MINIREVIEW

Genomics and Clinical Medicine: Rationale for Creating and Effectively Evaluating Animal Models

KELLY S. SWANSON,* MEREDITH J. MAZUR,* KAPIL VASHISHT,*,† LAURIE A. RUND,* JONATHAN E. BEEVER,‡ CHRISTOPHER M. COUNTER,§ AND LAWRENCE B. SCHOOK*,‡,[1]

Laboratory of Comparative Genomics, University of Illinois, Urbana, Illinois 61801; †Department of Veterinary Pathobiology, University of Illinois, Urbana, Illinois 61801; ‡Laboratory of Molecular Genetics, Department of Animal Sciences, University of Illinois, Urbana, Illinois 61801; and §Department of Pharmacology and Oncology, Duke University Medical Center, Durham, North Carolina 27710

Because resolving human complex diseases is difficult, appropriate biomedical models must be developed and validated. In the past, researchers have studied diseases either by characterizing a human clinical disease and choosing the most appropriate animal model, or by characterizing a naturally occurring or induced mutant animal and identifying which human disease it best resembled. Although there has been a great deal of progress through the use of these methods, such models have intrinsic faults that limit their relevance to clinical medicine. The recent advent of techniques in molecular biology, genomics, transgenesis, and cloning furnishes investigators with the ability to study vertebrates (e.g., pigs, cows, chickens, dogs) with greater precision and utilize them as model organisms. Comparative and functional genomics and proteomics provide effective approaches for identifying the genetic and environmental factors responsible for complex diseases and in the development of prevention and treatment strategies and therapeutics. By identifying and studying homologous genes across species, researchers are able to accurately translate and apply experimental data from animal experiments to humans. This review supports the hypothesis that associated enabling technologies can be used to create, *de novo*, appropriate animal models that recapitulate the human clinical manifestation. Comparative and functional genomic and proteomic techniques can then be used to identify gene and protein functions and the interactions responsible for disease phenotypes, which aids in the development of prevention and treatment strategies. Exp Biol Med 229:866–875, 2004

Key words: genomics; animal models; transgenesis; recombineering

Introduction

Understanding the effects of the environment, disease, or nutrition on human biological systems has dramatically increased over the past century. Public health programs and advances in clinical medicine have increased the human life span (all races, both genders) from 47.3 years in 1900 to 76.9 years in 2000 (1). Technological and medical advances have improved quality of life and provided prevention strategies and treatments for ailments and injuries that once were deadly. However, the incidence of chronic diseases due to complex genetic and environmental interactions has also dramatically increased in the past century. In fact, a majority of the top 10 causes of death in the United States in 2000 (i.e., heart disease, cancer, stroke, chronic lower respiratory disease, diabetes, Alzheimer disease, renal disease) represent diseases that have been shown to result from such interactions (2). Diseases including osteoarthritis, gastrointestinal disorders, mental or behavioral disorders, alcohol-

This work was partially supported by grants from the USDA-National Research Initiative (2002-35205-12712), the USDA Cooperative State Research Service (AG2002-34480-11828), and the USDA Agricultural Research Service (Agreement 58-5438-2-313).

[1] To whom correspondence should be addressed at 329 Edward R. Madigan Laboratory, University of Illinois, 1201 West Gregory Drive, Urbana, IL 61801. E-mail: schook@uiuc.edu

1535-3702/04/2299-0866$15.00

ism, and drug addiction are also of clinical importance because they are often comorbidities and affect quality of life.

Resolving human complex diseases is difficult (e.g., the time required to develop disease symptoms, expenses associated with human clinical experiments, ethical issues) and, thus, appropriate biomedical models must be developed and validated. *Biomedical models* have been defined as "surrogates for a human being, or a human biologic system, that can be used to understand normal and abnormal function from gene to phenotype and to provide a basis for preventive or therapeutic intervention in human diseases" (3, 4). In the past, researchers have used two approaches to study human diseases. One strategy fully characterizes a human clinical disease and chooses the most appropriate animal model based on criteria such as anatomical and/or physiological characteristics (i.e., biological relevance), cost, and animal husbandry required. Another tactic has been to fully characterize a naturally occurring or induced (by chemical or radiation exposure) mutant animal (most commonly the rat or mouse) and identify which human disease it resembles. Although a great deal of progress has been achieved using these methods, they have intrinsic faults that limit their relevance to clinical medicine. The recent advent of techniques in molecular biology, genomics, transgenesis, and cloning furnishes investigators with a new ability to study vertebrates such as pigs, cows, chickens, and dogs with greater precision and utilize them as model organisms.

Comparative and functional genomics and proteomics provide effective approaches in identifying the genetic and environmental factors responsible for complex diseases and in the development of prevention and treatment strategies and therapeutics. By identifying and studying homologous genes across species, researchers are able to accurately translate and apply experimental data from animal experiments to humans and vice versa. Although gene location and sequence is important, determining the functions and regulatory elements of genes is the critical step in understanding the biology of a disease and how it may be prevented or treated. This review supports the hypothesis that associated enabling technologies can be used to create, *de novo*, appropriate animal models that recapitulate the human clinical manifestation. Comparative and functional genomic and proteomic techniques can then be used to identify gene and protein functions and the interactions responsible for disease phenotypes, which aids in the development of prevention and treatment strategies.

Traditional View: Choosing Appropriate Animal Models

Traditionally, researchers have selected appropriate animal models by using various criteria with respect to the availability of resources including (i) relevance to the human disease (e.g., anatomical, physiological, and pathological similarities); (ii) practical issues pertaining to the model (e.g., physical and behavioral characteristics of animal, dietary, and housing requirements; cost; animal care expertise required; and (iii) ethical issues associated with the model, government laws, and restrictions (Fig. 1). Due to their small size, low cost, availability, and short breeding time and life span, rodents have historically been the most popular animal models. Conversely, research using primates and canines has declined due to ethical issues and cost associated with their use.

Steady progress has been made using traditional models to study human diseases. However, anatomical or physiological differences across species may negate the use of some species or generate misleading information. For example, due to low cost and ethical issues, the rat is commonly used for preclinical testing of drugs. However, because rats lack a homologous enzyme for human CYP3A4 (responsible for the metabolism of the majority of drugs tested), the generated data often do not accurately predict the human response (5). Given the recent developments in molecular biology and clinical medicine, a need to reevaluate and scrutinize the use of these traditional models exists.

Current View: Utilizing Genetic Information to Create Animal Models

The Human Genome Initiative is providing genetic information not only from humans, but also from animals traditionally used as models. In addition, related enabling technologies in transgenesis and animal cloning provide new approaches for designing and performing experiments to dissect complex biological systems. Because of these new technologies (e.g., transgenesis), scientists are no longer limited to the traditional methods of choosing naturally occurring models. Researchers can utilize genomic knowledge and available tools to create appropriate animal models. This approach is referred to as *reverse genetics*. In contrast to *forward genetics* in which the gene or genes responsible for a particular phenotype are identified by positional cloning (phenotype to genotype), the reverse genetics approach determines the function of a gene and predicts the phenotype of a cell, tissue, or organism (genotype to phenotype). The convergence of classical and reverse genetics, along with genomics, provides a working definition of the *genetic model organism* (6). According to Barr (7), an exemplary genetic model organism must (i) be genetically amenable to both forward and reverse genetic approaches; (ii) provide researchers with the ability to manipulate the genome (transgenesis); (iii) have a sequenced, or soon to be sequenced, genome; (iv) possess several practical characteristics (e.g., short generation time, small body size, ease and low cost of maintenance); (v) possess unique characteristics that simplify an analysis of the biological problem of interest; and (vi) be studied by a critical mass of investigators that has access to powerful resources. Thus, comparative genomics has added a new important criterion to animal models: In addition to closely

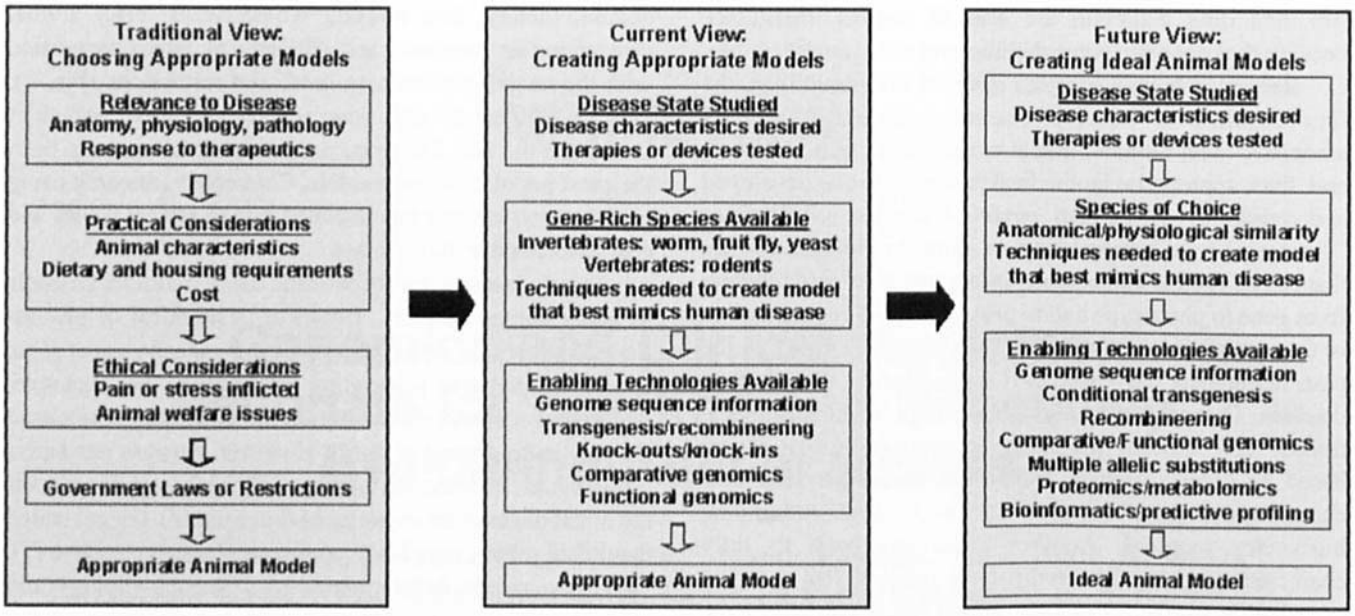

Figure 1. Methods to choose, create, and interpret data generated from animal models are expanding due to advances in molecular biology and bioinformatics.

resembling the human disease, an ideal animal model would also be amenable to a variety of genetic manipulations.

Until recently, most animals have not been amenable to both forward and reverse genetic manipulations and, thus, investigators were forced to choose between biomedical relevance and genetic power. Several invertebrate organisms possess many of the qualities of an ideal genetic model organism (e.g., short generation time, small size, sequenced genome, genetically amenable) and are useful in basic gene and protein functional analysis. However, invertebrates often lack a direct link to human medicine because of their simplistic anatomy and physiology and lack of gene homologs. For example, yeast strains are poor models for apoptosis because they lack endogenous caspases, or Bcl-2s genes, that are the main apoptosis regulators in mammals (8). An invertebrate's simple physiology and genetic regulation also negates its relevance to studying complex diseases. The fact that the genome of *Caenorhabditis elegans* (a roundworm composed of fewer than 1000 somatic cells and 1 mm in length) contains just one-third fewer genes than humans underscores the importance of gene regulation and gene-gene/protein-protein interactions rather than gene number. Although an invertebrate has an important role in fundamental gene analysis, its simplistic phenotype limits its use in studying the physiology of mammalian species phenotypically determined by complex genetic interactions.

Mammalian species have greater biomedical relevance but, until recently, have not possessed genetic power. Because sufficient genetic information has only been available in rodents, most genetic techniques have been limited to these species. Although expensive in comparison to invertebrates, murine mutagenic screens have been successful in revealing novel human genotypes (9, 10). Several mutagenic strategies have been used in the past and are briefly reviewed by Stanford *et al.* (11). Because of the pitfalls associated with spontaneous mutagenesis (e.g., low frequency, $\sim 5 \times 10^{-6}$ per locus, only visible phenotypes detected), methods able to induce mutagenesis such as x-ray exposure are utilized. The use of x-ray mutagenesis began in the 1930s because it produced mutations more frequently than spontaneously occurring mutagenesis (20–100× greater) and caused chromosomal rearrangements that provided landmarks for cloning (11). Because x-ray exposure is often difficult to interpret due to multiple genetic mutations, more powerful phenotype- or gene-driven approaches have been established for identifying gene function.

Phenotype-driven approaches (forward genetics) for identifying gene function are based on chemical mutagenesis (9, 10), while gene-driven approaches (reverse genetics) are based on insertional mutagenesis (11). The most common phenotype-driven approach in mice uses the potent chemical mutagen ethylnitrosourea (ENU) to generate random mutations in the genome. By mating ENU-exposed male mice to normal females, numerous mutant F1 progeny are produced and identified using various screening methods (e.g., relevant clinical symptoms). Mutated genes from abnormal (desired) phenotypes are identified and animals are archived by collecting and freezing sperm or ovaries from the F1 mutant.

Comparative mapping allows researchers to predict the identity and location of the corresponding human gene. Although a chemical mutagen could introduce multiple hits in a genome, ENU exposure usually results in monogenic mutations (9, 10). Although ENU exposure does not produce molecular landmarks for cloning, it has demon-

strated broad utility in several mutagenic screening programs because it produces single-gene mutations and is amenable to high-throughput screening and sequence validation (11).

The gene-driven approach to mutagenesis uses insertional mutagens to target or trap genes, which may result in more controlled and targeted mutations. These techniques also have a very high mutation rate, resulting in nearly 100% transgenic animals once prescreened *in vitro*. Access to sequence data and molecular biological techniques, insertional mutagenesis, or transgenesis enables researchers to create biomedical models intended for specific diseases rather than waiting for a random mutant to occur from chemical induction. Bockamp *et al.* (12) reviewed techniques used to generate transgenic mouse models. Although the first transgenic techniques resulted in gene knockouts (constitutive transgenesis), recent techniques permit conditional control of gene expression (13). Thus, transgenesis may enable the creation of the ideal model by causing the over- or underexpression of a given gene or by inserting a novel DNA sequence into an animal's germline.

The use of embryonic stem cells in culture (14, 15) and knowledge gained in regard to homologous recombination (HR) in mammalian cells (16) allows the targeting of specific genes. By establishing a precise site of integration and, consequently, influencing specific genes, HR in embryonic stem cells avoids the unfavorable effects that occur when sequences are randomly inserted. Selection markers have been developed to select cells that have incorporated exogenous DNA into their genomes (positive selection cassette) or that have excised the cassettes (negative marker) (17). Although gene targeting is more controlled and efficient than chemical mutagenesis, it still often fails to produce an animal model that resembles the desired human disease (11).

Many of the problems associated with traditional knockout strategies that exhibit constitutive expression of the transgene may be avoided with the conditional control of gene expression. In other words, investigators have the ability to activate or suppress gene expression without resulting in secondary pleiotropic effects. An ideal conditional transgenic animal would have the following characteristics: (i) reversible genetic switch, (ii) zero or low basal gene expression when the gene is switched "off," (iii) high and rapid induction of gene expression when the gene is switched "on," (iv) tightly controlled induction of gene expression without pleiotropic side effects, and (v) induction of the gene by highly specific nontoxic compounds (12, 13).

Binary transgenic systems that control the expression of a gene by the interaction of two components have been successfully used. With these systems, an "effector" transgene acts on a target gene to either activate or silence its expression. Site-specific recombination (effecting DNA) and transcriptional transactivation (effecting RNA) are two methods used to perform conditional transgenesis. DNA recombinases may be used to rearrange a target gene, thereby silencing its expression. *Cre* from the bacteriophage P1 and *Flp* from *Saccharomyces cerevisiae* are members of the integrase family and are commonly used for site-specific DNA recombination (18, 19). These recombinases are suitable for use in mammalian cells because they do not require the presence of any accessory proteins or high-energy co-factors for their activity. In this system, the conditional allele in the F1 progeny will be inactivated only in the tissues that express *Cre*. Because DNA recombination permanently alters transgene activation, its major drawback is that it is irreversible. Transcriptional transactivation systems such as the tetracycline-dependent regulatory system (20) and Gal4/UAS system (21) only influence RNA and have several advantages: reversibility of the expression of the target gene, sensitivity of transgene activation levels to inducer concentration, and ability to control the expression of more than one target transgene (13).

Gene trapping, another insertional strategy, is more advantageous than gene targeting in many respects because it is not as labor intensive, it is able to report an endogenous gene-expression pattern, and it traps genes regardless of their transcriptional activity (11). Although three types of trap vectors exist (i.e., enhancer trapping, promoter trapping, gene trapping), gene traps seem to be the most effective. The gene-trap vectors contain a splice acceptor site immediately upstream from a promoterless reporter gene and are able to be inserted into a large collection of chromosomal sites (22). On the transcriptional activation of the endogenous *cis*-acting promoter and enhancer elements of the trapped gene, a fusion transcript is generated from the upstream coding sequence and the reporter gene, which simultaneously mutates the trapped gene and reports its expression pattern (11). The main disadvantage to gene trapping is that because it is inserted in an intron, alternative splicing may lead to lower levels of wild-type transcripts and result in hypomorphic alleles (23).

Chromosomal engineering, or *recombineering*, uses phage-based homologous recombination and has the capability to manipulate large segments of DNA such as those carried on by bacterial artificial chromosomes (BAC) or P1 artificial chromosomes, which is impossible with many cloning vectors (24). Several approaches involving the modification of the host bacterium have been developed to permit BAC manipulation, including inducible promoters permitting transient expression of bacterial *rec*E and *rec*T genes or other analogous bacteriophage lamda (λ) genes (*exo* and *bet*) (25). Recently, an *Escherichia coli* strain harboring a defective λ prophage was developed and promoted high BAC recombination frequencies. In this strain, the prophage provides the recombination genes *exo*, *bet*, and *gam* under the control of a temperature-sensitive λ cl-repressor (26). The *E. coli* strain DY380 was generated by introducing the λ prophage into the BAC host strain DH10B, providing a rapid, single-step method to generate subtle changes in any gene in BAC clones using oligonucleotides as targeting vectors (27). Therefore, in

contrast to most *E. coli* cloning methods that require the use of restriction endonucleases to cleave DNA and DNA ligases to join DNA fragments, recombineering can be accomplished without using these enzymes. Using a polymerase chain reaction (PCR)-based selective amplification screen to identify targeted clones, Swaminathan *et al.* (27) demonstrated the ability of this system to generate single-base changes, deletions (up to 1.93 kb), and insertions of unique sequences in different regions of a BAC. Advantages of these systems are that only short segments of homology are required to direct recombination, and the high-efficiency rates allow recombinants to be screened rather than selected (24). Because recombinants can be screened, only one recombination step is required to create the desired modification directly, without the use of selective markers (e.g., drug-selectable markers, loxP sites, FLP restriction target (FRT) sites).

The use of double-stranded RNA (dsRNA) to cause RNA interference results in loss of function and is a powerful screening technique used to identify genes associated with specific biological processes. Using dsRNA to inhibit the expression of specific genes was first used in *C. elegans* by Fire *et al.* (28). These researchers reported that dsRNA was substantially more effective at producing interference than either strand individually and that the effect was evident in both those injected and their progeny (28). Several recent publications have provided potential mechanisms by which dsRNA cause the degradation of targeted messenger RNA (29–32). Although the use of long dsRNA enables effective silencing of gene expression in lower organisms, it is of limited use in mammals because the introduction of dsRNA (longer than 30 nucleotides) induces an interferon response that is sequence nonspecific (33). To avoid the interferon response in mammalian systems, small interfering RNA or short hairpin RNA (shRNA) are used. Until recently, genome-wide RNA interference surveys of gene function were limited to nematode worms and fruit flies (34, 35). However, two groups have recently developed resources for large-scale, RNA interference–based screens in mammals (36, 37). Berns *et al.* (36) reported the construction of a set of retroviral vectors encoding 23,742 distinct shRNA, which target 7914 human genes for suppression. Similarly, Paddison *et al.* (37) reported the construction and application of an shRNA expression library (comprising ~28,000 sequence-verified, shRNA-expression cassettes contained within multifunctional vectors) targeting 9610 human genes and 5563 mouse genes. Given the recent advances in this area, there is no doubt that RNA interference screens will continue to be important in determining mammalian gene function.

Because sequence information is required for the aforementioned techniques, their use has been limited to species with annotated genome sequences (e.g., invertebrates, mice). Regardless of the techniques used to generate mutants or transgenics, this information can be used to develop a mutant map of that species. Although a murine mutant map will not provide a detailed, comprehensive picture of all the networks and interactions of genes that contribute to complex diseases, it will lay the groundwork for mammalian genetics by expanding our understanding of the role of specific genes (38). Remarkable progress has been made with regard to mutagenic and transgenic strategies in the past few decades. As these strategies continue to be used and improved, their utility in mice genomics will continue to improve and their applicability to other mammalian species will be realized.

Future View: The Marriage of Genetic Information and Biomedical Relevance

Although the mouse genome can now be manipulated with relative ease, a phenotype gap still remains. Brown and Peters (39) coined the term *phenotype gap* to refer to the gulf between available mouse mutant resources and the full range of phenotypes possible (only ~1%–2% of mouse loci have known phenotypic mutants). This definition can be further expanded to include the animal mutant resources that recapitulate human phenotypes (i.e., normal and diseased). Although the continued utilization of murine mutant screens and transgenesis will generate data and models of great value, genetic manipulation of other mammalian species is required to close the phenotype gap.

Murine transgenesis methodology has been heavily studied and is highly successful in creating transgenic strains; however, it often fails to produce a model that resembles the desired human disease phenotype (11). Although many factors are likely responsible for the lack of success, differences in human and mouse life span and/or anatomy and physiology may play larger roles. Because of the vast differences in certain human and mouse organ systems, researchers must utilize the strengths offered by other animal models. For example, because divergence in lung and pancreatic anatomy between mice and humans is largely to blame for the lack of pathological lesions present in mouse cystic fibrosis models, an ovine model is now being used to study this disease (40). Prostate cancer is another area of research that the mouse is not best suited to study, given the differences in gross anatomy and microanatomy of the human and mouse prostate. Conversely, the dog is the only species besides humans to frequently develop spontaneous prostate cancer and has several advantages over other models: (i) both human and canine prostate cancer are strongly associated with age; (ii) like the human version, canine prostate carcinoma has a high propensity for osseous metastases; and (iii) dogs provide a large animal model, which makes imaging and diagnostic studies possible (41). Finally, our laboratories are currently using the pig as a model for the devastating disease ataxia-telangiectasia (A-T) because it is poorly reproduced in transgenic mice. Loss of function in both alleles of the human *ATM* gene gives rise to A-T, resulting in a

progressive loss of motor control (ataxia) and early death (42). The absence of ataxia and only mild neuropathological defects in transgenic mice ($Atm^{-/-}$) demonstrates the need for other animal models (43, 44). In addition to anatomical and physiological similarities with humans, pigs, like mice, can be genetically manipulated to lack the functional alleles of *ATM*, which results in a large animal model of A-T.

In addition to differences in anatomy and physiology, rodents may not be the most desirable in terms of gene homology. Although rodents are evolutionarily closer to humans than those from the Orders Carnivora (e.g., dog, cat) and Artiodactyla (e.g., pig, cow), the higher rate of nucleotide substitution observed in rodents may diminish their relevance in regard to comparative genomic techniques (45, 46). In a recent experiment the dog, cat, pig, and cow all had a higher percentage (~60%) of sequence homology to humans than did rats or mice (~40%) (46). Kirkness *et al.* (47) recently compared the canine genome (6.22 million sequence reads, 1.5× coverage) with drafts of the human and mouse genomes (National Center for Biotechnology Information Builds 31 and 3, respectively) using BLASTN. Despite much lower sequence coverage of the dog (1.5×), alignments covered a similar number of human transcripts and genes as the mouse (8×) (47). Other analyses discovered that although the level of nucleotide substitution was similar in dogs and humans, a 1.6-fold higher substitution rate was measured in the mouse (47). The recent reports of Thomas *et al.* (46) and Kirkness *et al.* (47) identified some of the limitations associated with rodent research and advocated the use of nonrodent species.

The Human Genome project windfall has greatly benefited other mammalian species such as the dog, cow, chicken, and pig, which are now having their genomes sequenced. Genome sequencing will soon recategorize these species as "gene rich" and will allow the use of enabling technologies (e.g., recombineering, transgenesis) to create appropriate animal models that possess more biological relevance than invertebrates or rodents. Genome characteristics, sequencing status, and on-line resources of common model organisms are presented in Tables 1 and 2. By genetically manipulating animals that are more similar to humans in terms of anatomy and physiology, ideal animal models may be created.

Functional Genomics and Proteomics: Providing Mechanisms and Intervention

Now that scientists have the knowledge and tools to rapidly sequence model organisms and create biologically relevant biomedical models through genetic manipulation, the next step is to accurately interpret sequence data and animal experimentation. Although the genome structure is important in many regards, the function, regulation, and interaction of genes and gene products have the major influence on phenotypes. This concept may be demonstrated by comparing a caterpillar and a butterfly, which are genomically identical but anatomically and physiologically distinct (48). As the field of functional genomics and proteomics matures, our understanding of gene and protein function, cellular function, and physiology will be greatly enhanced. This new-found knowledge will allow scientists to fully utilize genome sequence data to study clinical disease.

Successfully understanding clinical disease using genomic and proteomic technologies is a lofty goal that will be difficult to attain, especially for complex phenotypes or diseases. The process of determining gene function is a much more daunting task than once thought, as the "one gene (mutant), one product (disease)" theory most often does not apply. The fact that most mammalian genomes are estimated to contain only about 30,000 to 40,000 genes suggests that, in addition to the overall number of genes present in a genome, other factors such as temporal and spatial gene-expression patterns, alternative splicing, post-transitional modification, and protein-protein interactions greatly influence phenotype. Bioinformatic programs that are able to accurately translate complex genotypes into phenotypes by predicting the occurrence and relevance of these factors will be required to fully understand the complex organ systems of the human body and detect the abnormalities responsible for complex diseases. Before complex gene interactions can be interpreted, information regarding each individual gene must be collected and understood. Several high-throughput methods of assessing gene products including affinity precipitation (protein-protein interactions) (49), two-hybrid techniques (50), synthetic rescue (51), lethality experiments (52), and DNA microarray analysis (53, 54) have been developed in the last 25 years. The recent advent and acceptance of microarray technology, in particular, has made a major impact on biological research and has taken us another step closer to understanding complex biological systems.

The concept behind DNA microarrays is the precise positioning of DNA fragments (probes) at a high density on a solid support so that they can act as molecular detectors (55–58). DNA microarray analysis can be used to identify sequence variations (e.g., single nucleotide polymorphisms, gene mutations) or determine the gene-expression level (abundance) of a set of messenger RNA molecules. DNA microarrays have broad utility because they can measure the expression of thousands of genes simultaneously, providing a global view of gene expression rather than only a few genes that limit classical techniques. To accurately interpret and compare microarray data, minimum information about a microarray experiment (MIAME) standards have been established for microarray experiments (59). Although criticisms still exist (e.g., expression levels are only relative to standard or reference, lack of standard methodology, quality control issues), microarray experiments that abide by the MIAME standards, are properly designed (60), and are validated using other molecular biological techniques (e.g.,

Table 1. Genomes of Common Model Organisms

Organism	Genome size, Mb	No. of chromosomes	No. of genes	Sequencing status
Saccharomyces cerevisiae (yeast)	12	16	6400	Completed (1996)
Caenorhabditis elegans (worm)	97	5 autosomes, plus X	20,000	Completed (1998)
Drosophila melanogaster (fruit fly)	180	3 autosomes, plus X and Y	13,600	Completed (2000)
Xenopus tropicalis (frog)	1700	20 diploid	Unknown	Currently being sequenced
Danio rerio (zebrafish)	1700	50 diploid	Unknown	Currently being sequenced
Mus musculus (mouse)	2600	40 diploid	Unknown	Completed (2002)
Rattus norvegicus (rat)	2800	42 diploid	Unknown	First draft completed
Felis catus (cat)	3300	38 diploid	Unknown	"Moderate priority" on NHGRI list
Canis familiaris (dog)	2500	78 diploid	Unknown	First draft completed
Gallus gallus (chicken)	1200	78 diploid	Unknown	First draft completed
Sus scrofa (pig)	3000	38 diploid	Unknown	"Moderate priority" on NHGRI list
Bos taurus (cow)	3000	60 diploid	Unknown	Currently being sequenced

real-time PCR) (61) will play a crucial role in understanding complex biological systems.

The assessment of proteins may be referred to as *proteomics* and includes the measurement of proteins produced, identification of protein functions, and identification of protein-protein interactions (62). Because proteomic techniques provide a measure of RNA translation and post-translational modifications, they can be used in combination with DNA microarrays (i.e., measures of transcription) to generate more informative data. In addition to validating gene-expression data, measuring protein profiles and protein-protein interactions may identify critical posttranslational modifications that influence phenotypes. Because no single technology platform exists that satisfies all of the desired proteomic measurements (i.e., identify proteins, determine function, identify and interpret protein interactions), numerous tools are used in the field (62).

To date, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and mass spectrometry used in tandem has been the most widely used method for protein analysis (63). Proteins are separated by 2D-PAGE according to charge (i.e., isoelectric point) by isoelectric focusing in the first dimension and according to size (i.e., molecular mass) by sodium dodecyl sulfate (SDS)-PAGE in the second dimension (64). Proteins of interest can be isolated from the gel and identified via mass spectrometry. Because matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy can determine the mass of a protein or peptide with increasing sensitivity and ease of use, it has become the method of choice for protein identification by peptide-mass mapping (62). Because gel-based systems are technically complex, labor intensive, cost intensive, and fundamentally limited, liquid chromatography-mass spectroscopy (LC-MS) systems and protein microarrays are also being developed (62). Protein microarrays provide excellent tools for proteomics research with a systems-oriented

Table 2. On-Line Resources for Model Organisms

Species	Websites
Saccharomyces cerevisiae (yeast)	http://www.yeastgenome.org
Caenorhabditis elegans (worm)	http://www.wormbase.org
Drosophila melanogaster (fruit fly)	http://www.flybase.org
Xenopus tropicalis (frog)	http://xenbase.org
Danio rerio (zebrafish)	http://zfin.org http://www.nih.gov/science/models/zebrafish
Mus musculus (mouse)	http://www.informatics.jax.org http://www.ensembl.org/Mus_musculus
Rattus norvegicus (rat)	http://www.hgsc.bcm.tmc.edu/projects/rat http://rgd.mcw.edu
Felis catus (cat)	http://rex.nci.nih.gov/lgd/cat/catgenome.htm
Canis familiaris (dog)	http://www.dogmap.ch http://www.fhcrc.org/science/dog_genome
Gallus gallus (chicken)	http://poultry.mph.msu.edu http://www.genome.iastate.edu/chickmap
Sus scrofa (pig)	http://www.ansci.uiuc.edu/labs/schook http://www.genome.iastate.edu/pig http://www.marc.usda.gov/genome/genome.html
Bos taurus (cow)	http://www.marc.usda.gov/genome/genome.html http://www.hgsc.bcm.tmc.edu/projects/bovine

approach because only proteins of interest are measured. However, these chip-based proteomic systems have unique challenges, as prior knowledge of the proteins to be studied and appropriate affinity reagents is required (63). *Shotgun proteomics,* the combination of LC-MS and sequence database searching, has also been used for analyzing complex mixtures of proteins (65). Although effective LC-MS requires fractionation techniques to reduce complexity and isotope-coded affinity tags for quantitation, it is another reasonable alternative to gel-based techniques.

As scientists identify gene products and functions, proper gene nomenclature and classification must be performed to link with specific phenotypes. A standard method of naming and classifying genes is crucial for comparative research. The field of comparative genomics and proteomics would be complete chaos without the use of standardized methodology because a single gene product can have several molecular functions (phenotype), many gene products can share a single molecular function, and a gene may play different roles in different organisms. The Gene Ontology (GO) consortium (http://www.geneontology.org) was established to produce a controlled vocabulary that is applicable to all organisms and to establish guidelines for classifying gene function. According to GO, the function of a gene product may be described by its role in a biological process, molecular function, or cellular component. Because a gene product may have one or more functions in regard to biological process, molecular function, or cellular component, it may be classified under one or more of these categories. The GO molecular function terms represent activities that perform actions, but do not specify the location or in what context it takes place. GO biological process terms refer to one or more ordered assemblies of molecular functions, but are different than a pathway (e.g., signal transduction). Finally, GO cellular component terms describe the component of a cell and anatomical structure or gene product group of which it is a part.

The implementation of high-throughput techniques generates genomic data that challenge both scientists and clinicians, transforming it into relevant biological resources. Effectively navigating information from the large number and size of the molecular databases currently on the Internet is daunting. The first challenge is finding the most appropriate database(s) for the task at hand. The second challenge is to keep up with the latest information uploaded onto that site, as many are regularly updated with considerable amounts of new data. In fact, the amount of genomic data in the public database at the National Center for Biotechnology Information doubles every 18 months (66). Although some databases have a broad scope and contain information that is useful for a wide range of biological scientists, others are designed to focus on metabolic pathways or genes of a specific disease and are useful to scientists studying in that field of research. Baxevanis (67) assembled a list of approximately 400 high-quality molecular biology databases that may help scientists and clinicians choose which database is the most appropriate.

Although it has not yet reached the bedside, functional genomics and proteomics have numerous clinical applications that will someday aid in the development of early detection devices, diagnostics, and therapeutics of complex diseases. Currently, most complex diseases are detected in late stages and, consequently, have poor prognoses. Therefore, a great need exists for screening tools that can detect diseases at early stages when more intervention strategies are available and survival rate is greater. The importance of early detection may be exemplified using ovarian cancer survival rates and stage detected: Patients having ovarian cancer detected in Stage I have a very high 5-year survival rate of 95%, compared with cases detected in late stages who have a 5-year survival rate of 35% to 40% (68–70). Although an effective clinical biomarker should be measurable in an accessible bodily fluid such as serum, urine, or saliva (71), tissue biopsies are often required as a starting point. Using advanced computer algorithms, researchers are already beginning to identify serum or tissue gene expression and protein profiles or "signatures" of diseases for early detection and diagnostic purposes. Researchers focused on diseases such as inflammatory bowel disease (72); hepatic carcinoma and liver diseases (73); and breast (74), prostate (75), and ovarian (76) cancers are already using gene expression and protein profiles for this purpose. In addition to early detection, disease signatures may provide information regarding the event(s) that initiated the disease and the movement toward developing effective prevention and intervention strategies.

Perspectives

Although the incidences of complex diseases continue to rise at alarming rates, recent advances in molecular biology are now providing the tools and knowledge required to effectively study them. Genome sequence data is aiding in the development of ideal animal models by improving the efficiency of current enabling technologies (e.g., conditional transgenesis). These advances will soon allow the conditional regulation of specific genes or loci in any desired species, creating an ideal animal disease model that is genetically amenable and biologically relevant to humans. By identifying homologous genes across species, comparative genomics enables the effective translation of animal model data to humans. Continued use of the GO guidelines will be essential in comparative genomic research. High-throughput functional genomic and proteomic techniques that generate vast amounts of data provide a global view of gene expression and protein profiles. As this field matures, the effects of complex gene and protein interactions on phenotype will be realized. Finally, bioinformatics and mathematical models that interpret functional data will be essential in identifying disease biomarkers for early use in detecting, diagnosing, and treating disease. If the remark-

able advances in molecular biology and genomics over the past decade are any indication of the future, clinicians may soon have the ability to detect and diagnose a complex disease in an initial stage and extend and improve quality of life through personalized therapies.

We thank E. Forsberg at Infigen for his contributions and assistance.

1. National Vital Statistics Reports. Estimated life expectancy at birth in years, by race and sex: death-registration states, 1900-28, and United States, 1929-2000. National Vital Statistics Report 52:33–34. 2004.
2. National Vital Statistics Reports 2000. Deaths: leading causes for 2000. National Vital Statistics Report 50:1–86. 2002.
3. Tumbleson M, Schook LB, Eds. Advances in Swine in Biomedical Research. New York: Kluwer Academic, 1996.
4. National Research Council. Biomedical Models and Resources: Current Needs and Future Opportunities. Washington, D.C.: National Academy Press, 1998.
5. Soucek P, Gut I. Cytochromes P-450 in rats: structures, functions, properties and relevant human forms. Xenobiotica 22:83–103, 1992.
6. Dow JAT, Davies SA. Integrative physiology and functional genomics of epithelial function in a genetic model organism. Physiol Rev 83:687–729, 2003.
7. Barr MM. Super models. Physiol Genomics 13:15–24, 2003.
8. Fleury C, Pampin M, Tarze A, Mignotte B. Yeast as a model to study apoptosis. Biosci Rep 22:59–79, 2002.
9. Hrabé de Angelis M, Flaswinkel H, Fuchs H, Rathkolb B, Soewarto D, Marschall S, Heffner S, Pargent W, Wuensch K, Jung M, Reis A, Richter T, Alessandrini F, Jakob T, Fuchs E, Kolb H, Kremmer E, Schaeble K, Rollinski B, Roscher A, Peters C, Meitinger T, Strom T, Steckler T, Holsboer F, Klopstock T, Gekeler F, Schindewolf C, Jung T, Avraham K, Behrendt H, Ring J, Zimmer A, Schughart K, Pfeffer K, Wolf E, Balling R. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat Genet 25:444–447, 2000.
10. Nolan PM, Peters J, Strivens M, Rogers D, Hagan J, Spurr N, Gray IC, Vizor L, Brooker D, Whitehill E, Washbourne R, Hough T, Greenaway S, Hewitt M, Liu X, McCormack S, Pickford K, Selley R, Wells C, Tymowska-Lalanne Z, Roby P, Glenister P, Thornton C, Thaung C, Stevenson J-A, Arkell R, Mburu P, Hardisty R, Kiernan A, Erven A, Steel KP, Voegeling S, Guenet J-L, Nickols C, Sadri R, Naase M, Isaacs A, Davies K, Browne M, Fisher EMC, Martin J, Rastan S, Brown SDM, Hunter J. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nat Genet 25:440–443, 2000.
11. Stanford WL, Cohn JB, Cordes SP. Gene-trap mutagenesis: past, present, and beyond. Nat Rev Genet 2:756–768, 2001.
12. Bockamp E, Maringer M, Spangenberg C, Fees S, Fraser S, Eshkind L, Oesch F, Zabel B. Of mice and models: improved animal models for biomedical research. Physiol Genomics 11:115–132, 2002.
13. Lewandoski M. Conditional control of gene expression in the mouse. Nat Rev Genet 2:743–755, 2001.
14. Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature 292:154–156, 1981.
15. Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78:7634–7638, 1981.
16. Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature 317:230–234, 1985.
17. Chen Y-T, Bradley A. A new positive/negative selectable marker, puΔtk, for use in embryonic stem cells. Genesis 28:31–35, 2000.
18. O'Gorman S, Fox DT, Wahl GM. Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science 251:1351–1355, 1991.
19. Sauer B, Henderson N. Cre-stimulated recombination at loxP-containing DNA sequences placed into the mammalian genome. Nucleic Acids Res 17:147–161, 1989.
20. Gossen M, Bujard H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A 89:5547–5551, 1992.
21. Ornitz DM, Moreadith RW, Leder P. Binary system for regulating transgene expression in mice: targeting int-2 gene expression with yeast GAL4/UAS control elements. Proc Natl Acad Sci U S A 88:698–702, 1991.
22. Gossler A, Joyner AL, Rossant J, Skarnes WC. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science 244:463–465, 1989.
23. McClive P, Pall G, Newton K, Lee M, Mullins J, Forrester L. Gene trap integrations expressed in the developing heart: insert site affects splicing of the PT1-ATG vector. Dev Dyn 212:267–276, 1998.
24. Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet 2:769–779, 2001.
25. Zhang Y, Buchholz F, Muyrers JPP, Stewart AF. A new logic for DNA engineering using recombination in *Escherichia coli*. Nat Genet 20:123–128, 1998.
26. Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc Natl Acad Sci U S A 97:5978–5983, 2001.
27. Swaminathan S, Ellis HM, Waters LS, Yu D, Lee E-C, Court DL, Sharan SK. Rapid engineering of bacterial artificial chromosomes using oligonucleotides. Genesis 29:14–21, 2000.
28. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811, 1998.
29. Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. Science 297:2056–2060, 2002.
30. Zeng Y, Cullen BR. RNA interference in human cells is restricted to the cytoplasm. RNA 8:855–860, 2002.
31. Kawasaki H, Taira K. Short hairpin type of dsRNAs that are controlled by $tRNA^{Val}$ promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells. Nucleic Acids Res 31:700–707, 2003.
32. Dykxhoorn DM, Novina CD, Sharp PA. Killing the messenger: short RNAs that silence gene expression. Nat Rev Mol Cell Biol. 4:457–467, 2003.
33. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494–498, 2001.
34. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperien P, Ahringer J. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature 421:231–237, 2003.
35. Kiger AA, Baum B, Jones S, Jones MR, Coulson A, Echeverri C, Perrimon N. A functional genomic analysis of cell morphology using RNA interference. J Biol 2:27, 2003.
36. Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL, Bernards R. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature 428:431–437, 2004.
37. Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, Balija V, O'Shaughnessy A, Gnoj L, Scoble K, Chang K, Westbrook T, Cleary M, Sachidanandam R, McCombie WR, Elledge SJ, Hannon GJ. A resource for large-scale RNA-interference-based screens in mammals. Nature 428:427–431, 2004.

38. Brown SDM, Balling R. Systematic approaches to mouse mutagenesis. Curr Opin Genet Dev 11:268–273, 2001.
39. Brown SDM, Peters J. Combining mutagenesis and genomics in the mouse—closing the phenotype gap. Trends Genet 12:433–435, 1996.
40. Harris A. Towards an ovine model of cystic fibrosis. Hum Mol Genet 6:2191–2193, 1997.
41. Waters DJ, Sakr WA, Hayden DW, Lang CM, McKinney L, Murphy GP, Radinsky R, Ramoner R, Richardson RC, Tindall DJ. Workgroup 4: spontaneous prostate carcinoma in dogs and nonhuman primates. Prostate 36:64–67, 1998.
42. Meyn MS. Ataxia-telangiectasia, cancer and the pathobiology of the ATM gene. Clin Genet 55:289–304, 1999.
43. Barlow C, Hirotsune S, Paylor R, Liyanage M, Eckhaus M, Collins F, Shiloh Y, Crawley JN, Ried T, Tagle D, Wynshaw-Boris A. Atm-deficient mice: a paradigm of ataxia telangiectasia. Cell 86:159–171, 1996.
44. Elson A, Wang Y, Daugherty CJ, Morton CC, Zhou F, Campos-Torres J, Leder P. Pleiotropic defects in ataxia-telangiectasia protein-deficient mice. Proc Natl Acad Sci U S A 93:13084–13089, 1996.
45. Li W-H, Ellsworth DL, Krushkal J, Chang BH-J, Hewitt-Emmett D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol Phylogenet Evol 5:182–187, 1996.
46. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Bechstrom-Sternberg SM, Margulies EH, Blanchette M, Siepel AC, Thomas PJ, McDowell JC, Maskeri B, Hansen NF, Schwartz MS, Weber RJ, Kent WJ, Karolchik D, Bruen TC, Bevan R, Cutler DJ, Schwartz S, Elnitski L, Idol JR, Prasad AB, Lee-Lin S-Q, Maduro VVB, Summers TJ, Portnoy ME, Dietrich NL, Akhter N, Ayele K, Benjamin B, Cariaga K, Brinkley CP, Brooks SY, Granite S, Guan X, Gupta J, Haghighi P, Ho S-L, Huang MC, Karlins E, Laric PL, Legaspi R, Lim MJ, Maduro QL, Masiello CA, Mastrian SD, McCloskey JC, Pearson R, Stantripop S, Tiongson EE, Tran JT, Tsurgeon C, Vogt JL, Walker MA, Wetherby KD, Wiggins LS, Young AC, Zhang L-H, Osoegawa K, Zhu B, Zhao B, Shu CL, De Jong PJ, Lawrence CE, Smit AF, Chakravarti A, Haussler D, Green P, Miller W, Green ED. Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424:788–793, 2003.
47. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC. The dog genome: survey sequencing and comparative analysis. Science 301:1898–1903, 2003.
48. Coleman RA. Of mouse and man—what is the value of the mouse in predicting gene expression in humans? Drug Discov Today 8:233–235, 2003.
49. Larsson PO, Mosbach K. Affinity precipitation of enzymes. FEBS Lett. 98:333–338, 1979.
50. Fields S, Song O. A novel genetic system to detect protein-protein interactions. Nature 340:245–246, 1989.
51. Novick P, Osmond BC, Botstein D. Suppressors of yeast actin mutations. Genetics 121:659–674, 1989.
52. Bender A, Pringle JR. Use of a screen for synthetic lethal and multicopy suppressee mutants to identify two new genes involved in morphogenesis in Saccharomyces cerevisiae. Mol Cell Biol 11:1295–1305, 1991.
53. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SPA. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A 91:5022–5026, 1994.
54. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470, 1995.
55. Bowtell DDL. Options available—from start to finish—for obtaining expression data by microarray. Nat Genet 21(Suppl 1):25–32, 1999.
56. Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J. A concise guide to cDNA microarray analysis. Biotechniques 29:548–562, 2000.
57. Holloway AJ, van Laar RK, Tothill RW, Bowtell DDL. Options available—from start to finish—for obtaining data from DNA microarrays II. Nat Genet 32(Suppl):481–489, 2002.
58. Stears RL, Martinsky T, Schena M. Trends in microarray analysis. Nat Med 9:140–145, 2003.
59. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FCP, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat Genet 29:365–371, 2001.
60. Churchill GA. Fundamentals of experimental design for cDNA microarrays. Nat Genet 32(Suppl):490–495, 2002.
61. Chuaqui RF, Bonner RF, Best CJM, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krisman DB, Tangrea MA, Ahram A, Linehan WM, Knezevic V, Emmert-Buck MR. Post-analysis follow-up and validation of microarray experiments. Nat Genet 32(Suppl):509–514, 2002.
62. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. Nat Genet 33(Suppl):311–323, 2003.
63. MacBeath G. Protein microarrays and proteomics. Nat Genet 32(Suppl):526–532, 2002.
64. Daniel H. Genomics and proteomics: importance for the future of nutrition research. Br J Nutr 87:S305–S311, 2002.
65. Wolters DA, Washburn MP, Yates JR III. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem 73:5683–5690, 2001.
66. Pennisi E. Systems biology: tracing life's circuitry. Science 302:1646–1649, 2003.
67. Baxevanis AD. The molecular biology database collection: 2003 update. Nucleic Acids Res 31:1–12, 2003.
68. Jacobs IJ, Skates SJ, MacDonald N, Menon U, Rosenthal AN, Davies AP, Woolas R, Jeyarajah AR, Sibley K, Lowe DG, Oram DH. Screening for ovarian cancer: a pilot randomized controlled trial. Lancet 353:1207–1210, 1999.
69. Menon U, Jacobs IJ. Recent developments in ovarian cancer screening. Curr Opin Obstet Gynecol 12:39–42, 2000.
70. Cohen LS, Escobar PF, Scharm C, Glimco B, Fishman DA. Three-dimensional power doppler ultrasound improves the diagnostic accuracy for ovarian cancer prediction. Gynecol Oncol 82:40–48, 2001.
71. Petricoin EF, Liotta LA. Clinical applications of proteomics. J Nutr 133(Suppl 7):2476S–2484S, 2003.
72. Barceló-Batllori S, André M, Servis C, Lévy N, Takikawa O, Michetti P, Reymond M, Felley-Bosco E. Proteomic analysis of cytokine induced proteins in human intestinal epithelial cells: implications for inflammatory bowel diseases. Proteomics 2:551–560, 2002.
73. Steel LF, Shumpert D, Trotter M, Seeholzer SH, Evans AA, London WT, Dwek R, Block TM. A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. Proteomics 3:601–609, 2003.
74. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530–536, 2002.
75. Malins DC, Johnson PM, Barker EA, Polissar NL, Wheeler TM, Anderson KM. Cancer-related changes in prostate DNA as men age and early identification of metastasis in primary prostate tumors. Proc Natl Acad Sci U S A 100:5401–5406, 2003.
76. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359:572–577, 2002.