

# Reconstruction of an Integrated Genome-Scale Co-Expression Network Reveals Key Modules Involved in Lung Adenocarcinoma

Gholamreza Bidkhor<sup>1</sup>✉, Zahra Narimani<sup>1</sup>✉, Saman Hosseini Ashtiani<sup>1</sup>, Ali Moeini<sup>2</sup>, Abbas Nowzari-Dalini<sup>3</sup>, Ali Masoudi-Nejad<sup>1\*</sup>

**1** Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran, **2** Department of Algorithms and Computation, College of Engineering, University of Tehran, Tehran, Iran, **3** School of Mathematics and Computer Science, University of Tehran, Tehran, Iran

## Abstract

Our goal of this study was to reconstruct a “genome-scale co-expression network” and find important modules in lung adenocarcinoma so that we could identify the genes involved in lung adenocarcinoma. We integrated gene mutation, GWAS, CGH, array-CGH and SNP array data in order to identify important genes and loci in genome-scale. Afterwards, on the basis of the identified genes a co-expression network was reconstructed from the co-expression data. The reconstructed network was named “genome-scale co-expression network”. As the next step, 23 key modules were disclosed through clustering. In this study a number of genes have been identified for the first time to be implicated in lung adenocarcinoma by analyzing the modules. The genes EGFR, PIK3CA, TAF15, XIAP, VAPB, App1, Rab5a, ARF4, CLPTM1L, SP4, ZNF124, LPP, FOXP1, SOX18, MSX2, NFE2L2, SMARCC1, TRA2B, CBX3, PRPF6, ATP6V1C1, MYBBP1A, MACF1, GRM2, TBXA2R, PRKAR2A, PTK2, PGF and MYO10 are among the genes that belong to modules 1 and 22. All these genes, being implicated in at least one of the phenomena, namely cell survival, proliferation and metastasis, have an over-expression pattern similar to that of EGFR. In few modules, the genes such as CCNA2 (Cyclin A2), CCNB2 (Cyclin B2), CDK1, CDK5, CDC27, CDCA5, CDCA8, ASPM, BUB1, KIF15, KIF2C, NEK2, NUSAP1, PRC1, SMC4, SYCE2, TFDP1, CDC42 and ARHGEP9 are present that play a crucial role in cell cycle progression. In addition to the mentioned genes, there are some other genes (i.e. DLGAP5, BIRC5, PSMD2, Src, TTK, SENP2, PSMD2, DOK2, FUS and etc.) in the modules.

**Citation:** Bidkhor G, Narimani Z, Hosseini Ashtiani S, Moeini A, Nowzari-Dalini A, et al. (2013) Reconstruction of an Integrated Genome-Scale Co-Expression Network Reveals Key Modules Involved in Lung Adenocarcinoma. *PLoS ONE* 8(7): e67552. doi:10.1371/journal.pone.0067552

**Editor:** Robert A. Winn, University of Colorado School of Medicine, United States of America

**Received:** April 6, 2013; **Accepted:** May 18, 2013; **Published:** July 11, 2013

**Copyright:** © 2013 Bidkhor et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** No current external funding sources for this study.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: amasoudin@ibb.ut.ac.ir

✉ These authors contributed equally to this work.

## Introduction

Lung cancer is the major cause of cancer-related deaths in the world and is not easily diagnosed. At most, 15% of patients sustain life for no more than five years [1]. Lung cancer is categorized into two groups i.e. small-cell lung cancer (SCLC) that includes 20% of lung cancers and non small cell lung cancer (NSCLC) that comprises 80% of lung cancers. NSCLC is shown to initiate from lung epithelial cells that leads to various histological subvarieties including adenocarcinoma (ADC), large cell carcinoma (LCC) and squamous cell carcinoma (SCC). The incidence of ADC subtypes has relatively increased in the last few decades [2].

The variation in the human genome is diverse, including alterations in the human karyotype, point mutations, single nucleotide polymorphisms (SNPs) and so forth. The duplication, insertion and deletion variations within the range of one kilo-base to several mega-bases that wildly take place in human and other mammalian genomes are termed copy number variations (CNVs). Although there are powerful repair mechanisms in the human genome, the odds of CNVs are high in such a way that they are 100–10000 times higher than point mutations [3]. Recently, the

role of CNV in various cancers has been thoroughly investigated in many ways [4].

Array based comparative genomic hybridization (array-CGH) is capable of supplying high-resolution identification of CNVs. Single nucleotide polymorphism arrays (SNP array) can quantify cancer-specific loss of heterozygosity (LOH) and CNV with high precision in a genome-wide manner [5]. Through the use of the two approaches, several genomic regions frequently showing DNA gains (3q, 5p, 7q, 8q, 11q and 16p) and losses (3p, 4q, 5q, 6q, 8p 9p and 13q, 17q) have been detected in NSCLC patients. Genetic aberrations in NSCLC are potently associated with tumor histology and a comparison of the two histological subtypes of squamous cell carcinoma and adenocarcinoma has disclosed clear dissimilarities in the recurrence of genomic aberrations [6]. Being influenced by tumor subtype, gender, ethnicity and exposure to carcinogens (e.g. smoking, radon gas, asbestos and cooking oil fumes), built up genomic alterations apply different tumorigenic mechanisms bringing about triggering of oncogenic signaling pathways and unrestrained tumor growth and metastasis [7].

Previous studies on NSCLC, concentrating on oncogenic point mutations, have recognized repetitious mutations that lead to aberrant activation of EGFR, KRAS, PIK3CA, ERBB2, BRAF,

and some other genes. In addition, inactivating point mutations and deletions in TP53, STK11, NF1, CDKN2A and PTEN have been demonstrated. Mutations in several tyrosine kinase genes including PDGFRA and KDR have also been identified. Compared with lung adenocarcinoma, the range of genetic alterations in lung squamous cell carcinoma is less known [8].

Recently, genome-wide association studies (GWAS) have proven that three human genomic regions in chromosomes 5p15, 15q25, and 6p21 are associated with vulnerability to lung cancer in European and American populations. Besides, 3q29 and 18p11.22 are associated with susceptibility to lung cancer in Korean population [9].

The modern advancement of cDNA and oligonucleotide microarray analysis has enabled us to broadly analyze gene expression profiles in NSCLC cells and classify lung cancers at the molecular scale [10]. Gene expression profiling has given us many genes over- or down-expression in different disease states, but again a small number of the genes have been shown to be clearly functionally relevant to the tumorigenic procedure [11].

Our goal was to construct a “genome-scale co-expression network” for lung adenocarcinoma using all available and relevant data. The rationale behind our idea is the fact that simultaneous use of SNP array, gene expression microarray, array-CGH, CGH, GWAS and gene mutation data can give a more comprehensive understanding of the whole genome of the cancerous cells. Inasmuch as the genomic variations are different in various NSCLC subtypes, we focused on lung adenocarcinoma to achieve more precise results. In this study, through the integration of data obtained from the analyses mentioned above, it is feasible to deduce an integrated genome wide perspective of the mutated genes, the gene dosage aberrations and their effect on gene expression. Such achievements may help us in identifying the significant genes in lung adenocarcinoma. Another main aim of our study was to find the key modules in lung adenocarcinoma. Accordingly, through clustering of a “genome-scale co-expression network”, lung adenocarcinoma modules were revealed. Another objective achieved by analyzing the modules was to identify the significant genes implicated in lung adenocarcinoma.

## Materials and Methods

Figure 1 depicts a framework for construction of a “genome-scale co-expression network” in lung adenocarcinoma which includes different integrated data. Gene mutations, GWAS, array-CGH, CGH and SNP array data were used to identify the loci and the genes that exist in lung adenocarcinoma, with high precision. In other words, through integration of the mentioned data, lung adenocarcinoma was examined in genome-scale. Subsequently, gene expression microarray data were used to integrate with other data in order to construct “genome-scale co-expression network”. In the next step, using clustering, the key modules in lung adenocarcinoma were revealed and analyzed.

### Identification of the genes located on CNV segments in lung adenocarcinoma

Datasets for SNP microarray (SNP array) related to adenocarcinoma were obtained from the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo>). The accession numbers were GSE33848 [12] and GSE36363 [13], and there were 216 samples in total. The applied datasets have Affymetrix Genome-Wide Human SNP 6.0 array platform. The genome-wide Human SNP Array 6.0 consists of SNPs and CN probe sets related to two enzyme sets, namely Nsp and Sty.

The information regarding CNV regions was obtained by Affymetrix® Genotyping Console™ software (GTC). A number of SNPs and CN probe sets are merely located on fragments made by one of the enzymes, whereas the rest of SNPs and CN probe sets are located on fragments made by both of the enzymes. Genotyping Console 4.0 (GTC 4.0) harbors in its interface the possibility of choosing between genotyping SNP 6.0 array data with the Birdseed (v1) and the Birdseed (v2) algorithms. Birdseed v2 applies EM to generate a maximum likelihood fit of a 2-dimensional Gaussian mixture model in A\*B space. V1 applies SNP-specific models merely as an initial condition that allows the Expectation-Maximization (EM) fit to wander more freely leading to probable misleading of the clusters. On the other hand, Birdseed v2 uses SNP-specific priors in likelihood as Bayesian priors in addition to initial priors. This is considered as an advantage over v1 because the EM cannot freely wander with such constrains. SNP 6.0 CN/LOH analysis takes the advantage of the BRLMM-P+ algorithm, which is comparable with BRLMM-P, though with a few dissimilar parameters. GTC 4.0 was run with its default parameters. After GTC 4.0 run, only the regions (loci) that had been observed in at least 15% of the cancerous samples, were selected. Eventually, we determined the genes located in the mentioned CNV regions on the basis of NCBI Gene (<http://www.ncbi.nlm.nih.gov/gene>).

Array-CGH and CGH datasets related to adenocarcinoma were obtained from the source (<http://www.cghtmd.jp/CGHDatabase>). The selected cell lines are RERF-LC-MS, ABC-1, RERF-LC-OK, PC-14, HUT-29, SK-LC-3, VMRC-LCD, 11-18 and A549, which are related to lung adenocarcinoma.

### Gene mutation and GWAS data

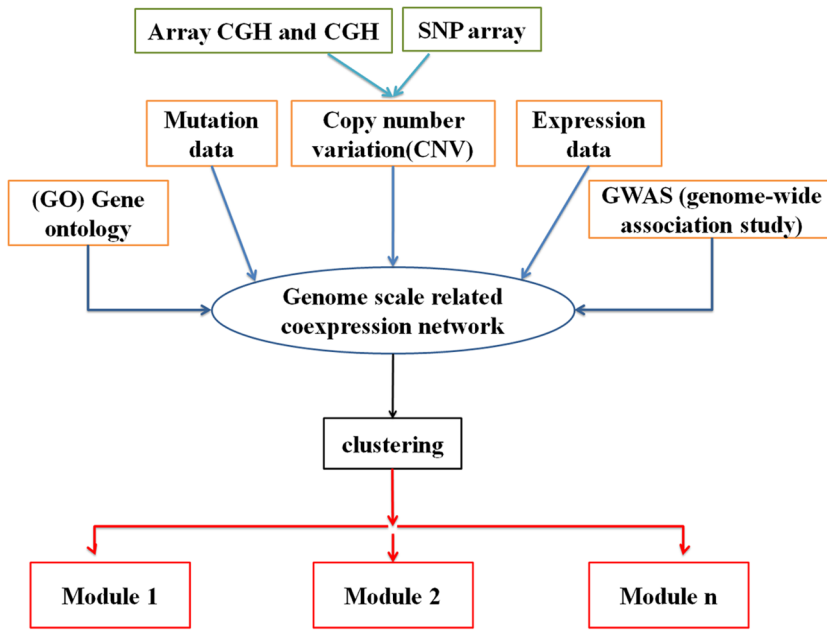
Gene mutation data regarding adenocarcinoma were obtained from HGMD (<http://www.hgmd.cf.ac.uk/ac/indidx.php>), RCGDB (<http://rcgdb.bioinf.uni-sb.de/MutomeWeb/>) and COSMIC (Catalogue of Somatic Mutations in Cancer) ([http://www.sanger.ac.uk/genetics/CGP/cosmic/add\\_info/](http://www.sanger.ac.uk/genetics/CGP/cosmic/add_info/)). GWAS data were gathered from a literature search, from which, some genes and loci were chosen [14,15,16,17].

### Gene expression microarray data

Datasets for Gene expression microarray of adenocarcinoma were obtained from GEO database. For including the most number of genes possible and in order to maximize the number of databases, the Affymetrix Human Genome U133 plus 2.0 Array platform datasets were used to build the co-expression network and various datasets were applied to denote multiple perturbed states in adenocarcinoma. In total, 158 samples were chosen from three datasets i.e. GSE12667 [18], GSE10245 [19] and GSE28571 [6]. GSE28571 and GSE10245 datasets also involve histological subtypes such as large cell carcinoma and squamous cell carcinoma in addition to adenocarcinoma, nonetheless, only adenocarcinoma samples were chosen.

### Reconstruction of “genome-scale co-expression network”

There are many existing methods to reconstruct a biological network from microarray data. The Methods based on machine learning, for example Bayesian network [20,21] and clustering algorithms, or methods based on information theory [22,23,24] are some of methods used in reconstruction of gene regulatory networks. ARACNE [24] is one of the popular statistical algorithms for the reconstruction of accurate cellular networks using microarray expression profiles. ARACNE is also flexible to



**Figure 1. A framework with the purpose of “genome-scale co-expression network” construction and corresponding analysis.**  
doi:10.1371/journal.pone.0067552.g001

work on complex mammalian cell data, and it uses statistical methods to eliminate indirect links between genes. It is, therefore, fast and efficient enough to reconstruct “genome-wide co-expression networks”.

Candidate gene-gene interactions are estimated by pairwise analysis of the expression profile using the mutual information.  $I(g_i, g_j) = I_{ij}$  is an information theoretic measure of relatedness which is zero if  $P(g_i)$  and  $P(g_j)$  are independent variables, i.e.  $P(g_i, g_j) = P(g_i) \cdot P(g_j)$ . Choosing an appropriate threshold for the mutual information can determine which gene expressions can be considered related to each other. The mutual information in ARACNE is computed using formula 1, where  $x_i$  and  $y_i$  represent expression levels and  $P(x_i)$  and  $P(y_i)$  represent the probability that  $X = x_i$ . The mutual information threshold can be imported as an input of ARACNE using a P-value parameter. This factor alone suffers from the problem of considering indirect interactions.

$$MI(X, Y) = \sum_{ij} P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \quad (1)$$

Then, ARACNE removes candidate indirect interactions using Data Processing Inequality (DPI) property, which is a well-known information theoretic measure. If formula 2 holds for the mutual information values between  $g_1$ ,  $g_2$ , and  $g_3$ , then DPI states that genes  $g_1$  and  $g_3$  only have relations because of their relations with the third gene,  $g_2$ , and they do not have a direct relation between themselves.

$$I(g_1, g_3) \leq \min[I(g_1, g_2); I(g_2, g_3)] \quad (2)$$

ARACNE was used for construction of a co-expression network for the expressed data. For construction of a “genome-scale co-expression network”, the co-expression network was constructed according to the genes obtained from gene mutation data, GWAS

data, array-CGH, CGH and SNP-array analysis. These genes (obtained through data integration) were used as a hub at the entrance of ARACNE so that the co-expression network is constructed on the basis of these genes.

#### Clustering of “genome-scale co-expression network”

Topological characteristics of the co-expression network were examined by Cytoscape 2.8.3 [25] and for clustering, ClusterONE [26] and MCODE [27] were used. ClusterONE, a Cytoscape plugin for clustering, was used as the clustering method in this section. This algorithm is fast and can be run in a command-line mode, which does not need to load the large genome wide network in Cytoscape. ClusterONE is designed to find densely connected subgraphs of a network by maximizing edges (weights) within a cluster and minimizing edges (weights) between different clusters. It allows the overlapping of subgraphs (clusters), which are necessary in gene co-expression networks, since a gene may take part in more than one functional module.

MCODE is another method for clustering that was used herein. MCODE is a clustering algorithm, which can be used for directed or undirected graphs. With our undirected co-expression graph, we can summarize MCODE algorithm in three steps: vertex-weighting, complex prediction, and the optional post-processing phase. The vertex-weighting function is defined as the product of the vertex core-clustering coefficient and the highest k-core level of the immediate neighborhood of the vertex. This weighting scheme defines a measure of local density for a vertex’s neighborhood. In the second stage, complexes with high vertex weight are used as seed and the complex neighbor vertices are checked to see if they are a part of this complex or not. This check is done using a weight threshold on the percentage weight vertex, which, is away from the weight of the seed vertex. This process is repeated until no other vertex can be added to this complex. In this way, complexes are detected. In the third phase, a post-processing is done in which some complexes may be removed (if they do not have a minimum degree of 2), and some complexes may enlarge according to a given fluff parameter. MCODE main algorithm – until step 2 –

results in non-overlapping subclusters, which are not suitable for clustering co-expression networks in which a gene may take part in more than one module. The only possible overlap in MCODE can happen in the third phase of the extending complexes.

## Results

### Mutated genes and identified genes by GWAS in lung adenocarcinoma

All 141 mutated genes selected from gene mutation databases were related to lung adenocarcinoma. In addition, 41 genes were chosen from the studies related to GWAS, which were previously published. The number of the genes reached 181 unique genes in total, as listed in the Table S1.

### CNV analysis

Our studies, along with several other ones, have shown that some of the regions including loss and gain in all types of NSCLC are quite different. For this reason, we focused on lung adenocarcinoma to get better results, considering a subtype from NSCLC. CNV analysis was done through the integration of SNP-array, CGH and array-CGH data. Only those loci from gain and loss were chosen which were already present in both. Array-CGH data were obtained from CGMD database with nine lung adenocarcinoma cell lines. Figure 2 depicts the regions having gain (duplication) and loss (deletion) in different cell lines. Gain and loss items were chosen from only those common in at least 40 percent of all cell lines. Gain items were 2p13, 3q25-29, 5p12-15, 5q23, 5q33-35, 7p11-22, 8q22-24, xq12-28, xp11-22 and 4q13, and loss items were 3p12-21, 8p22-23, 9p13, 9p21-23, 9q13, 9q21 and 13q21-34.

SNP-array analyses were performed using Genotyping console 4.0 and Gain and loss items were chosen in such a way that only those common in at least 15 percent of all cell lines were collected. Gain items involved some regions of 1q, 2p11.2, 3q26.1, 5p15.33, some regions of 5q, 6p25.3, 7q21.11, 14q13.2, some regions of 8, 10q11-22, 12p11-12, some regions of 14q, 15q11.2, 16p11, some regions of 17p and 18q, 19q12, some regions of 20, 22q11, some regions of Xq and Yp11.2 and loss items included 8p11.22, 9p11.2, 3q29, 17p11.2, 1q21.1-2, 3p12.3, some regions of 15q and xp22. Figure 3 shows gain and loss regions caused by SNP-array analysis. Then, another filtering was carried out to increase the precision of the loci including gain and loss items, so that finally

those loci were chosen that existed among the output of array-CGH, CGH and SNP-array analyses. These loci, along with their genes downloaded from Entrez Gene, are provided in Table S2. The total number of all these genes is 3652.

### Reconstruction of “genome-scale co-expression network”

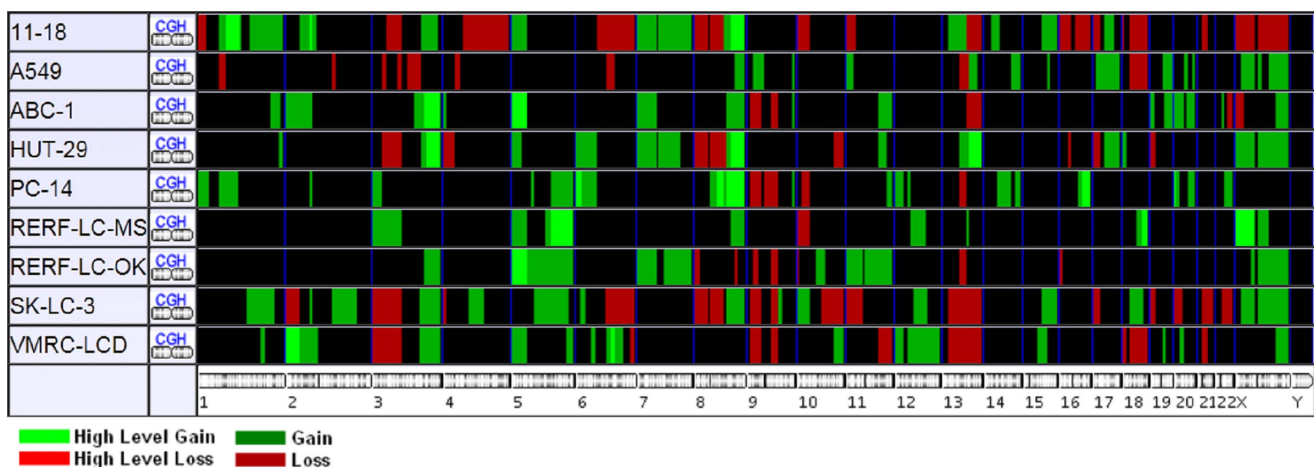
For construction of a “genome-scale co-expression network”, using ARACNE, the co-expression network was constructed according to the genes obtained from gene mutation data, GWAS data, array-CGH and SNP-array analysis in such a way that the genes obtained from gene mutation data and GWAS data (Table S1) were merged with the genes obtained from CNV analysis (Table S2). After the merger of the obtained 3833 genes, no more than 1588 ones included Probe and GO term and we only used those genes, which contained probe and GO term. These 1588 genes (Table S3) were used as hubs at the entrance of ARACNE so that the co-expression network is constructed on the basis of these genes. This co-expression network was named “genome-scale co-expression network”, because the co-expression network was constructed on the basis of the genome-scale data. We used ARACNE with the following parameter settings in Table 1.

Since we used “-s” option (Table 1), the resulting components each has at least one gene from our manually created gene set, and the components not including these genes were removed from the results. This is because we didn’t want to interpret any result that was not included in our gene set, therefore, this helped to reduce the computation needed to get our final co-expression network. ARACNE was implemented with P-values of 0.03 and 0.05, however, it led to no significant difference in the final clustering results and this is why from this point ahead, the results with P-value of 0.05 were employed.

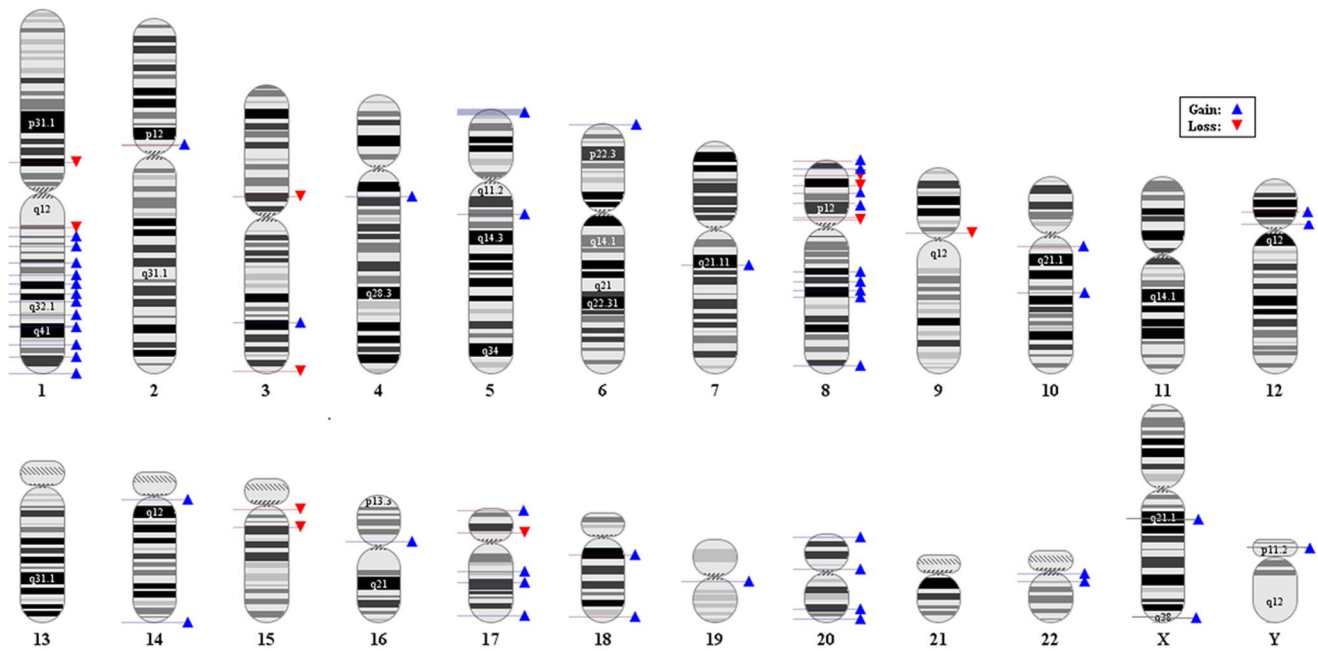
The resulting ARACNE co-expression network on our microarray expression data of 54675 probes has 43058 nodes and 2015975 edges. This large network is comprised of 12 connected components as follows: One component (the largest) with 43031 nodes and 2015959 edges, eight components containing only 2 nodes (therefore 1 edge), two components with 4 nodes and 3 edges, and one component with 3 nodes and two edges.

### Clustering of “genome-scale co-expression network”

To analyze the results, we applied clustering on the “genome-scale co-expression network” generated by ARACNE. Clustering



**Figure 2. Gain and loss comparison obtained from CGH and CGH-array analysis in different lung adenocarcinoma cell lines.**  
doi:10.1371/journal.pone.0067552.g002



**Figure 3. A schema of the loss and gain obtained from SNP-array analysis.**  
doi:10.1371/journal.pone.0067552.g003

algorithms are used to find important sub networks or modules. The results were first obtained from clustering with MCODE but the results were not well-clustered because the clustering was performed on the basis of network topology regardless of the edge weight. Therefore, clusterONE was used which performs clustering on the basis of the edge weight (MI).

We ran ClusterONE with its default parameter settings on the basis of MI. ClusterONE output was 91 clusters or modules including 2237 genes (considering overlapping cluster genes). Some of the clusters turned out to be subclusters of other larger clusters, therefore omitting such subclusters the final number of clusters was reduced from 91 to 23, including 972 genes. Many of the genes are repeated in more than one cluster and overall, there were 450 unique genes out of 23 modules. These clusters are available in Table S4. The co-expression network is made around the axis of the hubs, where all the resulting modules contain at least on hub. For each module, the full data (probe ID, gene full name, synonym, GO...) are provided in Table S5.

**Discussion**

Our results have detected 23 key modules in lung adenocarcinoma, which are accessible in detail in Table S5. Since EGFR signaling plays a key role in NSCLC, we have mainly chosen

modules 1 and 22 (which contain EGFR) as representative modules for our discussion. The merger of the two modules with the most identical genes (named “Merged-module” in this study) is depicted in Figure 4 and Figure 5. Figure 5, illustrates the first neighborhood of an EGFR node in Merged-module whose genes were underlined. It is explained below that many of the identified genes play critical roles in different types of cancer. Additionally, other selected modules will be discussed briefly.

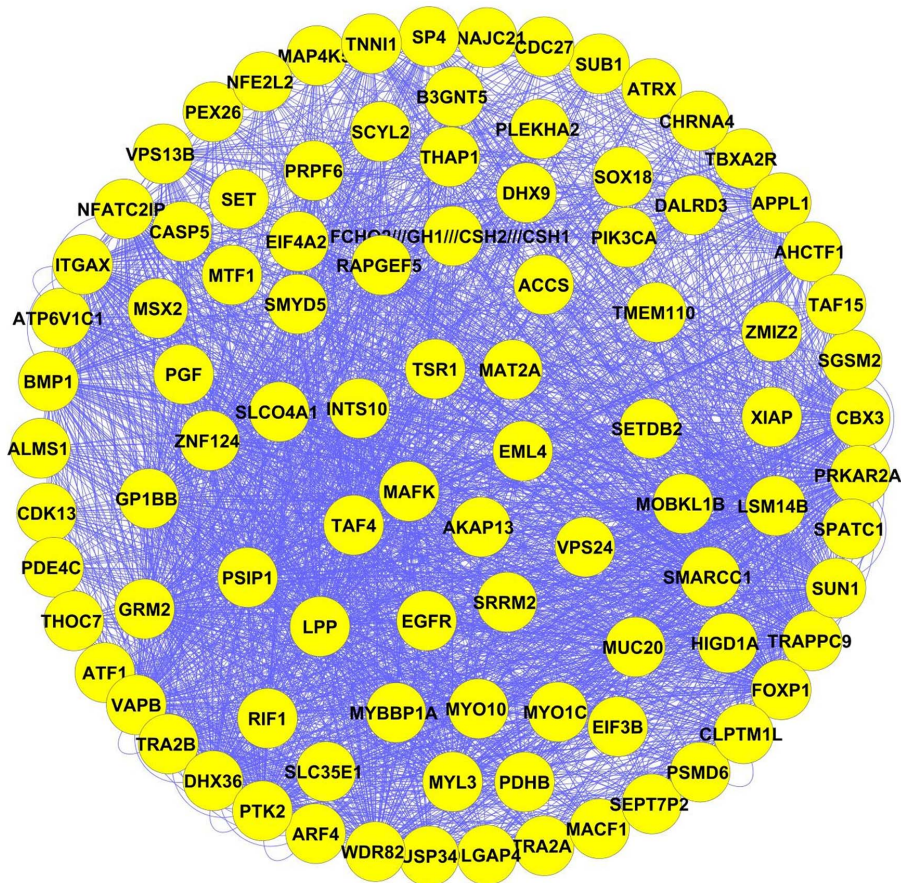
Various studies have shown that EGFR signaling plays a very important role in NSCLCs [28,29,30,31,32]. Since EGFR signaling is crucial for the cell survival and proliferation, it might be the main reason for tumor progression in NSCLC. EGFR signaling activates Ras/ERK, PI3K/Akt and STAT activation pathways. These three pathways are the main routes for the cell proliferation and survival [33,34,35]. Therefore, mutations that lead to excessive activation of these pathways may cause cancer. There are many reports on EGFR over-expression in NSCLC [30]. Hirish et al. [36], Mukohara et al. [31] and Rush et al. showed that the over-expression occurred in 62%, 78% and 98% of NSCLCs, respectively. Hirsh et al. via FISH [36], Gandi et al. via qPCR, CGH and FISH [29] showed a significant correlation between EGFR gene copy number increase with increasing EGFR expression. Knowing that the clustering was performed on the

**Table 1. Parameter settings used for ARACNE.**

Parameter	Value
List of probes for which a subnetwork will be constructed (-s)	1588 probes (Table S13)
P-value for MI threshold (-p)	9.14511467E-7 (0.05/number-of-probes)
Algorithm (accurate fast : -a)	Accurate
Kernel width (-k)	default: determined by program
MI threshold (-t)	0

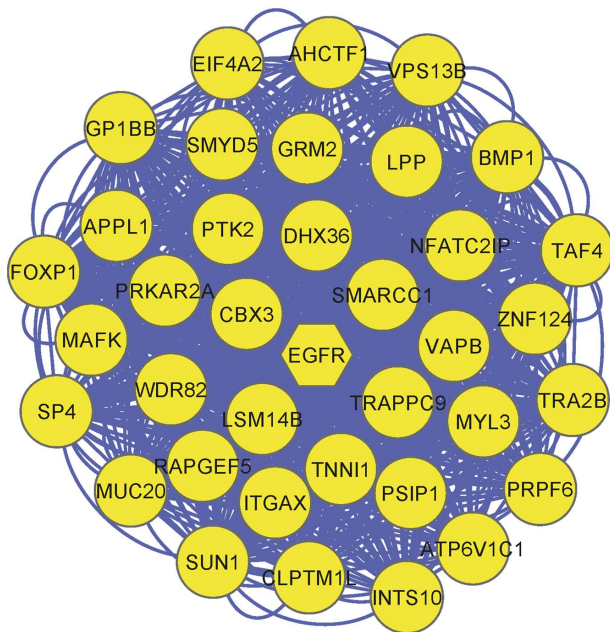
doi:10.1371/journal.pone.0067552.t001





**Figure 4. 95 genes in Merged-module; each node denotes a gene in Merged module and edges depict co-expressed genes relationships.**

doi:10.1371/journal.pone.0067552.g004



**Figure 5. The first neighborhood of EGFR node in Merged-module.**

doi:10.1371/journal.pone.0067552.g005

basis of MI, it is concluded that other genes present in the Merged-module are in accordance with the same expression pattern as that of EGFR. As such, all of these genes show over-expression in lung adenocarcinoma.

The PI3K-Akt pathway is deemed a potential regulator of the cell survival and proliferation. Some of the genes such as PIK3CA, TAF15, VAPB, App1, Rab5a, ARF4, and XIAP in Merged-module activate the PI3K-Akt pathway and are overexpressed in a manner similar to that of EGFR in lung adenocarcinoma. The Phosphoinositide-3-kinase catalytic alpha (PIK3CA) increased activity has been observed in a number of human cancer types such as breast, colon, liver, brain, stomach and lung [37]. TAF15 gene is also upregulated in liposarcoma [38]. XIAP protein over-expression has been identified in six NSCLC cell lines [39] and its inhibition triggers apoptosis in human lung adenocarcinoma A549 cells [40]. A genome-wide microarray analysis demonstrated that VAPB has been frequently overexpressed or amplified in breast cancer. Over-expression of VAPB in MCF10A-HER2 cells increases Akt phosphorylation [41]. App1 is known as an adaptor that takes part in cell signaling through interaction with different signaling molecules involving Akt, PI3-kinase (PI3K), Rab5, adiponectin receptor and TrkA [42,43]. Rab5a being activated by App1 is significantly overexpressed in ovarian cancer and is connected with lung, hepatocellular and stomach carcinomas [44]. Another gene in Merged module is ADP-ribosylation factor 4 (ARF4) that activates EGFR signaling has an anti-apoptotic function in human glioblastoma-derived U373MG cells [45].

Shigematsu et al. [46] and Sordella et al. [47] disclosed the activity of PI3K/Akt pathway in NSCLC. Herein, we have unraveled the over-expression of the above genes, which cause Akt over-activation. PIK3CA, XIAP and Rab5a have previously been reported in lung cancer and we have reported the effect of the rest of the mentioned genes for the first time.

Furthermore, two genes in Merged-module i.e. CLPTM1L and NFE2L2 are implicated in activation of the anti-apoptotic BCL2 family. The gain of CLPTM1L gene frequently occurs in the first stages of NSCLC. CLPTM1L is an overexpressed protein in lung tumor cells. Therefore, it denotes anti-apoptotic CLPTM1L function as a probable mechanism of vulnerability to lung tumorigenesis and resistance against chemotherapy [48]. NFE2L2 is a transcription factor up-regulated in the cell lines of pancreatic cancer, ductal adenocarcinoma [49] and NSCLC [50,51]. We have also revealed the over-expression of these two important genes in lung adenocarcinoma.

In Merged-module, there are some transcription factors as follows: SP4, ZNF124, LPP, FOXP1, SOX18, MSX2 and NFE2L2. The mentioned transcription factors show the same expression pattern as that of EGFR in lung adenocarcinoma. The SP family members such as SP4 are responsible for the regulation of the expression of a few genes such as EGFR. They are connected to cancer cell proliferation, differentiation and metastasis [52,53]. ZNF124 (ZK7) mRNA expression was detected in several human tissues such as a number of leukemia cell lines [54]. With High-resolution array-CGH, Kang et al. [55] and Choi et al. [56] have demonstrated the lipoma preferred partner (LPP) over-expression in NSCLC. FoxP1 (Proteins of the Forkhead-box family) has also been revealed to be expressed in many types of human malignant tumors, while it is along with metastasis. Elevated expression of FoxP1 has been demonstrated in NSCLC [57], hepatocellular carcinoma and breast cancer [58,59]. SOX18, another gene from the above transcription factors, is essential for tumor-induced lymph-angiogenesis and metastasis and a probable target for anti-angiogenic therapy of human cancers [60]. MSX2 is likewise a critical regulator of embryonic development that is assumed to play a role in pancreatic and breast cancer [61]. Regarding the above explanation, these transcription factors, being related to proliferation, survival and angiogenesis, play an important role in lung adenocarcinoma. LPP, FOXP1 and NFE2L2 have previously been reported in lung cancer and we have reported the rest of the genes for the first time in lung adenocarcinoma.

There are some genes in Merged-module that their products take part in the transcription or splicing. These genes include SMARCC1, TRA2B, CBX3 and PRPF6 that show the same upregulation pattern as EGFR in lung adenocarcinoma and we have reported all the mentioned genes for the first time in lung adenocarcinoma. SMARCC1 protein upregulation has been reported in prostate cancer [62], colorectal cancer [63] and cervical intraepithelial neoplasia [62]. TRA2B over-expression was observed in endometrial cancers [64], Gastric cancer cells [65] and cervical cancer [66]. In addition, elevated levels of CBX3 expression in tumor stem cell (TSC)-enriched osteosarcoma cultures was detected [67]. PRPF6 is excessively expressed in the lymph node of lymphoma and is believed to be a probable target for tumor metastasis studies [68]. Regarding the roles of these factors in other cancers and their upregulation in Merged-module, it is possible to infer an important role for them in lung adenocarcinoma.

Four genes i.e. ATP6V1C1, MYBBP1A, MACF1 and MYO10 in Merged-module are connected to metastasis and migration. The increased levels of these gene expressions may be among the

triggering factors of lung adenocarcinoma metastasis. The ATP6V1C1 has been shown to be involved in metastasis and multiple drug resistance. The ATP6V1C1 level is considerably high in oral squamous cell carcinoma [69,70]. MYBBP1A is another gene in Merged module besides being a key regulator in tumor cell proliferation and migration e.g. head and neck squamous cell carcinoma [71]. MACF1 (ACF7) which is brought to the cell membrane via APC (adenomatous polyposis coli) in response to ERBB2 is the key factor for microtubule capture. MYO10 is the myosinX coding gene which is associated with Filopodia formation. This phenomenon has been detected in basal-type breast carcinoma [72]. We have discovered ATP6V1C1, MYBBP1A, MACF1 and MYO10 upregulation in lung adenocarcinoma and accordingly, it is concluded that these genes have an important role in lung adenocarcinoma metastasis and we have reported all the mentioned genes for the first time in lung adenocarcinoma.

Two other important genes i.e. GRM2 and TBXA2R in Merged-module being from the GPCR superfamily, showed over-expression in lung adenocarcinoma. PKA, another important gene found in GPCR signaling showed over-expression as well. Subtypes of GRM2 (mGluR2) are involved in the pathogenesis of diverse cancer types like breast cancer [73], medulloblastomas and gliomas in such a way that GRM2 is overexpressed in all of these cancers [74]. GRM2 keeps the activity of ERK and PI3K pathways. Both pathways are activated in response to EGF [75]. For this reason, we could observe EGFR-like GRM2 over-expression in lung adenocarcinoma. Breast tumor tissues express higher levels of TBXA2R [76]. One of the important factors in GPCR signaling is the regulatory subunit of PKA named PRKAR2A, which is overexpressed in lung adenocarcinoma [77,78]. Elevated expression of Protein kinase A regulatory subunit has been found in primary tumors [79], AML and colorectal cancer [80]. We have reported all the mentioned genes for the first time in lung adenocarcinoma.

PTK2 [81,82], PGF [83,84], SLCO4A1 [85] and Cdc27 [86] are among the genes in Merged-module whose over-expression has been reported in different cancers including lung cancer and we have shown it as well. CDK13, BMP1, RNF13, MAT2A, CHRNA4 are also among the genes in Merged-module whose over-expression has been reported in different cancers, however, their over-expression has been demonstrated in lung adenocarcinoma in this study.

It is commonly accepted that the unleashed proliferation of cancerous cells is contingent on rising protein synthesis and the number of ribosomes. EIF3B that codes one of the EIF3A subunits is one of the Merged-module genes. EIF3B is a crucial part of the EIF3 complex that is implicated in tumor formation [87]. Therefore, it is concluded that EIF3B plays a critical factor in enhancement of protein synthesis in lung adenocarcinoma. In some other modules, ribosomal protein genes named RPS20, RPL8 and RPS4X besides EIF3E, EIF3H, EIF1B, EIF4A2 and EIF2B5 are present. On the basis of these results, the current studies suggest that translation plays a pivotal role in tumor progression.

It is evident that cell cycle regulatory factors play vital roles in different types of cancer. We have observed similar expression patterns in modules 3 and 12 in the genes whose products play important roles in cell cycle regulation. These genes are CCNA2 (Cyclin A2), CCNB2 (Cyclin B2), CDK1, CDK5, CDC27, CDCA5, CDCA8, ASPM, BUB1, KIF15, KIF2C, NEK2, NUSAP1, PRC1, SMC4, SYCE2, TFDP1, CDC42 and ARHGAP9 (CDC42 regulator) that show over-expression in lung adenocarcinoma. These factors are all connected with cell cycle

progression and their over-expression leads to tumor progression in lung adenocarcinoma.

Cyclin A2 (CCNA2) takes control of both S phase and G2/M transition regarding Cdk2 and Cdk1, respectively. In S phase, Cyclin A2 undertakes the initiation and progression regulation of DNA synthesis. Through the G2/M transition, Cyclin A2 has a crucial role in triggering Cyclin B1–Cdk1 activation [88,89]. CDK1 and CDK2 are CDK partners of A- and B- cyclins. A-type cyclins are capable of binding both CDK1 and CDK2, however, B-type cyclins are associated with CDK1 [90]. CDCA5 (Sororin) joins the cohesin complex to regulate the segregation of sister chromatids. Sororin undergoes phosphorylation in mitosis. Sororin is one of the phosphoproteins and protein kinases such as Cdk1/cyclinB and ERK2 regulates its dynamic localization and function [91]. On the basis of what previously mentioned, CDCA5, CDK1, CCNA2 and CCNB2 functions are dependent to each other and we have identified a similar expression pattern in lung adenocarcinoma in a way that all were upregulated together. CDCA8 (Borealin), ASPM, BUB1, KIF15, KIF2C, NEK2, NUSAP1, PRC1, SMC4 and SYCE2 genes are involved in cell division. Since the functions of the mentioned genes are related, their expression patterns turned out to be similar in our study.

CDCA7 expression is controlled by E2F1 and MYC factors that play important roles in cell cycle [130]. CDCA7 is frequently overexpressed in human cancers such as chronic myelogenous leukemia and lung cancers [92]. We have unraveled its over-expression in lung adenocarcinoma. Cdc42 is overexpressed in many of primary lung cancer patients, and Cdc42 over-expression is significantly associated with high TNM stages and lymph node metastasis [93] and its role has recently been proved in lung cancer [94]. ARHGEP9 (Cdc42 guanine nucleotide exchange factor 9) selectively activates Cdc42 [95]. We have proved the over-expression of CDC42 and ARHGEP9 factors in lung adenocarcinoma.

## Conclusion

Various studies have shown that EGFR signaling plays a very important role in NSCLC [28,29,30,31,36]. Since, EGFR signaling is crucial for cell survival and proliferation; it might be the main reason for tumor progression in NSCLC. EGFR signaling activates Ras/ERK, PI3K/Akt and STAT activation pathways. These three pathways are the main routes for cell proliferation and survival [33,34,35,96,97,98]. Four different reports have revealed an increase in PI3K/Akt and pSTAT3 activation pathways in NSCLC with EGFR mutation [46,47,99,100]. Shigematsu et al. [46] and Sordella et al. [47] believe that EGFR mutation is specially exerting effects on PI3K/Akt and STAT3 having a minute effect on ERK activation (Ras/ERK pathway). But Mukohara et al. [31] through wet-lab and

Bidkhorri et al. [32] through *in silico* investigations have shown that in NSCLC, frequency levels of all three pSTAT3, pAkt and pERK are high. Besides, Amann et al. [101] and Vicent et al. [102] have shown that in NSCLC samples with EGFR mutation, pERK level is high. Our results prove that most of the overexpressed genes whose products take part in EGFR signaling pathway can activate PI3K/Akt pathway much more than the other two pathways. Furthermore, most of the mentioned genes products mainly belong to the PI3K/Akt pathway compared to the other two pathways.

In this study, a number of genes have been identified for the first time to be implicated in lung adenocarcinoma. Some of these genes play pivotal roles in other cancer types such that they are considered as therapeutic targets. To manage and contain the investigation, we selected some modules among 23 modules as mentioned above. In addition, there are some genes in other modules that can have very important roles in lung adenocarcinoma, namely DLGAP5, BIRC5, PSMD2, Src, TTK, SENP2, PSMD2, DOK2, FUS among others. We suggest that the genes discussed here can also be used as potential leads for wet lab investigations.

## Supporting Information

**Table S1 A list from 181 genes obtained from gene mutation and GWAS data.**

(TXT)

**Table S2 Loci and genes obtained from CNV analysis.**

(TXT)

**Table S3 Final genes obtained from genome-scale analysis that used at the ARACNE input.**

(TXT)

**Table S4 23 key modules in lung adenocarcinoma.**

(CSV)

**Table S5 Full data for 23 key modules in lung adenocarcinoma.**

(ZIP)

## Acknowledgments

We would like to appreciate the great help of Nadia Karimpour during the course of this study.

## Author Contributions

Conceived and designed the experiments: GB ZN AMN AND. Performed the experiments: GB ZN SHA. Analyzed the data: GB ZN. Contributed reagents/materials/analysis tools: AM AMN AND. Wrote the paper: GB ZN SHA. Checked the algorithm and protocol and controlled the results and edited the manuscript: AND.

## References

- Jemal A, Siegel R, Xu J, Ward E (2010) Cancer statistics, 2010. *CA Cancer J Clin* 60: 277–300.
- Pore MM, Hiltermann TJ, Kruyt FA (2010) Targeting apoptosis pathways in lung cancer. *Cancer Lett*.
- Ahn MJ, Won HH, Lee J, Lee ST, Sun JM, et al. (2012) The 18p11.22 locus is associated with never smoker non-small cell lung cancer susceptibility in Korean populations. *Hum Genet* 131: 365–372.
- Lu TP, Lai LC, Tsai MH, Chen PC, Hsu CP, et al. (2011) Integrated analyses of copy number variations and gene expression in lung adenocarcinoma. *PLoS One* 6: e24829.
- Zhao X, Weir BA, LaFramboise T, Lin M, Beroukhir R, et al. (2005) Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res* 65: 5561–5570.
- Micke P, Edlund K, Holmberg L, Kultima HG, Mansouri L, et al. (2011) Gene copy number aberrations are associated with survival in histologic subgroups of non-small cell lung cancer. *J Thorac Oncol* 6: 1833–1840.
- Kao S, Shiau CK, Gu DL, Ho CM, Su WH, et al. (2012) IGDB.NSCLC: integrated genomic database of non-small cell lung cancer. *Nucleic Acids Res* 40: D972–977.
- Ramos AH, Dutt A, Mermel C, Perner S, Cho J, et al. (2009) Amplification of chromosomal segment 4q12 in non-small cell lung cancer. *Cancer Biol Ther* 8: 2042–2050.
- Janne PA, Li C, Zhao X, Girard L, Chen TH, et al. (2004) High-resolution single-nucleotide polymorphism array and clustering analysis of loss of heterozygosity in human lung cancer cell lines. *Oncogene* 23: 2716–2726.
- Ju Z, Kapoor M, Newton K, Cheon K, Ramaswamy A, et al. (2005) Global detection of molecular changes reveals concurrent alteration of several



- biological pathways in non-small cell lung cancer cells. *Mol Genet Genomics* 274: 141–154.
11. Dong J, Kislinger T, Jurisica I, Wigle DA (2009) Lung cancer: developmental networks gone awry? *Cancer Biol Ther* 8: 312–318.
  12. Broct P, Dalmaso C, Tan EH, Alifano M, Zhang S, et al. (2011) Genomic profiles specific to patient ethnicity in lung adenocarcinoma. *Clin Cancer Res* 17: 3542–3550.
  13. Wilkerson MD, Yin X, Walter V, Zhao N, Cabanski CR, et al. (2012) Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PLoS One* 7: e36530.
  14. Yoon KA, Park JH, Han J, Park S, Lee GK, et al. (2010) A genome-wide association study reveals susceptibility variants for non-small cell lung cancer in the Korean population. *Hum Mol Genet* 19: 4948–4954.
  15. Galvan A, Falvella FS, Spinola M, Frullanti E, Leoni VP, et al. (2008) A polygenic model with common variants may predict lung adenocarcinoma risk in humans. *Int J Cancer* 123: 2327–2330.
  16. Blons H, Pallier K, Le Corre D, Danel C, Tremblay-Gravel M, et al. (2008) Genome wide SNP comparative analysis between EGFR and KRAS mutated NSCLC and characterization of two models of oncogenic cooperation in non-small cell lung carcinoma. *BMC Med Genomics* 1: 25.
  17. Frullanti E, Galvan A, Falvella FS, Manenti G, Colombo F, et al. (2011) Multiple genetic loci modulate lung adenocarcinoma clinical staging. *Clin Cancer Res* 17: 2410–2416.
  18. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, et al. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455: 1069–1075.
  19. Kumer R, Muley T, Meister M, Ruschhaupt M, Bunes A, 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.
  20. Gholami AM, Fellenberg K (2010) Cross-species common regulatory network inference without requirement for prior gene affiliation. *Bioinformatics* 26: 1082–1090.
  21. Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED (2004) Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* 20: 3594–3603.
  22. Altay G, Emmert-Streib F (2011) Structural influence of gene networks on their inference: analysis of C3NET. *Biol Direct* 6: 31.
  23. Olsen C, Meyer PE, Bontempi G (2009) On the impact of entropy estimation on transcriptional regulatory network inference based on mutual information. *EURASIP J Bioinform Syst Biol*: 308959.
  24. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, et al. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7 Suppl 1: S7.
  25. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431–432.
  26. Nepusz T, Yu H, Paccanaro A (2012) Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods* 9: 471–472.
  27. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2.
  28. Yasuda H, Kobayashi S, Costa DB (2011) EGFR exon 20 insertion mutations in non-small-cell lung cancer: preclinical data and clinical implications. *Lancet Oncol*.
  29. Gandhi J, Zhang J, Xie Y, Soh J, Shigematsu H, et al. (2009) Alterations in genes of the EGFR signaling pathway and their relationship to EGFR tyrosine kinase inhibitor sensitivity in lung cancer cell lines. *PLoS One* 4: e4576.
  30. Tsakiridis T, Cutz JC, Singh G, Hirte H, Okawara G, et al. (2008) Association of phosphorylated epidermal growth factor receptor with survival in patients with locally advanced non-small cell lung cancer treated with radiotherapy. *J Thorac Oncol* 3: 716–722.
  31. Mukohara T, Kudoh S, Yamauchi S, Kimura T, Yoshimura N, et al. (2003) Expression of epidermal growth factor receptor (EGFR) and downstream-activated peptides in surgically excised non-small-cell lung cancer (NSCLC). *Lung Cancer* 41: 123–130.
  32. Bidkhori G, Moeini A, Masoudi-Nejad A (2012) Modeling of Tumor Progression in NSCLC and Intrinsic Resistance to TKI in Loss of PTEN Expression. *PLoS One* 7: e48004.
  33. Steelman LS, Chappell WH, Abrams SL, Kempf RC, Long J, et al. (2011) Roles of the Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR pathways in controlling growth and sensitivity to therapy—implications for cancer and aging. *Aging (Albany NY)* 3: 192–222.
  34. Catalanotti F, Reyes G, Jesenberger V, Galabova-Kovacs G, de Matos Simoes R, et al. (2009) A Mek1-Mek2 heterodimer determines the strength and duration of the Erk signal. *Nat Struct Mol Biol* 16: 294–303.
  35. Gupta A, Yang Q, Pandita RK, Hunt CR, Xiang T, et al. (2009) Cell cycle checkpoint defects contribute to genomic instability in PTEN deficient cells independent of DNA DSB repair. *Cell Cycle* 8: 2198–2210.
  36. Hirsch FR, Varella-Garcia M, Bunn PA Jr, Di Maria MV, Veve R, et al. (2003) Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol* 21: 3798–3807.
  37. Chaff JE, Arcila ME, Paik PK, Lau C, Riely GJ, et al. (2012) Coexistence of PIK3CA and other oncogene mutations in lung adenocarcinoma—rationale for comprehensive mutation profiling. *Mol Cancer Ther* 11: 485–491.
  38. Spitzer JJ, Ugras S, Runge S, Decarolis P, Antonescu C, et al. (2011) mRNA and protein levels of FUS, EWSR1, and TAF15 are upregulated in liposarcoma. *Genes Chromosomes Cancer* 50: 338–347.
  39. Dean EJ, Ward T, Pinilla C, Houghten R, Welsh K, et al. (2010) A small molecule inhibitor of XIAP induces apoptosis and synergises with vinorelbine and cisplatin in NSCLC. *Br J Cancer* 102: 97–103.
  40. Liang H, Gu MN, Yang CX, Wang HB, Wen XJ, et al. (2011) Sevoflurane inhibits proliferation, induces apoptosis, and blocks cell cycle progression of lung carcinoma cells. *Asian Pac J Cancer Prev* 12: 3415–3420.
  41. Rao M, Song W, Jiang A, Shyr Y, Lev S, et al. (2012) VAMP-Associated Protein B (VAPB) Promotes Breast Tumor Growth by Modulation of Akt Activity. *PLoS One* 7: e46281.
  42. Tan Y, You H, Coffey FJ, Wiest DL, Testa JR (2010) Appl1 is dispensable for Akt signaling in vivo and mouse T-cell development. *Genesis* 48: 531–539.
  43. Tan Y, You H, Wu C, Altomare DA, Testa JR (2010) Appl1 is dispensable for mouse development, and loss of Appl1 has growth factor-selective effects on Akt signaling in murine embryonic fibroblasts. *J Biol Chem* 285: 6377–6389.
  44. Zhao Z, Liu XF, Wu HC, Zou SB, Wang JY, et al. (2010) Rab5a overexpression promoting ovarian cancer cell proliferation may be associated with APPL1-related epidermal growth factor signaling pathway. *Cancer Sci* 101: 1454–1462.
  45. Jang SY, Jang SW, Ko J (2012) Regulation of ADP-ribosylation factor 4 expression by small leucine zipper protein and involvement in breast cancer cell migration. *Cancer Lett* 314: 185–197.
  46. Shigematsu H, Gazdar AF (2006) Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int J Cancer* 118: 257–262.
  47. Sordella R, Bell DW, Haber DA, Settleman J (2004) Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science* 305: 1163–1167.
  48. James MA, Wen W, Wang Y, Byers LA, Heymach JV, et al. (2012) Functional characterization of CLPTM1L as a lung cancer risk candidate gene in the 5p15.33 locus. *PLoS One* 7: e36116.
  49. Lister A, Nedjadi T, Kitteringham NR, Campbell F, Costello E, et al. (2011) Nrf2 is overexpressed in pancreatic cancer: implications for cell proliferation and therapy. *Mol Cancer* 10: 37.
  50. Sasaki H, Shitara M, Yokota K, Hikosaka Y, Moriyama S, et al. (2012) Increased NRF2 gene (NFE2L2) copy number correlates with mutations in lung squamous cell carcinomas. *Mol Med Report* 6: 391–394.
  51. Hu Y, Ju Y, Lin D, Wang Z, Huang Y, et al. (2012) Mutation of the Nrf2 gene in non-small cell lung cancer. *Mol Biol Rep* 39: 4743–4747.
  52. Sankpal UT, Maliakal P, Bose D, Kayaleh O, Buchholz D, et al. (2012) Expression of specificity protein transcription factors in pancreatic cancer and their association in prognosis and therapy. *Curr Med Chem* 19: 3779–3786.
  53. Colon J, Basha MR, Madero-Visbal R, Konduri S, Baker CH, et al. (2011) Tolfenamic acid decreases c-Met expression through Sp proteins degradation and inhibits lung cancer cells growth and tumor formation in orthotopic mice. *Invest New Drugs* 29: 41–51.
  54. Kuramoto K, Uesaka T, Kimura A, Kobayashi M, Watanabe H, et al. (2000) ZK7, a novel zinc finger gene, is induced by vascular endothelial growth factor and inhibits apoptotic death in hematopoietic cells. *Cancer Res* 60: 425–430.
  55. Kang JU, Koo SH, Kwon KC, Park JW, Kim JM (2009) Identification of novel candidate target genes, including EPHB3, MASP1 and SST at 3q26.2–q29 in squamous cell carcinoma of the lung. *BMC Cancer* 9: 237.
  56. Choi YW, Choi JS, Zheng LT, Lim YJ, Yoon HK, et al. (2007) Comparative genomic hybridization array analysis and real time PCR reveals genomic alterations in squamous cell carcinomas of the lung. *Lung Cancer* 55: 43–51.
  57. Feng J, Zhang X, Zhu H, Wang X, Ni S, et al. (2012) High expression of FoxP1 is associated with improved survival in patients with non-small cell lung cancer. *Am J Clin Pathol* 138: 230–235.
  58. Ijichi N, Shigekawa T, Ikeda K, Horie-Inoue K, Shimizu C, et al. (2012) Association of double-positive FOXA1 and FOXP1 immunoreactivities with favorable prognosis of tamoxifen-treated breast cancer patients. *Horm Cancer* 3: 147–159.
  59. Katoh M, Igarashi M, Fukuda H, Nakagama H (2012) Cancer genetics and genomics of human FOX family genes. *Cancer Lett*.
  60. Young N, Hahn CN, Poh A, Dong C, Wilhelm D, et al. (2006) Effect of disrupted SOX18 transcription factor function on tumor growth, vascularization, and endothelial development. *J Natl Cancer Inst* 98: 1060–1067.
  61. Gremel G, Ryan D, Rafferty M, Lanigan F, Hegarty S, et al. (2011) Functional and prognostic relevance of the homeobox protein MSX2 in malignant melanoma. *Br J Cancer* 105: 565–574.
  62. Heeboll S, Borre M, Ottosen PD, Andersen CL, Mansilla F, et al. (2008) SMARCC1 expression is upregulated in prostate cancer and positively correlated with tumour recurrence and dedifferentiation. *Histol Histopathol* 23: 1069–1076.
  63. Andersen CL, Christensen LL, Thorsen K, Schepeler T, Sorensen FB, et al. (2009) Dysregulation of the transcription factors SOX4, CBFB and SMARCC1 correlates with outcome of colorectal cancer. *Br J Cancer* 100: 511–523.
  64. Ouyang YQ, zur Hausen A, Orlowska-Volk M, Jager M, Bettendorf H, et al. (2011) Expression levels of hnRNP G and hTra2-beta1 correlate with opposite outcomes in endometrial cancer biology. *Int J Cancer* 128: 2010–2019.

65. Takeo K, Kawai T, Nishida K, Masuda K, Teshima-Kondo S, et al. (2009) Oxidative stress-induced alternative splicing of transformer 2beta (SFRS10) and CD44 pre-mRNAs in gastric epithelial cells. *Am J Physiol Cell Physiol* 297: C330–338.
66. Gabriel B, Zur Hausen A, Bouda J, Boudova L, Koprivova M, et al. (2009) Significance of nuclear hTra2-beta1 expression in cervical cancer. *Acta Obstet Gynecol Scand* 88: 216–221.
67. Saini V, Hose CD, Monks A, Nagashima K, Han B, et al. (2012) Identification of CBX3 and ABCA5 as putative biomarkers for tumor stem cells in osteosarcoma. *PLoS One* 7: e41401.
68. Su L, Chen D, Zhang J, Li X, Pan G, et al. (2008) The expression and bioinformatic analysis of a novel gene C20orf14 associated with lymphoma. *J Huazhong Univ Sci Technol Med Sci* 28: 97–101.
69. Garcia-Garcia A, Perez-Sayans Garcia M, Rodriguez MJ, Antunez-Lopez J, Barros-Angueira F, et al. (2012) Immunohistochemical localization of C1 subunit of V-ATPase (ATPase C1) in oral squamous cell cancer and normal oral mucosa. *Biotech Histochem* 87: 133–139.
70. Perez-Sayans M, Reboiras-Lopez MD, Somoza-Martin JM, Barros-Angueira F, Diz PG, et al. (2010) Measurement of ATP6V1C1 expression in brush cytology samples as a diagnostic and prognostic marker in oral squamous cell carcinoma. *Cancer Biol Ther* 9: 1057–1064.
71. Acuna Sanhuesa GA, Faller L, George B, Koffler J, Misetic V, et al. (2012) Opposing function of MYBBP1A in proliferation and migration of head and neck squamous cell carcinoma cells. *BMC Cancer* 12: 72.
72. Arjonen A, Kaukonen R, Ivaska J (2011) Filopodia and adhesion in cancer cell motility. *Cell Adh Migr* 5: 421–430.
73. Speyer CL, Smith JS, Banda M, DeVries JA, Mekani T, et al. (2012) Metabotropic glutamate receptor-1: a potential therapeutic target for the treatment of breast cancer. *Breast Cancer Res Treat* 132: 565–573.
74. Teh J, Chen S (2012) mGlu Receptors and Cancerous Growth. *Wiley Interdiscip Rev Membr Transp Signal* 1: 211–220.
75. Arcella A, Carpinelli G, Battaglia G, D'Onofrio M, Santoro F, et al. (2005) Pharmacological blockade of group II metabotropic glutamate receptors reduces the growth of glioma cells in vivo. *Neuro Oncol* 7: 236–245.
76. Watkins G, Douglas-Jones A, Mansel RE, Jiang WG (2005) Expression of thromboxane synthase, TBXAS1 and the thromboxane A2 receptor, TBXA2R, in human breast cancer. *Int Semin Surg Oncol* 2: 23.
77. Skarpen E, Thoresen GH, Tasken K, Samuelsen JT, Jahnsen T, et al. (1998) Localization of cAMP-dependent signal transducers in early rat liver carcinogenesis. *Histochem Cell Biol* 109: 203–209.
78. Vincent-Dejean C, Cazabat L, Groussin L, Perlemoine K, Fumey G, et al. (2008) Identification of a clinically homogenous subgroup of benign cortisol-secreting adrenocortical tumors characterized by alterations of the protein kinase A (PKA) subunits and high PKA activity. *Eur J Endocrinol* 158: 829–839.
79. Cho-Chung YS, Nesterova M, Pepe S, Lee GR, Noguchi K, et al. (1999) Antisense DNA-targeting protein kinase A-RIA subunit: a novel approach to cancer treatment. *Front Biosci* 4: D898–907.
80. Bradbury AW, Carter DC, Miller WR, Cho-Chung YS, Clair T (1994) Protein kinase A (PK-A) regulatory subunit expression in colorectal cancer and related mucosa. *Br J Cancer* 69: 738–742.
81. Li X, Wang X, Ye H, Peng A, Chen L (2012) Barbigerone, an isoflavone, inhibits tumor angiogenesis and human non-small-cell lung cancer xenografts growth through VEGFR2 signaling pathways. *Cancer Chemother Pharmacol* 70: 425–437.
82. Siejka A, Barabutis N, Schally AV (2012) GHRH antagonist inhibits focal adhesion kinase (FAK) and decreases expression of vascular endothelial growth factor (VEGF) in human lung cancer cells in vitro. *Peptides* 37: 63–68.
83. Zhang L, Chen J, Ke Y, Mansel RE, Jiang WG (2005) Expression of Placenta growth factor (PlGF) in non-small cell lung cancer (NSCLC) and the clinical and prognostic significance. *World J Surg Oncol* 3: 68.
84. Woo IS, Park MJ, Byun JH, Hong YS, Lee KS, et al. (2004) Expression of placental growth factor gene in lung cancer. *Tumour Biol* 25: 1–6.
85. Watanabe T, Miura T, Degawa Y, Fujita Y, Inoue M, et al. (2010) Comparison of lung cancer cell lines representing four histopathological subtypes with gene expression profiling using quantitative real-time PCR. *Cancer Cell Int* 10: 2.
86. Whyte L, Huang YY, Torres K, Mehta RG (2007) Molecular mechanisms of resveratrol action in lung cancer cells using dual protein and microarray analyses. *Cancer Res* 67: 12007–12017.
87. Wang Z, Chen J, Sun J, Cui Z, Wu H (2012) RNA interference-mediated silencing of eukaryotic translation initiation factor 3, subunit B (EIF3B) gene expression inhibits proliferation of colon cancer cells. *World J Surg Oncol* 10: 119.
88. Arsic N, Bendris N, Peter M, Begon-Pescia C, Rebouissou C, et al. (2012) A novel function for Cyclin A2: control of cell invasion via RhoA signaling. *J Cell Biol* 196: 147–162.
89. Mo ML, Chen Z, Li J, Li HL, Sheng Q, et al. (2010) Use of serum circulating CCNB2 in cancer surveillance. *Int J Biol Markers* 25: 236–242.
90. Gong D, Ferrell JE Jr (2010) The roles of cyclin A2, B1, and B2 in early and late mitotic events. *Mol Biol Cell* 21: 3149–3161.
91. Zhang N, Pati D (2012) Sororin is a master regulator of sister chromatid cohesion and separation. *Cell Cycle* 11: 2073–2083.
92. Osthus RC, Karim B, Prescott JE, Smith BD, McDevitt M, et al. (2005) The Myc target gene JPO1/CDCA7 is frequently overexpressed in human tumors and has limited transforming activity in vivo. *Cancer Res* 65: 5620–5627.
93. Chen QY, Jiao DM, Yao QH, Yan J, Song J, et al. (2012) Expression analysis of Cdc42 in lung cancer and modulation of its expression by curcumin in lung cancer cell lines. *Int J Oncol* 40: 1561–1568.
94. Zhu X, Shen H, Li H, Long L, Hui L, et al. (2012) miR-137 inhibits the proliferation of lung cancer cells by targeting Cdc42 and Cdk6. *FEBS Lett*.
95. Tyagarajan SK, Ghosh H, Harvey K, Fritschy JM (2011) Collybistin splice variants differentially interact with gephyrin and Cdc42 to regulate gephyrin clustering at GABAergic synapses. *J Cell Sci* 124: 2786–2796.
96. Hynes NE, Lane HA (2005) ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* 5: 341–354.
97. Silva A, Yunes JA, Cardoso BA, Martins LR, Jotta PY, et al. (2008) PTEN posttranslational inactivation and hyperactivation of the PI3K/Akt pathway sustain primary T cell leukemia viability. *J Clin Invest* 118: 3762–3774.
98. Mertens C, Darnell JE Jr (2007) SnapShot: JAK-STAT signaling. *Cell* 131: 612.
99. Emery IF, Battelli C, Auclair PL, Carrier K, Hayes DM (2009) Response to gefitinib and erlotinib in Non-small cell lung cancer: a retrospective study. *BMC Cancer* 9: 333.
100. Ganti AK (2010) Epidermal growth factor receptor signaling in nonsmall cell lung cancer. *Cancer Invest* 28: 515–525.
101. Amann J, Kalyankrishna S, Massion PP, Ohm JE, Girard L, et al. (2005) Aberrant epidermal growth factor receptor signaling and enhanced sensitivity to EGFR inhibitors in lung cancer. *Cancer Res* 2005/01/25 ed. 226–235.
102. Vicent S, Lopez-Picazo JM, Toledo G, Lozano MD, Torre W, et al. (2004) ERK1/2 is activated in non-small-cell lung cancer and associated with advanced tumours. *Br J Cancer* 90: 1047–1052.