Screening and Role of circRNAs in Vulvar Lichen Sclerosus

min yang

Beijing Hospital

Kailu Sun

Beijing Hospital

Jianmin Chang (✉ 472520208@qq.com)

Beijing Hospital

Research

Keywords: Vulvar Lichen Sclerosus, circRNA, mRNA, lncRNA

Posted Date: September 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-789590/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective—The aim of this study was to investigate the expression profile of circRNA in Vulvar Lichen Sclerosus, and to identify the underlying core genes of Vulvar Lichen Sclerosus.

Methods—rRNA removal was used for sequencing, and mRNA, lncRNA and circRNA differentially expressed between 20 groups of VLS tissues and 20 groups of healthy female vulva skin tissues were screened. Bioinformatics analysis was used to analyze its potential function.

Results—A total of 1545 differentially expressed mRNAs were assessed in VLS patients, of which 1541 were upregulated and 1004 were down-regulated; A total of 1453 differentially expressed lncRNAs were assessed, of which 812 were up-regulated and 641 were down-regulated. A total of 79 differentially expressed circRNAs were assessed, of which 54 were upregulated and 25 were down-regulated. The differential expression of CircRNA is closely related to biological processes and molecular functions. The differences in circRNA were mainly related to the "human T-cell leukemia virus 1 infection" signaling pathway and the "axon guidance" signaling pathway.

Conclusion—This study identified the profile of abnormal regulation of circRNA in VLS. Biological informatics analysis showed that the dysregulation of circRNAs could be related to the pathogenesis and pathological process of VLS.

Background

Vulvar Lichen Sclerosus (VLS) was one of the most common clinical manifestations of vulva. Thirteen percent of women had symptomatic vulvar diseases [1]. Prior to the International Association for the Study of Vulvovaginal Diseases (ISSVD) classification in 1975, vague terms were used for vulva lichen sclerosis, such as leukoplakia, nodules, and vulva dystrophy [2]. The current classification of ISSVD included the disease entity with exogenous skin disease, which was non-neoplastic and non-infectious in nature [3]. They had the potential to transform into vulvar intraepithelial neoplasia (VIN) and keratinized vulvar carcinoma [4]. VLS is a chronic inflammation, and its pathogenic mechanism is still unclear. However, studies have speculated that there are three pathophysiological mechanisms of VLS: infectious disease etiology, primary immune disorders and isotraumatopic reactions [5]. Clinicopathological studies of VLS indicate that the progression of VLS is related to the density of inflammatory infiltration and the number of pathogens [6].

There were many treatments available for VLS, such as Local steroid therapy, intra-lesional steroid therapy, topical calcineurin inhibitors and so on [7]. However, there were relatively few studies on VLS gene expression profile. Therefore, it was an urgent task to identify molecular biomarkers and explore the potential mechanisms of VLS, which may contribute to the development of novel diagnostic and therapeutic approaches for VLS [8, 9]. In recent years, bioinformatics methods had been widely used to analyze microarray data in order to identify differentially expressed genes (DEG) and perform various analyses [10]. In our study, two groups of samples were sequenced, and VLS-related circRNAs were

screened and predicted by gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment.

Covalently closed, single-stranded circular RNA (circRNA) was a special kind of RNA. CircRNA has no free 5' cap structure and 3' poly (A) structure, and was insensitive to nucleases, so it was more stable than linear RNA[11]. CircRNAs were mainly localized in the cytoplasm and could be stored in exosomes [12, 13]. As early as 20 years ago, scientists had identified circRNAs in plant viroids, yeast mitochondrial RNA and hepatitis B virus. Subsequently, endogenous circRNAs had been identified in human ETS – 1 gene and mouse SRY gene, but only a small number of circRNAs had been identified at that time[14, 15]. Moreover, it was regarded as a by-product of abnormal shear without regulatory potential [16]. In recent years, studies had found that circRNA derived from human INK4A /ARF and CDR1 genes has an influence on human atherosclerosis and was involved in the regulation of mRNA expression, providing a new dawn for circRNA research [17]. With the rapid development of high-throughput sequencing and bioinformatics analysis, tens of thousands of circRNAs had been identified in cells and tissues of different species, most of which were conserved and stable across different species, and their expression was cell, tissue and developmental stage specific[18, 19]. Although a large number of circRNAs had been identified, the research on the function of circRNAs and the mechanism of circRNA formation has only just begun. The types of circRNAs could be classified into the following four types according to their sources: ◻ Full exon type circRNAs; ◻ EicircRNAs in intron-exon combinations; ◻ lasso-type circRNA composed of introns; ◻ CircRNAs produced by cyclization of viral RNA genomes, tRNA, rRNA, snRNA, etc.[20–22]. To date, little research has been done on the relationship between circRNA and VLS[23]. The correlation between the significant changes in the expression of circRNAs and the pathogenicity of VLS had not been clearly studied. The purpose of this study was to mine the differential expression of circRNAs in VLS and determine the potential role of selected genes in the circRNAs etiology of VLS. In addition, this article conducted further research on the potential impact of circRNAs on VLS.

Results

mRNA expression profile of VLS patients mRNA

Sequencing was performed in three samples of vulva lichen scleroid tissue and three samples of normal tissue. Through the analysis, 262,922 transcripts were identified, corresponding to 73,150 genes. FPKM (Fragments Per Kilobase of exon model Per Million mapped reads) was used to measure the abundance of gene expression, and the expression abundance of known genes in different samples was counted by FPKM. FPKM denotes the number of sequence fragments per thousand transcriptome per million sequence bases, and FPKM value denotes the gene expression level. The results are shown in Table 1. The expression values in the above table were distributed statistically, and the expression levels of VLS genes were understood from the overall level by using the FPKM box chart of the samples, and the reproducibility of the designed samples was preliminarily judged (Fig. 1A). The respective expression levels of different samples were processed to compare the changes of mRNA expression trends among different samples. The difference expression of mRNA was displayed with density distribution diagram

(Fig. 1B). According to the length distribution of the ORF region of mRNA, the statistical histogram was performed (Fig. 1C).

The mRNA expression was differentially expressed in VLS patients

Based on screening threshold, the figure in the volcano identified a total of 1545 differentially expressed mRNA, $|\text{LOGFC}| > 1$, $P \text{ values} < 0.05$, of which 1541 were raised, lowered 1004 (Fig. 2A). Table 2 listed the ten mRNAs with the largest expression differences (up-regulated and down-regulated). It showed a statistical histogram based on the down-regulation frequency of significantly differentially expressed genes in Fig. 2B. Cluster analysis was performed in the VLS chain specific library. In order to better intuitively reflect the clustering expression pattern, we used $\text{Log}_{10}(\text{FPKM}+1)$ to process the gene expression (Fig. 2C).

Pathway enrichment analysis of differentially expressed genes

The GO function enrichment analysis and KEGG pathway enrichment analysis of mRNA host genes showed significant difference ($P < 0.05$). A total of 902 GO BPs (biological processes), 109 GO CC (cellular components), 235 GO MF (molecular functions), 1246 functional and KEGG pathways were enriched. Differentially expressed mRNAs were involved in signal transduction and protein binding in the plasma membrane (Fig. 3A). The GO enrichment analysis results were statistically analyzed by GGplot2, and the results were presented in the form of scatter plots (Fig. 3B). 'Cytokine-cytokine receptor interactions' and 'chemokine signaling pathways' are significant pathways (Fig. 3C).

Expression profile of lncRNA in VLS patients

We removed known mRNAs and transcripts less than 200bp, and then used CPC (Coding Potential Calculator) and CNCI (Coding non-coding Index) software to predict the remaining transcripts by lncRNA. If these remaining transcripts had the potential to encode proteins, we classified them as NOVEL mRNAs and then defined them as mRNA and filtered them. The CPC and CNCI scores of lncRNAs in each sample were statistically analyzed, and the results were expressed in box diagram (Fig. 4A and B) for genomic visualization of lncRNA levels in different samples (Fig. 4C). We made statistics on the percentages of different class codes of lncRNAs in each sample, and the results were shown in a pie chart (Fig. 4D).

Differential expression of lncRNA in VLS patients

Based on screening threshold, the figure in the volcano identified a total of 1453 differentially expressed lncRNA, $|\text{LOGFC}| > 1$, $P \text{ values} < 0.05$, of which 812 were raised, lowered 641 (figure 2 a). Table 3 lists ten lncRNAs with the largest expression differences (up-regulated and down-regulated). Fig. 2B shows a statistical histogram based on the down-regulation frequency of significantly differentially expressed genes. Cluster analysis was performed in the VLS chain specific library. In order to better intuitively reflect the clustering expression pattern, we used $\text{Log}_{10}(\text{FPKM}+1)$ to process the gene expression (Fig. 2C).

circRNA expression profile in VLS patients

The FPKM box chart of the samples was used to understand the gene expression level of VLS from the overall level and preliminarily judge the repeatability of the designed samples (Fig. 1A). The respective expression levels of different samples were processed to compare the changes in the expression trend of circRNA among different samples. We had analyzed the expression distribution in circRNAs, and the results were displayed with a density distribution diagram (Fig. 1B). The pie chart analysis was performed according to the number distribution of circRNA types (Fig. 1c).

CircRNA was differentially expressed in VLS patients

Based on screening threshold, the figure in the volcano identified a total of 79 differentially expressed circRNA, $|\text{LOGFC}| > 1$, P values < 0.05 , of which 54 were raised, lowered 25 (FIG. 7 a). We summarized the 10 circRNAs with the most significant expression differences (5 up-regulated and 5 down-regulated) in Table 4. Fig. 7B shows a statistical histogram based on the down-regulation frequency of significantly differentially expressed genes. Cluster analysis was performed in the VLS chain specific library. In order to better intuitively reflect the clustering expression pattern, we used $\text{Log}_{10}(\text{FPKM}+1)$ to process the gene expression (Fig. 7C).

Pathway enrichment analysis of differentially expressed genes

The enrichment analysis between GO function and KEGG pathway of circRNAs host genes showed significant differences ($P < 0.05$). After research, we found that the differentially expressed circRNAs in VLS enriched a total of 229 GO-BPs, 22 GO-CCs and 22 GO-MFs. In addition, we also found that 273 functions were related to the KEGG pathway. The differentially expression of circRNAs were associated with negative regulation of DNA transcription and positive regulation of RNA polymerase II transcription in the nucleus (Fig. 3A). The GO enrichment analysis results were statistically analyzed by GGplot2, and the results were presented in the form of scatter plots (Fig. 3B). The 'human T-cell leukemia virus 1 infection' and 'axon-directed' signaling pathways are significant pathways (Fig. 3C).

Interactions between circRNA and miRNA were analyzed

The interaction of circRNA-miRNA was predicted by TargetScan and miRANDA. TargetScan predicted miRNA target based on seed region. MiRanda is mainly based on the binding free energy of circRNA and miRNA. The smaller the free energy, the stronger the binding ability of the two. The results are shown in Table 5.

Discussion

It has been proved that circRNA plays a significant role in the occurrence and progression of disease. In addition, circRNA has become a hotspot for transcriptome and RNA research. According to the idiographic structure and complicated functional mechanisms of circRNAs, the researches of circRNAs had only come into the field of view in recent years[25]. Recent studies have shown that circRNA is closely related to the occurrence and development of a variety of diseases and tumors[26–28]. To clarify

the potential role of circRNA in VLS, our experiment verified the differential expressions of circRNAs in VLS patient tissues and performed functional analysis. We randomly sampled vulvar skin tissues from 20 patients and 30 healthy women. All patients are untreated patients and their VLS is in the acute phase. All patients are untreated patients and their VLS is in the acute phase. As this study mainly studies VS, the study patients are all female patients and no male patients. All patients in the study underwent a biopsy, and samples were taken after the biopsy.

In this study, we observed that 79 circRNAs were differentially expressed in VLS tissues, of which 54 circRNAs expressions were significantly increased and 25 circRNAs expressions were significantly reduced. This study lists 5 significantly up-regulated (hsa_circ_0000211, hsa_circ_0001789, hsa_circ_0005729, hsa_circ_0000839, hsa_circ_0000344) and down-regulated circRNAs(hsa_circ_0002472, hsa_circ_0008059, hsa_circ_0004440, hsa_circ_0109021, hsa_circ_0001742). These circRNAs may be related to the pathogenesis and pathological process of VLS.

We found through volcano graph analysis that the expression of circRNAs in the figure was similar to the expression of circRNAs in the VLS sick person group. Through cluster analysis, we found that there were obvious discrepancies in circRNAs expressions between vulvar skin of healthy women and skin of VLS patients. We found that compared with healthy female vulvar skin, the expression of some circRNAs in the skin of VLS patients was significant differences. This indicated that there were significant differences in the expression of certain circulating RNAs between the vulva skin tissues of healthy women and the tissues of patients with VLS. Whether these differential circRNAs can affect the treatment of VLS still needs further research.

The enrichment analysis between KEGG pathway and GO indicated that the unusual expressions of circRNAs were participated in multiple biology-processes and signal-transduction-pathways. Biological-information analysis of abnormal genes found that differentially expressed circRNAs were related to the negative regulation of DNA transcription and the positive regulation of RNA polymerase II transcription in the nucleus. KEGG pathway analysis showed that the "human T cell leukemia virus type 1 infection" signaling pathway and the "axon guidance" signaling pathway were significantly enriched with differential circRNAs. This suggests that the "human T cell leukemia virus type 1 infection" signaling pathway and the "axon guidance" signaling pathway may be related to the occurrence and development of VLS.

An increasing number of studies have shown that circRNAs played crucial parts in many biological functions, which were as ceRNAs or miRNAs sponges, which regulated gene expression by competitively binding miRNAs with MRE (microRNA-responsive element, MRE). In recent years, studies had revealed that circRNAs upstream of target-genes could participate in the mediation of miRNAs, mRNAs and ceRNAs of circRNA-genes. Therefore, circRNAs could affect the occurrences and pathological processes of a variety of diseases. For example, circ-cdr1as could act as a sponge for hsa-miR-7 and modulated person epidermal GF receptors by competitively binding to miR-7[29]. In addition, circRNAs could specifically bind to many kinds of latent factors, for example splicing factors muscle (MBL/MBNL1) [30].

Our research shows that the following 10 pairs of circRNAs may interact with miRNAs: has_circ_0000110 and has-miR-19a-5p, has_circ_0000211 and has-miR-224-5p, has_circ_0000311 and has-miR-215-3p, has_circ_0000344 and has-miR-196a-3p, has_circ_0000592 and has-miR-181a-2-3p, has_circ_0000690 and has-miR-198, has_circ_0000839 and has-miR-221-5p, has_circ_0000941 and has-miR-29a-3p, has_circ_0001485 and has-miR-182-3p, has_circ_0001742 and has-miR-133a-5p. These interacting circRNAs and miRNAs may be related to the molecular mechanism of VLS pathogenicity.

Conclusion

In conclusion, the research identified the differential expression of circRNA in VLS(has_circ_0000211, has_circ_0001789, has_circ_0005729, has_circ_0000839, has_circ_000034, has_circ_0002472, has_circ_0008059, has_circ_0004440, has_circ_0109021, has_circ_0001742). Bioinformatics analysis shows that circRNAs-related diseases may be related to VLS, and the signal pathway of 'human T-cell leukemia virus type 1 infection' and the signal pathway of "axon guidance" may also affect the pathological process of VLS. This study found a total of ten pairs of circRNAs and miRNAs interacting. The specific mechanism of these interacting circRNAs-miRNAs affect the pathological findings of VLS by regulating downstream mRNAs remains to be unclear. This study could expand the perspective of VLS gene research and provide the theoretical basis in future research on the part of circRNAs in VLS.

Materials And Methods

Clinical specimen acquisition

Specimens of LS tissue and vulva skin tissue of healthy women were obtained from 20 LS patients and 20 healthy women treated in Beijing Hospital. The study was divided into experimental group and control group with age matching. This study was approved by the Ethics Committee of Beijing Hospital and a written informed consent was obtained from patients. See Attached Table 1 for details

Preparation of sequencing library and circRNA sequencing

Among the 20 specimens randomly extracted from the three LS tissues, total RNA was extracted from this sample, and the chain specific library was constructed by the method of rRNA depletion. After the library passed the quality inspection, Illumina Novaseq™ 6000 was used for sequencing, with a double-ended reading length of 2*150 bp (PE150) [10]. Raw data generated by sequencing needs to be preprocessed, cutadpter is used to filter unqualified sequences to get clean data, and then the next step is analyzed. The specific processing steps are as follows : (1) Remove reads with Adaptor; (2) Removal of the proportion of reads containing more than 5% N(N means that the base information cannot be determined); (3) Removed low quality Reads (the number of bases with mass value $Q \leq 10$ accounted for more than 20% of the whole read); (4) The original sequencing quantity, effective sequencing quantity, Q20, Q30 and GC content were counted and comprehensively evaluated.

Cluster analysis of differential gene expression levels

Differential gene clustering analysis was used to determine the clustering patterns of regulatory patterns of genes under different experimental conditions. Cluster analysis of genes was conducted according to the similarity of gene expression profiles of samples to intuitively display the expression of genes in different samples (or different treatments), thus obtaining biology-related information. In order to better intuitively reflect the clustering expression pattern, we used $\text{Log}_{10}(\text{FPKM}+1)$ for gene expression display. The horizontal axis is the sample and the vertical axis is the gene. Different colors represent different gene expression levels, and the color ranges from blue to white to red to indicate the expression level ($\text{log}_{10}(\text{FPKM}+1)$) from low to high. High expression genes are in red and low expression genes are in dark blue.

Enrichment analysis of differential genes

GO functional significance enrichment analysis

Gene Ontology(GO) was an international standardized system of Gene function classification. It provided a dynamically updated controlled vocabulary to comprehensively describe the properties of genes and Gene produced in an organism. There were three ontologies in GO, which described molecular function, cellular component and biological process of gene respectively. The basic unit of GO was term(term, node), and each term corresponded to an attribute. The GO functional significance enrichment analysis first mapped all the differentially expressed genes to each term in the Gene Ontology database, calculated the number of genes in each term, and then applied hypergeometric test to find the GO items significantly enriched in differentially expressed genes compared with the entire genome background.

KEGG enrichment analysis

In organisms, different genes coordinated their biological functions with each other, and Pathway-based analysis was helpful for further understanding the biological functions of genes. KEGG was the main public database on Pathways. Pathway significance enrichment analysis used KEGG Pathway as the unit, and hypergeometric tests were used to identify the pathways that were significantly enriched in differentially expressed genes compared with the overall genome background.

Declarations

Ethics approval and consent to participate

All applicable international, national, and institutional guidelines for the care and use of animals were followed.

Consent for publication

We agreed to publish our manuscript to BioMedical Engineering Online.

Competing interests

The authors declare no competing interest.

Funding

The research was provided by the Beijing Natural Science Foundation (7172192).

Authors' contributions

Min Yang: Data curation, Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Kailu Sun:** Formal analysis, Investigation, Methodology, Writing-review & editing. **Jianmin Chang:** Methodology, Writing - review & editing.

Acknowledgements

The authors will thank the editors and reviewers for their valuable comments.

Availability of data and material

The datasets used or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Bleeker, M.C., et al., *Lichen Sclerosus: Incidence and Risk of Vulvar Squamous Cell Carcinoma*. Cancer Epidemiol Biomarkers Prev, 2016. **25**(8): p. 1224-30.
2. Berrebi, A., et al., *[Virus diseases and cancer of the vulva]*. Rev Fr Gynecol Obstet, 1987. **82**(10): p. 583-5.
3. Lynch, P.J., et al., *2011 ISSVD Terminology and classification of vulvar dermatological disorders: an approach to clinical diagnosis*. J Low Genit Tract Dis, 2012. **16**(4): p. 339-44.
4. Østergård, S., C.S. Vorbeck, and M. Meinert, *[Vulvar intraepithelial neoplasia]*. Ugeskr Laeger, 2018. **180**(20).
5. Fergus, K.B., et al., *Pathophysiology, Clinical Manifestations, and Treatment of Lichen Sclerosus: A Systematic Review*. Urology, 2020. **135**: p. 11-19.
6. Carlson, J.A., et al., *Vulvar lichen sclerosis and squamous cell carcinoma: a cohort, case control, and investigational study with historical perspective; implications for chronic inflammation and sclerosis in the development of neoplasia*. Hum Pathol, 1998. **29**(9): p. 932-48.
7. Singh, N., N. Mishra, and P. Ghatage, *Treatment Options in Vulvar Lichen Sclerosus: A Scoping Review*. Cureus, 2021. **13**(2): p. e13527.
8. Kashgari, F.K., et al., *Identification and experimental confirmation of novel cGMP efflux inhibitors by virtual ligand screening of vardenafil-analogues*. Biomed Pharmacother, 2020. **126**: p. 110109.

9. Gonzalo-Gil, E., U. Ikediobi, and R.E. Sutton, *Mechanisms of Virologic Control and Clinical Characteristics of HIV+ Elite/Viremic Controllers*. Yale J Biol Med, 2017. **90**(2): p. 245-259.
10. Xu, H., et al., *RNA-seq analyses the effect of high-salt diet in hypertension*. Gene, 2018. **677**: p. 245-250.
11. Meng, S., et al., *CircRNA: functions and properties of a novel potential biomarker for cancer*. Mol Cancer, 2017. **16**(1): p. 94.
12. Chen, L.L. and L. Yang, *Regulation of circRNA biogenesis*. RNA Biol, 2015. **12**(4): p. 381-8.
13. Chen, X., et al., *Circular RNAs in immune responses and immune diseases*. Theranostics, 2019. **9**(2): p. 588-607.
14. !!! INVALID CITATION !!! {}.
15. Hentze, M.W. and T. Preiss, *Circular RNAs: splicing's enigma variations*. Embo j, 2013. **32**(7): p. 923-5.
16. Bailleul, B., *During in vivo maturation of eukaryotic nuclear mRNA, splicing yields excised exon circles*. Nucleic Acids Res, 1996. **24**(6): p. 1015-9.
17. Memczak, S., et al., *Circular RNAs are a large class of animal RNAs with regulatory potency*. Nature, 2013. **495**(7441): p. 333-8.
18. Huang, X., et al., *High Throughput Single Cell RNA Sequencing, Bioinformatics Analysis and Applications*. Adv Exp Med Biol, 2018. **1068**: p. 33-43.
19. Daugaard, I. and T.B. Hansen, *Biogenesis and Function of Ago-Associated RNAs*. Trends Genet, 2017. **33**(3): p. 208-219.
20. Gokool, A., et al., *Circular RNAs: The Brain Transcriptome Comes Full Circle*. Trends Neurosci, 2020. **43**(10): p. 752-766.
21. Kong, S., et al., *Translatable circRNAs and lncRNAs: Driving mechanisms and functions of their translation products*. Cancer Lett, 2020. **483**: p. 59-65.
22. Friedrich, S., et al., *AUF1 p45 promotes West Nile virus replication by an RNA chaperone activity that supports cyclization of the viral genome*. J Virol, 2014. **88**(19): p. 11586-99.
23. Rutsaert, S., et al., *Evaluation of HIV-1 reservoir levels as possible markers for virological failure during boosted darunavir monotherapy*. J Antimicrob Chemother, 2019. **74**(10): p. 3030-3034.
24. Morabito, C., et al., *Illumina and PacBio DNA sequencing data, de novo assembly and annotation of the genome of Aurantiochytrium limacinum strain CCAP_4062/1*. Data Brief, 2020. **31**: p. 105729.
25. Shafabakhsh, R., et al., *Could circRNA be a new biomarker for pre-eclampsia?* Mol Reprod Dev, 2019. **86**(12): p. 1773-1780.
26. Yang, Y., et al., *The roles of miRNA, lncRNA and circRNA in the development of osteoporosis*. Biol Res, 2020. **53**(1): p. 40.
27. Wu, P., et al., *Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA*. Mol Cancer, 2020. **19**(1): p. 22.
28. Li, X., et al., *NUDT21 regulates circRNA cyclization and ceRNA crosstalk in hepatocellular carcinoma*. Oncogene, 2020. **39**(4): p. 891-904.

29. Hansen, T.B., J. Kjems, and C.K. Damgaard, *Circular RNA and miR-7 in cancer*. Cancer Res, 2013. **73**(18): p. 5609-12.
30. Fu, Y., et al., *MBNL 1-RNA recognition: contributions of MBNL 1 sequence and RNA conformation*. Chembiochem, 2012. **13**(1): p. 112-9.

Tables

Due to technical limitations, tables are only available as a download in the Supplemental Files section.

Figures

Figure 1

mRNA expression profile in VLS tissues A) VLS samples of gene expression value distribution statistics. The abscissa was the sample name, the ordinate was $\log_{10}(\text{FPKM})$, and the box chart for each region corresponded to five statistics (top to bottom for maximum, upper quartile, median, lower quartile, and minimum). B) mRNA expression density distribution. In the ideal state, the expression density map of each sample should conform to the normal distribution, and the expression trend of biological duplicate

samples should be consistent. The abscissa was $\log_{10}(\text{FPKM})$ and the ordinate was the gene density. C) Distribution of mRNA ORF length. The horizontal axis showed the length of the mRNA ORF, and the y axis showed the quantity of mRNAs was equal to the ORF length.

Figure 2

VLS tissues and normal tissues directly differentially expressed mRNA A) Volcanic map of VLS differential gene expression level. The x axis represented the normalized difference (VLS group/normal group). The y axis represented the normalized P value. In the figure, the red represented the up-regulated

significantly differentially expressed genes, the blue represented the down-regulated significantly differentially expressed genes, and the gray dots represented the non-significantly differentially expressed genes. B) Statistical chart of down-regulated frequency of VLS genes with significantly different expression. The red column represented up-regulated gene frequency, and the blue column represented down-regulated gene frequency. C) Cluster analysis diagram of VLS differential gene expression level. The horizontal axis was the sample (VLS group/normal group), and the vertical axis is the gene. Different colors represent different gene expression levels, and the color ranges from blue to white to red to indicate the expression amount ($\log_{10}(\text{FPKM}+1)$) from low to high. High expression genes were in red and low expression genes were in dark blue.

Figure 3

Pathway enrichment analysis of differentially expressed genes A) GO enrichment analysis of VLS differential genes. The functional characterization message of the calculation was the amount of abnormal genes enriched into the entry. B) GO enrichment scatter plot of VLS differential genes. In the horizontal axis, Rich Factor represents the number of differential genes located in the GO/the total number of genes located in the GO (Rich Factor = S gene number/B gene number). The higher the Rich

Factor was the higher the enrichment degree of GO. The ordinate was GO_TERM, which was the GO function comment. In the scatter plot, the size of the dot represented the number of genes with significant differences that S gene number matches to a single GO, the color of the dot represented the P value of enrichment analysis, namely the significance of enrichment, and the pvalue less than 0.05 indicated significant enrichment. C) KEGG enrichment analysis of VLS differential genes. The value of $-\log_{10}(p \text{ value})$ was important for enrichment. The larger value, the more significant the consequence. The amount of genes was equal to the amount of genes enriched in the clause. The richness factors were the percentage of the amount of genes in the way, which was abnormally expressed in the overall amount of genes in the way. The greater enrichment factors, the higher enrichment degrees. Vertical axis was the appellation of enrichment project.

Figure 4

mRNA expression profile in VLS tissues A) Prediction analysis of lncRNA CPC in VLS samples. The abscissa was the sample name and the ordinate is $\log_{10}(\text{FPKM})$. B) Prediction analysis of lncRNA CNCI in VLS samples. The abscissa was the sample name and the ordinate was $\log_{10}(\text{FPKM})$. C) Visualization results of lncRNA genomes from different samples. Genome mapping was carried out according to different classification of lncRNAs, and every 25MB of each chromosome was taken as the basic unit,

and the expression level of lncRNAs in each segment of the lncRNA genome in different samples was counted and mapped during the visualization mapping. The larger inner green ring represented all lncRNAs detected. Smaller inner loops indicated differentially expressed lncRNA folding changes > 2 and P value <0.05. The up-down regulation of lncRNA has been marked with red and blue bars. D) Pie chart of different class code ratios of lncRNA.

Figure 5

Direct differentially expressed lncRNAs in VLS tissues and normal tissues Volcanic map of VLS differential gene expression level. The x axis represented the normalized difference (VLS group/normal group), and the y axis represented the normalized P value. In figure, the red represented the up-regulated significantly differentially expressed genes, the blue represented the down-regulated significantly differentially expressed genes, and the gray dots represented the non-significantly differentially expressed genes. B) Statistical chart of down-regulated frequency of VLS genes with significantly different expression. The red column represented up-regulated gene frequency, and the blue column represented down-regulated gene frequency. C) Cluster analysis diagram of VLS differential gene expression level. The horizontal axis was the sample (VLS group/normal group), and the vertical axis was the gene. Different colors represented different gene expression levels, and the color ranged from blue to white to red to indicate the expression amount ($\log_{10}(\text{FPKM}+1)$) from low to high. High expression genes were in red and low expression genes were in dark blue.

Figure 6

Expression profile of lncRNA in VLS tissues A) Statistical distribution of gene expression values in VLS samples. The abscissa was the sample name, the ordinate was $\log_{10}(\text{FPKM})$, and the box chart for each region corresponds to five statistics (top to bottom for maximum, upper quartile, median, lower quartile, and minimum). B) Expression density distribution of lncRNA. The abscissa was $\log_{10}(\text{FPKM})$ and the ordinate was the gene density. C) Distribution of lncRNA species. □ Full exon-type circRNAs; ▨ EicircRNAs

in intron-exon combinations; $\boxtimes$ lassotype ciRNA composed of introns; $\boxtimes$ CircRNA generated by viral RNA genomes, tRNA, rRNA, snRNA and other circRNA.

Figure 7

circRNA was directly differentially expressed in VLS tissues and normal tissues A) Volcanic map of VLS differential gene expression level. The horizontal axis represented the normalized difference (VLS group/normal group), and the vertical axis represented the normalized P value. In the figure, the red

represented the up-regulated significantly differentially expressed genes, the blue represented the down-regulated significantly differentially expressed genes, and the gray dots represented the non-significantly differentially expressed genes. B) Statistical chart of down-regulated frequency of VLS genes with significantly different expression. The red column represented up-regulated gene frequency, and the blue column represented down-regulated gene frequency. C) Cluster analysis diagram of VLS differential gene expression level. The horizontal axis was the sample (VLS group/normal group), and the vertical axis was the gene. Different colors represented different gene expression levels, and the color ranged from blue to white to red to indicate the expression amount ($\log_{10}(\text{FPKM}+1)$) from low to high. High expression genes were in red and low expression genes were in dark blue.

Figure 8

Pathway enrichment analysis of differentially expressed genes A) GO enrichment analysis of VLS differential genes. The functional characterization information of the calculation was the amount of abnormal genes enriched into the item. B) GO enrichment scatter plot of VLS differential genes. In the horizontal axis, Rich Factor represented the number of differential genes located in the GO/the total number of genes located in the GO (Rich Factor =S gene number/B gene number). The higher the Rich

Factor was, the higher the enrichment degree of GO. The ordinate was GO_TERM, which was the GO function comment. In the scatter plot, the size of the dot represented the number of genes with significant differences that S gene number matched to a single GO, the color of the dot represented the P value of enrichment analysis, namely the significance of enrichment, and the pvalue less than 0.05 indicates significant enrichment. C) KEGG enrichment analysis of VLS differential genes. The value of $-\log_{10}$ (p value) was important for enrichment. The larger value, the more significant the consequence. The amount of genes was equal to the amount of genes enriched in the clause. The richness factors were the percentage of the amount of genes in the way, which was abnormally expressed in the overall amount of genes in the way. The greater enrichment factors, the higher enrichment degrees. Vertical axis was the appellation of enrichment project.

Supplementary Files

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

- [Table1.jpg](#)
- [Table2.jpg](#)
- [Table3.jpg](#)
- [Table41.jpg](#)
- [Table51.jpg](#)