



Laboratory-Clinic Interface

Genetic network and gene set enrichment analysis to identify biomarkers related to cigarette smoking and lung cancer

Xiaocong Fang^a, Michael Netzer^b, Christian Baumgartner^b, Chunxue Bai^a, Xiangdong Wang^{a,c,*}^a Department of Pulmonary Medicine, Zhongshan Hospital, Fudan University, No.180, Fenglin Road, Shanghai 200032, China^b Research Group for Clinical Bioinformatics, Institute of Electrical and Biomedical Engineering, UMIT, 6060 Hall in Tirol, Austria^c Biomedical Research Center, Zhongshan Hospital, Fudan University, Shanghai, China

ARTICLE INFO

Article history:

Received 24 April 2012

Received in revised form 3 June 2012

Accepted 6 June 2012

Keywords:

Lung cancer

Genes

Networks

Biomarkers

Smoking

Treatment

ABSTRACT

Objectives: Cigarette smoking is the most demonstrated risk factor for the development of lung cancer, while the related genetic mechanisms are still unclear.

Methods: The preprocessed microarray expression dataset was downloaded from Gene Expression Omnibus database. Samples were classified according to the disease state, stage and smoking state. A new computational strategy was applied for the identification and biological interpretation of new candidate genes in lung cancer and smoking by coupling a network-based approach with gene set enrichment analysis.

Measurements: Network analysis was performed by pair-wise comparison according to the disease states (tumor or normal), smoking states (current smokers or nonsmokers or former smokers), or the disease stage (stages I–IV). The most activated metabolic pathways were identified by gene set enrichment analysis.

Results: Panels of top ranked gene candidates in smoking or cancer development were identified, including genes involved in cell proliferation and drug metabolism like cytochrome P450 and WW domain containing transcription regulator 1. Semaphorin 5A and protein phosphatase 1F are the common genes represented as major hubs in both the smoking and cancer related network. Six pathways, e.g. cell cycle, DNA replication, RNA transport, protein processing in endoplasmic reticulum, vascular smooth muscle contraction and endocytosis were commonly involved in smoking and lung cancer when comparing the top ten selected pathways.

Conclusion: New approach of bioinformatics for biomarker identification and validation can probe into deep genetic relationships between cigarette smoking and lung cancer. Our studies indicate that disease-specific network biomarkers, interaction between genes/proteins, or cross-talking of pathways provide more specific values for the development of precision therapies for lung.

© 2012 Elsevier Ltd. All rights reserved.

Abbreviations: NSCLC, non-small cell lung cancer; GSEA, gene set enrichment analysis; IG, information gain; WWTR1, WW domain containing transcription regulator 1; SEMA5A, sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A; Ppm1f, protein phosphatase 1F; Agtr1, angiotensin II receptor, type 1; Ramp2, receptor (G protein-coupled) activity modifying protein 2; GRK5, G protein-coupled receptor kinase 5; arap3, ArfGAP with RhoGAP domain-ankyrin repeat and PH domain 3; CRYAB, crystallin, alpha B; UBE2L, ubiquitin-conjugating enzyme E2L; ednrb, endothelin receptor B; LEPR, leptin receptor; Eif2α, eukaryotic initiation factor 2, α subunit; srpk1, SFRS protein kinase 1; MICAL, microtubule associated monooxygenase, calponin and LIM domain; nth1, nth endonuclease III-like 1 (*E. coli*).

* Corresponding author.

E-mail addresses: fangxiaocong@gmail.com (X. Fang), michael.netzer@umit.at (M. Netzer), christian.baumgartner@umit.at (C. Baumgartner), bai.chunxue@zs-hospital.sh.cn (C. Bai), xiangdong.wang@telia.com (X. Wang).

0305-7372/\$ - see front matter © 2012 Elsevier Ltd. All rights reserved.

<http://dx.doi.org/10.1016/j.ctrv.2012.06.001>

Introduction

It has become a major public health problem as a leading cause of cancer death in men and increasingly in women, for which over 1 million people died every year in the world.¹ Cigarette smoking has been considered as one of the most demonstrated risk factors for the development of lung cancer where genetic variants are also considered to play an essential role. Cigarette smoking could induce dysregulation of genes or mRNAs, associated with lung cancer, carcinogenesis, and a great number of genes or pathways as potential candidates of targeted therapies for lung cancer.^{2,3} However, the major challenge is the great variation between studies and the inconsistency of selected candidates which need further validation.

Bioinformatics is the application of omics science, information technology, mathematics and statistics in the field of biomarker

discovery. There has been a great increase in genomics and other molecular research to produce a tremendous amount of information related to molecular biology.⁴ Various powerful data mining and statistics methods have been propagated to handle the huge amount of data. Biomarker networks or dynamic biomarker networks were proposed as new way of biomarker identification and validation to emphasize the relationship between genes, proteins or metabolites, taking into account the kinetic characteristics.^{5,6}

Numerous candidate biomarkers were reported to be associated with the development or prognosis of lung cancer using bioinformatics approaches.^{7–9} Targeted therapy for lung cancer patients with epidermal growth factor receptor (EGFR) mutations was considered as an example of the promising development toward personalized therapy for lung cancer patients.¹⁰ Public expression data with bioinformatics approaches was applied to identify useful biomarkers for lung cancer, which could be further validated in clinical trials.¹¹ Genetic relationships between cigarette smoking and lung cancer, e.g. the significance of differentially expressed genes resulted from cigarette smoking and roles in the development of lung cancer, should be further clarified. Cigarette smoking and lung cancer may share the same pathophysiological pathways, where the expression profile of genes may change along with the disease progress. The present studies propose a network-based approach to identify genes related to lung cancer and cigarette smoking, including the differentially expressed genes in patients with different smoking histories at different disease stages. The approach with gene set enrichment analysis (GSEA) was coupled to search for the activated pathways, which are involved in the pathophysiological processes between cigarette smoking and lung cancer, to identify a panel of expected and unexpected genes and associated pathways in lung cancer.

Data selection and recruitment

The preprocessed microarray expression dataset from the study conducted by Landi et al.¹² available at Gene Expression Omnibus online database (GDS3257) was used. The platform ID is GPL96. The data set comprises 107 samples, including 58 tumor tissues and 49 normal tissues from never smokers, non-smokers and former smokers. These samples were also classified according to the disease stages and gender (Table 1). There are 22,283 genes with defined Affymetrix_3PRIME_IVT_ID for each sample.

Coupled computational approach

Genetic network analysis

A new network-based biomarker discovery approach, as described previously,^{6,13} was applied for the search for highly discriminatory genes according to their connectivity strength within the network. Inferring of the network for n features results in $\frac{n(n-1)}{2}$ comparisons. The network for $n = 22,283$ genes in reasonable time was calculated as a filtering step (Step 1).

Overall, the modality includes two steps:

Step 1: Feature selection: In order to reduce the number of features (genes) by excluding the genes that show no or minimal discriminatory ability, the information gain (IG),¹⁴ an entropy-based method, was applied for feature extraction. Reasons for choosing IG included (i) the ability to easily identify features (genes) with no discriminatory ability (i.e., $IG = 0$), (ii) the applicability to multiple class problems, and (iii) a low calculation time. The IG score for each gene in the dataset was calculated with respect to the classes (i.e., disease state, disease stage and smoking state). The IG score for a feature f is defined as follows¹⁵:

$IG = H(Y) - H(Y|X)$, where

$$H(Y) = -\sum_{y \in Y} p(y) \log_2(p(y)), \text{ and}$$

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2(p(y|x)),$$

where $H(Y)$ denotes the entropy for Y (class variable) and $H(Y|X)$ is the entropy of Y after observing X . A threshold of $IG = 0$ was set to remove all features (genes) with no discriminatory ability.

In the present study the network analysis was performed by pair-wise comparison according to the disease states (tumor or normal), smoking states (current smokers or nonsmokers or former smokers), or the disease stages (stages I–IV).

Step 2: Inferring the network: On the reduced gene sets the networks were inferred including three steps:

- (i) calculating all ratios R between genes G which represent gene g interactions, where $r_{ij} = \left| \log_2 \left(\frac{g_i}{g_j} \right) \right|$ with $i > j$, $r \in R$,
- (ii) computing IG scores s_{ij} ; $s \in S$ on the logarithmic ratios R , and
- (iii) constructing a graph G with: $G_{ij} = \begin{cases} 1 & \text{if } |s_{ij}| > \tau \\ 0 & \text{else} \end{cases}$ for $i, j \in g_1, g_2, \dots, g_n$.

After constructing the network, the genes are ranked according to their degrees (i.e., number of edges of each gene) in the network.

GSEA

Briefly, GSEA searched for groups of genes that shared common biological function, chromosomal location, or regulation.¹⁶ This analysis contained two steps:

Step 1: Feature selection by a non-specific filtering to remove those genes with high variants according to the disease state or the smoking history, respectively.

Step 2: (i) Selection of the activated metabolic pathways with more than 10 genes involved in the pathophysiological processes of cigarette smoking or lung cancer, respectively, according to Kyoto Encyclopedia of Genes and Genomes (KEGG) database; (ii) calculation of the T value for every gene and the mean T value for all genes involved in every specific pathway; (iii) ranking the pathways according to the mean value of T . Top ranked metabolic pathways were compared with regard to pathophysiological processes of cigarette smoking and lung cancer. Overlaid pathways were indicated.

Network analysis

Networks of different groups, i.e., disease state (tumor/normal), smoking history (current smoker/former smoker/nonsmoker), and disease stage (I/II/III/IV), respectively, were compared for lung cancer and cigarette smoking to identify and validate common genes by reviewing the online databases, including PubMed, Database for Annotation, Visualization and Integrated Discovery, and Kyoto Encyclopedia of Genes and Genomes database. R¹⁷ was used to perform network analysis and GSEA. The overall work flow for the present study was described in Fig. 1.

Genetic network

For the identification of relevant gene markers, the IG score for each gene between the tumor tissues and normal tissues (T/NT), as well as in the current smokers, former smokers and never smokers (C/F/N) or current smokers and never smokers (C/N) was calculated. Out of 22,283 genes in the dataset, over 9365 genes were different between tumor and normal patient having a IG score >0 ,

Table 1
The composition of the dataset (GDS3257).

Tissue	Gender	Individual	Disease state
Tumor (n = 58)	Male (n = 35) Female (n = 23)	Never smoker (n = 16)	Stage I (n = 6)
			Stage II (n = 4)
			Stage III (n = 5)
			Stage IV (n = 1)
	Former smoker (n = 18)		Stage I (n = 6)
			Stage II (n = 7)
Normal (n = 49)	Male (n = 34) Female (n = 15)	Current smoker (n = 24)	Stage III (n = 5)
			Stage I (n = 10)
			Stage II (n = 10)
			Stage III (n = 2)
	Former smoker (n = 18)		Stage IV (n = 2)
			Stage I (n = 7)
		Never smoker (n = 15)	Stage II (n = 2)
			Stage III (n = 5)
			Stage IV (n = 1)
			Stage I (n = 6)
	Former smoker (n = 18)		Stage II (n = 9)
			Stage III (n = 3)
		Current smoker (n = 16)	Stage I (n = 9)
			Stage II (n = 3)
			Stage III (n = 1)
			Stage IV (n = 2)
	Former smoker (n = 18)		Stage I (n = 6)
			Stage II (n = 7)

while only 548 and 20 differentially expressed genes were detected for smoking history and disease stages, which suggested that most of genes be influenced by the tumor.

The consistency rate between top 100 and 500 genes with high-information scores according to different classifications was calculated. The consistency rate above top 500 genes was 6.8% between T/NT and C/F/N, 6.8% between T/NT and C/N and 59.8% between C/F/N and C/N. Consistency rates were 0%, 2% and 52% among top 100 genes, respectively. The statistical significance of each gene between the T/NT groups, C/F/N groups and C/N groups and the accuracy of information gain scores were calculated using ANOVA. The interdependence of the IG scores was 81.5%. Genes with IG scores ≥ 0.75 for T/NT comparison were selected for further analysis, and 156 genes were included for the tissue dependent

network analysis. 148 features were selected for the C/N dependent network with $r = 0.22$. The former smokers group was excluded to identify smoking related genes in current smokers and never smokers. The thresholds for the IG scores were set between 0.8 and 0.32 in the further analysis to select only the important couples of genes. The resulting networks were shown in Figs. 2–4, as described previously.¹⁸ Detailed information for the top 10 genes in patients with lung cancer or smokers were listed in Tables 2 and 3, respectively. Common genes in both cancer and smoking, inherent in top 500 genes according to the information gain scores, were listed in Table 4, where two genes, semaphorin 5A (SEMA5A) and protein phosphatase 1F (Ppm1f), were involved in the tissue or individual dependent genetic network.

GSEA

A total of 6365 genes were included in the analysis after feature selection, of which 2474 genes and 183 pathways were finally included in GSEA. Of those activated pathways about 10 genes were identified to be involved in both the cigarette smoking and lung cancer-related processes. The top 10 ranked pathways associated with lung cancer and cigarette smoking were listed in Tables 5 and 6, respectively, where 6 pathways were common in both groups, including cell cycle, DNA replication, RNA transport, protein processing in endoplasmic reticulum, or vascular smooth muscle contraction and endocytosis (see also Table 7). The genetic networks of those six pathways were depicted in the supplement files (Fig. S1–S6).

Association between network and GSEA

Genes involved in both the genetic network and the ten top ranked metabolic pathways in lung cancer and smoking were compared, after the identification of gene markers separately by genetic network study and GSEA (Table 8). There were six common pathways, including 04110, 04270, 04144, 03030, 04141, and 03013. In lung cancer, three genes (angiotensin II receptor, type 1 (Agt1), natriuretic peptide receptor A/guanylate cyclase A, and

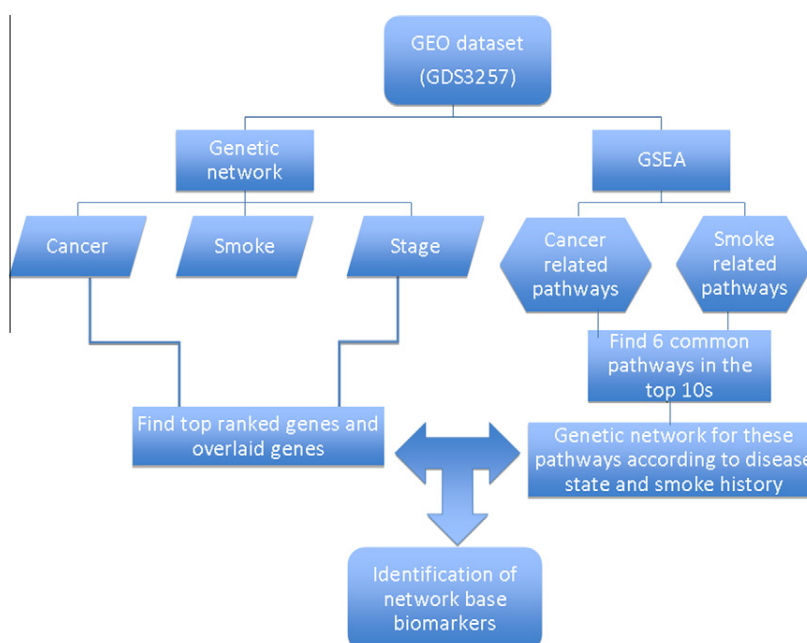


Fig. 1. The flowchart of the network-based biomarker discovery strategy in lung cancer and cigarette smoking.

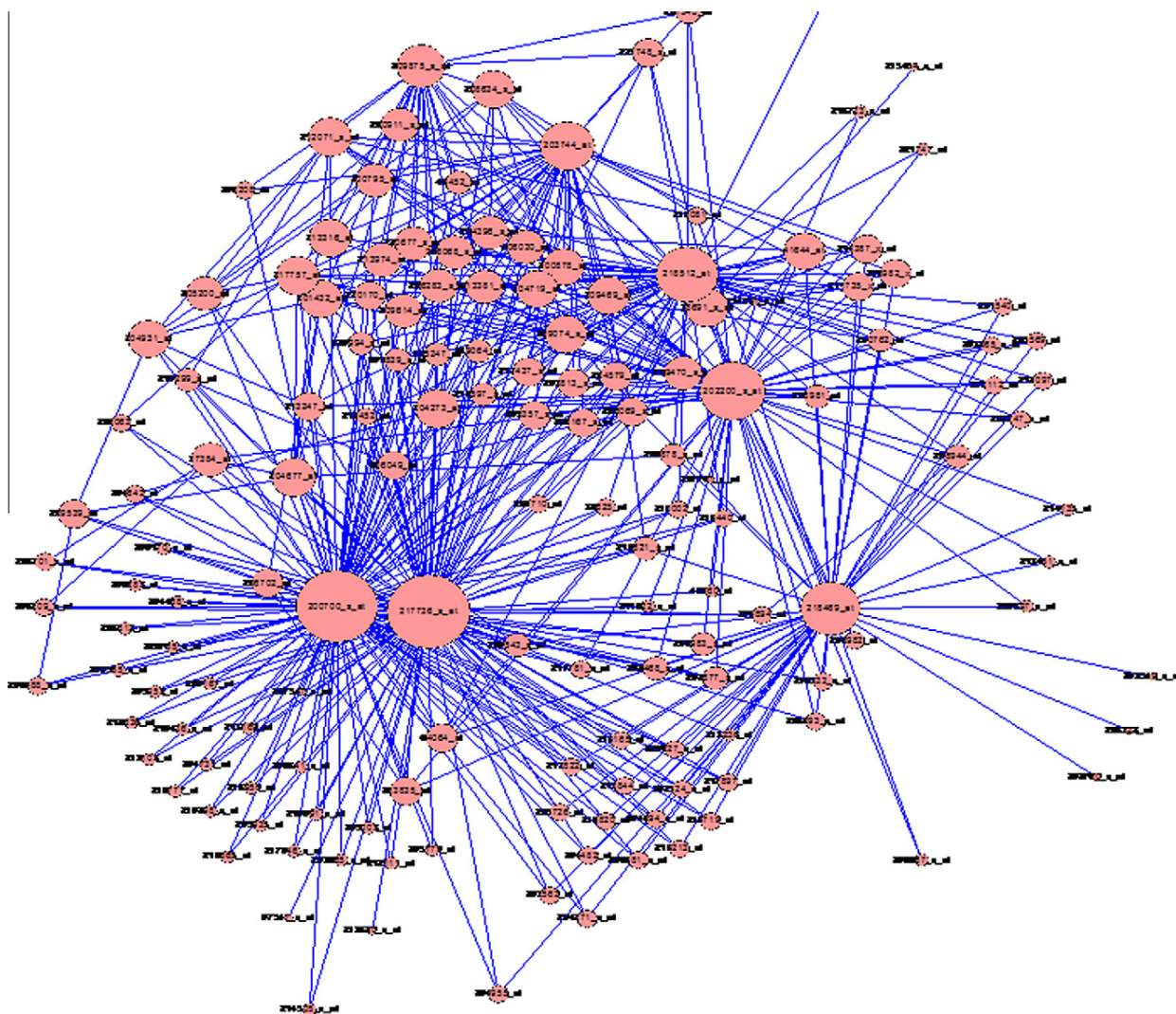


Fig. 2. Cancer related genetic network by comparing patients with cancer versus health people. The node size was proportional to the degree (i.e., number of edges) of specific genes which represent the magnitude of the ability associated with other genes. This figure was created using Cytoscape.¹⁸

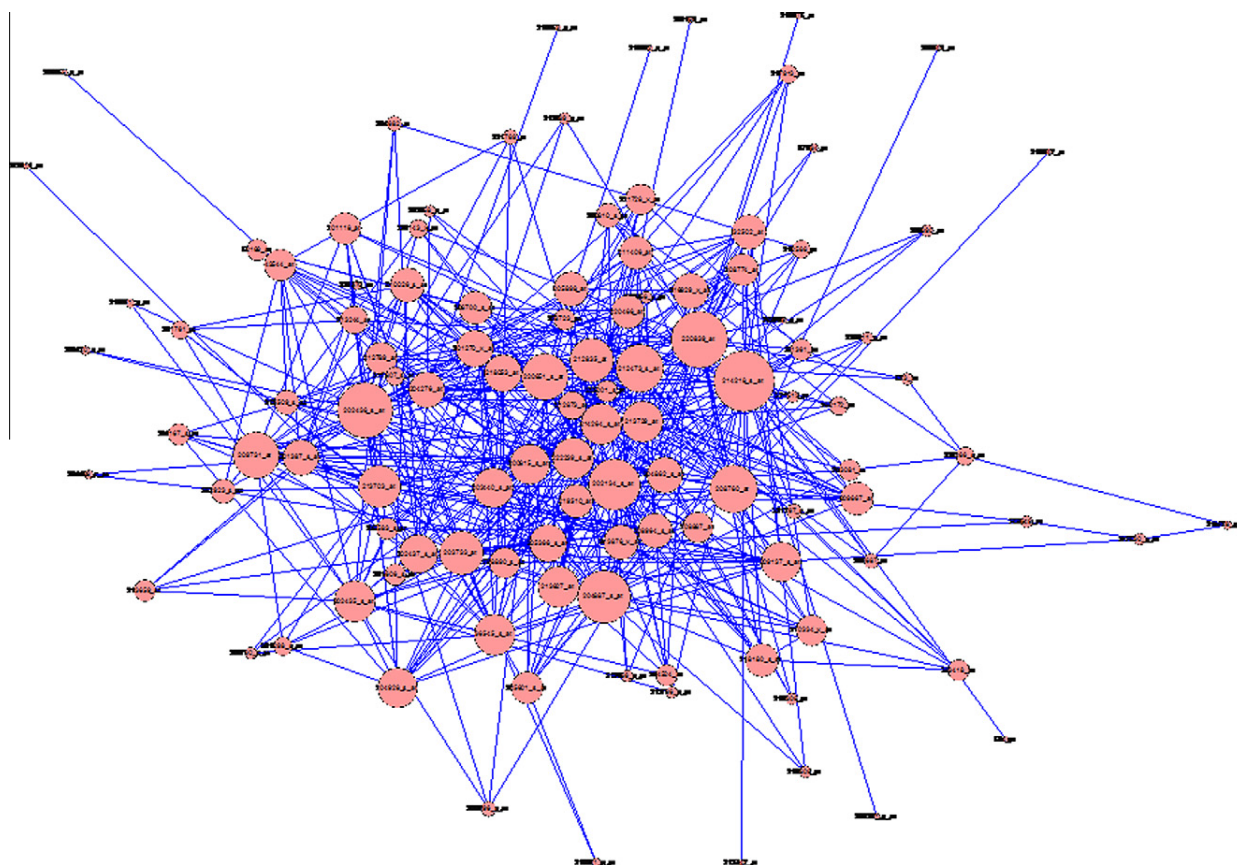
receptor (G protein-coupled) activity modifying protein 2 (Ramp2)) were involved in the “vascular smooth muscle contraction” pathway of the genetic network, six genes (caveolin 1, caveolin 2, fibroblast growth factor receptor 4 (Fgfr4), G protein-coupled receptor kinase 5 (GRK5), transforming growth factor, beta receptor II (Tgfb2), ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3 (arap3)) in “endocytosis” pathway, and two genes (crystallin, alpha B (CRYAB), eukaryotic translation initiation factor 2-alpha kinase 1 (Eif2ak1)) in the “protein processing in endoplasmic reticulum” pathway. In cigarette smoking, only one gene (Cyclin A2) was found to be involved in the “cell cycle” pathway, and two genes (ubiquitin-conjugating enzyme E2I (UBE2I) and eukaryotic translation initiation factor 1A, Y-linked (EIF1AY)) in the “RNA transport” pathway. None of those genes was found to coexist in the pathways and the two genetic networks. In other pathways listed in top ten activated pathways for lung cancer or cigarette smoking, 4 genes (Agtr1, sphingosine-1-phosphate receptor 1, endothelin receptor B (ednrb), leptin receptor (LEPR)) in the “neuroactive ligand-receptor interaction” pathway and three genes (acetyl-Coenzyme A carboxylase beta (Acacb), CD36 and LEPR) in “adipocytokine signaling” pathway were found to be involved in the cancer related genetic network, while 1 gene

(213675_at) in “focal adhesion” pathway was found to be involved in cigarette smoking related genetic network. No further related information was found about this particular gene.

Significance and importance of gene networks

Numerous genetic studies of human airway epithelial cells demonstrated gene expression signatures related to smoking or lung cancer.^{19,20} Differences in NRF2-mediated transcription profiles from bronchial airway epithelial cells were identified between nonsmokers, cigarette smokers with suspicion of lung cancer, and those with a subsequent diagnosis of lung cancer.¹⁹ In the present study, a new computational strategy comprising of a network based biomarker detection method combined with gene set enrichment analysis was used to identify related pathways. Several genes and pathways were identified to probably contribute to the relationship of cigarette smoking and lung cancer.

A number of common genes in Table 4 were noticed to be involved in top 500 genes with the highest IG scores on basis of the calculation of either according T/NT or C/N. SEMA5A and Ppm1f were involved in those two networks where SEMA5A may play a role in the development of neurological dysfunction.²¹



Down-regulation of transcriptional and translational genes in tumor tissue were associated with poor survival in smoking women

Lung cancer-specific networks

Secreted phosphoprotein 1 was demonstrated as a cancer biomarker for the progression and prognosis in multiple types of tumors³⁴ through regulating EGFR/ERK signal pathway.³⁵ The expression of secreted phosphoprotein 1 could be upregulated by nicotine and involved in the development of airway remodeling in chronic lung diseases.³⁶ Alpha-2-macroglobulin is a protease

Table 2

The top 10 ranked genes in the cancer related network.

Probe ID	Degree	Gene name	Gene symbol	Related pathophysiological processes
217736_s_at	72	Eukaryotic translation initiation factor 2- alpha kinase 1	eIF2αk1	Acts at the level of translation initiation to downregulate protein synthesis in response to stress
202200_s_at	43	SFRS protein kinase 1	srpk1	Plays a central role in the regulatory network for splicing
209875_s_at	16	Secreted phosphoprotein 1	SPP1	Plays an essential role in type I immunity. Also important to cancer-stroma interaction
217757_at	9	Alpha-2-macroglobulin	A2M	A protease inhibitor and cytokine transporter which can inhibit all four classes of proteinases
201432_at	8	Catalase	cat	Encodes catalase, a key antioxidant enzyme in the bodies defense against oxidative stress
204273_at	8	Endothelin receptor type B	ednrb	A G protein-coupled receptor which activates a phosphatidylinositol-calcium second messenger system
38691_s_at	7	Surfactant protein C	SFTPC	Secreted by the alveolar cells of the lung and maintains the stability of pulmonary alveolar
37384_at	6	Protein phosphatase 1F (PP2C domain containing)	Ppm1f	A member of the PP2C family which are negative regulators of cell stress response pathways
204677_at	5	Cadherin 5, type 2 (vascular endothelium)	cdh5	Calcium dependent cell adhesion proteins, play an important role in endothelial cell biology
41644_at	5	SAM and SH3 domain containing 1	SASH1	May have a role in a signaling pathway as a tumor suppressor

Table 3

The top 10 ranked genes in the smoking related network.

Probe ID	Degrees	Gene name	Gene symbol	Physiological functions
202436_s_at	17	Cytochrome P450, family 1, subfamily B, polypeptide 1	CYP1B1	Monooxygenases which involved in drug metabolism and synthesis of cholesterol, steroids and other lipids
202134_s_at	15	WW domain containing transcription regulator 1	WWTR1	Functions as a transcriptional coactivator
212473_s_at	12	Microtubule associated monooxygenase, calponin and LIM domain containing 2	MICAL2	A member of MICAL family act as a CasL-interacting molecule and intermediate filaments
220839_at	12	Methyltransferase like 5	mettl5	May belong to the methyltransferase superfamily
204887_s_at	11	Polo-like kinase 4 (Drosophila)	PLK4	Belong to the polo family of serine/threonine protein kinase and play a role in cell proliferation
208760_at	10	Ubiquitin-conjugating enzyme E2I (UBC9 homolog, yeast)	UBE2I	Belongs to the ubiquitin-conjugating enzyme family, associated with degradation of abnormal or short-lived proteins
209731_at	9	nth endonuclease III-like 1 (<i>E. coli</i>)	nthl1	A DNA N-glycosylase of the endonuclease III family with DNA glycosylase activity
204928_s_at	8	Solute carrier family 10 (sodium/bile acid cotransporter family), member 3	Slc10a3	A housekeeping gene which maps to a GC-rich region of the X chromosome
220651_s_at	8	Minichromosome maintenance complex component 10	Mcm10	Involved in DNA replication and participate in the activation of the pre-replication complex (pre-RC)
36545_s_at	8	Sfi1 homolog, spindle assembly associated (yeast)	SFI1	Plays a role in the dynamic structure of centrosome-associated contractile fibers via its interaction with CETN2

inhibitor and cytokine transporter and inhibits the four types of proteases,³⁷ associated with lung cancer,³⁸ ovarian cancer,³⁹ or cigarette pack-years.⁴⁰ Catalase is a key antioxidant enzyme and plays an essential role in various conditions such as cigarette smoking, diabetes, asthma, Alzheimer's disease, or cancer.⁴¹

The protein encoded by ednrb, a G protein-coupled receptor, activates a phosphatidylinositol-calcium second messenger system.⁴² Expressing of ednrb has been altered in several human tumors^{42,43} while the inactivation of the ednrb gene through epigenetic alteration was highly prevalent in lung cancer.^{44,45} The polymorphism of ednrb was also demonstrated to be associated with airway obstruction resulted from cigarette smoking.⁴⁶ Surfactant protein C (SFTPC) is produced by alveolar epithelial cells as an extremely hydrophobic surfactant protein to maintain lung function and homeostasis after birth. Expression of SPC was suggested as a marker for lung tumors originated from type II pneumocytes⁴⁷ and affected by cigarette smoking.⁴⁸ Cadherin 5 is a calcium-dependent cell-cell adhesion glycoprotein, functioning as a classic cadherin to assist cell adherence in a homophilic manner. Cadherin 5 plays an important role in endothelial cell biology through control of the cohesion and organization of the intercellular junctions,⁴⁹ involved in angiogenesis, tumor growth, or metastasis.⁵⁰ Tobacco smoking

could modulate the expression of VE-cadherin in endothelium.⁵¹ SAM and SH3 domain containing 1 (SASH1) was considered as a candidate tumor suppressor gene in breast cancer, lung tumor, or colon carcinoma.⁵²

Priority of selected genes from the networks

The top ten genes with highest scores involved in the smoking related genetic network and physiological functions are listed in Table 3. Cytochrome P450 protein, a monooxygenase, was reported to be involved in metabolism of tobacco carcinogens and associated with lung cancer.^{53,54} WW domain containing transcription regulator 1 (WWTR1) as a transcriptional coactivator was proposed to play important roles in oncogenesis⁵⁵ and be a biomarker for pancreatic cancer.⁵⁶ Microtubule associated monooxygenase, calponin and LIM domain (MICAL) family was initially identified as a CasL-interacting molecule, associated with vimentin, a cytoskeletal regulator that connects CasL to intermediate filaments.⁵⁷ MICAL2 was overexpressed in progression of prostate cancer.⁵⁸ Methyltransferase-like 5 belongs to the methyltransferase superfamily, and its function and roles in cigarette smoking and lung cancer remain unclear.

Table 4

The common involved genes in smoking and lung cancer related networks.

Probe ID	Gene symbol	Gene name	Physiological functions
213169_at	SEMA5A	Sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	Belongs to the semaphorin gene family that encodes membrane proteins, involved in axonal guidance during neural development
37384_at	Ppm1f	Protein phosphatase 1F (PP2C domain containing)	A member of the PP2C family of Ser/Thr protein phosphatases which are known to be negative regulators of cell stress response pathways
202177_at	gas6	Similar to growth arrest-specific 6; growth arrest-specific 6	Ligand for tyrosine-protein kinase receptors AXL, TYRO3 and MER whose signaling is implicated in cell growth and survival, cell adhesion and cell migration
202200_s_at	srpk1	SFRS protein kinase 1	Plays a central role in the regulatory network for splicing, controlling the intranuclear distribution of splicing factors in interphase cells and the reorganization of nuclear speckles during mitosis
202524_s_at	SPOCK2	Sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	Encodes a protein which binds with glycosaminoglycans to form part of the extracellular matrix. May also participate in diverse steps of neurogenesis
202908_at	wfs1	Wolfram syndrome 1 (wolframin)	Encodes a transmembrane protein, which is located primarily in the endoplasmic reticulum and ubiquitously expressed with highest levels in brain, pancreas, heart, and insulinoma beta-cell lines
203571_s_at	C10orf116	Chromosome 10 open reading frame 116	Exclusively expressed in adipose tissue. Its function is currently unknown
204271_s_at	ednrb	Endothelin receptor type B	A G protein-coupled receptor which activates a phosphatidylinositol-calcium second messenger system. Its ligand, endothelin, consists of a family of three potent vasoactive peptides: ET1, ET2, and ET3
204396_s_at	GRK5	G protein-coupled receptor kinase 5	encodes a member of the guanine nucleotide-binding protein (G protein)-coupled receptor kinase subfamily of the Ser/Thr protein kinase family, also play a role in regulating the motility of polymorphonuclear leukocytes
204731_at	Tgfr3	Transforming growth factor, beta receptor III	Binds to TGF-beta. Could be involved in capturing and retaining TGF-beta for presentation to the signaling receptors
205206_at	KAL1	Kallmann syndrome 1 sequence	May be an adhesion-like molecule with anti-protease activity. Also play a role in in neural cell adhesion and axonal migration
205779_at	Ramp2	Receptor (G protein-coupled) activity modifying protein 2	A member of the RAMP family of single-transmembrane-domain proteins which are required to transport calcitonin-receptor-like receptor (CRLR) to the plasma membrane
206170_at	ADRB2	Adrenergic, beta-2-, receptor, surface	A member of the G protein-coupled receptor superfamily
208056_s_at	Cbfa2t3	Core-binding factor, runt domain, alpha subunit 2; translocated to, 3	Functions as a transcriptional repressor
208789_at	PTRF	Polymerase I and transcript release factor	Regulates rRNA transcription by promoting the dissociation of transcription complexes and the reinitiation of polymerase I on nascent rRNA transcripts
209434_s_at	PPAT	Phosphoribosyl pyrophosphate amidotransferase	A member of the purine/pyrimidine phosphoribosyltransferase family, a regulatory allosteric enzyme that catalyzes the first step of de novo purine nucleotide biosynthetic pathway
209904_at	tnnc1	Troponin C type 1 (slow)	The central regulatory protein of striated muscle contraction
211066_x_at	PCDHGA12,3,4,5	Protocadherin gamma subfamily C, 3; protocadherin gamma subfamily C, 5; protocadherin gamma subfamily C, 4; protocadherin gamma subfamily A, 12	A member of the protocadherin gamma gene cluster, which have an immunoglobulin-like organization, suggesting that a novel mechanism may be involved in their regulation and expression
211986_at	ahnak	AHNAK nucleoprotein	May be required for neuronal cell differentiation
212071_s_at	SPTBN1	Spectrin, beta, non-erythrocytic 1	Involved in secretion, interacts with calmodulin in a calcium-dependent manner and is thus candidate for the calcium-dependent movement of the cytoskeleton at the membrane.
212552_at	hpcal1	Hippocalcin-like 1	May be involved in the calcium-dependent regulation of rhodopsin phosphorylation.
212796_s_at	TBC1D2B	TBC1 domain family, member 2B	May act as a GTPase-activating protein.
213620_s_at	Icam2	Intercellular adhesion molecule 2	A member of the intercellular adhesion molecule (ICAM) family
213656_s_at	KLC1	Kinesin light chain 1	A microtubule-associated force-producing protein that may play a role in organelle transport
217046_s_at	AGER	Advanced glycosylation end product-specific receptor	A member of the immunoglobulin superfamily of cell surface receptors.
217781_s_at	ZFP106	Zinc finger protein 106 homolog (mouse)	A member of the immunoglobulin superfamily of cell surface multiligand receptors, interacts with other molecules implicated in homeostasis, development, inflammation, and certain diseases
218418_s_at	KANK2	KN motif and ankyrin repeat domains 2	Phosphorylated upon DNA damage, probably by ATM or ATR
218950_at	arap3	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3	Phosphatidylinositol-3,4,5-trisphosphate-dependent GTPase-activating protein that modulates actin cytoskeleton remodeling by regulating ARF and RHO family members
219777_at	GIMAP6	GTPase, IMAP family member 6	Belongs to the IAN GTP-binding protein family
219888_at	SPAG4	Sperm associated antigen 4	Assist the organization and assembly of outer dense fibers (ODFs), a specific structure of the sperm tail
221747_at	TNS1	Tensin 1	Involved in cell migration, cartilage development and in linking signal transduction pathways to the cytoskeleton
43427_at	Acacb	Acetyl-Coenzyme A carboxylase beta	Involved in the provision of malonyl-CoA or in the regulation of fatty acid oxidation, rather than fatty acid biosynthesis
49452_at		Acetyl-Coenzyme A carboxylase beta	

Table 5

The 10 top activated pathways associated with lung cancer.

KEGG ID	Mean T value	Pathway name
04110	11.62343102	Cell cycle
03030	8.01760007	DNA replication
04510	-7.29863354	Focal adhesion
03013	7.12391298	RNA transport
04141	7.07550376	Protein processing in endoplasmic reticulum
04114	6.59124814	Oocyte meiosis
04270	-6.40126445	Vascular smooth muscle contraction
04810	-5.94453207	Regulation of actin cytoskeleton
04144	-5.87083652	Endocytosis
03440	5.83949963	Homologous recombination

Table 6

The 10 top activated pathways associated with cigarette smoking.

KEGG ID	Mean T value	Pathway name
04270	37.38459432	Vascular smooth muscle contraction
04110	-36.20249970	Cell cycle
04144	32.80585199	Endocytosis
00240	-30.98861570	Pyrimidine metabolism
03030	-29.90714777	DNA replication
03008	-29.79492651	Ribosome biogenesis in eukaryotes
04141	-27.67605174	Protein processing in endoplasmic reticulum
03013	-26.59872359	RNA transport
04080	26.43691493	Neuroactive ligand-receptor interaction
04920	25.36864988	Adipocytokine signaling pathway

Polo-like kinase 4 (PLK4) encodes a member of the polo family of serine/threonine protein kinases that regulate centriole duplication during the cell cycle and play an essential role in cell proliferation and tumorigenesis.^{59,60} It was suggested as a downstream protein of p53 and an appealing cancer drug target in lung cancer cell model.⁶¹ UBE2I encodes a member of the E2 ubiquitin-conjugating enzyme family that is important for targeting abnormal or short-lived proteins for degradation and involved in tumor growth and disease progression in human epithelial ovarian cancer.⁶² Nth endonuclease III-like 1 (nthl1)-encoded DNA N-glycosylase of the endonuclease III family has DNA glycosylase activity on DNA substrates containing oxidized pyrimidine residues and apurinic/apyrimidinic lyase activity. Human nthl1 was down-regulated in tumor tissues as compared to normal tissues from smoking patents.⁶³

Solute carrier family 10 (sodium/bile acid cotransporter family), member 3 (Slc10a3) was differentially expressed in ovarian cancers.⁶⁴ The protein encoded by minichromosome maintenance complex component 10 was essential for the initiation and elongation phases of DNA replication⁶⁵ and in carcinogenesis.⁶⁶ Sfi1 homolog, spindle assembly associated (yeast) plays a role in the dynamic structure of centrosome-associated contractile fibers via its interaction with CETN2.⁶⁷

Most of top ranked genes have been reported to correlate with cigarette smoking and lung cancer, including eIF2a kinase,^{26–31} secreted phosphoprotein 1,^{34–36} alpha-2-macroglobulin,^{37–40} catalase,⁴¹ ednrb,^{42–46} surfactant protein C,^{47,48} Ppm1f,^{24,25} cadherin 5,^{49–51} cytochrome P450, family 1, subfamily B, polypeptide 1^{53,54} and nthl1.⁶³ Of them, srpk1,^{32,33} SASH1,⁵² WWTR1,^{55,56} MICAL,⁵⁸ PLK4,^{59,60} UBE2I,⁶² Slc10a⁶⁴ or minichromosome maintenance complex component 10 (Mcm10)^{65,66} were reported to play roles in carcinogenesis, although the roles in cigarette smoking remain unclear. Methyltransferase-like 5 and Sfi1 homolog, spindle assembly associated (yeast)⁶⁷ were first suggested in present study as a candidate biomarker of lung cancer and the functions should be further validated as the potential therapeutic targets.

Candidate pathways from networks

GSEA is one of powerful analytic methods to reveal gene sets that share common biological function, chromosomal location, or regulation even when single-gene analysis reveals very few overlapping genes between groups.¹⁶ Using the GSEA, six pathways were identified to be involved in the processes of cigarette smoking and lung cancer, including cell cycle, DNA replication, RNA transport, protein processing in endoplasmic reticulum, vascular smooth muscle contraction and endocytosis. Our results were at least partially supported by the previous finding that cell cycle genes were differentially expressed in C/N samples and T/NT¹² and suggested the practicality of this approach and verification of the pathway. The application of GSEA could also identify disease-specific targets for therapy.

Among the 19 genes identified as the common genes both involved in the top ranked metabolic pathways and the genetic networks (Table 8), GRK5, Ramp2, acetyl-Coenzyme A carboxylase beta and arap3 were found as the common genes involved in both cancer- and smoking-related genetic networks. GRKs, originally recognized for their conserved role in G-protein-coupled receptor signal control, were found to play an immediate role in the regulation of human colon cancer or prostate cancer cells.^{68,69} Ramp2 was reported to be involved in multistep process of angiogenesis and tumoral cell proliferation including lung cancer.^{70,71}

Arap3, a multimodular signaling protein that is a substrate of Src family kinases, was suggested to play a role in the regulation of cell adhesion and invasion and inhibit cancer cell dissemination.⁷² Acab was reported to play a role in metabolic diseases,⁷³ while Elf2a kinase,^{26–31} ednrb^{42–46} and UBE2I⁶² were preliminarily identified as the top ranked genes with high scores in lung cancer or cigarette smoking related genetic networks and were discussed before.

The rest 12 genes were demonstrated to play a role in tumor genesis, prognosis, dissemination, drug resistance or radio-resistance in lung or non-lung cancers. For example, Agtr1 was demonstrated to play a role in breast cancer progression and considered as a therapeutic target in ER-positive and ERBB2-negative breast cancer.^{74,75} Natriuretic peptide receptor A methylation was associated with survival in colorectal cancer⁷⁶ and hepatocellular carcinoma.⁷⁷ Caveolin 1 is an essential protein constituent of caveolae and was associated with drug resistance and poor prognosis in NSCLC.⁷⁸ Caveolin 2 was also reported to serve as a tumor marker in lung cancer.⁷⁹ The polymorphism of fibroblast growth factor receptor 4 might be associated with prognosis in a wide variety of cancer types and lung cancer.^{80,81} Transforming growth factor, beta receptor II was a defined gene to play an important role in carcinogenesis⁸² and CRYAB could suppress the progression of nasopharyngeal carcinoma by associating with the cadherin/catenin adherens junction and modulating the β -catenin function.⁸³ Fractionated ionizing radiation had identified up-regulation of CRYAB in radioresistant of NSCLC.⁸⁴

The expression of sphingosine-1-phosphate receptor 1, a G protein-coupled receptor for the lysophospholipid sphingosine-1-phosphate, was elevated in STAT3-positive tumors.⁸⁵ The polymorphisms of Gln223Arg of LEPR was proposed to be a molecular marker for the progression and prognosis of NSCLC.⁸⁶ CD36 was involved in the growth and metastasis of breast cancer through regulating endothelial cell apoptosis.⁸⁷ Eukaryotic translation initiation factor 1A, Y-linked was found to appeal a significant different expression pattern in patients with tumors.⁸⁸

Challenges and opportunities

A number of discriminative sub-networks were proposed, including graph theory-based methods,^{89,90} protein association models,⁹¹ or transcriptome network.⁹² There was no information

Table 7

The five top ranked genes involved in lung cancer or cigarette smoking related pathways.

Pathway	Tumor/normal		Current smoker/never smoker	
	Probe ID	Gene name	Probe ID	Gene name
04270 "Vascular smooth muscle contraction"	205357_s_at	Angiotensin II receptor, type 1	205779_at	Receptor (G protein-coupled) activity modifying protein 2
	32625_at	Natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)	201957_at	Protein phosphatase 1, regulatory (inhibitor) subunit 12B
	205326_at	Receptor (G protein-coupled) activity modifying protein 3	206917_at	Guanine nucleotide binding protein (G protein), alpha 13
	205779_at	Receptor (G protein-coupled) activity modifying protein 2	206811_at	Adenylate cyclase 8 (brain)
04144 "Endocytosis"	201496_x_at	Myosin, heavy chain 11, smooth muscle	202660_at	Inositol 1,4,5-triphosphate receptor, type 2
	204396_s_at	G protein-coupled receptor kinase 5	203105_s_at	Dynamin 1-like
	203065_s_at	Caveolin 1, caveolae protein, 22 kDa	206254_at	Epidermal growth factor (beta-urogastrone)
	203323_at	Caveolin 2	218679_s_at	Vacuolar protein sorting 28 homolog (<i>S. cerevisiae</i>)
	204579_at	Fibroblast growth factor receptor 4	218085_at	Chromatin modifying protein 5
	220318_at	Epsin 3	215044_s_at	Signal transducing adaptor molecule (SH3 domain and ITAM motif) 2
04110 "Cell cycle"	200953_s_at	Cyclin D2	204822_at	TTK protein kinase
	201700_at	Cyclin D3	209642_at	Budding uninhibited by benzimidazoles 1 homolog (yeast)
	202870_s_at	Cell division cycle 20 homolog (<i>S. cerevisiae</i>)	210240_s_at	Cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4)
03030 "DNA replication"	207574_s_at	Growth arrest and DNA-damage-inducible, beta	203418_at	Cyclin A2
	222036_s_at	Minichromosome maintenance complex component 4	203362_s_at	MAD2 mitotic arrest deficient-like 1 (yeast)
	221143_at	Replication protein A4, 34 kDa	204128_s_at	Replication factor C (activator 1) 3, 38 kDa
	222036_s_at	Minichromosome maintenance complex component 4	221143_at	Replication protein A4, 34 kDa
	201528_at	Replication protein A1, 70 kDa	209507_at	Replication protein A3, 14 kDa
	204023_at	Replication factor C (activator 1) 4, 37 kDa	204023_at	Replication factor C (activator 1) 4, 37 kDa
	204767_s_at	Flap structure-specific endonuclease 1	203209_at	Replication factor C (activator 1) 5, 36.5 kDa
03013 "RNA transport"	208740_at	Sin3A-associated protein, 18 kDa	204409_s_at	Eukaryotic translation initiation factor 1A, Y-linked
	201738_at	Eukaryotic translation initiation factor 1B	208760_at	Ubiquitin-conjugating enzyme E2I (UBC9 homolog, yeast)
	208707_at	Eukaryotic translation initiation factor 5	208887_at	Eukaryotic translation initiation factor 3, subunit G
	208760_at	Ubiquitin-conjugating enzyme E2I (UBC9 homolog, yeast)	219252_s_at	Gem (nuclear organelle) associated protein 8
	208051_s_at	Poly(A) binding protein interacting protein 1; similar to poly(A) binding protein interacting protein 1	218894_s_at	Mago-nashi homolog B (<i>Drosophila</i>)
04141 "Protein processing in endoplasmic reticulum"	209283_at	Crystallin, alpha B	202908_at	Wolfram syndrome 1 (wolframin)
	202908_at	Wolfram syndrome 1 (wolframin)	212902_at	SEC24 family, member A (<i>S. cerevisiae</i>)
	202014_at	Protein phosphatase 1, regulatory (inhibitor) subunit 15A	200805_at	Lectin, mannose-binding 2
	218918_at	Mannosidase, alpha, class 1C, member 1	218436_at	SIL1 homolog, endoplasmic reticulum chaperone (<i>S. cerevisiae</i>)
	211048_s_at	Protein disulfide isomerase family A, member 4	218257_s_at	UDP-glucose ceramide glucosyltransferase-like 1

Table 8

Common genes both involved in the genetic network and GSEA identified pathways.

KEGG ID	Pathway names	Cancer	Smoking
04270	Vascular smooth muscle contraction	205357_s_at 32625_at 205779_at	Angiotensin II receptor, type 1 (Agtr1) Natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A) (NPR1) Receptor (G protein-coupled) activity modifying protein 2 (Ramp2)
04110	Cell cycle	None	203418_at
04144	Endocytosis	203065_s_at 203323_at 204579_at 204396_s_at 208944_at 218950_at	Caveolin 1, caveolae protein, 22 kDa (cav1) Caveolin 2 (cav2) Fibroblast growth factor receptor 4 (Fgfr4) G protein-coupled receptor kinase 5 (GRK5) Transforming growth factor, beta receptor II (70/80 kDa) (Tgfr2) ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3 (arap3)
03030	DNA replication	None	None
04141	Protein processing in endoplasmic reticulum	209283_at 217736_s_at	Crystallin, alpha B (CRYAB) Eukaryotic translation initiation factor 2-alpha kinase 1 (Eif2ak1)
03013	RNA transport	None	208760_at 204409_s_at
04510	Focal adhesion	— [#]	213675_at
04080	Neuroactive ligand-receptor interaction	205357_s_at 204642_at 204273_at 209894_at	Angiotensin II receptor, type 1 (Agtr1) Sphingosine-1-phosphate receptor 1 (S1PR1) Endothelin receptor type 8 (ednrB) Leptin receptor (LEPR)
04920	Adipocytokine signaling pathway	49452_at 206488_s_at 209894_at	acetyl-Coenzyme A carboxylase beta (AcacB) CD36 molecule (thrombospondin receptor) (CD36) Leptin receptor (LEPR)
00240	Pyrimidine metabolism	None	None
03008	Ribosome biogenesis in eukaryotes	None	None
04810	Regulation of actin cytoskeleton	—	None
03440	Homologous recombination	—	None

“None” means there is no common gene between the metabolic pathway and the genetic network.

[#] “—” means this metabolic pathway is not among the ten top activated pathways for cancer or smoking.

on the disease stage or smoking condition, since those studies mainly included groups with or without cancer. With the protein association network analysis, gene expression profiles were overlaid to the corresponding proteins resulting in a neglecting of information on posttranslational modification.⁹¹ Instead of existed methods, we take a more direct approach to identify potential disease-specific genes which appealed a higher change fold compared to other genes according to different disease conditions or smoking history. It will be a great challenge to confirm the expression of those identified genes in the formation of lung cancer, the production of correspondent proteins in the tissue, or interaction between identified genes and targeted proteins within the network, due to the variation between gene expression and protein production and between genomics and proteomics.^{93,94} Our results initiated the understanding of smoking- or lung cancer-associated genes and networks, as an example of cancer bioinformatics approaching systems clinical medicine.⁹⁵

The GSEA-based network analysis considers differentially expressed pathways and the genes-associated network structure. The present study evidenced that applied method was useful and helpful, while further improvements are still needed. The restriction in the size of the genetic network from the sample size may exclude some interested genes. For example, over 9365 of 22,283 genes in the network were significantly different between tumor and normal patient with IG scores >0, while only the most relevant gene sub-groups and pathways were selected using the proposed network-based computational analysis strategy. It is important to

apply clinical bioinformatics by integrating genomics/proteomics, bioinformatics or computational systems biology with clinical diagnosis, examination, imaging, functional tests, monitoring, and therapies.^{96–99} Future studies are expected to explore the potential mechanisms and correlation of identified networks with disease-specific severities, durations, clinical phenotypes, responses to therapies, or prevention.^{100–102}

Conclusion

A new and powerful network-based biomarker detection approach coupled with gene set enrichment analysis was used to identify and validate genes associated with lung cancer and relating pathways. Results from the present study revealed a panel of new and unexpected genes with potential physiological functions in cigarette smoking or lung cancer using the coupled discovery strategy, in addition to those consistent with previous findings in cigarette smoking and lung cancer. Thus, our studies indicate that disease-specific network biomarkers, interaction between genes/proteins, or cross-talking of pathways provide more specific values for the development of precision therapies for lung.

Authors' contributions

X.C.F. performed the analysis, evaluated the results and wrote the manuscript. M.N. developed the method. C.B. and X.D.W.

provided essential guidance and revised the manuscript. C.X.B. contribute to revision of the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

This work was supported by Shanghai Leading Academic Discipline Project (Project Number: B115), Fudan University (Distinguished Professor Grant), Shanghai Science & Technology Committee (08PJ1402900, 9540702600, 08DZ2293104); and partly by the Tiroler Wissenschaftsfond, Austria.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ctrv.2012.06.001>.

References

- Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* 2008;**83**:584–94.
- Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, et al. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prev Res (Phila)* 2011;**4**:803–17.
- Begum S. Molecular changes in smoking-related lung cancer. *Expert Rev Mol Diagn* 2012;**12**:93–106.
- Baumgartner C, Osl M, Netzer M, Baumgartner D. Bioinformatic-driven search for metabolic biomarkers in disease. *J Clin Bioinform* 2011;**1**:1–10.
- Wang X. Role of clinical bioinformatics in the development of network-based Biomarkers. *J Clin Bioinform* 2011;**1**:28.
- Netzer M, Weinberger KM, Handler M, Seger M, Fang X, Kugler KG, et al. A computational strategy for the identification and kinetic analysis of metabolic biomarkers: a pilot study for profiling human response to physical exercise. *J Clin Bioinform* 2011;**1**:34.
- Sienel W, Dango S, Ehrhardt P, Eggeling S, Kirschbaum A, Passlick B. The future in diagnosis and staging of lung cancer. Molecular techniques. *Respiration* 2006;**73**:575–80.
- Goldenberg A, Mostafavi S, Quon G, Boutros PC, Morris QD. Unsupervised detection of genes of influence in lung cancer using biological networks. *Bioinformatics* 2011;**27**:3166–72.
- Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, et al. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prev Res (Phila)* 2011;**4**:803–17.
- Wen J, Fu J, Zhang W, Guo M. Genetic and epigenetic changes in lung carcinoma and their clinical implications. *Mod Pathol* 2011;**24**: 932–43.
- Kim B, Lee HJ, Choi HY, Shin Y, Nam S, Seo G, et al. Clinical validity of the lung cancer biomarkers identified by bioinformatics analysis of public expression data. *Cancer Res* 2007;**67**:7431–8.
- Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* 2008;**3**:e1651.
- Netzer M, Fang X, Handler M, Baumgartner C. A coupled three-step network-based approach to identify genes associated with breast cancer. *BIOTECHNO 2012, The Fourth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies* 2012;1–5.
- Quinlan JR. Induction of decision trees. *Mach Learn* 1986;**1**(1):81–106.
- Hall M.A. Correlation-based feature selection for machine learning. Ph.D. dissertation, University of Waikato, 1999.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545–50.
- R Development Core Team, R. The R Project for Statistical Computing. Available: <http://www.R-project.org>.
- Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, et al. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2007;**2**(10):2366–82.
- Wang X, Chorley BN, Pittman GS, Kleeberger SR, Brothers 2nd J, Liu G, et al. Genetic variation and antioxidant response gene expression in the bronchial airway epithelium of smokers at risk for lung cancer. *PLoS One* 2010;**5**:e11934.
- Wang R, Wang G, Ricard MJ, Ferris B, Strulovici-Barel Y, Salit J, et al. Smoking-induced upregulation of AKR1B10 expression in the airway epithelium of healthy individuals. *Chest* 2010;**138**:1402–10.
- Ding H, Wang F, Ding X, Song X, Lu X, Zhang K, et al. Association study of semaphorin 5A with risk of Parkinson's disease in a Chinese Han population. *Brain Res* 2008;**1245**:126–9.
- Lu TP, Tsai MH, Lee JM, Hsu CP, Chen PC, Lin CW, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev* 2010;**19**:2590–7.
- Sadanandam A, Varney ML, Kinarsky L, Ali H, Mosley RL, Singh RK. Identification of functional cell adhesion molecules with a potential role in metastasis by a combination of in vivo phage display and in silico analysis. *OMICS* 2007;**11**:41–57.
- Rose JE, Behm FM, Drgon T, Johnson C, Uhl GR. Personalized smoking cessation: interactions between nicotine dose, dependence and quit-success genotype score. *Mol Med* 2010;**16**:247–53.
- Susila A, Chan H, Loh AX, Phang HQ, Wong ET, Tergaonkar V, et al. The POPX2 phosphatase regulates cancer cell motility and invasiveness. *Cell Cycle* 2010;**9**:179–87.
- Chen JJ, London IM. Regulation of protein synthesis by heme-regulated eIF-2 alpha kinase. *Trends Biochem Sci* 1995;**20**:105–8.
- Koumenis C, Wouters BG. "Translating" tumor hypoxia: unfolded protein response (UPR)-dependent and UPR-independent pathways. *Mol Cancer Res* 2006;**4**:423–36.
- Liu Y, László C, Liu Y, Liu W, Chen X, Evans SC, et al. Regulation of G(1) arrest and apoptosis in hypoxia by PERK and GCN2-mediated eIF2alpha phosphorylation. *Neoplasia* 2010;**12**:61–8.
- Pataer A, Raso MG, Correa AM, Behrens C, Tsuta K, Solis L, et al. Prognostic significance of RNA-dependent protein kinase on non-small cell lung cancer patients. *Clin Cancer Res* 2010;**16**:5522–8.
- Hengstermann A, Müller T. Endoplasmic reticulum stress induced by aqueous extracts of cigarette smoke in 3T3 cells activates the unfolded-protein-response-dependent PERK pathway of cell survival. *Free Radic Biol Med* 2008;**44**:1097–107.
- Tagawa Y, Hiramatsu N, Kato H, Sakoh T, Nakajima S, Hayakawa K, et al. Toxicology: Induction of CCAAT/enhancer-binding protein-homologous protein by cigarette smoke through the superoxide anion-triggered PERK-eIF2α pathway. *Toxicology* 2011;**287**:105–12.
- Hayes GM, Carrigan PE, Beck AM, Miller LJ. Targeting the RNA splicing machinery as a novel treatment strategy for pancreatic carcinoma. *Cancer Res* 2006;**66**:3819–27.
- Hayes GM, Carrigan PE, Miller LJ. Serine-arginine protein kinase 1 overexpression is associated with tumorigenic imbalance in mitogen-activated protein kinase pathways in breast, colonic, and pancreatic carcinomas. *Cancer Res* 2007;**67**:2072–80.
- Weber GF. The cancer biomarker osteopontin: combination with other markers. *Cancer Genomics Proteomics* 2011;**8**:263–88.
- Das R, Mahabeshwar GH, Kundu GC. Osteopontin induces AP-1-mediated secretion of urokinase-type plasminogen activator through c-Src-dependent epidermal growth factor receptor transactivation in breast cancer cells. *J Biol Chem* 2004;**279**:11051–64.
- Delimpoura V, Bakakos P, Tseliou E, Bessa V, Hillas G, Simoes DC, et al. Increased levels of osteopontin in sputum supernatant in severe refractory asthma. *Thorax* 2010;**65**:782–6.
- Sottrup-Jensen L. Alpha-macroglobulins: structure, shape, and mechanism of proteinase complex formation. *J Biol Chem* 1989;**264**:11539–42.
- Marchandise FX, Mathieu B, Francis C, Sibille Y. Local increase of antiprotease and neutrophil elastase-alpha 1-proteinase inhibitor complexes in lung cancer. *Eur Respir J* 1989;**2**:623–9.
- Lim W, Jeong W, Kim JH, Lee JY, Kim J, Bazer FW, et al. Differential expression of alpha 2 macroglobulin in response to diethylstilbestrol and in ovarian carcinomas in chickens. *Reprod Biol Endocrinol* 2011;**9**:137.
- Suriyaprom K, Harnroongroj T, Namjuntun P, Chantarapapong Y, Tungtrongchitr R. Effects of tobacco smoking on alpha-2-macroglobulin and some biochemical parameters in Thai males. *Southeast Asian J Trop Med Public Health* 2007;**38**:918–26.
- Goyal MM, Basak A. Human catalase: looking for complete identity. *Protein Cell* 2010;**1**:888–97.
- Wülfing P, Diallo R, Kersting C, Wülfing C, Poremba C, Rody A, et al. Expression of endothelin-1, endothelin-A, and endothelin-B receptor in human breast cancer and correlation with long-term follow-up. *Clin Cancer Res* 2003;**9**:4125–31.
- Zhou L, Feng X, Shan W, Zhou W, Liu W, Wang L, et al. Epigenetic and genetic alterations of the EDNRB gene in nasopharyngeal carcinoma. *Oncology* 2007;**72**:357–63.
- Knight LJ, Burrage J, Bujac SR, Haggerty C, Graham A, Gibson NJ, et al. Epigenetic silencing of the endothelin-B receptor gene in non-small cell lung cancer. *Int J Oncol* 2009;**34**:465–71.
- Chen SC, Lin CY, Chen YH, Fang HY, Cheng CY, Chang CW, et al. Aberrant promoter methylation of EDNRB in lung cancer in Taiwan. *Oncol Rep* 2006;**15**:167–72.
- Taillé C, Guénégou A, Almolki A, Piperaud M, Leynaert B, Vuillaumier S, et al. ETB receptor polymorphism is associated with airway obstruction. *BMC Pulm Med* 2007;**7**:5.
- Wikenheiser KA, Vorbroke DK, Rice WR, Clark JC, Bachurski CJ, Oie HK, et al. Production of immortalized distal respiratory epithelial cell lines from surfactant protein C/simian virus 40 large tumor antigen transgenic mice. *Proc Natl Acad Sci U S A* 1993;**90**:11029–33.
- Stenger PC, Alonso C, Zasadzinski JA, Waring AJ, Jung CL, Pinkerton KE. Environmental tobacco smoke effects on lung surfactant film organization. *Biochim Biophys Acta* 2009;**1788**:358–70.

49. Dejana E, Tournier-Lasserre E, Weinstein BM. The control of vascular integrity by endothelial cell junctions: molecular basis and pathological implications. *Dev Cell* 2009;**16**:209–21.
50. Li H, Shi X, Liu J, Hu C, Zhang X, Liu H, et al. The soluble fragment of VE-cadherin inhibits angiogenesis by reducing endothelial cell proliferation and tube capillary formation. *Cancer Gene Ther* 2010;**17**:700–7.
51. Barbieri SS, Ruggiero L, Tremoli E, Weksler BB. Suppressing PTEN activity by tobacco smoke plus interleukin-1 β modulates dissociation of VE-cadherin/ β -catenin complexes in endothelium. *Arterioscler Thromb Vasc Biol* 2008;**28**:732–8.
52. Zeller C, Hinzmann B, Seitz S, Prokoph H, Burkhard-Goettges E, Fischer J, et al. SASH1: a candidate tumor suppressor gene on chromosome 6q24.3 is downregulated in breast cancer. *Oncogene* 2003;**22**:2972–83.
53. Timofeeva MN, Kropp S, Sauter W, Beckmann L, Rosenberger A, Illig T, et al. CYP450 polymorphisms as risk factors for early-onset lung cancer: gender-specific differences. *Carcinogenesis* 2009;**30**:1161–9.
54. Wenzlaff AS, Cote ML, Bock CH, Land SJ, Santer SK, Schwartz DR, et al. CYP1A1 and CYP1B1 polymorphisms and risk of lung cancer among never smokers: a population-based study. *Carcinogenesis* 2005;**26**:2207–12.
55. Errani C, Zhang L, Sung YS, Hajdu M, Singer S, Maki RG, et al. A novel WWTR1–CAMTA1 gene fusion is a consistent abnormality in epithelioid hemangioendothelioma of different anatomic sites. *Genes Chromosomes Cancer* 2011;**50**:644–53.
56. Balasenthil S, Chen N, Lott ST, Chen J, Carter J, Grizzle WE, et al. A migration signature and plasma biomarker panel for pancreatic adenocarcinoma. *Cancer Prev Res (Phila)* 2011;**4**:137–49.
57. Suzuki T, Nakamoto T, Ogawa S, Seo S, Matsumura T, Tachibana K, et al. MICAL, a novel CasL interacting molecule, associates with vimentin. *J Biol Chem* 2002;**277**:14933–41.
58. Ashida S, Furihata M, Katagiri T, Tamura K, Anazawa Y, Yoshioka H, et al. Expression of novel molecules, MICAL2-PV (MICAL2 prostate cancer variants), increases with high Gleason score and prostate cancer progression. *Clin Cancer Res* 2006;**12**:2767–73.
59. Habedanck R, Stierhof YD, Wilkinson CJ, Nigg EA. The Polo kinase Plk4 functions in centriole duplication. *Nat Cell Biol* 2005;**7**:1140–6.
60. Kleylein-Sohn J, Westendorf J, Le Clech M, Habedanck R, Stierhof YD, Nigg EA. Plk4-induced centriole biogenesis in human cells. *Dev Cell* 2007;**13**:190–202.
61. Sun Y. P53 and its downstream proteins as molecular targets of cancer. *Mol Carcinog* 2006;**45**:409–15.
62. Guo Y, Fu P, Zhu H, Reed E, Remick SC, Petros W, et al. Correlations among ERCC1, XPB, UBE2I, EGF, TAL2 and ILF3 revealed by gene signatures of histological subtypes of patients with epithelial ovarian cancer. *Oncol Rep* 2012;**27**:286–92.
63. Radak Z, Goto S, Nakamoto H, Udud K, Papai Z, Horvath I. Lung cancer in smoking patients inversely alters the activity of hOGG1 and hNTH1. *Cancer Lett* 2005;**219**:191–5.
64. Cheng L, Lu W, Kulkarni B, Pejovic T, Yan X, Chiang JH, et al. Analysis of chemotherapy response programs in ovarian cancers by the next-generation sequencing technologies. *Gynecol Oncol* 2011;**117**:159–69.
65. Taylor M, Moore K, Murray J, Aves SJ, Price C. Mcm10 interacts with Rad4/Cut5 (TopBP1) and its association with origins of DNA replication is dependent on Rad4/Cut5 (TopBP1). *DNA Repair* 2011;**10**:1154–63.
66. Lim HJ, Jeon Y, Jeon CH, Kim JH, Lee H. Targeted disruption of Mcm10 causes defective embryonic cell proliferation and early embryo lethality. *Biochim Biophys Acta* 2011;**1813**:1777–83.
67. Kilmartin JV. Sfi1p has conserved centrion-binding sites and an essential function in budding yeast spindle pole body duplication. *J Cell Biol* 2003;**162**:1211–21.
68. Wu CC, Tsai FM, Shyu RY, Tsai YM, Wang CH, Jiang SY. G protein-coupled receptor kinase 5 mediates Tazartene-induced gene 1-induced growth suppression of human colon cancer cells. *BMC Cancer* 2011;**11**:175.
69. Kim JJ, Chakraborty P, Wang Z, Daaka Y. G-protein coupled receptor kinase 5 regulates prostate tumor growth. *J Urol* 2012;**187**:322–9.
70. Yue W, Dacic S, Sun Q, Landreneau R, Guo M, Zhou W, et al. Frequent inactivation of RAMP2, EFEMP1 and Dutt1 in lung cancer by promoter hypermethylation. *Clin Cancer Res* 2007;**13**:4336–44.
71. Kaafarani I, Fernandez-Sauze S, Berenguer C, Chinot O, Delfino C, Dussert C, et al. Targeting adrenomedullin receptors with systemic delivery of neutralizing antibodies inhibits tumor angiogenesis and suppresses growth of human tumor xenografts in mice. *FASEB J* 2009;**23**:3424–35.
72. Yagi R, Tanaka M, Sasaki K, Kamata R, Nakanishi Y, Kanai Y, et al. ARAP3 inhibits peritoneal dissemination of scirrhous gastric carcinoma cells by regulating cell adhesion and invasion. *Oncogene* 2011;**30**:1413–21.
73. Ma L, Mondal AK, Murea M, Sharma NK, Tönjes A, Langberg KA, et al. The effect of ACACB cis-variants on gene expression and metabolic traits. *PLoS One* 2011;**6**:e23860.
74. Ateeq B, Tomlins SA, Chinnaiyan AM. AGTR1 as a therapeutic target in ER-positive and ERBB2-negative breast cancer cases. *Cell Cycle* 2009;**8**:3794–5.
75. Xi B, Zeng T, Liu L, Liang Y, Liu W, Hu Y, et al. Association between polymorphisms of the renin-angiotensin system genes and breast cancer risk: a meta-analysis. *Breast Cancer Res Treat* 2011;**130**:561–8.
76. Cavalieri D, Dolara P, Mini E, Luceri C, Castagnini C, Toti S, et al. Analysis of gene expression profiles reveals novel correlations with the clinical course of colorectal cancer. *Oncol Res* 2007;**16**:535–48.
77. Deng YB, Nagae G, Midorikawa Y, Yagi K, Tsutsumi S, Yamamoto S, et al. Identification of genes preferentially methylated in hepatitis C virus-related hepatocellular carcinoma. *Cancer Sci* 2010;**101**:1501–10.
78. Ho CC, Kuo SH, Huang PH, Huang HY, Yang CH, Yang PC. Caveolin-1 expression is significantly associated with drug resistance and poor prognosis in advanced non-small cell lung cancer patients treated with gemcitabine-based chemotherapy. *Lung Cancer* 2008;**59**:105–10.
79. Wikman H, Seppänen JK, Sarhadi VK, Kettunen E, Salmenkivi K, Kuosma E, et al. Caveolins as tumour markers in lung cancer detected by combined use of cDNA and tissue microarrays. *J Pathol* 2004;**203**:584–93.
80. Matakidou A, El Galt R, Rudd MF, Webb EL, Bridle H, Eisen T, et al. Further observations on the relationship between the FGFR4 Gly388Arg polymorphism and lung cancer prognosis. *Br J Cancer* 2007;**96**:1904–7.
81. Frullanti E, Berking C, Harbeck N, Jézéquel P, Haugen A, Mawrin C, et al. Meta and pooled analyses of FGFR4 Gly388Arg polymorphism as a cancer prognostic factor. *Eur J Cancer Prev* 2011;**20**:340–7.
82. Borczuk AC, Sole M, Lu P, Chen J, Wilgus ML, Friedman RA, et al. Progression of human bronchioloalveolar carcinoma to invasive adenocarcinoma is modeled in a transgenic mouse model of K-ras-induced lung cancer by loss of the TGF- β type II receptor. *Cancer Res* 2011;**71**:6665–75.
83. Huang Z, Cheng Y, Chiu PM, Cheung FM, Nicholls JM, Kwong DL, et al. Tumor suppressor Alpha B-crystallin (CRYAB) associates with the cadherin/catenin adherens junction and impairs NPC progression-associated properties. *Oncogene* 2011 Dec 12. <http://dx.doi.org/10.1038/ncr.2011.529>. [Epub ahead of print].
84. Xu QY, Gao Y, Liu Y, Yang WZ, Xu XY. Identification of differential gene expression profiles of radioresistant lung cancer cell line established by fractionated ionizing radiation in vitro. *Chin Med J (Engl)* 2008;**121**:1830–7.
85. Lee H, Deng J, Kujawski M, Yang C, Liu Y, Herrmann A, et al. STAT3-induced S1PR1 expression is crucial for persistent STAT3 activation in tumors. *Nat Med* 2010;**16**:1421–8.
86. Li Y, Geng J, Wang Y, Lu Q, Du Y, Wang W, et al. The role of leptin receptor gene polymorphisms in determining the susceptibility and prognosis of NSCLC in Chinese patients. *J Cancer Res Clin Oncol* 2012;**138**:311–6.
87. Koch M, Hussein F, Woeste A, Gründker C, Frontzek K, Emons G, et al. CD36-mediated activation of endothelial cell apoptosis by an N-terminal recombinant fragment of thrombospondin-2 inhibits breast cancer growth and metastasis in vivo. *Breast Cancer Res Treat* 2011;**128**:337–46.
88. Tabernero MD, Espinosa AB, Maillou A, Rebelo O, Vera JF, Sayagues JM, et al. Patient gender is associated with distinct patterns of chromosomal abnormalities and sex chromosome linked gene-expression profiles in meningiomas. *Oncologist* 2007;**12**:1225–36.
89. Rapaport F, Zinoviyev A, Dutreix M, Barillot E, Vert JP. Classification of microarray data using gene networks. *BMC Bioinformatics* 2007;**8**:35.
90. Tian Z, Hwang T, Kuang R. A hypergraph-based learning algorithm for classifying gene expression and array CGH data with prior knowledge. *Bioinformatics* 2009;**25**:2831–8.
91. Wang YC, Chen BS. A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. *BMC Med Genomics* 2011;**4**:2.
92. Bai J, Hu S. Transcriptome network analysis reveals potential candidate genes for squamous lung cancer. *Int J Mol Med* 2012;**29**:95–101.
93. Zhao H, Adler KB, Bai C, Tang F, Wang XD. Epithelial proteomics in multiple organs and tissues: similarities and variations between cells, organs, and diseases. *J Proteome Res* 2006;**5**:743–55.
94. Wang XD, Adler KB, Chaudry IH, Ward PA. Better understanding of organ dysfunction requires proteomic involvement. *J Proteome Res* 2006;**5**:1060–2.
95. Wu D, Rice CM, Wang X. Cancer bioinformatics: a new approach to systems clinical medicine. *BMC Bioinformatics* 2012;**13**:71.
96. López E, Wesselink JJ, López I, Mendieta J, Gómez-Puertas P, Muñoz S. Technical phosphoproteomic and bioinformatic tools useful in cancer research. *J Clin Bioinform* 2011;**1**:26.
97. Chen H, Wang XD. Significance of bioinformatics in research of chronic obstructive pulmonary disease. *J Clin Bioinform* 2011;**1**:35.
98. Wu DJ, Zhu BJ, Wang XD. Metabonomics-based omics study and atherosclerosis. *J Clin Bioinform* 2011;**1**:30.
99. Yeh SH, Yeh HY, Soo VW. A network flow approach to predict drug targets from microarray data, disease genes and interactome network – case study on prostate cancer. *J Clin Bioinform* 2012;**2**:1.
100. Chen H, Song Z, Qian M, Bai C, Wang XD. Selection of disease-specific biomarkers by integrating inflammatory mediators with clinical informatics in AECOPD patients: a preliminary study. *J Cell Mol Med* 2012;**16**:1286–97.
101. Wang XD. A new vision of clinical and translational medicine: definition, commentary, understanding. *Clin Transl Med* 2012;**1**:5.
102. Wang XD, Marincola FM. A decade plus of translation: what do we understand? *Clin Transl Med* 2012;**1**:3.