— Invited Talk —

Human Genome Analysis and Medicine in the 21^{st} Century

Yusuke Nakamura
yusuke@ims.u-tokyo.ac.jp

Director, Human Genome Center
Institute of Medical Science, The University of Tokyo
4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

The human genome project is now considered to be the most important project in biological and medical research. The discovery of entire human genes that are estimated to be 70,000-150,000 in our genome, through this project must revolutionize biological medicine including molecular diagnosis of various diseases and development of novel treatment. DNA sequences of an entire human genome consisting of 3×10^9 nucleotides will be completely determined by 2003, and 90% of our genes will be identified by 2001 although it will take 10-20 years to obtain the information of their biological functions. Such information will accelerate discovery of genes susceptible to or causing various diseases and should contribute to screening of novel drugs that target these disease-gene products.

In this regard, analysis of expression profiles and SNPs (single nucleotide polymorphisms) using "microarray" or "DNA-chip" is quite important. "Microarray" or "DNA-chip" technology has made it possible to examine expression levels of thousands of genes and genotype a huge number of SNPs by a single experiment. We have been applying microarray analysis for screening of genes involving in colorectal, hepatocellular, and ovarian carcinogenesis as well as those related to responsiveness to anti-cancer drugs and those involving in various signal transduction pathways of medical importance. We have so far established a system to analyze 15,000 genes and are accumulating the expression profile data of various types of cancer cells. For example, we analyzed cancer tissues of 13 esophageal cancer patients who were treated by the same chemotherapy after their operation. Although all of them had an advanced cancer and could not have curative operation, four patients achieved a very long survival of 43-103 months, indicating that the chemotherapy was very efficient to these four patients. In contrast, four patients had very short survival of 4-12 months. A comparison of expression profiles of the patients with very short or very long survivals has disclosed that the expression levels of nearly 50 genes may be associated with responsiveness of the chemotherapy. This result implies that examination of expression levels of a set of genes may be a good predictor for a certain anti-cancer treatment. At present, a large number of cancer patients are treated with anti-cancer drugs without any knowing whether the drugs are effective to their cancer cells and a significant proportion of them suffered from side-effects without no effect. Hence, our approach must contribute to predict the effect before the patients start to have treatment with anti-cancer drugs.

Secondary, we have also been examining genes which are up-regulated or down-regulated in a certain biological condition by means of microarray coupled with laser-captured microdissection (LCM). A comparison of expression levels of normal mucosal cells, adenoma cells, and cancer cells from the same patients with colorectal tumor, we identified dozens of genes whose expression levels were significantly increased or decreased in tumor cells. The results clearly indicated that the microarray analysis is a very powerful tool to examine genes involving in carcinogenesis.

In addition to the expression profile analysis, the recent world-wide effort of the SNP (single nucleotide polymorphism) project that aims to discover 300,000 or more genetic variations in our genome will generate very valuable resources. We undertook a systematic survey of genomic DNA for SNPs located not only in coding sequences but also in non-coding regions (e.g introns and 5' flanking regions) of genes of medical interest. Using DNA samples from Japanese patients with rheumatoid arthritis (RA) or myocardial infraction (MI) as templates, we surveyed 82 genes that represent candidates for RA or MI, screening a total of 224 kb of DNA (107 kb of coding sequences and 117 kb of non-coding DNA). Within this 224-kb genomic sequences we identified 329 SNPs (1 per 680 bases on average), 50 insertions or deletions. Fifty-two percent of the coding SNPs were non-synonymous substitutions, and non-conservative amino acid changes were observed in a quarter of those. Allelic frequencies of some of the polymorphisms were significantly different from those reported in European populations. For example, the Q506R substitution in the coagulation factor V gene, the so-called "Leiden mutation", has a reported frequency of 2.3% in Europeans, but we detected the Leiden mutation in none of the Japanese genomes we investigated. The allelic frequencies of the -33A>G SNP in the thrombomodulin gene were also very different; this allele occurred at 12% frequency in the Japanese patients we examined, although it had been detected in none of 82 Caucasians reported previously. These data support the hypothesis that some SNPs are specific to particular ethnic groups.

In the meeting, I introduce the recent progress and future direction of human genome analysis and its impact on the medical science.