lncRNA LINC00652 is a significant prognostic and diagnostic biomarker in the Isfahan Breast cancer patients, correlated to ID1: integrated bioinformatics and experimental approach

Mahdies Saeri
School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran
Zahra Ramezani
Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran
Samaneh Abbasi
Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran
Shamim Foroutan
Zist Fanavari Novin Biotechnology Institute, Isfahan, Iran
Mansoureh Azadeh (mazadeh@phd.iaurasht.ac.ir)
Department of Biology, Rasht Branch, Islamic Azad University, Rasht, Iran

Research Article

Keywords: Breast cancer, Bioinformatics, Biomarker analysis, real-time PCR

Posted Date: September 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2064409/v1

Abstract

Background

The primary goal of this study is to compare the levels of expression of ID1, LINC00101, and LINC00652 in Isfahan Breast cancer samples to control samples. Understanding the relationship between variations in RNA expression levels and the clinicopathological statuses of Isfahan patients is another important goal of this research.

Method

R Studio evaluated 104 breast cancer tissue and 17 control samples in the GSE42568 microarray datasets to identify a relevant dysregulated gene. The limma package was used to conduct differential expression analysis. The DEGs were used to test mRNA-lncRNA interactions using the co-lncRNA online database. The qRT-PCR experiment aimed to determine the degree of expression of ID1, LINC001013, and LINC00652.

Result

Based on bioinformatics and experimental analyses, ID1 has a significant low-expression in the BC samples. Furthermore, lncRNAs LINC00101 and LINC00652 have significant high expression in the BC samples. ID1 and LINC00652 might be the two potential diagnostic (ROC analysis) and prognostic (Clinicopathological analysis) biomarkers of BC.

Discussion

According to ROC and clinicopathological analyses, ID1 and LINC00652 could be the two significant prognosis and diagnosis biomarkers in Isfahan Breast cancer patients.

3. Introduction

Over the last 50 years, most cancer research has focused on determining how tumor cells vary from normal cells in gene expression (Zhang et al., 1997). These distinctions might have a significant impact on several human illnesses, including Hepatitis (Asselah et al., 2009), Alzheimer (Theuns and Van Broeckhoven, 2000), Diabetes (Das and Rao, 2007), MS (Tajouri, Fernandez and Griffiths, 2007), and some different human cancer types, including Liver cancer (Chen et al., 2002), Colorectal cancer (Kheirelseid et al., 2013), Retinoblastoma (Kapatai et al., 2013), Lung cancer (Petty et al., 2004), Head and Neck cancer (Nagai, 1999), and Breast cancer (Bao and Davidson, 2008; Reis-Filho and Pusztai, 2011; Arpino et al., 2013; Guler, 2017). An extensive study on gene expression patterns in many pathogenic

illnesses, such as breast cancer, can assist in speeding up the treatment of these diseases and forecast their emergence.

Breast cancer is the most common malignancy among women, with 1.8 million new cases identified in 2013, accounting for 12 percent of all malignancies. Although breast cancer incidence is lower among Iranian women, epidemiological research shows that the number of newly diagnosed breast cancer patients has lately grown (Assad Samani et al., 2019). According to Cancer Statistics 2018, BC is the most frequent female malignancy, and the primary cause of cancer mortality in women, with over 2.1 million females diagnosed each year and over 62,000 fatalities. Developing countries account for over 60% of BC deaths (Bray et al., 2018). While early detection and recent advances in anti-cancer therapy have improved crucial outcomes for BC patients, the recurrence rate of the disease remains high (Kim et al., 2018; N. Li et al., 2019; Siegel, Miller, and Jemal, 2020). As a result, measuring the expression of genes linked to breast cancer, discovering diagnostic and prognostic biomarkers, and comprehending gene expression patterns in various clinical and pathological circumstances linked to breast cancer might provide valuable information about the illness and aid in its prevention.

Scientists have created several techniques to monitor gene expression. Two powerful technologies for assessing gene expression are real-time PCR and microarray. Gene expression profiling and genome-wide gene expression analysis utilizing DNA microarray might provide data on the expression level and relative expression of genes and RNAs in different groups, such as "tumor" and "control," or "treated" and "untreated" (Dufva, 2009).

Among some various genes involved in Breast cancer, ID1 was selected by integrated microarray analysis. The roles of ID1 in regulating cell differentiation during neurogenesis, lymphoid, angiogenesis, cell growth, and cell cycle progression have been studied. There are four inhibitors of DNA-binding (ID1, 2, 3, 4) proteins which are members of basic helix-loop-helix (bHLH), as they form transcriptionally inactive Id-bHLH protein complexes (Maw, Fujimoto and Tamaya, 2008). Based on bioinformatics analyses, it is hypothesized that abnormal changes in ID1 expression in the Breast could effectively develop breast cancer status.

LncRNAs are RNA molecules with a length of more than 200 nucleotides that can be found in the nucleus or cytoplasm. Only a few can encode a tiny number of polypeptides, while the remainder does not code for proteins. LncRNAs influence a wide range of biological events at the pre-transcriptional and transcriptional phases, including tumor invasion, metastasis, and apoptosis. As a result, lncRNA abnormalities in BC patients' peripheral blood may aid in the development and progression of the disease, allowing for early detection and therapy (Mercer, Dinger, and Mattick, 2009; Fang and Fullwood, 2016; Bin et al., 2018). In this study, evaluation of the ID1 and two relevant lncRNAs was the main aim. Also, investigation about the correlation of this three RNA and the possible roles of breast cancer.

4. Material And Methods

4.1. Bioinformatics analyses

GSE42568 microarray dataset was download from GEO online database and analyzed by R Studio (4.0.2) software. Statistical analyses and drawing the plots of microarray analysis were performed by Bioconductor packages. GEOquery (Sean and Meltzer, 2007) package was used for downloading the microarray data from GEO database. The Limma (Ritchie et al., 2015) package was used for statistical analyses and DEG analysis. Ggplot2 and pheatmap packages was used to drawing the plots and figures. Log2 fold-change (FC) ≥ 1 or ≤ -1 was considered as the expression threshold. P Value < 0.05 was considered as the significance level. lncRNA-mRNA interaction analysis was performed by co-lncRNA online software.

4.2. Tissue sample collection:

Twenty-one breast cancer tissue samples were collected from the Al-Zahra Hospital of Isfahan, where the cancer was confirmed. Patients with breast cancer were sampled, and breast cancer tissue was used as a tumor and adjacent tissue as normal tissue.

4.3.RNA extraction and cDNA synthesis:

For total RNA extraction of both tissues (tumor and control), TRIZOL was used (Invitrogen, Carlsbad, CA, USA). The quality of extracted RNAs was determined according to the 260/280 absorbance ratio, measured by the Nanodrop spectrometer.

Then cDNA synthesis was performed by Takara kit following the protocol included the initial melting step of 95°C for 4min, 35 cycles of denaturation in 94°C for 30 s, annealing temperature of 57°C for 30 s, the extension period in 72°C for 30 s, and final extension in 72°C for 5 min. (Takara, Tokyo, Japan).

4.4. Primers:

The primers were synthesized (Metabion company – Germany), which are listed in Table 1.

Table 1
Table of primer sequences.

Name	Primer sequence (5' – 3')
ID1	F - TGTTTCAGCCAGTCGCCAAG
	R - CGTTCATGTCGTAGAGCAGCAC
LINC00101	F - ATGGGAGTTGTCATGTTGGGA
	R - ACATCACTCGTCACTCTTGGTC
LINC00652	F - CCCAACCAGAGCAAACCCAG
	R - CCTGCTTCTTCCCATAAGGAC

4.5. Real time PCR:

To Real-Time PCR, Sybergreen (Amplicon Company, Denmark) was used, and RT-qPCR carried out using MIC Real-Time PCR device (Australia). Under the protocol, all samples were run in duplicate in MIC q-PCR device. First activation at 95 ° C for 20 s and 40 cycles of denaturation at 95 ° C for 5 s, primer annealing at 61 ° C for 15 s, and extension at 72 ° C for15 s.

4.6. Statistical analysis:

Graphs were created by Graph Pad Prism software (version 8.0). qRT-PCR data were analyzed using the -ΔΔCT method. Statistical analysis was performed using paired t-test and Wilcoxon test for comparing and analyzing expression data between normal and breast cancer tissue samples by Graph Pad prism eight and GenEx (version 6). Kruskal-Wallis test was performed for clinic pathological analysis. DGE analysis was assessed using the RStudio software (version 4.0.2). ROC curves were generated by Graph Pad Prism software.

5. Results

5.1. Bioinformatics analyses

Microarray analysis was performed on 104 Breast cancer samples and 17 control microarray samples from the GSE42568 dataset (Fig. 1). Based on microarray analysis, ID1 was considered as a significantly down-regulated gene in Breast cancer samples, as compared to control samples (logFC > 1, adj. P. Value < 0.05, Figs. 2 and 3). This differentially expressed gene was considered for the subsequent interaction analyses and further investigations.

According to co-lncRNA online database analysis, an interaction between ID1 and two LINC00101 and LINC00652 was found. Based on these bioinformatics analyses, a real-time PCR experiment was designed to investigate the role of these three RNA expressions in Isfahan's Breast cancer status.

5.2. Real time PCR analysis

Real-time PCR experiment revealed that ID1 has a significant down-regulation in Isfahan Breast cancer samples. Also, two LINC00101 and LINC00652 lncRNAs significantly low expression in Isfahan Breast cancer samples (Fig. 4).

ROC analysis was performed to investigate the prognosis biomarker capability of these three RNAs, based on sensitivity and specificity. Based on this analysis, ID1 and LINC 00652 could be considered as acceptable prognosis biomarkers in Isfahan Breast cancer (AUC > 0.7, P. Value < 0.05, Fig. 5)

5.3. Clinicopathological analyses

The expression of ID1, LINC00101, and LINC00652 was evaluated in the samples with different clinicopathological features, including Stage, Age, Lymph node metastasis, Menopausal status, Tumor size, ER receptor, PR receptor, and HER2/neu receptor. Based on this analysis, the expression of ID1 is

significantly decreased in the sample with the tumor size lower than 5cm and the samples that have no lymph node metastasis compared to the control sample (Fig. 6). Also, the expression of LINC00652 had a significantly high expression in persons older than 45 years (Fig. 7).

6. Discussion

To design a reasonable investigation and achieve a reliable result, we perform an integrated bioinformatics analysis, including microarray data analysis and lncRNA-mRNA interaction analysis, by R Studio and co-lncRNA database. These bioinformatical analyses demonstrated that ID1 might have a significantly low expression in Breast cancer samples. Also, we find that two LINC00101 and LINC00652 have interaction with ID1. Based on this bioinformatics research, we hypothesized that the three RNAs could affect breast cancer by altering their expression, and altering the expression of either lncRNAs could affect ID1 expression level. We designed a real-time PCR experiment to investigate every three RNAs' expression levels based on these results.

Our real-time PCR experiment indicated that ID1 could be a tumor suppressor in Isfahan Breast cancer samples. This gene had a significantly low expression in Isfahan tumor samples compared to controls. Furthermore, we demonstrated that LINC00652 and LINC00101 had significantly high expression in tumor samples and could be the Breast cancer oncogenes. ID1 and LINC00652 could be two prognosis biomarkers based on ROC analysis. Our experiment represented a significant correlation between the size of Breast cancer tumor and lymph node invasion and the expression level of ID1. According to this analysis, a significant reduction in ID expression is observed in non-invasive and small-sized tumor samples.

Previous studies about the two investigated lncRNAs tried to clear the role of these RNAs in several human diseases, including cancer. According to previous studies, up-regulation of LINC00101 (SNHG11) could promote Gastric cancer progression by activating the Wnt/β-Catenin Pathway and Oncogenic Autophagy (Wu et al., 2021). Another study demonstrated that SNHG11 has a significantly high expression in colorectal cancer samples by promoter hypomethylation. This up-regulation could facilitate metastasis by interacting with and stabilizing HIF-1α (L. Xu et al., 2020). Also, another investigation revealed that circulating lncRNA SNHG11 could be a novel diagnosis and prognosis biomarker for colorectal cancer (W. Xu et al., 2020). About hepatocellular carcinoma, previous studies revealed that SNHG11 could promote proliferation and migration by interaction with hsa-miR-184/AGO2 (Huang et al., 2020). In this experiment, we examined for the first time the expression of this RNA in Breast cancer. The result of our research was almost in line with previous results. In breast cancer specimens, increased expression of this RNA was associated with the progression of Breast cancer. About the LINC00652, some previous bioinformatical investigations revealed that this RNA could have a potential role in Lung cancer (J. Li et al., 2019), Renal cell carcinoma (D. Xu et al., 2020), Papillary thyroid cancer (You et al., 2018). We perform a real-time PCR experiment for the evaluation of LINC00652 for the first time.

Clinicopathological analysis revealed a significant correlation between lymph node metastasis and tumor size with the expression level of ID1. Also, LINC00652 had a significant up-regulation in the samples older than 45 years. It is suggested that similar experiments be performed on the expression of these RNAs in samples with different clinicopathological statuses with a larger size to obtain more and more reliable results from the effect of different pathological conditions on the expression of these RNAs.

7. Conclusion

Our experiment has introduced the two ID1 and LINC00652 RNAs as the diagnostic and prognostic biomarkers of breast cancer. Mentioned coding and non-coding RNAs have significant dysregulation in the BC patients, specially in the samples with the tumor size lower than 5cm, samples with no lymph node metastasis, and the patients older than 45 years.

Abbreviations

BC
Breast cancer
GEO
Gene expression omnibus
AUC
Area under curve
Adj. P. Value
Adjusted P. Value
LINC
Long Intergenic Non-coding
FC
Fold Change
qRT-PCR
Quantitative Real-Time Reverse Transcription PCR
PCR
Polymerase chain reaction
ID1
Inhibitor of DNA Binding 1
DEG
Differentially expressed gene

Declarations

Ethics approval and consent to participate

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Ethics Committee of Isfahan University of Medical Sciences (approval number: 3838988).

Declaration of interests: none.

Acknowledgements: We wish to extend our special thanks to Mohammad Rezaei for helping for data analysis, bioinformatics and reviewing the manuscript.

References

1. Arpino, G. *et al.* (2013) 'Gene expression profiling in breast cancer: A clinical perspective', *Breast*. Churchill Livingstone, pp. 109–120. doi: 10.1016/j.breast.2013.01.016.
2. Assad Samani, L. *et al.* (2019) 'TP53 rs1625895 is Related to Breast Cancer Incidence and Early Death in Iranian Population', Indian Journal of Clinical Biochemistry, 34(4), pp. 485–489. doi: 10.1007/s12291-018-0774-6.
3. Asselah, T. *et al.* (2009) 'Gene expression and hepatitis C virus infection', *Gut*. Gut, pp. 846–858. doi: 10.1136/gut.2008.166348.
4. Bao, T. and Davidson, N. E. (2008) 'Gene Expression Profiling of Breast Cancer', *Advances in Surgery*. NIH Public Access, pp. 249–260. doi: 10.1016/j.yasu.2008.03.002.
5. Bin, X. *et al.* (2018) 'Research progresses in roles of LncRNA and its relationships with breast cancer', Cancer Cell International. BioMed Central Ltd. doi: 10.1186/s12935-018-0674-0.
6. Bray, F. *et al.* (2018) 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', CA: A Cancer Journal for Clinicians, 68(6), pp. 394–424. doi: 10.3322/caac.21492.
7. Chen, X. *et al.* (2002) 'Gene Expression Patterns in Human Liver Cancers', Molecular Biology of the Cell, 13(6), pp. 1929–1939. doi: 10.1091/mbc.02-02-0023.
8. Das, U. N. and Rao, A. A. (2007) 'Gene expression profile in obesity and type 2 diabetes mellitus', Lipids in Health and Disease, 6, p. 35. doi: 10.1186/1476-511X-6-35.
9. Dufva, M. (2009) 'Introduction to microarray technology.', *Methods in molecular biology (Clifton, N.J.)*. Methods Mol Biol, pp. 1–22. doi: 10.1007/978-1-59745-538-1_1.
10. Fang, Y. and Fullwood, M. J. (2016) 'Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer', *Genomics, Proteomics and Bioinformatics*. Beijing Genomics Institute, pp. 42–54. doi: 10.1016/j.gpb.2015.09.006.
11. Guler, E. N. (2017) 'Gene Expression Profiling in Breast Cancer and Its Effect on Therapy Selection in Early-Stage Breast Cancer', European Journal of Breast Health, 13(4), pp. 168–174. doi: 10.5152/ejbh.2017.3636.
12. Huang, W. *et al.* (2020) 'LncRNA SNHG11 promotes proliferation, migration, apoptosis, and autophagy by regulating hsa-miR-184/AGO2in HCC', OncoTargets and Therapy, 13, pp. 413–421. doi:

10.2147/OTT.S237161.

13. Kapatai, G. *et al.* (2013) 'Gene expression profiling identifies different sub-types of retinoblastoma', British Journal of Cancer, 109(2), pp. 512–525. doi: 10.1038/bjc.2013.283.
14. Kheirelseid, E. A. H. *et al.* (2013) 'Clinical applications of gene expression in colorectal cancer', Journal of Gastrointestinal Oncology, 4(2), pp. 144–157. doi: 10.3978/j.issn.2078-6891.2013.010.
15. Kim, Y. A. *et al.* (2018) 'Doxorubicin-induced heart failure in cancer patients: A cohort study based on the Korean National Health Insurance Database', Cancer Medicine, 7(12), pp. 6084–6092. doi: 10.1002/cam4.1886.
16. Li, J. *et al.* (2019) 'Screening of important lncRNAs associated with the prognosis of lung adenocarcinoma, based on integrated bioinformatics analysis', Molecular Medicine Reports, 49(5), pp. 4067–4080. doi: 10.3892/mmr.2019.10061.
17. Li, N. *et al.* (2019) 'Global burden of breast cancer and attributable risk factors in 195 countries and territories, from 1990 to 2017: Results from the Global Burden of Disease Study 2017', Journal of Hematology and Oncology, 12(1), pp. 1–12. doi: 10.1186/s13045-019-0828-0.
18. Maw, M. K., Fujimoto, J. and Tamaya, T. (2008) 'Expression of the inhibitor of DNA-binding (ID)-1 protein as an angiogenic mediator in tumour advancement of uterine cervical cancers', British Journal of Cancer, 99(10), pp. 1557–1563. doi: 10.1038/sj.bjc.6604722.
19. Mercer, T. R., Dinger, M. E. and Mattick, J. S. (2009) 'Long non-coding RNAs: Insights into functions', Nature Reviews Genetics. Nat Rev Genet, pp. 155–159. doi: 10.1038/nrg2521.
20. Nagai, M. A. (1999) 'Genetic alterations in head and neck squamous cell carcinomas', Brazilian Journal of Medical and Biological Research, 32(7), pp. 897–904. doi: 10.1590/S0100-879X1999000700015.
21. Petty, R. D. *et al.* (2004) 'Gene expression profiling in non-small cell lung cancer: From molecular mechanisms to clinical application', Clinical Cancer Research. Clin Cancer Res, pp. 3237–3248. doi: 10.1158/1078-0432.CCR-03-0503.
22. Reis-Filho, J. S. and Pusztai, L. (2011) 'Gene expression profiling in breast cancer: Classification, prognostication, and prediction', *The Lancet*. Elsevier B.V., pp. 1812–1823. doi: 10.1016/S0140-6736(11)61539-0.
23. Ritchie, M. E. *et al.* (2015) 'Limma powers differential expression analyses for RNA-sequencing and microarray studies', Nucleic Acids Research, 43(7), p. e47. doi: 10.1093/nar/gkv007.
24. Sean, D. and Meltzer, P. S. (2007) 'GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor', Bioinformatics, 23(14), pp. 1846–1847. doi: 10.1093/bioinformatics/btm254.
25. Siegel, R. L., Miller, K. D. and Jemal, A. (2020) 'Cancer statistics, 2020', CA: A Cancer Journal for Clinicians, 70(1), pp. 7–30. doi: 10.3322/caac.21590.
26. Tajouri, L., Fernandez, F. and Griffiths, L. (2007) 'Gene Expression Studies in Multiple Sclerosis', Current Genomics, 8(3), pp. 181–189. doi: 10.2174/138920207780833829.

27. Theuns, J. and Van Broeckhoven, C. (2000) ‘Transcriptional regulation of Alzheimer’s disease genes: Implications for susceptibility’, *Human Molecular Genetics*. Oxford University Press, pp. 2383–2394. doi: 10.1093/hmg/9.16.2383.
28. Wu, Q. *et al.* (2021) ‘lncRNA SNHG11 Promotes Gastric Cancer Progression by Activating the Wnt/β-Catenin Pathway and Oncogenic Autophagy’, Molecular Therapy, 29(3), pp. 1258–1278. doi: 10.1016/j.ymthe.2020.10.011.
29. Xu, D. *et al.* (2020) ‘An optimal prognostic model based on gene expression for clear cell renal cell carcinoma’, Oncology Letters, 20(3), pp. 2420–2434. doi: 10.3892/ol.2020.11780.
30. Xu, L. *et al.* (2020) ‘LncRNA SNHG11 facilitates tumor metastasis by interacting with and stabilizing HIF-1α’, Oncogene, 39(46), pp. 7005–7018. doi: 10.1038/s41388-020-01512-8.
31. Xu, W. *et al.* (2020) ‘Circulating lncRNA SNHG11 as a novel biomarker for early diagnosis and prognosis of colorectal cancer’, International Journal of Cancer, 146(10), pp. 2901–2912. doi: 10.1002/ijc.32747.
32. You, X. *et al.* (2018) ‘Integrated analysis of long noncoding RNA interactions reveals the potential role in progression of human papillary thyroid cancer’, Cancer Medicine, 7(11), pp. 5394–5410. doi: 10.1002/cam4.1721.
33. Zhang, L. *et al.* (1997) ‘Gene expression profiles in normal and cancer cells’, Science, 276(5316), pp. 1268–1272. doi: 10.1126/science.276.5316.1268.

Figures

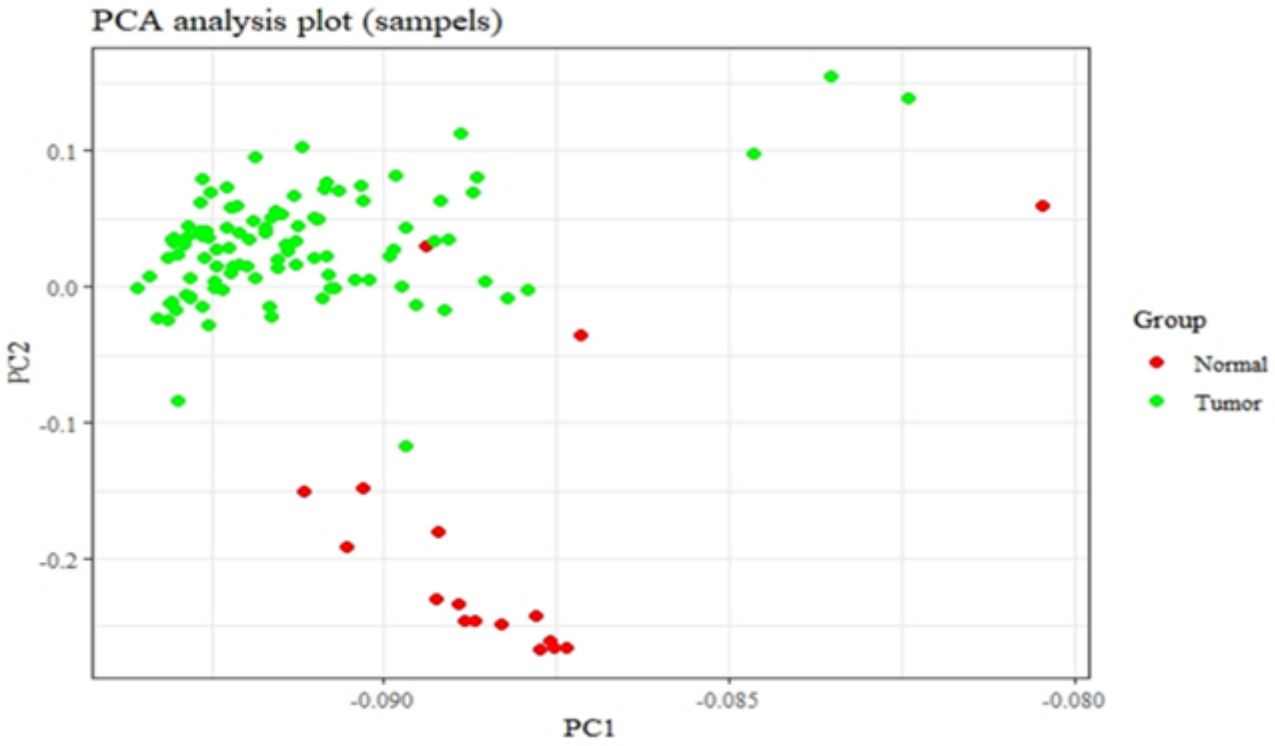

Figure 1

Principal component analysis of GSE42568 samples.

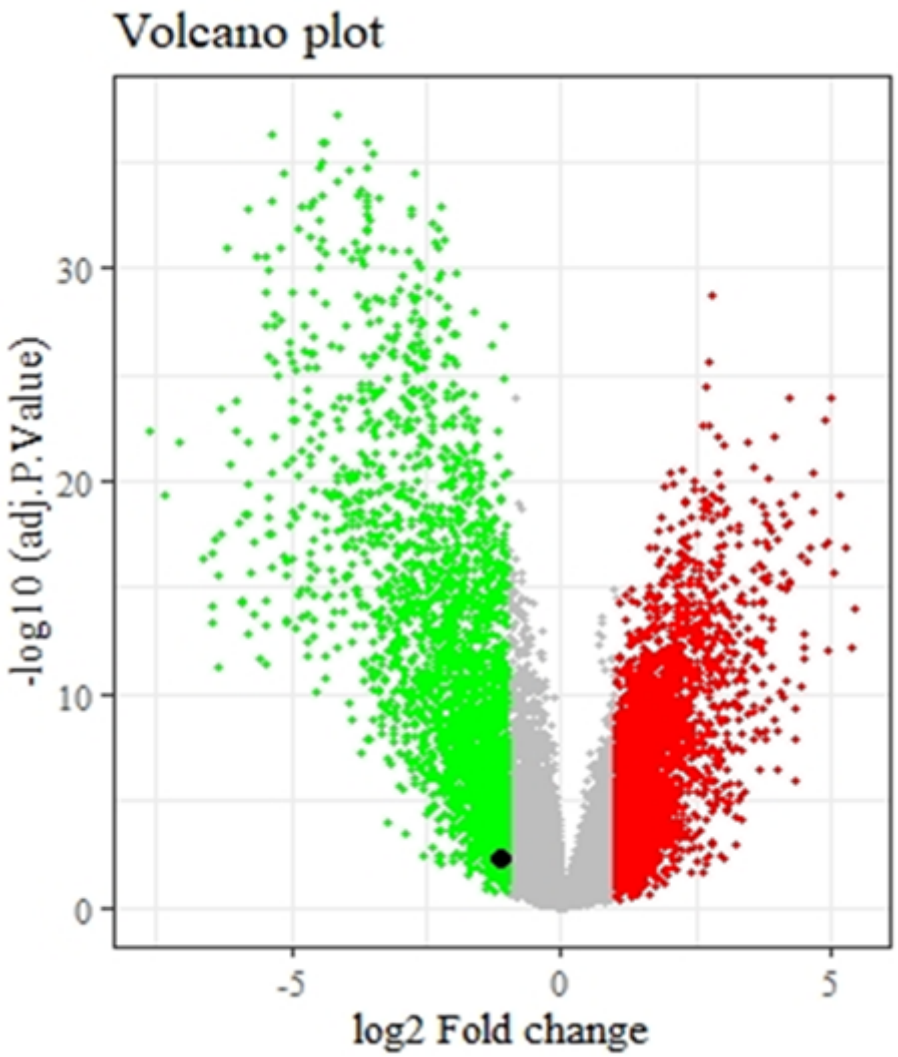

Figure 2

Volcano plot showing the differentially expressed genes between tumor and control samples of GSE42568. The ID1 gene is indicating by a black point.

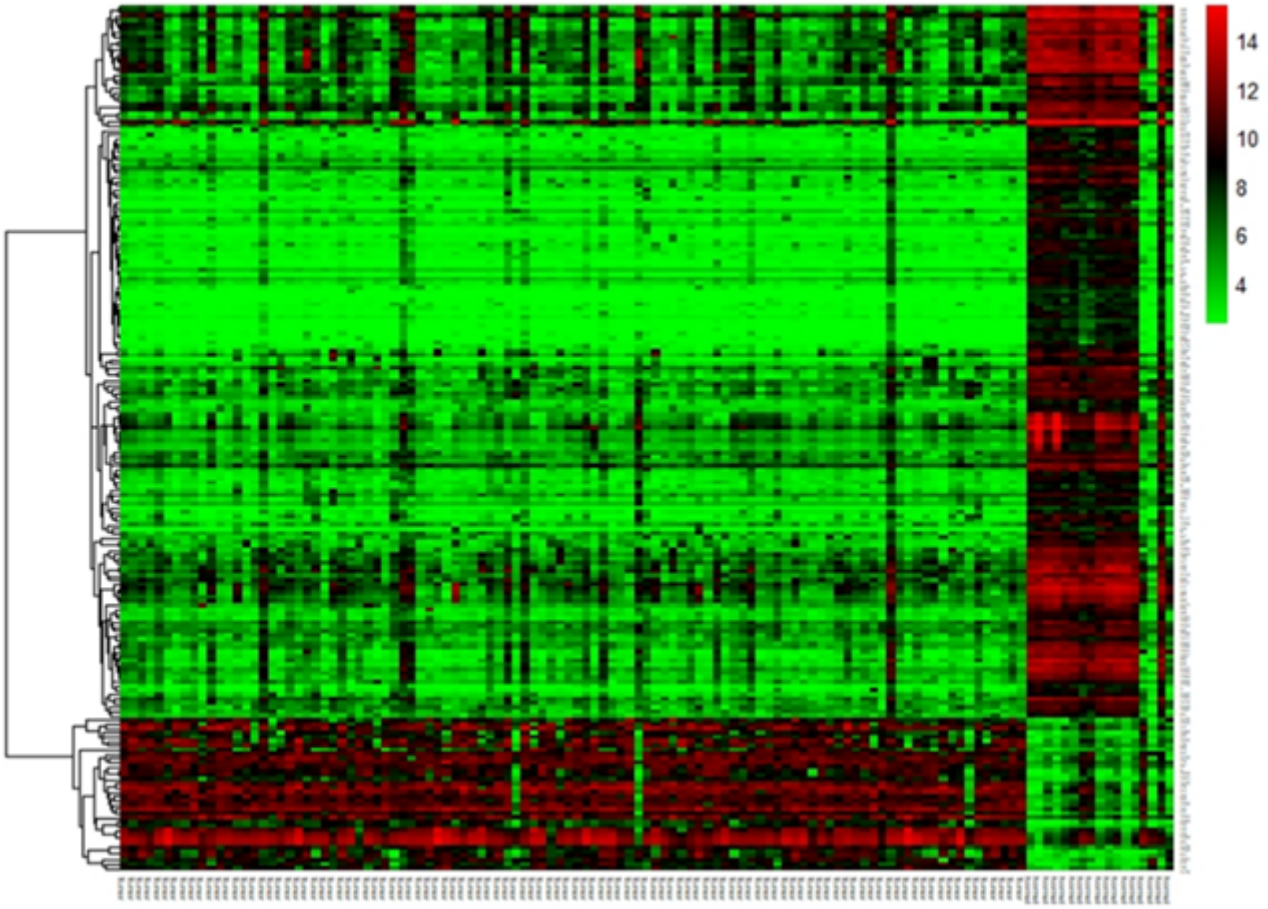

Figure 3

Heatmap from differentially expressed genes in GSE42568. The 104 left columns are tumor samples, and the following 17 columns are normal samples. The ‘correlation’ clustering method was used.

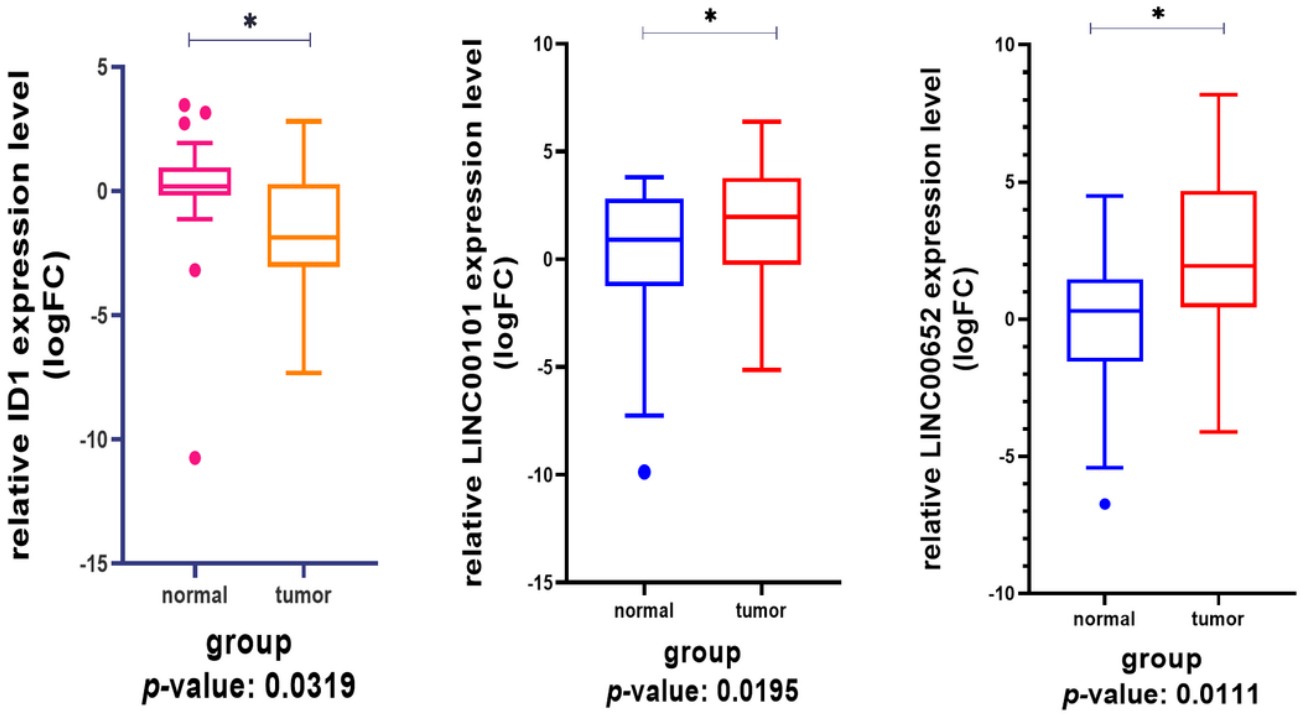

Figure 4

Relative expression analysis of ID1, LINC00101, and LINC00652 in Isfahan Breast cancer samples and paired normal tissue.

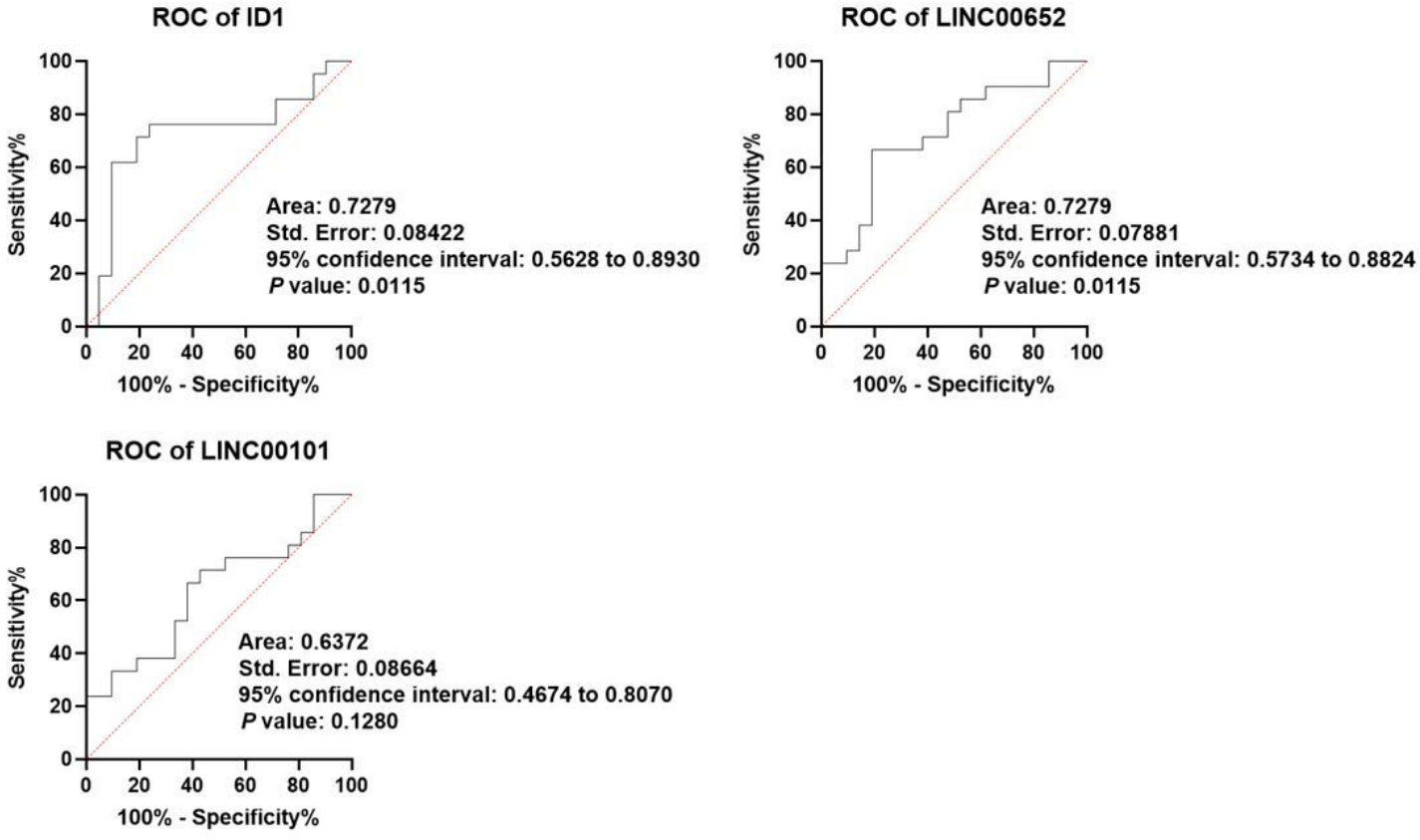

Figure 5

ROC curve of ID1, LINC00101 and LINC00652.

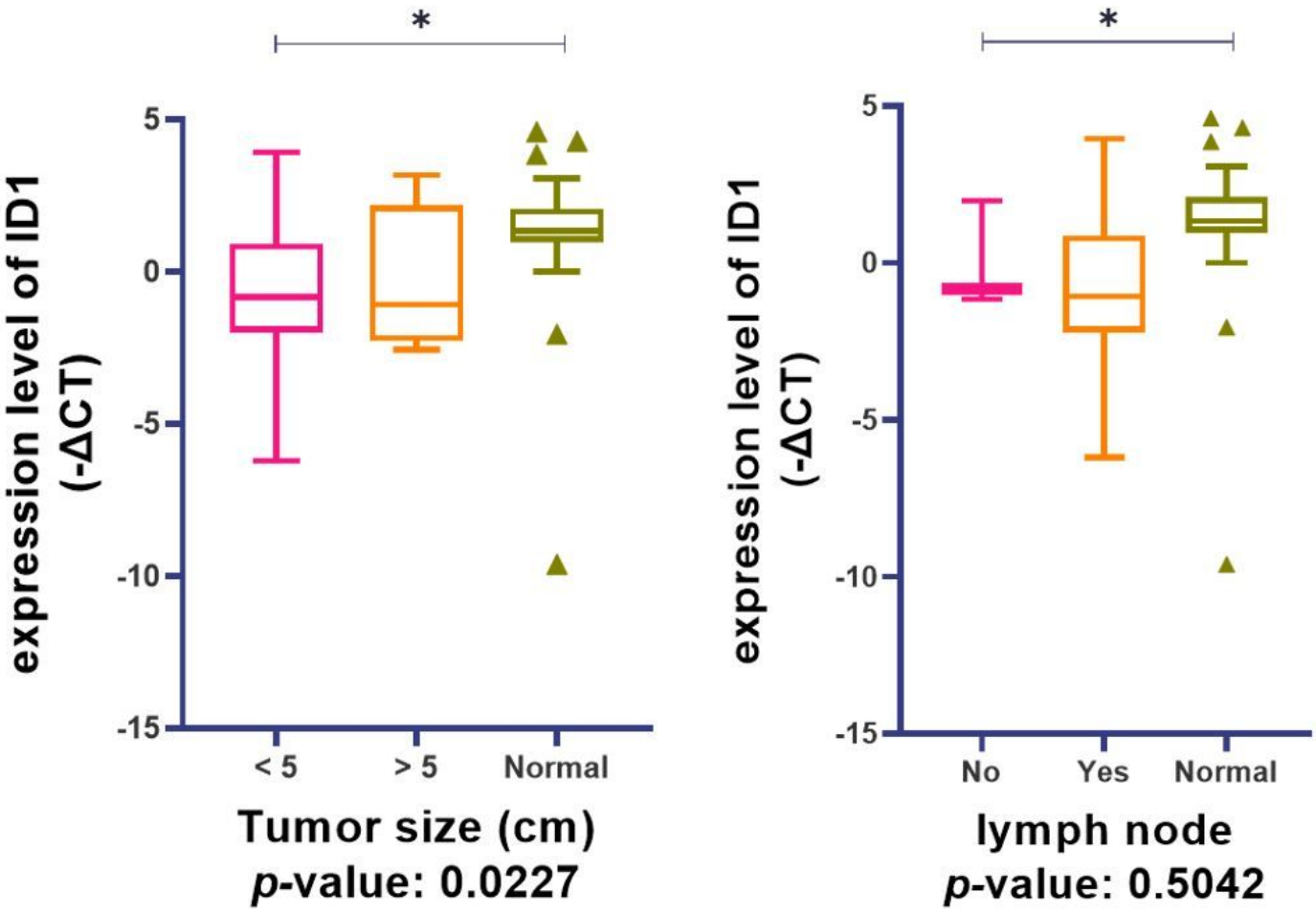

Figure 6

Clinicopathological analysis of ID1 expression level in different Tumor size and Lymph node metastasis status.

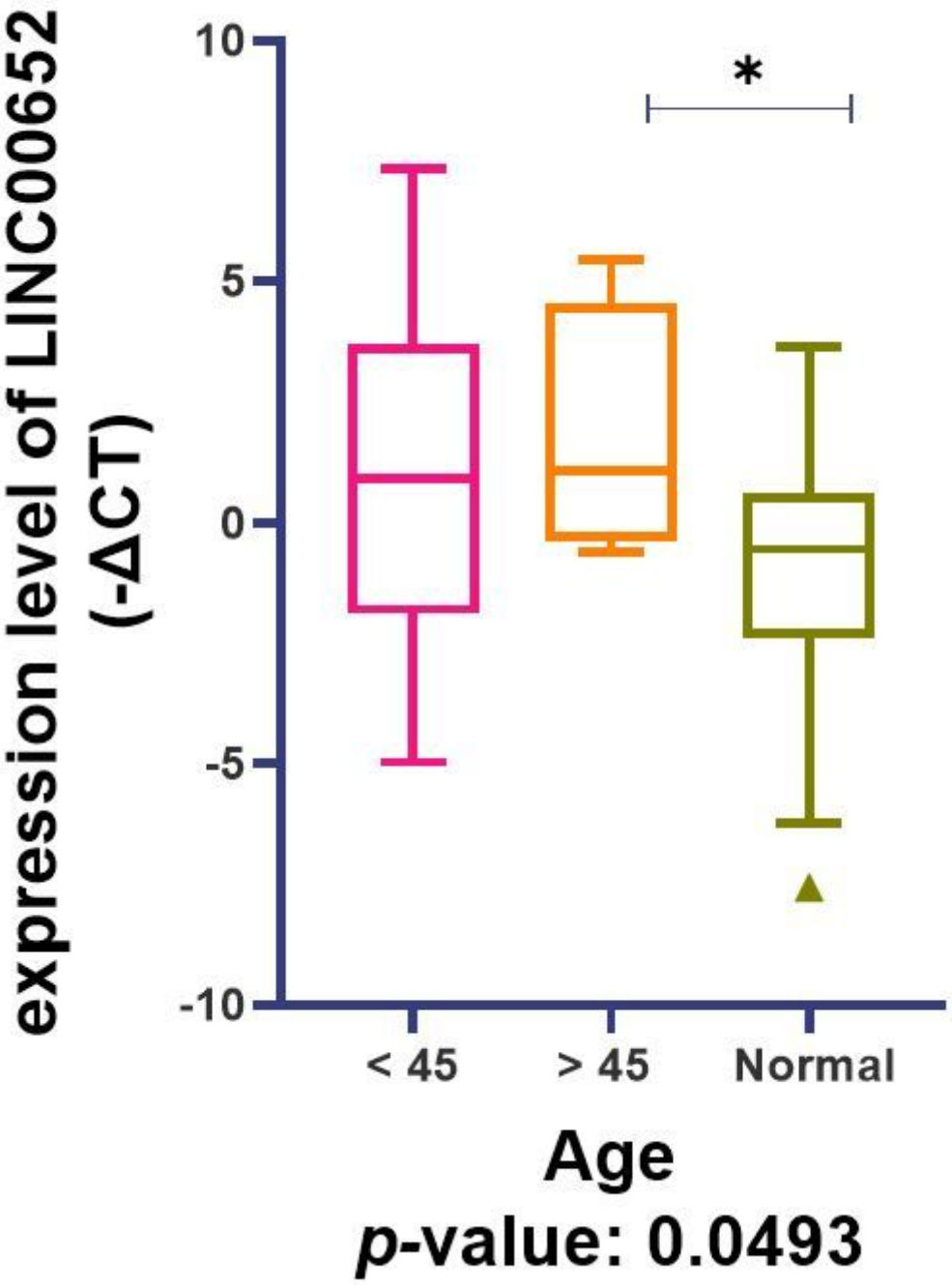

Figure 7

Analysis of LINC00652 expression level in the different age status.