Review

Systems medicine: the future of medical genomics and healthcare

Charles Auffray*, Zhu Chen† and Leroy Hood‡

Addresses: *Functional Genomics and Systems Biology for Health, LGN-UMR 7091, CNRS and Pierre & Marie Curie University of Paris VI, 7 rue Guy Moquet, 94801 Villejuif, France. †Center for Systems Biomedicine, Jiao-Tong University, 200240 Shanghai, China. ‡Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103-8904, USA.

Correspondence: Charles Auffray. Email: charles.auffray@vjf.cnrs.fr

Published: 20 January 2009

Genome Medicine 2009, 1:2 (doi:10.1186/gm2)

The electronic version of this article is the complete one and can be found online at http://genomemedicine.com/content/1/1/2

© 2009 BioMed Central Ltd

Abstract

High-throughput technologies for DNA sequencing and for analyses of transcriptomes, proteomes and metabolomes have provided the foundations for deciphering the structure, variation and function of the human genome and relating them to health and disease states. The increased efficiency of DNA sequencing opens up the possibility of analyzing a large number of individual genomes and transcriptomes, and complete reference proteomes and metabolomes are within reach using powerful analytical techniques based on chromatography, mass spectrometry and nuclear magnetic resonance. Computational and mathematical tools have enabled the development of systems approaches for deciphering the functional and regulatory networks underlying the behavior of complex biological systems. Further conceptual and methodological developments of these tools are needed for the integration of various data types across the multiple levels of organization and time frames that are characteristic of human development, physiology and disease. Medical genomics has attempted to overcome the initial limitations of genome-wide association studies and has identified a limited number of susceptibility loci for many complex and common diseases. Iterative systems approaches are starting to provide deeper insights into the mechanisms of human diseases, and to facilitate the development of better diagnostic and prognostic biomarkers for cancer and many other diseases. Systems approaches will transform the way drugs are developed through academy-industry partnerships that will target multiple components of networks and pathways perturbed in diseases. They will enable medicine to become predictive, personalized, preventive and participatory, and, in the process, concepts and methods from Western and oriental cultures can be combined. We recommend that systems medicine should be developed through an international network of systems biology and medicine centers dedicated to inter-disciplinary training and education, to help reduce the gap in healthcare between developed and developing countries.

Systems biology is developing rapidly. It is an integrative research strategy designed to tackle the complexity of biological systems and their behavior at all levels of organization (from molecules, cells and organs to organisms and ecosystems) in normal and perturbed conditions. It is based on an understanding of biological functions as system properties that are different from those of the individual interacting components (reviewed in [1-9]). It integrates the mass of data that has been collected with various global measurement technologies (techniques that look at the complete set of genes, proteins or other features in an organism), in order to formulate predictive mathematical and computational models of functional and regulatory biological networks. Specific biological hypotheses can thus be tested by designing a series of perturbation experiments. It thus combines data-driven (bottom-up) [10] and model-driven (top-down) [11] approaches into a question-driven (middle-out) inquiry in search of basic principles [12-14]. In the end,
systems approaches must be driven by high-quality hypothesis-driven biology and not just by data-accumulating technologies or high-performance computational modeling.

Although systems biology has ancient roots in physiology, biochemistry, and molecular and cellular biology, its current development is the result of recent advances in genomics and bioinformatics, which were made possible by the continuous development of high-throughput experimental and computational platforms. The field is also revisiting previous attempts at modeling biological complexity, by taking advantage of insights from system theory [15] and engineering sciences [16,17]. It forms the basis for an extension of genetic engineering into synthetic biology: designing and building biological systems with new properties from modular components [18-21].

Evolution, development, physiology and disease are viewed in systems biology as dynamic processes that operate on widely different scales in space and time between biological states that are constrained by inter-relationships among pathway and network components. In this context, detecting, understanding and treating disease translates into identifying and manipulating global perturbed networks rather than focusing only on unique failing components.

Here we review how medical genomics [22,23], based on recent advances in high-throughput experimental and computational technologies, is evolving in the context of systems biology into a more prospective systems medicine [24-31]. This new kind of medicine will be able to overcome the current limitations of disease complexity (through stratification of patients and diseases by molecular diagnostics) and drug discovery (through the analysis and targeting of disease-perturbed networks) [32-35]. We discuss some of the technological, conceptual and organizational challenges that we will face in implementing this new vision and practice of biology and medicine, and we argue that it offers new opportunities to more efficiently tackle key medical problems in both developed and developing countries.

**Technology is moving genomics from structure to function**

The initial sequencing of the human genome was made possible by the automation of the DNA sequencing chemistry and by the development of data-acquisition tools and software for the reliable interpretation and assembly of the DNA sequence. It required a multi-billion-dollar investment and the participation of thousands of researchers in the public and private sectors over more than a decade. It came together with the sequencing of genomes in a variety of microorganisms, animals and plants. All these efforts combined served as test cases to trigger sustained technology developments.

With the next-generation DNA mega-sequencing technologies currently available, which enable the collection of billions of nucleotides in single instrument runs [36], it is now possible to sequence and assemble a human genome in a matter of weeks at a small fraction of the cost of the reference genome [37]. With both incremental progress and the introduction of third-generation sequencing technologies, it may soon become possible to collect large numbers of individual genomes in days for US$1,000 or less and ascertain their unique variations. This opens up the possibility of a Personal Genome Project to find correlations between genotypes and normal or diseased phenotypes [38]. In parallel, over a period of more than 30 years, successive generations of increasingly miniaturized DNA arrays have been used for expression profiling, benefiting from the extensive sequencing of partial and complete cDNA collections. Microarray technology, because of its intrinsic complexity and that of the transcriptome, has reached an intermediate stage of maturity compared with sequencing [39]; it is possible to detect variations in expression of many but not all gene transcripts under normal and perturbed conditions.

Early on, insufficient attention was paid by users of current microarray platforms to proper design and quality assessment, which is needed to control for variation in the large number of biological and experimental parameters involved. This compromised the usefulness of these platforms, for example in the development of classification and predictive biomarkers [40-42]. The introduction of standards and guidelines for complete microarray workflows [43] is helping to rectify the problem; these need to cover all aspects, from RNA integrity assessment [44,45] to data analysis and reporting [46,47]. There has also been constant progress in the use of advanced statistical methods for multivariate classification [48] and for gene-set enrichment analysis [49] of expression profiles. At the same time, it has become clear through the combination of tiling arrays and systematic sequencing that a larger fraction of the human genome is transcribed into diverse types of RNA than was previously thought [50,51]. The increased power and reduced cost of deep sequencing thus means that it is starting to compete with high-density microarrays [52,53], to the extent that some believe this is marking the beginning of the end of microarrays [54].

However, given that each new generation of tools takes several years to mature, it is most likely that sequencing and microarrays will continue to coexist. Microarrays will probably be increasingly used for specialized applications, such as those related to transcription regulation, epigenetic modifications, and selection of subfractions of individual genomes for sequencing (for example, exons, highly conserved regions, and so on); whereas mega-sequencing will be used for deep exploration of transcriptomes. The results of transcriptome analyses will increasingly be validated by emerging technologies that
use miniaturized high-throughput reverse-transcriptase PCR [55] or multiplex direct visualization and counting of RNA molecules; the latter technology has the added advantage of avoiding biases resulting from reverse transcription [56].

From the genome and transcriptome sequences, it has been possible to derive a relatively complete parts list of genes and, by extension, of proteins, thus revolutionizing the field of proteomics. It is crucial to note that mass spectrometry is effective in the identification of peptides, and not of complete proteins. In order to identify and quantify interesting proteins by mass spectrometry, either through shotgun or directed approaches, an investigator therefore needs to know the sequence of the peptides obtained by enzymatic digestion of those proteins. The current generation of proteomic tools that are based on high-performance combinations of chromatography and mass spectrometry thus enable the identification of a growing number of proteins, and can also identify them over a wide range of abundances and when they have complex secondary modifications. The technology can achieve this using fragmentation, peptide sequencing and, as noted above, comparison with proteins that have been predicted from genome and transcriptome sequences [57-59].

Recent results indicate that the description of complete reference proteomes is now within reach in advanced centers, using multiple reaction monitoring combined with mass spectrometry; this combination is the most powerful and rapid targeted approach currently available [60-62]. These complete proteomes will probably serve as a reference for the subsequent development of simpler targeted assays [63], which will be complemented by array-based global surveys using affinity-based protein-specific reagents [64-66], and in certain cases, by single-cell proteomics using high-speed flow cytometry [67,68]. Furthermore, ongoing developments using nanomaterials are expected to provide next-generation proteomic analysis tools [69].

In addition to using chromatography and mass spectrometry, metabolomics is also taking advantage of nuclear magnetic resonance to analyze complex sets of metabolites in body fluids and tissues that reflect normal and disease states, and to study interactions with the gut microbial flora and environment factors [70-72]. Special attention is being paid increasingly to lipidomics [73,74] and glycomics [75,76] as complementary sources of biomarkers.

The development of each of these global high-throughput technologies has triggered implementation of standard operating procedures, ontologies and quality-assurance pipelines for data collection and analysis using dedicated software and databases, and this has required a change in culture in biological laboratories [77-81]. In turn, the need for independent validation of the results obtained with these ‘omics’ technologies has stimulated the emergence of large-scale chemical-genetics and functional screens using cell microarrays and RNA interference [82-84].

Computational and mathematical tools empower systems biology

With the increasing availability of large amounts of data and curated information on all types of biological system components, the focus has progressively shifted to identifying the interactions they make, forming transient or permanent macromolecular structures with particular biological functions, and to looking at how the interactions can be represented computationally as metabolic, protein, microRNA and gene-regulatory networks [85,86]. This emerging ‘network biology’ is taking advantage of advances in functional genomics, computational methods, computing power, and network and graph theories. It is reviving the advanced toolbox that has been published in textbooks and illustrated in static wall charts for decades [87-89]. Network biology is revealing the existence of modular structures in biological networks that may explain the robustness of biological systems when they are exposed to changing environments [90-93].

The initial attempts to identify biologically relevant protein-protein interactions using the yeast two-hybrid technology were plagued by high rates of artifactual events. Thanks to methodological improvements, the rate of false positives has been reduced. The careful curation of these interactions from targeted assays reported in the literature has led to high quality but incomplete maps of the human interactome, which are now available and which are expected to be extended to more complete coverage in the future [94-96]. Given that biological networks change their architectures dynamically during biological processes, such as development, physiological responses and disease, their complete determination will continue to be an enormous scientific and technological challenge.

Similar progress is being made in assembling human signaling, metabolic and gene regulatory networks that are based on metabolites, RNA and microRNA expression, protein-protein and protein-DNA interactions [97-100]. This has required the development of standardized languages and software tools for graphical representation of molecular interaction maps and computation of predictive and dynamic models [101-105]. Integration methodologies have also been essential to combine diverse types of data that have been collected with different platforms and in many laboratories, and thus to generate testable hypotheses [106-110]. A limitation that is often overlooked is that the quality of the annotation resources is very variable [111-114]. This has triggered sustained community efforts for integrative annotation, which combine automated computation with human-supervised curation, the use of...
quality indices, text-mining tools, biological ontologies and the semantic web [115-120].

In general, the models derived from these integrated methodologies have not yet reached the level of detail and precision of those obtained through highly focused systems biology approaches, such as those that describe the transcriptional control in a free living microorganism under changing environmental conditions [121] or the early phases of development of the sea urchin [122]. It seems likely that the same operating principles of network structure and dynamics that have been revealed in these latter model systems will be relevant to human physiology and pathology [123].

In a parallel track, the Physiome Project is building on over half a century of molecular modeling of excitable cells that used ordinary and partial differential equations and is also using finite element lattices for geometric modeling of complete human organs. This project has steadily developed a computational physiology framework with its own modeling language [124,125], and initial models of the beating heart, the contracting muscle and the breathing lung are already available [126,127]. Cell and development simulation efforts use yet other types of modeling formalisms and languages, including Boolean networks, cellular automata and process algebra [128-135], and many others are being developed in computational neuroscience, which has yet to merge with systems biology [136].

This diversity of approaches for modeling biological systems highlights the renewed importance of the contributions of mathematics, informatics and physics to systems biology [137-139]. Despite the introduction of novel computational methods, given that they are often based on distinct or incompatible principles, it is difficult or impossible to integrate these methods across the multiple levels of organization and time-scales characteristic of living systems [140-142]. Thus, multi-scale integration of different types of biological information (DNA, RNA, protein, networks, organelles, cells, tissues, organs, higher level phenotypes, and so on) remains a major challenge in systems biology. The plea for more theory by some of the founders of systems biology must be tempered by the fundamental need to have theories that closely reflect biological data through hypothesis-driven model testing [143]. Recent proposals based on allometric scaling [144] and scale relativity theory [8,145] may provide the theoretical framework and mathematical tools required to overcome some of these limitations, and may reveal an important role for small fluctuations in driving the behavior of biological systems [146,147].

The transition from medical genomics to systems medicine

With the availability of increasingly powerful high-throughput technologies, computational tools and integrated knowledge bases, it has become possible to establish new links between genes, biological functions and a wide range of human diseases [148-153]. This is providing signatures of pathological biology [154] and links to clinical research [155] and drug discovery [156,157]. These are the hallmarks of systems medicine as it is emerging from the initial, more targeted efforts of medical genomics.

Success in the identification of mutations affecting the hundreds of genes involved in inherited disorders has been a major outcome of the first generations of genetic maps of the human genome. In contrast, the reported associations between genetic polymorphisms and common complex traits have rarely been confirmed in independent studies. The situation has changed in the past two years [158], with the availability of dense maps of single nucleotide polymorphisms and the adoption by the community of medical geneticists of consensus guidelines for the optimal design of genome-wide or targeted association studies, including rules for independent replication [159,160]. Despite the very significant problems with signal-to-noise ratios that still severely limit the conclusions that can be drawn from such studies, progress has been made in identifying susceptibility loci involved in, for example, diabetes [161-163], obesity [164], and breast or lung cancer [165-167]. In the case of lung cancer, however, different scientific groups interpret the functional significance of the results differently. Further progress is expected now that the important role of other forms of genomic polymorphisms between individuals, including monozygotic twins, has been recognized; these include the effects of copy number variations and epigenetic modifications [168-170].

Taking advantage of expression-profiling surveys performed in extended human populations [171-174], systems biologists have started integrating physiopathology, network biology and DNA variations [175-177], providing novel insights into the mechanisms of various diseases, such as diabetes [178] and obesity [179]. Cancer, which can be considered as a prototypical systems disease, has benefited greatly from systems approaches and has served to a large extent as a test case to develop them [180-184]. This work has highlighted the importance of epigenetic variations in controlling transcriptional programs sustaining differentiation of normal and cancer stem cells [185,186].

Transcriptome and proteome analyses of collections of cancer samples, combined with functional annotation and modeling of modulated molecular pathways and networks, have revealed useful biomarkers for the classification and diagnosis of cancer subtypes, the prognosis of patient outcomes, the prediction of treatment responses and the identification of perturbation targets for drug development [187-196]. As an illustration of the value of systems approaches, the predictive power and robustness of biomarkers can be significantly increased by integrating transcriptome profiles with
interactome data to reveal more relevant functional subnetwork modules [197]. In a similar way, systems approaches are starting to have an impact on the study of immunological diseases [198], inflammation [199], infectious diseases such as tuberculosis [200], neurological diseases such as autism [201] and Alzheimer's [202], respiratory diseases such as asthma [203], cardiovascular and metabolic diseases [204-206] and many others. A common biological theme that emerges from many of these studies is that the control and dysfunction of energy metabolism has a central role. This is illustrated in cardiac system bioenergetics by the Frank-Starling law of cardiac muscle contraction [207, 208], in cancer by the Warburg effect (the dependence of cancer cells on aerobic glycolysis) [209], and in neurodegenerative diseases and aging by increases of oxidative stress [210, 211].

When East and West, North and South meet to develop systems medicine

Systems approaches are likely to help elucidate the mechanisms underlying the fundamental biological processes perturbed in human diseases and, in doing so, enable more efficient therapeutic interventions. They will change how drug targets are identified. Novel treatments will include multiple drugs interacting with key interconnected components within functional network modules, each contributing a fraction of the effects of perturbations that cause disease. It is likely that they will be effective only when combined with the multiple interactions of other drugs. This reflects the way that biological systems function and are organized to maintain themselves and constantly adapt to developmental, environmental, physiological or pathological changes. It is also reminiscent of the principles underlying traditional medicines developed empirically within Chinese or Indian cultures for the past several thousand years. Initial attempts at systems approaches, using transcriptome and proteome analyses to study the synergistic effect of combining Western drugs with Chinese medicine components in the treatment of leukemia, are starting to bear fruit [192, 212, 213]. Similarly, metabolome studies are being used to analyze the composition of herbal medicines and explain their properties [214], and to establish how gut microorganisms modulate human metabolic phenotypes and respond to the health or disease state of their host [215].

Systems approaches are also providing evidence on the effects of stress, relaxation, nutrition and lifestyle on the course of health and diseases [216, 217]. Systems studies need to pay greater attention to gender, age and time differences in diet, disease development and treatment administration and responses [218-221]. These factors can be monitored, for example, using non-invasive metabolomics surveys of urine [222, 223], and they will increasingly also be monitored using molecular fingerprints of blood proteins that indicate relevant physiological or disease states. Other important contextual phenomena that also need to be taken into account in future studies include the effects of the mother's genetic makeup or feeding habits on the development of the fetus and the timing of its biological clock, which have been observed in animal models [224, 225], and the central role of the major histocompatibility complex in the development and control of disease through immunity and inflammation [226-229].

Thus, systems biology will provide the foundation for a prospective medicine that will be predictive, personalized, preventive and participatory [230], and that takes into account the multiple components of the healthcare system, including disease outcomes as reported by the patients themselves, and public and private organizations involved in healthcare management. [231]. In addition to genomics and systems biology, the key components that will ensure the successful development of systems medicine are the modeling of physiopathology in a clinical-practice context [232], imaging [233], and bio-banking that complies with strictly enforced ethical regulations [234-236]. These intrinsically interdisciplinary endeavors will require dedicated centers and networks in which scientists of all disciplines can work together [237-239], with careful attention to clinical practice and education [240, 241].

In order to implement this vision, academia and industry will have to work closely together in an open-access and open-source environment focusing on the initial, pre-competitive phase of the drug discovery process. This will enable the subsequent development of valuable intellectual property that will result in more effective diagnostic and therapeutic approaches. Such developments might seem very far from the priorities of the less developed countries, in which the majority of the population is excluded from basic medical care. These countries are facing major challenges to their ability to fight infectious diseases and malnutrition, a situation aggravated by the shortage of safe drinking water and economic poverty [242]. International initiatives are underway to tackle these challenges in global health, such as support for engagement of communities in research and formulation of a research and development treaty that will redefine the rules for clinical trials and management of intellectual property rights [243, 244].

Strategic partnerships, such as the Systemscope Consortium, propose guidelines and strategies for 'rethinking research, understanding life, improving health' [245]. We support the view that leaders of the developing countries should consider establishing integrative systems biology and medicine centers networked with those emerging in the developed countries. Implementing such centers at the heart of their much-needed healthcare infrastructures would ensure immediate access to the most advanced technologies, and allow developing countries to build an essential
knowledge base centered on the analyses of their populations. These centers would provide a route to the adequate healthcare that is required to reduce the ever-growing gap between the developed and underdeveloped nations.

Competing interests
CA is Research Director at CNRS, and a consultant for bioMérieux and Mérieux Alliance. ZC is the Director of the Shanghai Center for Systems Biomedicine and is supported by the National Key Program for Basic Research (1973), China. LH is the President of the Institute for Systems Biology.

Acknowledgements
We thank our colleagues of the Genexpress team, the Center for Systems Biomedicine, the Institute for Systems Biology, and Haim Ben-dayan, Samir Brahmacari, Anthony Brooks, Dominique Charron, Eric Eveno, David Galas, Takashi Gojobori, Sandrine Imbeaud, Doron Lancet, Jacques Mallet, Xavier Leverve, Laurent Nottale, Denis Noble, Christophe Pison, Valdur Saks, Marc Vidal, John Weinstein for insightful discussions.

References
1. Ideker T, Galitski T, Hood L: A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2001, 2:343-372.
2. Kitano H: Systems biology: a brief overview. Science 2002, 295:1662-1664.
3. Noble D: Modeling the heart - from genes to cells to the whole organ. Science 2002, 295:1678-1682.
4. Auffray C, Imbeaud S, Roux-Rouquie M, Hood L: From functional genomics to systems biology: concepts and practices. C R Biol 2003, 326:879-892.
5. Nicholson JK, Holmes E, Lindon JC, Wilson ID: The challenges of modeling mammalian biocomplexity. Nat Biotechnol 2004, 22:1268-1274.
6. Westerhoff HV, Palsson BO: The evolution of molecular biology into systems biology. Nat Biotechnol 2004, 22:1249-1257.
7. Kirschner MW: The meaning of systems biology. Cell 2005, 121:503-504.
8. Auffray C, Nottale L: Scale relativity theory and integrative systems biology I: Founding principles and scale laws. Prog Biophysics Mol Biol 2008, 97:79-114.
9. Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Antohe CR, Elston TC, Collins JJ: A bottom-up approach to gene regulation. Nature 2006, 439:856-860.
10. Bray D: Molecular networks: the top-down view. Science 2003, 301:1864-1865.
11. O'Malley MA, Dupre J: Fundamental issues in systems biology. Bioessays 2005, 27:1270-1276.
12. Noble D: Claude Bernard, the first systems biologist, and the future of physiology. Exp Physiol 2008, 93:16-26.
13. Wilson I: Top-down versus bottom-up-rediscovering physiology via systems biology. Mol Syst Biol 2007, 3:113.
14. Wolkenhauer O: Systems biology: the reincarnation of systems theory applied in biology? Brief Bioinform 2001, 2:259-270.
15. Ideker T, Winston LR, Lauffenburger DA: Bioengineering and systems biology. Annu Biomed Eng 2006, 34:1226-1223.
16. Yildirim MA, Vidal M: Systems engineering to systems biology. Mol Syst Biol 2008, 4:185.
17. Brent R: A partnership between biology and engineering. Nat Biotechnol 2004, 22:1211-1214.
18. Church GM: From systems biology to synthetic biology. Mol Syst Biol 2005, 1:2005.0033.
19. van Ommen GJ: Medical genomics. Eur J Hum Genet 2001, 9:729.
20. Brand A, Brand H, Schulte in den Bäumen T: The impact of genetics and genomics on public health. Eur J Hum Genet 2008, 16:5-13.
21. Ahn AC, Tewari M, Poon CS, Phillips RS: The clinical applications of a systems approach. PLoS Med 2006, 3:e208.
22. Hood L, Heath JR, Phelps ME, Lin B: Systems biology and new technologies enable predictive and preventative medicine. Science 2004, 306:60-64.
23. Schmidt EE, Lamb J, Yang J, Zhu J, Edwards S, GuhaThakurta D, Sieberts SK, Sons A, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kasif S, Drake TA, Sachs A, Luiss AJ: An integrative genomics approach to infer functional associations between gene expression and disease. Nat Genet 2005, 37:10-117.
24. Zerhouni EA: Translational and clinical science - time for a new vision. N Engl J Med 2005, 353:1621-1623.
25. Nicholson JK: Global systems biology, personalized medicine and molecular epidemiology. Mol Syst Biol 2006, 252.
26. Butte AJ: Medicine. The ultimate model organism. Science 2008, 320:325-327.
27. Butcher EC: Can cell systems biology rescue drug discovery? Nat Rev Drug Discov 2005, 4:461-467.
28. Hood L, Perlmutter RM: The impact of systems approaches on biological problems in drug discovery. Nat Biotechnol 2004, 22:1215-1217.
29. Weinstein JN, Pommier Y: Connecting genes, drugs and diseases. Nat Biotechnol 2006, 24:1365-1366.
30. Yan Q: The integration of personalized and systems medicine: bioinformatics support for pharmacogenomics and drug discovery. Methods Mol Biol 2008, 448:1-19.
31. Strausberg RL, Levy LS, Rogers YH: Emerging DNA sequencing technologies for human genomic medicine. Drug Discov Today 2008, 13:569-577.
32. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y, Malikapati V, Roth GT, Gomes X, Turturco CL, Iryzk GP, Lupski JR, Chinnault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Roithberg JM: The complete genome of an individual by massively parallel DNA sequencing. Nat Genet 2008, 40:922-926.
33. Church GM: The personal genome project. Mol Syst Biol 2005, 1:2005.0030.
34. Quackenbush J: Weighing our measures of gene expression. Mol Syst Biol 2005, 263.
35. Ransohoff DF: Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer 2004, 4:309-314.
36. Ransohoff DF: Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer 2005, 5:142-149.
37. Ioannidis JP: Why most published research findings are false. PLoS Med 2005, 2:e124.
38. Imbeaud S, Auffray C: ‘The 39 steps’ in gene expression profiling: critical issues and proposed best practices for microarray experiments. Drug Discov Today 2005, 10:1175-1182.
39. Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E, Mueller O, Schroeder A, Auffray C: Towards standardization of RNA quality assessment using user-independent classifiers of micro-capillary electrophoresis traces. Nucleic Acids Res 2005, 33:e56.
40. Schroeder A, Mueller O, Stocker S, Salowski R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Grometz M, Ragg T: The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 2006, 7:3.
41. Dupuy A, Simon RM: Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Nat Cancer Inst 2007, 99:147-157.
148. Giallourakis C, Henson C, Reich M, Xie X, Mootha VK: Disease gene discovery through integrative genomics. *Annu Rev Genomics Hum Genet* 2005, 6:381-406.

149. Goh KI, Cusick ME, Childs D, Bidal V, Mital M, Barabasi AL: The human disease network. *Proc Natl Acad Sci USA* 2007, 104:8685-8690.

150. Lamb J, Crawford ED, Peck D, Modell JW, Blatt IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Herndon Y, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR: The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006, 313:1929-1935.

151. Liang M: Integrative pathway knowledge bases as a tool for systems molecular medicine. *Physiology Genomics* 2007, 32:209-212.

152. Schuster-Bockler B, Bateman A: Protein Interactions in human genetic diseases. *Genome Biol* 2008, 9:95.

153. Wu X, Jiang R, Zhang MQ, Li S: Network-based global inference of human disease genes. *Nat Syst Biol* 2008, 4:189.

154. Nicolau M, Tibshirani R, Borresen-Dale AL, Jeffery SS: Disease-specific genomic analysis: identifying the signature of pathobiological pathology. *Bioinformatics* 2007, 23:957-965.

155. Pepperkok R, Wiemann S: Integrating systems biology with clinical research. *Genome Biol* 2008, 9(S1):14.

156. Sears DB: Data integration: challenges for drug discovery. *Nat Rev Drug Discov* 2008, 7:455-458.

157. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M: Drug-target network. *Nat Biotechnol* 2007, 25:1119-1126.

158. van Ommen GJ: Popper revisited: GWAS here, last year. *Eur J Hum Genet* 2008, 16:75-77.

159. NCI-NHGRI Working Group on Replication in Association Studies, Skol AD, Scott LJ, Abecasis GR, Boehnke M: Optimal designs for genome-wide association studies. *Genome Medicine* 2009, Volume 1, Issue 1, Article 2.

160. Auffray et al: Completing the map of human genetic variation. *Nature* 2007, 447:161-165.

161. Fraas MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Sueter D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M: Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 2007, 104:1064-1069.

162. Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. *Nat Rev Genet* 2006, 7:85-97.

163. Cuthbert A, Lai SL, Sebat J, Lakshmi B, Leal SM, Scherer SW, Donis-Keller H, Hudson T, Lee C, Rong S, Stoffel M, Sunyaev S, Venter JC, Gabriel SB, Schaffner SF, Lee L, Peckham HE, et al.: Somatic copy number changes induced in cancer cells by culture. *Nature* 2007, 445:542-546.

164. Sninsky JJ, Iannone MF, Kircher MT, Morley G, Sabben-Clare LS, Sibson R, Greenleaf J, Smith GR, de la Chapelle A, Longley CS, Ulman C: Copy-number variation of a novel 7p15 locus in breast cancer. *Nature Genet* 2006, 38:1067-1073.

165. Auffray et al: Mapping the epigenome. *Nature Genet* 2007, 39:1202-1207.

166. Elemento O: Developing a systems-level understanding of gene expression. *Genome Biol* 2007, 8:304.

167. Wolfe CJ, Kohane IS, Butte AJ: Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics* 2005, 6:227.

168. Degnan JH, Lasky-Su J, Raby BA, Xie X, Melnyk SB, Smith AV, Zuberi K, Masterson S, Wang H, Levy A, et al.: Missing heritability and rare allele sharing in extended families. *Am J Hum Genet* 2009, 84:21-33.

169. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne C, O'Connell J, Hinder MT, O'Kelly ME, Russell IM, Weatherall DJ, Downey PJ, Frayling TM, Le Page AM, Cross MR, Zeggini E, DerSimonian R, Cooper R, et al.: Population variation in the human genome. *Science* 2007, 315:2148-2153.

170. Marioni JC, H seasoning C, Spencer N, Ganguly S, Cordero J, Ayres M, Dyer TD, Charlesworth J, Cole SA, Wray NR, Gieger C, Williams R, et al.: Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007, 315:84-88.

171. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardi B, et al.: Genetic network inference using large-scale transcriptional profiling in human lymphocytes. *Nature Genet* 2007, 39:1208-1216.

172. Stranger BE, Forrest MS, Dunning M, Igel G, Beutler AJ, Johnson J, Huse K, Mohlke KL, Easton D, Segrer SW, Tavare S, Deloukas P, Hurles ME, Dermitzakis ET: Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007, 315:84-88.

173. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardi B, et al.: Genetic network inference using large-scale transcriptional profiling in human lymphocytes. *Nature Genet* 2007, 39:1208-1216.

174. Wolfe CJ, Kohane IS, Butte AJ: Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics* 2005, 6:227.

175. Degnan JH, Lasky-Su J, Raby BA, Xie X, Melnyk SB, Smith AV, Zuberi K, Masterson S, Wang H, Levy A, et al.: Missing heritability and rare allele sharing in extended families. *Am J Hum Genet* 2009, 84:21-33.

176. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne C, O'Connell J, Hinder MT, O'Kelly ME, Russell IM, Weatherall DJ, Downey PJ, Frayling TM, Le Page AM, Cross MR, Zeggini E, DerSimonian R, Cooper R, et al.: Population variation in the human genome. *Science* 2007, 315:2148-2153.

177. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne C, O'Connell J, Hinder MT, O'Kelly ME, Russell IM, Weatherall DJ, Downey PJ, Frayling TM, Le Page AM, Cross MR, Zeggini E, DerSimonian R, Cooper R, et al.: Population variation in the human genome. *Science* 2007, 315:2148-2153.

178. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardi B, et al.: Genetic network inference using large-scale transcriptional profiling in human lymphocytes. *Nature Genet* 2007, 39:1208-1216.

179. Stranger BE, Forrest MS, Dunning M, Igel G, Beutler AJ, Johnson J, Huse K, Mohlke KL, Easton D, Segrer SW, Tavare S, Deloukas P, Hurles ME, Dermitzakis ET: Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007, 315:84-88.

180. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardi B, et al.: Genetic network inference using large-scale transcriptional profiling in human lymphocytes. *Nature Genet* 2007, 39:1208-1216.
