This report describes the activities and results of my sabbatical leave during the 2007 Fall semester. The purposes of the sabbatical were to refresh my knowledge in the recent development of statistical genetics and biological sciences and to visit some of the other statistical geneticists to exchange ideas and to stimulate new research topics. In the following, I will describe in detail the activities of my sabbatical leave.

1. Refreshing my knowledge in the recent development of statistical genetics and redirecting my research efforts from candidate gene association studies to whole genome association studies required extensive reading, and they occupied a large portion of the sabbatical leave time. This reading lead to the research projects described below.

   a) With the availability of large-scale genotyping technologies, the genotyping cost of genome-wide association (GWA) studies has been largely reduced and a boom of large-scale GWA studies is underway. Nevertheless, the success of most association studies partially depended on the development of the statistical methods to analyze the data. Currently, the statistical methods are extended from candidate gene to GWA studies through uniform-sized sliding-window approaches with window-size 3-5. This approach is problematic because the LD patterns certainly vary frequently over the genome. Therefore, I proposed the idea of the variable-sized sliding-window approach and the window-size being decided by the local LD pattern. This idea has resulted in a paper entitled “A Variable-sized Sliding-Window Approach for Genetic Association Studies via Principal Component Analysis” which was accepted for publication by the journal of Annals of Human Genetics.

   b) Currently, the statistical methods that are applicable to GWA studies are essentially single-marker methods. However, complex diseases are
presumed to be the results of interactions of several genes and environmental factors, with each gene only having a small effect on disease. Thus, the methods that can account for gene-gene interactions to search for a set of marker loci in different genes or across genome and to analyze these loci jointly are critical. Therefore, I proposed to use the linear combination of many weak signals to predict the trait. This idea resulted in a paper entitled “An Ensemble Learning Approach Jointly Modeling Main and Interaction Effects in Genetic Association Studies” which was published in Genetic Epidemiology.

2. The purpose 2 did not take exactly the course as I expected, but was successful nevertheless. Because of my health problem, I cannot visit Columbia University and Case Western Reserve University. Thus, I invited associate professor Xiaofeng Zhu from Case Western Reserve University to visit me instead. During this visit, we mainly discussed the statistical methods for GWA studies under family-based design. This visit sparked several new research topics which included “Joint analysis for genome-wide association studies in family-based designs”, “Multi-marker two-stage joint analysis for genome-wide association studies under family-based design”, and “Multi-marker two-stage joint analysis for genome-wide association studies under family-based design”.