



Bicluster and regulatory network analysis of differentially expressed genes in adenocarcinoma and squamous cell carcinoma

L. Li¹, J. Zhu¹, S.-X. Guo² and Y. Deng²

¹Department of Thoracic Oncology, Cancer Center, West China Hospital, Medical School, Sichuan University, GuoXueXiang, Chengdu, Sichuan, China

²West China Medical School, Sichuan University, Chengdu, Sichuan, China

Corresponding author: L. Li
E-mail: lu_lill@hotmail.com

Genet. Mol. Res. 12 (2): 1710-1719 (2013)

Received August 29, 2012

Accepted February 20, 2013

Published May 21, 2013

DOI <http://dx.doi.org/10.4238/2013.May.21.2>

ABSTRACT. Squamous cell carcinoma (SCC) and adenocarcinoma (AC) are the major histological types of non-small cell lung cancer (NSCLC). Although differences in molecular, histological, and clinical characteristics have been reported for both subtypes, no specific therapy exists thus far. The aim of this analysis was to identify potential therapeutic target genes that are specific to SCC and AC. We used microarray data to analyze the global gene expression profile of 58 human NSCLC specimens. We identified more than 2400 genes that were differentially expressed in SCC and AC. Bicluster analysis with iterative signature algorithm revealed 22 biclusters that were strongly associated with the histological subtypes AC and SCC of NSCLC. We also built a regulatory network of genes differentially expressed in SCC and AC. Some transcription factors and target genes related to lung cancer are linked in our network. Furthermore, we used the Database for Annotation, Visualization and Integrated Discovery to identify the main pathways in which these differentially expressed genes were involved. Eight pathways were enriched by this analysis. Our data provide a comprehensive transcriptional profile of candidate genes that

may be involved in the complex regulatory networks underlying the different NSCLC subtypes.

Key words: Iterative signature algorithm; Squamous cell carcinoma; Adenocarcinoma

INTRODUCTION

According to World Health Organization, lung cancer is the most common type of cancer in the world and the most common cause of cancer-related deaths in men and women, with 1.2 million new cases diagnosed every year (<http://www.who.int/mediacentre/news/releases/2003/pr27/en/>). Lung carcinomas are generally classified as either small-cell lung carcinoma (SCLC) or non-SCLC (NSCLC). Within these groups, further distinctions are made, with NSCLC being subclassified into adenocarcinoma (AC), squamous cell carcinoma (SCC), and large-cell carcinoma. After AC, SCC is the most frequent form of cancer, accounting for approximately 30% of all lung cancers. Inhaled carcinogens, such as cigarette smoke, are a major risk factor for the disease (Inamura et al., 2005). Because AC and SCC differ in their histopathological and clinical characteristics and in their relationship with smoking, their etiologies may be different. For example, different tumor suppressor genes may be related to the etiology of each type of cancer. Transcriptional differences between and within these NSCLC subtypes have been reported in several studies using microarray-based gene expression profiling. However, not all of these studies accounted for the potential confounding effects of NSCLC subtype-dependent transcriptomic variability. Furthermore, it remains a challenge to identify prognostically relevant transcriptomic differences between and within these subtypes that highlight the biological processes and molecular pathways underlying these cancers, which may harbor potential targets for therapeutic intervention.

DNA microarray analysis, which monitors the expression levels of thousands of genes simultaneously, has been used as a global approach to investigate the physiological mechanisms involved in health and disease (Spies et al., 2002). A high-throughput microarray experiment was designed to analyze genetic expression patterns of and identify potential therapeutic target genes in AC (Li et al., 2006). Genomic expression profiling has been proven to be a useful tool in identifying novel pathological mechanisms in human cancer (Wittchen et al., 2007).

Large sets of data, such as expression profiles from many samples, require analytic tools to reduce their complexity (Csardi et al., 2010). The iterative signature algorithm (ISA) (Ihmels et al., 2002; Bergmann et al., 2003; Ihmels et al., 2004) is a biclustering method that tries to find blocks that are different from the rest of the matrix. For example, the values covered by the bicluster are all above or below the background. It can be applied to identify coherent substructures (or modules) from any rectangular matrix of data. Its input is a matrix and its output is a set of biclusters, which are blocks of potentially reordered input matrices that fulfill some predefined criteria. As ISA is an unsupervised algorithm, it performs very well in finding modules, even in the presence of noise in the input matrix.

In the present study, we report a correlation between histopathological subtypes of NSCLC and global gene expression profiles using transcriptome analysis of 58 NSCLC patient samples. We performed bicluster analysis using ISA to classify the differentially expressed genes (DEGs) in SCC and AC into functional subgroups, and constructed a transcrip-

tome network linking transcription factors (TFs) and their target genes. Furthermore, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al., 2009) to identify the significant pathways in which these DEGs are involved.

MATERIAL AND METHODS

Microarray data selection and preprocessing

The gene expression profile of NSCLCs was obtained from the public functional genomics data repository Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) using the accession number GSE10245 (Kuner et al., 2009). These data are based on the Affymetrix Human Genome U133 plus 2.0 Array (Affymetrix, Santa Clara, CA, USA). Data from 40 AC and 18 SCC samples were used.

Pathway data

The Kyoto Encyclopedia of Genes and Genomes is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals (Kanehisa, 2002). The 'pathway' database records networks of molecular interactions in the cells, and variants of those specific to particular organisms (<http://www.genome.jp/kegg/>).

Regulation data

There are approximately 2600 proteins in the human genome that contain DNA-binding domains, and most of these are presumed to function as TFs (Wachi et al., 2005). The combinatorial use of a subset of the approximately 2000 human TFs easily accounts for the unique regulation of each gene in the human genome during development (Brivanlou and Darnell, 2002). These transcription factors are grouped into 5 superclass families, based on the presence of conserved DNA-binding domains. The TRANSFAC database contains data on transcription factors, and their experimentally proven binding sites and regulated genes (Wingender, 2008). Transcriptional Regulatory Element Database (TRED) has been built to accommodate the need for an integrated repository for both cis- and trans-regulatory elements in mammals (Jiang et al., 2007). TRED curates transcriptional regulation information, including transcription factor binding motifs and experimental evidence. The curation is currently focused on the target genes of 36 cancer-related TF families. A total of 227 regulatory relationships between 59 TFs and 136 target genes were collected from the 2 regulatory datasets.

Analysis of DEGs

The Limma library (Diboun et al., 2006) based on R (Team, 2011) and Bioconductor (Gentleman et al., 2004) was used to identify genes that were differentially expressed between AC and SCC subtypes. To control for the multiple testing issue, which might introduce too many false-positive results, the Benjamini-Hochberg (BH) method (Benjamini et al., 2001) was used to adjust the raw P values into a false-discovery rate. The original expression datasets from all conditions were processed into expression estimates, and these were used for con-

structing a linear model. Only genes with a P value less than 0.05 (2493 genes) were selected to perform the bicluster analysis. One hundred and eighty-nine DEGs showing a fold-change greater than 1.5 and a P value less than 0.05 were selected for the construction of the regulatory network.

Bicluster analysis

ISA (Ihmels et al., 2002) was developed to find biclusters (or modules as most of the ISA papers call them) that have correlated rows and columns. More precisely, the rows in the bicluster need to only be correlated across the columns of the bicluster and vice versa. Bicluster analysis with ISA using the ISA2 (Csardi et al., 2010) packages was performed to find enriched biclusters. After removing the biclusters whose sample count was less than 4 and gene count was less than 6, we were left with only 22 biclusters. The biclusters that had an absolute value of SCC sample numbers/AC sample numbers equal to or greater than 2/3 were considered to be significant.

Gene ontology (GO)-enrichment analysis

Functional annotation of DEGs was performed using the GO software BiNGO (Maere et al., 2005). We selected the DEGs and calculated a significance value for over- and under-represented GO categories using a cut-off criterion ($P < 0.05$). Furthermore, hypergeometric tests were used for statistical analysis and the BH method (Benjamini et al., 2001) was used for multiple testing correction with the hypergeometric distribution.

Pathway-enrichment analysis

DAVID (Huang et al., 2009), a high-throughput and integrated data-mining environment, analyzes gene lists derived from high-throughput genomic experiments. We used DAVID to identify over-represented pathways with P values less than 0.01.

RESULTS

GO-enrichment analysis of genes in all biclusters

By analyzing the microarray data from 40 AC and 18 SCC samples, we identified 2493 DEGs that showed statistically significant differences between the 2 groups. We selected these genes for further analysis. Bicluster analysis of all 58 NSCLC tumors using the 2493 most variably expressed transcripts revealed 22 biclusters (detailed information is shown in Table 1), which were strongly associated with the histological subtypes AC and SCC of NSCLC. We performed GO-enrichment analysis for each bicluster, and selected the most significant GO term associated with each bicluster as its GO term annotation (Table 1). As seen in Table 1, several biological process, such as “mitotic cell cycle” (biclusters 1-4, 10, 12, 15, 17-19, 21-22), “epidermis development” (biclusters 9, 16), “response to interleukin-6” (biclusters 7, 13), “epithelial tube branching involved in lung morphogenesis” (biclusters 6, 8, 11, 14), etc., were enriched. The GO enrichment of bicluster 9 is shown as an example in Figure 1.

Table 1. Gene ontology-enrichment analysis in the total 22 biclusters.

Bicluster ID	Gene count	Sample count			Description	P value	FDR
		All	SCC	AC			
1	349	5	0	5	Mitotic cell cycle	5.81E-24	1.39E-20
2	263	8	2	6	Mitotic cell cycle	5.17E-29	1.10E-25
3	278	8	3	5	Mitotic cell cycle	2.77E-34	6.53E-31
4	315	7	0	7	Mitotic cell cycle	3.75E-42	8.42E-39
5	115	22	12	10	Regulation of insulin receptor signaling pathway	2.75E-04	1.93E-01
6	96	20	12	8	Epithelial tube branching involved in lung morphogenesis	6.50E-04	1.94E-01
7	118	20	11	9	Response to interleukin-6	4.96E-05	7.20E-02
8	94	24	14	10	Epithelial tube branching involved in lung morphogenesis	5.33E-04	2.84E-01
9	127	16	13	3	Epidermis development	8.39E-08	7.91E-05
10	230	20	9	11	Mitotic cell cycle	8.10E-41	1.51E-37
11	113	20	11	9	Epithelial tube branching involved in lung morphogenesis	8.98E-04	2.14E-01
12	250	20	8	12	Mitotic cell cycle	8.95E-56	1.67E-52
13	128	18	10	8	Response to interleukin-6	7.08E-05	1.11E-01
14	116	20	11	9	Epithelial tube branching involved in lung morphogenesis	9.40E-04	2.89E-01
15	1108	8	6	2	Mitotic cell cycle	1.92E-15	7.69E-12
16	1053	9	8	1	Epidermis development	9.12E-07	3.62E-03
17	1064	10	4	10	Mitotic cell cycle	6.63E-42	2.60E-38
18	1096	10	4	6	Mitotic cell cycle	7.21E-42	2.94E-38
19	1120	9	4	5	Mitotic cell cycle	1.72E-34	7.01E-31
20	1059	8	8	0	Circulatory system development	7.79E-06	1.53E-02
21	1093	9	7	2	Mitotic cell cycle	1.19E-07	2.59E-04
22	1127	23	12	11	Mitotic cell cycle	7.81E-33	3.16E-29

Gene count = number of differentially expressed genes associated to each bicluster; sample count = number of samples associated to each bicluster; P value = Fisher exact test. FDR = false-discovery rate; SCC = squamous cell carcinoma; AC = adenocarcinoma.

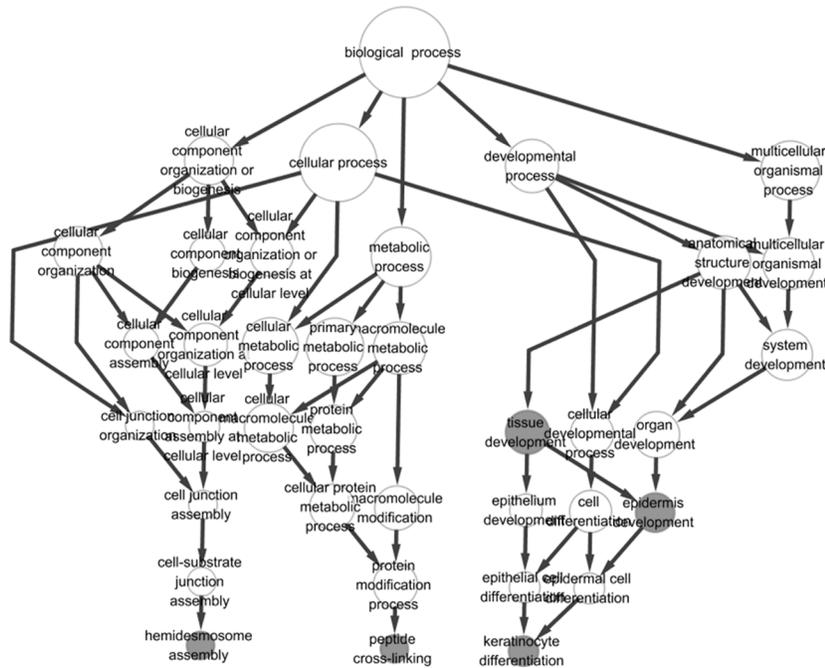


Figure 1. Gene ontology (GO)-enrichment analysis of bicluster 9. The gray nodes represent the significant GO terms (FDR < 0.05). The node size is proportional to the number of genes in the node.

Construction of a regulatory network of DEGs in SCC and AC

Using regulatory data collected from the TRANSFAC and TRED databases, we identified 227 relationships between 59 TFs and their 136 target genes. By integrating these regulatory relationships, a regulatory network was built between TFs and their target genes (Figure 2). In this network, estrogen receptor 1 (ESR1) was an important hub TF that regulates a large number of genes. ESR1 may be important in distinguishing between the 2 subtypes of lung cancer and may also be an important factor in determining the fate of an NSCLC tumor cell. The TFs FOXA2, SP1, and NFIC were also found to regulate several DEGs (marked in red or green) and with higher degrees formed a local network. Furthermore, our network suggests that ESR1 regulates its downstream target genes by regulating E2F1, which activates FOXA2.

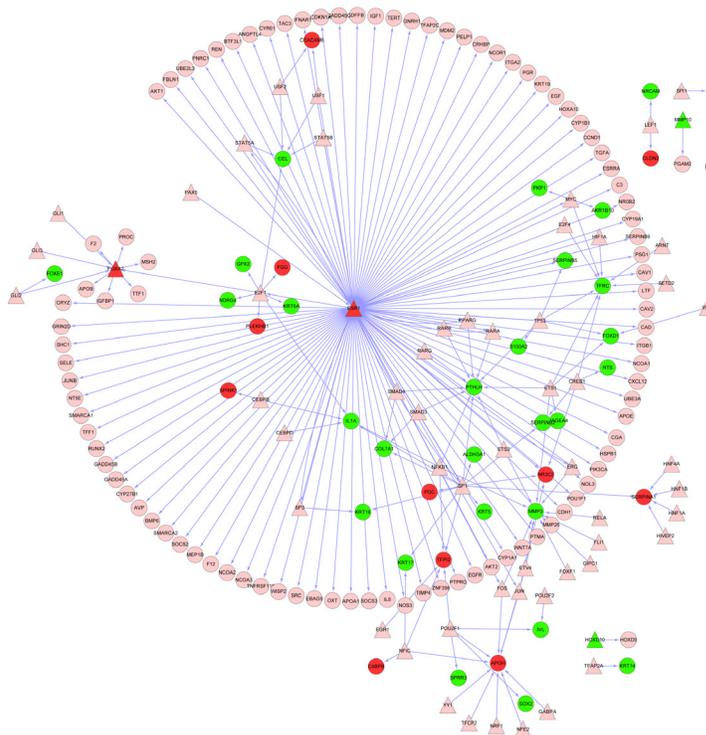


Figure 2. Regulatory network of transcription factors (TFs) and their target genes in squamous cell carcinoma (SCC) and adenocarcinoma (AC). The triangles represent TFs. The red points represent differentially expressed genes (DEGs) upregulated in SCC compared to AC. The green points represent DEGs downregulated in SCC compared to AC. Lines represent the regulatory relationships.

Pathway-enrichment analysis

A total of 189 genes expressed differentially in SCC and AC (fold-change >1.5 times) were further analyzed. We used DAVID to identify the pathways in which the DEGs may be involved. A P value less than 0.01 was chosen as the cut-off for this analysis. We identified 8

pathways in which the DEGs were involved (Table 2), including “tyrosine metabolism”, “tight junction”, and “cell adhesion molecules” among others.

Table 2. Pathway enrichment analysis.

Category	Description	P value	FDR
hsa00350	Tyrosine metabolism	0.014029	13.52789
hsa04530	Tight junction	0.0189183	17.83913
hsa00980	Metabolism of xenobiotics by cytochrome P450	0.03183	28.30744
hsa00982	Drug metabolism	0.0346141	30.40026
hsa00340	Histidine metabolism	0.0439794	37.0418
hsa04610	Complement and coagulation cascades	0.0453189	37.94338
hsa04514	Cell adhesion molecules	0.0660438	50.48606
hsa02010	ABC transporters	0.0918569	62.88899

FDR = false-discovery rate.

DISCUSSION

The recent development of cDNA microarray or cDNA chip technology, a high-throughput method of monitoring gene expression, has made it possible to analyze the expression of thousands of genes at once (Sчена et al., 1995; DeRisi et al., 1996; Ermolaeva et al., 1998; Khan et al., 1998). This genome-wide microarray study of 58 NSCLCs revealed that the 2 NSCLC subtypes SCC and AC have different transcriptional profiles. This conclusion is in line with previous microarray studies, which reported large transcriptomic differences between NSCLC subtypes (Bhattacharjee et al., 2001; Garber et al., 2001; Kuner et al., 2009). In this study, we performed bicluster analysis using ISA to classify the DEGs in SCC and AC into functional subgroups, and constructed a transcriptome network between TFs and their target genes. Furthermore, we used DAVID to identify pathways involving the DEGs.

Before cells undergo DNA replication, they enter the G1 phase of their cell cycle during which they interpret a flood of signals that influence cell division and cell fate. Mistakes in this process can lead to cancer (Massague, 2004). From the results of our GO-enrichment analysis in the 22 biclusters that we identified, we found that most biclusters were associated with the process of “mitotic cell cycle”, suggesting that most of the DEGs in the 2 NSCLC subtypes are involved in the mitotic cell cycle. Therefore, differential expression of genes controlling cell cycle regulation may distinguish AC from SCC. Consistent with this result, a more detailed view of G1 signaling networks, which coordinate cell growth, proliferation, stress management, and survival, will be helpful in defining the causes of malignancies and identifying better cancer therapies (Massague, 2004). In addition, subsets of genes that encode cell junction proteins (*DSG3*, *GBJ5*, *GBJ6*, *PVRL1*, *CLDN1*, *DSC2*, *CLDN3*, *CLDN12*, *CLDN23*, *PARD6B*, *CGN*, and *PARD6G*) were found in these biclusters. Some of these were discovered by a previous study to be upregulated in SCC as compared to AC (Kuner et al., 2009).

Another enriched biological process, as shown in Table 1 was “epidermis development”, suggesting that several DEGs in AC and SCC take part in the epidermis development process. These genes may also have a critical impact in affecting the fate of tumor cells. We also found genes such as *DSG3*, *GJB5*, *CLDN3*, and *CGN*, which encode cell junction proteins, in these biclusters. In tumorigenesis, a loss of junctional complexes has been reported to be involved in the epithelial-mesenchymal transition, an essential event for tumor progression

and metastasis (Thiery and Sleeman, 2006). In contrast, both, the higher and lower abundance of cell junction genes was described in diverse neoplastic tissues (Morin, 2005; Hewitt et al., 2006). Several TFs implicated in lung cancer have been linked in our regulatory network. The genes *ESR1*, *FOXA2*, *SPI1*, and *NFIC* are also hub nodes in our transcriptome network, indicating that these TFs may be useful in distinguishing between these 2 NCSLC subtypes and determining tumor cell fate. Furthermore, our network suggests that *ESR1* regulates its downstream target genes by regulating *E2F1*, which activates *FOXA2*. Some of these TFs have been implicated in lung cancer by previous studies (Stabile and Siegfried, 2004; Basseres et al., 2012; Hsu et al., 2012).

ESR1 is a nuclear receptor that is activated by the sex hormone estrogen. Data suggest that estrogens and growth factors promote tumor progression (Márquez-Garbán et al., 2007). Estrogens may be involved in lung carcinogenesis, and estrogen receptors are present and function in normal lung and tumor cell lines and tissues. Estrogen can directly stimulate the transcription of estrogen-responsive genes in the nucleus of lung cells, and it can transactivate growth factor signaling pathways, in particular the epidermal growth factor pathway (Stabile and Siegfried, 2004).

Sp1, a member of the Sp protein family, is overexpressed in a variety of cancers including gastric, colorectal, pancreatic, epidermal, thyroid, breast, and lung. Previous studies showed that Sp proteins mediate the expression of a number of genes involved in cell proliferation, survival, and angiogenesis (*Sp1* and Krüppel-like factor family of transcription factors in cell growth regulation and cancer), suggesting that targeting transcription factors that activate Sp protein expression can serve as a powerful strategy for the development of anti-cancer agents.

From the results of the pathway-enrichment analysis, we find that SCC and AC show differences in the following 8 pathways: 1) tyrosine metabolism: several proteins that have tyrosine kinase activity are encoded by oncogenes, and oncogenes play an important role in cell growth, proliferation, and differentiation (Arora and Scholar, 2005). The DEGs in this pathway include *HGD*, *ADH7*, *ALDH3B1*, and *ALDH3A1*. 2) Tight junction: the dysfunction of cell junction complexes is known to be important for tumorigenesis, tumor progression, and metastasis formation (Beaudry et al., 2010). Changes in the transcription of cell junction genes can reduce cell-cell connectivity during tumorigenesis. The DEGs in this pathway include *CLDN8*, *RAB3B*, *CGN*, *CLDN3*, *CLDN1*, and *CLDN2*. 3) Metabolism of xenobiotics by cytochrome P450: the DEGs in this pathway include *CYP2S1*, *ADH7*, *ALDH3B1*, and *ALDH3A1*. 4) Drug metabolism: the 2 subtypes of NSCLC are different with respect to this pathway. This may explain the differences between the 2 subtypes in terms of drug metabolism, excretion, and tolerability. The DEGs in this pathway include *FMO5*, *ADH7*, *ALDH3B1*, and *ALDH3A1*. 5) Histidine metabolism: loss of heterozygosity and exon deletions within the fragile histidine triad gene are associated with smoking habits in lung cancer patients. Thus, this gene has been proposed as a target for tobacco smoke carcinogens (Hecht, 1999). Therefore, histidine metabolism may be more important in SCC than in AC. Loss of fragile histidine triad function could initiate malignancy by stimulating the proliferation of cells that are the precursors of digestive tract cancer and lung cancer cells (Sozzi et al., 1996). The DEGs in this pathway include *HAL*, *ALDH3B1*, and *ALDH3A1*. 6) Complement and coagulation cascades: differences in this pathway may explain the differences in the therapeutic response of the 2 subtypes. 7) Cell adhesion molecules (CAMs). CAMs are proteins located on the tumor cell surface that are involved in the interaction of the tumor cell with the extracellular matrix,

vascular endothelial cells, and other tumor cells. CAMs are involved in tumor invasion and metastasis (Cavallaro and Christofori, 2001). The DEGs in this pathway include *CLDN8*, *NRCAM*, *CLDN3*, *CLDN1*, and *CLDN2*. 8) ABC transporters: ATP-binding cassette (ABC) transporters are a family of transporter proteins that contribute to drug resistance via ATP-dependent drug efflux pumps (Leonard et al., 2003). The 2 subtypes show differences in this pathway, indicating that their drug tolerances may be different. The DEGs in this pathway include *ABCC5*, *ABCA13*, and *ABCC6*. These genes are members of the *ABCC* and *ABCA* subfamilies of the ABC superfamily.

In conclusion, our data provide a comprehensive transcriptional profile of candidate genes that may underlie the complex transcriptional/regulatory networks in NSCLC subtypes. We identified several genes and pathways that are differentially expressed in SCC and AC; however, further analysis will be required to unravel their influences on the progression of human lung cancer.

REFERENCES

- Arora A and Scholar EM (2005). Role of tyrosine kinase inhibitors in cancer therapy. *J. Pharmacol. Exp. Ther.* 315: 971-979.
- Basseres DS, D'Alò F, Yeap BY, Löwenberg EC, et al. (2012). Frequent downregulation of the transcription factor Foxa2 in lung cancer through epigenetic silencing. *Lung Cancer* 77: 31-37.
- Beaudry VG, Jiang D, Dusek RL, Park EJ, et al. (2010). Loss of the p53/p63 regulated desmosomal protein perp promotes tumorigenesis. *PLoS Genet.* 6: e1001168.
- Benjamini Y, Drai D, Elmer G, Kafkafi N, et al. (2001). Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* 125: 279-284.
- Bergmann S, Ihmels J and Barkai N (2003). Iterative signature algorithm for the analysis of large-scale gene expression data. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 67: 031902.
- Bhattacharjee A, Richards WG, Staunton J, Li C, et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. U. S. A.* 98: 13790-13795.
- Brivanlou AH and Darnell JE Jr (2002). Signal transduction and the control of gene expression. *Science* 295: 813-818.
- Cavallaro U and Christofori G (2001) Cell adhesion in tumor invasion and metastasis: loss of the glue is not enough. *Biochim. Biophys. Acta* 1552: 39-45.
- Csardi G, Kutalik Z and Bergmann S (2010). Modular analysis of gene expression data with R. *Bioinformatics* 26: 1376-1377.
- DeRisi J, Penland L, Brown PO, Bittner ML, et al. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* 14: 457-460.
- Diboun I, Wernisch L, Orengo CA and Koltzenburg M (2006). Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC Genomics* 7: 252.
- Ermolaeva O, Rastogi M, Pruitt KD, Schuler GD, et al. (1998). Data management and analysis for gene expression arrays. *Nat. Genet.* 20: 19-23.
- Garber ME, Troyanskaya OG, Schluens K, Petersen S, et al. (2001). Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. U. S. A.* 98: 13784-13789.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5: R80.
- Hecht SS (1999). Tobacco smoke carcinogens and lung cancer. *J. Natl. Cancer Inst.* 91: 1194-1210.
- Hewitt KJ, Agarwal R and Morin PJ (2006). The claudin gene family: expression in normal and neoplastic tissues. *BMC Cancer* 6: 186.
- Huang dW, Sherman BT and Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44-57.
- Hsu TI, Wang MC, Chen SY, Yeh YM, et al. (2012). Sp1 expression regulates lung tumor progression. *Oncogene* 31: 3973-3988.
- Ihmels J, Friedlander G, Bergmann S, Sarig O, et al. (2002). Revealing modular organization in the yeast transcriptional network. *Nat. Genet.* 31: 370-377.

- Ihmels J, Bergmann S and Barkai N (2004). Defining transcription modules using large-scale gene expression data. *Bioinformatics* 20: 1993-2003.
- Inamura K, Fujiwara T, Hoshida Y, Isagawa T, et al. (2005). Two subclasses of lung squamous cell carcinoma with different gene expression profiles and prognosis identified by hierarchical clustering and non-negative matrix factorization. *Oncogene* 24: 7105-7113.
- Jiang C, Xuan Z, Zhao F and Zhang MQ (2007). TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res.* 35: D137-D140.
- Kanehisa M (2002). The KEGG database. *Novartis Found. Symp.* 247: 91-101.
- Khan J, Simon R, Bittner M, Chen Y, et al. (1998). Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.* 58: 5009-5013.
- Kuner R, Muley T, Meister M, Ruschhaupt M, et al. (2009). Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. *Lung Cancer* 63: 32-38.
- Leonard GD, Fojo T and Bates SE (2003). The role of ABC transporters in clinical practice. *Oncologist* 8: 411-424.
- Li R, Wang H, Bekele BN, Yin Z, et al. (2006). Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach. *Oncogene* 25: 2628-2635.
- Maere S, Heymans K and Kuiper M (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448-3449.
- Márquez-Garbán DC, Chen HW, Fishbein MC, Goodglick L, et al. (2007). Estrogen receptor signaling pathways in human non-small cell lung cancer. *Steroids* 72: 135-143.
- Massague J (2004). G1 cell-cycle control and cancer. *Nature* 432: 298-306.
- Morin PJ (2005). Claudin proteins in human cancer: promising new targets for diagnosis and therapy. *Cancer Res.* 65: 9603-9606.
- Schena M, Shalon D, Davis RW and Brown PO (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470.
- Sozzi G, Veronese ML, Negrini M, Baffa R, et al. (1996). The FHIT gene 3p14.2 is abnormal in lung cancer. *Cell* 85: 17-26.
- Spies M, Dasu MR, Svrakic N, Nestic O, et al. (2002). Gene expression analysis in burn wounds of rats. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 283: R918-R930.
- Stabile LP and Siegfried JM (2004). Estrogen receptor pathways in lung cancer. *Curr. Oncol. Rep.* 6: 259-267.
- Team RDC (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
- Thiery JP and Sleeman JP (2006). Complex networks orchestrate epithelial-mesenchymal transitions. *Nat. Rev. Mol. Cell Biol.* 7: 131-142.
- Wachi S, Yoneda K and Wu R (2005). Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* 21: 4205-4208.
- Wingender E (2008). The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief. Bioinform.* 9: 326-332.
- Wittchen F, Suckau L, Witt H, Skurk C, et al. (2007). Genomic expression profiling of human inflammatory cardiomyopathy (DCMi) suggests novel therapeutic targets. *J. Mol. Med.* 85: 257-271.