

## REVIEW

# Systems biomedicine: It's your turn

## —Recent progress in systems biomedicine

Zhuqin Zhang, Zhiguo Zhao, Bing Liu, Dongguo Li, Dandan Zhang, Houzao Chen and Depei Liu\*

State Key Laboratory of Medical Molecular Biology, Department of Biochemistry and Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China

\* Correspondence: liudp@pumc.edu.cn

Received December 25, 2012; Revised April 9, 2013; Accepted April 19, 2013

**The concept of “systems biology” is raised by Hood in 1999. It means studying all components with a systematic view. Systems biomedicine is the application of systems biology in medicine. It studies all components in a whole system and aims to reveal the patho-physiologic mechanisms of disease. In recent years, with the development of both theory and technology, systems biomedicine has become feasible and popular. In this review, we will talk about applications of some methods of omics in systems biomedicine, including genomics, metabolomics (proteomics, lipidomics, glycomics), and epigenomics. We will particularly talk about microbiomics and omics for common diseases, two fields which are developed rapidly recently. We also give some bioinformatics related methods and databases which are used in the field of systems biomedicine. At last, some examples that illustrate the whole biological system will be given, and development for systems biomedicine in China and the prospect for systems biomedicine will be talked about.**

### INTRODUCTION

The concept of “systems biology” is raised by Hood in 1999, when the human genome draft was nearly completed. It means studying all components with a systematic view. Systems biomedicine is the application of systems biology in medicine. Its aim is to reveal the patho-physiologic mechanisms of disease. It studies all components in a biological system holistically (including the DNA, mRNA, proteins, and small biological molecules in a cell, tissue or body) under defined conditions, and reveals the complex interactions between these components [1].

Systems biomedicine requires a multidisciplinary approach, with cooperation from experts in different fields including life sciences, information science, mathematics, computer science and other disciplines. Actually, this idea of multidisciplinary cooperation was proposed by Kamada [2] and Zeng [3] in 1992. Because of the lack of high-throughput biologic research tools, systems biomedicine research was hindered for many years. In recent years, with the development of both theory and technology, such as omics and bioinformatics,

plus cooperation between different fields, including biology, mathematics and computer science, systems biomedicine has received renewed interest.

In this review, we will talk about applications of omics methods in systems biomedicine, including genomics, metabolomics (proteomics, lipidomics, glycomics), and epigenomics. Microbiomics and common diseases omics are two fields which are developed rapidly. We will talk particularly about these two fields. As for the tools for systems biomedicine, bioinformatics is necessary and important. We will give some bioinformatics related methods and databases. At last, some examples that illustrate the whole biological system at the integrative level will be discussed.

### APPLICATIONS OF OMICS

#### Genomics

Genomics is a broad field. It focuses on the study of the genomes, includes whole-genome DNA sequencing and genetic mapping of organisms, and includes studies of multiple genomic phenomena, such as gene expression

and gene-gene interaction [4]. Rapid advances in genetic sequencing techniques have revolutionized genetic analysis, and DNA sequencing at the whole-genome level is now possible. Despite the completion of the Human Genome Project in 2003, which revolutionized human genetics, linkages between genomic data and phenotypic information are still limited. For this reason, further approaches are preformed to investigate the relationships between genotype and phenotype in organisms.

### GWAS

Genome-wide association study (GWAS) is a method to study the association between phenotype and single nucleotide polymorphisms (SNPs), which are polymorphic markers found quite evenly throughout the genome. GWAS is a revolutionary approach because it can estimate the genetic association between the entire genome and a disease in numerous unrelated individuals at high resolution, and is not affected by previous hypotheses about genetic associations with disease [5]. Early in 1996, Risch and Merikangas first suggested using GWAS approach to study the complex causes of human disease [6]. GWAS results were published in 2005 [7] and 2006 [8], and the technique was given a major kick-start by the Wellcome Trust Case Control Consortium (WTCCC) [9].

### Monogenic and oligogenic diseases

Monogenic and oligogenic diseases are generally thought of as simple Mendelian diseases, such as sickle cell disease (SCD) and  $\beta$ -thalassemia. In 2008, a major GWAS showed that rs11886868 located in BCL11A gene strongly affected fetal hemoglobin levels in 4305 Sardinians and in a large number of sickle cell patients [10]. Another report in 2008 identified an SNP located in the BCL11A gene that was associated with fetal hemoglobin levels and pain crises in sickle cell disease [11]. These findings suggested that BCL11A plays an important role in regulation of fetal globin expression. As expected, further research has identified BCL11A as a specific regulator of the expression of human hemoglobin [12,13].

Although exon regions only occupy 1% of the genome, 85% of the exon mutations will lead to Mendelian-inherited disorders. In recent years, whole-exome sequencing (WES) has been a powerful research tool to reveal novel exon mutations in Mendelian disorders with obscure etiologies [14–16]. In 2010, a WES study found a link between Miller syndrome and mutations in DHODH [17]. Recent WES studies have revealed that *de novo* germline SNPs in single genes is the major cause of rare sporadic malformation syndromes such as Schinzel-

Giedion syndrome, Kabuki syndrome and Bohring-Opitz syndrome [18–20]. Therefore, the widespread availability of exome sequencing would promote the study of diseases, especially for the monogenic and oligogenic diseases.

### Complex diseases

Coronary artery disease (CAD) and myocardial infarction (MI) are the leading causes of disability and mortality worldwide [21,22]. Recently, GWAS was used to identify loci for CAD and discover its risk factors in humans, and the results were recently reviewed [23]. Several studies have identified SNP 9q21 as a high risk locus for CAD [9,24–29]. Further studies have confirmed that the risk SNP 9q21 could regulate the expression of CDKN2A/B in humans [30–32]. Recently, a GWAS discovered four novel risk loci named 2q24.1, 4p32.1, 6p21.32 and 12q21.33 in 33000 Han Chinese cohorts. These findings could provide new insights into the pathways that contribute to CAD susceptibility in the Han Chinese population [33]. As one of the most important risk factors of CAD, hypertension is also widely studied. There were two major studies of hypertension using the GWAS approach in 2007 [9,34]. However, both studies did not identify markers associated with hypertension on a genome-wide level. Nevertheless, the following two studies have identified 8 genetic loci associated with systolic or diastolic blood pressure and 10 SNPs with hypertension in subjects of European ancestry [35,36]. GWASs have concluded that the 12q21 locus is associated with systolic blood pressure in non-European ethnic groups [37,38]. Moreover, the strengths and weaknesses of GWAS on hypertension research were reviewed recently [39].

During the past few years, GWAS has identified numerous strong associations between genetic loci and different types of cancer. Some loci, such as 8q24, were identified as cancer-susceptibility regions for many unrelated cancers, including prostate cancer [40], glioma [41], breast cancer [42], colorectal cancer [43], bladder cancer [44], ovarian cancer [45] and pancreatic cancer [46]. Therefore, an investigation into those loci may reveal new mechanisms of carcinogenesis.

Type 2 diabetes (T2D) is one of the most prevalent metabolic diseases. Before the GWAS era, only one locus located in glucokinase gene (GCK) had been strongly associated with fasting glucose levels [47]. GWASs have identified about 50 risk loci of T2D to date [48–51].

GWASs have also identified a number of SNPs with other complex diseases, such as auto-immune disease and psychiatric disorders.

In recent years, many important biologic discoveries have been made via GWAS. This approach has revealed

the associations between genomic variants and complex or monogenic diseases in large populations, especially the linkages between SNPs and common complex diseases, such as diabetes, cardiovascular disease (CVD), autoimmune disease, psychiatric disorders. To date, more than 2000 genetic loci have been identified with significant association to one or more complex traits [52]. Although association between risk loci and diseases are well documented, the mechanisms of these associations require further research. Furthermore, GWAS may also discover new targets of complex diseases for drugs [53–55]. Based on GWAS results, deeper sequencing and analysis of the risk loci, and integration of the results with metabolomics or even cellular pathways may reveal some new, potential avenues for making targeted drugs and clinical interventions [56].

## ENCODE

The aim of the Encyclopedia of DNA Elements (ENCODE) Project is to determine all the functional elements in the human DNA sequence. The whole project is divided into three phases. The initial pilot phase of the project ran from 2003 to 2007, focusing on the sequence elements and discovered their biological function (approximately 1% of human genome sequence) [57]. The second phase covered the whole genome, including 70000 promoter regions and 400000 enhancer regions [58]. Results of the second phase were published in September 2012, including 6 papers in *Nature*, 18 papers in *Genome Research* and 6 papers in *Genome Biology*. ENCODE's analysis revealed that about 80% of the human genome has a "biochemical function." It also revealed some new aspects of gene expression and regulation, and the organization of related information [58]. The last phase, which is still under way, will finish the annotation of all human DNA elements, aiding our understanding of normal life processes, and of mechanisms of disease.

## Metabolomics

Metabolomics, one of the most important fields in systems biomedicine, is the study of biochemical processes. The metabolome consists of many metabolites, which are the small molecules produced by enzymes in cells, tissues, organs or organisms [59]. Thanks to recent technological advancements, thousands of metabolites can now be measured, fast and quantitatively. Biomarker discovery by metabolomics can help patients and doctors make better drug choices.

## Proteomics

Proteomics is the study of the properties of proteins, such

as their structures, expression levels, post-translational modifications, and interactions [60]. In 1989, Fields and Song devised the yeast two-hybrid (Y2H) method to probe for protein–protein interactions [61]. High-throughput Y2H maps have since been generated for many different species.

An alternative approach, high-throughput co-affinity purification followed by mass spectrometry (AP/MS), can also be used to detect protein–protein interactions. In 2008, Yu et al. produced a high-quality binary interactome network (a protein interaction map) in yeast supported by literature-curated protein interaction data sets [62]. In 2010, a genetic interaction map of the entire *Saccharomyces cerevisiae* genome displayed more details of protein–protein interactions in cells, and it also identified that extensive and precise genetic landscape mapping could help to explain genetic interactions and help with drug target identification [63]. Several studies have highlighted proteins that interact directly with proteins already known to be implicated in pathogenesis [64–67]. Therefore, a thorough understanding of the function and structure of biological networks can reveal how diseases arise and progress.

## Lipidomics

Cellular lipids, which are generated and metabolized by enzymes, are small molecules with great chemical diversity. All biological membranes contain amphiphilic lipids, including glycerophospholipids, sterols, and sphingolipids. Cellular organelles usually have different, organelle-specific lipids. Mitochondria, for instance, are enriched with cardiolipin. Some diffusible and soluble lipids are usually considered as signal molecules, such as arachidonic acid, lipoxin B<sub>4</sub>, prostaglandin H<sub>2</sub> and platelet-activating factor. However, highly non-polar lipids, which are usually synthesized in the endoplasmic reticulum (ER), are usually stored in lipid bodies as energy stores.

In recent years, lipidomics has benefited from novel analytical approaches, particularly liquid chromatography (LC) and mass spectrometry (MS) [68]. MS is often coupled with LC, which can separate a lipid before their introduction to the ionization source of the mass spectrometer. High-resolution hybrid systems couple the advantages of several mass analyzers into a single instrument and allow for highly accurate mass measurements of ion species [69,70]. LC/MS has revealed over 500 different types of lipids in human plasma alone, such as fatty acyls, glycerolipids, and glycerophospholipids [71].

Clinically, the most important plasma lipids are cholesterol and triglyceride (TG). Due to the non-polar character of cholesterol and TG, the lipoproteins (spheroidal macromolecules) are required for their

secretion into the plasma. Studies have shown that levels of these plasma lipids and lipoproteins are linked to the susceptibility of CVD [72,73]. A comprehensive analysis of these lipids shows that the compositions of lipids in atherosclerotic plaques are different from normal plasma lipids [74]. Further studies have shown that plasma lipid profiling could be used to test for risk of unstable CAD [75]. In obese (ob/ob) mice, lipid profiles in liver and their correlation networks were significantly different when compared to the control mice [76]. Statins are the most commonly prescribed drugs for the prevention of CVD, and work by reducing the level of low density lipoprotein (LDL) cholesterol in the plasma. Individuals with different lipid metabolisms respond differently to statin treatment, and lipidomics could be used to guide the dosage [77].

Lipid profiling chromatography, MS and nuclear magnetic resonance (NMR) provide compositional identification of lipid samples. Proton magnetic resonance spectroscopy permits the study of biochemistry and metabolism of lipids in a living cell, and then provides a specific index to detect the abnormalities and progression of diseases *in vivo* [78]. Recently, a novel approach for lipid profiling, named single-cell laser-trapping Raman spectroscopy, may directly quantify lipid profiling in single cell *in vivo*. This new method could be very useful for a diverse range of applications in lipidomics [79].

#### Glycomics

Glycomics is the quantification of the glycome of a cell, tissue or organism [80]. Most cells, from prokaryotic bacteria to mammals, are coated with a glycocalyx (or 'sugar coat'). In eukaryotes, glycans bond covalently with proteins and lipids to form the dynamic, structurally diverse family of glycoconjugates. Glycoproteins and glucolipids are responsive to a wide range of intercellular and intracellular biological processes [81].

In recent years, studies have already revealed some potential biomarkers for multiple sclerosis [82], cancer [83] and inflammation-related diseases [84] in serum. Callewaert et al. showed that glycome profiles vary significantly in different stages of fibrosis [85].

Importantly, glycoproteins are also used in clinical cancer diagnosis. Therefore, discovering the biomarkers using glycomics is very beneficial for the early detection of cancer [86]. Recently, fucosylated  $\alpha$ -fetoprotein has been used as a diagnostic marker of primary hepatocarcinoma [81]. In the future, discovering biomarkers using glycomics will not only provide a new paradigm for understanding the role of the glycome in many biologic areas, but will also aid the process of clinical disease diagnosis.

#### Epigenomics

Maunakea et al. define epigenomes as "the combination of entire genome-wide chromatin modifications in any given cell types that directs its unique gene expression pattern" [87]. Unlike the genome, epigenomes are highly dynamic in different cell types. Epigenomes, characterized by DNA methylation, histone modifications, post-transcriptional regulation via miRNAs, and post-translational regulation via protein modification, establish and maintain cell type-specific gene expression states [88]. Moreover, studies have shown that the epigenotype plays a critical role in different types of diseases.

##### DNA methylation

DNA methylation of cytosine at position C5 in CpG dinucleotides is the major target for DNA modification in mammals [89]. DNA methylation is involved in the regulation of cellular processes, such as embryonic development, and genomic imprinting [90]. As early as 1983, a linkage between DNA hypomethylation and cancer was discovered [91]. Recent studies have suggested that loss of genomic methylation is an early event in cancer [92], and it is known that specific and global DNA methylation is usually disordered during carcinogenesis. A recent study showed that the change of DNA methylation in p15 (INK4b) is strongly associated with expression of ANRIL on chromosome 9p21 and CAD [93]. Further studies of methylation could provide new insights in disease and novel strategies of diagnosis and therapy.

##### Histone modifications

In 2008, Wang et al. analyzed 39 histone modifications in human CD4 + T cells [94]. Histone modifications play a key role in the regulation of gene expression. For this reason, histones are increasingly being recognized as dynamic regulators of gene activity, which are controlled by several post-translational chemical modifications, such as acetylation, methylation, phosphorylation, ubiquitylation and sumoylation [95]. Studies showed that altered histone modifications can lead to many human diseases, mostly related to cancers. It has been identified that the global loss of acetylation and trimethylation in H4K16 and H4K20 respectively, for instance, was associated with breast and liver cancer in studies [96,97]. Recently, a novel therapeutic strategy for aggressive B cell lymphomas through altered histone modifications was identified [98]. In diabetes development, key genes Pdx1 and Glut4 were proved to be regulated by DNA methylation and histone modifications [99,100]. Further studies of histone



modifications are useful to the understanding of both cell development and the disease therapy.

#### RNA editing and miRNA

Post-transcriptional modification of mRNA plays a key role in gene expression regulation and cell development. Early in 1991, RNA editing was identified as a determinant controller of ion flow in glutamate-gated channels in the brain [101]. Recently, common, tissue-specific methylations of the N6 position of adenosine (m6A) of mammalian mRNA were detected, and methylations of m6A were found to be specifically enriched near stop codons and in 3'UTR [102]. Moreover, RNA editing is also involved in some disease progress. Studies of 2C-subtype serotonin receptor RNA editing patterns in psychiatric disorders were recently reviewed [103].

The noncoding region includes several types such as microRNA, siRNA, piRNA and long noncoding RNA. These noncoding RNAs are functional in chromatin modification, transcription and post-transcription modification. Researches demonstrated that noncoding RNAs played a role in disease and could be diagnostic markers or therapeutic targets [104,105]. ANRIL, a long noncoding RNA, was found to be associated with CVD by GWAS. Research demonstrated that ANRIL can regulate the expression of the CDKN2A/CDKN2B locus and thus influence the proliferation of vascular smooth muscle cell and coronary heart disease [31,106]. BACE1-AS level was found to be elevated in Alzheimer's disease patients, and BACE1-AS can regulate BACE mRNA expression involved in the Alzheimer's disease pathology [107].

miRNAs relevant to epileptogenesis contain m6A sites and could be regulated by RNA epigenetic modification [108]. Mutation and dysfunction of miRNA may lead to various diseases [109]. Due to the informative nature of circulating miRNAs, many miRNAs are used as biomarkers for tumors. And different biomarkers may reflect presence of CVD, or tumors in specific tissues and differentiation of the states [110,111]. One recent study showed that miRNA-21 could affect myocardial disease by regulating MAP kinase signaling pathway in fibroblasts [112].

#### Protein modification

Post-translational modifications of protein regulates protein activation and other cell processes. One recent study identified nearly 200000 protein post-translational modification of the sites across 11 eukaryotic species [113]. Many of these studies suggested that the mutation of the post-translational target sites were directly or indirectly involved in disease. For instance, the abnormal post-translational modifications in prion protein were

shown to be associated with autosomal dominant spongiform encephalopathy [114]. The study of post-translational modifications will improve our fundamental understanding of the mechanisms of disease.

#### Mitochondrial genome

Human mitochondria contain a compact circular genome [115]. The regulation and expression of the human mitochondrial genome is unique. The first complete map of the human mitochondrial transcriptome was supplied in 2011 [116], and this map could enhance the research of disease-associated variants [116]. Mutations in the mitochondrial genome have been linked with many common diseases [117]. Studies showed that dysfunction of mitochondrial DNA (mtDNA) was involved in diabetes [118,119], and a recent study identified that mtDNA mutation C3256T in white blood cells is involved in atherosclerosis and CAD [120]. Although there is a clear association between mtDNA mutations and some diseases, the true relationship between mtDNA mutations and human health needs further investigation.

With the breakthrough of technology, especially next-generation sequencing technology, epigenomics is entering a new era. It is now possible to map dynamic epigenetic information with precision and speed. Epigenomic research may be highly beneficial to our understanding of the disease pathology in the future.

#### Human Microbiome Project

Animals and microorganisms have existed together for hundreds of millions of years. The harmonious relationship between animals and microorganisms is closely related to men's health. Researchers gradually realized the importance of the microbiome after the implementation of the Human Genome Project which was launched by the US National Institute of Health (NIH) in 2007. Metagenomics of the Human Intestinal Tract (MetaHIT) was then initiated in 2008 [121], which aims to sequence 3000 bacterial genomes found in the human gut [122]. NCBI and DACC have published 800 bacterial genomes so far [123].

The human body contains various microbiotas, and each microbiota contains various microbial cells [124]. Researchers have found that these microbiome genomes have a great influence on human health.

#### Microbiome diversity and similarity

Intestinal microbiota are dynamic and instable. Almost 99.9% of the human genome is identical, however, there are extreme differences in the human microbiome [124]. Since some diseases are associated with the microbiome,

these differences offer profound insights in the field of personalized medicine.

One of the most important features of the microbiome is its diversity. Microbiomes vary with the geography, age, and lifestyle of their human hosts [125,126]. Older people have a more diverse microbiome. A person's microbiome is also different in different regions of the body. Technologies such as 16S rRNA, stool transplantation [124] and sequencing of DNA are very powerful. Ever-decreasing sequencing costs provide large amounts of microbiomic data, which will aid in our understanding of the diversity of the microbiome.

In spite of a large amount of differences in human microbiome, several metabolic pathways are identical [127]. Stable metabolic pathway is the distinguishing feature of healthy people which can be employed to diagnose diseases. If the metabolic pathways deviate from the normal, a poor health will be the result. This suggests that it is very important to treat diseases with metabolic pathways of microorganisms.

#### Microbiome and diseases

Pathogenesis research at the genome level is insufficient. From former observations in mouse models to late studies in human volunteers, a series of evidence has proven that intestinal microbes play a critical role in diseases.

A complex interaction between bacteria and intestinal tissue could play an important role in pathogenesis [128]. Germ-free mice receiving microbiota from conventional mice presented obesity, and this was shown to be attributed to energy deposition into host adipocytes by microbiota [129]. The gut microbiome is a marker for diabetes. Research has indicated that microorganisms in type 2 diabetes (T2D) patients are more abundant than in controls [130]. Metabolic disease is affected by a multitude of factors, but diet plays the biggest role [131]. Lipopolysaccharides (LPS) and other bacterial fragments can lead to the development of metabolic and cardiovascular diseases [132]. Commensal microorganisms have also been linked to asthma. Disrupting the microbiota of babies has been shown to enhance the morbidity of asthma [133]. Wang et al. found that gut flora metabolism of phosphatidylcholine promoted cardiovascular diseases [134].

#### New model for diseases and prospects

The gut microbiome should be considered as a risk factor for the development of metabolic diseases. So a new model comes out to describe the influencing factors of diseases [135]. The new model emphasizes the importance of the interaction between gut microbiota and the host, whereas the old model is only based on environ-

mental and genetic factors. We should attach the importance of dietary, lifestyle and physiologic characteristics to health and take microbiome more seriously. We hope that we can understand the human health more systematically and use more effective ways to treat diseases.

#### Common diseases

One of the aims of systems biomedicine is to reveal the patho-physiologic mechanisms of diseases. Genetic diseases are categorized into three types: single-gene disorders, common complex diseases, and oligogenetic diseases. The common complex traits are influenced by many factors, each of which has only a modest effect. GWAS, epigenomics, metabolomics and microbiomics are playing important roles in studying the mechanisms of common disease.

#### Genome-wide association studies (GWAS)

Genome-wide association studies (GWAS) could be a powerful instrument to understand the network of common diseases, and an increasingly large body of evidence is supporting this theory. A majority of variants lie within noncoding sequences or within introns, which are often marked by deoxyribonuclease I (DNase I) hypersensitive sites (DHSs) [136]. The WTCCC [9] pointed out that many previously identified genetic loci were connected in a network of seven common diseases, and some loci were linked to more than one disease. Torkamani et al. suggested that due to the genetic links among diseases, a new way to categorize disease should be explored [137]. Harold et al. found that Alzheimer's disease was associated with *CLU* and *PICALM* gene by a two-stage GWAS [138]. Chen et al. found that rs3803662 and rs10941679 were associated with breast cancer [139]. Yeager et al. identified that in 10286 cases and 9135 controls of European ancestry, chromosome 8q24a was associated with prostate cancer [140]. Schnabel et al. reported that thanks to the development of novel technologies in GWAS, we could understand the complexity of CVD more systematically [141].

The GWAS approach is very promising. It could be further improved by integrating other omics and technologies to develop a new strategy to study the network of common diseases [142].

#### Epigenomics

A large proportion of mechanisms among common diseases are elusive only by the strategy of GWAS [143]. Epigenetic modifications are heritable alterations of the genome and its feature is that it can change the gene

expression, but cannot change the DNA sequence. Epigenetic modifications play pivotal roles in studying the networks of common diseases, and due to the development of high-throughput arrays, the high-throughput study of epigenetic modifications is now possible [144].

Toyota et al. reported that DNA methylation regulated diverse functions such as imprinting, genomic stability, and gene transcription, and it might even be associated with human tumors [145]. Ivanova et al. reported that BMP4 promoter methylation levels were directly correlated with tumors and led to bad prognosis [146]. Epigenetic modifications may also have a role in increasing CVD risk [147].

New opportunities will be provided by epigenome-wide association studies (EWASs), however, challenges will also be presented [148].

#### Metabolomics

Metabolic disturbances can lead to increased morbidity of common diseases [149]. GWAS could identify the differences between unhealthy and healthy patients.

Quinones and Kaddurah-Daouk identified that metabolic pathways such as fatty acids, oxidative stress and mitochondrial function were highly associated with central nervous system (CNS) disorders [150].

Metabolomics is a new strategy to study metabolites in urine, blood and tissue and it could be affected by the microbiome [151]. Metabolic differences between healthy and unhealthy people can lead to a systematic understanding of the network of common diseases.

### BIOINFORMATICS: COMPUTING SYSTEMS BIOMEDICINE

Recent years saw rapid advances in genome sequencing technology, which boasted the researches in identification of genes that potentially lead to such “complex diseases” as hypertension, cardiovascular disease, asthma, and cancer [152,153]. These complex diseases were proposed to be caused by a combination of multiple genes and a multitude of environmental factors. Genetically, their complexity lies in multiple aspects related to genetic polymorphisms (e.g., single-nucleotide polymorphisms or SNPs), gene expression, biological pathway, epigenetic modifications, noncoding RNA and gene-environment interactions. Nowadays, by adopting high-throughput methods, scientists can obtain much information about SNP genetic spectrum, gene expression profile, and protein spectrum. Accordingly, bioinformatics methods, tools, and databases have been developed and used to aid in the researches to understand these complex diseases at the systemic level.

Researches in this field are focused on the following questions: (1) How can genetic susceptibility markers be identified? (2) How can relationships between epigenetic markers and diseases be identified? (3) How can chromatin modification, RNA expression and protein expression data be combined to analyze the mechanisms of disease? (4) How can functional modules and biological pathways be elucidated? (5) How can all the data in a whole intact system, such as a cell, an organ, or a whole body be integrated? (6) What databases are available to guide our research in systems biomedicine?

#### 1. How can genetic susceptibility markers be identified?

Since the completion of the HapMap project, SNP markers have been widely applied, and GWASs have been adopted to test the association between SNPs and a specific disease. A GWAS research often involves 300000 or more SNP markers, evenly spreading across a genome [154]. GWAS involves powerful statistical test to locate many common disease genes by constructing and analyzing high-density SNP maps. By performing the WTCCC study [9] and German MI (Myocardial Infarction Family) Study [155–157], Samani et al. identified SNPs with the strongest associations with coronary artery disease [24]. Lu et al. applied meta-analysis to identify susceptibility loci related to coronary artery disease in the Han Chinese, and identified four new loci related to coronary artery disease [33]. These findings provide new insights into pathways contributing to the susceptibility for coronary artery disease in the Han Chinese population.

#### 2. How can relationships between epigenetic markers and diseases be identified?

Gene expression must take place accurately at the right time and place. Normal phenotypic changes of DNA and histones regulate the expression of gene function. Systematic study of the important roles of DNA methylation in the human epigenome, embryonic development, gene imprinting, allele inactivation and tumorigenesis has become a new research hot spot.

DNA methylation is a crucial epigenetic modification of the genome. DNA methylation and RNA-Seq data have implied the correlations between DNA methylation and transcriptional levels in both the kidney and the liver [158]. Robertson et al. showed that many human diseases are associated with aberrant DNA methylation patterns [90]. Yi et al. described the relationship between tumor occurrence and the abnormal methylation [159]. Uhlmann et al. found that tumor marker genes of different pathological types of glioma cells have different levels of methylation [160]. Recently, a large number of high-throughput data sets of methylation and histone modifica-

tion in normal tissues and diseases have been available thanks to the usage of next-generation sequencing technologies, and were collected in databases, such as e.g., the Cancer Genome Atlas (TCGA) [161], Human Histone Modification Database (HHMD) [162], Disease-Meth [163] and MethDB [164–167]. Accordingly, novel computational tools, including Batman [168], meth-BLAST [169], MethTools [170] and QUMA [171] to perform methylation analysis have been developed [172]. Notably, a comprehensive tool—CpG\_MPs has been applied to identify and analyze the methylation patterns of genomic regions from bisulfite sequencing data [173], whereas QDMR (quantitative differentially-methylated regions) implemented a new means of data analysis to identify DMRs (differentially-methylated regions) from genome-wide methylation profiles by adapting Shannon entropy theory [174].

### **3. How are chromatin modification, RNA expression and protein expression data combined to analyze mechanisms of disease?**

DNA microarray technology has offered the possibility to monitor the activities of thousands of genes simultaneously. To identify disease-risk biomarkers, multiple data mining approaches have been applied to analyze gene expression profile. According to the models of learning algorithms, current analytical strategies can be classified into two groups: unsupervised and supervised learning methods [175].

Unsupervised learning methods, such as K-means clustering, hierarchical clustering and principal component analysis method, can be used for gene clustering or sample clustering. For example, gene clustering is often performed to highlight some functionally related gene groups and predict the function of those genes [176], while sample clustering is performed to find different disease subtypes [177]. In 2000, based on the hierarchical clustering algorithm, Alizadeh et al. identified two molecularly distinct spliceosomes in diffuse large B cell lymphoma tumor patients using DNA microarray data [178].

Supervised learning methods, including linear discrimination, decision trees, artificial neural networks and support vector machines (SVM), can be used for the identification of the characteristics of genes and sites [179]. In 2004, Li, X. et al. developed a novel tree-based ensemble decision approach, analyzed two publicly available data sets [175] and identified 20 genes which are highly significant as related to colon cancer and 23 genes which are molecular signatures of the acute leukemia phenotype. Furthermore, in 2012, by adopting learning algorithms, Chen et al. combined genomic, transcriptomic, proteomic, metabolomic, and autoanti-

body profiles over a 14-month period from a single individual, and presented an integrative personal omics profile (iPOP), and revealed various disease risks [180]. With further accumulation of omics profiles from larger number of individuals, more sophisticated supervised learning methods are needed to aid in our understanding of the mechanisms of disease.

### **4. How can functional modules and biological pathways be elucidated?**

The development of complex diseases is initiated by multiple etiologies. Analysis of modules and pathways reflects the trend of system biology, which is about integration and interpretation beyond individual genes. However, it is a challenging task to identify common functional modules and biological pathways associated with complex diseases.

Statistical methods and software implementing them have been developed to predict gene functions and have accelerated the study of genes and their products for clinical purposes. Based on iPOP-analysis, integrating clustering analysis and pathway enrichment, Chen et al. analyzed various omics sets, identified differentially expressed components and elucidated several important pathways [180]. By analyzing the contribution of genetic factors and biological network of pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG), Chen et al. proposed an approach to prioritize risk pathways by fusing SNPs and pathways for common diseases. This approach was applied to five common diseases, and the result revealed that the five diseases not only share common risk pathways, but also have their own specific risk pathways [181].

### **5. How can all the data in a whole intact system, such as a cell, an organ, or a whole body be integrated?**

To explore human genetic disorders and the relationship among multiple genes at a higher level of cellular and organismal organization, Goh et al. constructed the Human Disease Network and Disease Gene Network [65]. To explore regulatory network models, Califano et al. proposed an integrative approach requiring the simultaneous reconstruction of context-specific gene regulatory network based GWAS data set [182]. By integrating the disease states of miRNAs, Li et al. constructed a network of bipartite miRNAs and sub-pathways [183]. Ye et al. constructed an miRNA-TF co-regulatory network specifically for T-ALL(T-ALL), inferred some hub regulators, genes, and their regulation in the network, and identified important miRNAs and regulatory modules in T cell acute lymphoblastic leukemia [184].



In summary, bioinformatic studies have begun to dig into the systematic biomedicine and get some promising results. With the help of bioinformatic approaches and tools, we can establish biological models and construct the biological control network, and better understand context-specific nature of biological process regulation, thus to gain a new insight into normal cell physiology and its dysregulation in disease. Biological system is a dynamic system, and with the development of systems biomedicine, we can observe the dynamics of the biological systems further. We assume that reconstruction of realistic biological model may be one of the most critical challenges of quantitative biology.

## 6. What databases are available to guide our research in systems biomedicine?

Up to now, numerous biological databases have been developed, which may help researchers in systematic biomedicine. Here, we introduce some databases related to complex diseases.

**OMIM** (Online Mendelian Inheritance in Man). OMIM is the most authoritative database of human genetic diseases. It classifies the genetic diseases and provides creditable and detailed information about disease heredity and related disease gene loci. (<http://www.ncbi.nlm.nih.gov/omim>, <http://omim.org/>).

**GAD** (Genetic Association Database). GAD contains a number of gene and polymorphism information of human complex diseases. This information comes from the research and arrangement of previous association analysis, which is convenient for researchers to quickly identify polymorphism of diseases from many multiple data. Moreover, this database allows the users to review submission records. (<http://geneticassociationdb.nih.gov/>).

**CGAP** (Cancer Genome Anatomy Project). The aim of CGAP is to produce the information by unscrambling the molecular structure and establish a series of analysis technologies to dig tumor-related genes, proteins, and other biological markers. Moreover, it provides information resources and technological methods for the study of tumors. This database contains 7 related modules to share data, bioinformatics analytical tools and biological related resources. (<http://cgap.nci.nih.gov/>).

**GeneCards**. GeneCards is an integrated database of human genes, which provides genomic, proteomic, transcriptomic, and functional information of known and predicted human genes. This database focuses on diseases, mutations and SNPs, gene expression, gene functions, pathways and protein interactions. This database particularly emphasizes the overall information, but there is less details about human diseases. It is a powerful functional genome data and contains external

linkage for related databases, which can provide chromosomal localization, gene expression information, homologous genes and corresponding proteins for human diseases. (<http://www.genecards.org/>).

**KEGG DISEASE**. The KEGG DISEASE is an online database to store the information about genes, pathways, drugs and diagnosis markers of diseases. It includes information about genetic and environmental perturbations. In this database, every disease has a certain H number for entry. According to the list of known genetic factors, environmental factors, diagnostic markers, or therapeutic drugs, diseases are being organized in KEGG DISEASE. (<http://www.genome.jp/kegg/disease/>).

**miR2Disease**. miR2Disease is a manually built database that provides information about miRNA deregulation in various human diseases. It is convenient for users to get the detailed information on an miRNA-disease relationship by each entry (<http://www.mir2disease.org/>) [185].

**HEP** (Human Epigenome Project-Data). The goals of HEP are to define, record, and explain the DNA methylation patterns of the whole human genome in the major tissue. To date, investigators can freely obtain DNA methylation spectral data of human 6, 20, and 22 chromosomes. (<http://www.epigenome.org/>).

**HHMD** (Human Histone Modification Database). HHMD is the most comprehensive and systematic database, which is based on the various experiments about human genome histone modifications. It contains the high-throughput experimental data of 43 human genome histone modifications, and it also provides the histone modifications information of 9 cancers obtained by literatures. (<http://bioinfo.hrbmu.edu.cn/hhmd>).

**MethyCancer**. This database is a multiple information database including DNA methylation, mutations, cancer-related gene, cancer information and CpG island. The aim of the database is to study the interaction among the DNA methylation, gene expression and cancer. (<http://methy-cancer.psych.ac.cn/>).

## INTEGRATION: RESEARCH FROM SYSTEMIC VIEW

As mentioned above, large quantities of omics data are accumulating. The final aim of systems biomedicine is to integrate all these components and to reconstruct a system. Understanding how complex phenotypes originate from a certain system is a difficult problem. Fortunately, some researches into integration have emerged in recent years.

One report utilized a computational model to simulate the process of cell divisions of the human pathogen *Mycoplasma genitalium* [186]. It first defined 4 types of substances (metabolites, RNA, protein, and DNA) and 28

submodels. Multiple external variables affected a cell, including the geometry, the mass of the cell, the stimuli that the cell accepted, and the type of host that the cell lived in. All variables influenced 4 types of substances and 28 submodels of the cell, and then determined whether the cell was to divide. This study demonstrated the application of systematic and quantitative methods to predict a cell's status.

Integrative personal omics profile (iPOP) used multiple omics methods, including genomics, transcriptomics, proteomics, and metabolomics to study a single individual [180]. It analyzed risk factors for some diseases, such as diabetes and coronary artery disease. It also analyzed profile transcription, protein, and RNA editing at the systemic level at different states. By comparing changes between healthy and diseased states, it revealed systemic signatures associated with disease. This study is a good example for the application of systems biomedicine in the human body.

## DEVELOPMENT IN CHINA AND PROSPECT OF SYSTEMS BIOMEDICINE

Systems biomedicine is an emerging field. In China, systems biomedicine is developing rapidly in recent years. Some centers for systems biomedicine have been built. Specific examples for systems biomedicine include Peking University Institute of Systems Biomedicine and Shanghai Center for Systems Biomedicine. These centers are pursuing interdisciplinary research, focusing on basic research from the systematic view, and encouraging applications in clinical medicine.

In addition to the rapid development of systems biomedicine in recent years, the idea of traditional Chinese medicine (TCM), which has a long history, resembles systems biomedicine in many aspects. TCM focuses on health maintenance, diagnosis and treatment of diseases on a systemic level, whereas systems biomedicine aims to integrate individual molecules and their interactions to understand how complex phenotypes arise. Systems biomedicine bridges TCM and western medicine. In the future, the systems approach will combine principles of TCM with that of western medicine, and pave the way for predictive, preventive and personalized medical practice.

Systems biomedicine will greatly facilitate our understanding of disease pathogenesis in the future. It helps to discover the disease mechanism and helps to prevent, diagnose and treat diseases in clinics. As we know, diseases are caused by multiple genetic and environment factors. For a certain person, disease is caused by some certain factors. Knowing the detailed information about one's genetic background and environment factors from the systematic view, we could evaluate the disease

susceptibility, offer the proper prevention, and use exact treatment. The development of systems biomedicine makes it possible to realize "3P" medicine, that is, predictive, preventive and personalized medicine.

## Conflict of interest

The authors Zhuqin Zhang, Zhiguo Zhao, Bing Liu, Dongguo Li, Dandan Zhang, Houzao Chen and Depei Liu declare that they have no conflict of interests.

## REFERENCES

1. Ideker, T., Galitski, T. and Hood, L. (2001) A new approach to decoding life: systems biology. *Annu. Rev. Genomics. Hum. Genet.*, 2, 343–372.
2. Kamada, T. (1992) System biomedicine: a new paradigm in biomedical engineering. *Front. Med. Biol. Eng.*, 4, 1–2.
3. Zeng, B. Z. (1992) On the holographic model of human body. in 1st National Conference of Comparative Studies Traditional Chinese Medicine and West Medicine (Medicine and Philosophy, Guangzhou, April 1992).
4. National Human Genome Research Institute (2010–11–08). FAQ About Genetic and Genomic Science. *Genome.gov*. Retrieved 2011–12–03.
5. Pearson, T. A. and Manolio, T. A. (2008) How to interpret a genome-wide association study. *JAMA*, 299, 1335–1344.
6. Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science*, 273, 1516–1517.
7. Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., et al. (2005) Complement factor H polymorphism in age-related macular degeneration. *Science*, 308, 385–389.
8. Dewan, A., Liu, M., Hartman, S., Zhang, S. S., Liu, D. T., Zhao, C., Tam, P. O., Chan, W. M., Lam, D. S., Snyder, M., et al. (2006) HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science*, 314, 989–992.
9. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447, 661–678.
10. Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V. G., Chen, W., Usala, G., Busonero, F., Maschio, A., Albai, G., et al. (2008) Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc. Natl. Acad. Sci. U.S.A.*, 105, 1620–1625.
11. Lettre, G., Sankaran, V. G., Bezerra, M. A., Araújo, A. S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., Costa, F. F., Hirschhorn, J. N., et al. (2008) DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc. Natl. Acad. Sci. U.S.A.*, 105, 11869–11874.
12. Sankaran, V. G., Menne, T. F., Xu, J., Akie, T. E., Lettre, G., Van Handel, B., Mikkola, H. K., Hirschhorn, J. N., Cantor, A. B. and Orkin, S. H. (2008) Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science*, 322, 1839–1842.
13. Sankaran, V. G., Xu, J., Ragoczy, T., Ippolito, G. C., Walkley, C. R.,

- Maika, S. D., Fujiwara, Y., Ito, M., Groudine, M., Bender, M. A., et al. (2009) Developmental and species-divergent globin switching are driven by BCL11A. *Nature*, 460, 1093–1097.
14. Ng, S. B., Nickerson, D. A., Bamshad, M. J. and Shendure, J. (2010) Massively parallel sequencing and rare disease. *Hum. Mol. Genet.*, 19, R119–R124.
  15. Gilissen, C., Hoischen, A., Brunner, H. G. and Veltman, J. A. (2011) Unlocking Mendelian disease using exome sequencing. *Genome Biol.*, 12, 228.
  16. Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A. and Shendure, J. (2011) Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.*, 12, 745–755.
  17. Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.*, 42, 30–35.
  18. Hoischen, A., van Bon, B. W., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., de Vries, P., de Reuver, R., Wieskamp, N., Mortier, G., et al. (2010) De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.*, 42, 483–485.
  19. Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., et al. (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.*, 42, 790–793.
  20. Hoischen, A., van Bon, B. W., Rodríguez-Santiago, B., Gilissen, C., Vissers, L. E., de Vries, P., Janssen, I., van Lier, B., Hastings, R., Smithson, S. F., et al. (2011) De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat. Genet.*, 43, 729–731.
  21. Kathiresan, S., Voight, B. F., Purcell, S., Musunuru, K., Ardissino, D., Mannucci, P. M., Anand, S., Engert, J. C., Samani, N. J., Schunkert, H., et al., and the Myocardial Infarction Genetics Consortium, and the Wellcome Trust Case Control Consortium. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.*, 41, 334–341.
  22. McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., Hinds, D. A., Pennacchio, L. A., Tybjaerg-Hansen, A., Folsom, A. R., et al. (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science*, 316, 1488–1491.
  23. O'Donnell, C. J. and Nabel, E. G. (2011) Genomics of cardiovascular disease. *N. Engl. J. Med.*, 365, 2098–2109.
  24. Samani, N. J., Erdmann, J., Hall, A. S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R. J., Meitinger, T., Braund, P., Wichmann, H. E., et al., and the WTCCC and the Cardiogenics Consortium. (2007) Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.*, 357, 443–453.
  25. Coronary Artery Disease (C4D) Genetics Consortium. (2011) A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.*, 43, 339–344.
  26. Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., Preuss, M., Stewart, A. F., Barbalic, M., Gieger, C., et al., and the Cardiogenics, and the CARDIoGRAM Consortium. (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.*, 43, 333–338.
  27. Takeuchi, F., Yokota, M., Yamamoto, K., Nakashima, E., Katsuya, T., Asano, H., Isono, M., Nabika, T., Sugiyama, T., Fujioka, A., et al. (2012) Genome-wide association study of coronary artery disease in the Japanese. *Eur. J. Hum. Genet.*, 20, 333–340.
  28. Wild, P. S., Zeller, T., Schillert, A., Szymczak, S., Sinning, C. R., Deiseroth, A., Schnabel, R. B., Lubos, E., Keller, T., Eleftheriadis, M. S., et al. (2011) A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease. *Circ. Cardiovasc. Genet.*, 4, 403–412.
  29. Slavin, T. P., Feng, T., Schnell, A., Zhu, X. and Elston, R. C. (2011) Two-marker association tests yield new disease associations for coronary artery disease and hypertension. *Hum. Genet.*, 130, 725–733.
  30. Jarinova, O., Stewart, A. F., Roberts, R., Wells, G., Lau, P., Naing, T., Buerki, C., McLean, B. W., Cook, R. C., Parker, J. S., et al. (2009) Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler. Thromb. Vasc. Biol.*, 29, 1671–1677.
  31. Visel, A., Zhu, Y., May, D., Afzal, V., Gong, E., Attanasio, C., Blow, M. J., Cohen, J. C., Rubin, E. M. and Pennacchio, L. A. (2010) Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature*, 464, 409–412.
  32. Harismendy, O., Notani, D., Song, X., Rahim, N. G., Tanasa, B., Heintzman, N., Ren, B., Fu, X. D., Topol, E. J., Rosenfeld, M. G., et al. (2011) 9p21 DNA variants associated with coronary artery disease impair interferon- $\gamma$  signalling response. *Nature*, 470, 264–268.
  33. Lu, X., Wang, L., Chen, S., He, L., Yang, X., Shi, Y., Cheng, J., Zhang, L., Gu, C. C., Huang, J., et al., and the Coronary ARtery Disease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Consortium. (2012) Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat. Genet.*, 44, 890–894.
  34. Levy, D., Larson, M. G., Benjamin, E. J., Newton-Cheh, C., Wang, T. J., Hwang, S. J., Vasan, R. S. and Mitchell, G. F. (2007) Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC Med. Genet.*, 8, S3.
  35. Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M. D., Bochud, M., Coin, L., Najjar, S. S., Zhao, J. H., Heath, S. C., Eyheramendy, S., et al., and the Wellcome Trust Case Control Consortium. (2009) Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.*, 41, 666–676.
  36. Levy, D., Ehret, G. B., Rice, K., Verwoert, G. C., Launer, L. J., Dehghan, A., Glazer, N. L., Morrison, A. C., Johnson, A. D., Aspelund, T., et al. (2009) Genome-wide association study of blood pressure and hypertension. *Nat. Genet.*, 41, 677–687.
  37. Wain, L. V., Verwoert, G. C., O'Reilly, P. F., Shi, G., Johnson, T., Johnson, A. D., Bochud, M., Rice, K. M., Henneman, P., Smith, A. V., et al., and the LifeLines Cohort Study, and the EchoGen consortium, and the AortaGen Consortium, and the CHARGE Consortium Heart Failure Working Group, and the KidneyGen consortium, and the CKDGen consortium, and the Cardiogenics consortium, and the CardioGram. (2011) Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.*, 43, 1005–1011.
  38. Cho, Y. S., Go, M. J., Kim, Y. J., Heo, J. Y., Oh, J. H., Ban, H. J., Yoon, D., Lee, M. H., Kim, D. J., Park, M., et al. (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.*, 41, 527–534.
  39. Delles, C. and Padmanabhan, S. (2012) Genetics and hypertension: is it time to change my practice? *Can. J. Cardiol.*, 28, 296–304.
  40. Schumacher, F. R., Berndt, S. I., Siddiq, A., Jacobs, K. B., Wang, Z., Lindstrom, S., Stevens, V. L., Chen, C., Mondul, A. M., Travis, R. C.,



- et al. (2011) Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum. Mol. Genet.*, 20, 3867–3875.
41. Sanson, M., Hosking, F. J., Shete, S., Zelenika, D., Dobbins, S. E., Ma, Y., Enciso-Mora, V., Idubai, A., Delattre, J. Y., Hoang-Xuan, K., et al. (2011) Chromosome 7p11.2 (EGFR) variation influences glioma risk. *Hum. Mol. Genet.*, 20, 2897–2904.
  42. Fletcher, O., Johnson, N., Orr, N., Hosking, F. J., Gibson, L. J., Walker, K., Zelenika, D., Gut, I., Heath, S., Palles, C., et al. (2011) Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J. Natl. Cancer Inst.*, 103, 425–435.
  43. Cui, R., Okada, Y., Jang, S. G., Ku, J. L., Park, J. G., Kamatani, Y., Hosono, N., Tsunoda, T., Kumar, V., Tanikawa, C., et al. (2011) Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut*, 60, 799–805.
  44. Rothman, N., Garcia-Closas, M., Chatterjee, N., Malats, N., Wu, X., Figueroa, J. D., Real, F. X., Van Den Berg, D., Matullo, G., Baris, D., et al. (2010) A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat. Genet.*, 42, 978–984.
  45. Goode, E. L., Chenevix-Trench, G., Song, H., Ramus, S. J., Notaridou, M., Lawrenson, K., Widschwendter, M., Vierkant, R. A., Larson, M. C., Kjaer, S. K., et al., and the Wellcome Trust Case-Control Consortium, and the Australian Cancer Study (Ovarian Cancer), and the Australian Ovarian Cancer Study Group, and the Ovarian Cancer Association Consortium (OCAC), and the Ovarian Cancer Association Consortium (OCAC). (2010) A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat. Genet.*, 42, 874–879.
  46. Low, S. K., Kuchiba, A., Zembutsu, H., Saito, A., Takahashi, A., Kubo, M., Daigo, Y., Kamatani, N., Chiku, S., Totsuka, H., et al. (2010) Genome-wide association study of pancreatic cancer in Japanese population. *PLoS ONE*, 5, e11824.
  47. Weedon, M. N., Clark, V. J., Qian, Y., Ben-Shlomo, Y., Timpson, N., Ebrahim, S., Lawlor, D. A., Pembrey, M. E., Ring, S., Wilkin, T. J., et al. (2006) A common haplotype of the glucokinase gene alters fasting glucose and birth weight: association in six studies and population-genetics analyses. *Am. J. Hum. Genet.*, 79, 991–1001.
  48. McCarthy, M. I. (2010) Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.*, 363, 2339–2350.
  49. Kooner, J. S., Saleheen, D., Sim, X., Sehmi, J., Zhang, W., Frossard, P., Been, L. F., Chia, K. S., Dimas, A. S., Hassanali, N., et al., and the DIAGRAM, and the MuTHER. (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat. Genet.*, 43, 984–989.
  50. Yamauchi, T., Hara, K., Maeda, S., Yasuda, K., Takahashi, A., Horikoshi, M., Nakamura, M., Fujita, H., Grarup, N., Cauchi, S., et al. (2010) A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat. Genet.*, 42, 864–868.
  51. Shu, X. O., Long, J., Cai, Q., Qi, L., Xiang, Y. B., Cho, Y. S., Tai, E. S., Li, X., Lin, X., Chow, W. H., et al. (2010) Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet.*, 6.
  52. Visscher, P. M., Brown, M. A., McCarthy, M. I. and Yang, J. (2012) Five years of GWAS discovery. *Am. J. Hum. Genet.*, 90, 7–24.
  53. National Human Genome Research Institute. A Catalog of Published Genome-Wide Association Studies. <http://www.genome.gov/gwastudies>.
  54. The Pharmacogenomics Knowledge Base. <http://www.pharmgkb.org>.
  55. Literature analysis by R. Ranganathan and C. Woods.
  56. Collins, F. S. (2011) Reengineering translational science: the time is right. *Sci. Transl. Med.*, 3, 90cm17.
  57. The ENCODE Project Consortium. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, 306, 636–640.
  58. Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., et al., and the ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57–74.
  59. Patti, G. J., Yanes, O. and Siuzdak, G. (2012) Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.*, 13, 263–269.
  60. Blackstock, W. P. and Weir, M. P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.*, 17, 121–127.
  61. Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, 340, 245–246.
  62. Yu, H., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., et al. (2008) High-quality binary protein interaction map of the yeast interactome network. *Science*, 322, 104–110.
  63. Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L., Toufighi, K., Mostafavi, S., et al. (2010) The genetic landscape of a cell. *Science*, 327, 425–431.
  64. Gandhi, T. K., Zhong, J., Mathivanan, S., Karthick, L., Chandrika, K. N., Mohan, S. S., Sharma, S., Pinkert, S., Nagaraju, S., Periaswamy, B., et al. (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat. Genet.*, 38, 285–293.
  65. Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M. and Barabási, A. L. (2007) The human disease network. *Proc. Natl. Acad. Sci. U.S.A.*, 104, 8685–8690.
  66. Oti, M., Snel, B., Huynen, M. A. and Brunner, H. G. (2006) Predicting disease genes using protein-protein interactions. *J. Med. Genet.*, 43, 691–698.
  67. Xu, J. and Li, Y. (2006) Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics*, 22, 2800–2805.
  68. Wenk, M. R. (2010) Lipidomics: new tools and applications. *Cell*, 143, 888–895.
  69. Thomas, A., Lenglet, S., Chaurand, P., Deglon, J., Mangin, P., Mach, F., Steffens, S., Wolfender, J. L. and Staub, C. (2011) Mass spectrometry for the evaluation of cardiovascular diseases based on proteomics and lipidomics. *Thromb. Haemost.*, 106, 20–33.
  70. Shevchenko, A. and Simons, K. (2010) Lipidomics: coming to grips with lipid diversity. *Nat. Rev. Mol. Cell Biol.*, 11, 593–598.
  71. Quehenberger, O., Armando, A. M., Brown, A. H., Milne, S. B., Myers, D. S., Merrill, A. H., Bandyopadhyay, S., Jones, K. N., Kelly, S., Shaner, R. L., et al. (2010) Lipidomics reveals a remarkable diversity of lipids in human plasma. *J. Lipid Res.*, 51, 3299–3305.
  72. Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, N. and Stokes, J. 3rd. (1961) Factors of risk in the development of coronary heart disease—six year follow-up experience. The Framingham Study. *Ann. Intern. Med.*, 55, 33–50.
  73. Pilia, G., Chen, W. M., Scuteri, A., Orrù, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P., et al. (2006) Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.*, 2, e132.
  74. Stegmann, C., Drozdov, I., Shalhoub, J., Humphries, J., Ladroue, C.,



- Didangelos, A., Baumert, M., Allen, M., Davies, A. H., Monaco, C., et al. (2011) Comparative lipidomics profiling of human atherosclerotic plaques. *Circ. Cardiovasc. Genet.*, 4, 232–242.
75. Meikle, P. J., Wong, G., Tsorotes, D., Barlow, C. K., Weir, J. M., Christopher, M. J., MacIntosh, G. L., Goudey, B., Stern, L., Kowalczyk, A., et al. (2011) Plasma lipidomic analysis of stable and unstable coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.*, 31, 2723–2732.
  76. Yetukuri, L., Katajamaa, M., Medina-Gomez, G., Seppänen-Laakso, T., Vidal-Puig, A. and Oresic, M. (2007) Bioinformatics strategies for lipidomics analysis: characterization of obesity related hepatic steatosis. *BMC Syst. Biol.*, 1, 12.
  77. Kaddurah-Daouk, R., Baillie, R. A., Zhu, H., Zeng, Z. B., Wiest, M. M., Nguyen, U. T., Watkins, S. M. and Krauss, R. M. (2010) Lipidomic analysis of variation in response to simvastatin in the Cholesterol and Pharmacogenetics Study. *Metabolomics*, 6, 191–201.
  78. Cobbold, J. F., Patel, J. H., Goldin, R. D., North, B. V., Crossey, M. M., Fitzpatrick, J., Wylezinska, M., Thomas, H. C., Cox, I. J. and Taylor-Robinson, S. D. (2010) Hepatic lipid profiling in chronic hepatitis C: an *in vitro* and *in vivo* proton magnetic resonance spectroscopy study. *J. Hepatol.*, 52, 16–24.
  79. Wu, H., Volponi, J. V., Oliver, A. E., Parikh, A. N., Simmons, B. A. and Singh, S. (2011) *In vivo* lipidomics using single-cell Raman spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 108, 3809–3814.
  80. Bertozzi, C. R. and Sasisekharan, R. (2009) Glycomics. In: Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., (eds). *Essentials of Glycobiology*. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, Chapter 48.
  81. Hart, G. W. and Copeland, R. J. (2010) Glycomics hits the big time. *Cell*, 143, 672–676.
  82. Villoslada, P. and Baranzini, S. (2012) Data integration and systems biology approaches for biomarker discovery: challenges and opportunities for multiple sclerosis. *J. Neuroimmunol.*, 248, 58–65.
  83. An, H. J., Kronewitter, S. R., de Leoz, M. L. and Lebrilla, C. B. (2009) Glycomics and disease markers. *Curr. Opin. Chem. Biol.*, 13, 601–607.
  84. Angata, T., Fujinawa, R., Kurimoto, A., Nakajima, K., Kato, M., Takamatsu, S., Korekane, H., Gao, C. X., Ohtsubo, K., Kitazume, S., et al. (2012) Integrated approach toward the discovery of glyco-biomarkers of inflammation-related diseases. *Ann. N. Y. Acad. Sci.*, 1253, 159–169.
  85. Callewaert, N., Van Vlierberghe, H., Van Hecke, A., Laroy, W., Delanghe, J. and Contreras, R. (2004) Noninvasive diagnosis of liver cirrhosis using DNA sequencer-based total serum protein glycomics. *Nat. Med.*, 10, 429–434.
  86. Taylor, A. D., Hancock, W. S., Hincapie, M., Taniguchi, N. and Hanash, S. M. (2009) Towards an integrated proteomic and glycomic approach to finding cancer biomarkers. *Genome Med.*, 1, 57.
  87. Maunakea, A. K., Chepelev, I. and Zhao, K. (2010) Epigenome mapping in normal and disease States. *Circ. Res.*, 107, 327–339.
  88. American Association for Cancer Research Human Epigenome Task Force, and the European Union, Network of Excellence, Scientific Advisory Board. (2008) Moving AHEAD with an international human epigenome project. *Nature*, 454, 711–715.
  89. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, 16, 6–21.
  90. Robertson, K. D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, 6, 597–610.
  91. Feinberg, A. P. and Tycko, B. (2004) The history of cancer epigenetics. *Nat. Rev. Cancer*, 4, 143–153.
  92. Widschwendter, M., Jiang, G., Woods, C., Müller, H. M., Fiegl, H., Goebel, G., Marth, C., Müller-Holzner, E., Zeimet, A. G., Laird, P. W., et al. (2004) DNA hypomethylation and ovarian cancer biology. *Cancer Res.*, 64, 4472–4480.
  93. Zhuang, J., Peng, W., Li, H., Wang, W., Wei, Y., Li, W. and Xu, Y. (2012) Methylation of p15INK4b and expression of ANRIL on chromosome 9p21 are associated with coronary artery disease. *PLoS ONE*, 7, e47193.
  94. Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Peng, W., Zhang, M. Q., et al. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, 40, 897–903.
  95. Wang, Y., Fischle, W., Cheung, W., Jacobs, S., Khorasanizadeh, S. and Allis, C. D. (2004) Beyond the double helix: writing and reading the histone code. *Novartis Found. Symp.*, 259, 3–17, (discussion 17–21, 163–169).
  96. Suh, Y. A., Lee, H. Y., Virmani, A., Wong, J., Mann, K. K., Miller, W. H. Jr, Gazdar, A. and Kurie, J. M. (2002) Loss of retinoic acid receptor beta gene expression is linked to aberrant histone H3 acetylation in lung cancer cell lines. *Cancer Res.*, 62, 3945–3949.
  97. Fraga, M. F. and Esteller, M. (2005) Towards the human cancer epigenome: a first draft of histone modifications. *Cell Cycle*, 4, 1377–1381.
  98. Zhang, X., Zhao, X., Fiskus, W., Lin, J., Lwin, T., Rao, R., Zhang, Y., Chan, J. C., Fu, K., Marquez, V. E., et al. (2012) Coordinated silencing of MYC-mediated miR-29 by HDAC3 and EZH2 as a therapeutic target of histone modification in aggressive B-Cell lymphomas. *Cancer Cell*, 22, 506–523.
  99. Park, J. H., Stoffers, D. A., Nicholls, R. D. and Simmons, R. A. (2008) Development of type 2 diabetes following intrauterine growth retardation in rats is associated with progressive epigenetic silencing of Pdx1. *J. Clin. Invest.*, 118, 2316–2324.
  100. Pinney, S. E. and Simmons, R. A. (2010) Epigenetic mechanisms in the development of type 2 diabetes. *Trends Endocrinol. Metab.*, 21, 223–229.
  101. Sommer, B., Köhler, M., Sprengel, R. and Seeburg, P. H. (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell*, 67, 11–19.
  102. Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E. and Jaffrey, S. R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, 149, 1635–1646.
  103. O'Neil, R. T. and Emeson, R. B. (2012) Quantitative analysis of 5HT (2C) receptor RNA editing patterns in psychiatric disorders. *Neurobiol. Dis.*, 45, 8–13.
  104. Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M. and Mattick, J. S. (2010) Non-coding RNAs: regulators of disease. *J. Pathol.*, 220, 126–139.
  105. Esteller, M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, 12, 861–874.
  106. Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M. J. and Zhou, M. M. (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a.

- Mol. Cell, 38, 662–674.
107. Faghihi, M. A., Modarresi, F., Khalil, A. M., Wood, D. E., Sahagan, B. G., Morgan, T. E., Finch, C. E., St Laurent, G. 3rd, Kenny, P. J. and Wahlestedt, C. (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat. Med.*, 14, 723–730.
  108. Rowles, J., Wong, M., Powers, R. and Olsen, M. (2012) FTO, RNA epigenetics and epilepsy. *Epigenetics*, 7, 1094–1097.
  109. Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W. and Cui, Q. (2008) An analysis of human microRNA and disease associations. *PLoS ONE*, 3, e3420.
  110. Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., et al. (2005) MicroRNA expression profiles classify human cancers. *Nature*, 435, 834–838.
  111. Li, C., Pei, F., Zhu, X., Duan, D. D. and Zeng, C. (2012) Circulating microRNAs as novel and sensitive biomarkers of acute myocardial infarction. *Clin. Biochem.*, 45, 727–732.
  112. Thum, T., Gross, C., Fiedler, J., Fischer, T., Kissler, S., Bussen, M., Galuppo, P., Just, S., Rottbauer, W., Frantz, S., et al. (2008) MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature*, 456, 980–984.
  113. Beltrao, P., Albanese, V., Kenner, L. R., Swaney, D. L., Burlingame, A., Villén, J., Lim, W. A., Fraser, J. S., Frydman, J. and Krogan, N. J. (2012) Systematic functional prioritization of protein posttranslational modifications. *Cell*, 150, 413–425.
  114. Grasbon-Frodl, E., Lorenz, H., Mann, U., Nitsch, R. M., Windl, O. and Kretschmar, H. A. (2004) Loss of glycosylation associated with the T183A mutation in human prion disease. *Acta Neuropathol.*, 108, 476–484.
  115. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature*, 290, 457–465.
  116. Mercer, T. R., Neph, S., Dinger, M. E., Crawford, J., Smith, M. A., Shearwood, A. M., Haugen, E., Bracken, C. P., Rackham, O., Stamatoyannopoulos, J. A., et al. (2011) The human mitochondrial transcriptome. *Cell*, 146, 645–658.
  117. Taylor, R. W. and Turnbull, D. M. (2005) Mitochondrial DNA mutations in human disease. *Nat. Rev. Genet.*, 6, 389–402.
  118. Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003) PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.*, 34, 267–273.
  119. Petersen, K. F., Befroy, D., Dufour, S., Dziura, J., Ariyan, C., Rothman, D. L., DiPietro, L., Cline, G. W. and Shulman, G. I. (2003) Mitochondrial dysfunction in the elderly: possible role in insulin resistance. *Science*, 300, 1140–1142.
  120. Sobenin, I. A., Sazonova, M. A., Ivanova, M. M., Zhelankin, A. V., Myasoedova, V. A., Postnov, A. Y., Nurbaev, S. D., Bobryshev, Y. V. and Orekhov, A. N. (2012) Mutation C3256T of mitochondrial genome in white blood cells: novel genetic marker of atherosclerosis and coronary heart disease. *PLoS ONE*, 7, e46573.
  121. Relman, D. A. (2012) Microbiology: learning about who we are. *Nature*, 486, 194–195.
  122. Human Microbiome Project Consortium. (2012) A framework for human microbiome research. *Nature*, 486, 215–221.
  123. Relman, D. A. (2012) The human microbiome: ecosystem resilience and health. *Nutr. Rev.*, 70, S2–S9.
  124. Ursell, L. K., Metcalf, J. L., Parfrey, L. W. and Knight, R. (2012) Defining the human microbiome. *Nutr. Rev.*, 70, S38–S44.
  125. McNiven, E. M., German, J. B. and Slupsky, C. M. (2011) Analytical metabolomics: nutritional opportunities for personalized health. *J. Nutr. Biochem.*, 22, 995–1002.
  126. Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., et al. (2012) Human gut microbiome viewed across age and geography. *Nature*, 486, 222–227.
  127. Human Microbiome Project Consortium. (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, 486, 207–214.
  128. Burcelin, R. (2012) Regulation of metabolism: a cross talk between gut microbiota and its human host. *Physiology (Bethesda)*, 27, 300–307.
  129. Coons, M. J., Demott, A., Buscemi, J., Duncan, J. M., Pellegrini, C. A., Steglitz, J., Pictor, A. and Spring, B. (2012) Technology interventions to curb obesity: a systematic review of the current literature. *Curr. Cardiovasc. Risk Rep.*, 6, 120–134.
  130. Kahrstrom, C. T. (2012) Microbiome: gut microbiome as a marker for diabetes. *Nat. Rev. Microbiol.*, 10, 733.
  131. Burcelin, R., Garidou, L. and Pomié, C. (2012) Immuno-microbiota cross and talk: the new paradigm of metabolic diseases. *Semin. Immunol.*, 24, 67–74.
  132. Greiner, T. and Bäckhed, F. (2011) Effects of the gut microbiota on obesity and glucose homeostasis. *Trends Endocrinol. Metab.*, 22, 117–123.
  133. Azad, M. B. and Kozyskyj, A. L. (2012) Perinatal programming of asthma: the role of gut microbiota. *Clin. Dev. Immunol.*, 2012, 932072.
  134. Wang, Z., Klipfell, E., Bennett, B. J., Koeth, R., Levison, B. S., Dugar, B., Feldstein, A. E., Britt, E. B., Fu, X., Chung, Y. M., et al. (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*, 472, 57–63.
  135. Harris, K., Kassis, A., Major, G. and Chou, C. J. (2012) Is the gut microbiota a new factor contributing to obesity and its metabolic disorders? *J. Obes.*, 2012, 879151.
  136. Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., et al. (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337, 1190–1195.
  137. Torkamani, A., Topol, E. J. and Schork, N. J. (2008) Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics*, 92, 265–272.
  138. Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., Pahwa, J. S., Moskvin, V., Dowzell, K., Williams, A., et al. (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.*, 41, 1088–1093.
  139. Chen, F., Chen, G. K., Stram, D. O., Millikan, R. C., Ambrosone, C. B., John, E. M., Bernstein, L., Zheng, W., Palmer, J. R., Hu, J. J., et al. (2013) A genome-wide association study of breast cancer in women of African ancestry. *Hum. Genet.*, 132, 39–48.
  140. Yeager, M., Chatterjee, N., Ciampa, J., Jacobs, K. B., Gonzalez-Bosquet, J., Hayes, R. B., Kraft, P., Wacholder, S., Orr, N., Berndt, S., et al. (2009) Identification of a new prostate cancer susceptibility locus

- on chromosome 8q24. *Nat. Genet.*, 41, 1055–1057.
141. Schnabel, R. B., Baccarelli, A., Lin, H., Ellinor, P. T. and Benjamin, E. J. (2012) Next steps in cardiovascular disease genomic research—sequencing, epigenetics, and transcriptomics. *Clin. Chem.*, 58, 113–126.
  142. Ndiaye, N. C., Azimi Nehzad, M., El Shamieh, S., Stathopoulou, M. G. and Visvikis-Siest, S. (2011) Cardiovascular diseases and genome-wide association studies. *Clin. Chim. Acta*, 412, 1697–1701.
  143. Rakyán, V. K., Down, T. A., Balding, D. J. and Beck, S. (2011) Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.*, 12, 529–541.
  144. Bell, C. G. and Beck, S. (2010) The epigenomic interface between genome and environment in common complex diseases. *Brief Funct. Genomics*, 9, 477–485.
  145. Toyota, M., Suzuki, H., Yamashita, T., Hirata, K., Imai, K., Tokino, T. and Shinomura, Y. (2009) Cancer epigenomics: implications of DNA methylation in personalized cancer therapy. *Cancer Sci.*, 100, 787–791.
  146. Ivanova, T., Zouridis, H., Wu, Y., Cheng, L. L., Tan, I. B., Gopalakrishnan, V., Ooi, C. H., Lee, J., Qin, L., Wu, J., et al. (2013) Integrated epigenomics identifies BMP4 as a modulator of cisplatin sensitivity in gastric cancer. *Gut*, 62, 22–33.
  147. Ordovás, J. M. and Smith, C. E. (2010) Epigenetics and cardiovascular disease. *Nat. Rev. Cardiol.*, 7, 510–519.
  148. Trojan, H. J. (1975) Eye complications in the long-term treatment with chloroquin. Study carried out in Togo. *Rev. Int. Trach. Pathol. Ocul. Trop. Subtrop.*, 52, 129–137.
  149. Adamski, J. (2012) Genome-wide association studies with metabolomics. *Genome Med.*, 4, 34.
  150. Quinones, M. P. and Kaddurah-Daouk, R. (2009) Metabolomics tools for identifying biomarkers for neuropsychiatric diseases. *Neurobiol. Dis.*, 35, 165–176.
  151. Madsen, R., Lundstedt, T. and Trygg, J. (2010) Chemometrics in metabolomics—a review in human disease diagnosis. *Anal. Chim. Acta*, 659, 23–33.
  152. Ashley, E. A., Butte, A. J., Wheeler, M. T., Chen, R., Klein, T. E., Dewey, F. E., Dudley, J. T., Ormond, K. E., Pavlovic, A., Morgan, A. A., et al. (2010) Clinical assessment incorporating a personal genome. *Lancet*, 375, 1525–1535.
  153. Grayson, B. L., Wang, L. and Aune, T. M. (2011) Peripheral blood gene expression profiles in metabolic syndrome, coronary artery disease and type 2 diabetes. *Genes Immun.*, 12, 341–351.
  154. Masood, E. (1999) As consortium plans free SNP map of human genome. *Nature*, 398, 545–546.
  155. Broeckel, U., Hengstenberg, C., Mayer, B., Holmer, S., Martin, L. J., Comuzzie, A. G., Blangero, J., Nürnberg, P., Reis, A., Riegger, G. A., et al. (2002) A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat. Genet.*, 30, 210–214.
  156. Fischer, M., Broeckel, U., Holmer, S., Baessler, A., Hengstenberg, C., Mayer, B., Erdmann, J., Klein, G., Riegger, G., Jacob, H. J., et al. (2005) Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction. *Circulation*, 111, 855–862.
  157. Wichmann, H. E., Gieger, C. and Illig, T., and the MONICA/KORA Study Group. (2005) KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*, 67, S26–S30.
  158. Xie, L., Weichel, B., Ohm, J. E. and Zhang, K. (2011) An integrative analysis of DNA methylation and RNA-Seq data for human heart, kidney and liver. *BMC Syst. Biol.*, 5, S4.
  159. Yi, J. M., Tsai, H. C., Glockner, S. C., Lin, S., Ohm, J. E., Easwaran, H., James, C. D., Costello, J. F., Riggins, G., et al. (2008) Abnormal DNA methylation of CD133 in colorectal and glioblastoma tumors. *Cancer Res.*, 68, 8094–8103.
  160. Uhlmann, K., Rohde, K., Zeller, C., Szymas, J., Vogel, S., Marczynek, K., Thiel, G., Nürnberg P. and Laird, P. W. (2003) Distinct methylation profiles of glioma subtypes. *Int. J. Cancer*, 106, 52–59.
  161. Cancer Genome Atlas Research Network. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455, 1061–1068.
  162. Zhang, Y., Lv, J., Liu, H., Zhu, J., Su, J., Wu, Q., Qi, Y., Wang, F. and Li, X. (2010) HHMD: the human histone modification database. *Nucleic Acids Res.*, 38, D149–D154.
  163. Lv, J., Liu, H., Su, J., Wu, X., Liu, H., Li, B., Xiao, X., Wang, F., Wu, Q. and Zhang, Y. (2012) DiseaseMeth: a human disease methylation database. *Nucleic Acids Res.*, 40, D1030–D1035.
  164. Grunau, C., Renault, E., Rosenthal, A. and Roizes, G. (2001) MethDB—a public database for DNA methylation data. *Nucleic Acids Res.*, 29, 270–274.
  165. Grunau, C., Renault, E. and Roizes, G. (2002) DNA Methylation Database “MethDB”: a user guide. *J. Nutr.*, 132, 2435S–2439S.
  166. Amoreira, C., Hindermann, W. and Grunau, C. (2003) An improved version of the DNA Methylation database (MethDB). *Nucleic Acids Res.*, 31, 75–77.
  167. Negre, V. and Grunau, C. (2006) The MethDB DAS server: adding an epigenetic information layer to the human genome. *Epigenetics*, 1, 101–105.
  168. Down, T. A., Rakyán, V. K., Turner, D. J., Flicek, P., Li, H., Kulesha, E., Gräf, S., Johnson, N., Herrero, J., Tomazou, E. M., et al. (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.*, 26, 779–785.
  169. Pattyn, F., Hoebeeck, J., Robbrecht, P., Michels, E., De Paepe, A., Bottu, G., Coornaert, D., Herzog, R., Speleman, F. and Vandesompele, J. (2006) methBLAST and methPrimerDB: web-tools for PCR based methylation analysis. *BMC Bioinformatics*, 7, 496.
  170. Grunau, C., Schattevoy, R., Mache, N. and Rosenthal, A. (2000) MethTools—a toolbox to visualize and analyze DNA methylation data. *Nucleic Acids Res.*, 28, 1053–1058.
  171. Kumaki, Y., Oda, M. and Okano, M. (2008) QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.*, 36, W170–5.
  172. Bock, C. and Lengauer, T. (2008) Computational epigenetics. *Bioinformatics*, 24, 1–10.
  173. Su, J., Yan, H., Wei, Y., Liu, H., Liu, H., Wang, F., Lv, J., Wu, Q. and Zhang, Y. (2013) CpG\_MPs: identification of CpG methylation patterns of genomic regions from high-throughput bisulfite sequencing data. *Nucleic Acids Res.*, 41, e4.
  174. Zhang, Y., Liu, H., Lv, J., Xiao, X., Zhu, J., Liu, X., Su, J., Li, X., Wu, Q., Wang, F., et al. (2011) QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic Acids Res.*, 39, e58.
  175. Li, X., Rao, S., Wang, Y. and Gong, B. (2004) Gene mining: a novel and powerful ensemble decision approach to hunting for disease genes using microarray expression profiling. *Nucleic Acids Res.*, 32, 2685–2694.
  176. Li, X., Rao, S., Jiang, W., Li, C., Xiao, Y., Guo, Z., Zhang, Q., Wang, L., Du, L., Li, J., et al. (2006) Discovery of Time-Delayed Gene

- Regulatory Networks based on temporal gene expression profiling. *BMC Bioinformatics*, 7, 26.
177. Zhou, X., Kao, M. C. and Wong, W. H. (2002) Transitive functional annotation by shortest-path analysis of gene expression data. *Proc. Natl. Acad. Sci. U.S.A.*, 99, 12783–12788.
  178. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403, 503–511.
  179. Butte, A. J., Ye, J., Häring, H. U., Stumvoll, M., White, M. F. and Kohane, I. S. (2001) Determining significant fold differences in gene expression analysis. *Pac. Symp. Biocomput.*, 6–17.
  180. Chen, R., Mias, G. I., Li-Pook-Than, J., Jiang, L., Lam, H. Y., Chen, R., Miriami, E., Karczewski, K. J., Hariharan, M., Dewey, F. E., et al. (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*, 148, 1293–1307.
  181. Chen, L., Zhang, L., Zhao, Y., Xu, L., Shang, Y., Wang, Q., Li, W., Wang, H. and Li, X. (2009) Prioritizing risk pathways: a novel association approach to searching for disease pathways fusing SNPs and pathways. *Bioinformatics*, 25, 237–242.
  182. Califano, A., Butte, A. J., Friend, S., Ideker, T. and Schadt, E. (2012) Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat. Genet.*, 44, 841–847.
  183. Li, X., Jiang, W., Li, W., Lian, B., Wang, S., Liao, M., Chen, X., Wang, Y., Lv, Y., Wang, S., et al. (2012) Dissection of human MiRNA regulatory influence to subpathway. *Brief. Bioinformatics*, 13, 175–186.
  184. Ye, H., Liu, X., Lv, M., Wu, Y., Kuang, S., Gong, J., Yuan, P., Zhong, Z., Li, Q., Jia, H., et al. (2012) MicroRNA and transcription factor co-regulatory network analysis reveals miR-19 inhibits CYLD in T-cell acute lymphoblastic leukemia. *Nucleic Acids Res.*, 40, 5201–5214.
  185. Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G. and Liu, Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, 37, D98–D104.
  186. Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B. Jr, Assad-Garcia, N., Glass, J. I. and Covert, M. W. (2012) A whole-cell computational model predicts phenotype from genotype. *Cell*, 150, 389–401.