Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics

Edited by James F. Crow and William F. Dove

One Hundred Years of Mouse Genetics: An Intellectual History. II. The Molecular Revolution (1981–2002)

Kenneth Paigen

The Jackson Laboratory, Bar Harbor, Maine 04609-1517

To a Well-Connected Mouse
(*Upon reading of the genetic closeness of mice and men.*)

Wee, sleekit, cow'rin, tim'rous beastie,
Braw science says that at the leastie
We share full ninety-nine per cent
O'genes, where'er the odd ane went.

A previous *Perspectives* (Paigen 2003) described the development of mouse genetics from its beginnings in 1902 until the beginning of the molecular revolution. Here, we pick up the story and follow it through to the end of its first century.

Understanding human biology in both its normal and pathological aspects requires experimental material. Despite their best efforts, clinical scientists face some intrinsic difficulties. Deliberately inducing pathology or toxicity is unacceptable; patients are understandably reluctant to provide serial tissue samples; a critical portion of human life, the embryonic/fetal period, is not very accessible; genetically defined lines cannot be created; the generation time is too long for extended genetic studies; and individuals exhibit a stubborn tendency to choose their own mating partners, frustrating geneticists.

In searching for an experimental surrogate, we are fortunate that the basic features of mammalian biology have changed little over the 75+ million years since the major orders of mammals diverged. This is true not only of gross anatomy and the major physiological systems, but at the molecular level as well. The recent genomic sequencing efforts suggest that we share 99% of our genes, and hence our molecular building blocks, with another mammal, the mouse.

In vitro and cell culture systems are very powerful, but they do not allow us to study physiological systems in their entirety, and certainly not their interactions, which are many. For these we need the whole animal.

In choosing a mammalian model, we want a creature that breeds rapidly, can be inbred (not all mammals can), whose husbandry is facile, and that is small and hence cheap to maintain. This last characteristic is of no small importance. Enter the mouse. Sixty years of classical genetics (Paigen 2003) resulted in the development of a very sophisticated genetic system with hundreds of inbred (RI) strains, even more mutants, a dense genetic map, sophisticated mapping resources such as recombinant inbred lines and their derivatives, and a well-established system of husbandry.

The dramatic transformation of mouse genetics into a truly powerful system for understanding mammalian biology at the molecular level and its relevance to the human state came when the newly emerging techniques of molecular biology were added to the well-established genetic systems of mice. We can date the birth of that union to a few months centered around New Year's Day of 1981, when the first transgenic mice were created. Since then, two complementary lines of experimentation have dominated, developing either our ability to begin with a known DNA sequence and answer questions about its functions in the organism or our ability to begin from a particular phenotype and discover the genes required for its normal or abnormal expression. Enhancing both genotype- and phenotype-driven approaches has been the ability to do this iteratively, first discovering genes underlying phenotypes and then using the powers of molecular gene manipulation to study their function and search for additional related genes or, conversely, beginning from the DNA sequences themselves and deducing from directed mutations the range of phenotypes controlled by each gene.

FROM GENOTYPE TO PHENOTYPE

Transgenic mice: Five laboratory groups more or less simultaneously showed that viral or mammalian DNA injected into mouse oocytes could be incorporated into the genome (Gordon *et al.* 1980; Brinster *et al.* 1981; Costantini and Lacy 1981; Harbers *et al.* 1981; Wagner *et al.* 1981). Chronologically, Jon Gordon and his colleagues (Gordon *et al.* 1980) were the first, but the experiment that forced a dramatic change in our thinking was that of T. E. Wagner and associates (Wagner *et al.* 1981). The latter group showed that an intact rabbit β-globin gene introduced into the mouse genome was expressed in a correct tissue-specific manner, resulting in the presence of rabbit protein in mouse erythrocytes. The implications of this finding were stunning. Not only did it make clear that correct expression of a gene is not limited to the environment of a unique chromosomal location, but for the first time we saw that developmental, tissue-specific regulatory systems are conserved and can function across orders of mammals, implying that the function and regulation of human as well as rabbit genes could be studied in an experimental species as tractable as the mouse. This was soon followed by a dramatic demonstration that captured the imagination of the lay public as well as scientists: Richard Palmiter, Ralph Brinster, and colleagues fused the rat growth hormone gene to a mouse metallothionein promoter and inserted this as a transgene, obtaining very high levels of gene expression and giant mice that made the cover of Nature (Palmiter *et al.* 1982). A flood of experimentation followed from many laboratories, and Palmiter has provided a detailed review of those early years (see Palmiter and Brinster 1986).

Homologous recombination and knockouts: The next dramatic advance, which opened the field to much broader applications, came a few years later when gene replacement became possible in mice. Initially, this led to the creation of absolute knockouts and later to tissue-specific and conditional knockouts. Two lines of experimentation had to merge for this to become possible. One was the development of cultured pluripotent embryonic stem cells (ES cells), which could be reimplanted into mouse blastocysts, giving rise to chimeric mice whose adult tissues, including their germ line, were derived from the inserted ES cells and thus were capable of transmitting ES cell genetic material to their offspring. The other line of experimentation was the achievement of homologous recombination in ES cells in a manner that left them pluripotent and germ-line competent. Together, these made gene-specific, genetic engineering possible. This molecular revolution took nearly a decade, and it is a tribute to the perseverance of all involved (as a reading of the articles published during this period makes clear) that the ultimate goal was apparent from the outset.

ES cell cultures were first obtained in 1981 by Martin Evans and Matt Kaufman (Evans and Kaufman 1981) and by Gail Martin (Martin 1981), following on the intellectual and experimental foundations laid by Roy Stevens and Barry Pierce. When these cells were introduced into blastocysts, they gave rise to a variety of adult tissues, including the germ line (Bradley *et al.* 1984). In 1986, two groups showed that genetically altered ES cells, either neomycin-resistant mutants (Gossler *et al.* 1986) or cells carrying retroviral insertions (Robertson *et al.* 1986), could transmit these traits to offspring. In the following year, HPRT-deficient mice were created from ES cells selected for this mutation (Hooper *et al.* 1987; Kuehn *et al.* 1987).

The challenge then was to achieve targeted homologous recombination in ES cells. In 1985, Smithies and his group (Smithies *et al.* 1985) successfully targeted a normally resident gene, the human β-globin locus, in other cultured cells, and two years later the Smithies (Doetschman *et al.* 1987) and Capecchi (Thomas and Capecchi 1987) groups both achieved gene targeting in ES cells. Two technical advances, positive/negative selection (Mansour *et al.* 1988) and the development of an appropriate enhancer/promoter system (Thomas and Capecchi 1987), brought the efficiency of targeting up and enhanced selection in ES cells, and in 1989–1990 three groups reported the first gene-targeted mice (Koller *et al.* 1989; Thompson *et al.* 1989; Thomas and Capecchi 1990).

A practical system had come into existence, and a virtual land rush began to knock out genes of biological interest. Many of these early experiments are well described in the addresses given by Evans, Smithies, and Capecchi when they jointly received the 2001 Lasker Award (Evans *et al.* 2001).

The early knockout results were sobering; the phenotypes of homozygous null mutations were often quite unexpected. When *MyoD*, a gene thought to be essential for muscle differentiation (Weintraub *et al.* 1989) was knocked out, there was little if any reduction in muscle mass (Rudnicki *et al.* 1992). *MyoD* turned out to be redundant to *Myf-5*, and both must be inactivated to affect muscle differentiation (Rudnicki *et al.* 1993). Similarly, when *p53*, the gene most frequently mutated in human tumors and presumed to be essential for orderly cell division, was knocked out, it was found not to be a developmental lethal (Donehower *et al.* 1992; Jacks *et al.* 1994). The resulting homozygotes were born alive and well. The importance of *p53* was vindicated when the mice came down with multiple tumors months later. Despite these frequent lessons in humility to the research community, knockouts have provided an extraordinarily powerful research tool, making possible the analysis of many physiological processes in ways that were previously unimaginable.

That power soon stimulated refinements to overcome the developmental lethality many knockouts exhibited. The Cre/loxP recombinase system was stolen from bac-

teriophage P1 and introduced into mammalian systems to allow tissue-specific gene inactivation. When mice received a target gene flanked by loxP sites from one parent and Cre recombinase under the control of a tissue-specific promoter from the other, the target gene is deleted from chromosomes only in tissues where the recombinase is expressed (Lakso *et al.* 1992; Orban *et al.* 1992). The microbial Flp system has been similarly employed (Wigley *et al.* 1994).

Later, to satisfy the desire to modulate gene function in a reversible manner, target genes were put under the control of the *tet* repressor, again a regulatory system borrowed from microorganisms, allowing them to be reversibly turned on and off by the presence or absence of tetracycline analogs in drinking water (Furth *et al.* 1994). An early demonstration of just how powerful tissue-specific reversible control could be appeared in the experiments of Mayford *et al.* (1996), who examined the role of CaMKII in the molecular basis of learning and memory by reversibly controlling the presence of this protein in specific neural tissues.

The recent discovery of RNA interference (dubbed RNAi), in which small, double-stranded RNA sequences attenuate the expression of genes containing the same sequence, holds the promise of advancing matters further by offering a much simpler technology for eliminating the functions of particular genes. Obtaining homozygotes of conventional knockouts requires isolating homologous recombinants in ES cells, injecting these into blastocysts that are then implanted in pseudo-pregnant females, obtaining chimeric offspring, and then breeding these for at least two generations to obtain homozygotes. If transgenes expressing the requisite RNA sequences as palindromes can be used as dominant negative regulators of gene function through RNA interference, this would provide a technically far easier means of generating knockouts. Combining this with tissue-specific and/or regulatable promoters would provide a very facile experimental system. Already, success has been achieved in cell culture systems (Elbashir *et al.* 2001; Brummelkamp *et al.* 2002; Paddison *et al.* 2002; Sui *et al.* 2002), and it seems likely that we shall soon see success in mice.

Lessons: Transgenic and knockout technologies have been critical in answering many questions about the role of specific genes in the biology of the organism. Beyond these individual answers obtained over the past 20 years, there have been some general lessons. One was, of course, the recurring difficulty of predicting phenotype. Another was just how pleiotropic mutations are, emphasizing what some classical geneticists had been saying for years: we do not have genes "for" a particular phenotype, but rather a particular phenotypic change is likely to be only one among the many consequences of the mutation in question, a point well understood by human geneticists and reflected in the fact that so many human gene defects are called "syndromes."

FROM PHENOTYPE TO GENOTYPE

Spontaneous and induced mutations: The obvious requirement of gene to phenotype studies is knowing which gene to begin with. And so there was a major simultaneous push to go in the other direction, starting from mice with altered or different phenotypes and then identifying the genes responsible.

This began with efforts to clone already existing mouse mutations by using a "positional" cloning strategy. The logic was straightforward. Carry out genetic crosses on a scale sufficient to refine the location of a mutation down to molecular dimensions, something less than a megabase, and then explore the genes within that region for sequence and/or expression changes in the mutant. The fact that these mutations arose on inbred strain backgrounds meant that the only sequence or expression differences present should be related to the original mutation, and having multiple alleles at the mutant locus was a definite assist.

An early and important success came with the cloning of the *W* and *Steel* mutations, showing that *W* encodes the c-kit tyrosine kinase receptor (Chabot *et al.* 1988; Geissler *et al.* 1988) and *Steel* encodes its ligand (Brannan *et al.* 1991; Flanagan *et al.* 1991). This was a fine confirmation of earlier transplantation studies between these two mutants and with wild-type animals showing that *Steel* and *W* determine, respectively, the ability to send and receive signals essential for hematopoietic stem cell differentiation. Elizabeth Russell (1979) has provided a review of these initial efforts.

Other early successes encouraged others by showing that gene cloning can help explicate biological phenomena. Two examples were the cloning of the *agouti* (*A*) gene in Rick Woychik's laboratory (Bultman *et al.* 1992) and the *obese* (*ob*) gene in Jeff Friedman's laboratory (Zhang *et al.* 1994). The latter identified leptin, a new peptide hormone acting as a satiety factor. Shortly thereafter, the *db* gene was also cloned, proving to be the leptin receptor. Together, these confirmed Doug Coleman's original hypothesis regarding these mutations (Paigen 2003).

Many additional mutations were soon cloned; many were classical mutations, and others arose as accidental consequences of transgene insertions. Indeed, a small industry quickly emerged, providing multiple new insights into old biological problems. (For a review of these many efforts, see Meisler 1992; Bedell *et al.* 1997; and Woychik and Alagramam 1998.) So successful were these efforts that researchers soon sought a broader supply of mutants to challenge their efforts, especially mutants altered in specific phenotypes. They turned to ethylnitrosourea (ENU), a powerful mutagen in male mice whose very high mutation rates make it feasible to seek changes in specific phenotypes. The discovery that ENU is an exceptionally powerful mutagen in mice had come some years earlier in the work

of the Russells at Oak Ridge, using the "seven locus" test for chemical mutagenesis (Russell *et al.* 1979; Davis and Justice 1998). Early success showing that it was now feasible to screen for and then clone mutations with specific phenotypes came from Vernon Bode's laboratory (Bode 1984; McDonald *et al.* 1988). Soon the Dove laboratory created and cloned a mutation causing colon cancer (Moser *et al.* 1990, 1993), which proved to be the ortholog of the human *Apc* gene with the same phenotype, and Joe Takahashi's laboratory, working with Dove, created and cloned the *Clock* mutation affecting circadian locomotor activity (Vitaterna *et al.* 1994; King *et al.* 1997).

These efforts had a considerable impact on thinking within the research community. In 1998 the National Institutes of Health (NIH) convened the first of several meetings of the mouse genetics community to recommend important new initiatives that NIH might undertake to further the use of mice in biological research. Stimulated by the early successes with ENU, one of the major recommendations was the undertaking of large-scale mutagenesis programs. (For a full report of that and later meetings go to http://www.nih.gov/science/models/mouse/.) Similar mutagenesis initiatives were launched elsewhere; more than a dozen such efforts are under way in Germany, the United Kingdom, Australia, Japan, Canada, and the United States, providing hundreds of new mutants per year. Many of these mutations affect phenotypes that were not possible to search for previously, such as seizure susceptibility, hypertension, fear-potentiated startle response, immune responses, and blood chemistry, phenotypes that would be too difficult or too expensive to test for without the very high mutant frequency provided by ENU.

Natural polymorphisms: The other phenotype-driven strategy for finding the genes and their products underlying complex traits depends on exploiting the remarkable phenotypic diversity found among inbred mouse strains, a diversity that often exceeds that of the entire human population. Indications that such diversity might exist came in earlier, smaller strain surveys and has become abundantly apparent with the onset of a systematic effort, the phenome project, to characterize a broad array of phenotypes across a standard set of inbred strains and then provide these data in the form of a searchable database (http://www.jax.org/phenome). The data on blood cholesterol levels serve to illustrate this diversity: among the tested set of inbred strains, total cholesterol levels vary fivefold, HDL-cholesterol levels tenfold, and HDL cholesterol as a percentage of total cholesterol ranges from 5 to 75%.

It is interesting to speculate on the sources of this exceptional diversity. Two obvious possibilities suggest themselves. First, inbred strains of laboratory mice represent a mosaic of genetic material from at least three wild species or subspecies, *Mus musculus musculus*, *M. m. domesticus,* and *M. m. castaneus*, and perhaps others (Beck *et al.* 2000). They represent a geographic range from the Atlantic shores of Western Europe to Japan. It may well be that what were once physiologically compatible sets of alleles at multiple loci in the original wild populations have been scrambled into less compatible, new arrangements in the course of the forced inbreeding used to create our present-day inbred strains. Second, our inbred mice are exactly that, homozygous at virtually every gene, having lost all the physiological buffering effects of heterozygosity that are obtained in natural populations and strain crosses.

This diversity among inbred strains can be exploited to find the genes determining physiologically important phenotypes, especially those relevant to human health. The conceptual basis for doing so stems from the landmark paper by Eric Lander and David Botstein (Lander and Botstein 1989) proposing a methodology for using a dense array of genetic markers to localize the genetic determinants for a polymorphic quantitative trait to particular chromosomal regions. The paper from John Todd's group (Todd *et al.* 1991) examining diabetes susceptibility not only initiated efforts to identify such quantitative trait loci (QTL) in mice, but also introduced two major advances in the approach. One was technical, for the first time employing as genetic markers in mice the polymorphic microsatellite markers (simple sequence repeats) detected the year before in human populations (Litt and Luty 1989; Tautz 1989; Weber and May 1989). The other advance was conceptual, breaking a complex phenotype into subphenotypes that are likely to have a simpler genetic causality. In this case, the subphenotype was insulitis, which turned out to be determined by one, but not the other, of the two new diabetes loci they had discovered.

Since these initial efforts, a wide variety of phenotypes representing multiple physiological systems have been mapped in mice (Paigen 2002). Initially, molecular identification of the underlying genes was difficult, but that is changing rapidly with the availability of the mouse genomic DNA sequence and the use of large-scale expression assays relying on arrays, chips, and especially real-time PCR. The year 2001 saw almost as many naturally polymorphic genes successfully cloned as in all previous years combined (Korstanje and Paigen 2002).

Human-mouse concordance: For years it has been an open question whether the genetic polymorphisms governing phenotypic diversity in the mouse and human populations would be the same. One can make theoretical arguments for either answer. The recently emerging experimental data, which have enormous implications for understanding our own species, indicate that a rather limited set of genes is responsible for most of the natural polymorphisms underlying many complex pathologies and that the salient loci are highly concordant in their chromosome locations between mouse and humans.

In the mouse, 27 chromosome regions affecting plasma HDL-cholesterol levels have been mapped, many recurring in crosses between different mouse strains. From the known map relationships between the human and mouse genomes, it is apparent that 18 of the 22 human QTL so far identified occur at corresponding chromosome locations (Wang and Paigen 2002). Recognizing that human variation largely occurs among the same set of genes requires that most of the important genes have already been identified in both species; *i.e.*, not more than a few dozen genes account for most of the genetic variation in both the mouse and the human.

Concordance is not limited to HDL levels, but is also observed for hypertension (Sugiyama *et al.* 2001) as well as asthma, inflammatory bowel disease, non-insulin-dependent (Type II) diabetes, and osteoporosis (B. Paigen, personal communication). The significance of this generality is great. If the same limited set of genes is responsible for most population variation in the human and the mouse, this suggests that these genes represent a special subset of all genes that could potentially influence a phenotype and that this subset codes for proteins that play critical regulatory roles. Because we already know that variation in these proteins does affect phenotype, these naturally polymorphic proteins present particularly inviting targets for drug intervention.

Finding and eventually cloning these genes in humans are now both difficult and expensive, either requiring extensive family studies, which are inefficient when a gene is only one among the several that contribute to the phenotype, or requiring linkage studies on populations, looking for statistical association between phenotype measures and genetic markers. Since haplotype blocks in non-African populations are on the order of 20 kb in size, a genome-wide scan for one individual could easily cost upward of $50,000 and for a statistically useful population, many millions. Finding these natural polymorphisms in mice and then using that information to guide human studies presents a low-cost, rapid shortcut.

WHAT HAVE WE LEARNED?

The union of molecular biology, genomics, and mouse genetics has led to major advances in our understanding of mammalian biology. A few of the many notable examples include insights into the molecular circuitry underlying embryonic and fetal development, the existence of multiple stem cell types in the adult organism, the genetic changes in tumorigenesis, and the emergence of molecular pathology as a full-fledged discipline exploring a wide variety of human ills. Detailed discussions of these are far beyond the scope of this *Perspectives*, but several points deserve comment.

Evolutionary conservation: The power of model systems for understanding human biology obviously depends on their degree of similarity to the human state. We have long known that all mammals share their basic anatomy and physiology. Now, as our knowledge of genomics has advanced over these last 20 years, we have learned just how conservative evolution has been at the molecular level as well. It has been startling to realize that not only are metabolic pathways conserved over evolutionary time, but genetic regulatory systems, the intracellular signaling pathways, and the molecules that carry out these functions are conserved as well. We are less surprised, then, when a comparison of the genomic sequences of mice and humans indicates that we share 99% of our genes and confirms earlier genetic studies showing that the arrangement of genes along chromosomes is largely preserved among mammals. Additionally, the discovery of strong sequence conservation in some noncoding regions is proving a useful means of identifying sequences whose evolution has been constrained by the need to preserve important regulatory and chromosomal structural elements encoded in DNA.

Imprinting: The first indication that the maternal and paternal genomes contributed to the zygote do not function equivalently came from nuclear transplantation experiments (McGrath and Solter 1984). Diploid zygotes created by combining a male and a female pronucleus gave rise to viable embryos, but combining two male or two female pronuclei did not. The nature of the abnormalities depended on the parental source of the pronuclei, making clear that there is a reciprocal, essential difference between the two. This was soon confirmed in genetic crosses that used translocations to produce animals in which both copies of a chromosome segment came from the same parent (Cattanach and Kirk 1985) and was put on a molecular basis when the *Tgf2r* (Barlow *et al.* 1991) and *H19* (Bartolomei *et al.* 1991) loci were shown to be expressed from only one parental allele. We now understand that the physical basis for imprinting is differential DNA methylation, which is replicated during mitosis and reset in the germ line. Almost simultaneously with these studies in mice, parallel investigations showed imprinting to be a significant phenomenon in human genetics (Hall 1997).

Informatics: The average quantity of data generated per scientist per year has been increasing exponentially for decades. This is true in many areas of science and is apparent whether we look at the numbers of new organic molecules synthesized, genes mapped, protein X-ray crystal structures solved, or megabases of new DNA sequence deposited in the GenBank. One suspects that this is true because a substantial fraction of our total research budgets are spent on the research and development of the research process itself, developing new technologies as well as increasing the efficiency of older ones. DNA sequencing is the example that immediately comes to mind. In not much more than a decade we have gone from laboriously sequencing a few hundred base pairs at a time by hand to implementing automated machines that deliver tens of kilobases of sequences in

a day. Storing the new volumes of information, making them accessible in a useful manner to researchers, and mining them to gain insight into biological processes require computers, very large, sophisticated databases, and a new science of computational biology. Providing some sense of scale, Baxevanis (2002) has provided summaries of 335 databases relevant to molecular genetics available at http://nar.oupjournals.org. More specific access to mouse-related information is available at http://www.informatics.jax.org.

Phenotyping: At the present time, the major limiting factor in our ability to apply the powers of mammalian genetics toward understanding mammalian biology is our capacity to measure phenotype. True, new experimental resources and improvements in genetic technologies that will further improve our discovery tools are on the way, but what limits their application is the ability to measure appropriate phenotypes. Developing means for measuring atherosclerotic plaques in mice opened a new field of research, as did learning to measure intraocular pressure, hyperactive airways, and fear-potentiated startle responses. The steady miniaturization of physiological measuring techniques and the introduction of imaging technologies (MRI, CT, PET, DEXA, and the like) are making it increasingly feasible to use mice and their genetic systems for the exploration of phenotype, and we can expect this to continue.

Pharmaceuticals: All of the drugs now in common use are targeted at fewer than 400 proteins, and these are largely surface receptors. In this context, consider that the number of mammalian genes is currently estimated at 30,000, that splice variants raise the potential number of polypeptides to something approaching 100,000, and that our knowledge of molecular pathology, while growing, is still rudimentary. All this suggests that genetic approaches directed at furthering our understanding of the molecular basis of disease are likely to have a profound impact on the development of new and more effective classes of drugs. Genetics will also help to resolve the problem that what is ostensibly one disease to the pathologist can arise from a variety of molecular causes, that each of these may require its own therapy, and that we also vary in our responses to pharmaceuticals.

THE PRESENT STATE OF AFFAIRS

We largely understand the basic mechanisms of genetics, how traits are passed from one generation to the next, as well as the fundamental mechanisms of gene action. The major challenges we now face are in deciphering the regulatory circuits that control gene expression; understanding the relationship between gene function and the higher order, three-dimensional structure of chromosomes; and continuing to improve our technical capabilities in multiple ways.

Importantly, mouse molecular genetics has moved beyond being a subject of research to becoming a research tool in its own right. Our existing capabilities make it possible to manipulate the genome and analyze its contributions in determining phenotype, both normal and abnormal. The great challenge is in applying these capabilities to improving the human condition. It also seems reasonable to expect that, as our knowledge of the molecular regulatory systems underlying development increases, so will our ability to alter developmental processes, providing additional and potentially powerful means of perturbing biological systems for discovery purposes.

Our present circumstances are virtually unique. Traditionally, advances in scientific knowledge do not come at a steady pace. Long periods are spent consolidating sporadic major advances, exploiting the new possibilities until the next breakthrough arrives. As time passes, advance becomes increasingly constrained until either a new technology emerges, opening up entirely new realms of investigation (think of the first light and electron microscopes), or a conceptual advance provides a new way of thinking about old problems (think of Mendelian genetics or the DNA double helix). But mouse genetics now finds itself in the unprecedented circumstance of being neither technically nor conceptually limited. We face highly productive years, even using only what we already understand conceptually and what we already know how to do technically. Further, the absence of either limitation is generating a positive feedback loop, stimulating both new conceptual frameworks and new technologies. It is a circumstance virtually unprecedented in the history of the biological sciences.

In the past few years, we have been introduced to multipotential adult stem cells, insulator sequences, and haplotype analysis, and on the technical side we have seen the introduction of gene expression profiling, SNPs and RNAi, and now have nearly the entire mouse genomic DNA sequence available. There is no expectation of a placid future, and if anything, the rate of change is accelerating. One can appreciate this sea change via a simple thought experiment. Place yourself mentally in the year 1988 and ask how far back in time it was possible to go and still be able to reasonably predict the 1988 state of affairs. Then do the same in 1995, and then now. This expansion of possibilities goes beyond our intellectual desire to understand biological systems; it goes to the heart of the reason society supports our efforts: understanding the nature of our physical being toward the goal of improving the quality and duration of human life.

WHERE IS ALL THIS TAKING US?

Certainly, we are moving toward a far more profound understanding of molecular pathology: how disease comes about at the cellular and molecular level, what the critical molecular circuits and events are, and how

we differ from each other in our genetic bases. It is also driving us to move beyond the very reductionist approach of one gene, one protein, one phenotype that has been so powerful the last half-century and increasingly to frame our thinking around entire physiological systems considered at the molecular level.

As an outcome, we can expect dramatic changes in our strategies for human intervention, both therapeutic and preventive. Some of these will be idiosyncratic to the particular ill they address, but others will be generic. These latter are likely to include drugs directed at the control of gene expression, manipulation of the immune system, and the use of stem cell derivatives in replacement therapy. At the least, we can expect conventional drugs that are directed at much more rationally chosen protein targets. Much of this effort will be directed, in one form or other, at modulating the regulatory systems controlling physiological systems.

We are also coming to an increased understanding of our genetic/environmental interactions. Genetics makes sense only in an environmental context, and while we cannot readily alter the genomes our parents gave us, we can alter our lifestyle, diet, and exposure to toxins.

It is difficult to predict in which physiological systems our knowledge, and hence our ability to intervene in terms of therapy and prevention, will advance most rapidly. But considered across the broad spectrum of human disease, we can confidently expect remarkable advances in our fundamental understanding.

Enormous challenges lie in translating this understanding into societally acceptable, practical advances and, even more, in accommodating the altered perceptions that will arise in our views of normality and pathology, for both physiology and behavior. In this translation of basic knowledge into practice, it is not clear that our present pharmaceutical/biotech, for-profit structure will entirely suffice. It may well be that for some classes of problems the nonprofit sector will need to assume additional roles. And as to our cultural perceptions, consider only how we will address the issue of personal responsibility for antisocial, aggressive behavior if, as seems likely, that turns out to contain a significant genetic component. Conflicts between new, scientifically based insights and historic cultural norms can be painful to resolve: witness over a century of tension between evolutionary concepts and religious beliefs and our present societal conflicts over reproductive rights and stem cell research.

CODA

It would be hard to devise a more emblematic close to the first century of mouse genetics than the December 5, 2002, issue of Nature, containing as it does a detailed annotation of the public draft sequence covering some 96% of the mouse genome (Waterston *et al.* 2002), a very broad and deep survey of mouse cDNA clones extracted from a wide variety of embryonic and adult tissue samples (Okazaki *et al.* 2002), and a provocative sampling of the haplotype structure of inbred mouse strains (Wade *et al.* 2002). These descriptions of the mouse genome are followed by two prophetic surveys of gene expression for nearly all the mouse orthologs of the genes on human chromosome 21, where trisomy leads to Down syndrome (Gitton *et al.* 2002; Reymond *et al.* 2002).

The details of the sequence annotation are fascinating, and several highlights are exceptional (Waterston *et al.* 2002):

Some 99.5% of mouse genes have a clear human counterpart, and nearly all fall within the predicted syntenic interval.

In support of previous evidence, the two genomes are indeed organized in a highly parallel fashion. Something over 90% of the respective genomes are contained in 216 conserved syntenic blocks whose median size is 23 MB; within each block there are occasional rearrangements giving rise to a total of 326 segments, each with entirely conserved gene order.

Evolutionary conservation goes further. Some 5% of the genome shows significant sequence conservation, suggesting common functions in the mouse and human; yet, at most, only 2.5% is accounted for by coding and associated sequences. What is the other conserved 2.5%?

The two genomes have not been evolving at equivalent rates. As judged from divergence among members of lineage-specific repeat families, since the mammalian radiation the average rate of base substitution has been twice as high in the rodent lineage, and this ratio may be as high as 5:1 in the recent past. However, after correcting for generation times the picture reverses, and the rate per generation in primates is probably an order of magnitude higher than that in rodents. Since nearly all human point mutations arise in males (Crow 2000), probably because of the larger number of cell divisions in the male germ line, it is possible that the difference between mice and humans simply reflects our relative size and time to sexual maturity.

The mammalian genome is hardly uniform in its properties. Beyond variation in base composition, as Waterston *et al.* (2002) note: "It is clear that the mammalian genome is evolving under the influence of non-uniform local forces." There is substantial variation across the genome in neutral base substitution rates, SNP density (in humans), insertion and deletion rates, and the local frequency of meiotic recombination. Remarkably, these rates are correlated (but not with base composition), reflecting as yet unappreciated features of chromosome organization.

The FANTOM consortium (Okazaki *et al.* 2002) re-

ported their analysis of 61K full-length cDNA sequences clustered into 33K "transcriptional units" providing alternate splice forms. Derived from over 200 embryonic and adult tissue samples, many normalized, this is by far the most extensive sampling of the mouse transcriptome so far. Among its highlights is the suggestion, supported by the genomic sequencing data, that a significant class of genes whose ultimate products are not proteins, but rather RNAs with novel functions, may exist.

The provocative suggestion in the report of Wade *et al.* (2002) on the haplotype structure of inbred mice is the possibility that our inbred mouse strains may each be mosaics of a very limited number of ancestral haplotypes, perhaps only two, representing *M. m. musculus* and *M. m. domesticus*. To the extent that this is true (and additional data are urgently required), it means that once they are haplotyped, the existing inbred mouse strains can be used as if they were a very large set of recombinant inbred strains with at least an order of magnitude increase in genetic resolution over existing RI lines. However, to the extent that this is true, it also requires that the progenitor populations of *M. m. musculus* and *M. m. domesticus* are virtually monomorphic, which would be quite remarkable.

The past 100 years have taken us from the first demonstration of Mendelian inheritance in mice to a nearly complete catalog of all mouse genes (many of whose functions are known), the mouse genomic sequence, and an extraordinarily powerful experimental system for understanding mammalian (*i.e.*, human) biology and pathology. And the pace of advance has steadily accelerated. It is impossible to predict what discoveries will come in the next 25, much less the next 100 years; the prospect is awesome.

LITERATURE CITED

Barlow, D. P., R. Stoger, B. G. Herrmann, K. Saito and N. Schweifer, 1991 The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature **349:** 84–87.

Bartolomei, M. S., S. Zemel and S. M. Tilghman, 1991 Parental imprinting of the mouse H19 gene. Nature **351:** 153–155.

Baxevanis, A. D., 2002 The Molecular Biology Database Collection: 2002 update. Nucleic Acids Res. **30:** 1–12.

Beck, J. A., S. Lloyd, M. Hafezparast, M. Lennon-Pierce, J. T. Eppig *et al.*, 2000 Genealogies of mouse inbred strains. Nat. Genet. **24:** 23–25.

Bedell, M. A., D. A. Largaespada, N. A. Jenkins and N. G. Copeland, 1997 Mouse models of human disease. II. Recent progress and future directions. Genes Dev. **11:** 11–43.

Bode, V. C., 1984 Ethylnitrosourea mutagenesis and the isolation of mutant alleles for specific genes located in the T region of mouse chromosome 17. Genetics **108:** 457–470.

Bradley, A., M. Evans, M. H. Kaufman and E. Robertson, 1984 Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature **309:** 255–256.

Brannan, C. I., S. D. Lyman, D. E. Williams, J. Eisenman, D. M. Anderson *et al.*, 1991 Steel-Dickie mutation encodes a c-kit ligand lacking transmembrane and cytoplasmic domains. Proc. Natl. Acad. Sci. USA **88:** 4671–4674.

Brinster, R. L., H. Y. Chen, M. Trumbauer, A. W. Senear, R. Warren *et al.*, 1981 Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. Cell **27:** 223–231.

Brummelkamp, T. R., R. Bernards and R. Agami, 2002 A system for stable expression of short interfering RNAs in mammalian cells. Science **296:** 550–553.

Bultman, S. J., E. J. Michaud and R. P. Woychik, 1992 Molecular characterization of the mouse agouti locus. Cell **71:** 1195–1204.

Cattanach, B. M., and M. Kirk, 1985 Differential activity of maternally and paternally derived chromosome regions in mice. Nature **315:** 496–498.

Chabot, B., D. A. Stephenson, V. M. Chapman, P. Besmer and A. Bernstein, 1988 The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. Nature **335:** 88–89.

Costantini, F., and E. Lacy, 1981 Introduction of a rabbit beta-globin gene into the mouse germ line. Nature **294:** 92–94.

Crow, J. F., 2000 The origins, patterns and implications of human spontaneous mutation. Nat. Rev. Genet. **1:** 40–47.

Davis, A. P., and M. J. Justice, 1998 An Oak Ridge legacy: the specific locus test and its role in mouse mutagenesis. Genetics **148:** 7–12.

Doetschman, T., R. G. Gregg, N. Maeda, M. L. Hooper, D. W. Melton *et al.*, 1987 Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature **330:** 576–578.

Donehower, L. A., M. Harvey, B. L. Slagle, M. J. McArthur, C. A. Montgomery, Jr. *et al.*, 1992 Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature **356:** 215–221.

Elbashir, S. M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber *et al.*, 2001 Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature **411:** 494–498.

Evans, M. J., and M. H. Kaufman, 1981 Establishment in culture of pluripotential cells from mouse embryos. Nature **292:** 154–156.

Evans, M., O. Smithies and M. R. Capecchi, 2001 Mouse gene targeting. Nat. Med. **7:** 1081–1090.

Flanagan, J. G., D. C. Chan and P. Leder, 1991 Transmembrane form of the kit ligand growth factor is determined by alternative splicing and is missing in the Sld mutant. Cell **64:** 1025–1035.

Furth, P. A., L. St. Onge, H. Boger, P. Gruss, M. Gossen *et al.*, 1994 Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. Proc. Natl. Acad. Sci. USA **91:** 9302–9306.

Geissler, E. N., M. A. Ryan and D. E. Housman, 1988 The dominant-white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. Cell **55:** 185–192.

Gitton, Y., N. Dahmane, S. Baik, A. Ruizi Altaba, L. Neidhardt *et al.*, 2002 A gene expression map of human chromosome 21 orthologues in the mouse. Nature **420:** 586–590.

Gordon, J. W., G. A. Scangos, D. J. Plotkin, J. A. Barbosa and F. H. Ruddle, 1980 Genetic transformation of mouse embryos by microinjection of purified DNA. Proc. Natl. Acad. Sci. USA **77:** 7380–7384.

Gossler, A., T. Doetschman, R. Korn, E. Serfling and R. Kemler, 1986 Transgenesis by means of blastocyst-derived embryonic stem cell lines. Proc. Natl. Acad. Sci. USA **83:** 9065–9069.

Hall, J. G., 1997 Genomic imprinting: nature and clinical relevance. Annu. Rev. Med. **48:** 35–44.

Harbers, K., D. Jahner and R. Jaenisch, 1981 Microinjection of cloned retroviral genomes into mouse zygotes: integration and expression in the animal. Nature **293:** 540–542.

Hooper, M., K. Hardy, A. Handyside, S. Hunter and M. Monk, 1987 HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature **326:** 292–295.

Jacks, T., L. Remington, B. O. Williams, E. M. Schmitt, S. Halachmi *et al.*, 1994 Tumor spectrum analysis in p53-mutant mice. Curr. Biol. **4:** 1–7.

King, D. P., Y. Zhao, A. M. Sangoram, L. D. Wilsbacher, M. Tanaka *et al.*, 1997 Positional cloning of the mouse circadian clock gene. Cell **89:** 641–653.

Koller, B. H., L. J. Hagemann, T. Doetschman, J. R. Hagaman, S. Huang *et al.*, 1989 Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene

by homologous recombination in embryonic stem cells. Proc. Natl. Acad. Sci. USA **86:** 8927–8931.

Korstanje, R., and B. Paigen, 2002 From QTL to gene: the harvest begins. Nat. Genet. **31:** 235–236.

Kuehn, M. R., A. Bradley, E. J. Robertson and M. J. Evans, 1987 A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature **326:** 295–298.

Lakso, M., B. Sauer, B. Mosinger, Jr., E. J. Lee, R. W. Manning *et al.*, 1992 Targeted oncogene activation by site-specific recombination in transgenic mice. Proc. Natl. Acad. Sci. USA **89:** 6232–6236.

Lander, E. S., and D. Botstein, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

Litt, M., and J. A. Luty, 1989 A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. **44:** 397–401.

Mansour, S. L., K. R. Thomas and M. R. Capecchi, 1988 Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature **336:** 348–352.

Martin, G. R., 1981 Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. USA **78:** 7634–7638.

Mayford, M., M. E. Bach, Y. Y. Huang, L. Wang, R. D. Hawkins *et al.*, 1996 Control of memory formation through regulated expression of a CaMKII transgene. Science **274:** 1678–1683.

McDonald, J. D., R. G. Cotton, I. Jennings, F. D. Ledley, S. L. Woo *et al.*, 1988 Biochemical defect of the hph-1 mouse mutant is a deficiency in GTP-cyclohydrolase activity. J. Neurochem. **50:** 655–657.

McGrath, J., and D. Solter, 1984 Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell **37:** 179–183.

Meisler, M. H., 1992 Insertional mutation of 'classical' and novel genes in transgenic mice. Trends Genet. **8:** 341–344.

Moser, A. R., H. C. Pitot and W. F. Dove, 1990 A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. Science **247:** 322–324.

Moser, A. R., E. M. Mattes, W. F. Dove, M. J. Lindstrom, J. D. Haag *et al.*, 1993 ApcMin, a mutation in the murine Apc gene, predisposes to mammary carcinomas and focal alveolar hyperplasias. Proc. Natl. Acad. Sci. USA **90:** 8977–8981.

Okazaki, Y., M. Furuno, T. Kasukawa, J. Adachi, H. Bono *et al.*, 2002 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature **420:** 563–573.

Orban, P. C., D. Chui and J. D. Marth, 1992 Tissue- and site-specific DNA recombination in transgenic mice. Proc. Natl. Acad. Sci. USA **89:** 6861–6865.

Paddison, P. J., A. A. Caudy, E. Bernstein, G. J. Hannon and D. S. Conklin, 2002 Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes Dev. **16:** 948–958.

Paigen, K., 2002 Understanding the human condition: experimental strategies in mammalian genetics. ILAR J. **43:** 123–135.

Paigen, K., 2003 One hundred years of mouse genetics: an intellectual history. I. The classical period (1902–1980). Genetics **163:** 1–7.

Palmiter, R. D., and R. L. Brinster, 1986 Germ-line transformation of mice. Annu. Rev. Genet. **20:** 465–499.

Palmiter, R. D., R. L. Brinster, R. E. Hammer, M. E. Trumbauer, M. G. Rosenfeld *et al.*, 1982 Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. Nature **300:** 611–615.

Reymond, A., V. Marigo, M. B. Yaylaoglu, A. Leoni, C. Ucla *et al.*, 2002 Human chromosome 21 gene expression atlas in the mouse. Nature **420:** 582–586.

Robertson, E., A. Bradley, M. Kuehn and M. Evans, 1986 Germline transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature **323:** 445–448.

Rudnicki, M. A., T. Braun, S. Hinuma and R. Jaenisch, 1992 Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. Cell **71:** 383–390.

Rudnicki, M. A., P. N. Schnegelsberg, R. H. Stead, T. Braun, H. H. Arnold *et al.*, 1993 MyoD or Myf-5 is required for the formation of skeletal muscle. Cell **75:** 1351–1359.

Russell, E. S., 1979 Hereditary anemias of the mouse: a review for geneticists. Adv. Genet. **20:** 357–459.

Russell, W. L., E. M. Kelly, P. R. Hunsicker, J. W. Bangham, S. C. Maddux *et al.*, 1979 Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc. Natl. Acad. Sci. USA **76:** 5818–5819.

Smithies, O., R. G. Gregg, S. S. Boggs, M. A. Koralewski and R. S. Kucherlapati, 1985 Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature **317:** 230–234.

Sugiyama, F., G. A. Churchill, D. C. Higgins, C. Johns, K. P. Makaritsis *et al.*, 2001 Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. Genomics **71:** 70–77.

Sui, G., C. Soohoo, B. Affarel, F. Gay, Y. Shi *et al.*, 2002 A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. Proc. Natl. Acad. Sci. USA **99:** 5515–5520.

Tautz, D., 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. **17:** 6463–6471.

Thomas, K. R., and M. R. Capecchi, 1987 Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51:** 503–512.

Thomas, K. R., and M. R. Capecchi, 1990 Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. Nature **346:** 847–850.

Thompson, S., A. R. Clarke, A. M. Pow, M. L. Hooper and D. W. Melton, 1989 Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell **56:** 313–321.

Todd, J. A., T. J. Aitman, R. J. Cornall, S. Ghosh, J. R. Hall *et al.*, 1991 Genetic analysis of autoimmune type 1 diabetes mellitus in mice. Nature **351:** 542–547.

Vitaterna, M. H., D. P. King, A. M. Chang, J. M. Kornhauser, P. L. Lowrey *et al.*, 1994 Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. Science **264:** 719–725.

Wade, C. M., E. J. Kulbokas, III, A. W. Kirby, M. C. Zody, J. C. Mullikin *et al.*, 2002 The mosaic structure of variation in the laboratory mouse genome. Nature **420:** 574–578.

Wagner, T. E., P. C. Hoppe, J. D. Jollick, D. R. Scholl, R. L. Hodinka *et al.*, 1981 Microinjection of a rabbit beta-globin gene into zygotes and its subsequent expression in adult mice and their offspring. Proc. Natl. Acad. Sci. USA **78:** 6376–6380.

Wang, X., and B. Paigen, 2002 Quantitative trait loci and candidate genes regulating HDL cholesterol: a murine chromosome map. Arterioscler. Thromb. Vasc. Biol. **22:** 1390–1401.

Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril *et al.*, 2002 Initial sequencing and comparative analysis of the mouse genome. Nature **420:** 520–562.

Weber, J. L., and P. E. May, 1989 Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. **44:** 388–396.

Weintraub, H., S. J. Tapscott, R. L. Davis, M. J. Thayer, M. A. Adam *et al.*, 1989 Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proc. Natl. Acad. Sci. USA **86:** 5434–5438.

Wigley, P., C. Becker, J. Beltrame, T. Blake, L. Crocker *et al.*, 1994 Site-specific transgene insertion: an approach. Reprod. Fertil. Dev. **6:** 585–588.

Woychik, R. P., and K. Alagramam, 1998 Insertional mutagenesis in transgenic mice generated by the pronuclear microinjection procedure. Int. J. Dev. Biol. **42:** 1009–1017.

Zhang, Y., R. Proenca, M. Maffei, M. Barone, L. Leopold *et al.*, 1994 Positional cloning of the mouse obese gene and its human homologue. Nature **372:** 425–432.