Phasing haplotypes of HLA genes from Next Generation Sequencing data at individual level

Chenxi Zhao[1],
zhao.bio@tmd.ac.jp

Nay Chi Htun[1],
nhtun.epi@mri.tmd.ac.jp

Noriko Sato[1],
nsato.epi@mri.tmd.ac.jp

Masaaki Muramatsu[1]
muramatsu.epi@mri.tmd.ac.jp

[1] Department of Molecular Epidemiology, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

Keywords: Next Generation Sequencing, Human leukocyte antigen, Haplotype

Objective: Haplotype information of an individual is valuable for practical applications of personalized medicine, such as pharmacogenomics. For instance, haplotypes of HLA (human leukocyte antigen) genes, which are essential for immune functions, are reported to be associated with Stevens–Johnson syndrome (SJS), a drug induced severe adverse reaction of the skin. In order to facilitate pharmacogenomics application of NGS (Next Generation Sequencing) data, we employed these data to define haplotypes of HLA genes.

Method: Whole-Genome Sequencing (WGS) and Whole-Exome Sequencing (WES) data was collected from public database. Initially, we focused on HLA-B haplotypes, which are involved in SJS induced by Carbamazepine and Allopurinol. Fastq data was mapped against hg19 reference genome and HLA-B gene data was extracted. Genotypes and haplotypes of HLA-B were detected by applying the linkSNPs program [1]. Integrative Genomics Viewer (IGV) was used for visualizing short reads.

Results: Haplotype phasing results showed that, 22 haplotype clusters were found. Combined information from the short reads, all 7 exons of HLA-B gene can be phased into 4 major parts, with 16 possible haplotype combinations. Protein blast result of the 16 combinations above showed that 10 candidate haplotypes showed the highest scores. We further used haplotype frequency table, to rank the probability of haplotype combinations in the individual [2]. As a result, we inferred the most likely HLA-B gene haplotypes.

Conclusion: Haplotype information of HLA genes from an individual NGS data is an important issue in personalized pharmacogenomics. As a practical example, the method introduced in this study may be useful in estimation of adverse reaction risks of SJS.

[1] Sasaki E, Sugino RP, Innan H, The linkage method: a novel approach for SNP detection and haplotype reconstruction from a single diploid individual using next-generation sequence data, *Mol Biol Evol*, 30(9):2187-2196, 2013.

[2] Maiers M, Gragert L, Klitz W, High-resolution HLA alleles and haplotypes in the United States population, *Hum Immunol*, 68(9):779-788, 2007.