

Naoyuki Kamatani, M.D., Ph.D.

Advisor, StaGen Co. Ltd

Professor and Director, Institute of Rheumatology, Tokyo Women's Medical University  
Professor, Division of Genomic Medicine, Department of Advanced Biomedical Engineering and Science,  
Tokyo Women's Medical University  
Group Director, Research Group for Medical Informatics, SNP Research Center (SRC), RIKEN

## 1. Statistics will become important

StaGen Co. LTD is the company focused on both genetics and statistics. It was established for the purpose of contributing to personalized medicine and pharmacogenomics (or pharmacogenetics) from the aspect of statistical genetics.

Pharmacogenomics and personalized medicine are the fields that will become very important in the near future. Since you may not understand the link between genetics and statistics, I will explain why.

Statistics will become important not only in the field of medicine but also in other fields as well. Japan is a special country among many modernized countries in that it is very weak in statistics. There are no departments or divisions with the name of statistics in Japan. In most of the modernized countries, large universities almost always have departments or divisions specialized in statistics. For example, the number of the members of American Society of Statistics is about 17,000 while that in Japan is about 1,500. We have to promote the education and the research in statistics in Japan. One of the reasons why statistics has not been regarded as important in Japan is that Japan has focused its target on secondary industry (making manufactured products) rather than tertiary industry (service). The direct target of the tertiary industry is humans although that of the secondary industry is products. For manufactured products, variations are to be avoided; however, variations are the essence of humans which can never be avoided. Variant products should be thrown away; however, human variations should be the center of the consideration in service industries. This is why Japan has not regarded statistics which deals with variations as important.

Statistics is becoming very important in medicine. Since 5-10 years ago, EBM (evidence-based medicine) has been proposed as the concept which is the most

important in medicine. EBM means the medicine based on the evidence; however, the center of the evidence is statistics. Since the education of statistics has been very weak in Japan, good drugs have become very difficult to be developed in Japan. Foreign countries including US and European countries have developed good drugs based on statistics.

As stated above, insufficient education in statistics has become a big obstacle to the industrial development in Japan especially in service industries in which humans are the direct targets. Statistics is often regarded as a very difficult science to learn. Because of the difficulty, it has been abandoned from the main parts in the educations in junior and senior high schools. However, the difficult and useful items in science can be the strongest tools for developed countries.

## 2. Why genetics and statistics?

Statistics is becoming important in the field of genetics.

The causes of most of human inherited diseases had not been clear until about 15 years ago; however the majority of them have been clarified by now. Major part of it is due to linkage analysis, a tool of statistical genetics. The problem in Japan is that, because of the lack of the ability in statistics, researchers can understand sequences but not the logics of the methods which link the sequences to the causes of the diseases.

Most of the genomic sequences of humans have been clarified. However, the sequences do not tell very much. For example, the person without the knowledge of English never understands the meaning of the sentences even if he (or she) understands ABC. However, in case of English, you will understand the meaning of the sentences if you ask them to native speakers.

However, the problem in genome sequences is that nobody will tell you the meaning of them. It is why the logics based on mathematics are necessary to understand the meanings from the sequences. These logics are based on statistical genetics, and algorithms are developed using such logics to clarify the meanings of the sequences. In other words, statistics (or statistical genetics) plays an important role in linking the genomic sequences to diseases or drug effects. The genomic sequences in individual subjects can now be read by machines; however, statistics is necessary to link the sequences to the diseases and drug adverse events or effects.

### 3. What does order-made (personalized) medicine mean?

I will explain the necessity of order-made (personalized) medicine.

The researches in US have clarified that the number of deaths due to drug adverse events exceeds the number of deaths due to traffic accidents in US. Needless to say, however, the lives of the majority of the patients are being saved by the drugs as well. It has also been clarified that many drugs are not effective to some of the patients. It is often observed that the effectiveness of a commercial drug either for cancer and rheumatoid patients is below 50%. The problem is that whether a drug is effective to a patient or whether the drug causes adverse reactions is not known before the patient takes the drug. If it is known, even in part, the drug should not be given to the patient to whom the drug is ineffective or who will develop major adverse reactions. If there are other drugs which are more suitable to the patient, he (or she) can take one of them.

A typical example of personalized medicine is blood transfusion. It is unthinkable that the doctors order blood transfusion without the test of blood types (ABO etc). It would cause a major problem if they give patients the suitable blood without testing for the blood types. The blood types are, of course, determined by the genomic sequence. It means that, in human blood transfusion, fundamental process of personalized medicine is being done; i.e. the treatments are different between the patients with different genomic sequences. Thus, in case of blood transfusion, the treatments are different between different patients; however, the treatments by drugs are not being done in such a way. Indeed, the drug treatment also has to differ between different patients if necessary just as the different treatments are necessary in case of blood transfusion.

Such an idea had been a dream until recently when a rapid progress of genetics was made in humans. For example, the blood concentrations of INH a drug for tuberculosis is shown to differ greatly between different subjects with different genomic sequences coding for an enzyme N-acetyltransferase. It is also known that the differences in the genomic sequence coding for pseudocholine esterase cause major differences in the effect of a muscle relaxant used for anesthesia and is related to the prolonged effect of the drug after surgical operations. You probably know about the differences in the effect of alcohol between different Japanese subjects. About half of the oriental people have variant gene at alcohol dehydrogenase 2 (ALDH2) locus and because of the possession, they are quite sensitive to alcohol. Since alcohol is a foreign compound, it is quite similar to a drug which is also a compound that is unfamiliar to the body. Considering

the case of alcohol, you probably understand the possible differences in the effects of many drugs between different individuals. It is known that people with a variance in thiopurine methyltransferase gene are very likely to develop severe adverse events by 6-mercaptopurine, a frequently used anti-cancer drug. It is also known that people who possess a variation in the genome at dihydropyrimidine dehydrogenase gene are quite likely to develop severe adverse events by 5-fluorouracil and similar anti-cancer agents. Examples in antirheumatic drugs are the association between genomic variations at MTHFR gene and adverse events by methotrexate and the association between genome variations at NAT2 gene and adverse events by sulfasalazine. Other cases include the drugs for asthma and hypercholesterolemia.

#### 4. Era of order-made (personalized) medicine

In US, personalized medicine is becoming a real story. In 2004, Food and Drug Administration (FDA) in US has publicized a draft guidance for the submission of the genomic data to FDA in new drug development. In 2005, Eli Lilly has commercialized a drug Strattera for the treatment of ADHD (attention-deficit/hyperactivity disorder). The information that the differences in genomic sequences in CYP2D6 gene cause different probabilities of adverse events was included in the label of Strattera. In accordance with the drug label, diagnostic kits for the polymorphisms in CYP2D6 gene are available. At least in US, the era of personalized medicine is becoming real. When people understand that the personalized medicine can reduce adverse events and decrease unnecessary drug administration, then Japan would not be able to refuse it.

There is one reason why a part of the pharmaceutical companies are reluctant to incorporate the idea of personalized medicine. They are afraid that the overall drug market may decrease if the patients who either do not get benefit or develop adverse events avoid taking medicine. On the other hand, good pharmaceutical companies understand that such an idea is against the public benefit and cannot be allowed in the society.

Efforts to decrease severe adverse events are an extremely big issue which should be taken by all the medical society. We have to construct the system in which good pharmaceutical companies that promote such efforts can get profit.

#### 5. New drug development based on pharmacogenomics

Pharmacogenomics (or pharmacogenetics) is the field of science in which effects of drugs are studied based on individual genomic sequences. Order-made medicine or personalized medicine is considered to be the ultimate goal of pharmacogenomics.

However, pharmacogenomics includes not only the therapies based on genome sequences but also the new drug development based on individual genome sequences. Recent use of the terminology pharmacogenomics is focused on the 'drug development based on individual genome sequences'. There is a reason why individual genome sequence data are used for the new drug development. Until several years ago, the idea of 'new drug development based on genome research' was advocated. Many people expected the occurrence of plenty of new good drugs based on such a wonderful idea. However, the fact is that the number of the new drugs has decreased. Although the seeds of new drugs have increased due to genome research; however, the process from seeds to commercialization is not efficiently progressed. This led to the development of the research on 'translational research' or 'critical path' whose target is the process from seeds to commercialization of drugs. Somewhere in the process, the drug development is ceased because of either 'lack of safety' or 'lack of usefulness'. Some drugs are withdrawn from the market because of the lack of safety even after commercialization. It is a major problem in medicine that the new drug development is so inefficient.

Pharmacogenomics is attracting attention because it may increase the efficacy of the drug development. For example, if there are differences in either efficacy or adverse events between subjects with different genome sequences, targeting the patients who are more suited to the drug may lead to the development of more safe and useful drugs. Furthermore, if the difference in the blood concentrations of a drug is evident between the subjects with different genome sequences, then we may propose the doses of the drugs based on the knowledge, or we may be able to decrease the sample size in subsequent clinical trials.

In any case, many of the pharmaceutical companies have begun banking of DNA along with clinical trials. Although it is not perfectly clear how the DNA collected will be used, new drug development based on genome sequences has already started. If pharmacogenomics can increase the efficacy of the new drug development and increase the safety during the clinical trials, then we should promote such science.

## 6. Our efforts

Several years ago, we expected that the era of personalized medicine and pharmacogenomics would come. Since then, we have discovered SNPs and haplotypes related to either adverse events or efficacy of drugs. We have made a large database necessary for the personalized medicine (PSC database). This database was made in collaboration with Pharma SNP Consortium (PSC; a consortium constructed by most of Japanese pharmaceutical companies), RIKEN (Prof. Yusuke Nakamura at Tokyo University), and Tokyo Women's Medical University (Prof Naoyuki Kamatani). The size

of data base is very big which is composed of SNPs at over 5,000 loci in 200 drug-related genes for 752 separate individuals. Using the database, haplotype blocks were determined and htSNPs were extracted. In addition, B cell lines were established from > 1,000 Japanese subjects, and were donated to a public cell bank (Foundation of Human Science). We have also developed algorithms to analyze the data from the database. The database has been put in a web site (<http://www.jpma.or.jp/psc/index.html>) and is being used by many researchers in the world. The haplotype data are not open to the public; however, they would be an important resource for the personalized medicine.

In addition, we have developed many algorithms for statistical genetics and published papers about them. They include POPSTRUCT, LDSUPPORT, PENHAPLO, QTLHAPLO, LDPOOLED and others that are widely used by researchers. We are proud to say that no other groups in Japan have equivalent career in this field.

The reason why we are ahead of other research groups is that we have through knowledge and technology both in genetics and statistics. Another reason is that we have sophisticated technology of computer programming. Both genetics and statistics are difficult fields of science to learn; however, because of this reason, the developed countries are eager to promote science in those fields to become ahead of other countries.

We plan to construct a further integrated system that can cope with the era of personalized medicine and pharmacogenomics which is expected to come in the near future.