Translational genomics in medical genetics

Edward R.B. McCabe, MD, PhD

I want to thank you very much for giving me the honor of serving you as the President of the American College of Medical Genetics (ACMG) during 2001 and 2002. I consider myself extremely privileged for you to have placed your trust in me.

The ACMG is the organization within clinical genetics that is responsible for education and policy. Last year in my presidential address I argued that we blend exacting science with compassion and advocacy in the care of our patients, and we must work individually and collectively to ensure full and equal access to genetic information and genetic health professionals.[1] This year I will discuss the critical need for medical geneticists to be the leaders in translating genomic information for practical medical use.

The premise underlying my presentation is that for the Human Genome Project to have its fullest impact, its products in the form of sequence and concepts must be translated for use in clinical medicine. Translational genomics will be the result of this endeavor, and medical geneticists should be the leaders in this arena.

The success of translational genomics will depend on our ability to discern the complexity of interactions among genes, gene products, and the environment. A full understanding of this complexity will depend upon new ways of considering genetics in the genomic era. This will require an intermingling of concepts from biology, engineering, and information systems, and comparisons within and between genomes, transcriptomes, proteomes, and metabolomes.

We are already seeing a shifting focus in medical genetics from rare to common diseases. I will argue, however, that translational genomics teaches us that the differences between rare and common diseases are not so extreme, because in fact there are a multitude of genotypes for "common disease" phenotypes. If we are to improve significantly the care of our patients with rare and common genetic diseases, then we must develop collaborative, multi-institutional, protocol-driven clinical studies.

Human Genome Project and translational genomics

Genetics has come a long way in less than 50 years, from the original identification of the double helical structure of DNA by Watson and Crick[2] in April 1953, to the drafts of the human genome published by the public[3] and private[4] efforts in February 2001.

Initiated in 1990, the completion of the public Human Genome Project was originally planned for 2005. However, the finished sequence is anticipated for spring 2003, to commemorate the 50th anniversary of the Watson and Crick publication.[2]

The results of the Human Genome Project will involve not only the complete sequencing of the human genome, but also the creation of a new branch of science and medicine, genomics, as well as a series of derivative disciplines. If a genome is all of the DNA for an organism, then genomics is the study of that genome and its implications. Derivative disciplines include transcriptomics, the study of the transcriptome, representing all of the transcripts or RNA copies of the genes in a cell, tissue, or individual; proteomics, the study of the proteome, representing all of the proteins in a cell, tissue, or individual; and metabolomics, the study of the metabolome, representing all of the molecular components of a cell, tissue, or individual that are produced by the proteins of the proteome. Additional derivative disciplines are being and will continue to be described.

Genomics, its derivative disciplines, and the synthesis of this information will generate the complete parts' lists and the parts' assembly instructions for a fully functioning organism.

The promise of the Human Genome Project is improved diagnosis and treatment through the application of genomic information and technologies, leading to predictive medicine and individualized medical care.[5] Genomic medicine will be predictive rather than reactive. It will be preventive in its philosophical orientation, rather than the more traditional medical approach of responding only after acute presentation. The predictive and preventive nature of genomic medicine will lead to screening of populations, subpopulations, and individuals for genetic predispositions to rare and common disorders.

Medical geneticists should be the translators for genomic medicine. The ACMG, with its roles in education and policy formation, gives our membership the venue in which to organize these translational efforts. We must prepare ourselves for our roles in translational genomics by addressing the complexity of biological systems.

Complexity of interactions between genes, gene products, and environment

The disciplines of genomics, transcriptomics, proteomics, and metabolomics are already changing our concepts regarding the organization of biological systems. Many of us thought of biochemical pathways as an interconnection of enzymes and

metabolites in a relatively homogeneously connected network. We have learned that such systems are referred to as exponential networks, in which the probability of a node's having increasing connectivity falls as an exponential function.[6,7] Robust biological and human-designed systems have the hub-and-spoke architecture of scale-free networks. These scale-free networks are more heterogeneous in their connectivity, with most nodes having one or a very few contacts, and only rare nodes being highly connected. Examples of scale-free networks include the Internet[6] and natural proteomes.[8]

Scale-free networks are highly robust, meaning that it is difficult to fragment such networks.[7] Communications between nodes are unaffected by very high failure rates, since most of the nodes have very low connectivity. If a node is lost, then there is a very high probability that this node will have been connected to only one other node, and while this connection will fail, there is a very low probability that the overall function of the network will be compromised.

The structural features that give scale-free networks their robust properties, however, are also the source of their vulnerability. The highly connected nodes, although quite rare, are critical to network function. Failure of these highly connected nodes, such as by a mutation in a biological scale-free network, will fragment the network and threaten the survival of the network. Considerations of the error and attack tolerance, as well as the vulnerabilities of scale-free networks, are essential to our understanding of the pathogenesis of genetic diseases.[7,9] Such considerations require a global view of an organism's functional components and their integration, informed, for example, by a knowledge of that organism's genome, transcriptome, proteome, and metabolome.

Intermingling of biology, engineering, and information systems

Our thinking about the complexity of cellular and organismic biology will require us to move from molecular to modular concepts.[10] Within this modular view of biology, the components will have properties that we will recognize to be similar to those in engineering and computer science. Recent evidence has shown that a biological system may be constructed that will function as an oscillator[11] or a toggle switch.[12] This modular view of biology will require the formulation of new models and general principles if we are to achieve an improved ability to appreciate and manipulate these functional modules.[10]

The Human Genome Project is also bringing about major changes in bioinformatics. Many think that, while the sequencing efforts represent a major scientific accomplishment, the real "genomics revolution" is in the ability to manage and analyze the information contained within the sequence. The ongoing efforts in sequencing and informatics will permit translation for improved medical care, although there is a substantial amount of work to be done before practicable translation is available to individuals.

If the sequencing of an individual's genome would be available for $1,000, then let us consider some of the issues that would accompany the translation of this information for medical decision-making by that individual. For any data acquisition there is inevitably an error rate, and this will be true for genomic sequence. Even if the error rate is quite low, the size of the genome is such that numerous errors are anticipated until the cost is so low that multiple repetitions are possible. For example, if sequencing is 99.99% accurate, this would result in an error every 10,000 nucleotides, or with a human genome of approximately 3.2 billion base pairs the number of errors would exceed 300,000. Interpretation of sequence variation between individuals would have to consider differences due to technical errors as well as natural polymorphisms. In addition, interpretation of an individual's genomic sequence will require a better understanding of complexity in biology, from the multiple functions of individual proteins to a modular view of biological networks. Beyond the technical and biological considerations in the use of the genomic sequence information are ethical, legal, and social issues. The cost of $1,000 that we have set arbitrarily would be affordable for some and absolutely inaccessible for others. For those who could afford to have their genomes sequenced, there would be the issues of privacy and discrimination, and how this information could be stored in a manner that would ensure fidelity and accessibility without compromising insurability.[1]

Comparisons within and between genomes, transcriptomes, proteomes, and metabolomes

Comparisons of sequences within and between organisms may be quite powerful. Such comparisons are quite common and useful for DNA coding sequences and deduced amino acid sequences, for which similarities and differences define and distinguish, for example, functional domains and gene families. Comparisons among noncoding sequences may identify and distinguish regulatory elements. Other conserved blocks of noncoding sequences are being recognized, but for many of these conserved regions their functions remain to be identified. A fuller understanding of function and interrelationships throughout the genome, transcriptome, and proteome and their impact on the metabolome will be required for interpretation of sequence differences within an individual's genome.

Functional comparisons must be made at all levels, but at this time are focusing on the proteome and metabolome. Studies in microorganisms can give us insight into principles underlying the robustness of all biological systems. Growth of *Escherichia coli* on glucose was analyzed to determine which enzymes contributed most strongly to the robust properties of this metabolic system.[13] Flux limits for enzymatic reactions were examined below which growth on glucose was compromised, and these enzymes were classified into three groups: pentose phosphate pathway, three-carbon glycolytic pathway, and tricarboxylic acid cycle. For the pentose phosphate pathway and tricarboxylic acid cycle, growth was not compromised until residual flux fell to ≈15–30% of normal. For the three-carbon glycolytic pathway, however, growth was compromised when flux fell to only 63% of normal, indicating that this region of metabolism was much more sensitive to variation. Therefore, while growth of *E. coli* on glucose is in general a

highly robust property, not all components are equally resistant to perturbation.

Comparisons of the large-scale organization of metabolic networks between organisms also may give insight into the robustness of biological systems. The metabolic networks of 43 organisms were compared quantitatively for 43 organisms representative of archae, bacteria, and eukaryotes.[14] Significant differences in the pathways were observed among these organisms, but the overall design properties were quite similar. The investigators speculated that the scale-free topological organization was the common design-element in these organisms and that this organization contributed to the common feature of robust tolerance to variation among these very different organisms.

DAX1 *deficiency and perturbation in a complex network*

Dosage-sensitive sex reversal, adrenal hypoplasia congenta, on the X chromosome, gene 1 (*DAX1*), is the nuclear receptor family member encoded by the *NROB1* gene.[15] Loss of *DAX1* function by deletion or point mutation causes disruption of the normal development of the steroidogenic axis leading to adrenal hypoplasia congenita associated with hypogonadotropic hypogonadism. The key question is how the loss of a single gene product results in the compromise of a robust developmental system.

Traditionally we have thought of blocks in metabolic pathways as analogous to the damming of a flow of water with a buildup behind the dam and overflow into side channels, and of blocks in transcriptional cascades as interruption of signal transfer. Consideration of the complexity of developmental networks, however, gives us a different view of the effects of mutations. *DAX1* is a node in a complex network with a high level of connectivity.[16] Loss of *DAX1* function would not simply compromise a one-dimensional transcriptional cascade, but we would speculate that mutation of *DAX1* would compromise the function of each of the transcription factors distal to it in the network, and therefore an entire sector of the transcriptional network would be lost.[16]

Shifting focus from rare to common diseases

Some consider the shift in focus within medical genetics from rare, Mendelian diseases to common, complex disorders as a paradigm shift. We must consider, however, that rare and common disorders, or "single" gene and multigenic diseases, respectively, represent points on a continuum and not distinct entities.[9,17–19]

Multitude of genotypes even for common disease phenotypes

Complexity is the rule for the phenotypes of "simple" Mendelian traits, as well as for common, multigenic disorders.[9,17–20] In a Mendelian or "single" gene disorder, there is one gene that exerts a primary effect, but the phenotype in the individual patient is influenced by additional modifier genes.[17–19] As a single gene loses its primary effect and the influences of two or more genes begin to be approximately co-equal, the disorder is recognized as multigenic. Such multi-

genic diseases may also be referred to as "complex" diseases, but this terminology denies the fundamental complexity of the "single" gene diseases. The major portion of the common disorders, such as cancer, cardiovascular disease, and diabetes mellitus, are presumed to be multigenic with superimposition of environmental influences.

Since the phenotypes of the "simple" Mendelian disorders represent the influences of the primary genetic abnormality and additional modifier genes, and those for multigenic disorders clearly involve the influences of multiple genes, and both have additional environmental influences, therefore the phenotypes of individuals in all categories of genetic disorders will be determined by the individuals' extended genotypes. In other words, each individual, regardless of the type of genetic disease they experience, are influenced by their unique genotype and environmental experience.

For Mendelian disorders, although there may be a group of modifier genes that may influence the phenotype, not every modifier gene will be involved in each patient. Similarly, for the multigenic disorders, if there are large groups of genes that may be involved in the pathogenesis, only a subset of these will be involved in any individual patient. Therefore, any genetic disease will be composed of individuals with rare composite genotypes, and we can develop similar approaches for the acquisition of an evidence base across all genetic disorders.

Development of collaborative, multi-institutional, protocol-driven clinical studies

The use of clinical studies involving collaborations between multiple institutions with strict adherence to protocols developed by consensus among the participating investigators represents the approach that will optimize progress toward a sound evidence base when individual patients are rare. We have determined that even for common disorders, those with an identical composite genotype will be rare; therefore, such an approach will be just as valid for rare disorders as for common diseases.

One example from genetics is the multi-institutional, randomized, double-blind, placebo-controlled investigation of the value of penicillin prophylaxis in patients with sickle cell disease.[21] Accumulation of a significant number of patients with sickle cell disease from a number of centers in a carefully controlled, protocol-driven study permitted the very clear conclusion to be drawn that penicillin did benefit patients with sickle cell disease. In fact, the statistical power of the study was so strong that the study was terminated earlier than anticipated when the data indicated that the patients on placebo were at risk for sepsis and death far in excess of those on penicillin prophylaxis. The consequence of this well-controlled study was the conclusion by a National Institutes of Health Consensus Development Conference that all neonates should be screened for sickle cell disease in order to identify those with this disorder and initiate penicillin prophylaxis to prevent the morbidity and mortality associated with clinical presentation in the absence of a diagnosis.[22]

Perhaps the most successful example of a multi-institutional, protocol-driven clinical collaboration is the Children's Oncology Group (COG). It is estimated that at least 85% of children with cancer are enrolled in COG protocols, and if those involved with natural history studies are included, the enrollment may be as high as 95%. The involvement of such a high proportion of children with cancer in clinical trials by COG and its predecessors permits an iterative approach to improvements in interventions and has been credited with the remarkable success that has been achieved in pediatric cancer outcomes.

The highly individualistic approach that characterizes much of American medicine, including medical genetics, provides the practitioner with management autonomy. If we argue that pooling of patients into collaborative, protocol-driven clinical studies will facilitate progress, then the conclusion would be that the lack of collaboration by autonomous physicians delays progress unnecessarily. To be successful, an organized national approach to develop collaborative studies in medical genetics will require "buy-in" by the clinical genetics community (with associated loss of practitioner autonomy) and adequate resources to support the protocols (including patient enrollment and support, data analyses, and iterative interventional changes).

Summary and conclusions

The need for translational research has received considerable recent attention. Medical geneticists and the ACMG have the opportunity to establish a model for translational studies for the medical community through collaborative natural history studies and interventional clinical trials. Medical genetics will become the clinical embodiment of the Human Genome Project if we are successful in transliterating genomic sequences and concepts into the language of medicine. Optimal care for our patients and those of our colleagues in other specialties of medicine will require that medical geneticists continue to accept the responsibility for translational genomics.

References

1. McCabe ERB. Clinical genetics: compassion, access, science, and advocacy. *Genet Med* 2001;3:426–429.
2. Watson JD. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 1953;171:737–738.
3. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304–1351.
5. Collins F. Shattuck lecture: medical and societal consequences of the Human Genome Project. *N Engl J Med* 2000;341:28–37.
6. Tu Y. How robust is the Internet? *Nature* 2000;406:353–354.
7. Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. *Nature* 2000;406:378–382.
8. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO. Protein function in the post-genomic era. *Nature* 2000;405:823–826.
9. Dipple KM, Phelan JK, McCabe ERB. Consequences of complexity within biological networks: robustness and health, or vulnerability and disease. *Mol Genet Metab* 2001;74:45–50.
10. Hartwell L, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature* 1999;402:C47–C52.
11. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature* 2000;403:335–338.
12. Judd EM, Laub MT, McAdams HH. Toggles and oscillators: new genetic circuit designs. *Bioessays* 2000;22:507–509.
13. Edwards JS, Palsson BO. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol Prog* 2000;16:927–939.
14. Jeong H, Tombor B, Albert R, Oltval ZN, Barabasi AL. The large-scale organization of metabolic networks. *Nature* 2000;407:651–654.
15. McCabe ERB. Adrenal hypoplasia and aplasia. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Vogelstein B, editors. The metabolic and molecular basis of inherited diseases, 8th ed. New York: McGraw-Hill, 2001:4263–4274.
16. Clipsham R, Zhang Y-H, Huang B-L, McCabe ERB. Genetic network identification by high density, multiplexed reversed transcriptional (HD-MRT) analysis in steroidogenic axis model cell lines. *Mol Genet Metab* 2002;77:159–178.
17. Dipple KM, McCabe ERB. Phenotypes of patients with "simple" Mendelian disorders are complex traits: thresholds, modifiers and systems dynamics. *Am J Hum Genet* 2000;66:1729–1735.
18. Dipple KM, McCabe ERB. Modifier genes convert "simple" Mendelian disorders to complex traits. *Mol Genet Metab* 2000;71:43–50.
19. Beaudet A, Scriver C, Sly W, Valle D. Genetics, biochemistry and molecular bases of variant human phenotypes. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Vogelstein B, editors. The metabolic and molecular basis of inherited diseases, 8th ed. New York: McGraw-Hill, 2001:3–125.
20. Scriver CR, Waters PJ. Monogenic traits are not simple: lessons from phenylketonuria. *Trends Genet* 1999;15:267–272.
21. Gaston MH, Verter JI, Woods G, Pegelow C, Kelleher J, Presbury G, Zarkowsky H, Vichinsky E, Iyer R, Lobel JS, Diamond S, Holbrook CT, Gill FM, Ritchey K, Fallette JM. Prophylaxis with oral penicillin in children with sickle cell anemia: a randomized trial. *N Engl J Med* 1986;314:1593–1599.
22. Wethers D. NIH Consensus Development Conference Statement on newborn screening for sickle cell disease and other hemoglobinopathies. 1987;6.