

Meeting Report

New Roles for Model Genetic Organisms in Understanding and Treating Human Disease: Report From The 2006 Genetics Society of America Meeting

Allan Spradling,^{1,2} Barry Ganetsky, Phil Hieter, Mark Johnston, Maynard Olson, Terry Orr-Weaver, Janet Rossant, Alejandro Sanchez and Robert Waterston

ABSTRACT

Fundamental biological knowledge and the technology to acquire it have been immeasurably advanced by past efforts to understand and manipulate the genomes of model organisms. Has the utility of bacteria, yeast, worms, flies, mice, plants, and other models now peaked and are humans poised to become the model organism of the future? The Genetics Society of America recently convened its 2006 meeting entitled "Genetic Analysis: Model Organisms to Human Biology" to examine the future role of genetic research. (Because of time limitations, the meeting was unable to cover the substantial contributions and future potential of research on model prokaryotic organisms.) In fact, the potential of model-organism-based studies has grown substantially in recent years. The genomics revolution has revealed an underlying unity between the cells and tissues of eukaryotic organisms from yeast to humans. No uniquely human biological mechanisms have yet come to light. This common evolutionary heritage makes it possible to use genetically tractable organisms to model important aspects of human medical disorders such as cancer, birth defects, neurological dysfunction, reproductive failure, malnutrition, and aging in systems amenable to rapid and powerful experimentation. Applying model systems in this way will allow us to identify common genes, proteins, and processes that underlie human medical conditions. It will allow us to systematically decipher the gene-gene and gene-environment interactions that influence complex multigenic disorders. Above all, disease models have the potential to address a growing gap between our ability to collect human genetic data and to productively interpret and apply it. If model organism research is supported with these goals in mind, we can look forward to diagnosing and treating human disease using information from multiple systems and to a medical science built on the unified history of life on earth.

MODEL system research has long contributed to basic biological knowledge and its application to human medicine. We have learned that humans share with other animals the same life processes stemming from the common evolutionary origins of all living metazoans. Studies on the basic biology of model organisms first identified the key genetic pathways, such as transcription control circuits, signal transduction pathways, cell cycle regulators, etc., that have since been shown to play critical roles in disease. Moreover, studies using model systems continue to uncover the basic principles as well as the specific details of gene interactions, population variation, and developmental plasticity. There is still an immense amount to learn in all areas of fundamental biology, and approaches using model organisms continue to be among the most productive. We must not relax in our efforts to understand the roles that

genes play in all fundamental life processes, including development, growth, physiology, behavior, and aging.

The human genomics revolution has raised the realistic prospect of using information about a patient's genome to identify predisposition to disease, to predict disease severity, and to select propitious therapies. Human genomic information promises to provide genetic markers that would allow health problems to be detected well before they become apparent. Gene-based medicine might even allow prophylactic and curative interventions to be tailored to our individual characteristics. However, major obstacles must still be overcome if the theoretical advantages of genetic medicine are to be realized on a large scale. Patient-specific genetic information is being obtained at an exponentially growing rate, but we still face unresolved technical, ethical, and practical difficulties in deciphering and applying it.

This report highlights work presented at the 2006 Genetics Society of America Meeting that suggests how model organism research can contribute to resolving these obstacles. Few, if any, processes at the level of genes are known to be unique to humans, although the

¹Chair of the Meeting Report Committee.

²Corresponding author: Department of Embryology/HHMI, Carnegie Institution of Washington, 3520 San Martin Dr., Baltimore, MD 21218. E-mail: spradling@ciwemb.edu

details of cellular and organismal physiology certainly differ. Consequently, we argue that key aspects of most human disorders ultimately can be modeled in organisms in which the compromised processes are controlled by closely related genes, thereby offering the potential for rapid and inexpensive genetic analyses. Genes that accelerate or retard the disease process may be identified rapidly, opening up the prospect of new therapeutic approaches. Finally, it may even be possible to test the functional consequences of specific human genetic variants in these models, providing information that cannot practically be obtained in any other way. (In this article, names in parentheses refer to a presenter at the meeting; see Meeting Program at <http://www.gsa-modelorganisms.org/>.)

CELL STRUCTURE AND FUNCTION

We still understand only a small fraction of the genes, pathways, and molecular mechanisms that sustain individual cells. For example, organelles such as mitochondria, nucleoli, or flagella each contain many hundreds of gene products, most of whose roles remain obscure. The yeast *Saccharomyces cerevisiae* continues to be a pre-eminent contributor to cell biology, and results of research on yeast frequently provide insight into the cells of humans and other metazoan animals. Multicellular models remain essential as well. Many genes play different roles in different cell types and at different developmental stages, based on complex cell and tissue interactions that cannot be mimicked in unicellular organisms.

Yeast, worms, flies, zebrafish, Arabidopsis, and mice are the members of the Security Council of the United Nations of species. These model organisms can be used to study a wide range of biological problems. Other model organisms have proven their value in specialized research areas. *Chlamydomonas reinhardtii*, for example, is the premier system for studying flagellar structure and function. Likewise, *Schizosaccharomyces pombe* is particularly valuable for studies of the cell cycle, *Zea mays* for studies of epigenetic regulation, and the planarian worm *Schmidtea mediterranea* is exceptionally well suited for studies of cell and organ regeneration based on stem cells, to name but a few.

The study of flagella in *Chlamydomonas* (Susan Dutcher, Washington University) illustrates the value of single-cell models for dissecting mechanisms that underlie human disease. Flagella grow from their distal ends by utilizing an internal transport network: the intraflagellar transport (IFT) system. Many human organs contain ciliated cells or modified cilia that employ IFT genes, including rod and cone photoreceptors, renal cells, Fallopian tube cells, airway cells, and many others. Ciliary defects contribute to common human diseases, including inherited mutations in the human homologs of various flagellar proteins that give rise to Bardet-

Biedel syndrome. These observations have been confirmed and extended by studying orthologous mutations in *Caenorhabditis elegans*, *Drosophila melanogaster*, and mice.

DEVELOPMENT AND PHYSIOLOGY

Genes often show similarities in their expression and function in different tissues in ways that are conserved between distant phyla. It is now clear that the developmental mechanisms and ground plans used by model organisms and humans descended from a common bilaterian Precambrian metazoan ancestor, sometimes referred to as Urbilateria. This organism had cell types, tissues, developmental processes, and gene functions that are still utilized by modern organisms. These underlying similarities have been dramatically documented through studies of Hox gene clusters (Mario Capecchi, University of Utah Medical School) and continue to be revealed as genes regulating tissue and organ development become characterized in many organisms. Evolutionary conservation among phyla is far more extensive than anyone anticipated prior to the genomics era and goes beyond conserved genes, pathways, and cell types to the level of tissues and organs. We are learning that tissue formation as well as function and dysfunction frequently occur for related reasons in different organisms.

The genetic heritage that we share with other organisms is most obvious in the case of vertebrate models such as mice, rats, and monkeys. We feel a kinship with warm-blooded animals that is largely absent in the case of "lower" organisms. However, recent research makes clear that common biological mechanisms can be widely spread across the phylogenetic tree. One need look no further for examples than the development of the blood and guts of *Drosophila*. Both the absorptive and the enteroendocrine cells of the *Drosophila* and vertebrate intestine are produced by the division of pluripotent stem cells (Allan Spradling, Carnegie Institution). The decision of stem cell daughters in both groups of organisms to differentiate along the enteroendocrine or enterocyte pathways depends on Notch signaling, suggesting that this entire system has been evolutionarily conserved. Blood cells in both *Drosophila* and vertebrates arise at multiple sequential embryonic sites, employ homologous transcription factors such as Runx, and utilize specific cell lineages such as the hemangioblast (Utpal Banerjee, University of California at Los Angeles). Expansion of blood cell number in both groups prior to adulthood likely depends on stem cells. However, analogs of only part of the human blood system (macrophages, but not lymphocytes or erythrocytes) are present in *Drosophila*. Clearly, to learn which aspects of human biology are shared with each model system, we must carry out experiments and not rely on our emotional affinities.

An important way to identify homologous tissues and processes among organisms is by mapping how homologous genes are expressed throughout the life cycles of

the different organisms. Such studies can also identify novel cell types that were overlooked by conventional methods. Deriving regulatory information from genomic DNA sequence data represents a challenge that is unsolved, however. Comparisons of the mouse and human sequence indicate that only ~5% is under selection, comprising 1.5% that codes for protein and 3.5% that is noncoding and presumably involved in gene regulation (Eric Green, National Human Genome Research Institute). Of particular interest are small sequence blocks of ~60 bp that are surprisingly well conserved throughout the vertebrates. Seventy-five percent of these blocks reside in non-protein-coding regions; their function remains a mystery. Consequently, experimental methods are essential to define and map genomic regulatory information. One of the most advanced such methods currently exists in the nematode, where computer analyses of four-dimensional microscopic images make it possible to map transgene expression patterns onto the fixed embryonic cell lineage (Zhirong Bao, Robert Waterston lab, University of Washington). Large collections of *Drosophila* strains genetically engineered to express different proteins fused to green fluorescent protein under normal gene controls constitute another versatile system for defining expression patterns (Allan Spradling, Carnegie Institution). Publicly available collections of BAC-transgenic lines and gene-trap reporter embryonic stem cells provide the tools for similar studies in the mouse.

At present there is no public database exploring the relationships between the genes, expression patterns, cell types, and tissues of model organisms and humans. The annotation of model organism sequences is still at an early stage and our knowledge of gene function remains so limited that efforts are still largely focused on single organisms. Organism-specific genome projects have spawned efforts in mice, worms, and flies to characterize the spatial expression patterns of gene transcripts during embryonic development. Grouping the function of genes in these organisms into “gene ontologies” (Jonathan Bard, University of Edinburgh) represents a starting point for determining orthologies between organisms. Current ontologies are largely language based but are beginning to incorporate graphic displays through links to other projects. Understanding the parallels among the cells and tissues of humans and model organisms, as well as their genetic underpinnings, would constitute one of the greatest intellectual achievements of science and would be of inestimable value for human medicine.

CELL GROWTH AND CANCER

The genes controlling the eukaryotic cell cycle were characterized in large part through genetic studies of the yeasts *S. cerevisiae* and *Schizosaccharomyces pombe*. Using high-resolution genome tiling arrays, the pattern of

DNA replication origin firing can now be mapped on each *S. pombe* chromosome during mitosis and meiosis (P. Nurse, Rockefeller University). Multiple origins are utilized, but the same origins are not always activated during each cell cycle, and the last 1% of replication may take place in the G₂ phase of the cell cycle. Different replication patterns in mitosis and meiosis may result from different levels of proteins that initiate replication.

Chromosome telomeres must be properly maintained and duplicated to stabilize chromosomes and suppress tumor formation. In budding yeast, *S. cerevisiae*, a specific protein complex (the CST complex) binds each telomere and controls its replication by telomerase. During replication the complex interacts with checkpoint proteins and prevents telomeres from being recognized and repaired as double strand breaks (Vicki Lundblad, Salk Institute). A similar complex known as shelterin carries out analogous functions at mammalian telomeres (Titia de Lange, Rockefeller University). POT1 proteins (encoded by one gene in humans and two related genes in mice) interact with the Ku70/80 repair protein to control 3' overhang length and to suppress the end-to-end chromosome fusions made famous by Barbara McClintock. Such fusions are thought to initiate the gene amplification observed in some human tumor cells.

Some of the same pathways that regulate growth and cell proliferation in humans operate similarly in *Drosophila* (Iswar Hariharan, University of California at Berkeley). Alterations in these genes are frequently involved in oncogenesis. In *Drosophila*, ribosome production and cell growth are controlled by Myc expression levels. A key regulator of Myc-mediated growth is an associated protein complex containing the Trithorax class gene *lid*, which has histone lysine methyltransferase activity (Julie Seacomb, Robert Eisenman lab, Fred Hutchinson Cancer Center). It has long been suspected that tumor-associated stromal cells act nonautonomously to stimulate the proliferation of epithelial cancers. Studies of a *Drosophila* model have identified mutations in the human “tumor susceptibility gene 101 (Tsg101)” as a possible cause of such stromal stimulation. Mutations in the *Drosophila* homolog of Tsg101 lead to excess Notch signaling and ectopic secretion of a JAK-STAT ligand, causing adjacent cells to overgrow (Iswar Hariharan, University of California at Berkeley).

Studies of the cell cycle and cell growth in yeast and invertebrates have generally focused on fundamental control mechanisms, rather than directly modeling diseases where cell cycle control is disrupted. However, several presentations at the meeting demonstrated that invertebrates can serve as models of cancer. Gliomas are among the most common malignancies of the human central nervous system, and effective treatments are lacking. The *Drosophila* brain contains glial cells that strongly resemble the major types of vertebrate glia. Activating two pathways associated with human gliomas

(*i.e.*, EGFR signaling via Ras and PI3 kinase) in *Drosophila* glia using the GAL4 system caused glial cells to proliferate rapidly and invade tissues throughout the brain and optic lobes (Renee Read, John Thomas lab, Salk Institute). Acute myelogenous leukemia, one of the most common human blood malignancies, is caused by a rearrangement of the AML1 gene. When the rearranged human protein is expressed in *Drosophila* blood cell precursors, they proliferate extensively within the hemolymph, providing a *Drosophila* model for a common childhood disease (Utpal Banerjee, University of California at Los Angeles). Human neuroendocrine tumors of the intestine have been associated with alterations in Notch signaling. When Notch signaling was disrupted in *Drosophila* intestinal progenitors, cells expressing the enteroendocrine cell marker Prospero proliferated, leading to death (Allan Spradling, Carnegie Institution). Finally, germ cells in *C. elegans* adults with mutations in the differentiation gene *gld-1* form germline tumors that proliferate, break through the gonad wall, fill the body cavity, and kill the worm (Cynthia Kenyon, University of California at San Francisco).

STEM CELLS

Stem cells are critically important for homeostasis and acute repair of human blood, immune system, epithelia, gut, brain, breast, cornea, lung, and probably many other tissues. In addition, embryonic cells converted to the stem cell state in culture are promising sources of replacement cells for a wide variety of human disorders. Studies of stem cells in *Drosophila*, mouse, and other model systems have greatly advanced our understanding of their regulation within stem cell niches generated by nearby supporting stromal cells. The male germline stem cells in the *Drosophila* testis form adhesive junctions to the hub cells that establish their niche (Margaret Fuller, Stanford). They divide asymmetrically, and the daughter centriole always segregates into the cell that will differentiate. This is the first known case of differential segregation of centrioles during stem cell division. New genes required for stem cell maintenance continue to be discovered in model organisms. Recently, there has been considerable interest in genes that affect stem cell chromatin, and the nuclear protein Stonewall, a suppressor of position-effect variegation, can now be added to this list (Jean Maines, Dennis McKearin lab, Utah Southwestern Medical School). Unlike the situation in other characterized stem cells, the germline stem cells of *C. elegans* are maintained by a probabilistic choice between regenerative and differentiative divisions (Judith Kimble, University of Wisconsin). The differentiation of downstream germ cells is controlled by derepressing translation of the mRNA encoding the Gld-1 regulator, which takes place by reversing the deadenylation activity

of the Pumilio-like gene FBF and by activating the cytoplasmic poly(A) polymerase Gld-2.

Mechanisms regulating the interactions between stem cells and their niche can be elucidated in great detail using genetics in model organisms and can inform the study of stem cell–niche interactions in humans as well. Vertebrate hematopoietic stem cells (HSCs) are maintained in niches created by partner osteoblasts (Linheng Li, Stowers Institute). Upon activation, HSCs detach from the osteoblast, accelerate their division rate, and associate with blood vessels. The same mechanisms identified in model organism stem cell niches, including anchoring to the niche via adherens junctions, maintenance of gene repression by local BMP, Wnt, and Notch signals, and activation of daughters that have left the niche, are observed in vertebrate HSCs and other stem cells. Vertebrate and invertebrate stem cells from corresponding tissues appear to share even more detailed mechanisms. For example, human, *Drosophila*, and worm germline stem cells all express and appear to depend on Pumilio-like genes. Both *Drosophila* and vertebrate epithelial stem cells require hedgehog signals, while Notch signals play a central role downstream of intestinal stem cells.

Studies of gene expression and cellular differentiation during the early stages of mouse embryogenesis are providing new insight into the origin and potentials of embryonic stem (ES) cells (Janet Rossant, University of Toronto). Three basic cell types are found in the mouse blastocyst—epiblast, trophoblast, and primitive endoderm—and each can give rise to stem cell cultures, including pluripotent ES cells. The expression patterns of the homeobox genes Oct4 or Nanog suggest that they are involved in determining cell fates. Blastocysts mutant for Oct4 or Nanog were unable to form one of the stem cell types. Primitive endoderm and epiblast cells may sort out following asymmetric divisions that give rise to daughters expressing either Gata6 or Nanog, reminiscent of many instances of cell fate specification in invertebrates.

NEUROLOGICAL DISEASES

Great progress has been made in modeling neurodegenerative diseases using model organisms. Misexpression of proteins containing engineered triplet repeats encoding amino acids such as glutamine cause neural defects reminiscent of human triplet repeat diseases. Refinements in the expressed protein and genetic background of these models are achieving greater disease specificity. Moreover, other neurodegenerative diseases can also be mimicked. For example, important aspects of Parkinson's disease are recapitulated in yeast that overexpress α -synuclein, leading to cytotoxicity (Susan Lindquist, Whitehead Institute). Study of toxicity suppressors led to a model explaining why Parkinson's disease selectively affects dopaminergic neurons. Suppressor

genes are involved in ER-to-Golgi transport, highlighting a new mechanism of action of α -synuclein distinct from other neurodegenerative diseases such as Huntington's disease. Studies of Huntington's disease in yeast have identified novel genes required for Huntingtin protein toxicity (Flaviano Giorgini, University of Washington). Studies in flies expressing the N terminus of the human Huntingtin protein have allowed for rapid screening of drug candidates, some of which are now in phase 1 clinical trials (Laszlo Bodai, Lawrence Marsh lab, University of California at Irvine). Studies of protein folding, aggregation, and proteolysis, including of yeast prion proteins, are also highly relevant to prion diseases (Susan Lindquist, Whitehead Institute). New models of Alzheimer's disease, the most common form of human dementia, have implicated abnormal activity of the neuron-specific protein kinase Cdk5 (Li-Huei Tsai, Harvard Medical School). In its normal role in synaptic plasticity and hippocampus-dependent memory, Cdk5 activity is regulated by the relative amount of p25 protein generated by calpain-dependent cleavage. Mice expressing elevated levels of p25 showed many hallmarks of Alzheimer's disease. Finally, systematic study of Hox gene functions has identified Hoxb8 as a gene whose loss causes mice to exhibit behaviors that are highly characteristic of obsessive-compulsive disorder (Mario Capecchi, University of Utah Medical School).

HEART DISEASE

Important models of heart development and function are now provided by studies using zebrafish and mouse. Remarkably, heart function in *Drosophila* is similar enough to that in vertebrates to provide insight into heart failure and arrhythmias (Karen Ocorr, Rolf Bodmer lab, The Burnham Institute). M-mode traces of *Drosophila* heartbeats reveal a gradual onset of episodes of arrhythmia and an increased susceptibility to stress-induced cardiac arrest as a function of age. Some of the same genes that cause human congenital heart disease increase the frequency of heart failure in *Drosophila*! Mutations within the single *Drosophila* KCNQ potassium channel gene affect myocardial repolarization in a manner similar to the effects of KCNQ1 mutants in humans, where ventricular repolarization is impaired, leading to arrhythmia and ventricular fibrillation. This model system is likely to be useful in identifying additional genes that predispose to heart failure, including genes involved in mitochondrial function, which has been implicated in congestive heart failure.

EPIGENETIC DISEASES

Epigenetics refers to the ability of cells to faithfully transmit to their progeny information that is not en-

coded in the nucleotide sequence of their genomes. The epigenetic inheritance of chromatin states during cell division is the universal basis for cell specialization and differentiation during metazoan development. In humans and other mammals, male and female gametes contain epigenetically imprinted genes. Inheritance of defective or deleted versions of particular genes or small gene regions subject to imprinting gives rise to Prader-Willi syndrome, Angelman syndrome, and Beckwith-Wiedemann syndrome. Mixed epigenetic and genetic factors may be responsible for the complicated etiologies associated with autism and many other conditions (Arthur Beaudet, Baylor College of Medicine).

Model system studies have greatly advanced our understanding of the molecular mechanisms that underlie epigenetic marking and imprinting. The long-range modulation of X chromosome activity that underlies dosage compensation in *C. elegans* has been particularly instructive (Barbara Meyer, University of California at Berkeley). A complex of proteins that effects dosage compensation is related to complexes that normally condense chromosomes during the cell cycle. Sequence motifs that recruit dosage compensation proteins specifically to the X chromosome can be mapped at very high resolution by determining if introduced copies recruit the protein complex. The ability of specific alleles of certain genes to interact and heritably change when present in *trans* has been termed "paramutation" and is well characterized in maize (Vicki Chandler, University of Arizona). An 853-bp tandem repeat located 100 kb 5' to the transcription start site controls paramutation at the *b* locus and is affected in *trans* by a gene encoding an RNA-dependent RNA polymerase. This represents one more hint from model organism studies that small RNAs play a role in chromatin programming. The stability of epigenetic states is frequently ensured through the action of Polycomb complex genes that affect methylation of histones (Nathan Montgomery, Terry Magnuson lab, University of North Carolina). Insight into the multiple protein isoforms produced by the Eed gene encoding a histone H3-lysine 27 methyl transferase was obtained by engineering the locus to express particular isoforms. The results show that monomethylation of H3 K27 might not require other members of Polycomb complex 1, in contrast to di- and trimethylation.

A key issue concerns how cells maintain their epigenetic states following DNA replication. Growing evidence implicates the highly conserved histone variant H3.3 in this process, which, unlike normal H3, is produced outside of S phase (Steve Henikoff, Fred Hutchinson Cancer Center). Active loci regularly displace and reassemble nucleosomes and would preferentially accumulate H3.3-containing nucleosomes except during S phase. Thus, the pattern of H3.3-enriched regions would epigenetically mark active genes. Studies of H3.3 enrichment levels across the *Drosophila* genome using

tiling arrays revealed a striking correspondence with active genes.

AGING AND AGE-RELATED DISEASES

Research on *C. elegans* as well as other model organisms has solidified the connection between metabolism and aging (Cynthia Kenyon, University of California at San Francisco). Active energy metabolism as reflected by insulin signaling stimulates growth and reproductive activity. Low-energy resources produce low insulin signaling and shift metabolism to cell maintenance and stress resistance. This shift also slows the aging process. These changes are mediated by the Daf-16 transcription factor that acts on multiple targets. These include DNA repair genes, stress resistance genes, Sirtuins, and other factors, some of which have been implicated separately as affecting life span.

The incidence of many diseases, including cancer, neurodegeneration, heart failure, deafness, cataracts, and arthritis increases steeply in older individuals. Age-dependent increases correlate with fractional life span, not with absolute age or the absolute number of cell divisions. In model organisms, the onset of many diseases with a high incidence in old individuals is delayed by mutations in the insulin-mediated aging pathway. For example, the onset of *gld-1*-induced lethal germline tumors was greatly retarded in *daf-2* mutants with extended life span, due to a slower proliferation of tumor cells and greater p53-mediated apoptosis (Cynthia Kenyon, University of California at San Francisco). It is often thought that aging research is an activity supported by aging baby boomers determined to postpone the inevitable. However, this research suggests that understanding aging using model organisms might reveal therapies that could be shown to postpone the onset of particularly devastating diseases, rather than promoting extended life span *per se*.

GENETIC VARIATION, GENE INTERACTIONS, AND MULTIGENIC TRAITS

The genetic structure of human populations is critically important for virtually all applications of human genomic information to medical practice. Studies of model organism populations have established the fundamental principles of population structure and methods for their analysis. These studies provide crucial guidance in this aspect of human genetics. Many human traits are genetically complex, being influenced by multiple genes (QTL), each with only a partial effect on the trait in question. Studies of *Drosophila* using >1000 isogenic strains bearing single identifiable changes provide new insight into the nature of QTL (Trudy Mackay, North Carolina State). Many QTL correspond to genes that already have been discovered through genetic studies of development. Alleles of these genes that are found

in natural populations affect multiple processes, however, including phenotypes not obvious from the nature of their major developmental effect. Frequently, these QTL affect the two sexes differently and exhibit epistatic and gene–environment interactions. This work points to the possibility of using model organisms systematically to map genes, gene interactions, and gene–environment linkages that will inform the complex genetic basis of human disease. The mouse and the rat are other organisms in which extensive use of QTL analysis among different strains has made it possible to map important traits that underlie complex diseases such as diabetes, hypertension, and obesity, with obvious implications for human disease.

Understanding how multigenic diseases arise requires understanding how defects in one pathway can interact with defects in other pathways to influence a phenotype. In yeast, genomewide synthetic lethal screens and two-hybrid protein–protein interaction screens are already used extensively to detect gene interactions. Another powerful, general, and systematic method for identifying interactions takes advantage of the highly efficient, genomewide RNA interference screens that can be carried out in *C. elegans* (Anthony Fraser, Wellcome Trust Sanger Institute). Genes that interact with a starting mutation affecting virtually any process or disease model can be mapped throughout the genome. A test using the well-studied Ras-signaling pathway rapidly detected 60% of known pathway members as well as several novel components. More than 50 such screens have already been carried out, revealing, among many other things, the existence of a small number of genes, including *HMG1*, *HMG2*, *TTRAP*, *cdc2*, and *dpy-20*, that interact with a wide variety of cellular processes. The human homologs of these “highly connected” genes would be excellent candidates for causative genes in QTL mapping in their vicinity.

It has long been postulated that an important mechanism of speciation is reproductive isolation via hybrid sterility or inviability between two closely related populations. However, the nature of such “speciation” genes has remained largely a mystery. Advances in model organism genetics and comparative genome sequencing are now making it possible to learn the molecular nature of such genes (Allen Orr, University of Rochester). Hybrid inviability can result from the rapid evolution of a nucleoporin gene that is part of the highly conserved nuclear pore complex. Presumably, two components of the complex encoded on different chromosomes have co-evolved rapidly under selection and can no longer interact properly with the alleles in another species, despite overall conservation of nuclear pore function. Changes in gene location may also affect species hybrids. Most variation in gene position between closely related *Drosophila* species results from inversions within a single chromosome arm rather than from transpositions (William Gelbart, Harvard). The gene encoding a

sperm-specific Na/K ATPase α -subunit is an exception, as it is located on chromosome 3 in *D. simulans*, but on chromosome 4 in *D. melanogaster* and most other related species. This explains Muller's classic observation that *D. melanogaster* males bearing a *D. simulans* fourth chromosome are sterile; such flies lack the ATPase gene entirely. Loss of the human homolog of this gene causes a sperm motility phenotype in humans identical to that seen in the affected *Drosophila* hybrids.

Genes responsible for rapid morphological evolution are of great interest for humans, where substantial change has come about in a short time. Stickleback fish have evolved numerous species in lakes just in the 10,000 years since the last ice age. They provide a powerful model for defining genes causing large changes in skeletal structure and behavior (David Kingsley, Stanford). Genes identified in this manner, such as the homolog of human ectodysplasin (the mouse Tabby gene), represent potential sources of human population variation. The number of generations involved in these speciation events is similar to the number of human generations since our ancestors dispersed out of Africa, highlighting the potential importance of such studies. Comparisons of genome sequences and expression array data between humans and closely related primates have begun to reveal intriguing differences, such as a twofold change in the level of phytanic acid hydroxylase, which have played a role in the rapid transition by humans from a vegetarian to a meat-rich diet (Joseph Hacia, University of Southern California).

TECHNOLOGY DEVELOPMENT AND THE EXPANDING "DATA GAP"

Our ability to ask biological questions has always been limited by technology. Improvements in our ability to acquire large amounts of raw genetic data are continuing and are accelerating, enabling researchers to further expand their horizons. The cost of (re)sequencing a new human genome may soon fall into the 200,000 dollar range through developments such as "polony" sequencing (George Church, Harvard Medical School) and could drop further due to the development of nanodetector arrays that can directly convert biomolecular interactions at each spot on a chip into easily amplified electrical signals (Ron Davis, Stanford). Chip-based assays based on proximity ligation (Ron Davis, Stanford) or novel capture agents such as click chemistry (Lee Hood, Institute for Systems Biology) may soon allow the blood levels of as many as 2000 selected proteins to be rapidly (and cheaply) determined. The completion of the first stage of the HapMap project (Peter Donnelly, University of Oxford) is allowing the variation in DNA sequences among patient families and human subpopulations to be characterized and correlated with disease states (Mary Clare King, University of Washington). There are a growing number of candidate disease genes and disease-associated polymor-

phisms that have been implicated by classical human genetics and by these newer approaches. It may become possible in the near future to map multigenic factors involved in human health to small chromosome regions, and even to individual genes.

Unfortunately, obtaining information that can be reliably applied in the diagnosis and treatment of disease from sequence data is proving to be more difficult than initially hoped. Human subpopulations, like natural populations of model organisms (Trudy Mackay, North Carolina State), show widely varying amounts of linkage disequilibrium. Simply identifying that a patient has a polymorphism that has been linked to disease in one test population may be insufficiently predictive to justify prophylactic treatment. A large discrepancy is therefore likely to arise between available human genetic data and clinically useful data, a situation that we term the "data gap." Bioinformatic analyses of sequence data and high-throughput gene expression data alone cannot bridge the gap, because we are nowhere close to understanding theoretically the link between genotype and phenotype even for most changes within a coding sequence. Experimental tests in living cells and animals remain one of the most promising avenues for evaluating the significance of genotypic variation in humans.

A variety of tests will be needed to evaluate experimentally the significance of human genetic variation. Some functional tests can be carried out in short-term human cell cultures. Others may be more rapidly and cheaply studied in yeast. For example, natural human polymorphisms in genes encoding human-cofactor-containing enzymes can be functionally tested by expressing the human proteins in yeast (Jasper Rine, University of California at Berkeley). Up to 50% of polymorphisms that reduced activity could be fully rescued by raising the level of the cofactor (such as folate). However, the great majority of genes do not encode proteins with activities that are easily scored, and, moreover, cultures of individual cells cannot recapitulate the complex developmental regulation and the multiple, context-dependent functions commonly observed for genes *in vivo*. It might be possible to accurately assess the functional significance of gene polymorphisms by testing them in a multicellular model organism, where appropriately expressed human genes or their homologs have been documented to carry out a conserved process.

When some aspect of human physiology can be simulated in a model organism, it can be used productively for early stage screening of drug candidates. With the availability of cell lines from humans, such as embryonic stem cells, that can differentiate into many cell and tissue types *in vitro*, these first-pass screens could then be rapidly confirmed using relevant human cells before moving to expensive preclinical and clinical models. Model organism genetics can thus first help define new targets for disease therapy and then combine with stem

cell biology to reduce the time and expense of moving from initial target to a final therapeutic outcome.

Developing such functional tests will require that human polymorphic mutations or even entire human genes be introduced rapidly and efficiently into the genomes of test cells or organisms. Fortunately, new methods for modifying mammalian genomes are becoming available. Targeted DNA cleavage using designer zinc-finger endonucleases (Dana Carroll, University of Utah Medical School) catalyze high-frequency homologous gene replacement. The efficient, highly sophisticated genetic manipulations that can be carried out in *Drosophila* using engineered transposable elements will soon become possible in the mouse and other mammalian genomes using the PiggyBac transposon (Tian Xu, Yale). This includes an enhanced ability to molecularly identify genes affecting myriad biological and disease processes using forward genetic screens. The ability to correctly assign gene orthologs between humans and model systems is also critically important. Rapidly increasing knowledge of genome sequences across a broad range of phylogenetic groups is greatly improving our ability to make such assignments (Richard Durbin, Sanger Center). This publicly available database is based on curated trees of alignments (TreeFam). Phylogeny-based alignments are likely to be extremely valuable in light of the fact that genomes do not evolve at constant rates or in uniform ways. Nematode species (Rhabditids) are now known to undergo high rates of gene gain, loss, intron evolution, and neutral sequence substitution, and their 18S rDNA sequences are as diverged as those of all deuterostomes.

CONCLUSIONS

Clearly, model organism genetics continues to hold great promise for advancing biology, medicine, and human genetics. However, the potential benefits high-

lighted here will not be realized if we apply only what we already know. Discovery research must continue to be our highest priority, as it will remain the foundation of everything else we undertake. Much more work is needed simply to identify orthologies across the biological spectrum and to develop valid models of normal physiology and disease pathology. A critical need is better cross-organism databases that enable one to compare the genes, expression patterns, gene functions, cell types, tissue organization, and biological subprocesses across organisms, including humans. Maintaining and expanding our community resources, such as mutant collections and siRNA libraries for many organisms, including those not amenable to standard genetic techniques, is crucial. They provide access to the genetic power of the different model organisms and enable investigators to take full advantage of whole-genome sequence information. Finally, we must look for ways to interact with clinician scientists and human geneticists and bring their knowledge and perspectives to the modeling efforts.

In closing, we would be remiss if we did not mention that support for model organism research seems increasingly at risk due to the burgeoning quest for direct translational research on human disease. Long-standing researchers are being forced from the field, and excellent young scientists are being shut out. This disturbing trend must be reversed and the approaches described here need to be enthusiastically supported. The golden age of biology is upon us! Now is not the time to retreat from the approaches that brought us here. By redoubling our efforts to understand cell and tissue function, we can develop experimental tests that will improve diagnosis of disease, that will reveal patient-specific therapies, and that will teach us about our genetic heritage as a species. We have never been more excited about the benefits that genetic research can bring to the human population.