

# Center for Health Informatics and Bioinformatics Committees/Initiatives Progress Report & Update

---

*Program Director: Constantin Aliferis, MD, PhD*

*January 4, 2010*

## **Committee/Initiative: The informatics Best Practices Integrative Consultation Core (BPIC)**

**Members:** Constantin Aliferis, Alexander Alekseyenko, Yin Aphinyanaphongs, Stuart Brown, John Chelico, Lawrence Fu, Yuval Kluger, Jim Robinson, Ross Smith, Alexander Statnikov, Marc Triola, Jinhua Wang, Jiri Zavadil

**Update:** The informatics faculty has met to design the structure of BPIC on several occasions during regular faculty meetings. Constantin Aliferis has met at least 10 times with business management staff to develop business plans and reimbursement mechanisms. Dr Aliferis has discussed BPIC structure and operations with Vice Dean Lee during several of their regular informatics center progress meetings. Dr Aliferis has also discussed BPIC operations during two meetings with Associate Vice Dean Levy and on many one-to-one meetings with department/division/CoE chairs or directors and other PIs.

BPIC members have held 4 meetings since November 2009 to discuss BPIC goals and to establish guidelines and implementation of the Core. An additional meeting is scheduled for January 2010 to finalize and further refine BPIC protocol through conduct of mock consultation sessions. The BPIC has already provided an estimated 100 consultations to various researchers in the Medical Center since November 2009. 5 of these consultations have utilized the new consultation forms and 3-stage process that are in development.

**Next Steps:** A BPIC operations manual and official policies will be drafted and presented to Vivian Lee, Vice Dean for Science, for approval and medical center-wide implementation. The BPIC is slated to launch its walk-in format stage 1 consultation which will include thrice-weekly, open-door consultation sessions for both Bioinformatics and Health Informatics consultation needs for researchers at the Medical Center.

## **Committee/Initiative: Comprehensive educational activities**

Members: Drs. Aliferis, Statnikov, Kluger, Brown.

### Update:

- we offered 3 informatics tutorials to faculty, students, and researchers; online materials are available at <http://www.nyuinformatics.org/education/seminars-and-tutorials>.
- We launched an invited speaker series with 4 speakers in 2009; online slides and videos are available at <http://www.nyuinformatics.org/education/speakers>.
- We launched a research colloquium series with 4 talks delivered in 2009; online slides can be found at <http://www.nyuinformatics.org/education/research-colloquia>.
- Informatics faculty are offering 2 graduate level courses and one Residency training module in translational informatics. Information can be found at <http://www.nyuinformatics.org/education/courses>.
- Informatics faculty are primary mentors for trainees at the following levels: one research Fellow, two post-doctoral fellows, one MD/PhD student, one PhD student, one MS student, one lab rotation student and one high school intern.

Next Steps: In 2010 we will offer two tutorials in Next generation sequencing informatics and Statistical genetics. We will continue all the above programs and continue training students at all levels.

## **Committee/Initiative: NYU Graduate Training Program in Biomedical Informatics (GTP)**

Members: Dr. Aliferis is leading this effort with guidance by Dean Lee. Several faculty will be closely involved in several new sub-committees, starting with the steering committee and the curriculum committee.

Update: Dr Aliferis has had three meetings with Dean Abramson, two meetings with Dean Davidson, two meetings with Dr. Cardoso, two sessions in the TREC executive steering committee, one meeting with the Vice Deans for medical school curriculum and student affairs, and two meetings with Library leadership. Several meetings are scheduled to take place to coordinate the program development with Currant, Computational Biology, Nursing Informatics, and other relevant programs and complementary initiatives. The steering committee and the curriculum committee will start operations in January 2010.

## **Committee/initiative: High-Performance Computing Parallel Cluster Facility (HPFC)**

Members: Constantin Aliferis, Alexander Alekseyenko, Stuart Brown, Nader Mherabi, Ross Smith, Robert Sprinkle, Alexander Statnikov.

Update: Phase I design and implementation and Phase II design have been completed. Two candidates for HPCF Director have been interviewed and currently a hiring decision is being made.

Next Steps: Hire an HPCF Director and implement Phase II.

## **Committee/Initiative: Laboratory Information Management System (LIMS)**

Members: Constantin Aliferis, Steven Blais, Stuart Brown, Jane Carlton, Ramanuj Dasgupta, Richard Hayes, David Levy, Thomas Neubert, Carol Oddoux, Harry Ostrer, James Robinson, Baljit Singh, Ross Smith, Robert Sprinkle, Fred Valentine, Jiri Zavadil

Update: The LIMS committee has held 4 meetings since June 2009 to determine criteria for purchase of a LIMS and identify appropriate vendors for product review. Proposals have been solicited and received from 7 vendors. These proposals have been reviewed by the committee and recommendations have been gathered to determine which vendors to invite for formal in-house presentations.

Next Steps: Invite LIMS vendors to present products in-house beginning in January 2010.

## **Committee/Initiative: Center for Health Informatics and Bioinformatics Website -**

Members: Constantin Aliferis, Alexander Alekseyenko, Yin Aphinyanaphongs, Stuart Brown, John Chelico, Lawrence Fu, Yuval Kluger, Jim Robinson, Ross Smith, Alexander Statnikov, Marc Triola, Jinhua Wang, Jiri Zavadil, Michael Mainiero, Matthew Lisowski, Heewon Chu, Huming Tang

Update: A comprehensive website for the Center for Health Informatics and Bioinformatics, [www.nyuinformatics.org](http://www.nyuinformatics.org), was launched in October 2009.

Next Steps: This website is continually updated as new information and initiatives develop.

## **Committee/Initiative: Faculty Hires and Faculty Development**

**Members:** Dr Aliferis, Dean Lee, Dr Blaser, Dr Cronstein, Dr Sabbatini, Dr Roth, Dr Carroll, with input from several Division Chiefs, users or informatics services, and heads of units that house informatics faculty.

**Update:** To-date we have hired 5 new informatics faculty in critical new methods innovation and service areas. This has involved bringing 10 candidates to NYUMC for extensive interview schedules, scheduling and advertising of candidate talks, developing competitive 3-year business plans for each candidate, and negotiating hiring and offer details with various academic departments where new faculty hold appointments.

In addition we brought together 17 total faculty (12 full-time, 3 part-time, 2 consulting) and fostered a common academic research, teaching and educational environment.

The PD is mentoring several new and more senior faculty members as well as staff and have helped them improve their dossiers. Dr Aliferis is a member of 9 faculty mentoring committees and serves as primary mentor to 5 of them. Dr Aliferis is meeting with mentored faculty from once/week to once/month depending on faculty needs.

As a result of the active mentoring of faculty and staff we have achieved a number of career advancements including:

- (a) exceeding the Dean's standards for grant support for 2 faculty members that had previously very limited grant support;
- (b) meeting national informatics standards for promotion for 2 faculty;
- (c) diversifying the activities and skillset of several staff and giving merit raises to 2 of them;
- (d) ensuring appropriate reimbursments for year#1 of the operation of the Clinical and Research Informatics Data Management Unit;
- (e) correcting salary disparities for 2 pre-existing faculty and giving them merit increases/AAMC salary matching increases;
- (f) one of the new faculty hires has already met his minimal 3-year goals in terms of grant support as well as academic productivity;
- (g) all new faculty members have very promising and expanding academic productivity, project involvement/collaborations and grant participation.

**Next Steps:** We have approval from Vice Dean Lee to hire an additional 3 faculty members. In addition Dr Aliferis has accepted a charge by Dean Abrmson to lead an advising committee for informatics-specific RPT criteria suitable for NYU SoM. This group will start its operation in February 2010. We will also proceed to work with Department Chairs for finalizing our two recommended promotions to Associate Professor for Drs Kluger and Triola.

## **Committee/Initiative: Federated Data Warehouse (FDW) user needs assessment**

**Members:** Constantin Aliferis, John Chelico, Ross Smith

**Update:** The FDW committee has held 4 meetings since August 2009 to determine data warehousing needs at the Medical Center. Constantin Aliferis has also met at least 4 more times with Heidi Fitterling and Jackie Price to coordinate data collection and analysis. Two surveys were conducted, one sent to a select group of data users and the second sent to the entire Medical Center community eliciting feedback and one face-to-face information gathering session with super-users was held. Dr Aliferis had 4 additional meetings to discuss with Dr Chelico his role in providing FDW user support needs. A meeting with Dean Conocenti, and Drs Gourevitch, Triola, Chelico discussed Dr Chelico's potential roles. Dr Aliferis and Dean Lee have developed a straw man plan for FDW/data mining capabilities.

**Next Steps:** From the information gathering phase, it was determined that a team will be assembled to manage ad-hoc queries, which will be used to document data warehousing needs. Several of these queries will be piloted using a couple of potential data warehousing models to prepare real-life model comparisons with respect to cost, time, quality, etc. Drs Aliferis, Chelico and Smith are preparing a report to Dean Lee and an update to the RCIT committee.

## **Committee/Initiative: Molecular Signatures Lab**

**Members:** Constantin Aliferis, Alexander Alekseyenko, Jizhou Ai

**Collaborators:** Alexander Statnikov, Douglas Hardin, Frank Harrell, Jonathan Schildcrout, Isabelle Guyon, Ioannis Tsamardinos, Subramani Mani

**Update:** The Molecular Signatures Lab at CHIBI (MSL) researches data analytic/bioinformatics theory, algorithms, data analysis protocols and software systems for the development of optimally predictive, safe, parsimonious and cost-effective molecular signatures. Research carried out by Dr. Aliferis and his colleagues includes proving theorems that connect predictivity with mechanistic (causal interpretation) when selecting genes and proteins to build signatures, creation of new sample and computationally efficient algorithms for discovery of biomarkers and development of maximally compact and predictive signatures, identifying sources of data analysis errors and ways to correct them in major studies, algorithms to deal with the multiplicity of molecular signatures and biomarkers, algorithms for discovery of biologically interpretable biomarkers, intelligent analysis systems for the automatic development of reproducible signatures (GEMS and FAST-AIMS), extensive benchmarking studies of classifier and gene/protein selection algorithms, causal characterization of various data analysis algorithms used in molecular signature research and information-retrieval models for storing and retrieving information about molecular signatures and other high-throughput molecular medicine modalities.

The MSL pursues these areas of research, as well as the development and study of guidelines for reliable data analysis of omics data. The MSL collaborates very closely with the Computational Causal Discovery lab (led by Dr. Statnikov) and with the departments of Pathology and Medicine, The Genome Center and Genetics for developing novel signatures, as well as to eventually deliver molecular signatures at the bedside. Research findings from MSL research inform the recommendations of the BPIC and software produced is deployed by BPIC when appropriate in research projects on campus. Software and methods from MSL are made available broadly to the scientific community outside NYU.

## **Committee/Initiative: Computational Causal Discovery Lab**

**Members:** Alexander Statnikov, Constantin Aliferis (advisor), Nikita Lytkin, Jizhou Ai, Varun Narendra

**Collaborators:** Laura Brown, Gregory Cooper, Isabelle Guyon, Douglas Hardin, Frank Harrell, Lewis Frey, Subramani Mani, Pierre Massion, Ioannis Tsamardinos

**Update:** Extensive research in recent years has shown the critical importance of graph-theoretic, computational causal discovery methods for many fields – and, in particular, with respect to the analyses undertaken within NYU Center for Health Informatics and Bioinformatics on behalf of NYULMC researchers. Several varieties exist with causal Bayesian Networks being a dominant paradigm. For example, causal methods can identify biomarkers that are targets for new drug development or SNPs that likely cause disease; they can indicate membership of genes and proteins in disease-related pathways; and they can lead to more compact and better generalizable molecular signatures compared to the first-generation and less sophisticated differential expression/associative/ad hoc or clustering methods. The purpose of the Computation Causal Discovery laboratory is to develop, test and apply causal methods suitable for clinical, molecular, imaging and multi-modal data of high-dimensionality.

## **Committee/Initiative: Evidence Based Medicine Information Retrieval and Scientometrics Lab**

**Members:** Lawrence Fu, Yindalon Aphinyanaphongs, Constantin Aliferis (advisor)

**Collaborators:** Karen Brewer, Alexander Statnikov, Brian Haynes

**Update:** The sheer volume of scientific information in the biomedical literature and on the Internet makes manual inspection of literature searches prohibitive even for narrow clinical and research questions. This abundance of information slows the rate of research discoveries (especially translational ones that are slower by virtue of having to cross disciplinary boundaries) and propagates low-quality information to patients via the internet. EIRSL work has recently developed pattern recognition-based filtering methods that can automatically identify the content and quality of both web pages and scientific articles. The models hold the promise of expediting literature search and synthesis by focusing on content-specific articles of the highest methodological quality.

Other EIRSL work has shown that it is possible to augment traditional bibliometric quality measures such as citation count and impact factor using machine learning approaches. In particular, machine learning methods accurately predict citation counts in a deep horizon (10 years after publication) using only data available at publication time. Predicted citations counts could be a powerful filter to focus attention on recent publications that are more likely to influence new scientific and clinical developments.

Machine learning methods can also accurately characterize the nature of a citation as being essential or not. Citation counts can be adjusted by discarding citations that are not important to the citing papers since many highly-cited papers gather numerous citations due to factors other than quality (e.g., may be cited by rebuttal and refutation papers etc.). EIRSL will advance the state of the art in these exciting areas by extending the number and scope of these models and delivering them to both researchers and patients through collaborations with the Ehrman Medical Library to evaluate the effectiveness of these methods and to guide their future improvement.

EIRSL will also develop new technology using pattern-recognition approaches to accurately identify content and quality of MEDLINE and web documents as well as to enhance traditional bibliometric criteria such as the impact factor and citation counts. The motivation for the development of these techniques is to identify higher-impact, higher-quality research findings more quickly to accelerate their use for both research and clinical care.

## **Committee/Initiative: Clinical and Research Informatics Data Management Unit**

**Members:** Constantin Aliferis, James Robinson, Ross Smith, Stuart Brown, Paul Conocenti, Nader Mherabi, Bobby Sprinkle, Bruce Cronstein, Judy Hochman).

**Update:** We announced the establishment of the Clinical and Research Informatics Data Management Unit on November 3, 2009 to provide comprehensive data management and database development capabilities at NYULMC.

Leading to launching the new unit, Velos, a commercial clinical research data system widely used by medical centers and academic institutions was purchased to be the core clinical research software product across the NYULMC. Velos is a comprehensive product with several clinical research modules and functions. The goal of this new unit is to provide an infrastructure that allows individual clinical research investigators as well as departments, divisions, and labs to develop study databases, data acquisition and management systems, and research registries using software that is compliant with all state-of-the-art requirements and national standards for secure data management. The Clinical and Research Informatics Data Management Unit consists of experienced data managers and programmers working under the direction of Jim Robinson. The new unit will function as a part of the Center for Health Informatics and Bioinformatics within the broad umbrella of CTSI activities and in close collaboration with the MCIT department which was instrumental in developing the unit's function and shares implementation responsibilities for Velos.

Because of the size and scope of the project and the comprehensiveness of Velos, the first several months will be dedicated to internal Velos training and, developing a comprehensive external training program for departments, research investigators and their teams. During this time we will also establish policies and procedures, develop standards, begin a "Forms Library" that allows researchers to select previously developed and programmed forms, and conduct a small number of projects to fully explore as many of the Velos capabilities as possible.

This unit is available to work with investigators in the preparation of grants and proposals. The unit can also provide comprehensive clinical and research database systems development, and data management services on a fee for service basis for investigators who want to outsource clinical research study support. Finally for researchers with multi-modal data that span both clinical and molecular information the unit works with CHIBI faculty with experience in molecular data storage and management to provide custom solutions.