Genetic variants in regulatory regions of microRNAs are associated with lung cancer risk

Kaipeng Xie1,2,*, Cheng Wang1,2,*, Na Qin1,2, Jianshui Yang1,2, Meng Zhu1,2, Juncheng Dai1,2, Guangfu Jin1,2, Hongbing Shen1,2, Hongxia Ma1,2, Zhibin Hu1,2

1Department of Epidemiology and Biostatistics, Collaborative Innovation Center of Cancer Medicine, School of Public Health, Nanjing Medical University, Nanjing, 211166, China
2Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, School of Public Health, Nanjing Medical University, Nanjing, 211166, China
*These authors contributed equally to this work

Correspondence to: Zhibin Hu, email: zhibin_hu@njmu.edu.cn
Hongxia Ma, email: hongxiama@njmu.edu.cn

Keywords: non-small cell lung cancer, microRNA, susceptibility, survival

Received: February 18, 2016 Accepted: May 28, 2016 Published: June 25, 2016

ABSTRACT

Genetic variants in regulatory regions of some miRNAs might be associated with lung cancer risk and survival. We performed a case-control study including 1341 non-small cell lung cancer (NSCLC) cases and 1982 controls to evaluate the associations of 7 potentially functional polymorphisms in several differently expressed miRNAs with NSCLC risk. Each SNP was also tested for the association with overall survival of 1001 NSCLC patients. We identified that rs9660710 in miR-200b/200a/429 cluster and rs763354 in miR-30a were significantly associated with NSCLC risk (odds ratio (OR) = 1.17, 95% confidence interval (CI) = 1.06–1.30, P = 0.002; OR = 0.88, 95% CI = 0.80–0.98, P = 0.017; respectively). However, no significant association between variants and NSCLC death risk was observed in survival analysis. Functional annotation showed that both rs9660710 and rs763354 were located in regulatory elements in lung cancer cells. Compared to normal tissues, miR-200a-3p, miR-200a-5p, miR-200b-3p, miR-200b-5p and miR-429 were significantly increased in The Cancer Genome Atlas (TCGA) Lung Adenocarcinoma (LUAD) tumors, whereas miR-30a-3p and miR-30a-5p were significantly decreased in tumors (all P < 0.05). Furthermore, we observed that rs9660710 is an expression quantitative trait locus (eQTL) or methylation eQTL for miR-429 expression in TCGA normal tissues. Our results indicated that rs9660710 in miR-200b/200a/429 cluster and rs763354 in miR-30a might modify the susceptibility to NSCLC.

INTRODUCTION

Non-small cell lung cancer (NSCLC) accounts for over 80% of lung cancer, which remains the leading cause of cancer-related death worldwide [1, 2]. Although tobacco consumption is a major risk factor for lung cancer, accumulating evidence suggests that genetic variations may play a vital role in tumor development and progression [3–6]. Therefore, identification of these genetic biomarkers may enhance risk prediction and improve clinical outcomes of NSCLC patients.

Emerging evidence has suggested that some key microRNAs (miRNAs) are consistently dysregulated in lung cancer tissues and function as oncogenes or tumor suppressor genes in the initiation and progression of lung cancer [7–10]. Notably, Guan et al. synthesized fourteen microarray-based human lung cancer miRNA expression profiling studies including a total of 585 tumor and 519 non-cancerous control samples, and employed a vote-counting strategy to identify several consistent differentially miRNAs, such as up-regulated miR-210, miR-21, miR-183, miR-182, miR-200b, and down-regulated miR-30a, miR-145, miR-143 and miR-451a [8]. Furthermore, using a meta-analysis with more than 1100 lung cancer and noncancerous samples from 20 original studies, another study also confirmed these abnormally expressed miRNAs [9]. Meanwhile, previous studies have suggested that the high expression of miR-210, miR-21,
miR-183, miR-182, and low expression of miR-30a, miR-145, miR-143 were associated with poor prognosis of lung cancer [11–16]. These studies provided the promising candidates for further researches on miRNAs in lung cancer development and progression.

Interestingly, the biological function of these miRNAs has also been explored in lung cancer. For example, miR-21 located in the region of 17q23.1 promotes lung tumorigenesis through inhibition of negative regulators of the Ras/MEK/ERK pathway and apoptosis [17]. MiR-210 localizes in chromosome 11p15.5 and acts as an important regulator of the cellular response to hypoxia [18]. MiR-182 and miR-183 belong to the members of the miR-183/96/182 cluster, the overexpression of which has been identified as a potential risk factor for lung cancer development [19]. MiR-200b/200a/429 clustered at chromosome 1p36 and miR-30a located at 6q13 are key regulators of epithelial-mesenchymal transition (EMT) and thus affect the cell migration, invasion and metastasis in lung cancer [20–22]. MiR-451a inhibits ras-related protein 14 (RAB14) expression and functions as a tumor suppressor in human non-small cell lung cancer [10]. MiR-145 and miR-143 located in 5q33.1 inhibit NSCLC cell growth and metastasis by targeting epidermal growth factor receptor (EGFR) and nucleoside diphosphate linked moiety X-type motif 1 (NUDT1) [23, 24]. Taken together, these miRNAs have been identified to play the important role in the NSCLC tumorigenesis and progression.

In recent years, comprehensive analyses based on genome-wide association studies (GWAS) have indicated that disease/trait-associated loci are enriched in the genomic regulatory regions [25–27]. Thus, it is biologically plausible that single nucleotide polymorphisms (SNPs) in the regulatory regions of miRNA may also contribute to phenotypic variations in the human population [25]. Therefore, we evaluate the effect of potentially functional polymorphisms within the 10 kb upstream region of the above-mentioned key miRNAs cluster or single miRNA (miR-143/145, miR-183/96/182 cluster, miR-21, miR-210, miR-200b/200a/429 cluster, miR-30a, miR-451a) on the NSCLC risk and survival in a Chinese Han population.

RESULTS

Subjects characteristics

Characteristics of all subjects (1341 cases and 1982 controls) have been described in Supplementary Table S1. The age and gender were comparable between cases and controls ($P_{\text{age}} = 0.972$, $P_{\text{gender}} = 0.179$). The cases have a higher rate of smoking (61.1%) compared with the controls (48.5%), and most cases (64.1%) are diagnosed with adenocarcinoma.

Associations of 7 SNPs with NSCLC risk

Based on the bioinformatics methods (see Materials and Methods), 7 potentially functional SNPs (rs3733846 in miR-143/145, rs12538588 in miR-183/96/182, rs1292060 in miR-21, rs763354 in miR-30a, rs9660710 in miR-200b/200a/429, rs12286521 in miR-210 and rs901975 in miR-451a) were genotyped in our study. The genotype distributions of these SNPs between cases and controls are shown in Table 1. After adjusting for age, gender, and smoking, the minor allele of rs9660710 (C > A) in miR-200b/miR200a/429 demonstrated a significant association with an increased risk of NSCLC [additive model: Odds ratios (OR) = 1.17, 95% confidence intervals (CI) = 1.06–1.30, $P = 0.002$], while the variant allele of rs763354 (G > A) in miR-30a was associated with a decreased risk (additive model: adjusted OR = 0.88, 95%CI = 0.80–0.98, $P = 0.017$). No significant association was observed between other five SNPs and NSCLC risk.

Furthermore, we performed stratification analysis on the associations of rs9660710 and rs763354 with NSCLC risk by age, gender, smoking and histological subtype (Table 2). The heterogeneity test showed that the difference was significant in the subgroup of age for rs763354 ($P = 0.040$), and the protective effect of rs763354 A allele in subjects aged ≥ 60 years (adjusted OR = 0.80, 95%CI = 0.70–0.92, $P = 0.002$) was prominent as compared with subjects aged < 60 years (adjusted OR = 0.99, 95%CI = 0.85–1.15, $P = 0.847$).

Relationships between 7 SNPs and NSCLC survival

We further investigated the relationships between these SNPs and NSCLC survival. The clinical characteristics of 1001 NSCLC patients are described in Supplementary Table S2. We found that some factors, including gender, smoking, surgery, clinical stage and chemotherapy or radiotherapy had significant influences on patient prognosis (log-rank $P < 0.05$). Associations of these SNPs with overall survival in different genetic models are shown in Supplementary Table S3. Log-rank test indicated that no variant showed a significant association with overall survival time in additive, dominant or co-dominant models (all $P > 0.05$). The results were similar after adjusting for age, gender, smoking, clinical stage, chemotherapy or radiotherapy, surgery status and histological types (all $P > 0.05$, Supplementary Table S4).

In silico analysis of promising variants and miRNAs

We annotated two significantly risk-associated variants (rs763354, rs9660710) in regulatory elements cataloged in Encyclopedia of DNA Elements (ENCODE)
**Table 1: Primary information of 7 potentially functional SNPs and associations with NSCLC risk**

<table>
<thead>
<tr>
<th>miRNAs</th>
<th>SNP</th>
<th>Allele</th>
<th>Cases</th>
<th>Controls</th>
<th>MAF*</th>
<th>Additive model</th>
<th>Dominant model</th>
<th>Codominant model</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-143/145</td>
<td>rs3733846</td>
<td>A/G</td>
<td>580/597/164</td>
<td>839/925/218</td>
<td>0.34/0.34</td>
<td>1.01 (0.91–1.13)</td>
<td>0.95 (0.84–1.11)</td>
<td>0.93 (0.80–1.08)</td>
<td>1.11 (0.88–1.40)</td>
</tr>
<tr>
<td>miR-183/96/182</td>
<td>rs12538588</td>
<td>G/A</td>
<td>1229/112/0</td>
<td>1818/163/0</td>
<td>0.04/0.04</td>
<td>0.98 (0.76–1.26)</td>
<td>0.98 (0.76–1.26)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>miR-21</td>
<td>rs1292060</td>
<td>A/G</td>
<td>435/645/261</td>
<td>648/928/397</td>
<td>0.44/0.44</td>
<td>1.00 (0.90–1.10)</td>
<td>1.01 (0.87–1.18)</td>
<td>1.03 (0.87–1.20)</td>
<td>0.99 (0.81–1.21)</td>
</tr>
<tr>
<td>miR-30a</td>
<td>rs763354</td>
<td>G/A</td>
<td>486/640/210</td>
<td>650/965/352</td>
<td>0.40/0.42</td>
<td>0.88 (0.80–0.98)</td>
<td>0.85 (0.74–0.99)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>miR-200b/200a/429</td>
<td>rs9660710</td>
<td>C/A</td>
<td>431/669/240</td>
<td>746/930/306</td>
<td>0.43/0.39</td>
<td>1.17 (1.06–1.30)</td>
<td>1.26 (1.09–1.46)</td>
<td>1.23 (1.05–1.44)</td>
<td>1.34 (1.09–1.65)</td>
</tr>
<tr>
<td>miR-210</td>
<td>rs12286521</td>
<td>A/G</td>
<td>752/510/79</td>
<td>1126/738/118</td>
<td>0.25/0.25</td>
<td>1.02 (0.91–1.14)</td>
<td>1.03 (0.90–1.19)</td>
<td>1.04 (0.90–1.21)</td>
<td>0.98 (0.72–1.32)</td>
</tr>
<tr>
<td>miR-451a</td>
<td>rs901975</td>
<td>G/A</td>
<td>571/601/169</td>
<td>861/873/245</td>
<td>0.35/0.34</td>
<td>1.03 (0.93–1.14)</td>
<td>1.04 (0.90–1.20)</td>
<td>1.04 (0.90–1.21)</td>
<td>1.04 (0.83–1.30)</td>
</tr>
</tbody>
</table>

*Major/minor allele.

*Major homozygote/heterozygote/minor homozygote in cases and controls.

*Minor allele frequency among cases/controls.

*Logistic regression with adjustment for age, gender and smoking.

*het: heterozygote versus major homozygote; hom: minor homozygote versus major homozygote.

*P for additive model.

Bold values are statistically significant.

---

... project (https://www.encodeproject.org/), and HaploReg (http://compbio.mit.edu/HaploReg), respectively. Visual inspection of the SNPs and histone modification peaks showed that rs763354 is situated within the enhancer element (H3K4me1 and H3K27ac histone mark) in both normal lung cell line (Normal Human Lung Fibroblasts, NHLF) and lung cancer cell line (A549), and fall into the promoter element (H3K4me3) in A549 cell line (Figure 1A). Rs9660710 is situated in the enhancer element (H3K27ac) in A549 cell line (Figure 1B). Furthermore, based on the HaploReg data, we found that rs763354, rs9660710 and their correlated variants within a LD block are positioned in regulatory regions (histone promoter or enhancer marks in several tissues), and alter multiple regulatory motifs or disturb some proteins bound (Supplementary Tables S5–S6). Taken together, our results indicated that these SNPs were probably involved in the regulation of miRNA expression.

To further explore the influence of rs763354 and rs9660710 on their related miRNAs (miR-30a-3p, miR-30-5p, miR-200a-3p, miR-200a-5p, miR-200b-3p, miR-200b-5p and miR-429) in NSCLC development, we used publicly available TCGA Lung Adenocarcinoma (LUAD) datasets and firstly evaluated the association of rs763354 and miR-30a expression. To support the hypothesis, we extracted the TCGA LUAD DNA methylation data, which covers 51 CpG sites in the 10kb 5ʹ upstream region of miR-200b/200a/429 cluster. We observed that rs763354 on their related miRNAs (miR-30a-3p, miR-30-5p, miR-200a-3p, miR-200a-5p, miR-200b-3p, miR-200b-5p and miR-429) were significantly increased in tumor tissues compared with paired normal tissues while others (miR-210, miR-200a, miR-200b, miR-200b-5p and miR-429) were significantly increased in tumor tissues (Figure 2). Taken together, the minor allele of rs9660710 may increase the miR-429 expression and thus contribute to lung tumorigenesis.

Interestingly, rs9660710 (C > A) located in the CpG island in the promoter of miR-200b/200a/429 can lead to the loss of the CpG site (Figure 1B), which may prevent the CpG site from being methylated and increasing the miR-429 expression. To support the hypothesis, we extracted the TCGA LUAD DNA methylation data, which covers 51 CpG sites in the 10kb 5ʹ upstream region of miR-200b/200a/429 cluster. We observed that rs9660710 was negatively associated with the methylation level of cg00152708 in 32 normal tissues (rho = –0.427, P = 0.017), indicating that rs9660710 may be an expression quantitative trait locus (eQTL) for miR-429. However, due to the limited normal tissues (6 samples) with both DNA methylation...
and miRNA expression information, we cannot directly analyze the relationship between the methylation level of cg00152708 and miR-429 expression. Therefore, we speculated that rs9660710 (C > A) may decrease cg00152708 methylation level and thus increase the miR-429 expression in lung carcinogenesis, indicating a role of rs9660710 as a methylation quantitative trait locus (meQTL) for miR-429 expression.

**DISCUSSION**

In the present study, we explored the associations of 7 SNPs in regulatory regions of candidate miRNAs with the NSCLC risk and survival. Our findings suggested that rs9660710 (C > A) may decrease cg00152708 methylation level and thus increase the miR-429 expression in lung carcinogenesis, indicating a role of rs9660710 as a methylation quantitative trait locus (meQTL) for miR-429 expression.

Table 2: Stratification analysis of rs9660710 and rs763354 genotypes associated with NSCLC risk

<table>
<thead>
<tr>
<th>Variables</th>
<th>rs9660710 (CC/CA/AA)</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>rs763354 (GG/GA/AA)</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>Controls</td>
<td></td>
<td></td>
<td>Cases</td>
<td>Controls</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 60</td>
<td>187/293/109</td>
<td>334/397/142</td>
<td>1.18</td>
<td>(1.01–1.37)</td>
<td>210/276/100</td>
<td>309/416/147</td>
</tr>
<tr>
<td>≥ 60</td>
<td>244/376/131</td>
<td>412/533/164</td>
<td>1.18</td>
<td>(1.02–1.35)</td>
<td>276/364/110</td>
<td>341/549/205</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>310/477/161</td>
<td>504/649/205</td>
<td>1.13</td>
<td>(1.00–1.27)</td>
<td>357/439/149</td>
<td>446/677/233</td>
</tr>
<tr>
<td>Female</td>
<td>121/192/79</td>
<td>242/281/101</td>
<td>1.27</td>
<td>(1.06–1.53)</td>
<td>129/201/61</td>
<td>204/288/119</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>175/254/93</td>
<td>389/469/162</td>
<td>1.14</td>
<td>(0.98–1.33)</td>
<td>186/251/85</td>
<td>342/492/175</td>
</tr>
<tr>
<td>Ever</td>
<td>256/415/147</td>
<td>357/461/144</td>
<td>1.20</td>
<td>(1.05–1.38)</td>
<td>300/389/125</td>
<td>308/473/177</td>
</tr>
<tr>
<td>Histological types</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>151/231/99</td>
<td>746/930/306</td>
<td>1.28</td>
<td>(1.10–1.49)</td>
<td>172/234/73</td>
<td>650/965/352</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>280/438/141</td>
<td>746/930/306</td>
<td>1.13</td>
<td>(1.00–1.27)</td>
<td>314/406/137</td>
<td>650/965/352</td>
</tr>
</tbody>
</table>

a Major homozygote/heterozygote/minor homozygote.

b Adjusted for age, gender and smoking where appropriate in additive model.

P values were from heterogeneity test based on χ²-based Q test. Bold values are statistically significant.
of AP-2α suppressed the binding of miR-200b/200c/429 to AP-2α, and might be a potential prognostic marker for cisplatin treatment [32]. Another study showed that miR-143/145 rs3733845 confers protective effect on the risk of colorectal cancer in Chinese population [33]. In the present study, we failed to analyze the associations between these two SNPs (rs61768479 and rs1045385) and NSCLC risk and survival, owing to the low frequency

Figure 1: Chromatin features of risk-associated SNPs. Functional annotation in proximity to miR-30a rs763354 location (A) and miR-200b/200a/429 cluster rs9660710 location (B) in A549 and NHLF cells. ChIP-seq tracks for promoter histone marks (H3K4me3) and enhancer histone marks (H3K4me1, H3K27ac) in A549 and NHLF cells are present along with DNAse hypersensitivity tracks from ENCODE data. The black vertical lines guide the position of risk-associated SNPs. The blue stripe represents the normal lung fibroblasts (NHLF). The red stripe represents the lung cancer cell line (A549). The green stripe represents the CpG island in the region.
in Chinese Han population and the reason that rs1045385 is located in the target gene. But for rs3733845, we did not observe the significant association of this SNP with lung cancer risk in our study, which may be due to the complex effects of polymorphisms on different cancer types. Our study showed for the first time that rs9660710 in miR-200b/200a/429 cluster conferred the increased risk of NSCLC in Chinese population, but miR-30a rs763354 showed a protective effect on NSCLC risk.

In silico analysis suggested that rs763354 and rs9660710 were situated in different histone modification patterns between normal and lung cancer cells, highlighting the possibility that these loci may regulate the transcription of miRNAs. Functional annotations from HaploReg indicated that rs763354 might increase the binding of Sox2, which might bind to the promoter regions of miRNAs [34]. However, we did not observe significant association between rs763354 and miR-30a-3p or miR-30a-5p expression in 38 TCGA adjacent normal tissues. The variant allele of rs9660710 (C > A) is predicted to strengthen the affinity motifs for BRCA1 (Supplementary Table S5), which participates in transcriptional regulation by interacting with a large amount of different transcription factors, including TP53 and CMYC [35]. In our study, we observed that the variant allele of rs9660710 is an eQTL for miR-429 expression and significantly increase the miR-429 expression in normal tissues. Moreover, rs9660710 is a CpG site related SNP (cgSNP) [36] and meQTL for miR-429 expression. Consequently, we speculated that rs9660710 may disrupt transcription factor response elements or DNA methylation to promote the expression of miR-429, respectively. However, further functional characterizations are warranted to validate the genotype-phenotype correlation.

In conclusion, our study provided evidence that SNPs in miRNAs regulatory regions were significantly associated with NSCLC risk, indicating the importance of the functional variants of miRNAs in NSCLC tumorigenesis. Larger studies with functional investigations are needed to validate the findings.

MATERIALS AND METHODS

Study subjects

All subjects in this study have been described previously [37]. In brief, 1341 patients with NSCLC were newly diagnosed and recruited from the First Affiliated Hospital of Nanjing Medical University and the Cancer Hospital of Jiangsu Province since 2003. All patients were histopathologically or cytologically confirmed NSCLC. The 1982 controls were selected from healthy individuals participated in screening for non-infectious diseases during the same time period. The controls were frequency-matched to cases on age and gender. After having signed the informed consent, each participant was
face-to-face interviewed by trained interviewers to collect personal information on demographics data. Furthermore, we collected patients’ clinical information from patients’ medical records and subsequent personal or family contacts, including the date of diagnosis, surgery status, clinical stage, chemotherapy or radiotherapy, histological types and the date of death during follow-up (last follow-up in August 2013). Finally, a total of 1001 cases (74.6%) had sufficient information about survival time to permit statistical analysis with the median survival time (MST) of 26.0 months. This study was approved by the institutional review board of Nanjing Medical University.

SNPs selection and genotyping

Based on the UCSC Genome Browser database (http://genome.ucsc.edu/) and HapMap SNP database (phase II + III Feb 09, on NCBI B36 assembly, dbSNP b126), 38 common SNPs (MAF ≥ 0.05 in Chinese Han population) were selected in the 10kb 5’ upstream region of each pre-miRNA. If the candidate miRNA is located in a cluster, we then screened the upstream region of the cluster instead of the single miRNA (miR-143/145, miR-183/96/183, miR-21, miR-210, miR-200b/200a/429, miR-30a, miR-451a). The potentially functional SNPs were predicted in silico by using the SNPinfo web server (http://snpinfo.niehs.nih.gov/). Among 16 potentially functional SNPs that might disturb the binding of transcription factors, 9 SNPs were excluded because of high linkage disequilibrium (LD) analyzed by (r² > 0.8). As a result, 7 SNPs (rs3733846 in miR-143/145, rs12538588 in miR-183/96/183, rs1292060 in miR-21, rs763354 in miR-30a, rs9660710 in miR-200b/200a/429, rs9660710 in miR-210 and rs901975 in miR-451a) were genotyped in our study.

Genomic DNA was isolated from a leukocyte pellet by proteinase K digestion, followed by phenol-chloroform extraction and ethanol precipitation. All SNPs were genotyped by Illumina Infinium Human Exome BeadChip (Illumina Inc., San Diego, CA). All SNPs were successfully genotyped with call rates > 95%.

In silico analysis

In order to evaluate whether promising SNPs harbor regulatory elements, we downloaded Chromatin Immunoprecipitation sequencing (ChiP-seq) and DNaseI Hypersensitivity (DNaseI HS) data in normal lung cell line (Normal Human Lung Fibroblasts, NHLF) and lung cancer cell line (A549) from Encyclopedia of DNA Elements (ENCODE) project (https://www.encodeproject.org/), and visualized these regulatory elements using Integrative Genomics Viewer (IGV) (http://www.broadinstitute.org/igv/). We also used the HaploReg resource V2 (http://compbio.mit.edu/HaploReg/) to examine the risk loci and variants in high LD (r² > 0.8 in Asian from the 1000 Genomes Project) for functional elements available from the Epigenome roadmap project [38].

To assess miRNA expression, level 3 miRNA isoform quantification data (reads per million of total reads mapping to a mature microRNA) in LUAD tumors and adjacent normal tissues were retrieved from TCGA portal on date 07/15/2014 (http://cancergenome.nih.gov/). We filtered out cross-mapped regions and then summed over the reads per million miRNAs mapped (RPM) values for each mature miRNA according to the previous study [37]. After normalization by using the EdgeR package [39], the expression data of interest miRNAs were extracted and were log2-transformed for further analysis.

In order to bridge the gap between the genotype and phenotype, we have applied for the Affymetrix SNP 6.0 Level 1 data of TCGA LUAD samples to extract genotypes of rs9660710 and rs763354. Because the chip didn’t include the genotype information of rs9660710, we further imputed the genotypes of rs9660710 for TCGA samples by using IMPUTE2 (1000 Genomes Phase I integrated variant set b37 March 2012 release). The level 3 TCGA methylation data (Illumina’s Infinium HumanMethylation450 Beadchip) for 32 LUAD normal tissues and paired 29 tumor tissues were downloaded from the TCGA data portal on date 07/15/2014.

Statistical analysis

Differences between cases and controls in the distribution of demographic characteristics were evaluated by the χ² test (for categorical variables) or Student’s t-test (for continuous variables). Genotype frequencies in control group were tested for the Hardy-Weinberg equilibrium by using goodness-of-fit χ² test. OR and their 95% CI were calculated by using logistic regression analysis to evaluate the associations between genotypes and lung cancer risk with adjustment for age, gender and smoking. For survival analysis, the distribution of demographic and clinical variables according to death status was tested by the χ² test. Log-rank test was applied to compare the survival time in different subgroups categorized by patient characteristics and genotypes. Univariate and multivariate Cox proportional hazard regression analysis were performed to estimate the crude or adjusted hazard ratio (HR) and 95% CI with adjustment of age, gender, smoking, clinical stage, chemotherapy or radiotherapy, surgery status and histological types. The Chi-square-based Q test was applied to test the heterogeneity of associations between subgroups. The paired two-sample Student’s t-test was used to compare the miRNA expression or methylation levels of CpG sites in paired tumor and adjacent normal tissues from TCGA dataset. Spearman’s rank correlation was used to evaluate genotype-expression and genotype-methylation associations. All of the statistical analyses were performed with R software (version 3.2.2; The R Foundation for Statistical Computing). A two-sided P value of < 0.05 was considered statistically significant.
ACKNOWLEDGMENTS

We thank the ENCODE and TCGA Projects for making the data public.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

GRANT SUPPORT

This work was supported by the National Science Foundation for Distinguished Young Scholars of China (81225020, 81270044, 81521004, 81230067), Jiangsu Specially-Appointed Professor project, National Program for Support of Top-notch Young Professionals from the Organization Department of the CPC Central Committee, Jiangsu Province Clinical Science and Technology Projects (BL2012008) and the Priority Academic Program for the Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine).

REFERENCES

20. Gill BJ, Gibbons DL, Roudsari LC, Saik JE, Rizvi ZH, Roybal JD, Kurie JM, West JL. A synthetic matrix with independently tunable biochemistry and mechanical...


