Mitochondrion 11 (2011) 27–32

Contents lists available at ScienceDirect

Mitochondrion

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

Association of mitochondrial haplogroup D and risk of esophageal cancer in Taihang Mountain and Chaoshan areas in China

Xiao-Yun Li [a,1], Yu-Bai Guo [a,c,1], Min Su [a,*,1], Lu Cheng [b], Zu-Hong Lu [b], Dong-Ping Tian [a]

[a] Department of Pathology, The Key Immunopathology Laboratory of Guangdong Province, Shantou University Medical College, Shantou, Guangdong Province 515031, China
[b] State Key Laboratory of Bioelectronics, Southeast University, Nanjing, Jiangsu Province 210096, China
[c] Department of Pathology, The First Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong Province 515031, China

ARTICLE INFO

Article history:
Received 27 January 2010
Received in revised form 23 June 2010
Accepted 23 June 2010
Available online 1 July 2010

Keywords:
Esophageal cancer
Risk
mtDNA haplogroup
Susceptibility markers

ABSTRACT

Both the Taihang Mountain area in north-central China and Chaoshan area in the southeastern littoral of China are areas with high risk of esophageal cancer (EC). Our previous study confirmed that populations from the two areas might share similar matrilineal backgrounds and found that mitochondrial DNA (mtDNA) haplogroup D, especially subhaplogroups D4a and D5a, might be genetic background markers of EC in Chaoshan area. Here, to further determine whether D4a, D5a, and D might be susceptibility markers for EC in the two high-risk areas, we performed a case–control study with larger samples and analyzed the distributions of these three haplogroups in subjects (controls [$n=898$] and patients [$n=768$]) from the two areas. D4a haplogroup was significantly associated with increased risk of EC in Taihang Mountain subjects, especially women. D5 haplogroup was associated with EC at the general population level in the Taihang Mountain area and in subjects $\leq$60 years, especially women $\leq$60 years, in the Chaoshan area. D haplogroup was associated with EC only in subjects $\leq$60 years, especially men $\leq$60 years, in the Chaoshan area. D4a and D5 showing positive association with EC in the Taihang Mountain area became the predominant subhaplogroups of D in Chaoshan controls. In conclusion, D, D4a, and D5 haplogroups might be susceptibility markers for EC in the two high-risk areas in China, particularly D4a and D5 for the Taihang Mountain area and D and D5 for the Chaoshan area.

1. Introduction

Esophageal cancer (EC) is one of the most common cancers with high mortality worldwide. The proportion of EC in China comprises 70% of all cases of the disease in the world (Day and Varghese, 1994; Parkin et al., 1993). China has geographical "hot spots" with high EC incidence. A well-known region at high risk for EC is in the Taihang Mountain area between Henan, Hebei, and Shanxi provinces in north-central China. In 2003, the world standardized incidence for EC in Linxian of Henan province was 133.23/100,000 people (Sun et al., 2007). In the past, the Han inhabitants of north-central China (Henan and Shanxi Hans) continuously migrated into the Chaoshan area in Guangdong province via Fujian province because of warfare and famine and gradually became the predominant inhabitants of the Chaoshan area (Su et al., 2001). The Chaoshan area, located in the southeastern littoral of China and surrounded by the Lianhua Mountains and South China Sea and thus relatively isolated geographically from the inner part of China, is a high-risk EC region. From 1995 to 2004, the world standardized incidence for EC for Nanao Island, a relatively isolated district within the Chaoshan area, was 109.28/100,000 people (Su et al., 2007).

Mitochondrial DNA (mtDNA), strictly maternally inherited, has been widely used for studying evolutionary relationships among human ethnic groups (Pakendorf and Stoneking, 2005; Wallace, 1994, 2005; Wallace et al., 1999). Our previous study of mtDNA polymorphism (Li et al., 2007) confirmed that Chaoshan and Taihang Mountain populations at high risk for EC might share similar matrilineal genetic backgrounds, which might be an importance genetic factor contributing to the high incidence of EC in the Chaoshan area. This previous study found that the mtDNA haplogroup D, especially subhaplogroups D4a and D5a, might be candidate genetic background markers for screening populations susceptible to EC in the Chaoshan area.

Many studies of genetic susceptibility and risk of EC focused on the nuclear genome (Ding et al., 2009; Hu et al., 2003, 2005; Pan et al., 2009; Yue et al., 2004), but few concerned mtDNA. Besides our study, only 1 report exists of EC risk and maternal genetic background represented by mtDNA haplogroup and suggested that haplogroup N was a risk factor for EC in north India (Darvishi et al., 2007). mtDNA

Abbreviations: EC, esophageal cancer; mtDNA, mitochondrial DNA; OXPHOS, oxidative phosphorylation.

* Corresponding author. Fax: +86 754 88900429.
E-mail address: minsu@stu.edu.cn (M. Su).

[1] These authors are co-first authors of this article.

1567-7249/$ – see front matter © 2010 Elsevier B.V. and Mitochondria Research Society. All rights reserved.
doi:10.1016/j.mito.2010.06.005

haplogroups, defined by unique sets of mtDNA polymorphisms that have accumulated sequentially along radiating maternal lineages, can be understood as a monophyletic clade in the rooted mtDNA tree and provide evolutionary information about human ethnic groups (Torroni et al., 2000; Wallace, 1994; Wallace et al., 1999). Certain mtDNA haplogroups may predispose to or protect against some nonmalignant mitochondrial diseases such as metabolic syndrome (Tanaka et al., 2007), Parkinson's disease (Ghezzi et al., 2005), Leber's hereditary optic neuropathy (Herrnstadt and Howell, 2004; Wallace et al., 1999), and schizophrenia (Magri et al., 2007). Thus, the population distribution of population-specific mtDNA haplogroups might not simply reflect human evolution; it might be associated with susceptibility to diseases such as tumors. Haplogroup D was found likely to play a genetic role in endometrial cancer in southwest China (Xu et al., 2006). Haplogroup U was associated with increased risk of prostate cancer and renal cancer in white North Americans (Booker et al., 2006). In European–American women, haplogroup K was associated with significant increased risk of breast cancer, whereas haplogroup U was associated with significantly decreased risk of breast cancer (Bai et al., 2007).

Because of their similar matrilineal genetic backgrounds, Taihang Mountain and Chaoshan populations might share mtDNA lineages associated with increased risk of EC. Our previous study (Li et al., 2007), although revealing that haplogroups D, especially D4a and D5a, are associated with high risk of EC in the Chaoshan area, belonged to population genetics study in which the sample size was small, and such study didn't investigate EC patients in the Taihang Mountain area, thereby, we now used a case–control study with a larger sample size to detect the distribution of D, D4a and D5 in high-risk EC populations and patients from the Taihang Mountain and Chaoshan areas to validate an association between these haplogroups and increased risk of EC in the two areas. We also investigated whether these haplogroups might be susceptibility markers for EC in the two areas. According to the phylogenetic tree of East Asian mtDNA (Kivisild et al., 2002), haplogroup D was categorized into subhaplogroups D4 and D5 and further classified into D4a, D4b and D5a. D5a, a subhaplogroup of D5, is further characterized by mutation of site 16266 in the control region of the mtDNA. However, some individuals with D5 without the 16266 mutation also likely belong to D5a because the 16266 mutation is recurrent. In this study, to avoid information loss, we analyzed the D5 clade, which includes D5a.

2. Materials and methods

2.1. Patients and controls

Blood samples of 1666 unrelated adult Han subjects from the Taihang Mountain and Chaoshan areas in China were collected from 2002 to 2006. Informed consent was obtained from each subject. These subjects were from (1) the Chaoshan high-risk area: 433 patients with EC confirmed on pathology (329 men) who were 28 to 83 years old (mean age 58 ± 10 years), and 418 control subjects (230 men) who were 20 to 96 years old (mean age 49 ± 15 years) but without EC; and (2) Taihang Mountain area: 335 patients with EC confirmed on pathology (291 men) who were 37 to 84 years old (mean age 62 ± 9 years), and 480 control subjects (152 men) who were 40 to 69 years old (mean age 51 ± 7 years) but without EC. The controls from the two EC high-risk areas represented EC high-risk populations.

2.2. DNA extraction and single-nucleotide polymorphism (SNP) genotyping

Genomic DNA was extracted from peripheral blood by use of the SiMax™ Genome DNA Kit (SBS Genetech, China). On the basis of the phylogenetic tree of East Asian mtDNA, we selected for genotyping the defining SNPs for haplogroups D, D4a and D5 of C5178A, C3206T, and A5301G, respectively. Fragments containing these 3 SNPs were amplified by PCR, then 3-D polyacrylamide gel-based DNA microarray assay as described (Xiao et al., 2006, 2007).

Primers and probes were designed by use of Primer 5.0 software (PREMIER Biosoft International, Canada) (Table 1). SNPs 5178 and 5301 were amplified together with 1 set of primers. One of the pair of primers was modified with acrylamide phosphoramidite (Acrydite™; Matrix Technologies) at its 5′ terminal. Each pair of probes was labeled with Cy3 and Cy5 fluorescent dyes at the 5′ terminal. The PCR reactions were performed in a 30 μl reaction solution containing 10 pmol primer and 50 ng genomic DNA and consisted of an initial step at 95 °C for 5 min, then 35 cycles of denaturing at 94 °C for 30 s, annealing at 60 °C for 30 s and extension at 72 °C for 40 s, with a final extension at 72 °C for 5 min.

PCR products were processed by ethanol precipitation, evaporation or left untreated. Solutions containing acrylamide-modified PCR products, glycerol, ammonium persulfate and acrylamide monomers were prepared, spotted, and polymerized onto acryl-modified slides. Tetramethylethylenediamine (TEMED) was introduced onto the microarray to immobilize the modified nucleic acids. Following the attachment on slides, to obtain single-stranded DNA for hybridization analysis, double-stranded DNA on the slides was denatured in 0.1 M NaOH for 10 min. After hybridization, the slides underwent electrophoresis at 5–30 V/cm for 5–20 min in 1 × Tris–borate–EDTA buffer at 4 °C. Images of the slides were captured by use of the LuxScan™-10 K Confocal Scanner (Packard Bioscience, USA) and analyzed by use of Genepix Pro 3.0 software (Axon Instruments Inc, USA).

To confirm the results of microarray hybridization, several PCR products were randomly selected for sequencing by use of ABI Prism 377 DNA Sequence (Applied Biosystems, USA).

2.3. Data analysis

According to the phylogenetic tree of East Asian mtDNA (Kivisild et al., 2002) and with the polymorphism results of the 3 SNPs (5178, 5301, and 3206), subjects with 5178A were categorized as belonging to haplogroup D, those with 5178A–3206C–5301G to D5, and those with 5178A–3206T–5301A to D4a.

The frequencies of the 3 haplogroups (D, D4a, and D5) for patients and controls were calculated separately. The frequency of haplogroup D was the total frequency of all haplogroups in the D clade. The frequency of individual haplogroups in patients was compared to that in controls by χ^2 or Fisher's exact test. To estimate the relative risk of EC for the 3 haplogroups, odds ratios (ORs) with 95% confidence intervals (95% CIs) were calculated for patients in each haplogroup by use of binary logistic regression analysis. We analyzed all subjects and adjusted for age and sex (≤60 years, men ≤60 years, and women ≤60 years) in comparing patients and controls for chi-square test and logistic regression analysis, and each haplogroup was compared to all other haplogroups pooled. Statistical analysis involved SPSS v13.0 (Chicago, IL, USA). A $p<0.05$ was considered statistically significant.

The attributable risk percent among subjects exposed to a given risk factor or set of risk factors is a useful measure of the disease risk associated with the exposure, and can be calculated as follows:

$$A_e\% = \frac{R_e - R_o}{R_e} \times 100\%, \quad (1)$$

where $A_e\%$ is the attributable risk percent among exposed subjects, R_e the absolute risk among exposed subjects, and R_o the absolute risk among subjects unexposed (Cole and MacMahon, 1971). Case–control studies can obtain relative risks but not absolute disease rates; thus $A_e\%$ can be estimated on the basis of the following formula (Cole and MacMahon, 1971):

$$A_e\% = \frac{OR-1}{OR} \times 100\%. \quad (2)$$