Diagnostic value of GDF10 for the tumorigenesis and progression in lung squamous cell carcinoma

Peiyuan Mei
Huazhong University of Science and Technology

Jiaping Chen
Huazhong University of Science and Technology

Wangyang Meng
Huazhong University of Science and Technology

Yangwei Wang
Huazhong University of Science and Technology

Yunchong Meng
Huazhong University of Science and Technology

Rong Zhao
Huazhong University of Science and Technology

Wei Lin
Huazhong University of Science and Technology

Yongde Liao
Huazhong University of Science and Technology

Han Xiao (✉️ 13260536972@163.com)
Huazhong University of Science and Technology

Research Article

Keywords: lung squamous cell carcinoma, tumorigenesis, progression, BMP, GDF10

Posted Date: November 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2131584/v1

License: ©  This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background

Lung squamous cell carcinoma (LUSC) remains a poor survival rate, calling for a novel molecular with diagnostic and treatment value. Accumulative evidence found bone morphogenetic proteins (BMPs) and their receptors (BMPRs) play important roles in tumorigenesis and progression, however, was lack of comprehensive analysis of their expression in LUSC.

Methods

R/Limma package was performed to analyze the differential expression of BMPs/BMPRs in combination of TCGA and GTEx, and explore their expression characteristics with LUSC tumorigenesis in GSE33479. Meanwhile, survminer packages were performed to explore their prognostic value and correlation of clinical features in LUSC. Then, the potential diagnostic biomarkers and mechanisms associated with LUSC progression were further explored through weight gene correlation network analysis (WGCNA). At the same time, LASSO analysis was performed to construct a prognostic risk model for LUSC with the differential expression of BMPs/BMPRs as the core. Finally, the specimens were collected from 33 patients with LUSC and detected by IHC to confirm the relationship between protein levels of the above diagnostic BMPs/BMPRs and progression of LUSC.

Results

On the whole, 2 upregulated genes (BMP8A, BMP7) and 8 downregulated genes (BMP2, BMP5, BMP6, GDF5, GDF7, GDF10, ACVRL1 and BMPR2) were identified differentially expressed genes in LUSC. In these differentially expressed genes, GDF10 was only a significant correlation with pathological T stage of LUSC (p < 0.001). The co-expressed network showed that the positively related magenta module (Coefficient:0.93, p = 1.4e-70 < 0.001) and the negatively correlated turquoise module (Coefficient:0.89, p = 1e-200) are significantly associated with GDF10. Meanwhile, combining 72 significantly down-regulated genes in magenta module and 351 significantly up-regulated genes in turquoise module together, a prognostic risk model was constructed with GDF10 as the core gene and 5 hub genes (HRASLS, HIST1H2BH, FLRT3, CHEK2 and ALPL) (HR:1.73, 95%CI:1.32–2.28, p = 1e-04 < 0.001). At last, immunohistochemical results verified that the protein expression level of GDF10 decreased with the tumorigenesis and progression of LUSC.

Conclusion

Both mRNA and protein expression levels of GDF10 acted as an independent protective factor in the tumorigenesis and progression of lung squamous cell carcinoma. As a result, it may be a potential diagnostic biomarker and a new therapeutic target for LUSC.

Background

Lung squamous cell carcinoma (LUSC) is the secondly major subtype of non-small cell lung cancer (NSCLC) after lung adenocarcinoma (LUAD) with multi-step progression\cite{1},which has high morbidity and mortality worldwide\cite{2}.
Compared to LUAD, the 5-year survival rate for LUSC remains poor and the targeted therapies available for LUSC are limited[3]. So far, there has not been an ideal molecular target with diagnostic value for the tumorigenesis and progression in lung squamous cell carcinoma.

As early as 1999, the important clinical significance of early diagnosis has been highlighted for prevention and treatment of lung cancer[4]. Since then, researchers have been searching for molecular diagnostic biomarkers, which could make an accurate assessment of the tumorigenesis and progression of lung cancer[5]. Two opportunities have driven the development of diagnostic biomarkers of LUSC: rapid advances in deep sequencing have made it possible to comprehensively screen diagnostic biomarkers, and invasive LUSC in smokers preceded by a range of consecutive developmental stages have made it a convenient model for mechanistically studying the early evolution of cancer[6]. The search for quantifiable biomarkers with more important diagnostic value, to improve the diagnostic accuracy and guide more effectively treatment of lung squamous cell carcinoma, which is also the original purpose of the current research.

To find such quantifiable diagnostic biomarkers of LUSC, we put our sights on the role of bone morphogenetic proteins (BMPs). As a subfamily of the transforming growth factor β (TGF-β) family, BMPs play a diverse range of roles in different tissue types including cell proliferation, apoptosis and differentiation control, not just bone metabolism[7]. They also play important roles in the development and progression of diverse cancers, including prostate cancer, breast cancer, lung cancer and so on[8]. Recently, more evidence found that BMPs contribute to the process of tumorigenesis and regulate progression in various cancers, binding to two types of BMP receptors (BMPRs), type I (ACVRL1, ACVR1, BMPR1A and BMPR1B) and type II (BMPR2, ACVR2A and ACVR2B)[9]. As for lung cancer, studies have reported that BMP2/BMP4 could regulate the development and even promote tumorigenesis of lung[10, 11]. Our previous research found that BMP5 is a potential crucial target for lung adenocarcinoma treatment based on significant differential expression and superior prognostic value[12]. BMPs/BMPRs is expected to be the novel diagnostic biomarker in lung cancer. However, BMPs family containing more than 20 ligands is large and BMPs signaling in cancer including lung cancer is complex. Besides, complexity and even contradiction still surround the function of BMPs in lung cancer. At present, there is still a lack of comprehensive analysis of BMPs/BMPRs expression levels in LUSC. And almost all relevant studies have not adequately considered the critical effects of histopathological types. Because of BMPs/BMPRs is still a long way to become a quantifiable biomarker with more important diagnostic value of lung squamous cell carcinoma, we should actively explore the relationship between BMPs/BMPRs expression levels and the tumorigenesis and progression of lung squamous cell carcinoma.

In the current study, we aimed to analyze the mRNA expression profile of BMPs/BMPRs in a range of consecutive developmental stages of LUSC and tried to screen out diagnostic molecules that as an accurate indicator to predict tumorigenesis and progression of LUSC, and there is systematic description of diagnostic molecules expression levels in lung squamous cell carcinoma tissue (Fig. 1). Such a diagnostic marker would undoubtedly serve as an indicator to help pathologists and clinicians more accurately distinguish the stages of tumorigenesis, and greatly advance diagnostic and prognostic value in LUSC.

Methods

**Differential expression analysis of BMPs/BMPRs in TCGA-LUSC dataset**
The lung squamous cell carcinoma cohort in The Cancer Genome Atlas (TCGA-LUSC) provides the current study the RNA-seq transcriptome, phenotype and survival data of 501 tumor tissues and 49 normal tissues. As a complement, the Genotype-Tissue Expression (GTEx) provides gene expression profile data of 288 normal lung tissues (GTEx-LUNG). All the above data was obtained from the University of California Santa Cruz (UCSC) Xena platform (https://xenabrowser.net/datapages/). In the first, the R/Bioconductor package “Limma” was used for analysis of differentially expressed genes (DEGs) in 501 tumor tissues compared to 337 normal tissues. The expression profile data of 11 BMPs (BMP2-7, BMP8A, BMP8B, GDF5, GDF7 and GDF10) and 7 BMPRs (ACVR1, ACVR2A, ACVR2B, ACVRL1, BMPR1A, BMPR1B and BMPR2) was extracted from the tumor-normal DEGs. The genes with the absolute value of logFC (the logarithm of fold change) > 1 and p < 0.05 were defined as the DEGs. The results of differential expression analysis were presented in the form of heatmaps, and violin plot visualized the overall expression levels of BMPs/BMPRs in LUSC and normal lung tissues.

Multi-group Differential Expression Analysis Of Bmps/bmprs In Geo Dataset

In order to further explore the expression of BMPs/BMPRs in consecutive developmental stages of LUSC, we obtained a LUSC dataset (GSE33479) from the Gene Expression Omnibus (GEO) database. The dataset provided the gene expression profile data and phenotype data of 9 morphological stages (stages 0–8) of the tumorigenesis of LUSC from 122 biopsies of 77 patients. Stages 0–8 are respectively: normal bronchial tissue with normal fluorescence, normal with low fluorescence, hyperplasia, metaplasia, mild dysplasia, moderate dysplasia, severe dysplasia, carcinoma in situ (CIS) and LUSC. According to the results of Céline Mascaux et al., the 9 consecutive developmental stages were divided into 4 distinct and successive molecular steps (Step A: stage 0–2 subsumed under “normal bronchial tissue”, Step B: stage 3–5 subsumed under “low-grade lesions”, Step C1: stage 6–7 subsumed under “high-grade lesions”, and Step C2: stage 8 subsumed under “LUSC”). Based on the 4 molecular steps, the “Limma” package was also applied for multi-group differential expression analysis of BMPs/BMPRs of LUSC. Samples at the lower molecular step were used as the control group, and the pairwise comparison of samples of four molecular steps was carried out for multi-group difference analysis. The genes with the absolute value of logFC greater than mean of all genes logFC and p < 0.05 were defined as the DEGs. The volcano plots presented the results of six paired differential expression analysis, each of which was controlled by the low-level step among 4 molecular steps. The results of differential expression analysis among 4 steps and 9 stages were showed in the form of heatmaps.

Correlation And Survival Analysis Of Differentially Expressed Bmps/bmprs

The BMPs/BMPRs with both differential expression in TCGA-GTEx-LUSC dataset and regular expression among 4 molecular steps in GEO dataset was the focus of our subsequent analysis. After removing 337 normal tissues, the prognostic values of such differentially expressed BMPs/BMPRs were evaluated in 501 LUSC tissues. The “survminer” and “survival” R packages were performed for drawing of Kaplan-Meier curve. The Kaplan Meier plotter (http://kmplot.com/), an online database capable of assessing the association of genes on survival in 4 types of cancer (lung, breast, gastric and ovarian cancer), was also applied to verify the prognostic value of BMPs/BMPRs in patients with LUSC (n = 524)[13]. The “corrplot” package was used for visualization of correlation with each other among 11 BMPs and 7 BMPRs. The correlation between differentially expressed BMPs/BMPRs and clinicopathologic feature (age, gender, smoke, T-stage, N-stage, M-stage, stage and status) was presented in the form of heatmap.
The Weight Gene Correlation Network Analysis (Wgcna)

To explore the mechanism of the differentially expressed BMPs/BMPRs with the potential as a biomarker assessing the tumorigenesis and progression of LUSC, the weight gene correlation network was constructed by the “WGCNA” package in R[14]. As for RNA-seq transcriptome data of TCGA-LUSC, 5000 genes with the top high median absolute deviation (MAD) value were selected out, and their clustering modules were identified based on clinicopathologic features including the expression of BMPs/BMPRs selected as the above[15]. The highly positively and negatively relevant modules with BMPs/BMPRs were identified for the subsequent function analysis.

Go/kegg Function Enrichment Of Module Genes

The genes in the highly positively and negatively relevant modules with BMPs/BMPRs as the potential biomarker were selected for differential expression analysis to confirm the expression consistency between BMPs/BMPRs and their related module genes. The R package “ClusterProfiler” was performed for Gene Ontology (GO) and KEGG ((Kyoto Encyclopedia of Genes and Genomes) enrichment of the module genes.

Recognition Of Hub Genes And Construction Of The Risk Model

We respectively selected significantly down-regulated genes and up-regulated genes as hub genes in positively and negatively module with BMPs/BMPRs. Based on the hub genes, we implemented LASSO analysis to establish the prognostic risk model. Then, Cox multivariate regression analysis were carried out to further screen the hub genes with differential expression and significant prognostic value (p < 0.001) from the module genes.

Patients And Tumor Samples

Retrospectively collected 33 LUSC patients who received surgical resection at Union Hospital of Huazhong University of Science and Technology, Tongji Medical College (Wuhan, China) from January 1, 2019 to December 30, 2020. Paraffin tissues from these patients were collected at department of Pathology, Union Hospital. All 33 patients were histologically diagnosed with primary lung squamous cell carcinoma with differently pathological stage. Meanwhile, we collected basic clinical information, including age, gender, smoking history, TNM stage. The study was approved by the Institutional Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology, and written informed consent was obtained from all of the patients before surgery.

Immunohistochemistry Staining

Immunohistochemistry (IHC) staining was performed for the paraffin tissue, which contain 33 lung squamous cell carcinoma tissues. Primary antibody of rabbit anti-GDF10 antibody (dilution 1:400; cat.no.CSB-PA009343LA01HU) was purchased from Cusabio Technology (Wuhan, China). The IHC secondary antibody kit (abs957; Absin, Shanghai, China) was used according to the instructions of the manufacturer. We evaluated GDF10 expression levels in each specimen with a semiquantitative immunoreactivity scoring system, which ranged from 1 to 8 and was equal to the sum of the intensity of IHC staining score and the percentage of positive cells score. Protein
expression levels of GDF10 were independently and blindly evaluated via two experienced graduate students of surgery.

**Statistics**

Respectively analyzed Normally and non-normally distributed data using student’s t test and Mann-Whitney tests after a normality check. To evaluate the correlation with each other among BMPs/BMPRs by the Pearson test, and to check the correlation of these DEGs with clinicopathologic features by Kendall rank correlation coefficient tests. To evaluate the prognostic value of BMPs/BMPRs using Kaplan-Meier survival analysis and Cox multivariate regression analysis. IHC score with clinical features analyses were conducted using GraphPad Prism 8. Differences were considered to be statistically significant at $p < 0.05$, and the results were expressed as mean ± S.E.M.

**Results**

**Differential expression characteristics of BMPs/BMPRs in LUSC**

To confirm the role of BMPs/BMPRs in LUSC, we compared their transcript expression levels between 501 tumor tissues (all from TCGA-LUSC) and 337 normal tissues (49 from TCGA-LUAD and 288 from GTEx-LUNG) in the first. Then, 11 BMPs and 7 BMPRs were included in the above differential expression analysis. The results of differential expression analysis for BMPs showed that there is a total of 8 significantly differentially expressed genes (2 upregulated vs. 6 downregulated) in addition to 3 stable expressed genes (BMP3, BMP4, and BMP8B) (Table 1). Compared to BMP8A, BMP7 was the more significantly upregulated BMP ($\logFC = 4.06$, $p \approx 0$); Among 6 downregulated BMPs (BMP2, BMP5, BMP6, GDF5, GDF7, and GDF10), GDF10 was the most significantly downregulated one ($\logFC = -5.87$, $p \approx 0$) (Fig. 2A). As for 7 BMPRs, both ACVRL1 and BMPR2 were significantly downregulated and the other BMPRs belonged to the stable genes in tumor tissues comparing to normal tissues (Table 2), and ACVRL1 ($\logFC = -3.57$, $p \approx 0$) showed the more significantly downregulated than BMPR2 ($\logFC = -1.22$, $p \approx 0$) (Fig. 2B). The violin plot describing the distribution of BMP expression in tumor and normal tissues also supported that GDF10 was the most significantly downregulated BMP (Fig. 2C), and the down-regulating degree of ACVRL1 was obvious from the violin plot of BMPRs (Fig. 2D).

**Expression Characteristic Of Bmps/bmprs With The Tumorigenesis Of Lusc**

As a convenient model, invasive LUSC in smokers made it possible to analyze the expression rules of BMPs/BMPRs in a range of consecutive developmental stages and explore their potential roles of the early evolution of LUSC. In the study of early lung squamous carcinogenesis, Céline Mascaux at el collected and made public a very high-quality dataset GSE33479 containing transcriptome data of 4 molecular steps (9 morphological stages) of the tumorigenesis of LUSC from 122 biopsies of 77 patients.

Taking a qualitative overview of the 4 steps, GDF10 showed the most significant downregulation trend and a gradual decline along with the progress among all BMPs and BMPRs (Fig. 3A). Refined to 9 specific morphological stages (normal group including stage 0–1), GDF10 expression still showed a significant gradient descent with tumor development (Fig. 3B). According to the GDF10 expression levels from high to low, the 9 stages could be reclassified as 3 phases by the boundary of mild dysplasia (stage 4) and carcinoma in situ (stage 7). Before mild
dysplasia, normal, hyperplasia, and metaplasia showed extremely high GDF10 expression levels; In contrast, the expression levels of GDF10 in both carcinoma in situ and LUSC were extremely low; The expression level of GDF10 at dysplasia stages were in the transition of the above two phases.

In order to quantitatively determine the expression pattern of GDF10 in different stages, low-grade lesions were used as the control group and pairwise comparisons were conducted for the four steps (Supplementary Table S1). The results of difference analysis were presented in the form of volcano plots. Consistently with the results of qualitative analysis, no matter how pairwise compare, GDF10 in the late tumorigenesis was downregulated compared to the early molecular steps (Fig. 3C-H). Base on the above, we thought that GDF10 has the potential to quantitatively evaluate the tumorigenesis of LUSC.

**The Prognostic Value Of Gdf10 In Lusc**

As the above, GDF10, as the most down-regulated member of BMP family in tumor tissue, was gradually down-regulated with tumorigenesis of LUSC. Once LUSC was developed, whether GDF10 still maintained the same expression characteristics as above with tumor progression and affects prognosis need to be further confirmed.

To better understand the effect of GDF10 on the clinical outcome of LUSC, we systematically analyzed the correlation between differentially expressed genes including GDF10 and the clinical features including age, gender, smoking, stage, T-stage, N-stage, and M-stage (Fig. 4A). Among the numerous features, GDF10 was only significantly correlated with T stage (p < 0.001), and showed a trend of gradual down-regulation with the progression of T stage, which was consistent with the expression characteristics during tumorigenesis.

In the evaluation of prognostic value of GDF10, the Kaplan-Meier curve (Fig. 4B) based on 501 LUSC survival data from TCGA showed that GDF10 high expression indicated the poor prognosis of LUSC (HR: 1.34, 95% CI: 1.02 ~ 1.76, p = 0.034). The results from the Kaplan Meier plotter from 524 patients with lung squamous cell carcinoma (Fig. 4C) showed that GDF10 high expression had the same prognostic effect but without significance (HR: 1.1, 95% CI: 0.87 ~ 1.40, p = 0.43). The prognostic value of GDF10 in lung squamous cell carcinoma is not significant, which may be affected by clinicopathological factors and related molecules.

Before exploring other GDF10 related molecules, we firstly analyzed the correlation between other deBMPs/BMPRs and GDF10 expression (Fig. 4D). Among 6 downregulated BMPs and 2 BMPRs, GDF10 had the most significant positive correlation with the receptor ACVRL1, and there was only a significant negative correlation with BMP7 in 2 upregulated BMPs.

**The Characteristics Of Gdf10 Co-expression Modules In Lusc**

As described above, the current study calculated the median absolute deviation (MAD) of each gene in 501 tumor tissues and 49 normal tissues of TCGA-LUSC, and the top 5,000 genes with the highest MAD values were selected for co-expression network construction. While moderately retaining the average connectivity of each gene node, we chose the appropriate weighting factor $\beta$ to construct a scale-free network. Finally, the $\beta$ value was determined to be 4 for co-expression network construction (Fig. 5A and B), and a total of 13 modules were then identified (Fig. 5C).
Based on the correlation between modules and clinicopathologic features, the modules significantly associated with GDF10 expression were selected. The magenta module was the most positively correlated one (Coefficient: 0.77, p = 1e-100) and the turquoise module was the most negatively correlated one (Coefficient: -0.84, p = 1e-138) with GDF10 expression (Fig. 5D). Then, both the magenta module including 160 genes and the turquoise module including 1627 genes were chosen as the focus of subsequent research. As the above results show, the genes in the both modules may have certain expression correlation with GDF10, which may play an important role in LUSC.

Function Enrichment Of Gdf10 Co-expression Modules

As the most positively correlated module of GDF10, the magenta module contains 160 genes. Considering the effect of the collinearity of genes in modules, we re-evaluated the correlation between module genes and GDF10 in the form of scatter plot (Fig. 6A). There was a high linear correlation between the absolute expression value of magenta module genes and GDF10 (Coefficient: 0.93, p = 1.4e-70). As the module most positively related to the down-regulated gene GDF10, 72 of the 160 genes in magenta module were significantly down-regulated but none showed an up-regulation trend (Fig. 6B). 72 genes were extracted for subsequent function enrichment analysis as the genes most consistent with the downregulated trend of GDF10. GO function enrichment found that these genes were most significantly enriched in tissue homeostasis, carbohydrate binding and coated vesicle in three terms of biological process, cell component and molecular function, respectively (Supplementary Table S2 and Fig. 6C).

Although there was also a significant linear correlation between the expression absolute value of GDF10 and genes of the turquoise module (Coefficient: 0.89, p = 1e-200) (Fig. 6D), the turquoise module was larger and more complex than magenta module. As the most negatively correlated module, turquoise module contains 1627 genes. Different expression analysis found that 299 genes were significantly down-regulated in addition to 351 significantly up-regulated genes, of which expression difference were inconsistent with their correlation with GDF10 (Fig. 6E). 351 up-regulated genes were extracted for subsequent GO function enrichment analysis as the genes most consistent with the downregulated trend of GDF10 (Supplementary Table S3 and Fig. 6F).

Meanwhile, KEGG function enrichment found that 72 down-regulated genes of magenta module and 351 up-regulated genes of turquoise module were mainly enriched on the cell cycle and DNA function (Fig. 6G). The 72 significantly down-regulated genes in magenta module positively correlated with GDF10 and 351 significantly up-regulated genes in turquoise module negatively correlated with GDF10, they were selected as hub genes for subsequent analysis.

Establishment And Prognosis Value Of Risk Model Based On Gdf10 And Its Co-expressed Modules

Based on the 423 hub genes and GDF10 as the core gene, we implemented LASSO analysis to establish a prognostic risk model (Fig. 7A and B). And the hub genes as potential prognostic factors were filtered out from the 423 genes (Supplementary Table S4). At last, the risk model was constructed with GDF10 as the core gene and 5 others as hub genes (Table 3). The risk model showed better correlation with prognosis, and higher risk scores suggested poorer OS (HR: 1.73, 95% CI: 1.32–2.28, p = 1e-04 < 0.001) in LUSC (Fig. 7C).
To confirm the prognostic value of the 5 hub genes (HRASLS, HIST1H2BH, FLRT3, CHEK2 and ALPL) in the risk model in LUSC. Besides GDF10, the prognostic value of every hub gene in the risk model was evaluated, and Kaplan–Meier curves shown that almost all hub genes had a high prognostic value, especially HRASLS and HIST1H2BH (Figs. 7D–H). The high expression of the 2 genes indicated a better prognosis consistent with GDF10, suggesting that HRASLS and HIST1H2BH may act as the key molecules for prognosis influenced by GDF10 in LUSC.

Expression Characteristic Of Gdf10 Protein With The Progression Of Lusc

According to the above results, expression level of GDF10 as down-regulated member of BMP family in LUSC, was gradually down-regulated with tumorigenesis of LUSC. In order to verify its protein expression characteristics with the progression of LUSC, we collected paraffin specimens from 33 patients who were histologically diagnosed with primary lung squamous cell carcinoma with differently pathological stage for immunohistochemical analysis.

As shown in Table 4, among 33 patients with lung squamous cell carcinoma, the median age of the patients was 63 years (range 52–77), 32 (97%) patients were male and 26 (78.8%) patients had a history of smoking. Except for 11 (33.3%) patients with well-differentiated tumors, the rest (66.7%) patients were diagnosed with poorly differentiated, moderately differentiated, or poorly moderately differentiated tumors. According to the eighth edition of UICC/AJCC lung cancer stage classification (2017), 12 (36.3%) patients were classified as pathological T1 stage and the T2, T3, T4 stage included 10 (30.3%), 7 (21.2%), 4 (12.1%), respectively.

As the above, GDF10 mRNA expression still showed a significant negative correlation with T stage of LUSC. From immunohistochemical of paraffin specimens, representative IHC staining of GDF10 also instructed a significant negative correlation with different pathological T stage in primary tumors (Figs. 8A). Meanwhile, based on the correlation between the IHC score of GDF10 protein expression and clinicopathologic features, GDF10 protein expression was higher in T1 stage than T2-4 stage in LUSC (Figs. 8B). However, GDF10 protein expression was no significance in different pathological TNM stage, N stage, M stage and differentiation grade (Figs. 8C and Supplementary Figure S1).

Discussion

BMPs/BMPRs are belong to the transforming growth factor-beta (TGF-β) that is a superfamily of cytokines[16]. TGF-β/BMP signaling pathway played a diverse range of roles in different tissue types including cell proliferation, apoptosis and differentiation control, not just bone metabolism [17]. Now, many studies have been reported about BMPs/BMPRs to act as both a tumor suppressor or a tumor promoter in cancer [18, 19]. There are BMP2, BMP4, BMP7, and BMP5 that have been demonstrated that played an important in tumorigenesis and progression of lung cancer. However, hardly studies have been known about the molecular mechanism of BMPs/BMPRs in LUSC. Thus, we implemented a study to explore the relationship between BMPs/BMPRs expression levels and the tumorigenesis and progression of lung squamous cell carcinoma. We expected to search for quantifiable biomarkers with more important diagnostic value and improve the diagnostic accuracy of lung squamous cell carcinoma by the study.

Compared the two the RNA-seq transcriptome cohorts of TCGA-LUSC and GTEx-LUNG, we systematically screened out the BMPs/BMPRs differentially expression in LUSC (8 down-regulated vs. 2 up-regulated, with GDF10 down-regulated most significantly). Then, we found that GDF10 mRNA expression showed regular down-regulation with
tumorigenesis of LUSC from GEO database, which had potential diagnostic value in the future. Thus, we
subsequently analyzed the correlation between GDF10 mRNA expression level and the clinical features of LUSC
patients, the result showed that GDF10 was particularly associated with T staging in clinicopathological features.
But, the prognostic value of GDF10 is not significant in lung squamous cell carcinoma, which may be affected by
other related molecules. So, we constructed the weight gene correlation network to screen out the hub genes co-
expressed with GDF10 which were most significantly enriched in biological process and cell cycle. Meanwhile, we
established the risk model based GDF10 as the core gene by LASSO analysis, which showed better correlation with
prognosis and higher risk scores suggested poorer OS of LUSC patients. At last, in order to verify the role of GDF10
protein level in the progression of LUSC, we systematically collected 34 clinical specimens of LUSC to implement
immunohistochemical analysis. We also found that GDF10 protein expression was also negatively correlation with T
age stage of LUSC.

As we all known, BMPs/BMPRs signaling is regulated at multiple levels[20]. Meanwhile, we confirmed that both
mRNA and protein expression of GDF10 were negatively correlation with the tumorigenesis and progression
function of LUSC in our study. Besides, we established a prognostic risk model with the GDF10 as the core gene
and acquired intimately related 5 hub genes (HRASLS, HIST1H2BH, FLRT3, CHEK2 and ALPL). For the reason that,
we further discussed the current research progress of GDF10 and the 5 hub genes in tumor and particularly
explored its potential value in the tumorigenesis and prognosis of lung squamous cell carcinoma.

GDF10(also known as BMP3b) with bone morphogenetic protein-3 (BMP-3) together form a special subgroup of
the BMP family[21]. GDF10 gene encodes a secreted ligand of the TGF-β superfamily of proteins. Ligands of this
family bind various TGF-β receptors leading to recruitment and activation of SMAD family transcription factors
that regulate gene expression[22]. Increasing evidence demonstrated that this protein may act as a tumor
suppressor and reduced expression of this gene is associated with tumor cell growth, invasion and metastasis[23–
26]. These findings demonstrate that GDF10 serve as tumor suppressor gene in breast cancer and prostate cancer
[23, 27]. However, the role of GDF10 in oral squamous cell carcinoma was controversial. GDF10 acts as oncogene
by activating the TGFβRI/Smad3/ERK signaling and promotes the transformation of normal fibroblasts into
cancer-associated fibroblasts[28]. It has also been shown downregulation of GDF10 enhance chemotherapy
resistance and epithelial-mesenchymal transition, GDF10 acts as a prognostic biomarker during
carcinogenesis[29]. It was worth noting that previous studies declared that downregulation of GDF10 transcription
due to DNA hypermethylation in the BMP3B promoter promotes progression in lung cancer cell lines [30–32].
While, the research field is currently unclear the diagnostic value of GDF10 for the tumorigenesis and progression
in lung squamous cell carcinoma, as well as it may in turn provide new insights as prognosis biomarker and
therapeutic target for LUSC patients.

Intimately related with GDF10, the HRAS-like suppressor(HRLSLS) enzymes family (also known as the H-rev107
family) consisted of five member (HRASLS1-5) as tumor suppressor genes negatively regulating the activity of
oncogene Ras and already known to play diverse biological roles in vivo [33]. Meanwhile, previous studies
generally demonstrated the HRASLS family as a direct anti-cancer role in colon cancer cells cervical cancer
cells[34, 35]. Then, HRASLS methylation was closely associated with poor survival in gastric cancer[36]. Like our
study, HRASLS also as a tumor suppressor has been found that could predict prognosis of LUSC[37]. It is
important for HRASLS to demonstrate the regulatory relationship of HRASLS with GDF10 and its role in LUSC.

Now, Targeted programmed cell death protein 1 (PD-1) therapy has been used to improve the long-term prognosis
of patients with advanced lung cancer[38]. Some studies have shown that the expression level of HIST1H2BH was
overexpressed in LUSC. And it was significantly correlated with the expression level of PD-1 in LUSC tissue, but the interaction relationship between HIST1H2BH and PD-1 has not been verified [39]. Meanwhile, HIST1H2BH was also regarded as prognostic marker to predict the survival of LSCC patients [40]. The HIST1H2BH as a subtype of histone H2B performed diverse biological functions, such as DNA repair and posttranslational modifications in tumorigenesis[41]. Meanwhile, we also found that HIST1H2BH as a GDF10 co-expressed hub gene may be play an important role in cell cycle and DNA function by analysis of GO/KEGG function enrichment, and its high expression predicted better prognosis in patients with lung squamous cell carcinoma. Thus, the HIST1H2BH and GDF10 may have a potential to regulate PD-1 expression in immunotherapy, it was worth to explore in our further investigation.

Many studies have linked FLRT3, a type I transmembrane protein containing extracellular leucine-rich repeats, and TGF-β signaling to tumorigenesis and metastasis[42]. Consistently, the studies have demonstrated that inhibition of FLRT3 expression significantly regulates BMP2 expression, and FLRT3 negatively relates to BMP signaling [43]. CHEK2, encodes a protein that serves an important role in DNA repair, is a multiple cancer-predisposing gene, and its mutations lead to increase breast cancer risk and multiple cancer development[44]. ALPL may be play a important role in protein activity regulation and intracellular signaling [45]. It is worth noting that upregulation of ALPL can inhibit invasion and metastasis of LUAD cell [46]. As the co-expression gene of GDF10, the diagnostic value and regulatory relationship with GDF10 are unclear. Thus, we should accomplish prefect experimental results to illustrate their relationship in future exploration in LUSC.

To compare with other studies concentrated on BMPs/BMPRs in lung cancer, the present research is a systematic analysis of BMPs/BMPRs based on large-sample RNA-seq data. In this study, with the fully consideration of the significant differences in histological types of lung cancer, we paid attention to analyze the expression differences and diagnostic value of BMPs/BMPRs in lung squamous cell carcinoma which the 5-year survival rate remains poor and the targeted therapies available are limited. First of all, we found that GDF10 has superior diagnostic value, and obtained the regularly expression of GDF10 with tumorigenesis and progression of lung squamous cell carcinoma by the data from normal tissue to LUSC. Then, we preferentially explored the correlation between GDF10 mRNA expression and clinical characteristics of LUSC. Meanwhile, we further constructed a risk model with good prognostic value based the co-expressed gene network in LUSC.

It is the most important that we retrospectively collected clinical paraffin tissues to verify that GDF10 was also the negatively correlation with tumorigenesis and progression of LUSC at the protein expression level.

While, the limitations of this study must be acknowledged. Firstly, the main data results come from bioinformatics analysis databases (TCGA and GEO databases) with LUSC tumors relative to the non-matching tissues and did not detect the protein expression level. Secondly, although we retrospectively collected LUSC paraffin tissues with different T stage to detect the protein expression, the sample size still needed to be expanded to validate at the protein and mRNA expression levels. Thirdly, the relational molecular mechanism of GDF10 promoting the tumorigenesis and progression of LUSC and its regulatory relationship with co-expressed hub genes need further experimental exploration in the future.

Although there were still many shortcomings, the value of this research should not be ignored. GDF10 consistently demonstrated superior diagnostic potential for the tumorigenesis and progression of LUSC at both mRNA and protein expression levels, providing a new target for diagnosis and therapy of LUSC in the future.

**Conclusions**
Taken together, we found GDF10 mRNA and protein expression levels from high to low accurately reflect the tumorigenesis and progression of lung squamous cell carcinoma in external database and LUSC paraffin tissues. GDF10 may become a diagnostic biomarker and a new therapeutic target for LUSC patients by clinical translation.

**Abbreviations**

BMP, bone morphogenetic protein  
GDF10, growth differentiation factor 10, known as BMP3B

**Declarations**

**Acknowledgments**

The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

**Authors' Contributions**

Peiyuan Mei and Han Xiao contributed to the study design and the provider of important information. Jiaping Chen, Wangyang Meng, Yangwei Wang, Yunchong Meng, Rong Zhao, Wei Lin performed data analysis. Peiyuan Mei, Han Xiao and Yongde Liao wrote the main manuscript texts. All authors revised and edited the manuscript and gave the final approval for the version to be published.

**Funding**

This study was supported by grants from National Natural Science Foundation of China (81572277, 81272590).

**Availability of data and materials**

All data from this study are included in this published article.

**Ethics approval and consent to participate**

The present study was performed with the approval of the Institutional Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology. All aspects of the study complied with the Declaration of Helsinki. All participants provided written informed consent before surgical treatment.

**Patient consent for publication**

Not applicable.

**Competing interests**

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, and the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
References


**Tables**

Table 1 to 3 are available in the Supplementary Files section.

**Table 4**: Clinical characteristics of 33 patients with LUSC.
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No. of patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>33</td>
</tr>
<tr>
<td>Age (median, range)</td>
<td>63 52, 77</td>
</tr>
<tr>
<td>≤63</td>
<td>17 52%</td>
</tr>
<tr>
<td>&gt;63</td>
<td>16 48%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>32 97%</td>
</tr>
<tr>
<td>Female</td>
<td>1 3%</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>26 78.8%</td>
</tr>
<tr>
<td>No</td>
<td>7 21.2%</td>
</tr>
<tr>
<td>Differentiation grade</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>11 33.3%</td>
</tr>
<tr>
<td>Poor-moderate and moderate</td>
<td>18 54.5%</td>
</tr>
<tr>
<td>Moderate-well and well</td>
<td>4 12.1%</td>
</tr>
<tr>
<td>Pathological tumor (T) stage</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>12 36.4%</td>
</tr>
<tr>
<td>2</td>
<td>10 30.3%</td>
</tr>
<tr>
<td>3</td>
<td>7 21.2%</td>
</tr>
<tr>
<td>4</td>
<td>4 12.1%</td>
</tr>
<tr>
<td>Pathological node (N) stage</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>26 78.8%</td>
</tr>
<tr>
<td>1</td>
<td>4 12.1%</td>
</tr>
<tr>
<td>2</td>
<td>3 9.1%</td>
</tr>
<tr>
<td>Pathological metastasis (M) stage</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>30 90.9%</td>
</tr>
<tr>
<td>1</td>
<td>3 9.1%</td>
</tr>
<tr>
<td>Pathological TNM stage</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>14 42.4%</td>
</tr>
<tr>
<td>II</td>
<td>10 30.3%</td>
</tr>
<tr>
<td>III</td>
<td>6 18.2%</td>
</tr>
<tr>
<td>IV</td>
<td>3 9.1%</td>
</tr>
</tbody>
</table>
Figures

Figure 1

Workflow of differential expression of BMPs/ BMPRs and evaluation of their diagnostic and prognostic value in LUSC.
Figure 2

The identification of differentially expressed BMPs/BMPRs in normal lung tissues and lung squamous cell carcinoma. (A) Heatmap of differential expression analysis for 11 BMPs in the TCGA and GTEx database. (B) Heatmap of differential expression analysis for 7 BMPRs in the TCGA and GTEx database. (C) Violin plot indicating the expression difference of 11 BMPs in tumor and normal tissues in the TCGA and GTEx database. (D) Violin plot indicating the expression difference of 7 BMPRs in tumor and normal tissues in the TCGA and GTEx database.
Figure 3

Heatmap of differential expression analysis for BMPs/BMPRs in morphological stages of the carcinogenesis of LUSC. (A) Heatmap of differential expression analysis for 11 BMPs and 7 BMPRs in four distinct and successive molecular steps of progression of LUSC. (B) Heatmap of differential expression analysis for GDF10 in morphological stages of the carcinogenesis of LUSC (stages 0–8). (C) Volcano plots of the expression pattern of BMPs/BMPRs in Step B-A stages of LUSC. (D) Volcano plots of the expression pattern of BMPs/BMPRs in Step C1-A stages of LUSC. (E) Volcano plots of the expression pattern of BMPs/BMPRs in Step C2-A stages of LUSC. (F) Volcano plots of the expression pattern of BMPs/BMPRs in Step C1-B stages of LUSC. (G) Volcano plots of the
expression pattern of BMPs/BMPRs in Step C2-B stages of LUSC. (H) Volcano plots of the expression pattern of BMPs/BMPRs in Step C2-C1 stages of LUSC.

**Figure 4**

The construction and association of significantly differentially expressed genes with clinicopathological features and survival analysis. (A) Heatmap of the association between significantly differentially expressed genes and clinicopathological features including age, gender, smoking, stage, T-stage, N-stage, and M-stage; Heatmap of the association between GDF10 and T-stage (**p < 0.001). (B) The Kaplan–Meier curve of patients with low expression and high expression based on the risk model in TCGA-LUSC (HR: 1.34, 95% CI: 1.02–1.76, p=0.034). (C) The Kaplan–Meier curve of patients with low expression and high expression based on the risk model in online
database (HR: 1.1(0.87-1.4), p=0.43). (D) The correlation among ten differentially expressed BMPs/BMPRs with coefficients in the form of circle size and p value (*p<0.05, **p<0.01, ***p<0.001).

Figure 5

The recognition of modules and their correlation with GDF10 in WGCNA. (A, B) The scale-free index and the average connectivity were both calculated under different $\beta$. The approximate scale-free topology was achieved at a soft threshold power of 4. (C) Gene clustering tree diagram of 13 modules, and each module respectively contains a series of highly connected genes. (D) Correlation heatmap of different modules associated with GDF10 expression and clinicopathologic features in LUSC (the numbers in brackets indicate the p-value, and the numbers without brackets indicate the correlation).
Figure 7

The establishment of the risk model and the evaluation of its prognostic value. (A) Distribution of least absolute shrinkage and selection operator (LASSO) coefficients for GDF10 and 423 hub genes. (B) Partial likelihood deviation of the LASSO coefficient distribution. (C) The Kaplan–Meier curve of patients with low risk and high risk based on the established risk model in TCGA-LUSC (HR: 1.73, 95% CI: 1.32–2.28, p=1e-04<0.001). (D-H) The Kaplan–Meier curves of HRASLS, HIST1H2BH, FLRT3, CHEK2 and ALPL in the tumor samples of TCGA-LUSC (n=493). HRASLS (HR: 0.67, 95% CI: 0.51–0.88, p=0.0045<0.01), HIST1H2BH (HR: 0.69, 95% CI: 0.53–0.91, p=0.0086<0.01), FLRT3 (HR: 1.41, 95% CI: 1.08–1.86, p=0.013<0.05), CHEK2 (HR: 0.73, 95% CI: 0.56–0.96, p=0.024<0.05) and ALPL (HR: 1.33, 95% CI: 1.02–1.75, p=0.0377<0.05).
Figure 8

Protein expression levels of GDF10 in patients with lung squamous cell carcinoma. (A) Representative IHC staining of GDF10 in primary tumors from different pathological T stage. (B) The IHC score of GDF10 in T1 stage was higher than T2-4 stage (*p < 0.05). (C) The IHC score of GDF10 in different pathological TNM stage (ns. no significance).

Supplementary Files

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

- SupplementaryFigureS1.docx
- SupplementaryMaterialslegends.docx
• Supplementary Table S1: The differentially expressed BMPs and BMPRs by pairwise comparisons in the four steps.xlsx
• Supplementary Table S2: GO enrichment results of downregulated DEGs in magenta.xlsx
• Supplementary Table S3: GO enrichment results of upregulated DEGs in turquoise.xlsx
• Supplementary Table S4: The hub genes as potential prognostic factors in magenta and turquoise module genes.xlsx
• Table 1: DEGBMPs.csv
• Table 2: DEGBMPRs.csv
• Table 3: The established riskscore including GDF10 and 5 hub genes1.csv