RHO Guanine Nucleotide Exchange Factors Predict Prognosis of Non-Small Cell Lung Cancer: A Comprehensive Bioinformatics Analysis

CURRENT STATUS: POSTED

Rui-Jie Zeng  
Shantou University Medical College

Chun-Wen Zheng  
Shantou University Medical College

Yi-Wei Xu  
Shantou University Medical College

Li-Yan Xu  
Shantou University Medical College

En-Min Li  
Shantou University Medical College

nmli@stu.edu.cn  
Corresponding Author

ORCiD: https://orcid.org/0000-0001-6375-3614

DOI: 10.21203/rs.3.rs-18164/v1

SUBJECT AREAS  
Cancer Biology

KEYWORDS  
RHO GEFS, ABR, PREX1, DOCK2, DOCK4, NSCLC
Abstract

Background: Conventionally, RHO GEFs are known as activators for RHO GTPases which promote tumorigenesis. However, the role of RHO GEFs in non-small cell lung cancer (NSCLC) remains largely unknown.

Methods: A comprehensive bioinformatics analysis of protein structure, transcriptional expression, survival, methylation, mutation and gene-set enrichment data was performed using multiple databases.

Results: Through the screening of 81 RHO GEFs for their expression profiles and correlations with survival, four of them are identified with strong significance for predicting the prognosis of NSCLC patients. The four RHO GEFs, namely ABR, PREX1, DOCK2 and DOCK4, are downregulated in NSCLC compared to normal tissue. The downregulation of ABR, PREX1, DOCK2 and DOCK4, which can be contributed by promoter methylation, is correlated with unfavorable prognosis. Moreover, the underexpression of the four key RHO GEFs upregulates MYC signaling and DNA repair pathways, leading to carcinogenesis and poor prognosis.

Conclusions: The data unveil the unprecedented role of ABR, PREX1, DOCK2 and DOCK4 as tumor suppressor in NSCLC. The previously unnoticed functions of RHO GEFs in NSCLC will inspire researchers to investigate the distinct roles of RHO GEFs in cancers, in order to provide critical strategies in clinical practice.

1. Background

Lung cancer is one of the most common malignancies and represents a leading cause of cancer deaths (> 1 million annually) worldwide [1, 2]. Despite therapeutic advances, the prognosis of lung cancer remains unfavorable, and more than half of the patients diagnosed with lung cancer die within one year [3]. Non-small cell lung cancer (NSCLC) constitutes approximately 85% of all lung cancers, of which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most common subtypes [3, 4]. A comprehensive understanding of genetic alterations and associated mechanisms in the development of NSCLC is required to more effectively predict the prognosis of patients and identify druggable targets in cancer therapeutics.
RHO guanosine triphosphatases (GTPases) are essential molecular switches involved in the regulation of numerous downstream pathways in various types of cancer [5, 6]. The cycling between guanosine triphosphate (GTP)-bound (active) state and guanosine diphosphate (GDP)-bound (inactive) state contributes to the activation or inactivation of downstream effectors [7]. In the “on” state (GTP-bound) of RHO GTPases, target proteins are recognized and subsequently a response is generated until the RHO GTPases are turned to “off” state (GDP-bound) [5]. Ras-related C3 botulinum toxin substrate 1 (RAC1), Ras homolog family member A (RHOA) and cell division control protein 42 homolog (CDC42), as the most important and extensively-studied members of RHO GTPases, have been identified to regulate actin cytoskeleton reorganization, migration, metastasis and promote the development of lung cancer [8-11].

The activity of RHO GTPases is principally regulated by three types of regulatory proteins, including guanine nucleotide exchange factors (GEFs) for activation, GTPase-activating proteins (GAPs) for inactivation and GDP-dissociation inhibitors (GDIs) for GDP-bound form stabilization [12]. GEFs catalyze GDP release and subsequently assist in its binding of GTP, converting the inactive RHO GTPases to their active states for further functioning [13]. As activators of RHO GTPases, RHO GEFs has attracted the attention of researchers in recent years. RHO GEFs can be divided into two distinct families: diffuse B-cell lymphoma (DBL) family and dedicator of cytokinesis (DOCK) family [14]. There are 70 members of DBL family GEFs and 11 members of DOCK family GEFs [14, 15]. The altered expression or mutation of RHO GEFs has been identified in human cancers [16]. However, compared to the roles of RHO GTPase family members, the roles of RHO GEF family members in NSCLC remain largely unclear. Consequently, exploring the altered expression and relevant mechanisms of RHO GEFs leading to NSCLC development is significant.

Herein, we analyze the expressions of RHO GEFs and their correlations with clinical parameters in NSCLC patients. Among 81 members of RHO GEF family, four of the members, namely active breakpoint cluster region-related (ABR), phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1 (PREX1), dedicator of cytokinesis protein 2 (DOCK2) and dedicator of cytokinesis protein 4 (DOCK4) are selected for their observably altered expression and significant prediction values for
the prognosis in NSCLC patients. Moreover, methylation and mutation profiles are analyzed to interpret the observed altered expression of the selected members. Potential mechanisms related to ABR, PREX1, DOCK2 and DOCK4 in NSCLC development are also investigated.

2. Materials And Methods

2.1 Phylogenetic analysis

The protein sequences of human DBL family RHO GEFs (70 members) and human DOCK family RHO GEFs (11 members) were obtained from UniProt (https://www.uniprot.org/). The phylogenetic tree was constructed by MEGA7.0 software using neighbor-joining method with default parameters and 1000 bootstrap replicates [17]. The final tree containing the information of domain organization was visualized by the online tool Evolview [18].

2.2 Oncomine database analysis

Oncomine (https://www.oncomine.org/) is an online mining platform by which the expression analyses on comparing the transcriptome data in most types of cancer with their corresponding normal tissues can be performed [19, 20]. The mRNA expression levels of genes encoding RHO GEFs in NSCLC tissues, represented by LUAD and LUSC, as well as normal lung tissues were analyzed by Oncomine. In this study, P-value < 0.05, fold change > 2.0 and top 5% gene rank were selected as the thresholds. The resulting data were input into and visualized by Microsoft Office Excel 16.0 software (Microsoft Corporation, Redmond, CA).

2.3 Gene Expression Profiling Interactive Analysis (GEPIA) database analysis

GEPIA (http://gepia.cancer-pku.cn/) is an online tool containing differential expression analysis between tumors and normal tissues based on The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) data [21]. To date, TCGA has produced RNA-Seq data including 9736 tumor samples across 33 cancer types, as well as the data containing 726 adjacent normal tissues. The GTEx project contains RNA-Seq data for more than 8000 normal samples from unrelated donors. GEPIA integrates the information from cancer genomics big data for end users. The GEPIA database was used to validate the transcriptional profiles of RHO GEFs in NSCLC patients.

2.4 UALCAN database analysis

UALCAN (http://ualcan.path.uab.edu/) is a web server using TCGA RNA-seq and clinical data from 31 cancer types [22]. This database provides a platform for identifying candidate biomarkers specific for
tumor sub-groups. The UALCAN database was utilized to analyze the mRNA expression of normal tissues and NSCLC specimens from different sub-groups based on nodal metastasis status and tumor stages.

2.5 Kaplan-Meier Plotter database analysis
Kaplan-Meier Plotter (https://kmplot.com/) is an online database capable for assessing the effects of 54,000 genes on survival in 21 types of cancer. This system includes gene chip and RNA-seq data from Gene Expression Omnibus (GEO), European Genome-Phenome Archive (EGA) and TCGA database. In this database, data of lung cancer are available [23]. The patient specimens were divided into high expression and low expression groups using JetSet best probe set and auto-selected best cutoff [24]. Outlier arrays were excluded to control array quality [25]. Overall survival (OS) and first progression (FP) were analyzed in NSCLC patients. Log-rank P-value and hazard ratio (HR) with 95% confidence interval (CI) were calculated and displayed on the web server. The Kaplan-Meier Plotter database was used to evaluate the prognostic value of genes encoding RHO GEFs in NSCLC.

2.6 MEXPRESS database analysis
MEXPRESS (https://mexpress.be/) is an online data visualization tool for the visualization and integration of TCGA expression, DNA methylation and clinical data [26, 27]. The MEXPRESS database was employed to investigate the promoter methylation status of selected genes in NSCLC specimens compared to normal tissues, and only CpG sites with statistically significant results were reported.

2.7 cBioPortal for Cancer Genomics (cBioPortal) database analysis
cBioPortal (http://cbioportal.org) is a web resource for the exploration, visualization and analysis of cancer genomics data [28]. The datasets selected in our study were LUAD (TCGA, Firehose Legacy, containing 580 samples) and LUSC (TCGA, Firehose Legacy, containing 503 samples). The platform for methylation sequencing information was Illumina Human Methylation 450 (HM450). The correlation between the expression of the selected genes and DNA methylation in LUAD and LUSC was determined by cBioPortal. Moreover, the mutations of the selected genes and their relationship with gene expression were investigated using cBioPortal.

2.8 Gene Set Enrichment Analysis (GSEA)
GSEA is used for determining whether a defined gene set is expressed with significant differences
under two different biological conditions [29]. The TCGA data regarding LUAD (n = 479) and LUSC (n = 501) patients were downloaded from the Genomic Data Commons (GDC) Data Portal website (https://portal.gdc.cancer.gov/). Subsequently, the patients were classified into two groups (high vs. low expression) for each selected gene respectively, and the cutoffs were determined as medians. GSEA was conducted on the mRNA expression data of the selected genes using GSEA 4.0.3 software (http://software.broadinstitute.org/gsea/). A list of protein-coding genes of hallmark gene sets (H) from the Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb/) was obtained for GSEA analysis, with 1,000 permutations for calculating the normalized enrichment score (NES) and P values. Gene sets with a false discovery rate (FDR) of < 0.25 and P < 0.05 were regarded as significantly enriched gene sets.

3. Results
3.1 The human RHO GEFs are divided into two families.

The human RHO GEFs are comprised of 81 members, which can be mainly divided into two families, namely DOCK family (Fig. 1A) and DBL family (Fig. 1B). The structures of RHO GEFs are various, but conserved domains are also identified. Proteins of the DBL RHO GEFs are defined by the presence of conserved Dbl-homology (DH) domains. The DBL family of RHO GEFs contains 70 members with 72 DH domains, among which TRIO and KALIRIN possess two DH domains, and the other members all contain one DH domain, which is responsible for stimulating guanine nucleotide exchange. In addition to the DH domain, most of the DBL RHO GEFs contain a pleckstrin-homology (PH) domain following the DH domain. The PH domains are able to bind to phospholipids and essential to properly mediate the localization of DBL RHO GEFs to its action sites via lipid binding. Proteins of the DOCK RHO GEFs are classified by the presence of DOCK-homology region 1 (DHR-1) and DOCK-homology region 1 (DHR-2) domains. Similar to the DH and PH domains, DHR-2 domains are responsible for guanine nucleotide exchange, and DHR-1 domains make interaction with phospholipids, which aid in targeting DOCK RHO GEFs to the plasma membrane. Apart from DH, PH, DHR-1 and DHR-2 domains, other domains of RHO GEFs, for example, Src homology 2 (SH2) and Src homology 3 (SH3) domains, are important in mediating protein-protein interactions, protein-lipid interactions or messenger binding. In
order to highlight the highly conserved domains of RHO GEFs, the details for the other domains displayed in Fig. 1 are illustrated in Supplementary Fig. 1.

3.2 The expression levels of RHO GEFs are altered in lung cancer.
Oncomine database analysis is performed to investigate the altered mRNA expression in lung cancer. There are 43 genes encoding DBL GEFs (Supplementary Fig. 2A) and 7 genes (Supplementary Fig. 2B) encoding DOCK GEFs showing remarkable alterations of mRNA expression levels in lung cancer compared to normal tissues. For the DBL family of RHO GEFs, 24 genes are downregulated in lung cancer, whereas 13 genes are upregulated, and the other 6 genes show contradictory results of aberrant expression in different datasets. For the DOCK family of RHO GEFs, 6 genes are underexpressed in lung cancer, while only 1 gene is identified to be overexpressed. Moreover, GEPIA database analysis is used to verify the altered mRNA expression of genes encoding RHO GEFs in NSCLC. Due to the fact that gene expression patterns can be variant in LUAD and LUSC, the following analyses are performed independently in LUAD and LUSC [30]. Most of the genes encoding DBL RHO GEFs are downregulated in LUAD (Fig. 2A) and LUSC (Fig. 2B). For all of the DOCK RHO GEFs except DOCK7, the median expressions of genes are also decreased in LUAD (Fig. 2C) and LUSC (Fig. 2D) compared to normal samples. Furthermore, the UALCAN database and Kaplan Meier Plotter database analyses for all genes encoding RHO GEFs are performed (data not shown). The genes showing contradictory effects within or across different databases are excluded. Eventually, ABR and PREX1 in the DBL family of RHO GEFs, as well as DOCK2 and DOCK4 in the DOCK family of RHO GEFs, are selected for this study.

3.3 ABR, PREX1, DOCK2 and DOCK4 are downregulated in NSCLC.
In order to obtain an overview of the expression patterns of ABR in various tumor types and in the corresponding normal tissues, a body map is generated by GEPIA (Fig. 3A). Compared to normal lung tissue, the expression of ABR is lower in lung cancer. To further validate the downregulation of ABR in lung cancer, specific datasets in Oncomine analysis are illustrated (Fig. 3A). From Garber Lung dataset, the expression of ABR in LUSC is markedly lower than normal lung tissue (P < 0.001, Fold change − 2.435), while the datasets of LUAD with statistical significance are lacking in Oncomine
database. Therefore, GEPIA database is used to confirm our results. From the GEPIA database analysis of both LUAD and LUSC, ABR is significantly downregulated in tumors (P < 0.05). The downregulations of PREX (Fig. 3B), DOCK2 (Fig. 3C) and DOCK4 (Fig. 3D) are also validated by Oncomine and GEPIA database analysis. These results imply that ABR, PREX1, DOCK2 and DOCK4 can function as tumor suppressors in NSCLC, while more specific data on tumor subgroups and the corresponding survival data are needed for confirmation. Consequently, the expression levels of the selected genes in tumor subgroups divided by nodal metastasis status or individual stages are investigated.

From UALCAN database analysis, ABR (Supplementary Fig. 3A), PREX1 (Supplementary Fig. 3C), DOCK2 (Supplementary Fig. 3E) and DOCK4 (Supplementary Fig. 3G) are significantly downregulated in each nodal metastasis status (N0, N1, N2 and N3) of LUAD and LUSC when compared to normal tissues. In addition, the expressions of ABR (Supplementary Fig. 3B), PREX1 (Supplementary Fig. 3D), DOCK2 (Supplementary Fig. 3F) and DOCK4 (Supplementary Fig. 3H) are observably lower in each stage (stage1, stage2, stage3 and stage 4) of LUAD and LUSC in comparison to normal samples. Although the median expressions of the selected genes are generally decreased as nodal metastasis status or tumor stage advances, the majority of these data show no statistical significance. Taken together, these consistent results confirm that ABR, PREX1, DOCK2 and DOCK4 are notably downregulated in NSCLC, and their prognostic values are subsequently investigated.

### 3.4 Downregulations of ABR, PREX1, DOCK2 and DOCK4 are unfavorable prognostic factors for NSCLC.

Using the Kaplan Meier Plotter database, survival analysis is performed to determine the prognostic values of the selected genes in NSCLC patients. In regard to overall survival, ABR (Fig. 4A), PREX1 (Fig. 4C), DOCK2 (Fig. 4E) and DOCK4 (Fig. 4G) all demonstrate significant value in the prognosis of LUAD and LUSC patients, and a lower expression of these genes predicts an unfavorable overall survival in the patients. For LUAD patients, the expression of DOCK4 is more significantly correlated with overall survival (HR = 0.33; 95% CI: 0.24–0.44; P = 1.4E-14) compared to that of ABR, PREX1 and DOCK2. For LUSC patients, DOCK4 expression is also more markedly correlated to overall survival (HR = 0.64; 95% CI: 0.49–0.82; P = 0.00055) than ABR, PREX1 and DOCK2 expression. However, for the
prognosis of tumor progression, ABR shows no significance in both LUAD and LUSC patients (Fig. 4B). In addition, lower DOCK4 expression is associated with earlier first progression in LUAD patients, but not in LUSC patients (Fig. 4H). The downregulation of PREX1 (Fig. 4D) and DOCK2 (Fig. 4F) significantly correlates with earlier tumor progression in LUSC and LUAD patients. Among the four RHO GEFs, DOCK4 predicts the progression in LUAD patients with a higher significance (HR = 0.42; 95% CI: 0.29–0.62; P = 5.7E-06), while DOCK2 more observably predicts LUSC progression (HR = 0.48; 95% CI: 0.29–0.8; P = 0.0041). Collectively, through the analysis using Oncomine, GEPIA, UALCAN and Kaplan Meier Plotter databases, consistent results highlight the downregulations of ABR, PREX1, DOCK2 and DOCK4 in NSCLC, and they notably predict the unfavorable prognosis of NSCLC patients.

3.5 Promoter methylation levels contribute to the aberrant expression of ABR, PREX1, DOCK2 and DOCK4 in NSCLC

DNA methylation in promoter regions is important in modulating gene expression via inducing stable epigenetic inhibition of gene expression. Thereby, whether methylation is correlated with the expression of ABR, PREX1, DOCK2 and DOCK4 is examined. Using TCGA data from MEXPRESS database, significantly higher methylation levels of ABR (P = 0.004 for LUAD vs. normal tissue [cg13655563]; P = 9.335E-14 for LUSC vs. normal tissue [cg13655563]) (Fig. 5A), PREX1 (P = 4.322E-24 for LUAD vs. normal tissue [cg03361585]; P = 6.439E-5 for LUSC vs. normal tissue [cg27511255]) (Fig. 5B), DOCK2 (P = 2.146E-14 for LUAD vs. normal tissue [cg21233722]; P = 7.347E-26 for LUSC vs. normal tissue [cg21233722]) (Fig. 5C) and DOCK4 (P = 5.099E-4 for LUAD vs. normal tissue [cg16341945]; P = 1.302E-7 for LUSC vs. normal tissue [cg06869407]) (Fig. 5D) in tumor tissues compared to normal specimens are identified. Accordingly, whether the increased promoter methylation levels in tumor lead to the lower expression of genes is further investigated. cBioPortal database analysis reveals that the promoter methylation levels of ABR (P = 6.91E-20, r = -0.41 for LUAD; P = 7.31E-8, r = -0.28 for LUSC) (Fig. 5E), PREX1 (P = 1.786E-5, r = -0.20 for LUAD; P = 2.689E-3, r = -0.16 for LUSC) (Fig. 5F), DOCK2 (P = 5.12E-10, r = -0.29 for LUAD; P = 4.785E-5, r = -0.21 for LUSC) (Fig. 5G) and DOCK4 (P = 1.638E-6, r = -0.22 for LUAD; P = 1.609E-5, r = -0.22 for LUSC) (Fig. 5H) are negatively correlated with mRNA expression in NSCLC. The above results suggest that
the hypermethylation of ABR, PREX1, DOCK2 and DOCK4 is not only identified in NSCLC in comparison to normal tissues, but also negatively correlated with their mRNA expression in NSCLC.

3.6 Analysis of ABR, PREX1, DOCK2 and DOCK4 mutations in NSCLC.

Genetic alterations of ABR, PREX1, DOCK2 and DOCK4 are studied by cBioPortal database. The proportion and distribution of TCGA samples with genetic alterations in LUAD (Supplementary Fig. 4A) and LUSC (Supplementary Fig. 4B) are shown, and DOCK2 possesses the highest proportion of genetic alterations in both LUAD (13%) and LUSC (10%) among the four selected RHO GEFs. The locations of mutation are further illustrated in Supplementary Fig. 4C and Supplementary Fig. 4D. The majority of the mutations for the four selected RHO GEFs are missense mutations, and several mutations are located in the regions encoding for the conserved domains of RHO GEFs. However, the mutations of the selected genes have no statistically significant impact on the mRNA expression in LUAD (Supplementary Fig. 4E) and LUSC (Supplementary Fig. 4F).

3.7 GSEA identifies potential mechanisms by which ABR, PREX1, DOCK2 and DOCK4 regulate NSCLC development and progression

Based on the data from multiple analyses, the lower expression of ABR, PREX1, DOCK2 and DOCK4 can lead to NSCLC development and progression, and therefore GSEA by TCGA data is employed to explain the underlying mechanisms. The expression matrix of NSCLC patients is divided into high-expression and low-expression groups by the median expression levels ABR, PREX1, DOCK2 and DOCK4, respectively. GSEA identifies that, in LUAD and LUSC, patients with lower expressions of ABR (nominal $P = 0.002$, FDR $= 0.011$ for LUAD; nominal $P < 0.001$, FDR $< 0.001$ for LUSC) (Fig. 6A), PREX1 (nominal $P = 0.002$, FDR $= 0.004$ for LUAD; nominal $P < 0.001$, FDR $= 0.003$ for LUSC) (Fig. 6C), DOCK2 (nominal $P < 0.001$, FDR $< 0.001$ for LUAD; nominal $P < 0.001$, FDR $= 0.001$ for LUSC) (Fig. 6E) and DOCK4 (nominal $P = 0.002$, FDR $= 0.008$ for LUAD; nominal $P = 0.017$, FDR $= 0.053$ for LUSC) (Fig. 6G) have higher expression of genes associated with MYC signaling, which are essential for the development of lung cancer [31–33]. Moreover, our results also reveal that the DNA repair pathway is significantly enriched under the downregulation of ABR (nominal $P = 0.024$, FDR $= 0.056$), PREX1 (nominal $P = 0.002$, FDR $= 0.006$), DOCK2 (nominal $P < 0.001$, FDR $= 0.004$) and DOCK4 (nominal $P = 0.006$, FDR $= 0.012$) in LUAD. Similarly, DNA repair pathway enrichment is identified under ABR
(nominal $P = 0.008$, FDR = 0.027), DOCK2 (nominal $P = 0.012$, FDR = 0.022) and DOCK4 (nominal $P = 0.020$, FDR = 0.039) downregulation in LUSC. However, the DNA repair pathway is not enriched for the PREX1-low-expression group in LUSC (nominal $P = 0.075$, FDR = 0.115). The details of the above data are summarized in Table 1.

Table 1: Gene Set Enrichment Analysis (GSEA) of gene sets significantly enriched in LUAD and LUSC.

| Tumor Types | Genes Encoding GEFs | Enriched Gene Sets       | Size | ES  |
|-------------|---------------------|--------------------------|------|-----|
| LUAD        | ABR                 | HALLMARK_MYC_TARGETS_V1  | 199  | -0.73912 |
|             |                     | HALLMARK_DNA_REPAIR      | 150  | -0.53791 |
|             | PREX1               | HALLMARK_MYC_TARGETS_V1  | 199  | -0.75718 |
|             |                     | HALLMARK_DNA_REPAIR      | 150  | -0.61628 |
|             | DOCK2               | HALLMARK_MYC_TARGETS_V1  | 199  | -0.7828 |
|             |                     | HALLMARK_DNA_REPAIR      | 150  | -0.63067 |
|             | DOCK4               | HALLMARK_MYC_TARGETS_V1  | 199  | -0.73671 |
|             |                     | HALLMARK_DNA_REPAIR      | 150  | -0.57858 |
| LUSC        | ABR                 | HALLMARK_MYC_TARGETS_V1  | 199  | -0.7608 |
|             |                     | HALLMARK_DNA_REPAIR      | 150  | -0.54108 |
|             | PREX1               | HALLMARK_MYC_TARGETS_V1  | 199  | -0.72201 |
|             |                     | HALLMARK_DNA_REPAIR      | 150  | -0.43784 |
|             | DOCK2               | HALLMARK_MYC_TARGETS_V1  | 199  | -0.77524 |
|             |                     | HALLMARK_DNA_REPAIR      | 150  | -0.54779 |
|             | DOCK4               | HALLMARK_MYC_TARGETS_V1  | 199  | -0.63924 |
|             |                     | HALLMARK_DNA_REPAIR      | 150  | -0.49595 |

ES: enrichment score; NES: normalized enrichment score; FDR: false discovery rate; FWER: family-wise error rate.

4. Discussions

RHO GEFs are known for their critical roles as molecular switches in activating RHO GTPases, and therefore function as regulators for various diseases not limited to cancer [5, 6]. ABR, as is indicated by its name, was first identified as a breakpoint cluster region (BCR)-related protein that shared high homology with BCR in human (68% amino acids identity) [34]. ABR maintains the normal reactivity of innate immune system, and the altered function of ABR can lead to the development of leukemia [35, 36]. PREX1 was first discovered in the cytosol of neutrophils [37]. PREX1 is important in regulating ROS production, migration and chemotaxis of neutrophils [38, 39]. Elevated expression of PREX1 has been associated with the development of melanoma, prostate cancer and breast cancer [40–42].

DOCK2 expression was initially deemed to be restricted to hematopoietic cells [43]. Although it is predominantly expressed in lymphocytes and hematopoietic tissues, recent research has revealed the tumor-promoting role of DOCK2 in lymphoma and colorectal cancer [44–46]. Distinct from the identification of ABR, PREX1 and DOCK2 in non-cancer cells, DOCK4 identification was initially reported in osteosarcoma cells, in which DOCK4 was deleted during tumor progression [47]. However, DOCK4 was also reported to promote breast cancer development and associated with bone
metastasis [48, 49]. To date, the roles of ABR, PREX1, DOCK2 and DOCK4 in lung cancer are largely unknown, and the underlying mechanisms remain to be explored.

Through the integration of data from multiple databases, our study demonstrates that ABR, PREX1, DOCK2 and DOCK4 are downregulated in NSCLC. In addition, NSCLC subgroups with higher nodal metastasis levels or cancer stages generally have lower median expressions of ABR, PREX1, DOCK2 and DOCK4, whereas statistical significance for the comparison among these subgroups remains to be testified with a larger number of samples. Notably, the downregulations of ABR, PREX1, DOCK2 and DOCK4 are associated with unfavorable overall survival of NSCLC patients, suggesting that these four novel RHO GEFs can serve as promising biomarkers to predict the overall survival of the patients. For the prognostic value for disease progression, only PREX1 and DOCK2 are with statistical significance in both LUAD and LUSC patients. Although DOCK4 is not a candidate marker to predict the disease progression of LUSC patients, its higher expression is markedly correlated with delayed tumor progression in LUAD patients. Hence, ABR, PREX1, DOCK2 and DOCK4 downregulations are characteristic of NSCLC, and they are promising in predicting the prognosis of NSCLC patients.

Gene expression and repression within cancer cells can be controlled by the epigenetic mechanisms of DNA methylation, which is the interaction between genes and phenotype without causing mutations in DNA sequence [50, 51]. DNA methylation involves the covalent addition of methyl groups to the C-5 position of the cytosine rings, especially in a CpG dinucleotide [52]. In mammals, approximately 70% of the promoters are rich in unmethylated CpG [53]. Hypermethylation in CpG sites of promoters is typically associated with gene silencing in transcriptional levels [54]. In lung cancer, DNA hypermethylation of tumor suppressors represents a hallmark and an early event in tumorigenesis [55]. Our study reveals that NSCLC samples from TCGA database contain higher methylation levels in the promoters of ABR, PREX1, DOCK2 and DOCK4 genes compared to the normal samples. Moreover, in NSCLC specimens, elevated levels of promoter methylation are markedly associated with lower expression of ABR, PREX1, DOCK2 and DOCK4. Thus, DNA hypermethylation might contribute to the downregulation of ABR, PREX1, DOCK2 and DOCK4 in NSCLC, and the methylation profiles of the four key RHO GEFs may be novel biomarkers for lung cancer screening.
In addition to epigenetic alterations, gene mutations can also affect gene expression levels [56]. For the types of mutation for ABR, PREX1, DOCK2 and DOCK4 in NSCLC, the majority of them are missense and truncating mutations, which mean a change of a single amino acid into another and a change in the DNA that can shorten the protein, respectively. Our results illustrate that the expressions of ABR, PREX1, DOCK2 and DOCK4 in NSCLC are not correlated with mutation. It should be noted that loss-of-function or gain-of-function mutations can lead to potential inhibitory or tumorigenic effects [57]. From our analysis, a few mutations occur in the RHOGEF (DH), PH, DOCK-C2 (DHR-1) and DHR-2 domains, indicating that the critical functions of RHO GEF can be altered by the mutations, and this is yet to be confirmed.

Mechanistically, the downregulation of ABR, PREX1, DOCK2 and DOCK4 upregulates MYC signaling and DNA repair pathways, as is identified by GSEA using TCGA data. MYC oncogene encodes a transcription factor and triggers gene expressions in cancer cells [58]. MYC signaling is implicated in the pathogenesis of most human cancers, and its deregulation is correlated with poor survival of patients [59]. MYC activation is associated with many features of cancer, including protein synthesis, proliferation and altered cellular pathways [59, 60]. MYC upregulation is detected in > 40% NSCLC, and it is related to the loss of cell differentiation and tumor progression [61, 62]. Therefore, ABR, PREX1, DOCK2 and DOCK4 downregulation may lead to the development of NSCLC via upregulating the MYC protein expression and its downstream targets.

In clinical practice, chemotherapy and radiotherapy are gold standards for the treatment of patients with lung cancer, and they have largely prolonged the survival of patients [63]. The therapeutic effects of platinum-based chemotherapeutic drugs and ionizing radiation are achieved by DNA damage [64–66]. Nevertheless, the enhanced DNA repair mechanisms counteract the therapeutic benefits to patients, and thereby lead to poor survival of the patients. Hence, the downregulations of ABR, PREX1, DOCK2 and DOCK4 promote cancer development and lead to an unfavorable prognosis by activating MYC and DNA repair signaling pathways in NSCLC, while further in vitro and in vivo studies are necessary to further confirm their association. Moreover, agents targeting DNA damage repair mechanisms have shown promise in NSCLC clinical models [65]. These four key RHO GEFs are
also promising as predicting biomarkers for the response to DNA-damage regimens.

In spite of the conventional perspective that RHO GEFs are activators for RHO GTPases, our study unexpectedly but reasonably reveals that the RHO GEFs: ABR, PREX1, DOCK2 and DOCK4 are tumor suppressors in NSCLC. In addition to the above findings, it should be noticed that the following evidence might also support our results. ABR contains a RHO GAP domain in addition to its RHO GEF domain, functioning as a dual RHO GEF/GAP [67]. Additionally, the prominent expression of ABR, PREX1 and DOCK2 in cells of the immune system suggests their correlations with lymphocyte infiltration via cytokine secretion in NSCLC tissues. Tumor-infiltrating lymphocytes can attack and eliminate tumor cells, and therefore contribute to a better prognosis of NSCLC patients [68, 69]. The association between DOCK2 expression and lymphocyte infiltration has been identified in colorectal cancer, and their correlation in NSCLC remains to be confirmed [45]. Moreover, DOCK4 was initially identified as a tumor suppressor in osteosarcoma cells, and latter studies also indicate the tumor suppressive role of DOCK4 in ovarian cancer [47, 70]. For future studies, researchers should not simply regard all RHO GEFs as tumor promoters on account of their involvement in activating RHO GTPases. Instead, the functions of RHO GEFs other than activating RHO GEFs in cancers should be noticed.

5. Conclusions

In conclusion, our study identifies that the downregulation of ABR, PREX1, DOCK2 and DOCK4 promotes NSCLC development, and is associated with unfavorable effects on the overall survival of patients. Therefore, they can serve as promising biomarkers to predict the prognosis of NSCLC patients. The downregulation of ABR, PREX1, DOCK2 and DOCK4 in NSCLC can be induced by promoter methylation, and their methylation profiles might be potential indicators for lung cancer screening. Moreover, the underexpression of the four key genes upregulates the tumorigenic MYC signaling and leads to tumor development. The DNA repair pathway is also induced by their downregulation, which contributes to chemotherapeutic and radiotherapeutic resistance, thereby resulting in a poor prognosis. Notably, our study highlights the previously unnoticed functions of RHO GEFs in NSCLC. In light of these findings, the distinct roles and mechanisms of RHO GEFs in
tumorigenesis should be further clarified, in order to provide comprehensive understanding of RHO GEFs and critical guidance for clinical practice.

**Abbreviations**

ABR, Active breakpoint cluster region-related; BCR, Breakpoint cluster region; CDC42, Cell division control protein 42 homolog; CI, Confidence interval; DH, Dbl-homology; DOCK, Deducator of cytokinesis; DBL, Diffuse B-cell lymphoma; DHR, DOCK-homology region; EGA, European Genome-Phenome Archive; FDR, False discovery rate; FP, First progression; GDI, GDP-dissociation inhibitor; GEO, Gene Expression Omnibus; GEPIA, Gene Expression Profiling Interactive Analysis; GSEA, Gene Set Enrichment Analysis; GDC, Genomic Data Commons; GTEx, Genotype-Tissue Expression; GAP, GTPase-activating protein; GEF, Guanine nucleotide exchange factor; GDP, Guanosine diphosphate; GTPases, Guanosine triphosphatases; GTP, Guanosine triphosphate; HR, Hazard ratio; LUAD, Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma; MSigDB, Molecular Signatures Database; NES, Normalized enrichment score; OS, Overall survival; PREX1, Phosphatidylinositol 3,4,5-trisphosphate-dependent Rac exchanger 1; PH, Pleckstrin-homology; RHOA, Ras homolog family member A; RAC1, Ras-related C3 botulinum toxin substrate 1; SH, Src homology; TCGA, The Cancer Genome Atlas.

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and materials**

All the analysis data are available from the database mentioned in the method section.

**Competing interests**

The authors declare no conflict of interest.
Funding

This work was supported by the Natural Science Foundation of China-Guangdong Joint Fund (U1601229) and the Special Funds for the Cultivation of Guangdong College Students’ Scientific and Technological Innovation (“Climbing Program” Special Funds, pdjh2019a0182).

Authors’ contributions

Conceptualization – RJZ, LYX and EML; Methodology – RJZ, CWZ, LYX and EML; Investigation – RJZ and CWZ; Writing Original Draft – RJZ; Writing Review & Editing – CWZ, YWX, LYX and EML; Funding Acquisition – RJZ, CWZ, LYX and EML; Supervision – LYX and EML.

Acknowledgments

The authors thank Yang Chen for technical support.

References

1. de Groot PM, Wu CC, Carter BW, Munden RF: The epidemiology of lung cancer. *Transl Lung Cancer Res* 2018, 7(3):220-233.
2. Barta JA, Powell CA, Wisnivesky JP: Global Epidemiology of Lung Cancer. *Ann Glob Health* 2019, 85(1).
3. Zappa C, Mousa SA: Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res* 2016, 5(3):288-300.
4. Herbst RS, Morgensztern D, Boshoff C: The biology and management of non-small cell lung cancer. *Nature* 2018, 553(7689):446-454.
5. Etienne-Manneville S, Hall A: Rho GTPases in cell biology. *Nature* 2002, 420(6916):629-635.
6. Haga RB, Ridley AJ: Rho GTPases: Regulation and roles in cancer cell biology. *Small GTPases* 2016, 7(4):207-221.
7. Kumawat A, Chakrabarty S, Kulkarni K: Nucleotide dependent switching in rho GTPase: conformational heterogeneity and competing molecular interactions. *Scientific reports* 2017, 7(1):1-11.

8. Zou T, Mao X, Yin J, Li X, Chen J, Zhu T, Li Q, Zhou H, Liu Z: Emerging roles of RAC1 in treating lung cancer patients. *Clin Genet* 2017, 91(4):520-528.

9. Konstantinidou G, Ramadori G, Torti F, Kangasniemi K, Ramirez RE, Cai Y, Behrens C, Dellinger MT, Brekken RA, Wistuba II et al: RHOA-FAK is a required signaling axis for the maintenance of KRAS-driven lung adenocarcinomas. *Cancer Discov* 2013, 3(4):444-457.

10. Xiao X-H, Lv L-C, Duan J, Wu Y-M, He S-J, Hu Z-Z, Xiong L-X: Regulating Cdc42 and its signaling pathways in cancer: small molecules and MicroRNA as new treatment candidates. *Molecules* 2018, 23(4):787.

11. Porter AP, Papaioannou A, Malliri A: Deregulation of Rho GTPases in cancer. *Small GTPases* 2016, 7(3):123-138.

12. Sosa MS, Lopez-Haber C, Yang C, Wang H, Lemmon MA, Busillo JM, Luo J, Benovic JL, Klein-Szanto A, Yagi H: Identification of the Rac-GEF P-Rex1 as an essential mediator of ErbB signaling in breast cancer. *Molecular cell* 2010, 40(6):877-892.

13. Cardama GA, González N, Maggio J, Menna PL, Gomez DE: Rho GTPases as therapeutic targets in cancer. *International journal of oncology* 2017, 51(4):1025-1034.

14. Lawson CD, Ridley AJ: Rho GTPase signaling complexes in cell migration and invasion. *Journal of Cell Biology* 2018, 217(2):447-457.

15. Cook DR, Rossman KL, Der CJ: Rho guanine nucleotide exchange factors: regulators of Rho GTPase activity in development and disease. *Oncogene* 2014, 33(31):4021-4035.
16. Lazer G, Katzav S: Guanine nucleotide exchange factors for RhoGTPases: good therapeutic targets for cancer therapy? *Cellular signalling* 2011, 23(6):969-979.

17. Kumar S, Stecher G, Tamura K: MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 2016, 33(7):1870-1874.

18. He Z, Zhang H, Gao S, Lercher MJ, Chen W-H, Hu S: Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic acids research* 2016, 44(W1):W236-W241.

19. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barret TR, Anstet MJ, Kincaid-Beal C, Kulkarni P: Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia (New York, NY)* 2007, 9(2):166.

20. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia (New York, NY)* 2004, 6(1):1.

21. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic acids research* 2017, 45(W1):W98-W102.

22. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK, Varambally S: UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* 2017, 19(8):649-658.

23. Győrffy B, Surowiak P, Budczies J, Lánczky A: Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* 2013, 8(12):e82241-e82241.

24. Li Q, Birbkab NJ, Gyorffy B, Szallasi Z, Eklund AC: Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics* 2011, 12:474-474.
25. Győrffy B, Benke Z, Lánczky A, Balázs B, Szállási Z, Timár J, Schäfer R: RecurrenceOnline: an online analysis tool to determine breast cancer recurrence and hormone receptor status using microarray data. *Breast Cancer Res Treat* 2012, 132(3):1025-1034.

26. Koch A, De Meyer T, Jeschke J, Van Criekinge W: MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics* 2015, 16(1):636.

27. Koch A, Jeschke J, Van Criekinge W, van Engeland M, De Meyer T: MEXPRESS update 2019. *Nucleic Acids Research* 2019, 47(W1):W561-W565.

28. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013, 6(269):pl1-pl1.

29. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 2005, 102(43):15545-15550.

30. Faruki H, Mayhew GM, Serody JS, Hayes DN, Perou CM, Lai-Goldman M: Lung Adenocarcinoma and Squamous Cell Carcinoma Gene Expression Subtypes Demonstrate Significant Differences in Tumor Immune Landscape. *J Thorac Oncol* 2017, 12(6):943-953.

31. Brägelmann J, Böhm S, Guthrie MR, Mollaoglu G, Oliver TG, Sos ML: Family matters: How MYC family oncogenes impact small cell lung cancer. *Cell Cycle* 2017, 16(16):1489-1498.

32. Soucek L, Whitfield JR, Sodir NM, Massó-Vallés D, Serrano E, Karnezis AN, Swigart LB, Evan GI: Inhibition of Myc family proteins eradicates KRas-driven lung cancer in mice. *Genes Dev* 2013, 27(5):504-513.
33. Yang S, Liu Y, Li M-Y, Ng CSH, Yang S-L, Wang S, Zou C, Dong Y, Du J, Long X et al: FOXP3 promotes tumor growth and metastasis by activating Wnt/β-catenin signaling pathway and EMT in non-small cell lung cancer. Mol Cancer 2017, 16(1):124-124.

34. Tan EC, Leung T, Manser E, Lim L: The human active breakpoint cluster region-related gene encodes a brain protein with homology to guanine nucleotide exchange proteins and GTPase-activating proteins. J Biol Chem 1993, 268(36):27291-27298.

35. Cunnick JM, Schmidhuber S, Chen G, Yu M, Yi S-J, Cho YJ, Kaartinen V, Minoo P, Warburton D, Groffen J et al: Bcr and Abr cooperate in negatively regulating acute inflammatory responses. Mol Cell Biol 2009, 29(21):5742-5750.

36. Namasu CY, Katzerke C, Bräuer-Hartmann D, Wurm AA, Gerloff D, Hartmann J-U, Schwind S, Müller-Tidow C, Hilger N, Fricke S: ABR, a novel inducer of transcription factor C/EBPα, contributes to myeloid differentiation and is a favorable prognostic factor in acute myeloid leukemia. Oncotarget 2017, 8(61):103626.

37. Welch HC, Coadwell WJ, Ellson CD, Ferguson GJ, Andrews SR, Erdjument-Bromage H, Tempst P, Hawkins PT, Stephens LR: P-Rex1, a PtdIns (3, 4, 5) P3-and Gβγ-regulated guanine-nucleotide exchange factor for Rac. Cell 2002, 108(6):809-821.

38. Welch HCE, Condliffe AM, Milne LJ, Ferguson GJ, Hill K, Webb LMC, Okkenhaug K, Coadwell WJ, Andrews SR, Thelen M et al: P-Rex1 regulates neutrophil function. Curr Biol 2005, 15(20):1867-1873.

39. Cervantes-Villagrana RD, Adame-García SR, García-Jiménez I, Color-Aparicio VM, Beltrán-Navarro YM, König GM, Kostenis E, Reyes-Cruz G, Gutkind JS, Vázquez-Prado J: Gβγ signaling to the chemotactic effector P-REX1 and mammalian cell migration is directly regulated by Gαq and Gα13 proteins. Journal of Biological Chemistry 2019, 294(2):531-546.

40. Lindsay CR, Lawn S, Campbell AD, Faller WJ, Rambow F, Mort RL, Timpson P, Li A,
Cammareri P, Ridgway RA: P-Rex1 is required for efficient melanoblast migration and melanoma metastasis. *Nature communications* 2011, 2(1):1-9.

41. Qin J, Xie Y, Wang B, Hoshino M, Wolff DW, Zhao J, Scofield MA, Dowd FJ, Lin M-F, Tu Y: Upregulation of PIP3-dependent Rac exchanger 1 (P-Rex1) promotes prostate cancer metastasis. *Oncogene* 2009, 28(16):1853-1863.

42. Dillon LM, Bean JR, Yang W, Shee K, Symonds LK, Balko JM, McDonald WH, Liu S, Gonzalez-Angulo AM, Mills GB: P-REX1 creates a positive feedback loop to activate growth factor receptor, PI3K/AKT and MEK/ERK signaling in breast cancer. *Oncogene* 2015, 34(30):3968-3976.

43. Nishihara H, Kobayashi S, Hashimoto Y, Ohba F, Mochizuki N, Kurata T, Nagashima K, Matsuda M: Non-adherent cell-specific expression of DOCK2, a member of the human CDM-family proteins. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1999, 1452(2):179-187.

44. Gadea G, Blangy A: Dock-family exchange factors in cell migration and disease. *European journal of cell biology* 2014, 93(10-12):466-477.

45. Miao S, Zhang RY, Wang W, Wang HB, Meng LL, Zu LD, Fu GH: Overexpression of dedicator of cytokinesis 2 correlates with good prognosis in colorectal cancer associated with more prominent CD8+ lymphocytes infiltration: a colorectal cancer analysis. *Journal of cellular biochemistry* 2018, 119(11):8962-8970.

46. Wang L, Nishihara H, Kimura T, Kato Y, Tanino M, Nishio M, Obara M, Endo T, Koike T, Tanaka S: DOCK2 regulates cell proliferation through Rac and ERK activation in B cell lymphoma. *Biochemical and biophysical research communications* 2010, 395(1):111-115.

47. Yajnik V, Paulding C, Sordella R, McClatchey AI, Saito M, Wahrer DC, Reynolds P, Bell DW, Lake R, van den Heuvel S: DOCK4, a GTPase activator, is disrupted during
tumorigenesis. *Cell* 2003, 112(5):673-684.

48. Kobayashi M, Harada K, Negishi M, Katoh H: Dock4 forms a complex with SH3YL1 and regulates cancer cell migration. *Cellular signalling* 2014, 26(5):1082-1088.

49. Westbrook JA, Wood SL, Cairns DA, McMahon K, Gahlaut R, Thygesen H, Shires M, Roberts S, Marshall H, Oliva MR: Identification and validation of DOCK4 as a potential biomarker for risk of bone metastasis development in patients with early breast cancer. *J Pathol* 2019, 247(3):381-391.

50. Pfeifer GP: Defining Driver DNA Methylation Changes in Human Cancer. *Int J Mol Sci* 2018, 19(4):1166.

51. Li L, Li C, Mao H, Du Z, Chan WY, Murray P, Luo B, Chan AT, Mok TS, Chan FK et al: Epigenetic inactivation of the CpG demethylase TET1 as a DNA methylation feedback loop in human cancers. *Scientific reports* 2016, 6:26591-26591.

52. Jin B, Li Y, Robertson KD: DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer* 2011, 2(6):607-617.

53. Saxonov S, Berg P, Brutlag DL: A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103(5):1412-1417.

54. Belinsky SA: Silencing of genes by promoter hypermethylation: key event in rodent and human lung cancer. *Carcinogenesis* 2005, 26(9):1481-1487.

55. Langevin SM, Kratzke RA, Kelsey KT: Epigenetics of lung cancer. *Transl Res* 2015, 165(1):74-90.

56. Zhang Z, Miteva MA, Wang L, Alexov E: Analyzing effects of naturally occurring missense mutations. *Comput Math Methods Med* 2012, 2012:805827-805827.

57. Kim KH, Roberts CWM: Targeting EZH2 in cancer. *Nat Med* 2016, 22(2):128-134.
58. Stine ZE, Walton ZE, Altman BJ, Hsieh AL, Dang CV: MYC, Metabolism, and Cancer. 
   *Cancer Discov* 2015, 5(10):1024-1039.

59. Chen H, Liu H, Qing G: Targeting oncogenic Myc as a strategy for cancer treatment. 
   *Signal Transduct Target Ther* 2018, 3:5-5.

60. Gabay M, Li Y, Felsher DW: MYC activation is a hallmark of cancer initiation and 
    maintenance. *Cold Spring Harb Perspect Med* 2014, 4(6):a014241.

61. Lorenz J, Friedberg T, Paulus R, Oesch F, Ferlinz R: Oncogene overexpression in non-
    small-cell lung cancer tissue: prevalence and clinicopathological significance. *Clin 
    Investig* 1994, 72(2):156-163.

62. Mustachio LM, Roszik J, Farria AT, Guerra K, Dent SY: Repression of GCN5 expression 
    or activity attenuates c-MYC expression in non-small cell lung cancer. *Am J Cancer 
    Res* 2019, 9(8):1830-1845.

63. Lemjabbar-Alaoui H, Hassan OU, Yang YW, Buchanan P: Lung cancer: Biology and 
    treatment options. *Biochim Biophys Acta* 2015, 1856(2):189-210.

64. Woods D, Turchi JJ: Chemotherapy induced DNA damage response: convergence of 
    drugs and pathways. *Cancer Biol Ther* 2013, 14(5):379-389.

65. Sen T, Gay CM, Byers LA: Targeting DNA damage repair in small cell lung cancer and 
    the biomarker landscape. *Translational lung cancer research* 2018, 7(1):50-68.

66. Butkiewicz D, Drosik A, Suwiński R, Krześniak M, Rusin M, Kosarewicz A, Rachtan J, 
    Matuszczyk I, Gawkowska-Suwińska M: Influence of DNA repair gene polymorphisms 
    on prognosis in inoperable non-small cell lung cancer patients treated with 
    radiotherapy and platinum-based chemotherapy. *Int J Cancer* 2012, 131(7):E1100-
    E1108.

67. Vaughan EM, Miller AL, Yu H-YE, Bement WM: Control of local Rho GTPase crosstalk 
    by Abr. *Curr Biol* 2011, 21(4):270-277.
68. Yu P, Fu Y-X: Tumor-infiltrating T lymphocytes: friends or foes? Laboratory Investigation 2006, 86(3):231-245.

69. Geng Y, Shao Y, He W, Hu W, Xu Y, Chen J, Wu C, Jiang J: Prognostic Role of Tumor-Infiltrating Lymphocytes in Lung Cancer: a Meta-Analysis. Cellular Physiology and Biochemistry 2015, 37(4):1560-1571.

70. Kuo K-T, Guan B, Feng Y, Mao T-L, Chen X, Jinawath N, Wang Y, Kurman RJ, Shih I-M, Wang T-L: Analysis of DNA copy number alterations in ovarian serous tumors identifies new molecular genetic changes in low-grade and high-grade carcinomas. Cancer research 2009, 69(9):4036-4042.

Figures
Figure 1

The human DBL family and DOCK family of RHO GEFs. (A) The 70 members of DBL family RHO GEFs and (B) 11 members of DOCK family RHO GEFs are arranged in phylogenetic tree by MEGA7.0 software, using the protein sequences obtained from UniProt. The protein domain architecture is visualized by Evolview. Each color represents one distinct structural domain. The DH, PH domains of DBL RHO GEF family and DHR-1, DHR2 domains of DOCK RHO GEF family with the corresponding colors are highlighted in the left panel. All proteins are visualized in the same scale except OBSCN. For other structural domains, please refer to Supplementary Fig.1. DH: Dbl homology; PH: Pleckstrin homology; DHR-1: DOCK homology region 1; DHR-2: DOCK homology region 2.
Figure 2

Expression of RHO GEFs in NSCLC. The expression levels of RHO GEFs in NSCLC, represented by LUSC and LUAD, are analyzed by GEPIA database. (A) The median expression of DBL RHO GEFs in LUAD compared to normal tissues. (B) The median expression of DBL RHO GEFs in LUSC in comparison with normal samples. (C) The median expression of DOCK RHO GEFs in LUAD compared to normal tissues. (D) The median expression of DOCK RHO GEFs in LUSC in comparison with normal specimens.
Figure 3

Expression of ABR, PREX1, DOCK2 and DOCK4 in NSCLC. (A) The body map of various organs (GEPIA), the cohort of Garber Lung (Oncomine), and the LUSC and LUAD patient specimens with normal tissues for the expression of ABR (GEPIA) are illustrated. (B) The body map (GEPIA), the cohorts of Hou Lung and Okayama Lung (Oncomine), and the LUSC and LUAD patient samples with normal tissues for PREX1 expression (GEPIA) are demonstrated. (C) The body map including different organs (GEPIA), the cohorts of Bhattacharjee Lung and Selamat Lung (Oncomine), and the LUSC and LUAD patient samples with normal tissues for DOCK2 expression (GEPIA) are shown. (D) The body map (GEPIA), the cohorts of Bhattacharjee Lung, Hou Lung, Wachi Lung, Landi Lung, Su Lung and Stearman Lung (Oncomine), and the LUSC and LUAD patient samples with normal tissues for DOCK4 expression (GEPIA) are illustrated. The darker color in the body map represents higher expression. *P < 0.05.

Fig.4

ABR (212895_s_at) - Overall Survival (OS)

|       | LUAD | LUSC |
|-------|------|------|
| HR    | 0.62 (0.49-0.79) logrank P = 9.9E-05 |
| HR    | 0.75 (0.56-1) logrank P = 0.847 |

ABR (212895_s_at) - First Progression (FP)

|       | LUAD | LUSC |
|-------|------|------|
| HR    | 0.78 (0.67-1.07) logrank P = 0.13 |
| HR    | 0.71 (0.60-1.00) logrank P = 0.19 |

PREX1 (224925_at) - Overall Survival (OS)

|       | LUAD | LUSC |
|-------|------|------|
| HR    | 0.62 (0.49-0.79) logrank P = 0.0015 |
| HR    | 0.72 (0.53-0.98) logrank P = 0.036 |

PREX1 (224925_at) - First Progression (FP)

|       | LUAD | LUSC |
|-------|------|------|
| HR    | 0.47 (0.34-0.85) logrank P = 4.3E-06 |
| HR    | 0.50 (0.31-0.81) logrank P = 0.013 |

DOCK2 (213180_at) - Overall Survival (OS)

|       | LUAD | LUSC |
|-------|------|------|
| HR    | 0.61 (0.47-0.78) logrank P = 0.0053 |
| HR    | 0.69 (0.52-0.92) logrank P = 0.01 |

DOCK2 (213180_at) - First Progression (FP)

|       | LUAD | LUSC |
|-------|------|------|
| HR    | 0.64 (0.47-0.88) logrank P = 0.0054 |
| HR    | 0.48 (0.39-0.68) logrank P = 0.9041 |
Figure 4

Prognostic value of ABR, PREX1, DOCK2 and DOCK4 expression in NSCLC. (A, C, E, G)

Survival curves referring to OS are plotted for LUAD and LUSC patients. (A, C, E, G) Survival curves with regard to FP are generated for patients with LUAD and LUSC. Log-rank P values and HRs with 95% CIs are displayed. FP: first progression; HR: hazard ratio; OS: overall survival.
Figure 5

Impact of DNA methylation on ABR, PREX1, DOCK2 and DOCK4 expression in NSCLC.

Visualization of TCGA data for the promoter region methylation and mRNA expression of (A) ABR, (B) PREX1, (C) DOCK2 and (D) DOCK4 in LUAD and LUSC, compared to those in normal tissues using MEXPRESS. The correlation of (E) ABR, (F) PREX1, (G) DOCK2 and (H) DOCK4 expression with promoter methylation is analyzed by cBioPortal database. Spearman’s correlation analysis and Pearson’s correlation analysis are performed in the database. The regression line is used to illustrate the correlation trend.
Fig. 6

A  Enrichment Plot: HALLMARK_MYC_TARGETS_V1
   LUAD  LUSC
   
B  Enrichment Plot: HALLMARK_DNA_REPAIR
   LUAD  LUSC
   
C  Enrichment Plot: HALLMARK_MYC_TARGETS_V1
   LUAD  LUSC
   
D  Enrichment Plot: HALLMARK_DNA_REPAIR
   LUAD  LUSC
   
E  Enrichment Plot: HALLMARK_MYC_TARGETS_V1
   LUAD  LUSC
   
F  Enrichment Plot: HALLMARK_DNA_REPAIR
   LUAD  LUSC
   
G  Enrichment Plot: HALLMARK_MYC_TARGETS_V1
   LUAD  LUSC
   
H  Enrichment Plot: HALLMARK_DNA_REPAIR
   LUAD  LUSC

Nominal p-value < 0.002
FDR p-value = 0.011

Nominal p-value < 0.001
FDR p-value = 0.001

Nominal p-value < 0.024
FDR p-value = 0.006

Nominal p-value < 0.008
FDR p-value = 0.077

Nominal p-value < 0.002
FDR p-value = 0.004

Nominal p-value < 0.001
FDR p-value = 0.003

Nominal p-value < 0.001
FDR p-value = 0.004

Nominal p-value < 0.001
FDR p-value = 0.004

Nominal p-value < 0.002
FDR p-value = 0.006

Nominal p-value < 0.001
FDR p-value = 0.003

Nominal p-value < 0.005
FDR p-value = 0.002

Nominal p-value < 0.001
FDR p-value = 0.004

Nominal p-value < 0.002
FDR p-value = 0.006

Nominal p-value < 0.003
FDR p-value = 0.012

Nominal p-value < 0.001
FDR p-value = 0.004

Nominal p-value < 0.002
FDR p-value = 0.006

Nominal p-value < 0.022
FDR p-value = 0.009
Figure 6

GSEA analysis between the high-expression group and low-expression group in ABR, PREX1, DOCK2 and DOCK4 for NSCLC. The correlation between the enrichment of MYC signaling gene set and (A) ABR, (C) PREX1, (E) DOCK2 and (G) DOCK4 expression in LUAD and LUSC. The association between the expression of (B) ABR, (D) PREX1, (F) DOCK2 and (H) DOCK4 expression with DNA repair gene set enrichment. The barcode plot demonstrates gene positions in each gene set. The horizontal bar indicates positive (red) and negative correlation (blue) with gene expression. FDR: false discovery rate.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

Supplementary Fig4.TIF
Supplementary Fig3.TIF
Supplementary Fig1.TIF
Supplementary Fig2.TIF