Whole genome sequencing approach to identify rare allele and risk of lung adenocarcinoma

Hsuan-Yu Chen (陳璿宇)
Institute of Statistical Science, Academia Sinica

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 findings were further validated in a large external normal cohort ($n > 1000$) and a lung adenocarcinoma cohort ($n > 1000$). The findings of this study showed that the rare allele was identified and the carrier frequency was higher in lung adenocarcinoma than the general normal population. The disease associated risk alleles have to be rare and traditional GWAS studies could not explore rare allele due to original probe design of the SNP array. Whole genome sequencing can resolve the limitations of GWAS studies and find more rare and novel disease associated alleles.