genes Involved in Anticoagulant Biosynthesis

Chun-Xu Fan

Shaanxi Normal University

Nibras Najm Abbood

Shaanxi Normal University

Jun-Hui Lu

Shaanxi Normal University

Le Wu

Shaanxi Normal University

De-Long Guan

Shaanxi Normal University

Huateng Huang

Shaanxi Normal University

Sheng-Quan Xu (✉ xushengquan@snnu.edu.cn)

College of Life Sciences, Shaanxi Normal University <https://orcid.org/0000-0001-7630-632X>

Research article

Keywords: *Hirudinaria manillensis*, Transcriptome sequencing, comparative transcriptomic analysis, Anticoagulation genes, Differential gene expression

Posted Date: September 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-67619/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The Asian buffalo leech *Hirudinaria manillensis* is a valuable animal widely used in traditional Chinese medicine to cure blood-clotting. However, although the content of the anticoagulant genes in *H. manillensis* were identified, their actual expression profile remains undetermined. Herein, in this study we conducted transcriptome analysis on this species to address this subject.

Results: A high qualified accurate gene expression profile of *H. manillensis* transcripts was obtained. By identifying genes upregulated during feeding, anticoagulant-related genes and signaling pathways were identified. The assembled Unigenes were compared in various mainstream databases, and a total of 16,682 Unigenes received annotation information. In addition, the Unigenes were evaluated in terms of length distribution, GC content, and expression level. The data showed good sequencing quality and high reliability. A total of 155 anticoagulant-related genes were found in this study, including those involved in different types of degradation of the extracellular matrix and inhibition of platelet aggregation.

Conclusions: Substantial transcriptome information was obtained by transcriptome sequencing of *H. manillensis*. This information should help to provide a further molecular theoretical basis for functional gene analysis, genomics, genetic diversity analysis, and molecular marker development of *H. manillensis* and for the anticoagulation-related genes of this species and its medicinal value as an anticoagulant.

Background

Hirudinaria manillensis (common name Asian buffalo leech, *Poecilobdella manillensis*) is a leech of the Arhynchobdellida from the order Hirudinea. It is also a type species of the family *Poecilobdella*. These large bloodsucking leeches live in paddies, ditches, rivers, or ponds and survive by feeding on the blood of mammals, which has high nutritional value. They also exhibit clinical anticoagulant, thrombolytic, and antiplatelet activities. They are animals with medicinal value and have great potential for further uses because of their effects on aggregation, lowering blood lipids, and improving blood circulation [1].

In recent years, a new generation of RNA sequencing technology (RNA-seq) has changed the way research is performed away from the model of traditional molecular biology. This technology provides an important means of analyzing gene expression levels, and performing differential gene expression screening, functional gene mining, and evolutionary analysis [2–7]. Transcriptome sequencing can enrich our understanding of the leech genome. For example, as described in reports published in 2013 and 2016, Kvist et al. used 454 pyrosequencing to obtain transcriptome data from the salivary secretions of *Haemadipsa interrupta* [8] and *Pontobdella macrothela*, revealing the correlations of anticoagulants contained in these two leeches and exploring the evolution of these anticoagulants in different leeches.

Leeches have become increasingly widely used in medical applications due to their natural and high-efficiency anticoagulant capacity [9–12]. However, to date, research on *H. manillensis* has only focused on its classification, breeding, and exploration of the level of anticoagulant activity. The lack of transcriptome and genome-related research has restricted our understanding of this important medicinal animal and medicinal research on it. In view of this, we carried out transcriptome sequencing of *H. manillensis*, the results of which

can supplement the information in genetic databases on this species and strongly promote related research on other leeches. The identified anticoagulant-related genes can also be investigated by others. Related research on bloodsucking animals provides a reference for promoting the development and utilization of medicines by humans.

Results

Transcriptome assembly

A total of 41.80 Gb of clean data was obtained from the transcriptome sequencing analysis of six transcriptome samples of *H. manillensis*. The clean data of each sample reached 6.88 Gb, the expected coverage exceeded 200×, and the Q20 base percentage was over 94.90%. The Q30 base percentage was over 88.83%. Clean reads were subjected to filtering out of the joints with the joints, as well as the removal of low-quality reads (unknown N > 1% and quality values Q < 15 reads), so there was no need to further use Trimmomatic or remove other impurities. Sequencing statistics of all six transcriptome samples are shown in Table 1.

Table 1
Data on *H. manillensis* transcriptome sequencing results

| Sample ID | Raw Reads | Clean Reads | Raw Bases (Gb) | Clean Bases (Gb) | Effective Rate(%) | Read Length | Q20 (%) | Q30 (%) | GC (%) |
|-----------|------------|-------------|----------------|------------------|-------------------|-------------|---------|---------|--------|
| ST1 | 46,812,410 | 45,953,050 | 7.02 | 6.89 | 98.16 | 150×2 | 97.70 | 94.19 | 43.74 |
| ST2 | 47,683,582 | 46,840,394 | 7.15 | 7.03 | 98.23 | 150×2 | 97.76 | 94.29 | 43.51 |
| ST3 | 48,116,230 | 47,844,864 | 7.58 | 7.11 | 93.77 | 150×2 | 97.43 | 93.69 | 43.05 |
| AF1 | 47,116,220 | 46,275,578 | 7.07 | 6.94 | 98.22 | 150×2 | 97.82 | 94.44 | 42.65 |
| AF2 | 46,678,946 | 45,885,834 | 7.00 | 6.88 | 98.30 | 150×2 | 97.45 | 93.66 | 43.44 |
| AF3 | 48,826,824 | 47,301,362 | 7.55 | 7.13 | 94.43 | 150×2 | 94.90 | 88.83 | 42.66 |

Note: Sample ID: sample name ST refers to before blood sucking, while AF refers to after blood sucking. Effective rate: The amount of valid data, the ratio of clean reads to raw reads. Q20: The number of reads achieving a sequencing error quality value of less than 1%. Q30: The number of reads achieving a sequencing error quality value of less than 0.1%.

Functional annotation of genes

To obtain complete functional annotations, all assembled single gene sequences were used as search queries in various databases, and the assembled Unigene uses GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes), KOG (Clusters of Orthologous Groups), P-fam (Protein family), Swiss-

Prot (a manually annotated and reviewed protein sequence database), and Nr (NCBI Non-redundant Protein Sequences) to understand the gene types and expression. The notes on each database are shown in Table 2; according to this table, a total of 16,682 Unigenes received annotation information, of which for 9397 information was posted to the Unigenes in the GO database; 7838 in the KEGG database; 11,534 in the KOG database; 14,273 Unigene in the P-fam database; whilst 15,933 Unigene in the Swiss-Prot database and 13,461 Unigenes annotated to the Nr database, accounting for 56.33%, 46.98%, 69.14%, 85.56%, 95.51% and 80.69% of the total number of annotated Unigenes respectively. The clean reads of each sample were sequence-matched with the designated reference genome. Based on the alignment results, alternative splicing prediction analysis, gene structure optimization analysis, and new gene discovery were performed, and 992 new genes were discovered, of which 394 were functionally annotated. The structure of 7,592 genes was optimized(table S1).

Table 2
Gene function annotation statistics

| Database | Annotated number | Percentage (%) |
|---------------|------------------|----------------|
| COG | 5837 | 34.99 |
| GO | 9397 | 56.33 |
| KEGG | 7838 | 46.98 |
| KOG | 11,534 | 69.14 |
| P-fam | 14,273 | 85.56 |
| Swiss-Prot | 15,933 | 95.51 |
| eggNOG | 14,604 | 87.54 |
| Nr | 13,461 | 80.69 |
| All Annotated | 16,682 | 100 |

Note: Annotation database: Functional annotation database; Annotated number: Number of genes annotated to the corresponding database; Percentage (%): Percentage of genes annotated in the database as a percentage of total genes.

KOG, GO, and KEGG analyses

Classification of KOG functions

The *H. manillensis* Unigenes were annotated into the KOG database and classified based on the orthologous relationship. A total of 11,534 Unigenes were annotated in this database and are involved in 25 functional categories. There was a large number of Unigene annotations with only general functional predictions. This may be because the current database and transcriptome data of leeches are very limited. Among the KOG functional categories of *H. manillensis* of all Unigenes, the top ten most involved in the number of Unigenes

show signal transduction mechanisms. These categories include signal transduction mechanisms; General function prediction only; Post-translational modification, protein turnover chaperones; Function unknown; Intracellular trafficking, secretion, and vesicular transport; Inorganic ion transport and metabolism; RNA processing and modification; Cytoskeleton; and Translation, Ribosome structure and biogenesis, with corresponding numbers of Unigenes of 2184, 2127, 1183, 871, 856, 608, 502, 499, 471, and 464 respectively (Table S2). Among them, 856 Unigenes with unknown function were found. We speculate that these sequences may be new genes unique to *H. manillensis*.

Classification of GO functions

A total of 9,397 Unigenes were annotated in the GO database and divided into three categories: cellular components, molecular functions, and biological processes (Figure 2). In the 17 functional categories of the main category cellular component (table S2), the main annotations of the Unigenes were cell, cell part, and organelle components. In the 15 functional categories of the main category molecular function, the main annotations of the Unigenes were binding, catalytic activity, and transporter activity. Finally, in the 16 functional categories in the main category of biological process, the main annotations of the Unigenes were metabolic processes, cellular processes, and single-organism processes.

KEGG pathway enrichment analysis

The obtained *H. manillensis* Unigene sequences were analyzed in the KEGG database for comprehensive metabolic pathway analysis. A total of 7,838 Unigenes were annotated in this database, which participate in 253 KEGG pathways (table S2). The top ten pathways most strongly associated with the Unigenes were Spliceosome (199 Unigenes); Purine metabolism (196 Unigenes); Ubiquitin-mediated proteolysis (178 Unigenes); RNA transport (175 Unigenes); Endocytosis (158 Unigenes); Protein processing in endoplasmic (151 Unigenes); Ribosome (135 Unigenes); Neuroactive ligand-receptor interaction Ribosome (131 Unigenes); Pyrimidine metabolism (123 Unigenes); and Lysosome (122 Unigenes).

Analysis of differentially expressed genes of *H. manillensis* before and after bloodsucking

DEGs screening

The screening of differentially expressed genes (DEGs) before and after vaccination of *H. manillensis* was carried out in accordance with strict screening criteria ($FDR \leq 0.01$ and $FC \geq 2$). The screening results obtained are shown in Figure 4. The results show that there are 3048 DEGs (table S3) between before and after bloodsucking (ST and AF). Compared with the levels in the ST group, the expression levels of 2020 genes were upregulated in AF, while the expression levels of 1028 genes were downregulated. After the

feeding of *H. manillensis*, the number of upregulated genes was higher than that of the downregulated genes. This indicates that the bodily functions of this species do not decrease to reduce consumption after a single feeding, and more genes are instead activated to participate in metabolic mechanisms. The identified DEGs should provide an abundance of molecular information for further exploration of the bloodsucking habits of *H. manillensis*.

GO function analysis of DEGs

The selected DEGs were annotated with GO functions to reveal the associations between DEGs and certain biological functions. The results showed that a total of 1210 DEGs were annotated in the GO data (table S3). The GO annotation information results of the DEGs are shown in Figure 5. According to these results, DEGs were classified into 48 functional categories in the three major categories of biological processes, cellular components, and molecular functions. Within the category of biological processes, involving 16 GO function categories, the differentially expressed genes were particularly associated with metabolic processes, cellular processes, and single-organism processes, while the remaining DEGs were mainly involved in response to stimulus, among others. Within the category of cellular component, involving 17 GO function categories, the categories of membrane, membrane part, cell, and cell part had the most differentially expressed genes. Within the 15 functional categories of molecular function, the differentially expressed genes were mainly involved in binding, catalytic activity, and transporter activity. The remaining DEGs were mainly involved in enzyme regulatory activity and electron carrier activity, among others. From this, we can conclude that metabolic activity, protein synthesis activity, and stress response of *H. manillensis* change significantly between before and after blood sucking.

Then, the study enriched the GO of the differential gene in this process, the results of which are shown in Figure 6. In the results of the GO term enrichment of *H. manillensis*, the main identified categories were protein degradation, plasma membrane, serine endopeptidase, response to wounding and cell redox homeostasis. These uniquely enriched GO terms are related to damage repair and inflammation of the body. These results show that there are diverse anti-inflammatory and analgesic effects on the host in *H. manillensis*. It is speculated that these may be related to the size of the wound and the physiological characteristics of the bloodsucking.

Functional analysis of KEGG pathways associated with DEGs

Using the KEGG database, pathway enrichment analysis on the DEGs was performed. The results revealed a total of 889 DEGs that participate in 173 metabolic pathways (table S3). There was a total of 20 KEGG pathways with significant enrichment, including Metabolism of xenobiotics by cytochrome, Apoptosis, Glutathione metabolism, Vitamin B6 metabolism, VEGF signaling pathway, Phenylalanine and so on. The statistical results are shown in Table 3.

For the KEGG pathway results, this study found that, among the KEGG pathways enriched with the differentially expressed genes of *H. manillensis*, the particularly well-represented pathways included Neuroactive ligand-receptor interaction, Dorsoventral axis formation, cysteine and methionine metabolism.

Most of these genes are not directly or only indirectly related to anticoagulation, but are related to developmental regulation. Therefore, it is speculated that the activated metabolism of *H. manillensis* is the main genetic change that occurs after feeding, and there is relatively limited regulation of anticoagulation-related genes.

Table 3
KEGG pathways with significant enrichment of differentially expressed genes

| Pathway-term | Gene_number |
|-----------------------------------------------------|-------------|
| Metabolism of xenobiotics by cytochrome | 6 |
| Apoptosis | 5 |
| Glutathione metabolism | 13 |
| Vitamin B6 metabolism | 2 |
| VEGF signaling pathway | 2 |
| Ubiquinone and other terpenoid-quinone biosynthesis | 3 |
| Starch and sucrose metabolism | 6 |
| Phenylalanine, tyrosine and tryptophan biosynthesis | 3 |
| Phenylalanine metabolism | 4 |
| Nitrogen metabolism | 6 |
| Neuroactive ligand-receptor interaction | 20 |
| MAPK signaling pathway | 8 |
| Glyoxylate and dicarboxylate metabolism | 7 |
| Drug metabolism-cytochrome P450 | 4 |
| Dorso-ventral axis formation | 10 |
| DNA replication | 7 |
| Cysteine and methionine metabolism | 11 |
| Biosynthesis of unsaturated fatty acids | 5 |
| Arginine and proline metabolism | 11 |
| Adipocytokine signaling pathway | 3 |

Note: Pathway-term: KEGG path name; Gene number: Number of genes enriched

Functional gene mining related to bloodsucking characteristics of *H. manillensis*

Screening of anticoagulant genes in *H. manillensis*

The P-fam protein family were annotated with antithrombin collected by querying the interproscan website [13, 14]. All genes annotated to the same P-fam family were then screened as anticoagulant-related genes in *H. manillensis*. The results showed that there were 155 anticoagulant-related gene sequences in the *H. manillensis* genome. These genes were categorized using anticoagulant and anticoagulant-related gene types (analgesic and anti-inflammatory, inhibition of platelet, extracellular matrix degradation, and direct anticoagulation). The results are shown in Table 4. The number of anticoagulant and anticoagulant-related loci in *H. manillensis* suggests the possible anticoagulant effects and medicinal value.

Table 4
Classification of anticoagulant and anticoagulant genes in *H. manillensis*

| Functional classifications | Gene description | Obtained gene number |
|----------------------------------|------------------------------------------------------|----------------------|
| Analgesic and anti-inflammatory | Antistasin family | 34 |
| | Interleukin-17 | 2 |
| | Kunitz/Bovine pancreatic trypsin inhibitor domain | 6 |
| | Hemerythrin HHE cation binding domain | 3 |
| | Macin | 2 |
| | Spider toxin | 1 |
| | Potato inhibitor I family | 17 |
| Extracellular matrix degradation | Inter-alpha-trypsin inhibitor heavy chain C-terminus | 1 |
| | Matrixin | 21 |
| Inhibition of platelets | Apyrase | 8 |
| | Destabilase | 8 |
| | Disintegrin | 51 |
| Anticoagulant | Hirudin | 1 |

Note: Functional classification: Functional annotation classification; Obtained gene number: Number of genes belonging to this classification

Anticoagulant-related genes of *H. manillensis*

Antistasin family-related genes

The antistasin family of proteins is a class of cysteine-rich serine protease inhibitors that can markedly inhibit the coagulation process, albeit through different mechanisms of action [15-17]. Antistasin can also respond to host inflammatory responses, exerting an anti-inflammatory effect. Among the genes related to the antistasin family identified in this study, 10 related genes were upregulated and 5 were downregulated upon blood sucking in *H. manillensis*. This indicates that the antistasin protein family plays an important role in the anticoagulant process after the blood sucking of *H. manillensis*.

ATP diphosphate hydrolase (apyrase)-related genes

ATP diphosphate hydrolase (apyrase) is an antiplatelet factor. The mechanism by which apyrase in the saliva of bloodsucking arthropods works is as follows: ADP present in the host after tissue damage promotes platelet activation, but apyrase can hydrolyze ATP and ADP to AMP and inorganic phosphorus, which hinders platelet aggregation [18-22]. This process makes it difficult for a bitten host to form a thrombus, which facilitates the feeding of bloodsucking animals.

In this study, three apyrase-related genes were shown to be downregulated after blood sucking by *H. manillensis* and one was downregulated. This indicates that apyrase participates in the anticoagulation process after *H. manillensis* bloodsucking.

Destabilase-related genes

Destabilase acts as a thrombolytic agent in leech; it cleaves peptide bonds and promotes fibrinolysis, thereby inhibiting coagulation activity [23-26]. As a multifunctional enzyme, destabilase can also exert antibacterial function. In addition, the salivary glands of the leech can also secrete other kinds of enzyme, which together with destabilase play a bacteriostatic role.

This study identified three genes related to destabilase that were upregulated after the blood sucking of *H. manillensis* and two that were downregulated. The results show that destabilase plays a role in the anticoagulant and antibacterial processes upon the sucking of blood by *H. manillensis*.

Disintegrin-related genes

Disintegrin contains a sperm-glycoside-aspartic acid sequence or a lysine-aspartate sequence and is rich in cysteine. It binds to the glycoprotein fibrinogen receptor on the platelet membrane, competitively antagonizing the binding of fibrin to platelets, and prevents platelet activation and changes in glycoprotein receptor conformation, blocking the binding of receptors to multiple ligands. In this way, it inhibits the final common pathway of platelet aggregation and blocks the formation of a thrombus.

This study identified 15 genes related to disintegrin that were upregulated after blood sucking by *H. manillensis* and 9 that were downregulated. This indicates that disintegrin plays an important role in the anticoagulation process after the blood sucking of this species.

Gene analysis of bloodsucking characteristics of *H. manillensis*

Finally, 155 genes related to anticoagulation were screened, corresponding to 13 active substances, including antistasin, apyrase, destabilase, and disintegrin. According to their functions and effects, these genes are classified into those involved in platelet inhibition and those directly encoding anticoagulants. Among them, the expression level of the antistasin family, which has both anticoagulant and anti-inflammatory effects, was generally increased. In terms of the expression trends of genes related to the inhibitors of platelet aggregation apyrase, destabilase, and disintegrin, these were upregulated. At the same time, the expression levels of related genes involved in the anti-inflammatory response of host wounds also showed an upward trend. This shows that, in the process of blood sucking, *H. manillensis* produces a variety of active anticoagulant substances, and directly or indirectly, these together inhibit the host's blood coagulation process, accompanied by an anti-inflammatory response. Taken together, these findings indicate that the anticoagulant process is a complex biological pathway in which multiple substances cooperate and interact.

Upon consuming blood, bloodsucking animals can also take in infectious bacteria from the surface of the wound and immune proteins from the blood of the host. Therefore, they need to effectively protect themselves against bacterial invasion and immune protein attack. Destabilase has a strong inhibitory effect on bacteria. Hyaluronidase also acts as an active bacteriostatic protein, which dissolves the envelope of bacteria and forms antibodies, effectively inhibiting the activity of various bacteria and fungi. Some of the functional genes associated with the immune response, such as those annotated as hyaluronidase, were shown to be upregulated after the blood sucking of *H. manillensis*.

After bloodsucking animals ingest host blood, the hemoglobin in the blood tends to hydrolyze to produce a large amount of hemoglobin. Heme is a potent oxidant that produces hydroxyl radicals with excellent oxidizing ability. Studies have shown that heme can cause DNA damage in mitochondria and alter the expression of apoptotic proteins in these organelles. Hemoglobin also causes a certain degree of damage to the genome. In *H. manillensis*, hemoglobin is likely to cause great damage during the process of digesting blood. Therefore, this species needs to cope with the oxidative stress generated during blood digestion. Citelli et al. [15] asserted that a variety of proteins with oxidoreductase activity, such as superoxide dismutase, glutathione S-transferase, and thioredoxin, are involved in this antioxidant mechanism. In the process of blood sucking by animals such as ticks, the stress response and non-specific immunity play important roles. After the blood sucking of *H. manillensis*, the expression of related genes involved in the antioxidant process was found to be upregulated and the antioxidant capacity was significantly enhanced.

Upon blood sucking, *H. manillensis* needs to produce a large amount of active anticoagulant substances to ensure the fluidity of the host blood. At the same time, it also produces antibacterial and immune-related substances to protect against foreign substances, and proteins having oxidoreductase activity are produced by antioxidation to cope with oxidative stress generated by hemoglobin during blood digestion.

Functional analysis of KEGG pathway in anticoagulation-related DEGs

We found 155 anticoagulant-related genes based on the transcriptome data of *H. manillensis* and explored their related pathways. The Notch signaling pathway, nicotinate, and nicotinamide metabolism associated with anticoagulant genes are involved in cell growth. It is of great significance to help to detect the occurrence and development of tumors and to treat them in a targeted manner by excavating these pathways.

(1) Notch signaling pathway

In the KEGG database annotation, there are 48 Unigene annotations in the Notch signaling pathway, with three DEGs participating in the pathway. The Notch signaling pathway is important for cell-to-cell communication, which regulates cell differentiation, apoptosis, proliferation, and morphogenesis. It is also critical for the growth and development of organisms [27-32]. The Notch signaling pathway is mainly composed of Notch-related receptors, Notch-associated ligands, Notch downstream signal transduction molecules, and nuclear response factors. TACE (ADAM17) is a key enzyme in the activation of the Notch signaling pathway and has a significant limiting effect on its activation [33-35, 39]. Simultaneous activation of ADAM17 induces GPIIb digestion [36], and platelet membrane glycoprotein (GP) Iba digestion reduces the expression of platelet surface functional receptors, resulting in weakened platelet adhesion. This affects the formation of blood clots and inhibits the aggregation function of platelets, being a recognized negative regulatory mechanism of platelets. After the feeding of *H. manillensis*, the expression of TACE-related gene (ADAM 17-like protease) is upregulated, which inhibits platelet aggregation, blocks thrombus formation, affects the activation of the Notch signaling pathway, and participates in the anticoagulant process.

Notch-mediated signaling plays a key role in the development of the cardiovascular system and cardiovascular disease, and is closely related to the regulation of immune system function, pancreatic cancer, and medulloblastoma [37-39]. Notch signaling may play different roles in different environments, times, or cell lines, and its over- or under-expression may have a negative impact on the organism. Therefore, an in-depth understanding of Notch signaling and its interaction with various pathways can provide a foundation for more research on the treatment of cardiovascular diseases and tumors, and for carrying out targeted and efficient clinical treatment.

(2) Nicotinate and nicotinamide metabolism

ATP diphosphate hydrolase (Apyrase) blocks platelet aggregation. The apyrase-related gene detected in this study is involved in the nicotinate and nicotinamide metabolism pathway, and for three enzymes there was upregulated gene expression after vaccination with *H. manillensis*, while for one enzyme there was downregulated gene expression.

In the KEGG database annotation, there are 27 Unigene annotations on the Notch signaling pathway, with four DEGs participating in the pathway. Nicotinamide phosphoribosyltransferase (Nampt) is the rate-limiting enzyme of this pathway [40-42]; it also has physiological functions such as promoting angiogenesis, anti-apoptosis, participating in the body's inflammatory response, and promoting the proliferation, differentiation, and maturation of various cells. After the feeding of *H. manillensis*, apyrase-related genes are involved in the anticoagulant process, inhibiting platelet aggregation, and downregulating the expression of Nampt-related genes, affecting the niacin and nicotinamide metabolic pathways.

The concentration of Nampt in the blood also significantly increases in many tumor patients, so an inhibitor of the activity of this enzyme is an anti-tumor drug with great potential for clinical application [41-43]. However, in this context, there are still many problems to be resolved. The role of anticoagulation genes related to *H. manillensis* before and after blood sucking is discussed to provide a basis for the better application of tumor treatment.

SNP analysis

Single-nucleotide polymorphism (SNP) is mainly a DNA sequence polymorphism caused by DNA replication, a genetic marker formed by a single-nucleotide variation in the genome. These variations have become commonly used as molecular markers, and in disease diagnosis, paternity testing, plant breeding, and genetic mapping [44, 45]. By identifying potential SNP sites, it is possible to analyze whether these sites affect the expression level of a gene of interest or the particular protein that is produced.

SNPs can be divided into two types: transitions and transversions. Depending on the number of alleles at the SNP site, the SNP can be divided into a homozygous SNP site (only one allele) or a heterozygous SNP site (two or more alleles). There is a difference in the proportion of heterozygous SNPs in different species. SNP molecular markers were developed from *H. manillensis* transcriptome data. In the obtained 17,051 Unigenes of *H. manillensis* transcripts, an average of 154,399 SNP sites were found, of which 109,917 were transitions, accounting for 71.19%, and 44,482 were transversions, accounting for 28.81% (Table 5).

Table 5
SNP statistics

| Sample ID | SNP Number | Genic SNP | Intergenic SNP | Transition | Transversion | Heterozygosity |
|-----------|------------|-----------|----------------|------------|--------------|----------------|
| ST1 | 147115 | 124889 | 22226 | 71.41% | 28.59% | 51.76% |
| ST2 | 169399 | 140997 | 28402 | 70.70% | 29.30% | 49.96% |
| ST3 | 132107 | 111179 | 20928 | 71.54% | 28.46% | 51.54% |
| AF1 | 194328 | 159091 | 35237 | 70.21% | 29.79% | 52.70% |
| AF2 | 173142 | 145623 | 27519 | 70.70% | 29.30% | 51.13% |
| AF3 | 110302 | 94053 | 16249 | 72.57% | 27.43% | 47.29% |

Note: Sample ID: Sample analysis number; SNP Number: Total number of SNP sites; Genic SNP: Total number of SNP sites in the gene region; Intergenic SNP: Total number of SNP sites in the intergenic region; Transition: The percentage of the number of conversion-type SNP sites within the total number of SNP sites; Transversion: The percentage of the number of SNP sites of the transversion type within the total number of SNP sites; Heterozygosity: The percentage of heterozygous SNP sites in the total number of SNP sites.

Discussion

H. manillensis from the order Hirudinea is a bloodsucking leech. It has a wide range of feeding habits, a large body size, and sucks a large amount of blood. We used the Illumina Hiseq sequencing platform to sequence the transcriptome of this species. We obtained a total of 41.80 Gb of clean data for analysis, with 16,682 annotated Unigenes. We used the six databases GO, KEGG, KOG, P-fam, Swiss-Prot, and Nr to annotate the *H. manillensis* transcript Unigenes, which corresponded to 9397, 7838, 11,534, 14,273, 15,933, and 13,461, respectively, and found 155 anticoagulant-related genes. Using the KEGG database, we also performed pathway enrichment analysis on the DEGs. Finally, we identified a total of 889 DEGs participating in 173 metabolic pathways. We explored the anticoagulant-related genes and pathways based on *H. manillensis* transcriptome data. After sucking blood, *H. manillensis* produces active anticoagulant substances, and antibacterial and immune-related substances. The protein having oxidoreductase activity produces an oxidative stress by hemoglobin during blood digestion by antioxidation.

Transcriptomes are important resources for genetic diversity analysis, comparative genomics, and potential molecular marker-assisted selection. SNP molecular markers were developed from *H. manillensis* transcriptome data. In the obtained 17,051 Unigenes of *H. manillensis* transcripts, an average of 154,399 SNP sites were found, of which 109,917 were transitions, accounting for 71.19%; and 44,482 were transversions, accounting for 28.81%. Among all variant types, G/A and C/T occurred most frequently. The 154,399 SNP sites identified in this transcriptome should be a valuable resource as molecular markers in *H. manillensis*.

The transcriptome dataset of *H. manillensis* is valuable for further research on functional gene analysis and molecular marker development of this species. At the same time, it provides a theoretical molecular basis for the study of anticoagulant-related genes and the anticoagulant medicinal value of *H. manillensis*.

Materials And Methods

Material

All leeches collected in this study were in vivo. At least 20 individuals of *H. manillensis* were collected from Hechi City, Guangxi Province. They were kept in mineral water and brought back to the laboratory. Then, each individual was subjected to morphological identification, and after species determination they were selected for the experiment and reared in a dark and ventilated place.

RNA extraction, detection, library construction, and sequencing

The RNA required for transcriptome construction was extracted using Qiagen kit (QIAGEN RNeasy mini kit). First, the live leeches were directly frozen in liquid nitrogen, then pulverized in a mortar, and a small amount of RNAiso lysate was added and centrifuged immediately. After the precipitate had been obtained, a sufficient amount of lysate was added and digested for at least 2 h. Next, total RNA was extracted from the lysed sample in accordance with the standard procedure on the kit. The quality of the final extracted RNA was accurately quantified using 1% agarose gel electrophoresis, qubit, and an Agilent 2100 bioanalyzer. Since the transcriptome was designed for two expression states per species, which were set to pre-feeding and 3 h after feeding, for each transcriptome sequencing sample, the total RNA of an individual was extracted, which was stored in liquid nitrogen and used for subsequent sequencing.

The transcriptome sequencing was carried out by Beijing Baimaike Biotech Co., Ltd. For transcriptome sequencing, two RNA libraries were constructed for *H. manillensis* using the Hiseq 4000 platform. Library construction used oligomagnetic beads to physically enrich mRNA; the mRNA was then eluted and the random six-mer primers were reverse-transcribed to form a cDNA library. The cDNA library was then subjected to ultrasonication, fragment recovery (300 bp), end repair, poly-A addition, ligation, library recovery, and PCR preamplification. After the library quality test had been passed, bridge PCR was carried out for chip fixation, and PE 150 bp (PE 150) was sequenced. The generated data were then used for subsequent analysis.

Functional annotation of genes

The NR[46], Swiss-Prot[47], GO[48], COG[49], KOG[50], Pfam[51] and KEGG[52] gene annotations were used for Blast[53] alignment using the downloaded databases, and the E-value threshold was set to $1e - 01$. The KEGG pathway map was obtained using KAAS[54].

Screening and analysis of anticoagulant genes in *H. manillensis*

A number of records related to anticoagulant-related proteins and polypeptides in leech are available for download. By downloading all of these previously reported leeches with anticoagulant function and protein

and polypeptide sequences associated with anticoagulation, from NCBI's protein. DESeq[55] was used for differential expression analysis between sample groups to obtain the differentially expressed gene set between the two biological conditions

SNP detection

In this study, GATK[56] software was used to identify the single base mismatch between the sequencing sample and the reference genome based on the comparison result of TopHat2 of reads of each sample and the reference genome sequence, so as to identify potential SNP loci. Then it can be analyzed whether these SNP sites affect the expression level of genes or the type of protein products. Meanwhile, SnpEff [57] were by annotated variation (SNP) and predict variation effects.

Abbreviations

NR:non-redundant protein sequence database;Swiss-Prot:A manually annotated, non-redundant protein sequence database;GO:Gene Ontology database;COG:The database of Clusters of Orthologous Groups of proteins;KOG:The database of Clusters of Protein homology;Pfam:The database of Homologous protein family;KEGG:The database of Kyoto Encyclopedia of Genes and Genomes;DEGs:differentially expressed genes; Nampt: Nicotinamide phosphoribosyltransferase ;SNP: Single-nucleotide polymorphism

Declarations

Acknowledgements

Not applicable.

Funding

This work was supported financially by the Excellent Doctor Innovation Project of Shaanxi Normal University (S2015YB03), Fundamental Research Funds for the Central Universities (2019CSLZ009), and Fundamental Research Funds for the Central Universities (GK201604008; GK201702010; GK201701006;). In addition, this work was partly supported by the National Natural Science Foundation of China (no. 31372250).

Availability of data and materials

The RNA sequence dataset supporting the present results is available in the repository of NCBI Sequence Read Archive (SRA) with the following accession Number: SRX6035209 ~ SRX6035214.

Authors' contributions

S-Q. X, and H-H. T conceived the study and designed the experiments. NNA and J-H L performed sequencing and bioinformatics analysis, C-X F, L-W and D-L. G analyzed the data. C-X F and D-L. G wrote the manuscript.

S-Q. X and H-H. T revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

No permission was required for Shaanxi Normal University employees to perform insect collections.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

Author details

College of Life Sciences, Shaanxi Normal University, Xi'an, China

References

1. Hussein M , Ammar A E , Chin M Y , et al. Abstract 5130: Assessment of the antitumor activity of leech (*hirudinaria manillensis*) saliva extract in prostate cancer[J]. *Journal of Clinical Oncology*, 2015, 33:e16120.
2. Mittempergher L , Spangler J B , Snel M H , et al. Abstract 5409: Assessment of the MammaPrint 70-gene profile using RNA sequencing technology[J]. *Cancer Research*, 2017, 77(13 Supplement):5409-5409.
3. Todd E V , Black M A , Gemmell N J . The power and promise of RNA-seq in ecology and evolution[J]. *Molecular Ecology*, 2016, 25(6):1224-1241.
4. Mittempergher L , Spangler J B , Snel M H , et al. Abstract 5409: Assessment of the MammaPrint 70-gene profile using RNA sequencing technology[J]. *Cancer Research*, 2017, 77(13 Supplement):5409-5409.
5. Rau A , Maugis-Rabusseau C , Martin-Magniette M L , et al. Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models[J]. *Bioinformatics*, 2015, 31(9):1420-1427.
6. Ran D , Daye Z J . Gene expression variability and the analysis of large-scale RNA-seq studies with the MDSeq[J]. *Nucleic Acids Research*, 2017, 45(13).
7. Yijun Z , Fei G , Ran L , et al. Referenced and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in *Ammopiptanthus mongolicus*[J]. *BMC Genomics*, 2012, 13.
8. Kvist S , Brugler M R , Goh T G , et al. Pyrosequencing the salivary transcriptome of *Haemadipsa interrupta* (Annelida: Clitellata: Haemadipsidae): anticoagulant diversity and insight into the evolution of anticoagulation capabilities in leeches[J]. *Invertebrate Biology*, 2014, 133(1):74-98.

9. Tessler M , Marancik D , Champagne D , et al. MARINE LEECH ANTICOAGULANT DIVERSITY AND EVOLUTION.[J]. Journal of Parasitology, 2018.
10. Salzet M . Anticoagulants and inhibitors of platelet aggregation derived from leeches[J]. Febs Letters, 2001, 492(3):187-192.
11. Abdulkader A M , Ghawi A M , Alaama M , et al. In vivo anti-hyperglycemic activity of saliva extract from the tropical leech *manillensis* [J]. Chinese Journal of Natural Medicines, 2013, 11(5):488-493.
12. Alaama M . Characterization and optimization of lyophilization and storage conditions of Leech saliva extract from the tropical leech *manillensis* .[J]. Pakistan Journal of Pharmaceutical Sciences, 2013, 26(3):525-535.
13. InterPro in 2017—beyond protein family and domain annotations[J]. Nucleic Acids Research, 2017, 45(D1):D190-D199.
14. Mulder N J . InterPro[M]// Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics. John Wiley & Sons, Ltd, 2005.
15. Christa SÖLLNER, M7ntele R, Eckerskorn C , et al. Isolation and characterization of hirustasin, an antistatin-type serine proteinase inhibitor from the medical leech *Hirudo medicinalis*[J]. European Journal of Biochemistry, 1994, 219(3):937-943.
16. Mittl P R , Marco S D , Fendrich G , et al. A new structural class of serine protease inhibitors revealed by the structure of the hirustasin–kallikrein complex[J]. Structure, 1997, 5(2):253-264.
17. Uson I , Sheldrick G M , De L F E , et al. The 1.2 Å crystal structure of hirustasin reveals the intrinsic flexibility of a family of highly disulphide-bridged inhibitors[J]. Structure, 1999, 7(1):55-63.
18. Faudry E , Santana J , Ebel C , et al. Salivary apyrases of *Triatoma infestans* are assembled into homo-oligomers[J]. Biochemical Journal, 2006, 396(3):509.
19. Adalberto A. Pérez de León, Tabachnick W J . Apyrase activity and adenosine diphosphate induced platelet aggregation inhibition by the salivary gland proteins of *Culicoides variipennis*, the North American vector of bluetongue viruses[J]. Veterinary Parasitology, 1996, 61(3-4):327.
20. Goto S , Salomon D R , Ikeda Y , et al. Characterization of the Unique Mechanism Mediating the Shear-dependent Binding of Soluble von Willebrand Factor to Platelets[J]. Journal of Biological Chemistry, 1995, 270(40):23352-23361.
21. Smartt C T , Kim A P , Grossman G L , et al. The Apyrase Gene of the Vector Mosquito, *Aedes eegypti*, Is Expressed Specifically in the Adult Female Salivary Glands[J]. Experimental Parasitology, 1995, 81(3):239-248.
22. Koyamada, Miyatake, Tsukasa, et al. APYRASE ADMINISTRATION PROLONGS DISCORDANT XENOGRAFT SURVIVAL1,2,3,4[J]. Transplantation, 1996, 62(12):1739-1743.
23. Zavalova L L , Kuzina E V , Levina N B , et al. Monomerization of fragment DD by destabilase from the medicinal leech does not alter the N-terminal sequence of the gamma-chain[J]. Thrombosis Research, 1993, 71(3):241.
24. Baskova, Zavalova L , Berezhnoy S , et al. Inhibition of induced and spontaneous platelet aggregation by destabilase from medicinal leech[J]. Platelets, 2000, 11(2):83-86. Baskova L P , Nikonov G I . Destabilase, the novel ϵ -(Glu)-Lys isopeptidase with thrombolyt43

25. activity[J]. *Blood Coagulation & Fibrinolysis*, 1991, 2(1):167-172.
26. B456`+va I P , Kalabushev S N , Akhaev D N , et al. Role of isopeptidolysis in the process of thrombolysis[J]. *Thrombosis Research*, 2018, 165:18-23.
27. Ji P , Yu L , Guo W C , et al. Doxorubicin Inhibits Proliferation of Osteosarcoma Cells Through Upregulation of the Notch Signaling Pathway.[J]. *Oncology Research*, 2014, 22(4):185.
28. Lu Z , Liu H , Xue L , et al. An activated Notch1 signaling pathway inhibits cell proliferation and induces apoptosis in human esophageal squamous cell carcinoma cell line EC9706[J]. *International Journal of Oncology*, 2008, 32(3):643—651.
29. Jung S G , Kwon Y D , Song J A , et al. Prognostic significance of Notch 3 gene expression in ovarian serous carcinoma[J]. *Cancer Science*, 2010, 101(9):1977-1983.
30. Urbanek K , Lesiak M , Krakowian D , et al. Notch signaling pathway and gene expression profiles during early in vitro differentiation of liver-derived mesenchymal stromal cells to osteoblasts[J]. *Laboratory Investigation*, 2017.
31. Wang X D , Leow C C , Zha J , et al. Notch signaling is required for normal prostatic epithelial cell proliferation and differentiation[J]. *Developmental Biology*, 2006, 290(1):0-80.
32. Suwanjune S , Wongchana W , Palaga T . Inhibition of gamma-secretase affects proliferation of leukemia and hepatoma cell lines through Notch signaling[J]. *Anti-Cancer Drugs*, 2008, 19(5):477-486.
33. Paryan M , Mohammadiyeganeh S , Samiee S M , et al. Investigation of deregulated genes of Notch signaling pathway in human T cell acute lymphoblastic leukemia cell lines and clinical samples.[J]. *Molecular Biology Reports*, 2013, 40(10):5531-5540.
34. Tian L , Wu X , Chi C , et al. ADAM10 is essential for proteolytic activation of Notch during thymocyte development[J]. *International Immunology*, 2008, 20(9):1181-1187.
35. Yong L , Zhen R , Yu W , et al. ADAM17 promotes cell migration and invasion through the integrin β 1 pathway in hepatocellular carcinoma[J]. *Experimental Cell Research*, 2018:S0014482718303811-.
36. Pidard D , Renesto P , Berndt M C , et al. Neutrophil proteinase cathepsin G is proteolytically active on the human platelet glycoprotein Ib-IX receptor: characterization of the cleavage sites within the glycoprotein Ib\, α r, subunit[J]. *Biochemical Journal*, 1994, 303(2):489-498.
37. Zhang P , Yang Y , Nolo R , et al. Regulation of NOTCH signaling by reciprocal inhibition of HES1 and Deltex 1 and its role in osteosarcoma invasiveness[J]. *Oncogene*, 2010, 29(20):2916-2926.
38. Saravanakumar M , Devaraj H . Notch signalling in cardiogenesis: Insight into their role in early cardiovascular development[J]. *Molecular Biology Reports*, 2013, 40(5):3537-3547.
39. Yang Y , Duan W , Jin Z , et al. New role of Notch-mediated signaling pathway in myocardial ischemic preconditioning[J]. *Medical Hypotheses*, 2011, 76(3):427-428.
40. Laiguillon M C , Bougault C , Priam S , et al. AB0131?NAMPT/VISFATIN in osteoarthritis: An adipokine involved in the communication between cartilage and bone[J]. *Annals of the Rheumatic Diseases*, 2013, 71(Suppl 3):645.4-645.
41. Wang J Z , Liu Y , Wang X Y , et al. Abstract 3006: Arsenic trioxide inhibits overexpression of NAMPT in oral cancer[J]. *Cancer Research*, 2016, 76(14 Supplement):3006-3006.

42. Wang J Z , Liu Y , Wang X Y , et al. Abstract 3006: Arsenic trioxide inhibits overexpression of NAMPT in oral cancer[J]. *Cancer Research*, 2016, 76(14 Supplement):3006-3006.
43. Ju H Q , Zhuang Z N , Li H , et al. Regulation of the Nampt-mediated NAD salvage pathway and its therapeutic implications in pancreatic cancer[J]. *Cancer* Alderborn, A. Determination of Single-Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing[J]. *Genome Research*, 2000, 10(8):1249-1258.
44. Liu X P , Hou J L , Liu J H . A novel single nucleotide polymorphism detection of a double-stranded DNA target by a ribonucleotide-carrying molecular beacon and thermostable RNase HIII[J]. *Analytical Biochemistry*, 2010, 398(1):83-92.
45. Ju H Q , Zhuang Z N , Li H , et al. Regulation of the Nampt-mediated NAD salvage pathway and its therapeutic implications in pancreatic cancer[J]. *Cancer Letters*, 2016, 379(1):1-11.
46. Deng YY, Li JQ, Wu S F, Zhu YP, et al. Integrated NR Database in Protein Annotation System and Its Localization. *Computer Engineering* 2006.,32(5):71-74.
47. Apweiler R, Bairoch A, Wu CH, Barker WC, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research* 2004 Jan 1;32(Database issue):D115-9.
48. shburner M, Ball C A, Blake J A, Botstein D, et al. Gene ontology: tool for the unification of biology. *Nature genetics* 2000, 25(1): 25-29.
49. Tatusov RL, Galperin MY, Natale DA. The COG database: a tool for genome scale analysis of protein functions and evolution. *Nucleic Acids Research* 2000, 28(1):33-36.
50. Koonin EV, Fedorova ND, Jackson JD, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome biology*, 2004, 5(2): R7.
51. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. *Nucleic acids research*, 2013: gkt1223.
52. Kanehisa M, Goto S, Kawashima S, Okuno Y, et al. The KEGG resource for deciphering the genome. *Nucleic Acids Research* 2004, 32(Database issue):D277-D280.
53. Altschul S F, Madden T L, Schäffer A A, Zhang J, et al. Gapped BLAST and PSI BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research* 1997, 25(17): 3389-3402.
54. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;35(Web Server issue):W182-W185.
55. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology* 2010, 11:R106.
56. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 2010, 20(9): 1297-1303.
57. Cingolani P, Platts A, Wang L L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012, 6(2): 80-92.

Figures

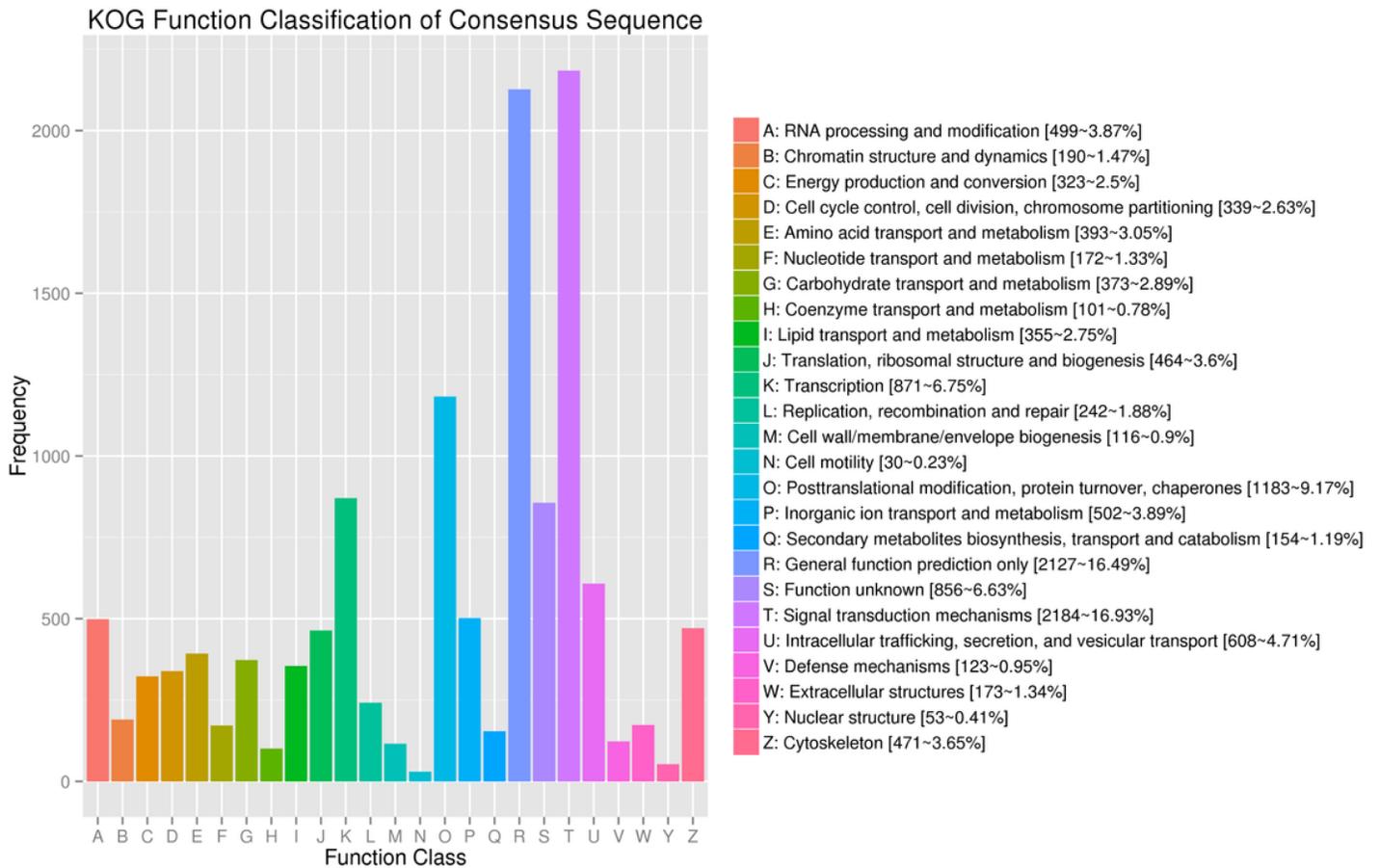


Figure 1

Functional annotation classification chart of *H. manillensis* Note: The abscissa represents the KOG classification, while the ordinate represents the number of genes.

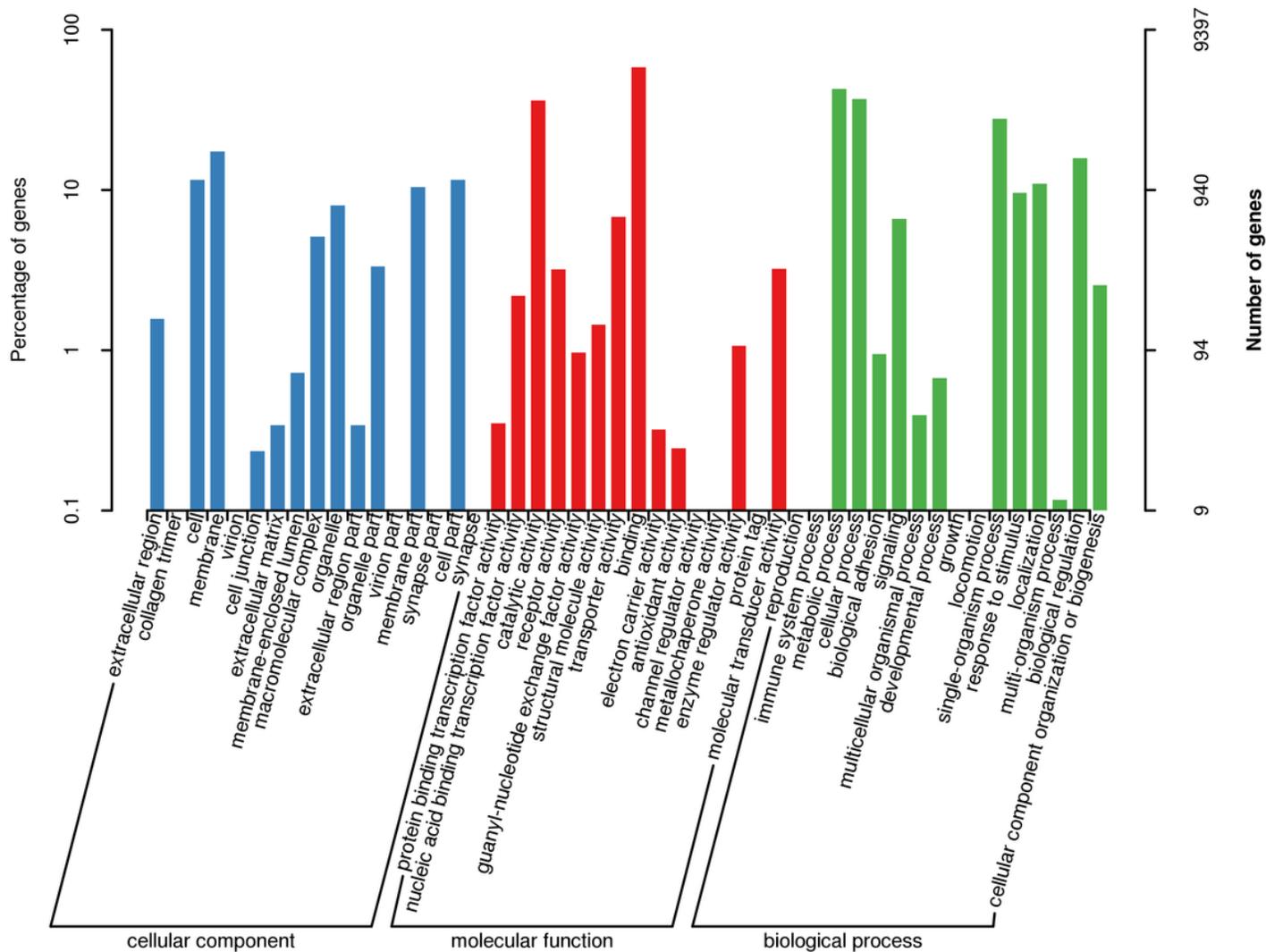


Figure 2

Taxonomic chart of GO annotation of *H. manillensis* Note: The abscissa presents the GO classification content (term), the left ordinate presents the percentage of the number of genes (%), and the right ordinate presents the number of genes.

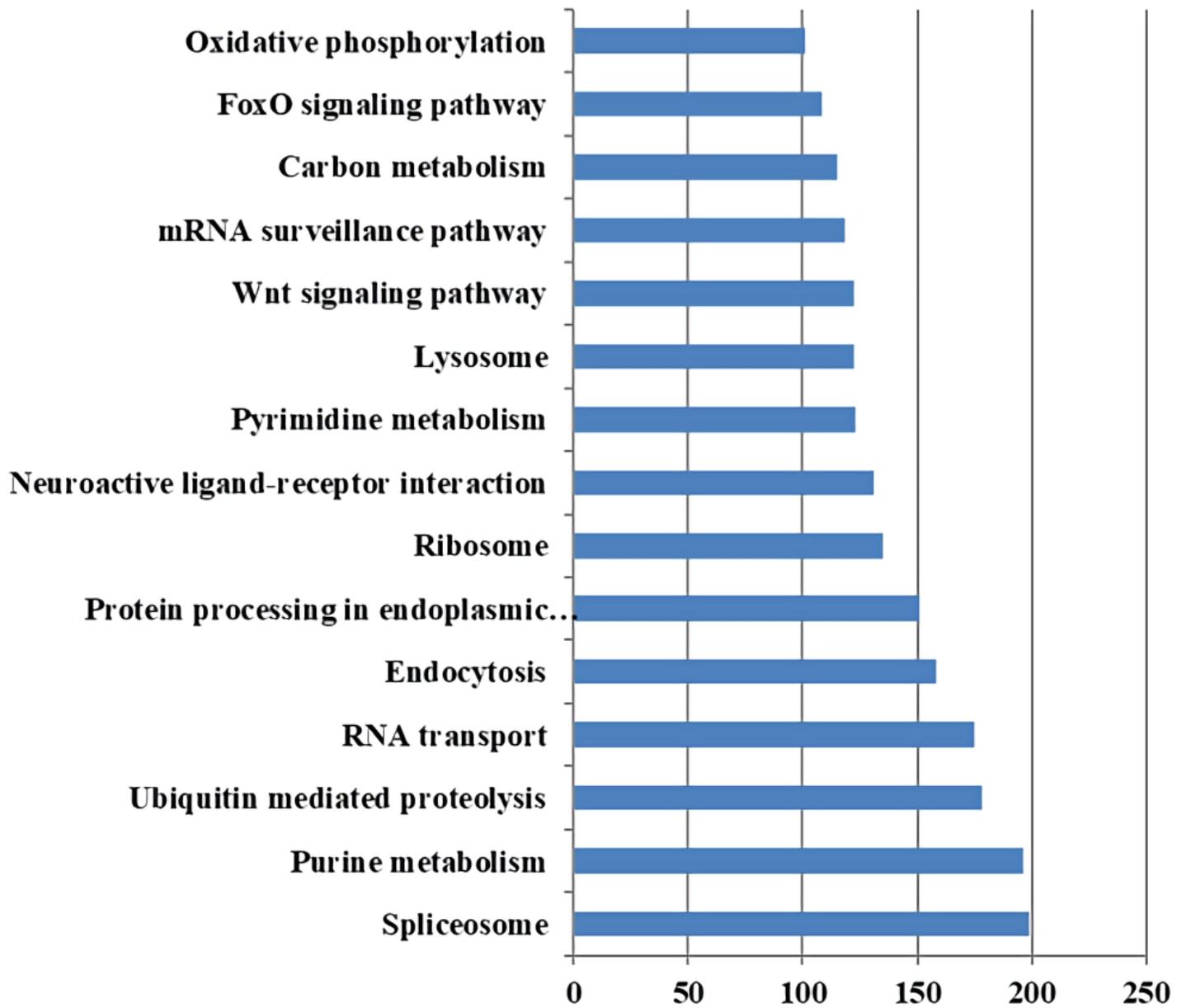


Figure 3

H. manillensis KEGG annotation classification chart Note: The abscissa presents the number of genes and the ordinate presents the KEGG path.

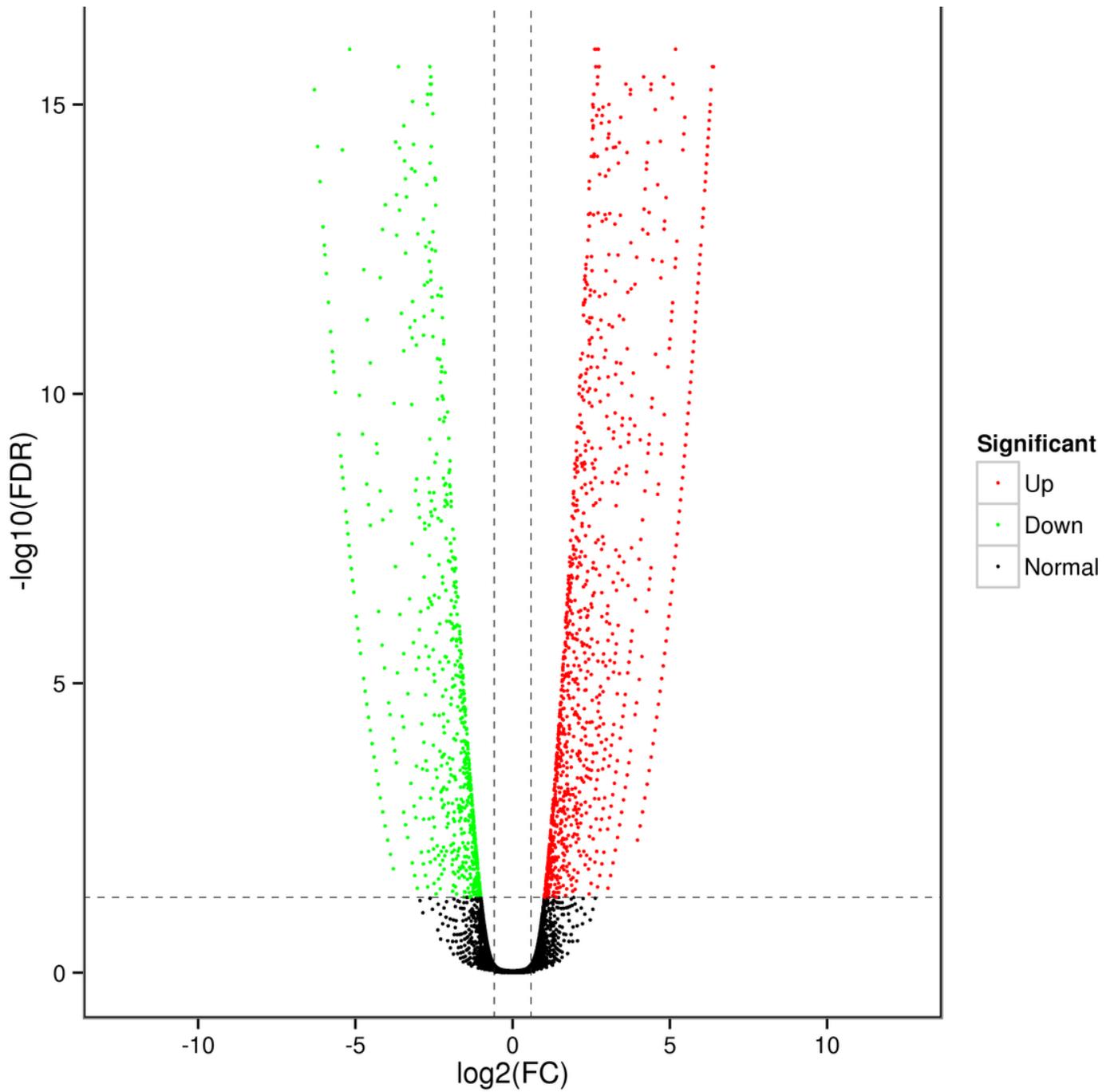


Figure 4

Volcano map of the differential gene distribution of *H. manillensis*. Note: Red represents up, green represents down. The abscissa shows the \log_2 value of fold change, while the ordinate shows the \log_{10} value of FDR.

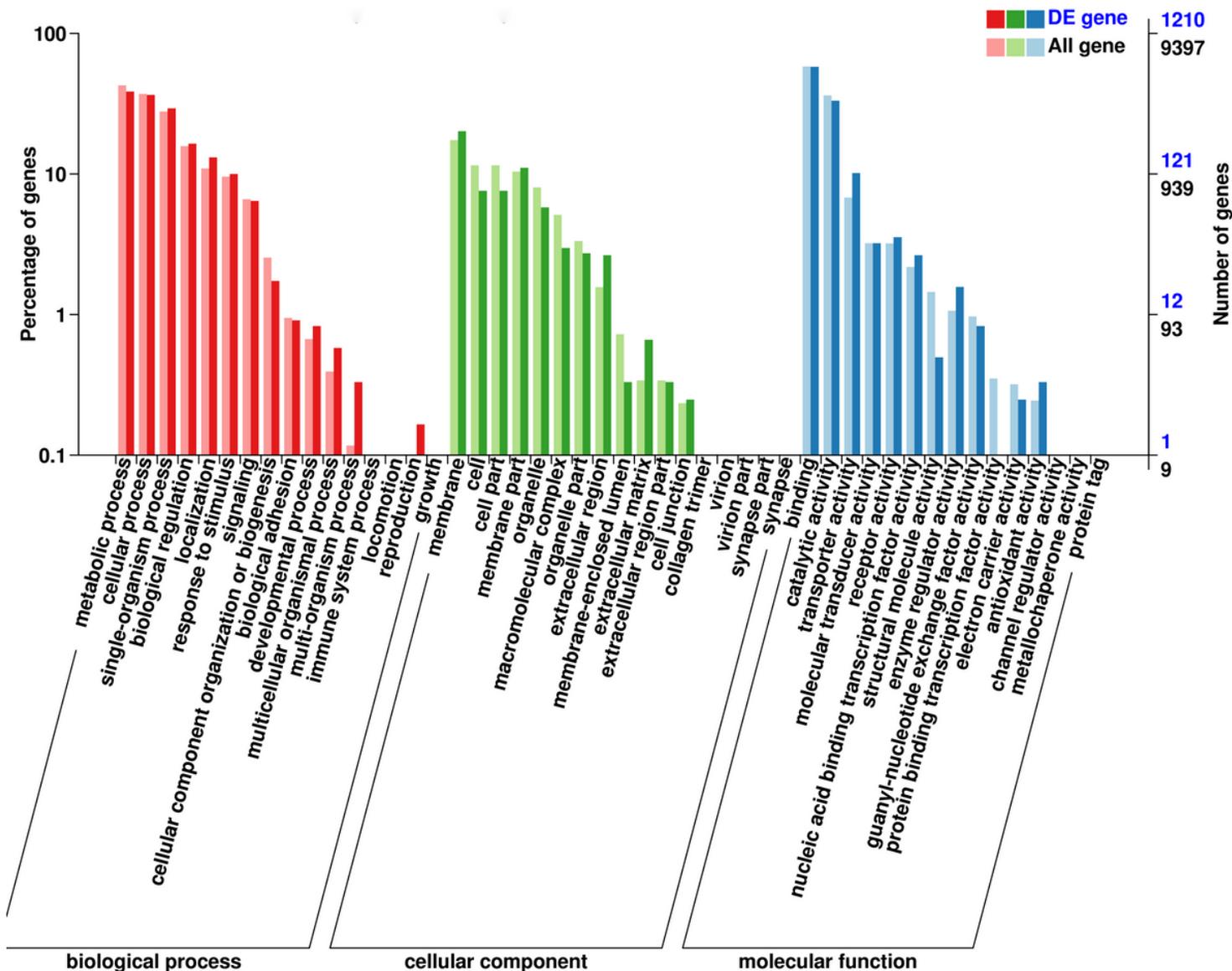


Figure 5

GO functional classification of differentially expressed genes in *H. manillensis*. Note: The abscissa presents the GO classification (terms), the left ordinate presents the percentage of the number of genes (%), while the right ordinate presents the number of genes.

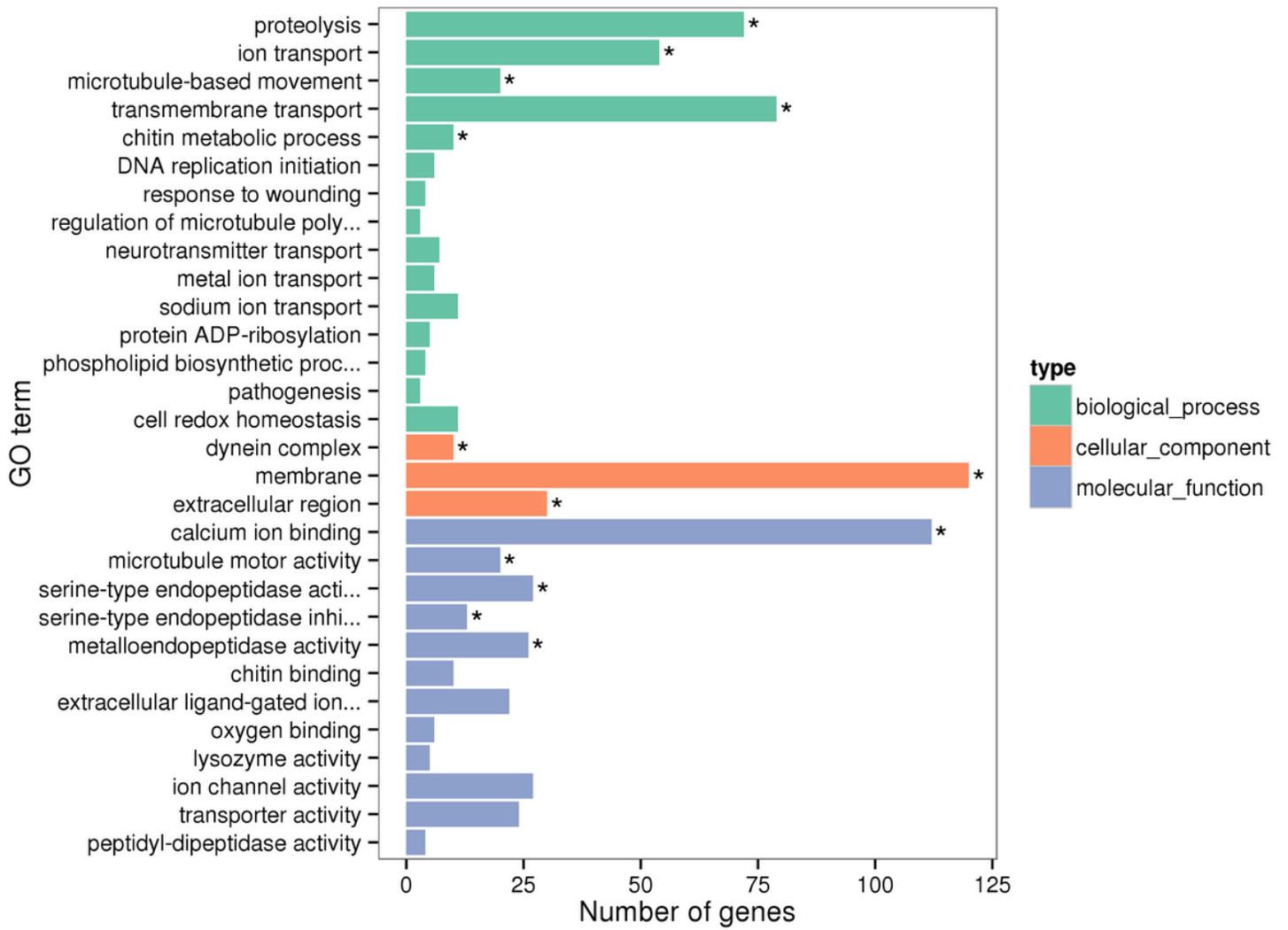


Figure 6

GO enrichment histogram of differentially expressed genes in *H. manillensis*. Note: The ordinate presents the enriched GO terms and the abscissa presents the number of differentially expressed genes in the term. Green, orange, and purple represent biological process, cellular component, and molecular function, respectively.

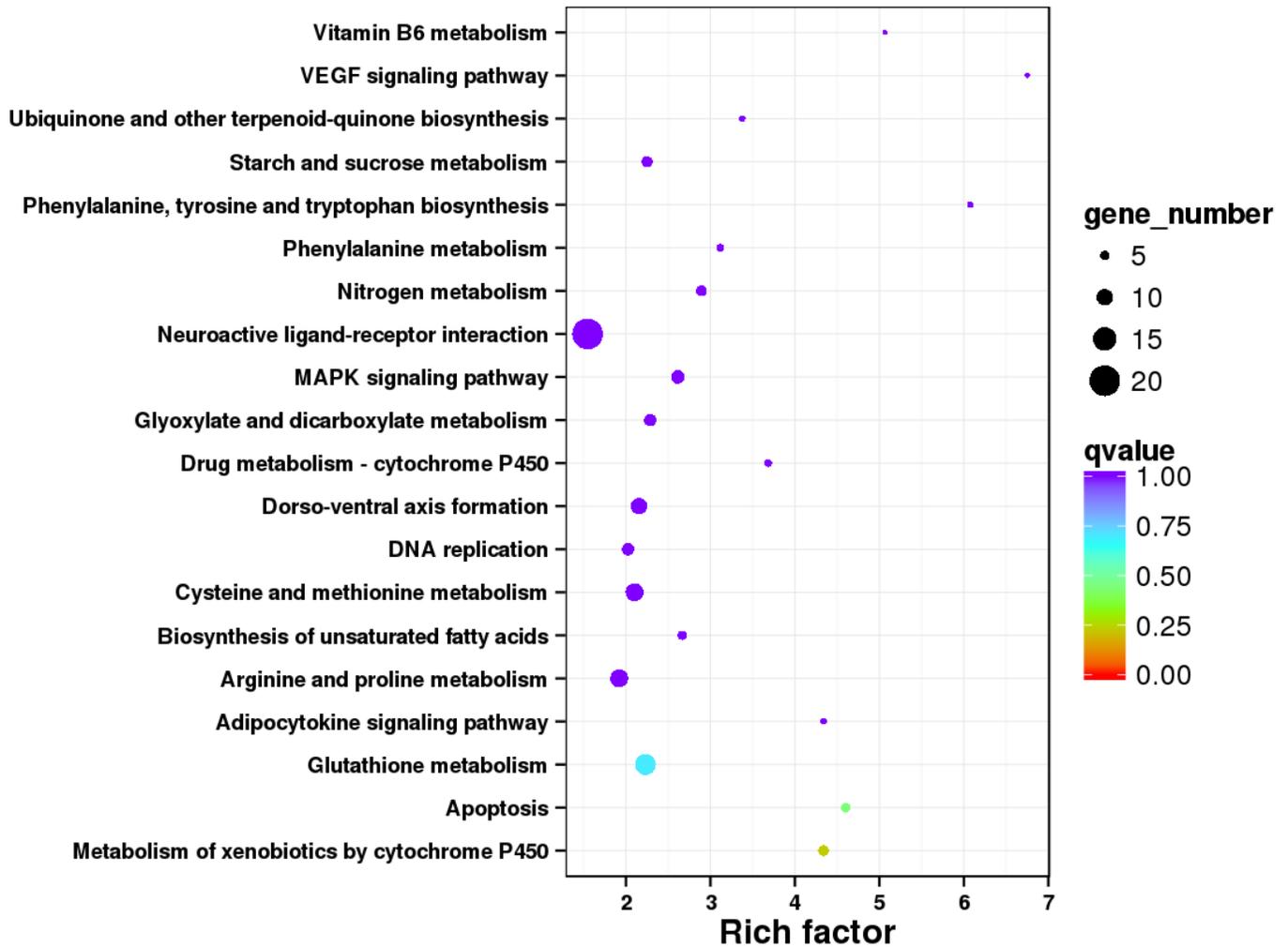


Figure 7

Map of KEGG pathways with enrichment of differentially expressed genes in *H. manillensis*. Note: The vertical axis represents the path name, while the horizontal axis represents the rich factor. The size of the dot indicates the number of differentially expressed genes in this pathway, and the color of the point corresponds to a different Qvalue range.

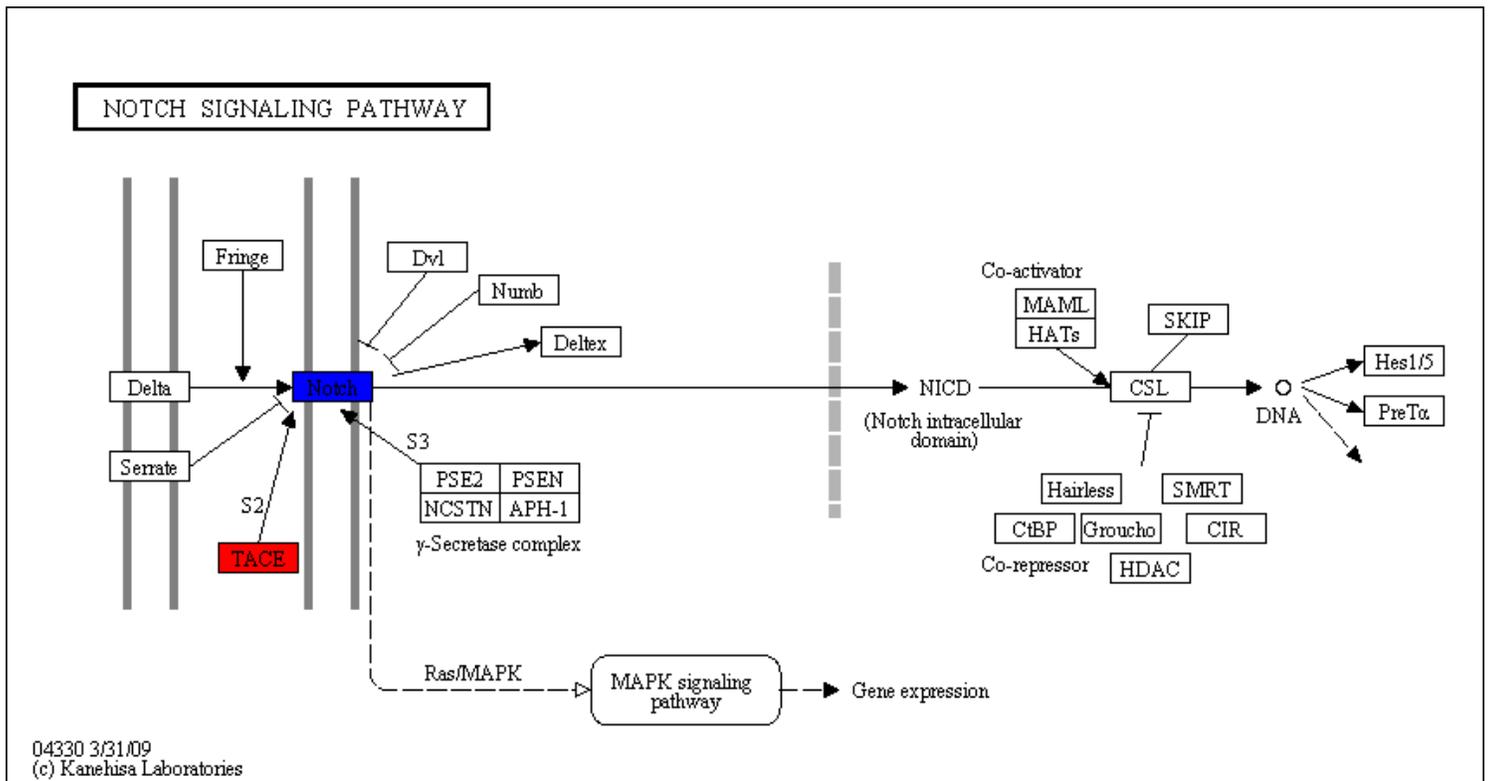


Figure 8

Notch signaling pathway Note: Relative to the control group, the red box-labeled enzyme is associated with the upregulated gene, and the green-box-labeled enzyme is associated with the downregulated gene. The blue box-labeled enzyme is related to both upregulated and downregulated genes, and the numbers in the box represent the numbers of enzymes (EC numbers).

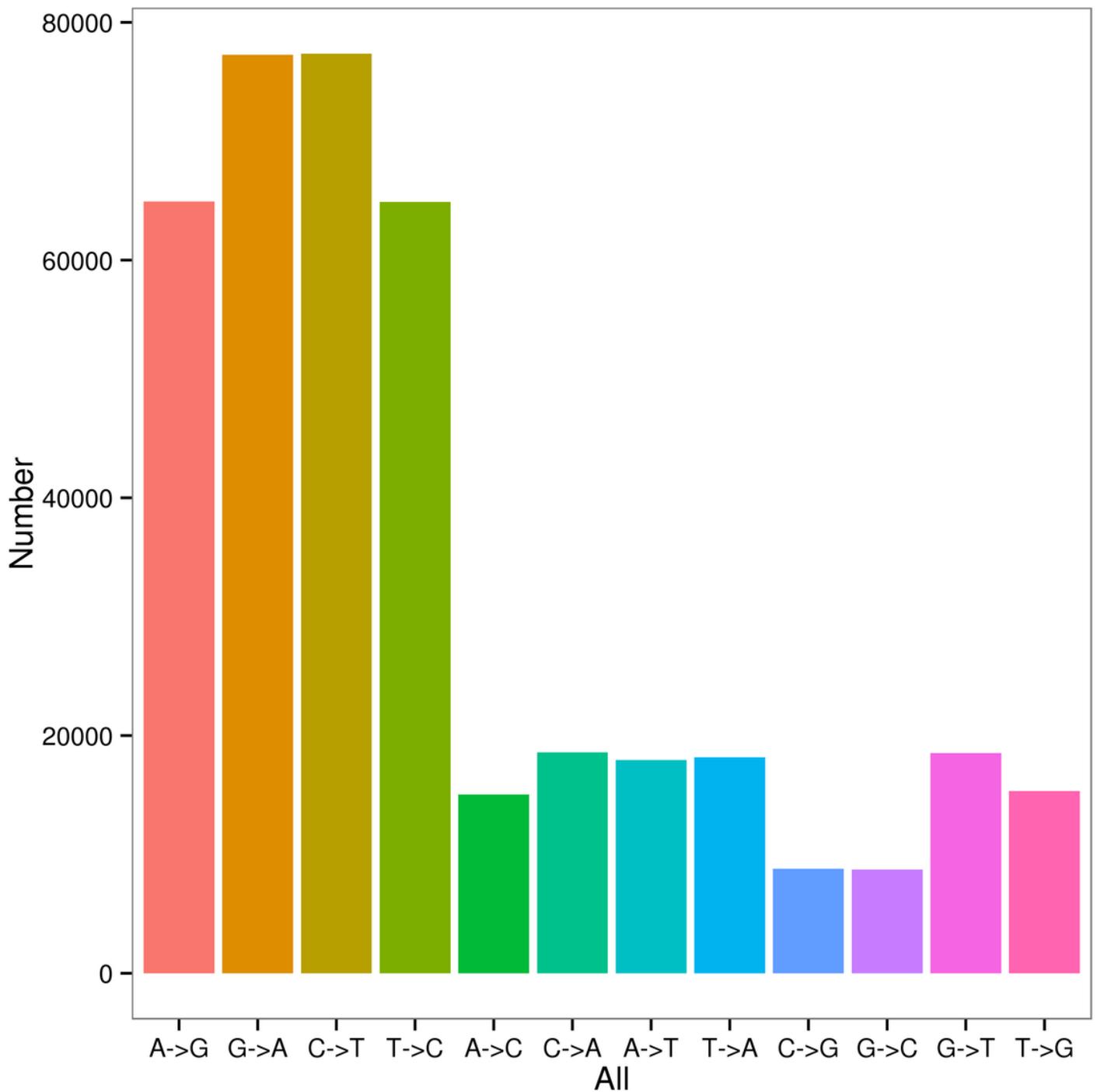


Figure 10

10 SNP mutation-type density distribution map Note: The horizontal axis presents the number of SNPs distributed per 1000 bp in the gene, and the vertical axis presents the number of genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [tableS3.xlsx](#)
- [tableS2.xlsx](#)
- [tableS1.xlsx](#)