RNA sequencing revealed the multi-stage transcriptome transformations during the development of gallbladder cancer associated with chronic inflammation

Sen Yang
Department of Hepatobiliary and Pancreatic Surgery, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Litao Qin
Medical Genetic Institute of Henan Province, Henan Provincial Key Laboratory of Genetic Diseases and Functional Genomics, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Pan Wu
Department of Hepatobiliary and Pancreatic Surgery, Henan Provincial People's Hospital, School of Clinical Medicine, Henan University

Yanbing Liu
Department of Hepatobiliary and Pancreatic Surgery, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Yanling Zhang
Department of Pathology, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Bing Mao
People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Yiyang Yan
Department of Hepatobiliary and Pancreatic Surgery, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Shuai Yan
Department of Hepatobiliary and Pancreatic Surgery, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Feilong Tan
Department of Hepatobiliary and Pancreatic Surgery, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Xueliang Yue
Department of Hepatobiliary and Pancreatic Surgery, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Hongshan Liu

Department of Hepatobiliary and Pancreatic Surgery, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

Huanzhou Xue (xhuanzhou@126.com)

Department of Hepatobiliary and Pancreatic Surgery, People's Hospital of Zhengzhou University, Henan Provincial People's Hospital

---

Research Article

Keywords: gallbladder carcinoma, RNA sequencing, inflammation, long non-coding RNA, immunity, metabolism

Posted Date: April 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1516146/v1

Abstract

Purpose

Gallbladder cancer (GBC) is a highly malignant tumor with extremely poor prognosis. Previous studies have suggested that the carcinogenesis and progression of GBC is a multi-stage and multi-step process, but most of them focused on the genome changes. The transcriptome changes, relating to every stage of GBC evolution, have rarely been studied.

Methods

We selected three cases of normal gallbladder, four cases of gallbladder with chronic inflammation induced by gallstones, five cases of early GBC, and five cases of advanced GBC, using next-generation RNA sequencing to reveal the changes in mRNAs and lncRNAs expression during the evolution of GBC.

Results

In-depth analysis of the sequencing data indicated that transcriptome changes from normal gallbladder to gallbladder with chronic inflammation were distinctly related to inflammation, lipid metabolism, and sex hormone metabolism; transcriptome changes from gallbladder with chronic inflammation to early GBC were distinctly related to immune activities and connection between cells; and the transcriptome changes from early GBC to advanced GBC were distinctly related to transmembrane transport of substances and migration of cells.

Conclusion

Expression profiles of mRNAs and lncRNAs change significantly during the evolution of GBC, in which lipid-based metabolic abnormalities play an important promotive role, inflammation and immune activities play a key role, and changes in membrane proteins are the most highlighted molecular changes.

1. Introduction

As a common malignant carcinoma of the biliary tree, gallbladder cancer (GBC) still has an extremely poor prognosis, with a median survival time of < 1 year. This is because of the low early diagnosis rate and high ability to invade and metastasize. Currently, effective chemotherapies and targeted drugs are lacking (Boutros et al. 2012, Hundal and Shaffer 2014).

Gallbladder stones are the most important risk factors for GBC. Gallbladder stones stimulate the wall of the gallbladder for a long time and cause chronic inflammation, eventually leading to carcinogenesis. This is the most commonly recognized carcinogenesis pathway in GBC(Espinoza et al. 2016). Similar to

other tumors, such as colon cancer, the formation and progression of GBC is a multi-stage and multi-step process with the accumulation of multiple changes in the genome and transcriptome (Wistuba and Gazdar 2004, Jain et al. 2014, Bizama et al. 2015). Regarding the molecular events in the process, previous studies have mostly focused on the changes in the genome, but there are few studies on the changes in the transcriptome(Srivastava et al. 2011, Li et al. 2014, Mhatre et al. 2017).

Non-coding RNAs, especially lncRNAs, have been a research hotspot in recent years. Except for an extremely small number of regions that encode mRNA, most parts of the human genome are still poorly understood, producing a large number of non-coding RNAs including microRNAs and lncRNAs(Stein 2004, Kapranov et al. 2007). lncRNAs are a type of non-coding RNA with a length of more than 200 nucleotides. It can perform physiological functions through various mechanisms, such as trans- and cis-regulation. An increasing number of studies has shown that lncRNAs are associated with a variety of diseases, including tumors (Perez et al. 2008, Gupta et al. 2010, Loewer et al. 2010). However, little is known about its role in carcinogenesis and progression of GBC (Khandelwal et al. 2017).

The improvement of the treatment effect of GBC depends on the development of more effective drugs, which requires further understanding of the molecular mechanism of carcinogenesis and progression of GBC. Therefore, using next-generation sequencing technology, we studied the expression profiles of mRNA and lncRNAs in the four stages of this process: normal gallbladder, gallbladder with chronic inflammation, early GBC, and advanced GBC. Through cluster analysis, we identified the highlighted molecular category changes during GBC development.

2. Materials And Methods

2.1. Case selection and sample processing

A number of gallbladder and GBC samples were collected from Henan Provincial People's Hospital in 2019 and 2020, following informed consent under a protocol approved by the Ethics Committee of Henan Provincial People's Hospital. Specimens were immediately frozen in liquid nitrogen after resection and stored at -80°C for long-term storage. Based on certain principles, we selected three cases of normal gallbladder (N8 N10 N20), four cases of gallbladder with chronic inflammation (Y8 Y12 Y13 Y16), five cases of GBC in the early stage (T5 T12 T13 T18 T31), and five cases of GBC in the advanced stage (T1 T19 T22 T27 T32), and performed transcriptome sequencing (Shanghai Biotechnology Company, Shanghai, China). The sample selection principles were as follows: normal gallbladder specimens were obtained from patients who underwent hepatectomy or pancreaticoduodenectomy without stones, polyps, obstructive jaundice, or cholangitis; the chronic inflamed gallbladder specimens were surgically removed from patients with calculous cholecystitis, excluding acute cholecystitis; GBC specimens should be adenocarcinoma pathologically. All specimens were pathologically confirmed. GBC samples were staged according to the AJCC 8th edition TNM staging method. Stages 0, Ⅰ, and Ⅱ were defined as early stage, and stages III–IV were considered as late stage. Clinicopathological data of the selected samples are shown in Tables S1 and S2.

2.2. RNA extraction and quality inspection

The TransZol Up Plus RNA Kit (Cat#ER501-01, Trans, Beijing, China) was used for total RNA extraction according to the manufacturer's instructions. Total RNA was purified using an RNAClean XP Kit (Cat A63987, Beckman Coulter, Inc. Kraemer Boulevard Brea, CA, USA) and RNase-Free DNase Set (Cat#79254, QIAGEN, GmBH, Germany) after passing quality inspection using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, US). Purified total RNA was subjected to quality inspection using a NanoDrop ND-2000 spectrophotometer and Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, US). Finally, only qualified total RNA was used for subsequent sequencing experiments.

2.3. Sequencing experiments

First, the purified total RNA was subjected to rRNA removal, fragmentation, first-strand cDNA synthesis, second-strand cDNA synthesis, end repair, 3'end addition, connector ligation, and enrichment to build a sequencing sample library following the experimental instructions. The concentration of the constructed library was detected using a Qubit® 2.0 Fluorometer, and the size of the library was detected using the Agilent 2100. The reagents used for library construction and quality inspection are listed in Table S3. The quality inspection results are presented in Table S4.

Then, cluster generation and first-direction sequencing primer hybridization were performed on the cBot equipped with the Illumina sequencer, following the cBot User Guide.

Finally, the flow cell with the cluster was placed on the sequencing machine using the prepared sequencing reagents, according to the Illumina User Guide. A paired-end program was used to perform paired-end sequencing. The sequencing process was controlled by data collection software provided by Illumina, and real-time data analysis was performed. The quality control standard for sequencing results was as follows: the amount of data was about 10G/sample, and the ratio of base quality in each direction greater than 20 (Q20) was not less than 85%. Sequencing quality was evaluated by the Q value, and the relationship between the Q value and sequencing error rate E value is

$$\mathrm{Q} = -10Log_{10}E$$

The sequencing quality of all samples was excellent and the base distribution was balanced. The quality control results are presented in Table S5.

2.4. RNA sequencing data analysis

The raw reads obtained by sequencing may contain unqualified reads with low end quality and sequencing primers. These unqualified reads may have a certain impact on the quality of the analysis; therefore, they must be filtered to obtain clean reads for data analysis. We used Seqtk (https://github.com/lh3/seqtk) to filter raw reads, according to the following procedure: 1. removal of the ligation sequence; 2. removal of the bases whose 3'end quality Q is less than 20; 3. removal of reads with

a length of less than 25 bp; 4. removal of the ribosome RNA reads from each species. The pre-processed statistics are presented in Table S6.

Genome mapping was performed on the pre-processed reads using a spliced mapping algorithm from Hisat2 (version:2.0.4) (Kim et al. 2015). The genome version used was GRCh38. The mapping process adopted the default parameters. The mapping results are listed in Table S7.

To make the gene expression levels of different genes and samples comparable, the reads were converted into FPKM (fragments per kilobase of exon model per million mapped reads) to standardize gene expression (Mortazavi et al. 2008). We first used Stringtie (version: 1.3.0) (Pertea et al. 2015, Pertea et al. 2016) to count the number of fragments of each gene after Hisat2 alignment, then used the trimmed mean of M values (TMM) method to normalize them (Robinson and Oshlack 2010), and finally calculated the FPKM value of each gene through a Perl script. The FPKM formula is as follows:

$$\mathrm{FPKM} = \frac{\mathrm{totalexonfragments}}{\mathrm{mappedreads(millions)} \times \mathrm{exonlength(KB)}}$$

Total exon fragments are the number of fragments aligned to the gene exon (fragment: a pair of reads); exon length is the total length of the gene exon; and mapped reads are the total number of reads aligned to the reference genome.

Differentially expressed genes were analyzed using edgeR (Robinson et al. 2010). The obtained p-value was subjected to multiple hypothesis tests, and the adjusted p-value was called the q-value. The p-value threshold was determined by controlling the false discovery rate. Furthermore, we calculated the multiple of differential expression based on the FPKM value, namely fold-change. The screening conditions for the differentially expressed genes were as follows: 1. q-value $\leq$ 0.05; 2. fold-change $\geq$ 2.

2.5. Function analysis for differentially expressed genes

Using GO (Gene Ontology) (http://www.geneontology.org/) analysis, the number of differentially expressed genes with the same function item was calculated. KEGG (Kyoto Encyclopedia of Genes and Genomes) (http://www.kegg.jp/) analysis was used to count the number of differentially expressed genes in each pathway. Furthermore, GO and KEGG enrichment analyses were performed to screen for significantly enriched GO and KEGG items from the differentially expressed genes. The calculation formula for the p-value is as follows:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

The calculation formula for rich factor was as follows: rich factor = (m/ n)/ (M/ N).

N is the number of genes with GO or KEGG annotation among all genes, n is the number of differentially expressed genes in N, M is the number of genes annotated as a specific GO or KEGG term among all genes, and m is the number of differentially expressed genes annotated as a specific GO or KEGG term. The q-value was obtained from the p-value obtained after the multiple hypothesis test. With q-value ≤ 0.05, the GO or KEGG terms that satisfied this condition were defined as significantly enriched in differentially expressed genes. The smaller the q-value, the more significant the enrichment. The greater the rich factor, the greater the degree of enrichment.

2.6. LncRNA analysis

The spliced results of Stringtie (version 1.3.0) were compared with the reference annotations using gffcompare (version 0.9.8), and new transcripts that failed to match the known annotations were obtained. Three types of transcripts (i.e., i, u, and x) were extracted for lncRNA prediction. The specific steps were as follows: step1: transcription length ≥ 200bp and exon ≥ 2; step2: predicted ORF < 300bp; step3: predict using Pfam (Sun et al. 2012), CPC (Kong et al. 2007), CNCI (Sun et al. 2013); and select the transcripts with CPC score < 0 and CNCI score < 0 and insignificant Pfam comparison as the potential lncRNAs; and step 4: compare with known lncRNAs and remove the same sequence. Remarks: i: a transfrag falling entirely within a reference intron; u: unknown, intergenic transcript; x: exonic overlap with reference on the opposite strand.

Expression quantification was performed for the predicted novel and known lncRNAs from the NONCODE and Ensembl database. The ID starting with MSTRG is a novel lncRNA, the ID starting with NON is the known lncRNA in the NONCODE database, and the ID starting with ENS is the known lncRNA in the Ensembl database.

Trans- and cis-regulation was used to predict target genes. The mRNA database of this species was used for trans-prediction. First, BLAST was used to select complementary or similar sequences, then RNAplex (Tafer and Hofacker 2008) was used to calculate the complementary energy between the two sequences, and finally, sequences above the threshold were selected. Genes whose distance from the lncRNA was less than 10 kb were selected as the target genes for cis-regulation.

2.7. Quantitative real-time PCR

To further verify the accuracy of the RNA sequencing experiment, quantitative real-time PCR was performed. Two differentially expressed mRNAs and two differentially expressed lncRNAs were selected for each comparison, and a total of 12 genes were determined for this test. The specimens used were the same as those used in the sequencing experiment.

RNA was extracted as described above. Quantitative real-time experiments were performed using Power SYBR Green PCR Master Mix (Cat#4368708, ABI, USA) according to the manufacturer's instructions. The primer sequences of the related genes are listed in Table S8. β-Actin was used as the reference gene. Each reaction was performed in triplicates. Relative expression of each gene was quantified using the gene's $2^{-\Delta Ct}$.

2.8. Statistical analysis

All statistical analyses were performed using SPSS for Windows, version 24.0. The analytical methods used in the sequencing experiments were described above. Quantitative real-time PCR results were compared between groups using an independent sample t-test. The expression levels of the genes in each group are shown as mean ± standard deviation. Statistical significance was set at $P \leq 0.05$.

3. Results

3.1 Overview of sequencing results and verification by quantitative real-time PCR

Compared with the human genome, the ratio of reads aligned to gene regions, coding regions, splice sites, introns, and non-coding regions was normal, genome coverage was good, and sequencing quantity was sufficient, as shown in Fig. S1.

A total of 84043 lncRNAs were detected, of which 1030 lncRNAs were newly predicted. By observing the differences in transcript length, number of exons, and expression levels between lncRNAs and mRNAs, it was shown that the lncRNAs conformed to the general characteristics, as shown in Fig. S2.

The quantitative real-time PCR results of the selected 12 genes were highly consistent with the sequencing results, suggesting that the sequencing experiment had high reliability. As shown in Fig. 1.

3.2 Transcriptome changes from normal gallbladder to gallbladder with chronic inflammation

A total of 851 different mRNAs were identified, of which 385 were upregulated and 466 were downregulated. There were 322 different lncRNAs, of which 103 were upregulated and 219 were downregulated. The expression of mRNAs and lncRNAs showed obvious differences between the two groups, and the expression of samples in the same group showed good homogeneity, as shown in Fig. 2.

GO enrichment of differentially expressed mRNAs

There were 759 GO items with q-value ≤ 0.05. By further summarizing the top 100 mRNAs with a larger rich factor, it was found that different mRNAs were distinctly related to inflammation (35 items), metabolism (24 items) including lipid metabolism (10 items) and sex hormone metabolism (four items). The top three enriched factors were estrogen 16-α-hydroxylase activity, lipid hydroxylation, and the omega-hydroxylase P450 pathway, as shown in Fig. 3.

KEGG enrichment of differentially expressed mRNAs

There were 28 KEGG items with q-value ≤ 0.05, which were distinctly related to inflammation (six items), lipid metabolism (three items), steroid hormones metabolism (three items), amino acid and foreign substance metabolism (seven items). The top three enriched factors were phenylalanine, tyrosine and tryptophan biosynthesis, synthesis and degradation of ketone bodies, and steroid hormone biosynthesis, as shown in Fig. 3.

GO and KEGG enrichment of differentially expressed lncRNAs

Target genes were predicted by trans- and cis-regulation. There were 877 predicted target genes for the differentially expressed lncRNAs, of which 59 showed significant differences in expression.

There were 0 GO items with q-value ≤ 0.05 for target genes, and there were two KEGG items with q-value ≤ 0.05, which were homologous recombination, valine, leucine, and isoleucine degradation.

GO and KEGG enrichment analyses was also performed for differentially expressed target genes. There were 28 GO items with q-value ≤ 0.05, which were distinctly related to inflammation (17 items) and foreign substance metabolism (four items). There were seven KEGG items with q-value ≤ 0.05, and 16 items with p-value ≤ 0.05, which were mostly related to inflammation (nine items), lipid metabolism (two items), and tumor-related pathways (three items), as shown in Fig. S3.

3.3 Transcriptome changes from gallbladder with chronic inflammation to early GBC

A total of 176 different mRNAs were identified, of which 58 were upregulated and 118 were downregulated. There were 84 different lncRNAs that were identified, of which 20 were upregulated and 60 were downregulated. The expression of mRNAs and lncRNAs showed obvious differences between the two groups, and the expression of samples in the same group showed good homogeneity, as shown in Fig. 4.

GO enrichment of differentially expressed mRNAs

There were 116 GO items with q-value ≤ 0.05. Further cluster analysis revealed that the different mRNAs were distinctly related to immune activity (63 items) and connection between cells (30 items). The top three enriched factors were regulation of B cell receptor signaling pathway, regulation of humoral immune response, and regulation of complement activation, as shown in Fig. 5.

KEGG enrichment of differentially expressed mRNAs

There were seven KEGG items with q-value ≤ 0.05, and 24 items with p-value ≤ 0.05, which were mostly related to metabolism (13 items) and immune activity (six items), as shown in Fig. 5.

GO and KEGG enrichment of differentially expressed lncRNAs

Target genes were predicted by trans- and cis-regulation. There were 54 predicted target genes for differentially expressed lncRNAs, of which six showed significant differences in expression.

There were 0 GO items with q-value ≤ 0.05 for target genes and 76 items with p-value ≤ 0.05, which were mostly related to the modification and polymerization of proteins (36 items), connection and signal transduction between cells (23 items), and immune activity (seven items). There were 0 KEGG items with q-value ≤ 0.05, and 17 items with p-value ≤ 0.05, which were mostly related to immune activity (eight items) and signal transduction (six items), as shown in Fig. S4.

GO and KEGG enrichment analyses were also performed for the differentially expressed target genes. There were 0 GO items with q-value ≤ 0.05, and five items with p-value ≤ 0.05, which were related to development (four items) and connection between cells (one items). There was 0 KEGG items with q-value ≤ 0.05, and 0 items with p-value ≤ 0.05.

3.4 Transcriptome changes from early GBC to advanced GBC

A total of 26 different mRNAs were identified, of which 20 were upregulated and six were downregulated. There were 18 different lncRNAs, of which seven were upregulated and 11 were downregulated. The expression of mRNAs and lncRNAs showed obvious differences between the two groups, and the expression of samples in the same group showed good homogeneity, as shown in Fig. 6.

GO enrichment of differentially expressed mRNAs

There were 11 GO items with q-value ≤ 0.05. Further cluster analysis revealed that different mRNAs were all related to the transmembrane transport of substances (11 items), including the transmembrane transport of carboxylic acids (three items), ions (six items), and phospholipids (two items). There were 25 items with p-value ≤ 0.05, which were distinctly related to transmembrane transport of substances (17 items), cell membrane components (three items), and cell migration (three items). The top three enriched factors were carboxylic acid transmembrane transporter activity, carboxylic acid transmembrane transport, and organic anion transmembrane transporter activity, as shown in Fig. 7.

KEGG enrichment of differentially expressed mRNAs

There was only 1 KEGG item with q-value ≤ 0.05 or p-value ≤ 0.05, that was bile secretion, as shown in Fig. 7.

GO and KEGG enrichment of differentially expressed lncRNAs

Target genes were predicted by trans- and cis-regulation. There were 14 predicted target genes for the differentially expressed lncRNAs, none of which showed significant differences in expression.

There were three GO items with q-value ≤ 0.05 for target genes, and 50 items with a p-value ≤ 0.05, which were mostly related to RNA expression regulation (19 items), cell proliferation (five items), and cell

migration (three items). There was only one KEGG item with q-value ≤ 0.05 or p-value ≤ 0.05, that was miRNAs in cancer, as shown in Fig. S5.

3. Discussion

Previous studies have suggested that the carcinogenesis and progression of GBC is a multi-stage and multi-step process, but most of them focused on the genome level. The transcriptome level, such as changes in the expression profiles of mRNAs and lncRNAs, have rarely been studied. To this end, we selected normal human gallbladder, chronically inflamed gallbladder, early GBC, and advanced GBC tissue samples; performed transcriptome sequencing of mRNAs and lncRNAs; and explored the expression profile transformations of GBC during the evolution of GBC. For differentially expressed genes, we performed GO and KEGG enrichment analyses. Generally, the adjusted q-value ≤ 0.05 was used as the significance threshold. If there were fewer corresponding items, then p-value ≤ 0.05 was used as the significance threshold instead, although it is not stricter than the q-value standard. For the significant items, we further classified them one by one to discover the underlying meaning behind the differentially expressed genes.

Analysis of sequencing data showed that the transcriptome differences between normal gallbladder and gallbladder with chronic inflammation were distinctly related to inflammation, lipid metabolism, and sex hormone metabolism; the transcriptome differences between gallbladder with chronic inflammation and early GBC were distinctly related to immune activities and connection between cells; the transcriptome differences between early and advanced GBC were distinctly related to transmembrane transport of substances and migration of cells.

Our study found that lncRNA transcription changed significantly during the formation and evolution of GBC, which is consistent with previous studies (Wu et al. 2017, Hu et al. 2019). Trans and cis regulation are important regulatory methods for lncRNAs (Guttman and Rinn 2012). We used this mechanism to predict the target genes of differentially expressed lncRNAs and then performed GO and KEGG analyses. These results were consistent with the mRNA analysis results in some aspects, but there were also differences. The reason may be that current research on lncRNAs is still in its infancy, and the functions of most lncRNAs are still unclear. Thus, it was not possible to perform functional analysis of differentially expressed lncRNAs directly; we could only use lncRNA target genes for indirect analysis, which was not rigorous in fact.

It is believed that gallbladder stones are the most important cause of GBC, and obesity, metabolic syndrome, and sex are also important risk factors for GBC(Wistuba and Gazdar 2004, Espinoza et al. 2016). We found that the transcriptome differences between the inflammatory gallbladder and the normal gallbladder were distinctly related to inflammation, lipid metabolism, and sex hormone metabolism, which is consistent with previous studies. However, it was not clear whether the changes in metabolism-related genes that were mainly related to lipid metabolism were secondary changes after the formation of stones or whether such populations had changes in these metabolic genes, which led to the

formation of stones. After all, obesity and metabolic syndrome are also risk factors for gallstone formation(Wistuba and Gazdar 2004). In addition, the metabolism of sex and other steroid hormones is a type of lipid metabolism. It is not clear whether the changes in metabolism-related genes contributed to changes in the levels of sex and other steroid hormones, or whether changes in sex and other steroids hormones affected the lipid-based metabolic changes. It may be a causal relationship, or it may be a kind of synergistic effect.

We found that changes related to inflammation were important in gallbladders with chronic inflammation, and changes related to immune activity were important in early GBC. Inflammation and immune activity are closely related and share many similarities in many ways. Therefore, it can be said that inflammation and immune activity play a key role in the formation of GBC. In addition, the connection between cells is an important molecular event in the process of inflammatory and immune activity. Membrane proteins are the most important participants in communication between cells and transportation of substances across the membrane. Therefore, it can be said that membrane proteins are the most highlighted molecular changes in the formation and evolution of GBC.

To elucidate the molecular mechanism of GBC carcinogenesis and progression, we propose the following hypotheses. The metabolic changes mainly related to lipid induce gallbladder stones, which stimulate the gallbladder wall for a long time, cause damage to the gallbladder mucosa, and lead to chronic inflammation of the gallbladder wall. Chronic inflammation further induces the transformation of genes related to immune activity and connection between cells, leading to malignant proliferation of the gallbladder mucosa and evasion of the body's immune surveillance. Further changes in membrane proteins mainly related to substance transportation, lead to changes in the internal and external environment of cells and changes in the nature of cell migration, which promotes cancer cells to spread far away, especially through lymphatic metastasis. Inflammation plays a key role in these processes, and changes in membrane proteins are the most distinct molecular changes, as shown in Fig. 8.

Our research highlights the roles of metabolism, inflammation, immunity, and membrane proteins in GBC development. However, it only provides an overview landscape; the specific detailed molecular mechanisms still require further study, which will help the development of targeted drugs and improve the prognosis of GBC.

Declarations

Acknowledgements: We thank Shengxian Yuan (Eastern Hepatobiliary Surgery Hospital, Shanghai, China) for giving us some useful advises on the study. We would like to thank Editage (www.editage.cn) for English language editing.

Funding

This work was supported by Henan Province Medical Science and Technology Research Plan (SB201901079).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Sen Yang designed and organized the study, analyzed the data, and wrote the manuscript. Litao Qin contributed to sequencing data analysis, statistical processing and manuscript writing. Pan Wu performed experiments and participated in the acquisition of samples. Yanbing Liu collected the clinicopathological data of the patients. Yanling Zhang contributed to analyze the pathological data of the patients. Bing Mao contributed to figures and tables processing. Yiyang Yan, Shuai Yan and Feilong Tan contributed to samples collection. Xueliang Yue and Hongshan Liu commented on and revised the manuscript. Huanzhou Xue supervised the study, and revised the manuscript. All authors read and approved the final manuscript.

References

1. Bizama, C., P. Garcia, J. A. Espinoza, H. Weber, P. Leal, B. Nervi and J. C. Roa (2015). Targeting specific molecular pathways holds promise for advanced gallbladder cancer therapy. Cancer Treat Rev 41: 222-234. https://doi.org/10.1016/j.ctrv.2015.01.003
2. Boutros, C., M. Gary, K. Baldwin and P. Somasundar (2012). Gallbladder cancer: past, present and an uncertain future. Surg Oncol 21: e183-191. https://doi.org/10.1016/j.suronc.2012.08.002
3. Espinoza, J. A., C. Bizama, P. Garcia, C. Ferreccio, M. Javle, J. F. Miquel, J. Koshiol and J. C. Roa (2016). The inflammatory inception of gallbladder cancer. Biochim Biophys Acta 1865: 245-254. https://doi.org/10.1016/j.bbcan.2016.03.004
4. Gupta, R. A., N. Shah, K. C. Wang, J. Kim, H. M. Horlings, D. J. Wong, M. C. Tsai, T. Hung, P. Argani, J. L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R. B. West, M. J. van de Vijver, S. Sukumar and H. Y. Chang (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464: 1071-1076. https://doi.org/10.1038/nature08975
5. Guttman, M. and J. L. Rinn (2012). Modular regulatory principles of large non-coding RNAs. Nature 482: 339-346. https://doi.org/ 10.1038/nature10887
6. Hu, Y. P., Y. P. Jin, X. S. Wu, Y. Yang, Y. S. Li, H. F. Li, S. S. Xiang, X. L. Song, L. Jiang, Y. J. Zhang, W. Huang, S. L. Chen, F. T. Liu, C. Chen, Q. Zhu, H. Z. Chen, R. Shao and Y. B. Liu (2019). LncRNA-HGBC stabilized by HuR promotes gallbladder cancer progression by regulating miR-502-3p/SET/AKT axis. Mol Cancer 18: 167.https://doi.org/10.1186/s12943-019-1097-9
7. Hundal, R. and E. A. Shaffer (2014). Gallbladder cancer: epidemiology and outcome. Clin Epidemiol 6: 99-109. https://doi.org/10.2147/CLEP.S37357
8. Jain, K., T. Mohapatra, P. Das, M. C. Misra, S. D. Gupta, M. Ghosh, M. Kabra, V. K. Bansal, S. Kumar, V. Sreenivas and P. K. Garg (2014). Sequential occurrence of preneoplastic lesions and accumulation of

loss of heterozygosity in patients with gallbladder stones suggest causal association with gallbladder cancer. Ann Surg 260: 1073-080. https://doi.org/10.1097/SLA.0000000000000495

9. Kapranov, P., J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermuller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana and T. R. Gingeras (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316: 1484-1488. https://doi.org/10.1126/science.1138341
10. Khandelwal, A., A. Malhotra, M. Jain, K. M. Vasquez and A. Jain (2017). The emerging role of long non-coding RNA in gallbladder cancer pathogenesis. Biochimie 132: 152-160. https://doi.org/10.1016/j.biochi.2016.11.007
11. Kim, D., B. Langmead and S. L. Salzberg (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12: 357-360. https://doi.org/10.1038/nmeth.3317
12. Kong, L., Y. Zhang, Z. Q. Ye, X. Q. Liu, S. Q. Zhao, L. Wei and G. Gao (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector ma chine. Nucleic Acids Res 35: W345-349. https://doi.org/10.1093/nar/gkm391
13. Li, M., Z. Zhang, X. Li, J. Ye, X. Wu, Z. Tan, C. Liu, B. Shen, X. A. Wang, W. Wu, D. Zhou, D. Zhang, T. Wang, B. Liu, K. Qu, Q. Ding, H. Weng, Q. Ding, J. Mu, Y. Shu, R. Bao, Y. Cao, P. Chen, T. Liu, L. Jiang, Y. Hu, P. Dong, J. Gu, W. Lu, W. Shi, J. Lu, W. Gong, Z. Tang, Y. Zhang, X. Wang, Y. E. Chin, X. Weng, H. Zhang, W. Tang, Y. Zheng, L. He, H. Wang, Y. Liu and Y. Liu (2014). Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. Nat Genet 46: 872-876. https://doi.org/10.1038/ng.3030
14. Loewer, S., M. N. Cabili, M. Guttman, Y. H. Loh, K. Thomas, I. H. Park, M. Garber, M. Curran, T. Onder, S. Agarwal, P. D. Manos, S. Datta, E. S. Lander, T. M. Schlaeger, G. Q. Daley and J. L. Rinn (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet 42: 1113-1117. https://doi.org/10.1038/ng.710
15. Mhatre, S., Z. Wang, R. Nagrani, R. Badwe, S. Chiplunkar, B. Mittal, S. Yadav, H. Zhang, C. C. Chung, P. Patil, S. Chanock, R. Dikshit, N. Chatterjee and P. Rajaraman (2017). Common genetic variation and risk of gallbladder cancer in India: a case-control genome-wide association study. The Lancet Oncology 18: 535-544. https://doi.org/10.1016/s1470-2045(17)30167-5
16. Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and B. Wold (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5: 621-628. https://doi.org/10.1038/nmeth.1226
17. Perez, D. S., T. R. Hoage, J. R. Pritchett, A. L. Ducharme-Smith, M. L. Halling, S. C. Ganapathiraju, P. S. Streng and D. I. Smith (2008). Long, abundantly expressed non-coding transcripts are altered in cancer. Hum Mol Genet 17: 642-655. https://doi.org/10.1093/hmg/ddm336
18. Pertea, M., D. Kim, G. M. Pertea, J. T. Leek and S. L. Salzberg (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11: 1650-1667. https://doi.org/ 10.1038/nprot.2016.095

19. Pertea, M., G. M. Pertea, C. M. Antonescu, T. C. Chang, J. T. Mendell and S. L. Salzberg (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33: 290-295. https://doi.org/ 10.1038/nbt.3122
20. Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140. https://doi.org/10.1093/bioinformatics/btp616
21. Robinson, M. D. and A. Oshlack (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11(3): R25. https://doi.org/10.1186/gb-2010-11-3-r25
22. Srivastava, K., A. Srivastava, K. L. Sharma and B. Mittal (2011). Candidate gene studies in gallbladder cancer: a systematic review and meta-analysis. Mutat Res 728: 67-79. https://doi.org/10.1016/j.mrrev.2011.06.002
23. Stein, L. D. (2004). Human genome: end of the beginning. Nature 431: 915-916. https://doi.org/10.1038/431915a
24. Sun, L., H. Luo, D. Bu, G. Zhao, K. Yu, C. Zhang, Y. Liu, R. Chen and Y. Zhao (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res 41: e166. https://doi.org/10.1093/nar/gkt646
25. Sun, L., Z. Zhang, T. L. Bailey, A. C. Perkins, M. R. Tallack, Z. Xu and H. Liu (2012). Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse Klf1 knockout study. BMC Bioinformatics 13: 331. https://doi.org/10.1186/1471-2105-13-331
26. Tafer, H. and I. L. Hofacker (2008). RNAplex: a fast tool for RNA-RNA interaction search. Bioinformatics 24: 2657-2663. https://doi.org/10.1093/bioinformatics/btn193
27. Wistuba, II and A. F. Gazdar (2004). Gallbladder cancer: lessons from a rare tumour. Nat Rev Cancer 4: 695-706. https://doi.org/10.1038/nrc1429
28. Wu, X. S., F. Wang, H. F. Li, Y. P. Hu, L. Jiang, F. Zhang, M. L. Li, X. A. Wang, Y. P. Jin, Y. J. Zhang, W. Lu, W. G. Wu, Y. J. Shu, H. Weng, Y. Cao, R. F. Bao, H. B. Liang, Z. Wang, Y. C. Zhang, W. Gong, L. Zheng, S. H. Sun and Y. B. Liu (2017). LncRNA-PAGBC acts as a microRNA sponge and promotes gallbladder tumorigenesis. EMBO Rep 18(10): 1837-1853. https://doi.org/10.15252/embr.201744147

Figures

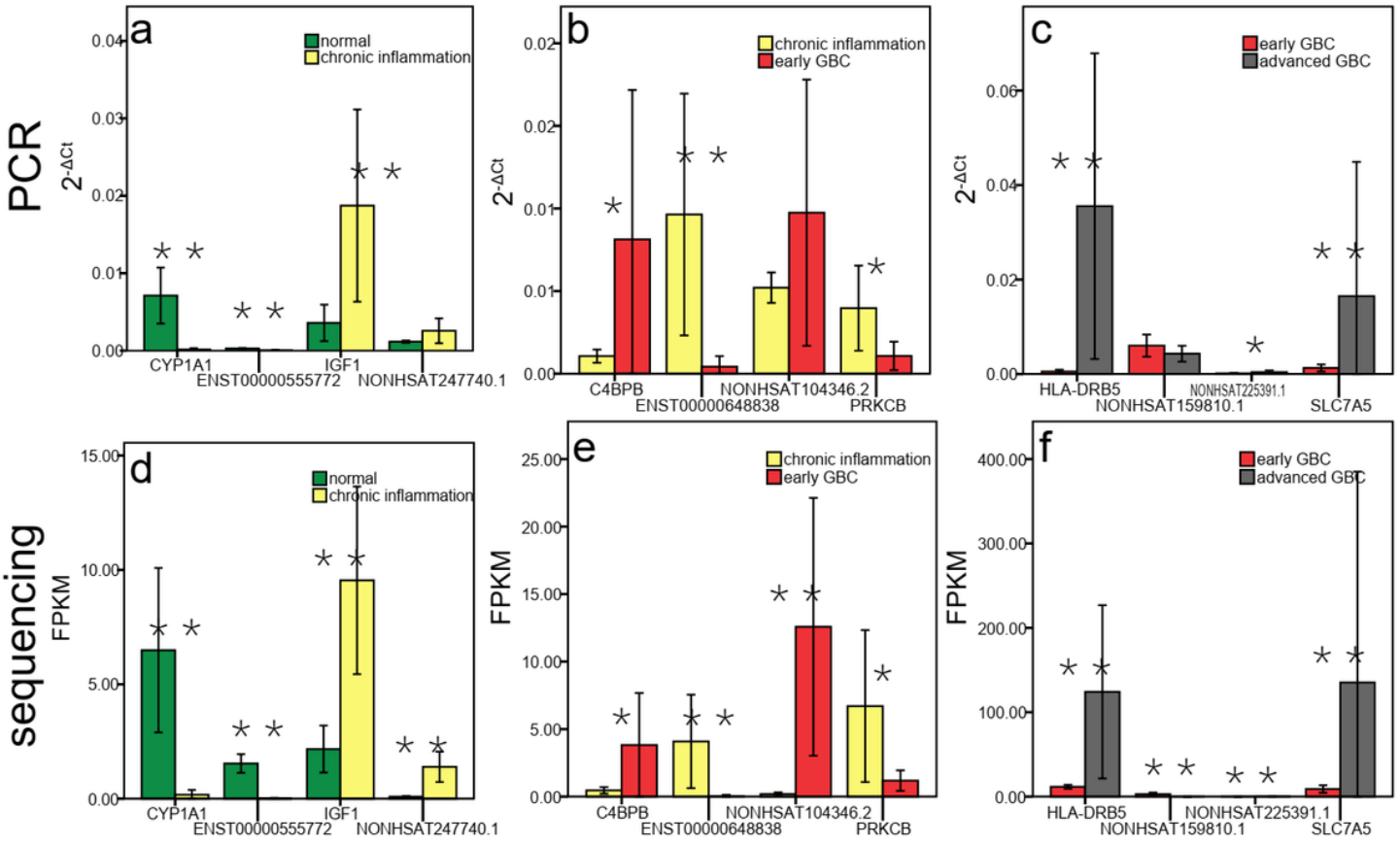

Figure 1

Verification of quantitative real-time PCR experiment

The first row of three graphs are the results of the PCR experiment. The X axis represents each gene, and the Y axis represents the relative expression of each gene in the PCR experiment, represented as $2^{-\Delta Ct}$ ± mean ± standard deviation. The second row of three graphs are the results of RNA sequencing experiments. The X axis represents each gene, and the Y axis represents the relative expression of each gene in the sequencing experiment, represented as FPKM±mean ± standard deviation. a and d are the results of the four genes selected in the comparison between normal gallbladder and chronic inflammation gallbladder. b and e are the results of the four genes selected in the comparison between chronic inflammation gallbladder and early GBC. c and f are the results of the four genes selected in the comparison between early GBC and advanced GBC. *P≤0.05 **P≤0.01

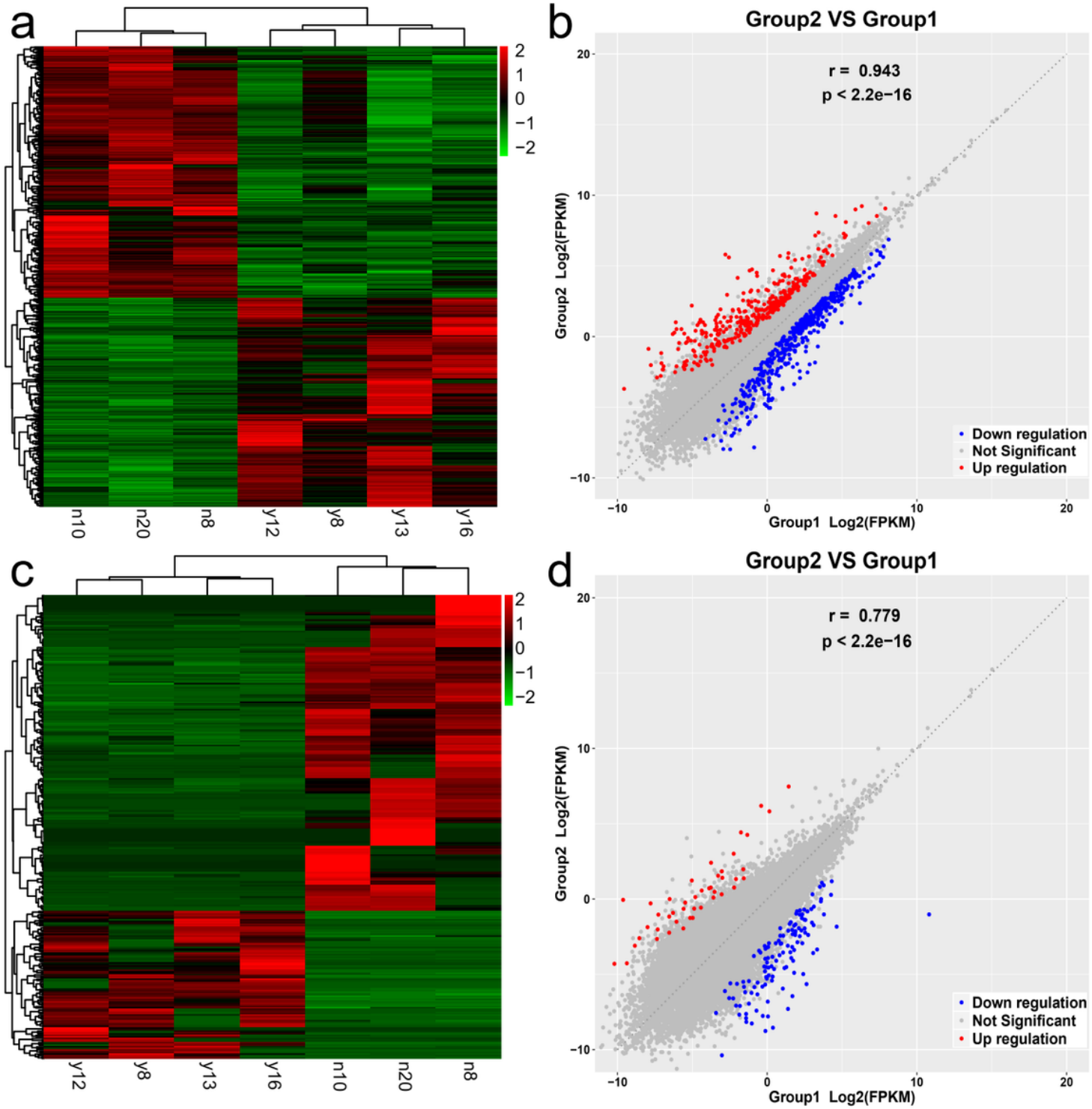

Figure 2

Expression differences of mRNAs and lncRNAs between normal gallbladders and gallbladders with chronic inflammation

Group 1 is normal gallbladder including N8 N10 N20; Group 2 is gallbladder with chronic inflammation including Y8 Y12 Y13 Y16. a The heatmap figure of mRNA expression between the two groups. The deeper the red, the higher the expression, and the darker the green, the lower the expression. b The

correlation scatter diagram of mRNA expression between the two groups, the red dots are the upregulated mRNAs of gallbladder with chronic inflammation relative to the normal gallbladder, the green dots are the downregulated mRNAs, and the gray dots indicate the differences are not significant. c The heatmap figure of lncRNA expression between the two groups. d The correlation scatter diagram of lncRNA expression between the two groups.

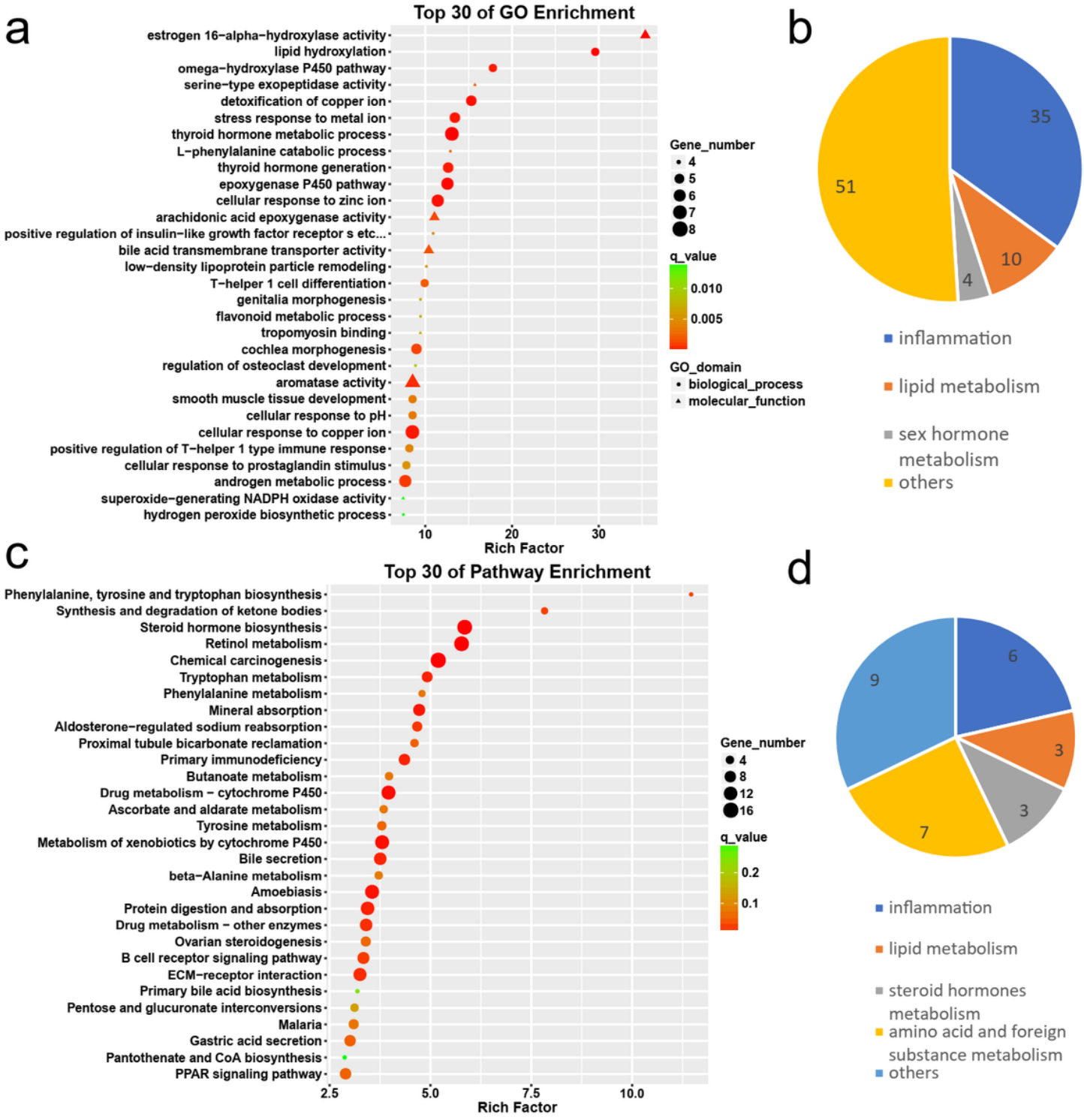

Figure 3

GO and KEGG enrichment analysis of differentially expressed mRNAs between normal gallbladders and gallbladders with chronic inflammation

a The top 30 GO items with a high degree of enrichment, the shapes of icons represent different GO categories, the size represents the number of gene items, the color depth represents the size of q-value and the X axis indicates the value of the rich factor. b The top 100 GO items with a larger enrichment factor of q-value ≤ 0.05 is further classified, the number on the disc represents the number of GO items corresponding to the category. c The top 30 KEGG entries with a high degree of enrichment. d The 28 KEGG items with q-value ≤ 0.05 are further classified, and the number on the disc represents the number of KEGG items corresponding to the category.

Figure 4

Expression differences of mRNAs and lncRNAs between gallbladders with chronic inflammation and early GBC

Group 2 is gallbladder with chronic inflammation including Y8 Y12 Y13 Y16; group 4 is early GBC including T5 T12 T13 T18 T31. a The heatmap figur of mRNA expression between the two groups. The deeper the red, the higher the expression, and the darker the green, the lower the expression. b The correlation scatter diagram of mRNA expression between the two groups, the red dots are the upregulated mRNAs of gallbladder with early GBC relative to gallbladder with chronic inflammation, the green dots are the downregulated mRNAs, and the gray dots indicate the differences are not significant. c The heatmap figure of lncRNA expression between the two groups. d The correlation scatter diagram of lncRNA expression between the two groups.

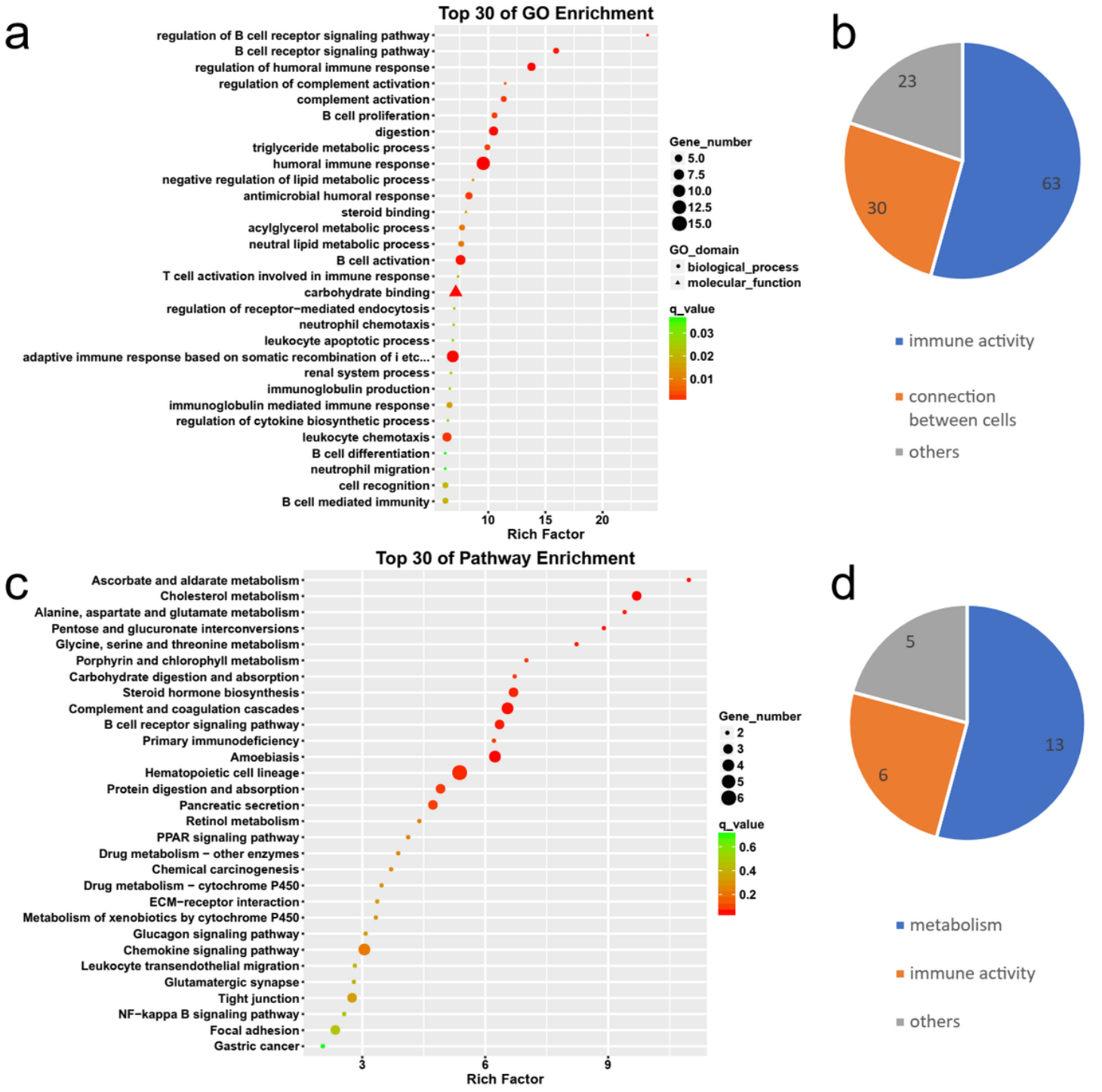

Figure 5

GO and KEGG enrichment analysis of differentially expressed mRNAs between gallbladders with chronic inflammation and early GBC

a The top 30 GO items with a high degree of enrichment, the shapes of icons represent different GO categories, the size represents the number of gene items, the color depth represents the size of the q-value and the X axis indicates the value of the rich factor. b The 116 GO items with q-value ≤ 0.05 is further classified, the number on the disc represents the number of GO items corresponding to the

category. c The top 30 KEGG items with a high degree of enrichment. d The 24 KEGG items with p-value ≤ 0.05 is further classified, and the number on the disc represents the number of KEGG items corresponding to the category.

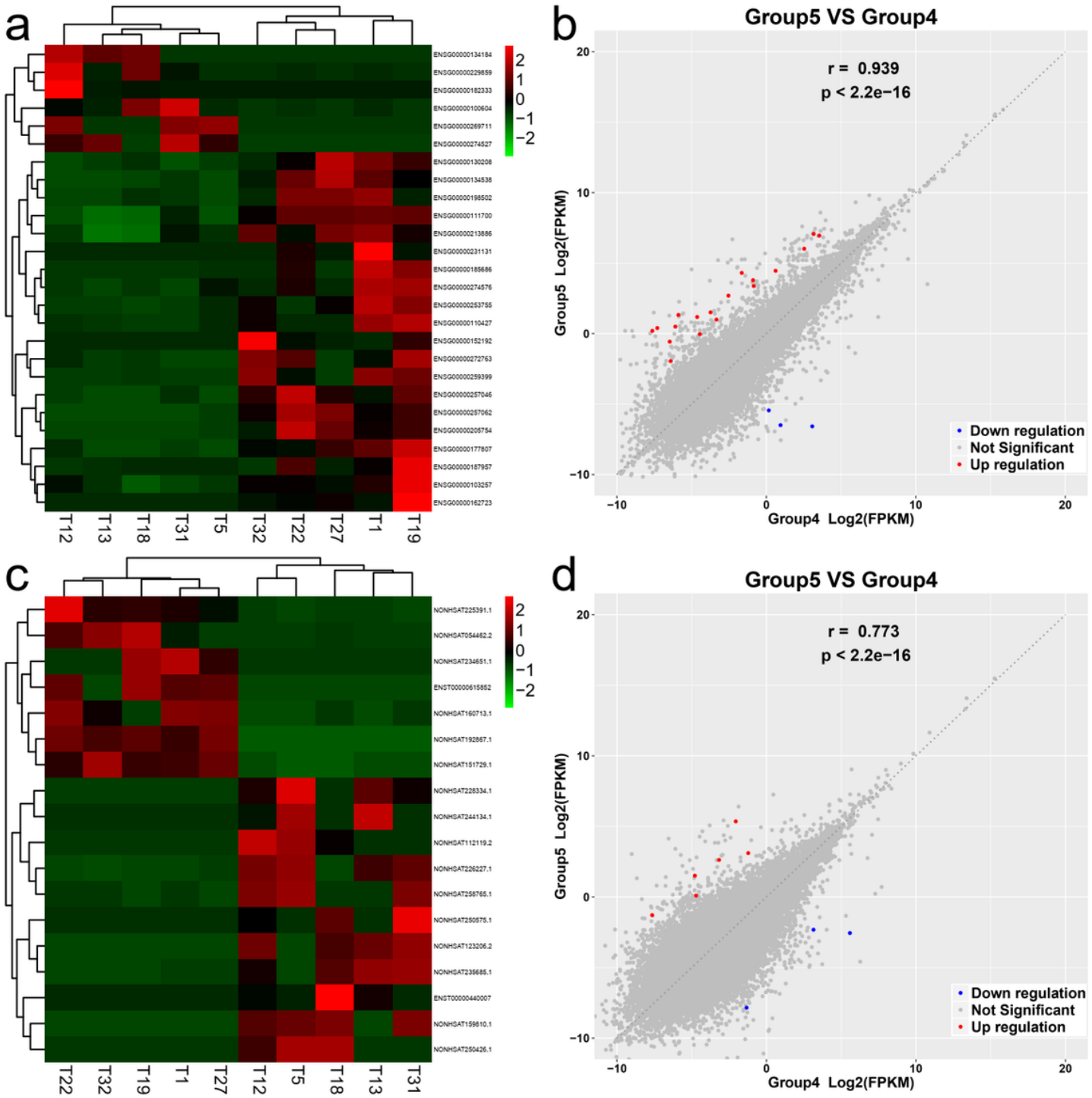

Figure 6

Expression differences of mRNAs and lncRNAs between early GBC and advanced GBC

Group 4 is early GBC including T5 T12 T13 T18 T31. Group 5 is advanced GBC including T1 T19 T22 T27 T32. a The heatmap figure of mRNA expression between the two groups. The deeper the red, the higher the expression, and the darker the green, the lower the expression. b The correlation scatter diagram of mRNA expression between the two groups, the red dots are the upregulated mRNAs of advanced GBC relative to early GBC, the green dots are the downregulated mRNAs, and the gray dots indicate the differences are not significant. c The heatmap figure of lncRNA expression between the two groups. d The correlation scatter diagram of lncRNA expression between the two groups.

Figure 7

GO and KEGG enrichment analysis of differentially expressed mRNAs between early GBC and advanced GBC

a The top 30 GO items with a high degree of enrichment, the shapes of icons represent different GO categories, the size represents the number of gene items, the color depth represents the size of the q-value and the X axis indicates the value of rich factor. b The 11 GO items with q-value ≤ 0.05 is further classified, the number on the disc represents the number of GO items corresponding to the category. c The top 30 KEGG items with a high degree of enrichment.

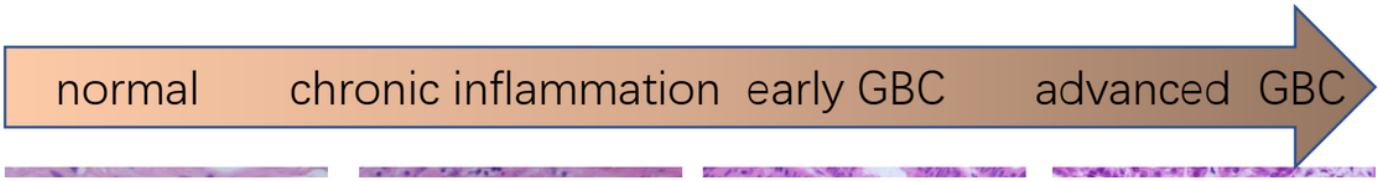

Figure 8

Schematic diagram of the multi-stage development of GBC

The top arrow indicates the four stages in the development of GBC. The four pictures in the middle were taken from typical pictures of each stage from our specimens. The text below indicates the prominent gene expression changes during the phase transition.

Supplementary Files

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

- SupplementaryInformation.pdf