Atherosclerosis 218 (2011) 151–155

Contents lists available at ScienceDirect

Atherosclerosis

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

Novel association analysis between 9 short tandem repeat loci polymorphisms and coronary heart disease based on a cross-validation design

Liu Hui*, Ye Jing, Mu Rui, Yu Weijian

College of Medical Laboratory, Dalian Medical University, Dalian 116044, China

ARTICLE INFO

Article history:
Received 7 April 2011
Received in revised form 8 May 2011
Accepted 20 May 2011
Available online 30 May 2011

Keywords:
Coronary heart disease
STR
Heredity

ABSTRACT

Objective: To investigate genes associated with coronary heart disease (CHD) screened with a novel cross-validation design.
Methods: On the basis of age at the onset of the first episode of CHD, stratified sampling by age (<50 years, 50–59 years, 60–69 years, 70–79 years and >80 years) was performed. Alleles of the nine CODIS STR loci including D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, and D7S820, were determined using the STR Profiler Plus PCR amplification kit. Allele frequencies were compared with a control population. The mean age of patients with and without the alleles was compared. Cross-validation was based on differences in both frequency values and ages instead of adjustment procedure for multiple testing.
Results: There were statistical differences in frequency values between the CHD group and the control population for three alleles, and also statistical differences in the age at first onset of CHD for two alleles; at least one allele, D21S11-28.2, was statistically different with regards to both frequency values and age. It was confirmed that D21S11-28.2 is truly related with CHD.
Conclusions: A single true CHD-related allele could be discriminated from the sampling errors through cross-validation. It appears that CHD-related genes may be located near to loci D21S11.

1. Introduction

Coronary heart disease (CHD) is a leading cause of morbidity and mortality worldwide, affecting millions of people in both developed and developing countries. Despite much investigation, the causes of CHD are not yet fully identified. Apart from some rare Mendelian forms of CHD, most cases of CHD are believed to have a multifactorial basis involving a number of genes and environmental factors that interact to determine whether a person will develop the disease [1,2]. Several risk factors for atherogenesis have been recognized, such as old age, male sex, family history, hypertension, dyslipidemia, smoking history, and presence of diabetes mellitus [3]. However, the identification of the genetic component of the etiopathogenesis of CHD remains a great challenge.

In these association analyses, the genes associated with susceptibility or resistance to CHD were defined mainly by investigating the differences between the frequency distributions of relevant genes in CHD patients and controls [4–6]. However, a gene defined as those associated with susceptibility to CHD cannot be a true CHD-related gene, because we cannot confirm whether a statistical difference is due to a 5% sampling error when many alleles are observed in same time. Although the Bonferroni correction is used for multiple comparison, some recent studies provide arguments that Bonferroni correction is too rigid [7]. Association analysis of disease-related gene almost cannot accept from this problem. These considerations necessitate the development of a new research design.

CHD mostly occurs in the population aged over 50 years. Both its incidence and fatality rate increase with age. There are complex interactions between the various etiopathogenic factors including increasing age and genes. We believe that difference in the age at CHD onset between the individuals who carry a certain gene and those who do not is a measure of strength of the genetic effect. Thus, the pathogenic effects of a gene on individuals should include 2 factors: (1) the differences between the frequency distributions of the gene in the disease and control groups (differences in frequency distribution increase with an increase in the effect of a gene) and (2) the age at disease onset in gene carriers (increase in pathogenic effects corresponds to earlier onset). These 2 factors provide information in 2 dimensions; therefore, cross-validation of this information should improve the efficacy of association analysis for the identification of complex disease including CHD related genes.

Microsatellite or short tandem repeats (STR) consist of tandemly repeated DNA units ranging from 2 to 6 nucleotides [8,9]. Because

* Corresponding author. Tel.: +86 411 86110383; fax: +86 411 86110392.
E-mail addresses: liuhui60@mail.dlptt.ln.cn, liuhui60@sina.com (L. Hui).

0021-9150/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.atherosclerosis.2011.05.024

of the variable number of highly polymorphic tandem repeats in humans, the investigation of STR is a powerful tool in human genetics. Forensic laboratories currently use the following STR loci listed in the US national combined DNA indexing system (CODIS) database: D3S1358 (3p), vWA (12p13.3), FGA (4q28), D8S1179 (8pter), D21S11 (21q21), D18S51 (18q21.33), D5S818 (5qter), D13S317 (13q22), and D7S820 (7pter). The methodology is more mature for the detection of these STR loci, we selected these loci because reliable genotyping assays for them are commercially available with international standards [10–12]. This would ensure consistency in allele designation across different laboratories [13]. These nine CODIS STR loci were randomly selected to evaluate CHD; if some CHD-associated STR-alleles were identified, it might be suggested that extensive existence of CHD-associated genes in the genome. Here, we have used a novel cross-validation design based on differences in the age at onset of CHD between individuals who carried a specific gene and those who did not in patients with CHD, and difference of frequency between the distributions of relevant genes in CHD and healthy control participants, in order to identify CHD-associated genes.

2. Materials and methods

2.1. Subjects

All enrolled subjects were unrelated ethnic Han Chinese and signed informed consent. The patients, who were enrolled from Affiliated Hospital of Dalian University, were diagnosed with CHD according to the WHO criteria or by coronary angiography (significant coronary artery stenoses $\geq$50% in at least one major coronary artery) [14]. The age of first onset should be recorded, and those patients hospitalized multiple times, the ages when they were hospitalized due to CHD for the first time should be recorded as the first onset of CHD. There were two exclusion criteria: (1) patients who had previously received intravenous thrombolysis, coronary stent implantation or coronary artery bypass grafting; and (2) patients who could not clearly remember the hospitalization time during their first onset of CHD.

On the basis of their age at the first onset on CHD, stratified sampling by age was performed to secure a uniform distribution of patients. The age distribution of patients was stratified as follows: <50 years (7 men and 4 women), 50–59 years (9 men and 11 women), 60–69 years (8 men and 12 women), and 70–79 years (10 men and 10 women) and >80 years (2 men and 5 women). A total of 78 patients (156 chromosomes) were enrolled.

2.2. Sample preparation and genotyping

DNA was extracted from 3 ml of peripheral blood by the Chelex 100 procedure [15]. Polymerase chain reaction (PCR) was performed using the AmpFLSTR Profiler plus PCR Amplification Kit (Perkin Elmer, Foster City, CA, USA) under conditions recommended by the manufacturer in a reaction volume of 50 μl, using a 9600 Perkin Elmer thermal cycler. Amplification products (1.5 μl) were added to 10 μl formamide and 1 μl of an internal size standard (Genescan-500 ROX, Applied Biosystems). The samples were heat denatured at 95 °C for 5 min and chilled for 5 min in an ice-water bath before performing capillary electrophoresis by using an ABI 310 automated sequencer (Applied Biosystems). We used the Genescan Analysis 2.1 software (Applied Biosystems) to determine fragment sizes. Allele identification was achieved by comparison of the amplified fragments with the allelic ladders included in the reagent set, and the alleles were labeled according to the international nomenclature using the Genotyper Software package (Perkin Elmer).

2.3. Statistical analysis on the gene frequency distributions

In heterozygotes, 2 types of STR are identified at a single STR locus, whereas in homozygotes both the STRs at a single locus are identical. The distributions of genotypes of these polymorphisms were examined using the Chi-square test to determine whether they followed the Hardy–Weinberg equilibrium [16].

Allele frequencies were calculated and compared with control population, who was from same geographic region with subjects in this study (n = 4422) [17]. Constituent ratio of allele frequencies in the CHD group compared to that in control population for each locus by using the chi-square test. The significant loci were chosen, and examined using the chi-square test to determine significant difference of allele frequencies between CHD group and control group.

A p value $\leq$ 0.05 indicated significant difference. The analyses were performed by the SPSS 13.0 statistical software package.

2.4. Statistical analysis on the age differences

The basic concept of the analysis is to consider the age difference of first onset between the individuals who carry a certain gene and those who do not as a CHD effect of a gene. The association between the gene and the CHD can be ascertained by comparing the average age of individuals carrying a gene and that of individuals not carrying the gene. In the present study, the age of the individuals in successive age group was considered as the dependent variable and that of the gene carriers was considered as the grouping variable; this analysis was performed using the SPSS 13.0 statistical software package.

The significant difference of mean age among the subjects carrying the genes for each locus were chosen with Mann–Whitney U test, and examined using the Mann–Whitney U test to determine significant difference between the mean age of the subjects carrying the genes and that of subjects that did not carry the genes. A p value of $\leq$0.05 indicated significant difference.

If the average age of the subjects who carry a certain gene is significantly lower (susceptibility for CHD) or higher (resistance for CHD) than that of the subjects who did not carry the gene, then the gene can be regarded as a CHD-related gene. If the total number of genotype-positive chromosomes in the subjects was less than five, the sample size is considered too small for statistical analysis; this allele was be excluded from the observed alleles.

2.5. Cross-validation of the 2 sets of analytical results

According to the probability principle of multiplication, the p value (type I error) is as follows:

$$p = 1 - (1 - p1) \times (1 - p2)$$

where $p1$ and $p2$ are the p values obtained with the results of the 2 analyses, i.e., the method of gene frequency distributions and the method of age differences, respectively, for 1 gene. A p value of $\leq$0.05 indicated significant difference.

3. Results

The number of alleles observed in our experiment, data regarding the STRs, and chromosomal localization for these 9 STR loci are summarized in Table 1 [17]. For each locus, the Hardy–Weinberg equilibrium was tested by comparing the observed genotype numbers with those expected under the hypothesis of panmixia (Hardy–Weinberg equation); no deviations from the Hardy-Weinberg equilibrium were observed in control group, as shown in Table 1.