Association of PD-1 polymorphisms with the risk and prognosis of lung adenocarcinoma in the northeastern Chinese Han population

**CURRENT STATUS:** ACCEPTED

**BMC Medical Genetics**

Kun Huang
the Second Affiliated Hospital of Harbin Medical University

Erqiang Hu
Harbin Medical University

Wan Li
Harbin Medical University

Junjie Lv
Harbin Medical University

Yuehan He
Harbin Medical University

Gui Deng
Harbin Medical University

Jinling Xiao
the Second Affiliated Hospital of Harbin Medical University

Chengcheng Yang
the Second Affiliated Hospital of Harbin Medical University

Xinyu Zhao
the Second Affiliated Hospital of Harbin Medical University

Lina Chen chenlina@ems.hrbmu.edu.cn
Harbin Medical University

*Corresponding Author*

**ORCiD:** 0000-0002-4576-6814

Xinyan Wang
the Second Affiliated Hospital of Harbin Medical University

**DOI:**
10.21203/rs.2.11652/v1
SUBJECT AREAS
  Medical Genetics

KEYWORDS
  Lung adenocarcinoma staging, PD-1, polymorphism, SNPs
Abstract

Introduction Lung cancer is a leading cause of death from cancer worldwide, especially non-small cell lung cancer (NSCLC). The marker of progression in lung adenocarcinoma has been rarely studied. PD-1 is an effective drug target for the treatment of non-small cell lung cancer. The study of the effect of polymorphism on the progression of lung adenocarcinoma in the Han population of Northeast China may provide a valuable reference for the research and application of these drugs. Methods Chi-square test, Wilcoxon rank sum test, and classification efficacy assessment were used to teste SNPs of PD-1 in 287 patients and combined with clinical information. Results We successfully identified biomarkers (rs2227981, rs2227982, and rs3608432) that could distinguish the early stages and late stages of lung adenocarcinoma. Multiple clinical indicators showed significant differences among different SNP typing and cancer stages. Furthermore, this gene was confirmed to effectively distinguish the staging of lung adenocarcinoma with RNA-seq data in TCGA. Conclusions This indicated that the PD-1 gene and the SNPs on it could be used as markers for distinguishing lung adenocarcinoma staging in the Northeast Han population.

Introduction

Lung adenocarcinoma is the most common type of lung cancer and the leading cause of cancer death worldwide may develop as a result of the interaction between environmental risk factors and individual genetic susceptibility (Cancer Genome Atlas Research 2018). Therefore, genetic susceptibility to lung adenocarcinoma has become an area of interest in lung cancer research (McKay et al. 2017) (Lu et al. 2019). Almost all lung adenocarcinoma patients were diagnosed in advanced stages (IIIB and IV) (Group 2017). Lung adenocarcinoma develops silently and has no specific symptoms. Therefore, early
lung cancer (Stage I and Stage II) is difficult to find. So accurate pathological staging is an important factor in the treatment of lung adenocarcinoma and plays an important role in the selection of postoperative treatment and prognosis.

PD-1 is currently the drug target of non-small cell lung cancer (Huang et al. 2018). Various studies have evaluated the relationship between PD-1 polymorphism and various cancer risks, such as rs2227981 polymorphism was associated with breast cancer, colon cancer, and Head and neck squamous cell carcinomas (Salmaninejad et al. 2018)(Qiu et al. 2014)

Whether PD-1 gene polymorphism has an effect on the therapeutic effect and disease progression of lung adenocarcinoma remains to be studied.

In order to combine the lung adenocarcinoma staging information and clinical information, four polymorphisms of the PD-1 gene, including rs2227981 (PD-1.5), rs2227982 (PD-1.9), rs36084323 (PD-1.1) and rs7421861, were selected, and the genotypes of 287 patients were determined by SNaPshot primer extension assays.

Materials And Methods

Subjects

The subjects comprised 111 healthy controls and 287 patients who were diagnosed with lung adenocarcinoma at the Department of Respiratory Medicine of the Second Affiliated Hospital of Harbin Medical University (Harbin, Heilongjiang, China) between 2013 and 2015. The patients were diagnosed by surgery and pathological assessment. The healthy control group was randomly selected from 111 age-matched healthy subjects without any history of familial or personal autoimmune diseases or malignancies, who received annual physical examinations at the same hospital. According to lung cancer classification developed by WHO in 2009, the histological classification and stage were assessed. These samples of patients were collected prior to treatment. All patients and controls were
recruited from the northeastern Chinese Han population. The smoking status of each patient was recorded, and the subjects were divided into non-smokers and smokers. Table 1 (in the Supplementary Files) lists the clinicopathological features of the patients and the controls. The written informed consent for blood collection and subsequent analysis was provided by each participant. The ethics committee of the same hospital approved the study.

**DNA extraction and genotyping**

The TIANamp Blood DNA Kit (TIANGEN, Beijing, China) was used to extract genomic DNA from 500 μl of EDTA-anticoagulated venous blood samples. Four SNPs of the PD-1 gene were genotyped: rs2227981, rs2227982, rs36084323 and rs7421861. Genotype was assayed by SNaPshot Multiplex Kit (PE Applied Biosystems, Warrington, UK and Foster City, CA, USA). Primer 3.0 was used to design the primers for use in PCR amplification. The primer sequences for each SNP were as follows: rs2227981, 5’-

TCTCCTGAGGAAATGCGCTGAC-3’ (forward) and 5’-TGGTGTCCCGATGCAGCTACAGA-3’
(reverse); rs2227982, 5’-TCTCCTGAGGAAATGCGCTGAC-3’ (forward) and 5’-

TGGTGTCCCGATGCAGCTACAGA-3’ (reverse); rs36084323, 5’-CTCCCATTCTGTCGGAGCCTCT-3’
(forward) and 5’-GAAGGGGAGGTCAGCCTCACAG-3’ (reverse); and rs7421861, 5’-

CCCAGCTGGAATGTCAATGAGAA-3 (forward) and 5’-TTACACTCCCTGTGAGGAGGC-3
(reverse).

PCR was performed with 1 L of DNA sample, 3.0 mM Mg^{2+}, 1× GC-I buffer (Tahara), 1 L multiple PCR primers, 0.3 mM dNTP and 1 unit HotStarTaq polymerase (Qiagen, Inc.) in a total volume of 20 L. The PCR cycling program was as follows: 95°C for 2 min; followed by 11 cycles of 94°C for 20 sec, 65°C (decreased 0.5°C per cycle) for 40 sec and 72°C for 90 sec; plus 24 cycles of 94°C for 20 sec, 59°C for 30 sec and 72°C for 90 sec; with a final
extension at 72°C for 2 min and 4°C forever. Next, 5 units shrimp alkaline phosphatase and 2 units Exonuclease I was added to the PCR product, incubated at 37°C for 1 h and inactivated at 75°C for 15 min for purification. SNaPshot multiple single base extension reaction was performed using 5 μL SNaPshot Multiplex Kit (Applied Biosystems), 0.5 L 5’ ligase primer mixture (1.2 μM), 0.5 L 3’ ligase primer mixture (1.6 μM), 2 L ddH₂O and 2 L purified PCR product in a final volume of 10 L. The reactions were cycled as follows: 28 cycles of 96°C for 10 sec, 55°C for 5 sec and 60°C for 30 sec, with the products subsequently kept at 4°C. Purified extension product 0.5 L was then combined with 0.5 μL Liz120 Size Standard and 9 μL Hi-Di, inactivated at 95°C for 5 min and then sequenced and analyzed using an ABI 3730XL DNA Analyzer and GeneMapper 4.1 (Applied Biosystems Co.Ltd.USA), and the nucleotide at each SNP site was identified and recorded.

Statistical analysis

1) Hardy-Weinberg equilibrium

The statistical test of Hardy-Weinberg equilibrium has been an important tool for detecting genotyping errors in the past and is still important in the quality control of next-generation sequence data. The chi-square test was used to test whether the 4 SNPs were in Hardy–Weinberg equilibrium with Excel.

2) Chi-square test

The chi-square test gives evidence of association or no association. The difference between the observed frequency and the expected frequency can be assessed by a statistical test called 2 (Pandis 2016). The statistical formula for this test is the following:

\[ 2 = \sum (O - E)^2 / E \]  

Where O is the observed cell frequency, E is the expected cell frequency, and S is the sum of all cells in the table. The P value of the test was calculated by R program.

3) Wilcoxon rank sum test
The Wilcoxon rank sum test is a method often used in statistical practice to compare position measurements where the underlying distribution is far from normal or not known in advance (Rosner et al. 2003). We used the Wilcoxon rank sum test to analyze the differences between six clinical indicators with different SNP genotyping.

4) Logistic regression analyses

Logistic regression analyses were performed using IBM SPSS (Statistical Package for the Social Sciences) Statistics ver. 17.0.

5) Classification efficacy assessment

To evaluate the classification efficiency of genes and clinical indicators, the Support Vector Machine (SVM) method was employed to construct a classifier for cancer and normal samples or early stage and late stage of cancer. Leave-one-out cross-validation (LOOCV) was carried out to assess the performance. The receiver operating characteristic (ROC) curves were plotted and the areas under the curves (AUC) were computed.

Result

Association of TNM stages and six clinical indicators

For the six clinically indicators related to lung adenocarcinoma, the TNM staging of the patients was investigated, and the Wilcoxon rank sum test was used to obtain the significant p values of the indicators at different stages (Figure 1). The results indicated that there were significant differences in the 6 indicators between the early stage (TNM I and TNM II) and the late stage (TNM III and TNM IV). The significance of two stages (early and late) was obviously better than that of four stages. So the subsequent analysis was carried out in two stages.

Figure 1: Association of TNM stages and six clinical indicators

The higher the column in the graph, the more significant it is. The red line represents the
0.05 threshold.

**Association of PD-1 SNPs with lung adenocarcinoma in different models**

For the four SNPs (rs2227981, rs2227982, rs3608432, rs7421861), we used chi-square test of three models (allele model, dominant model, recessive model) to test the difference of SNP genotypes between two stages in all lung adenocarcinoma samples, male samples, and female samples separately (Figure 2). These results demonstrated that rs2227981 was significantly correlated with lung adenocarcinoma stages in allele model and recessive model. The women samples were significantly correlated with lung adenocarcinoma stages in all three models. rs2227982 and rs36084323 were significantly correlated with lung adenocarcinoma staging in all three models. The rs7421861 in male samples were significantly correlated with lung adenocarcinoma stages in allele model and dominant model. Therefore, rs2227981, rs2227982, rs3608432, and rs7421861 were expected to be markers to distinguish lung adenocarcinoma stages.

**Figure 2: Association of PD-1 SNPs with lung adenocarcinoma in different models**

A: allele model  B: dominant model  C: recessive model

The dotted red line indicates p-value=0.05

Furthermore, the correlation between lung adenocarcinoma stages and SNP genotypes, smoking, sex, and age were analyzed by logistic regression model under dominant and recessive models, respectively (Figure 3). Except for rs7421861, the other three SNPs were significantly correlated with the staging of lung adenocarcinoma in both dominant and recessive models.

**Figure 3: P_value of logistic regression model under dominant and recessive models**

The dotted red line indicates p-value=0.05

In addition, we also used Haplovew software to test the correlation between SNP haplotypes and lung adenocarcinoma staging (Table 3 and Table 4 in the Supplementary...
Files). Among the four haplotypes, three showed a significant correlation with lung adenocarcinoma stages. This further indicated that the four SNPs on PD-1 can be used as potential markers for lung adenocarcinoma staging.

The difference of Six Clinical Indicators in different Stages and different SNP genotyping

We examined the difference of six clinical indicators in the staging of lung adenocarcinoma (Figure 4). There were significant differences in the six indicators between the early stage and late stage samples, either using the whole sample or the male and female samples were used separately.

Figure 4: Difference of Six Clinical Indicators in different Stages
A: CEA B: NLR C: LYM D: GRAN E: WBC F: LDH

Then the association between each SNP and clinical indicators were examined in the whole samples, early stage samples, and late stage samples respectively (Figure 5).

Figure 5: Association between SNP genotyping and six indicators
A: rs2227981 B: rs2227982 C: rs36084323 D: rs7421861

The dotted red line indicates p_value=0.05

It can be seen that in lung adenocarcinoma, the correlation between four SNPs and six clinical indicators shows some differences between men and women samples.

For all the samples, there was a significant correlation between rs2227981 and three indicators: CEA, NLR, and GRAN. The genotyping of rs2227982 and rs36084323 were significantly correlated with NLR, GRAN, and WBC. The rs7421861 typing was significantly correlated with LDH.

For male samples alone, there was a significant correlation between rs2227981 and three indicators: CEA, LYM, and LDH. The genotyping of rs2227982 and rs36084323 were
significantly correlated with LYM, GRAN, and WBC. The rs7421861 typing was significantly correlated with LDH.

For female samples alone, there was a significant correlation between rs2227981 and five indicators: CEA, NLR, GRAN, WBC, and LDH. The rs2227982 and rs36084323 were significantly correlated with CEA, NLR, and LYM. The rs7421861 was significantly correlated with LDH.

The differences in gender between the correlations of these SNPs and clinical indicators may provide a reference for the clinical test results of patients.

Classification efficacy evaluation of genes and clinical indicators

To evaluate the classification efficiency of PD-1, its target gene PD-L1 and coding genes of CEA and LDH, the Support Vector Machine (SVM) method was employed to construct a classifier for early stage and late stage samples, based on these genes. The gene expression data and clinical data were obtained from TCGA(https://cancergenome.nih.gov/). Leave-one-out cross-validation (LOOCV) was carried out to assess the performance. The receiver operating characteristic (ROC) curves were plotted and the areas under the curves (AUC) were computed(Table 5 in the Supplementary Files). It can be seen from the results that when PD-1 and PD-L1 genes are used, AUC is greater than 0.75 under different gender conditions, and was greater than that of LDH and CEA-related genes.

Then the six indicators were also tested (Table 6 in the Supplementary Files). The result indicated that CEA, GRAN, LDH, and NLR have good classification efficiency in different genders. LYM has good classification efficiency in female samples.

Discussion

Programmed death 1 (PD-1) and its ligand, programmed death ligand 1 (PD-L1) were drug
targets of lung adenocarcinoma. In-depth study of their SNPs in the staging of lung adenocarcinoma may be helpful for the diagnosis and treatment of patients. Therefore, we selected 111 normal and 287 patients for the DNA extraction and genotyping, and the corresponding clinical indicators and stage information were analyzed. There was a certain correlation between four SNPs and the six clinical indicators. The significance of the correlation between clinical indicators and staging of lung adenocarcinoma of the two stages was higher than that of four stages. For two stages, PD-1 gene polymorphisms at three investigated positions and their haplotypes were associated with the risk and prognosis of lung adenocarcinoma in three genetic models (allele model, dominant model, and recessive model).

Based on SVM and LOOCV, CEA, LDH, and NLR could distinguish stages among the six clinical indicators. PD-1 and PD-L1 could distinguish stages significantly, and their classification efficiency was higher than that of CEA and LDH related genes. In view of the high classification efficiency of PD-1 and PD-L1 for staging and the significant correlation between SNPs, haplotypes on them and staging, the polymorphism of PD-1 may be used as a marker for the staging of lung adenocarcinoma. Nevertheless, the data may be limited, for clinical data and SNP genotyping were obtained by our experiments, and the RNA-seq data were downloaded from TCGA, further evidence from other studies across different populations incorporating with the stage of cancers is required in order to confirm or refute the findings of this study.

Declarations

Funding

This work was supported, in part, by the National Natural Science Foundation of China (Grant No. 61272388, 81627901, 61702141, 81627901, 81471724, 31210103913)
81101088, and 81130028); the Innovative Scientific Research Funding Project of Harbin Medical University (2017JCZX46); the Harbin Applied Technology Research and Development Project (Grant No. 2016RQQXJ105); the National Basic Research Program of China (2015CB931800); and the University Student Innovation and Entrepreneurship Training Program in Heilongjiang Province (Grant No. 201610226012)

Acknowledgments

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

References

Blank C, Mackensen A (2007) Contribution of the PD-L1/PD-1 pathway to T-cell exhaustion: an update on implications for chronic infections and tumor evasion Cancer immunology, immunotherapy : CII 56:739-745 doi:10.1007/s00262-006-0272-1

Blank CU, Haanen JB, Ribas A, Schumacher TN (2016) CANCER IMMUNOLOGY. The "cancer immunogram" Science 352:658-660 doi:10.1126/science.aaf2834

Cancer Genome Atlas Research N (2018) Author Correction: Comprehensive molecular profiling of lung adenocarcinoma Nature 559:E12 doi:10.1038/s41586-018-0228-6

Group GBDNDC (2017) Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015 The Lancet Neurology 16:877-897 doi:10.1016/S1474-4422(17)30299-5

Hua Z et al. (2011) PD-1 polymorphisms are associated with sporadic breast cancer in Chinese Han population of Northeast China Breast cancer research and treatment 129:195-201 doi:10.1007/s10549-011-1440-3

Huang Q, Zhang H, Hai J, Socinski MA, Lim E, Chen H, Stebbing J (2018) Impact of PD-L1 expression, driver mutations and clinical characteristics on survival after anti-PD-1/PD-L1
immunotherapy versus chemotherapy in non-small-cell lung cancer: A meta-analysis of randomized trials Oncoimmunology 7:e1396403 doi:10.1080/2162402X.2017.1396403

Jin G et al. (2009) Common genetic variants on 5p15.33 contribute to risk of lung adenocarcinoma in a Chinese population Carcinogenesis 30:987-990 doi:10.1093/carcin/bgp090

Lu S et al. (2019) EGFR and ERBB2 Germline Mutations in Chinese Lung Cancer Patients and Their Roles in Genetic Susceptibility to Cancer Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer 14:732-736 doi:10.1016/j.jtho.2018.12.006

McKay JD et al. (2017) Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes Nature genetics 49:1126-1132 doi:10.1038/ng.3892

Pandis N (2016) The chi-square test American journal of orthodontics and dentofacial orthopedics : official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics 150:898-899 doi:10.1016/j.ajodo.2016.08.009

Qiu H, Zheng L, Tang W, Yin P, Cheng F, Wang L (2014) Programmed death-1 (PD-1) polymorphisms in Chinese patients with esophageal cancer Clinical biochemistry 47:612-617 doi:10.1016/j.clinbiochem.2013.12.023

Rosner B, Glynn RJ, Lee ML (2003) Incorporation of clustering effects for the Wilcoxon rank sum test: a large-sample approach Biometrics 59:1089-1098

Salmaninejad A et al. (2018) PD-1 and cancer: molecular mechanisms and polymorphisms Immunogenetics 70:73-86 doi:10.1007/s00251-017-1015-5

Ye L, Leslie C, Jacques A, Mesbah Ardakani N, Amanuel B, Millward M (2019) Programmed death ligand-1 expression in non-small cell lung cancer in a Western Australian population and correlation with clinicopathologic features Modern pathology : an official journal of
Tables
Due to technical limitations, tables 1 through 6 are only available as downloads in the supplemental files section.

Figures
Figure 1

Association of TNM stages and six clinical indicators. The higher the column in the graph, the more significant it is. The red line represents the 0.05 threshold.
Figure 2

Assocation of PD-1 SNPs with lung adenocarcinoma in different models A: allele model B: dominant model C: recessive model The dotted red line indicates p_value=0.05
Figure 3

P_value of logistic regression model under dominant and recessive models. The dotted red line indicates p_value=0.05.
Figure 4: Difference of Six Clinical Indicators in different Stages A: CEA B: NLR C: LYM D: GRAN E: WBC F: LDH Then the association between each SNP and clinical indicators were examined in the whole samples, early stage samples, and late stage samples respectively (Figure 5).
Figure 5

Association between SNP genotyping and six indicators A: rs2227981 B: rs2227982 C: rs36084323 D: rs7421861 The dotted red line indicates

$p\_value=0.05$

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

Table 1.jpg
Table 5.jpg
Table 3.jpg
Table 6.jpg
