

## Research Article

# Drug Repositioning Discovery for Early- and Late-Stage Non-Small-Cell Lung Cancer

Chien-Hung Huang,<sup>1</sup> Peter Mu-Hsin Chang,<sup>2</sup> Yong-Jie Lin,<sup>1</sup> Cheng-Hsu Wang,<sup>3</sup>  
Chi-Ying F. Huang,<sup>4,5</sup> and Ka-Lok Ng<sup>6,7</sup>

<sup>1</sup> Department of Computer Science and Information Engineering, National Formosa University, 64 Wen-Hwa Road, Hu-Wei, Yun-Lin 632, Taiwan

<sup>2</sup> Division of Hematology and Oncology, Department of Medicine, Taipei Veterans General Hospital, Faculty of Medicine, National Yang-Ming University, Taipei 112, Taiwan

<sup>3</sup> Cancer Center, Keelung Chang Gang Memorial Hospital, Keelung 204, Taiwan

<sup>4</sup> Institute of Biopharmaceutical Sciences, National Yang-Ming University, No. 155, Section 2, Linong Street, Taipei 112, Taiwan

<sup>5</sup> Genome Research Center, National Yang-Ming University, Taipei 112, Taiwan

<sup>6</sup> Department of Biomedical Informatics, Asia University, 500 Lioufeng Road, Wufeng Shiang, Taichung 41354, Taiwan

<sup>7</sup> Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 40402, Taiwan

Correspondence should be addressed to Chi-Ying F. Huang; [bpshuang@gmail.com](mailto:bpshuang@gmail.com) and Ka-Lok Ng; [ppiddi@gmail.com](mailto:ppiddi@gmail.com)

Received 10 April 2014; Revised 7 July 2014; Accepted 12 July 2014; Published 18 August 2014

Academic Editor: X. Li

Copyright © 2014 Chien-Hung Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Drug repositioning is a popular approach in the pharmaceutical industry for identifying potential new uses for existing drugs and accelerating the development time. Non-small-cell lung cancer (NSCLC) is one of the leading causes of death worldwide. To reduce the biological heterogeneity effects among different individuals, both normal and cancer tissues were taken from the same patient, hence allowing pairwise testing. By comparing early- and late-stage cancer patients, we can identify stage-specific NSCLC genes. Differentially expressed genes are clustered separately to form up- and downregulated communities that are used as queries to perform enrichment analysis. The results suggest that pathways for early- and late-stage cancers are different. Sets of up- and downregulated genes were submitted to the cMap web resource to identify potential drugs. To achieve high confidence drug prediction, multiple microarray experimental results were merged by performing meta-analysis. The results of a few drug findings are supported by MTT assay or clonogenic assay data. In conclusion, we have been able to assess the potential existing drugs to identify novel anticancer drugs, which may be helpful in drug repositioning discovery for NSCLC.

## 1. Introduction

Lung cancer is the leading cause of death worldwide [1, 2]. According to medical classification, lung cancer can be divided into two major classes: small cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC). NSCLC accounts for more than 85% of all lung cancer cases, and adenocarcinoma is the most common subtype. The question of how to search for suitable potential drugs for NSCLC is an important issue in biomedical research. However, the process of new drug development is cost-intensive and time-consuming.

A previous study [3] established a systematic strategy to identify potential drugs and target genes for lung cancer. The findings from this study suggested that eight drugs from DrugBank and three drugs from NCBI could potentially reverse the expression of certain up- and downregulated genes. These results are supported by IC50 experimental data. However, the previous study can be extended in several aspects that were addressed in the present study.

Cancer is a multistage progression process that results from genetic sequences mutations, where early- and late-stage cancer-associated genes (CAG) are potentially very different. Therefore, the aim of this paper is to explore

TABLE 1: Summary of microarray datasets.

GEO ID	Organization name	Number of samples (early-stage)	Number of samples (late-stage)
GSE7670	Taipei Veterans General Hospital	8	11
GSE10072	National Cancer Institute, NIH	15	9
GSE19804	National Taiwan University	35	13
GSE27262	National Yang-Ming University	25	n/a

a strategy to identify stage-specific potential drugs for NSCLC through an integrated analysis of microarray profiling. In order to reduce the effect of biological heterogeneity among different individuals, normal as well as cancer tissues were taken from the same patient.

To address the target drug problem, there is a need to address the following issues. First, there is concern that different individuals may correspond to different sets of differentially expressed genes. Second, it is known that cancer is a heterogeneous disease; different stages of cancer correspond to different drug targets involving stage-specific CAG. Third, results derived from different microarray profiling vary from study to study; therefore, a rigorous approach is needed to address this problem. Fourth, reliability of drug finding prediction remains to be verified.

In order to reduce the biological heterogeneity effect among different individuals, tumor/adjacent nontumor pairwise arrays for NSCLC were employed in the present study, thus allowing pairwise statistical tests. To deal with the second issue, the samples were divided into early-stage and late-stage ones, which are denoted as stage IA/IB and stage III/IV, respectively. For the third issue, meta-analysis was adopted to integrate multiple microarray profiles. Finally, potential drug predictions were validated via biochemical assays.

Many proteins are associated with human diseases, although very often their precise functional role in disease pathogenesis remains unclear. A strategy to gain a better understanding into the interaction and function of these proteins is to make use of the protein-protein interaction (PPI) data and construct a network for disease-associated proteins. In our previous work [3, 4], it was hypothesized that the PPI networks, derived from differentially expressed genes (DEGs), could be analyzed topologically to prioritize potential drug targets.

We performed gene set enrichment analysis (GSEA) for pathway analysis and then made use of drug-gene interaction databases and the Connectivity Map (cMap) to find potential drugs for the treatment of NSCLC. It is conjectured that a small drug molecule may potentially reverse the disease signature if the molecule-induced signature is significantly negatively correlated with the disease-induced signature found in the cMap [4]. In fact, potential new treatments for cancers have been successfully identified via the cMap, including acute leukemia, colon cancer, hepatocellular carcinoma, neuroblastoma, NSCLC, and renal cell carcinoma [5–7]. Both up- and downexpressed genes are potential therapeutic targets; therefore, identification of potential drugs to treat lung cancer by using an *in silico* screening approach followed by MTT assay or clonogenic assay validation might accelerate drug discovery.

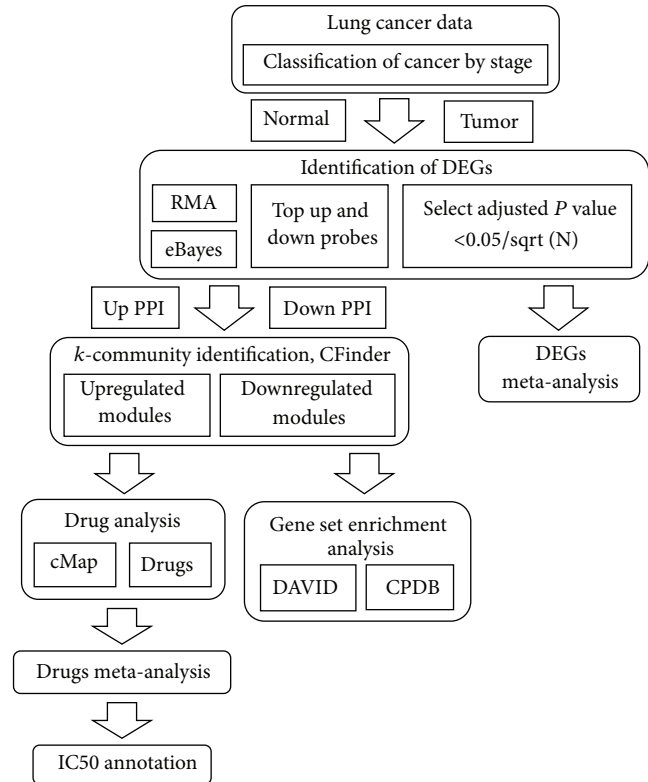


FIGURE 1: Workflow of this study.

In Section 2, we give a description of the input data and the methods used in this paper. In Section 3, results for cluster analysis, enriched pathways, and cMap drug predictions are reported. We conclude in the final section.

## 2. Methods

This study proposes an *in silico* strategy to narrow down the search for lung cancer genes for target identification and drug discovery; the workflow of this study is shown in Figure 1.

**2.1. Input Data Set.** The microarray data for lung cancer was downloaded from GEO [8] and summarized in Table 1. Experiments GSE7670 [9] and GSE10072 [10] use the HG-U133A array, where GSE19804 [11] and GSE27262 [12] use HG-U133 plus 2.0 chip.

Each sample consisted of cancerous and noncancerous lung tissues obtained from a cohort of patients. To infer differentially expressed genes (DEGs), two pair tests (normal as well as cancer tissues are taken from the same patient) were conducted. The main advantage of using paired samples

is that it could reduce the biological heterogeneity effect. In the late stages of cancer, it is very common to find cell invasion, metastasis, and drug-resistance related genes [13]. To investigate this issue, we divided the samples into early- and late-stage ones. Early-stage samples were taken from patients with stage I, IA, and IB cancers, whereas late-stage data were obtained from stage III and IV patients.

**2.2. Microarray Data Analysis.** Microarray technology allows for high-throughput screening and analysis of tens of thousands of genes at the same time. Some genes are activated or inhibited, and some are DEGs, which due to certain regulatory factors, result in changes in gene expression levels by a few times, ten times, or more. Given sets of microarray data, one can identify DEGs among a large number of gene expressions and understand the mechanism of lung cancer formation induced by these DEGs.

There are many microarray data analysis methods, such as using the concept of false discovery rate (FDR) to screen for significant genes [14], using ANOVA to explore the impact of microarray gene expression values within a single factor [15], and clustering analysis. Among the many statistical methods, significance analysis of microarray (SAM) [16, 17], empirical Bayes analysis of microarrays (EBAM) [18], and empirical Bayes statistics (eBayes) [19] are three commonly employed approaches to screen DEGs. The publicly available microarray data analysis package *Bioconductor* [20, 21] was adopted to perform such calculations.

The statistical method eBayes was chosen in this study because it was found that eBayes, SAM, and EBAM achieve a similar level of cancer gene prediction accuracy [22]. The selected DEGs were divided into two groups, an upregulated group (up probes in Figure 1) and a downregulated group (down probes), according to the gene expression fold change (FC) values.

Among the DEGs, genes were classified as either up- or downregulated genes if the  $\log_2FC$  was less than or greater than zero, respectively. Any gene expression level with fold change less than 5.64 ( $\log_2 50$ ) was reset to 5.64 in order to facilitate the cMap search.

**2.3. Cluster Analysis.** We adopted BioGrid version 3.2.101 in our analysis, which consists of 209,838 PPI records. In a PPI network, a densely connected area is referred to as a cluster, which is a functional module. Nodes having a high degree of connection are defined as hubs and are more likely to be essential. Members of a cluster are usually involved in similar biological processes, and protein complexes can be identified through the clustering of a network [23, 24]. It is suggested that a protein complex is a biologically functional module composed of subunits performing similar functions [25]. Given two proteins, *A* and *B*, with a PPI, if both *A* and *B* are obtained from the eBayes prediction as upregulated, then the PPI among *A* and *B* is the so-called up PPI. Communities constructed from up PPI are called upregulated communities.

To investigate the functional modules in which potential lung cancer related proteins are involved, a set of highly confident human PPIs were input to the CFinder software [26] to perform an analysis based on the clique percolation

clustering approach [27]. A *k*-community was set with *k* being equal to three (complete subgraphs of size *k*). Any two *k*-communities are adjacent if they share *k*−1 common nodes. A *k*-community (*k* ≥ 4) is constructed by merging all possible adjacent (*k* − 1)-communities.

**2.4. Gene Set Enrichment Analysis (GSEA).** DAVID [28] is a web-based resource which provides batch annotation and GO [29] term enrichment analysis to highlight the most relevant GO terms associated with a given gene list. The ConsensusPathDB (CPDB) [30] tool provides gene set analysis and metabolite set analysis. The DAVID tool is based on the Fisher exact test, while the CPDB tool is based on the Wilcoxon test. To find the enriched pathways of our lung cancer gene signature, we performed an overrepresentation pathway analysis using both DAVID and CPDB. Under the threshold of a *P* value of less than 0.005, enriched pathways from the overrepresentation analysis including up- and downregulated *k*-communities were obtained from CFinder analysis. Significant pathway results were ranked according to the *P* value. Thus, enriched GO terms for these two protein groups were obtained. We used both tools in this stage for cross-verification.

**2.5. Potential Target Genes and Drug Discovery.** Both of the up- and downregulated communities are derived from the CFinder tool and were used to query the cMap database, where potential drugs with *P* values of less than 0.05 are retained. To identify target genes, the FDA-approved drugs and the chemical-protein links data from STITCH [31] were merged. The Gene Name Service was then used to translate the protein ID to its corresponding HUGO-approved gene symbol and Entrez gene ID. Drugs obtained from the cMap output were mapped and finally identified with known drug targets in the cancer up- or downregulated PPI network.

The idea of drug repositioning is a recently developed approach in the pharmaceutical industry that endeavors to identify new uses for existing drugs and has achieved certain successes [32]. Furthermore, this approach has the potential to accelerate the development time for drugs, as well as reducing side effects. There are many works on identifying repositioned drugs, which are based on various methods: the graph-based inference method [33, 34], the microarray expression method [35], the differential expressed correlation method [36], and the integration of phenotypic, chemical indexes and PPI method [37], and using the drug-gene-disease relationship [38]. We also note that CancerResource [39] is a very comprehensive resource for drug repositioning study.

Several issues arise from combining different datasets, such as the problem of data heterogeneity, different sample sizes, and the data dependence problem. In principle, these issues can be tackled by employing a meta-analysis approach. Meta-analysis (MA) [40, 41] is a set of statistical methods for summarizing the results of several studies into a single estimate. The strength of MA is that it is capable of identifying relationships across a number of different studies.

For the drug prediction study, cMap provides an enrichment score,  $\rho$ , and a *P* value to quantify each cMap drug.

The  $\rho$  value lies between  $-1$  and  $1$ ; therefore it can be treated as a sample correlation coefficient and serve as an effect size index for MA [41]. In practice,  $\rho$  is transformed to the Fisher  $z$  scale, and all the analyses are conducted using the converted values. After the analyses are completed, the  $z$  values are transformed back to the original metric. The transformation to Fisher's  $z$  is given by

$$z = \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho} \quad (1)$$

and the variance of  $z$  is defined by  $V_z = 1/(N - 3)$ , where  $N$  denotes the sample size.

The weight assigned to each study in a fixed-effect model is given by

$$W_i = \frac{1}{V_{Y_i}}, \quad (2)$$

where  $W_i$  is the within-study variance for study  $i$ . The weighted mean ( $M$ ) is computed as

$$M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i}. \quad (3)$$

For unweighted calculations,  $W_i$  equals one. The variance of the summary effect ( $V_M$ ) is given by

$$V_M = \left( \sum_{i=1}^k W_i \right)^{-1}. \quad (4)$$

For unweighted calculations, the Z-score for normal distribution is defined by

$$Z = \frac{M}{SE_M}, \quad (5)$$

where  $SE_M$  denotes the standard error and is equal to  $\sqrt{V_M}$ .

For weighted calculations, the Z-score is defined by

$$Z = \frac{\sum_{i=1}^k W_i Y_i}{\sqrt{\sum_{i=1}^k W_i^2}}. \quad (6)$$

From (7), one can determine the one-tailed test  $P$  value.

The 95% lower and upper limits for the summary effect would be computed as

$$\begin{aligned} LL_M &= M - 1.96 \times SE_M, \\ UL_M &= M + 1.96 \times SE_M. \end{aligned} \quad (7)$$

The formula for the random-effects model is given in a monograph written by Borenstein et al. [41]. The above analyses allow us to determine the confidence interval of the CC,  $r$ .

Besides the use of  $\rho$ , the use of the Fisher combined test (FCT) [40] is another option. The Fisher summary statistic method combines the  $P$  values and is defined by

$$F_i = -2 \sum_{j=1}^N \log(p_{ij}) \quad (8)$$

which tests (chi-square  $\chi^2$ ) the null hypothesis for gene  $i$ , where indices  $i$  and  $j$  denote the  $i$ th gene from the  $j$ th dataset, respectively. However, cMap may return a zero  $P$  value, hence rendering (8) infinite; therefore, it was not used in the present analysis.

There are two models in meta-analysis: the fixed-effect model and random-effect model [41]. In the fixed-effect model it is assumed that there is only one true effect size and that all differences among the studies or batches are due to sampling errors only. In contrast, the random-effect model allows the effect size to vary from study to study. Each study estimates a different effect size. These two models are considered in our work.

In other words, a test for homogeneity of distribution was performed. As it is rather common to find that the effect size may vary from one study to the next, we employ the MA method, such as the  $Q$  statistics and  $I^2$  statistics, to quantify the heterogeneity, to test it, and to incorporate it into the weighting scheme. We use a  $P$  value of 0.1 for  $I^2$  statistics as the criterion for statistical significance. A  $P$  value larger than or equal to 0.1 means that there is little variation between batches; then a fixed-effect model might be appropriate; otherwise choose random-effect model [41]. Degree of heterogeneity is characterized by the  $I^2$  value. A value of  $I^2$  less than 25% implies no heterogeneity, whereas a value larger than 75% means extremely high heterogeneity.

If the studies are homogenous, then it is likely that the various studies are testing the same hypothesis. If these estimates are heterogeneous, then it is probable that each study is not testing the same hypothesis. Therefore, it may not be appropriate to combine all the study results into one meta-analysis. In such case, we would need to conduct a separate meta-analysis, such as meta-regression analysis for different subsets of studies [41].

**2.6. Cell Culture.** All cell-culture-related reagents were purchased from Invitrogen. Human lung cancer cell lines A549 and H460 were purchased from the American Type Culture Collection/Bioresource Collection and Research Center (BCRC) (Taiwan). These cells have performed STR-PCR profile at BCRC. A14 was a derivative of A549 cells stably selected with a p53 shRNA construct. Human lung adenocarcinoma cell lines, CL<sub>1-0</sub> and CL<sub>1-5</sub>, were kind gifts from Dr. Pan-Chyr Yang. H1299 stable clones (transfected with EGFR-WT (wild type) and EGFR-L858R mutant) were kindly provided by Dr. Yi-Rong Chen. All cells were cultured in RPMI 1640 with 10% fetal bovine serum (FBS), 2 mM of L-glutamine, and 1% penicillin/streptomycin and maintained in a 37°C, 5% CO<sub>2</sub> incubator.

**2.7. MTT Cell Viability Test.** Cell viability was determined using an MTT assay. Cells were seeded in a 96-well microplate for 16~20 hrs and treated with the indicated drugs. After drug treatment for 72 hrs, 50  $\mu$ L MTT solution (2 mg/mL) per well was added and incubated at 37°C. Two hours later, 150  $\mu$ L liquid per well was removed and DMSO was added and the absorbance at 570 nm was detected using an ELISA reader (Infinite M1000, TECAN, Switzerland). The untreated groups were considered to be 100% viable.



TABLE 2: Statistics of DEGs for early- and late-stage NSCLC.

Early-stage NSCLC						
	GSE7670	GSE10072	GSE19804	GSE27262	Common DEGs	#PPI
DEGs	642	642	642	642		
UP DEGs	213	213	214	212	211	41
DOWN DEGs	429	429	428	430	426	105
Late-stage NSCLC						
	GSE7670	GSE10072	GSE19804	GSE27262	Common DEGs	#PPI
DEGs	780	780	780	n/a		
UP DEGs	257	258	257	n/a	254	166
DOWN DEGs	523	522	523	n/a	520	133

#PPI denotes the total number of PPI among common DEGs.

TABLE 3: Total number of  $k$ -communities identified by CFinder.

Early-stage NSCLC				
$k$	Up group	Number of genes	Down group	Number of genes
3	2	6	6	12
4	0	0	1	4
Late-stage NSCLC				
$k$	Up group	Number of genes	Down group	Number of genes
3	31	37	6	16
4	3	8	0	0

2.8. *Clonogenic Assay.* Two NSCLC cell lines, A549 and H460, were seeded in 6-well plates with 500 cells/well for 7–10 days. Each well contained 1.5 mL RPMI medium as culture condition and tested drugs were added 24 hrs after the seeding of the cells. The medium and drugs were changed once on day four. After treatment, cells were washed with PBS, and the colonies were fixed (acetic acid:methanol, 1:3) and stained with 0.5% crystal violet in methanol. After removing the excess crystal violet and rinsing with tap water, the colonies were counted manually.

### 3. Results

3.1. *Microarray Data Analysis.* In this study multiple microarray source data were employed for analysis. Robust multiarray average (RMA) was used for gene expression normalization. The eBayes analysis was subsequently conducted on the previous results. DEGs were predicted by an eBayes with an adjusted  $P$  value of 0.005. By integrating with the BioGrid [42] PPI data, a list of binary interactions among DEGs was determined for the up and down groups.

There may be concern regarding the use of different microarray platforms being subjected to heterogeneity problem. We note that the following two steps can tackle such concern: (i) selecting common DEGs among all the platforms for further analysis and (ii) employing meta-analysis approach and performing test of heterogeneity to determine whether the fixed-effect model or random-effect model is needed.

A total of 642 and 780 genes were identified as the common DEGs for the early- and late-stage cancer, respectively. The results of the total number of DEGs, “UP” and “DOWN” DEGs for early- and late-stages of cancer, are reported in

Table 2. The second last column in the table denotes the total number of UP and DOWN DEGs for different GSE platforms. It is noted that the number of “DOWN” DEGs identified is larger than “UP” DEGs in both of early- and late-stages, in which the ratio is about 2 to 1.

3.2. *Cluster Analysis.* Genes which do not highly interact with other genes are assumed to be less important and consequently such genes were removed before the subsequent analysis. Hence, by CFinder, any gene which did not belong to a  $k$ -community was excluded. We also counted the number of  $k$ -communities in the NSCLC PPI network and found there was no community with  $k$  larger than five. Table 3 summarizes the number of  $k$ -communities identified by CFinder. For early-stage, a total of six and sixteen genes belong to the two up- and seven downregulated  $k$ -communities, respectively, whereas a total of forty-five and sixteen genes belong to the thirty-four up- and six downregulated  $k$ -communities, respectively, for late-stage.

Only genes belonging to the communities identified by CFinder were selected for the next stage of analysis.

3.3. *Enriched Biological Pathways.* Pathway annotation of communities was given by implementing DAVID and CPDB. According to REACTOME [43] and KEGG [44] databases, pathways with their  $P$  values less than 0.05 and ranked among the top ten are reported. Using the annotation tool in DAVID database REACTOME, Table 4 lists the enriched pathways information for early- and late-stage NSCLC. The “Count” and “%” columns denote the number of overlapped genes in the filtered community genes and the corresponding pathway and the percentage of overlapped genes, respectively.

TABLE 4: Summary of pathways returned by REACTOME using DAVID for early- and late-stage NSCLC.

Early-stage NSCLC			
Term	Count	%	<i>P</i> value
Hemostasis	6	27.3	$2.87E - 04$
Signaling in immune system	6	27.3	$7.22E - 04$
Integrin cell surface interactions	3	13.6	0.0223
Metabolism of carbohydrates	3	13.6	0.0295
Late-stage NSCLC			
Term	Count	%	<i>P</i> value
Cell cycle, mitotic	21	34.4	$8.49E - 12$
Cell cycle checkpoints	10	16.4	$3.08E - 06$
Cdc20: phospho-APC/C-mediated degradation of cyclin A	6	9.8	$8.20E - 04$
APC-Cdc20-mediated degradation of Nek2A	4	6.6	0.00211
DNA replication	5	8.2	0.0239

TABLE 5: Summary of the top ten pathways returned by REACTOME using CPDB for early- and late-stage NSCLC.

Early-stage NSCLC			
Pathway name	Effective size	% of overlap	<i>P</i> value
Cell surface interactions at the vascular wall	94	6.4%	$1.53E - 08$
Nephrin interactions	23	13%	$1.02E - 05$
Hemostasis	463	1.5%	$1.45E - 05$
Glycolysis	28	10.7%	$1.87E - 05$
Gluconeogenesis	32	9.4%	$2.82E - 05$
PECAM1 interactions	12	16.7%	0.00022
Integrin cell surface interactions	66	4.5%	0.00025
Glucose metabolism	67	4.5%	0.00026
Regulation of signaling by CBL	18	11.1%	0.00052
CD28 dependent PI3K/Akt signaling	21	9.5%	0.00071
Late-stage NSCLC			
Pathway name	Effective size	% of overlap	<i>P</i> value
Cell cycle	442	5.9%	$5.15E - 21$
Cell cycle, mitotic	355	6.8%	$1.02E - 20$
Mitotic M-M/G1 phases	214	8.4%	$3.88E - 17$
APC/C-mediated degradation of cell cycle proteins	38	26.3%	$4.87E - 15$
Regulation of mitotic cell cycle	38	26.3%	$4.87E - 15$
M phase	183	8.2%	$3.57E - 14$
Mitotic prometaphase	110	10.9%	$5.59E - 13$
Mitotic anaphase	130	9.2%	$4.22E - 12$
Mitotic metaphase and anaphase	131	9.2%	$4.63E - 12$
Resolution of sister chromatid cohesion	101	10.9%	$5.80E - 12$

As we noted from Table 4, GSEA suggested that hemostasis, signaling in immune system, integrin cell surface interaction, and metabolism of carbohydrates are enriched pathways for early-stage cancer, whereas cell cycle and DNA replication pathways are ranked among the top for late-stage cancer. It is noted that these late-stage cancer pathways are dominated by cell-cycle related processes.

Cancer is a multistage progression process that results from mutations in genetic sequences. Accumulation of genetic mutations could lead to a defective DNA repair mechanism, consequently giving rise to genetic instability and uncontrolled cell growth [45].

Numerous studies have reported that homeostasis and cancer formation are related [46–49]. Integrins are the receptors that mediate cell adhesion to the extracellular matrix (ECM). Varner and Cheresch [50] pointed out in 1996 that ECM receptors, integrins, regulate the cellular proliferation machinery in tumor cells. In the seminal review paper written by Hanahan and Weinberg [51], it was reported that integrin can influence cell behavior and transform cells into an active proliferative state. Recent studies have also suggested that integrins are involved in cancer progression [52, 53] and lung squamous cell carcinoma [54]. Furthermore, elevated glucose consumption is observed in tumor formation [55–57].

TABLE 6: Summary of top pathways returned by KEGG using DAVID for early- and late-stage cancer.

Early-stage NSCLC			
Term	Count	%	P value
Leukocyte transendothelial migration	4	18.2	0.00294
Glycolysis/gluconeogenesis	3	13.6	0.00982
Epithelial cell signaling in <i>Helicobacter pylori</i> infection	3	13.6	0.0125
Cell adhesion molecules (CAMs)	3	13.6	0.0432
Tight junction	3	13.6	0.0444
Focal adhesion	3	13.6	0.0911
Late-stage NSCLC			
Term	Count	%	P value
Cell cycle	12	19.7	3.22E – 10
Oocyte meiosis	8	13.1	7.47E – 06
Progesterone-mediated oocyte maturation	6	9.8	2.68E – 04
DNA replication	3	4.9	0.0250
p53 signaling pathway	3	4.9	0.0791

TABLE 7: Summary of top pathways returned by KEGG using CPDB for early- and late-stage cancer.

Early-stage NSCLC			
Term	Effective size	% of overlap	P value
Leukocyte transendothelial migration	118	3.4%	6.37E – 05
Glycolysis/gluconeogenesis	66	4.5%	0.000250
Epithelial cell signaling in <i>Helicobacter pylori</i> infection	68	4.4%	0.000273
HIF-1 signaling pathway	106	2.8%	0.00100
Tight junction	134	2.2%	0.00197
Cell adhesion molecules (CAMs)	147	2.0%	0.00257
Late-stage NSCLC			
Term	Effective size	% of overlap	P value
Cell cycle	124	9.7%	2.39E – 12
Oocyte meiosis	110	7.3%	1.38E – 07
Progesterone-mediated oocyte maturation	86	7.0%	7.25E – 06
DNA replication	36	8.3%	0.00100
Epstein-Barr virus infection	203	2.5%	0.00512
Viral carcinogenesis	206	2.4%	0.00544
p53 signaling pathway	68	4.4%	0.00620
Measles	134	3.0%	0.00630
Hepatitis B	146	2.7%	0.00849

Late-stage cancer patients commonly have cell invasion and metastasis. Malignant cells have the ability to invade adjacent normal tissue structures. Malignant tumor cells break off from the tumor and enter blood vessels or the lymphatic system and migrate to other parts of the body and initiate another tumor. Biomedical studies have suggested that the development of the metastatic process involves an interaction between cell cycle signaling, adhesion pathways, and epithelial-mesenchymal transition program [13]. It is also known that signal transduction pathways, such as p53, MAPK, Notch, and ROS, are heavily involved in metastasis

[58]. In particular, mutations in p53 and K-RAS appears only later in tumor progression [45].

Defects in the cell cycle mitotic checkpoint generate aneuploidy and might facilitate tumorigenesis [59]. Mitotic progression and sister-chromatid segregation are controlled by the anaphase promoting complex/cyclosome (APC/C). APC/C forms a protein complex with its mitotic coactivator, CDC20, which controls mitotic progression [59]. CDC20 protein level may directly affect cell fate during prolonged mitotic arrest [60] and its turnover rate may be a key factor in cancer patient response to antimitotic therapies [61].

Using the CPDB tool, the top ten most significant pathways for early-stage NSCLC and late-stage NSCLC returned by REACTOME are listed in Table 5. Again, GSEA suggested that hemostasis, cell surface interaction, and metabolism of carbohydrates, that is, glycolysis and gluconeogenesis, are the enriched pathways for early-stage cancer. For late-stage cancer, again it is found that the cell cycle pathways are ranked among the top pathways. Essentially, results returned by DAVID and CPDB are consistent with each other.

From Table 5, it is found that PECAM1 [62–64] and CBL are frequently altered in lung cancer [65], and CD28 is associated with NSCLC formation [66]. PECAM1 interactions are related to angiogenesis.

Using the KEGG database, pathways with  $P$  value less than 0.05 returned by DAVID are listed for early-stage and late-stage cancer in Table 6. Again, enrichment analysis suggested that glycolysis/gluconeogenesis and cell signaling are the enriched pathways for early-stage cancer. It is known that integrin is a key regulator of cell adhesion [53].

It was also found that the cell cycle pathway and DNA replication pathway were ranked among the top pathways for late-stage cancer. It is known that cancer is due to uncontrolled cell mitosis, and this uncontrolled process is a common element in all types of cancer.

Cell adhesion molecules (CAMs), a diverse system of glycoproteins, have been found to play an important role in cancer progression and in the application of cancer therapy [67–69]. Tight junctions are cellular structures located at the apicobasal region of epithelial cell membranes [70]. It has been experimentally found that lung tumors show changes in the expression in tight junction proteins [71]. Other studies have also indicated that tight junction proteins show aberrant expression in breast cancer [72] and correlate with metastasis [73–75].

We noted that the significant enriched pathways found in Table 6 (late-stage) are also identified in the work by Liu et al. [76]. Except oocyte meiosis, the other four pathways are involved in two NSCLC subtypes: adenocarcinoma and squamous cell carcinoma.

Using the CPDB tool, significant pathways returned by KEGG are listed in Table 7. Again it was found that the cell cycle pathway and DNA replication pathway are ranked among the top pathways for late-stage cancer.

The hypoxia-inducible factor-1 (HIF-1) is an oxygen-sensitive transcriptional activator and is causally involved in NSCLC [77–79].

Again, the cell cycle pathway ranked first (among the top of the list) both in REACTOME and KEGG using CPDB. In other words, analyses using DAVID and CPDB are in good agreement. Relative to DAVID, CPDB tends to return more pathway information.

Integrins are the receptors that mediate cell adhesion to ECM. The extracellular matrix (ECM) is a network of macromolecules that underlies all epithelia and endothelia and that surrounds all connective tissue cells. This matrix provides mechanical strength and also influences the behavior and differentiation state of cells in contact with it.

TABLE 8: The number of IC50 verified drugs and potential drugs identified by using ES, cMap  $P$  value less than 0.1 and 0.5 for early- and late-stage NSCLC.

Effect size	Early-stage	Late-stage
ES	2/24	6/30
cMap $P$ value < 0.1	3/13	5/26
cMap $P$ value < 0.5	8/56	7/65

Numbers before and after the slash sign (/) denote the numbers of IC50 verified drugs and potential drugs, respectively.

TABLE 9: IC50 values of potential drugs for early- and late-stage NSCLC.

Effect size	Stage	cMap drug name	MTT	Clonogenic		
ES	Early	Mebendazole	<1	>10		
		Prenylamine	>5	>10		
		Mebendazole	<1	>10		
		Spiperone	>10	<10		
	Late	Anisomycin	<0.1	<0.1		
		Pyrvinium	<0.1	<0.1		
		Mefloquine	>5	>5		
		Niclosamide	>5	>5		
		$P$ value < 0.1	Early	Trichostatin A	<1	<1
				Monensin	<1	<1
Cloperastine	<10			<10		
Late	Trichostatin A		<1	<1		
	Mefloquine		>5	>5		
	Pyrvinium		<0.1	<0.1		
		Securinine	>5	>5		
		Nortriptyline	<10	<10		

3.4. Potential Drugs and Their Target Genes for NSCLC. Both the up- and downregulated communities extracted from CFinder were analyzed by cMap. Under the constraint of an enrichment score (ES) of less than zero, and cMap drugs associated with  $P$  value, that is, cMap  $P$  value less than 0.1 or 0.5, potential drugs were inferred by performing MA. Fisher's summary statistic method was used for combining cMap  $P$  value.

After performing meta-analysis using ES as the effect size, twenty-four potential drugs were found with a  $P$  value for MA being less than 0.05 for early-stage cancer. The results are listed in Table 8. Among the twenty-four drugs, two drugs tested by MTT or clonogenic assay were validated as effective (i.e., mebendazole and prenylamine).

From Table 8, among the 30 potential drugs ( $P$  value for MA is less than 0.05) for late-stage cancer, there were six drugs tested by MTT or clonogenic assay and validated as effective, that is, mebendazole, spiperone, anisomycin, pyrvinium, mefloquine, and niclosamide.

We performed the heterogeneity test on the 24 drugs for early-stage and the 30 drugs for the late-stage cancer using the  $I^2$  statistics. It is found that both of the fix-effect model and the random-effect model are required according to the  $I^2$  statistics test with a  $P$  value less than 0.1 [41].

We used available drugs in the list for *in vitro* cytotoxic validations (Table 9). Certain drugs showed effective



TABLE 10: The number of common drugs for early-stage and late-stage using the enrichment score (ES) and cMap  $P$  value (less than 0.1) for meta-analysis.

Effect size	ES		$P$ value	
	Early-stage	Late-stage	Early-stage	Late-stage
ES				
Early-stage		0.421	0.156	0.136
Late-stage	16		0.132	0.217
$P$ value < 0.1				
Early-stage	5	5		0.182
Late-stage	6	10	6	

cytotoxic effects for lung cancer cells. However, the very limited data showed that there were inconsistencies in MTT and clonogenic assays. For example, mebendazole showed a good IC50 in MTT ( $<1\mu\text{M}$ ) but not in the clonogenic assay ( $>10\mu\text{M}$ ). On the other hand, spiperone showed a relatively effective IC50 ( $<10\mu\text{M}$ ) in clonogenic assay rather than in MTT assay ( $>10\mu\text{M}$ ). This phenomenon is still hard to explain in the current status.

Dose-dependent figures for four of the representable drugs are shown in Figure 2. The reasons to use two most commonly used lung cancer cell lines A549 and H460 include the following: (i) they have different histologic subtypes, that is, A549 is adenocarcinoma and H460 is large cell carcinoma, although both belong to non-small-cell carcinoma; (ii) the origin of A549 cell was obtained from lung tissue and H460 was from lung pleural effusion, which may represent different stages of lung cancer; and (iii) both cells are EGFR wild type that could be tested by the drugs potentially effective for intrinsic EGFR-TKI resistance.

We conducted meta-analysis using the  $P$  values from cMap drugs obtained from individual arrays. As shown in Table 10, the first row lists the early- and late-stage ES and  $P$  value used for meta-analysis. Entries in the lower diagonal denote the number of common drugs for MA choosing ES and  $P$  value as the effective sizes, and, in contrast, entries in upper-diagonal show its Jaccard index (JI) score. Given two sets  $A$  and  $B$ ,  $\text{JI}(A, B)$  is defined as  $|A \cap B| / (|A| \cup |B| - |A \cap B|)$ , where  $|A \cap B|$ ,  $|A|$  and  $|B|$  denote the cardinality of  $A \cap B$ ,  $A$  and  $B$ , respectively.

For early-stage cancer, there are five common drugs (JI is 0.156) predicted by both ES and cMap  $P$  value meta-analysis, whereas there are ten common drugs (JI is 0.217) for late-stage cancer. The number of common drugs for both early- and late-stage cancer are around five or six, assuming ES versus  $P$  value. There are sixteen (JI = 0.421) and six (JI = 0.182) common drugs predicted by ES and  $P$  value meta-analysis, respectively. This seems to indicate that MA tends to return a higher overlapping between early- and late-stage results.

We submitted the selected drugs to DrugBank and NCBI to search for up- and downregulated target genes. The results of the number of target genes are summarized in Table II, which are potential therapeutic targets for future lung cancer clinical trials. For early-stage cancer, no target gene is reported by using the GSE7670 platform; therefore, we report common drug target genes among the rest of

TABLE 11: The results of up- and downregulated target genes within  $k$ -communities obtained from DrugBank and NCBI.

	GSE7670	GSE10072	GSE19804	GSE27262
	Early-stage			
Up	0	7	10	15
Down	0	5	7	14
	Late-stage			
Up	1	4	7	n/a
Down	2	1	7	n/a

the other three microarray platforms. For late-stage cancer, we report common drug target genes among any two of the three platforms.

Table 12 summarizes the up- and down communities' drug target genes. As it is shown in the table, certain genes are predicted by both effect size studies. For instance, up community genes, RPL26L1, FEN1, and IDH1, are found in both studies.

Figure 3 depicts the PPI network of upregulated target genes using Cytoscape [80]. The upregulated target gene RPL26L1 directly interacts with six proteins; Figure 4 represents the PPI network of downregulated target gene, PPARG. This gene directly interacts with eleven proteins.

## 4. Conclusion

We applied the meta-analysis technique to infer therapeutic drugs for NSCLC treatment by integrating microarray expression profiles. Since cancer is a multistage progressive disease, early- and late-stage CAG are potentially very different; therefore, stage-specific DEGs were identified. PPI data were then employed to construct dense PPI modules. The up- and downregulated communities were used as queries to perform functional enrichment analysis and potential drug identification using cMap. Drugs can act on not merely the transcription level, but rather on the protein, posttranscription, or posttranslation levels. Large-scale drug screening needs fast and efficient ways. In the current status, using gene expression change to infer drug repositioning is the most suitable way, which has been claimed in the rationale of cMap original paper [81]. It is still difficult to see the modulations of protein level in such a large-scale, high throughput method.

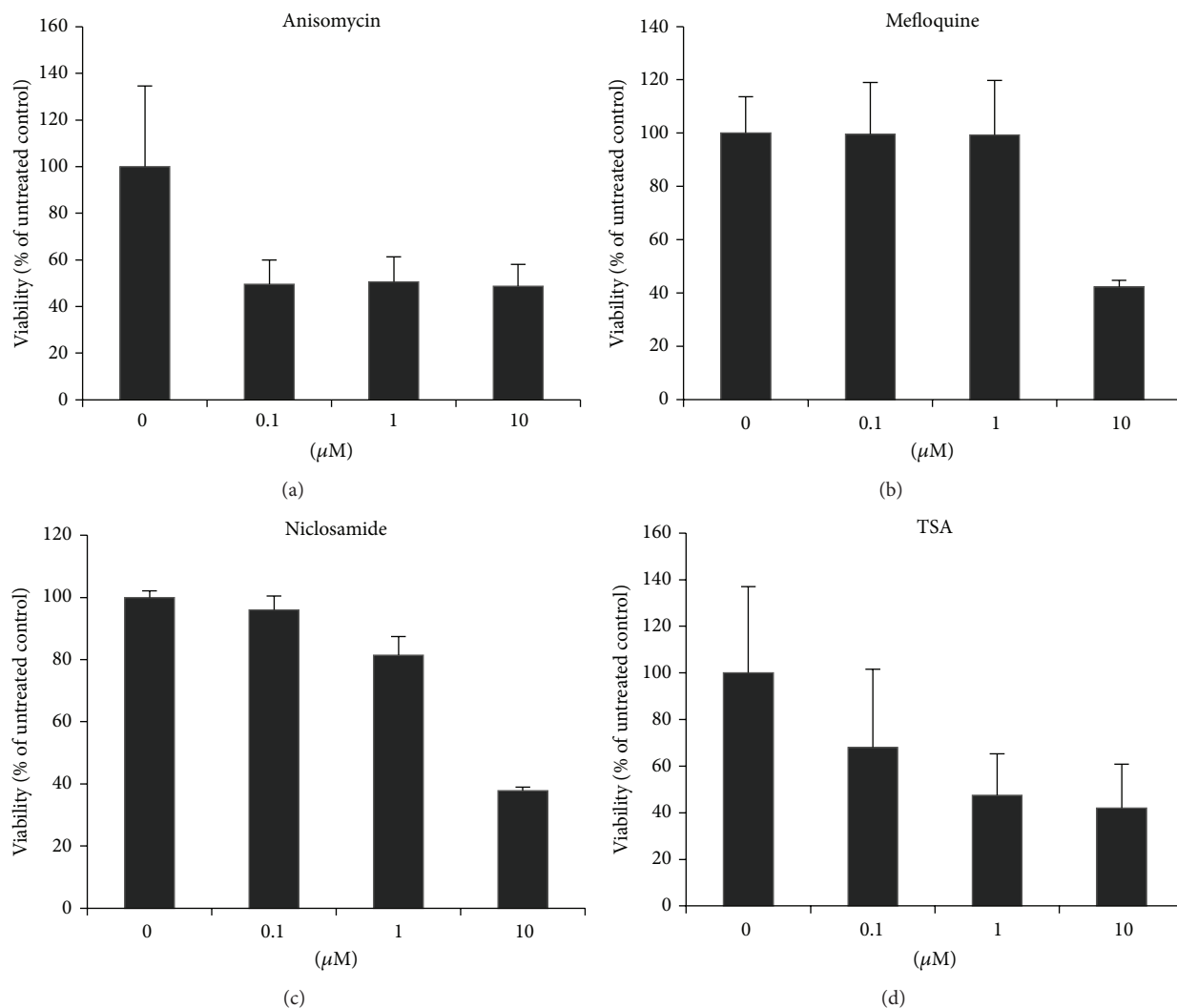


FIGURE 2: Dose-dependent plots for four of the representable drugs, that is, anisomycin, mefloquine, niclosamide, and trichostatin A (TSA).

TABLE 12: The results of up- and downregulated target genes within  $k$ -communities using ES and cMap  $P$  value as the effect size.

Effect size	Up-community gene	Down-community gene
ES		
Early-stage	FEN1, IDH1, PSMB2, PSMB5, RPL26L1	NR3C1, PPARG
Late-stage	PSMB2	NR3C1, PPARG
$P$ value < 0.1		
Early-stage	EZH2, FEN1, IDH1, and RPL26L1	NR3C1, PPARG
Late-stage	Not available	NR3C1, PPARG

Enrichment analysis suggests pathways that are early- and late-stage specific. This supports the use of the meta-analysis technique to derive reliable results when combining multiple gene expression datasets.

Enrichment scores and  $P$  values obtained from cMap were adopted as the effect size indices for target drug meta-analysis. Certain common drugs were found by using the enrichment score and  $P$  value meta-analysis technique.

A fraction of our drug findings results are supported by IC50 experimental data.

Our findings suggest that certain up- and downregulated genes are potential drug targets. Furthermore, the drugs derived from DrugBank and NCBI are potential lung cancer therapeutic drugs.

In summary, we have developed a pipeline to infer therapeutic drugs for disease treatment by integrating microarray,

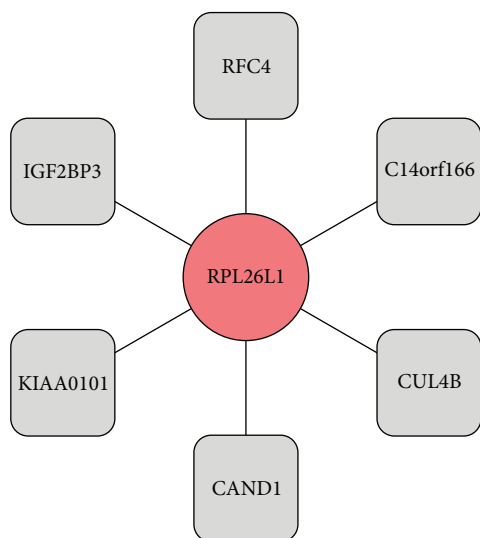


FIGURE 3: The upregulated target genes (circles) PPI partners (squares); solid line represents PPI.

PPI, and the cMap resources. Meta-analysis was adopted to integrate multiple datasets. Up- and downregulated communities were used as queries to perform functional enrichment analysis and potential drug prediction. Overrepresented cancer stage-specific pathways are determined. The target drug results are supported by IC50 measurement data. It is expected that the approach developed in the current work should be of value for future studies into understanding the molecular mechanism of lung cancer formation and identifying therapeutic drug targets.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

The work of Chien-Hung Huang is supported by the National Science Council of Taiwan under Grants NSC 101-2221-E-150-088-MY2, and the work of Ka-Lok Ng is supported by NSC 102-2221-E-468-024 and NSC 102-2632-E-468-001-MY3. The work of Chi-Ying F. Huang is supported by the grant from the National Science Council (NSC102-2627-B-010-001); by the grant from Center of Excellence for Cancer Research at Taipei Veterans General Hospital, Health and Welfare Surcharge of Tobacco Products (MOHW103-TD-B-111-02); by the grant from Ministry of Education, Aim for the Top University Plan (103AC-T503). The work of Cheng-Hsu Wang and Peter Mu-Hsin Chang is supported by the grant from the Keelung Chang Gung Memorial Hospital (CMRPG2D0041). The work of Peter Mu-Hsin Chang is supported by the grant from the National Science Council, NTUH SPARK Research Program (NSC 102-3114-B-002-001). The authors' gratitude goes to Dr. Timothy Williams, Department of Foreign Languages and

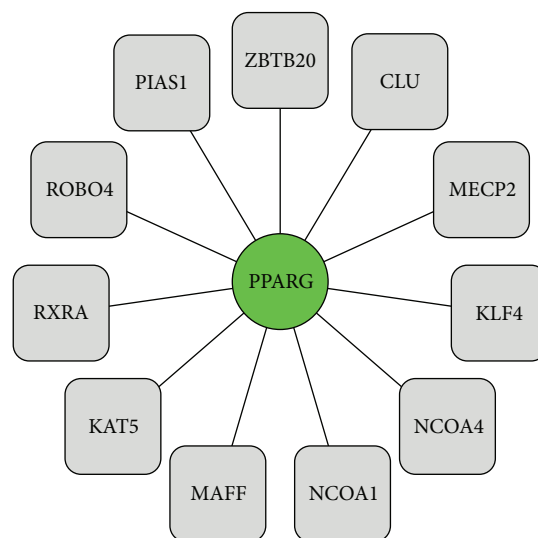


FIGURE 4: The downregulated target genes (circles) PPI partners (squares); solid line represents PPI.

Literature, Asia University, for his help in proofreading the paper.

### References

- [1] A. Jemal, R. Siegel, E. Ward et al., "Cancer statistics, 2008," *CA: A Cancer Journal for Clinicians*, vol. 58, no. 2, pp. 71–96, 2008.
- [2] Department of Health, *Cancer Registry Annual Report in Taiwan Area*, Department of Health, Executive Yuan, China, 2007.
- [3] C. H. Huang, M. Y. Wu, P. M. H. Chang, C. Y. Huang, and K. L. Ng, "In silico identification of potential targets and drugs for non small cell lung cancer," *IET Systems Biology*, vol. 8, no. 2, pp. 56–66, 2014.
- [4] M. Y. Lan, C. L. Chen, K. T. Lin et al., "From NPC therapeutic target identification to potential treatment strategy," *Molecular Cancer Therapeutics*, vol. 9, no. 9, pp. 2511–2523, 2010.
- [5] D. C. Hassane, M. L. Guzman, C. Corbett et al., "Discovery of agents that eradicate leukemia stem cells using an in silico screen of public gene expression data," *Blood*, vol. 111, no. 12, pp. 5654–5662, 2008.
- [6] J. M. Rosenbluth, D. J. Mays, M. F. Pino, L. J. Tang, and J. A. Pietsenpol, "A gene signature-based approach identifies mTOR as a regulator of p73," *Molecular and Cellular Biology*, vol. 28, no. 19, pp. 5951–5964, 2008.
- [7] S. R. Setlur, K. D. Mertz, Y. Hoshida et al., "Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer," *Journal of the National Cancer Institute*, vol. 100, no. 11, pp. 815–825, 2008.
- [8] T. Barrett, S. E. Wilhite, P. Ledoux et al., "NCBI GEO: archive for functional genomics data sets—update," *Nucleic Acids Research*, vol. 41, no. 1, pp. D991–D995, 2013.
- [9] L.-J. Su, C.-W. Chang, Y.-C. Wu et al., "Selection of *DDX5* as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme," *BMC Genomics*, vol. 8, article 140, 2007.
- [10] M. T. Landi, T. Dracheva, M. Rotunno et al., "Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival," *PLoS ONE*, vol. 3, no. 2, Article ID e1651, 2008.

- [11] T. Lu, M. Tsai, J. Lee et al., "Identification of a novel biomarker, *SEMA5A*, for non-small cell lung carcinoma in nonsmoking women," *Cancer Epidemiology Biomarkers and Prevention*, vol. 19, no. 10, pp. 2590–2597, 2010.
- [12] T. Y. Wei, C. C. Juan, J. Y. Hisa et al., "Protein arginine methyltransferase 5 is a potential oncoprotein that upregulates G1 cyclins/cyclin-dependent kinases and the phosphoinositide 3-kinase/AKT signaling cascade," *Cancer Science*, vol. 103, no. 9, pp. 1640–1650, 2012.
- [13] R. A. Weinberg, *The Biology of Cancer*, Garland Science, New York, NY, USA, 2nd edition, 2013.
- [14] B. Efron and R. Tibshirani, "Empirical Bayes methods and false discovery rates for microarrays," *Genetic Epidemiology*, vol. 23, no. 1, pp. 70–86, 2002.
- [15] M. K. Kerr, C. A. Afshari, L. a. Bennett, J. Martinez, and N. J. Walker, "Statistical analysis of a gene expression microarray experiment with replication," *Statistica Sinica*, vol. 12, no. 1, pp. 203–217, 2002.
- [16] V. G. Tusher, R. Tibshirani, and G. Chu, "Significance analysis of microarrays applied to the ionizing radiation response," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 9, pp. 5116–5121, 2001.
- [17] S. Zhang, "A comprehensive evaluation of SAM, the SAM R-package and a simple modification to improve its performance," *BMC Bioinformatics*, vol. 8, article 230, 2007.
- [18] B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher, "Empirical Bayes analysis of a microarray experiment," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1151–1160, 2001.
- [19] B. Efron, "Robbins, empirical Bayes and microarrays," *The Annals of Statistics*, vol. 31, no. 2, pp. 366–378, 2003.
- [20] <http://www.bioconductor.org>.
- [21] R. A. Irizarry, "From CEL files to annotated lists of interesting genes," in *Bioinformatics and Computational Biology Solutions Using R & Bioconductor*, pp. 431–442, Springer, New York, NY, USA, 2005.
- [22] S. T. Chen, H. F. Wu, and K. L. Ng, "A platform for querying breast and prostate cancer-related microRNA genes," in *Proceeding of the International Conference on Bioinformatics and Biomedical Engineering (ICBBE '12)*, pp. 271–274, Shanghai, China, 2012.
- [23] A. D. King, N. Pržulj, and I. Jurisica, "Protein complex prediction via cost-based clustering," *Bioinformatics*, vol. 20, no. 17, pp. 3013–3020, 2004.
- [24] Y. Qi, F. Balem, C. Faloutsos, J. Klein-Seetharaman, and Z. Bar-Joseph, "Protein complex identification by supervised graph local clustering," *Bioinformatics*, vol. 24, no. 13, pp. 250–268, 2008.
- [25] J. B. Pereira-Leal, E. D. Levy, and S. A. Teichmann, "The origins and evolution of functional modules: lessons from protein complexes," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, no. 1467, pp. 507–517, 2006.
- [26] B. Adamcsek, G. Palla, I. J. Farkas, I. Derényi, and T. Vicsek, "CFinder: locating cliques and overlapping modules in biological networks," *Bioinformatics*, vol. 22, no. 8, pp. 1021–1023, 2006.
- [27] J. Wang, B. Liu, M. Li, and Y. Pan, "Identifying protein complexes from interaction networks based on clique percolation and distance restriction," *BMC Genomics*, vol. 11, article S10, supplement 2, 2010.
- [28] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.
- [29] Gene Ontology Consortium, "The gene ontology (GO) project in 2006," *Nucleic Acids Research*, vol. 34, pp. D322–D326, 2006.
- [30] A. Kamburov, C. Wierling, H. Lehrach, and R. Herwig, "ConsensusPathDB—a database for integrating human functional interaction networks," *Nucleic Acids Research*, vol. 37, no. 1, pp. D623–D628, 2009.
- [31] M. Kuhn, C. von Mering, M. Campillos, L. J. Jensen, and P. Bork, "STITCH: interaction networks of chemicals and proteins," *Nucleic Acids Research*, vol. 36, no. 1, pp. D684–D688, 2008.
- [32] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nature Reviews Drug Discovery*, vol. 3, no. 8, pp. 673–683, 2004.
- [33] F. Iorio, R. Bosotti, E. Scacheri et al., "Discovery of drug mode of action and drug repositioning from transcriptional responses," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 33, pp. 14621–14626, 2010.
- [34] Z. Wu, Y. Wang, and L. Chen, "Network-based drug repositioning," *Molecular BioSystems*, vol. 9, no. 6, pp. 1268–1281, 2013.
- [35] Z. Wu, Y. Wang, and L. Chen, "A new method to identify repositioned drugs for prostate cancer," in *Proceedings of the 6th IEEE International Conference on Systems Biology (ISB '12)*, pp. 280–284, Xian, China, August 2012.
- [36] S. Y. Sun, Z. P. Liu, T. Zeng, Y. Wang, and L. Chen, "Spatio-temporal analysis of type 2 diabetes mellitus based on differential expression networks," *Scientific Reports*, vol. 3, article 2268, 2013.
- [37] S. Zhao and S. Li, "Network-based relating pharmacological and genomic spaces for drug target identification," *PLoS ONE*, vol. 5, no. 7, Article ID e11764, 2010.
- [38] S. Zhao and S. Li, "A co-module approach for elucidating drug-disease associations and revealing their molecular basis," *Bioinformatics*, vol. 28, no. 7, pp. 955–961, 2012.
- [39] J. Ahmed, T. Meinel, M. Dunkel et al., "CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge," *Nucleic Acids Research*, vol. 39, no. 1, pp. D960–D967, 2011.
- [40] F. M. Wolf, *Meta-Analysis: Quantitative Methods for Research Synthesis*, Sage, Thousand Oaks, Calif, USA, 1986.
- [41] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, *Introduction to Meta-Analysis*, John Wiley & Sons, London, UK, 2009.
- [42] B. Breitkreutz, C. Stark, T. Reguly et al., "The BioGRID interaction database: 2008 update," *Nucleic Acids Research*, vol. 36, no. 1, pp. D637–D640, 2008.
- [43] D. Croft, G. O'Kelly, G. Wu et al., "Reactome: a database of reactions, pathways and biological processes," *Nucleic Acids Research*, vol. 39, no. 1, pp. D691–D697, 2011.
- [44] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori, "The KEGG resource for deciphering the genome," *Nucleic Acids Research*, vol. 32, pp. D277–D280, 2004.
- [45] S. A. Frank, *Dynamics of Cancer: Incidence, Inheritance, and Evolution*, Princeton University Press, Princeton, NJ, USA, 2007.
- [46] C. Boccaccio and P. M. Comoglio, "A functional role for hemostasis in early cancer development," *Cancer Research*, vol. 65, no. 19, pp. 8579–8582, 2005.
- [47] M. Franchini, M. Montagnana, E. J. Favaloro, and G. Lippi, "The bidirectional relationship of cancer and hemostasis and the potential role of anticoagulant therapy in moderating thrombosis and cancer spread," *Seminars in Thrombosis and Hemostasis*, vol. 35, no. 7, pp. 644–653, 2009.
- [48] S. Jain, J. Harris, and J. Ware, "Platelets: linking hemostasis and cancer," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 30, no. 12, pp. 2362–2367, 2010.



- [49] D. Garnier, N. Magnus, E. D'Asti et al., "PL-05 genetic pathways linking hemostasis and cancer," *Thrombosis Research*, vol. 129, no. 1, pp. S22–S29, 2012.
- [50] J. A. Varner and D. A. Cheresh, "Integrins and cancer," *Current Opinion in Cell Biology*, vol. 8, no. 5, pp. 724–730, 1996.
- [51] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *Cell*, vol. 100, no. 1, pp. 57–70, 2000.
- [52] R. Rathinam and S. K. Alahari, "Important role of integrins in the cancer biology," *Cancer and Metastasis Reviews*, vol. 29, no. 1, pp. 223–237, 2010.
- [53] D. Subramani and S. K. Alahari, "Integrin-mediated function of Rab GTPases in cancer progression," *Molecular Cancer*, vol. 9, article 312, 2010.
- [54] L. F. Stead, S. Berri, H. M. Wood et al., "The transcriptional consequences of somatic amplifications, deletions, and rearrangements in a human lung squamous cell carcinoma," *Neoplasia*, vol. 14, no. 11, pp. 1075–1086, 2012.
- [55] R. J. Gillies, I. Robey, and R. A. Gatenby, "Causes and consequences of increased glucose metabolism of cancers," *Journal of Nuclear Medicine*, vol. 49, supplement 2, pp. 24S–42S, 2008.
- [56] R. B. Hamanaka and N. S. Chandel, "Targeting glucose metabolism for cancer therapy," *Journal of Experimental Medicine*, vol. 209, no. 2, pp. 211–215, 2012.
- [57] A. Annibaldi and C. Widmann, "Glucose metabolism in cancer cells," *Current Opinion in Clinical Nutrition and Metabolic Care*, vol. 13, no. 4, pp. 466–470, 2010.
- [58] R. J. B. King and M. W. Robins, *Cancer Biology*, Prentice Hall, Upper Saddle River, NJ, USA, 3rd edition, 2006.
- [59] G. J. P. L. Kops, B. A. A. Weaver, and D. W. Cleveland, "On the road to cancer: aneuploidy and the mitotic checkpoint," *Nature Reviews Cancer*, vol. 5, no. 10, pp. 773–785, 2005.
- [60] J. Nilsson, "Cdc20 control of cell fate during prolonged mitotic arrest: Do Cdc20 protein levels affect cell fate in response to antimitotic compounds?" *BioEssays*, vol. 33, no. 12, pp. 903–909, 2011.
- [61] A. M. Fry, "Cdc20 turnover rate: a key determinant in cancer patient response to anti-mitotic therapies?" *BioEssays*, vol. 35, no. 9, pp. 762–762, 2013.
- [62] L. A. Schimmenti, H.-C. Yan, J. A. Madri, and S. M. Albelda, "Platelet endothelial cell adhesion molecule, PECAM-1, modulates cell migration," *Journal of Cellular Physiology*, vol. 153, no. 2, pp. 417–428, 1992.
- [63] H. M. DeLisser, M. Christofidou-Solomidou, R. M. Strieter et al., "Involvement of endothelial PECAM-1/CD31 in angiogenesis," *The American Journal of Pathology*, vol. 151, no. 3, pp. 671–677, 1997.
- [64] N. Ilan and J. A. Madri, "PECAM-1: old friend, new partners," *Current Opinion in Cell Biology*, vol. 15, no. 5, pp. 515–524, 2003.
- [65] Y. H. Tan, S. Krishnaswamy, S. Nandi et al., "CBL is frequently altered in lung cancers: its relationship to mutations in met and EGFR tyrosine kinases," *PLoS ONE*, vol. 5, no. 1, Article ID e8972, 2010.
- [66] L. Karabon, E. Pawlak, A. Tomkiewicz et al., "CTLA-4, CD28, and ICOS gene polymorphism associations with non-small-cell lung cancer," *Human Immunology*, vol. 72, no. 10, pp. 947–954, 2011.
- [67] T. Okegawa, R. Pong, Y. Li, and J. Hsieh, "The role of cell adhesion molecule in cancer progression and its application in cancer therapy," *Acta Biochimica Polonica*, vol. 51, no. 2, pp. 445–457, 2004.
- [68] N. Makrilia, A. Kollias, L. Manolopoulos, and K. Syrigos, "Cell adhesion molecules: role and clinical significance in cancer," *Cancer Investigation*, vol. 27, no. 10, pp. 1023–1037, 2009.
- [69] M. Zigler, A. S. Dobroff, and M. Bar-Eli, "Cell adhesion: Implication in tumor progression," *Minerva Medica*, vol. 101, no. 3, pp. 149–162, 2010.
- [70] N. Sawada, M. Murata, K. Kikuchi et al., "Tight junctions and human diseases," *Medical Electron Microscopy*, vol. 36, no. 3, pp. 147–156, 2003.
- [71] Y. Soini, "Tight junctions in lung cancer and lung metastasis: a review," *International Journal of Clinical and Experimental Pathology*, vol. 5, no. 2, pp. 126–136, 2012.
- [72] K. Brennan, G. Offiah, E. A. McSherry, and A. M. Hopkins, "Tight junctions: a barrier to the initiation and progression of breast cancer?" *Journal of Biomedicine and Biotechnology*, vol. 2010, Article ID 460607, 16 pages, 2010.
- [73] T. A. Martin and W. G. Jiang, "Tight junctions and their role in cancer metastasis," *Histology and Histopathology*, vol. 16, no. 4, pp. 1183–1195, 2001.
- [74] T. A. Martin and W. G. Jiang, "Loss of tight junction barrier function and its role in cancer metastasis," *Biochimica et Biophysica Acta*, vol. 1788, no. 4, pp. 872–891, 2009.
- [75] T. A. Martin, M. D. Mason, and W. G. Jiang, "Tight junctions in cancer metastasis," *Frontiers in Bioscience*, vol. 16, no. 3, pp. 898–936, 2011.
- [76] J. Liu, X. Y. Yang, and W. J. Shi, "Identifying differentially expressed genes and pathways in two types of non-small cell lung cancer (NSCLC): adenocarcinoma and squamous cell carcinoma," *Genetics and Molecular Research*, vol. 13, no. 1, pp. 95–102, 2014.
- [77] T. Yohena, I. Yoshino, T. Takenaka et al., "Upregulation of hypoxia-inducible factor-1 $\alpha$ mRNA and its clinical significance in non-small cell lung cancer," *Journal of Thoracic Oncology*, vol. 4, no. 3, pp. 284–290, 2009.
- [78] A. L. Jackson, B. Zhou, and W. Y. Kim, "HIF, hypoxia and the role of angiogenesis in non-small cell lung cancer," *Expert Opinion on Therapeutic Targets*, vol. 14, no. 10, pp. 1047–1057, 2010.
- [79] M. Ioannou, G. Simos, and G. K. Koukoulis, "HIF-1 $\alpha$  in lung carcinoma: histopathological evidence of hypoxia targets in patient biopsies," *Journal of Solid Tumors*, vol. 3, no. 2, pp. 35–43, 2013.
- [80] M. S. Cline, M. Smoot, E. Cerami et al., "Integration of biological networks and gene expression data using Cytoscape," *Nature Protocols*, vol. 2, no. 10, pp. 2366–2382, 2007.
- [81] J. Lamb, E. D. Crawford, D. Peck et al., "The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease," *Science*, vol. 313, no. 5795, pp. 1929–1935, 2006.