

ADVANCES IN DRUG DEVELOPMENT

Current Developments in Oncology Drug Research

Section Editor: Mark J. Ratain, MD

The BioBank Japan Project

Yusuke Nakamura, MD, PhD
Human Genome Center
Institute of Medical Science
The University of Tokyo
Tokyo, Japan

H&O How does the BioBank Japan Project operate?

YN The BioBank Japan Project started over 4 years ago with the goal of collecting DNA and serum samples, along with clinical information, from 300,000 patients. The project is organized by our group in the Institute of Medical Science at The University of Tokyo. The BioBank Project grew out of previous experience with the 5-year Japanese Millennium Project, as well as the 3-year International HapMap Project, to which our research group contributed nearly 25% of the genotype data. Included in these 300,000 patients are more than 50,000 cancer patients. Sixty-six hospitals affiliated with 12 institutes are participating in the BioBank Project, and we trained medical coordinators, mostly nurses or pharmacists, at these institutions in how to obtain consent from patients and how to record their clinical data and collect samples. Each day, 200 medical coordinators ask patients with any of 47 diseases, including major types of cancer, to provide blood for DNA and serum. Also, serum samples have been collected once a year to follow patients.

In addition to the BioBank Japan Project, we are constructing a large-scale expression profile database of more than 1,000 cancers. Also, we have been developing systems to predict the efficacy and safety of various anticancer drugs using a microarray system. We have been collecting tumor samples and analyzing the expression profiles using the cDNA microarray, extracting genes that are related to the efficacy of the different regimens.

H&O What are the goals of collecting these data?

YN The major goals of the BioBank Project are to identify the genes susceptible to various types of disease and to collect information on the drugs provided to patients in order to identify the genetic factors associated with Grade 3 or higher adverse reaction. It is difficult to measure drug response, but we think that genetic variations in each individual may affect the risk of adverse reactions. Of the more than 10,000 patients treated with various regimens of anticancer drugs whose data we have collected, several hundred have experienced Grade 3 or higher adverse reactions. We are screening the genotypes of those individuals and comparing these data with the severity of adverse reactions. Thus far, we have identified dozens of single nucleotide polymorphisms (SNPs) that may have very strong associations with various disease phenotypes or adverse reactions. However, these associations are not conclusive. Currently, we are in the process of validating these types of discoveries. Altogether, not including HapMap, we are presently working with 3 billion genotyping data points, combining the number of SNPs and individuals in the database. Using a combination of genomic information, expression profiles, and genetic variations like SNPs, it will be possible to predict which patients will respond and which will have a higher risk of adverse reaction to therapy. The overall goal of the BioBank Japan Project is to provide personalized dosing to patients such that only those patients likely to experience a positive response, with a very small risk of adverse reaction, receive a given therapy.

H&O How will the BioBank Project make these data available to clinicians?

YN Currently, we are discussing how to release the genotyping data. Due to the type of informed consent this project requires, it is very difficult to release the genotyping information from each individual. Instead, we are now summarizing many different kinds of data. We conduct a preliminary association analysis and select SNPs that show a *P* value of .01 or less. The number of positive SNPs in the first screening was expected to be between 4,000 and 6,000, based on our previous experience. Those SNPs are analyzed for a larger number of cases; we initially planned approximately 2,000 cases per SNP. Through additional SNP mapping of the disease loci, we will be able to narrow down the candidate genes to 10 or 20. Then we will release a summary of genotyping information at each locus for each disease. We are currently using the Illumina genotyping method, which is commercially available.

H&O Could you discuss the goal of point-of-care genotyping?

YN We are developing a genotyping device that would be potentially available at a hospital. Our group has worked together with Shimadzu and Toppan, two Japanese companies, to develop equipment that can genotype 32 different loci simultaneously, using one plastic plate, within 80 minutes. Our goal is to increase the number of loci and to be able to complete a genotype within 30 minutes. If we can genotype within 30 minutes, we can recommend a patient to have therapies based on the results of the genotyping (point-of-care treatment), using individual cassettes for each drug. With this system, our hope is that it will be possible to avoid severe adverse reactions.

Twenty prototype machines have been developed, and we have been verifying the quality of the data obtained by the new equipment. So far, using prototype machines, we have analyzed more than 100 individuals and compared the results to those examined with DNA sequencers. We have obtained 100% accurate genotyping with our prototype machines. It is our belief that these machines will have important clinical uses, and we are confident that if we can make the time shorter, these machines would be used widely at hospitals to provide useful information and discover potential adverse reactions.

H&O What are the next steps for the BioBank Japan Project?

YN As I mentioned, the goal is to collect data from 300,000 patients. At present, we have obtained data from 270,000 patients; within this year, we expect to reach the goal of 300,000. It is important then to follow these patients, and the project's follow-up will extend to the next 5 years. We will obtain yearly serum samples from the patients. If we can, for example, identify differences in protein or peptide levels in serum between a patient's good condition one year and bad condition the following year, then we can potentially point to predictive information that signals disease worsening. Similarly, we are trying to use serum samples for the development of different kinds of biomarkers. For example, if we follow 300,000 patients, we expect to have roughly 2,000 new cancer cases per year during the follow-up period. It will be possible to compare the serum samples and identify the differences between the samples. Such information should prove very useful. We are hoping to enlarge the scope of the project to include information not only for predicting adverse reactions to drugs but also to predicting prognosis.