Gene 646 (2018) 56–63

Contents lists available at ScienceDirect

Gene

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

Research paper

Regulatory variants in cancer-related pathway genes predict survival of patients with surgically resected non-small cell lung cancer

Kyung Min Shin[a,1], Mi Jeong Hong[b,e,1], Shin Yup Lee[c,d,1], Cheng Cheng Jin[b], Sun Ah Baek[b], Jang Hyuck Lee[b], Jin Eun Choi[b,e], Hyo-Gyoung Kang[b,e], Won Kee Lee[f], Yangki Seok[d,g], Eung Bae Lee[d,g], Ji Yun Jeong[h], Seung Soo Yoo[c,d], Jaehee Lee[c], Seung Ick Cha[c], Chang Ho Kim[c], Young Chul Kim[i], In Jae Oh[i], Kook Joo Na[j], Sukki Cho[k], Sanghoon Jheon[k], Jae Yong Park[b,c,d,e,*]

[a] Department of Radiology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea
[b] Department of Biochemistry and Cell Biology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea
[c] Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Republic of Korea
[d] Lung Cancer Center, Kyungpook National University Medical Center, Daegu, Republic of Korea
[e] Cell and Matrix Research Institute, School of Medicine, Kyungpook National University, Daegu, Republic of Korea
[f] Biostatistics, Medical Research Collaboration Center in Kyungpook National University Hospital and Kyungpook National University School of Medicine, Daegu, Korea
[g] Department of Thoracic Surgery, School of Medicine, Kyungpook National University, Daegu, Republic of Korea
[h] Department of Pathology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea
[i] Department of Internal Medicine, Chonnam National University Hwasun Hospital, Hwasun, Republic of Korea
[j] Department of Thoracic and Cardiovascular Surgery, Chonnam National University Hwasun Hospital, Hwasun, Republic of Korea
[k] Department of Thoracic and Cardiovascular Surgery, Seoul National University School of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Lung cancer
Polymorphism
RegulomeDB
Survival

ABSTRACT

Background: We conducted this study to identify genetic variants in cancer-related pathway genes which can predict prognosis of NSCLC patients after surgery, using a comprehensive list of regulatory single nucleotide polymorphisms (SNPs) prioritized by RegulomeDB.

Method: A total of 509 potentially functional SNPs in cancer-related pathway genes selected from RegulomeDB were evaluated. These SNPs were analyzed in a discovery set (n = 354), and a replication study was performed in an independent set (n = 772). The association of the SNPs with overall survival (OS) and disease-free survival (DFS) were analyzed.

Results: In the discovery set, 76 SNPs were significantly associated with OS or DFS. Among the 76 SNPs, the association was consistently observed for 5 SNPs (*ERCC1* rs2298881C > A, *BRCA2* rs3092989G > A, *NELFE* rs440454C > T, *PPP2R4* rs2541164G > A, and *LTBP4* rs3786527G > A) in the validation set. In combined analysis, *ERCC1* rs2298881C > A, *BRCA2* rs3092989, *NELFE* rs440454C > T, and *PPP2R4* rs2541164G > A were significantly associated with OS and DFS (adjusted HR ·aHR· for OS = 1.46, 0.62, 078, and 0.76, respectively; P = 0.003, 0.002, 0.007, and 0.003 respectively; and aHR for DFS = 1.27, 0.69, 0.86, and 0.82, respectively; P = 0.02, 0.002, 0.03, and 0.008, respectively). The *LTBP4* rs3786527G > A was significantly associated with better OS (aHR = 0.75; P = 0.003).

Conclusion: Our results suggest that five SNPs in the cancer-related pathway genes may be useful for the prediction of the prognosis in patients with surgically resected NSCLC.

Abbreviations: aHR, adjusted HR; *BRCA2*, breast cancer 2, early onset; CIs, confidence intervals; DFS, disease-free survival; ENCODE, encyclopedia of DNA Elements; *ERCC1*, excision repair cross-complementing group 1; *ERG*, V-ets avian erythroblastosis virus E26 oncogene related; ETS, E26 transformation-specific; eQTLs, expression quantitative trait loci; GWAS, genome-wide association study; HRs, hazard ratios; HWE, Hardy-Weinberg equilibrium; LD, linkage disequilibrium; *LTBP4*, latent transforming growth factor beta binding protein 4; MAF, minor allele frequency; *NELFE*, negative elongation factor complex member E; NSCLC, non-small cell lung cancer; OS, overall survival; *PPP2R4*, protein phosphatase 2 regulatory subunit 4; SNPs, single nucleotide polymorphisms; 3C, chromosome conformation capture

* Corresponding author at: Lung Cancer Center, Kyungpook National University Medical Center, 807, Hoguk-ro, Buk-gu, Daegu 702-210, Republic of Korea.
E-mail address: jaeyong@knu.ac.kr (J.Y. Park).

[1] These authors contributed equally to this paper.

https://doi.org/10.1016/j.gene.2017.12.055
Received 22 September 2017; Received in revised form 20 December 2017; Accepted 27 December 2017
Available online 29 December 2017

1. Introduction

Lung cancer is the leading causes of cancer-related mortality worldwide (Siegel et al., 2016). Recent advances in multimodality therapy have improved the prognosis of lung cancer, but the five-year survival rate of NSCLC still remains < 20% (Siegel et al., 2016). Although surgery is the best treatment modality for potentially curing early stage NSCLC, a significant proportion of patients die of recurred lung cancer. Pathologic stage is the most powerful determinant of prognosis in early stage NSCLC after surgery. However, given the remarkable variability in survival among NSCLC patients even at the same pathologic stage (Detterbeck et al., 2009), host genetic factors may have an impact on the prognosis. Therefore, identification of genetic biomarker is very important for more precise prognostication and identification of patients who are at higher risk of poor prognosis.

Genome-wide association study (GWAS) have permitted comprehensive evaluation of the associations between genome-wide genetic variants and the prognosis of lung cancers, and identified several genetic loci associated with the clinical outcomes of advanced NSCLC receiving chemotherapy (Lee et al., 2013; Hu et al., 2012; Sato et al., 2011) and early NSCLC after curative resection (Tang et al., 2015; Yoon et al., 2014). However, understanding the biological mechanism of GWAS results is incomplete since most of the variations identified in GWAS are not causal (Witte, 2010). Indeed, approximately 90% of single nucleotide polymorphisms (SNPs) associated with human diseases in GWAS have been found outside of protein-coding regions (Schaub et al., 2012; Boyle et al., 2012). Recently, the Encyclopedia of DNA Elements (ENCODE) project revealed that the majority (80%) of the genome, especially outside of protein-coding regions, contain elements linked to biochemical functions, and that genetic variants in non-coding DNA play an important role in the regulation of gene expression (Consortium, 2012). RegulomeDB is a database with a large collection of regulatory information from ENCODE and other sources, which may provide various testable hypotheses and putative mechanistic explanations for genetic association studies (Boyle et al., 2012). It contains a scoring system for prioritizing SNPs based on the degree of evidence of the regulatory function of a given SNP, which helps prioritize and select SNPs for genetic association studies to improve power to detect true causal variants.

We previously reported the association between potential regulatory SNPs in *ERCC1* from RegulomeDB and the survival of surgically resected NSCLC, suggesting that RegulomeDB is a practical tool for selecting potentially functional SNPs in the regulatory region (Lee et al., 2015). In this study, we investigated the associations of genetic polymorphisms in cancer-related pathway genes with survival of surgically resected NSCLC patients using a comprehensive collection of SNPs selected using RegulomeDB.

2. Materials and methods

2.1. Patients

The discovery set included 354 patients with pathologic stages I, II or IIIA NSCLC who underwent curative surgical resection at the Kyungpook National University Hospital (KNUH) from December 1997 to January 2010. Genomic DNA samples were provided by the National Biobank of Korea - KNUH, which is supported by the Ministry of Health, Welfare and Family Affairs. The validation set comprised 772 patients with surgically resected NSCLC at Seoul National University Bundang Hospital (SNUBH, n = 428) and Chonnam National University Hwasun Hospital (CNUHH, n = 344). In the present study, 292 among 354 patients in the discovery set and 173 among 772 patients in the validation set were also enrolled in our previous study (Lee et al., 2015). All the patients were ethnic Koreans, and none of them received chemotherapy or radiotherapy prior to surgery to avoid the effects on DNA. The pathologic stages were determined according to the seventh edition of Union for International Cancer Control and American Joint Committee on Cancer TNM (Goldstraw and International Association for the Study of Lung Cancer, 2009). Written informed consent was routinely obtained from all patients before surgery, and this study was approved by the institutional review board of the KNUH, SNUBH, and CNUHH.

2.2. Polymorphism selection and genotyping

RegulomeDB provides a scoring system, with categories ranging from 1 to 6 based on the degree of experimental or computational evidence of functional consequence of a given variant. Category 1 includes variants that are known expression quantitative trait loci (eQTLs) which have been shown to be associated with expression of target genes, and is most likely significant among the six categories (Boyle et al., 2012). We prioritized functional SNPs using RegulomeDB (http://regulome.stanford.edu) and selected 39,433 SNPs that were classified into category 1. Among those, we selected 1591 SNPs in 661 genes involved in cancer-related pathway using data sets from OncoSearch (http://oncosearch.biopathway.org) and Gene Ranker (http://cbio.mskcc.org/tcga-generanker). Based on data from the NCBI SNP database (http://ncbi.nlm.nih.gov/SNP) 1063 SNPs with the minor allele frequency (MAF) ≥ 0.1 in HapMap-JPT were collected, and finally 509 SNPs were selected for genotyping after excluding those in linkage disequilibrium (LD, $r^2 \geq 0.8$). The selected SNPs in the discovery set were genotyped using SEQUENOM's MassARRAY® iPLEX assay (SEQUENOM Inc., San Diego, CA, USA) and listed in Supplementary Table 1. For validation, we genotyped 76 SNPs that were significantly associated with overall survival (OS) or disease-free survival (DFS) in the discovery set ($P < 0.05$, Table 2). All genotyping analyses were conducted "blind" with regard to the case/control status to ensure quality control. Approximately 10% of the samples of the cohort were randomly selected and genotyped a second time by a different investigator and the results of the second round of genotyping were 100% concordant.

2.3. Data analysis

Differences in the distribution of genotypes according to clinicopathologic factors of the patients were compared using the *t*-test and chi-square test for continuous and categorical variables, respectively. The Hardy-Weinberg equilibrium (HWE) was tested by a goodness-of-fit chi-square test with 1 degree of freedom. Overall survival (OS) was measured from the date of surgery to the date of death or the last follow-up. Disease-free survival (DFS) was calculated from the date of surgery until first evidence of disease recurrence or last date of follow-up for patients who were free of disease. Kaplan-Meier method and log-rank tests were used to analyze the differences in OS and DFS across different genotypes. Hazard ratios (HRs) and 95% confidence intervals (CIs) were estimated using multivariable Cox proportional hazards models, with adjustment for age, gender, smoking status, tumor histology, pathologic stage, and adjuvant therapy. Statistical analyses were performed using statistics software (SAS, version 9.2, SAS institute, Cary, NC, USA).

3. Results

The baseline clinical and pathological characteristics of the patients in the discovery and validation sets and their association with OS and DFS are given in Table 1. Never smokers and adenocarcinomas were significantly more frequent in the validation set compared with the discovery set ($P = 0.001$ and < 0.001, respectively), which is probably because of the lower smoking rate in the validation set in which most patients were recruited from capital city area compared to discovery set in which many live in rural area. Heavier smokers (≥ 42 pack-years) among ever smokers and patients who received adjuvant therapy were significantly more frequent in the validation set ($P = 0.01$ and <