Agenda Item No. 22

AGENDA ITEM BRIEFING

Submitted by: Eddie J. Davis, Interim President
Texas A&M University

G. Kemble Bennett, Vice Chancellor and Dean of Engineering
Director, Texas Engineering Experiment Station

Elsa A. Murano, Vice Chancellor and Dean of Agriculture and Life Sciences
Director, Texas Agricultural Experiment Station

Subject: Establishment of the Center for Statistical Bioinformatics, a Joint TAMU, TEES and TAES Center

Proposed Board Action:

Authorizes the establishment of the Center for Statistical Bioinformatics, a joint TAMU, TEES and TAES Center.

Background Information:

The concept paper for the Center for Statistical Bioinformatics was discussed at the December 2006 Board of Regents meeting.

The proposed center will foster multi-disciplinary research and education in the area of statistical models and methods for the analysis of high throughput molecular data involved in life science research. The Center will initially involve researchers from the Department of Statistics (College of Science), the Department of Electrical Engineering (TEES/College of Engineering) and the Department of Veterinary Physiology and Pharmacology (TAES/College of Veterinary Medicine) and welcomes other members of the Texas A&M University community whose research focuses on statistical bioinformatics. The Center will allow Texas A&M University to compete more effectively for external resources where the evidence of institutional commitment and expertise is important. It will also provide a focus that will enhance both fundamental research in statistical bioinformatics and establish interdisciplinary links that will expedite discoveries by the university's life science community.

A&M System Funding or Other Financial Implications:

The establishment of this Center will not require any new employees or space allocation. The Center will not ask for funding from TAMU, TEES or TAES. Funding for Center operations will come from externally funded research projects.

Additional Information:

See attached proposal.

Agenda Item No. 22

TEXAS A&M UNIVERSITY
Office of the President
January 24, 2007

Members, Board of Regents
The Texas A&M University System

Subject: Establishment of the Center for Statistical Bioinformatics, a Joint TAMU, TEES and TAES Center

Authorization is respectfully requested to establish the Center for Statistical Bioinformatics, a joint TAMU, TEES and TAES Center, that will be administratively assigned to the Department of Statistics in the College of Science, as documented in the attached proposal.

The Center for Statistical Bioinformatics is designed to foster multi-disciplinary research and education in the area of statistical models for the analysis of high throughput molecular data involved in life science research.

I recommend adoption of the following minute order:

"The Center for Statistical Bioinformatics, a joint TAMU, TEES and TAES Center, is hereby established as an organizational unit of Texas A&M University in the Department of Statistics in the College of Science."

Respectfully submitted,

[ORIGINAL SIGNED BY]
Eddie J. Davis
Interim President
Texas A&M University

[ORIGINAL SIGNED BY]
G. Kemble Bennett
Vice Chancellor and Dean of Engineering
Director, Texas Engineering Experiment

[ORIGINAL SIGNED BY]
Elsa A. Murano
Vice Chancellor and Dean of Agriculture and Life Sciences
Director, Texas Agricultural Experiment Station

Approval Recommended:

[ORIGINAL SIGNED BY]
Michael D. McKinney
Chancellor

[ORIGINAL SIGNED BY]
James G. Hooton
Executive Vice Chancellor for Finance

Approved for Legal Sufficiency:

[ORIGINAL SIGNED BY]
Cullen M. Godfrey
General Counsel

[ORIGINAL SIGNED BY]
Leo Sayavedra
Vice Chancellor for Academic and Student Affairs

Proposal

Center for Statistical Bioinformatics
Texas A&M University, Texas Engineering Experiment Station, and
Texas Agricultural Experiment Station

1. Rationale for creation of the Center for Statistical Bioinformatics

Mission

The Center for Statistical Bioinformatics will foster multi-disciplinary research and education in the area of statistical models and methods for the analysis of high throughput molecular data involved in life science research. It will involve researchers from the Department of Statistics (College of Science), the Department of Electrical Engineering (TEES/College of Engineering) and the Department of Veterinary Physiology and Pharmacology (TAES/College of Veterinary Medicine), all of whom are active in the area of Statistical Bioinformatics. Researchers from TAES/College of Agriculture and Life Sciences have also expressed interest in the Center.

There will be 13 initial members of the Center for Statistical Bioinformatics. The goal is to be inclusive and to welcome other members of the Texas A&M University community whose research focuses on Statistical Bioinformatics.

The Center will allow Texas A&M University to compete more effectively for external resources where the evidence of institutional commitment and expertise is important. It will also provide a focus that will enhance both fundamental research in statistical bioinformatics and establish interdisciplinary links that will expedite discoveries by the university's life science community.

Statistical Bioinformatics And An Example of What It Is

Statistical bioinformaticians at Texas A&M develop new statistical methods to analyze genomic data. Besides pre-processing data, they develop statistical methods to predict who will get a disease, whether and how cancers fall into subtypes, how genes are regulated in a network fashion (how a change in one gene affects other genes), how genes interact with our environment and how the choices in lifestyle, such as nutrient intakes, affect disease, etc.

Bioinformatics refers to the processing and analysis of the massive amounts of data that are now being generated by genomic scientists. Technologies such as microarray, protein mass spectroscopy, gene sequencing, haplotypes and SNP (single nucleotide polymorphism) mapping have led to an explosion in the amount of data that are available. Bioinformatics is the catch-all phrase for computational, statistical and mathematical techniques that attempt to use this information in ways that advance basic science and human health.

Bioinformatics itself is an extremely broad term. *Statistical Bioinformatics* is the focus of the proposed center. It is the branch of Bioinformatics that focuses on the development and application of statistical approaches to genomics data, such as Bayesian methods, signal processing, wavelet methods, gene-environment interaction studies and functional data approaches. Statistical Bioinformatics is a young field, but it is extremely important for helping

biologists interpret their data in ways that take into account the inherent uncertainties arising from such data.

The biological terms listed above, such as haplotypes, can be fairly daunting and even their pronunciation can be a mystery. For example, for no obvious reason SNP is pronounced as "snip". Haplotypes are just as mysterious and have no obvious meaning in everyday English: they are the genetic constitution of an individual chromosome, or roughly what an individual inherits from her/his individual parents.

An example of Texas A&M work in statistical bioinformatics may help the Board understand what the university is doing in this area. With funding from NASA and the National Cancer Institute, researchers in the Department of Nutrition and Food Science have developed a novel way to measure gene expression (what is going on with the genes) non-invasively. Their goals are to understand the effects of diet on colon cancer, and even more exciting, their goal is to alert physicians early on to the possibility of colon cancer in an individual without having to perform unpleasant procedures such as a colonoscopy. Their patented technology uses fecal material to measure gene expression. The hope is that people who are on their way to developing colon cancer have distinct patterns of gene expression that are different from those who are at little risk for developing colon cancer.

Where Statistical Bioinformatics comes in is in translating this dream of non-invasive screening for colon cancer into the reality of actual numerical predictions of just who is likely to be at great risk. Three members of the proposed Center for Statistical Bioinformatics have been the primary statistical researchers analyzing these data. They have developed novel statistical ways of dealing with gene expressions that are missing because of technological issues involving the fecal material: without this advance the data could not have been used and hundreds of thousands of dollars in research expenditures would have been wasted. They have developed statistical methods that can determine which genes are affected by diet, and whether such effects depend on radiation exposure. Vitally, using rat models, Texas A&M researchers have developed statistical methods that can classify individuals early on whether they are at risk of progressing to polyps and eventually colon cancer. Human data is in the process of being collected so that the statistical methods can be applied to classification of humans.

Justification

Texas A&M University is already extremely strong in Statistical Bioinformatics, having faculty members residing in three colleges, and with principal investigators of 14 competitive National Institutes of Health and National Science Foundation Grants. Texas A&M runs the only National Cancer Institute training program focusing on Bioinformatics and Nutrition, are principals in a Genomics facility core within the Center for Environmental and Rural Health, and make up a significant portion of the faculty members at Texas A&M University identified through the University's Bioinformatics web site.

There are 13 faculty members at Texas A&M University who work in the area of Statistical Bioinformatics and who are interested in joining the proposed Center. More than half of these faculty members are involved in the Bioinformatics and Nutrition Training Program. Recognizing this core group of faculty, the aim of the Center is to help facilitate further development in order to help Texas A&M maintain its leadership in the area.

List of Current Members

The Center for Statistical Bioinformatics currently has identified the following members.

- Raymond J. Carroll (Statistics)
- David Dahl (Statistics)
- Aniruddha Datta (TEES/Electrical Engineering)
- Edward Dougherty (TEES/Electrical Engineering)
- Ruzong Fan (Statistics)
- Jeff Hart (Statistics)
- Ivan Ivanov (TAES/Veterinary Physiology and Pharmacology)
- Erning Li (Statistics)
- Bani Mallick (Statistics)
- Samiran Sinha (Statistics)
- Clifford Spiegelman (Statistics)
- Marina Vannucci (Statistics)
- Naisyin Wang (Statistics)
- Researchers from TAES/College of Agriculture and Life Sciences have also expressed interest in the Center.

Proposed Activities

A major activity of the Center will be to continue to run the currently established Bioinformatics Seminar Series, which brings in leading statistical bioinformaticians from around the world for visits. Financial support from the three Colleges involved in the Bioinformatics and Nutrition Training Program helps to support this series, with funding guaranteed through 2011.

In addition to the seminar series, the Center will establish a bi-yearly 2-day *Workshop on Recent Trends in Statistical Bioinformatics*. A prototype of this workshop occurred on October 28, 2006. The workshop in 2008 and 2010 will be funded by the Bioinformatics and Nutrition Training Program. The Center will seek additional financial support from the National Institutes of Health and the National Science Foundation for this workshop: no internal funds will be sought.

The Department of Electrical Engineering runs a bi-yearly workshop on *Gene Regulatory Networks*, organized by two members of the proposed Center. This workshop is funded by registration fees.

The Center will maintain a web page, a prototype of which is already online (http://statbio.stat.tamu.edu). This web page will provide announcements, webinars, etc. The Center will also help advertise the research work already ongoing in four major research groups:

- The Bayesian Bioinformatics Laboratory: http://www.stat.tamu.edu/~bmallick/Bayeslab/index.html
- Genomic Signal Processing Laboratory: http://gsp.tamu.edu
- Bayesian Methods for Genomics Research Group http://www.stat.tamu.edu/~mvannucci/webpages/research.html
- The Laboratory for Statistical Bioinformatics in Nutrition and Cancer http://www.stat.tamu.edu/~carroll/carroll.lab

Because of the extensive experience this group has in setting up web sites, and the resources of the Department of Statistics, the Center will also help Center members set up their own laboratory web pages.

2. Impact on Education and Training of Students

The Center's major impact on the education and training in Statistical Bioinformatics comes through its supervision of Ph.D. students and postdoctoral associates. The identified members of the Center currently supervise 26 Ph.D. students and 9 postdoctoral associates, see Table 1.

Table 1: Numbers of Ph.D. students and postdoctoral associates in Statistical Bioinformatics currently supervised by identified members of the Center.

Department	Number of supervised Ph.D. students	Number of postdoctoral associates
Electrical Engineering	6	3
Statistics	20	6

As described above, there are other explicit forms of training that is currently provided.

- The Bioinformatics seminar series meets thrice monthly, and is regularly attended by graduate students, who can register for 1-course credit.
- The Department of Electrical Engineering runs a bi-yearly workshop on *Gene Regulatory Networks*, organized by two members of the proposed Center.
- In the spring semester 2007, Center members in the Department of Electrical and Computer Engineering, the Department of Veterinary Physiology and Pharmacology, and the Department of Statistics will offer courses in Statistical Bioinformatics.
- The Genomic Signal Processing Laboratory within the Department of Electrical Engineering regularly runs workshops and conferences on genomic signal processing and statistical bioinformatics.
- The Department of Statistics organizes a bi-annual workshop on statistical bioinformatics.

3. Sources and Future Expectations of Financial Support

The Center for Statistical Bioinformatics would neither request nor require any additional financial support.

Sources of Financial Support

Administrative support for the Center would be provided by an existing staff member in the Department of Statistics who works as Dr. Raymond Carroll's administrative assistant and whose salary is currently $44,000 per year.

Faculty supporting the Center currently organize and run a Bioinformatics and Nutrition Training Program (http://www.stat.tamu.edu/b3nc/), funded by the National Cancer Institute of the National Institutes of Health. The Program receives approximately $30,000 per year in total in matching financial support from the Colleges of Science, Engineering and Agriculture and Life Sciences. The training grant and the matching funds provide financial support for various activities, especially the seminar series.

The Bioinformatics and Nutrition Training Program was renewed recently, and has received funding through at least the end of 2011. The University expects to apply for a further renewal in 2010.

External Financial Support

Currently, the identified members of the Center are principal investigators on 14 National Science Foundation (NSF) or National Institutes of Health (NIH) grants, contracts and training programs. In FY2006, these grants contracts and training programs received $2,885,620 from the NSF and NIH in total. See Table 2 for more details.

Table 2: Grant and contract funding in FY2006 for currently identified members of the Center for Statistical Bioinformatics. NIH = National Institutes of Health. This refers only to grants and contracts for which the members are the Principal Investigators, and does not include impact on such major grants as the Center for Environmental and Rural Health ($1,000,000+) or when members serves as investigators on others' grants.

Source of Funding	Number of Grants	Total Grant Awards in FY2006
NSF Research Grants	8	$ 680, 228
NIH Research Grants	4	1,009,720
NIH Training Grants (Bioinformatics and Nutrition Training Grant)	1	529,672
NIH Contracts (Proteomics Initiative)	1	666,000
Total	**14**	**$ 2,885,620**

Future Financial Support

Administrative support for the Center will be provided by the Department of Statistics. Future funding for seminar series and visitors is assured through the end of 2011. Over $2,800,000 in grant and contract support will only rise in the future.

4. Governance and Advisory Structure

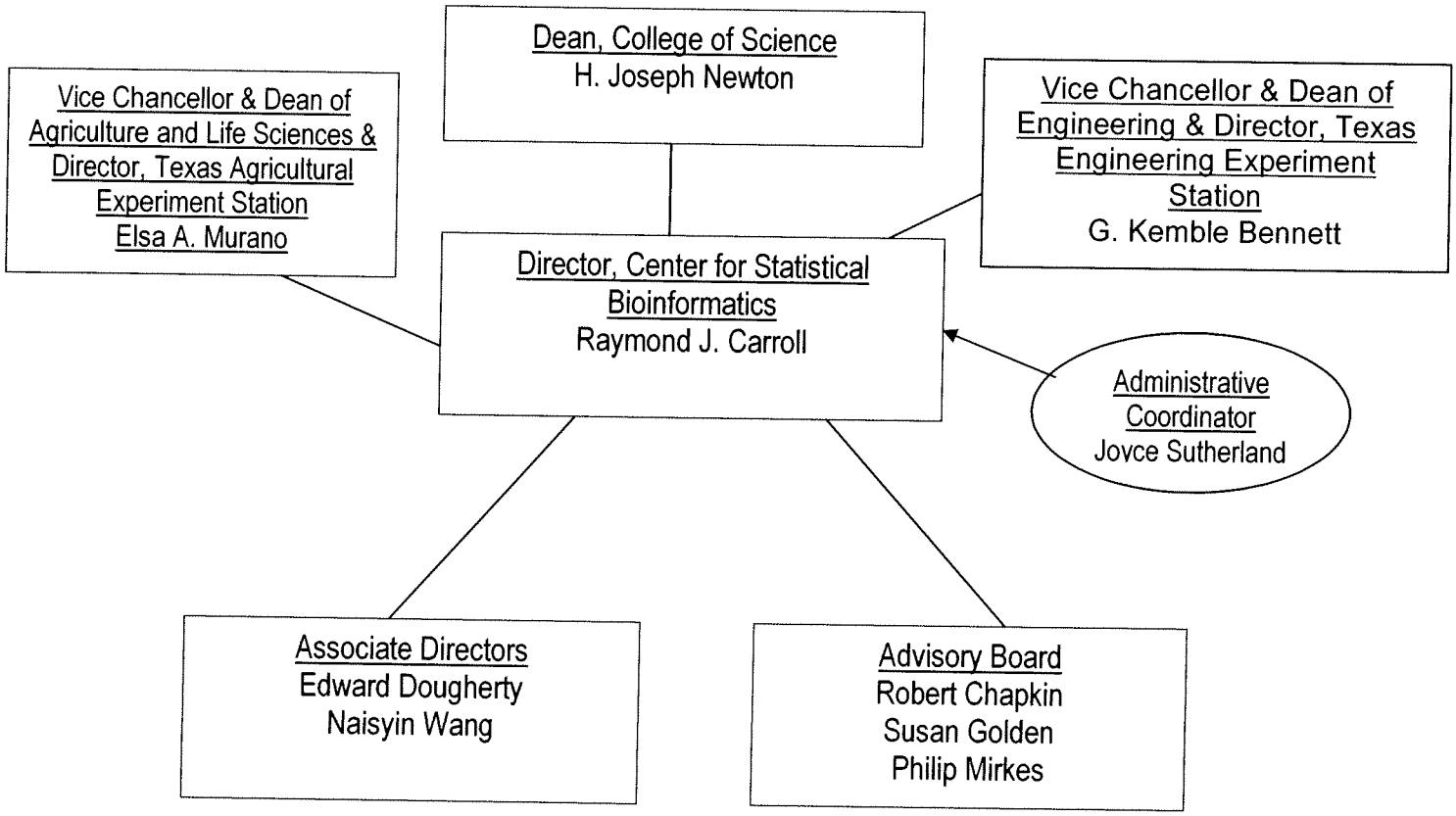
The Center will be led by a Director who will report to the Head of the Department of Statistics at Texas A&M University. The Director will be appointed by the Dean of the College of Science, in consultation with the Vice Chancellor and Dean of Engineering and Director, TEES and the Vice Chancellor and Dean of Engineering and Life Sciences and Director, TAES . The Center Director will meet yearly with the Advisory Board (see below), and provide on a yearly basis a written report to the Dean of the College of Science and the Vice Chancellor and Dean of Engineering and Director of TEES, the Vice Chancellor and Dean of Agriculture and Life Sciences and Director of TAES, the Advisory Board and the Center members.

The Center's initial director will be Dr. Raymond J. Carroll, Distinguished Professor of Statistics, Nutrition and Toxicology. Dr. Carroll is the P.I. on the Bioinformatics and Nutrition Training Grant, and is the only statistician in the country to have received a MERIT Award from the National Cancer Institute. Dr. Edward Dougherty (Department of Electrical Engineering) and

Dr. Naisyin Wang (Department of Statistics) will be the initial Associate Directors. Dr. Carroll will organize the day-to-day activities of the CSB, with regular input from the associate directors. Because many of the proposed members of the Center are also involved in the Bioinformatics and Nutrition Training Program, there will be considerable informal communications.

The Center will also have an Advisory Board, consisting of three internal biologists: (a) Dr. Robert Chapkin, Professor of Nutrition and Food Sciences and head of the Genomics Facility Core at the Center for Environmental and Rural Health; (b) Dr. Susan Golden, Distinguished Professor of Biology and Director of the Center for Biological Clocks Research; and (c) Dr. Phillip Mirkes, Professor of Veterinary Physiology and Pharmacology and Director of the Center for Environmental and Rural Health. These advisors will help in the selection of seminar speakers and speakers at workshops, and will also provide advice on the biological relevance of the Statistical Bioinformatics research in the Center. The Advisory Board members will serve for three years. Members can be reappointed, subject to the approval of the Dean of the College of Science and in consultation with the Vice Chancellor and Dean of Engineering and Director of TEES and the Vice Chancellor and Dean of Agriculture and Life Sciences and Director of TAES, as well as with the consultation of the heads of the Departments of Electrical and Computer Engineering and the Department of Statistics. New members will also serve three year terms, and will be selected in the same manner as for reappointments.

Proposed Administrative Structure, Center for Statistical Bioinformatics



5. Mechanism for Periodic Review

In addition to the annual report described above, every third year a panel will be convened to review the Center. The main purpose of this panel will be to decide whether the Center should be continued. It will also provide advice on how the Center might better function and whether its activities fit into the mission of the departments involved and more generally the College of Science, the College of Engineering, the TEES, the College of Agriculture and Life Sciences, and the TAES. The panel will report its recommendation on the continuation of the Center directly to the Dean of the College of Science in addition to the Vice Chancellor and Dean of Engineering and Director of TEES, and the Vice Chancellor of Agriculture and Life Sciences and Director of TAES. It will report its advice and suggestions to the Director and Associate Directors.

The panel will include the Advisory Board detailed above, one non-administrative member of the Center, and two statistical bioinformaticians from outside Texas A&M University[1]. The panel will be chaired by a member of the advisory board, to be selected by the dean of the College of Science in consultation with the Vice Chancellor and Dean of Engineering and Director of TEES, and the Vice Chancellor of Agriculture and Life Sciences and Director of TAES.

[1]The following statistical bioinformaticians have agreed to serve as the outside members of the panel: Dr. Xihong Lin of the Department of Biostatistics at Harvard University, and Dr. Jeffrey Morris of the Department of Biostatistics at the M. D. Anderson Cancer Center.