Perspectives **on Informatics**

Viewpoint Paper ■

Translational Bioinformatics: Coming of Age

Aʀᴜʟ J. Bᴜᴛᴛᴇ, MD, PʜD

A b s t r a c t   The American Medical Informatics Association (AMIA) recently augmented the scope of its activities to encompass translational bioinformatics as a third major domain of informatics. The AMIA has defined translational bioinformatics as "*. . . the development of storage, analytic, and interpretive methods to optimize the transformation of increasingly voluminous biomedical data into proactive, predictive, preventative, and participatory health.*" In this perspective, I will list eight reasons why this is an excellent time to be studying translational bioinformatics, including the significant increase in funding opportunities available for informatics from the United States National Institutes of Health, and the explosion of publicly-available data sets of molecular measurements. I end with the significant challenges we face in building a community of future investigators in Translational Bioinformatics.

■ **J Am Med Inform Assoc.** 2008;15:709–714. DOI 10.1197/jamia.M2824.

Introduction

Translational Medicine has been described as the effective transformation of information gained from the past fifty years of biomedical research into knowledge that can improve [the state of] human health and disease.[1] This transformation requires two processes to work effectively: first, taking basic biological findings and applying them to human biology, and second, taking clinical research findings and actually improving the health of populations. The specific development of information systems is a rate-limiting challenge for these two processes.[1] Many healthcare institutions are expanding the role of their operational information technology systems, such as electronic health record, decision support, and computerized provider-order-entry systems to include the mission of translational research.[2]

Achieving the impact of translational medicine requires expanding the role and scope of bioinformatics just as much as those for clinical informatics. In 1999, the Advisory Committee to the Director, National Institutes of Health

Affiliations of the author: Stanford Center for Biomedical Informatics, Department of Medicine and Department of Pediatrics, Stanford University School of Medicine, Stanford, CA; Lucile Packard Children's Hospital, Palo Alto, CA.

Correspondence: Atul Butte, MD, PhD, Stanford Center for Biomedical Informatics, 251 Campus Drive, Room X-215 MS-5479, Stanford, CA 94305-5479; e-mail: <abutte@stanford.edu>.

(NIH) Working Group on Biomedical Computing, co-chaired by David Botstein and Larry Smarr, released the Biomedical Information Science and Technology Initiative (BISTI) report, which recommended that NIH should be responsive to the growth in biological data and should apply funding resources to accelerate the development and application of computational tools to science. While the BISTI report certainly led to increased funding for bioinformatics research, in retrospect, the subsequent initiatives often led to the development of novel tools, perhaps at the expense of identifying novel questions. Perhaps there was no way for the BISTI authors to predict that a generation of scientists, asking medical questions at a molecular level solely using computational resources, could appear so quickly.

The circumstances are now such that it is time to recognize this new area of inquiry called *Translational Bioinformatics*. The American Medical Informatics Association (AMIA) recently added translational bioinformatics as one of its three major domains of informatics. The AMIA has defined translational bioinformatics as:

> "*. . . the development of storage, analytic, and interpretive methods to optimize the transformation of increasingly voluminous biomedical data into proactive, predictive, preventative, and participatory health. Translational bioinformatics includes research on the development of novel techniques for the integration of biological and clinical data and the evolution of clinical informatics methodology to encompass biological observations. The end product of translational bioinformatics is newly found knowledge from these integrative efforts that can be disseminated to a variety of stakeholders, including biomedical scientists, clinicians, and patients.*"[3]

Translational Bioinformatics involves the development and use of computational methods that can reason over the enormous amounts of life science data being collected and stored for the purpose of creating new tools for medicine. While bioinformatics methodologies have been used to

enable biological discoveries for decades, here the end product has to be translational, or applying to human health and disease.

Why should investigators in computer science, biomedical informatics, and biomedical research in general be interested in Translational Bioinformatics today? I will list eight reasons why now is an excellent time to be studying Translational Bioinformatics. Five of these reasons are intrinsic to this scientific discipline, while three are extrinsic, regarding the practice of this discipline in today's scientific, funding, and political context. I will end with the significant challenge of building a community of future investigators in Translational Bioinformatics.

Availability of Molecular Tools

First, many tools now exist that enable a large scale, parallel, quantitative, and inexpensive assessment of molecular states. Instead of thinking about molecular measurements one at a time, we now have many available tools in science that measure many molecules at a time; these tools are commonly described as being *high throughput* based on this feature. The premier example of this is the gene expression microarray,[4,5] which enables the measurement of gene expression (RNA) levels across tens of thousands of genes. Microarray technology has successfully quantitated differences between diseases and discovered novel sub-types of disease.[6,7] This one platform has provided the ability to quantify gene expression under differing experimental conditions. It can be used by various algorithms to classify, learn, or predict biologically relevant processes.

Beyond being large, these technologies are nearly comprehensive. It's one thing having a technology to measure ten genes, or 10,000 genes, but once you get close to 40,000 genes, there aren't many more genes left to measure. This may be only an illusion of stability, however, as demand increases for levels of resolution to improve. For instance, newer gene expression microarrays have evolved to measure exons, the individual components of RNA molecules, instead of entire transcripts.[8] Future technologies may enable faster measurements to be made, with less bias towards the known catalog, or with less measurement noise.

Another important point is the low cost of these modalities. Gene expression microarrays were a cost-prohibitive technology when they were developed 11 years ago, but now they are essentially commodity items. Microarrays that measure activity of every gene in the genome now cost only about $300 per sample (plus labor and supplies) for academics.

Other research modalities have also become inexpensive. Between any two individuals, there are an estimated 10 million differences, or single nucleotide polymorphisms (SNPs), in DNA.[9] The measurement of 1.8 million of these differences also costs about $300 per sample, in academia. Last year's analytic model had half a million SNPs for the same price, and the model from the previous year had about 10,000 SNPs for the same price, so there has been a geometrically progressive price reduction.

Public Availability of Molecular Measurement Data

Second, not only do technologies exist that enable the large scale, parallel, quantitative, and inexpensive assessment of molecular states, but also the data from these tools are now increasingly publicly available. A translational bioinformatician may not necessarily even have to run one of these data-generating machines.

The premier example of an internationally available data resource is GenBank, initially created in the early 1980s by Walter Goad.[10] Because so many investigators at the dawn of the sequencing era were generating DNA sequences, there was a need for a repository to centrally manage and use these sequences. Funding from the NIH for GenBank started in 1982, and in the subsequent quarter century, GenBank has grown to include 82 billion nucleotides in 78 million sequences.[11] At the time of this writing, hundreds of organisms have been completely sequenced including, of course, man and mouse. But a total of 270,000 species have had some sequence measured.[12] In this way, GenBank has both breadth and depth.

The equivalent of GenBank for gene expression microarrays is known as the Gene Expression Omnibus (GEO).[13] The GEO is also maintained by the National Center for Biotechnology Information at the National Library of Medicine. At the time of this writing, GEO has over 183,000 samples from over 7,200 experiments, an impressive growth in seven years. The number of samples either doubles or triples each year.

This availability of massive data sets is not just an American initiative. The European Bioinformatics Institute (EBI) has a similar web-based database called ArrayExpress[14] with over a hundred-thousand samples from over 3,000 experiments. All together, translational bioinformaticians can likely get their hands on more than a quarter million microarray samples today. This is more data than can be generated by any one biologist, and the results from analyzing these larger collections of samples are potentially enormous in impact. As of 2007, diseases contributing to nearly a third of human disease-related mortality in the United States have been studied by microarrays.[15]

This availability is not limited to gene expression data. The EBI also has a web-based database called PRIDE, which holds proteomics data.[16] The PRIDE database holds 3,200 independent samples with 2.6 million mass spectra freely available for download. Data from genome-wide association studies have their own repository, the NCBI Database of Genotype and Phenotype (dbGaP).[17] As of this writing, fourteen genetic studies are available for downloading in this one year old database, with over 40,000 human samples.

Culture of Sharing Molecular Data and Tools

Why are these measurements increasingly available? Availability is a function of both stick and carrot. Molecular data would not be available without some kind of mandate, and the strongest comes from academic journals. Top-tier journals require the deposition of molecular measurements into international repositories for manuscripts under consideration for publication.

Funding agencies also increasingly require the public availability of scientific data, such as the Wellcome Trust and NIH.[18] Grant proposals to NIH asking for over $500,000 per year need to have text describing how the data will be shared.[19] Though this requirement is new, policies in shar-

ing data from major projects, such as the Human Genome Project, go back more than a decade.[20]

Beyond these requirements, however, there is a culture of open sharing in molecular biology and bioinformatics that continues to grow: sharing of tools, data, findings, and publications. Important tools for bioinformatics, such as Significance Analysis of Microarrays (SAM),[21] TM4 Multiple Expression Viewer,[22] GenePattern,[23] GenMAPP,[24] and R and Bioconductor,[25,26] are downloadable for free, and in many cases, have available source code.

In many cases, biomedical research communities have come together and have realized that sharing takes more than just uploading files to a common website. Through contention and agreement, these communities are starting to standardize terminology, phenotypes, and gene names.[27] Challenges still remain in cataloging, calibrating, and normalizing data across experimenters, across measurement modalities, and across biological models; improper attention to these could lead to false positives and negatives. These biomedical research communities could benefit from learning how some of these challenges were addressed by the clinical informatics community.

Where those standards have not yet been reached, there is at least the understanding in the appropriate communities that standards must be reached. Increasingly, there is inter-community sharing, where one community will learn from the standardization efforts of another community. Examples of inter-community sharing include the design of the Minimum Information About a Proteomics Experiment (MIAPE)[28] using the Minimum Information About a Microarray Experiment (MIAME),[29] and partnerships between the FlyBase and ZFIN with the National Center for Biomedical Ontology to standardize phenotype descriptors.[30]

Curiously, this culture of sharing has not extended well to clinical research or clinical informatics. Clinical informatics tools, including vocabularies and text-parsing tools, are not always shared, or require signed licensing agreements. Clinical data, even de-identified subsets, are not as available on the Internet as molecular measurements. This could be due to fears of release of personal medical information, disclosure of evidence of culpability, or worries that one might miss a discovery in one's own patient cohort.[31]

Clinicians are Expected to Interpret Bioinformatics Methodologies

It is amazing how much bioinformatics physicians and other health professionals must know. Terms like "shrunken centroid," "unsupervised cluster analysis," "gene expression signature," "ten-fold cross validation," "global scaling," "q value," and even the "Cochran-Mantel-Henzel stratified analysis test" appear in journals read by healthcare providers. For example, all the preceding terms recently appeared in *The New England Journal of Medicine*. The growing importance of high-throughput molecular measurements in medicine even led to a thirteen article series in *The New England Journal of Medicine* between 2002 and 2003, with the lead article in the series by Alan Guttmacher and Francis Collins.[32] The role of Translational Bioinformatics now plays a front-page role in the top-most tier of journals. Even if clinicians do not know how to use or implement these

methods, they must understand how these methods are relevant to health care.

Question Asking in Translational Bioinformatics

This topic addresses the sustainability and growth of the field of Translational Bioinformatics. Certainly biologists now understand and apply high-throughput measurement modalities. A biologist running an expression microarray or proteomic study will generate a sizable amount of raw data. Distilling all the raw data and determining the relevant operational genes and findings clearly requires the proper application of bioinformatics. Over the past two decades, it has been clear that bioinformaticians can help biologists to analyze such complex data, given the primary questions that the biologists are asking.

In addition, bioinformatics must play a key role in the storage and retrieval of high-throughput data. A bioinformatician could work with a biologist to set up a web site and a standardized database for experimental measurements, facilitate the sharing of the measurements, and relate them to clinical outcomes.

Because of the public availability of raw high-throughput molecular data, roles for translational bioinformaticians can now change to beyond just providing a service. Translational bioinformaticians, given the data resources outlined above, have essentially more samples available regarding a given disease, e.g., breast cancer, than any individual biologist studying breast cancer might alone create. A translational bioinformatician can go to the NCBI GEO and download over 9,300 microarray studies on breast cancer (over 1,800 of them entered in 2007).

The availability of substantial public data enables bioinformaticians' roles to change. Instead of just facilitating the questions of biologists, the bioinformatician, adequately prepared in both clinical science and bioinformatics, can ask new and interesting questions that could never have been asked before. For example, Mootha et al. integrated four publicly available expression data sets with genetic linkage data and proteins identified from mitochondria to find the gene mutation associated with Leigh syndrome, French-Canadian type.[33] English collected 49 publicly available high-throughput experiments of multiple types, such as genetic scans, gene expression microarrays, proteomics, and RNA interference, all related to the study of obesity. She found that an integrative model across 49 experiments could statistically significantly outperform each of the independent experiments in rediscovering known obesity-associated genes and predicting novel ones.[34]

These examples demonstrate an approach to integrating public and private data sets to address an important question in medicine. There is a role for the translational bioinformatician as question-asker, not just as infrastructure-builder or assistant to a biologist.

Calls for Translational Medicine

The final three reasons to learn and participate in Translational Bioinformatics are extrinsic to the science and tools available in that discipline. They relate to current practices in Translational Bioinformatics. There are increasing calls from many vantage points for Translational Medicine. In the late 1990s, the NIH budget doubled, with grant proposal

funding rates rising to a peak of 32%.[35] The completion of the pilot phase of the Human Genome Project in 1999 and the release of the finished genomic sequence in 2003 provided visible evidence that increased NIH budgets yielded new research data and tools. Now, more than five years after the doubling many constituencies demand clinical and translational applicability from basic science research—for patients and the public,[36] the pharmaceutical industry,[37] clinical researchers,[1,38] basic science investigators,[39] funding agencies,[40] and NIH itself.[41,42] While the NIH budget doubling led to new data and knowledge, questions about which products from the genome era can help patients are fair.

Increasing Research Funding for Translational Bioinformatics

Calls for increasing translational research have led to greater financial support for Translational Bioinformatics. In May 2002, Elias Zerhouni, Director of the National Institutes of Health, outlined the NIH Roadmap, a plan to modernize the process of medical research for the 21st century. Dr. Zerhouni outlined three major themes as part of the initial Roadmap: (1) New Pathways to Discovery, which addressed Bioinformatics and Computational Biology as novel methods for molecular study, (2) Research Teams of the Future, which suggested that cell biologists and computational biologists should collaboratively accelerate movement of scientific discoveries from the bench to the bedside, and (3) Re-engineering the Clinical Research Enterprise, which tackled the transformation of basic research discoveries into drugs, treatments, and methods for prevention. There has been a role for Translational Bioinformatics in all three of these major themes of the NIH Roadmap.

Most importantly, Dr. Zerhouni wrote in *The New England Journal of Medicine* in 2005:

> *It is the responsibility of those of us involved in today's biomedical research enterprise to translate the remarkable scientific innovations we are witnessing into health gains for the nation . . . At no other time has the need for a robust, bidirectional information flow between basic and translational scientists been so necessary.*[41]

There are impressive informatics-related terms in that quote for a Director of NIH, such as "robust, bidirectional information flow." Coincident with this quote and publication, the push to reinvent clinical research reached a new peak with the release of the Request for Applications (RFA) for the NIH Roadmap Institutional Clinical and Translational Science Awards (CTSA).[43] These awards required that medical schools, research hospitals, and related institutions commit to reinventing how they perform and teach clinical and translational research. To enable this transformation, NIH planned to fund approximately 60 institutions at about $30 million each. As might be expected, the RFA for a $30 million grant spans over 50 printed pages. Unexpectedly, however, the word "informatics" in this RFA appeared 38 times.[43] An institution cannot apply for a CTSA grant without organizing a clear plan for informatics, and must include tools and infrastructure to enable Translational Medicine. Each institution is required to determine a local Biomedical Informatics Director, and each of these participates on a national committee to set standards for clinical and translational research. This was clear recognition by
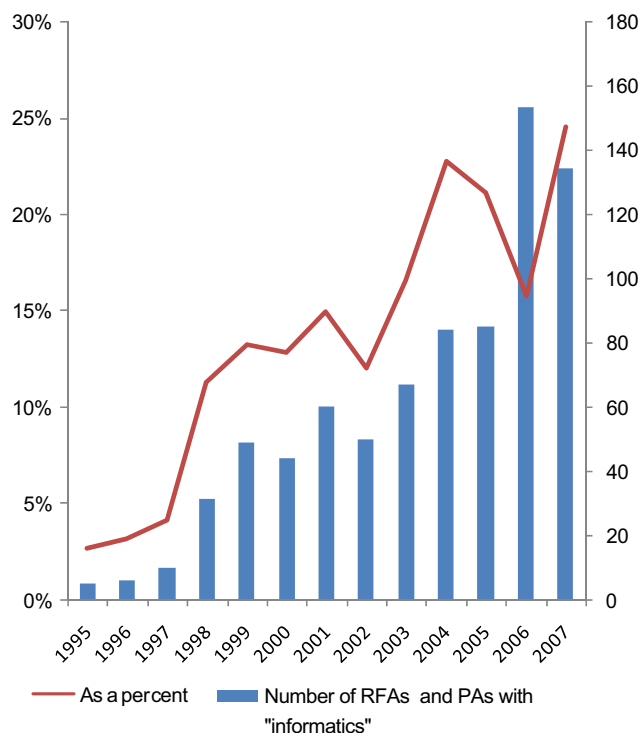

Figure 1. Bars represent the count per year of Request for Applications (RFAs) and Program Announcements (PAs) found through the NIH Advanced Funding Opportunities & Notices Search web-site[44] containing the term "informatics", either currently active or inactive. Line represents the fraction of this count over the total count of RFAs and PAs that year.

NIH that the problems of Translational Medicine will not be solved without the help of informatics, and substantial money backed up this statement. Beyond the CTSA, NIH has continued to support Translational Bioinformatics through its funding of other large programs, including seven National Centers for Biomedical Computing (NCBC) and the Cancer Bioinformatics Grid (caBIG).

The CTSA effort provides an example of the depth of funding available when NIH focuses on specific major problems. There is also breadth. Figure 1 shows the yearly count of how often the word "informatics" appears in Request for Applications (RFAs) and Program Announcements (PAs) issued in the *NIH Guide*, the weekly publication issued by NIH on new funding mechanisms and program announcements.

Between 2001 and 2005, mentions of "informatics" grew at a slow pace, with 40 and 80 mentions per year. But over the past two years, during a period when the NIH budget has otherwise been relatively flat, there has been a remarkable increase. In 2006, the number of RFAs "informatics" mentions jumped dramatically to nearly 160. Example RFAs in which "informatics" appears include the obviously related, such as "A Data Analysis & Coordination Center (DACC) for the Human Microbiome Project" (RFA-RM-08-007) and the "Biomedical Informatics Research Network Coordinating Center (U24)" (RFA-RR-08-002), but also the less obvious, such as "Continuation and Expansion of the Drug Induced Liver Injury Network" (RFA-DK-07-012) and "As-

say Development for High Throughput Molecular Screening (R21)" (PAR-08-024).

While the 2007 count of 136 was lower than 2006, we have now reached new watershed in that a quarter of the RFAs and PAs mentioned the term "informatics." This crude method of counting admittedly does not distinguish clinical informatics from bioinformatics, does not consider the dollars available for each RFA, ignores the duration, availability, and expiration of RFAs, and may even falsely count RFAs in which informatics is explicitly excluded. Yet this still remains a simple example of how broadly informatics is now considered across many funding mechanisms involving all institutes of NIH.

Few Investigators in Translational Bioinformatics

The final point is obvious. There is an absolute paucity of people trained to make use of these resources, to build the infrastructure, to ask these novel questions, and to even answer those questions. In the United States, the National Library of Medicine (NLM) University-based Biomedical Informatics Research Training Programs comprise the premier mechanism for students to combine advanced health professional degrees with doctoral level (PhD) studies in biomedical informatics, producing as a result MD/PhD, RN/PhD, PharmD/PhD, etc. trained individuals. While these programs were initially geared towards training individuals in healthcare informatics, they have diversified in scope and focus. Fifteen of the 18 funded training programs now emphasize training in bioinformatics or computational biology.[45] In fact, more training programs emphasize training opportunities in bioinformatics or computational biology than other areas of biomedical informatics, including healthcare informatics, clinical research informatics, and public health informatics. Of course, training of such individuals is not limited to those institutions with NLM-based training programs. Many other top-tier institutions train students in departmental programs, housed in Departments of Biomedical Informatics, Genetics, and Computer Science, as well as inter-departmental programs.

The future development of practitioners of Translational Bioinformatics will require that individuals enter this discipline from even more diverse backgrounds. For instance, it is still rare for a clinician-scientist, who has completed training in medicine, pediatrics, or surgery, to undergo joint training in a sub-specialty as well as bioinformatics. A quantitatively-thinking cardiologist-scientist in-training could be trained in both human physiological measurements as well as methods for multi-scale modeling of the heart. A quantitatively-thinking oncology research-nurse-in-training could be trained in both making molecular measurements and methods in machine learning to find genes that predict outcome. Success in these joint training programs will require vision, as well as bioinformatics training program directors that reach out to and work with traditional sub-specialty fellowship directors.

Conclusion

The role and future importance of the nascent field of Translational Bioinformatics appears promising. There are demonstrated needs, funding, resources, and roles. The most significant challenges to the future growth of Translational

Bioinformatics remain in education. Clinicians need to be educated so that they can understand Translational Bioinformatics methods as well as methods used in clinical trials. It is reasonable for educators in bioinformatics to expect more graduate students to take a greater interest in specific open clinical questions, in addition to the methods they are uniquely qualified to develop and apply to solve those questions.

Computer scientists, even at the undergraduate level, should be educated that the algorithms and methods they develop in machine learning, visualization, network modeling, and knowledge representation will find a receptive audience in biomedical research. Quantitative-thinking undergraduate and graduate students in biology and chemistry should be exposed to, and excited by, increasing digital sources of data. There is no single educational solution that spans these constituencies, but the pieces have to include web-based instruction, traditional lecture-based courses, graduate degree programs, research fellowships, and continuing medical education courses. Some of these educational opportunities might be most efficiently delivered when centralized within departmental structures, but clearly Translational Bioinformatics will be practiced outside existing department walls.

Despite these challenges in developing of a committed set of investigators in Translational Bioinformatics, this is clearly a unique and exciting time to be part of the growth phase of this new scientific discipline.

References ■

1. Sung NS, Crowley WF, Jr., Genel M, et al. Central challenges facing the national clinical research enterprise. JAMA 2003 Mar 12;289(10):1278–87.
2. Payne PR, Johnson SB, Starren JB, Tilson HH, Dowdy D. Breaking the translational barriers: the value of integrating biomedical informatics and translational research. J Investig Med 2005 May;53(4):192–200.
3. American Medical Informatics Association. AMIA Strategic Plan. 2006. Available at: http://www.amia.org/inside/stratplan/; accessed Mar 15, 2007.
4. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270(5235):467–70.
5. Lipshutz RJ, Morris D, Chee M, et al. Using oligonucleotide probe arrays to access genetic diversity. Biotechniques 1995 Sep;19(3):442–7.
6. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286(5439):531–7.
7. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403(6769):503–11.
8. Clark TA, Schweitzer AC, Chen TX, et al. Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol 2007;8(4):R64.
9. Kruglyak L, Nickerson DA. Variation is the spice of life. Nat Genet 2001 Mar;27(3):234–6.
10. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res 2007 Dec 11.
11. Anonymous. Genetic Sequence Data Bank NCBI-GenBank Flat File Release 162.0 Distribution Release Notes. October 15, 2007; 162.0:[This document describes the format and content of the flat files that comprise releases of the GenBank nucleotide sequence database.]. Available at: ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt. Accessed Dec 1, 2007.

12. Anonymous. Taxonomy Nodes (all dates). 2007. Available at: http://www.ncbi.nlm.nih.gov/Taxonomy/txstat.cgi. Accessed Dec 1, 2007.

13. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic Acids Res 2007 Jan;35(Database issue):D760–5.

14. Parkinson H, Sarkans U, Shojatalab M, et al. ArrayExpress—a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2005 Jan 1;33(Database issue):D553–5.

15. Dudley J, Butte AJ. Enabling integrative genomic analysis of high-impact human diseases through text mining. Pac Symp Biocomput 2008:580–591.

16. Martens L, Hermjakob H, Jones P, et al. PRIDE: the proteomics identifications database. Proteomics 2005 Aug;5(13):3537–45.

17. Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 2007 Oct;39(10):1181–6.

18. The Wellcome Trust. Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. Fort Lauderdale: Wellcome Trust; 2003 January 14–15, 2003.

19. NIH Data Sharing Policy and Implementation Guidance. March 5, 2003. Available at: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm. Accessed Dec 1, 2007.

20. NIH, DOE Guidelines Encourage Sharing of Data, Resources. Hum Genome News. 1993;4(5).

21. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001;98(9):5116–21.

22. Saeed AI, Bhagabati NK, Braisted JC, et al. TM4 microarray software suite. Methods Enzymol 2006;411:134–93.

23. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet 2006 May;38(5):500–1.

24. Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR. GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. Nat Genet 2002 May;31(1):19–20.

25. Gentleman R. Bioinformatics and computational biology solutions using R and Bioconductor. New York: Springer Science+Business Media; 2005.

26. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5(10):R80.

27. HUGO—a UN for the human genome. Nat Genet 2003 Jun;34(2):115–6.

28. Taylor CF, Paton NW, Lilley KS, et al. The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol 2007 Aug;25(8):887–93.

29. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 2001;29(4):365–71.

30. Rubin DL, Lewis SE, Mungall CJ, et al. National Center for Biomedical Ontology: advancing biomedicine through structured organization of scientific knowledge. OMICS. 2006 Summer;10(2):185–98.

31. Butte AJ. Medicine. The ultimate model organism. Science 2008 Apr 18;320(5874):325–7.

32. Guttmacher AE, Collins FS. Genomic medicine—a primer. N Engl J Med 2002 Nov 7;347(19):1512–20.

33. Mootha VK, Lepage P, Miller K, et al. Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. Proc Natl Acad Sci U S A. 2003 Jan 21;100(2):605–10.

34. English SB, Butte AJ. Evaluation and Integration of 49 Genome-wide Experiments and the Prediction of Previously Unknown Obesity-related Genes. Bioinformatics 2007 Oct 5;23(21):2910–7.

35. Zerhouni EA. Research funding. NIH in the post-doubling era: realities and strategies. Science 2006 Nov 17;314(5802):1088–90.

36. Waller D. Too Little Bang for the Buck in Drug Research? Time, December 27, 2006. Available at: http://www.time.com/time/nation/article/0,8599,1572974,00.html. Accessed Dec 1, 2007.

37. Hurko O, Rutkowski JL. Translational Medicine Breaks Bottlenecks. Drug Discovery & Development 2005;8(11):30–8.

38. Paul M. Translational investigators: life sciences' application engineers. Nat Biotechnol 2007 Jul;25(7):817–8.

39. Horton B. From bench to bedside . . . research makes the translational transition. Nature 1999;402(6758):213–5.

40. Carpenter S. Science careers. Carving a career in translational research. Science 2007;317(5840):966–7.

41. Zerhouni EA. Translational and clinical science—time for a new vision. New Engl J Med 2005 Oct 13;353(15):1621–3.

42. Wadman M. Research policy: the map man. Nature 2007;448(7152):406–7.

43. National Institutes of Health. Institutional Clinical and Translational Science Award. 2005. Available at: http://grants2.nih.gov/grants/guide/rfa-files/RFA-RM-06-002.html. Accessed Mar 15, 2007.

44. Office of Extramural Research. Advanced Funding Opportunities & Notices Search. January 10, 2008. Available at: http://grants.nih.gov/grants/guide/search_guide.htm. Accessed Jan 1, 2008.

45. NLM's University-based Biomedical Informatics Research Training Programs. January 7, 2008. Available at: http://www.nlm.nih.gov/ep/GrantTrainInstitute.html. Accessed Jan 1, 2008.