

David L. Mattson**1. INTRODUCTION**

The sequencing of the human genome has provided the scientific community with the “roadmap of life” (1,2), yet the understanding of this map is dependent on the elucidation of the function of the approx 30,000 to 35,000 genes that have been predicted/identified from the sequence of the human genome. As recently reviewed, however, only one-third of these genes have any inferred function ascribed to them, whereas approximately one-sixth or less have a confirmed action (3,4). Moreover, it is estimated that there are as many as 100,000 distinct proteins with presumably distinct biological function. The challenge facing scientists in the post-genome era is to construct a model of the organism that includes gene sequence and expression, protein structure and function, the molecular interactions that occur between proteins and other molecules to create pathways, and the combination of the many complex pathways that results in functioning cells, tissues, and organisms. Only by addressing and meeting this challenge can science utilize the vast potential of the human genome project to link gene to function in health and disease. To begin to address this problem, a combination of automated technologies, novel experimental strategies, and the newly available genomic sequence data are permitting scientists to begin to assemble the individual components of biological systems into a map of the human organism that will provide an understanding of the function of different genes and gene pathways.

The term “genomics” emerged 10 to 15 yr ago and describes the variety of technological and computational approaches employed to examine the structure and function of DNA of large numbers of genes in contrast to the study of a single gene or gene family (5). “Functional” or “physiological genomics” is a term that has been used to describe the multitude of approaches employed to identify and elucidate the functional importance of different genes in physiological and pathophysiological conditions. These approaches broadly encompass scientific disciplines aimed at linking nucleic acid sequence, protein structure, protein function, and complex, integrated biological function.

Function, in biological terms, is understood on many different levels depending on the context (6). A molecular biologist may consider the identification of an unknown gene that encodes a specific cell surface receptor as a functional analysis, but a biochemist might consider function to be the affinity of different ligands for the same receptor. In contrast, a structural biologist may consider the three-dimensional structure of the receptor protein

as function, whereas a cell biologist might consider the function of the receptor in terms of the intracellular pathways affected. Similarly, a physiologist might only consider the function of the receptor in terms of the integrated effects of stimulation of the receptor on a particular organ system. Each of these approaches, separately or combined, in the context of assigning function to genomic information, constitute studies in functional genomics. The integration of many or all of these disciplines and approaches, however, along with the computational proficiency to assemble the different pieces into a model of the whole organism most fully defines functional genomics.

Although a large variety of approaches are used to address questions in functional genomics, two general approaches are utilized (3). One approach is a continuation of classical, hypothesis-based research where the importance of a single gene, molecule, or pathway is specifically addressed by the experimental design of the study. A second approach, which takes particular advantage of high-throughput genomic resources, is known as discovery-based research and is performed to determine an unknown gene or genes that are linked to a problem of functional importance. Despite the fact that these two experimental approaches are quite different, they are not necessarily mutually exclusive. For instance, it would be predicted that a discovery-based project that identified a candidate gene as the cause of a disease process would perform a number of very specific, hypothesis-based studies to affirm the functional importance of the candidate gene. The discovery-based and hypothesis-based approaches can therefore complement each other under different circumstances.

Regardless of the experimental approach taken, the combined use of methodologies that manipulate or serve to otherwise describe the genome, along with assays of messenger RNA (mRNA) expression, protein function, protein pathways, cell-signaling pathways, cellular function, tissue function, or organism function have all come to be termed “physiological genomics” or “functional genomics.” Moreover, scientists performing functional genomics experiments utilize a wide variety of technologies and resources borrowed from a number of different fields of study in the life sciences. The need to utilize a diverse array of techniques in functional genomics experiments typically leads to large groups of scientists working in collaboration to address a specific question; this also requires biologists to have a working knowledge of the methodology and terminology of each of the different disciplines.

2. GENERAL APPROACHES TO PHYSIOLOGICAL GENOMICS: HYPOTHESIS VS DISCOVERY-BASED RESEARCH

Living organisms are complex biological systems with billions of molecules interacting simultaneously in an organized manner to sustain life. A traditional approach favored by scientists interested in understanding biological systems was to reduce these systems to their simplest common parts. By quantifying the actions and interactions of the individual components, it was hoped that the individual constituents could be reassembled and permit an understanding of the nature of the whole organism. This approach in which the biological role of an individual molecule is systematically addressed is often described as a hypothesis-based or reductionist approach. To address a specific disease process with this scientific approach, biologists first characterize the pathophysiology of a certain condition. Armed with this information, the genetic underpinnings of the disease process can then be identified by first deducing a number of genes/gene products that are likely candidates as

the mediators of the disease process. The role of the individual gene is then tested using a variety of methodologies to determine if the pharmacological or genetic deletion or manipulation of the normal gene leads to phenotypes consistent with the disease process. Finally, if there is a positive association between the disease phenotype and the blockade of the gene product, further study is performed to understand the particular genetic mutations responsible for the pathological phenotype.

This popular “reductionist” approach to biological science has revealed an enormous amount of information regarding the mechanisms whereby biological systems function. One weakness of this approach, however, has been the inability to fully understand the complicated nature of the integrated organism because of insufficient knowledge regarding the innumerable complex interactions that occur in an organism and the inability of currently available technology to reconstruct an organism from its individual components.

The public and privately funded projects to sequence the human genome and the evolution in technology that has driven and arisen from these projects have led to an alteration in this traditional experimental paradigm. Although many researchers continue a hypothesis-based approach to research in which experiments are specifically designed to determine the functional role of an individual gene or gene product, the genomic revolution has permitted scientists to begin discovery-based studies in which large parts of the entire genome are scanned to determine the genetic basis of disease. In discovery-based research, scientists utilize approaches and technologies that permit the entire genome to be scanned or randomly altered, and functional experiments are then performed to determine the association of the affected genes with disease. Two examples of discovery-based research are the genetic linkage analysis of quantitative trait loci (QTL) and mutagenesis screens.

In the QTL segregation analysis, two inbred animal strains, one normal and one affected with the disease of interest (the phenotype), are initially bred (7–9). An example of this breeding strategy in inbred rats is presented in Fig. 1, with each parent homozygous for a different allele at a representative genetic locus (AA or BB). The offspring in the first generation of this genetic cross, known as the F_1 generation, are identical because both parental strains are inbred; this is represented by heterozygosity, AB, at the representative locus. Further intercrossing within the F_1 generation produces an F_2 generation, in which the genetic information from the two parental strains has been scrambled due to recombination during meiosis (9). These rats in the F_2 generation are then carefully phenotyped for traits associated with the disease. In addition, a systematic, genome-wide scan is performed on each rat of the F_2 generation utilizing anonymous genetic markers that are polymorphic between the two parental strains. A mathematical analysis, known as a linkage analysis, is then performed on the data to determine if the inheritance of the genetic markers correlates with a given quantitative trait in the F_2 generation. The linkage analysis utilizes many genetic markers distributed throughout the genome and provides a powerful tool for the discovery of novel genes or gene pathways in animal models of disease.

A second example of discovery-based research in physiological genomics are the mutagenesis projects that have been carried out in mice, zebrafish, and other species. An example of this approach, as performed in zebrafish to screen for recessive mutations, is schematized in Fig. 2 (10). In this particular mutagenesis screen, a parental male zebrafish was exposed to ethylnitrosurea (ENU), a mutagenic agent. The dose of ENU, an alkylating agent that acts as a powerful mutagen, is titrated to induce random point mutations in the spermatogonia of the treated animal. Subsequent breeding of the mutagenized male to

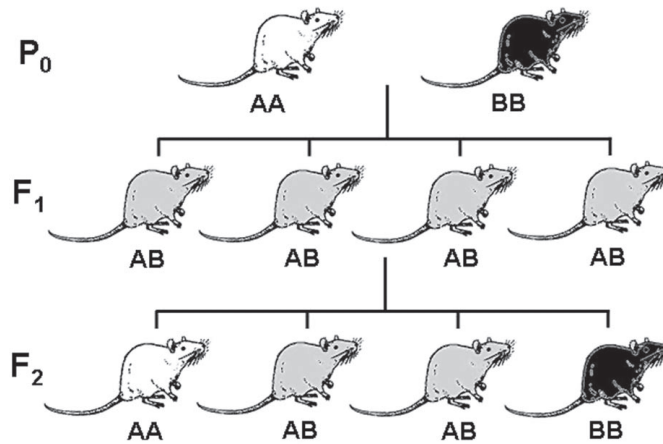


Fig. 1. Schematic illustrating an F₂ intercross. Two inbred parental strains (P₀) with homozygous genotypes for different alleles are initially crossed to yield an F₁ generation. The F₁ animals will be identical, but intercrossing of the F₁ generation yields an F₂ generation with the segregation of different alleles. Phenotyping and genotyping of the F₂ animals allows the mapping of quantitative trait loci. See text for details.

a normal female yields an F₁ generation. The F₁ generation of this cross was then interbred to yield heterozygote carriers of the mutated genes in the F₂ generation. The F₂ generation was then interbred to yield homozygotic and heterozygotic carriers of the mutations in the F₃ generation. In this mutagenesis study, careful screening of the F₃ embryos was then performed. Any signs of mutations were judged to be owing to a single-gene, recessive mutation if it was evident in 25% of the growing embryos and larvae. Once a mutation has been identified and confirmed, further phenotyping and genotyping protocols are performed to identify the gene or genes affected that lead to this altered trait (10).

Both the QTL analysis and the mutagenesis screen are examples of discovery-based science in which there are no initial assumptions in regard to the gene, molecule, cell, or tissue type affected. The entire experiment is performed to discover novel genes or gene pathways important in normal function and in cases of pathology. This approach contrasts with hypothesis-based research where the importance of a specific gene or molecule is specifically addressed by the experimental design of the study. Regardless of the experimental approach, the combined use of methodologies that manipulate or serve to otherwise describe the genome, along with functional assays of protein, pathway, cellular, tissue, or organism function have all come to be termed “physiological” or “functional genomics.”

3. MODELS AND APPROACHES COMMONLY USED IN PHYSIOLOGICAL GENOMICS EXPERIMENTS

As one might expect, there is a diverse array of methodology and technology utilized as well as a large assortment of different species that are studied. Examples of different methodologies employed include expression profiling, proteomics, chromosome transfer, gene deletion, gene delivery, gene silencing, mutagenesis, and a variety of gene-mapping studies. These types of experiments are performed in a number of species including rats, mice, primates, dogs, yeast, zebrafish, and *C. elegans*. These approaches and models have been

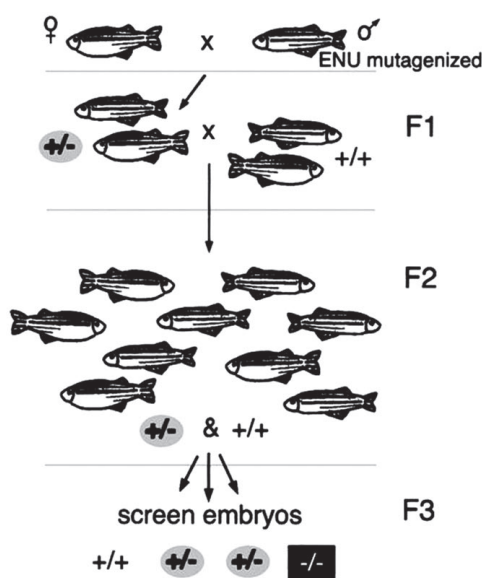


Fig. 2. Outline of zebrafish mutagenesis screens. *N*-Ethyl-*N*-nitrosourea (ENU) was used to mutagenize spermatogonia of parental males. Outcrosses were performed with wild-type females to produce an F_1 generation, each F_1 fish possessing a unique set of mutations. Sibling F_1 matings created the F_2 generation, and the mutations were driven to homozygosity in the F_3 embryos. (Reproduced from ref. 10 with permission.)

utilized to explore the basis of both monogenetic and polygenetic disease processes. A comprehensive review of all of the techniques and methodologies employed in physiological genomics studies is beyond the scope of this chapter. Outlined here are a number of examples of approaches successfully utilized in functional genomics applications. Also included is a concluding section that describes a general approach to integrate these approaches in order to begin to link genes with function. This summary is by no means exhaustive, but is designed to provide the reader with a sense of the approaches currently in use. Those interested in more specific details and a further discussion of the different methodologies are referred to the source material referenced throughout the chapter.

4. GENE-MAPPING STUDIES

Gene mapping refers to the localization of a disease gene or genes on the gene map of the organism of interest. A number of approaches are used to map disease genes, but the basis for any mapping study is the phenomenon of recombination which produces new arrangements of genes. During meiosis, pieces of DNA are exchanged between homologous chromosomes by crossing-over. Recombination that occurs between two distinct positions or loci on a gene map results in a new combination of alleles (different forms of the same gene).

The position of genetic loci on the gene map is determined through the use of genetic markers (11). Genetic markers are typically short-tandem repeats of DNA sequences or single nucleotide polymorphisms that can be readily detected by the polymerase chain reaction. Other genetic markers are genes with known phenotypic expression and restriction fragment length polymorphisms. The combination of alleles on the genetic map, known as

the genotype, is then correlated to the measured function in the animals, known as the phenotype. A mathematical analysis of the relationship between the genotype and phenotype allows one to determine the degree of association of a certain marker with the genotype of interest.

A large proportion of the genetic variation that underlies disease phenotypes is caused by complex, polygenic interactions and is controlled by loci that have quantitative effects on the phenotype of interest. A phenotype that can be quantified, or a quantitative trait, is one that has a measurable variation owing to genetic and/or environmental influences. This variation can consist of discrete values or can have a continuous distribution. Importantly, quantitative traits contrast with qualitative traits, which are either present or absent (e.g., cystic fibrosis or sickle cell anemia). A QTL is a genetic locus that contains alleles that affect this genetic variation. These genetic loci are identified through a statistical analysis of the complex traits. Typically, the traits are affected by environmental as well as genetic factors.

Genetic-mapping strategies in experimental animals generally involve specific animal breeding approaches in which a genetic cross of two inbred strains (usually a normal or unaffected strain and a strain that demonstrates a disease phenotype) is made to examine the correlation between genetic loci and the disease phenotype. An example of such an approach is provided in Fig. 1. Mapping studies in human families will typically involve the close genotypic and phenotypic examination of a family whose members are affected with a particular disease. Coarse mapping of genetic loci to a chromosome segment using standard breeding strategies in animals or in human family studies will normally narrow a region of interest to 10 to 30 centimorgans (cM). A centimorgan is a unit of measure of recombination frequency; 1 cM equals a 1% chance that a marker at one genetic locus will be separated from a marker at a second locus owing to crossing over in a single generation. In humans, 1 cM is equivalent to approx 1 million base pairs, and would be predicted to contain 11 to 12 separate genes. Coarse mapping would, therefore, narrow a region to between 100 and 300 genes. Although this is an enormous reduction from the approx 35,000 potential genes, the number of possible genes affecting the phenotype of interest is too large to begin a gene-by-gene functional evaluation. A more detailed analysis, often termed "fine mapping" is then employed (11,12).

5. CHROMOSOME TRANSFER: CONSOMIC AND CONGENIC RATS

As described previously, a co-segregation analysis and mapping of QTL is largely a coarse-mapping function that can narrow a region of interest (ROI) in the genome to a section of DNA of approx 30 cM (11,12). Since a section of genomic DNA of this size would be predicted to have up to 300 genes, additional work is needed to reduce the number of candidate genes in the ROI. To further narrow ROI, a number of different breeding approaches can be used in mice and/or rats. These breeding approaches include a number of different backcrossing and intercrossing strategies that produce advanced intercross lines, recombinant inbred strains, congenic strains, or consomic strains, to name a few (11). The reader is referred to the literature for a more comprehensive review of these different methodologies (7,9,11,12). This section focuses on the use of consomic strains, although each of the previously described approaches has proven valuable.

Inbred consomic strains are developed by marker-assisted breeding to insert a complete chromosome from one inbred strain into the genetic complement of a recipient strain (4,5,7).

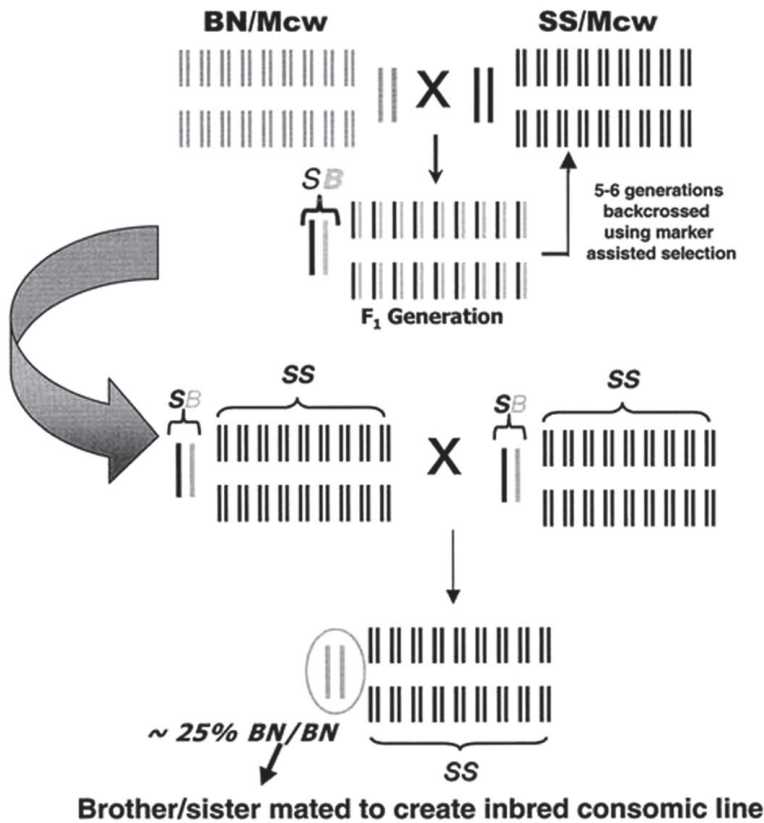


Fig. 3. The breeding and genotyping scheme required to develop an inbred consomic line. See text for details. The genome for the normotensive, salt-resistant BN/Mcw (Brown Norway) rat is shown in gray and the genome for the hypertensive SS/Mcw (Dahl Salt-Sensitive) rat is shown in black. (Reproduced from ref. 5 with permission.)

The breeding strategy to generate an inbred consomic rat line is illustrated in Fig. 3, in which an individual chromosome from the BN/Mcw (Brown Norway) rat is transferred to the SS/Mcw (Dahl Salt-Sensitive) background with the aid of marker-assisted breeding. In this strategy, the F₁ generation of an intercross between these two inbred parental lines is backcrossed to one of the parents, the SS/Mcw. Through the use of a total genome scan, which indicates which alleles are from which parental strain, subsequent progeny can be selected that possess the greatest numbers of homozygous alleles for the parental strain to which the backcross is occurring, the SS/Mcw, while maintaining the heterozygosity for the targeted chromosome. This strategy is employed with five to six subsequent backcrosses to the SS/Mcw parental strain until the progeny are isogenic for all SS/Mcw chromosomes except for the targeted chromosome, which is heterozygous. These rats can then be intercrossed to derive an F₂ generation, which has one chromosome homozygous for the BN/Mcw, whereas the rest of the genetic complement is homozygous for the SS/Mcw genetic background.

Once these consomic lines have been derived, they can be bred to provide stable inbred lines for physiological study. These consomic lines can be used to validate the significance of a QTL discovered in a mapping study, they permit the localization of physiological traits

to a chromosome without the necessity for further genotyping, they provide a genetically well-defined strain for physiological study, and they provide a basis for further fine mapping by the creation of congenic strains (4,5,7,13). Congenic strains are developed by integrating an individual piece of a chromosome containing a genomic region with genes of potential interest into the genomic background of the recipient strain (8,9). The creation of a set of congenic strains with overlapping regions can then be utilized to further refine the genetic region containing the QTL of interest. Although this and other breeding strategies can serve to narrow the QTL of interest to a genetic locus of 1–5 cM, there are still a relatively large number of possible candidate genes in the ROI. Further study is then necessary to identify the causative gene or genes.

The process of identifying the genes responsible for diseases in a QTL can utilize a number of different resources depending on the model (11). Positional cloning involves the use of linkage mapping to determine the location of a disease gene by employing additional genetic markers in the genetic ROI. This ROI is narrowed by relating genotype to phenotype with high resolution by utilizing congenic strains, recombinant inbred strains, or other breeding strategies. The DNA from the ROI can then be isolated with the use of various genomic resources and the disease gene can be sequenced with the mutation potentially identified.

6. GENE-EXPRESSION PROFILING

Changes in gene expression underlie many biological phenomena. The quantification of changes in gene expression by Northern blotting, *in situ* hybridization, reverse transcription coupled to the polymerase chain reaction, ribonuclease protection assays, and other methodologies have provided invaluable insight into the importance of mRNA changes during a variety of physiological maneuvers and under different pathophysiological conditions. These methods are limited for functional genomics applications, however, in that they are targeted to a single mRNA species which has been designated by the investigator as worthy of study. The advent of DNA microarray technology now permits the evaluation of gene expression changes in large numbers of genes in a single experiment (14–16).

The basic scheme behind a microarray experiment is quite simple and is schematized in Fig. 4 (14). Total RNA is isolated from two individual biological samples (organs, tissues, cells), which are to be compared. The total mRNA from each sample is reverse transcribed into individual complementary DNAs (cDNAs), and the cDNA from the different samples are labeled with different fluorescent dyes. The samples are then hybridized to a microarray slide which typically contains thousands of individual spots. Each spot on the array contains a single nucleic acid probe that is complementary to a known mRNA. The nucleic acid probes are cDNA or oligonucleotide probes that have been robotically arrayed on the glass slide. Following the hybridization and appropriate washing steps, a confocal laser scanner that can differentiate between the two labeling dyes is used to quantify the binding of the cDNAs from the different samples to the individual spots. Because the identity of the probe in each spot is known, and the samples have been labeled with different dyes, the ratio of intensity of binding to the different spots can then be used to assess the relative expression of individual transcripts in each sample.

This revolutionary technology permits the investigator to determine which genes are active in a specific cell or tissue at a specific time. Variations in mRNA expression can serve as

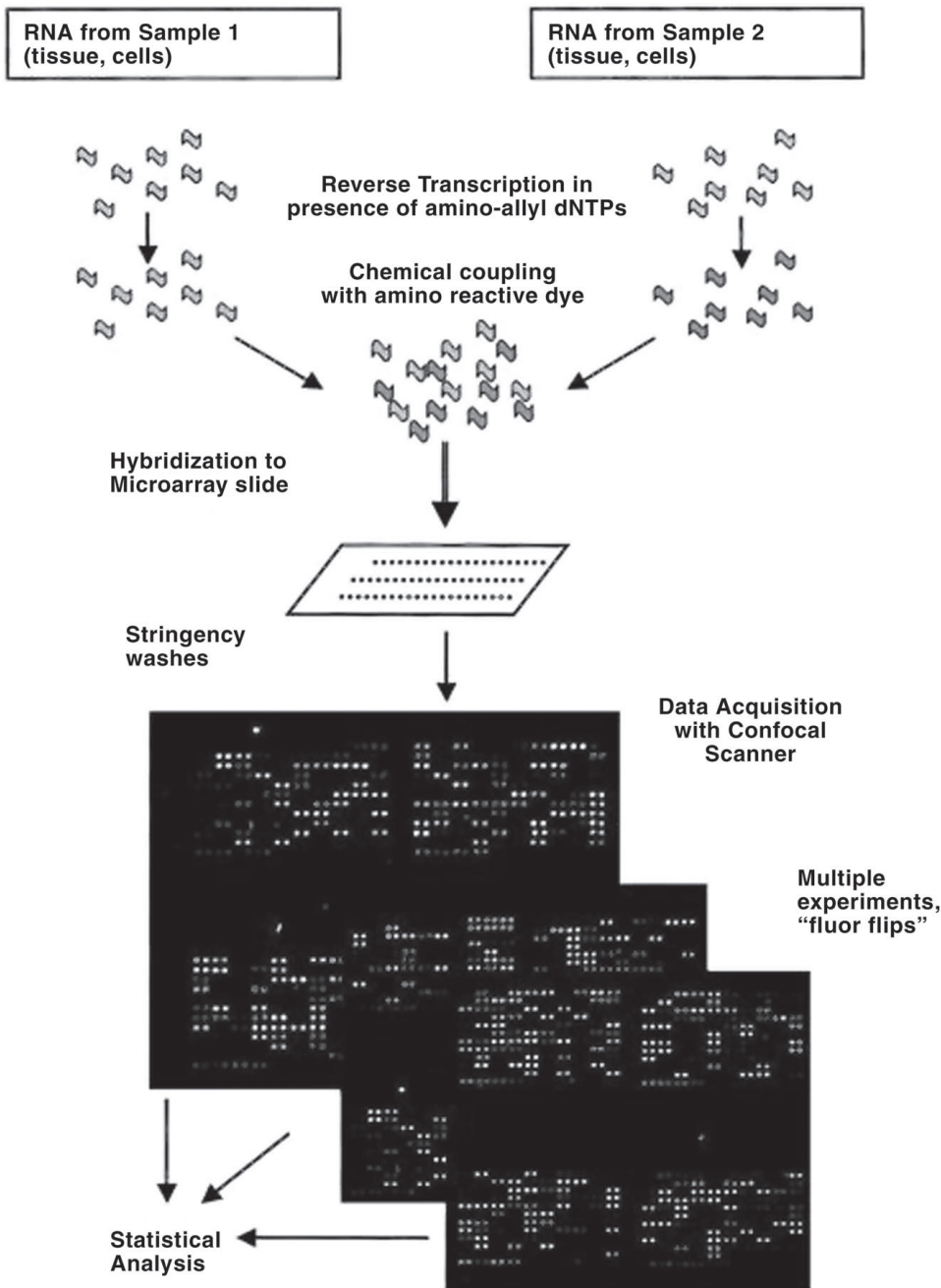


Fig. 4. Schematic demonstrating a DNA microarray experiment. Two RNA samples are reverse transcribed and labeled with fluorescent tags. The samples are then hybridized to a microarray slide and scanned to acquire the expression data. See text for details. (Reproduced from ref. 14 with permission.)

an important indicator of disease or predisposition to disease. By comparing gene expression patterns between cells or tissues from different environments, such as normal tissue compared to diseased tissue, specific genes, or gene pathways that play a role in pathophysiological processes can be identified. In the same manner, evaluation of expression profiles between parental and consomic or congenic animals can help identify the genes or gene pathways responsible for different disease processes (17).

Although this technology permits the rapid analysis of mRNA expression in different biological samples, it is important to note that the etiology of all disease processes cannot be explained on the basis of changes in RNA expression alone. Moreover, mRNA expression can be increased or decreased as a result of the disease process; expression profiling studies should therefore be carefully designed to discriminate between the changes that are causative in the disease process and those changes that result from the disease. Finally, it is important to verify the results obtained by expression profiling at the level of protein and biological function to confirm that the gene in question is indeed responsible for the observed change in function.

7. PROTEOMICS

Proteins are involved at some level in all aspects of biological function; the examination of the entire complement of proteins in a biological system has therefore become a rapidly expanding and evolving field of research. The field of study that addresses gene function at the level of proteins is known as proteomics. Proteomic studies include experiments aimed at understanding protein–protein interactions, studies designed to perform the large-scale identification of proteins and their numerous posttranslational modifications, and studies planned to quantify the individual levels of different proteins in health and disease. Each of these endeavors, but especially the identification and quantification of different proteins under different pathological conditions, can complement the other methodologies described in this chapter to address questions of scientific importance in functional genomic studies.

The identification and quantification of unknown proteins in a complex mixture is a multistep procedure that is approached using a number of different strategies depending upon the application (18–20). All of these procedures require initial steps to isolate the proteins from the tissue; a separate step involves the separation of the complex protein mixture into individual components; and the final step is then taken to identify the unknown protein or proteins. Although there are a variety of different approaches utilized to address this question, one commonly utilized procedure for protein identification is the coupling of two-dimensional (2D) gel electrophoresis with mass spectrometry to identify proteins expressed under different physiological conditions. The 2D gel electrophoresis process separates proteins based on two parameters, isoelectric point (charge) and molecular weight, in a polyacrylamide gel. The gel is then stained to reveal individual spots that contain a single protein with a unique isoelectric point and molecular weight. The intensity of staining of the individual spots may be quantified to provide an index of the amount of that particular protein present in the sample. The individual spots on the gel are then excised, digested with a protease such as trypsin, and subjected to mass spectrometry analysis. The mass spectrometry methodologies can then be used to identify the unknown protein using a process known as peptide mapping. Through this technique, the unknown protein, which has been digested with the protease, can be identified based on the

composition of its individual peptides. This identification is made by comparing the peptide composition of the unknown protein to the composition of proteins previously identified and stored in a database.

The use of 2D gel electrophoresis for protein separation, however, is problematic because there is often comigration of more than one protein to a single spot on a gel. In addition, only the most abundantly expressed proteins are resolved and detected by 2D gel electrophoresis so only a relatively small subset of the total cellular complement of proteins is detected by this technique. As an alternative to methods that employ 2D gel electrophoresis in the separation step, mass spectrometry methodologies have been coupled with liquid chromatography techniques to identify differentially expressed proteins. The application of different liquid chromatography separation strategies has greatly increased the number of proteins that can be identified.

Recent technological advancements have led to the possibility that unknown proteins can not only be identified but that the protein content of cells and tissues can be determined. Proteins have been incorporated with stable isotopes or labeled with different marker agents that enable the quantification of protein content by mass spectrometry. This approach takes advantage of the ability of mass spectrometry to differentiate chemicals of the same composition that are composed of stable isotopes of different masses; the ratio of the signal for such pairs can be used to calculate the abundance of each. This exciting approach, along with chip-based methods to determine differential expression of proteins, and strategies employed to examine protein–protein interactions and posttranslational modifications of proteins are some of the challenges facing the proteomics field (18,20).

Proteomics has an important place in the functional genomics approach to understand systems biology. The connection between the genes and gene expression and the function of biological systems and whole organisms are the proteins. A description of the proteome, the quantification of protein levels throughout development and during the cell cycle, and an understanding of protein pathways are all important components if the chart linking genes to biological function is to be assembled.

8. GENE TARGETING

Once a candidate gene has been identified, it is still imperative that the investigator confirm the role of this gene in the physiological or pathophysiological process under investigation. One experimental approach utilized in functional genomic studies to confirm the action of a candidate gene following a gene-mapping, mutagenesis, or other discovery-based study is gene targeting. In one of the most commonly used gene-targeting approaches, homologous recombination is performed in germ line-competent stem cells to disrupt a targeted gene in mice (21,22). Gene-disruption studies, also known as gene “knockout” experiments, are utilized to demonstrate the influence of complete deletion of a gene product. This approach permits scientists to examine function in animals which completely lack the functional gene product. Studies of this type can, therefore, be used to confirm the importance of a specific gene in biological processes. One other approach taken with homologous recombination strategies is to create animals in which a specific gene is duplicated or other genes are inserted into the genome of mice. Using this strategy, the functional influence of elevated levels of a native gene product, or the influence of elevated levels of a mutant gene product can be examined.

This innovative approach has led a large number of physiology labs to begin using mice as a model system. One concern with the use of gene-targeted mice, particularly the null mutant models, is that the targeted gene has been deleted from conception. There may, therefore, be developmental abnormalities in the affected animals or other components of the same or other biological pathways may be up- or downregulated to compensate for the genetic manipulation. In addition, the gene of interest will be manipulated in every cell of the animal's body, so the importance of a single molecular species in an individual cell type can be difficult to discern. To address this problem, strategies have been conceived that permit the knockout of a gene in a particular tissue. One approach to create a tissue-specific knockout utilizes the Cre/loxP system (23). Cre is a 38 kDa recombinase protein from a bacteriophage that mediates site-specific recombination between loxP sites (two 13 bp inverted repeats separated by an 8 bp spacer region). By combining this approach with a tissue-specific promoter, it is possible to create a mouse with a tissue-specific gene deletion. Moreover, this system can be combined with a tetracycline-sensitive element that will permit a tissue-specific deletion that can be controlled by the presence (administration) of tetracycline to the mouse (24–26). Using these strategies, the timing of the tissue-specific knockout condition can be controlled permitting the examination of genetic knockout animals that have had the gene product present throughout development and early life.

The advent of this revolutionary technology has permitted the examination of gene function in a large number of different fields of physiological study. The limiting factor to date in the use of gene-targeting strategies is that the mice are the only mammal from which germ line-competent stem cells have been isolated and successfully cultured. Although different strategies have been successfully used to create transgenic (9) and even knockout rat (with enu mutagenesis) models (27), the mouse is the only species that has been successfully and widely used for gene-targeting studies. Because mice have a small body size and most physiologists have limited experience in the study of function in these animals, the initial impact of this technology has been restricted to *in vitro* studies and simple, noninvasive, *in vivo* studies. Moreover, the availability of these models was limited to investigators with access to a transgenic facility and laboratory budgets that permitted the development of gene-targeted models. As such, the widespread application of mouse knockout and gene-targeted models to physiological questions has been somewhat limited. In the past several years, however, numerous technological advancements have occurred permitting the more thorough evaluation of physiological function *in vivo* in mice. Moreover, the Jackson Laboratory (www.jax.org/) and a number of other commercial vendors have served as repositories for gene targeted mice that can be purchased for study by the general scientific community.

9. SIRNA/ANTISENSE/PHARMACOLOGICAL APPROACHES

As discussed previously, gene targeting by homologous recombination has been successfully used to manipulate gene function in mice. Unfortunately, gene-targeting experiments are expensive and time-consuming and cannot be readily performed in all species. An alternative strategy to examine or confirm the functional importance of a candidate gene is an approach that uses standard pharmacological agents (i.e., receptor agonists and antagonists or enzyme inhibitors) to manipulate a particular gene product or pathway. This fairly straightforward approach can serve as a final confirmation of the importance of a particular molecule in normal function or in a disease process. The greatest weakness of this

traditional pharmacological approach, however, is that it is largely dependent on the availability, specificity, and selectivity of reagents specific for the candidate gene. It is, therefore, only possible to employ this approach when interested in examining a gene product or a biological system that has already been described in some detail and for which the appropriate reagents to manipulate the system have been developed. Because discovery research, by definition, should lead to the detection of novel genes and pathways, experimental approaches that can be applied to a greater number of possible candidate genes is required.

To circumvent the problems associated with the use of pharmacological agents, both small interference RNA (siRNA [28–30]) and antisense oligonucleotide technology (31) have been developed and utilized to selectively target genes or gene products for which standard pharmacological tools are not available. These reagents are small strands of RNA or DNA that are complementary to the mRNA of the targeted gene. These agents bind to the target RNA and create a double stranded complex. The double-stranded complex leads to either the arrest of translation (31) or cleavage of the double-stranded complex (28–30); regardless of the mechanism of action, the end result is gene silencing. This approach theoretically permits an investigator to create his or her own designer drugs to target any mRNA for which the sequence is known. The use of siRNA or antisense, along with the pharmacological approach described earlier, permits a phenotypic analysis both before and after loss of function in an otherwise normal animal without the often confounding developmental abnormalities that can occur in models in which a target gene has been deleted from conception.

10. BIOINFORMATICS

An additional and very important component of functional genomics studies is bioinformatics. Bioinformatics is the information science that employs computer technology, databases, and web-based resources to decipher and organize the vast amounts of information generated from genomics and proteomics projects (32). For example, bioinformatics scientists develop algorithms and other data management tools which facilitate functional genomics research by permitting the examination of large numbers of genes, proteins, and interacting networks of proteins. For any of the above described techniques and methodologies, if a high-throughput operation is desired, suitable data-processing, data-storage and -retrieval, and data-visualization systems must be in place. One of the most important tasks of the bioinformatics field is this critical task. As the volume and the complexity of the data sets increase, the role of bioinformatics will only increase. Without this critical resource, physiological genomics applications will not be possible.

11. COMPARATIVE GENOMICS

Animal models of disease are generally chosen based on pathophysiological characteristics shared with human disease. It is important, however, to ensure that the genetic basis for disease in the animal models correlates with the genetic basis of disease in humans. As such, comparative gene-mapping strategies, which are largely successful owing to the efforts of bioinformatics scientists, have been utilized to demonstrate that there are regions of the human genome that are evolutionarily conserved with other mammalian species. In fact, it has been demonstrated that QTLs for obesity, autoimmune disease, and hypertension lie in evolutionarily conserved regions in the rat, mouse, and human (7). These comparative mapping data justify the use of rodent and other species as models for human disease. Moreover,

these data indicate that scientists can utilize different species as they prepare to attach function to the genome. For instance, a mapping study performed in rats may indicate that a certain gene is present in a QTL judged important in a disease of interest. In the absence of reagents to manipulate the function of this gene in the rat, gene-targeting strategies could be justified in the mouse to demonstrate the importance of this gene to the disease phenotype. Finally, comparative mapping strategies could be utilized to demonstrate that a given QTL is conserved between the mouse, rat, and human to justify the experimental approach.

12. HIGH-THROUGHPUT PHENOTYPING

The term “phenotyping” has been used throughout this chapter to refer to the quantification of the observable or visible properties of an organism. As such, an organism’s phenotype is either its physical appearance and makeup or a specific trait. The ability to accurately and quickly measure a given phenotype is extremely important in functional genomics. All of the previously described techniques and methodologies are useful tools in functional genomic studies, but an experimental design which contains the most highly automated genotyping or expression array process will still be limited by the ability to accurately quantify the phenotype of the organisms to be studied. Some simple phenotypes such as height, weight, eye color, and coat color are easily measured and therefore may not impose a significant impediment to research progress. The quantification of more complex phenotypes, however, can pose a particular difficulty to those interested in quantifying complex traits. Examples of disease processes that require careful measurement include cardiovascular disease, diabetes, cancer, respiratory disease, sleep disorders, and other neurological disorders. The ability to accurately phenotype large numbers of animals for complex traits is therefore a tremendous challenge facing investigators participating in functional genomics studies.

As it has been recently noted, one of the greatest problems with obtaining accurate phenotypic information and comparing this information is the large number of different approaches to quantify similar phenotypes (33). Because phenotyping is often times based on methods that vary between different fields of study and different laboratories, there is a great need to systematically standardize phenotyping protocols. The scientific community will only be able to take advantage of the enormous wealth of genotypic data when comprehensive and truly high-throughput systems for phenotyping have been established. One such effort, known as the Mouse Phenome Project is headquartered at the Jackson Laboratory (34). In this project, quantifiable phenotypic characterization of a defined set of mouse strains by industrial and academic laboratories has begun with the data deposited in an accessible database (www.jax.org/phenome).

A number of groups have attempted to circumvent phenotyping limitations by developing new strategies and methods to measure functional parameters in different experimental models. The advent of gene manipulation in the mouse has led to a large effort to adapt functional measurements to this small mammal. Similarly, the relative advantages of performing ENU mutagenesis in zebrafish have by necessity led to advances in assays of function in these tiny animals. Several examples of groups developing phenotyping strategies that increase throughput in different models are the screening protocols developed in the National Institutes of Health Program for Genomic Applications to quantify heart, lung, and blood phenotypes in mice (<http://pga.jax.org/>) and rats (<http://pga.mcw.edu/>). Technological advances have, therefore, made the measurement of different indices of physiological function feasible even in minute animals and embryos. Despite the significant efforts and

advances that have increased the ability to characterize complex phenotypes in small animals (e.g., rats, mice, and zebrafish), however, the accurate quantification and measurement of different parameters in these small animals is fraught with difficulty. An extensive effort to increase the throughput of this important portion of the functional genomics strategy to a point that is equivalent to that which can be reached by many of the genomics and proteomics technologies described above will be required before functional genomics studies can reach their full potential.

13. INTEGRATIVE PHYSIOLOGY: PUTTING THE PIECES TOGETHER

It is clear that no single scientific field or methodology by itself will be sufficient to connect genes with proteins, cell function, organ and system function, and the intact organism. Instead, a combination of methodologies borrowed from different scientific disciplines will likely best serve the needs of an investigator. To fully integrate genomic data and functional data, we will initially require a structured list of genes and gene products. Following that enormous task, quantitative data regarding the complex interactions in cells will be desired; for instance, whole-scale analysis of mRNA and protein levels at different times throughout a cell's life and in different cell types will be necessary. Furthermore, knowledge in regard to the interactions between different gene products (i.e., rate constants, affinities, etc.) with subsequent mathematical modeling of the thousands of different parameters that vary at different times throughout the life of different cells is needed to begin to understand function in a single cell or cell type. Once that level of complexity has been mastered, the interactions among the trillions of cells and the myriad signaling mechanisms and pathways that combine to make up an organism must be similarly analyzed and quantified to be understood (5,19,35–37).

Although there are significant efforts underway to understand and quantify gene expression and proteins on a large scale, the assembly of these individual parts into functional cells and into individual organisms will require a much greater understanding of cellular and systems physiology than is currently recognized. As a logical next step in the quest to link genes with function, the quantitative description of functional parameters in humans and in experimental models is being considered. In fact, as discussed previously, a database of physiological parameters from different mouse strains has recently been established (34). Despite this effort, however, the complex interactions among cells, organs, and organ systems in the intact organism must still be described and quantified in order to be placed into proper context in terms of organism function.

An early attempt to describe and quantify the numerous positive and negative feedback systems involved in the regulation of the circulation was made more than 30 yr ago by Guyton and colleagues (38,39). As depicted in Fig. 5, the complex interrelationship among blood vessels, the heart, the lungs, the autonomic nervous system, the kidneys, numerous endocrine controllers of blood pressure, the metabolic needs of tissues, and a host of other factors were included in the model. Using a systems analysis approach, the model is based on the quantitative analysis of positive and negative feedback systems, the temporal response of the different feedback pathways, and the relationship between the different pathways. This quantitative analysis of circulatory function, which involved approx 400 physiological phenomena and their interrelationships, was gradually constructed over a 12-yr period using the knowledge available at that time (38,39). Each of the blocks in the diagram is represented by one or more mathematical equations that were derived from experiments

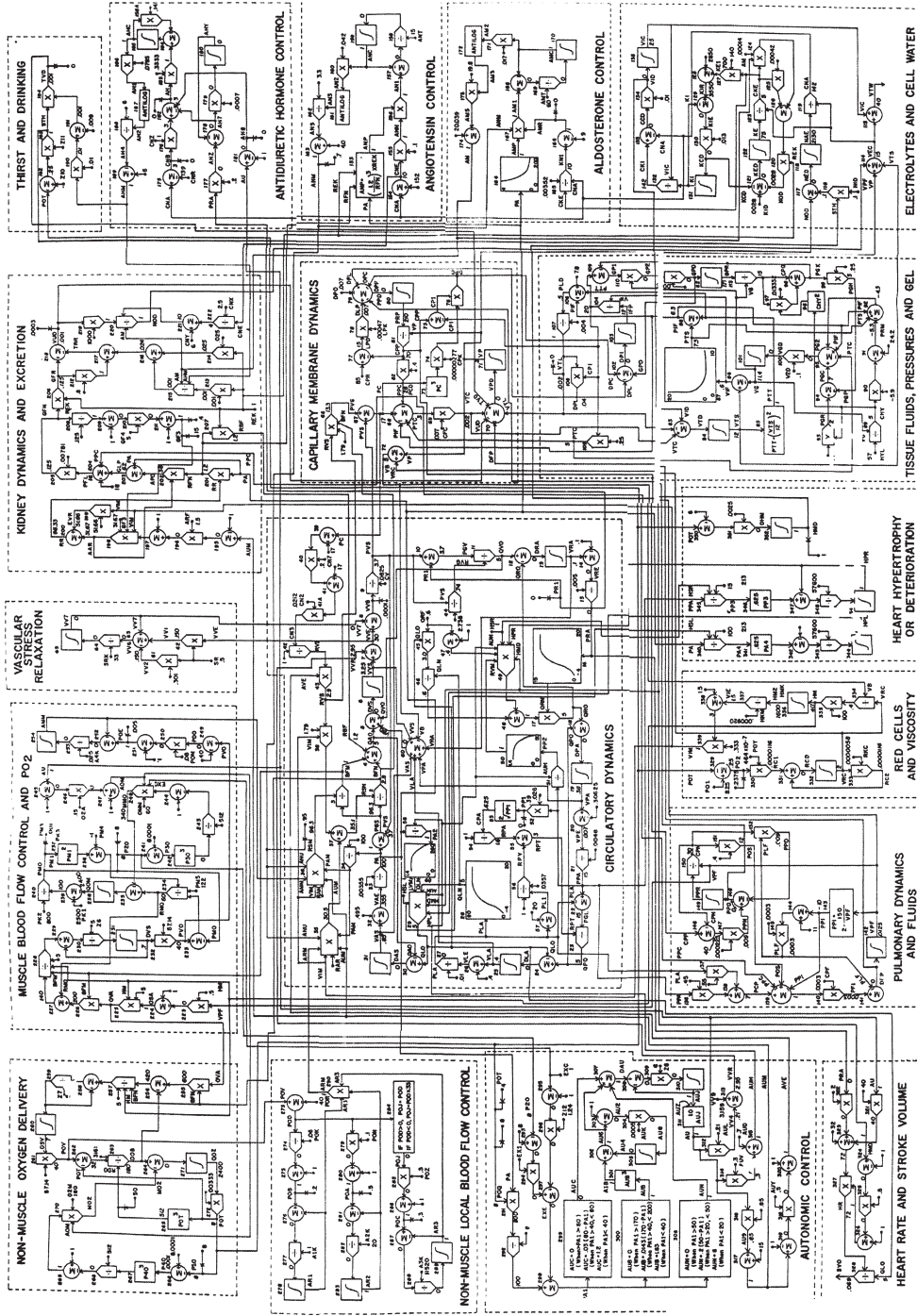


Fig. 5. Schematic illustrating systems analysis of the circulation. This analysis is composed of 354 blocks, each of which represents one or more mathematical equations describing the circulation. In general, each block represents the quantitative results of research performed by one or more investigators. (Reproduced from ref. 39 with permission.)

performed by different investigators. This type of complex modeling of quantitative physiological responses is one approach that could be utilized as part of the effort to understand complex biology from gene to function. When one considers the complexity of this diagram in the context of the roughly 35,000 genes, 100,000 proteins, and the trillions of cells that make up the body, the construction of a model that relates all of the genes to total organism function is an intimidating undertaking.

To perform this task, the basic quantitative information regarding the numerous interactions that occur between different organ systems in the normal control of the body's systems must be determined and systematically described. A far greater emphasis will need to be placed on the careful acquisition of these basic, quantitative types of physiological data. This information, once obtained, will then need to be analyzed and integrated for the construction of complex mathematical models by computational biologists, which should provide the final link between genes and function.

14. SUMMARY AND CONCLUSIONS

The sequencing of the human genome has provided 21st century physiologists with an unprecedented opportunity to explore and understand integrated organism physiology. The rapid evolution of genomic, proteomic, bioinformatic, and computational biology resources along with a renewed interest and recognition of the need and importance of quantitative measurements of cell and organism function places physiologists in an enviable position. Science is poised to begin the process of understanding the complex relationship between the approx 35,000 genes and organism structure and function. To accomplish this monumental charge, physiology as a field must step up to the task. This will likely require physiologists to utilize resources from multiple fields of investigation (genomics, proteomics, computational sciences, etc.), to adapt established techniques and develop new methodologies that enable true high-throughput phenotyping, to develop models and model systems that can provide functional information on a system-wide or even an organism-wide basis, and to begin thinking of new ways to describe and quantify physiological and pathophysiological processes. To accomplish this task, the future physiologist will likely work in a large collaborative group, embrace cutting-edge technology, and perform experiments in which enormous amounts of data are generated. At the same time, this physiologist will still be required to carefully design well-controlled studies, to produce high-quality, quantitative data, and to perform a rigorous statistical analysis to substantiate his or her conclusions. The future is bright for integrative physiologists; it is now up to the individual scientists to take advantage of this tremendous research opportunity.

REFERENCES

1. Lander, E.S., Linton, L.M., Birren, B., et al. (2001) International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
2. Venter, J.C., Adams, M.D., Myers, E.W., et al. (2001) The sequence of the human genome. *Science* **291**, 1304–1351.
3. Glueck, S.B. and Dzau, V.J. (2002) Physiological genomics: implications in hypertension research. *Hypertension* **39**, 310–315.
4. Roman, R.J., Cowley, A.W. Jr., Greene, A., Kwitek, A.E., Tonellato, P.J., and Jacob, H.J. (2002) Consomic rats for the identification of genes and pathways underlying cardiovascular disease. *Cold Spring Harbor Symposia on Quantitative Biology* **67**, 309–315.

5. Cowley, A.W. Jr. (2003) Genomics and homeostasis. *Am. J. Physiol.* **284**, R611–R627.
6. Vukmirovic, O.G. and Tilghman, S.M. (2000) Exploring genome space. *Nature* **405**, 820–822.
7. Jacob, H.J. and Kwikteck, A.E. (2002) Rat genetics: attaching physiology and pharmacology to the genome. *Nature Reviews Genetics* **3**, 33–42.
8. Kreutz, R. and Hubner, N. (2002) Congenic rat strains are important tools for the genetic dissection of essential hypertension. *Seminars in Nephrology* **22**, 135–147.
9. Rapp, J.P. (2000) Genetic analysis of inherited hypertension in the rat. *Physiol. Rev.* **80**, 135–72.
10. Warren, K.S. and Fishman, M.C. (1998) Physiological genomics: mutant screens in zebrafish. *Am. J. Physiol.* **275**, H1–H7.
11. Members of the Complex Trait Consortium. (2003) The nature and identification of quantitative trait loci: a community's view. *Nature Reviews Genetics*. **4**, 911–916.
12. Glazier, A.M., Nadeau, J.H., and Aitman, T.J. (2002) Finding genes that underlie complex traits. *Science* **298**, 2345–2349.
13. Greene, A.S. (2002) Application of physiological genomics to the microcirculation. *Microcirculation* **9**, 3–12.
14. Arcellano-Panlilio, M. and Robbins, S.M. (2002) Cutting-edge technology I. Global gene expression profiling using DNA microarrays. *Am. J. Physiol.* **282**, G397–G402.
15. Ruijter, J.M., Van Kampen, A.H.C., and Baas, F. (2002) Statistical evaluation of SAGE libraries: consequences for experimental design. *Physiol. Genomics* **11**, 37–44.
16. Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.
17. Liang, M., Yuan, B., Rute, E., et al. (2002) Renal medullary genes in salt-sensitive hypertension: a chromosomal substitution and cDNA microarray study. *Physiol. Genomics* **8**, 139–149.
18. Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207.
19. Bader, G.D., Heilbut, A., Andrews, B., Tyers, M., Hughes, T., and Boone, C. (2003) Functional genomics and proteomics: charting a multidimensional map of the yeast cell. *Trends in Cell Biology* **13**, 344–356.
20. Patterson, S.D. and Aebersold, R.H. (2003) Proteomics: the first decade and beyond. *Nature Genetics* **33**, 311–323.
21. Coffman, T.M. (1998) Gene targeting in physiological investigations: studies of the renin–angiotensin system. *Am. J. Physiol.* **274**, F999–F1005.
22. Smithies, O. (1997) A mouse view of hypertension. *Hypertension* **30**, 1318–1324.
23. Sauer, B. (1993) Manipulation of transgenes by site-specific recombination: use of Cre recombinase. *Methods of Enzymology* **225**, 890–900.
24. Furth, P.A., St. Onge, L., Boger, H., et al. (1994) Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. *Proc. Natl. Acad. Sci.* **91**, 9302–9306.
25. Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766–1769.
26. Gossen, M. and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci.* **89**, 5547–5551.
27. Zan, Y., Haag, J.D., Chen, K.S., et al. (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. *Nature Biotechnology* **21**, 645–651.
28. Dykxhoorn, D.M., Novina, C.D., and Sharp, P.A. (2003) Killing the messenger: short RNAs that silence gene expression. *Nature Reviews* **4**, 457–467.
29. Skipper, M. (2003) Have our dreams been shattered? *Nature Reviews Genetics* **4**, 671.
30. Tijsterman, M., Ketting, R.F., and Plasterk, R.H.A. (2002) The genetics of RNA silencing. *Annu. Rev. Genet.* **36**, 489–519.
31. Scalzitti, J.M., and Hensler, J.G. (1999) Design and efficacy of serotonin-2A receptor antisense oligodeoxynucleotide. *Methods Enzymology* **314**, 76–89.
32. Kanehisa, M. and Bork, P. (2003) Bioinformatics in the post-sequence era. *Nature Genetics* **33**, 305–310.

33. Freimer, N. and Sabatti, C. (2003) The human phenome project. *Nature Genetics* **34**, 15–21.
34. Bogue, M. (2003) Mouse Phenome Project: understanding human biology through mouse genetics and genomics. *J. Appl. Physiol.* **95**, 1335–1337.
35. Bassingthwaighe, J.B. (2000) Strategies for the physiome project. *Annals of Biomedical Engineering* **28**, 1043–1058.
36. Kitano, H. (2003) Computational systems biology. *Nature* **420**, 206–210.
37. Hood, L. and Galas, D. (2003) The digital code of DNA. *Nature* **421**, 444–448.
38. Guyton, A.C., Coleman, T.G., Cowley, A.W. Jr., Manning, R.D. Jr., Norman, R.A. Jr., and Ferguson, J.D. (1974) A systems analysis approach to understanding long-range arterial blood pressure control and hypertension. *Circ. Res.* **35**, 159–176.
39. Guyton, A.C., Coleman, T.G., and Granger, H.J. (1972) Circulation: overall regulation. *Ann. Rev. Physiol.* **34**, 13–46.



<http://www.springer.com/978-1-58829-315-2>

Integrative Physiology in the Proteomics and
Post-Genomics Age

Walz, W. (Ed.)

2005, X, 269 p., Hardcover

ISBN: 978-1-58829-315-2

A product of Humana Press