Identification of Susceptibility Allele of Lung Adenocarcinoma

Hsuan-Yu Chen
Institute of Statistical Science, Academia Sinica, Taiwan

Abstract

The whole genome re-sequencing analysis by using next generation sequencer was applied to a multiplex lung adenocarcinoma family with the affect mother, 4 affected daughters, and 1 non-affected son. A total of 438 billion bases were sequenced and the average coverage depth was 24X. In each sample, more than 90% coverage of human whole genome was sequenced. The average number of detected SNPs and small InDels of mother, son, and 4 daughters is around 3 million and 150,000, respectively. Under the dominant model assumption, there were 70,827 SNPs and 1,988 small InDels detected from the sequence data. Focusing on the 240 non-synonymous SNPs obtained, we genotyped these SNPs by MALDI-TOF MS and compared with 4 published individual genomes and 30 normal persons. The results showed that 6 SNPs had the same genotypes between NGS and MALDI-TOF MS, not found in 4 individual genomes, and absent in 30 normal persons. The findings were further validated in a large external normal cohort (n=1146) and a lung adenocarcinoma cohort (n=1055). One non-synonymous mutation site of an important transcription regulator was confirmed. In addition, we compared the results in Asian population of the 1000 genome project. Only 1 out of 286 Asian population samples carried this mutation and we obtained an odds ratio OR=4.1 (p value= 0.035) after adding Asian population data into the normal population (n=1146+286=1432). The findings showed that the allele frequency of risk allele was higher in lung adenocarcinoma than the general normal population.