Gene 587 (2016) 178–182

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Genetic variants in chromatin-remodeling pathway associated with lung cancer risk in a Chinese population

Liguo Geng [a,1], Meng Zhu [a,1], Yuzhuo Wang [a], Yang Cheng [a], Jia Liu [a], Wei Shen [a], Zhihua Li [a], Jiahui Zhang [a], Cheng Wang [a], Guangfu Jin [a,b], Hongxia Ma [a,b], Hongbing Shen [a,b], Zhibin Hu [a,b,*], Juncheng Dai [a,b,*]

[a] Department of Epidemiology and Biostatistics, Collaborative Innovation Center For Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing 211166, China
[b] Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center of Cancer Medicine, Nanjing Medical University, Nanjing 211166, China

ARTICLE INFO

Article history:
Received 8 March 2016
Received in revised form 10 May 2016
Accepted 10 May 2016
Available online 12 May 2016

Keywords:
Lung carcinoma
Chromatin remodeling
Polymorphism
Association study

ABSTRACT

Chromatin remodeling complexes utilize the energy of ATP hydrolysis to remodel nucleosomes and have essential roles in transcriptional modulation. Increasing evidences indicate that these complexes directly interact with numerous proteins and regulate the formation of cancer. However, few studies reported the association of polymorphisms in chromatin remodeling genes and lung cancer. We hypothesized that variants in critical genes of chromatin remodeling pathway might contribute to the susceptibility of lung cancer. To validate this hypothesis, we systematically screened 40 polymorphisms in six key chromatin remodeling genes (*SMARCA5*, *SMARCC2*, *SMARCD2*, *ARID1A*, *NR3C1* and *SATB1*) and evaluated them with a case-control study including 1341 cases and 1982 controls. Logistic regression revealed that four variants in *NR3C1* and *SATB1* were significantly associated with lung cancer risk after false discovery rate (FDR) correction [For *NR3C1*, rs9324921: odds ratio (OR) = 1.23, *P* for FDR = 0.029; rs12521436: OR = 0.85, *P* for FDR = 0.040; rs4912913: OR = 1.17, *P* for FDR = 0.040; For *SATB1*, rs6808523: OR = 1.33, *P* for FDR = 0.040]. Combing analysis presented a significant allele-dosage tendency for the number of risk alleles and lung cancer risk ($P_{trend} < 0.001$). Moreover, expression quantitative trait loci (eQTL) analysis revealed that these two genes were differently expressed between lung tumor and adjacent normal tissues in the database of The Cancer Genome Atlas (TCGA) ($P = 0.009$ for rs6808523). These findings suggested that genetic variants in key chromatin remodeling genes may contribute to lung cancer risk in Chinese population. Further large and well-designed studies are warranted to validate our results.

1. Introduction

Lung cancer is the most frequent cause of cancer-related death worldwide (Siegel et al., 2015). In China, the estimated number of new cases and deaths in 2015 were 733,300 and 610,200, respectively, being the leading one of all cancers (Chen et al., 2016). Although most lung cancer cases can be attributed to tobacco consumption, previous candidate gene studies and subsequent genome-wide association studies (GWAS) indicated that genetic components also played an important role on lung cancer carcinogenesis (Ma et al., 2009; Amos et al., 2008; Hu et al., 2011; Wang et al., 2008).

Chromatin remodeling complexes mainly consist of SWI/SNF (switching defective/sucrose non-fermenting) family, ISWI (imitation switch) family, INO80 (inositol requiring 80) family and CHD (chromodomain, helicase, DNA-binding) family (Gaspar-Maia et al., 2011). Each family utilizes the energy of ATP hydrolysis to mobilize nucleosomes along DNA, expels histones from DNA, promotes the exchange of histone variants, and eventually alters the structure of nucleosome and modulates DNA accessibility (Becker and Horz, 2002). These ATP-dependent chromatin remodeling complexes regulate a wide range of cellular processes, including DNA replication, DNA-damage response and transcription regulation, and are associated with neoplastic transformation (Wang et al., 2007). Increasing evidences have shown the contribution of polymorphisms in chromatin remodeling genes to the cancer carcinogenesis, such as *ARID1A* for gastric cancer (Wang et al., 2011). However, few studies referred to the association of chromatin remodeling genes polymorphisms and lung cancer, particularly in Chinese population. Therefore, we hypothesized that polymorphisms in chromatin remodeling genes might contribute to the susceptibility of lung cancer.

To validate this hypothesis, we systematically selected six key genes (*SMARCA5*, *SMARCC2*, *SMARCD2*, *ARID1A*, *NR3C1* and *SATB1*) in chromatin remodeling pathway, which had been reported associated with cancer in previous studies (Jin et al., 2014; Kim et al., 2013; Cruickshank

Abbreviations: TCGA, The Cancer Genome Atlas; GWAS, genome-wide association studies; eQTL, expression quantitative trait loci; SNP, single nucleotide polymorphism; MAF, minor allele frequency; HWE, Hardy–Weinberg equilibrium; LD, linkage disequilibrium; CI, confidence interval; OR, odds ratio; FDR, false discovery rate.

* Corresponding authors at: Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, 101 Longmian Rd., Nanjing 211166, China.
E-mail addresses: zhibin_hu@njmu.edu.cn (Z. Hu), djc@njmu.edu.cn (J. Dai).

[1] Liguo Geng and Meng Zhu contributed equally to this work.

http://dx.doi.org/10.1016/j.gene.2016.05.013
0378-1119/

et al., 2015; Han et al., 2015; Pan et al., 2011; Cheng et al., 2014), and evaluated the association of these genetic variants and lung cancer risk with a case-control design including 1341 cases and 1982 controls.

2. Materials and methods

2.1. Study population

This study was approved by the Institutional Review Board of Nanjing Medical University. The recruitment of all individuals was described previously (Shen et al., 2015). Briefly, 1341 lung cancer cases and 1982 cancer-free controls were recruited, the cases were newly diagnosed and consecutively recruited from the Cancer Hospital of Jiangsu Province and the First Affiliated Hospital of Nanjing Medical University since 2003. All cases were histopathologically or cytologically (sputum, bronchial brush/wash/lavage, pleural effusions and lymph nodes) confirmed by at least 2 local pathologists. We excluded those who had a history of malignancy, metastasized cancer from other organs and received chemotherapy or radiotherapy. Cancer-free controls were selected from a community-based screening program for non-infectious disease in Jiangsu Province and were frequency-matched to the cases by age and gender.

After signing informed consent, all subjects offered 5-ml venous blood and were interviewed face-to-face by trained interviewers to collect demographic data (e.g. age and gender) and exposure information (e.g. smoking history). Individuals who smoked >1 cigarette per day for >1 year in their lifetime were defined as current smokers; smokers who had quit smoking for >1 year were classified as former smokers; all others were regarded as nonsmokers. Pack-years of smoking were defined as packs per day × smoking years. According to the median pack-years of smoking among smokers in control group (25 pack-years), smokers were divided into light and heavy smokers.

2.2. Polymorphisms selection and genotyping

We selected single nucleotide polymorphisms (SNPs) in six key genes by combining potentially functional SNPs as well as tagging SNPs. The detailed strategies were similar with previous study (Shen et al., 2015). Firstly, the public HapMap Project database (phase II + III Feb. 09, on NCBI B36 assembly, dbSNP b126) was used to search all SNPs (in Chinese Han population) in respective gene region (including 10 kb up-stream region of each gene), following a series of criteria: minor allele frequency (MAF) ≥ 0.05; Hardy–Weinberg equilibrium (HWE) ≥ 0.05; Call rate > 90%. Secondly, the potentially functional SNPs were marked by SNPinfo Web Server (http://snpinfo.niehs.nih.gov/). Finally, the HaploView 4.2 software was used to select tagging SNPs with r^2 of pairwise linkage disequilibrium (LD) <0.8. All potentially functional SNPs were forced to include as tagging SNPs. As a result, a total of 47 SNPs were finally chosen for genotyping.

The isolation of genomic DNA and the method of genotyping were described in previous study (Shen et al., 2015). Briefly, genomic DNA was isolated from leukocyte pellets of venous blood by proteinase K digestion and extracted by phenolchloroform. The genotyping was performed by Illumina Infinium® BeadChip (Illumina Inc.) without knowing the status of case and control. Genotype calling was performed using the GenTrain version 1.0 clustering algorithm in GenomeStudio V2011.1 (Illumina). Finally, 40 SNPs were successfully genotyped with call rate > 95% (Table 2).

2.3. The Cancer Genome Atlas (TCGA) database for lung cancer

TCGA Database contains various types of genomic data from a wide variety of cancers including lung cancer. Thus, we downloaded the genotype data and Expectation-Maximization (RSEM) normalized read counts (level 3) in lung adenocarcinoma (LUAD) and lung squamous carcinoma (LUSC) from TCGA released at July, 2014 (https://tcga-data.nci.nih.gov/tcga/). The RSEM included tumor tissues and corresponding adjacent normal tissues. We referred to RSEM as relative mRNA expression levels.

After integrating RSEM data, a total of 107 samples (57 LUAD and 50 LUSC) with both tumor tissues and adjacent normal tissues were finally used to explore the expression difference of specific gene (tumor tissues vs. corresponding adjacent normal tissues).

Meanwhile, through integrating genotype data and RSEM matrix (in tumor tissues), we obtained 962 samples (481 LUSC and 481 LUAD) to perform expression quantitative trait loci (eQTL) analysis on the SNPs vs. expression level (log10 transformed) of corresponding genes.

2.4. Statistical analysis

The χ^2 test for categorical variables and student's *t*-test for continuous variables were used to evaluate the differences of demographic information and smoking consumption between the cases and controls. HWE was evaluated using the goodness-of-fit χ^2 test among controls. Odds ratios (ORs) and their 95% confidence intervals (CIs) were calculated by performing logistic regression analyses to evaluate the association between genotypes and lung cancer risk, with adjustments for age, gender and pack-years of smoking. Gene-based analysis was performed using the sequence kernel association test (SKAT). The heterogeneity between subgroups was assessed by χ^2-based Cochran's *Q*-test. For eQTL analysis, a linear regression analyses was conducted to assess the correlation between genotypes and transcript expression levels (log10 transformed) of genes. Paired Student's *t*-test was done to compare expression difference of gene between tumor tissues and corresponding adjacent normal tissues. We defined statistical significance by using false discovery rate (FDR) and set the significance level at 0.05 for single-variant analysis. All of the statistical analyses were performed with R software (version 3.2.0; The R Foundation for Statistical Computing, http://www.cran.r-project.org/).

3. Results

The distributions of variables between cases and controls were summarized in Table 1. Briefly, the cases and controls were comparable in age and gender ($P > 0.05$). Compared to cases, controls had a lower rate of smoking (current & former: 48.54% vs 61.08%, $P < 0.001$).

The genotype distributions of the 40 SNPs and their associations with lung cancer risk were presented in Table 2. Logistic regression analyses revealed that the A allele of rs12521436 (*NR3C1*) was significantly associated with the decreased risk of lung cancer (OR = 0.85, 95% CI: 0.77–0.94, *P* for FDR = 0.040). Additionally, A allele of rs9324921 (*NR3C1*), G allele of rs4912913 (*NR3C1*) and G allele of rs6808523 (*SATB1*) were significantly correlated with the increased

Table 1
Distributions of select variables in lung cancer cases and cancer-free controls.

Variables	Case (N = 1341)	Control (N = 1982)	*P*
Age(mean ± sd)	61.06 ± 10.15	61.32 ± 11.07	0.473
≤60	596 (44.44%)	883(44.55%)	0.980
>60	745(55.56%)	1099(55.45%)	
Gender			
Male	949(70.77%)	1358(68.52%)	0.179
Female	392(29.23%)	624(31.48%)	
Smoking status			
Current	634(47.28%)	876(44.20%)	<0.001
Former	185(13.80%)	86(4.34%)	
Never	522(38.93%)	1020(51.46%)	
Pack-year(py)			
≤25	774(57.72%)	1505(75.93%)	<0.001
>25	567(42.28%)	477(24.07%)	
Histology type			
Squamous cell carcinoma	481(35.87%)		
Adenocarcinoma	860(64.13%)		