Pathol. Oncol. Res. (2014) 20:293–299
DOI 10.1007/s12253-013-9694-1

RESEARCH

Functional Modules Analysis Based on Coexpression Network in Pancreatic Ductal Adenocarcinoma

Baomin Shi · Xiuyan Wang · Xujie Han · Pengfei Liu · Weiwei Wei · Yan Li

Received: 6 May 2013 / Accepted: 30 August 2013 / Published online: 2 November 2013

Abstract Pancreatic ductal adenocarcinoma (PDAC) is the most common epithelial, exocrine pancreatic malignancy, accounting for more than 80 % of the malignant neoplasms of the pancreas. Although the molecular basis of pancreatic cancer is now better understood than ever before, there remains a long distance from being completely understood. In this study, we identified the differentially expressed genes (DEGs) in PDAC tissue compared with normal tissue and constructed a co-expression network by computing the pairwise correlation coefficient between the DEGs. We applied a statistical approach of MCODE to cluster genes in the coexpression network. Ten functional modules were identified in this network. Our results strongly suggest that dysregulations of immune response, homeostasis and cell adhesion may significantly contribute to the development and progression of PDAC. Results from this study will provide the groundwork for the understanding of PDAC. Future studies are needed to confirm some of the possible interactions suggested by this study.

Keywords Co-expression network · Functional modules · Pancreatic ductal adenocarcinoma

B. Shi and X. Wang contributed equally to this work.

B. Shi · X. Han · P. Liu · W. Wei · Y. Li
Department of General Surgery,
Tongji Hospital of Tongji University, Shanghai, China

X. Wang
Department of Ultrasonography,
Tongji Hospital of Tongji University, Shanghai, China

B. Shi (✉)
Department of General Surgery, Tongji Hospital of Tongji University, 389 xincun road, Putuo district, Shanghai 200065, China
e-mail: shibaominsbm@hotmail.com

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most common epithelial, exocrine pancreatic malignancy, accounting for more than 80 % of the malignant neoplasms of the pancreas [1]. It remains the fourth leading cause of cancer-related death in the world, with a less than 6 months median survival and a 5 year survival rate of about 5 % [2, 3]. The majority of cases are diagnosed in the advanced stages, making curative therapy impossible and leading to poor prognosis and incidence equaling mortality [4]. Therefore, it still requires efforts for a more detailed molecular-level understanding of its evolution to enable the development of more effective therapeutics.

Genome-wide transcriptome analysis using expression arrays recently has gained popularity as a means to better understand the molecular characteristics of pancreatic cancer. Jones et al. screened over 21000 genetic alterations in 24 different PDAC and found these genetic alterations mostly affected 12 signaling pathways. These dysregulated pathways mainly involved in specific cellular functions, such as apoptosis, DNA damage repair, G1/S phase cell cycle progression, cell adhesion and invasion [5, 6]. However, despite these very important advances, the precise molecular basis of the disease is incompletely understood.

A number of high throughput microarrays have been deposited into Gene Expression Omnibus in recent years [7–10]. A central problem is to infer functional molecular modules underlying cellular alterations from these high throughput data, such as differential gene and protein concentrations. In this present study, we downloaded gene expression profile of PDAC and identified the differentially expressed genes between PDAC tissue and normal tissue. Further, we constructed a co-expression network and identified functional modules in this network. We anticipate our result may shed new lights on PDAC study.

Materials and Methods

Affymetrix Microarray Data

We downloaded the gene expression profile data on pancreatic cancer patients with normal controls from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database, a public functional genomics data repository. This expression data were deposited by Badea and colleague [11] (ID: GSE15471). Expression analysis of 36 pancreatic ductal adenocarcinoma tumors and matching normal pancreatic tissue samples from pancreatic cancer patients of the Clinical Institute Fundeni (ICF) using Affymetrix U133 Plus 2.0 whole-genome chips. Pairs of normal and tumor tissue samples were obtained at the time of surgery from resected pancreas of 36 pancreatic cancer patients. Three of the 36 normal-tumor sample pairs were carried out replicates in order to gauge the technical measurement errors. Thus there were a total of 78 genechip hybridizations for further analysis.

We preprocessed the CEL source files by RMA (Robust Multichip Averaging) algorithm [12] with defaulted parameters in R bioconductor package [13]. Probe sets were mapped to NCBI entrez genes using DAVID [14]. If there were multiple probe sets that correspond to the same gene, the expression values of those probe sets were averaged. The expression dataset, as a result, led to 20283 genes.

Identification of Differentially Expressed Genes

For GSE15471, we used the SAM 4.0 (Significant Analysis of Microarrays) methods to identify the differentially expressed genes (DEGs) between pancreatic cancer tissue and normal controls [15]. SAM identifies genes with statistically significant changes in expression by assimilating a set of gene-

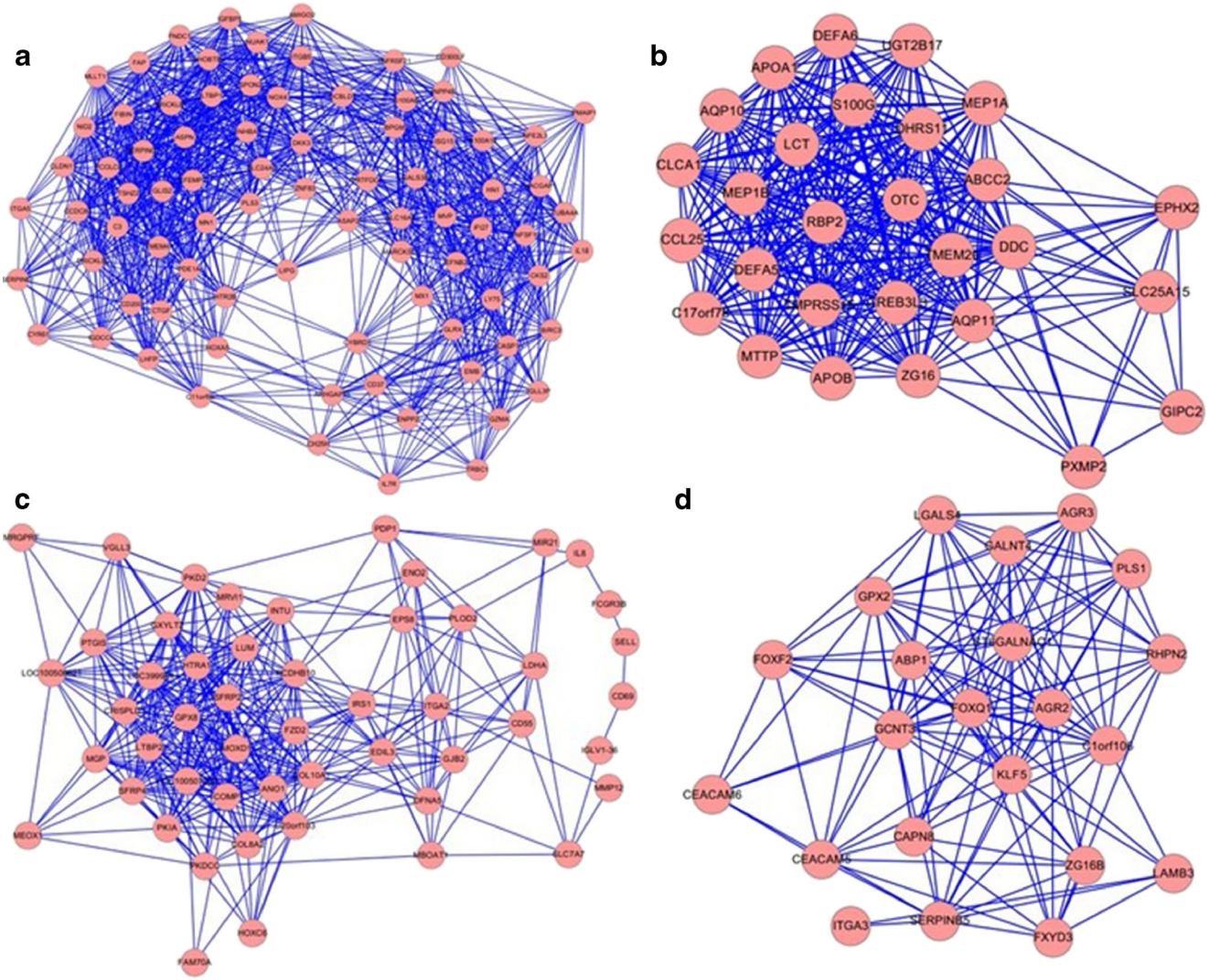

Fig. 1 Functional modules identified from the co-expression network. Subfigures **a**)-**d**) correspond to Module 3 to Module 6

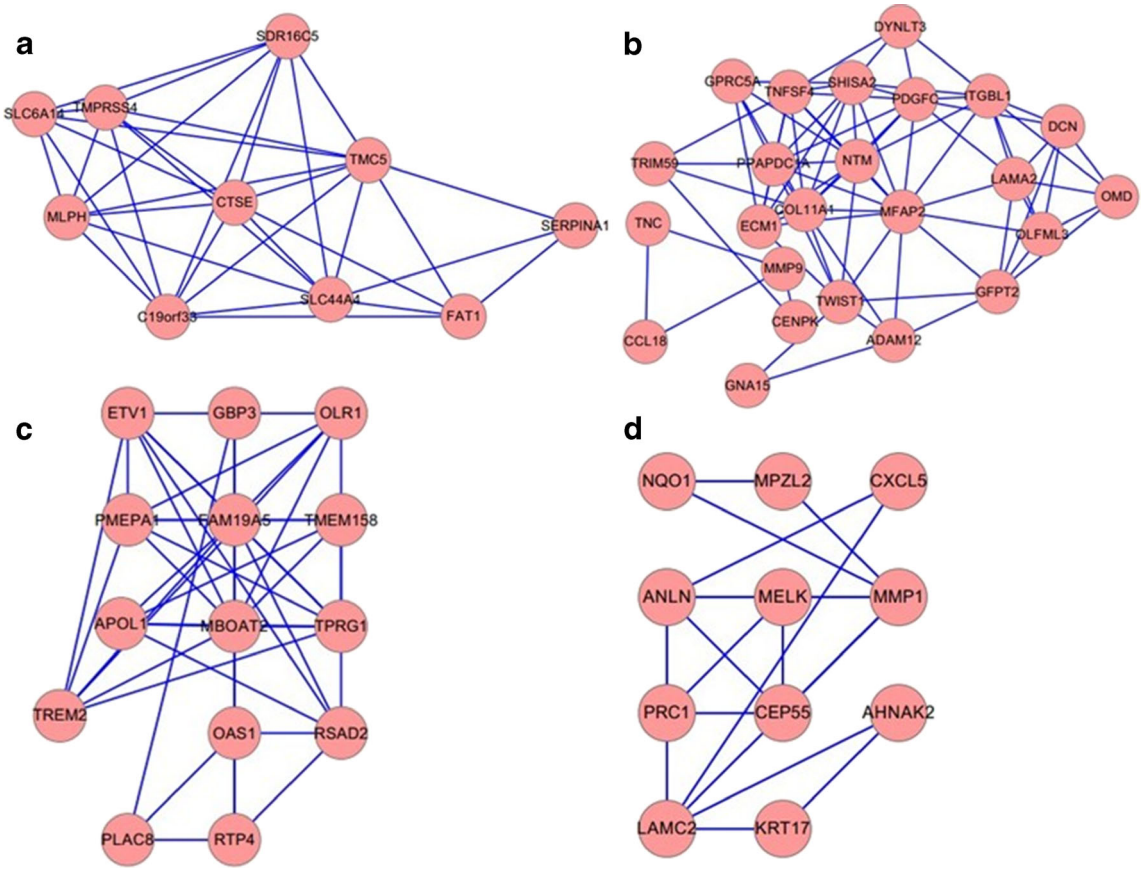

Fig. 2 Functional modules identified from the co-expression network. Subfigures **a**)-**d**) correspond to Module 7 to Module 10

specific *t* tests and uses permutations of the repeated measurements to estimate the percentage of genes identified by chance, the false discovery rate (FDR). The Fold Change value larger than 2 was selected as cutoff criterion for DEGs.

Construction of Differentially Co-Expressed Genes Network

To construct the differentially co-expressed genes network, we calculated the Pearson correlation coefficient (PCC) of all pairwised DEGs. The DEG pairs whose absolute PCC were equal or larger than 0.7 were considered as coexpressed relationships. We constructed an ensemble differentially co-expressed genes network by integrating the co-expressed gene pairs.

Identification of Functional Modules in the Network

MCODE (Molecular Complex Detection) effectively finds densely connected regions of a molecular interaction network, many of which correspond to known molecular complexes, based solely on connectivity data [16]. MCODE detects protein complexes that are with the highest quality, in terms of the function and localization similarity of proteins within predicted complexes. We used MCODE to identify the functional modules with the Degree cutoff=2, K-core=2 and Max.depth=100.

Correlation Test of the Functional Modules with Phenotype

To verify whether the functional modules were significantly correlated with clinical phenotype (normal or disease), we first calculated the average expression values of all genes in each module across subjects. Then, the probability of average expression values of each module being different between normal tissue and cancer tissue was computed using *t*-test.

Table 1 Correlation test of the functional modules with phenotype

Module	Node number	*p*-value
Module_1	401	2.84E-10
Module_2	120	7.41E-05
Module_3	82	4.99E-12
Module_4	28	0.00083
Module_5	51	8.56E-15
Module_6	22	6.63E-06
Module_7	10	3.64E-09
Module_8	24	3.17E-16
Module_9	14	3.86E-16
Module_10	11	1.04E-11

Gene Ontology Analysis

DAVID(The Database for Annotation, Visualization and Integrated Discovery) consists of an integrated biological knowledgebase and analytic tools aimed at systematically extracting biological meaning from large gene/protein lists [14]. To explore the biological function of genes in each module, we searched for over-representation in gene ontology (GO) categories. The GO terms only with p-value less than 0.01 were selected.

Results

Differential Gene Expression Between Pancreatic Cancer Tissue and Normal Controls

The SAM4.0 was used to compare gene expression profiles of pancreatic cancer tissue and normal controls. At a fold change value=2 (FDR=0.0012), a total of 936 genes were differentially expressed, including 766 up-regulated genes and 170 down-regulated genes.

Table 2 Functional annotation of modules (p-value<0.01, module_1 only list the top10 GO terms)

Module	GO ID	GO name	p-value
Module_1	GO:0006955	immune response	6.63E-17
	GO:0009611	response to wounding	2.55E-14
	GO:0006952	defense response	1.59E-11
	GO:0001568	blood vessel development	3.38E-10
	GO:0001944	vasculature development	5.76E-10
	GO:0007155	cell adhesion	1.26E-09
	GO:0022610	biological adhesion	1.29E-09
	GO:0006954	inflammatory response	4.20E-09
	GO:0019882	antigen processing and presentation	1.16E-08
	GO:0030199	collagen fibril organization	2.07E-08
Module_2	GO:0007586	digestion	1.68E-05
	GO:0006508	proteolysis	2.34E-04
	GO:0009070	serine family amino acid biosynthetic process	0.00225536
	GO:0046942	carboxylic acid transport	0.00254912
	GO:0015849	organic acid transport	0.002626762
	GO:0006575	cellular amino acid derivative metabolic process	0.00400803
	GO:0015837	amine transport	0.006883094
	GO:0000096	sulfur amino acid metabolic process	0.009858611
Module_3	GO:0006955	immune response	9.01E-04
	GO:0007155	cell adhesion	0.001709549
	GO:0022610	biological adhesion	0.00172842
Module_4	GO:0048878	chemical homeostasis	9.41E-04
	GO:0055092	sterol homeostasis	0.001655677
	GO:0042632	cholesterol homeostasis	0.001655677
	GO:0055088	lipid homeostasis	0.002965614
	GO:0042158	lipoprotein biosynthetic process	0.003690958
	GO:0009636	response to toxin	0.004081603
	GO:0042592	homeostatic process	0.005155151
	GO:0042157	lipoprotein metabolic process	0.006976634
Module_5	GO:0007155	cell adhesion	0.002810985
	GO:0022610	biological adhesion	0.002834405
Module_8	GO:0007155	cell adhesion	1.25E-04
	GO:0022610	biological adhesion	1.26E-04
	GO:0043062	extracellular structure organization	0.001079713
	GO:0030198	extracellular matrix organization	0.007642476
Module_9	GO:0006955	immune response	3.22E-05

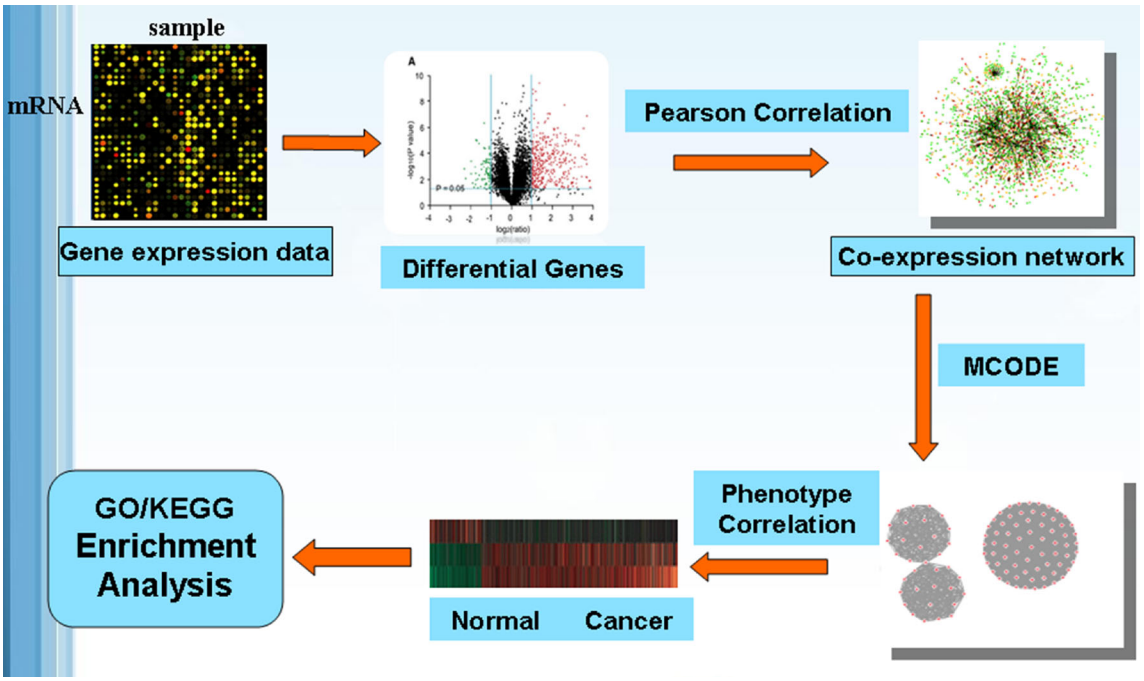

Fig. 3 Flowchart in this study. Geneexpression data were downloaded from GEO and differentially expressed genes were identified. Pearson correlation test were then applied to identify co-expression DEGs and co-expression network were constructed. After that, MCODE were used to identify functional modules from the co-expression network and correlations of these modules with phenotype were verified. Finally, GO enrichment analysis was performed to biologically characterize those modules

Co-Expression Network Construction and Functional Modules Identification

It is thought that genes with similar patterns of mRNA expression and genes with similar functions are likely to be regulated via the same mechanisms [17]. To measure similarity in expression, we calculated the pairwise correlation coefficient between the DEGs. After excluding the gene pairs with PCC<0.7, we constructed a co-expression network which including 933 nodes and 117651 edges using Cytoscape [18].

Because coexpressed genes are biologically related, grouping these highly connected genes by network analysis may shed light on underlying functional processes in a manner complementary to standard differential expression analyses. In order to identify the functional molecular complexes from this coexpression network, we employed a statistical approach of MCODE to cluster genes. We set the parameters as degree cutoff=2, k-core=2 and max.depth=100 to detect functional modules and finally got 10 modules with gene number larger than ten. Figures 1 and 2 shows the functional module 3 to module 10.

Correlation Test of the Functional Modules with Phenotype

To examine if these modules were associated with phenotype, we correlated the average expression values of all genes in each module in cancer tissue with normal tissue. The result showed that all modules were significantly correlated with phenotype with p-value<0.05 (Table 1).

Functional Annotation of the Modules

To biologically characterize these modules, we used the online biological classification tool DAVID to classify these genes in each module and observed various level of GO category enrichment in all 10 modules (Table 2). Total 89 GO terms were enriched in module 1. The most significant enrichment was the GO category of immune response with p-value=6.63E-17. The other significant GO categories included response to wounding (p-value=2.55E-14) and defense response (p-value=1.59E-11). Total 8 GO terms were enriched in module 2. The most significant enrichment was the GO category of digestion with FDR= 1.68E-05. Total 3 GO terms were enriched in module 3, with the most significant enriched GO category of immune response (FDR=9.01E-04). The GO functions of module 4 were most related to homeostasis, such as chemical homeostasis, sterol homeostasis and cholesterol homeostasis. Module 5 and 8 were significantly related to cell adhesion. Only one GO category was enriched in module 10, which was immune response. The rest modules, module 6, module 7 and module 10, did not enriched in any category.

Discussion

Pancreatic cancer continues to pose an enormous challenge to clinicians and cancer scientists [3]. Although the molecular basis of pancreatic cancer is now better understood than ever

before, there remains a long distance from being completely understood. In this study, we constructed a co-expression network in PDAC by computing the pairwise correlation coefficient between the DEGs, and analyzed the properties of this network by identifying functional modules: sets of genes that together involved in a biological process (Fig. 3).

Understanding the structure and function of coexpression network is essential for study the pathogenesis of diseases. In this work, we identified ten significant functional modules using a statistical approach of MCODE and explore the function of each module using the online biological classification tool DAVID.

Our result showed that 3 of 10 modules were related to immune response, they were module 1, module 3 and module 9. This result confirmed the significance of immune system in the development and progression of PDAC. Patients with cancer can develop tumor-specific immune responses, although established cancer usually progresses despite the antitumor immune response [19, 20]. Therefore, it is important to develop strategies that harness the molecules and cells of the immune system to treat this disease.

The module 2 and module 4 were significantly related to digestion and homeostasis, respectively. The pancreas is comprised of separate functional units that regulate two major physiological processes: protein and carbohydrate digestion and glucose homeostasis [21]. Our result suggested these two major physiological processes were dysfunctional in PDAC patients. However, we are unaware of any microarray analysis that reported these two dysregulated functions in PDAC. Further studies are needed to confirm our result.

The DEGs in module 5 and module 8 were enriched in cell adhesion and biological adhesion. Cell to cell adhesion and interaction play an important role in carcinogenesis and largely determine metastatic potential [22, 23]. Cell adhesion is mediated by the interaction of extracellular matrix (ECM) components with cell surface molecules [24]. Similar with our result, a phosphoproteomic analysis by Zhou et al. revealed that differential phosphorylation of many proteins involved in cell adhesion, cell junction [25]. An understanding of interactions among genes in module 5 and module 8 will certainly help in comprehending the complex dynamics of tumor invasion and metastasis in PDAC ecology.

Overall, this study used a systems biology approach to identify functional modules that were closely related to phenotype of PDAC. Our results strongly suggest that dysregulations of immune response, homeostasis and cell adhesion may significantly contribute to the development and progression of PDAC. Results from this study will provide the groundwork for the understanding of PDAC. Future studies are needed to confirm some of the possible interactions suggested by this study.

References

1. Alexakis N, Halloran C, Raraty M, Ghaneh P, Sutton R, Neoptolemos JP (2004) Current standards of surgery for pancreatic cancer. Br J Surg 91(11):1410–1427. doi:10.1002/bjs.4794
2. Siegel R, Ward E, Brawley O, Jemal A (2011) Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. CA Cancer J Clin 61(4):212–236
3. Ghaneh P, Costello E, Neoptolemos JP (2007) Biology and management of pancreatic cancer. Gut 56(8):1134–1152
4. Li D, Xie K, Wolff R, Abbruzzese JL (2004) Pancreatic cancer. Lancet 363(9414):1049–1057. doi:10.1016/S0140-6736(04)15841-8
5. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321(5897):1801–1806
6. Carter H, Samayoa J, Hruban RH, Karchin R (2010) Prioritization of driver mutations in pancreatic cancer using cancer-specific high-throughput annotation of somatic mutations (CHASM). Cancer Biol Ther 10(6):582–587
7. Zhang G, Schetter A, He P, Funamizu N, Gaedcke J, Ghadimi BM, Ried T, Hassan R, Yfantis HG, Lee DH, Lacy C, Maitra A, Hanna N, Alexander HR, Hussain SP (2012) DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. PLoS One 7(2):e31507. doi:10.1371/journal.pone.0031507
8. Perez-Mancera PA, Rust AG, van der Weyden L, Kristiansen G, Li A, Sarver AL, Silverstein KA, Grutzmann R, Aust D, Rummele P, Knosel T, Herd C, Stemple DL, Kettleborough R, Brosnan JA, Morgan R, Knight S, Yu J, Stegeman S, Collier LS, ten Hoeve JJ, de Ridder J, Klein AP, Goggins M, Hruban RH, Chang DK, Biankin AV, Grimmond SM, Wessels LF, Wood SA, Iacobuzio-Donahue CA, Pilarsky C, Largaespada DA, Adams DJ, Tuveson DA (2012) The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. Nature 486(7402):266–270. doi:10.1038/nature11114
9. Vincent A, Omura N, Hong SM, Jaffe A, Eshleman J, Goggins M (2011) Genome-wide analysis of promoter methylation associated with gene expression profile in pancreatic adenocarcinoma. Clin Cancer Res 17(13):4341–4354. doi:10.1158/1078-0432.CCR-10-3431
10. Badea L, Herlea V, Dima SO, Dumitrascu T, Popescu I (2008) Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. Hepatogastroenterology 55(88):2016–2027
11. Badea L, Herlea V, Dima SO, Dumitrascu T, Popescu I (2008) Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. Hepatogastroenterology 55(88):2015–2026
12. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4(2):249–264. doi:10.1093/biostatistics/4.2.249
13. Team RDC (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing
14. da Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4(1):44–57
15. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98(9):5116–5121. doi:10.1073/pnas.091062498

16. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4:2
17. Allocco DJ, Kohane IS, Butte AJ (2004) Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 5:18. doi:10.1186/1471-2105-5-18
18. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504. doi:10.1101/gr.1239303
19. Hiraoka N, Onozato K, Kosuge T, Hirohashi S (2006) Prevalence of FOXP3+ regulatory T cells increases during the progression of pancreatic ductal adenocarcinoma and its premalignant lesions. Clin Cancer Res 12(18):5423–5434
20. Gross S, Walden P (2008) Immunosuppressive mechanisms in human tumors: why we still cannot cure cancer. Immunol Lett 116(1):7–14
21. Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, Depinho RA (2006) Genetics and biology of pancreatic ductal adenocarcinoma. Genes Dev 20(10):1218–1249
22. Gloushankova NA (2008) Changes in regulation of cell-cell adhesion during tumor transformation. Biochemistry (Mosc) 73(7):742–750
23. Tlsty TD (1998) Cell-adhesion-dependent influences on genomic instability and carcinogenesis. Curr Opin Cell Biol 10(5):647–653
24. Deer EL, Gonzalez-Hernandez J, Coursen JD, Shea JE, Ngatia J, Scaife CL, Firpo MA, Mulvihill SJ (2010) Phenotype and genotype of pancreatic cancer cell lines. Pancreas 39(4):425–435. doi:10.1097/MPA.0b013e3181c15963
25. Zhou W, Capello M, Fredolini C, Racanicchi L, Piemonti L, Liotta LA, Novelli F, Petricoin EF (2011) Phosphoproteomic analysis of pancreatic ductal adenocarcinoma cells reveals differential phosphorylation of cell adhesion, cell junction and structural proteins. J Proteomics Bioinforma 4:170–178