外国語要旨

Bioinformatic analysis of the effect of single nucleotide polymorphism of human drug metabolism genes on function and conformation of protein

Mika Sakamoto

Recently, the technique for analyzing the genomic sequence has been improved, and we found numerous variation of the human genome and the genes. This study focused on the genomic variation of cytochrome P450 and ABC transporter. Human cytochrome P450 (P450) is an enzyme that is associated with the oxidative metabolism of a large number of xenobiotics and endogenous organic compounds. P450 is known for its diversity, and we found numerous genetic polymorphism in genes coding P450. Genetic polymorphism of P450 has been extensively studied with regard to alteration in enzymatic activity. The distribution of genetic polymorphism of P450 shows significant difference among geological regions. Genetic polymorphism is an important factor that influences drug concentrations and has the potential to predict the optimal dosage of drug in personalized medicine. GnomAD, database of human genetic variant, and PharmVar, a database of pharmacogene variant, contains information of many variant of P450 genes. ClinVar, a database of human genetic variation with information of clinical significance, also contains variant information of of many P450 genes, but there are few information on variants that are identified.

In this study, I focused on genes related to drug transport and metabolism among the diversity found in the human genome sequence. When drugs are metabolized in cells, cytochrome P450 (P450) oxidizes, reduces, hydrolyzes the drug, ABC transporter pumps the metabolite out of the cell.

This study was to clarify the characteristics of genetic variants of P450 genes included in gnomAD exome. First, the relationship between the characteristics of amino acid substitution and the minor allele frequencies in three human population; East Asian, African and Caucasian, was examined. The result showed that there were no significant differences in the missense variant with different allelic frequencies in the three human population. Next, to clarify the three-dimensional structural features of the variation, the amino acid substitution sites by missense variant of P450 genes were mapped on the three-dimensional structures of P450 proteins. Features of the amino acid substitution sites were acquired from the coordinates of amino acid residue on the three-dimensional structure, and the information of protein-protein interaction related to redox and known pathogenic/drug response mutation. With classification by principal component analysis, the features of missence variants of P450 genes whose clinical significance is known

was analyzed. The result of principal component analysis was able to classify the variants by the features computed from the coordinates of three-dimensional structure, such as the distance of known pathogenic/drug response mutation to the target residue and the distance of the residue related to protein-protein interaction to the target residue. Furthermore, the set of features of known pathogenic and drug response mutation of P450 genes computed from the coordinates of three-dimensional structure was used to train classifiers using several algorithms of machine learning, such as logistic regression classifier, support vector machine classifier and random forests classifier. The accuracy of these classifiers were compared with one another and it was found that the random forests model was the best model to classify variants of P450. The pathogenicity and drug responsibility of the missense variants of P450 genes of unknown clinical significance were predicted using the random forests classifier. Approximately one third of missense variant of P450 genes that were unknown for clinical significance were classified as pathogenic or drug responsible. This result may lead that the information of three-dimensional structure of protein is a valid source for predicting the effect of missense variant of P450 genes.
ABC transporter family is a huge group in the transporter membrane proteins and actively transports the substrates using the energy derived from ATP hydrolysis. A variation of a single amino acid in the amino acid sequence of ABC transporter has been known to be linked with certain disease. The mechanism of the onset of the disease by the variation is, however, still unclear. I compared the structures of ABC transporter in apo and ATP-binding forms and found a possible conformation shift around pivot-like residues in the transmembrane domains. When this conformation change in ABC transporter and the location of pathogenic variation were compared, I found a reasonable match between the two, explaining the onset of the disease by the variation. This study provided a new approach using three-dimensional structure data to clarify the function of the pathogenic variant. And it is expected to clarify individual differences caused by genetic variants and to promote preventive and personalized medicine.