Thoracic Cancer
Open Access
Thoracic Cancer ISSN 1759-7706

ORIGINAL ARTICLE

# Metabolic gene *NR4A1* as a potential therapeutic target for non-smoking female non-small cell lung cancer patients

Rong Sun[1†], Min-Yue Bao[2†], Xin Long[1], Yuan Yuan[1], Miao-Miao Wu[1], Xin Li[2] & Jin-Ku Bao[1,2]

1 Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China
2 State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu, China

**Keywords**
Gene co-expression network; microarray data; nilotinib; non-small cell lung cancer; non-smoking female.

**Correspondence**
Jin-Ku Bao, Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, No. 29, Wangjiang Road, Chengdu, Sichuan 610064, China.
Tel: +86 133 0822 2306
Fax: +86 28 8541 2279
Email: baojinku@scu.edu.cn

†These authors contributed equally to this work.

Received: 13 October 2018;
Accepted: 5 January 2019.

doi: 10.1111/1759-7714.12989

Thoracic Cancer **10** (2019) 715–727

## Abstract

**Background:** Although cigarette smoking is considered one of the key risk factors for lung cancer, 15% of male patients and 53% of female patients with lung cancer are non-smokers. Metabolic changes are critical features of cancer. Therapeutic target identification from a metabolic perspective in non-small cell lung cancer (NSCLC) tissue of female non-smokers has long been ignored.

**Results:** Based on microarray data retrieved from Affymetrix expression arrays E-GEOD-19804, we found that the downregulated genes in non-smoking female NSCLC patients tended to participate in protein/amino acid and lipid metabolism, while upregulated genes were more involved in protein/amino acid and carbohydrate metabolism. Combining nutrient metabolic co-expression, protein–protein interaction network construction and overall survival assessment, we identified *NR4A1* and *TIE1* as potential therapeutic targets for NSCLC in female non-smokers. To accelerate the drug development for non-smoking female NSCLC patients, we identified nilotinib as a potential agonist targeting *NR4A1* encoded protein by molecular docking and molecular dynamic stimulation. We also show that nilotinib inhibited proliferation and induced senescence of cells in non-smoking female NSCLC patients in vitro.

**Conclusions:** These results not only uncover nutrient metabolic characteristics in non-smoking female NSCLC patients, but also provide a new paradigm for identifying new targets and drugs for novel therapy for such patients.

## Introduction

In 2000, it was reported that lung cancer in 15% of men and 53% of women worldwide was not the result of smoking.[1] Furthermore, Chinese women exhibit higher rates of lung cancer (20.4 cases per 100 000 women) than European women despite a lower prevalence of smoking.[2] Currently, there is increased interest in the potential for targeting nutrition metabolism.[3–7] This potential treatment strategy originated from the fact that most tumors are highly dependent on dysregulated carbohydrate,[8] lipid[9] and protein/amino acid metabolism.[10] Striking differences in the epidemiological, clinical, and molecular characteristics of lung cancers arising in never smokers versus smokers have been identified, suggesting that they are separate entities.[11] To date no systematic study on the molecular mechanism of tumor metabolism in non-smoking female NSCLC patients has been conducted. We believe that it may be feasible to use relevant bioinformatics methods to explore therapeutic targets related to NSCLC in female non-smokers from the perspective of nutritional metabolism.

*TIE1* may play a role in angiogenesis associated with tumorigenesis.[12,13] It has also been reported that *TIE1* is an independent factor that has a negative impact on the survival of gastric cancer,[14] and a cleaved form is also overexpressed in breast cancer cells.[15] Although the role of the *TIE1* in lung cancer is still unclear, studies have shown that the *TIE1* is enhanced in the developing lung compared to other organs, and the expression of *TIE1* is markedly aggravated and continues in the late stage of pulmonary

  

vascular development.[16] This indicates that *TIE1* may play an important role in the pathogenesis of lung cancer.

*NR4A1* (Nur77) is a member of the nuclear receptor superfamily of transcription factors, and the reported roles of *NR4A1* in cancer are paradoxical. Several studies have reported that *NR4A1* is overexpressed in multiple types of tumors, and is critical for survival and/or cell proliferation in pancreatic, cervical, lymphoma, lung, and colon cancer cells.[17] However, several studies have reported that *NR4A1* is a tumor suppressor gene. Low *NR4A1* expression in triple-negative breast cancer is associated with advanced tumor stages, lymph node metastasis, and poor relapse-free survival.[18] These controversial results regarding *NR4A1* highlight the possibility that *NR4A1* plays a specific role according to the type or subtype of cancer.

We conducted data mining of microarray data to identify potential therapeutic targets for female NSCLC patients. The study was based on the idea of drug repurposing: to screen FDA-approved drugs to treat non-smoking female NSCLC patients to save time and cost (Fig 1).[19]

# Methods

## Microarray data

The following bioinformatics analysis was based on the processed microarray data of E-GEOD-19804, submitted by Lu *et al.*[20] The microarray data included 60 frozen NSCLC and 60 frozen adjacent normal lung tissues from female non-smokers, and was downloaded from the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress).

Gene expression profiles of GSE31210, submitted by Kohno *et al.*,[21,22] were downloaded from the Gene Expression Omnibus (GEO: https://www.ncbi.nlm.nih.gov/geo/) as second microarray data to test the expression tendencies of differentially expressed genes (DEGs) in the E-GEOD-19804 dataset. Although the GSE31210 dataset contained 246 samples, only 8 met our analytical criteria. The eight included samples contained four frozen NSCLC and four frozen adjacent normal lung tissues from female non-smokers.

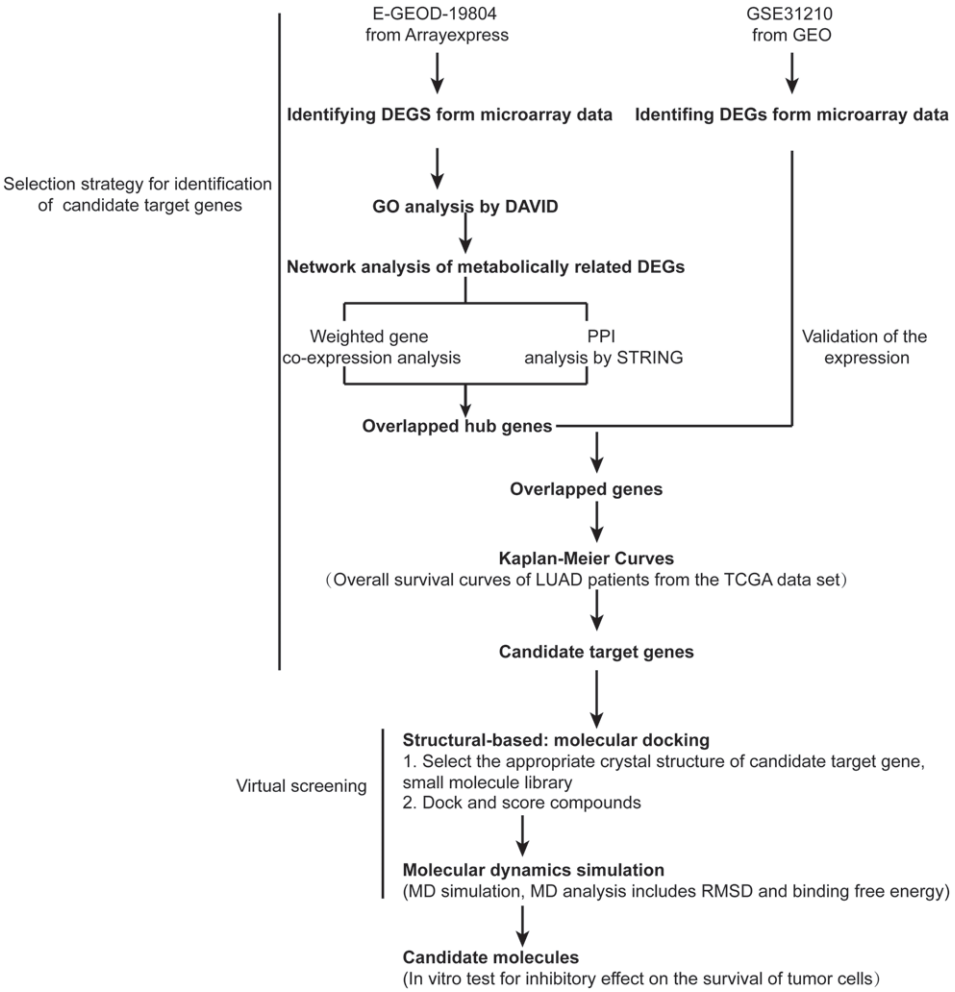


**Figure 1** Flow diagram of data mining and virtual screen process. DAVID, Database for Annotation, Visualization and Integrated Discovery; DEGs, differentially expressed genes; GEO, Gene Expression Omnibus; GO, Gene Ontology; LUAD, lung adenocarcinoma; MD, molecular dynamic; PPI, protein-protein interaction; RMSD, root mean square deviation; TCGA, The Cancer Genome Atlas.

 

## Identification of differentially expressed genes (DEGs) based on microarray data

DEG analysis was carried out using Bioconductor packages[23–25] in R software[26] with processed data of E-GEOD-19804 and GSE31210. We categorized 120 samples into two groups, which included 60 NSCLC samples as the experiment and 60 adjacent normal tissues as the control. Eight samples from GSE31210 were also categorized into two groups, including four NSCLC samples as the experiment and four adjacent normal tissues as the control. We used the Moderated T statistic method[27] to identify DEGs by Limma packages[28] in R software (R Foundation for Statistical Computing, Vienna, Austria). We finally identified DEGs with an absolute value of logarithmic fold change (logFC) $\geq 1$ and defined a cutoff of $P < 0.01$ as statistically significant.

## DEG enrichment analysis

Gene ontology (GO) analysis is a common and useful method for annotating gene products and identifying characteristic biological attributes of high-throughput genome or transcriptome data.[29] An Expression Analysis Systematic Explorer (EASE) score < 0.1 was adopted to refine the GO terms set in major clusters through Database for Annotation, Visualization and Integrated Discovery Bioinformatics Resources version 6.8 (DAVID: https://david.ncifcrf.gov/).[30,31]

## Weighted gene co-expression analysis of DEGs related to metabolism

Highly co-expressed genes are connected in the network and therefore can be grouped into modules. Within the modules, weighted gene co-expression analysis allows the identification of the central connected genes as so called "hub" genes.[32] Co-expression networks were built in an R environment[33] with a measure of similarity based on a matrix of pairwise Pearson's correlation coefficients. The similarity matrix was then transformed into an adjacency matrix using a power adjacency function. The co-expression modules were then identified by raising the soft thresholding power to 14 to produce a weighted network, which is the lowest power to guarantee that the scale-free topology fit index curve flattens out upon reaching a high value.[34]

## Protein–protein interaction (PPI) data of DEGs from STRING

The protein–protein interaction (PPI) network provides a valuable framework to better understand the functional organization of the proteome. In this network, proteins interact with several other proteins, suggesting a central regulatory role, and are likely to be regulatory "hubs."[35] The PPI network was retrieved from STRING version 9.0 (http://STRING.embl.de/). To evaluate the interactive relationships of DEGs, we mapped the 508 metabolically related DEGs to STRING, and only human experimentally-validated interactions with a combined score > 0.4 were selected as significant.

## Intersection of co-expression and PPI networks of DEGs related to metabolism

To find the genes with connections with both co-expression and PPI pairs, we intersected the data between co-expression and PPI data of DEGs related to metabolism. We then selected candidate hub genes from DEGs with "co-expression degree $\geq 1$," "STRING degree $\geq 1$," and "absolute value of logFC $\geq 1$."

## Clinical samples

We searched The Cancer Genome Atlas (TCGA) database of 389 lung adenocarcinoma (LUAD) patients on 8 July 2018 (https://cancergenome.nih.gov/) for overall survival (OS) assessment. All patients were white, 44.7% were male and 55.3% were female, including 70 non-smoking female LUAD patients. Detailed patient information including gender, age at diagnosis, tumor stage, days to death, days to last follow-up, cigarettes per day, and years smoked, is listed in Table S9.

## Identification of DEGs based on TCGA transcriptome profiling data

DEG analysis was carried out using Bioconductor packages[23–25] in R software[26] with transcriptome profiling data downloaded from TCGA. Genes with very low read counts are usually not included.[36] Thus, $\log_2$ (average count/million) was calculated to determine whether a gene was reasonably expressed or not.[36]

## Overall survival assessment of two independent cohorts

The median expression value was a cutoff to define high versus low candidate DEGs. To model survival, patients were split into two groups based on the median expression value of each candidate DEG. Gene expression at or below median was considered low expressing tumors and above median was considered high expressing tumors. Kaplan–Meier plots and landmark analysis were conducted by R[26]

  

and Empower (R) (www.empowerstats.com, X&Y Solutions, Inc., Boston, MA, USA), respectively.

## Docking experiment

*NR4A1* downregulation is associated with a poor clinical outcome, while *TIE1* downregulation is associated with a better clinical outcome. Therefore, from the perspective of drug screening, we needed to screen for small molecular agonists to increase Nur77 activity and small molecular antagonists to decrease Tie-1 activity. In order to improve the accuracy of molecular docking results, we needed an accurate binding site of the agonist/antagonist in the target protein. There are two ligand binding domains in the asymmetric subunit of Nur77, and one of the domains forms an agonist binding domain of Nur77.[37] In contrast, the protein structure of Tie-1 fibronectin type III domains (Fn3) was only purified and crystallized in 2017.[38] The Fn3 domain in Tie-1 contributes to the heterodimerization of Tie-2/Tie-1.[38] However, no reports have examined the antagonist binding site in the structure of Tie-1. Therefore, we finally selected *NR4A1* as a drug-screening target.

The crystal structure of Nur77 (PDB entry: 3V3E) in complex with glycerol was derived from the Protein Data Bank (PDB: http://www.rcsb.org/pdb/home/home.do);[39] 2748 commercially available molecules in the FDA-approved database were obtained from a subset in the ZINC database (http://zinc.docking.org/catalogs/dbap; pH 4.5~9.5).[40] Molecular docking between the compounds and Nur77 was carried out using Dock6[41] and AutoDock Vina[42] Docking scores were calculated to represent binding affinities. For the sake of brevity, the details of the docking process are presented in the supplementary files.

To further improve the accuracy of docking results, we selected the molecules that existed in the top 10% lowest GB/SA, descriptor, and autodock scores. The three-dimensional diagrams of the interactions between Nur77 and ligands were presented by UCSF Chimera (Resource for Biocomputing, Visualization, and Informatics, San Francisco, CA, USA)[43] and two-dimensional putative docking pose of candidate ligand in the binding pocket was calculated and exhibited by LigPlot$^+$ software (European Bioinformatics Institute, Hinxton, UK).[44]

## Molecular dynamic (MD) simulations

Before performing the molecular dynamic (MD) simulations, the top-ranked ligands were selected as the initial structure for the simulations. The MD simulations were performed using the GROMACS package version 4.6.7 (University of Groningen, Groningen/Netherlands; KTH Royal Institute of Technology, Stockholm, Sweden; Uppsala University, Uppsala, Sweden)[45] for 30 nanoseconds. For the sake of brevity, the details of the MD process are presented in the supplementary material.

After the MD process, root mean square deviation (RMSD) analysis was carried out using the g_rmsd programs in the GROMACS package (version 4.6.7)[45] and was performed on the entire 30 nanosecond trajectory for each receptor-ligand system. The potential, kinetic, and total energy was calculated using the g_energy program available in GROMACS. The total energy ($E_{total}$) is the sum of the potential energy ($E_{potential}$) and the kinetic energy ($E_{kinetic}$) (Eqn (1). Additionally, to intuitively compare the energy levels between the different MD simulation systems, we normalized the three energy component values as Equation (2:

$$E_{total} = E_{potential} + E_{kinetic} \tag{1}$$

$$Energy^{i}_{nor} = \frac{Energy^{i}}{SUM\left(Energy^{i}_{tar}\right)} \tag{2}$$

$Energy_{nor}$ indicates the normalized energy and $i$ indicates the specific kind of three energies. $Energy^{i}$ indicates the original value of energy. $SUM\left(Energy^{i}_{tar}\right)$ indicates the sum of the same kind of energy of Nur77-ligand complexes.

## The effect of nilotinib on proliferation

A non-smoking female NSCLC cell line, H1975, was purchased from KeyGEN Biotech. Co., Ltd. (KG342, Nanjing, China). NSCLC cell lines H1975, A549, and H1299 were used as the experimental groups, and human kidney epithelial cell line 293T as a normal control for testing the effect of nilotinib on cells. All cell lines were cultured in RPMI-1640 medium containing 10% fetal bovine serum, 100 U/mL penicillin, and 100 μg/mL streptomycin at 37°C in an atmosphere containing 5% $CO_2$. Nilotinib was purchased from TargetMol (Catalog No. T1524 CAS 641571-10-0, Shanghai, China), dissolved in dimethyl sulfoxide, and stored at −20°C. Cell viability was measured in a 96-well plate using Cell Counting Kit 8 (CCK-8 Kit, KGA317, KeyGEN Biotech. Co., Ltd.).

## Senescence β-galactosidase assay

Cells were cultured for 24 hours in 12-well plates at a density of approximately 45 000 cells/well and treated with nilotinib for 12 and 24 hours, and then fixed and stained using a senescence β-galactosidase (SA-β-gal) staining kit (C0602, Beyotime Biotechnology, Shanghai, China).

 

### Statistical analysis

All experimental data and results were confirmed in at least three independent experiments. Statistical differences were determined using multiple comparisons via GraphPad Prism 6.0 (La Jolla, CA, USA). Statistical significance was defined as $P < 0.05$. Error bars indicate the standard error of the mean unless otherwise indicated.

## Results

### Identification of DEGs

A total of 1454 DEGs were identified after analyses of E-GEOD-19804, including 976 downregulated and 478 upregulated genes (Table S1). A total of 6087 genes were identified after analyses of GSE31210, of which 3239 were upregulated and 2848 were downregulated (Table S2).

### Gene Ontology enrichment analysis of DEGs

The upregulated and downregulated genes were separately submitted to the functional annotation tool provided by DAVID. DAVID clustered 969 downregulated genes into 1951 GO terms (Table S3), and 476 upregulated genes into 784 GO terms (Table S4).

We found that the associated metabolism of downregulated and upregulated genes was different. For downregulated genes, carbohydrate metabolism accounted for only 12% of the GO terms and genes (Fig 2a,b). Whereas for lipid and protein/amino acid metabolism, although the number of genes participating in lipid metabolism was 10% lower than the genes participating in protein/amino acid metabolism (Fig 2b), the GO terms of lipid metabolism were 26% higher than that of protein/amino acid metabolism (Fig 2a). For downregulated genes, tumor metabolic changes tended to involve protein/amino acid metabolism and lipid metabolism. Most upregulated genes and their metabolic terms were mainly focused on protein/amino acid and carbohydrate metabolism (Fig 2c,d).

To analyze the relationship between metabolically related DEGs, we extracted genes with GO entries associated with carbohydrate, lipid, and protein/amino acid metabolism as metabolically related DEGs. We then removed the repeated genes contained in 106 metabolically related GO entries (Table S5) to obtain 508 metabolically related DEGs (Table S6).

### Weighted gene co-expression analysis

We constructed co-expression networks based on the selected 508 metabolically related DEGs (Table S6).

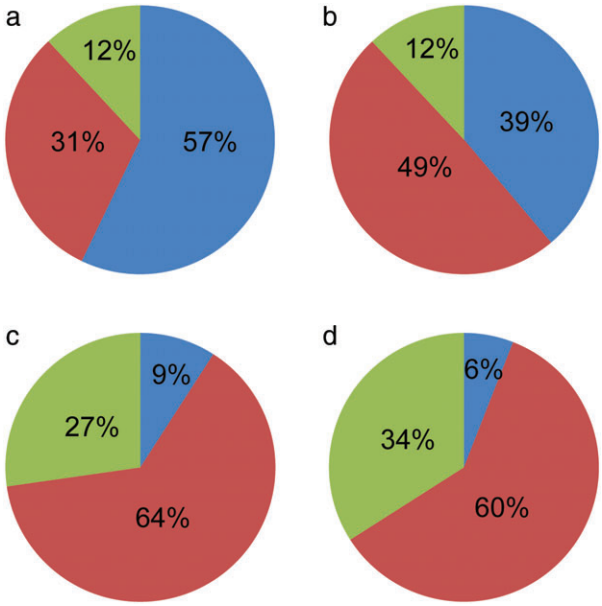


**Figure 2** The metabolically related Gene Ontology (GO) terms and gene proportions of upregulated and downregulated genes by GO functional annotation. The GO term proportions of three metabolic components for (**a**) downregulated and (**c**) upregulated genes. The proportions of gene numbers of three metabolic components for (**b**) downregulated and (**d**) upregulated genes. (■) Lipid metabolism, (■) Protein/amino acid metabolism, (■) Carbohydrate metabolism.

Consequently, 82 DEGs with 171 co-expression pairs were identified (Table S7). Our analysis found that 12 DEGs were co-expressed with other metabolically related DEGs among lipid and protein/amino acid metabolically related downregulated genes (Fig 3a). *TEK* showed the largest degree of co-expression (Fig 3b). Ten metabolic-related DEGs were co-expressed with *TEK*. The co-expression degree of *CAV1* reached 7. *FABP4*, *LOC102724428*, *NR4A1*, *SIK1*, and *TIE1* had the least co-expression degrees with only one gene association. In contrast, no gene pairs showed co-expression associations for carbohydrate and protein/amino acid metabolism in upregulated genes.

### Construction of PPI network

STRING furnished original reliable protein data for consequent analysis. In order to generate PPI networks, 426 out of 508 metabolically related DEGs were mapped. A PPI network was formed with 130 upregulated DEGs and 296 downregulated DEGs containing 426 nodes and 2404 edges (Table S8, Fig 4a). Based on the total PPI network, we then analyzed whether 12 DEGs (proteins) identified from the co-expression network interacted with other metabolically related DEGs (proteins). Interestingly, the 12 DEGs identified overlapped with 10 genes (proteins) in

  

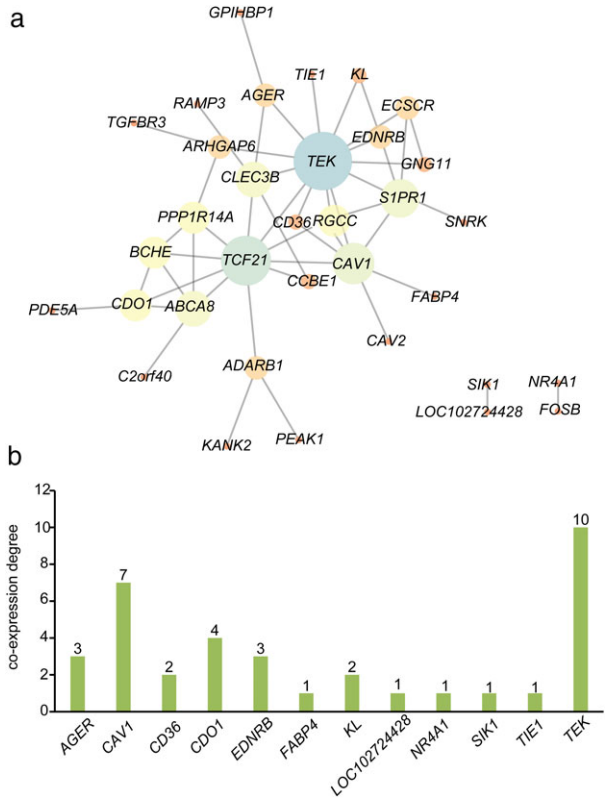


**Figure 3** Co-expression network of metabolically related differentially expressed genes (DEGs). (**a**) Co-expression network of downregulated metabolically related DEGs presented by nodes and edge by analyzer of cytoscape. The deeper the blue and the larger the node indicates more proteins that could interact with this node, while the deeper the yellow and the smaller the node indicates less protein interactors. (**b**) The exact number of co-expression degree numbers for the 12 DEGs, which are associated with both lipid and protein/amino acid metabolism.

155 PPI pairs (Fig 4b,c), including *AGER, CAV1, CD36, EDNRB, FABP4, KL, NR4A1, SIK1, TEK*, and *TIE1*.

The 10 downregulated DEGs obtained by co-expression and PPI network analyses presented consistent expression tendencies in gene expression profiles of GSE31210. That is, the 10 downregulated DEGs were reliable for further OS assessment.

## Deregulated expression of *NR4A1* and *TIE1* in tissues is associated with poor clinical outcome

We further analyzed the relationship between candidate targets and proliferation levels among LUAD and non-smoking female LUAD patients. However, because we had no access to the clinical data of the patients who provided samples for microarray experiments, we chose other independent cohorts from TCGA to further investigate the association between the expression of candidate targets and OS. To reduce the variations resulting from different NSCLC subtypes, only patients with LUAD were examined (Table S9). Kaplan–Meier plots showed no difference in OS between low and high *TIE1/NR4A1* expression groups (Figs 5a, 6a) (including smoking and non-smoking patients). However, female non-smoking LUAD patients in the high *TIE1* expression group had significantly poorer OS than the low *TIE1* expression group ($P < 0.05$) (Fig 5b).

Although no significant difference in OS ($P = 0.483$) was found between female non-smoking LUAD patients in the low and high *NR4A1* expression groups, we noted that the Kaplan–Meier curves crossover and separate after 3.89 years (Fig 6b). Thus, 3.89 years was chosen as a "landmark" to conduct further analysis. Standard OS assessment techniques before and after 3.89 years were used (Fig 6c). Although no significant difference in OS ($P = 0.84$) was found between low and high *NR4A1* expression groups during the first 3.89 years (Fig 6c), the low *NR4A1* expression group had significantly poorer OS after 3.89 years ($P < 0.05$) (Fig 6c). The Kaplan–Meier plots of the other eight candidate targets are presented in Figure S1, which shows no difference in the OS of non-smoking female LUAD patients between low and high candidate expression target groups. Accordingly, *NR4A1* and *TIE1* may be identified as potential therapeutic targets for NSCLC in female non-smokers.

## Screening of potential ligands targeting Nur77 by molecular docking and AutoDock Vina

Based on an RMSD value < 2 Å, which is the standard criterion for successful molecular docking,[46] we ranked compounds by descriptor and Hawkins GB/SA scores, respectively. The detailed screening information of compounds with the top 10% lowest score are presented in Table S10. After overlapping the lists of AutoDock Vina and two Dock6 scoring, we identified the top 6 scored compounds that can target the agonist binding site of Nur77. These compounds may have high potential to serve as agonists for Nur77. The detailed docking results of these six compounds and crystal ligand are listed in Table 1.

## Structural stability of the backbone of three models during MD simulations

We selected the top 2 scoring Nur77 candidates for further MD simulations. For comparison, we also performed 30 nanosecond Nur77-glycerol MD simulation. The

 

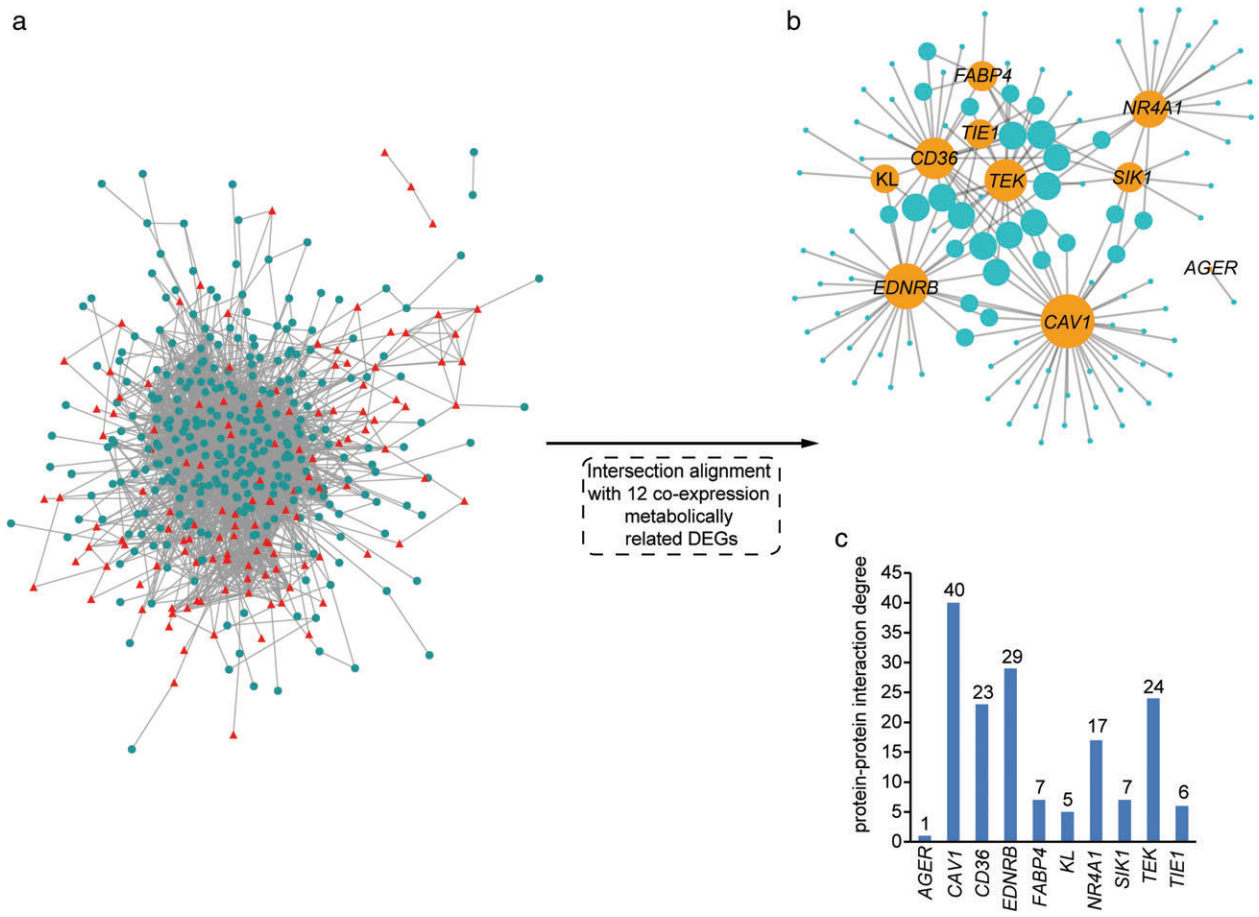


**Figure 4** Protein-protein interaction (PPI) network of metabolically related differentially expressed genes (DEGs). (**a**) PPI networks of 426 out of 508 metabolically related DEGs. The red triangle nodes represent upregulated genes and the green circle represents downregulated genes. The gray edge indicates the protein interaction between both ends of the edge. (**b**) The PPI network is represented by a node and edge graph. The 10 orange nodes indicate that the genes overlapped with the 12 DEGs, which were co-expressed with other metabolically related DEGs. The blue nodes indicate other metabolically related DEGs. The larger node indicates more interaction pairs. (**c**) The exact number of PPI degree numbers for 10 DEGs.

backbone RMSD values with respect to the X-ray structure were calculated to assess the stability of the three protein-ligand binding models during MD simulations. Figure 7a shows that three binding systems of Nur77 remained stable with average RMSD fluctuations of 0.179 nm (Nur77-ZINC64622551), 0.233 nm (Nur77-ZINC06716957), and 0.216 nm (Nur77-glycerol) during the 30 nanosecond MD simulation process.

## Binding free energy of MD simulation systems

To intuitively compare the energy levels between different MD simulation systems, we normalized the three energy component values. The Nur77-ZINC06716957 binding system presented the lowest value of the three types of energy (Table 2). Figure 7b also shows that ZINC06716957 had the highest normalized value of potential and total energy, and the lowest normalized value of kinetic energy in Nur77-ligand binding systems, indicating that ZINC06716957 might have a better affinity than other ligands.

## Nilotinib inhibits cell proliferation and induces senescence in H1975 cells

ZINC06716957 refers to the old drug nilotinib, which is a small molecular tyrosine kinase inhibitor approved for the treatment of imatinib-resistant chronic myelogenous leukemia.[47] To examine the effect of nilotinib on NSCLC cells, human NSCLC lines H1975, H1299, and A549 were treated with different concentrations of nilotinib for 24 hours, and the cell viability was measured by CCK-8 assay. As shown in Figure 8a, nilotinib

  

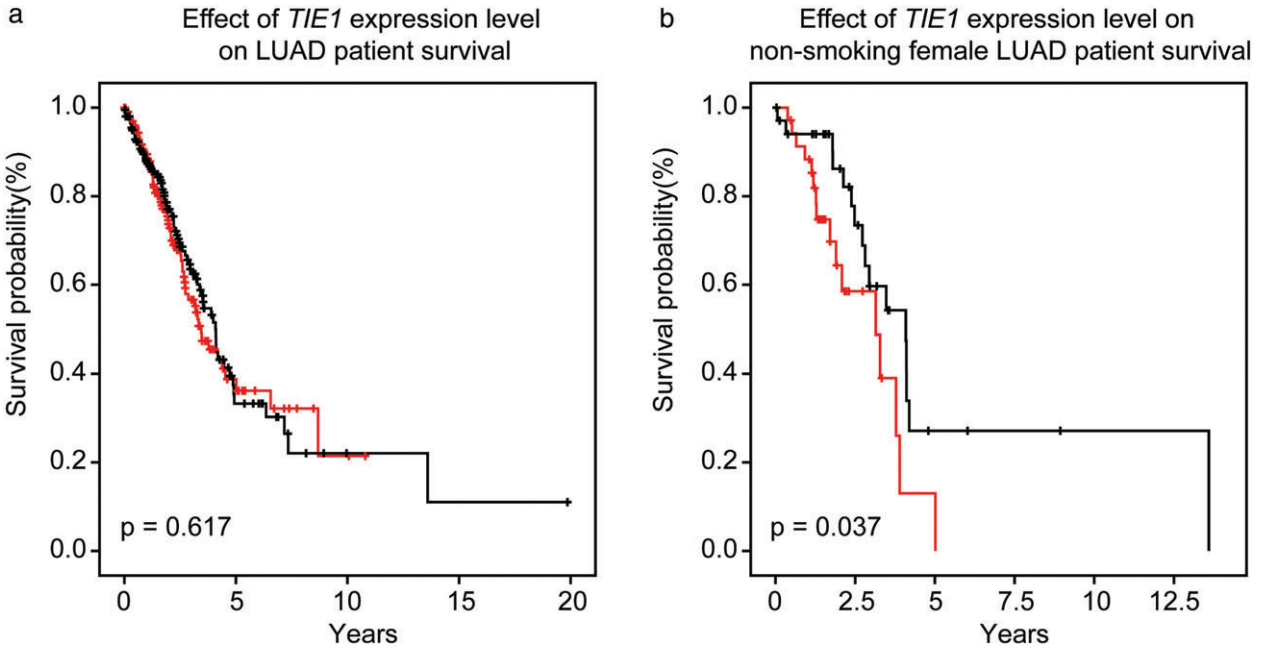


**Figure 5** Kaplan–Meier estimates of *TIE1* expression with clinical outcome in two independent cohorts of lung adenocarcinoma (LUAD) patients. (**a**) LUAD patients include smoking and non-smoking patients ($n$ = 406). (+) High *TIE1* expression ($n$ = 203), and (+) low *TIE1* expression ($n$ = 203). (**b**) Non-smoking female LUAD patients ($n$ = 70), (+) high *TIE1* expression ($n$ = 35), (+) low *TIE1* expression ($n$ = 35).

treatment markedly decreased cell viability in a dose-dependent manner in both H1975 and H1299 cells, but had less cytotoxic effect in 293T and A549 cells. We observed a characteristic change in H1975 cell shape to an enlarged and flattened phenotype, which is reminiscent of senescence.[48] Consistently, SA-β-gal assays showed that nilotinib significantly enhanced senescence in dose and time-dependent manners, as indicated by a significant increase in the percentage of SA-β-gal-positive cells (Fig 8b,c).

## Discussion

Network-based biological analysis is a promising approach to uncover key genes and biological processes from a network viewpoint that cannot be recognized from individual gene-based signatures.[49,50] Using weighted gene co-expression analysis methodology and constructing a PPI network, we found that 10 downregulated genes in the metabolic-related DEGs had both co-expressive and interactive relationships with other metabolic-related DEGs, respectively. In addition to the GSE31210 dataset, Lee *et al.* provided an RNA-sequencing study of LUAD and adjacent normal tissues of six female Korean never-smoker patients.[51] These 10 downregulated DEGs presented consistent expression tendencies in gene expression profiles of GSE31210 and RNA-seq study. That is, the 10 downregulated DEGs were reliable for further survival analyses.

Although a number of agonists targeting Nur77 have been reported in the database,[52] they have not been approved as anti-tumor clinical drugs. There are more than 2000 US Food and Drug Administration approved drugs available in the DrugBank, with diverse structural types and well-known pharmacological features. Re-evaluation of these drugs might identify a new use for the treatment of female non-smoking NSCLC patients with low *NR4A1* expression. We identified nilotinib, which combines well into the agonist-binding pocket of Nur77 (Fig 7c). Although nilotinib was less stable than glycerol during the 30 nanosecond MD simulation process (Fig 7a), the results of three different energies (Fig 7b) showed that nilotinib had higher affinity than glycerol. Additionally, nilotinib formed two hydrogen bonds with Thr48 outside of binding pockets (Fig 7d), which improved the binding affinity of nilotinib with Nur77. Therefore, nilotinib may be a promising new agonist targeting Nur77.

Striking differences in response rates to EGFR-tyrosine kinase inhibitors have been observed between lung cancer in never smokers and smokers. It has been reported that gefitinib, an EGFR-TKI, responds approximately fourfold higher in non-smoking metastatic NSCLC.[53,54] Similar differential effects of EGFR-TKIs

 

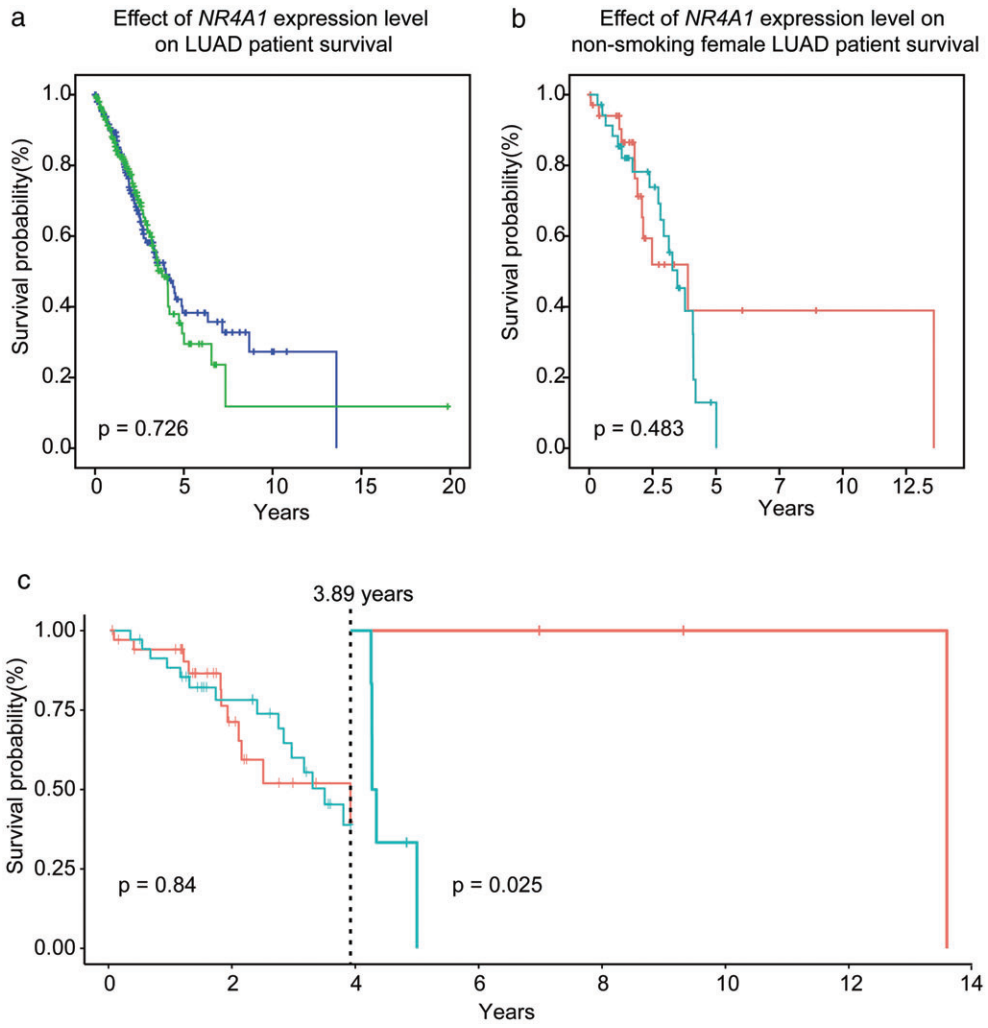


**Figure 6** Kaplan–Meier estimates of *NR4A1* expression with clinical outcome in two independent cohorts of lung adenocarcinoma (LUAD) patients. (**a**) LUAD patients include smoking and non-smoking patients ($n$ = 406). (+) High *NR4A1* expression ($n$ = 203), (+) low *NR4A1* expression ($n$ = 203). (**b**) Non-smoking female LUAD patients ($n$ = 70). (+) High *NR4A1* expression ($n$ = 35), (+) low *NR4A1* expression ($n$ = 35). (**c**) Landmark analysis discriminating between events occurring before and after 3.89 years in non-smoking female LUAD patients ($n$ = 70). (+) High *NR4A1* expression ($n$ = 35), (+) low *NR4A1* expression ($n$ = 35).

**Table 1** The docking results of the six identified compounds

| Compounds | Descriptor score (kcal/mol) | Hawkins GB/SA score (kcal/mol) | AutoDock score (kcal/mol) |
|---|---|---|---|
| ZINC06716957 | −52.19 | −46.90 | −7.70 |
| ZINC64622551 | −48.40 | −52.43 | −7.30 |
| ZINC12503300 | −48.67 | −43.33 | −7.50 |
| ZINC12503303 | −49.03 | −46.40 | −7.50 |
| ZINC13818943 | −49.03 | −47.52 | −7.60 |
| ZINC19796084 | −48.61 | −45.17 | −7.40 |
| Glycerol | −23.15 | −23.25 | −3.80 |

Glycerol is the compound in complex with ligand-binding domain in the crystal structure of Nur77.

have also been observed in subgroup analyses from large, multicenter, randomized studies of gefitinib and erlotinib.[55,56] However, almost all patients with NSCLC acquire resistance to EGFR-TKIs after a period of time, leading to disease progression.[57–59] Thus, the identification of new therapeutic targets for non-smoking with NSCLC and the development of new drugs is critical. However, bringing new drugs to the market is a lengthy and expensive process. Nilotinib significantly inhibits the proliferation of H1975 cells in a dose-dependent manner, and significantly induces the senescence of H1975 cells in time and dose-dependent manners. The bioinformatics analysis methods and results detailed herein could potentially provide a new paradigm for

  

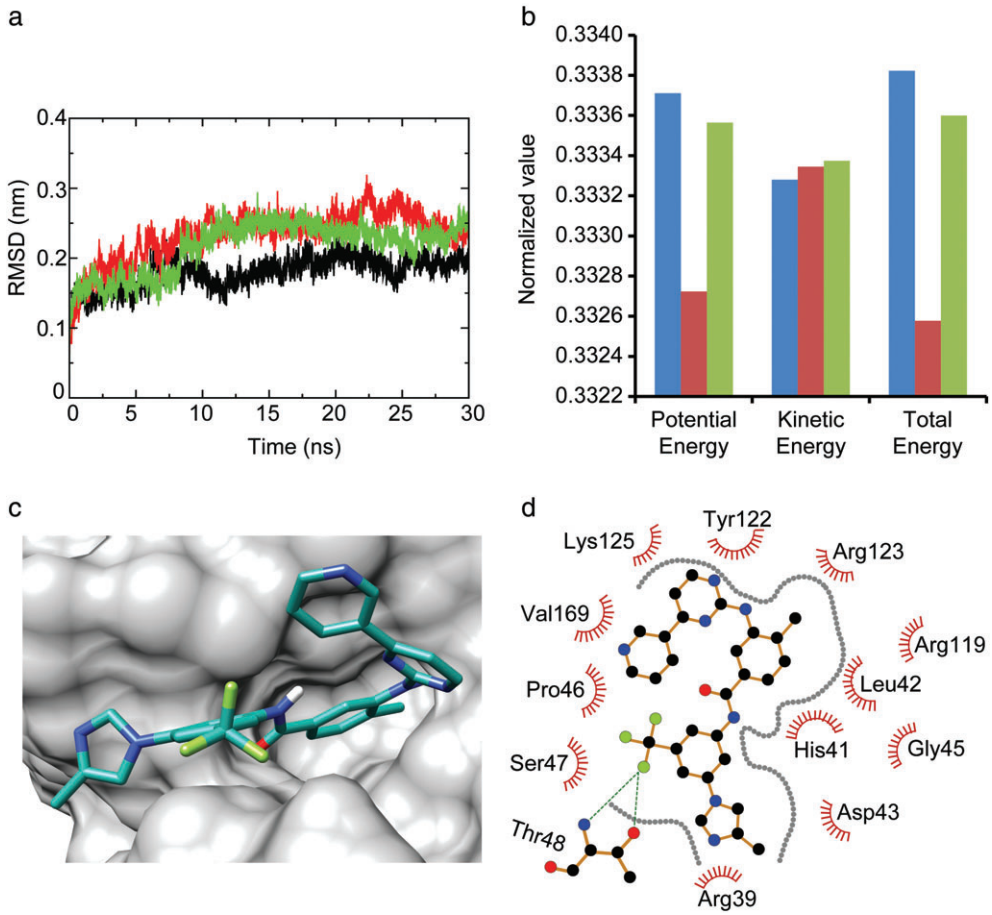


**Figure 7** (**a,b**) Structural analysis of molecular dynamic (MD) simulation trajectories of Nur77-ligand complexes and (**c,d**) diagrams of the interactions between Nur77 and ZINC06716957 (nilotinib). (**a**) Comparison of backbone root mean square deviation (RMSD) values for Nur77-ligand complexes. (——) Nur77-ZINC64622551, (——) Nur77-ZINC06716957, (——) Nur77-glycerol. (**b**) Comparison of the normalized potential energy, kinetic energy and total energy values in Nur77-ligand complexes. (■) Nur77-ZINC6716957, (■) Nur77-ZINC64622551, (■) Nur77-glycerol. The ordinate represents the normalized value of three kinds of energy. For the kinetic energy normalized value, the lower the value, the higher the affinity of Nur77-ligand binding complexes, while the higher the value, the higher the affinity of Nur77-ligand binding complexes for potential and total energy. (**c**) The binding model of Nur77-nilotinib. The ligand is displayed as stick model, and the receptor is represented as a dark gray surface model. (**d**) Hydrophobic interactions are represented by dark red arcs. Hydrogen bonds are indicated by broken green lines. Ligands are represented in dark golden rod. C, N, O, and F atoms are shown in black, blue, red, and green, respectively. The gray circle indicates the general outline of the binding pocket.

**Table 2** The calculated binding free energies of protein-ligand associations

| | Average energy (kJ/mol) | | | | | |
|---|---|---|---|---|---|---|
| MD simulation systems | Potential energy | Potential energy† | Kinetic energy | Kinetic energy† | Total energy | Total energy† |
| Nur77-ZINC06716957 | −760 337 | 0.333 711 811 | 138 799 | 0.333 280 508 | −621 579 | 0.333 822 95 |
| Nur77-ZINC64622551 | −758 085 | 0.332 723 409 | 138 826 | 0.333 345 339 | −619 259 | 0.332 576 979 |
| Nur77-Glycerol | −760 002 | 0.333 564 78 | 138 838 | 0.333 374 153 | −621 164 | 0.333 600 071 |

†Indicates the normalized binding free energies of protein-ligand associations.

identifying new targets and drugs as novel therapy for non-smoking female NSCLC patients. Future work should concentrate on in vivo experiments of nilotinib to determine the mechanism of antitumor effect.

## Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No. 81373311, 81173093,

 

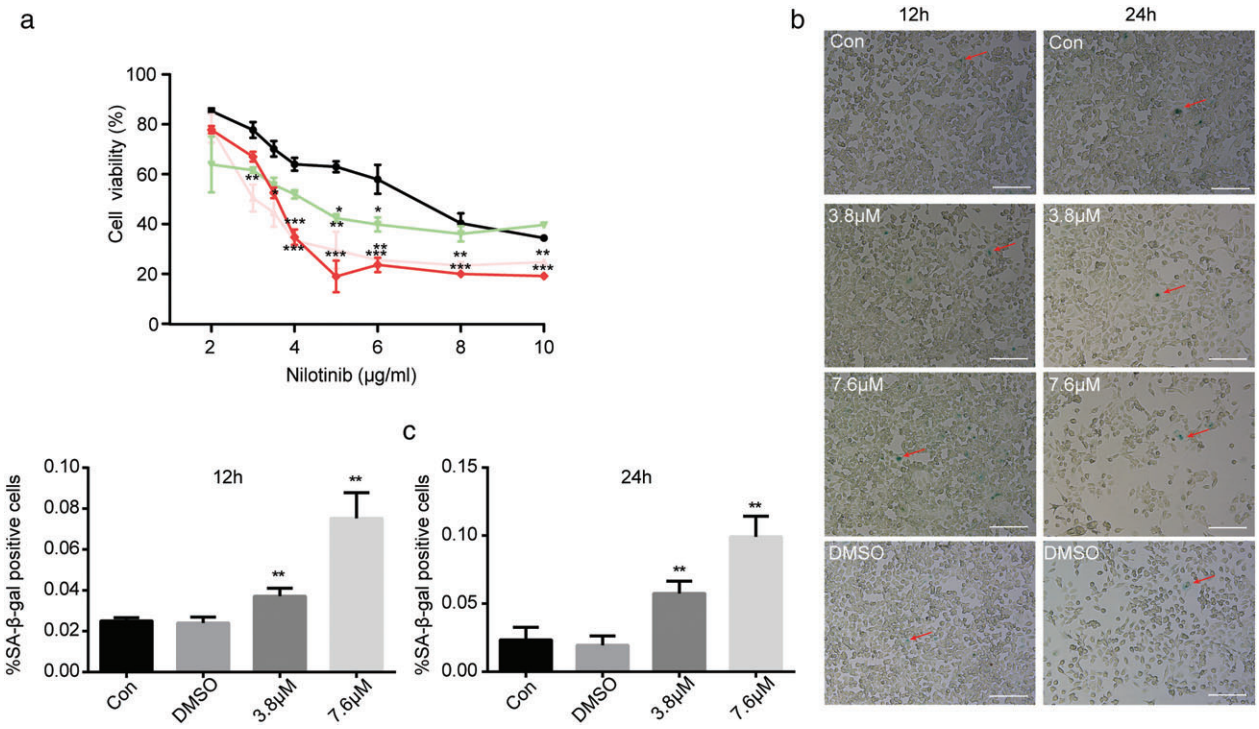


**Figure 8** Nilotinib inhibits cell proliferation and induces senescence in H1975 cells. (**a**) Cell viability was measured using Cell Counting Kit 8 assays in 293T, A549, H1299, and H1975 cells treated with the indicated concentrations of nilotinib for 24 hours. (—●—) 293T (IC50 = 12.6 μm), (—▲—) H1975 (IC50 = 6.1 μm), (—▼—) A549 (IC50 = 8.3 μm), (—◆—) H1299 (IC50 = 6.6 μm). (**b**) Senescence β-galactosidase (SA-β-gal) staining showed increased β-gal positivity (blue cells, as indicated by arrow) with nilotinib treatment compared to dimethyl sulfoxide (DMSO) and control groups (white scale bars represent 200 μm for all micrographs) (**c**) Quantification of β-gal positivity shows an increased percentage of aged cells in the nilotinib treated group. $0.01 < *P < 0.05$, $0.001 < **P < 0.01$, $***P < 0.001$. IC50, half maximal inhibitory concentration.

30970643, 31300674, and J1103518); and the Youth Science Foundation of West China Hospital of Stomatology (No. 2017-3).

## Disclosure

No authors report any conflict of interest.

## References

1 Frank F, Pereira JR, Joachim VP *et al.* Randomized, multinational, phase III study of docetaxel plus platinum combinations versus vinorelbine plus cisplatin for advanced non-small-cell lung cancer: The TAX 326 study group. *J Clin Oncol* 2003; **21**: 3016–24.
2 Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015; **65**: 87–108.
3 Vander Heiden MG. Targeting cancer metabolism: A therapeutic window opens. *Nat Rev Drug Discov* 2011; **10**: 671–84.
4 Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: The metabolic requirements of cell proliferation. *Science* 2009; **324**: 1029–33.
5 Lorenzo G, Oliver K, Heiden MGV *et al.* Metabolic targets for cancer therapy. *Nat Rev Drug Discov* 2013; **12**: 829–46.
6 Pathania D, Neamati MN. Opportunities in discovery and delivery of anticancer drugs targeting mitochondria and cancer cell metabolism. *Adv Drug Deliv Rev* 2009; **61**: 1250–75.
7 Tennant DA, Durán RV, Eyal G. Targeting metabolic transformation for cancer therapy. *Nat Rev Cancer* 2010; **10**: 267–77.
8 Muñoz-Pinedo C, El Mjiyad N, Ricci JE. Cancer metabolism: Current perspectives and future directions. *Cell Death Dis* 2012; **3**: e248.
9 Santos CR, Schulze A. Lipid metabolism in cancer. *FEBS J* 2012; **279**: 2610–23.
10 Landel AM, Hammond WG, Meguid MM. Aspects of amino acid and protein metabolism in cancer-bearing states. *Cancer* 2015; **55**: 230–7.
11 Sophie S, Joan HS, Adi FG. Lung cancer in never smokers: A different disease. *Nat Rev Cancer* 2007; **7**: 778–90.
12 Sato TN, Qin Y, Kozak CA, Audus KL. Tie-1 and tie-2 define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system. *Proc Natl Acad Sci U S A* 1993; **90**: 9355–8.

  

13 Korhonen J, Polvi A, Partanen J *et al.* The mouse tie receptor tyrosine kinase gene: Expression during embryonic angiogenesis. *Oncogene* 1994; **12**: 395–403.
14 Lin WC, Li AF, Chi CW *et al.* Tie-1 protein tyrosine kinase: A novel independent prognostic marker for gastric cancer. *Clin Cancer Res* 1999; **5**: 1745–51.
15 Yang XH, Hand RA, Livasy CA, Cance WG, Craven RJ. Overexpression of the receptor tyrosine kinase Tie-1 intracellular domain in breast cancer. *Tumor Biol* 2003; **24**: 61–9.
16 Taichman DB, Schachtner SK, Yixun L *et al.* A unique pattern of Tie1 expression in the developing murine lung. *Exp Lung Res* 2003; **29**: 113–2.
17 Syng Ook L, Xi L, Shaheen K *et al.* Targeting NR4A1 (TR3) in cancer cells and tumors. *Expert Opin Ther Targets* 2011; **15**: 195–206.
18 Wu H, Bi J, Yan P *et al.* Nuclear receptor NR4A1 is a tumor suppressor down-regulated in triple-negative breast cancer. *Oncotarget* 2017; **8**: 54364–77.
19 Chong CR, Sullivan DJ Jr. New uses for old drugs. *Nature* 2007; **448**: 645–6.
20 Lu TP, Tsai MH, Lee JM *et al.* Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev* 2010; **19**: 2590–7.
21 Hirokazu O, Takashi K, Yuko I *et al.* Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* 2012; **72**: 100–11.
22 Mai Y, Rui Y, Asuka N *et al.* Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLOS One* 2012; **7**: e43923.
23 Huber W, Carey VJ, Gentleman R *et al.* Orchestrating high-throughput genomic analysis with bioconductor. *Nat Methods* 2015; **12**: 115–21.
24 Kannan L, Ramos M, Re A *et al.* Public data and open source tools for multi-assay genomic investigation of disease. *Brief Bioinform* 2016; **17**: 603–15.
25 Lawrence M, Huber W, Pagã SH *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013; **9**: e1003118.
26 R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria 2013. URL http://www.R-project.org/.
27 Yu L, Gulati P, Fernandez S *et al.* Fully moderated T-statistic for small sample size gene expression arrays. *Stat Appl Genet Mol Biol* 2011; **10**: 1–22.
28 Ritchie ME, Belinda P, Di W *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**: e47.
29 Ashburner M, Ball CA, Blake JA *et al.* Gene ontology: Tool for the unification of biology. *Nat Genet* 2000; **25**: 25–9.
30 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**: 44–57.
31 Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**: 1–13.
32 Liu W, Huang X, Liang X *et al.* Identification of key modules and hub genes of keloids with weighted gene coexpression network analysis. *Plast Reconstr Surg* 2017; **139**: 391–2.
33 Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinf* 2008; **9**: 559.
34 Plachetzki DC, Sabrina PM, Johnson BR *et al.* Gene co-expression modules underlying polymorphic and monomorphic zooids in the colonial hydrozoan, *Hydractinia symbiolongicarpus*. *Integr Comp Biol* 2014; **54**: 276–83.
35 Scardoni G, Laudanna C. Centralities based analysis of complex networks. In: *New Frontiers in Graph Theory*. InTechOpen, London 2012, Rijeka, Croatia.
36 Robinson MD, McCarthy DJ, Smyth GK. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; **26**: 139–40.
37 Zhan YY, Chen Y, Zhang Q *et al.* The orphan nuclear receptor Nur77 regulates LKB1 localization and activates AMPK. *Nat Chem Biol* 2012; **8**: 897–904.
38 Leppanen VM, Saharinen P, Alitalo K. Structural basis of Tie2 activation and Tie2/Tie1 heterodimerization. *Proc Natl Acad Sci U S A* 2017; **114**: 4376–81.
39 Berman HM, Westbrook J, Feng Z *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000; **28**: 235–42.
40 Irwin JJ, Shoichet BK. ZINC-a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005; **45**: 177–82.
41 Allen WJ, Balius TE, Mukherjee S *et al.* DOCK 6: Impact of new features and current docking performance. *J Comput Chem* 2015; **36**: 1132–56.
42 Oleg T, Arthur JO. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010; **31**: 455–61.
43 Pettersen EF, Goddard TD, Huang CC *et al.* UCSF Chimera: A visualization system for exploratory research and analysis. *J Comput Chem* 2004; **25**: 1605–12.
44 Laskowski RA, Swindells MB. LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* 2011; **51**: 2778–86.
45 Sander P, Szilárd P, Roland S *et al.* GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 2013; **29**: 845–54.
46 Brozell SR, Sudipto M, Balius TE *et al.* Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J Comput Aided Mol Des* 2012; **26**: 749–73.
47 Deremer DL, Ustun C, Natarajan K. Nilotinib: A second-generation tyrosine kinase inhibitor for the treatment of chronic myelogenous leukemia. *Clin Ther* 2008; **30**: 1956–75.
48 Sebastian H, Andreas B, Christian A *et al.* Vemurafenib induces senescence features in melanoma cells. *J Invest Dermatol* 2013; **133**: 1601–9.

 

49 Calvano SE, Wenzhong X, Richards DR *et al.* A network-based analysis of systemic inflammation in humans. *Nature* 2005; **437**: 1032–7.
50 Fu Dong Y, Shao You Y, Yuan Yuan L *et al.* Co-expression network with protein-protein interaction and transcription regulation in malaria parasite *Plasmodium falciparum*. *Gene* 2013; **518**: 7–16.
51 Kim SC, Jung Y, Park J *et al.* A high-dimensional, deep-sequencing study of lung adenocarcinoma in female never-smokers. *PLOS One* 2013; **8**: e55596.
52 Zhan Y, Du X, Chen H *et al.* Cytosporone B is an agonist for nuclear orphan receptor Nur77. *Nat Chem Biol* 2008; **4**: 548–56.
53 Miller VA, Kris MG, Neelam S *et al.* Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non-small-cell lung cancer. *J Clin Oncol* 2004; **22**: 1103–9.
54 Lim ST, Wong EH, Chuah KL *et al.* Gefitinib is more effective in never-smokers with non-small-cell lung cancer: Experience among Asian patients. *Br J Cancer* 2005; **93**: 23–8.
55 Shepherd FA, Rodrigues Pereira J, Ciuleanu T *et al.* Erlotinib in previously treated non–small-cell lung cancer. *N Engl J Med* 2005; **353**: 123–32.
56 Herbst RS, Prager D, Hermann R *et al.* TRIBUTE: A phase III trial of erlotinib hydrochloride (OSI-774) combined with carboplatin and paclitaxel chemotherapy in advanced non-small-cell lung cancer. *J Clin Oncol* 2005; **23**: 5892–9.
57 Camidge DR, Pao W, Sequist LV. Acquired resistance to TKIs in solid tumours: Learning from lung cancer. *Nat Rev Clin Oncol* 2014; **11**: 473–81.
58 Jie W. Screening for epidermal growth factor receptor mutations in lung cancer. *N Engl J Med* 2010; **361**: 958–67.
59 Oxnard GR, Arcila ME, Sima CS *et al.* Acquired resistance to EGFR tyrosine kinase inhibitors in EGFR-mutant lung cancer: Distinct natural history of patients with tumors harboring the T790M mutation. *Clin Cancer Res* 2011; **17**: 1616–22.


## Supporting Information

Additional Supporting Informationmay be found in the online version of this article at the publisher's website:

**Appendix S1.** Supplementary method details of docking and molecular dynamic (MD) simulation process.

**Figure S1.** Kaplan–Meier estimates of the target gene expression of eight candidates with clinical outcome in non-smoking female lung adenocarcinoma (LUAD) patients ($n$ = 70).

**Table S1.** Finally identified differentially expressed genes (DEGs) of E-GEOD-19804 dataset by R software.

**Table S2.** Finally identified differentially expressed genes (DEGs) of GSE31210 by R software.

**Table S3.** Gene functional annotation results of downregulated genes provided by Database for Annotation, Visualization and Integrated Discovery (DAVID).

**Table S4.** Gene functional annotation results of upregulated genes provided by Database for Annotation, Visualization and Integrated Discovery (DAVID).

**Table S5.** Nutrition metabolic associate Gene Ontology (GO) terms of differentially expressed genes (DEGs).

**Table S6.** The 508 selected metabolic-related differentially expressed genes (DEGs) for gene co-expression analysis.

**Table S7.** The results of co-expression pairs.

**Table S8.** The results of protein-protein interaction pairs.

**Table S9.** Clinical characteristics of lung adenocarcinoma (LUAD) patients.

**Table S10.** The information of ranked top 10% lowest score compounds.

  