Comprehensive analysis of pseudogene *LDHAP5* expression level and its potential pathogenesis in ovarian serous cystadenocarcinoma

Shitong Lin
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Yifan Meng
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Canhui Cao
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Ping Wu
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Peipei Gao
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Wenhua Zhi
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Ting Peng
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Peng Wu (pengwu8626@tjh.tjmu.edu.cn)
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
https://orcid.org/0000-0002-0737-2785

Primary research

Keywords: ovarian serous cystadenocarcinoma, pseudogene, LDHAP5, EGFR

Posted Date: March 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-17362/v1

Version of Record: A version of this preprint was published at Cancer Cell International on June 10th, 2020. See the published version at https://doi.org/10.1186/s12935-020-01324-6.

Abstract

Background We aim to find out differentially expressed pseudogenes and explore their potential functions in four types of common gynecologic malignancies (cervical squamous cell carcinoma, ovarian serous cystadenocarcinoma, uterine corpus endometrial carcinoma and uterine carcinosarcoma) using bioinformatic technology. Materials & methods: We identify up-regulated or down-regulated pseudogenes and build the pseudogenes-miRNA-mRNA regulatory network through public datasets to explore their potential functions in carcinogenesis and cancer prognosis.

Results LDHAP5 was selected as the most potential candidate pseudogene among 63 up-regulated pseudogenes for it was significantly associated with poor overall survival in ovarian serous cystadenocarcinoma. KEGG pathway analysis revealed that LDHAP5 was most enrichment in microRNAs in cancer, pathway in cancer and PI3K-AKT signaling pathway. Further analysis revealed that EGFR was the potential target mRNA of LDHAP5 which may play a great role in ovarian serous cystadenocarcinoma.

Conclusion LDHAP5 was first discovered to be associated with the occurrence and prognosis of ovarian serous cystadenocarcinoma, and it may be used as a novel specifically therapeutic target against ovarian serous cystadenocarcinoma.

1. Background

Gynecological malignancies account for a large part of women's tumors and seriously endanger women's health. It is estimated that there will be approximately 13,800 new cases of uterine cervical cancer, 65,620 cases of uterine corpus cancer, and 21,750 cases of ovarian cancer in the United States in 2020, and it is speculated that there will be 4,290, 12,590 and 13,940 cancer deaths, respectively.(1) Advanced gynecological malignancies have a poor prognosis for lacking effective treatments to control distant metastasis.(2) However, most current clinical drugs are non-specific, and their therapeutic effects are limited.(3) Therefore, it is urgent to identify novel biomarkers for gynecological tumors, so as to improve drug efficacy and prolong survival.

Pseudogene was first discovered and named by Jacp et al in 1977.(4) Pseudogenes usually originate from paralogous functional genes (“parent gene”), but they lost the capacity of encoding functional proteins for accumulation of mutation (frameshift mutations, early or delayed stop codons, etc.).(5) Pseudogenes have not been paid attention until PTEN pseudogene 1 (PTENP1) was found to share the same microRNA response elements (MREs) with its homologous functional parent gene, PTEN.(6)

With the advancement of next-generation sequencing (NGS), approximately 20,000 pseudogenes have been discovered in the human genome, and the role of pseudogenes as a long non-coding RNAs (lncRNAs) in the development of disease has been further revealed.(7–9) The current research results showed that pseudogenes mainly regulate gene expression at the post-transcriptional levels through the following two pathways.(10) The first way is that pseudogenes can be used as competitive endogenous

RNAs (ceRNAs) to competitively bind miRNAs with coding gene, thereby positively regulating gene expression.(11–13) For example, PTEN pseudogene 1 (PTENP1) can competitively bind miRNA-17, miRNA-21, miRNA-19 and other miRNAs through the ceRNA mechanism, thereby preventing its parent gene PTEN from being degraded by miRNAs and increasing the expression of PTEN gene.(6) Pseudogenes play a negative role in another regulatory pathway, which can complete with their parent gens for destabilizing RNA binding proteins (RBPs), resulting in a decrease in the expression of parent genes.(14)

In our study, we attempted to identify differentially expressed pseudogenes in four gynecological malignancies through the pseudogene database dreamBase, and then use pseudogenes-miRNA-mRNA regulatory network to further explore the potential function and mechanism in gynecological malignancies.

2. Materials & Methods

2.1 Screening for dysregulated pseudogenes in four gynecological malignancies.

We obtained RNA-seq data of pseudogenes in 32 common human cancers from the pseudogene online public database dreamBase (http://rna.sysu.edu.cn/dreamBase/pancancer.php?SClade=mammal&SOrganism=hg38).(15) |Log2FC|>2.0 was set as cutoff to identify differentially expressed pseudogenes. R v 3.5.1 and EXCEL v2016. were used to further analyze their expression landscape.

2.2 Prognostic analysis of upregulated expressed pseudogenes.

Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia.cancer-pku.cn/) was used to evaluate prognostic values (overall survival) of upregulated pseudogenes in 32 kinds of common human cancers.(16) The group thresholds were as follows: the group cut-off was ‘Median’, the ‘cutoff-high’ and ‘cutoff-low’ were 50%, axis units were ‘Months’, and p value < 0.05 was considered statistically significant.

2.3 Screening for pseudogene- regulated miRNAs and miRNA-target mRNAs.

The public online datasets of starBase v2.0 and miRTarBase were used to identify pseudogenes-binding miRNAs and miRNA-target mRNAs, respectively.(17, 18) The network of pseudogenes-miRNA-mRNA was constructed using Cytoscape v_3.7.2.(19)

2.4 KEGG pathways and Gene oncology (GO) enrichment analysis of target mRNAs.

The list of miRNA-target genes was imported into the STRING v_11.0, and the top five significantly GO terms and KEGG pathways were selected according to the values of false discovery rate (FDR), and then were visualized by GraphPad PRISM Version 6.02.(20)

2.5 Construction of protein-protein interaction network and screening for hub genes.

STIRNG v_11.0 was used to construct the regulatory network of protein-protein, and then visualized by Centiscape plugin of Cytoscape v_3.7.2.(19-21) The top ten hub genes were identified according to the values of Degree unDir.

2.6 Hub genes expression and mutations analysis.

Hub genes expression and mutations analysis in ovarian serous cystadenocarcinoma were analyzed using the online database cBioPortal.(22) 489 patients (TCGA, Nature 2011) with ovarian serous cystadenocarcinoma were selected for further analysis. The select genomic profiles were as follows: 'Mutations'; 'Putative copy-number alterations (GISTIC)'; 'mRNA/miRNA expression Z-scores (all genes)', and the Z-scores threshold were ± 2. Finally, OncoPrint was obtained under the guidance of online database at c-BioPortal.

2.7 Identification of potential target gene of *LDHAP5*.

Pearson correlation analysis between *LDHAP5* and the top ten hub genes expression in ovarian serous cystadenocarcinoma was performed using GEPIA.(16) Kaplan–Meier overall survivals of target genes were analyzed by Kaplan–Meier Plotter.(23) The mRNA expression levels of ten hub genes in TCGA patients were further measured using the dataset of Oncomine Main.(24)

3. Results

3.1 identification of dysregulated pseudogenes in four common gynecological malignancies.

According to epidemiological statistics, cervical squamous cell carcinoma, ovarian serous cystadenocarcinoma, uterine corpus endometrial carcinoma and uterine carcinosarcoma are still lethal diseases in women.(1) In order to explore the potential role of pseudogenes in the carcinogenesis and cancer prognosis of four gynecological malignancies, we used the public database dreamBase to identify differentially expressed pseudogenes. As shown in Fig. 1A and Table 1, we identified 63 up-regulated and 0 down-regulated pseudogenes simultaneously in four gynecological malignancies after preliminary screening. We further measured the expression levels of 63 up-regulated pseudogenes in 32 types of human cancers (Fig. 1B). Finally, 40 pseudogenes were thought to play potential roles in gynecological malignancies after removing the pseudogenes that were less highly expressed in 32 types of human cancers.

Table 1
Numbers of downregulated pseudogenes among the four types of common gynecological malignancies from dreamBase.

Tumor types	Numbers of downregulated pseudogenes
cervical & endocervical cancer	140
uterine carcinosarcoma	0
ovarian serous cystadenocarcinoma	0
uterine corpus endometrioid carcinoma	103

3.2 Prognostic analysis of upregulated expressed pseudogenes in 32 types of human cancers.

Next, we explored the prognostic values of 40 up-regulated pseudogenes in 32 kinds of human cancers using GEPIA. As shown in Fig. 2, only KRT8P3, KRT8P45 and LDHAP5 predicted poor overall survivals in ovarian serous cystadenocarcinoma (HR = 1.3, P = 0.046; HR = 1.3, P = 0.019; HR = 1.3, P = 0.03). There were no other pseudogenes that were significantly correlated with poor prognosis in the four types of gynecological malignancies.

3.3 Investigation of pseudogenes-miRNA-mRNA regulatory network.

By searching the database of starBase v2.0, we found only LDHAP5 had its corresponding miRNAs. The specific characteristics of 9 retrieved miRNAs were shown in additional file 1 (Table S1). As shown in additional file 2 (Table S2), only hsa-miR-181d-5p, hsa-miR-181c-5p, hsa-miR-7-5p, hsa-miR-543, hsa-miR-151a-5p and hsa-miR-181b-5p had their own target gene, and 148 miRNA target genes that were validated at least by one of three strongly experimental method (reporter assay, Western blot and qRT-PCR) were identified through retrieving miRTarBase. The network of pseudogenes-miRNA-mRNA was constructed using Cytoscape v_3.7.2 (Fig. 3A).

3.4 KEGG pathways and Gene oncology (GO) enrichment analysis of miRNA target mRNAs.

148 miRNA target genes were imported into the STRING v_11.0, and then we performed GO and KEGG pathway enrichment analysis under the operation guidance of the website. We selected the top five significantly GO terms and KEGG pathways according to the values of FDR. The top five Biological Process (GO), Molecular Function (GO) and Cellular Component (GO) with their corresponding FDR values were shown in (Fig. 3B). The top five significantly KEGG pathways were MicroRNAs in cancer (hsa05206, FDR = 4.32E-26), Pathway in cancer (hsa05200, FDR = 6.77E-18), PI3K-AKT signaling pathway (hsa04151, FDR = 9.95E-16), Endocrine resistance (hsa01522, FDR = 9.95E-16), Foxo signaling pathway (hsa04068,

FDR = 2.65E-15) (Fig. 3C). These findings confirmed that pseudogene LDHAP5 may mediate the occurrence and progression of ovarian serous cystadenocarcinoma.

3.5 EGFR was identified as the target mRNA of LDHAP5 in ovarian serous cystadenocarcinoma.

We used the Centiscape plugin of Cytoscape v_3.7.2 to visualize the regulatory network of protein-protein constructed by STRING v_11.0 (Fig. 4). Then the top ten hub genes (TP53, MYC, EGFR, PTEN, HRAS, SIRT1, TNF, RELA and CREB1) were identified according to the values of Degree unDir (Table 2). We then further explored the sequence mutations and copy-number alterations of ten hub genes in ovarian serous cystadenocarcinoma using cBioportal. The group (TCGA, Nature 2011) which contained 489 patients was selected. However, only 361 complete patients (64.6%) were suitable for further analysis. The mutation frequencies of the ten hub genes were TP53 (96%,), MYC (34%), EGFR (9%), PTEN (14%), HRAS (9%), KRAS (24%), SIRT1 (10%), TNF (24%), RELA (11%) and CREB1 (10%), respectively (Fig. 5). Pearson correlation analysis showed that EGFR (R = 0.16, P = 0.00072), PTEN (R = 0.098, P = 0.043), SIRT1 (R = 0.094, P = 0.013), RELA (R = 0.18, P = 0.00013) and CREB1 (R = 0.16, P = 0.00094) were significantly correlated with LDHAP5 expression level in ovarian serous cystadenocarcinoma (Table 3). We found only EGFR (fold change = 1.192, P = 0.001), PTEN (fold change = 1.214, P = 0.007) and CREB1 (fold change = 1.723, P = 1.66E-04) mRNA were highly expressed in TCGA ovarian patients (n = 594) than normal patients (n = 8) using the database of Oncomine Main (Fig. 6A), and then we further analyzed the prognostic values (overall survival) of five hubs in ovarian serous cystadenocarcinoma using Kaplan–Meier plotter (Table 4, Fig. 6B). Only the EGFR was significantly correlated with poor outcome (HR = 1.51, 95%CI = 1.15-2, P = 0.0033) in ovarian serous cystadenocarcinoma, while the SIRT1 predicted good outcome (HR = 0.75, 95%CI = 0.57-1, P = 0.047). According to the pseudogene-miRNA-mRNA regulatory mechanism, we finally concluded that LDHAP5 may play potential roles in ovarian serous cystadenocarcinoma through targeting EGFR.

Table 2
The ten hub genes with their characters identified by cytoscape v_3.7.2.

Gene name	Betweenness unDir	Closeness unDir	Degree unDir
TP53	3022.961	0.006211	77
MYC	2046.146	0.005882	67
EGFR	999.453	0.005348	53
PTEN	813.2333	0.005348	53
HRAS	604.1613	0.005291	51
KRAS	636.6394	0.005208	48
SIRT1	272.0242	0.004808	37
TNF	406.4497	0.004808	36
RELA	260.4591	0.004785	35
CREB1	479.3882	0.004651	32

Table 3
Pearson correlation analysis between LDHAP5 and ten hub genes expression in ovarian serous cystadenocarcinoma using GEPIA.

Gene names	R	P
TP53	-0.022	0.65
MYC	0.0044	0.93
EGFR	0.16	0.00072
PTEN	0.098	0.043
HRAS	0.089	0.065
KRAS	0.0073	0.88
SIRT1	0.094	0.013
TNF	0.038	0.43
RELA	0.18	0.00013
CREB1	0.16	0.00094
*GEPIA: Gene expression profiling interactive analysis.		

Table 4
Prognostic values of five candidate hub genes in ovarian serous cystadenocarcinoma using Kaplan-Meier plotter.

Gene names	HRs with 95% CIs	P	Poor/good	FDR (%)
EGFR	1.51 (1.15-2)	0.0033	Poor	100
PTEN	1.29 (0.97–1.71)	0.08	Poor	> 50
SIRT1	0.75(0.57-1)	0.047	Good	> 50
RELA	1.22(0.94–1.59)	0.13	Poor	100
CREB1	0.85(0.64–1.13)	0.27	Good	100
CI: Confidence interval; HR: Hazard ratio; FDR: False discovery rate.				

4 Discussion

With the deepening of research, we have a better understanding of pseudogenes. Currently, there are two major classifications of pseudogenes. Firstly, pseudogenes are divided into three categories based on differences in structure and origin, namely duplicated pseudogenes, unitary pseudogenes and processed pseudogenes, respectively. Duplicated pseudogenes are caused by the mutation of gene coding region or regulatory region in the process of genome DNA tandem replication or chromosome unequal exchange.(25) Unitary pseudogenes cannot be transcribed and translated because of the spontaneous mutations in the coding region or regulatory region of a single copy gene with coding function.(26) Both duplicated pseudogenes and unitary pseudogenes are collectively called unprocessed pseudogenes. Processed pseudogenes are formed by the random integration of mRNA transcripts into cDNA, and lose their normal functions due to improper insertion site or sequence mutation.(27, 28) Secondly, according to the function of pseudogenes, pseudogenes can be classified into three categories: pseudogenes that can be transcribed, pseudogenes that cannot be transcribed, and pseudogenes that can encode short-chain peptides or truncated proteins, respectively. They play great roles in carcinogenesis and cancer prognosis in different ways.(29–31)

Based on the ceRNA hypothesis, our research focused on pseudogenes that can be transcribed into mRNA. We then further used the pseudogene-miRNA-mRNA regulatory network to identify pseudogenes that may play potential roles in common gynecological malignancies and explore their mechanism.

The initial goal of our study was to find pseudogenes that differentially expressed simultaneously in four common gynecological malignancies. However, we only found three significantly up-regulated pseudogenes (KRT8P3, KRT8P45 and LDHAP5) that predicted poor prognosis in ovarian serous cystadenocarcinoma after Kaplan–Meier survival analysis. With the deepening of our research, LDHAP5 was selected as the candidate pseudogenes for it has corresponding miRNAs. There are two reasons accounting for this phenomenon, the first one is that many pseudogenes remain unidentified so far. After

all, pseudogenes were considered as “junk” or “fossil” DNA at first, and many methods had been invented to avoid detecting pseudogenes.(32–36) The second possibility is that the current ceRNAs hypothesis is not yet perfect, and it needs to be further demonstrated to build a more comprehensive regulatory network.(37)

In our study, 148 potential target mRNA were identified. Functional enrichment analysis showed that MicroRNAs in cancer (hsa05206), Pathway in cancer (hsa05200), PI3K-AKT signaling pathway (hsa04151), Endocrine resistance (hsa01522), Foxo signaling pathway (hsa04068) were the top five significantly enriched gene sets. Interestingly, epithetlial ovarian cancer, bladder cancer, lung cancer, colorectal cancer, et. were enriched in the pathway of MicroRNAs in cancer (hsa05206). PI3K-AKT signaling pathway has been extensively studied and proven to play a great role in a variety of cancers. Studies have shown that activated AKT mediates downstream reactions through phosphorylation of a range of intracellular proteins, which include cell survival, growth, proliferation, cell migration, and angiogenesis.(38, 39) More significantly, many studies have shown that EGFR was dysregulated expressed in many solid tumors, and the PI3K-AKT signaling pathway can be used as a downstream regulatory pathway for EGFR to mediate the occurrence and progression of diseases, which has been confirmed in many cancers.(40, 41)

The shortcoming of our research is that our conclusion is mainly based on the analysis of existing databases. In order to further confirm the role of pseudogene LDHAP5, we need to construct ovarian cancer cell lines that differentially express LDHAP5 in future. We then will confirm our previous theoretical results in vivo and in vitro level, and even use clinical pathological specimens of ovarian cancer patients to further confirm. EGFR antagonists (gefitinib, lapatinib, erlotinib, etc.) have been used in a variety of cancers, such as pancreatic cancer, small cell lung cancer, colorectal cancer and so on.[42–44] Once our research is successfully validated, it may be used in ovarian cancer in future. With the deepening of research work, more functions of pseudogenes and corresponding mechanisms will be further revealed, and they will make contributes to identify more biomarkers, specific drug design, and the adoption of personalized treatment in the future.

5 Conclusion

To summarize, our study for the first time systematically elucidated the high expression of pseudogene LDHAP5 in ovarian serous cystadenocarcinoma, and it may lead to poor prognosis through targeting EGFR. It may serve as a new therapeutic target, and thereby improving the prognosis of patients with ovarian cancer in future.

Abbreviations

MREs
microRNA response elements; lncRNAs:long non-coding RNAs; ceRNAs:competitive endogenous RNAs; RBPs:RNA binding proteins; GEPIA:Gene Expression Profiling Interactive Analysis; GO:Gene oncology;

FDR:false discovery rate.

Declarations

Acknowledgements

Not applicable.

Ethical approval and informed consent

Not applicable.

Funding

This work was supported by National Natural Science Foundation of China (81630060), and the research-oriented clinician funding program of Tongji Medical College, Huazhong University of Science and Technology.

Availability of data and materials

Not applicable.

Conflict of interest statement

The author(s) declare no competing interests.

Consent for publication

Not applicable.

Authors' contribution

Peng Wu was responsible for the study concept and design; Shitong Lin, Canhui Cao, Ping Wu, Peipei Gao, Wenhua Zhi, Ting Peng were involved in data collection, data screening and statistical analysis;

Shitong Lin wrote the manuscript, and Yifan Meng took charge of supervising the manuscript. The final manuscript was approved by all the authors above.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: A Cancer Journal for Clinicians. 2020;70(1):7-30.
2. Fotopoulou C, Neumann U, Kraetschell R, Schefold JC, Weidemann H, Lichtenegger W, et al. Long-term clinical outcome of pelvic exenteration in patients with advanced gynecological malignancies. Journal of surgical oncology. 2010;101(6):507-12.
3. Tan XJ, Lang JH. [Key problems of chemotherapy for gynecological malignancies]. Zhonghua fu chan ke za zhi. 2012;47(8):561-3.
4. Jacq C, Miller JR, Brownlee GG. A pseudogene structure in 5S DNA of Xenopus laevis. Cell. 1977;12(1):109-20.
5. Ding W, Lin L, Chen B, Dai J. L1 elements, processed pseudogenes and retrogenes in mammalian genomes. IUBMB life. 2006;58(12):677-85.
6. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. 2010;465(7301):1033-8.
7. Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Carter DR. Pseudogenes: pseudo-functional or key regulators in health and disease? Rna. 2011;17(5):792-8.
8. Y T. Pseudogenes. Comp Funct Genomics. (2012:6–9).
9. Korrodi-Gregorio L, Abrantes J, Muller T, Melo-Ferreira J, Marcus K, da Cruz e Silva OA, et al. Not so pseudo: the evolutionary history of protein phosphatase 1 regulatory subunit 2 and related pseudogenes. BMC evolutionary biology. 2013;13:242.
10. Muro EM, Mah N, Andrade-Navarro MA. Functional evidence of post-transcriptional regulation by pseudogenes. Biochimie. 2011;93(11):1916-21.
11. Poliseno L, Marranci A, Pandolfi PP. Pseudogenes in Human Cancer. Frontiers in medicine. 2015;2:68.
12. Poliseno L. Pseudogenes: newly discovered players in human cancer. Science signaling. 2012;5(242):re5.
13. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146(3):353-8.
14. Hu X, Yang L, Mo YY. Role of Pseudogenes in Tumorigenesis. Cancers. 2018;10(8).
15. Zheng LL ZK, Liu S et al. DreamBase: DNA modification, RNA regulation and protein binding of expressed pseudogenes in human health and disease. Nucleic Acids Res. (46(D1), D85–D91 (2018)).
16. Tang Z LC, Kang B et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. . Nucleic Acids Res. (45(W1), W98–W102 (2017)).

17. Li JH LS, Zhou H et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. (42(Database issue), D92–D97 (2014)).
18. Chou CH SS, Yang CD et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res. (46(D1), D296–D302 (2018).).
19. Shannon P MA, Ozier O et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. (13(11), 2498–2504 (2003)).
20. Szklarczyk D MJ, Cook H et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. (45(D1), D362–D368 (2017)).
21. Scardoni G PM, Laudanna C. Analyzing biological network parameters with CentiScaPe. Bioinformatics. (25(21), 2857–2859 (2009)).
22. Cerami E GJ, Dogrusoz U et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery. (2(5), 401–404 (2012]).
23. Gyorffy B LaA, Sz´all´asi Z. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. Endocr Relat Cancer (19(2), 197–208 (2012)).
24. Rhodes DR K-SS, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P, et al. Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia. (9: 166-180, 2017.).
25. Mighell AJ, Smith NR, Robinson PA, Markham AF. Vertebrate pseudogenes. FEBS letters. 2000;468(2-3):109-14.
26. Zhang ZD, Frankish A, Hunt T, Harrow J, Gerstein M. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. Genome biology. 2010;11(3):R26.
27. Maestre J, Tchenio T, Dhellin O, Heidmann T. mRNA retroposition in human cells: processed pseudogene formation. The EMBO journal. 1995;14(24):6333-8.
28. D'Errico I, Gadaleta G, Saccone C. Pseudogenes in metazoa: origin and features. Briefings in functional genomics & proteomics. 2004;3(2):157-67.
29. Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature. 2008;453(7194):534-8.
30. Eichenlaub MP, Ettwiller L. De novo genesis of enhancers in vertebrates. PLoS biology. 2011;9(11):e1001188.
31. Kandouz M, Bier A, Carystinos GD, Alaoui-Jamali MA, Batist G. Connexin43 pseudogene is expressed in tumor cells and inhibits growth. Oncogene. 2004;23(27):4763-70.
32. Andersson JO, Andersson SG. Pseudogenes, junk DNA, and the dynamics of Rickettsia genomes. Molecular biology and evolution. 2001;18(5):829-39.

33. Williams TK YC, Brody J. . Does this band make sense? Limits to expression based cancer studies. Cancer Lett Brent MR. (2008;271:81–4. 8.).
34. MR B. Steady progress and recent breakthroughs in the accuracy of automated genome annotation. Nat Rev Genet (2008;9:62–73.).
35. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science. 1969;165(3891):349-57.
36. Britten RJ, Davidson EH. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. The Quarterly review of biology. 1971;46(2):111-38.
37. Cheetham SW, Faulkner GJ, Dinger ME. Overcoming challenges and dogmas to understand the functions of pseudogenes. Nature reviews Genetics. 2019.
38. Hoxhaj G, Manning BD. The PI3K–AKT network at the interface of oncogenic signalling and cancer metabolism. Nature Reviews Cancer. 2019;20(2):74-88.
39. Murugan AK. Special issue: PI3K/Akt signaling in human cancer. Seminars in cancer biology. 2019;59:1-2.
40. Fu X, Cui G, Liu S, Zhao S. Linc01014 regulates gefitinib resistance in oesophagus cancer via EGFR-PI3K-AKT-mTOR signalling pathway. Journal of cellular and molecular medicine. 2020;24(2):1670-5.
41. Zhang F, Xu M, Yin X, Guo H, Zhang B, Wang Y, et al. TWEAK promotes hepatic stellate cell migration through activating EGFR/Src and PI3K/AKT pathways. Cell biology international. 2019.
42. Wang Z, Cheng Y, An T, Gao H, Wang K, Zhou Q, et al. Detection of EGFR mutations in plasma circulating tumour DNA as a selection criterion for first-line gefitinib treatment in patients with advanced lung adenocarcinoma (BENEFIT): a phase 2, single-arm, multicentre clinical trial. The Lancet Respiratory Medicine. 2018;6(9):681-90.
43. Pyrotinib Tops Lapatinib in Metastatic Breast Cancer. Cancer discovery. 2019;9(11):OF3.
44. Sidaway P. Erlotinib effective in the neoadjuvant setting. Nature reviews Clinical oncology. 2018;15(11):654.

Figures

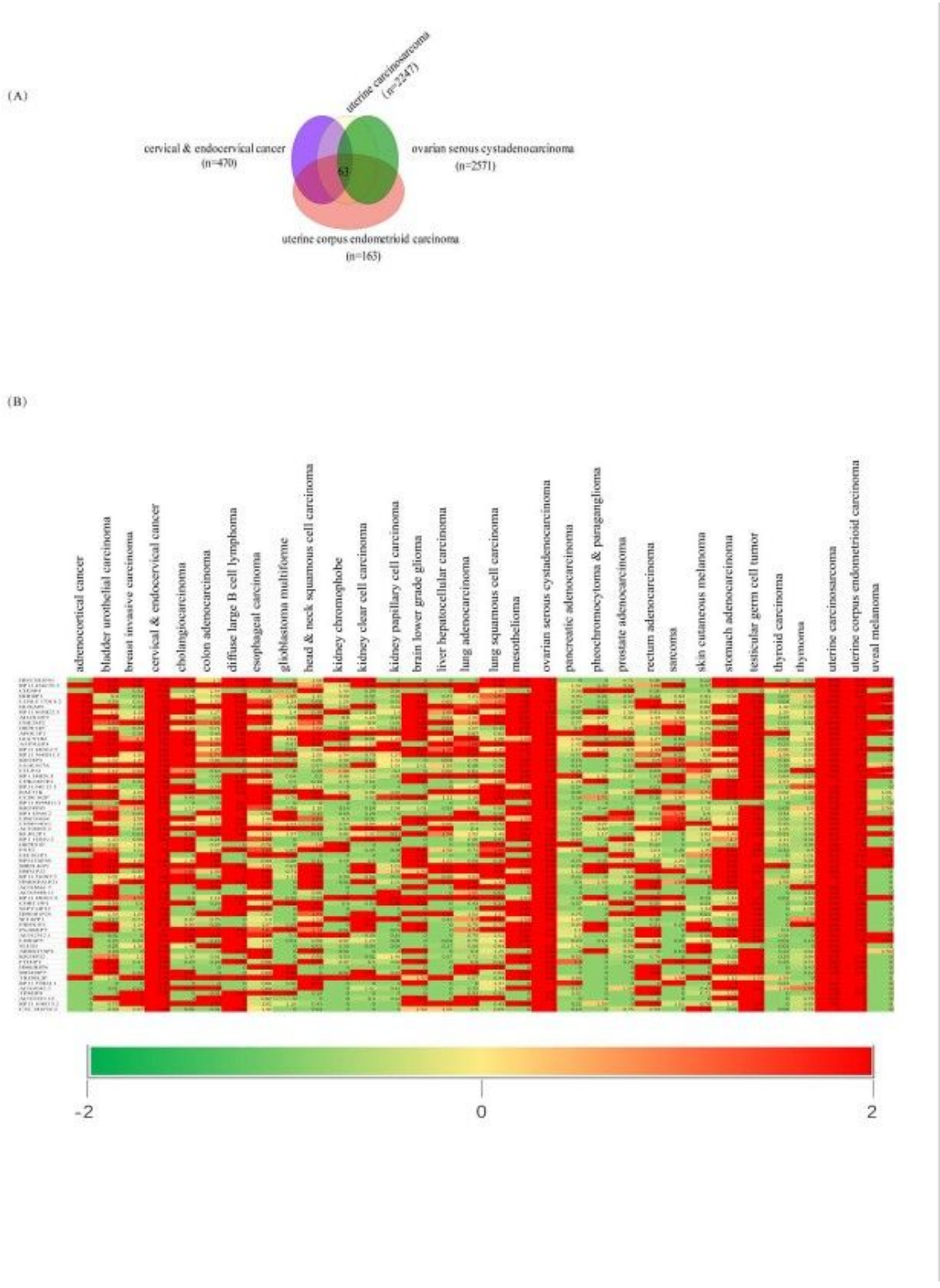

Figure 1

Identification of differentially expressed pseudogenes in four types of gynecological malignancies. RNA-Seq expression data of pseudogenes in 32 kinds of common human cancers were downloaded from the dataset of dreamBase (http://rna.sysu.edu.cn/dreamBase/index.php). |log2 FC| > 2.0 was set as the cutoff to distinguish differentially expressed pseudogenes. (A) 63 pseudogenes that were simultaneously expressed in four common gynecological malignancies (cervical & endocervical cancer, ovarian serous

cystadenocarcinoma, uterine corpus endometrioid carcinoma and uterine carcinosarcoma). (B) Heat map of 63 frequently upregulated pseudogenes in 32 types of human cancers. Red represents upregulated genes and green represents downregulated genes. The values in the boxes represent the |log2 FC| values.

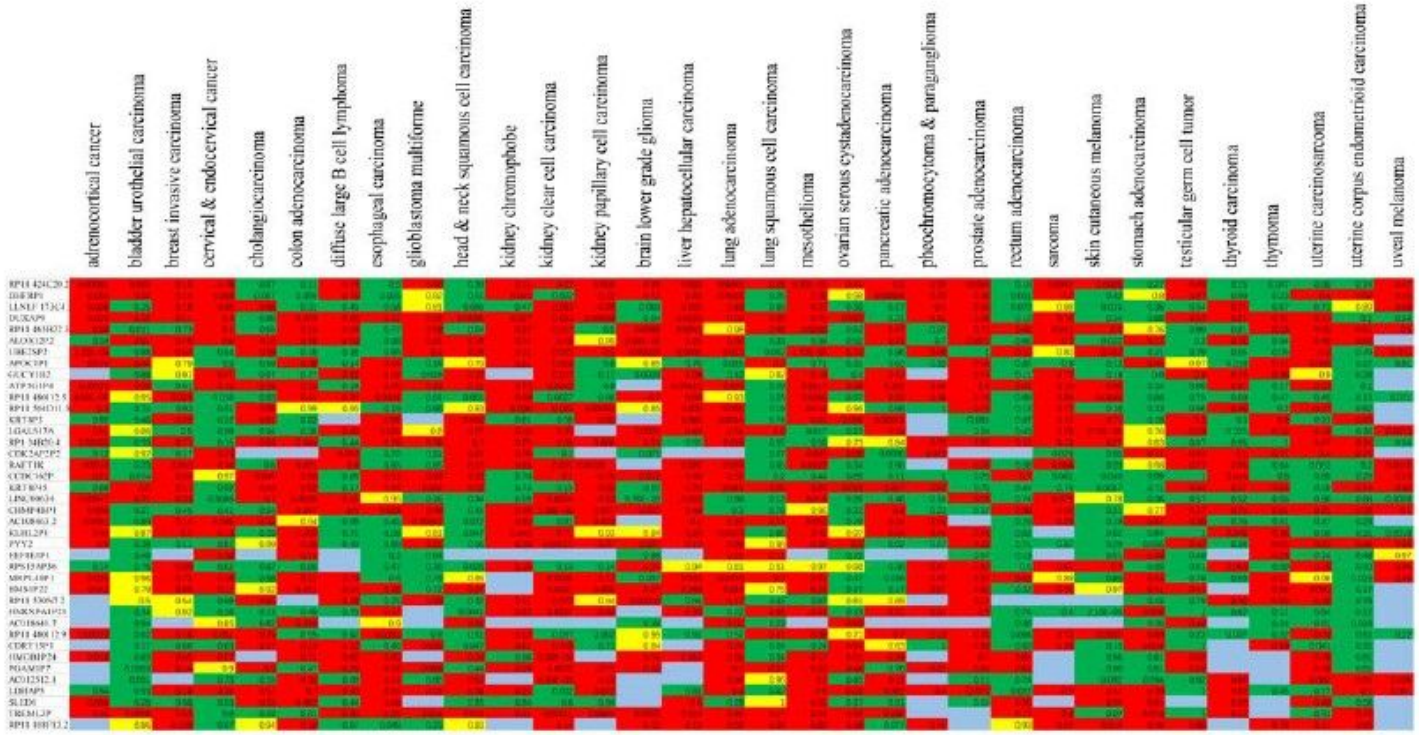

Figure 2

Prognostic values of 41 upregulated pseudogenes in 32 kinds of human cancers using GEPIA. The group thresholds were as follows: the group cut-off was ‘Median’, the ‘cutoff-high’ and ‘cutoff-low’ were 50%, axis units were ‘Months’, and p value < 0.05 was considered statistically significant. Red represents poor outcome, green represents good outcome, yellow represents neutral outcome (hazard ratio=1), and wathet means that “The group thresholds you set are too strict. The sample size is insufficient at your custom thresholds”. The values in the boxes represent the P values. GEPIA: gene expression profiling interactive analysis.

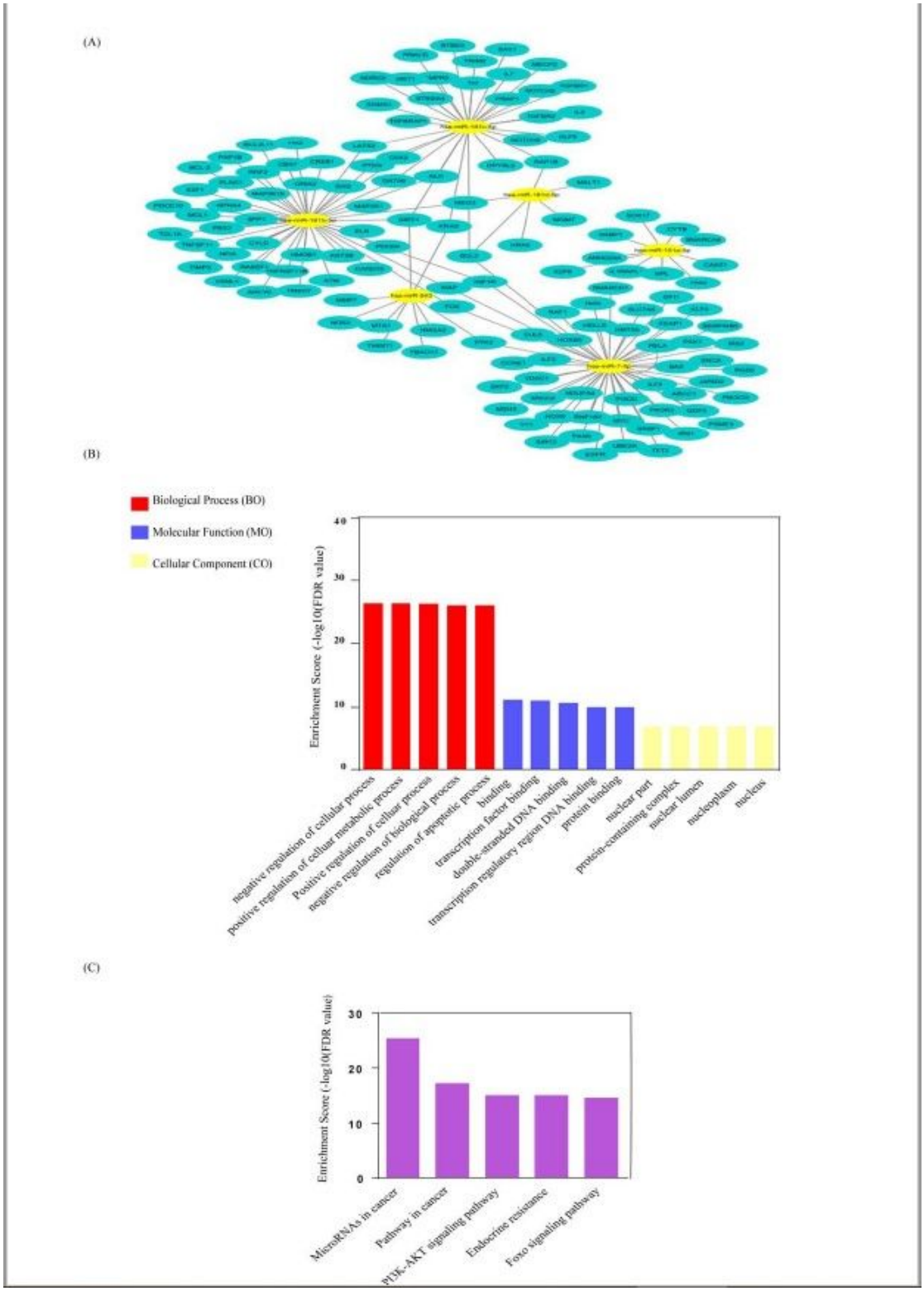

Figure 3

Construction of pseudogenes-miRNA-mRNA regulatory network and KEGG pathways and GO enrichment analysis of 148 miRNA target mRNAs. (A) The candidate miRNAs were identified by searching starBase v2.0. 148 target genes that were validated at least by one of three strongly experimental method (reporter assay, Western blot and qRT-PCR) were identified through retrieving miRTarBase. The network of pseudogenes-miRNA-mRNA was constructed using Cytoscape v_3.7.2. (B) 148 miRNA target mRNAs

were divided into three functional groups: biological processes, cellular components and molecular functions. The top five GO enrichments were shown according to the values of FDR. (C) The top five KEGG pathways according to the values of FDR. GO, Gene oncology; FDR, false discovery rate.

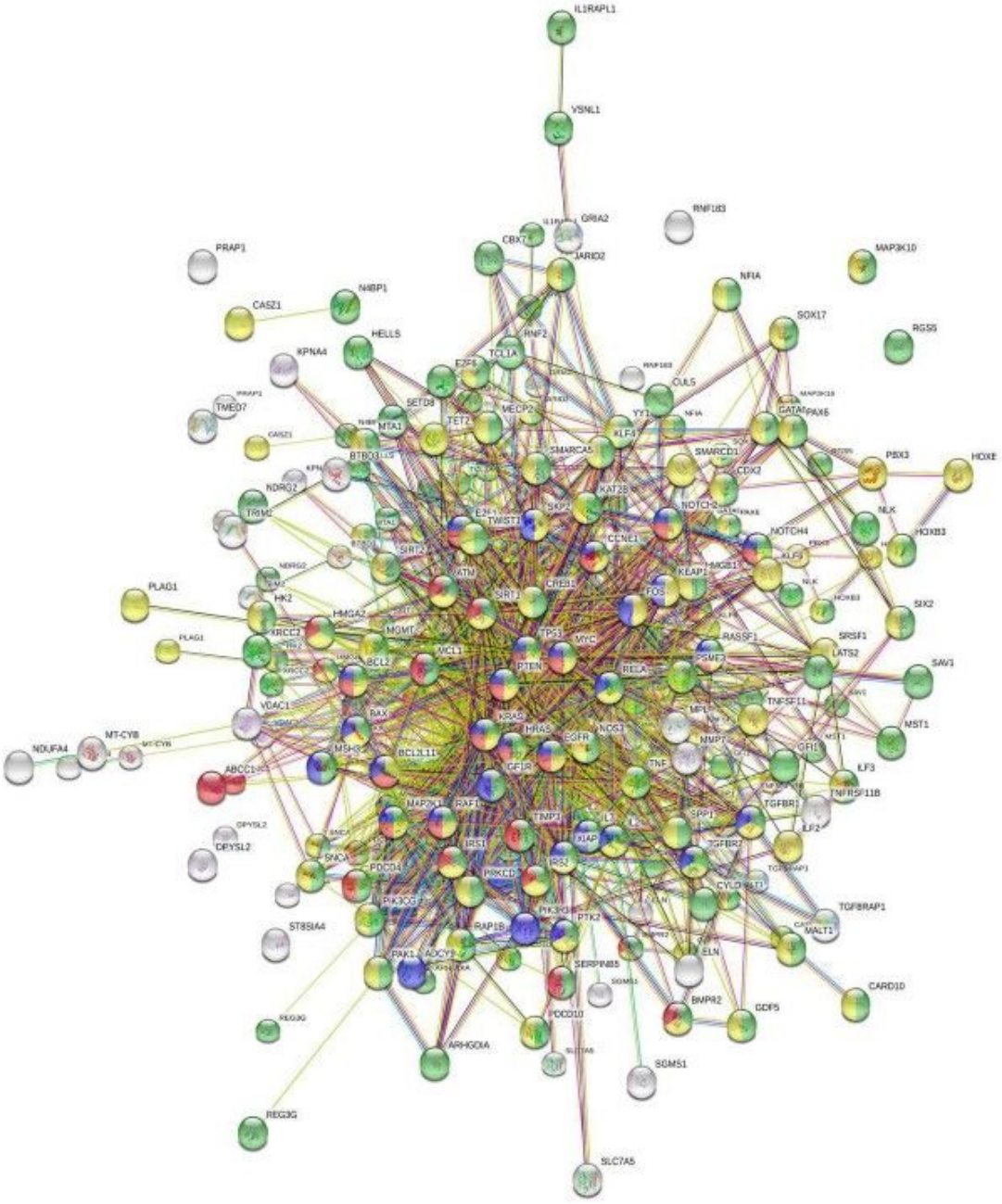

Figure 4

Identification of potential target genes of LDHAP5. The protein-protein interaction network of 148 genes was constructed using STRING v_11.0.

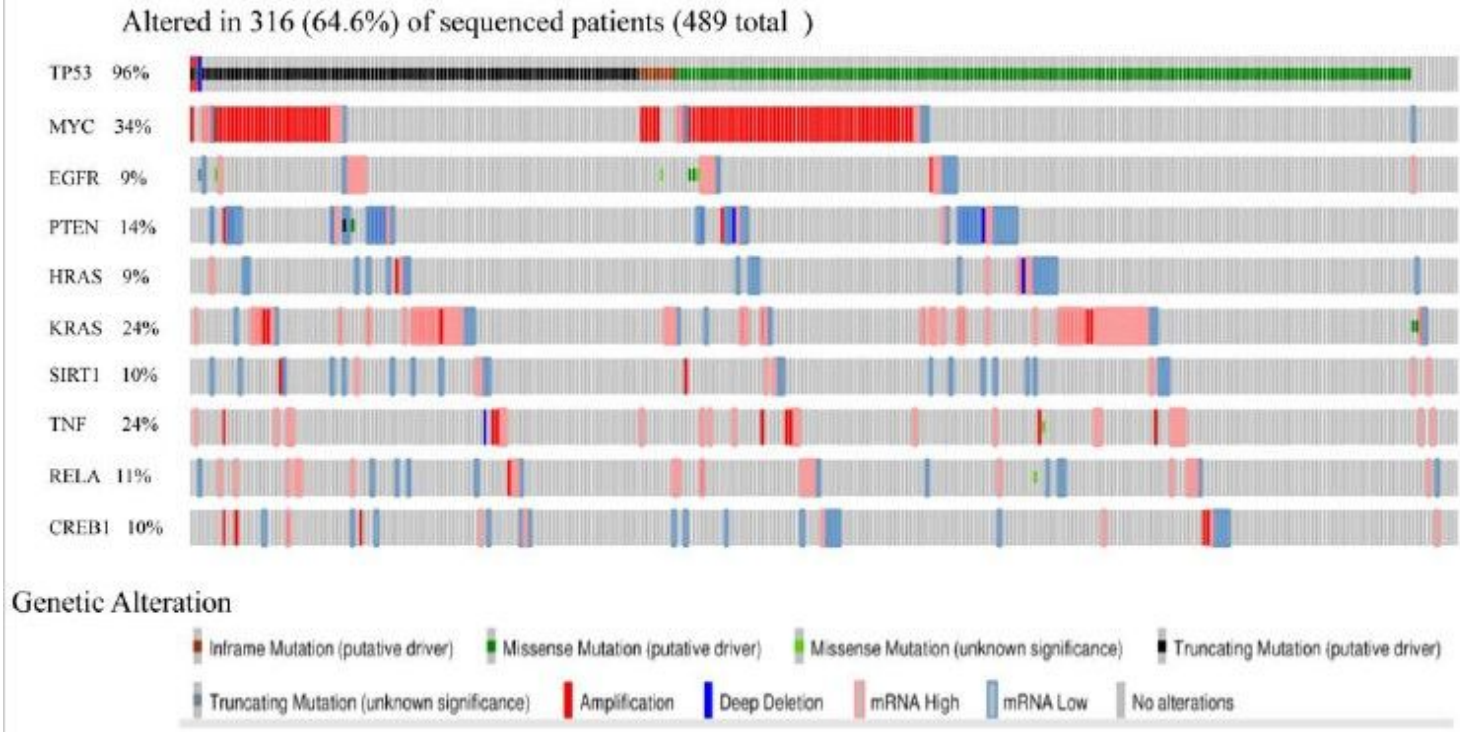

Figure 5

Ten hub genes expression and mutation analysis in ovarian serous cystadenocarcinoma (TCGA, Nature 2011) using cBioportal. OncoPrint in c-Bioportal showed the types of ten hub genes variations in ovarian serous cystadenocarcinoma and their corresponding proportions. TCGA: The Cancer Genome Atlas.

Supplementary Files

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

- TableS1.docx
- TableS2.docx