Genome-wide association study and whole-genome sequence analysis

Tatsuhiko Tsunoda, Group Director, RIKEN Center for Genomic Medicine

Whole-genome sequence analyses including genome-wide association studies (GWAS), which exhaustively explore disease-related genes in the human genome, have revolutionized medical research. In 2002, our center reported the first GWAS results in the world (*Nature Genetics* **32**: 650-654, 2002). In 2004, using the gene-based data, I constructed the first linkage disequilibrium (LD) map in the world, and found exotic patterns of natural selection on genes (*Human Molecular Genetics* **13**: 1623-1632, 2004). Thereafter, we participated in the International HapMap project to construct an LD map and select tagging SNPs, which have been used for chips/arrays (*Nature* **449**:851-861, 2007). This has resulted in a large increase in the number of GWAS, further accelerated by the BioBank Japan project, revealing many genes related to common diseases, cancers, and drug responses. To extend the coverage, we are now enlarging sample sizes using disease cohorts and performing meta-analysis through collaborations around the world. To explore hidden SNPs with lower allele frequency, as well as to combine data from chips with different marker sets, we use the imputation technique with reference haplotypes. One promising approach for developing higher quality reference haplotypes and for exploring unknown variation is analyzing lower frequency variations, e.g. SNVs and CNVs, through next-generation sequencing (NGS). To develop our analytical pipeline for NGS, we sequenced a single genome at high coverage, resulting in the first report of a Japanese individual's whole-genome sequence (*Nature Genetics* **42**:931-936, 2010). That work allowed us to establish methodologies for detecting multiple types of variations: single nucleotide variations (SNVs), structural variations including copy number variations (CNVs), and novel sequences. Using experimental validation, we confirmed that our methods could detect these multiple types of variation with high accuracy. In addition to those methodological analyses, examination of the first Japanese whole-genome sequence revealed that considerable variation, particularly potentially-functional rare variation, remains undiscovered in the human genome. Based on our methodology, we have constructed a pipeline for analyzing cancer genomes (*Nature* **464**:993-998, 2010) as well as whole-exome analysis for common/monogenic diseases, which I will discuss further.