



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

REVIEW ARTICLE

Translational research in infectious disease: current paradigms and challenges ahead

JUDITH M. FONTANA, ELIZABETH ALEXANDER, and MIRELLA SALVATORE

NEW YORK, NY

In recent years, the biomedical community has witnessed a rapid scientific and technological evolution after the development and refinement of high-throughput methodologies. Concurrently and consequentially, the scientific perspective has changed from the reductionist approach of meticulously analyzing the fine details of a single component of biology to the “holistic” approach of broadmindedly examining the globally interacting elements of biological systems. The emergence of this new way of thinking has brought about a scientific revolution in which genomics, proteomics, metabolomics, and other “omics” have become the predominant tools by which large amounts of data are amassed, analyzed, and applied to complex questions of biology that were previously unsolvable. This enormous transformation of basic science research and the ensuing plethora of promising data, especially in the realm of human health and disease, have unfortunately not been followed by a parallel increase in the clinical application of this information. On the contrary, the number of new potential drugs in development has been decreasing steadily, suggesting the existence of roadblocks that prevent the translation of promising research into medically relevant therapeutic or diagnostic application. In this article, we will review, in a noninclusive fashion, several recent scientific advancements in the field of translational research, with a specific focus on how they relate to infectious disease. We will also present a current picture of the limitations and challenges that exist for translational research, as well as ways that have been proposed by the National Institutes of Health to improve the state of this field. (Translational Research 2012; 159:430–453)

Abbreviations: 2-DE = 2-dimensional electrophoresis; 2-D DIGE = 2-dimensional differential in-gel electrophoresis; CF = cystic fibrosis; CTSA = Clinical and Translational Science Awards program; EBV = Epstein-Barr virus; FDA = U.S. Food and Drug Administration; GWAS = genome-wide association studies; HCV = hepatitis C virus; HMP = Human Microbiome Project; HPLC = high-pressure liquid chromatography; LC = liquid chromatography; LSB = Laboratory of Systems Biology; mAb = monoclonal antibody; MRM/SRM = multiple reaction monitoring/selective reaction monitoring; MS = mass spectrometry; MS/MS = tandem mass spectrometry; NCATS = National Center for Advancing Translational Sciences; NCRR = National Center of Research Resources; NIAID = National Institute of Allergy and Infectious Disease; NIH = National Institutes of Health; NME = new molecular entity; NMR = nuclear magnetic resonance; PBMC = peripheral blood mononuclear cell; PCR = polymerase chain reaction; PRR = patho-

From the Department of Public Health; Medicine, Weill Cornell Medical College, New York, NY.

Supported by U.S. NIAID Grant 1R21AI073926-01A2 and by NIH Grant UL1RR024996 (for work in the MS laboratory); and by NIH KL2RR024997 (to E.A.).

Submitted for publication October 5, 2011; revision submitted December 23, 2011; accepted for publication December 24, 2012.

Reprint requests: Mirella Salvatore, MD, Department of Public Health and Medicine, Weill Cornell Medical College, 402 East 67th Street, LA-249, Box 74, New York, NY 10065; e-mail: mis2053@med.cornell.edu.

1931-5244/\$ - see front matter

© 2012 Mosby, Inc. All rights reserved.

doi:10.1016/j.trsl.2011.12.009

gen recognition receptor; QQQ = triple quadrupole mass spectrometry; SARS-CoV = coronavirus associated with severe acute respiratory syndrome; SNP = single nucleotide polymorphism; TB = tuberculosis; UTI = urinary tract infection; YFV = yellow fever virus

Translational research is defined precisely by the National Institutes of Health (NIH) as “the process of applying ideas, insights and discoveries generated through basic scientific inquiry to the treatment or prevention of human disease.”¹ As the field of translational research has become increasingly popular in recent years, it has undergone numerous reiterations, such that the specific meaning of the term “translational research” has itself been redefined several times.^{1,2} The recent evolution in next-generation sequencing techniques and the introduction of high-throughput methods has resulted in an explosive cascade of research applications, spanning from target identification to diagnostics and therapeutics. These technical advances have provided the impetus for some radical changes in the way research itself is conceived and performed. As a result, enhanced interactions and broader collaborations among researchers with different expertise will be required just to keep up with the rapidly changing state of science. For a multidisciplinary approach to be effective, better ways to collect and share data (eg, biorepositories) must be identified. In addition, a more rapid translation of information from basic science into useful clinical applications will require the removal of communication barriers and financial roadblocks that currently prevent basic science teams from working with each other and with clinical researchers. Finally, regulatory changes will be necessary to promote faster approval of new molecular entities (NMEs) resulting from such scientific collaborations.

A major turning point in the validation of the field of translational research came with the creation of the NIH’s Clinical and Translational Science Awards program (CTSA) in 2006, which supports a national consortium of 60 medical research institutions in 30 states.³ CTSA is the largest program of the National Center of Research Resources (NCRR), which was established initially in 1990 for the purpose of providing resources, tools, and networks to NIH researchers nationwide as well as support for other programs, such as biomedical technology and research infrastructure.^{4,5} In a climate of rapid scientific development, these programs aim to provide infrastructure, resources, training, and opportunities for collaboration among scientists and clinicians from a wide variety of disciplines and across different institutions. The common goal of these programs is to facilitate the transition from basic science discoveries to clinical studies, community practice, and health care decision making.^{4,7} Although

the programs of the NCRR have promoted higher quality training, improved research management, and enhanced community participation in clinical research, many roadblocks remain for translational research.⁶ In March 2011, Francis Collins, director of the NIH, announced the creation of a new National Center for Advancing Translational Sciences (NCATS), which will assume the CTSA and dissolve the NCRR by redistributing its other programs throughout other institutes and centers. The aim of this controversial new program will be to catalyze the progression from basic scientific discovery to clinically useful therapeutic or diagnostic tools by identifying bottlenecks in translational research and exploring innovative methods to approach and remove them.⁸ Thus, the creation of the NCATS is an attempt to shift the current paradigm in which most basic science discoveries take more than 10 years before they are translated into a clinically relevant drug.

The long time periods and costs necessary for the development and approval of useful drugs have also affected pharmaceutical research and development negatively, particularly in the area of infectious disease, where drug companies are curtailing or ceasing antimicrobial development despite increasing antibiotic resistance and “bad bugs.”⁹⁻¹¹ Similarly, despite the Orphan Drug Act, which was passed in 1983 to stimulate the development of drugs for rare diseases,¹² the pipeline for new drugs targeting diseases not of interest to pharmaceutical companies is also dry. Ultimately, whereas the creation of programs like the NCATS is a step in the right direction, the changes proposed by this organization still leave several problems unresolved and generate a new series of concerns.

In this article, we will review, in a noninclusive fashion, several examples of basic science achievements that have been brought to the forefront of translational research. A complete review of the translational applications of all recent scientific discoveries transcends the purpose of this article, so we will limit our discussion to the field of infectious disease. Overall, we draw attention to the general challenges facing translational research and define some of the recently realized achievements and future prospects of this field.

THE “OMICS” REVOLUTION OF TRANSLATIONAL RESEARCH

The Human Genome Project was conceived initially in 1985¹³ and is considered to be one of the most important steps in the start of the genomic era. By definition,

genomics characterizes the hereditary information of an organism. Research in proteomics, which defines the structure and function of proteins encoded by the genome of an organism, is the next logical step after genomics for the study of biological systems, because proteins are the major catalysts of intracellular processes.¹⁴ Metabolomics uses the systematic analysis of the complete set of chemical fingerprints left behind by certain cellular processes to determine key aspects of the function and regulation of those processes.¹⁵ Together, genomics, proteomics, and metabolomics comprise the 3 major systems biology approaches at the forefront of translational research (Fig 1).

Genomics. Since it was first invented in the late 1970s, DNA sequencing has been one of the most influential innovations in biological research.¹⁶⁻¹⁸ This rapidly evolving technique has provided a feasible approach to a multitude of problems concerning biology, including infectious disease. Polymerase chain reaction (PCR) is the most common method for amplifying nucleotide sequences based on defined primer sets that recognize specific regions of the DNA of a sample. Physicians routinely refer to data from PCR and sequencing methods as an integral part of the standard of care. For example, HIV medications are chosen after analyzing the virus genotype for the presence of mutations that would confer resistance to some of the antiretroviral drugs on the market, and viral loads (viral levels measured by PCR) are used to help predict the responses to therapy.¹⁹ Similar approaches are also being used for other viral diseases such as hepatitis B virus,²⁰⁻²² hepatitis C (HCV) virus,²³⁻²⁶ and influenza.²⁷⁻²⁹ PCR and sequencing techniques also play a crucial role in pathogen discovery and outbreak investigation. In the recent years, the importance of this field has been highlighted by the growing number of emerging pathogens and their rapid spread as a consequence of globally mobile populations, climatic changes, improved surveillance and threats of bioterrorism. Both basic research and clinical microbiology laboratories use uniplex and multiplex PCR assays for the rapid identification and typing of many bacterial and viral pathogens. Conventional uniplex PCR methods use specifically targeted sets of primers (1 set per reaction) to amplify conserved genetic regions of a sample and identify close variants of known pathogens. Multiplex PCR assays, which first emerged in the late 1990s,³⁰ are based on the same theory but can amplify multiple regions of DNA at once from a single sample, thus producing additional information that would otherwise require several times the amount of reagents to run using uniplex PCR.³¹ Although multiplex PCR techniques were a significant improvement compared with conventional PCR, both

approaches are still limited in their ability to detect new pathogens and to screen large numbers of sequences in an efficient manner.

The introduction of microarray technology in the middle to late 1990s allowed for the development of microarray-based pathogen detection platforms and high-density, pan-microbial arrays, which truly made rapid screening for a large number of pathogens feasible.³² In this application of microarray technology, sample material suspected to contain pathogen genomes is hybridized to a 2-dimensional (2-D) array (or gene chip) of hundreds or thousands of miniaturized spots containing nucleic acid probes specific for various pathogen sequences. Microarray-based pathogen-detection platforms, such as the GreeneChip and the Virochip,³³⁻³⁶ use 70-mer oligonucleotide probes of conserved genetic regions within each taxonomic group of viruses. Because of their length, these probes are more tolerant of sequence mismatches and therefore allow for the detection of unknown microbial targets.³⁷ Furthermore, the probes are updated continuously when new viruses and sequences are added to their associated databases. To identify a pathogen from a clinical specimen, nucleic acids are amplified randomly and then hybridized to the chips. Given the large number of probes and possible hybridization successes, the results are analyzed with the help of computer programs that use algorithms and statistics to identify virus hybridization patterns. Although microarray-based pathogen-detection platforms have been instrumental in the detection of numerous pathogens, including novel ones like the coronavirus associated with severe acute respiratory syndrome (SARS-CoV),^{34,38} they still rely on some prior knowledge of relevant genome sequences.³⁹⁻⁴¹ The introduction of high-throughput sequencing has revolutionized the ability to detect novel infectious agents whose genomic sequences are completely unknown or are present in extremely low numbers. The high-throughput sequencing revolution started in 1987 with the introduction of the first automatic sequencer.⁴² The first of the next-generation sequencing technologies was developed only a few years later,⁴³ but it took nearly 15 years of technologic improvements before it became commercially available.^{44,45} Currently, the most frequently used platforms of this type include the 454 large-scale parallel pyrosequencing system (454 Life Sciences/Roche, Branford, Conn) and the Solexa (Illumina, Inc, San Diego, Calif) sequencing-by-synthesis system, but this field is evolving rapidly with new platform generations regularly being developed. These techniques allow for the “deep” sequencing of clinical samples. The term “deep” refers to the depth of coverage or the average number of times that a single nucleotide is sequenced, and it allows for high levels of accuracy in

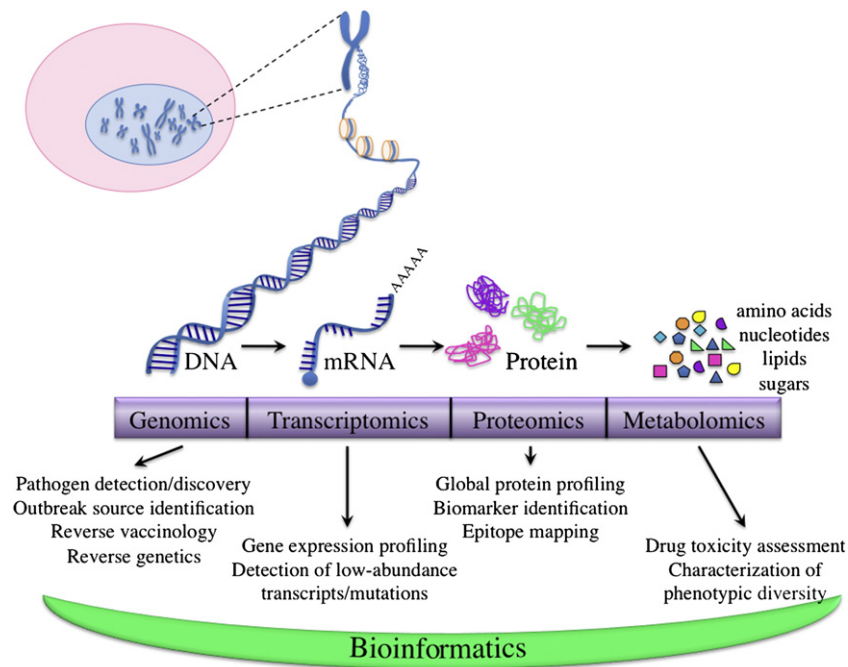


Fig 1. A “top-down” explanation of “omics.” Genomics is the study of the complete set of hereditary genetic information in an organism. With the advent of microarray technology and next-generation sequencing, numerous applications have arisen from the field of genomics, including pathogen discovery, epidemiologic advances, and a variety of molecular techniques that allow for precise manipulation of microbial genomes. As DNA begets RNA and protein, so do the fields of transcriptomics and proteomics follow logically from genomics. Transcriptomics, which is a subset of the field of genomics and concerns the collection of messenger RNA transcripts expressed within an organism, has emerged as a result of more sophisticated techniques that allow for the highly sensitive determination of low-abundance mutations, transcripts, and SNPs. Proteomics is the next major “omics” field after genomics, and it focuses on the complement of proteins, their modifications, and their interactions within an organism. With the aid of analytical techniques such as 2-DE, MS/MS, and shotgun proteomics, the field of proteomics has found useful application for the identification of biomarkers and mapping of epitopes that may provide targets for antimicrobial drug development. The field of metabolomics is a natural offshoot of proteomics as it uses many of the same MS techniques. A major difference, however, is that whereas genomics and proteomics provide insight into the potential of cellular processes, metabolomics gives an instantaneous snapshot of what is actually happening in a cell. Each one of these “omics” fields generates massive amounts of rich, detailed data that surpass the capabilities of manual data analysis. This therefore necessitates the incorporation of bioinformatics for the development of computer algorithms that are used to analyze and model data. (Color version of figure is available online.)

sequence determination.⁴⁶⁻⁴⁸ “Deep” sequencing methods can, therefore, achieve thousands to millions of simultaneous sequence reads per run, allowing for the precise determination of the entire genome of an organism. “Ultra-deep” sequencing, which is largely being enabled by platforms such as the 454 sequencing system, provides an even deeper sequence coverage and allows for several additional applications, including amplicon and transcript sequencing, which can detect extremely low-abundance genetic variations.⁴⁹ Because pyrosequencing techniques are based on the detection of light produced whenever a nucleotide is incorporated and do not rely on physical separation of DNA bases, these platforms can be run in parallel and miniaturized to any reaction volume that generates

detectable levels of light, thus driving down the overall cost of sequencing. Nevertheless, the shorter reads necessitated by the deeper sequencing capacity of next-generation sequencing creates additional challenges for sequence assembly and gene annotation.⁴⁹ Although the cost of sequencing is decreasing, the analysis of the accumulated data is costly, time consuming, and computationally challenging. The analysis is complicated by the fact that computer systems will have to discriminate between the nucleic acids of the host and those of the newly discovered infectious agent. Despite these challenges, several new viruses have been identified using “deep” sequencing systems.⁵⁰⁻⁵³ Additionally, these systems have made subsequent steps in pathogen discovery, including sequencing the entire genome to

facilitate the development of diagnostic techniques and the identification of potential therapeutic targets, more time and cost effective.

Beyond pathogen discovery, sequencing-based methods have the ability to discriminate between closely related strains and can therefore be employed successfully to study viral variations and evolution. Next-generation sequencing techniques also provide an important tool that enables epidemiologists to follow person-to-person transmission more closely and eventually identify the source of an outbreak. This was the case in the recent cholera outbreak in Haiti, where the sequences of the *Vibrio cholerae* genome from the Haitian outbreak were compared with the sequences of previously isolated strains and found to be more closely related to strains isolated in Bangladesh in 2002 and 2008 than to the South American isolates.⁵⁴ Similarly, high-throughput sequencing techniques have allowed for the fast sequencing of the 2009 H1N1 “swine” flu isolates and for the determination of the exact origin of each of its gene segments.⁵⁵

New sequencing techniques, in combination with other advances in molecular medicine, have also improved the field of vaccine discovery drastically. In the past, vaccines were constituted by crude or partially purified, inactivated viral or bacterial preparations, and often carried the risk of serious allergic reactions or autoimmune diseases because of impurities in the formulation.⁵⁶ Furthermore, administration of the vaccine itself could sometimes cause disease, as illustrated by vaccine-induced paralytic polio.⁵⁷ Although some of these problems still exist with present-day vaccines, the field of genomics has paved the way for more powerful and focused techniques that allow for a better understanding of the nature of pathogen antigenicity and how our immune system can be tuned to respond optimally.

With the explosion of sequence data from the genomic era, the genomes of a large number of medically important microbes have been decoded and made readily available. As a result, techniques such as reverse vaccinology have emerged that allow for the design of more specifically targeted vaccines.⁵⁸ The power of reverse vaccinology lies in its ability to identify novel vaccine targets based on computer-facilitated predictions of the immunogenicity and prevalence of all antigens in a given pathogen without ever having to cultivate the specific microorganism. Once potentially suitable vaccine targets are identified, each gene from the infectious agent that encodes one of these targets is cloned into an expression vector, and the resulting proteins are expressed *in vitro* and used to immunize an animal model from which antisera will be derived. Each corresponding antiserum is then tested for its ability to neutralize or protect against the original infectious agent. Using this

information, vaccines can be designed to elicit specific immune responses against the specified target and are combined with rationally designed adjuvants whose mechanisms of action could improve vaccine immunogenicity. One example of the successful use of this process is the recently developed vaccine for serogroup B *Neisseria meningitidis*.⁵⁹⁻⁶¹ An array of other potential vaccines, including those targeting group B *Streptococcus* and *Streptococcus pneumoniae* species, are currently under study.⁶²

One technique that has derived from the newly sequenced genomes of viral pathogens is that of reverse genetics, which has made it possible to generate viruses entirely by co-transfection of specifically engineered plasmid DNA in cell culture. This technique has very important implications for vaccine production because it means that very specific changes can be reverse engineered into the sequences of the plasmid DNA used for co-transfection, allowing extremely fine control over the pathogenicity and immunogenicity of the resulting virus.⁶³ Reverse genetics has been used successfully in the development of promising candidates for vaccines targeting influenza viruses. Several groups have explored reverse genetics strategies to design live-attenuated H5N1 influenza virus vaccines,⁶⁴ and a recent phase I clinical trial has demonstrated the safety and immunogenicity of a reverse genetics-derived NS1-truncated variant of an H1N1 influenza virus in male volunteers.⁶⁵ Reverse genetics has also been used in pathogenetic studies. In this case, the insertion, mutagenesis, or deletion of specific sequences in the plasmids before transfection can clarify the effect of such modifications on viral infectivity and can help to elucidate the function of any modified protein in the context of the viral life cycle.⁶⁶ Finally, the ability to exert such exquisite control over a viral sequence would permit the identification of specific sequence alterations that promote viral attenuation, which is information that can be used for the design and generation of new vaccines.⁶⁷⁻⁷¹

Additional applications that have risen from the genomic revolution include the use of genetically altered viral vectors as an alternative approach to vaccination^{72,73} or as a conduit for delivering antibodies as passive immunotherapy^{74,75} and the use of sequence data in the design of specific inhibitors and antivirals. These approaches are currently under study and are likely to come to the forefront in the near future.

The genomic revolution has improved significantly the way we detect, prevent, and treat infectious diseases. Future advancements in this field will continue to decrease the time necessary to identify newly emerging human pathogens and produce new vaccines for clinically relevant infections.

Proteomics. Proteomics is a predominantly mass spectrometry (MS)-based technique that deals with the large-scale evaluation of cellular function at the protein level. MS analysis is accomplished by ionizing a sample into its molecular components (eg, with an electron beam, laser, etc) and then passing them through an electromagnetic field that separates them by their ratio of molecular mass to charge (m/z).⁷⁶ Proteomic analysis (as opposed to other MS-based techniques) seeks specifically to elucidate the entire complement of proteins and protein modifications and/or interactions in a specimen at a given time. Because proteins are the major effectors of cellular pathways, proteomics provides a way to interpret information encoded within the genome of an organism. Since the term proteomics was first coined in 1995,¹⁴ the field has expanded tremendously, largely because of the development of ionization and analytic methodologies, which have allowed for the more efficient separation and detection of almost all protein components within a sample of interest.⁷⁷ Unlike genomics, no method or instrument currently exists that can identify and quantify the components of a complex protein sample in a single step.^{77,78} Instead, a combination of steps with varied methodologies and instrumentation are used.

Proteomic platforms are characterized by 2 key steps: (1) protein separation, mostly obtained through either electrophoretic or chromatographic methods, and (2) protein analysis using primarily MS-based techniques. Two-dimensional electrophoresis (2-DE), which was developed in the 1970s, is a method that allows all proteins contained within a complex biological mixture to be separated by their isoelectric points and molecular weights, and then visualized by staining.^{79,80} Today, 2-DE-based proteomic approaches remain the gold standard for the separation of complex protein mixtures despite several limitations, including the detection of only relatively abundant proteins in a complex mix, poor separation of insoluble and/or hydrophobic proteins, protein comigration, and the labor-intensive nature of identifying multiple protein spots on the resulting gels.⁷⁹ To address the problem of comigration, a modification of 2-DE, known 2-D differential in-gel electrophoresis (2-D DIGE) was developed. In this technique, multiple protein samples are labeled separately with fluorescent dyes, mixed together, and then resolved on a single gel, eliminating the need for cross-gel comparisons and increasing the overall statistical power by the inclusion of a greater number of samples.^{77,81,82} Whereas the development of 2-D DIGE represented some improvement, the need for the increased sensitivity of detection and more efficient separation of proteins within a complex

mixture remained and was the driving force behind the development of “gel-free” technologies.⁷⁹

Gel-free proteomic approaches were made possible by coupling the physical separation capabilities of chromatographic techniques to the mass analysis capabilities of tandem mass spectrometry (MS/MS).^{79,80} Chromatography is a technique that involves moving the components to be separated (mobile phase) across an immobile support (stationary phase) such that the migration rates of solutes (ie, proteins) vary depending on their chemical properties.⁸³ Liquid chromatography (LC), in which the mobile phase is liquid, is one of the most widely used protein separation techniques in proteomics.⁸³ With the use of smaller particles (less than 2 μm), monolithic columns (using porous channels rather than beads) and extremely high pressures (up to 19,000 p.s.i.), high-performance LC (HPLC) and ultrahigh-performance LC have improved the efficiency and resolution of protein separation drastically.⁸⁴ In HPLC, complex protein mixtures are digested by trypsin into their peptide components, which are then moved in a liquid mobile phase at high pressures across a stationary phase that allows for their separation based on polarity. Next, the sorted peptide components are analyzed by MS/MS. The first mass spectrometer is programmed to detect peptides over a range of predetermined masses to distinguish targets of interest from the vast array of proteins typically present in complex biological fluids. The resulting MS1 data are used to pinpoint ions of interest, which are then fragmented, measured, and recorded by the second mass spectrometer (MS2 data). Finally, computer algorithms are used to aid researchers in identifying individual proteins from combined MS1 and MS2 data.^{77,79,81,85} As an example, the SEQUEST algorithm, developed in 1994, uses the fragmentation spectra of peptides to infer amino acid sequences and match them to translated genomic sequences. This approach has an identification rate of greater than 100 proteins per run.⁸⁶ The newest iteration of this software, MacroSEQUEST, features enhanced capabilities (eg, posttranslational modification scanning) and a dramatic improvement in processing speed compared with the original version.⁸⁷

The combination of HPLC and MS/MS, which is now known as “shotgun proteomics,” has proven to be a fast and highly sensitive approach, especially when scaled down to microflow or nanoflow rates.^{88,89} As such, shotgun proteomics has not only allowed for the identification of thousands of individual proteins within a given sample with essentially no prior knowledge of the component peptides but also has greatly facilitated the analysis of entire proteomes. Although shotgun proteomics represents a significant technological advancement in the analysis of complex

biological mixtures, the concentrations of individual proteins within such samples can be low enough to require levels of sensitivity beyond what even standard HPLC conditions can achieve. By decreasing flow rates and column diameters, this level of sensitivity can be obtained, but only by adapting the system to the special requirements of nanoliter flow rates (20–1000 nL/min).⁸⁸ Another shortcoming of shotgun proteomics is the inability of one-dimensional HPLC to provide sufficient separating power to manage the vast complexity of proteins inherent in most biological specimens. To address this issue, multidimensional separating approaches, such as multidimensional protein identification technology (MudPit), which combines strong cation exchange chromatography with HPLC, were developed.^{79,85} Even with such improvements, these advanced shotgun proteomics platforms can still identify only approximately 40% to 70% of the proteins predicted by a microbial genome (less for proteomes of higher complexity). Furthermore, the actual proteome size may be smaller or larger than predicted by the genome because of the absence of transcription of some proteins and the alternative splicing and/or posttranslational modification of others.^{77,81} Therefore, whereas shotgun proteomics does enable the identification and relative abundance of a large number of proteins in a sample, it is not optimal for absolute protein quantitation.

In addition to identifying proteins within complex biological mixture, the need to quantitate these components can provide information regarding their dynamics within the context of a biological system. This added focus has led to the emergence of a derivative field known as quantitative proteomics. Protein quantitation can be relative or absolute. Relative protein abundance may be estimated by several methods, including by measuring the total number of repeat observations of MS/MS spectra from all peptides in a protein (spectral counting), calculating the percentage of possible peptides per protein observed (peptide counting), or by comparing directly the extracted ion chromatograms (XIC, aligned using both mass and elution time) of one sample with that of another sample.^{90,91} These label-free techniques have been facilitated greatly by the development of new software programs, such as MZmine, MapQuant, and SuperHirn.⁹¹⁻⁹³ Alternatively, protein quantification may be accomplished by combining shotgun proteomic techniques with stable isotope labeling.^{90,91} Using this approach, samples are labeled differentially with stable isotopes (heavy vs light), then pooled and analyzed together. Peptides common to both samples, but labeled with different isotopes, behave the same on analysis and are detected as peak pairs. By comparing the sizes of these peaks, the relative abundance of the

peptides in each sample can be compared.⁷⁹ The methods for absolute quantitation of protein abundance are more desirable in many cases (eg, for determining copies per cell). One technology that addresses this need and is now commonplace in clinical settings is called multiple reaction monitoring, also known as selective reaction monitoring (MRM/SRM). This technique exploits the unique capabilities triple quadrupole (QQQ) MS to analyze precursor and fragment ion pairs. QQQ MS uses 3 sets of 4 circular metal rods (quadrupoles) that run parallel to each other. The first quadrupole (Q1) acts specifically to select peptide ions with predefined m/z values. These peptides are then fragmented by the second quadrupole (Q2), which serves as a collision cell. The third quadrupole (Q3) can be set either to scan the entire m/z range to provide information about the structure of the original ion or to filter out fragments with specific m/z values that may be indicative of particular characteristics (eg, functional groups). MRM/SRM can provide targeted information (sensitive to the attomole level) about multiple peptides across a wide concentration scale within a single sample.^{90,94,95}

The sensitivity of the MRM/SRM technique can also be improved by enriching the target from a complex background using antibody-mediated approaches such as stable isotope capture by anti-peptide antibodies and immuno-matrix-assisted laser desorption/ionization.⁹⁰ As with relative protein quantitation, absolute protein quantitation possesses its own arsenal of computer programs designed to assist in the analysis of MRM/SRM data, including Skyline MRM/SRM builder, MRMpilot (Applied Biosystems), Peptide Optimizer, and Dynamic MRM software (Agilent Technologies, Santa Clara, Calif), MaRiMba, MR Maid, and MRMer.⁹⁰

To date, these analytical platforms have given rise to 4 major proteomic applications: (1) global protein profiling—analysis of all proteins within a given biological unit or sample, (2) analysis of protein modifications (phosphorylation, ubiquitination, nitrosylation, etc), (3) analysis of protein–protein interactions, and (4) analysis of protein–gene interactions.^{77,78} With regard to the study of infectious diseases, global protein profiling using shotgun proteomics has been used recently to identify and quantitate protein expression in the malaria parasite, *Plasmodium falciparum*, according to the stage of its life cycle and its host compartment. These studies resulted in more than 200 candidate proteins for possible stage-specific drug or vaccine targets. In addition, they generated information on a large set of “orphan” proteins (proteins that could be mapped to the genome but were not found in the set of previously known proteins), protein interactions, and protein expression at each stage of the parasite’s life cycle.^{96,97} An analysis of protein modifications has been used to

identify phosphoprotein signaling pathways that are altered in Rift Valley fever virus-infected airway epithelial cells and to uncover potential targets of reactive nitrogen intermediates in *Mycobacterium tuberculosis* (*M. tuberculosis*), thus informing on potential targets of novel antiviral and antimycobacterial compounds.^{98,99} Finally, analyses of protein-protein and protein-gene interactions have been used to study the relationship between Epstein-Barr virus (EBV) infection and nasopharyngeal carcinoma, as well as alterations in host cellular processes associated with HIV-1 viral replication.^{100,101} Recent studies have aimed to establish the context in which these types of interactions occur by combining proteomic-based techniques with either cryo-electron tomography or “membrane shaving” with proteinase K digestion to examine specific cellular localization.⁸¹ Such approaches have been used successfully to map the spatial proteome of *Leptospira interrogans*¹⁰² and the membrane proteome of *Staphylococcus aureus* (*S. aureus*).¹⁰³ Studies such as these ultimately might lead to an improved understanding of microbial pathogenesis and provide important clues for the identification of novel biomarkers and antivirals.

Because proteins carry out many functional biological activities, the proteomic analysis of an organism provides a layer of detail beyond that of genomics. Although it is still an emerging technology, proteomics has already had a profound impact on clinical and biological research by informing on the pathogenesis of disease, facilitating biomarker discovery and aiding in the identification of potential drug targets. New developments in MS platforms, MS-based tissue imaging techniques, and combined proteomic/genomic approaches will likely extend the impact of proteomics to drug design and testing as well as pathogen identification and discovery, which will continue to advance our understanding of complex biological processes.^{77,78}

Metabolomics. The newest field of systems biology, metabolomics, has been defined as “the comprehensive and simultaneous systematic determination of metabolite levels in whole organisms and their changes over time as a consequence of stimuli such as diet, lifestyle, environment, genetic effect and pharmaceutical intervention, both beneficial and adverse.”¹⁰⁴ A more concise definition of metabolomics might be the global profiling of all small molecule metabolites contained within a sample of interest at a given time. Metabolites, which are the intermediates and products of all cellular processes, typically fall into one of 4 categories: amino acids, nucleotides, lipids, and sugars. Moreover, metabolites represent the level at which most pharmaceuticals exert their effects. Therefore, by studying the production and consumption of these

chemical fingerprints, one can gain insight into the dynamic functions of any biological system.^{104,105}

The analytic techniques employed in metabolomic studies are largely similar to those used in proteomic studies and are based primarily on either nuclear magnetic resonance (NMR) spectroscopy or MS. The typical workflow involves mechanical extraction coupled to a rapid cooling step that halts metabolism in its native state, followed by detection (either via NMR or MS) and data analysis. MS-based metabolomics requires the additional separation of components by either gas chromatography or LC for enhanced molecular identification. For each molecule within a sample of interest, LC-MS-based metabolomics generates 3 pieces of data: (1) chromatographic retention time, which informs on the chemical structure; (2) a m/z ratio, which informs on the molecular mass; and (3) abundance. The retention time and m/z ratio may then be queried against a library of known standards to determine the identity of each metabolite.¹⁰⁶ An NMR-based metabolomic analysis, by contrast, provides a nondestructive way to view changes in metabolite levels either *in vitro* or *in vivo* more broadly. Moreover, it provides more detailed structural information than MS. Because MS analysis provides better resolution and sensitivity, it remains the predominant methodology for analyzing metabolic data. All metabolomic analyses result in complex, multivariable data sets that require visualization software for spectral analysis as well as bioinformatics and statistical methodologies for interpretation.^{104,107}

Metabolomic analysis of data sets produced by NMR and MS yields 2 major applications, steady-state metabolite profiling and labeled flux analysis. Steady-state metabolite profiling may be global (profiling of all small molecule metabolites within a sample) or targeted (analysis of only prespecified metabolites of interest), and it has been used to discover new biomarkers, inform on dysregulated pathways associated with disease, and generate information about cellular functions. Labeled flux analysis, in which radiolabeled isotopes are used to analyze changes in the amount of a given metabolite that is used and produced in a particular metabolic pathway, is an approach that yields more detailed information about the dynamic nature of biological pathways and can be used to better identify potential therapeutic targets.¹⁰⁷ Recently, Henderson et al¹⁰⁸ used a steady-state LC-MS-based metabolomics approach to identify how bacterial strains associated with urinary tract infections (UTIs) differed from those that colonize the gastrointestinal tract. Their results indicated that urinary *Escherichia coli* (*E. coli*) strains preferentially express 2 small molecules, yersiniabactin and salmochelin, which are known to be important in iron scavenging and support bacterial survival and growth. These

findings suggest that new antibiotics directed against yersiniabactin- or salmochelin-producing *E. coli* strains might be an improved and more targeted strategy to prevent recurrent UTIs. In another example, Munger et al¹⁰⁹ used changes in metabolic flux between uninfected and human cytomegalovirus-infected cells to identify pathways that were upregulated by viral infection, thus providing important information regarding potential targets for novel antivirals aimed at blocking viral replication.

Lipidomics and glycomics are two growing fields related to metabolomics that merit special mention.^{110,111} Lipidomics is the large-scale study of the networks and pathways employed by all cellular lipids, and it employs largely the same tools as those used for metabolomics, but with specific modifications that accommodate the unique characteristics of lipids.¹¹² Lipids have key functions in signaling pathways, energy storage, and the structural integrity of cell membranes. They also play important roles in host–pathogen interactions and immunomodulation.^{111,113} Glycomics is the comprehensive study of all sugars in an organism.^{114,115} Because cellular sugars can be simple sugars, glycoproteins, or glycolipids, there is no common experimental approach for their analysis. Instead, glycomic studies use a combination of the techniques found in metabolomics, proteomics, and lipidomics.¹¹⁰ Glycoconjugates participate in a variety of biological processes associated with cell adhesion and migration, bacterial and viral recognition, signaling pathways, and innate immunity.¹¹⁵ Therefore, the study of changes in the lipid and sugar components of a biological system can provide powerful clues about disease-related mechanisms and novel therapeutic targets. Indeed, these approaches have shown particular promise in the study of infectious disease. Lipidomics has been used to uncover the mechanisms involved in enhancing the drug susceptibility of *Candida albicans*¹¹⁶ to gain insight regarding the intracellular processes and lipid composition of HIV,^{117,118} CMV,¹¹⁹ and HCV,^{120,121} and to identify lipidomic modifications that occur in the host as a result of bacterial or viral infection.¹²² The utility of glycomics has also been demonstrated recently in the characterization of specific interactions between viral proteins and various forms of sialic acid,^{123,124} the detailed analysis of how lipopolysaccharide from *Coxiella burnetii* may contribute to pathogenesis,¹²⁵ and the discovery that the glycan coats of HIV-1 virions match that of the T cell immunomodulatory microvesicles from which they derive, suggesting a mechanism by which the virus can evade the antiviral immune response.¹²⁶

The aforementioned examples demonstrate the potential of metabolomics as a tool for enhancing development of novel diagnostics and therapeutics.

Like proteomics, metabolomics has already been used as a vehicle for investigating disease pathogenesis, potential biomarkers, and therapeutic targets. Moving forward, improvements in automated metabolite identification will expand exponentially the potential uses of metabolomics-based approaches. Additionally, the integration of metabolomics data with other global profiling techniques, such as genomics and proteomics, will allow for a deeper understanding of biological pathways than possible to date, with unparalleled possibilities for clinical application.

THE “OMICS” OF INFECTIOUS DISEASE

With the advent of genomics, proteomics, and metabolomics came an explosion of other “omics” sciences. This convention arose, in part, because the suffixes “ome” and “omics” succinctly describe a holistic way of looking at relationships that exist within a relatively complex scientific domain; however, this trend also reflects the growing involvement of bioinformatics and the internet as a way to integrate complex biological data.¹²⁷ Concurrent to the significant advances in “omics” sciences that have allowed for a better understanding of the molecular and cellular processes that occur within humans, researchers also began to appreciate the number and complexity of organisms that resided on and in the human body. After the publication of the human genome sequence in 2001,^{128,129} scientists began to call for a “second human genome project” that would catalog the complete inventory of microbial genomes at major sites of colonization on the human body.^{130,131} Because microbiota are believed to play a fundamentally important role in human health and disease,¹³² a microbial inventory in combination with a comprehensive analysis of host gene expression would provide insights into the mechanisms of this interaction. A landmark study that was published in 2005 by Eckburg et al¹³³ used the 16S ribosomal RNA sequencing technique to characterize the endogenous gastrointestinal microbial flora in 3 individuals. The authors concluded that the bacterial communities varied tremendously from one individual to the next. This and other studies that came after indicated that a larger scale investigation of the human microbial ecosystem was warranted. Thus, the NIH initiated a 5-year project called the Human Microbiome Project (HMP) in 2007.¹³⁴

The HMP uses various “omics” approaches to characterize the complexity of microbial communities, determine the existence of a core microbiome at various colonization sites on and in the human body, and examine the relationship between changes in the microbiome and human health (Fig 2).^{135,136} To date, the HMP has

determined that substantial alterations in the human microbiome are important for a variety of disease states, including psoriasis,¹³⁷ sexually transmitted infection of the male urogenital tract,¹³⁸ Crohn disease,¹³⁹ gastroesophageal reflux disease,¹⁴⁰ and others. There is also a growing interest regarding the effects of the gut microbiome on the development of the immune system. A recent study showed that mice treated with antibiotics that depleted specific gut bacterial populations would have an impairment of virus-specific cell-mediated and humoral responses in the lung after infection with influenza virus.¹⁴¹ These data will require confirmation and more in-depth studies that address mechanism; however, they raise a series of new questions about the role of the bacterial microbiota in human health. Not only has the HMP made major advances in our understanding of the human microbiome, but also it has contributed significantly to the growing field of metagenomics by developing broadly applicable techniques for analyzing massive amounts of sequence data. For example, Langmead et al¹⁴² designed a cloud-computing software tool called Crossbow that can analyze large DNA sequence data sets more rapidly (38-fold coverage of the human genome in 3 h) to perform the alignment and detection of single-nucleotide polymorphisms (SNPs) without sacrificing accuracy. Another group developed a robust software program called QIIME (pronounced “chime” and stands for quantitative insights into microbial ecology) that can analyze pyrosequencing data from thousands of heterogeneous experimental data sets in order to acquire information rapidly about various microbial communities.¹⁴³

Although the HMP does fund at least 1 project aimed at characterizing the relationship between viruses and human illness,¹⁴⁴ the primary focus of this project is currently with bacterial populations. Therefore, although not as well-defined as the HMP, at least 2 other infectious disease-related “omes” have emerged in parallel: the human virome and the human mycobiome, which comprise the myriad of viruses and fungi, respectively, that inhabit the human biosphere and play roles in health and disease (Fig 2). In 2003, Anderson et al³⁶ proposed a system for the global screening of viruses in large human populations. Although this global screening is not limited to viruses that persist chronically in the human body as is the case for the commensal bacteria populations included in the HMP, such a system could lay the groundwork for the rapid detection of outbreaks and the systematic discovery of novel human viruses, as well as provide better clues about etiologies in human disease. For example, 1 study investigating the human virome uncovered differences in the metabolic states of DNA viromes in the respiratory tract of cystic fibrosis (CF) compared with non-CF patients.¹⁴⁵

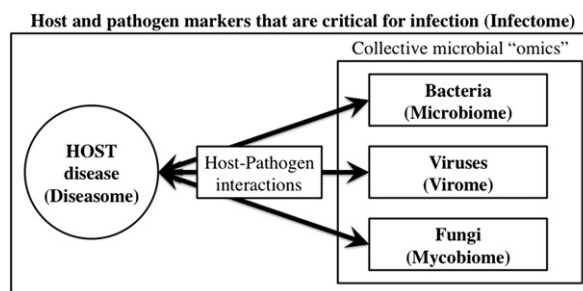


Fig 2. The “omics” paradigm for host–microbe interactions. State-of-the-art “omics” technologies have paved the way for the comprehensive characterization of a variety of hosts and microbes. The microbiome, virome, and mycobiome are 3 microorganism-specific data sets that have arisen from the “omics” revolution. Although each subset comprises distinct information about its target (bacteria, viruses, fungi, etc), they do not exist in isolation. An increasing number of studies is focused on the intersection between the collective set of microbial “omics” and the profile of gene expression that is associated with human disease (diseasome). Thus, investigation into the complex dynamics that occur during infection (infectome) can provide significant insight regarding the etiology and progression of disease, as well as lead to the identification of targets at which to aim therapeutic intervention.

Another group demonstrated that bacteriophage populations became more similar in the guts of individuals who were on a diet of similar fat content, thus raising the possibility that virus populations could be engineered to combat obesity.¹⁴⁶ In contrast, studies regarding the human mycobiome seem to still be in their infancy as the first study describing a baseline mycobiome in humans was published only recently.¹⁴⁷ In this study, the authors used multitag pyrosequencing to characterize the fungi present in the oral cavity of healthy individuals and identified differences and similarities in white and Asian populations.

In light of major advancements in the genotypic characterization of bacterial, viral, and mycotic pathogens associated with humans, it is clear that an additional holistic approach will be required to understand the phenotype of this interaction more fully and how it may affect human health and disease. Thus, the emerging field of infectomics will aim to study the structural and functional genomics and proteomics of microbial infections more efficiently, accurately, and integratively (Fig 2). Whereas the microbiome, virome, and mycobiome will identify the comprehensive inventories of microbes that inhabit the human body, the infectome will examine the distinct signatures of each microbe as they vary depending on host or microbial gene expression, tissue type infected, and other dynamic characteristics of both host and pathogen.¹⁴⁸ By networking the infectome with the human diseasome (Fig 2), which links certain diseases together based on common gene expression profiles,¹⁴⁹ it will be possible to identify links

between microbial infection and the etiology of human disease, which could inform the design of better vaccines and therapeutics.¹⁵⁰

TAKING IT ALL TOGETHER: A LOOK AT SYSTEMS BIOLOGY

The need to manage the rapidly accumulating number of sequences and massive amounts of data from high-throughput platforms has required the development of more sophisticated computer programs to analyze and integrate the data. Thus, systems biology has developed into a new field that aims to understand the complexity of pathogen–host interactions by using computational integration of high-throughput experimental data and by modeling molecular networks via bioinformatics. Intrinsic to this approach is the idea that biological systems display “emergent properties” which are complex patterns that arise from a multiplicity of relatively simple interactions. Therefore, a major objective of the systems biology approach with regard to infectious disease is to make predictions about the dynamic behavior of biochemical networks involved during infection with any pathogen.¹⁵¹ Thus, whereas small-scale studies look only at one side of the interaction, system biology gives a global perspective of the events.

Under the umbrella of systems biology, 4 main approaches are relevant to the study of infectious diseases: the systems biology of viral and bacterial pathogens, systems immunology, systems vaccinology, and high-throughput drug discovery. Common to each approach is the goal of integrating high-throughput multiomics data to construct predictive models of the networks and dynamic interactions between pathogens and their hosts to identify preventive and therapeutic targets for clinical development more quickly.

Systems biology of viral and bacterial pathogens. To recognize the importance of systems biology for the understanding, integration and practical application of the results of high-throughput screenings in the field of infectious disease and host–pathogen interaction, the National Institute of Allergy and Infectious Disease (NIAID) has created the Systems Biology for Infectious Diseases Research program. The research activities of this program were separated initially into 4 centers: The Tuberculosis (TB) Systems Biology Center, The Systems Virology Center, The Center for Systems Influenza, and The Center for Systems Biology for Enteropathogens. Each center will use both computational and experimental methodologies to analyze and model host–pathogen interactions comprehensively to determine the extent to which molecular and cellular networks are induced or altered during the course of infection. The overarching

goal of this program is to generate and develop reagents, protocols, and statistical models that will be rapidly made available to the scientific community to promote swift advancements in scientific research and translation to clinical practice.¹⁵²

The purpose of the TB Systems Biology Center is to model the molecular pathways of *M. tuberculosis* under conditions relevant to TB pathogenesis using a combination of experimental and computational techniques. For example, chromatin immunoprecipitation followed by next-generation DNA sequencing analysis (ChIP-seq) is a powerful technique that can be used to map global DNA binding sites precisely for any protein of interest so that the regulation of gene expression within the context of pathogenesis can be more fully understood.¹⁵³ By combining this technique with a variety of “omics” data (transcriptomics, proteomics, glycomics, metabolomics, lipidomics, etc) during *in vitro* and *in vivo* growth of *M. tuberculosis*, the TB Systems Biology Center aims to measure the baseline metabolic and topologic state of the gene regulatory network of *M. tuberculosis*, delineate global alterations that occur in both the pathogen and host during infection of host macrophages, and integrate profiling data to develop predictive computational models of bacterial regulatory and metabolic networks. Using these models, predictions can be made regarding the state of these networks during infection with *M. tuberculosis*, leading to additional hypothesis-driven research.^{151,154}

The Systems Virology Center and the Center for Systems Influenza are similarly aimed at using systems biology to analyze and model alterations comprehensively that occur in molecular and cellular events during the course of virus infection and to provide deeper understanding into the processes that lead to the initiation and progression of infectious disease. The Systems Virology Center will focus on creating a unified view of the mechanisms of pathogenesis for highly pathogenic respiratory viruses including the 2009 pandemic H1N1 influenza virus, the H5N1 avian influenza virus, and SARS-CoV.^{151,155} In contrast, the Center for Systems Influenza will focus more specifically on defining host innate immune responses to infection with influenza virus strains of varying pathogenicities, as well as characterizing regulatory networks associated with concurrent secondary infections like *S. aureus*.^{151,156} To date, several specific achievements have come out of these centers. Using a technique called weighted gene correlation network analysis, in which relationships between genes that are coexpressed are weighted based on their strengths or capacities,¹⁵⁷ the transcriptional response of human bronchial epithelial cells infected with a highly

pathogenic avian H5N1 influenza virus was profiled and compared with cells infected with a less virulent strain to identify potential novel mediators of pathogenesis.¹⁵⁸ Another group performed a comprehensive proteomic analysis using LC-MS/MS to identify a large number of human host proteins associated with the polymerase complex of an H5N1 influenza virus. In addition to previously published interactions, this method identified novel interactions of the viral polymerase with host mitochondrial, apoptosis-inducing, innate antiviral, and RNA polymerase accessory proteins that could provide new insight into mechanisms by which the viral polymerase may contribute to host cell pathogenicity.¹⁵⁹ Additionally, this center has used next-generation sequencing and analysis of the host transcriptome in response to influenza virus and SARS-CoV infections to investigate beyond the proteome and characterize the roles of non-protein-coding RNAs in regulating the host innate immune response.¹⁶⁰

Finally, the Center for Systems Biology for Enteropathogens uses computational and experimental “omics” methodologies to analyze and model the interactions that occur between a host macrophage and the *Salmonella enterica* (*S. enterica*) and *Yersinia pestis* bacterial species to shed light on mechanisms behind the outcome of infection (either bacterial replication and host cell death or containment of the pathogen).^{151,161} Researchers affiliated with this center have developed and used constraint-based reconstruction and analysis methods to simulate, analyze, and predict the way metabolic processes behave in various microorganisms.^{162,163} In another study, novel virulence factors of *S. enterica*, including a new class of translocated effectors, were discovered and characterized through sample-matched multiomic measurements of various mutant strains.¹⁶⁴ Other computational approaches, such as support vector machine-based identification and evaluation of virulence effectors, have also been used to facilitate the identification of secreted bacterial effector proteins from genomics sequence information.¹⁶⁵

SYSTEMS IMMUNOLOGY

To understand fully the mechanisms by which drugs and vaccines combat disease, one must also grasp the way in which the human immune system responds to its infectome. Unfortunately, as science analyzes the fine details of immunity with higher and higher resolution, there is actually a decreasing ability to predict how the system behaves as a whole in the context of infection. To that end, researchers have begun to take systems biology approaches to immunology. One example of an effort to meet the challenge of amassing high-quality

quantitative data using the available technical advances that “omics” sciences have provided while organizing and analyzing that data in ways that are meaningful to the study of immunology is the development of the NIAID Program in Systems Immunology and Infectious Disease Modeling in 2006, which became the Laboratory of Systems Biology (LSB) in 2011.^{166,167} Rather than being a collection of independent laboratories, the LSB is an integrated group of scientists and staff who play a major role in fostering the growth of systems biology efforts both within the United States and abroad.¹⁶⁷

At the forefront of this field is a rapidly increasing collection of systems biology techniques that have been applied to the unique set of problems encountered at the intersection between the human immune system and infectious disease. For example, computer simulations of the molecular dynamics of association, dissociation and posttranslational modifications of large peptide folding processes are used frequently in the field of systems immunology to characterize receptor–ligand interactions that occur between the host and pathogen. One study used this technique and the model system of EBV to demonstrate that some T cell receptor structures appear more frequently than others because of certain pattern recognition receptor (PRR)-like traits that occur within the T cell repertoire.¹⁶⁸ Another group used constant pH molecular dynamics simulations to investigate the mechanism behind the ligand-binding activity of an innate immune system PRR that recognizes a pH-sensitive microbial protein to understand more completely how this type of interaction plays a role in immune surveillance and clearance of apoptotic cells.¹⁶⁹ In addition to simulations that model the dynamics of molecular-level interactions, several programs have emerged that combine the techniques of systems biology with information provided by data-driven prediction methods. Because of advances in the predictive capacity of systems biology research, the newest version of the computer simulation software, C-ImmSim, can model the immune response *in silico* more accurately. Recently, the creators of C-ImmSim demonstrated a variety of new applications now accessible by their improved program, including the modeling of primary and secondary responses to prime/boost immunization, the way in which immunodominant peptides lead to affinity maturation and the relationship of diversity in MHC to “time to AIDS” in HIV-infected patients.¹⁷⁰ Another simulation framework that has made its way into the public domain is the Multiscale Systems Immunology platform, which aims to model the early immune response to vaccination and natural infection by incorporating realistic biophysics and intracellular dynamics.¹⁷¹

SYSTEMS VACCINOLOGY

The growing field of systems vaccinology combines modern analytic tools of systems biology with data on human immunological response patterns to vaccines in order to uncover molecular signatures of vaccine efficacy and to guide the design and evaluation of new vaccines. Most recently, systems vaccinology has been used to study the immune responses to vaccines against yellow fever virus (YFV).^{172,173} Although the YFV vaccine has been used in humans since 1937, little was known about how the vaccine protected from infection. With the advent of systems vaccinology, scientists examined comprehensively the initial molecular signatures in individuals vaccinated against YFV not only to elucidate the mechanisms behind the vaccine's efficacy but also to establish predictive correlates for the later development of protective responses.^{172,173} Systems vaccinology was also used to establish the predictive correlates of a protective antibody response after administration of the trivalent inactivated influenza virus vaccine,¹⁷⁴ and similar approaches have been applied to the study of immune responses to *Brucella melitensis* and fungal infections.¹⁷⁵⁻¹⁷⁷ Ideally, subsequent development of these models will define inherent and infection-mediated perturbations in the genetic regulatory networks of the host and will enable predictions about how the host will respond to vaccines. Based on these models, we can design vaccines that induce optimal immune responses without toxic effects, thus improving vaccine safety profiles.^{75,178}

Once the correlates of vaccine-induced protective immune responses against infection are well defined with the help of systems vaccinology, new methodologies must be identified that can apply this information to the design and development of new vaccines. The likelihood that commonalities exist among many different vaccines regarding the correlates of successful immune responses will drive faster and more accurate ways of screening vaccine candidates for their effectiveness. For example, Pulendran et al¹⁷⁹ predicted the development of a vaccine chip microarray, similar to the MammaPrint prognostic chip that was developed for breast cancer, which will be able to predict the immunogenicity of any vaccine. Although this specific technology may so far still exist only in theory, several systems biology approaches are being used currently to aid with antigen discovery and vaccine development. Phage display is one such technique that enables the screening and identification of unique molecules with highly specific affinities for targets of interest.¹⁸⁰ This approach works by introducing a foreign sequence at an appropriate site within a bacteriophage coat protein gene that

results in the display of that protein on the phage surface. If random sequences are inserted, then it is possible to obtain a library containing a large number of distinct proteins, each displayed on a different phage. By immobilizing a DNA or protein target to a solid surface, libraries can be "panned" for phages that recognize the target. After washing, phages that do not bind to the target will be washed away, whereas specific phages can be eluted and used to produce more phage, thus enriching the population for relevant binding phages and enabling the isolation of specific proteins of interest for use in subsequent applications.¹⁸¹ Phage display is a powerful technique because of its many applications for vaccine and drug development. One application of phage display technology is the identification of the epitope repertoire of antibodies in postvaccination human sera. To this end, Khurana et al¹⁸² used H5N1 influenza virus whole genome-fragment phage display libraries to evaluate the quality of antibody responses in humans after immunization with a novel virus-like particle vaccine. This approach also enabled the authors to determine the most appropriate dosage of the vaccine to induce the desired antibody repertoire and will help them to evaluate the usefulness of incorporating an adjuvant into the vaccine.

HIGH-THROUGHPUT DRUG DISCOVERY

In the war against infectious disease, the ability to develop nontoxic and fast-acting therapeutic and prophylactic drugs is just as important as being able to design safe and efficacious vaccines. Besides its usefulness in evaluating the efficacy of vaccine-induced immune responses, phage display can also be used to identify and synthesize monoclonal antibodies (mAbs) that can be used for therapeutic purposes in a variety of infections. An additional benefit of antibody identification through phage display is that because phage libraries can be generated from any set of sequences, it is possible to display human antibody fragments, thus avoiding the side effects inherent in humanized or chimeric antibodies.¹⁸³ This approach has produced several mAbs with potential therapeutic applications against agents of infectious disease, including influenza A virus,¹⁸⁴ *Clostridium difficile*, HIV, viral hepatitis, rabies, *Pseudomonas aeruginosa*, methicillin-resistant *S. aureus*, and *Bacillus anthracis* (*B. anthracis*).¹⁸⁵ Of these, the mAb targeting the protective antigen of *B. anthracis*, raxibacumab, is so far the only one that has met all criteria for approval by the U.S. Food and Drug Administration (FDA).¹⁸⁶

With the growing emergence of antibiotic-resistant bacterial strains and the ever-present public health

threat of pandemic viral infections, there has been an increasing pressure on the scientific community to discover, test, and produce novel antimicrobial drugs. High-throughput methods for synthesis and screening an increased number of targets resulting from the “omic” revolution have significantly advanced the search for new therapeutics. Combinatorial chemistry and diversity-oriented synthesis are 2 of the most important new methodologies in drug discovery because they have drastically reduced the time and costs associated with producing effective, marketable, and competitive new drugs. The power of these techniques comes from the ability to synthesize extremely large libraries of high-quality compounds with complex molecular diversity that increases their likelihood of interacting with biological macromolecules in a selective manner with high affinity.¹⁸⁷ Phage display libraries have also been used widely in screening for novel drug candidates based on specific protein–protein interactions. Although they are frequently cheaper than most chemical synthesis methodologies, there are still several limitations regarding stop codon placement within displayed sequences, inefficient transformation methods, peptide length, and restricted use of naturally occurring amino acids. Nevertheless, several proof-of-principle studies have been performed recently to overcome these limitations, including 1 study that demonstrated that the incorporation of genetically encoded unusual amino acids might be feasible.¹⁸⁸

After the synthesis of potential molecular targets, high-throughput screening is necessary to test the compounds in these large, natural, or synthetic libraries in *in vitro* or *in silico* assays and identify those that are highly specific and biologically active against a defined target of interest. Although many different assays are used in high-throughput screening, the general principle is that the specific molecular target or pathway of interest is combined systematically with each possible drug compound, then verified in manual or automated platforms for a positive result, or “hit.” The resulting “hits” would then be selected based on how strongly they interacted with the target, and additional confirmatory experiments would be performed to characterize the interaction.¹⁸⁹ The incorporation of automation, computational methods, and nanotechnology has allowed for enormously increased efficiency in both the development of combinatorial libraries and the high-throughput screening of potentially useful drugs. To illustrate the rapidly accelerating success of this method, a recent paper has reported achieving ultrahigh-throughput screening speeds of 100 million individual enzyme reactions in 10 h using only small reagent volumes; this represents a 1000-fold increase in speed and a 1 million-fold reduction in cost compared with

standard high-throughput techniques.¹⁹⁰ A negative consequence of the massive amounts of data generated from high- and ultrahigh-throughput technologies is that bottlenecks will occur at the level of drug development and optimization. In the past decade, more than 3000 expressed proteins have been described as potential drug targets, yet only 20% of those have been exploited commercially.¹⁹¹ Therefore, despite the optimism engendered by high-throughput screening and other advances in drug discovery and validation, numerous obstacles remain for the translation of this vast amount of data into clinical application.

FROM THE SIDE OF THE HOST: GENETIC PREDISPOSITION TO INFECTIOUS DISEASE AND PERSONALIZED MEDICINE

The human genome project brought about a conceptual shift in the focus of medicine, from treatment to the prediction of disease and prevention of side effects. The resulting principle is that if the manifestation of disease is influenced by interactions between the host and the infectious agent, then characterizing the host genetic correlates to disease would help to predict individual susceptibility to specific infection-related pathologies. The recent technological advances in sequencing and computational biology have facilitated the identification of host factors that predispose or affect the response to disease. Therefore, therapies can now be tailored according to the genetic makeup of the host and the characteristics of the microbe responsible for disease. Francis Collins suggested that genetics could be integrated with medicine either by using classic or “old” genetics to develop genetic tests for monogenic diseases and prenatal diagnostics or by using “new” genetics to predict patient responses and implement personalized medicine.¹⁹² The clinical application of personalized medicine principles may have seemed almost visionary a few years ago, but now this approach is used routinely in the treatment of some diseases. For example, genomic data are used to make accurate predictions about the susceptibility of a patient to disease or about how he/she will respond to treatment. Specifically, the application of personalized medicine is common in diseases such as cancer caused by some solid tumors or hematologic malignancies where patients are genetically tested to determine their likelihood of responding to therapy or having a serious adverse reaction to various drugs.^{193,194}

Personalized medicine has also been applied to the field of infectious disease to define host predisposition to some infections and to predict successful responses as well as adverse reactions to treatment. Evidence that certain infectious diseases have a genetic predisposition initially came from a study that showed an

approximately 6-fold increase in the risk of mortality from infectious diseases in adopted children if the biological parents also died from an infection.¹⁹⁵ Other groups later confirmed these results, and since then, the susceptibility to several infectious diseases, including *S. aureus* and *M. tuberculosis*, has been linked to rare defects in genes encoding effectors of the immune response.^{152,196-203} The completion of the Human Genome Project and the map of human genetic variation²⁰⁴ has promoted the evolution of high-density SNP arrays that analyze the association of certain polymorphisms with susceptibility to infectious disease, as well as the linkage disequilibrium, or correlation between 2 SNPs. These increasingly common genome-wide association studies (GWAS) have identified a series of genes implicated in the predisposition of different bacterial and viral diseases.^{205,206}

Although GWAS was first applied to the fields of rheumatic and autoimmune diseases, recently this approach has identified novel genetic loci involved in susceptibility to infectious disease. In addition, the use of GWAS is beginning to be demonstrated in the identification of markers that indicate the quality of a patient's response to treatment. For example, a recent GWAS showed that an SNP located upstream of the IL28B gene, which codes for type III interferon, is associated strongly with more than a 2-fold difference in response to anti-HCV drug treatment.²⁰⁷ Other studies showed a relationship between IL28B gene polymorphisms and the spontaneous clearance of acute HCV.²⁰⁸ Because HCV, which almost invariably leads to a chronic disease with liver failure in approximately 70% of individuals, is treated most effectively during the acute phase of the disease, this study was a significant step toward allowing physicians to predict spontaneous clearance accurately and to identify a subgroup of patients that will benefit from early treatment.²⁰⁸

Microarray technology has paved the way for the development of the field of transcriptomics, which is the large-scale profiling of all RNA transcripts in a cell to characterize gene expression (Fig 1). Transcriptomics has been applied widely to the study of the host response and identification of biomarkers after an infection or treatment. In this case, patterns of gene expression in human cells, most frequently peripheral blood mononuclear cells (PBMCs), are studied by hybridizing biological material to defined gene chips.²⁰⁹⁻²¹¹ With regard to infectious disease, microarray technology has made a huge contribution to our understanding of the changes evident in the transcriptome of humans during infection. Lempicki et al²¹² used microarray-derived gene expression patterns to demonstrate that the increased expression of interferon-stimulated genes in patient PBMCs could be one explanation for why

anti-HCV therapy in patients coinfecting with HCV and HIV has lower success rates. In another study, Jacobsen et al²¹³ used microarray technology to identify the biomarkers expressed during infection with *M. tuberculosis* that would allow for more accurate predictions regarding the outcome of infection. Thus, microarray technology has been broadly applicable in studies aimed at characterizing the infection profiles of a variety of microorganisms, including SARS-COV,²¹⁴ dengue,²¹⁵ influenza,²¹⁶ *S. aureus*,²¹⁷ and *S. enterica*.²¹⁸ The advent of next-generation deep and ultra-deep sequencing techniques has also improved the ability of researchers to identify splice variants and low abundance transcriptomic changes in response to infectious disease.⁴⁹ Data such as these can be used to inform improvements in the diagnosis and prognosis of disease, the design of antiviral and antibiotic therapies, and the construction of vaccines that can target susceptible populations more specifically.

The application of new genomic techniques in predicting the response of a patient to drug therapy based on his/her genetic profile (pharmacogenomics) has found practical application in identifying populations that are susceptible to short- and long-term toxicities of specific HIV antiretroviral drugs. One example is the reaction to abacavir, which is a nucleoside analog reverse transcriptase inhibitor. The major toxicity manifests as a hypersensitivity reaction that occurs in approximately 5% to 8% of recipients within 6 weeks of commencing therapy.²¹⁹ Mallal et al²²⁰ showed that patients carrying the *HLA-B5701* allele are more likely to experience this hypersensitivity reaction to abacavir. This marker had a negative predictive value of 100%, indicating that genetic tests can provide unequivocal information that can help to predict and prevent otherwise unpredictable drug reactions. Another example of the use of pharmacogenomics to guide therapy is in the case of the antiretroviral drug maraviroc, which is a CCR5-specific HIV entry inhibitor. This drug is effective only in patients who are infected with strains of HIV that use the CCR5 coreceptor for entry, but not in patients who are infected with the CCR4-tropic viruses. Therefore, by screening for the specific tropism of the strain of HIV with which the patient is infected, clinicians can target drug therapy more specifically, for example, by using maraviroc only in patients who are infected with the susceptible strain of HIV.²²¹ These results underline the success of GWAS and personalized medicine in predicting which populations should avoid treatment with specific drugs because they are more prone to side effects.

Although the clinical application of "omics" and systems biology holds great potential for medical discovery, the introduction of high-throughput technologies

has also introduced new questions and challenges regarding the transformation of the large volumes of data produced by these approaches into useful information for the development of diagnostic, prognostic, and therapeutic devices. The “large n, small p” paradigm, in which the number of study subjects (n) is higher than the variables (p) generated from the study itself, currently serves as a fundamental analytic principle for data modeling. With a shift in this ratio toward smaller n values compared with the large number of variables amassed through high-throughput data acquisition, it is becoming clear that this paradigm will need to be revised to preserve statistical accuracy.²²² To address this problem, some researchers have proposed changes in the mathematical principles used for data analysis, such as using continuous variables²²²⁻²²⁴ and nonlinear modeling.^{222,225,226} Some also recommend switching from deterministic to probability-based modeling strategies to account for the uncertainty that is typical of common clinical scenarios and to predict the relative probabilities of various outcomes given the same initial situation.^{222,227} In addition to the necessary evolution of data modeling strategies, another critical factor in the successful translation of research data is the capability to access and analyze large sets of data that originated from different sources and contain diverse types of information. There is, therefore, a need to develop methods that enable the construction of hybrid data sets that contain both clinical and biological data, and are organized in ways that are accessible to all translational researchers and compatible among different systems. Translation bioinformatics is an emerging area that proposes to offer a holistic approach to this issue by providing integrative methods for incorporating multiple data modalities and enabling the development of predictive models for therapeutic responses.²²⁸ Although a detailed description of the methods used for data integration and modeling is beyond the scope of our article, other articles by Yan,²²⁸ Kanehisa et al,²²⁹ and Sarkar et al²³⁰ offer a more complete review of the topic. These developments are still the object of research and far from completion, but as the data accumulate, only the full development of new analytic strategies will enable their full diagnostic and clinical use.

Here, we mentioned some advances in translational research that are applicable to infectious disease studies. There have been several other major achievements in translational research with other diseases; some have been especially relevant in the fields of cancer and neuroscience, specifically with regard to epigenomics and stem cell research. Although a thorough discussion of these fields is beyond the scope of this article, their impact on translational research has also been profound.

Regardless of the field of study or research technique, however, the promise of translational research remains hindered by a few key factors, which we will now discuss.

BRIDGING THE GAPS: TURNING SCIENTIFIC PROMISE INTO CLINICAL PRACTICE

Whereas the aforementioned studies illustrate the incredible promise of new technologies in translational research, their full potential remains largely untapped partly because of the high failure rates, long timelines, and costs of converting basic science research into clinically applicable methodologies. The recently proposed NCATS, a new NIH entity that would include the CTSA (and eliminate the NCRR), is an effort to circumvent these roadblocks. In the remainder of this review, we will discuss how time frames, costs, and poor success rates have persistently hindered the translation of basic research advances into clinically useful products, how the newly proposed NCATS aims to circumvent these roadblocks, and the additional challenges facing the field of translational medicine.^{8,11}

Failure rates. Over the past decade, the number of NMEs approved by the FDA and comparable regulatory bodies around the world has decreased progressively. In the second half of the 1990s, for example, the FDA approved more than 30 NMEs per year.²³¹ By contrast, the number of NMEs approved since 2001 has averaged slightly less than 23 per year, with the exception of 36 approved in 2004, despite the increasing number of potential therapeutic compounds identified from high-throughput technologies.²³¹⁻²³³ In addition to the decline in approvals, the number of NME applications submitted to the FDA over the past decade has decreased, with 2010 producing one of the lowest numbers in more than 15 years.²³³ Because of this decline in approvals and applications, as well as the increasing use of generic drugs, it is estimated that large cap pharmaceutical companies only can replace, on average, 26 cents to every dollar lost to patent expirations by new product revenues.²³⁴ According to current estimates, only 8% of NMEs will make it from candidate selection (in the preclinical stage) to a successful launch.^{232,235} This decline in drug development threatens to take not only an economic toll but also a medical one, as approximately 40% of the 2-year increase in life expectancy from 1986 to 2000 was attributable to the introduction of new drugs.²³⁶ This lack of new drugs is especially worrisome in the field of infectious disease, where the emergence of multidrug resistant pathogens has led to a situation in which formerly curable infections are now untreatable. Paradoxically, some pharmaceutical companies have indicated that they are curtailing or

curbing anti-infective research.¹⁰ A major reason for this is that anti-infective agents, which are only needed for a short time in smaller populations, make less economically attractive targets than drugs targeting chronic diseases that appear in the larger aging population (eg, antihypertensives and antidepressants).

At this time, approximately 85% of the failures in translating promising NMEs into clinical application occur during early clinical trials. According to a recent editorial,²³⁷ many of these NMEs fail because of outdated clinical trial designs. Drug testing requires the study of large sample populations to show statistically significant benefits compared with existing pharmaceuticals. The introduction of specific markers for personalized medicine makes it even more difficult to recruit patients who qualify for these tests. Therefore, many clinical trials take years to complete their enrollment, or they are cancelled because of the inability to recruit a sufficient number of subjects. These failures therefore contribute to the total costs and failure rates of translational research. As a result, many NMEs are never developed, and companies are removing more products from their pipelines than ever before.²³⁷

To circumvent this problem and increase the success rate of proposed NMEs, Francis Collins has proposed several priorities of the new NCATS. First, the NCATS will support studies that investigate broadly applicable, rather than disease-specific, target validation approaches (ie, targets that may be relevant to multiple diseases), as well as targets that are considered to be “too risky” for industry investment, such as those identified by GWAS. Second, the NCATS will encourage the use of “omics” and systems biology approaches in order to help design and validate new diagnostics and therapeutics for translation into clinical application. Third, the NCATS would also house the recently established NIH-FDA Regulatory Science Initiative, which was launched in 2010 with the goals of facilitating improvements in regulatory science and providing new insights to broadly benefit the field of translational research.^{8,238-240} Finally, together with the NIH-EPA-FDA Tox21 Consortium, the NCATS would support the use of new cell-based approaches as a means of pursuing preclinical toxicology studies.²⁴¹

Timelines. By current estimates, the average timeline for discovering and developing an NME is approximately 13.5 years. This estimate includes the regulatory review but not the time it takes to identify and validate the target pathway.²³² Time spent in clinical trials, unnecessary bureaucratic procedures associated with research and design management, and inefficient regulatory review have all been identified as contributors to this problem.^{9,232,242,243} To minimize these problems, the proposed NCATS would seek to

encourage the “reengineering of the translational process, from initial target identification to first-in-human application of small molecules, biologics, diagnostics, and devices” to streamline this process and attract pharmaceutical involvement.⁸ Additionally, the center would aim to act as an honest broker, hastening Institutional Review Board (IRB) approval for multicenter trials, further decreasing the time that NMEs spend in clinical trials and facilitating the rescue and repurposing of abandoned compounds. Last, the center would support collaborative efforts for the testing of new entities to act as a “proof of concept” in the hopes that promising new drugs would be picked up by industry.^{8,244}

Costs. The cost of a translational endeavor is inexorably linked to both high failure rates and long timelines. Indeed, the average cost to bring an NME to market is now estimated to be \$1.8 billion. This estimate seems to be the same regardless of whether the NME is produced by a big pharmaceutical company or small biotechnology firm.²³² The NCATS aim to decrease these costs in a number of ways. First, it would encourage the use of early human trials, also called “phase zero” clinical trials, in which “microdoses” of a new drug are administered in a small number of patients.²³⁷ In these trials, the absorption, tissue distribution, and toxicity of the drug are assessed by highly sensitive methodologies, such as molecular imaging, metabolomics, or proteomics.⁸ The feasibility of such endeavors has been demonstrated already by the Consortium for Metabonomic Toxicology at the Imperial College in London, which showed it is possible to design a prediction model for clinical toxicity using metabolomic data.¹⁰⁴ Another way in which the NCATS would decrease development costs is by encouraging innovative models, including human tissue biobanks, stem cell models, and tissue-engineered organoids. These models would potentially allow researchers to decrease or omit the use of animal models, which have been criticized as both costly and not accurately predictive of efficacy in humans.^{8,237} Finally, the NCATS would encourage the development of technologies aimed at biomarker identification. These have been used already by researchers to design a new style of “adaptive” clinical trial, in which the study design changes as new data are collected, allowing patients who are more likely to respond to be assigned to treatment with a given compound.²³⁷ Beyond encouraging the development of new technologies aimed at decreasing costs, the NCATS would promote cross-collaborations between academic institutions, biotechnology firms, philanthropic organizations, and pharmaceutical companies to split research and development costs.

Last, the proposed Cures Acceleration Network, which will be assessed for congressional funding in the next fiscal year, would provide the NIH with the ability to award up to \$15 billion per year to academic and private consortia to direct the development of certain especially promising therapeutics.⁸

ARE NCATS THE FUTURE OF TRANSLATIONAL RESEARCH?

Although the need to encourage the development of new drugs and therapeutics is clear, some have expressed concerns that the proposed NCATS represents a change in priority for the NIH away from basic, discovery-based science and toward application-focused research, thus neglecting the fact that basic science lays the foundation for translational research. Indeed, this has been a criticism since the development of the CTSA.²⁴⁵ Another criticism is that the support of GWAS and other high-throughput methodologies causes a dramatic shift from research that is hypothesis based to that which is data driven, placing an overemphasis on screening for drugs and biomarkers at the expense of discovery-based science.²⁴⁶ For example, studies based on high-throughput technologies are typically not the result of experiments designed to test a hypothesis. Rather, they aim to identify systematically all the molecular effectors that are involved in a specific system in the most comprehensive and efficient way possible. To address this critique, some authors have added an extra step of complexity to their high-throughput methods by determining a mechanistic function for the molecules they have identified, switching in this way from “fishing expedition” to discovery-based research. In the end, despite assertions that the establishment of the NCATS will reinforce, rather than detract from basic science, many investigators continue to have concerns, mostly related to money. In today’s climate of funding cuts, some wonder: Will it be possible to receive funding for more basic (ie, less obviously applicable) research? Will the emphasis on collaboration and multidisciplinary approaches shift most of the funding toward multicenter groups of established scientists, making it difficult for new investigators to survive? Will the current emphasis on clinical applicability result in a narrowing of ideas and perspectives that will result in most funding being allocated to well-established scientists pursuing traditional approaches? Beyond these concerns regarding a change in the focus of NIH, concerns regarding policy and infrastructure exist, such as whether funding NCATS programs will take money away from other programs that were previously housed within the NCR. As the NCR currently awards funds of approximately \$1.25 billion annually, this is a substantial concern.²⁴⁷

Although none of the NCR programs are currently slated to be discontinued, the perceived lack of transparency surrounding the establishment of the NCATS, the dissolution of the NCR, and where programs formerly housed within the NCR will now be located has resulted in persistent scrutiny from both NIH stakeholders and Congress.²⁴⁸⁻²⁵⁰

Perhaps the greatest issue of debate is where to draw the line between public and private industry. Many criticize the proposed NCATS because they believe it is designed to help companies develop new drugs by decreasing their financial risks, and it is believed that it will do so by using taxpayers’ money. Also, they argue that the NIH only has experience with basic science research and not with drug development. Indeed, some have argued that the decline in research and development productivity is not related to increasing costs, because public investment in the pharmaceutical industry has increased proportionally since 1970, nor is it to the result of time or attritional rates, which have been steady since the 1970s. Instead, they argue that it may be caused by poor prioritization and mismanagement within the pharmaceutical industry.⁹ Specifically, changes in management techniques in the pharmaceutical industry hierarchy have led to corporate policies that discourage innovation in favor of cost cutting and risk minimization.^{9,246} This together with a shift of control from scientists and researchers to marketers, a focus on short-term profits and fast sales growth, and the discontinuation of compounds for nontechnical reasons, have led to the decreased translation of scientific discoveries to clinical applications.⁹ As such, some have argued that the development of the NCATS might make it even less likely that the pharmaceutical industry will try to solve these problems. Ultimately, although the NCATS might serve as a champion for key therapeutic compounds, the pharmaceutical industry itself will need to reform certain practices if the pace of drug development is to keep up with the pace of translational research.

Translational research has made incredible progress in recent years, both with the development of new techniques and with the expansion of more holistic ways of viewing the data that stem from those approaches. For the most part, the discovery of novel technologies, the development of new infrastructures, and the training of budding scientists have supported this evolution. The transition is not complete and roadblocks still exist on the path to scientific progress. It remains to be seen whether the newly proposed NCATS, which has raised as many objections as it has hopes, will be the answer to these problems. What is evident, however, is that translational research must be reprioritized and bolstered in a way that increases the speed at which

promising drugs and technologies move through the pipeline into clinical application, while at the same time refraining from stifling the breadth of ideas and creativity of basic researchers, which is necessary for the progression of science itself.

We thank Ann Beeder, MD for useful discussion.

REFERENCES

- [No authors listed] Lost in clinical translation. *Nat Med* 2004; 10:879.
- Woolf SH. The meaning of translational research and why it matters. *JAMA* 2008;299:211–3.
- National Center for Research Resources. Clinical and Translational Science Awards. 2011.
- National Institutes of Health. National Center for Research Resources. Bethesda, Md: NIH, 2011.
- Steele SJ. Working with the CTSA Consortium: what we bring to the table. *Sci Transl Med* 2010;2:63mr5.
- Pienta KJ, Spork AL, Scheske J. The Clinical and Translational Science Awards (CTSAs) are transforming the way academic medical institutions approach translational research: the University of Michigan Experience. *Clin Transl Sci* 2011;4:233–5.
- Schubert C. Four years on, clinical partnerships program proves worth. *Nat Med* 2010;16:612–3.
- Collins FS. Reengineering translational science: the time is right. *Sci Transl Med* 2011;3:90cm17.
- Cuatrecasas P. Drug discovery in jeopardy. *J Clin Invest* 2006; 116:2837–42.
- Spellberg B, Powers JH, Brass EP, Miller LG, Edwards JE Jr. Trends in antimicrobial drug development: implications for the future. *Clin Infect Dis* 2004;38:1279–86.
- Harris G. Federal research center will help develop medicines. *New York Times* January 23, 2011;A1.
- Office of Inspector General. The orphan drug act: implementation and impact. Washington, DC: Department of Health and Human Services, 1983;13.
- Watson JD, Cook-Deegan RM. Origins of the human genome project. *FASEB J* 1991;5:8–11.
- Proteome Research: New Frontiers in Functional Genomics. Berlin, Germany: Springer, 1997.
- Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. *Trends Biotechnol* 1998; 16:373–8.
- Hutchison CA 3rd. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res* 2007;35:6227–37.
- Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 1977;74:560–4.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;74: 5463–7.
- Gulick RM. Antiretroviral treatment 2010: progress and controversies. *J Acquir Immune Defic Syndr* 2010;55:S43–8.
- Deng L, Tang H. Hepatitis B virus drug resistance to current nucleos(t)ide analogs: mechanisms and mutation sites. *Hepatology* 2011;41:1017–24.
- Lin CL, Kao JH. The clinical implications of hepatitis B virus genotype: recent advances. *J Gastroenterol Hepatol* 2011;26: 123–30.
- Vivekanandan P, Singh OV. Molecular methods in the diagnosis and management of chronic hepatitis B. *Expert Rev Mol Diagn* 2010;10:921–35.
- Hiraga N, Imamura M, Abe H, et al. Rapid emergence of telaprevir resistant hepatitis C virus strain from wildtype clone in vivo. *Hepatology* 2011;54:781–8.
- Susser S, Vermehren J, Forestier N, et al. Analysis of long-term persistence of resistance mutations within the hepatitis C virus NS3 protease after treatment with telaprevir or boceprevir. *J Clin Virol* 2011;52:321–7.
- Kieffer TL, Kwong AD, Picchio GR. Viral resistance to specifically targeted antiviral therapies for hepatitis C (STAT-Cs). *J Antimicrob Chemother* 2010;65:202–12.
- Rosen HR. Clinical practice. Chronic hepatitis C infection. *N Engl J Med* 2011;364:2429–38.
- Duan S, Boltz DA, Li J, et al. Novel genotyping and quantitative analysis of neuraminidase inhibitor resistance-associated mutations in influenza A viruses by single-nucleotide polymorphism analysis. *Antimicrob Agents Chemother* 2011;55:4718–27.
- Tong SY, Dakh F, Hurt AC, et al. Rapid detection of the H275Y oseltamivir resistance mutation in influenza A/H1N1 2009 by single base pair RT-PCR and high-resolution melting. *PLoS One* 2011;6:e21446.
- Wijaya L, Chua YY, Cui L, Chan K, Tan BH. Intravenous zanamivir in critically ill patients due to pandemic 2009 (H1N1) influenza A virus. *Singapore Med J* 2011;52:481–5.
- Grondahl B, Puppe W, Hoppe A, et al. Rapid identification of nine microorganisms causing acute respiratory tract infections by single-tube multiplex reverse transcription-PCR: feasibility study. *J Clin Microbiol* 1999;37:1–7.
- Edwards MC, Gibbs RA. Multiplex PCR: advantages, development, and applications. *PCR Methods Appl* 1994;3:S65–75.
- Uttamchandani M, Neo JL, Ong BN, Moochhala S. Applications of microarrays in pathogen detection and biodefence. *Trends Biotechnol* 2009;27:53–61.
- Wang D, Coscoy L, Zylberberg M, et al. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A* 2002;99:15687–92.
- Wang D, Urisman A, Liu YT, et al. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 2003;1:E2.
- Palacios G, Quan PL, Jabado OJ, et al. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg Infect Dis* 2007;13:73–81.
- Anderson NG, Gerin JL, Anderson NL. Global screening for human viral pathogens. *Emerg Infect Dis* 2003;9:768–74.
- Lipkin WI. Microbe hunting. *Microbiol Mol Biol Rev* 2010;74: 363–77.
- Rota PA, Oberste MS, Monroe SS, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003;300:1394–9.
- Mihindukulasuriya KA, Wu G, St Leger J, Nordhausen RW, Wang D. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J Virol* 2008;82:5084–8.
- Kistler A, Avila PC, Rouskin S, et al. Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. *J Infect Dis* 2007;196:817–25.
- Chiu CY, Greninger AL, Kanada K, et al. Identification of cardioviruses related to Theiler's murine encephalomyelitis virus in human infections. *Proc Natl Acad Sci U S A* 2008;105:14124–9.
- Connell C, Fung S, Heiner C, et al. Automated DNA-sequence analysis. *BioTechniques* 1987;5:342–8.
- Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 2000;18:630–4.
- Stein RA. Next-generation sequencing update. *Genetic Engineering & Biotechnology News* 2008;28.

45. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; 437:376–80.
46. Ansorge WJ. Next-generation DNA sequencing techniques. *N Biotechnol* 2009;25:195–203.
47. Nakamura S, Yang CS, Sakon N, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 2009;4:e4219.
48. Sorber K, Chiu C, Webster D, et al. The long march: a sample preparation technique that enhances contig length and coverage by high-throughput short-read sequencing. *PLoS One* 2008;3: e3495.
49. Rothberg JM, Leamon JH. The development and impact of 454 sequencing. *Nat Biotechnol* 2008;26:1117–24.
50. Palacios G, Druce J, Du L, et al. A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med* 2008;358: 991–8.
51. Briese T, Paweska JT, McMullan LK, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog* 2009;5:e1000455.
52. Greninger AL, Runkel C, Chiu CY, et al. The complete genome of klassevirus - a novel picornavirus in pediatric stool. *Virology* 2009;6:82.
53. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008;319:1096–100.
54. Chin CS, Sorenson J, Harris JB, et al. The origin of the Haitian cholera outbreak strain. *N Engl J Med* 2011;364:33–42.
55. Garten RJ, Davis CT, Russell CA, et al. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* 2009;325:197–201.
56. Bell JF, Wright JT, Habel K. Rabies vaccine freed of the factor causing allergic encephalitis. *Proc Soc Exp Biol Med* 1949;70: 457–61.
57. Alexander LN, Seward JF, Santibanez TA, et al. Vaccine policy changes and epidemiology of poliomyelitis in the United States. *JAMA* 2004;292:1696–701.
58. Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* 2001;19:2688–91.
59. Danzig L. Reverse vaccinology—in search of a genome-derived meningococcal vaccine. *Vaccine* 2006;24:11–2.
60. Yogev R, Tan T. Meningococcal disease: the advances and challenges of meningococcal disease prevention. *Hum Vaccin* 2011; 7:828–37.
61. Pizza M, Scarlato V, Maignani V, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 2000;287:1816–20.
62. Tettelin H, Medini D, Donati C, Maignani V. Towards a universal group B *Streptococcus* vaccine using multistrain genome analysis. *Expert Rev Vaccines* 2006;5:687–94.
63. Subbarao K, Katz JM. Influenza vaccines generated by reverse genetics. *Curr Top Microbiol Immunol* 2004;283:313–42.
64. Steel J. New strategies for the development of H5N1 subtype influenza vaccines: progress and challenges. *BioDrugs* 2011;25: 285–98.
65. Wacheck V, Egorov A, Groiss F, et al. A novel type of influenza vaccine: safety and immunogenicity of replication-deficient influenza virus created by deletion of the interferon antagonist NS1. *J Infect Dis* 2010;201:354–62.
66. Emonet SE, Urata S, de la Torre JC. Arenavirus reverse genetics: new approaches for the investigation of arenavirus biology and development of antiviral strategies. *Virology* 2011;411:416–25.
67. McGettigan JP. Experimental rabies vaccines for humans. *Expert Rev Vaccines* 2010;9:1177–86.
68. Horimoto T, Kawaoka Y. Designing vaccines for pandemic influenza. *Curr Top Microbiol Immunol* 2009;333:165–76.
69. Richt JA, Garcia-Sastre A. Attenuated influenza virus vaccines with modified NS1 proteins. *Curr Top Microbiol Immunol* 2009;333:177–95.
70. Bouloy M, Flick R. Reverse genetics technology for Rift Valley fever virus: current and future applications for the development of therapeutics and vaccines. *Antiviral Res* 2009;84:101–18.
71. Billeter MA, Naim HY, Udem SA. Reverse genetics of measles virus and resulting multivalent recombinant vaccines: applications of recombinant measles viruses. *Curr Top Microbiol Immunol* 2009;329:129–62.
72. Draper SJ, Heeney JL. Viruses as vaccine vectors for infectious diseases and cancer. *Nat Rev Microbiol* 2010;8:62–73.
73. Rollier CS, Reyes-Sandoval A, Cottingham MG, Ewer K, Hill AV. Viral vectors as vaccine platforms: deployment in sight. *Curr Opin Immunol* 2011;23:377–82.
74. Sofer-Podesta C, Ang J, Hackett NR, et al. Adenovirus-mediated delivery of an anti-V antigen monoclonal antibody protects mice against a lethal *Yersinia pestis* challenge. *Infect Immun* 2009;77: 1561–8.
75. Kasuya K, Boyer JL, Tan Y, et al. Passive immunotherapy for anthrax toxin mediated by an adenovirus expressing an anti-protective antigen single-chain antibody. *Mol Ther* 2005;11: 237–44.
76. Sparkman D. Mass spectrometry desk reference. 2nd ed. Pittsburgh, Pa: Global View Publishing, 2006.
77. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198–207.
78. Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol* 1999;17:121–7.
79. Roe MR, Griffin TJ. Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes. *Proteomics* 2006;6:4678–87.
80. Brewis IA, Brennan P. Proteomics technologies for the global identification and quantification of proteins. *Adv Protein Chem Struct Biol* 2010;80:1–44.
81. Malmstrom L, Malmstrom J, Aebersold R. Quantitative proteomics of microbes: principles and applications to virulence. *Proteomics* 2011;11:2947–56.
82. Alban A, David SO, Bjorkesten L, et al. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* 2003;3:36–44.
83. Ly L, Wasinger VC. Protein and peptide fractionation, enrichment and depletion: tools for the complex proteome. *Proteomics* 2011;11:513–34.
84. Carr PW, Stoll DR, Wang X. Perspectives on recent advances in the speed of high-performance liquid chromatography. *Anal Chem* 2011;83:1890–900.
85. Link AJ, Eng J, Schieltz DM, et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999;17: 676–82.
86. Eng JKM, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences is a protein database. *J Am Soc Mass Spectrom* 1994;5:976–89.
87. Faherty BK, Gerber SA. MacroQUEST: efficient candidate-centric searching and high-resolution correlation analysis for large-scale proteomics data sets. *Anal Chem* 2010;82:6821–9.
88. Frohlich T, Arnold GJ. Proteome research based on modern liquid chromatography–tandem mass spectrometry: separation,

- identification and quantification. *J Neural Transm* 2006;113:973–94.
89. Chiou SH, Wu CY. Clinical proteomics: current status, challenges, and future perspectives. *Kaohsiung J Med Sci* 2011;27:1–14.
 90. Elliott MH, Smith DS, Parker CE, Borchers C. Current trends in quantitative proteomics. *J Mass Spectrom* 2009;44:1637–60.
 91. Vogel C, Marcotte EM. Calculating absolute and relative protein abundance from mass spectrometry-based protein expression data. *Nat Protoc* 2008;3:1444–51.
 92. Matthiesen R, Carvalho AS. Methods and algorithms for relative quantitative proteomics by mass spectrometry. *Methods Mol Biol* 2010;593:187–204.
 93. Mueller LN, Rinner O, Schmidt A, et al. SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* 2007;7:3470–80.
 94. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* 2008;4:222.
 95. Kuzyk MA, Smith D, Yang J, et al. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol Cell Proteomics* 2009;8:1860–77.
 96. Lasonder E, Ishihama Y, Andersen JS, et al. Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 2002;419:537–42.
 97. Florens L, Washburn MP, Raine JD, et al. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 2002;419:520–6.
 98. Rhee KY, Erdjument-Bromage H, Tempst P, Nathan CF. S-nitroso proteome of *Mycobacterium tuberculosis*: enzymes of intermediary metabolism and antioxidant defense. *Proc Natl Acad Sci U S A* 2005;102:467–72.
 99. Popova TG, Turell MJ, Espina V, et al. Reverse-phase phosphoproteome analysis of signaling pathways induced by Rift valley fever virus in human small airway epithelial cells. *PLoS One* 2010;5:e13805.
 100. Chan EY, Qian WJ, Diamond DL, et al. Quantitative analysis of human immunodeficiency virus type 1-infected CD4+ cell proteome: dysregulated cell cycle progression and nuclear transport coincide with robust virus production. *J Virol* 2007;81:7571–83.
 101. Feng X, Zhang J, Chen WN, Ching CB. Proteome profiling of Epstein-Barr virus infected nasopharyngeal carcinoma cell line: identification of potential biomarkers by comparative iTRAQ-coupled 2D LC/MS-MS analysis. *J Proteomics* 2011;74:567–76.
 102. Malmstrom J, Beck M, Schmidt A, et al. Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* 2009;460:762–5.
 103. Francois P, Scherl A, Hochstrasser D, Schrenzel J. Proteomic approaches to study *Staphylococcus aureus* pathogenesis. *J Proteomics* 2010;73:701–8.
 104. Lindon JC, Holmes E, Nicholson JK. Metabonomics in pharmaceutical R&D. *Febs J* 2007;274:1140–51.
 105. Nicholson JK, Wilson ID, Lindon JC. Pharmacometabonomics as an effector for personalized medicine. *Pharmacogenomics* 2011;12:103–11.
 106. van der Werf MJ, Jellema RH, Hankemeier T. Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. *J Ind Microbiol Biotechnol* 2005;32:234–52.
 107. Vinayavekhin N, Homan EA, Saghatelian A. Exploring disease through metabolomics. *ACS Chem Biol* 2010;5:91–103.
 108. Henderson JP, Crowley JR, Pinkner JS, et al. Quantitative metabolomics reveals an epigenetic blueprint for iron acquisition in uropathogenic *Escherichia coli*. *PLoS Pathog* 2009;5:e1000305.
 109. Munger J, Bennett BD, Parikh A, et al. Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nat Biotechnol* 2008;26:1179–86.
 110. Aoki-Kinoshita KF. An introduction to bioinformatics for glycomics research. *PLoS Comput Biol* 2008;4:e1000075.
 111. Wenk MR. The emerging field of lipidomics. *Nat Rev Drug Discov* 2005;4:594–610.
 112. Wenk MR. Lipidomics: new tools and applications. *Cell* 2010;143:888–95.
 113. Wenk MR. Lipidomics of host-pathogen interactions. *FEBS Lett* 2006;580:5541–51.
 114. Srivastava S. Move over proteomics, here comes glycomics. *J Proteome Res* 2008;7:1799.
 115. Hart GW, Copeland RJ. Glycomics hits the big time. *Cell* 2010;143:672–6.
 116. Hameed S, Dhamgaye S, Singh A, Goswami SK, Prasad R. Calcineurin signaling and membrane lipid homeostasis regulates iron mediated multidrug resistance mechanisms in *Candida albicans*. *PLoS One* 2011;6:e18684.
 117. Brugger B, Glass B, Haberkant P, et al. The HIV lipidome: a raft with an unusual composition. *Proc Natl Acad Sci U S A* 2006;103:2641–6.
 118. Brugger B, Krautkramer E, Tibroni N, et al. Human immunodeficiency virus type 1 Nef protein modulates the lipid composition of virions and host cell membrane microdomains. *Retrovirology* 2007;4:70.
 119. Liu ST, Sharon-Friling R, Ivanova P, et al. Synaptic vesicle-like lipidome of human cytomegalovirus virions reveals a role for SNARE machinery in virion egress. *Proc Natl Acad Sci U S A* 2011;108:12869–74.
 120. Diamond DL, Syder AJ, Jacobs JM, et al. Temporal proteome and lipidome profiles reveal hepatitis C virus-associated reprogramming of hepatocellular metabolism and bioenergetics. *PLoS Pathog* 2010;6:e1000719.
 121. Merz A, Long G, Hiet MS, et al. Biochemical and morphological properties of hepatitis C virus particles and determination of their lipidome. *J Biol Chem* 2011;286:3018–32.
 122. van der Meer-Janssen YP, van Galen J, Batenburg JJ, Helms JB. Lipids in host-pathogen interactions: pathogens exploit the complexity of the host cell lipidome. *Prog Lipid Res* 2010;49:1–26.
 123. Song X, Yu H, Chen X, et al. A sialylated glycan microarray reveals novel interactions of modified sialic acids with proteins and viruses. *J Biol Chem* 2011;286:31610–22.
 124. Yu X, Coulson BS, Fleming FE, et al. Novel structural insights into rotavirus recognition of ganglioside glycan receptors. *J Mol Biol* 2011;413:929–39.
 125. Toman R, Skultety L, Ihnatko R. *Coxiella burnetii* glycomics and proteomics—tools for linking structure to function. *Ann N Y Acad Sci* 2009;1166:67–78.
 126. Krishnamoorthy L, Bess JW Jr, Preston AB, Nagashima K, Mahal LK. HIV-1 and microvesicles from T cells share a common glycome, arguing for a common origin. *Nat Chem Biol* 2009;5:244–50.
 127. Bhak J. History of omics. Available from: <http://omics.org>. Accessed September 28, 2011.
 128. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
 129. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001;291:1304–51.
 130. Relman DA, Falkow S. The meaning and impact of the human genome sequence for microbiology. *Trends Microbiol* 2001;9:206–8.
 131. Relman DA. New technologies, human-microbe interactions, and the search for previously unrecognized pathogens. *J Infect Dis* 2002;186:S254–8.

132. Hooper LV, Gordon JI. Commensal host-bacterial relationships in the gut. *Science* 2001;292:1115–8.
133. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. *Science* 2005;308:1635–8.
134. National Institutes of Health. NIH launches human microbiome project. Bethesda, MD: National Institutes of Health, 2007.
135. National Institutes of Health. Human microbiome project. Bethesda, MD: National Institutes of Health, 2007.
136. Peterson J, Garges S, Giovanni M, et al. The NIH human microbiome project. *Genome Res* 2009;19:2317–23.
137. Gao Z, Tseng CH, Strober BE, Pei Z, Blaser MJ. Substantial alterations of the cutaneous bacterial biota in psoriatic lesions. *PLoS One* 2008;3:e2719.
138. Nelson DE, Van Der Pol B, Dong Q, et al. Characteristic male urine microbiomes associate with asymptomatic sexually transmitted infection. *PLoS One* 2010;5:e14116.
139. Frank DN, Robertson CE, Hamm CM, et al. Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* 2011;17:179–84.
140. Yang L, Lu X, Nossa CW, et al. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. *Gastroenterology* 2009;137:588–97.
141. Ichinohe T, Pang IK, Kumamoto Y, et al. Microbiota regulates immune defense against respiratory tract influenza A virus infection. *Proc Natl Acad Sci U S A* 2011;108:5354–9.
142. Langmead B, Schatz MC, Lin J, Pop M, Salzberg SL. Searching for SNPs with cloud computing. *Genome Biol* 2009;10:R134.
143. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–6.
144. Storch G, Weinstock G, Sodergren E, et al. The human virome in children and its relationship to febrile illness. *Human Microbiome Project* 2010.
145. Willner D, Furlan M, Haynes M, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One* 2009;4:e7370.
146. Minot S, Sinha R, Chen J, Li H, et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 2011;21:1616–25.
147. Ghannoum MA, Jurevic RJ, Mukherjee PK, et al. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog* 2010;6:e1000713.
148. Huang SH, Triche T, Jong AY. Infectomics: genomics and proteomics of microbial infections. *Funct Integr Genomics* 2002;1:331–44.
149. Goh KI, Cusick ME, Valle D, et al. The human disease network. *Proc Natl Acad Sci U S A* 2007;104:8685–90.
150. Navratil V, de Chasse B, Combe CR, Lotteau V. When the human viral infectome and disease networks collide: towards a systems biology platform for the aetiology of human diseases. *BMC Syst Biol* 2011;5:13.
151. Aderem A, Adkins JN, Ansong C, et al. A systems biology approach to infectious disease research: innovating the pathogen-host research paradigm. *MBio* 2011;2: e00325–10.
152. Haverkamp MH, van Dissel JT, Holland SM. Human host genetic factors in nontuberculous mycobacterial infection: lessons from single gene disorders affecting innate and adaptive immunity and lessons from molecular defects in interferon-gamma-dependent signaling. *Microbes Infect* 2006;8:1157–66.
153. Kidder BL, Hu G, Zhao K. ChIP-Seq: technical considerations for obtaining high-quality data. *Nat Immunol* 2011;12:918–22.
154. National Institute of Allergy and Infectious Disease. TB systems biology program. Bethesda, Md: NIAID, 2011.
155. National Institute of Allergy and Infectious Disease. Systems virology. Bethesda, Md: NIAID, 2011.
156. National Institute of Allergy and Infectious Disease. Systems influenza. Bethesda, Md: NIAID, 2011.
157. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005;4:17.
158. Li C, Bankhead A 3rd, Eisfeld AJ, et al. Host regulatory network response to infection with highly pathogenic H5N1 avian influenza virus. *J Virol* 2011;85:10955–67.
159. Bradel-Tretheway BG, Kelley Z, Chakraborty-Sett S, et al. The human H5N1 influenza A virus polymerase complex is active in vitro over a broad range of temperatures, in contrast to the WSN complex, and this property can be attributed to the PB2 subunit. *J Gen Virol* 2008;89:2923–32.
160. Peng X, Gralinski L, Armour CD, et al. Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *MBio* 2010;1:e00206–10.
161. National Institute of Allergy and Infectious Disease. Center for Systems Biology of Enteropathogens. Bethesda, Md: NIAID, 2011. Available from: <http://www.sysbep.org>. Accessed September 28, 2011.
162. Schellenberger J, Que R, Fleming RM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 2011;6:1290–307.
163. Becker SA, Feist AM, Mo ML, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2007;2:727–38.
164. Yoon H, Ansong C, McDermott JE, et al. Systems analysis of multiple regulator perturbations allows discovery of virulence factors in Salmonella. *BMC Syst Biol* 2011;5:100.
165. Buchko GW, Niemann G, Baker ES, et al. A multi-pronged search for a common structural motif in the secretion signal of Salmonella enterica serovar Typhimurium type III effector proteins. *Mol Biosyst* 2010;6:2448–58.
166. Germain R. Ron Germain: towards a grand unified theory. Interview by Amy Maxmen. *J Exp Med* 2010;207:266–7.
167. National Institute of Allergy and Infectious Disease. Laboratory of Systems Biology (LSB). Bethesda, Md: NIAID, 2011. Available from: <http://www.niaid.nih.gov/labsandresources/labs/about/labs/lsb/pages/default.aspx?wt.ac=bcLsb>. Accessed September 30, 2011.
168. Miles JJ, Bulek AM, Cole DK, et al. Genetic and structural basis for selection of a ubiquitous T cell receptor deployed in Epstein-Barr virus infection. *PLoS pathogens* 2010;6:e1001198.
169. Yang L, Zhang J, Ho B, Ding JL. Histidine-mediated pH-sensitive regulation of M-ficolin:GlcNAc binding activity in innate immunity examined by molecular dynamics simulations. *PLoS One* 2011;6:e19647.
170. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One* 2010;5:e9862.
171. Mitha F, Lucas TA, Feng F, Kepler TB, Chan C. The multiscale systems immunology project: software for cell-based immunological simulation. *Source Code Biol Med* 2008;3:6.
172. Pulendran B. Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology. *Nat Rev Immunol* 2009;9:741–7.
173. Querec TD, Akondy RS, Lee EK, et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. *Nat Immunol* 2009;10:116–25.
174. Nakaya HI, Wrannert J, Lee EK, et al. Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol* 2011;12:786–95.

175. Santamaria R, Rizzetto L, Bromley M, et al. Systems biology of infectious diseases: a focus on fungal infections. *Immunobiology* 2011;216:1212–27.
176. Liang L, Tan X, Juarez S, et al. Systems biology approach predicts antibody signature associated with *Brucella melitensis* infection in humans. *J Proteome Res* 2011;10:4813–24.
177. Adams LG, Khare S, Lawhon SD, et al. Enhancing the role of veterinary vaccines reducing zoonotic diseases of humans: linking systems biology with vaccine development. *Vaccine* 2011;29:7197–206.
178. Rappuoli R, Black S, Lambert PH. Vaccine discovery and translation of new vaccine technology. *Lancet* 2011;378:360–8.
179. Pulendran B, Li S, Nakaya HI. Systems vaccinology. *Immunity* 2010;33:516–29.
180. Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 1985;228:1315–7.
181. Wang LF, Yu M. Epitope identification and discovery using phage display libraries: applications in vaccine development and diagnostics. *Curr Drug Targets* 2004;5:1–15.
182. Khurana S, Wu J, Verma N, et al. H5N1 virus-like particle vaccine elicits cross-reactive neutralizing antibodies that preferentially bind to the oligomeric form of influenza virus hemagglutinin in humans. *J Virol* 2011;85:10945–54.
183. Barbas CF 3rd, Kang AS, Lerner RA, Benkovic SJ. Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc Natl Acad Sci U S A* 1991;88:7978–82.
184. Sun L, Lu X, Li C, et al. Generation, characterization and epitope mapping of two neutralizing and protective human recombinant antibodies against influenza A H5N1 viruses. *PLoS One* 2009;4:e5476.
185. Nelson AL, Dhimoleda E, Reichert JM. Development trends for human monoclonal antibody therapeutics. *Nat Rev Drug Discov* 2010;9:767–74.
186. Migone TS, Subramanian GM, Zhong J, et al. Raxibacumab for the treatment of inhalational anthrax. *N Engl J Med* 2009;361:135–44.
187. Galloway WR, Bender A, Welch M, Spring DR. The discovery of antibacterial agents using diversity-oriented synthesis. *Chem Commun (Camb)* 2009;2446–62.
188. Bratkovic T. Progress in phage display: evolution of the technique and its applications. *Cell Mol Life Sci* 2010;67:749–67.
189. Pereira DA, Williams JA. Origin and evolution of high throughput screening. *Br J Pharmacol* 2007;152:53–61.
190. Agresti JJ, Antipov E, Abate AR, et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci U S A* 2010;107:4004–9.
191. Perrior T. Overcoming bottlenecks in drug discovery. *Drug Discov World* 2010;29–33.
192. Collins FS. The Genome Project and human health. *FASEB J* 1991;5:77.
193. Schilsky RL. Personalized medicine in oncology: the future is now. *Nat Rev Drug Discov* 2010;9:363–6.
194. Cronin M, Ross JS. Comprehensive next-generation cancer genome sequencing in the era of targeted therapy and personalized oncology. *Biomark Med* 2011;5:293–305.
195. Sorensen TI, Nielsen GG, Andersen PK, Teasdale TW. Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med* 1988;318:727–32.
196. Casanova JL, Abel L, Quintana-Murci L. Human TLRs and IL-1Rs in host defense: natural insights from evolutionary, epidemiological, and clinical genetics. *Annu Rev Immunol* 2011;29:447–91.
197. Guo Y, Audry M, Ciancanelli M, et al. Herpes simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant in protective immunity. *J Exp Med* 2011;208:2083–98.
198. Picard C, Casanova JL, Puel A. Infectious diseases in patients with IRAK-4, MyD88, NEMO, or IkappaBalpha deficiency. *Clin Microbiol Rev* 2011;24:490–7.
199. Tabarsi P, Marjani M, Mansouri N, et al. Lethal tuberculosis in a previously healthy adult with IL-12 receptor deficiency. *J Clin Immunol* 2011;31:537–9.
200. Burbelo PD, Browne SK, Sampaio EP, et al. Anti-cytokine autoantibodies are associated with opportunistic infection in patients with thymic neoplasia. *Blood* 2010;116:4848–58.
201. Holland SM. Interferon gamma, IL-12, IL-12R and STAT-1 immunodeficiency diseases: disorders of the interface of innate and adaptive immunity. *Immunol Res* 2007;38:342–6.
202. Vinh DC, Schwartz B, Hsu AP, et al. Interleukin-12 receptor beta1 deficiency predisposing to disseminated Coccidioidomycosis. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 2011;52:e99–102.
203. Holland SM, DeLeo FR, Elloumi HZ, et al. STAT3 mutations in the hyper-IgE syndrome. *N Engl J Med* 2007;357:1608–19.
204. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–320.
205. Vannberg FO, Chapman SJ, Hill AV. Human genetic susceptibility to intracellular pathogens. *Immunol Rev* 2011;240:105–16.
206. Oki NO, Motsinger-Reif AA, Antas PR, et al. Novel human genetic variants associated with extrapulmonary tuberculosis: a pilot genome wide association study. *BMC Res Notes* 2011;4:28.
207. Rauch A, Kutalik Z, Descombes P, et al. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* 2010;138:1338–45. 45 e1-7.
208. Thomas DL, Thio CL, Martin MP, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. *Nature* 2009;461:798–801.
209. Poustka A, Pohl T, Barlow DP, et al. Molecular approaches to mammalian genetics. *Cold Spring Harb Symp Quant Biol* 1986;51:131–9.
210. Cantor CR, Mirzabekov A, Southern E. Report on the sequencing by hybridization workshop. *Genomics* 1992;13:1378–83.
211. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–70.
212. Lempicki RA, Polis MA, Yang J, et al. Gene expression profiles in hepatitis C virus (HCV) and HIV coinfection: class prediction analyses before treatment predict the outcome of anti-HCV therapy among HIV-coinfecting persons. *J Infect Dis* 2006;193:1172–7.
213. Jacobsen M, Reipsilber D, Gutschmidt A, et al. Candidate biomarkers for discrimination between infection and disease caused by *Mycobacterium tuberculosis*. *J Mol Med (Berl)* 2007;85:613–21.
214. Reghunathan R, Jayapal M, Hsu LY, et al. Expression profile of immune response genes in patients with severe acute respiratory syndrome. *BMC Immunol* 2005;6:2.
215. Simmons CP, Popper S, Dolocek C, et al. Patterns of host genome-wide gene transcript abundance in the peripheral blood of patients with acute dengue hemorrhagic fever. *J Infect Dis* 2007;195:1097–107.
216. Zaas AK, Chen M, Varkey J, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe* 2009;6:207–17.

217. Ardura MI, Banchereau R, Mejias A, et al. Enhanced monocyte response and decreased central memory T cells in children with invasive *Staphylococcus aureus* infections. *PLoS One* 2009;4:e5446.
218. Thompson LJ, Dunstan SJ, Dolecek C, et al. Transcriptional response in the peripheral blood of patients infected with *Salmonella enterica* serovar Typhi. *Proc Natl Acad Sci U S A* 2009; 106:22433–8.
219. Hetherington S, McGuirk S, Powell G, et al. Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. *Clin Ther* 2001;23:1603–14.
220. Mallal S, Phillips E, Carosi G, et al. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 2008;358:568–79.
221. Genebat M, Ruiz-Mateos E, Leon JA, et al. Correlation between the Trofile test and virological response to a short-term maraviroc exposure in HIV-infected patients. *J Antimicrob Chemother* 2009;64:845–9.
222. Marko NF, Weil RJ. Mathematical modeling of molecular data in translational medicine: theoretical considerations. *Sci Transl Med* 2010;2:56rv4.
223. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;25: 127–41.
224. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods* 2002;7:19–40.
225. Zhao X, Cheung LW. Kernel-imbedded Gaussian processes for disease classification using microarray gene expression data. *BMC Bioinformatics* 2007;8:67.
226. Kong W, Vanderburg CR, Gunshin H, Rogers JT, Huang X. A review of independent component analysis application to microarray gene expression data. *BioTechniques* 2008;45:501–20.
227. Lawler GF. Introduction to stochastic processes. 2nd ed. Boca Raton, FL: CRC, 2006.
228. Yan Q. Translational bioinformatics and systems biology approaches for personalized medicine. *Methods Mol Biol* 2010; 662:167–78.
229. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2011;40:D109–14.
230. Sarkar IN, Butte AJ, Lussier YA, Tarczy-Hornoch P, Ohno-Machado L. Translational bioinformatics: linking knowledge across biological and clinical realms. *J Am Med Inform Assoc* 2011;18:354–7.
231. [No authors listed] Raising the game. *Nat Biotechnol* 2008; 26:137.
232. Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 2010;9:203–14.
233. US Food and Drug Administration. NME drug and new biologic approvals. Silver Spring, Md: FDA, 2011.
234. Goodman M. Market watch: pharma industry strategic performance: 2007–2012E. *Nat Rev Drug Discov* 2008;7:967.
235. Munos B. Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov* 2009;8:959–68.
236. Lichtenberg FR. The impact of new drug launches on longevity: evidence from longitudinal, disease-level data from 52 countries, 1982–2001. *Int J Health Care Finance Econ* 2005;5:47–73.
237. Ledford H. Translational research: 4 ways to fix the clinical trial. *Nature* 2011;477:526–8.
238. National Institutes of Health. Regulatory science. Bethesda, Md: NIH, 2011.
239. US Food and Drug Administration. FDA-NIH Joint Leadership Council Charter. Silver Spring, Md: FDA, 2011.
240. US Food and Drug Administration NIOH. FDA-NIH News Release: NIH and FDA announce collaborative initiative to fast-track innovations to the public. Silver Spring, Md: FDA, 2010.
241. Collins FS, Gray GM, Bucher JR. Toxicology. Transforming environmental health protection. *Science* 2008;319:906–7.
242. Kennedy D. FDA centennial. *Science* 2006;312:19.
243. Wood AJ. A proposal for radical changes in the drug-approval process. *N Engl J Med* 2006;355:618–23.
244. Collins FS. Mining for therapeutic gold. *Nat Rev Drug Discov* 2011;10:397.
245. Fang FC, Casadevall A. Lost in translation—basic science in the era of translational research. *Infect Immun* 2010;78:563–6.
246. Petsko GA. Herding CATS. *Sci Transl Med* 2011;3:97cm24.
247. Redhead CS, Smith PW. Public Health Service (PHS) Agencies: Overview and Funding, FY2010–FY2012. Washington, DC: Congressional Research Service, 2011.
248. Smrb Hears From The Public About Ncats. Available from: <http://feedback.nih.gov/index.php/ncats/smrh-hears-from-the-public-about-ncats/>. Accessed March 1, 2011.
249. Kaiser J. NIH reorganization draws scrutiny from Congress. *Science Insider* January 27, 2011.
250. FitzGerald GA. NCATS purrs: emerging signs of form and function. *Sci Transl Med* 2011;3:83ed2.