

Omics-based Medicine and Systems Pathology

A New Perspective for Personalized and Predictive Medicine

H. Tanaka

Graduate School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan

Keywords

Omics-based medicine, systems pathology, genomics, omics, pharmacogenomics

Summary

Objectives: Recent important advances in the human genomics and post-genomic “omics” are now bringing about a new medical care which we call “omics-based medicine”. In this article, we investigated the development and future possibilities of omics-based medicine.

Methods: We divided the development of omics-based medicine into three generations in order to clarify the main clinical goals and characteristics of informatics method of each generation, together with its future possibilities.

Results: The first generation of omics-based medicine started with “genomic medicine” based on the inborn individual differences of genome. It has opened the study of genetic polymorphism of the diseases and promoted

the personalized medication based on the pharmacogenetic/pharmacogenomic difference of the drug response. In the second generation of omics-based medicine, owing to the advances in the high-throughput technology, vast amount of the various post-genomic disease omics data containing comprehensive molecular information of diseased somatic cells has become available. It reflects the ongoing state of diseases more closely and enables the predictive medicine such as prognosis prediction of disease by applying the data-driven analysis. Finally, due to the rapidly growing knowledge about the cellular molecular network, system-level understanding of the disease, called systems pathology, becomes possible. It can fully exploit the substantial contents of disease omics and will lead to a comprehensive understanding of disease process by using model-driven analysis.

Conclusion: Omics-based medicine and systems pathology will realize a new personalized and predictive medicine.

6], which aims to realize “personalized medical care”, based on the inborn (germline) individual differences, or “polymorphisms”, of the patient’s genomic information. Some personalized medical cares have already been put into clinical practice. For example, “personalized medication” is now in clinical use for several drugs; for which pre-prescription genotyping or protein assay is necessary. This pre-diagnostic test prevents undesirable side effects and ensures the effectiveness of those drugs to the individual patients.

Meanwhile, owing to the recent progresses in high-throughput technologies such as various types of DNA microarrays [7–11] or mass spectrometers (MS) [12, 13], *post-genomic omics information*, or simply “omics” data have become available in clinical context. A lot of studies have revealed that omics information provides more comprehensive and substantial information as to the ongoing process of diseases than just “genetic disposition” of them, so that it could be used in actual clinical practice such as early diagnoses, subtype classifications, and prognoses of diseases [14–16]. With this background, omics information observed in diseased state (“disease omics”) is expected to bring about a new kind of medical care, which would be more predictive or preventive than conventional genomic medicine. This new stage of molecular medicine needs a new term to distinguish itself from genomic medicine. We may call it simply “omics-based medicine” [17].

In use of disease omics for predictive or preventive medicine, omics data have been first analyzed on their direct level by applying data-mining or exploratory statistics (“data-driven analysis of omics data”). Taking a typical example, efficient sets of genes called “signature” have been deter-

Correspondence to:

Hiroshi Tanaka, MD, PhD, Professor
Dean, Graduate school of biomedical science Ph.D.
program
Tokyo Medical and Dental University Graduate School
1-5-45 Yushima Bunkyo-ku
Tokyo 113-8510
Japan
E-mail: tanaka@cim.tmd.ac.jp

Methods Inf Med 2010; 49: 173–185

doi: 10.3414/ME9307

prepublished: February 22, 2010

1. Introduction

Recent drastic advances in the human genomics and subsequent studies of post-genomic comprehensive molecular information, collectively called “omics” [1, 2] such as transcriptomics, proteomics, and metabolomics, are giving rise to new possibilities of medicine. In realizing this pos-

sibility, a rapidly progressing informatics, called “clinical bioinformatics” [3], or in a more recent term, “translational informatics” [4] is playing an indispensable role by deriving clinically meaningful information from the vast amount of omics data.

To date, application of comprehensive molecular information to medicine has been referred to as “genomic medicine” [5,

mined by applying data-mining or exploratory statistics to gene expression profiles of diseased cells in order to make a prediction of disease prognosis such as the recurrence of cancers within several years after the surgery [14, 16]. Another example is statistical pattern analysis of the proteomic data. Extraction of the characteristic patterns from the protein mass spectra by using statistical feature analysis is conducted for an early detection of cancer [18, 19]. These attempts have in some cases gained remarkable successes, but in most cases the results were not sufficient for further clinical application. It becomes soon recognized that vastness of omics data with its extremely high inter-dependency often prevents data-driven approaches from obtaining decisive results.

On account of the rapidly growing knowledge about the cellular molecular network and its alternations in diseases, it becomes clear that, except for rare genetic diseases, most of the diseases are not caused by mutation of one or two genes, but rather caused by the cooperative effects of aberrations of several genes and proteins. These cooperative effects result in alternations in signaling pathways or gene regulatory networks, producing the disease phenotypes. Hence diseases would be better understood as a phenotype caused by “systems distortion of the molecular network” due to the interrelated malfunction of genes and proteins. This view of diseases is particularly true for common complex diseases, like ordinary cancers which are now considered as “pathway disease” [20]. Thus, system-level understanding of diseases based on the knowledge of alternated molecular pathway is now thought to be crucial to fully exploit the substantial contents of disease omics and to comprehensively understand what is going on underlying them (“model-driven analysis of omics data”).

We have been proposing this kind of systems approach to diseases as “systems pathology” for several years [21, 22] in the sense that it is a proper application of systems biology [23–25] to diseases. This stage of omics-based medicine, namely that based on the systems-pathological understanding of diseased pathway, could be called “*omics-based systems medicine*” or we may call it simply “*systems medicine*” like

other advocates [26] if the aimed meaning of this concept is correctly understood.

Anyhow, we would like to emphasize that this new type of medicine, based on the vast amounts of omics data and system-level understanding of the whole disease process, would not have been possible and will not be progressed in the future without cooperation of, so to speak, “(post-genomic) omics informatics” or “systems pathology”.

In this article, we describe the essential features and development of the omics-based medicine along with those of informatics used for it as we depicted briefly above. Needless to say, genomics is, ordinarily, included in the omics. Hence, in describing the development of the omics-based medicine, we start with a concise review of genomic medicine as its first generation, and then we proceed to the post-genomic omics-based medicine as the second and current generation, and finally describe omics-based systems medicine as the third generation with its future possibilities. We think this “three-generation paradigm” would be quite suitable for discussing the characteristics and main clinical goals of each generation, together with the generation-specific informatics used to attain these goals.

2. Genomic Medicine as the First Generation of Omics-based Medicine – Concise Review

2.1 Genomic Medicine for Personalized Care

The molecular medicine has, as well known, begun with “genomic medicine”, an advocacy of a new medicine which aims to realize personalized care based on the individual differences of genomic information. As the human genome project [27] proceeded, it became clear that human genome has individual congenital differences. There are various types of polymorphisms in human genome, such as variable number of tandem repeat (VNTR) of microsatellites (repeat unit is 1 to several bps) or minisatellites (9 to tens bps), in-

sertion and deletion (indel) polymorphism, copy number variation (CNV), and so on [28]. But now most widely used polymorphisms are single nucleotide polymorphisms (SNPs), which are observed, on average, every 300–1000 nucleotides in human genome (about 0.1–0.3%); thus totally 3–10 million SNPs exist for an individual human genome.

From the clinical viewpoint, there are mainly two important kinds of clinical phenotypes, brought by the above genomic polymorphisms:

1. Disease susceptibility (causality), and
2. Drug responsiveness.

The former means the genetic risk of disease occurrence, whereas the latter means personalized difference in response to drugs. Since there have been many excellent reviews such as [5, 6] explaining “genomic medicine”, we will not describe it here in detail, but introduce it briefly with emphasis on its difference from the subsequent (second or third) generations of the omics-based medicine, which would be more central subject of this article.

2.2 Genomic Medicine and Disease Gene Study

Genomic study about human diseases substantially started in 1980s, when both genetic maps of whole genome (though they might be coarse) and sufficient number of DNA markers (randomly spread over the genome) became available by using, first, RFLP (restriction fragment length polymorphism) [28] and then, microsatellites. Genetic maps and DNA markers provide a framework for determining the loci of disease-causative genes in the human genome (“positional cloning”) [29]. Together with advances in the statistical genetics [30] such as Mendelian linkage analysis based on family trees, several genetic studies succeeded in positional cloning of disease-causative genes like those of Huntington disease [29, 31], Duchenne muscular dystrophy [32] or cystic fibrosis [33]. As a matter of fact, such successes in identification of disease-causative genes provided motivation to begin the human genome project.

Although early genetic studies have succeeded in identifying various causative genes of monogenic diseases, their prevalence is relatively low even altogether. Clinically more important is to clarify the genetic background of the polygenic or multifactorial diseases like common complex diseases such as asthma, cancer, diabetes, heart disease and mental disorders. But unlike the monogenic diseases, we sometimes observe several hundreds of “disease susceptibility genes” with low genotype relative risk (GRR) [34], which means the risk associated with a certain genotype relative to that with the other genotype (non-carrier), for the polygenic diseases.

A certain SNP may be linked to a disease susceptibility gene through the mechanism of “linkage disequilibrium” [35], which brings about the high co-occurrence between that SNP and the disease susceptibility gene. In such a case, this SNP can be used as a marker which, if it is found, provides the increased genotype-relative risk of that disease. But it is generally seldom to find markers for susceptibility genes with relatively high GRR, especially for common complex diseases.

To overcome these obstacles, a new type of a large-scale trial called genome-wide association studies (GWA studies or GWAS) [36, 37] has now been undertaken widely. GWA studies involve rapidly scanning hundred thousands of markers across genomes of many people to find genetic variations associated with a particular disease. There have been many large-scaled GWA studies conducted, among which, for example, studies on the diabetes are well known because the different GWA studies discovered a new common susceptibility gene TCF7L2 [37–39]. GWA studies become possible due to a new kind of DNA array which covers more than 500,000 SNPs in one chip. The large size of the population and comprehensiveness of the SNP probes on the chip [40] provide GWAS with a higher detection power of disease susceptibility genes.

Another approach to gain higher disease susceptibility for polygenic diseases is to use the combined pattern of multiple SNPs, called “SNP haplotype”. The term “haplotype”, shortening of “haploid geno-

type”, is defined as a set of alleles of closely linked loci on the same chromosome that tends to be inherited together. Likewise, a SNP haplotype is defined as a particular pattern of sequential SNPs on one chromosome which are statistically associated. In order to obtain the efficient estimation of disease risk for the polygenic diseases, single SNP is not sufficient. By using SNP haplotype the more exact relative risk of the disease can be gained by the SNP haplotype-disease association study. This kind of information is now being energetically collected by the international Hapmap project [41].

The disease gene study to determine mutations or polymorphisms of disease-related gene (causative or susceptibility gene) reveals the detailed individual difference in the pathogenesis of diseases. Hence, it would contribute to “personalized medicine” in predicting a future possibility of the disease occurrence for specific individual. However, this field of genomic medicine now does not seem to make an immediately effective contribution to clinical practice, seeing that even “high GRR” of the susceptibility gene for common diseases is ordinarily at most about 1.5, though various attempts like GWAS or SNP haplotype analysis are explored. It would be better considered that current disease gene studies contribute to the scientific clarification of genetic background of diseases rather than immediate applications to actual clinical practice.

It makes a quite contrast to another field of genomic medicine, “personalized medication”, where genomic polymorphisms of drug-related genes sometimes cause remarkable differences in the response to the drug. The reason would be that drug has appeared relatively recently in the human evolutionary history, so that genetic mutations or polymorphisms of drug-related genes causing fatal response to drugs still remain since there is not a sufficient time for human evolution to eliminate them. In contrast, genetic mutations and polymorphisms of disease genes which cause fatal effects have been considerably eliminated through the long term in human evolution [42].

2.3 Personalized Medication Based on the Genetic Polymorphism

On the contrary to the slow progress of personalized care based on disease gene study, personalized medication is an immediately effective and rapidly advancing field of the genomic medicine; tailor-made medication of some drugs is now put into clinical practice.

Certain genetic polymorphisms may be involved in the functions related to the drug effects, causing inter-individual differences in response to the drugs, and giving rise to the classification of “responder” and “non-responder”. The branch of pharmacology studying influences of genomic variation on drug response in patients is now called pharmacogenomics or pharmacogenetics (PGx, abbreviation for both). The concept of pharmacogenetics has been used more than 40 years to denote the study of single genes and their effects on inter-individual differences in drug metabolizing enzymes. In the contrary, pharmacogenomics is relatively a new concept after human genome project to denote the study of not just single genes but the functions and interactions of all genes in the genome in the overall variability of drug response [43]. But the distinction between the two terms is considered rather arbitrary, however, and now the two terms can be used interchangeably.

As mentioned above, PGx is one of the most successful fields among “genomic medicine”. *Journal of the American Medical Association (JAMA)* described in 1998 that about two million patients per year suffered from drug side-effects and among them, about 100,000 patients died, which amount to the 4th or 5th major cause of death in the United States [44]. Food and Drug Administration (FDA) in US [45] promotes to use the pre-diagnostic DNA tests to examine the drug responsiveness which are compulsory for some drugs to avoid useless side-effects.

Individual differences of drug response are mainly ascribed to two classes of genetic polymorphisms. One is called pharmacodynamics (PD) polymorphism and is due to the polymorphisms of genes for proteins related to drug action such as a drug recep-

tor or a protein kinase. The other is called pharmacokinetics (PK) polymorphism and is due to the polymorphisms of genes for proteins responsible for the absorption, distribution, metabolism and excretion (ADME) process of the drug, such as drug metabolizing enzymes or transporters.

The PD polymorphism of drug response is often found in the molecular target drugs of several recent anticancer agents. The noteworthy example is Gefitinib (Iressa), which is the anticancer drug to lung cancer (epidermal growth factor receptor-tyrosine kinase inhibitor; EGFR-TKI), especially effective against advanced non-small cell lung carcinoma [46]. This drug has very remarkable effect on a “super-responder” who dramatically improves with administration of Gefitinib even after the lung carcinoma has spread all over the one lung [47]. This drug is effective especially for adenocarcinomas of Asian female at the rate of 25–30%, but only 3% effective for US male and 17% for US female. It, however, may cause “serious side effects” contracting interstitial lung disorder for 2% of Japanese, and one third dies from this disorder. Hence, to avert the dangerous risk of medication to the non-responder, genetic screening is inevitable. The difference of response of this drug is now ascribed to the specific mutation in EGF receptor and it is reported that drug effect becomes 10 times greater for the patient with this mutation [48].

Another well-known example is Trastuzumab (Herceptin), an anticancer drug for breast cancer using a monoclonal antibody. This drug is effective for the patients of breast cancer who over-express HER2 (human epidermal growth factor type 2) receptor [49, 50]. FDA has decided that pre-diagnostic DNA test using immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH) test is compulsory for administration of this drug.

The PK polymorphism diversifies the patient's individual response of certain drugs, ranging from “extensive metabolizer” whose metabolizing enzyme eliminates the drug too fast to make it less effective, to “poor metabolizer” who metabolizes the drug poorly so that the drug remains too long within the patient to produce the excessive effect. The most well-

known polymorphism of drug metabolizing enzyme widely related to the drug effects is the family of cytochrome P450 (CYP). To date, more than 50 isoenzymes of CYP in human are known. The classification of this enzyme is based on the degree of homology of the genetic sequences. The drug effects of many important drugs such as anticancer agents like 5-fluorouracil (5-FU), Irinotecan, or anti-tuberculosis agent like Isoniazid and so forth, depend on the polymorphism of CYP. There are many excellent reviews on the polymorphism of CYP and its effect on drug metabolism. We will not describe it here in general but briefly introduce several aspects of the personalized medication, taking Warfarin as an example.

Warfarin is the most widely used anticoagulation drug prescribed to inhibit the synthesis of clotting factors, thus preventing heart attacks, strokes, and blood clotting. It is also known for its narrow therapeutic range; excess dose causes bleeding whereas insufficient dose causes thrombosis. Furthermore, this narrow therapeutic range widely varies among individuals, with more than 10 times variation: Hence, the maintenance dose of this drug has to be monitored carefully by frequent blood testing to ensure the adequate yet safe dose is taken. Warfarin inhibits the vitamin K epoxide reductase complex subunit 1 (VKORC1) from synthesizing vitamin K necessary for the blood coagulation, thus exhibiting the anticoagulation action. As to the polymorphism for the drug effect of Warfarin are CYP2C9 (*1, wild type and *3, low activity type) for PK polymorphism and VKORC1 (-1639G > A, G allele, high transcript activity) for that of PD. Recent study reported the patient's initial response is determined largely by VKORC1 genotype [51]. As mentioned before, FDA guidance recommended pre-diagnostic DNA tests; some are compulsory and others are either recommended or referential, to examine the drug responsiveness. As for Warfarin, the pre-diagnostic DNA test is recommended.

Tailor-made medication of drugs is thought to be prevailing along with the advances of the pharmacogenomics and pharmacogenomics [52].

3. Advantages and Possibilities of Medicine Based on Post-genomic Omics – the Second Generation of Omics-based Medicine

As early as the midst of 1990s, while the human genome project (HGP) was still in progress, post-genomic omics such as the transcriptomics, or proteomics had become available due to the rapid advances in high-throughput technologies. Although each of the post-genomic omics is originally derived from the same genome sequence, it possesses own original information, through various *post-genomic* (post-transcriptional, or post-translational) processes such as alternative splicing, post-translational modification or epigenetic modification. Now post-genomic omics has been developing prominently; some of its applications have been approved by FDA for the clinical use.

There is a remarkable difference between genomics and post-genomic omics in relation to clinical medicine. The genomic medicine which utilizes congenital polymorphism of the “germ-line” genome sequence that is identical all over the tissues and remains same during the whole life. In contrast, most of post-genomic omics is related to comprehensive molecular information of “diseased somatic cell”, which might vary during the time course of disease and also differ among the tissues, so that omics is thought to much more reflect the current state of the disease, providing more direct information than genomic information.

Thus it is expected that, if we make the best use of these omics data in clinical medicine, it would innovate on the conventional medicine, opening a new stage of medical care, which we could call “(post-genomic) omics-based medicine”. In this section, we will describe what kind of new clinical possibilities omics information actually brings about to medical care in addition to the conventional genomic medicine.

3.1 Omics Provides Information for Detailed Mechanism of Disease

3.1.1 Subtyping and Prognosis Prediction of Disease

Post-genomic omics information provides detailed information to reveal the structure of molecular process of the diseased cells. Omics data of diseased cells provides clinically and pathologically unobservable information which would be utilized for detailed classification of disease. The best example is subtyping of diseases based on the gene-expression profiling (transcriptomics) observed by DNA microarrays. The study by Golub et al. [14] is the first attempt for this; they, taking an example of human acute leukemia, demonstrated the feasibility of cancer sub-classification based solely on gene expression pattern without referring to any previous medical knowledge. Alizadeh et al. [15] investigated gene expression patterns of diffuse large B-cell lymphoma (DLBCL), and found the two subtypes which could not be found by conventional clinicopathological observation and revealed that each of them showed quite different prognoses.

Gene expression patterns do not just provide the microscopic information about the molecular activities of diseased cells, but they also show the tight relation to the most macroscopic characterization of diseases, namely, prognosis of a disease, which can not be correctly predicted solely based on the “mesoscopic” observations such as clinical and pathological findings. Rosenwald et al. [16] used gene expression profiling of DLBCL to identify three subgroups in addition to the already known two subgroups [15] and, based on such subgrouping, molecular predictor of patient’s survival after chemotherapy was formulated, which was found to be independent from previous clinicopathological prognosis index. Likewise, there have been many studies about the predictability of disease prognosis based on the transcriptomics [53]. Sørli et al. [54] combined several population studies conducted independently about the gene expression profiling of breast cancer to create a classification of patient subgroups having distinct prognosis. Conventional pathological

tests of diseased tissue and clinical observation were unable to give an exact prediction of prognosis. In recent years, two gene-expression-based prognostic tests (“MammaPrint” [55] and “OncotypeDX” [56]) to predict prognoses of patients of breast cancer after the surgery were approved by FDA and are now commercially available.

3.1.2 Response to the Anticancer Drug

Gene expression profiles can also be used for prediction of the response to the anticancer drug. Cheng et al. [57] used biopsy samples from primary breast tumors before treatment in order to predict the response to treatment by docetaxel based on the gene expression profiles of the samples. It was concluded that the molecular profiles could be used to develop a clinical test for docetaxel sensitivity. Ayers et al. [58] examined the feasibility of developing the multigene predictor of pathologic complete response (pCR) to neoadjuvant chemotherapy combining fluorouracil, doxorubicin, cyclophosphamide and paclitaxel for breast cancer. Based on the gene expression profiling of the tumor, overall 78% predictive accuracy was obtained. Differences in the response of anticancer drug are ascribed not to the congenital polymorphism of the drug-metabolizing enzyme described in the previous section of genomic medicine, but to the characteristics of the disease pattern, or more precisely, the underlying diseased molecular pathway of the cancer, which exhibits itself in its gene expression profile.

3.2 Diagnostic Ability of Proteomics

Since omics information is more directly related to the state of diseased cells, it can also be used for an early detection of diseases. As for the cancer, there are many well-established tumor markers but they are useful in relatively later stage only to confirm that cancer is already established. On the contrary to the conventional tumor marker, disease proteomics measured by mass spectrometry could be utilized for an early detection of cancer. Many studies

have revealed that not just a specific mass spectral peak but a whole mass spectral pattern contains the information to detect cancer in very early stage, which could be extracted by statistical or data mining method. For example, it is reported that ovarian tumors can be discovered in very early stage with the precision of 99%, by proteomic examination of patients’ serum using SELDI-TOF-MS [18].

We conducted a study to compare the discrimination power between the proteomic information and conventional tumor markers with respect to detection of hepatocellular carcinoma (HCC) [59]. It is well known that patients suffering from liver cirrhosis develop HCC. We collected serum specimens from liver cirrhosis patients (LC: 20 subjects), early HCC (E-HCC: 20 subjects) and advanced HCC (A-HCC: 20 subjects). For proteomic measurement, SELDI-TOF-MS was used to analyze the patients’ serum, and for the tumor marker, AFP and PIVKA-II were measured. Three statistical components were extracted from mass spectral pattern by using the partial least square method.

Classification results (▶ Table 1) showed that the advanced HCC could be diagnosed at least as HCC by all of the three models. On the other hand, the conventional tumor markers had poor performance in detecting early HCC compared to peak intensities obtained by SELDI-TOF-MS. Furthermore, the highest sensitivity (0.875) for detection of HCC from LC was achieved by combining proteomic data and tumor markers. Thus, proteomic examination is complementary to the well-established tumor markers and could be used together for the development of more accurate early detection method for LC patients.

3.3 Omics-based Medicine Brings about Personalized and Predictive Medicine

Based on the various clinical competence of disease omics mentioned in the above sections, we can describe the essential features of the post-genomic omics-based medicine more decisively. As we mentioned at the beginning of this section, post-genomic disease omics could provide the comprehen-

a) Prediction model using peak intensities (sensitivity: 0.850, specificity: 0.550 for HCC diagnosis)			
Prediction	Diagnosis		
	LC	E-HCC	A-HCC
LC	11	3	3
E-HCC	8	12	10
A-HCC	1	5	7
b) Prediction model using tumor markers (sensitivity: 0.800, specificity: 0.800 for HCC diagnosis)			
Prediction	Diagnosis		
	LC	E-HCC	A-HCC
LC	16	8	0
E-HCC	4	8	2
A-HCC	0	4	18
c) Prediction model using both peak intensities and tumor markers (sensitivity: 0.875, specificity: 0.700 for HCC diagnosis)			
Prediction	Diagnosis		
	LC	E-HCC	A-HCC
LC	14	4	1
E-HCC	6	14	4
A-HCC	0	2	15

sive molecular information of “diseased somatic cell”. Hence, it would reflect more directly the current state of disease, varying in the course of disease. In contrast, the “germ-line” genome sequences in genomic medicine which remain the same during the whole life.

Omics information is considered to lie at the intermediate level of the hierarchy from the genotype of the genome to the clinical phenotype, so that the relation to the clinical phenotype is much closer than genomic information (► Fig 1). This “intermediateness” feature of omics data (close to clinical phenotype and also to the ongoing process of disease) provides its essential clinical competence. On the contrary to the genomic medicine which suggests “only” the possibility of occurrence of a disease in the future, the (post-genomic) omics-based medicine contains more direct disease-related information from diseased cells and is able to predict “when disease will occur” and “how disease will develop” in more exact way.

Another feature of the omics data other than its “intermediateness” is its comprehensiveness or its totality, because omics data originally comes from the genome sequence so that it is essentially “genome-wide”. Most of the current measurement technologies like gene expression microarray or SNP chip cover the whole genome, though some sort of omics like proteomics are currently lacking in some part (very large protein molecule information), they could practically be used to extract the alternation of its overall pattern by disease. The “comprehensiveness” feature of the omics data guarantees the validity of the predictions of diseases which are made based on the disease omics.

The genomic medicine is thought to bring about the personalized medicine, whereas the second generation of omics-based medicine brings about the “predictive medicine”. Disease omics provides detailed information to identify the subtype or fine molecular characterization of disease, which enables the more exact predic-

Table 1

Classification accuracy estimated by leave-one-out cross-validation using a) peaks-only, b) markers-only, and c) peaks-and-markers models for diagnosis of hepatocellular carcinoma

tion of disease or early detection of disease. Of course, when using the term of “omics”, we also include genomic information, so that the omics-based medicine, if we do not restrict it to post-genomic omics, includes the advantages of the first generation. It might be better to say that omics-based medicine in its broad sense (including the genomic medicine) is expected to bring about “personalized and predictive medicine”. Our characterization of omics-based medicine has some common features with 4P medicine (personalized, preventive, predictive and participatory medicine) which L. Hood stated concerning his concept of systems medicine [26].

Furthermore, recent biotechnological advances have revolutionarily changed the situations around the omics-based medicine, where the next generation sequencer enables a genome to be sequenced within hours to days [60]. The sequencing speed which was two million base pairs (bps) per day at the end of the human genome project, in 2003, is now over one billion bps per day. Next generation sequencers are not only for DNA sequencing of genomes but also it could be used for other kinds of sequencing such as mRNA-sequencing (mRNA-seq) or ChIP-sequencing [61]. The mRNA-seq can be used for direct digital quantification of the sequences instead of using conventional hybridization technique. Hence, it could avoid the hybridization errors and will allow us to obtain quantification data in higher precision. We could call these digitally sequenced molecular information “digital omics”.

Along with the biotechnological advances, cost of sequencing also becomes rapidly reduced, expecting the thousand dollar genome, which would mean the arrival of “personal genome” era.

3.4 Omics Informatics in the Second Generation

Typical computational methods used in this generation could be called “data-driven approach”. Vast amounts of omics data are analyzed, with or without reference to external criteria (mostly, clinicopathological characteristics and patient outcomes). To do this, statistical or data-mining (machine

learning) methods such as Cox regression model or support vector machine are used to extract the gene set or proteomic pattern which could provide clinically useful prediction, early detection and sub-classification of disease.

But mostly the omics data are ill-conditioned where number of variables (e.g. number of genes in DNA microarray) by far exceeds that of the samples (e.g. number of patient cases). Hence, for example, quite different sets of genes were proposed best to predict a prognosis of the same cancer by different research groups. In this generation, there are several typical studies which gained successful analysis of omics data but in most cases vastness of omics data with its extremely high inter-correlation often prevent obtaining decisive results if using only the direct level analysis approach. This disadvantage of the data-driven approach to the omics data should be noticed, so it is widely recognized that any kind of knowledge about the disease process which omics data is utilized to analyze must be incorporated to complement the indecisiveness of the data-driven analysis of omics data. This trend in some aspect led to the concept of systems pathology of third generation of omics-based medicine and inclined to use the “model-driven approach” for the analysis of omics data.

In concluding the informatics of the second generation, we would suggest a promising approach in the data-driven analysis. In order to make the best use of the omics information, a new disease-oriented bioinformatics or translational informatics should be developed to deal “multi-omics” data, which could analyze not only each of omics data but also the interrelation between the different modalities of omics data. Since the different modality of omics has its own omics information, to use multi-modal omics data will give us new information to compensate the indecisiveness of the results obtained by single modal omics [62].

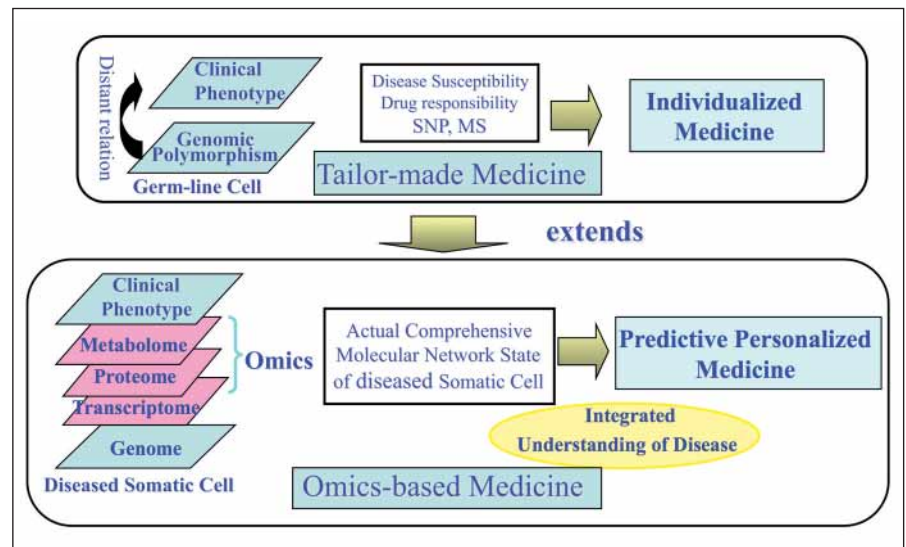


Fig. 1 Concept of omics-based medicine: Post-genomic disease omics provides comprehensive molecular information of “diseased somatic cell”. Hence, it varies during time course of diseases and differs among sites of diseases, unlike the “germline” genome sequences in tailored medicine which remain the same during the whole life. Omics information lies at the intermediate level of the hierarchy so that the relation to the clinical phenotype is much closer to bring about the predictive medicine.

4. Systems Pathology and Omics-based Systems Medicine – The Third Generation of Omics-based Medicine

4.1 Systems Pathology Comprehends the Whole Process of Diseases

As with the background leading to the emergence of systems biology in life science, the same situation has come up in omics-based medicine where systems approach is crucially needed to fully understand the whole process of disease. There is a situation for necessitating the systems approach to diseases. The “molecule-oriented” approach where the causes of diseases are sought for specific molecules such as a disease-causative gene or a mutant protein, which has been successful in the monogenic diseases studies, is not effective for polygenic multifactorial diseases like common complex disease. As we already mentioned, rapid advances in the knowledge about cellular molecular network and its alternations in diseases have made it clear that, except for rare genetic diseases,

most of diseases are substantially caused and supported by the “sustainably distorted” molecular networks of signaling pathways or gene regulatory networks. Although this distortion is induced by the cooperative effects of aberrations of genes and proteins, the distorted pathway determines the disease state more directly. Thus, system-level understanding of diseases based on the knowledge of diseased molecular pathway is now thought to be indispensable to understand the full information contents of disease omics and to comprehend the whole process of disease (“model-driven analysis of omics data”).

We have been proposing a new concept of “systems pathology” for denoting the counterpart in omics-based medicine to systems biology in life science. “Comprehensiveness” feature of omics data makes it possible to understand a disease as an integrated whole, underlying the diseased omics data. In developing the systems pathology, it would be appropriate to describe in what sense diseases should be dealt with as a unified system. We will describe several systems features of diseases, which make disease systems approach appropriate. The following features are mostly seen in common complex diseases. We will exclude the

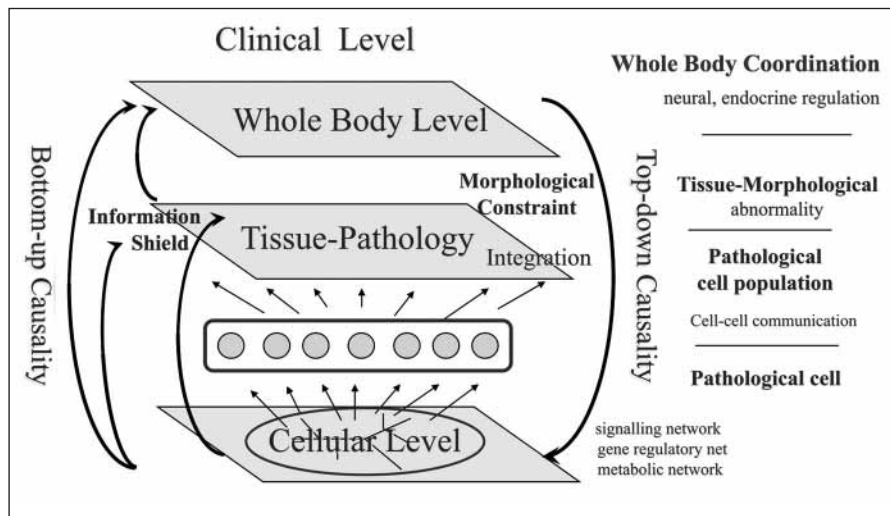


Fig. 2 Hierarchical organization of diseases and self-sustainability due to the bidirectional causation loop: Diseases are an integrated multi-hierarchical network system, comprising subcellular molecular network, cell-cell communication, tissue/organ linkage and whole body coordination. Diseases such as common complex diseases have a self-sustaining mechanism due to bidirectional causative loop.

rare genetic disease in the description of the features.

4.2 Systems Feature of Disease

4.2.1 Distorted Molecular Pathways as a Primary Cause of Disease

Owing to advances in the studies on the molecular networks, it becomes widely known that genes and proteins work together to form the molecular pathways, which accomplish the coherent and complex biological function. Therefore, it would be considered that, except for the rare genetic diseases, primary causes of most of diseases are not the direct effect of dysfunction of *gene* but the integrated effects of alternated or distorted *molecular pathways*, which are brought about by the cooperative effects of aberrations of several genes and proteins. In other words, not gene but pathway is a primary entity for disease.

Taking an example of cancer, ordinary cancer occurs by alternations of signaling pathway or gene regulatory network caused by aberrations of several (not just one or two) oncogenes or tumor suppressor genes, eventually leading to the uncontrolled cell proliferation. It should be noted that tu-

morigenesis is a multi-step process, where several essential alternations in tumorigenic pathways should participate.

By the experimental studies [63], introduction of simian virus 40 early region (SV40 ER), human telomere reverse transcriptase (hTERT), and H-RAS into the normal cultured cell defined a set of genetic changes that are sufficient to program the tumorigenic phenotype in most human cultured cells. The SV40 ER encodes both the SV40 large T antigen (LT) and small t antigen (ST), where LT inactivates the retinoblastoma protein (pRB) and p53 tumor suppressor pathways and ST suppresses the activity of protein phosphatase 2A (PP2A) which leads to the activation of MYC, a transcriptional factor to induce cell proliferation.

Hence it would be suggested that there is a minimum set of alternations in the molecular pathways such as immortalization (hTERT), cell cycle (pRB), apoptosis (p53), cell growth (MYC), angiogenesis, migration and adhesion (H-RAS) for causation of carcinoma. There are also many varieties of equivalent minimum gene sets to cause tumor, but it should be recognized that alternations in several crucial pathways such as those for cell growth, mitosis and invasion together with suppression of apoptosis should be necessary for activating tumorigenesis. Hence, we could consider

ordinary cancer is caused by simultaneous occurrence of aberrations of these pathways.

This is not only true for cancer but also applies to other diseases, especially common diseases, where hundreds of genes are related to disease causation like hypertension or diabetes. It is unlikely that each of these many genes directly influences the causation of disease but several genes might work together to form the alternation of certain pathway and joint effects of aberrations of these pathways cause and sustain the disease.

4.2.2 Hierarchical Organization and Self-sustainability of Disease

Aberrations of the molecular network are not restricted to those of subcellular network such as intracellular signaling pathway or gene-regulatory network. Cell-cell communication or intercellular network which is indispensable for coordination of the collective behavior of cells or regulation of tissue-level function is also the “seat” of diseases. For example, intercellular networks of cytokines such as interleukins (IL) and TNF- α cooperatively work in the immune system for host defense, but over-expression of these genes causes, for example, chronic inflammatory proliferative disease (CIPD) [64] such as rheumatic arthritis and psoriasis. Furthermore, larger scaled physiological control systems such as circulatory, nervous or endocrine system are also intrinsically related to disease to form systemic level clinical phenotype.

Hence, it would be appropriate to consider that diseases are an integrated multi-hierarchical network system, comprising subcellular molecular network, cell-cell communication, tissue/organ linkage and systemic coordination (► Fig. 2). There are also inter-hierarchical interactions. Bottom-up causality and top-down causality are acting together to make disease an integrated whole. The bottom-up causality could be paraphrased as “aggregative or integrative causation directed towards higher levels of tissues and bio-objects”. In contrast, top-down causality also could be paraphrased as “dis-aggregative or distributive causation across tissues which derive from the higher-order whole body, sys-

tem, or organ". Finally, environmental (lifestyle-related) factors also influence the disease process, especially in common diseases. Thus, diseases integrate such a multi-hierarchical distorted network and eventually manifest it at the level of systemic clinical phenotypes.

The common complex diseases such as hypertension or diabetes are the lifelong disease lasting 20–30 years or more. For the diseased state to be continued for such a long time there must be self-sustaining mechanism comprised of some kind of a feedback loop.

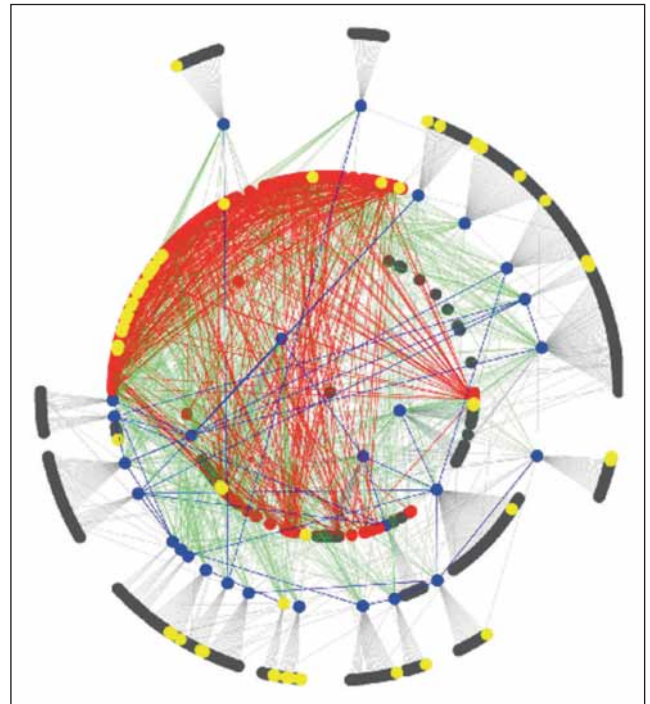
As an example for the self-sustaining mechanism in a hierarchically organized system of a disease, we take hypertension, where, other than the "bottom-up causality" which propagates from the molecular level to the systemic level, the whole body state of the disease drives gene expressions to maintain the systemic state of hypertension. In hypertension, if the blood pressure rises due to certain causes such as lifestyle-related factor or neural-humoral factors and if, even after several months, hypertension is continued, then the genetic activity of endothelial cell of capillary begins to start the remodeling process of capillary vessel to make them stiffer in order to sustain high blood pressure [65]. This could be a self-sustaining mechanism of hypertension. This phenomenon also shows an example where inter-hierarchical bidirectional activity is working to sustain the overall diseased state. This is some kind of "take over" of innate homeostatic morphogenetic function, used for maintaining the "diseased state".

4.2.3 Disease as a Dynamical System

So far we have discussed on "organizational aspects" of disease systems such as its hierarchical organization and self-sustainability. But an equally important systems feature of diseases is that it develops and changes itself in the course of time based on its intrinsic "dynamics". That is, diseases are dynamical systems. Take infectious diseases by retrovirus such as AIDS for example, temporal course of disease after its onset is considered to be a coevolutionary process between the host immune system and

Fig. 3

Cloud topology of protein interaction network (PPI): PPI comprises three layers: hub (high-degree nodes), middle layer (inner ring) and branch (low-degree nodes). Hub and branch nodes are depicted outer ring, and middle-degree nodes are depicted inner ring. Drug target proteins are depicted by yellow spots, which are concentrated on middle- and low-degree nodes layer (reprinted from PLoS Computational Biology [85]).



human immunodeficiency virus (HIV). HIV evolves within the host (patient) a million times faster than the evolution of normal organisms, and it evolves under the selection pressures of neutralization by host immune system and anti-viral agents. Many methods to describe the within-host viral evolutionary process have been developed such as longitudinal phylogenetic tree [66, 67], including our method based on quasi-species grouping of HIV [68]. These studies might lead to characterization of intrinsic dynamics common to retroviral infectious diseases.

Dynamics of cancer progression also could be modeled from the viewpoint of Darwinian evolution of cancer cells. In the recent cancer stem cell theory, mutations have already taken place in the stem cell stage, and it replicated through the developmental process; from stem cell, progenitor to matured cell. Cancer stem cells which subsequently gain the advantageous mutations for proliferation expand their clones (clonal expansion). Many dynamical models have been proposed to describe the cancer progression from this concept [69].

There are not so many diseases whose dynamics are known. But if system-level

understanding of disease dynamical structure is advanced, we could specify common characteristic dynamics for certain class of diseases, which might be used for qualitative prediction of course of diseases.

Finally, we will take a promising topic where the dynamical systems approach will be effective, which is "epithelial-mesenchymal transformation" (EMT), playing an important role in invasion or metastasis of cancer [70]. As we have already mentioned, cellular molecular network for gene expression regulation (transcriptional regulatory network) is mostly the same for all the tissues. But, depending on the cell type, the expressions of group of genes whose functions are necessary for that type of cells are promoted; whereas, those of other groups of genes which are not needed are suppressed, in order to attain the proper function of that type of cell. These combinations of up-regulated and down-regulated genes differ among cell types, but such combination of gene expression is quite normal and stable for each cell type.

Since we have 210 distinct cell types, if we consider the transcriptional regulatory network as a dynamical system, it could be considered that this network has 210 "stable solutions" corresponding to each

cell type. Each expression profile corresponding to a cell type is one of the stable solutions of transcriptional regulatory network.

EMT is a cellular process where cells in an epithelium acquire the ability to invade the underlying extracellular matrix or adjacent mesenchymal cells. During EMT, epithelial cells acquire a mesenchymal appearance with increased motility and invasiveness. It is a normal process in development and is used for the formation of three-dimensional structure in the embryo. However, in cancer, it is a key step for invasion and metastasis. EMT is characterized by the combined loss of epithelial cell junction proteins such as E-cadherin and integrin, and gain of mesenchymal markers such as vimentin and fibronectin related to cell motility.

From our standpoint of systems pathology, EMT can be considered as the “phase transition” (in large, coherent structural change) of the transcriptional regulatory network, which is the transition between different stable network solutions corresponding to epithelial and mesenchymal type of cell. Along with the phase transition of the transcriptional regulatory network, the change of genetic activities of several representative genes takes place such as E-cadherin (tightly connecting the neighboring cells), integrin (connecting to the basement membrane) or vimentin (related to the cell filament) and fibronectin (comprising extracellular matrix). EMT makes it possible for cancer cell of epithelial origin to invade mesenchymal tissues under the basement membrane such as connective tissue, muscular tissue and blood vessel, by “pretending” mesenchymal cell appearance. Based on the concept that EMT is “phase transition” of the molecular network, for example, we have a clue for cellular network approach to reveal what extent of aberrations of pathways is necessary and critical to evoke EMT phase transition for invasion and metastasis.

With the above system features in mind, we can comprehend the disease in the whole spectrum. Disease distorts bio-systems where normal pathway is modified to form “sustained disease pathway”.

4.3 Systems Pathology Informatics Fully Exploit the Disease Omics

Systems pathology, namely system-level understanding of diseases, is an effective framework to extract the full length of disease omics data. Based on the recognition of the above systems features of disease, various important clinical applications such as systems analysis, modeling, and simulation of disease are now in progress.

Among the applications of systems pathology, the most widely studied is the estimation of diseased molecular pathway underlying diseased omics. The approaches to estimate the molecular pathway from the gene expression profile, called “reverse engineering approach” [71], were developed first for the identification of the “normal” cellular functions. One of the first seminal studies promoting this approach aimed to reveal the gene regulatory networks in *Saccharomyces cerevisiae* from the gene expression profiles by using Bayesian network identification method [72]. Since then, many kinds of methods have been invented. Among them, mainly two kinds of methods are now mostly utilized.

One is the coexpression model which is an undirected probabilistic model based on the correlation coefficient, partial correlation coefficient (Gaussian graphical model [73]) or mutual information (Relevant Network [74], ARANCNE [75]) between any pairs of the genes. The other is Bayesian network model [76], which is a directed graphical model for representing probabilistic causality between gene activities. Although some of such purely data-driven approach showed successes in revealing the diseased pathways [77], it is considered [78] that there are many limitations in purely data-driven approaches for microarray data: they need a large sample size and capture the only parts of biologically relevant networks.

Combination of the transcriptomic data and biological knowledge of pathway is considered to be more feasible to estimate the diseased pathway. For example, gene expression profiles of diseased cells are superimposed on the protein-protein interaction network to identify the disease-causative subnetwork, which is called CGI (combining the gene expression and pro-

tein interaction data). Chuang et al. applied this approach to the gene expression profiles of metastatic or non-metastatic patients of breast cancer to identify the subnetwork whose expression levels correlate with the prognosis of metastasis [79]. We also superimposed the gene-expression profile on the BIND (Biomolecular Interaction Network, <http://bond.unleashedinformatics.com/index.jsp>) where protein-protein interaction is combined with gene (DNA)-protein (transcription factor) interaction, to identify the causative pathway for recurrence of the hepatocellular carcinoma after surgery [80].

Another promising approach is eQTL [81], which considers gene expression profile one of the quantitative traits (phenotypes) and applies the statistical genetics to explore the aberrations in genetic locus. The eQTL approach can relate gene expression profile to restricted number of causative genes. Besides the reverse engineering, the CGI and eQTL, a lot of methods have been reported for systems pathology approach to the diseases. The essentially similar concepts like “cancer systems biology” [82] or “systems genetics of disease” [83] are proposed on the common understanding that only the integrative approach to diseases like systems pathology can fully exploit the disease omics.

Network topological approach is also promising area of the systems pathology. Jeong [84] revealed that degree (number of connections of a node) of protein-protein interaction network follows “scale free” (power law) distribution which implies the hub-branch structure. Many network topological studies have been conducted such as those on the disease networks or disease gene network [42]. We applied the moving stratification method to extract the subnetworks within the specified degree range to find the densely connected middle-degreed layer (“backbone”) of protein-protein interaction network (PPI), beside the well known hub-branch structure [85]. On this three-layered structure of PPI, we mapped the target proteins of FDA-approved drugs. It was found that they are mainly located in middle layer of PPI network (► Fig. 3). This result means hub nodes (high-degree proteins) are not suitable for drug target because of possibility of broad side-effects.

Table 2 Features of the three generations of omics-based medicine

Generation	Name	Main information	Main goal	Informatics approach	Methods
1st	Genomic medicine	Mutations and polymorphisms of genome of germ line cells	Personalized (tailor-made) medicine	(Genetically) statistical	Statistical genetics, pharmacogenetics, pharmacogenomics
2nd	(Post-genomic) omics-based medicine	Comprehensive molecular profiles of somatic cells	Predictive and preventive medicine	Data-driven	Data-mining, machine learning, exploratory statistics
3rd	(Omics-based) systems medicine	Molecular pathways and networks and its alteration by diseases	Comprehensive (personalized, predictive, preventive) medicine	Model-driven	System identification, (reverse engineering), system simulation, system design

However, middle layer proteins are appropriate for drug targets with less serious side-effects and higher effectiveness as backbone proteins. Topological analysis does not require the numerical information of cellular network like a reaction coefficient, so that it is more feasible to draw useful findings about the cellular network where quantitative information is difficult to obtain.

Along with this line, what is equally important and have not been explored yet is the qualitative systems analysis and simulation [86] of diseased pathway. Several qualitative methods have once been explored in relation to the common sense physical reasoning. As mentioned above, quantitative information is difficult to obtain in the cellular network, whereas we have a lot of qualitative structural knowledge about the signaling pathway and gene-regulatory network. A new methodology would be expected to structurally or qualitatively analyze or predict the behavior of the disease pathway.

5. Discussions and Conclusion

In this article, we described the essential features and development of the omics-based medicine along with those of the informatics used for it. In describing the development of the omics-based medicine, we employed “three-generation paradigm”. Its first generation started with “genomic medicine” and then proceeded to the “(post-genomic) omics-based medicine” as the second generation, and finally ar-

rived at “omics-based systems medicine” as the third generation which is currently still in its beginning and is expected for future development

This three-generation paradigm would be a quite suitable concept for revealing the intrinsic development of omics-based medicine. Based on this paradigm, the typical informatics method used in each generation is clearly classified: classical statistics (statistical genetics) for the first generation was succeeded by the data-driven approach like machine learning, data-mining or exploratory statistics, which was eventually replaced by model (hypothesis)-driven approach.

Another advantage of this three-generation paradigm is that it provides the unified platform which can treat three concepts of molecular medicine inclusively: 1) “genomic medicine” aiming to realize the personalized medicine based on the inborn difference of genome, 2) “(post-genomic) omics-based medicine” aiming to realize the predictive medicine based on the gene expression profiles and protein mass spectra which vary depending on the states of the disease, and 3) “systems medicine” aiming to realize the wholistic understanding of disease based on the knowledge of cellular pathway and its alternations by diseases (► Table 2).

Although we rather overemphasized the newly acquired features of each generation, it should be noted that, in the actual medical activity, the subsequent generation involves that of the preceding generation. For example, the second generation of omics-based medicine involves the personalized care in genomic medicine. Hence, systems

medicine of the third generation is most comprehensive medicine involving the personalized and predictive care.

Disease is generally considered as a “complex system” in the sense that it is a hierarchically organized, self-sustainable, dynamical system which evolves on its own law. We have discussed several intrinsic features of disease as a dynamical complex system, such as self-sustainability of hypertension by bidirectional causality, evolutionary dynamics of HIV, cancer clone expansion, cancer metastasis (EMT) as systems phase transition, possibility of qualitative causal reasoning of disease dynamics, and topological location of drug target in the protein interaction network.

It would be expected that quite different view of the disease will emerge under the light of systems approach to it. However, systems approach to disease is theoretically in its beginning and much to be done in the future.

On the contrary, in practical application of system concept of disease, a number of informatics methods are now proposed and applied for the reverse engineering to estimate the cellular diseased network from omics data.

In conclusion, pathway-based understanding of disease, which we call the systems pathology, is the most promising framework to fully exploit the information contents of disease omics and comprehensively recognize the disease process. Together with its precedent generations, we could expect the omics-based systems medicine substantially revolutionize current medical care to make it more personalized and predictive.

Acknowledgment

This work was supported by Special Coordination Funds for Promoting Science and Technology (“Fostering talent in emergent fields” program and “Research and Development program for resolving critical issues”) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The author would like to express special thanks to Dr. Kaoru Mogushi for assisting in preparation of this article and also to the reviewers for their constructive comments for improving the article.

References

- Weinstein JN. Fishing expeditions. *Science* 1998; 282: 628–629.
- Weinstein JN. ‘Omic’ and hypothesis-driven research in the molecular pharmacology of cancer. *Curr Opin Pharmacol* 2002; 2: 361–365.
- Knaup P, Ammenwerth E, Brandner R, Brigl B, Fischer G, Garde S, Lang E, Pilgram R, Ruderich F, Singer R, Wolff AC, Haux R, Kulikowski C. Towards clinical bioinformatics: advancing genomic medicine with informatics methods and tools. *Methods Inf Med* 2004; 43: 302–307.
- Gaughan A. Bridging the divide: the need for translational informatics. *Pharmacogenomics* 2006; 7: 117–122.
- Guttmacher AE, Collins FS. Genomic medicine – a primer. *N Engl J Med* 2002; 347: 1512–1520.
- Kumar D. Genomic medicine: a new frontier of medicine in the twenty first century. *Genomic Med* 2007; 1: 3–7.
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270: 467–470.
- Fodor SP, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL. Multiplexed biochemical assays with biological chips. *Nature* 1993; 364: 555–556.
- Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. Genome-wide location and function of DNA binding proteins. *Science* 2000; 290: 2306–2309.
- Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engle P, McDonagh PD, Loerch PM, Leonardson A, Lum PY, Cavet G, Wu LF, Altschuler SJ, Edwards S, King J, Tsang JS, Schimmack G, Schelter JM, Koch J, Ziman M, Marton MJ, Li B, Cundiff P, Ward T, Castle J, Krolewski M, Meyer MR, Mao M, Burchard J, Kidd MJ, Dai H, Phillips JW, Linsley PS, Stoughton R, Scher S, Boguski MS. Experimental annotation of the human genome using microarray technology. *Nature* 2001; 409: 922–927.
- Xing Y, Kapur K, Wong WH. Probe selection and expression index computation of Affymetrix Exon Arrays. *PLoS ONE* 2006; 1: e88.
- Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T. Protein and polymer analyses up to m/z 100000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 1988; 2: 151–153.
- Hutchens TW, Yip TT. New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom* 1993; 7: 576–580.
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; 286: 531–537.
- Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JJ, Yang L, Marti GE, Moore T, Hudson J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403: 503–511.
- Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltman JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, López-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002; 346: 1937–1947.
- Tanaka H. Bioinformatics and genomics for opening new perspective for personalized care. *Stud Health Technol Inform* 2008; 134: 47–58.
- Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; 359: 572–577.
- Meani F, Pecorelli S, Liotta L, Petricoin EF. Clinical application of proteomics in ovarian cancer prevention and treatment. *Mol Diagn Ther* 2009; 13: 297–311.
- Grubb RL, Deng J, Pinto PA, Mohler JL, Chinnaiyan A, Rubin M, Linehan WM, Liotta LA, Petricoin EF, Wulfkuhle JD. Pathway biomarker profiling of localized and metastatic human prostate cancer reveal metastatic and prognostic signatures. *J Proteome Res* 2009; 8: 3044–3054.
- Tanaka H. System-theoretic approach to life, genome and disease. *Proc of 3rd annual conference of Chemo-Bioinformatics* 2002. p 3.
- Tanaka H, Nakaya J. Research and education for biomedical informatics at Tokyo Medical and Dental University. *Yearb Med Inform* 2007. pp 157–162.
- Kitano H. Systems biology: a brief overview. *Science* 2002; 295: 1662–1664.
- Kitano H. Computational systems biology. *Nature* 2002; 420: 206–210.
- Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. *Science* 2004; 306: 640–643.
- Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. *Genome Med* 2009; 1: 2.
- International human genome sequencing consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431: 931–945.
- Strachan T, Andrew PR. *Human molecular genetics* (3rd edition). New York: Garland Publishing; 2003.
- Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY. A polymorphic DNA marker genetically linked to Huntington’s disease. *Nature* 1983 306: 234–238.
- Balding DJ, Bishop M, Cannings C. *Handbook of Statistical Genetics* (3rd edition). Hoboken: Wiley; 2007.
- Huntington’s Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington’s disease chromosomes. *Cell* 1993; 72: 971–983.
- Hoffman EP, Brown RH, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* 1987; 51: 919–928.
- Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 1989; 245: 1059–1065.
- Schaid DJ, Sommer SS. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am J Hum Genet* 1993; 53: 1114–1126.
- Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SE, Ward R, Lander ES. Linkage disequilibrium in the human genome. *Nature* 2001; 411: 199–204.
- Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med* 2009; 360: 1759–1768.
- Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 2007; 8: 657–662.
- Diabetes Genetics Initiative. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; 316: 1331–1336.
- The Wellcome Trust case control consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447: 661–678.
- Teo YY, Small KS, Clark TG, Kwiatkowski DP. Perturbation analysis: a simple method for filtering SNPs with erroneous genotyping in genome-wide association studies. *Ann Hum Genet* 2008; 72: 368–374.
- International Hapmap consortium. The International HapMap Project. *Nature* 2003; 426: 789–796.
- Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci U.S.A.* 2007; 104: 8685–8690.
- Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. *Nature* 2004; 429: 464–468.
- Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 1998; 279: 1200–1205.

45. FDA. Guidance for Industry Pharmacogenomic Data Submissions, 2003. <http://www.fda.gov/cder/guidance/5900dft.pdf>
46. Blackledge G, Averbuch S. Gefitinib ('Iressa', ZD1839) and new epidermal growth factor receptor inhibitors. *Br J Cancer* 2004; 90: 566–572.
47. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004; 350: 2129–2139.
48. Asahina H, Yamazaki K, Kinoshita I, Sukoh N, Harada M, Yokouchi H, Ishida T, Ogura S, Kojima T, Okamoto Y, Fujita Y, Dosaka-Akita H, Isobe H, Nishimura M. A phase II trial of gefitinib as first-line therapy for advanced non-small cell lung cancer with epidermal growth factor receptor mutations. *Br J Cancer* 2006; 95: 998–1004.
49. Baselga J. Clinical trials of Herceptin (trastuzumab). *Eur J Cancer* 2001; 37 (Suppl 1): S18–24.
50. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001; 344: 783–792.
51. Schwarz UI, Ritchie MD, Bradford Y, Li C, Dudek SM, Frye-Anderson A, Kim RB, Roden DM, Stein CM. Genetic determinants of response to warfarin during initial anticoagulation. *N Engl J Med* 2008; 358: 999–1008.
52. Aneesh TP, Sonal SM, Jose A, Chandrana L, Zachariah SM. Pharmacogenomics: The Right Drug to the Right Person. *J Clin Med Res* 2009; 1: 191–194.
53. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerckhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415: 530–536.
54. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U.S.A.* 2003; 100: 8418–8423.
55. Slodkowska EA, Ross JS. MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev Mol Diagn* 2009; 9: 417–422.
56. Asad J, Jacobson AF, Estabrook A, Smith SR, Boolbol SK, Feldman SM, Osborne MP, Boachie-Adjei K, Twardzik W, Tartter PI. Does oncotype DX recurrence score affect the management of patients with early-stage breast cancer? *Am J Surg* 2008; 196: 527–529.
57. Cheng JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, Mohsin S, Osborne CK, Chamness GC, Allred DC, O'Connell P. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003; 362: 362–369.
58. Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K, Lecoche M, Metivier J, Booser D, Ibrahim N, Valero V, Royce M, Arun B, Whitman G, Ross J, Sneige N, Hortobagyi GN, Pusztai L. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* 2004; 22: 2284–2293.
59. Mogushi K, Nomura F, Tomonaga T, Sunaga M, Sogawa K, Tanaka H. Evaluation of feature selection and disease classification method for protein expression profile analysis using SELDI-TOF-MS. Proc. 2nd Annual Meeting of Japan Society for Clinical Proteomics 2006. p 20.
60. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008; 26: 1135–1145.
61. Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. *Nat Biotechnol* 2009; 27: 455–457.
62. Kleemann R, Verschuren L, van Erk MJ, Nikolsky Y, Cnubben NH, Verheij ER, Smilde AK, Hendriks HF, Zadelaar S, Smith GJ, Kaznacheev V, Nikolskaya T, Melnikov A, Hurt-Camejo E, van der Greef J, van Ommen B, Kooistra T. Atherosclerosis and liver inflammation induced by increased dietary cholesterol intake: a combined transcriptomics and metabolomics analysis. *Genome Biol.* 2007; 8: R200.
63. Zhao JJ, Roberts TM, Hahn WC. Functional genetics and experimental models of human cancer. *Trends Mol Med* 2004; 10: 344–350.
64. Hirano T. Cytokines in autoimmune disease and chronic inflammatory proliferative disease. *Cytokine Growth Factor Rev* 2002; 13: 297–298.
65. Schiffrin EL. Remodeling of resistance arteries in essential hypertension and effects of antihypertensive treatment. *Am J Hypertens* 2004; 17: 1192–200.
66. Drummond A, Rodrigo AG. Reconstructing genealogies of serial samples under the assumption of a molecular clock using serial-sample UPGMA. *Mol Biol Evol* 2000; 17: 1807–1815.
67. Ren F, Ogishima S, Tanaka H. Longitudinal phylogenetic tree of within-host viral evolution from noncontemporaneous samples: a distance-based sequential-linking method. *Gene* 2003; 317: 89–95.
68. Hasegawa N, Sugiura W, Shibata J, Matsuda M, Ren F, Tanaka H. Inferring within-patient HIV-1 evolutionary dynamics under anti-HIV therapy using serial virus samples with vSPA. *BMC Bioinformatics* 2009; 10: 360.
69. Michor F, Iwasa Y, Nowak MA. Dynamics of cancer progression. *Nat Rev Cancer* 2004; 4: 197–205.
70. Lee JM, Dedhar S, Kalluri R, Thompson EW. The epithelial-mesenchymal transition: new insights in signaling, development, and disease. *J Cell Biol* 2006; 172: 973–981.
71. Liang S, Fuhrman S, Somogyi R. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput* 1998; 18–29.
72. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol* 2000; 7: 601–620.
73. Edwards D. Introduction to Graphical Modelling. Heidelberg: Springer; 2000.
74. Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput* 2000. pp 418–429.
75. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006; 7 (Suppl 1): S7.
76. Friedman N, Koller D. Being Bayesian about network structure. *Math Learn* 2003; 50: 95–126.
77. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet* 2005; 37: 382–390.
78. Markowitz F, Spang R. Inferring cellular networks—a review. *BMC Bioinformatics* 2007; 8 (Suppl 6): S5.
79. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol* 2007; 3: 140.
80. Tanaka S, Mogushi K, Yasen M, Noguchi N, Kudo A, Kurokawa T, Nakamura N, Inazawa J, Tanaka H, Arii S. Surgical contribution to recurrence-free survival in patients with macrovascular-invasion-negative hepatocellular carcinoma. *J Am Coll Surg* 2009; 208: 368–374.
81. Göring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet* 2007; 39: 1208–1216.
82. Wang E, Lenferink A, O'Connor-McCourt M. Cancer systems biology: exploring cancer-associated genes on cellular networks. *Cell Mol Life Sci* 2007; 64: 1752–1762.
83. Sieberts SK, Schadt EE. Moving toward a system genetics view of disease. *Mamm Genome* 2007; 18: 389–401.
84. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization of metabolic networks. *Nature* 2000; 407: 651–654.
85. Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H. Structure of protein interaction networks and their implications on drug design. *PLoS Comput Biol* 2009; 5: e1000550.
86. Kuipers B. Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge (Artificial Intelligence). Cambridge: The MIT Press; 1994. p 13.