Polymorphisms associated with the risk of lung cancer in a healthy Mexican Mestizo population: Application of the additive model for cancer

Rebeca Pérez-Morales1,*, Ignacio Méndez-Ramírez2, Clementina Castro-Hernández1, Ollin C. Martínez-Ramírez1, María Eugenia Gonsebatt1 and Julieta Rubio1

1Departamento de Medicina Genómica y Toxicología Ambiental, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, México.
2Departamento de Probabilidad y Estadística, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Mexico City, México.

Abstract

Lung cancer is the leading cause of cancer mortality in Mexico and worldwide. In the past decade, there has been an increase in the number of lung cancer cases in young people, which suggests an important role for genetic background in the etiology of this disease. In this study, we genetically characterized 16 polymorphisms in 12 low penetrance genes (AhR, CYP1A1, CYP2E1, EPHX1, GSTM1, GSTT1, GSTP1, XRCC1, ERCC2, MGMT, CCND1 and TP53) in 382 healthy Mexican Mestizos as the first step in elucidating the genetic structure of this population and identifying high risk individuals. All of the genotypes analyzed were in Hardy-Weinberg equilibrium, but different degrees of linkage were observed for polymorphisms in the CYP1A1 and EPHX1 genes. The genetic variability of this population was distributed in six clusters that were defined based on their genetic characteristics. The use of a polygenic model to assess the additive effect of low penetrance risk alleles identified combinations of risk genotypes that could be useful in predicting a predisposition to lung cancer. Estimation of the level of genetic susceptibility showed that the individual calculated risk value (iCRV) ranged from 1 to 16, with a higher iCRV indicating a greater genetic susceptibility to lung cancer.

Key words: additive model, genetic polymorphism, lung cancer, Mexicans, molecular epidemiology.

Received: February 3, 2011; Accepted: July 18, 2011.

Introduction

Lung cancer (LC) is the major cause of mortality from neoplasias worldwide. In Mexico, LC was responsible for 11.45% of deaths by malignant neoplasia from 1998-2002, and a prospective study indicated that mortality from LC will be even greater in the future (Ruiz-Godoy et al., 2007). Lung cancer is a serious public health problem (Proctor, 2001). The World Health Organization (WHO) estimates that in 2030 the number of deaths attributable to the consumption of tobacco will be 100 million, accompanied by an increased incidence of LC (Xie and Minna, 2008).

Lung cancer is highly related to tobacco consumption, although only 20% of smokers develop LC. On the other hand, the number of cases among people under 50 years old has increased (Gemignani et al., 2007), highlighting the importance of genetic background in the etiology of LC. In this context, efforts have been made to identify polymorphisms associated with the risk of developing LC. Association studies have focused mainly on polymorphisms in genes coding for (1) enzymes involved in the bioactivation of carcinogens, such as cytochromes P450 CYP1A1 (Shah et al., 2008), CYP2E1 (Zhan et al., 2010) and CYP2D6 (Shaw et al., 1998), (2) enzymes involved in detoxification, e.g., glutathione transferases GSTM1, GSTT1 and GSTP1 (Sobti et al., 2008), N-acetyl transferase 2 (NAT2) and epoxide hydrolase (EPHX1) (Zhou et al., 2002), (3) proteins involved in DNA repair, such as XRCC1, XRCC3, ERCC2, ERCC4, MGMT and OGG1 (Zienolddiny et al., 2006; Hung et al., 2007) and (4) proteins implicated in the cell cycle and apoptosis, such as CCND1, CHEK2, MDM2, TP53 and EGF (Hosgood III et al., 2008).

In some populations, a significant association between gene polymorphisms and the risk of LC risk has been established, while in other populations no associations have been found, probably because of the low frequency of polymorphisms. For example, the CYP1A1 rs1048943
environmental and tobacco smoke carcinogens, with their potential to cause mutations leading to LC, particularly in specific genes can modulate the formation of DNA adducts (Kawajiri, 1999) and the repair of DNA. The study included 382 unrelated, healthy Mexican Mestizo individuals whose parents and grandparents were born in Mexico. After providing informed consent, the subjects answered a questionnaire that included information on their age, gender, smoking status and lifestyle.

Polymorphism analysis

Genomic DNA was obtained from blood samples by phenol-chloroform extraction and ethanol precipitation (Sambrook et al., 1987). The presence of the $AhR$ rs2066853 (Arg554Lys) polymorphism was determined by using the TaqMan probe C_11170747_20 from Applied Biosystems, according to the manufacturer’s recommendations. Restriction fragment length polymorphism (RFLP) analysis was used to assess the following polymorphisms: $CYP1A1$ rs4646903 (6235T > C) and $CYP1A1$ rs1800031 (5639T > C) (Kawaijiri, 1999), $CYP1A1$ rs1048943 (Ile462Val) and $CYP1A1$ rs1799814 (Thr461Asn) (Cascorbi et al., 1996), $CYP2E1$ rs2031920 (-1053C > T) (Morita et al., 1997), $EPHX1$ rs1051740 (Tyr113His) and $EPHX1$ rs2234922 (His139Arg) (Smith and Harrison, 1997), $GSTPI$ rs947894 (Ile105Val) (Harries et al., 1997), $XRCC1$ rs25487 (Arg399Gln) (Abdel-Rahman and El-Zein, 2000), $ERCC2$ rs13181 (Lys751Asn) (Lunn et al., 2000), $MGMT$ rs12917 (Leu84Phe) (Courtney et al., 2005), $CCND1$ rs603965 (870G > A) (Wang et al., 2002), and $TP53$ rs1042522 (Arg72Pro) (Irarrázabal et al., 2003). Polymorphisms of $GSTT1$ null and $GSTM1$ null were determined by multiplex PCR using a previously reported protocol (Abdel-Rahman et al., 1996).

Statistical analyses

The statistical package GenePop version 4.0.10 (http://genepop.curtin.edu.au) was used to assess whether the genotypes of each gene were in Hardy-Weinberg equilibrium and to determine the degree of linkage between the $EPHX1$ and $CYP1A1$ gene polymorphisms. Conglomerate and hierarchical clustering analyses were used to determine the genetic variability of the sample. The estimated cancer risk and the genotypic and allelic frequencies were determined using the statistical package JMP version 8.

Materials and Methods

Subjects

The research protocol was approved by the Committee of Bioethics of the Instituto de Investigaciones Biomédicas of the Universidad Nacional Autónoma de México, and the Hospital “20 de Noviembre” ISSSTE gave permission to use the buffy coat of blood bank samples as a source of DNA. The study included 382 unrelated, healthy Mexican Mestizo individuals whose parents and grandparents were born in Mexico. After providing informed consent, the subjects answered a questionnaire that included information on their age, gender, smoking status and lifestyle.
Results

Of the 382 Mexican Mestizo individuals studied, 29% were women and 71% were men. The age range was between 18 and 80 years (mean age: 39.2 ± 12.1 for men and 41.5 ± 13.1 for women) and 48% of the population were smokers.

The following polymorphisms were studied in candidate genes: Ahr rs2066853, CYP1A1 rs4646903, CYP1A1 rs1048943, CYP1A1 rs1799814, CYP2E1 rs2031920, EPHX1 rs1051740, EPHX1 rs2234922, GSTM1 null, GSTT1 null, GSTP1 rs947894, XRCC1 rs25487, ERCC2 rs13181, MGMT rs12917, CCND1 rs603965 and TP53 rs1042522. The genotypic and allelic frequencies of these polymorphisms are shown in Table 1. All genotypes were in Hardy-Weinberg equilibrium.

GSTM1 and GSTT1 were not analyzed for Hardy-Weinberg equilibrium because the methodology did not allow discrimination between heterozygous and homozygous positive genotypes.

Polymorphisms in the same gene may have synergistic or antagonistic effects, as in the case of CYP1A1 and epoxide hydrolase 1. This is functionally significant because CYP1A1 acts in association with EPHX1 to convert polyaromatic hydrocarbons to highly toxic, mutagenic and carcinogenic epoxides.

Linkage analysis of four CYP1A1 and two EPHX1 polymorphisms showed that the CYP1A1 rs1799814 and CYP1A1 rs1800031 genotypes were linked in all cases, with a probability of 1 (p < 0.001), whereas CYP1A1 rs4646903 and CYP1A1 rs1048943 polymorphisms were not linked, with a probability of 0 (p < 0.001). EPHX1 rs1051740 and EPHX1 rs2234922 were linked, with a probability of 0.03254 (p = 0.0025). Combinations of CYP1A1 polymorphisms are shown in Table 2.

Conglomerate analysis was used to determine the genetic variability and the possible grouping of the individuals analyzed. We identified six groups that clustered according to their genetic characteristics, although there was considerable heterogeneity (Figure 1). Additionally, hierarchical cluster analysis of the genotypes showed that CYP1A1 rs1800031 and CYP1A1 rs1799814 polymorphisms clustered together, whereas CYP1A1 rs4646903 and CYP1A1 rs1048943 were close to each other but clustered separately (Figure 2). The EPHX1 rs1051740 and EPHX1 rs2234922 polymorphisms were widely separated, as also indicated by linkage analysis.

To determine the theoretical levels of susceptibility to LC, a risk matrix was generated using a log-additive model in which a value of 0 was assigned to homozygous genotypes that produced no risk, 1 to heterozygous genotypes (medium risk) and 2 to homozygous genotypes that produced changes in the activity of the protein, considered to be high risk. The individual calculated risk values (iCRVs) were determined by adding the values of the log-additive model to each locus. The iCRVs ranged from 1 to 16, although there was a marked decrease in individuals with iCRV ≤ 5 and ≥ 12 (Figure 3).

Table 1 - Frequency of lung cancer risk polymorphisms among healthy Mexican Mestizos.

<table>
<thead>
<tr>
<th>Gene (rs)</th>
<th>Genotype frequency, n (%)</th>
<th>Allelic frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mutant allele</td>
<td>n</td>
</tr>
<tr>
<td>Ahr (2066853)</td>
<td>Lys</td>
<td>370</td>
</tr>
<tr>
<td>CYP1A1 (4646903)</td>
<td>C</td>
<td>382</td>
</tr>
<tr>
<td>CYP1A1 (1048943)</td>
<td>Val</td>
<td>382</td>
</tr>
<tr>
<td>CYP1A1 (1800031)</td>
<td>C</td>
<td>382</td>
</tr>
<tr>
<td>CYP1A1 (1799814)</td>
<td>Asn</td>
<td>382</td>
</tr>
<tr>
<td>CYP2E1 (2031920)</td>
<td>T</td>
<td>382</td>
</tr>
<tr>
<td>EPHX1 (1051740)</td>
<td>His</td>
<td>382</td>
</tr>
<tr>
<td>EPHX1 (2234922)</td>
<td>Arg</td>
<td>382</td>
</tr>
<tr>
<td>GSTP1 (947894)</td>
<td>Val</td>
<td>382</td>
</tr>
<tr>
<td>GSTM1 (deletion)</td>
<td>Null</td>
<td>382</td>
</tr>
<tr>
<td>GSTT1 (deletion)</td>
<td>Null</td>
<td>382</td>
</tr>
<tr>
<td>XRCC1 (25487)</td>
<td>Gln</td>
<td>382</td>
</tr>
<tr>
<td>ERCC2 (13181)</td>
<td>Gln</td>
<td>382</td>
</tr>
<tr>
<td>MGMT (12917)</td>
<td>Phe</td>
<td>382</td>
</tr>
<tr>
<td>CCND1 (603965)</td>
<td>A</td>
<td>382</td>
</tr>
<tr>
<td>TP53 (1042522)</td>
<td>Pro</td>
<td>382</td>
</tr>
</tbody>
</table>

aNot determined. b – wild type allele. c – mutant allele.
Discussion

Cancer is a polygenic disease, the risk of which may be related to the presence of low-penetrance genes that have additive effects. In this study, we examined the frequency of some polymorphisms possibly related to the risk of developing lung cancer in a sample of healthy Mexican Mestizos. These polymorphisms may be useful biomarkers of genetic susceptibility to lung cancer in specific populations.

We have previously reported on the high frequency of the CYP1A1 rs1048943 polymorphism in Mexican Mestizos (Pérez-Morales et al., 2008), a polymorphism that is highly represented in Amerindians (Kvitko et al., 2000). In some populations, the CYP1A1 rs1048943 polymorphism has been associated with a risk of LC (Quinoñes et al., 2001; Lee et al., 2008; Shah et al., 2008), although not all studies have found such an association, especially when the frequency of this polymorphism is very low, as in European and American Caucasian populations (Hung et al., 2003). However, in some populations the CYP1A1 rs1048943 polymorphism has a significant influence on the risk of developing LC because of the additive effect of other polymorphisms.

Although the CYP1A1 rs1799814 and CYP1A1 rs1800031 genotypes were found to be linked, the linkage analysis may have been affected by the high frequencies of...
the wild type allele at these loci in our population. On the other hand, in contrast to a previous report (Hayashi et al., 1991), the CYP1A1 rs4646903 and CYP1A1 rs1048943 polymorphisms were not linked. This observation is functionally significant because the CYP1A1 rs1048943 polymorphism increases the activity of the encoded enzyme (Hassett et al., 1994), which could suppress the low activity of the EPHX1 113His variant (Salam et al., 2007). For example, an individual with the CYP1A1 rs4646903, CYP1A1 rs1048943, EPHX1 rs1051740 and EPHX1 rs2234922 polymorphisms will produce reactive metabolites more efficiently and consequently have a higher risk of developing LC.

A conglomerate analysis of the genetic variability of the population studied here revealed six groups that clustered according to their genetic characteristics (Figure 1). In clusters 2 and 4, the polymorphism CYP1A1 rs1800031 was quite separate from the rest of the genotypes, although it should be noted that these two groups consisted of only two subjects each such that the probability of finding individuals belonging to these clusters in a given population is very low. In clusters 2, 3, 4 and 5, AhR rs2066853, EPHX1 rs2234922, GSTT1 null, ERCC2 rs13181, and CYP1A1 rs1799814 were over-represented with respect to the other polymorphisms analyzed, but the number of individuals was smaller than in clusters 1 and 6; there was no predominant genotype in the latter two clusters because of the combination of polymorphic alleles at each locus in each individual. Despite this clustering, our findings indicate that the population studied was highly heterogenous, as is characteristic of Mexican Mestizos.

Hierarchical cluster analysis showed that the genes were grouped according to their frequency. This analysis also revealed that CYP1A1 rs1800031 and CYP1A1 rs1799814 occurred together, in agreement with the linkage analysis, whereas CYP1A1 rs4646903 and CYP1A1 rs1048943 were found together in some individuals but were not linked (as indicated by the separation of their branches). Overall, this analysis showed that a portion of the population carried both risk alleles, although they were not linked. However, we have not determined whether these polymorphisms are in a cis or trans position, which could affect the linkage results.

Our analysis revealed individuals with a high susceptibility to LC based on the presence of risk genotypes, although different combinations of risk genotypes may confer varying degrees of susceptibility when combined with other components, such as environmental factors.

If the risk of developing LC is attributed to the interaction of several low-penetrance genes that exert an additive effect then we should be able to detect individuals with a high number of risk alleles and a greater genetic susceptibility to LC. We estimated the levels of susceptibility and found that the iCRV ranged from 1 to 16; there was a marked decrease in the susceptibility to LC among individ-
uals with an iCRV $\geq 12$. Application of the polygenic model (Fletcher and Houlston, 2010) (Figure 3) yielded a normal distribution of risk alleles in which low and high risk individuals occurred at the extremes of the distribution. Theoretically, high-risk individuals should be more susceptible to LC, but one cannot exclude the important role of genotype x environmental interactions to which individuals are exposed.

Based on the polygenic model of cancer, which takes into consideration the additive effect of multiple risk alleles, the high risk genotypes identified in this study included genes involved in phase I and II metabolism, DNA repair, oxidative stress and cell cycle regulation. All of these gene groups need to be considered when analyzing the efficiency of reactive metabolite generation, DNA-adduct repair, damage persistence and the cellular response to cycle arrest or apoptosis. However, as shown here, only a small minority of individuals actually possess a large number of risk alleles. Although studies of other populations have associated specific polymorphisms with a risk of LC, these relationships have not always been confirmed, possibly because of the pleiotropic or epistatic effects of one genotype on another within the same population.

Since LC is a multifactorial, polygenic disease, it is incorrect to attribute the susceptibility to LC to a single gene or group of related genes. In this context, determining the genetic background of a healthy population, such as done here for a Mexican Mestizo population, can provide a sound basis for subsequent studies on the association between these risk genotypes and LC. In such cases, the iCRV should be higher in patients than in the healthy controls.

Acknowledgments

This study was supported by a grant (PAPIIT-IN219807) from the Universidad Nacional Autónoma de México to J.R. and a post-graduate scholarship (no. 204279) from the Consejo Nacional de Ciencia y Tecnología to R.P. We thank Dr. Sergio García Méndez, head of the Blood Bank, Hospital “20 de Noviembre” ISSSTE in Mexico City, for providing the samples, and Dr. Pavel Petrosyan for critical reading of the manuscript.

References


Internet Resources