A Roadmap to functional genomics

Raymond B. Penn, Victor E. Ortega, and Eugene R. Bleecker

Department of Internal Medicine and the Center for Human Genomics, Wake Forest University Health Sciences, Winston-Salem, North Carolina

Submitted 10 January 2007; accepted in final form 21 February 2007

Penn RB, Ortega VE, Bleecker ER. A Roadmap to functional genomics. *Physiol Genomics* 30: 82–88, 2007. First published February 27, 2007; doi:10.1152/physiolgenomics.00010.2007.—In August 2006, the Center for Human Genomics of the Wake Forest University School of Medicine in Winston-Salem, NC, hosted the National Institutes of Health-sponsored Roadmap Course entitled *Models and Technologies for Defining Phenotype*. Twenty-four biomedical and genomic researchers from throughout the world and with varying degrees of experience in the genomics, biological, and biomedical engineering sciences were invited to participate as students in a comprehensive course dedicated to presenting and evaluating current and future approaches that can overcome the problems experienced to date in characterizing the functional consequences of gene variation. A total of 34 senior researchers from four different academic institutions served as course faculty and employed a pedagogical approach that emphasized hands-on workshops, demonstrations, and small group discussions and tasks. Through this report we convey the complex and formidable problems unique to genomics research as we attempt to link the field of genomic research to complex human diseases. Furthermore, we describe the logic and organization of a Roadmap Course designed to teach a diverse group of researchers a multi-disciplinary approach to addressing complex biomedical scenarios in the field of human genomics.

translational research; multidisciplinary; imaging sciences; NIH Roadmap

JULY 31–AUGUST 11, 2006, the Center for Human Genomics of the Wake Forest University School of Medicine in Winston-Salem, NC, hosted the National Institutes of Health (NIH)-sponsored “Roadmap Course” entitled “Models and Technologies for Defining Phenotype.” Twenty-four biomedical and genomic researchers from throughout the world and with varying degrees of experience in the genomics, biological, and biomedical engineering sciences were invited to participate as “students” in a comprehensive course dedicated to presenting and evaluating current and future approaches that can overcome the problems experienced to date in characterizing the functional consequences of gene variation. A total of 34 senior researchers from four different academic institutions served as course faculty and employed a pedagogical approach that emphasized hands-on workshops, demonstrations, and small group discussions and tasks.

Through this report we convey the complex and formidable problems unique to genomics research as we attempt to link the field of genomic research to complex human diseases. Furthermore, we describe the logic and organization of a Roadmap Course designed to teach a diverse group of researchers a multidisciplinary approach to addressing complex biomedical scenarios in the field of human genomics. Finally, we describe the utility of a grant-writing simulation that provides participants an opportunity to apply the methods learned to detail a comprehensive genomics research project replete with strategies for understanding the biological significance of gene variations linked to complex diseases.

WHY THIS COURSE WAS HELD: FUNCTIONAL GENOMICS

Genomics research has had tremendous success in identifying associations between gene polymorphisms and specific common diseases. Enthusiasm for most of these findings has, however, been tempered by not fully understanding the biological mechanisms through which gene variations influence disease development and severity. Although some of these associations are almost certainly spurious, many are likely genuine and occur as a result of novel and poorly understood processes. The current challenge for genomics research lies in validation of findings through application of clinical and basic science approaches to assess the biological relevance of gene variation. What is the significance of gene variants with respect to altered function of the gene product at the level of the molecule, the cell, the tissue, the organ system, and ultimately, the organism? Defining these characteristics or “phenotypes” has emerged as an important, albeit difficult and less well-developed component of genomics research. Accordingly, genomics research appears poised to evolve into a multidisciplinary or hybrid field by integrating various basic, clinical, and engineering sciences for the purpose of defining phenotype-gene interactions, i.e., pursuing “functional genomics.”

How exactly this evolution will come about is unclear. Currently, the instances in which successful fusion of genomics and biomedical research occurs seem largely dependent on the collaborative efforts of experienced investigators with the ability to integrate different scientific disciplines. Preferably there would be more broad-based participation in genomics research, in which large numbers of scientists sufficiently conversant in disciplines outside their expertise could collaborate to provide a strong functional genomics component to studies of gene-disease associations. Providing a formal education process to recruit greater numbers and depth of expertise to genomics research seems the more practical approach to the problems inherent in functional genomics.

FURTHER DEFINING THE PROBLEM: THE UNIQUE CHALLENGE OF FUNCTIONAL GENOMICS

Disease can be viewed as a complex process involving dysregulated organism function. A gene encodes an elemental building block of the organism. How gene variation contributes to a complex disease process depends on the impact of single
nucleotide polymorphisms (SNPs) or haplotypes on multiple functions in a spectrum ranging from the most reductionist to the most integrative. Molecular and intracellular events affect specific cell functions, which in turn affect tissue functions that impact one or more organ systems. Physiological processes, mechanisms of the human body responsible for the origin, development, and progression of life, are dependent on integration of organ system functions, fail to some degree, and cannot be adequately maintained by normal homeostatic control mechanisms. Appreciation of the effects at the level of the cell, tissue, organ system, and organ system integration is necessary to definitively link the effect of altered DNA sequence to altered physiology. Characterizing the effect of gene variation at any given level can provide clues to the mechanism by which physiology is affected, but connecting the dots from the initial effect (e.g., altered mRNA or protein expression/function) to the integrated endpoint (disease) is required to provide a full mechanistic explanation. In complex diseases, the contributions of numerous regulatory elements, which may be sufficient, required, or redundant, help determine whether variation in a specific gene is of any consequence. Addressing the issue of biological significance of a gene variation requires a spectrum of approaches, each with limitations and strengths (Fig. 1).

An important consideration when choosing analytic approaches for examining the consequence of gene variation is the trade-off between experimental control (the ability to directly control for a finite number of experimental parameters) and physiological relevance (the extent to which outcomes determined reflect the in vivo condition). Toward the more integrative end of the spectrum, approaches measure outcomes that are closely linked to the disease process and are likely to reflect the sum effect of gene variation on multiple, more elemental processes. Moreover, the outcomes occur in a genetic and biologic background that may be necessary to support the effect of the gene variation of interest and its causal relationship to disease (in this instance the variability helps promote the outcome but confounds interpretation). In addition, the specific data generated can often be associated with similar integrative data obtained from the same subject. The obvious limitation is the inability to control for numerous unknown biologic, genetic, and interactive gene-environment factors that contribute to individual variation. Thus, it is often unclear if the SNP of interest is necessary or sufficient to cause the change in outcome observed; it may simply be in linkage disequilibrium with another variation(s). A countervailing strength is the ability to observe effect sizes that are biomedically relevant, but of modest magnitude consistent with disorders of complex etiology in studies with large numbers of observations, e.g., subjects from an epidemiological cohort, that are both biomedically relevant.

Conversely, more reductionist approaches afford greater experimental control, and many experimental conditions can be matched between groups while specific experimental variables can be isolated and tested more readily. When the genetic and biologic background remain fixed (e.g., in artificial cell-based systems in which only the gene of interest is varied), it is easier to appreciate the sufficiency of a gene variation in causing the (reductionist) outcome. Yet the observed effects may still be dependent on the more comprehensive genetic context and may in fact be absent in similar experimental models, demonstrating that the variation tested can contribute but is not sufficient. Choice of species and cell model may be critical, and fishing among models can be time consuming, complex, and expensive. Equally, effects that are biomedically relevant when active over many years, such as the simple

![Fig. 1. Experimental approaches in functional genomics.](http://physiolgenomics.physiology.org/)
impact of aging on many disease processes, may not be large enough to measure confidently and reproducibly in model systems. Lastly, it is important to recognize that because reductionist outcomes feed into more complex processes with multiple layers of regulatory control, demonstration that the effect on the elementary biological/physiological process truly impacts the disease process requires further inquiry.

Thus given the limitations of various reductionist and integrative approaches to defining phenotype, multiple approaches appear necessary to get a true handle on the mechanisms by which gene variations impact complex diseases. Presently, the more integrative approaches for assessing phenotype are more frequently embraced by genomics, perhaps reflecting the inherent difficulty of appropriately designing, executing, and funding some of the more reductionist approaches, and the ability to make simple transitions of studies. For example identification of a diabetes-associated polymorphism lends itself quickly to assessing physiological parameters such as insulin resistance but is challenging to explore in appropriately designed experiments in tissue or cell-based models. In addition to the advantages of integrative approaches noted above, other features of physiological measures make them well suited for supporting genomics studies. Often they can be performed in a high(er) throughput manner and generate a larger number of observations, as in the case of collection of various biomarkers from blood, urine, or sputum, or with noninvasive imaging technologies. The data collected are often part of the basic clinical profile that helps to characterize the subject with/without disease or assess disease severity. Despite these advantages integrative data help to complete only part of the puzzle. To obtain a clear understanding of the causal effects of gene variation, it will be necessary to complement these studies with more reductionist approaches.

The other important concept is that genomic approaches in complex diseases (cardiovascular, metabolic, neurological, oncological, respiratory, etc.) are based on understanding the primary and associated phenotypes that are to be investigated. Limiting these studies to a single phenotype often leads to failure with a high cost of investment in these expensive and complicated genetic approaches.

Thus, the identification of intermediate phenotypes represents a key component of these genetic approaches. Intermediate or secondary phenotypes include risk factors and closely associated traits that are related to disease development or severity. Ideally, phenotypic evaluation should include molecular, cellular, and physiological approaches as well as the use of biomarkers that are used to characterize the trait or disorder under investigation. Moreover, the basic primary and secondary phenotypes used in genomics studies should be selected so that replication of findings is facilitated among different studies and population samples. Unfortunately, current training in molecular and analytic genomics often does not adequately address the important areas of phenotypic characterization using the type of rigorous standardization that is necessary for these studies. There have been major advances in molecular, cellular, proteomic, physiologic, and radiographic approaches that are figuratively begging for a mode of expression/application, primarily because of the communication barriers that exist among the engineering, clinical/basic science, and genomics worlds.
lyze and demonstrate specific techniques appropriate for characterizing gene variation and its biologic significance. Emphasis was placed on those techniques employed by analytic and molecular genetics and those of the physical and biomedical engineering sciences.

During the morning sessions of the first week of the course, each group undertook coursework in a “foreign” discipline (Table 1). For the group comprising biological or engineering scientists, a genomics curriculum was followed. For the group comprising genomics scientists, a curriculum in imaging technologies was followed. During the afternoon sessions, coursework focused on research design issues for integration of analytic and genomics and functional genomics, examples of how genomics studies of various complex diseases have attempted phenotype characterization, and current and emerging methodologies in the biological and physical sciences capable of serving functional genomics.

During the second week of the course, all participants followed a common curriculum that continued emphasizing applicable methodologies in the biological and physical sciences, clinical research design issues, and ethical and human subjects issues specific to genomics research. A major focus of the second week was the organization of small, multidisciplinary groups for the purpose of writing an NIH-style grant (in outline form) on a gene variation(s) in which the function genomics component was addressed by incorporating one or more of the various methodologies discussed during the course. Participants were given the option of working independently or preferentially forming multidisciplinary groups to take advantage of collective backgrounds. The last 3 days of the course were dedicated to grant-writing workshops in which course faculty were available for consultation. On the final day grant proposals were presented through PowerPoint presentations to all participants and faculty who, in turn, provided a peer review.

**COURSE CURRICULUM**

Coursework was presented in the form of lectures, workshops, group discussions, and demonstrations. Overview lectures for each topic where followed by demonstrations and workshops in which techniques were presented in greater detail, and when possible, hands-on data analysis was performed. For each methodology and technique discussed, examples of their application to specific genomics studies were discussed. Many of the lectures were given by local or visiting senior faculty whose area of expertise was the topic of discussion (see Table 2).

**Genomics curriculum.** Conveying the fundamentals of genomics research within a short time frame was particularly challenging for those tracked in the genomics curriculum. The genomics track participants were investigators in the biological and engineering sciences with little or no familiarity with the genomics. Moreover, the evaluation of gene variation has become increasingly complex through the recent development of newer technologies able to facilitate the measurement of multiple variations in genes rapidly and accurately in large population samples, as well as the emerging availability of genome-wide chip analyses. These tools permit investigation of gene-gene interactions (10) and the necessary analysis of haplotypes across a gene to best understand the phenotypic effects of gene variation (8–10, 14, 16). Furthermore, analytic approaches to evaluate multilocus gene interactions (17), haplotype analysis, and genome-wide high-density SNP analysis are constantly evolving and represent important didactic areas for molecular and cell biologists, imaging scientists, and scientists who study animal and human models.

The goal of the genomics curriculum was to render participants sufficiently conversant in the approaches, terminology, and interpretation of data of genomics studies and characterize existing limitations in genomics studies that can be served by the biological and engineering sciences. Coursework focused on the most commonly used approaches for determining gene variations associated with disease, understanding the logic and key elements of study designs, and multiple workshops in which data were analyzed and interpreted. Key concepts were reinforced in the form of group discussions and a journal club format through the presentation of specific, relevant studies employing each of the approaches.

**Imaging sciences curriculum.** Imaging techniques provide numerous advantages and the potential for future approaches to define novel phenotypes for genomics investigators. These tools should facilitate definition of phenotype and overcome many of the limitations inherent in the traditional biochemistry/cell biology strategies. With the advent of modern imaging, many of the biological phenomena previously examined through indirect methods are now directly visualized in cells,

**Table 1. Coursework**

<table>
<thead>
<tr>
<th>Genomics (Track 1)</th>
<th>Imaging Sciences (Track 2)</th>
<th>Biological/Physical Science (Track 1 and 2)</th>
<th>Synthesis/Overview (Track 1 and 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intro to Genomics Research</td>
<td>Overview of in vivo Imaging Technologies</td>
<td>Proteomics</td>
<td>Course Goals</td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>Ultrasound CT, microCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype Analysis</td>
<td>Family Studies, Linkage Analysis Pharmacogenetics</td>
<td>Cardiac Imaging</td>
<td></td>
</tr>
</tbody>
</table>

MRI, magnetic resonance imaging; PET, positron emission tomography; CT, computed tomography.
humans, and small animals. The imaging technologies applicable to both cells and the in vivo condition enable the biological sciences to expand greatly beyond reductionist models and examine protein function in a more integrative and physiological context. Moreover, imaging techniques have been critical in establishing the profound importance of compartmentalization of cell signaling events and the role of structural proteins as scaffolds and molecular sinks (previously only inferred from biochemical data), providing a new direction in the understanding of these processes.

Noninvasive imaging procedures are well suited for genomics studies, as they can be readily performed in a large number of genotyped patients with high compliance. This contrasts with the utility of more invasive measures of organ system function, or cell-based experiments requiring cell harvesting, which are more dependent on patient cooperation and suffer from poor compliance. Recent advances in imaging technologies also offer the opportunity to extend the value and expand the capabilities of small animal models. Small animal imaging can now be employed as a convenient and efficient means of screening transgene expression in mice when transgene expression is linked to a reporter such as luciferase or green fluorescent protein detectable by optical imaging (18). Moreover, the noninvasive nature of imaging modalities such as micropositron emission tomography (PET) eliminates the need to sacrifice animals, making longitudinal studies in small animal models more powerful (each animal serves as its matched control) and economically feasible (one animal, multiple time-dependent measurements).

Lectures, demonstrations, and hands-on workshops were used to present multiple in vivo imaging technologies, including magnetic resonance imaging (MRI), PET, ultrasound, and computed tomography (CT). Classes were held within the Center for Biomolecular Imaging at Wake Forest University Medical Center, where each of the techniques is routinely performed for clinical diagnostic services and ongoing clinical and genomics studies.

**Biological/physical sciences curriculum.** Classes focused on cutting-edge approaches that employed primarily cell-based assays for assessing mRNA and protein expression, structure,
various genotyped patients (representing those haplotypes shown to occur in regions capable of affecting intron/exon splicing. More reductionist studies were proposed to assess gene variation effects at the cellular level. Using both monocytes and T cells derived from whole blood from each of the subjects, characterization of candidate gene expression would be assessed by real-time PCR and basal or induced levels of each gene product through either ELISA or flow cytometry analysis. To assess the effects of ADRB2 variation on relevant signaling events, ADRB2 pathway activation would be assessed in both T cells and monocytes through analysis of phosphorylation of the protein kinase A substrate, vasodilator-stimulated phosphoprotein, by flow cytometry (13).

COURSE EVALUATION

Participants and teaching faculty engaged in a round table discussion of what was felt to be the strengths and weaknesses of the course and the extent to which the course met its goals.
Participants were nearly unanimous in their belief that they had absorbed a significant amount of information and had a much better understanding of several disciplines and their approaches, such that productive participation in a comprehensive genomics project involving functional genomics was likely. All participants felt that the grant-writing exercise was the most valuable component of the course, and both participants and faculty agreed that the quality of grant applications generated was high. Many participants articulated that they would have preferred the grant-writing exercise be initiated at the very start of the course.

Those with a genomics background felt the coursework in imaging and biological/physical sciences was challenging yet understandable and were confident in their ability to incorporate many of the techniques presented into future genomics studies. It was acknowledged that most of the physical science underlying many of the technologies was difficult and often impossible to grasp (in the time available) but was in fact not necessary to fully understand to appreciate the techniques, the types of data they generated, and their application to genomics studies.

Those with engineering or basic science backgrounds found the genomics coursework particularly challenging. The participants did appear to master basic genomics terminology and were able to characterize multiple approaches for associating gene variation with a disease, but most felt uncomfortable with their command of the field overall by the end of the course. Some felt the number of approaches presented was excessive in the time allotted and that more time should have been dedicated to their application to genomics studies.

Whereas few participants felt that by the end of the course they were able to lead an investigative team in a comprehensive genomics project, all felt they could be an important contributor to such a team and that the barriers of communication had been significantly reduced. All participants stated their participation in future collaborations on genomics studies was likely and had been significantly improved as a result of participating in the course. Several participants planned on incorporating specific components of their exercise grants into their own grants in the near future. All agreed to participate in postcourse assessment of their participation in collaborative endeavors and generation of genomics grants and manuscripts.

ACKNOWLEDGMENTS

The authors thank Don Bowden for helpful comments regarding the manuscript.

Authors’ correspondence: rpenn@wfubmc.edu; vortega@wfubmc.edu; ebleeck@wfubmc.edu.

GRANTS

This manuscript and the NIH Roadmap course “Models and Technologies for Characterizing Phenotype” were supported by National Institute of Diabetes and Digestive and Kidney Diseases Grant R31 DK-069505

REFERENCES


