1. Introduction

Telomeres, the nucleoprotein complexes at the end of eukaryotic chromosomes, are specialized structures that protect chromosome ends (Blackburn 1982). Telomeres are composed of TTAGGG repeats and a specific associated protein complex termed shelterin, which regulates telomere protection and length. Telomerase is the key telomere maintenance enzyme. Most adult human cells have limited amount of telomerase so that successive cell divisions lead to progressive telomere attrition due to the end-replication problems (Blassio MA et al., 1991; Maser RS et al., 2002).

Though it is widely considered that telomere dysfunction plays an important role in human carcinogenesis, the relationship between telomere health in somatic cells and the risk of developing non-small cell lung cancer is not well defined.

In this study, we tested the differences in demographic characteristics and several new telomere length features between non-small cell lung cancer cases and controls, and then selected significant predictors combined as a biomarker, which may apply to lung cancer detection.

2. Method

2.1 Data Description

The study population and chromosome preparation from blood lymphocyte and cell lines were described previously [1]. This analysis concentrates on a subset of subjects to whom the chromosome preparations from blood lymphocytes were available.

2.2 Definition of Telomere Features

Telomere length at each of the chromosome ends was measured by telomere quantitative fluorescent in situ hybridization (Q-FISH) as previously described [2]. The telomere lengths are defined as follows: (1) Sqi, Sqi, Sq2, and Sq3, defined as lengths of telomers from the ith autosome (i = 1, 2, 22, 23); (2) Sk1, Sk2, Sk3, and Sk4, defined as lengths of telomers from female’s sex chromosome; (3) Xp, Yq, Yp and Yq, defined as lengths of telomers from male’s sex chromosome. For each study subject, 30 metaphase cells were analyzed. For each of the telomere lengths, we calculate the average of the 30 metaphase cells from each subject, and then calculate the average of the telomere lengths of all subjects.

By Student t-test, there is no significant difference in any of the telomere lengths for any autosome or sex chromosome between cases and controls. Therefore, we define telomere length features as follows.

$\hat{\delta}_1 = \log(\text{Sqi}_1) + \log(\text{Sqi}_2)$ and $\hat{\delta}_2 = \log(\text{Sqi}_3) + \log(\text{Sqi}_4)$

$\hat{\delta}_3 = \log(\text{Sk1}) + \log(\text{Sk2})$ and $\hat{\delta}_4 = \log(\text{Sk3}) + \log(\text{Sk4})$

$\hat{\delta}_5 = \log(\text{Xp}) + \log(\text{Yq})$ and $\hat{\delta}_6 = \log(\text{Yp}) + \log(\text{Yq})$

$\hat{\delta}_7 = \log(\text{Xp1}) + \log(\text{Sk1})$ and $\hat{\delta}_8 = \log(\text{Xp2}) + \log(\text{Sk2})$

$\hat{\delta}_9 = \log(\text{Xp3}) + \log(\text{Sk3})$ and $\hat{\delta}_10 = \log(\text{Xp4}) + \log(\text{Sk4})$

2.3 Statistical Analyses

2.3.1 Preliminary Analyses

Epidemiology data: Chi-square test and Student t-test; Telomere features: Student t-test. All analyses are performed at significant level of 0.05 and p-values are two-sided.

2.3.2 Backward Stepwise Logistic Regression

First, a logistic regression with all potential predictors is conducted. Since we attempt to select the variables that make significantly contributions to fitting the model, a backward stepwise selection is then used by minimum Akaike Information Criterion (AIC) value. Next, we delete one insignificant predictor with the largest p-value by time, till each of the predictors was significant. Eventually, we get a revised logistic regression model and combine those significant predictors as a biomarker.

2.3.3 Receiver Operating Characteristic (ROC) Curve

ROC curve is a graphical plot which is created by plotting sensitivity (the ability to predict an event correctly) versus 1-specificity for the possible cut-off classification probability values.

3. Result

Table 1 showed the demographic characteristics; Table 2 showed the significant T-test results of telomere length features between non-small cell lung cancer cases and controls.

Table 1: Demographic characteristics of lung cancer cases and controls.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Case (n=137)</th>
<th>Control (n=54)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td>66 (33)</td>
<td>67 (33)</td>
<td>0.7134</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Female</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>92 (49.2)</td>
<td>54 (43.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>45 (29.5)</td>
<td>30 (21.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 (7.3)</td>
<td>10 (7.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 (2.2)</td>
<td>3 (2.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 (5.8)</td>
<td>8 (5.8)</td>
<td></td>
</tr>
</tbody>
</table>

4. Conclusion

The aim of this study is to examine and select disease-related demographic characteristics and new telomere length features as a biomarker to detect non-small cell lung cancer. According to the results, current smoking status, smoking years and the telomere length variations on the 3p, 7q, 8p, 13p and 14p chromosomes significantly contribute to non-small cell lung cancer. The ROC curve shows the prediction of this biomarker is accurate (AUC=0.862). However, biological procedures are needed to further identify and explain this biomarker.

Reference