

in this area have been remarkable, but are fraught with controversy. Couples considering pregnancy now have many options for genetic screening. In fact, those undergoing in vitro fertilization may now opt to have their embryos genetically screened before implantation (14). This can be helpful to couples whose offspring are known to be at risk for an inherited disease. Although some view this technology as a wonderful breakthrough, critics argue that it borders on eugenics.

In our lifetime, we have watched with amazement the progress of this field from the initial discovery of the structure of DNA in 1953 by Watson and Crick (15), to the present-day sequencing of the human genome. Increased understanding of the human genome may ultimately result in the eradication of common diseases, but in the meantime we need to be on guard against potential misuse of genetic information. This is an emerging technology, and we should proceed with caution. The science is expanding at a breathtaking pace, and the overwhelming amount of

new information puts governments under increasing pressure to pass legislation.

Eventually every country must decide what genetic information should be protected, who will have access to it, and how it may be used. In addition, governments must ensure that the public realizes practical gains from their investment in genetic technology, because much of the research is made possible by taxpayer-supported federal enterprises in partnership with academic and industrial institutions. Further, for this partnership to continue, the public must understand the new technologies so that unfounded fears will not develop and slow progress. Ultimately, the greatest difficulty will be for policy-makers to strike a balance between timely promotion and use of the best genetic research and careful protection of people from genetic discrimination.

Editor's note: The authors have chosen to express their individual views about future directions for legislation in the United States separately. See page 1250.

References and Notes

1. www.nhgri.nih.gov/ELSI/
2. The Gallup Organization for the Institute of Health Freedom, *Public Attitudes Toward Medical Privacy*, Survey 131007, 26 September 2000.
3. F. S. Collins, written testimony submitted to the U.S. Senate Health, Education, Labor, and Pensions Committee, Hearing on Genetic Information in the Workplace, 20 July 2000.
4. National Center for Genome Resources survey, *Attitudes Toward Genetic Testing*, March 4, 1998; see www.ncgr.org/about/news/1998/0304.html.
5. American Management Association, *The 2000 Workplace Testing Survey*, 31 March 2000; see www.amanet.org/research/summ.htm.
6. C. Venter, testimony before the U.S. Senate Health, Education, Labor, and Pensions Committee, Hearing on Genetic Information in the Workplace, 20 July 2000.
7. N. L. Jones, CRS Long Report for Congress, RL30006, *Genetic Information: Legal Issues Relating to Discrimination and Privacy*, 2 October 2000.
8. Executive Order 13145, 65 Fed. Reg. 6, 877, 10 February 2000.
9. SHRM position paper on genetic information (Society for Human Resource Management, Alexandria, VA, 2000).
10. P. S. Miller, *J. Health Care Law Policy* 3, 255 (2000).
11. B. P. Fuller et al., *Science* 285, 1359 (1999).
12. R. Gizbert, ABCNEWS.com, www.abcnews.go.com/on-air/WorldNewsTonight/wnt990218_iceland.html.
13. L. Frank, *Science* 290, 31 (1999).
14. D. S. King, *J. Med. Ethics* 25, 176 (1999).
15. J. D. Watson, F. H. C. Crick, *Nature* 171, 737 (1953).

FUTURE DIRECTIONS: SEQUENCE INTERPRETATION

Functional Annotation of Mouse Genome Sequences

The International Mouse Mutagenesis Consortium¹: Joseph H. Nadeau,* Rudi Balling [President, International Mammalian Genome Society (IMGS)], Greg Barsh, David Beier, S. D. M. Brown, Maja Bucan, Sally Camper (Secretariat, IMGS), George Carlson, Neal Copeland,† Janan Eppig, Colin Fletcher, Wayne N. Frankel, Detlev Ganten, Dan Goldowitz, Chris Goodnow, Jean-Louis Guenet (Secretariat, IMGS), Geoff Hicks, Martin Hrabe de Angelis, Ian Jackson (Vice President, IMGS), Howard J. Jacob, Nancy Jenkins,† Dabney Johnson, Monica Justice (Secretariat, IMGS), Steve Kay, David Kingsley, Hans Lehrach, Terry Magnuson (Secretariat, IMGS), Miriam Meisler (Past President, IMGS), AnneMarie Poustka, Eugene M. Rinchik, Janet Rossant, Lee B. Russell, John Schimenti, Toshihiko Shiroishi (Secretariat, IMGS), William C. Skarnes, Phil Soriano, William Stanford, Joseph S. Takahashi, Wolfgang Wurst, Andreas Zimmer

With the reports of the DNA sequence of the human genome and progress in sequencing the mouse genome, the first phase of the Human Genome Project is complete (1–3). Analysis of these DNA sequences will reveal the inventory of genes used for building these organisms, as well as many regulatory elements that compose the “instruction manual” for converting the genetic “parts list” into organismal form and function. Research attention is now beginning to shift from problems of gene struc-

ture and genome organization to questions of protein function and interactions, developmental and physiological pathways, and systems biology.

Various computational methods are being used to deduce functions for genes. Analyses of the genome sequences of species such as *Haemophilus influenzae*, *Helicobacter pylori*, *Caenorhabditis elegans*, and *Drosophila melanogaster*, and humans illustrate the power of these methods (1, 2). However, many fundamental aspects of biological functions are not directly evident in DNA sequences. It is not unusual to discover a gene sequence about which little functional information can be deduced. For example, sequence analysis leads to no prediction of function for as many as 30% of the genes in the human genome, and the inferred functions of

most of the remaining genes have yet to be proven (1, 2). Because of the striking sequence similarities between humans and mice (1), discoveries in one species lead to strong inferences in the other.

Laboratory mice and related species can make important contributions to functional genomics and identification of new models of human disease. Many spontaneous mutants have contributed profoundly to biomedical research and our understanding of disease etiology and pathogenesis. The ability to make crosses between genetically defined strains, to work with large sample sizes, to engineer mutations in specific genes, and to generate mice with induced mutations facilitates identification of genetic variants of biomedical interest. By including known single-gene mutants in surveys of mutagenized mice (also known as “sensitized surveys”), induced mutations that modulate the mutant phenotype can be identified, as was done with great success in the discovery of naturally occurring variants that suppress disease traits in *Apc* and *Cfr* mutant mice (4). These mouse models reveal new drug targets for adenomatous polyposis coli and cystic fibrosis, as well as provide ways to evaluate potential therapeutics, predict treatment effects, and prioritize treatments for clinical trials.

Remarkably, despite more than 100 years of research in mouse genetics, fewer than 5000 out of an expected total of 30,000 genes have functions attributed to them through direct experimental studies. Recent progress in mouse genetics and genomics has provided proof-of-principle for large-scale studies to produce compre-

¹The members of the IMMC are listed in (21).

*To whom correspondence should be addressed. Department of Genetics, BRB 624, Case Western Reserve University School of Medicine, 10900 Euclid Avenue, Cleveland, OH 44106, USA. E-mail: jhn4@po.cwru.edu

†The viewpoints stated here reflect their personal opinions and not those of the government.

hensive collections of mouse mutations. These mutant mice are a permanent resource for future biomedical research, raising the possibility of attributing functions to every gene in the genome.

International efforts to functionally annotate the DNA sequence of the mouse genome can only be achieved with a combination of phenotype- and gene-based approaches in small- and large-scale public and private projects. Genome-wide mutagenesis, gene-trap, insertional mutagenesis, and large-scale genome alterations, as well as many kinds of targeted mutagenesis, are the methods of choice. Mutagenesis with agents such as ethyl nitrosourea (ENU) can be carried out in a high-throughput fashion in phenotype-based surveys. Although ENU causes numerous simple molecular lesions in each mutagenized mouse, its random action as a chemical mutagen can be a disadvantage in trying to study genes that warrant immediate attention because of their perceived importance. Other genes, such as the last gene in the genome to be mutated, are difficult to obtain with agents that act randomly because of the properties of statistical sampling. In these cases, more directed strategies such as gene targeting and mutagenesis of embryonic stem (ES) cells *in vitro* are needed. Gene targeting has traditionally been used to obtain loss-of-function mutations, but recent developments (5–7) raise the possibility of obtaining many different kinds of mutations. In addition, systematic analysis of the effects of generalized or tissue-specific misexpression of gene products can also provide useful insights into gene function.

Public and private programs are under way in Australia, Canada, Germany, Japan, the United Kingdom, the United States, and other countries, with funding from governmental agencies, private foundations, and companies. These programs involve surveys for genes and traits affecting diverse biological traits and have already generated more than 1500 ENU-induced mutants (8, 9), 1500 gene-trap insertional mutations in ES cells (10), and roughly 4000 engineered mutants. Leaders of these programs now believe it is time to call for an International Mouse Mutagenesis Consortium (IMMC) that will systematically and comprehensively assign functions to every gene in the genome and will identify every gene that affects traits of high biomedical interest. This proposal is based on a public-private partnership with research groups in academia and industry working together to develop and apply genetic, genomic, phenotypic, bioinformatic, and computational methods to achieve the IMMC goals. Although we will focus here on commitments from the academic research community, we recognize that many companies are increasingly interested in the research opportu-

nities in this area. With the establishment of the IMMC, we extend an open invitation to commercial entities to join us in this endeavor or to create the enabling resources for the next generation of genomics research.

IMMC Goals

We propose these long-range goals:

- Produce at least one heritable mutation, in either ES cells or mice, in every gene in the genome.
- Identify every gene that affects key traits of biomedical interest.
- Establish an infrastructure for preserving and distributing mutant cells and mice.
- Enhance the informatics and database support for these functional studies.

We propose the following 10-year goals:

- Establish the International Mouse Mutagenesis Consortium. The IMMC mission is to encourage the exchange of information and resources among participating mutagenesis laboratories and centers, to facilitate the distribution of information and reagents to the wider community, and to demonstrate to funding agencies (public and private) that their local contributions are being leveraged into the larger international effort.
- Conduct a full-genome survey for key traits. The goal is to dig deep into the genome to find as many genes as possible that affect traits such as blood pressure, nociception, learning, sleep, respiration, fertility, energy metabolism, thermal regulation, apoptosis, neural tube development, social behavior, and olfaction. The traits selected for these studies should involve every major developmental and physiological system.
- Develop more precise and efficient phenotyping methods. Success of the IMMC depends on improved phenotyping. Although many of the proposed goals can be achieved with existing methods, improvements are needed to reduce the amount of sample material for testing, to enable continuous monitoring in conscious, unrestrained animals, and to provide high-resolution *in vivo* imaging. Several of these opportunities are discussed in more detail below.
- Establish standard operating procedures. Phenotyping centers will publish on their Web sites the detailed protocols for each assay. These centers also need to coordinate the assays so that similar testing procedures yield similar results in different laboratories. Such standardized protocols are also important for sharing with research laboratories the assay conditions that were used in the original discovery and characterization of each mutant.
- Assign at least one function to as many genes as possible. The goal is to explore the breadth of biological functions. The challenge is to identify assays that will probe the extraordinary diversity of biological

functions. This concerted effort will require ingenuity in devising many new screens for novel traits. These assays, which should be based on local research strengths and interests, will be applied to mutants on many different backgrounds. The need for diverse assays represents an opportunity for large and small research laboratories.

- Determine the chromosomal location of every mutant gene.
- Develop efficient mutation detection methods to identify mutated genes.
- Use “sensitized surveys” to find enhancers, suppressors, and interacting proteins for selected single-gene mutations.
- Establish a network of phenotyping centers with expertise in particular traits and biological functions.
- Develop a centralized mouse phenotype database.
- Improve methods for preserving mutant mice in order to generate a permanent resource.
- Establish a cost-effective infrastructure for preserving and distributing mutant mice.

Resources and Cost Estimates

A network of resource centers is needed to make certain that mutant gametes, embryos, and mice are readily available. The costs for these resource centers are substantial. Investments made today in mouse mutagenesis will be a foundation for future biomedical research.

The development of these comprehensive resources should be guided by principles that assure ready access and the broadest possible distribution. The value of the resources is their fundamental utility to enable subsequent research discoveries. Our goal is to create a “biological” operating system in support of research into mammalian biology. One of the most important principles of the Human Genome Project has been to assure the democratization of access to tools and reagents. The same standards that have been accepted by the community with respect to access to structural genomic reagents such as clones, libraries, genetic markers, maps, and sequences should be applied to the emerging functional genomics resources.

Activities of the IMMC will be coordinated under the auspices of the International Mammalian Genome Society (11). Being an IMMC member commits investigators to the common philosophy of open sharing of information and resources.

The costs include mutant production with chemical mutagenesis, gene-trapping, gene-targeting, phenotype surveys and evaluation, mutant gene mapping and cloning, phenotyping centers, informatics and databases, and preservation and resource centers. The total worldwide annual costs are estimated to be \$200 million.

These estimates are based on the community experience in the various mutagenesis programs that are ongoing in various countries. We note that mutant costs vary considerably, sometimes as much as 20-fold, depending on the nature of the assays used and differences in local expenses.

The following resources and technologies will enable the IMMC to pursue its goals:

Improved methods to map mutant genes. Improvements in technologies for genotyping single nucleotide polymorphisms (SNPs) may yield the necessary cost, scale, and efficiency to map large numbers of mutant genes.

Improved methods to identify mutant genes. Although the mouse genome sequence provides the information and reagents for efficient identification of mutated genes, finding them by means of sequence analysis, especially with analysis of genomic DNA, is expensive. Methods that directly identify mutated genes, through chip-based or molecular technologies, would have great utility.

Improved phenotyping technologies. Progress with DNA chips, microengineered machine systems (MEMs), nanotechnology, and imaging technologies provide enormous opportunities to revolutionize in vitro, ex vivo, and in vivo phenotyping methods by simultaneously increasing throughput and efficiency while reducing costs and the amount of sample material. The recently announced Nanotechnology Initiative (12) and other, similar multidisciplinary initiatives will contribute many of the devices and instrumentation that will be the foundation for the next generation of phenotyping technologies (13).

Improved informatics. A Web site has been established (11) that will act as a clearing-house for information on mutagenesis activities worldwide, including for each group their biological focus; genotype or phenotype screens; database access; and how to obtain mice, sperm, embryos, tissues, or cells. The site also provides a mechanism for individuals and institutions worldwide to join the consortium. There is a need to ensure that all mutants, however generated, are recorded with a unique identifier in a common database. The Mouse Genome Database (MGD) provides the best means for achieving this goal, as each allele is listed as a separate entity with its own accession number (14). IMMC members are committed to entering validated mutants into MGD as soon as they are available for community access. IMMC members are also committed to obtaining an accession number for any new mutation, however generated, at

the time of submission for publication. Mutations will be linked through the IMMC Web site to local sites for detailed phenotypic information and raw data.

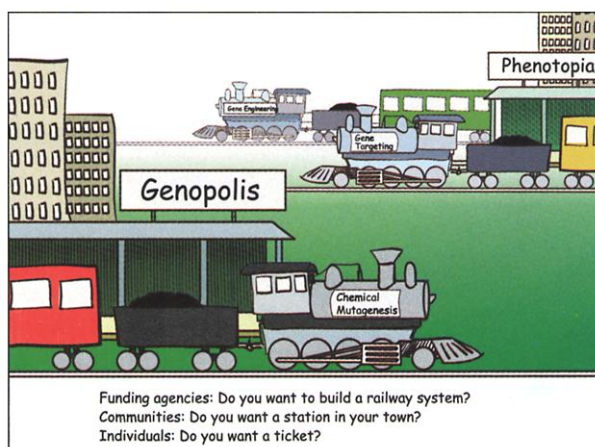
More efficient and reliable methods are needed for archiving, managing, analyzing, displaying, and disseminating the complex phenotype data sets resulting from mutagenesis programs. Unlike DNA sequence and genetic databases, there are no large-scale phenotype databases on which to model the databases for the mutagenesis centers, except perhaps those used to manage information for clinical and epidemiological studies. Several mutagenesis centers (8, 9) are exploring paradigms for these databases. In collaboration with the center activities, MGD will serve as the community resource for integrating and unifying phenotypic information with genetic information about the laboratory mouse. We also need tools to assist in the integration of the relevant

be required. Considerable improvements are obviously needed to identify the mutated genes more efficiently. In particular, we need to be able to map mutant genes to small intervals more efficiently, and we need to be able to identify the molecular lesion in the critical region more quickly.

Functionally characterizing trapped and engineered mutations. An important advantage of gene-based strategies, such as gene-trapping, gene-targeting, and chemical mutagenesis of ES cells in vitro, is that mutations can be obtained in specific genes of interest. The challenge with these gene-based methods is that the phenotypic consequences of the engineered mutations can be difficult to predict or detect. Moreover, the natural world requires individuals to respond, from birth until death, to disease, infection, weather, social interactions, and many other factors, conditions, and processes that are not encountered in the simplified world of a laboratory mouse colony. These responses require coordinated action of diverse developmental, neurological, physiological, metabolic, and immunological systems. Existing technologies and phenotypic assessment paradigms are often inadequate to assess the subtleties of the biological responses to these challenges. Technologies are urgently needed that assess organismal responses to challenges that are normally encountered in the real world and to assess subtle differences in diverse biological functions.

Related Areas of Genetic Research

We recognize that many other important genetic research activities deserve increased research support. These include the genetic and phenotypic analysis of naturally occurring polygenic traits, which is occurring within the Phenome Project (16), as well as studies of quantitative traits (QTLs) with chromosome substitution strains [(17) strains in which single chromosomes are replaced with the corresponding chromosome on a defined and inbred genetic background], deletion strains [(18) strains with deleted chromosome segments], and balancer chromosomes [(19) chromosomes with rearrangements that suppress recombination and with genetic markers so that inheritance can be readily followed]. Variation in naturally occurring traits is an important area, in part because many are models of human diseases. To date, however, finding QTLs has been difficult and additional research support is urgently need to map, clone, and characterize these genes in a more facile, high-throughput manner. In this article, we focused specifically on the challenge of annotating DNA sequences with functions when large numbers of mutants, of many different kinds, are needed. We recognize that many other important genetic research activities deserve increased support, including comparative genomics with other



databases that will allow queries of all aspects of the biology of laboratory mice from DNA sequence to phenotypes. The development of a Gene Ontology (15) will assist the development of the necessary phenotype vocabularies and the implementation of structured and meaningful data sets.

Key Challenges

At least two key challenges must be resolved for the proposed goals to be achieved:

High-throughput DNA-based mutation detection systems. Positional cloning has been a successful approach for identifying mutant genes, and the rate of success is expected to increase dramatically with the availability of finished genome sequences for humans and mice (1-3). However, if a laboratory could positionally clone one gene per week, they would only identify 52 genes after a year of hard work. Even if one gene could be cloned per day, an extraordinary rate with existing technologies, only 365 mutants would be cloned per year. With perhaps 30,000 genes in the mouse genome (3), a network of 100 cloning centers would

model systems, such as the rat. Also required are many other reagents, resources, and technologies, such as full-length cDNAs; validated expression and protein arrays; additional recombination systems; and improved methodologies for tissue-specific expression, overexpression, or inducible expression of gene products.

Conclusions

The availability of the mouse genome sequence and the development of high-throughput, gene-based and phenotype-based mutagenesis paradigms constitute a turning point in biomedical research. We now set challenging goals for the next 10 years. Achieving these goals will require the biomedical research community to improve efficiencies, to reduce costs, and to coordinate international expertise and resources. The impact of these activities will be enormous—deeper insights into functions of genes individually and collectively; fundamental biological and disease processes; and ultimately improved diagnosis, prevention, and treatment of birth defects and adult diseases.

References and Notes

1. J. C. Venter *et al.*, *Science* **291**, 1304 (2001).
2. International Human Genome Sequencing Consortium, *Nature* **409**, 860 (2001).
3. A mouse genome sequence is currently available to subscribers at www.celera.com/products/products.cfm. Information about publicly funded programs sequencing the mouse genome can be found at www.ncbi.nlm.nih.gov/Traces/trace.cgi.
4. J. H. Nadeau *Nature Rev. Genet.*, in press.
5. M. J. Justice, *Nature Rev. Genet.* **1**, 109 (2000).
6. B. Zheng *et al.*, *Mol. Cell Biol.* **20**, 648 (2000).
7. D. Metzger, R. Feil, *Curr. Opin. Biotechnol.* **10**, 470 (1999).
8. M. H. Hrabe de Angelis *et al.*, *Nature Genet.* **25**, 444 (2000).
9. P. M. Nolan *et al.*, *Nature Genet.* **25**, 440 (2000).
10. M. V. Wiles *et al.*, *Nature Genet.* **24**, 13 (2000).
11. www.imgs.org
12. www.nano.gov
13. www.nih.gov/science/models/mouse
14. www.informatics.jax.org/mgihome/nomen/allmut_form.shtml
15. M. Ashburner *et al.*, *Nature Genet.* **25**, 25 (2000).
16. K. P. Paigen, J. T. Eppig, *Mamm. Genome* **11**, 715 (2000).
17. J. H. Nadeau *et al.*, *Nature Genet.* **24**, 221 (2000).
18. H. Su *et al.*, *Nature Genet.* **24**, 92 (2000).
19. B. Zheng *et al.*, *Nature Genet.* **22**, 375 (1999).
20. We thank G. Duyk and K. Moore for their many thoughtful insights and suggestions during the preparation of this plan.
21. Members of the IMMC: R. Balling (President, International Mammalian Genome Society), German Center for Biotechnology, D-38124 Braunschweig, Germany. G. Barsh, Departments of Pediatrics and Genetics, Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. D. Beier, Genetics Division, Brigham and Women's Hospital, Boston, MA 02115, USA. S. D. M. Brown, MRC Mammalian Genetics Unit, UK Mouse Genome Centre, Harwell, OX11 0RD, UK. M. Bucan, Center for Neurobiology and Behavior, Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104, USA. S. Camper (Secretariat, International Mammalian Genome Society), Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA. G. Carlson, McLaughlin Research Institute, Great Falls, MN 59405, USA. N. Copeland, National Cancer Institute—Frederick, Frederick, MD 21702, USA. J. Eppig, Jackson Laboratory, Bar Harbor, ME 04609, USA. C. Fletcher, The Genomics Institute of the Novartis Research Foundation, San Diego, CA 92121, USA. W. N. Frankel, Jackson Laboratory, Bar Harbor, ME 04609, USA. D. Ganten, Max Delbrück Center for Molecular Medicine, 13092 Berlin-Buch, Germany. D. Goldowitz, Department of Anatomy and Neurobiology, University of Tennessee, Memphis, TN 38163, USA. C. Goodnow, Medical Genome Centre, John Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia. J.-L. Guenet (Secretariat, International Mammalian Genome Society), Unité de Génétique des Mammifères, Institut Pasteur, 75724 Paris Cedex 15, Paris, France. G. Hicks, MICB Center for Mammalian Functional Genomics, Manitoba Institute of Cell Biology, University of Manitoba, Winnipeg, Manitoba R3F 0V9, Canada. M. Hrabe de Angelis, Genome Project Group, GSF National Research Center for Environment and Health Institute of Experimental Genetics, D-85764 Neuherberg, Germany. I. Jackson (Vice President, International Mammalian Genome Society), MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, Scotland. H. J. Jacob, Human and Molecular Genetics Center, Milwaukee, WI 53226, USA. N. Jenkins, National Cancer Institute—Frederick, Frederick, MD 21702, USA. D. Johnson, Mammalian Genetics and Genomics, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. M. Justice (Secretariat, International Mammalian Genome Society), Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. S. Kay, Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. D. Kingsley, Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA. H. Lehrach, Max-Planck Institute of Molecular Genetics—Berlin, D-14195 Berlin, Germany. T. Magnuson (Secretariat, International Mammalian Genome Society), Department of Genetics, University of North Carolina—Chapel Hill, Chapel Hill, NC 27599, USA. M. Meisler (Past President, International Mammalian Genome Society), Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA. J. H. Nadeau (corresponding author), A. Poustka, Division of Molecular Genome Analysis, German Cancer Research Center, 69120 Heidelberg, Germany. E. M. Rinchik, Mouse Genetics and Mutagenesis, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. J. Rossant, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. L. B. Russell, Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. J. Schimenti, The Jackson Laboratory, Bar Harbor, ME 04609, USA. T. Shiroishi (Secretariat, International Mammalian Genome Society), Mammalian Genetics Laboratory, National Institute of Genetics, Mishima, Shizuoka-ken 411-8540, Japan. W. C. Skarnes, Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA 94720, USA. P. Soriano, Program in Developmental Biology, Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. W. Stanford, Lunenfeld Gene Trap Laboratory, Program in Development and Fetal Health, Samuel Lunenfeld Research Institute, Toronto, Ontario M5G 1X5, Canada. J. S. Takahashi, Howard Hughes Medical Institute, Northwestern University, Evanston, IL 60208, USA. W. Wurst, Molecular Neurogenetics, Max-Planck Institute of Psychiatry, 80804 Munich, Germany. A. Zimmer, Department of Psychiatry, University of Bonn, Bonn 53113, Germany.

FUTURE DIRECTIONS: GENE NUMBER

What If There Are Only 30,000 Human Genes?

Jean-Michel Claverie

The confirmation that there might be fewer than 30,000 protein-coding genes in the human genome is one of the key results of the monumental work presented in this issue of *Science* by Venter *et al.* (1). That a mere one-third increase in gene numbers could be enough to progress from a rather unsophisticated nematode [*Caenorhabditis elegans*, with about 20,000 genes (2)] to humans (and other mammals) is certainly quite provocative and will undoubtedly trigger scientific, philosophical, ethical, and religious questions throughout the beginning of this new century. By the same token, humans appear only five times as complex as a bacterium like *Pseudomonas aeruginosa* (3). Although a significant uncertainty is still attached to this low number (see below), it was not totally unexpected, after the downward trend initiated by the analysis of the first two complete human chromosomes (4, 5), as well as two independent statistical studies (6, 7), and the unexpectedly low

(14,000) *Drosophila* gene number (8).

After the older *C* value paradox (9), we now have an apparent *N* value paradox on our hands: Neither the cellular DNA content (in mass) nor its gene content appears directly related to our intuitive perception of organismal complexity. However, logic taught us that paradoxes often arise from the use of imprecise or ambiguous terminology. In a quick (admittedly nonrepresentative) survey among people in my laboratory, the answers to the question: "How much more complex is a human compared to a nematode?" ranged from a mere 100 to near infinity. Those widely different opinions were mostly the result of the lack of an objective (physical) measurement of what we mean by "biological complexity." Some only considered the diversity of cell types, others considered brain circuitry, and others went as far as including the cultural achievements of the human species as a whole. Thus, 30,000 human genes is not equally surprising to everybody.

Furthermore, any personal estimate of biological complexity *K* can be fitted to the gene number *N*, by arbitrarily choosing a suitable functional relationship $K = f(N)$: proportional: $K \sim N$, polynomial: $K \sim N^p$, exponential: $K \sim a^N$, or even factorial: $K \sim$

Structural & Genetic Information Laboratory, CNRS-AVENTIS UMR 1889 31 Chemin Joseph Aiguier, 13402, Marseille, France. E-mail: Jean-Michel.Claverie@igs.cnrs-mrs.fr