

Bioinformatics at the Albert Einstein College of Medicine

Kremena Star

Medical Scientist Training Program
Department of Molecular Genetics
Albert Einstein College of Medicine
Bronx, New York 10461

In the broad sense of the word, bioinformatics is the application of computer technology to the management of biological information. In particular, it is the science of developing databases and algorithms to facilitate and expedite biological research (Table 1). For example, computer searches allowed for identification of "cellular gene analogues of viral oncogenes" (Lesk, 1998). Other applications include protein and RNA folding pattern prediction, evolutionary trees construction, and nucleic acids sequences comparison.

The need for "sophisticated analysis of biological sequences" (Durbin et al., 1999) is the driving force behind the new and expanding field of bioinformatics. Since "molecular biology and computer science are the most rapidly evolving disciplines in the contemporary scientific scene" (Lesk, 1998), data (e.g., human genome) are being produced at an exponential rate. Therefore, no modern scientist or scientific institution can be successful without knowing how to control, manipulate, and interpret this ever-growing database of biological knowledge.

In this article, I outline the current state of readiness of the Albert Einstein College of Medicine (AECOM) to han-

dle pressing bioinformatics problems of our times and attempt to project gaps that must be filled in order to successfully enter into the era of infinite data. What is the bioinformatics' status quo at AECOM today? In terms of facilities there are a number of resources available to everyone (Table 2).

On February 1, 2004 the Genome Informatics Core Laboratory, largely unknown to the AECOM community, became the Bioinformatics Shared Resource (BISR). It provides bioinformatics support to individual researchers in both basic science and clinical areas. In addition, this shared resource will support the bioinformatics needs of other core facilities and centers of the medical school, including the chemical genome core facility, fluorescent in situ hybridization (FISH) core facility, microarray facility, proteomics core facility, and the cancer center. It can also develop application software tailored to specific needs. The BISR consists of a group of computer-savvy faculty members who will go out of their way to help you.

The DNA microarray facility analyzes cDNA microarray chips and has a bioinformatician, Dr. Kate Milova, on staff. Dr. Milova helps users with data acquisition, such as conversion of images of arrays into quantitative numbers

TABLE 1 | BIOINFORMATICS WEB SITES

Content	Web Site
Introduction to Bioinformatics (San Diego Supercomputer Center)	http://www.sdsc.edu/~gribskov/bimm140/lectures/12.IntroInfoResI.Num.pdf
Database Searching (San Diego Supercomputer Center)	http://www.sdsc.edu/~gribskov/bimm140/lectures/2003_apr21-23.pdf
Genomics/Proteomics (San Diego Supercomputer Center)	http://www.sdsc.edu/~gribskov/bimm140/lectures/2003_jun4.pdf
Gene Modeling (San Diego Supercomputer Center)	http://www.sdsc.edu/~gribskov/bimm140/lectures/2003_apr16.pdf
Computational Biology Lecture Notes (Carnegie Mellon University)	http://www.cmu.edu/bio/education/courses/03310/LectureNotes/
Albert Einstein College of Medicine Bioinformatics/Biostatistics Club	http://www.geocities/aecombbc
Biocomputing Tutorials	http://www.bio.cam.ac.uk/Embnetut/Universl/tutrial.html

Bioinformatics at the Albert Einstein College of Medicine

and normalization of values between channels of a single experiment and between different experiments. She also maintains the facility website and is creating a web-accessible database that will give information on the clones from the AECOM chips along with links to their gene products.

The Laboratory for Macromolecular Analysis and Proteomics (LMAP) provides state-of-the-art technologies for proteomics, mass spectrometry, protein identification, protein sequencing, and custom peptide synthesis. LMAP staff scientists assist in experimental design during early planning stages, carry out the experiments, and analyze data.

The Einstein Center for Synchrotron Biosciences gives AECOM investigators access to electromagnetic radiation in the infrared and x-ray regions. It provides a number of protein analysis technologies, including: x-ray footprinting, x-ray crystallography, infrared spectroscopy, and genomics services. The center's most recent activity was the bioinformatics workshop at the end of January, 2004. Due to its tremendous success, this workshop will be expanded and repeated next year from January 10 to January 12. According to Dr. Mark Chance, January 10 will be an intro to National Center for Biotechnology Information (NCBI) online resources, January 11 will focus on advanced topics (NCBI microarray database, structural bioinformatics, etc.) and January 12 will be designated for AECOM and New York City area bioinformatics experts.

The Division of Biostatistics in the Department of Epidemiology and Population Health provides statistical expertise in the analysis of genomic and proteomic data.

The Division, which was recently created in April 2003, is headed by Dr. Mimi Kim. Currently, it consists of six faculty members and four masters level biostatisticians. The faculty conduct research on new statistical approaches for the design and analysis of biomedical studies, including basic science, translational, clinical, and observational studies. They also collaborate with AECOM investigators on all stages of research, from experimental design and formulation of hypotheses to data analysis, interpretation, and reporting of results.

Finally, the Seaver Foundation Center for Bioinformatics is a "virtual" center headed by Dr. Steven Schwartz. At present it has only one faculty member: Andras Fiser. However, its future seems bright. It will have six faculty members (according to Dr. Schwartz, three of them should be hired by the summer) and will be located in the new Price Center for Genetic and Translational Medicine - to be built across the street from Forchheimer, with construction scheduled to begin by the end of this year. When this center for interdisciplinary science is finished it will house eight bioinformatics labs.

These high-quality shared resources at AECOM provide the infrastructure for integrating data producing facilities with data management, data mining (Bioinformatics Shared Resource) and analytical resources (Division of Biostatistics) (Figure 1). Up to now principal investigators (P.I.) had to coordinate all facets of experiment design, data production, and analysis. However, recent developments of high throughput technology (e.g., microarrays) have resulted in the production of large datasets whose analysis and management requires additional expertise.

This problem is being addressed by the development of

TABLE 2 AECOM BIOINFORMATICS RESOURCES		
Facility	Contact Person	Web address
Bioinformatics Shared Resource (BISR)	Dr. Xin Zheng	http://bioinformatics.aecom.yu.edu/
DNA Microarray Facility	Dr. Geoffrey Childs	http://www.aecom.yu.edu/cancer/new/cores/microarray/default.htm
Laboratory for Macromolecular Analysis and Proteomics (LMAP)	Dr. Ruth Angeletti	http://www.bioc.aecom.yu.edu/labs/angellab/
Einstein Center for Synchrotron Biosciences	Dr. Mark Chance	http://www.aecom.yu.edu/home/csb/
Division of Biostatistics	Dr. Mimi Kim	http://www.aecom.yu.edu/desm/Epi_biostat.htm
Seaver Foundation Center for Bioinformatics	Dr. Steven Schwartz	none

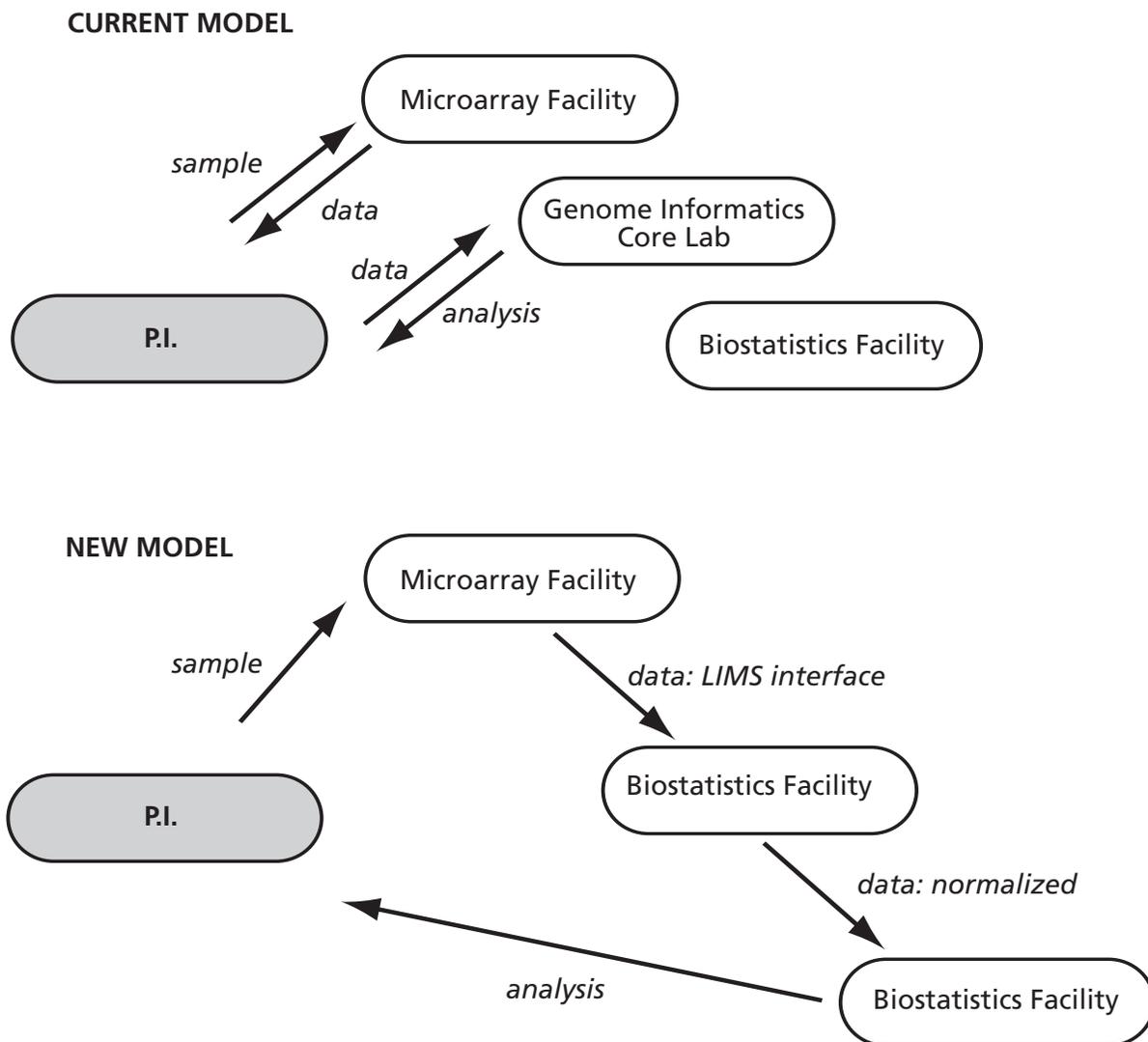


FIGURE 1 | Integration of AECOM Shared Resources (LIMS, Laboratory Information Management System; P.I., Principle Investigator) (courtesy of the Greally Laboratory).

a new model for applying high throughput technology to solve biological questions. In this model, faculty members have access to biostatistics and bioinformatics input during the experimental design stage. This ensures optimal structure of the data generated for further studies and stored in a central database at the BISR. Data analysis is performed with input from the investigator and the biostatistics facility. In a similar fashion, data mining and data management is done by the investigator and the bioinformatics resource. This integrated model gives the AECOM faculty the opportunity to obtain an optimal outcome from the shared facilities. It provides for efficient and effective experimental design and ensures

derivation of maximal information from the generated datasets.

When it comes to resources and faculty recruitment, AECOM is doing quite well. However, in terms of courses offered, the picture is not a pretty one. Currently, AECOM does not have a single course in bioinformatics, computational genomics, structural genomics, functional genomics... (nothing ending with genomics or bioinformatics for that matter). The sobering fact is that we are the only graduate school in New York City without such courses. Columbia University has courses entitled Introduction to Genomic Information Science and Technology and

Bioinformatics at the Albert Einstein College of Medicine

Computational Genomics. New York University has Computational Structural Biology, Intro to Computer Modeling and Macromolecules, Computational Molecular Biology, and Bioinformatics. City University of New York has ... you get the idea.

AECOM handles the issue by incorporating small bioinformatics topics into graduate courses, such as Molecular Genetics, Gene Expression, Mammalian Genetics, Proteomics Analysis, and Biophysics of Macromolecules. But are those bits and pieces sufficient to prepare students for tackling the complexity of something like the University of California Santa Cruz Genome Browser on our own? Discussions and meetings for organizing a bioinformatics course at AECOM date as early as January, 2000. Despite all good intentions, four years later we are still without such a course. Faculty members pointed out that AECOM has no Departments of Computer Science and Mathematics, which usually facilitate bioinformatics. However, when graduate students petitioned the Graduate Division to offer a programming course and solicited Dr. Cimino to teach it, the curriculum committee rejected the proposal. This proposal was never modified and re-submitted for further evaluation (Albert Einstein College of Medicine, 2004). As a result, the majority of AECOM students have no background in programming or mathematics. Is this a valid excuse to remain where we are and ignore the development of exciting new research tools and methods? Changes are coming whether we are ready or not.

At the same time AECOM faculty members are formulating plans to have the first AECOM bioinformatics course in the spring of 2005. However, this will depend on the successful recruitment of professors with expertise in this area. On the other hand, AECOM already has a very strong Division of Biostatistics. Its members are more than qualified to provide students with math and statistical background required for successful use of bioinformatics tools. There is also the idea of opening the Yeshiva University mathematics and computational courses to AECOM students. Even better, why can AECOM not become part of the Interuniversity Doctoral Consortium? It allows for cross-registration among member institutions: Columbia University, Fordham University, New School University, New York University, Princeton University, Rutgers, State University of New York at Stony

Brook, and City University of New York (Interuniversity Doctoral Consortium, 2004).

We all agree that changes should be made. In fact "there seems to be wide agreement within both industry and academia that there are not enough scientists adequately trained in bioinformatics" (Altman, 1998). As the current recruitment of bioinformatics faculty at AECOM shows, it is difficult to find scientists with appropriate skills for independent research in this field. In practical terms, it is expensive to train individuals in bioinformatics at the post-doctoral level because it requires taking numerous graduate school courses in addition to years of Ph.D. training. A better alternative is to teach bioinformatics at the graduate level. Due to the interdisciplinary nature of contemporary scientific research, the role of Bioinformatics will continue to increase. Therefore, it is important that the AECOM community not only acknowledges this fact but also implements measures to ensure its adequate preparation for the future.

ACKNOWLEDGMENTS

Special thanks to the people who took time to talk with me and share their opinions and ideas on the subject: Scott Emmons, John Grealley, Andras Fiser, Derek Amanatullah, Kate Milova, Lu Chen, Zara Mantirosyan, Julie Herrick, Zhong-Yin Zhang, Jonathan Warner, Shaoqing Yu, Xin Zheng, Mark Yin, Steven Schwartz, Mark Chance, Mimi Kim, Todd Evans, Ruth Angeletti, Michael Prystowsky, Christopher Cimino, Vinayaka Prasad, Oyekanmi Nash, and Kami Kim.

REFERENCES

- Altman, R. (1998) A curriculum for Bioinformatics: the time is ripe. *Bioinfo.* 14: 549-550.
- Durbin, R., Eddy, E., Krogh, A., and Mitchison, G. (1999) *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge University Press, United Kingdom.
- Lesk, A.M. (1998) *Computational Molecular Biology*. Oxford University Press, New York.
- Albert Einstein College of Medicine. (2004) Programming course proposal. Retrieved February 2004, <http://cobweb.aecom.yu.edu/ooe/courses/cbe/grad-comp/>.
- Interuniversity Doctoral Consortium. (2004) Retrieved February 2004 <http://www.nyu.edu/gsas/Programs/IUDC.html>.

EJBM Commentary

Commentary presents an author's opinion, explanation, or criticism of a specific topic within science, medicine, etc. Commentary can be submitted by electronic mail (ejbm@aecom.yu.edu) or regular mail (1300 Morris Park Avenue; Forchheimer Building, Room 306; Bronx, New York 10461). Receipt of Commentary is acknowledged. Commentary is not peer reviewed and represents the viewpoint of the author on a particular issue or topic. Commentary is edited for space and clarity.