Lung Cancer 114 (2017) 79–89

Contents lists available at ScienceDirect

Lung Cancer

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

Research paper

Genetic predisposition to lung adenocarcinoma among never-smoking Chinese with different epidermal growth factor receptor mutation status

Li Han[a], Cheuk-Kwong Lee[b], Herbert Pang[c], Hong-Tou Chan[d], Iek-Long Lo[d], Sze-Kwan Lam[a], Tak-Hong Cheong[d], James Chung-Man Ho[a,*]

[a] Division of Respiratory Medicine, Department of Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong Special Administrative Region
[b] Hong Kong Red Cross Blood Transfusion Service, Hong Kong Special Administrative Region
[c] School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region
[d] Pneumology Department, Centro Hospitalar C.S. Januario Macau, Macau Special Administrative Region

ARTICLE INFO

Keywords:
Lung adenocarcinoma
Epidermal growth factor receptor mutation
Single nucleotide polymorphisms
Gene-environment interaction
Never-smokers

ABSTRACT

Objectives: The inconsistent findings from genetic association studies may be related to the heterogeneity in different molecular subtypes of lung cancer. This study evaluated the predisposing single-nucleotide polymorphisms (SNPs) in epidermal growth factor receptor (EGFR) mutant and EGFR wild-type lung adenocarcinoma separately among never-smokers.

Materials and methods: This was a two-stage case-control study. Never-smokers with pathologically confirmed lung adenocarcinoma and healthy controls were recruited in Hong Kong and Macau. Genomic DNA was extracted and genotyped by MassARRAY. In the discovery stage, 51 SNPs were investigated at the SNP, gene and pathway level among 103 EGFR mutant and 78 EGFR wild-type lung adenocarcinoma cases compared with matched controls. In the validation stage, SNPs that were identified with significant lung cancer risk were replicated in a separate cohort of 84 lung adenocarcinoma cases and compared with 103 Chinese Han, Beijing and 105 Chinese Han, Southern public controls from the 1000 genome database.

Results and conclusion: The genetic association of IL-6 rs2069840 with EGFR mutant lung adenocarcinoma was ascertained. In the discovery stage, haplotype GGG in three SNPs (rs2069840, rs2069852, rs2066992) of IL-6, synergetic effects of IL-6 rs2069840 and environmental tobacco smoke in the workplace were found to be related to EGFR mutant lung adenocarcinoma. ERCC2 rs238406 showed a marginally significant association with EGFR mutant lung adenocarcinoma in the validation stage ($P = 0.096$). ERCC2 rs50871 and ATM rs611646 showed significant association with EGFR wild-type lung adenocarcinoma in the discovery stage. In conclusion, IL-6 rs2069840 conferred susceptibility to EGFR mutant lung adenocarcinoma in a Hong Kong and Macau never-smoking Chinese population.

1. Introduction

Lung adenocarcinoma (ADC) has become the predominant cell type in lung cancer cases throughout the world [1]. Although tobacco smoking is a dominant environmental risk factor for development of lung ADC, lung cancer in never-smokers (LCINS) has emerged as a distinct disease entity in terms of biological behaviour, molecular profile, and therapeutic options [2,3]. The predilection of epidermal growth factor receptor (EGFR) mutations and anaplastic lymphoma kinase (ALK) re-arrangement among LCINS has led to the rapid development of targeted therapies for lung ADC. Nonetheless EGFR mutations are found in almost 80% of never-smoking lung ADC in the Chinese population [4]. Therefore, even among never-smokers, lung ADC is now considered a heterogeneous disease with different molecular characteristics. There has been great interest in the identification of potential genes that are associated with a predisposition to development of LCINS. This may allow early recognition of never-smokers who have an increased susceptibility to NSCLC. Candidate gene studies and

Abbreviations: SNPs, single-nucleotide polymorphisms; EGFR, epidermal growth factor receptor; NSCLC, non-small cell lung cancer; ADC, lung adenocarcinoma; LCINS, lung cancer in never-smokers; ALK, anaplastic lymphoma kinase; GWAS, genome-wide association studies; HKU/HA HKW IRB, Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster; CHB, Chinese Han, Beijing; CHS, Chinese Han, Southern; MAF, minor allele frequency; SD, standard deviation; LD, linkage disequilibrium; MDR, multifactor dimensionality reduction; AIC, Akaike information criterion; TA, testing accuracy; CVC, cross-validation consistency; FDR, false discovery rate; HWE, Hardy-Weinberg equilibrium; TKI, tyrosine-kinase inhibitor; NER, nucleotide excision repair; FS, functional significance

* Corresponding author at: Department of Medicine, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong Special Administrative Region.
E-mail address: jhocm@hku.hk (J.C.-M. Ho).

http://dx.doi.org/10.1016/j.lungcan.2017.10.012
Received 10 July 2017; Received in revised form 23 October 2017; Accepted 28 October 2017

genome-wide association studies (GWAS) have identified a link between single nucleotide polymorphisms (SNPs) of genes involved in DNA repair, inflammation, carcinogen metabolism and tumour suppression and the development of lung cancer among never-smokers. The inconsistent results from previous reports may have been due to different genotyping methods [5,6], population demographics, family history of cancers and environmental exposure [6,7]. Nonetheless the differences among lung ADC with distinct driver oncogenes are likely crucial. This study was designed with the primary objective to investigate the predisposing SNPs in EGFR mutant and wild-type lung ADC among never-smokers. The secondary objectives included identification of environmental risk factors as well as possible gene-environment interactions that contribute to the development of EGFR mutant and wild-type lung ADC.

2. Materials and methods

2.1. Subject recruitment

This was a two-stage (discovery and validation stage) case-control study. The study was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (HKU/HA HKW IRB) (UW13-343) and the hospital ethics committee of the Centro Hospitalar C.S. Januario, Macau. Eligible ethnically Chinese participants with confirmed primary lung cancer were prospectively recruited at Queen Mary Hospital in Hong Kong starting from September 2006 to a HKU/HA HKW IRB-approved project to compile a prospective lung cancer database (UW06-151 T/1176). Cases and controls with Chinese ethnicity, age over 18 years, male or female, who had never smoked or had smoked fewer than 100 cigarettes during their lifetime were included. Cases with confirmed primary lung adenocarcinoma, with or without EGFR mutations were included. Those with a history of cancer other than lung in origin, and controls with a history of any cancer or respiratory disease were excluded. By February 2015, a total of 653 lung cancer patients had been recruited from Hong Kong and Macau, regardless of histological type, tumour molecular subtype or smoking status, of whom 299 never-smoking patients with lung adenocarcinoma served as cases: 146 (61%) EGFR mutants, 93 (39%) EGFR wild-type, 60 unknown EGFR status (54 were diagnosed before 2011 when the EGFR test was unavailable, 6 with insufficient samples for testing). The EGFR mutation test was performed by standard methodology, either by direct sequencing or allele-specific polymerase chain reaction, depending on the quality of tumour sample. 453 blood donors were recruited from the Hong Kong Red Cross from May 2013 to February 2015 of whom 332 never-smoking healthy donors were chosen as controls. Controls were randomly selected and individually matched with cases by gender and age ± 5 years in a 1:1 ratio. After matching, 103 EGFR mutant and 78 EGFR wild-type age- and gender-matched pairs were included in the discovery stage. In the validation stage, a separate cohort of 84 never-smoking Chinese patients with EGFR mutant lung adenocarcinoma recruited from Hong Kong and Centro Hospitalar C.S. Januario, Macau from March 2015 to December 2016 were chosen as independent cases, while 103 Chinese Han, Beijing (CHB) and 105 Chinese Han, Southern (CHS) public controls in 1000 genome database were selected as controls [9].

2.2. SNP selection, sample preparation and genotyping

Fifty-one SNPs in 14 genes belonging to four different pathways (DNA repair, carcinogen metabolism, inflammation, tumour suppression) were genotyped (Supplementary Table 1). The SNPs were selected based on the following criteria: (1) SNPs with known or putative functions from previously reported candidate genetic association studies or GWAS; (2) Tagger SNPs with pairwise linkage disequilibrium (LD) set as a squared correlation coefficient (r^2) more than 0.8 ($r^2 \geq 0.8$); (3) Minor allele frequency (MAF) ≥ 5% in CHS descendants or CHB descendant-based data from 1000 genome project [9]. A venous blood sample (10mLs) was taken from all subjects, with buffy coat separated by centrifugation and stored at −20 °C. Genomic DNA was extracted from the stored buffy coat using a DNA Blood Kit (Qiagen, Hilden, Germany) and genotyped using Sequenom's MassARRAY system (Sequenom, San Diego, California, USA).

2.3. Questionnaires

A structured face-to-face interview was conducted with each of the study subjects (cases and controls) by trained research assistants using a standard questionnaire (demographics, environmental exposures and family history of lung cancer/other cancer; and additional clinical characteristics for lung cancer for cases).

2.4. Statistical analysis

In the discovery stage, case-control comparisons of demographics and environment exposures were made separately between 103 EGFR mutant cases and matched controls as well as 78 EGFR wild-type cases and matched controls. Case-case comparisons were made between EGFR mutant and wild-type cases regarding clinical characteristics, presented as mean ± standard deviation (SD) or N (%) and compared by paired *t*-test or chi-square where appropriate. A *p*-value ≤0.05 was defined as statistically significant. Case-control comparisons of genotype frequencies were made between 103 EGFR mutant matched pairs and 78 EGFR wild-type matched pairs separately to identify the genetic association in relation to EGFR mutant or wild-type lung cancer risk. It was explored in the following manner: (a) at the SNP level for individual SNP association using SNPstats [10]; (b) at the gene level for linkage disequilibrium (LD) by haplotype analysis using Haploview for those SNPs that showed an individual significant association [11]; and (c) at the pathway level for gene–gene and gene-environment interaction using a multifactor dimensionality reduction (MDR) model among significant/marginally significant SNPs as shown in the individual genetic association analysis [12]. The best genetic model was selected based on the lowest Akaike information criterion (AIC) value from the three models (additive, dominant, recessive) in SNPstats. Testing accuracy (TA) and cross-validation consistency (CVC) were used to choose the best model of MDR with permutation testing to determine statistical significance [12]. The Benjamini and Hochberg method was employed to control false discovery rate (FDR) [13]. SNPs with genotypes that departed significantly from Hardy-Weinberg equilibrium (HWE) or those with a call rate ≤ 90% were excluded from the association analysis.

In the validation stage, control–control comparisons of allelic frequencies were made between 103 Red Cross controls (matched controls for 103 EGFR mutant cases) and 103 CHB and 105 CHS public controls in 1000 genome. They are presented as N (%) and the analysis was performed by Chi-square. Case-case comparisons of demographics and environmental exposures were made between EGFR mutant cases recruited in the discovery stage (n = 146) and EGFR mutant cases recruited in the validation stage (n = 84). They are presented as mean ± standard deviation (SD) or N (%) and the comparison was made using paired *t*-test or chi-square where appropriate. A *p*-value ≤0.05 was defined as statistically significant. Case-control comparisons were made between 84 EGFR mutant cases and 103 CHB public controls or 105 CHS public controls in 1000 genome separately to validate the genetic associations.

Sample size was calculated based on discordant pairs of matched cases and controls as shown in our pilot study by McNemar's Z-test. The proportion of discordant pairs of IL-6 rs 2069840 was 7.8% in matched cases and 18.4% in matched controls. Therefore, 89 EGFR mutant matched pairs were needed to detect an association of IL-6 rs2069840 (OR = 3.62) with 80% power, 2-sided at 0.05 significant level. The proportion of discordant pairs of ATM rs611646 was 13.3% in matched