European Review for Medical and Pharmacological Sciences 2013; 17: 311-317

Expression profiles analysis of pancreatic cancer

D. YANG, Z. ZHU, W. WANG, P. SHEN, Z. WEI, C. WANG, Q. CAI

Department of Gastrointestinal Surgery, Shanghai Changzheng Hospital affiliated to the Second Military Medical University, Shanghai, China

Abstract. – BACKGROUN: Pancreatic cancer is the fourth most common cause of cancer-related deaths across the globe and has a poor prognosis.

AIM: To investigate the characteristics of genomic expression profiles of pancreatic cancer and screen differentially expressed genes.

MATERIALS AND METHODS: Using GSE16515 dataset downloaded from GEO (Gene Expression Omnibus) database, we first screened the differentially expressed genes (DEGs) in pancreatic cancer by packages in R language. The key functions of DEGs were investigated by GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis. The potential functionally important SNP (Single Nucleotide Polymorphism) was selected from the dbSNP database.

RESULTS: A total of 1270 DEGs were identified. Most of them were predicted to be involved in pancreatic cancer development by sequence variant. Six genes (CDC42, STAT1, RALA, BCL2L1, TGFA, and EGF) were enriched in the known pancreatic cancer pathway. All these six genes had SNP, usually mutation at A/G and C/T point.

CONCLUSIONS: Our results provide some underlying biomarkers for early diagnosis of pancreatic cancer.

Key Words:
Pancreatic cancer, Genomic expression profiles, Single nucleotide polymorphism.

Introduction

Pancreatic cancer is a common gastrointestinal malignancy, with an estimated 44,030 new cases in 2011[1]. Early stages of pancreatic cancer are truly asymptomatic and the majority of patients are diagnosed at an advanced stage. Once cancer is diagnosed, the disease is associated with poor prognosis. The five-year survival rate is usually less than 5% and nearly zero in advanced cancer[2], although several anticancer regimens have been recommended to treat pancreatic cancer. Therefore, it is urgently to develop fine approaches for early diagnosis of pancreatic cancer.

Recently, the development of pancreatic cancer has been attributable to the over-expression of several oncogenes (c-myc[3,4] and K-ras[4,5]), inactivation tumor suppressor genes (Kruppel-like factor 6[6] and p53[7]) or the deregulation of various signaling pathway (Hedgehog[8] and PI3K/Akt[9]). In addition, several molecular markers have been applied for early prediction of pancreatic tumor. For example, CA19-9 (carbohydrate antigen 19.9) is a tumor marker that is frequently elevated in pancreatic cancer. Elevated preoperative CA19-9 is significantly associated with lymph node involvement, tumor ≥ 3 cm, and lack of tumor differentiation[10]. However, any sensitive enough and highly specific tumor markers for early pancreatic cancer screening have not been identified[11].

Global gene-expression profiling and the use of microarray databases have allowed simultaneous identification of hundreds of thousands of genes in pancreatic cancer[12], which is helpful for screening more molecular biomarkers. In this study, we aimed to investigate the characteristics of genomic expression profiles of pancreatic cancer and screen differentially expressed genes (DEGs) by using the GSE16515 microarray data. The underling functions of these crucial genes were explored by Gene Ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. In addition, we also analyzed the single nucleotide polymorphism (SNP) of these genes, which is beneficial to developing personalized therapeutic strategy.

Materials and Methods

Expression Profiles Related to Pancreatic Cancer

Expression profile of GSE16515[13] was obtained from a public functional genomics data repository National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO, http://www.ncbi.nlm.nih.gov/geo/) which

Corresponding Author: Qingping Cai, MD; e-mail: caiqingpingasdf@hotmail.com

is based on the Affymetrix GPL570 platform data (Affymetrix Human Genome U133A 2.0 Array). This chip contained 52 samples, of which 36 and 16 pancreatic samples were from pancreatic cancer patients and normal control, respectively. In all samples, the ratio of male to female was 17:9, and ages were from 49 to 84.

Pre-processing of Expression Profile

The original CEL (Celestia) files and the platform probe annotation information file were also downloaded for the next step of bioinformatics analysis. The raw data were transformed into expression values using the affy package in R language, and then the missing part of the data was filled[14]. Expression data were normalized using the global median normalization method[15].

Analysis of Differentially Expressed Genes (DEGs)

Significance of gene expression differences between normal and cancer sample was tested by using three statistical test methods: *t*-test, Wilcox-test and Fisher exact-test of multtest-package in R language[16]. To circumvent the multi-test problem which may induce too much false positive results, all p-values in three methods were adjusted into false discovery rate (FDR) used Benjamini and Hochberg method[17]. Only the genes with FDR < 0.05 and |logFC (fold change)| > 1 were selected as DEGs.

Enrichment Analysis of Gene Function

In order to find the function enrichment of DEGs on cell level, we used the database GO (Gene Ontology)[18] to classify the gene function and location information. We performed GO clustering analysis by cluster Profiler package[19,20], then deduced the affection of these DEGs to the cells by cluster the cells within the molecular functions and biological processes.

The database for annotation, visualization and integrated discovery (DAVID)[21] was used to identify over-represented GO terms in biological process. *p*-value < 0.05 and Fold Enrichment > 1 were set as the threshold for the analysis using the hypergeometric distribution.

Pathway Analysis

In order to understand the pathways and genes related to pancreatic cancer on the cellular level, all DEGs were mapped into the KEGG pathway database[20,22] for the signaling pathways and genes of pancreatic cancer.

The DAVID[21] was also used to identify over-represented KEGG pathway based on hypergeometric distribution. *p*-value < 0.05 and FDR < 0.01 were set as the threshold for the significant pathway analysis.

SNP Screening

The potential functionally important SNP was selected from the dbSNP NCBI database (http://www.ncbi.nlm.nih.gov/SNP) using pancreatic cancer related DEGs.

Results

Differentially Expressed Genes Analysis

Due to various reasons, such as background and the probe design, the original microarray data cause a great difference between the microarray data, so the data must be normalized. After data preprocessing, gene expression profile data with higher normalization was used for DEGs analysis (Figure 1). The differentially expressed genes were test by three statistical tests and multiple test screening. A total of 1,380 genes were selected as DEGs with the *p*-value < 0.05 and |logFC| > 1. All data are shown in Table I.

GO Enrichment Analysis

All DEGs were enriched into 19 functional clusters based on the cutoff of *p*-value < 0.05 and Fold Enrichment > 1. It can be seen from Figure 2, in which 1270 DEGs (accounting for more than 92% of the total genes) were enriched in sequence variant.

KEGG Pathways Analysis

All genes were mapped into KEGG database. According to the threshold of hypergeometric distribution $p < 0.05$ and FDR < 0.01, the DEGs were enriched into a total of 26 pathways (Table II). Among them, 9 pathways were associated with cancer. Hsa05212 (pancreatic cancer pathway) showed the minimum *p* value, in which five up-regulated genes (CDC42, STAT1, RALA, BCL2L1 and TGFA) and one down-regulated gene (EGF) were involved.

Comparing With the Known SNP

We searched the known SNPs of the above 6 genes in dbSNP of NCBI (National Center for Biotechnology Information) database. As shown in Table III, SNP phenomenon of these genes

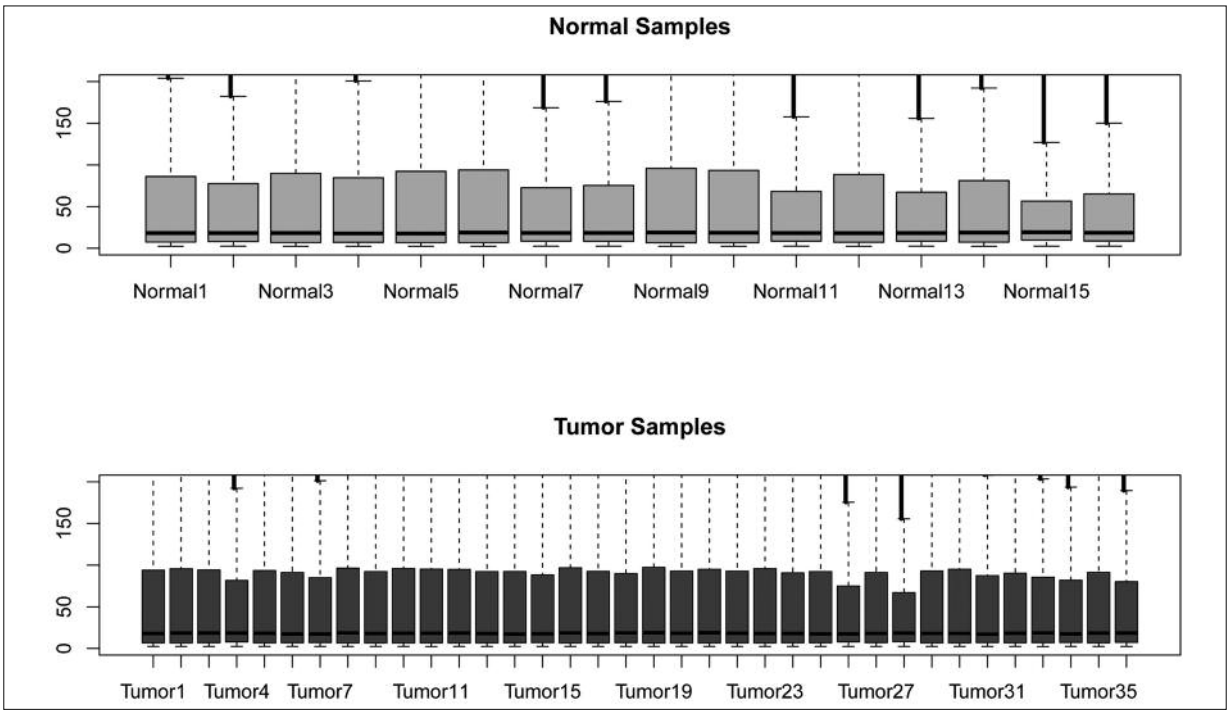

Figure 1. Normalized expression value data box plots. The gray boxes show 16 normal pancreatic sample; Black ones show cancer patients samples. Black line in each box represents the median of each sample. All the black lines are almost in the same position, which indicates high degree of standardization.

existed in humans. Most of them have been cited by PubMed, but not been confirmed by the clinical study.

After statistical analysis, we found that the SNP of 6 genes focused on mutation at A/G (4/6) and C/T (2/6) point (A, adenine; G, guanine; C, cytosine; T, thymine).

Discussions

In this study, we systemically investigated the differential gene expression profiling of pancreatic cancer patients and explore their possible roles by GO and pathway enrichment analyses. The silent results included 1270 DEGs had been identified and 6 genes of them (CDC42, STAT1, RALA, BCL2L1, TGFA, and EGF) were involved in pancreatic cancer pathway. They may participate in pancreatic cancer development and progression by sequence variant. These seemed to be in accordance with the previous studies as following:

CDC42 (cell division cycle 42), a Rho GTPase family member, has been implicated as an important regulator of cell transformation, proliferation, survival, invasion and metastasis of human cancer cells[23]. CDC42 has been reported to be

Table I. Some of differentially expressed genes.

Gene	*t*-test	adj	Wilcox	adj	Exact_test	adj	logFC
RBM20	0.019727	0.047474	0.000386	0.004237	3.22E-99	6.06E-96	-3.991628444
PM20D1	0.001608	0.007129	0.001788	0.010518	2.53E-99	4.93E-96	-3.958876991
GNMT	7.46E-05	0.001298	0.002417	0.012389	2.44E-113	8.88E-110	-3.940700753
TRHDE	0.000341	0.002598	0.002339	0.01217	2.13E-110	6.84E-107	-3.916501301
RBPJL	1.56E-05	0.000721	0.001937	0.01099	3.35E-105	9.15E-102	-3.760315752
ALB	0.00012	0.001533	0.004439	0.017196	1.13E-98	1.99E-95	-3.545989557
CTRL	1.51E-05	0.000712	0.005532	0.019305	1.37E-84	1.27E-81	-3.257324281
EGF	6.79E-05	0.001257	0.005677	0.019566	2.33E-75	1.72E-72	-3.118090632

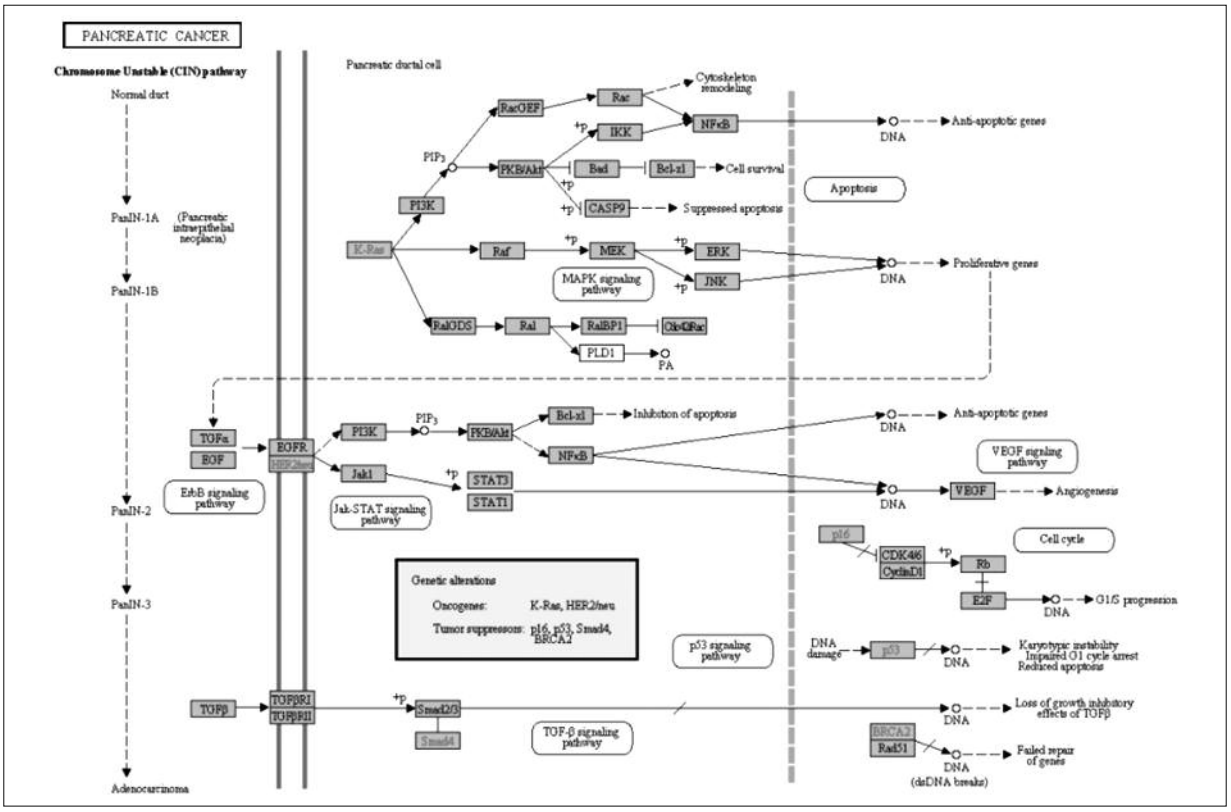

Figure 2. Proportion of genes from each functional cluster. The different colors indicated different functions.

Table II. Pathways analysis.

Term	*p*-value	FDR
hsa05212: Pancreatic cancer	2.03E-89	2.10E-86
hsa05200: Pathways in cancer	1.81E-51	1.87E-48
hsa05220: Chronic myeloid leukemia	2.72E-34	2.82E-31
hsa05223: Non-small cell lung cancer	1.06E-25	1.10E-22
hsa05210: Colorectal cancer	1.61E-23	1.67E-20
hsa05215: Prostate cancer	5.39E-23	5.60E-20
hsa05219: Bladder cancer	6.60E-22	6.85E-19
hsa05214: Glioma	7.23E-19	7.51E-16
hsa05218: Melanoma	5.17E-18	5.37E-15
hsa04010: MAPK signaling pathway	1.11E-13	1.15E-10
hsa05213: Endometrial cancer	2.49E-13	2.58E-10
hsa05221: Acute myeloid leukemia	9.04E-13	9.39E-10
hsa05222: Small cell lung cancer	2.54E-12	2.63E-09
hsa05211: Renal cell carcinoma	7.99E-12	8.30E-09
hsa04110: Cell cycle	4.91E-09	5.10E-06
hsa04370: VEGF signaling pathway	9.48E-09	9.84E-06
hsa04662: B cell receptor signaling pathway	9.48E-09	9.84E-06
hsa04062: Chemokine signaling pathway	3.24E-08	3.36E-05
hsa04012: ErbB signaling pathway	3.59E-08	3.73E-05
hsa04722: Neurotrophin signaling pathway	6.36E-08	6.60E-05
hsa04510: Focal adhesion	7.26E-08	7.54E-05
hsa04520: Adherens junction	2.18E-07	2.27E-04
hsa04810: Regulation of actin cytoskeleton	1.34E-06	0.00139
hsa04660: T cell receptor signaling pathway	3.00E-06	0.003112
hsa04210: Apoptosis	7.72E-06	0.008009
hsa05120: Epithelial cell signaling in Helicobacter pylori infection	2.16E-05	0.022467

Table III. 6 gene SNP search results.

Gene	Cited in PubMed	Clinical source	Human (active)
EGF	2	0	68
BCL2L1	1	0	43
CDC42	1	0	48
RALA	0	0	47
TGFA	11	0	74
STAT1	12	0	36

overexpressed in testicular cancer[24], colorectal cancer[25], lung cancer[26], and breast cancer[27]. There are few reports about CDC42 expression in pancreatic cancer. CDC42 may promotes migration of cancer cells through 1 integrin[28].

Mucin 4 (MUC4) is a transmembrane mucin, which is aberrantly expressed in pancreatic adenocarcinoma with no detectable expression in the normal pancreas. Recent study indicates that MUC4 expression is mediated by stimulation of STAT1 (signal transducer and activator of transcription 1) expression and this effect is abrogated by short interfering RNA-mediated inhibition of STAT1 expression[29]. Unphosphorylated or serine-phosphorylated STAT1 can act as transcription factors of MUC4, either alone by progressive binding to different sites in the promoter or both together[30].

RAL small GTPases, encoded by the RALA and RALB genes, are members of the RAS superfamily of small GTPases and can act as downstream effectors of RAS. RALA, but not Ralb expression has been demonstrated to promote the tumorigenic growth of the fibrosarcoma, bladder- and colon-cancer and pancreatic cancer cell lines[31,32]. However, using a Kras-driven non-small cell lung carcinoma mouse model, either RALA or RALB is found to be sufficient for tumor growth[33].

BCL2L1 protein (also known as Bcl-XL) belongs to the BCL-2 protein family and play important role in antiapoptosis. Bcl-XL is found to be high expressed in both pancreatic cancer tissues and cell lines[34]. Patients whose tumors exhibited no, faint, or weak bcl-xL expression tended to live longer after tumor resection (median 12 months) than patients whose tumors exhibited moderate bcl-XL mRNA expression (median 5 months)[35].

Mice with transgenic overexpression of transforming growth factor, alpha (TGFα) under control of the pancreatic Elastase promoter (*Ela-Tgfa*)

Table IV. Mutation and statistical number of each gene.

Gene	SNP	Num	Gene	SNP	Num
EGF	G/T	1	RALA	-/A/T/TAA	1
	A/C	4		C/G	1
	A/T	4		A/T	3
	C/G	6		A/C	4
	G/T	9		G/T	6
	C/T	16		C/T	15
BCL2L1	G/T	1	STAT1	A/C	2
	G/T	3		A/T	3
	A/C	5		G/T	5
	C/G	5		C/G	5
	A/G	9		C/T	10
	C/T	11		A/G	11
CDC42	C/G/T	1	TGFA	C/G/T	1
	-/A/G/T	1		A/C/G	1
	A/T	1		A/T	1
	A/C/T	1		-/A/C	2
	A/C	2		G/T	5
	G/T	5		C/G	6
	C/G	5		A/C	8
	C/T	14		A/G	20
	A/G	21		C/T	31

can develop pancreatic tumors[36]. TGF-α is able to promote tumor progression throughout induction of cyclin D1-Cdk4 without increase of cyclin E or proliferating cell nuclear antigen in ductal lesions[37]. Concomitant expression of TGFα and $Kras^{G12D}$ accelerates the progression of pancreatic intraepithelial neoplasia lesions to metastatic pancreatic cancer and leads to the development of cystic papillary lesions resembling human intraductal papillary mucinous neoplasms[38].

Conclusions

Present findings shed new light on the biology of pancreatic cancer and have implications for future research. The genes screened in this analysis may serve as a genetic marker for diagnosis of pancreatic cancer because they have SNP region. However, further clinical trials are needed to validate our conclusions.

References

1) Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA: A Cancer Journal for Clinicians. 2012.
2) Farlay J, Parkin D, Steliarova-Foucher E. Estimates of cancer incidence and mortality in europe in 2008. Eur J Cancer 2010; 46: 765-781.
3) Lewis BC, Klimstra DS, Varmus HE. The c-myc and PyMT oncogenes induce different tumor types in a somatic mouse model for pancreatic cancer. Genes Dev 2003; 17: 3127-3138.
4) Shi XH, Liang ZY, Ren XY, Liu TH. Combined silencing of K-ras and Akt2 oncogenes achieves synergistic effects in inhibiting pancreatic cancer cell growth *in vitro* and *in vivo*. Cancer Gene Ther 2008; 16: 227-236.
5) Collins MA, Bednar F, Zhang Y, Brisset JC, Galbán S, Galbán CJ, Rakshit S, Flannagan KS, Adsay NV, Pasca di Magliano M. Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice. J Clin Invest 2012; 122: 639-653.
6) Hartel M, Narla G, Wente MN, Giese NA, Martignoni ME, Martignetti JA, Friess H, Friedman SL. Increased alternative splicing of the KLF6 tumour suppressor gene correlates with prognosis and tumour grade in patients with pancreatic cancer. Eur J Cancer 2008; 44: 1895-1903.
7) Morton JP, Timpson P, Karim SA, Ridgway RA, Athineos D, Doyle B, Jamieson NB, Oien KA, Lowy AM, Brunton VG, Frame MC, Ashworth A, Oien KA, Evans TR, Sansom OJ. Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. Proc Natl Acad Sci 2010; 107: 246-251.
8) Tian H, Callahan CA, DuPree KJ, Darbonne WC, Ahn CP, Scales SJ, De Sauvage FJ. Hedgehog signaling is restricted to the stromal compartment during pancreatic carcinogenesis. Proc Natl Acad Sci 2009; 106: 4254-4259.
9) Ma J, Sawai H, Matsuo Y, Ochi N, Yasuda A, Takahashi H, Wakasugi T, Funahashi H, Sato M, Takeyama H. IGF-1 mediates PTEN suppression and enhances cell invasion and proliferation via activation of the IGF-1/PI3K/Akt signaling pathway in pancreatic cancer cells. J Surg Res 2010; 160: 90-101.
10) Martin LK, Wei L, Trolli E, Bekaii-Saab T. Elevated baseline CA19-9 levels correlate with adverse prognosis in patients with early-or advanced-stage pancreas cancer. Med Oncol 2012: 29: 3101-3107.
11) Sharma C, Eltawil KM, Renfrew PD, Walsh MJ, Molinari M. Advances in diagnosis, treatment and palliation of pancreatic carcinoma: 1990-2010. World J Gastroenterol 2011; 17: 867-897.
12) Sudo H, Mizoguchi A, Kawauchi J, Akiyama H, Takizawa S. Use of non-amplified RNA samples for microarray analysis of gene expression. PLoS One 2012; 7: e31397.
13) Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, Lingle W, Petersen G, Lou Z, Wang L. FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. Cancer Cell 2009; 16: 259-266.
14) Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. Bioinformatics 2001; 17: 520-525.
15) Fujita A, Sato JR, Rodrigues LO, Ferreira CE, Sogayar MC. Evaluating different methods of microarray data normalization. BMC Bioinformatics 2006; 7: 469.
16) Kim SY, Lee JW, Sohn IS. Comparison of various statistical methods for identifying differential gene expression in replicated microarray data. Stat Methods Med Res 2006; 15: 3-20.
17) Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc. Series B (Methodol) 1995; pp. 289-300.
18) Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. Gene ontology: tool for the unification of biology. Nat Genet 2000; 25: 25.
19) Da Wei Huang BTS, Lempicki RA. Systematic and integrative analysis of large gene lists using david bioinformatics resources. Nat Protoc 2008; 4: 44-57.
20) Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37: 1-13.
21) Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using david bioinformatics resources. Nat Protoc 2009; 4: 44-57.

22) Team RR. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2010.

23) Stengel K, Zheng Y. Cdc42 in oncogenic transformation, invasion, and tumorigenesis. Cell Signal 2011; 23: 1415-1423.

24) Kamai T, Yamanishi T, Shirataki H, Takagi K, Asami H, Ito Y, Yoshida KI. Overexpression of RhoA, Rac1, and Cdc42 GTPases is associated with progression in testicular cancer. Clin Cancer Res 2004; 10: 4799-4805.

25) Gómez DPT, Valdés-Mora F, Bandrés E, Pérez-Palacios R, Espina C, Cejas P, García-Cabezas MA, Nistal M, Casado E, González-Barón M. Cdc42 is highly expressed in colorectal adenocarcinoma and downregulates id4 through an epigenetic mechanism. Int J Oncol 2008; 33: 185-193.

26) Chen QY, Jiao DM, Yao QH, Yan J, Song J, Chen FY, Lu GH, Zhou JY. Expression analysis of Cdc42 in lung cancer and modulation of its expression by curcumin in lung cancer cell lines. Int J Oncol 2012; 40: 1561-1568.

27) Zuo Y, Wu Y, Chakraborty C. Cdc42 negatively regulates intrinsic migration of highly aggressive breast cancer cells. J Cell Physiol 2012; 227: 1399-1407.

28) Reymond N, Im JH, Garg R, Vega FM, d'Agua BB, Riou P, Cox S, Valderrama F, Muschel RJ, Ridley AJ. Cdc42 promotes transendothelial migration of cancer cells through β1 integrin. J Cell Biol 2012; 199: 653-668.

29) Andrianifahanana M, Singh AP, Nemos C, Ponnusamy MP, Moniaux N, Mehta PP, Varshney GC, Batra SK. IFN-γ-induced expression of MUC4 in pancreatic cancer cells is mediated by STAT-1 upregulation: a novel mechanism for IFN-γ response. Oncogene 2007; 26: 7251-7261.

30) Kossow C, Jose D, Jaster R, Wolkenhauer O, Rateitschak K. Mathematical modelling unravels regulatory mechanisms of interferon-γ-induced STAT1 serine-phosphorylation and MUC4 expression in pancreatic cancer cells. IET Syst Biol 2012; 6: 73-85.

31) Lim KH, Baines AT, Fiordalisi JJ, Shipitsin M, Feig LA, Cox AD, Der CJ, Counter CM. Activation of RalA is critical for Ras-induced tumorigenesis of human cells. Cancer Cell 2005; 7: 533-545.

32) Lim KH, O'Hayer K, Adam SJ, Kendall SDS, Campbell PM, Der CJ, Counter CM. Divergent roles for RalA and RalB in malignant growth of human pancreatic carcinoma cells. Curr Biol 2006; 16: 2385-2394.

33) Peschard P, McCarthy A, Leblanc-Dominguez V, Yeo M, Guichard S, Stamp G, Marshall CJ. Genetic deletion of RALA and RALB small GTPases reveals redundant functions in development and tumorigenesis. Curr Biol 2012; 22: 2063-2068.

34) Miyamoto Y, Hosotani R, Wada M, Lee JU, Koshiba T, Fujimoto K, Tsuji S, Nakajima S, Doi R, Kato M. Immunohistochemical analysis of bcl-2, bax, bcl-x, and mcl-1 expression in pancreatic cancers. Oncology 1999; 56: 73-82.

35) Friess H, Lu Z, Andrén-Sandberg A, Berberat P, Zimmermann A, Adler G, Schmid R, Büchler MW. Moderate activation of the apoptosis inhibitor bcl-xL worsens the prognosis in pancreatic cancer. Ann Surgery 1998; 228: 780-787.

36) Wagner M, Lührs H, Klöppel G, Adler G, Schmid RM. Malignant transformation of duct-like cells originating from acini in transforming growth factor α transgenic mice. Gastroenterology 1998; 115: 1254-1262.

37) Wagner M, Greten FR, Weber CK, Koschnick S, Mattfeldt T, Deppert W, Kern H, Adler G, Schmid RM. A murine tumor progression model for pancreatic cancer recapitulating the genetic alterations of the human disease. Genes Dev 2001; 15: 286-293.

38) Siveke JT, Einwächter H, Sipos B, Lubeseder-Martellato C, Klöppel G, Schmid RM. Concomitant pancreatic activation of Kras(G12D) and Tgfa results in cystic papillary neoplasms reminiscent of human IPMN. Cancer Cell 2007; 12: 266-279.