2015 Annual Report

Department of Biomedical Informatics
University of Pittsburgh School of Medicine

[School of Medicine Version]

Prepared For:

Arthur S. Levine, M.D.
Senior Vice Chancellor for Health Sciences and Dean,
University of Pittsburgh School of Medicine

October 1, 2015
A. TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mission Statement</td>
<td>3</td>
</tr>
<tr>
<td>Annual Report 2015-2015 Introduction</td>
<td></td>
</tr>
<tr>
<td>2014-2015 Introduction</td>
<td>4</td>
</tr>
<tr>
<td>Key Future Goals</td>
<td>6</td>
</tr>
<tr>
<td>Research &amp; Other Scholarly Activities</td>
<td></td>
</tr>
<tr>
<td>Trends in Research Support</td>
<td>7</td>
</tr>
<tr>
<td>Research Grant Growth</td>
<td>7</td>
</tr>
<tr>
<td>Summary of Research</td>
<td>8</td>
</tr>
<tr>
<td>New Research Initiatives</td>
<td>37</td>
</tr>
<tr>
<td>Major Collaborations (Outside of the University)</td>
<td>38</td>
</tr>
<tr>
<td>Research Faculty: Summary of Interests &amp; Initiatives</td>
<td>45</td>
</tr>
<tr>
<td>Summary of Research Grants by Faculty</td>
<td>58</td>
</tr>
<tr>
<td>Entrepreneurial Activities</td>
<td>61</td>
</tr>
<tr>
<td>Teaching Activities</td>
<td></td>
</tr>
<tr>
<td>Training Program / Graduate Student Education</td>
<td>64</td>
</tr>
<tr>
<td>PhD, Masters and Certificate Graduate Students</td>
<td>65</td>
</tr>
<tr>
<td>Departing Fellows</td>
<td>69</td>
</tr>
<tr>
<td>Graduate Student Researchers</td>
<td>70</td>
</tr>
<tr>
<td>Post-Docs</td>
<td>71</td>
</tr>
<tr>
<td>Departing Post-Docs</td>
<td>72</td>
</tr>
<tr>
<td>Faculty Data</td>
<td></td>
</tr>
<tr>
<td>Faculty Roster, AY 2014-2015</td>
<td>73</td>
</tr>
<tr>
<td>Faculty Biographies</td>
<td>75</td>
</tr>
<tr>
<td>Bibliography</td>
<td>99-116</td>
</tr>
<tr>
<td>Staff Listing</td>
<td></td>
</tr>
<tr>
<td>Administrative Staff</td>
<td>117</td>
</tr>
<tr>
<td>Senior Research Staff</td>
<td>118</td>
</tr>
<tr>
<td>Technical Staff</td>
<td>118</td>
</tr>
<tr>
<td>Financial Plan</td>
<td></td>
</tr>
<tr>
<td>FY 2015 Financial Summary</td>
<td>120-121</td>
</tr>
<tr>
<td>Five Year Plan</td>
<td>122</td>
</tr>
<tr>
<td>FY 2016 Financial Projections</td>
<td>123</td>
</tr>
<tr>
<td>Schedule of Revenue and Expenses FY 2015 Budget</td>
<td>124</td>
</tr>
</tbody>
</table>
B. ACTIVITIES AND MISSION

**DEPARTMENT OF BIOMEDICAL INFORMATICS**
**MISSION STATEMENT**

- To provide **national and regional leadership in innovation through research** in Informatics.

- To provide the **highest quality of training in Informatics** and to provide our students with a world-class education that prepares them to become outstanding leaders in biomedical informatics research, education, and practice.

- To provide the **highest quality of support for the clinical practice of medicine** through regional and nationally recognized Informatics leadership in Clinical Informatics.
2014-2015 Introduction

The Department of Biomedical Informatics (DBMI) at the University of Pittsburgh School of Medicine covers all of the research domains of biomedical informatics including bioinformatics, imaging informatics, clinical informatics and public health informatics as described by Friedman et al, JAMIA 2004:

a. Bioinformatics – The application of the science of Biomedical Informatics to cells and molecules
b. Imaging Informatics – The application of Biomedical Informatics to organs and tissues
c. Clinical Informatics – The application of Biomedical Informatics to patient care
d. Public Health Informatics - The application of the science of Biomedical Informatics to populations and society.

DBMI’s core faculty include significant national leadership and expertise in clinical and translational science informatics, biosurveillance and public health informatics, machine learning and biostatistics, pathology/oncology informatics, information extraction from free text, bioinformatics, vocabularies and ontologies, imaging informatics and human computer interaction.

As of June 30, 2014, the Department of Biomedical Informatics consisted of sixteen core faculty. Effective July 1, 2015, Dr. Steven Handler, M.D., Ph.D. transferred within the University of Pittsburgh to the Department of Geriatric Medicine, and Dr. Roger Day, Sc.D. retired, therefore the number of DBMI faculty reduced to fourteen for FY2015. However, the initiation of several junior faculty rising from Post-Doctoral or Research Scientist positions, as well as a major recruitment initiative, should produce a number of new faculty that will be reported in the FY2016 report.

For FY2015, the Department had $12,643,462 in research revenue and $425,813 in other operating revenue from teaching and service contracts. With research funding remaining at a stagnant level due to the difficult financial environment nationally, the department actually exceeded budgeted expectations for FY2015.
Direct grant expenditures for FY2015 were $3,412,489 higher than FY2014, and the indirect revenue associated with those grants was $786,666 higher.

The overall total of faculty members dropped from 16 faculty member in FY14 to a total of 14 members in FY15. The total funding of each faculty member increased by 71.30% from $513K per faculty member in FY2014, to $879K per faculty member in FY2015. Our current projected FY2016 budget is $15.9 M which is a greater than a 31% increase in annual grant revenue from actual FY2015.
Key Future Goals

1. Research goals are to continue to increase, or at least maintain current funding levels, especially in the current financial climate. A few new large grants for FY2016 are being launched including **Big Data for Better Health (BD4BH)** funded by the Commonwealth of Pennsylvania, Department of Health CURE grant program, (Cooper Co-PI) and **the Center for Commercial Applications (CCA) of Healthcare Data (Becich, Director)**, a UPMC Enterprises (UPMCE) funded academic commercialization effort making expectations for FY2016 higher than in previous years. The goal going forward will be to sustain NIH funding at approximately $15M per year and diversify our revenue streams to include licensing revenues, philanthropy and corporate sponsored research funding.

2. Training program goals include continued focus on writing plans for trainees and success in publications. In addition, plans are to continue to expand the Computer Science, Biology, and Biomedical Informatics (CoSBBI) Summer Academy for high school students. (See [http://www.upci.upmc.edu/summeracademy/](http://www.upci.upmc.edu/summeracademy/)). Dr. Becich is launching a new 501c3 not-for-profit to support this program.

3. Aggressively develop and cultivate relationships with non-traditional (i.e., non-NIH) sources of funding such as the Agency for Healthcare Research and Quality (AHRQ), National Science Foundation (NSF), the Commonwealth of Pennsylvania, industry partners, and Foundations in order to increase the breadth and depth of our research funding portfolio.

4. Strengthen ties with other leading BMI groups such as Vanderbilt (Kevin Johnson), and sustain valuable connections with our former DBMI faculty now at other institutions such as University of Utah (Wendy Chapman, Chair, Biomedical Informatics) and Regenstrief Institute at the Indiana University School of Medicine (Titus Schleyer, Director, Center for Biomedical Informatics).

5. Continue to grow and develop our relationships through collaborative research, quality improvement, and other developmental projects that directly benefit UPMC, the UPMC Health Plan, and The Center for Connected Medicine, and The Technology Development Center.

6. Recruit additional faculty and postdocs to meet the increasing demands of current and new clinical and translational informatics projects.
C. CLINICAL ACTIVITIES

NOT APPLICABLE

D. RESEARCH & OTHER SCHOLARLY ACTIVITIES

Trends in Research Support

During FY2015, the Department of Biomedical Informatics had $9,497,643 in direct research expenditures and $3,145,819 in indirect research revenue for a total of FY2015 research support of $12,643,462. Based on grant award notices received and estimates for pending awards, the approved FY2016 total research budget is $15,917,802, a 31.8 percent increase from the final FY2015 actuals. Although DBMI faculty have grant submissions pending, the uncertainty with Federal research budgets is such that it is difficult to predict approval likelihood for submitted grants.

Department faculty have successfully secured new and renewed funding, in terms of both individual submissions and collaborations with colleagues throughout the University of Pittsburgh community as well as outside universities.

The Department will work to maintain the current level of federal research support for at least the next few years. It is difficult to anticipate the federal resources for research, although recent experience suggests much higher thresholds to reach in order to secure funding.

Research Grant Growth:

<table>
<thead>
<tr>
<th>MEASURES</th>
<th>FY11</th>
<th>FY12</th>
<th>FY13</th>
<th>FY14</th>
<th>FY15</th>
<th>FY 15 Change From FY11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directs</td>
<td>$7,933,452</td>
<td>$6,235,831</td>
<td>$6,625,664</td>
<td>$6,085,154</td>
<td>$9,497,642</td>
<td>56.1%</td>
</tr>
<tr>
<td>Indirects</td>
<td>$2,844,350</td>
<td>$2,305,331</td>
<td>$2,494,800</td>
<td>$2,359,153</td>
<td>$3,145,819</td>
<td>33.3%</td>
</tr>
<tr>
<td>Total Grants</td>
<td>$10,777,802</td>
<td>$8,541,162</td>
<td>$9,120,465</td>
<td>$8,444,277</td>
<td>$12,643,462</td>
<td>31.8%</td>
</tr>
<tr>
<td>Revenue/FTE</td>
<td>$698,497</td>
<td>$551,042</td>
<td>$549,527</td>
<td>$513,689</td>
<td>$879,926</td>
<td>71.3%</td>
</tr>
</tbody>
</table>
Summary of Research

Michael J. Becich, M.D., Ph.D.

**Project Title:** National Mesothelioma Virtual Bank (NMVB) for Translational Research

**Project Background:** Malignant mesothelioma is a rare form of cancer that presents as a malignancy in the sac lining of the chest (the pleura), the abdominal cavity (the peritoneum) or the lining around the heart (the pericardium). Asbestos exposure through inhalation of asbestos fibers is the main cause of mesothelioma. For many years, American workers have been exposed to asbestos in the workplace, including workers in industrial and building trades and Navy personnel. Although the use of asbestos has been significantly reduced since the 1970s, mesothelioma is still a significant occupational health burden. Each year, in the United States, almost 3,000 people are diagnosed with mesothelioma.

The National Mesothelioma Virtual Bank (NMVB) for Translational Research has created and maintains a national virtual patient registry and research resource bank. The registry has been established and managed by the University of Pittsburgh team in collaboration with New York University, Roswell Park Cancer Institute (newest partner) and the University of Pennsylvania via a CDC NIOSH cooperative agreement. To date, the NMVB has made available over 1159 unique cases to share with the research community. The NMVB database is used for clinical and outcomes data collection related to biospecimen resources which include serum, plasma, fresh frozen tissue, formalin fixed paraffin embedded tissue, tissue microarrays and genomic data both from tumor DNA and buffy coat DNA.

**Research objectives:** The purpose of the NMVB is to provide high quality clinical and outcomes data and patient biospecimens to the research community in order to accelerate translational research. NMVB will serve as a resource that will allow researchers access to de-identified clinical data associated with a full range of biospecimens (listed above). Thus, NMVB will support scientific discovery, enhance detection and facilitate the development of effective treatments to maximally benefit the patients affected by this deadly disease. The specific aims of the NMVB are:

- To continue to serve the needs of the mesothelioma cancer research community by collecting tissue, blood, and clinical data and providing efficient access to these federated resources.
- To automate the biospecimens annotation through electronic extraction of clinical data from electronic health records (EHR), cancer registry system and integrating the Text Information Extraction System (TIES).
- To create a sustainable informatics model by deploying i2b2 (Informatics for Integrating Biology and the Bedside) and SHRINE (Shared Health Research Information Network) to maximize the efficiency and cost effectiveness of the data mining process across NMVB sites.

**Principal methods:** NMVB has been providing high quality and well characterized mesothelioma biospecimen to the research community for 8 years. The annotation process of biospecimen is a challenging task that consists of gathering patient level information from variety of data source even in a single hospital network. We have been
gathering patient level information including demographic, clinical, pathology and follow-up information manually by reviewing patient charts and pathology reports and storing this information in the centralized NMVB database. This process is expensive, laborious and consumes considerable resources available to sustain the operation of biorepository. To overcome these challenges, we are adopting a scalable solution for automating annotation process by extracting patient level information from EHR sources. Our proposed design is based on following components:

1) Core ontology adopting standard vocabularies (SNOMED, ICD-9, RxNORM, LOINC)
2) Implementation of TIES to structure text based pathology reports into machine readable case level data on biospecimens.
3) Adopt i2b2 and SHRINE to create a modular ontology-based, federated infrastructure that provides NMVB collaborators full ownership and access to their contributed data while supporting robust data sharing to mesothelioma research community.
4) Create an adapter between TIES and i2b2 to translate data extracted from clinical documents and load it into i2b2.

Recent results:


Conclusions:

We continue to expand NMVB to other funded networks including and NCATS CTSA program, NCI (through two funded U24s with Dr. Jacobson and a Mesothelioma SPORE with Mayo) and PCORI CDRN.

Future plans:

To expand the NMVB to the Rowell Park Cancer Institute (RPCI) as a collection site and continue collection at NYU, U Penn and U Pitt. Our goal is to expand the current 1159 patients in the NMVB database to over 1500 cases by 2016. We will also continue to document and to evaluate the usefulness of the NMVB to the scientific community and measure its impact in collaboration with the Mesothelioma Applied Research Foundation. This expansion of the NMVB will provide unique and innovative tools to aid in the prevention, early detection, and treatment of this uniformly fatal disease.

Project Title: A PaTH towards a Learning Health System in the Mid-Atlantic Region

Project Background: The PaTH Clinical Data Research Network (CDRN) comprises four Mid-Atlantic Health Systems: University of Pittsburgh (Pitt)/UPMC, Penn State College of Medicine/Hershey Medical Center, Temple University School of Medicine and Penn State College of Medicine/Hershey Medical Center.
Pathways for Patient-Centered Outcomes Research (PaTH)

Medicine/Temple Health, and Johns Hopkins University/ Johns Hopkins Health System/ Johns Hopkins Health Care (PaTH = Pitt/Penn State and Temple, Hopkins). The PaTH CDRN is funded by the Patient Centered Outcomes Research Institute (PCORI) and is led by Pitt. The Informatics Coordinating Center for the effort is supported by the Department of Biomedical Informatics at Pitt. PaTH aims study a longitudinal cohort of at least 1 million diverse individuals across a variety of health care settings to maximize our power to conduct meaningful patient-centered outcomes research. As PaTH is one of eleven funded CDRNs the goal is to link 11 million medical records of patients via open source software to enable pragmatic clinical trials.

Research Objectives
PaTH is committed to the following goals:
1. Defining data elements that allow data to be centrally aggregated and shared with researchers
2. Creating component systems that have the capacity to share data needed for a study using agreed-upon data standards to ensure interoperability
3. Making PaTH data accessible to researchers outside of the network for larger collaborations and creating a scalable, sustainable infrastructure

Principal methods: PaTH employs several software packages to meet its research objectives—i2b2 (integrating informatics from bench to bedside), SHRINE (shared research informatics network) and SHRINE+ (developed by Pitt DBMI). Each institution in PaTH deployed an i2b2 system as a research database loaded with data from their EMR. The site’s EMR data must be extracted, transformed, and loaded (i.e., the ETL process) in accordance with the Meaningful Use 2 terminologies (SNOMED, RxNorm, and LOINC). Clinicians conducting research within PaTH provide input and clinical expertise to guide the ETL process. In PaTH a researcher can perform federated queries (i.e., one data query is sent to all sites simultaneously and the results combined as a single response) using the SHRINE software package that interfaces with i2b2. PaTH has developed additional software, called SHRINE+, to enable exchange of limited datasets across the network. Requested limited datasets are sent to the Comparative Effectiveness Research Center Data Center (CERC-DC) at Pitt for analysis by the PaTH researchers.

Recent results: In the first 6 months of the project, PaTH has:
• Setup and configured i2b2 instances at all four sites
• Setup and configured SHRINE software at all four sites
• Connected all four i2b2 instances using SHRINE
• Defined a common ETL process for the first disease cohort, idiopathic pulmonary fibrosis (IPF)
• Loaded IPF data at all four sites
• Developed the SHRINE+ software
• Successfully installed and tested the SHRINE+ software at three of the four sites
• Written software that exports i2b2 data to the database schema used by PCORI
• Deployed and tested the PopMedNet software that connects PaTH to the larger PCORI network

Publications:
Conclusions:
This innovative program aims to accomplish the first demonstration of sharing medical records from 11 million patients within 18 months. Nothing of this magnitude in the area of secondary use of health data from electronic medical records has ever been accomplished. Pitt’s and DBMI’s leadership in this project has already created two significant additional research opportunities leading to additional grant funded efforts. This same infrastructure will also be critical to the success of our upcoming NCATS CTSAs renewal as well as the success of our newly funded Big Data to Knowledge (BD2K) Center of Excellence award.

Future plans:
The project is funded for the next 12 months. In months 6-12, PaTH will add the data elements for our second cohort Atrial Fibrillation. In months 12-18, PaTH will add data for the Weight cohort and the unselected longitudinal cohort. PaTH will continue to coordinate our development efforts with the other i2b2 based CDRNs in PCORI. We have recently been notified that we have a high likelihood of being funded for an additional eighteen month period (Sept 2015 to March 2017) to continue the work of the PaTH CDRN from PCORI. In addition, we will participate in a $10M Aspirin Trial which will be the first demonstration of the power of the eleven CDRN sites to do large scale pragmatic clinical trials (RFA to be released in Nov 2014 and funded by Jan 2015).

Project Title: Sarcoidosis and A1AT Genomics & Informatics Center (NHLBI, PIs: Kaminski, Wisniewski, and Becich)

Project Background: GRADS – genomic research in Alpha-1 Antitrypsin and Sarcoidosis - is a collaborative clinical research program that will collect detailed phenotype, gene expression, miRNA, and microbiome, with the goal of integrating clinical studies and molecular phenotyping to improve understanding of the two diseases and their various manifestations. The GRADS Genomic Information Center (GIC) will coordinate the data and provide collaborating institutions – and eventually the broader lung research community – with access to data for analysis. Our goal is to develop a database and web-based infrastructure that will support interactive exploration of the data store, allowing users to identify cohorts suitable for further analysis.

Research objectives: To develop a web-based architecture that will support cohort identification by combining dynamic interactive search histograms that will allow searching based on clinical and molecular features; a database-server infrastructure that will provide appropriate data for the interactive tools; GPU-optimized statistical comparisons for identifying highly-variant molecular markers.

Principal methods: GPU-based C code for rapid statistical analyses on molecular profiles; Hibernate/Tomcat/Java stack for provision of data via web-services; Javascript interactive tools based on the jQuery, D3, DC, and crossfilter libraries.

Recent results: Front-end tools are nearing completion; Database and statistical tools are complete.

Conclusions: Highly-interactive tools have the potential to facilitate cohort identification in clinical/molecular data sets.

Future plans: Integration of tools with broader GRADS infrastructure; Generalization of tools to support other data stores such as i2B2; Empirical evaluation of usability and utility.

Dr. Becich’s Collaborative Informatics Projects: Dr. Becich is Associate Director of the NCATS Clinical and Translation Science Institute (CTSI, see http://www.ctsi.pitt.edu) and Director of the Biomedical Informatics Core for important national research program
(DBMI’s support is currently over $1.2M in funding annually). Dr. Becich is also Associate Director of the University of Pittsburgh Cancer Institute (UPCI) and is Director of the Cancer Bioinformatics Core (with Uma Chandran, PhD) for the Cancer Center Support Grant which was recently resubmitted for renewal. Dr. Becich has been funded since 1997 by the CCSG program at UPCI. He also co-directs the Informatics Core for the NCI funded Melanoma and Skin Cancer SPORE with Melissa Saul, MS. This program was recently refunded for years 6 to 10 of funding and Dr. Becich has been involved in this program since its initial funding in 2008. Finally, Dr. Becich is a co-investigator in Dr. Jacobson’s “Advanced Development of TIES” and is linking his NMVB network with the TIES Cancer Research Network through the i2b2/SHRINE infrastructure the PCORI CDRN PaTH project provides which is described above. This effort is funded by NCI for four years by a U24 mechanism and began in 2014. These core facilities and collaborative NIH funded projects in informatics provide a solid foundation for multiple collaborative grants in which many DBMI faculty and senior staff scientists participate.

Tanja Bekhuis, Ph.D., M.S., M.L.I.S., A.H.I.P.

**Project Title:** NIH/R00, Screening Nonrandomized Studies for Inclusion in Systematic Reviews of Evidence

**Project Background:** We hypothesized that (a) methods based on natural language processing and machine learning (ML) can be used to automatically identify topically relevant studies with a mix of nonrandomized (NR) designs eligible for inclusion in systematic reviews of medical evidence; and (b) machine performance can consistently reach current human standards with respect to identifying eligible studies.

**Research objectives:** Evidence-based medicine depends on the timely synthesis of research findings. An important source of synthesized evidence resides in systematic reviews. However, a bottleneck in review production involves dual screening of citations with titles and abstracts to find eligible studies. In a recently published study, we tested the effect of various kinds of textual information (features) on performance of a machine learning classifier. Based on our findings, we proposed an automated system to reduce screening burden, as well as offer quality assurance.

**Principal methods:** We built a database of citations from 5 systematic reviews that varied with respect to domain, topic, and sponsor. Consensus judgments regarding eligibility were inferred from published reports. We extracted 5 feature sets from citations: alphabetic, alphanumeric+, indexing, features mapped to concepts in systematic reviews, and topic models. To simulate a two-person team, we divided the data into random halves. We optimized the parameters of a Bayesian classifier, then trained and tested models on alternate data halves. Overall, we conducted 50 independent tests.

**Recent results:** All tests of summary performance (mean F3) surpassed the corresponding baseline, P<0.0001. The ranks for mean F3, precision, and classification error were statistically different across feature sets averaged over reviews; P-values for Friedman’s test were .045, .002, and .002, respectively. Differences in ranks for mean recall were not statistically significant. Alphanumeric+ features were associated with best performance; mean reduction in screening burden for this feature type ranged from 88% to 98% for the second pass through citations and from 38% to 48% overall.
Conclusions: A computer-assisted, decision support system based on our methods could substantially reduce the burden of screening citations for systematic review teams and solo reviewers. Additionally, such a system could deliver quality assurance both by confirming concordant decisions and by naming studies associated with discordant decisions for further consideration.

Future plans: To extend this research by building profiles of electronic records for scientific articles based on criteria specified in the protocols for systematic reviews.

Richard D. Boyce, Ph.D.

Project Title: Knowledge-Based Approaches To Drug-Drug Interaction And Adverse Drug Event Prediction, And Identification

Project Background: The combination of poor quality evidence on potential drug-drug interactions (PDDIs), and a general lack of PDDI knowledge by prescribers, results in many thousands of preventable medication errors each year.

Research objectives: We propose a new paradigm that would reduce preventable medication errors by more effectively synthesizing existing PDDI knowledge, and more rapidly producing evidence to fill in knowledge gaps. We will advance three research aims while building the framework. Aim 1: Derive a new PDDI meta-data standard that can meet the information needs of pharmacist working in different care settings. Aim 2: Apply a novel evidence synthesis process to enhance product label PDDI information. Aim 3: Pilot test new methods for PDDI information retrieval supporting drug information experts.

Principal methods: Artificial intelligence, knowledge base development, mixed methods involving qualitative interviews and focus groups, information system design and evaluation.


Conclusions: This study is in the first of four years of funded work.

Future plans: Continue to make progress on the aims which will lead to other large grant proposals.

Project Title: Drug Safety and Decision Support for Older Adults

Project Background: Unfortunately, a high prevalence of psychotropic prescribing, and the large number PDDIs involving psychotropics, makes appropriate monitoring very challenging for NH clinicians.

Research objectives: The long-term goal of the proposed work is to develop an effective informatics intervention that prevents harm to NH residents from drug-drug interactions while avoiding known issues with PDDI alerting such as alert fatigue. Specific aims: Aim 1: Validate an automated falls prognostic model for NH patients exposed to psychotropic PDDIs. Aim 2: Identify modifiable potential barriers to the use of active PDDI monitoring in the NH and design a pilot intervention that addresses them.
Principal methods: Pharmacoepidemiology, clinical decisions support intervention design, mixed methods involving qualitative interviews and focus groups, quantitative surveys, and prognostic algorithm design and validation.


Conclusions: This study is nearing completion of the first of three years of funded work.

Future plans: Continue to make progress on the aims which will lead to a large grant proposal to test a newly designed clinical intervention.

Project Title: Pharmacogenomics Decision Support

Project Background: Pharmacogenomics provides a tremendous opportunity to make a measurable and positive impact on patient outcomes, while providing an exceptional return on investment. Many genetic variations are associated specific changes in medication effectiveness and/or risk of toxicity. As the costs of genotyping technology continue plummet, the need for a systematic approach to translate the pending deluge of pharmacogenomics data into meaningful drug and dose decisions is becoming more urgent.

Research objectives: Build an informatics framework and educational resources necessary for efficient clinical implementation.

Principal methods: Following identification of the core clinical data elements required for pharmacogenomics-based decision making, these variables will be mapped into the EDW and best practices for data flow and its presentation in Cerner® will be developed. Ongoing efforts towards developing a web platform that supports pharmacogenomics decision-making and provider education by integrating local drug use policies, data present within FDA drug product labels, and clinical annotations will also be extended.


Conclusions: We are making progress on the research objectives

Future plans: Complete the research aim. Technology transfer. Further extramural grant proposal submissions.
Gregory F. Cooper, M.D., Ph.D.

Project Title: Center for Causal Discovery

Project Background: Much of science consists of discovering and modeling causal relationships in nature. With rapid advancements in technology and networking, biomedical scientists increasingly generate multiple complex data types for a large number of samples, each of which has an enormous number of measurements recorded. Although statistical and machine learning methods can predict the value of a variable $X$ from observed predictors, the best predictors of $X$ are often poor models of the causes of $X$ (hence the slogan “correlation is not causation”), which motivated the development of algorithms specifically devoted to the discovery of valid causal models.

Indeed, tremendous progress has been made in developing computational methods for representing and discovering causal knowledge from data. These causal discovery methods have found applications in a wide range of fields, including econometrics, education, epidemiology, climate research, medicine, and biology. Current capabilities include (1) the representation of existing causal knowledge as a graphical network model with precisely defined semantics; (2) the discovery of causal networks of relationships from a combination of prior knowledge and experimental and observational data; and (3) the use of causal networks to suggest how changing one variable (e.g., a drug binding to a signaling protein and blocking a pathway) is likely to influence the state of another variable (e.g., cell apoptosis). While much progress has been made in the development of these computational methods and their application in biomedical science and other fields, they are not sufficiently efficient to analyze big datasets nor easy for biomedical scientists to access or apply to their data.

To fill this gap, the Center for Causal Discovery (CCD) is building on the extensive code base of causal modeling and discovery (CMD) algorithms that we have developed and implemented over the past 25 years and integrating new or improved algorithms as they are reported in the literature. Software products from the Center will allow biomedical and data scientists to select and apply one or more data-appropriate causal discovery algorithms to their biomedical datasets and compare the causal relationships that each algorithm predicts.

Research objectives:

Aim 1. Develop and implement state-of-the-art methods for causal modeling and discovery (CMD) of knowledge from biomedical big data
   - Make the best existing CMD methods available
   - Develop new CMD methods
Aim 2. Investigate three biomedical projects
   - Evaluate the usefulness of CMD methods on these problems
   - Drive further the development of the CMD methods
Aim 3. Disseminate CMD methods and knowledge widely to biomedical researchers and data scientists
   - Software
   - Training
   - Collaborative activities with other BD2K Centers

Principal methods: Algorithm development and optimization is at the core of CCD efforts, and we are focusing on the discovery of structural causal relationships that can be represented by CBNs. We are using two main classes of algorithms that model hidden
variables and sample selection and have the ability to discover them based on observational data, data from experimental interventions, or both: constraint-based algorithms, which use tests of conditional independence, and Bayesian algorithms, which allow the specification of structure and parameter prior probabilities. We are in the process of implementing and making available the best CMD algorithms as well as developing new algorithms to address pressing needs of causal discovery in biomedicine.


Conclusions: The CCD is developing and making available tools for causal modeling and discovery with large and complex data to generate new biomedical knowledge and to encourage and train both data and biomedical scientists in their use. It will serve as a central resource for anyone seeking causal discovery algorithms, software tools, training, and collaboration.

**Project Title:** Outlier-Based Clinical Alerting

**Project Background:** The goals of this project are to develop, implement, and evaluate computer-based methods that model usual clinical care and then apply those models to detect individual patient care that is anomalous. In the future, such a system may serve as a “safety net” that continuously monitors patient care, as documented in an electronic medical record (EMR) and raises an alert when such care appears to be anomalous.

**Research objectives:** A hypothesis of the project is that such anomalies correspond to medical errors often enough to make such alerting worthwhile. Within the ICU domain the project is investigating the extent to which this hypothesis is supported.

**Principal methods:** The method learns unusual patient-management actions from a training set of past patient cases that are recorded in an EMR system. A total of 16,500 ICU patient cases were used to train a system that models usual patterns of ordering medications and laboratory tests, in the context of a patient’s clinical history. The models were applied to a separate set of 8,158 ICU patient cases and used to generate alerts. A subset of the 240 alerts generated by the models were evaluated and assessed by 18 ICU clinicians.

**Recent results:** The results of the evaluation obtained on 240 medication and laboratory omission alerts show the true positive rates for the alerts to range from 0.44 to 0.71. The true positive rate for medication order alerts ranged from 0.31 to 0.61 and for laboratory order alerts from 0.44 to 0.75.

**Conclusions:** The true positive alert rates obtained by the outlier-based alerting approach compare favorably to the true positive alert rates of existing clinical alerting systems. Also, using this new approach, alerting systems can be constructed automatically from past EMR data. These results support outlier-based alerting systems as a promising new approach to clinical alerting.

**Ongoing work:** We are in the process of improving the machine-learning methodology. In 2016 we plan to evaluate it prospectively on ICU cases. We have an active NIGMS R01 grant that is funding this effort.
Project Title: Probabilistic Disease Surveillance

Project Background: We have developed an innovative probabilistic approach to disease surveillance and deployed it in an influenza monitoring system in Allegheny County (AC), PA (1). The approach differs from current approaches to disease surveillance in that it is a fully probabilistic method that computes three probabilities from patient data in electronic medical records (EMRs): (1) \( P(case \mid data_p) \), where \( data_p \) are data on patient \( p \) and \( case \) is a disease case specification (i.e., a diagnosis); (2) \( P(outbreak \mid data_{all}) \), where \( data_{all} \) are data on all patients over time and \( outbreak \) is a current disease outbreak in the population of a particular disease; and (3) \( P(epidemic\ models \mid data_{all}) \), which is derived based on patient data and epidemiological knowledge. A key advantage of the approach is its ability to integrate heterogeneous and potentially complex data types such as the information in emergency department (ED) dictated reports and coded ED laboratory results. Our integrated approach to computing these key probabilities from heterogeneous data sources is unique.

The approach naturally divides into (1) case detection, which is done in healthcare organizations and which we have implemented as a probabilistic case detection system (CDS), and (2) outbreak detection and characterization (OD&C), which is done regionally and which we have implemented as an outbreak detection and characterization system (ODS).

Research objectives: We are applying these probabilistic methods to other respiratory diseases to investigate their portability using cases and outbreaks from AC and Salt Lake County (SLC), Utah. The objective of the planned evaluations is to test whether the theoretical potential of integrating probabilistic case detection with probabilistic OD&C leads to more sensitive and specific individual patient case detection as well as earlier and more accurate outbreak detection and characterization.

The specific aims of the proposed research are to: 1.) significantly advance the development and integration of the system components. 2.) expand the disease models. We propose to extend the approach from influenza to six other respiratory diseases. 3.) evaluate the components and the system.

Principal methods: We are using Bayesian modeling methods. The ability to infer the above probabilities from data in EMRs has significant potential to improve disease surveillance and public health practice. \( P(outbreak \mid data_{all}) \) is an ideal quantity for epidemiologists to use when setting alert thresholds in disease surveillance systems and \( P(epidemic\ models \mid data_{all}) \) can inform decisions made by epidemiologists about disease control interventions, such as vaccination and school closure. Most significantly, our approach can develop information about an outbreak of an emerging epidemic or pandemic disease in Region A that can be shared in computable form with Region B to give it immediate disease surveillance capabilities for the new threat. This capability could be critical in mitigating an emerging disease with a high reproductive rate and case severity.

Recent results: Please see the following publication:

http://dx.doi.org/10.1016/j.jbi.2014.08.011

Conclusions: We have been able to use Bayesian methods to link patient disease diagnosis and epidemiological modeling of disease outbreaks. An evaluation of the method indicates that it is able to perform useful detection and characterization of influenza disease outbreaks.

Ongoing work: We have recently extended these methods to detect and characterize multiple, possibly overlapping infectious disease outbreaks. We are in the process of evaluating these extensions.
Gerald P. Douglas Ph.D.

Project Title: Understanding the Dispensary Workflow at the Birmingham Free Clinic. Responding to Challenges with Informatics Interventions.

Project Background: Pharmacists and other volunteers at the BFC recognize challenges in certain areas of their workflow surrounding the medication management process, such as the Patient Medication Assistance Program application process and working with a limited formulary. The recent introduction of an electronic medical record system in the clinic have improved aspects of patient care, but created some new challenges in the medication management process.

Research objectives: To understand process and workflow challenges associated with medication management in the clinic and develop health IT interventions to address these challenges. Specific aims are:

Aim 1: Quantifying the workflow challenges and inefficiencies identified and propose informatics interventions to alleviate these challenges.

Aim 2: Develop a functional prototype for the Prescription Management and General Inventory Control (RxMAGIC) system.

Aim 3: Conduct a usability evaluation of the functional prototype.

Principal methods: Contextual inquiry. Problem-driven development. Usability testing.

Recent results: A time and motion study identified significant areas for improvement of non-value-added tasks such as medication labeling, inventory stock management and general documentation. Usability testing of RxMAGIC suggests that the system will significantly reduce time wasted on non-value-added tasks in the clinic.

Conclusions: Inexpensive health IT interventions have the potential to address process and workflow challenges in low-resource settings.

Future plans: We are customizing RxMAGIC for use in a rural health center in Honduras, and will evaluate its impact on reducing medication stock-outs.

Project Title: Assessing challenges in Laboratory Testing at Kamuzu Central Hospital

Project Background: Laboratories in low- and middle-income countries (LMIC) strive to provide timely and accurate diagnosis for the care and treatment of the diseases burdening their hospitals. The pressure on laboratories to keep up with the demand for services, despite resource and infrastructure constraints, is causing many hospitals to seek solutions to overcome the challenges, alleviate the strain, and improve the quality of services. We believe that appropriate health IT interventions can be used in these setting to improve process and workflow. To better understand these challenges we conducted a qualitative assessment of challenges at Kamuzu Central Hospital, a tertiary referral hospital in Malawi.

Research objectives: To identify perceived challenges associated with routine laboratory-related processes to support patient care.

Principal methods: Key-informant interviews of hospital and laboratory staff with thematic analysis of transcripts.
Recent results: Twenty-one hospital and laboratory staff were interviewed. Challenges were identified in the pre-analytical, analytical and post-analytical phases of the diagnostic testing process. Four meta-themes emerged: Insufficient resources, Staff management challenges, workflow delays & Blood bank management.

Conclusions: While most laboratory management information systems look exclusively at the supporting the analytical phase of the diagnostic testing cycle, there may be value in developing information systems that support the all phases for hospitals in low-resource settings.

Future plans: We are customizing BLIS, and open source lab information management system to include bedside lab order entry and results review. We plan to evaluate the impact of this on challenges uncovered through this research.

Project Title: Improving Vital Statistics in Malawi.
Project Background: Most people in Africa and Asia are born and die without appearing in any legal record or official statistic. Malawi currently has no functional system for civil registration and vital statistics. Paper village registers are issued to village headmen with the goal of recording citizens of the village and births and deaths as they occur. However, this approach has been shown to generate incomplete data and is largely dysfunctional through the country as the government has no framework for routinely collecting data from these registers. In 2012 we developed a prototype electronic village register, which was installed in the home of a village headman. Early use of the system suggests that the technology solution can work in this setting but there is not sufficient value proposition for the village headman to use the system, hence the data completeness is still inadequate.

Research objectives: We aim to develop a technology platform that can run in the homes of village headman that provides a more comprehensive set of information service. This research seeks to identify a critical mass of value for the village headman such that he will incorporate the technology platform into his daily routine.

Principal methods: Focus groups, key informant interviews and contextual inquiry.

Recent results: Early indications have identified three functions of the technology platform that village headmen value: 1) an electric light powered by the same solar panel that runs the technology platform, 2) Voice over IP solution that allows village headmen to talk with one-another and with the group village headman, and 3) Access to daily newspaper headlines.

Conclusions: Early stage research.

Future plans: Incorporate alerts and reminders into the technology platform to increase the completeness of birth and death reporting.

Madhavi Ganapathiraju, Ph.D.
Project Title: Discovery of mental health and inflammation interactome
Project Background: Sickness behavior, depression, anxiety disorder, etc originate not only as outcomes of rational thinking but as a result of physiological and genetic factors interacting with environmental stressors. Conversely, psychological stress can influence patho-physiological processes and lead to brain and mind disorders, as well as influence systemic processes like inflammation and immunity. There is a vast body of literature that describes the modulations of psychological/neurological processes and inflammatory response on each other (see e.g. from Inflammation to sickness and depression and neural
origins of sickness in response to inflammation). We are working on discovering the molecular mechanisms behind the modulations of psycho-neuro-inflammo processes by each other through the discovery of novel protein-protein interactions (PPI) of genes involved in these processes.

**Research objectives:** The objective of this work is to carry out systematically designed computational work to discover the human mental health and inflammation (MHAIN) interactome. The MHAIN interactome refers to the network of PPIs where at least one of the two proteins is involved in either brain or inflammation.

**Principal methods:** We analyze the features of individual proteins that can contribute to the prediction of interactions between them and then develop algorithms to infer, predict, estimate and acquire such features when they are unavailable for proteins of interest. Next, we develop new algorithms by integrating various approaches from machine learning, including proactive and transfer learning, and multi-sensor fusion, to address each of the challenges in PPI prediction. Finally, we mine the predicted interactome for new biological insights.

**Recent results:** The PPI that we predicted between the genes OASL and RIGI was validated by our collaborator; further functional studies of this predicted interaction, lead to the discovery that OASL enhances cells’ ability to detect virus RNA, activating the immune system to sense the virus (by activating RIG-I pathway) and inhibiting replication (published in Immunity, PMID 24931123). This work helping biological discovery of the role of OASL gene by the scientific community and is getting cited well. Next, we constructed the interactome of schizophrenia genes with over 600 novel PPIs (this work is being submitted to a reputed journal). Finally, we constructed the interactome of genes identified to be associated with congenital heart disease by collaborator Dr. Cecilia Lo’s group, and showed the connection of these genes to several relevant pathways as revealed by novel PPIs (this work has been published in Nature.). All 14 PPIs that we studied experimentally in wet lab (through collaborations at Pitt) have been validated to be true interactions. We carried out the network analysis of cilia-associated genes and inferred relevant biological characterization of cilia genes (through computational network analyses) and this work has been prepared for submission to a reputed journal.

**Future plans:** This is ongoing work. Although we discovered several novel interactions at very high accuracy, there remain several thousand more interactions that remain to be discovered. We are developing algorithms to discover this. We are working on disseminating the novel predictions to the scientific community so that they may be translated to biomedical insights.

**Vanathi Gopalakrishnan, Ph.D.**

**Project Title:** Bayesian Rule Learning Methods for Disease Prediction and Biomarker Discovery

**Project Background:** The precision of predictive modeling tools is of utmost significance to disease classification and biomarker discovery from rapidly accumulating high-dimensional genomic and proteomic datasets. Although many machine learning methods have been applied to predict disease status from such biomarker profiling datasets, a single, learned model can be limited in its ability to predict well across a wide variety of patient cases, particularly given the sparse training data sizes of analyzed clinical samples. In contrast, by
using multiple models that each predicts a subset of the cases relatively well, overall predictive performance may be improved. This project will use multiple models to predict disease status from biomarker profiling data.

**Research objectives:** This project will develop, apply, evaluate, and refine algorithms that extend a novel Bayesian Rule Learning (BRL) system to search a richer space of model representations, and combine predictions from multiple models to improve the precision for disease classification from biomarker profiling datasets. BRL is a novel hybrid modeling technique that combines the mathematical rigor of Bayesian network learning with rule-based inference; this intersection is an underexplored area of fundamental research in informatics, facilitating prior knowledge incorporation. We further investigate a unified framework for ensemble classification with BRL (ecBRL) involving the use of multiple classifiers to capture different patterns within the same underlying data.

**Principal methods:** This project will test the hypothesis that the ecBRL methods developed herein for combining classification evidence from multiple models are more accurate for disease state prediction than is the application of the single best predictive model that can be found using BRL or other state-of-the-art classifiers. These methods will be applied to predictive modeling of biomarker data generated for early detection and staging of lung, breast, and esophageal cancers from the analysis of various biological materials. The predictive performance and generalizability of the models will be examined using held-out sets of cases and controls. Furthermore, we will validate candidate gene expression, DNA methylation and protein markers for breast cancer status using new sets of retrospectively obtained de-identified tissue and blood serum.

**Recent results:** We have made tremendous scientific progress toward the achievement of the aims involving the development and testing of our BRL methods. Firstly, we implemented a more general Bayesian scoring method along with new algorithms to perform search of local structures—Bayesian decision trees and graphs. We observed that BRL algorithms performed on par with a state-of-the-art decision tree learner while retaining parsimony. Then, we extended our methods to perform selective Bayesian Model Averaging, which will be further explored in the competing renewal proposal that we have submitted for this project. Furthermore, we developed new methods for computing informative priors from mining of the literature, wherein we have analyzed 5.3 million PUBMED abstracts, for lung and breast cancer biomarkers from across 14 different biofluids. Most recently, we have been able to validate BRL models learned on one set of data for prediction of esophageal cancer, using an independent validation dataset. This work has resulted in a provisional patent application filed by the University of Pittsburgh, and a publication that has just appeared in journal Cancer (Impact Factor of 5.2).

**Conclusions:** BRL methodology is useful for generating predictive models of disease from biomarker discovery data.

**Future plans:** A continuing application has been submitted to develop the ecBRL methods that extend current BRL methodology. This competing renewal NIH application has received a fundable score.

**Project Title:** Mining Biomedical Image Data for Actionable Knowledge

**Project Background:** The acquisition, processing and analyses of complex biomedical imaging data are a routine part of many clinical practice workflows used to diagnose, monitor and treat diseases. Due to the complex nature of individual workflows, there is a need to study each from the perspective of predictive modeling of such biomedical big data
from which the imaging biomarkers are typically derived and used for patient care. A major challenge in predictive modeling is data scarcity, which presents an important opportunity to make the most of existing data and transform it into actionable knowledge for enabling translational science. This would be of utmost significance in the clinical practice of non-invasive Cardiovascular Magnetic Resonance Imaging (CMRI) of pediatric patients wherein there is opportunity to produce actionable knowledge from data acquired in specialized medical centers and use it in non-specialized clinics to facilitate timely detection of disease.

**Research objectives:** This project will develop and apply a novel CMRI Biomarker Extraction and Discovery (CMRI-BED) workflow that embeds predictive modeling of retrospective pediatric cardiovascular imaging data to resolve the detection and classification of congenital heart disease, particularly cardiomyopathies. Predictive modeling includes the use of novel classification rule learning methods that can develop accurate and parsimonious classifiers from prior knowledge and data. Rules are easy to understand by clinicians, and represent actionable knowledge that describes non-linear relationships among measurable variables (biomarkers) and the clinical or demographic variables that help classify the outcome or diagnosis. Prior knowledge can be incorporated with the clinician in the loop, which is very important for refining the models and methods in the workflow.

**Principal methods:** This project will test the hypothesis that the CMRI-BED workflow developed and tested herein is feasible for accurate processing of CMRI data to extract biomarkers and discover their significance within actionable, quantifiable, predictive classification rules for identifying pediatric congenital heart disease. Models learned from retrospectively obtained de-identified datasets for pediatric cardiomyopathy case-control discrimination will be tested on an independent set of similarly generated data for cases and controls to establish the validity of the imaging biomarkers and their relationships within and across classification rule sets.

**Recent results:** We have generated several publications that have analyzed both publicly available data from the Cardiac Atlas Project as well as data that we collected from Children’s Hospital of Pittsburgh of UPMC. We have been able to demonstrate the usefulness of a novel imaging marker for ischemic cardiomyopathy for classification of symptomatic and asymptomatic patients using publicly available data. This was presented in the premier SPIE Medical Imaging Conference. We have more recently analyzed a CHP dataset containing a standard set of 32 CMRI markers for 83 cases and controls to learn predictive models for positive or negative findings of cardiomyopathy. The overall cross-validation performance is around 83%, which is very encouraging. Models learned on the entire dataset yield AUROC values of 90%.

**Conclusions:** Preliminary results show that there is merit to pursuing this workflow.

**Future plans:** We are currently in the process of writing up the results for publication in journal – this is an invited paper selected by conference organizers. Subsequent to that, we will be submitting a revised proposal to the NHLBI for this project.

**Project Title:** Transfer Rule Learning for Knowledge Based Biomarker Discovery and Predictive Biomedicine

**Project Background:** Predictive modeling of biomedical data arising from clinical studies for early detection, monitoring, and prognosis of diseases is a crucial step in biomarker discovery. Since the data are typically measurements subject to error, and the sample size of any study is very small compared to the number of variables measured, the validity and
verification of models arising from such datasets significantly impacts the discovery of reliable discriminatory markers for a disease. An important opportunity to make the most of these scarce data is to combine information from multiple related data sets for more effective biomarker discovery. Because the costs of creating large data sets for every disease of interest are likely to remain prohibitive, methods for more effectively making use of related biomarker discovery data sets continues to be important.

**Research objectives:** This project develops and applies Transfer Rule Learning (TRL), a novel framework for integrative biomarker discovery from related but separate data sets, such as those generated from similar biomarker profiling studies. TRL alleviates the problem of data scarcity by providing automated ways to express, verify and use prior hypotheses generated from one data set while learning new knowledge via a related data set. This is the first study of transfer learning for biomarker discovery. Unlike other transfer learning approaches, TRL takes knowledge in the form of interpretable, modular classification rules, and uses them to seed learning of a rule model on a new data set. Classification rules simplify the extraction of discriminatory markers, and have been used successfully for biomarker discovery and verification in a non-integrative fashion.

**Principal methods:** This project tests the main hypothesis that TRL provides a mechanism for transfer learning of classification rules between related source and target data sets that improve performance on the target data, compared to learning without transfer. TRL will be evaluated using cross-validation performance of classification accuracy and transfer measures, on related groups of existing biomarker discovery datasets obtained from multiple experimental platforms for lung cancer detection and prognosis. A new set of independent validation data will be generated for early detection of lung cancer to test the models generated on pilot data. Insights into the impact of different modeling algorithms on transfer outcomes will be gleaned.

**Recent results:** We have leveraged the notion of functional modules to serve as a bridge/pivot to facilitate the transfer of classification rules between different but related “omic” data. To test the concept of transfer learning of classification rules—through functional mapping—for integrative biomarker discovery, we experimented with data sets from three cancer types, namely brain, prostate, and lung cancer. We have also generated a new set of biomarker measurements on a new set of retrospectively collected sera samples from Vanderbilt University to mirror the study that was previously conducted for lung cancer case-control determination at the University of Pittsburgh Cancer Institute (UPCI).

**Conclusions:** Preliminary results show that, for more often than not, transfer learning of classification rules—through functional mapping—improve classification performance and learning efficiency.

**Future plans:** We are currently in the planning phase of the competing renewal proposal for this project. We have generated novel ideas and tested them, and are in the process of writing manuscripts that show the value of these methods for integrative biomarker discovery from large set of publicly available TCGA data across multiple types of biological samples that have been analyzed previously for diagnosis of breast and lung cancers.
Harry Hochheiser, Ph.D.

**Project Title:** Interactive Search and Review of Clinical Records with Multi-layered Semantic Annotation *(NLM, PI: Chapman)*

**Project Background:** Although Natural Language Processing (NLP) has proven useful for extracting structured annotations from clinical texts, the NLP process is often too far removed from the clinical researchers who are trying to use the results to guide the research. We hypothesize that interactive tools that provide users with the ability to review and interactively revise annotations extracted from clinical records will help clinical researchers more effectively use NLP to extract information relevant to clinical research challenges.

**Research objectives:** To develop interactive tools for review and revision of the results of using NLP to extract information, including visual displays of the results of variable annotations; tools for exploring the range of documents, terms, and spans associated with those annotations; and interactive features for providing feedback that might be used to revise NLP models to account for user feedback. To apply these tools to initial data challenges faced collaborators in the interpretation of colonoscopy records and in extracting opioid-related adverse drug events.

**Principal methods:** Contextual inquiry and collaborative design techniques have been used with clinical collaborators to inform design of the tool. Bag-of-words based NLP has been used to train models for colonoscopy data. Information visualization and interaction techniques for guiding the design of the tool. Web-based implementation with NLP services provided by a REST-based back-end provide a scalable architecture.

**Recent results:** A preliminary prototype is functional and awaiting usability evaluation.

**Conclusions:** The use of interactive tools for reviewing and revising NLP output shows preliminary promise.

**Future plans:** An empirical study comparing our tools to current approaches is planned for Spring 2015.

---

**Project Title:** Sarcoidosis and A1AT Genomics & Informatics Center *(NHLBI, PIs: Kaminski, Wisniewski, and Becich)*

**Project Background:** GRADS – genomics research in Alpha-1 Antitrypsin and Sarcoidosis is a collaborative clinical research program that will collect detailed phenotype, gene expression, miRNA, and microbiome, with the goal of integrating clinical studies and molecular phenotyping to improve understanding of the two diseases and their various manifestations. The GRADS Genomic Information Center (GIC) will coordinate the data and provide collaborating institutions – and eventually the broader lung research community – with access to data for analysis. Our goal is to develop a database and web-based infrastructure that will support interactive exploration of the data store, allowing users to identify cohorts suitable for further analysis.

**Research objectives:** To develop a web-based architecture that will support cohort identification by combining dynamic interactive search histograms that will allow searching based on clinical and molecular features; a database-server infrastructure that will provide appropriate data for the interactive tools; GPU-optimized statistical comparisons for identifying highly-variant molecular markers.
Principal methods: GPU-based C code for rapid statistical analyses on molecular profiles; Hibernate/Tomcat/Java stack for provision of data via web-services; Javascript interactive tools based on the jQuery, D3, DC, and crossfilter libraries.

Recent results: Front-end tools are nearing completion; Database and statistical tools are complete.

Conclusions: Highly-interactive tools have the potential to facilitate cohort identification in clinical/molecular data sets.

Future plans: Integration of tools with broader GRADS infrastructure; Generalization of tools to support other data stores such as I2B2; Empirical evaluation of usability and utility.

Project Title: Quantifying Electronic Medical Records Usability to Improve Clinical Workflow (AHRQ, PI Agha)

Project Background: Electronic Medical Records (EMRs) have the potential to improve quality of care but to date there is little research to quantify the effect of EMRs as barrier or facilitator of quality. Design and implementation of EMR's, should not be viewed as an end in distracting them with burdensome documentation or Human-Computer Interaction (HCI) limitations. In an ideal patient-centered process, the provider would focus primarily on the patient. However in the time-constrained framework of office consultations, the EMR competes for the provider's focus of attention. It is desirable to understand the degree to which EMR task complexity imposes a cognitive burden on providers.

Research objectives: Use a variety of data collection methods to paint a rich picture of the dynamics of EMR usage before, during, and after physician-patient encounters. Analyze these data to identify determinants of cognitive load and opportunities for redesigns that might provide greater usability and flexibility while reducing cognitive demands.

Principal methods: Detailed recording of EMR interaction usage, including keystrokes/screen capture, eye-tracking, motion tracking, and video; Assessments of cognitive load based on the NASA-TLX; Interviews with practitioners based on stimulated recall; Qualitative coding of interview content and of interaction events involving patients and physicians.

Recent results: Preliminary analyses have identified some critical incidents in physician/EMR interactions.

Conclusions: None as of yet

Future plans: Further data analysis and development of proposed redesigns

Project Title: Cancer Deep Phenotype Extraction from Electronic Medical Records (NCI, PI: Jacobson)

Project Background: Precise phenotype information is needed to advance translational cancer research, particularly to unravel the effects of genetic, epigenetic, and other factors on tumor behavior and responsiveness. Current models for correlating EMR data with -omics data largely ignore the clinical text, which remains one of the most important sources of phenotype information for cancer patients. Unlocking the value of clinical text has the potential to enable new insights about cancer initiation, progression, metastasis, and response to treatment. We propose further collaboration of two mature informatics groups with long histories of developing open-source natural language processing (NLP) software (Apache cTAKES, caTIES and ODIE) to extend existing software with new methods for cancer deep phenotyping.
Research objectives: To develop visual analytics tools that will assist cancer researchers in interpreting phenotype data as extracted from EMRs.
Principal methods: Contextual Inquiry; Iterative, user-centered design, development, and evaluation of visual analytics tools
Recent results: Development of stakeholder profiles; Development of interview guide
Conclusions: None yet
Future plans: Interview stakeholders; Develop initial designs; Review designs with users; Implement and iteratively refine visual analytics tools; Evaluate tools

Project Title: Addressing gaps in clinically useful evidence on drug-drug interactions (NLM R01 LM011838, Boyce PI):
Project Background: We propose a new knowledge representation paradigm for potential drug-drug interactions that will contribute to public health by making more effective use of drug-drug interaction evidence, filling in important gaps in drug safety knowledge, and spurring innovations in drug information retrieval.
Research Objectives: To understand clinical information needs regarding synthesis and review of information about drug-drug interactions and build tools that will provide this information in a manner that will support effective curation and creation of drug-drug interaction references.
Principal Methods: Interviews with pharmacists in a variety of settings regarding perceptions and information needs, presentation of prototype information systems for initial impressions and to elicit commentary regarding information needs and preferences and user-centered iterative redesign and evaluation.
Recent Results: Preliminary models of user needs based on initial interviews, candidate redesigns of prototypes based on preliminary models and preliminary literature review of drug-drug interaction information systems.
Conclusions: Crowd sourcing techniques may be effective means of curating drug-drug interaction statements:
Future Plans: Continued interviews and iterative design and refinement of search tools enabling compendia creators to interpret drug-drug interaction data.

Project Title: Semantic LAMHDI: Linking Diseases to Model Organism Resources (NIH OD, PI Haendel,), & Building Phenotype Comparison Tools for the Undiagnosed Disease Program . Subsequently renamed the Monarch Initiative

Project Background The goal of this work is to facilitate the identification of models for disease research, make better use of existing model organisms and in vitro resources and data about them, and provide the ability to uncover new relationships between disease, phenotypes and genes that will further our understanding of disease.
Research objectives: To combine an information resource that will leverage a federated database of databases containing extensive biological data with ontological methods for semantic similarity to develop tools for using phenotype data to identify models systems most similar to human phenotype profiles. To develop phenotype comparison tools that can
leverage this infrastructure for support in a wide variety of domains and problems, such as in the analytic workflows used by the NHGRI Undiagnosed Disease Program/Network.

**Principal methods:** Interactive web-based visualization developed in Javascript with the D3 Visualization library. Integration with Monarch Initiative web architecture and services; Use of public source code repositories (www.github.com) in support of adaptation and reuse.

**Recent results:** The Phenogrid phenotype similarity widget is available on the Monarch web site (www.monarchinitiative.org) and should soon be ready for deployment on other sites. Submission of a BD2K U01 Visualization proposal in June 2014 for further extension of this work.

**Conclusions:** The PhenoGrid widget may prove useful for interpreting phenotype-genotype relationships linking human diseases to animal models.

**Future plans:** Extension of the capability of the Phenogrid widget to provide additional feedback and context, and to support additional data types; Revision of the monarch initiative web site to use the Phenogrid widget as one component of a visual analytics framework.

**Rebecca S. Jacobson, M.D., M.S.I.S.**

**Project Title:** Advanced Development of TIES – Enhancing Access to Tissue for Cancer Research

**Project Background:** Archived human tissues are an essential resource for translational research. Formalin-fixed, paraffin-embedded (FFPE) tissues from cancer patients are used in a wide range of assays, including RT-PCR, SNP profiling, multiplex biomarkers, imaging biomarkers, targeted exome, whole exome, and whole genome sequencing. Remainder FFPE tissues generated during patient care are ‘retrospective’; use of these tissues under specific conditions does not require consent. For personalized medicine researchers, these specimens are vital resources enabling biomarker validation, detailed molecular analysis, and systems modeling before application is made to individual patients. But access to these human tissues is often a rate limiting factor in cancer research. The TIES software system automates the deidentification and phenotype annotation process from Pathology and radiology reports using its natural language processing (NLP) pipeline. TIES was one of the first NLP pipelines developed for clinical text.

**Research objectives:** We propose advanced development of the TIES software to (1) increase institutional capacity for using FFPE to support molecular characterization of human tumors, (2) increase access to tissues within cancer centers, and (3) improve the ability to share tissues and associated phenotype data among cancer centers.

**Principal methods:** Enhance the informatics technology to support inter-institutional “trust”, paraffin registry development, tissue microarray (TMA) development, and nondestructive tissue use; establish the TIES Cancer Research Network (TCRN) with four founding member institutions; develop governance, network agreements, and policies for operating the TCRN; recruit and support pilot scientific collaborations across the network; disseminate the software and measure its impact.

**Recent results:** Major activities in Y2 have focused on (1) the release of TIES v5.3, which includes support for virtual slides and auditing support for the TCRN. (2) Operationalizing the TCRN (3) assisting pilot projects between the centers. (4) Marketing and dissemination of TIES. As a result of our efforts, two new Cancer Centers have requested membership and are bringing up their TIES nodes.
Conclusions: We have created a model federated system for data and biospecimen sharing across the nation’s cancer centers.

Future plans: We plan to expand the current network, creating a unique resource for personalized medicine and translational research.

Project Title: Cancer Deep Phenotype Extraction from Electronic Medical Records

Project Background: Precise phenotype information is needed to advance translational cancer research, particularly to unravel the effects of genetic, epigenetic, and systems changes on tumor behavior and responsiveness. Examples of phenotypic variables in cancer include: tumor morphology (e.g. histopathologic diagnosis), co-morbid conditions (e.g. associated immune disease), laboratory findings (e.g. gene amplification status), specific tumor behaviors (e.g. metastasis) and response to treatment (e.g. effect of a chemotherapeutic agent on tumor). Current models for correlating EMR data with -omics data largely ignore the clinical text, which remains one of the most important sources of phenotype information for cancer patients. Unlocking the value of clinical text has the potential to enable new insights about cancer initiation, progression, metastasis, and response to treatment. This U24 brings together two mature academic development groups with complementary skills and extensive expertise in biomedical informatics, cancer informatics, and natural language processing. The software being developed in this project will advance the state-of-the-art for EMR based phenotyping, by transforming the paradigm of phenotyping from retrospective queries and dichotomous outputs to prospective pipeline systems and deep phenotype extraction.

Research objectives: Develop and evaluate new methods for deep phenotype extraction combining NLP, machine learning, ontologies and production rule systems.

Principal methods: The basic methods draw on efforts of both contributing labs to current methods in clinical NLP. We will develop methods for extracting general modifiers (negation, uncertainty, subject) and cancer specific characteristics (e.g. grade, invasion, lymph node involvement, metastasis, size, stage). Several aims propose investigation of biomedical information extraction where there has been little or no previous work (e.g. clinical genomic entities, and causal discourse). Visualization of extracted data, usability of the software, and dissemination are also emphasized. Three driving oncology projects led by accomplished translational investigators in Breast Cancer, Melanoma, and Ovarian Cancer will drive development of the software. The proposed research bridges novel methods to automate cancer deep phenotype extraction from clinical text with emerging standards in phenotype knowledge representation and NLP.

Recent results: We have adopted the Agile development methodology and have made steady progress in developing the base architecture for the DeepPhe platform. We have completed the Breast Cancer phenotype model and have identified and prioritized the data elements that we will extract. Extraction and summarization of TNM, Stage and biomarkers is completed.

Conclusions: The DeepPhe project will provide an entirely new paradigm for deep phenotyping, that builds on prospective and ubiquitous processing of clinical reports and ancillary data.

Future plans: We will further develop and text these methods, and will implement them at University of Pittsburgh as part of the Cancer Information Service. A first public release of the software is planned for Spring, 2016.
Xia Jiang, PhD

**Project Title:** Detecting Genome Wide Epistasis with Efficient Bayesian Network Learning  
**Project Background:** To discover epistatic interactions, ideally we would like to analyze the correlation of every subset of loci with the disease. However, this is not possible with the data sets in GWA studies. If we have 500,000 loci (which is typical in a GWAS), we would need to investigate $2^{500000}$ interactions. The potential of GWA studies will not be realized until we develop statistical methods to handle the large data sets currently available. Such methods would need to find the interacting loci while avoiding an exhaustive search.  
**Research objectives:** The objective is to develop new efficient methods for performing a GWAS by applying Bayesian Network learning methodology. This work builds on my previously developed combinatorial BN method which outperformed a well-known combinatorial method using simulated data, and that used real data to substantiate an epistatic relationship between the APOE and GAB2 genes with Alzheimer's disease. We will implement my method in a pilot system and use that system to investigating the genetic basis of breast cancer and lung cancer.  
**Principal methods:** Bayesian Network Learning; Efficient Search Algorithms  
**Recent results:** We introduced the Bayesian network posterior probability (BNPP) method which addresses the difficulties. The method represents the relationship between a disease and SNPs using a directed acyclic graph (DAG) model, and computes the likelihood of such models using a Bayesian network scoring criterion.  
**Conclusions:** Our results concerning simulated data sets indicate that the BNPP exhibits both better evaluation and discovery performance than does a p-value based method. For the real data sets, previous findings in the literature are confirmed and additional findings are found.  
**Future plans:** 1) Our causal knowledge comes from heterogeneous datasets. We can use the knowledge obtained in this research concerning the causal features to develop a single dataset that contains more of the discovered causal features and learn from this dataset. We may improve performance by so doing. 2) The knowledge base can be extended with other types of features such as protein, cell line and pathway features;

Xinghua Lu, M.D., Ph.D.

**Project Title:** Translational Bioinformatics Methodology  
**Project Background:** My group’s research concentrates on developing translational bioinformatics methodology for personalized cancer medicine.  
**Research objectives:** Identify driver somatic genomic alterations (SGAs) from each the tumor of each individual tumor; Detecting aberrant pathways that are suitable for existing molecularly targeted anti-cancer drugs; Systematically examine cancer genomic big data to search for potential synthetically lethal targets for tumors with specific tumors.  
**Principal methods:** Bayesian network, causal inference, deep learning.  
**Recent results:** We are able to identify driver mutations at individual tumors, which to our best knowledge is the first method that capable of achieve this goal. Using deep learning algorithm, we were able to identify genes regulated by common transcription factors and their upstream regulators.
Conclusions: We have made significant progress towards personalized medicine for cancer patients.

Future plans: We will apply the algorithms mentioned above to cancer drug response data set to develop computational models that are capable of predicting tumor responses to existing anti-cancer drugs to guide clinical decision for cancer patients.

Fu-Chiang (Rich) Tsui, Ph.D.

Project Title: Adult Readmission Reduction

Project Background: Hospital readmissions pose quality of patient care and financial burden to hospitals. Various payers such as Medicaid, Accountable Care Organizations (ACOs) and Highmark Blue Cross Blue Shield have non-payment policies for readmissions within 30 days from discharge. Moreover, readmission rate is one of metrics to evaluate quality of hospital’s patient care. We propose to develop a System for Hospital Adaptive Readmission Prediction and Management (SHARP) to reduce 30-day readmission rates efficiently and effectively. Current approach for readmission prediction is ad-hoc and lack of comprehensive modeling using both structured and unstructured EHR.

Research Objectives:
- Evaluate prediction performance of each data type among structured and unstructured EHR using Bayesian models for heart failure hospital admissions.
- Develop, refine and evaluate predictive Bayesian models that use multiple data types available from heart failure (HF) hospital admissions.
- Develop, refine, and evaluate predictive models from pneumonia and acute myocardial infarction (AMI) hospital admissions.

Principal Methods:
We propose to adopt big data approach by using all unstructured electronic health records (EHR), e.g., clinical freetext reports] and structured EHR, e.g., registration, laboratory and medication data, to predict pediatric readmissions in 30 days.

We plan to use natural language processing to extract findings, symptoms from freetext reports such as progress notes, and history and physical. Using NLP, we can unlock valuable patient information in narrative reports previously not being extensively utilized due to the lack of NLP. We plan to develop and evaluate Bayesian networks that analyze findings and symptoms extracted from unstructured and structured EHR to identify patients with risk of readmission.

Recent Results: Through extracting 30,000 more features from inpatient electronic health records in 13 UPMC hospitals, the HF predictive model reached AUROC (c-stat) that was significantly better than the performance of an ad-hoc algorithm published in a 2013 Journal of American Medical Association (p-value = 0.0002) and the performance of multiple physicians’ manual chart review. More than 20,000 HF visits were used to train the predictive model and the test data for HF evaluation comprised over 3,000 admissions.

Conclusions: SHARP system outperformed existing predictive models for 30-day hospital readmissions that were built through ad-hoc approach such as physician consultation and literature search. We attribute our superior performance from two aspects: big data approach (pure data mining) and disease specific.

Future Plans: We plan to resubmit a grant proposal to NLM for conducting the research. We plan to implement the system in UPMC adult hospitals.
**Project Background:** Pediatric Readmission Reduction

**Project Background:**
Hospital readmissions pose quality of patient care and financial burden to hospitals. Various payers such as Medicaid, Accountable Care Organizations (ACOs) and Highmark Blue Cross Blue Shield have non-payment policies for readmissions within 30 days from discharge. Moreover, readmission rate is one of metrics to evaluate quality of hospital’s patient care. We propose to develop a System for Hospital Adaptive Readmission Prediction and Management (SHARP) to reduce 30-day readmission rates efficiently and effectively.

Current approach for readmission prediction is ad-hoc and lack of comprehensive modeling using both structured and unstructured EHR.

**Research Objectives:**
- Evaluate prediction performance of each data type among structured and unstructured EHR using Bayesian models for seizure hospital admissions.
- Develop, refine and evaluate predictive Bayesian models that use multiple data types available from seizure hospital admissions.
- Develop, refine, and evaluate predictive models from non-disease specific hospital admissions and compare performance between models obtained in Aim3 and Aim2 for seizure admissions.

**Principal Methods:**
We propose to adopt big data approach by using all unstructured electronic health records (EHR), e.g., clinical freetext reports] and structured EHR, e.g., registration, laboratory and medication data, to predict pediatric readmissions in 30 days.

We plan to use natural language processing to extract findings, symptoms from freetext reports such as progress notes, and history and physical. Using NLP, we can unlock valuable patient information in narrative reports previously not being extensively utilized due to the lack of NLP. We plan to develop and evaluate Bayesian networks that analyze findings and symptoms extracted from unstructured and structured EHR to identify patients with risk of readmission.

**Recent Results:**
Through extracting 30,000 more features from inpatient electronic health records in Children’s Hospital of Pittsburgh of UPMC, the all-cause model reached AUROC (c-stat) 0.83. The seizure predictive model reached AUROC that was significantly better than the performance of the HOSPITAL index published in 2013 Journal of American Medical Association and the Rothman Index (RI), a product of Pera Health (p-value < 10^-5), and existing approach LACE model (p-value<0.002).

**Conclusions:**
SHARP system outperformed existing predictive models for 30-day hospital readmissions that were built through ad-hoc approach such as physician consultation and literature search. We attribute our superior performance from two aspects: big data approach and disease specific.

**Future Plans:**
SHARP will be fully integrated into Cerner EHR system and in production providing services to inpatient clinicians in 2016 Q2. We plan to submit R01 research proposals for prospective evaluation of SHARP in 2016.
Shyam Visweswaran, M.D., Ph.D.

**Project Title:** Development of learning electronic medical records (EMRs)

**Project Background:** Electronic medical records (EMRs) are capturing increasing amounts of patient data that can be leveraged by machine-learning methods for computerized decision support. This research focusses on the development of intelligent EMRs that contain adaptive and learning components to provide decision support using the right data, at the right time.

**Research objectives:** Develop a prototype EMR that 1) tracks which patient data are of clinical interest to a physician, 2) uses machine learning to learn models that predict data that are of clinical interest in a patient case, and 3) selectively highlights patient data that are predicted to be of interest by models.

**Principal methods:** We developed a prototype learning EMR system called LEMR. Using LEMR a physician user evaluated 58 critical care patient cases to indicate which laboratory tests were most relevant in a patient case. Logistic regression models were constructed using this data and their performance was evaluated using ROC curves.

**Recent results:** The ROC curves for 10 laboratory tests (out of 43 distinct laboratory tests) had areas under the ROC curve ranging from 0.53 to 1.00.

**Conclusions:** We successfully designed and built a prototype LEMR, and the results indicate that this approach can predict laboratory tests of clinical interest with high accuracy.

**Future plans:** Refine the LEMR prototype and evaluate it on a large set of critical care patient cases.

Michael M. Wagner, M.D., Ph.D.

**Project Title:** MIDAS Informatics Services Group (ISG)

**Project Background:** Models of Infectious Disease Agent Study (MIDAS)

Funded by the [National Institute of General Medical Sciences](https://www.nigms.nih.gov) at the NIH, MIDAS is a collaborative network of research scientists who use computational, statistical, and mathematical models to understand infectious disease dynamics and thereby assist the nation to prepare for, detect, and respond to infectious disease threats. The University of Pittsburgh in collaboration with Carnegie Mellon University, Pittsburgh Supercomputing Center, the University of Arkansas Medical Sciences, and Johns Hopkins University is forming the MIDAS Informatics Services Group (ISG), which will develop and provide services to researchers and practitioners in the field of infectious disease epidemiology.

The broad goal of the project is to catalyze research in infectious disease epidemiology and to improve the related practice of disease control. The project will use the methods of service-oriented architectures and ontologies to build an informatics infrastructure that will enable MIDAS researchers to develop larger and more complex models and larger and more capable systems.

**Research objectives:** The specific aims of the project are to: (1) Develop software for end users ranging from modelers to decision makers; (2) synthesize populations and environments for use by modelers; (3) significantly extend a prototype Apollo Library of standardized computable information; (4) significantly extend an ontology-based Information
Management System; (5) create an “On Demand” High Performance Computing service; and (6) play other catalytic roles expected of the informatics resource, including logistical support, data acquisition, model validation, software engineering and quality control.

**Principal methods:** The project itself will develop an ontology-based information management system that will index datasets, publications, existing models, and computer-interpretable information—the ‘raw materials’ of modeling. The project will also employ informatics methods from the field of knowledge representation to construct a library of computer-interpretable information that can be re-used. The re-use of information will enable the construction of potentially ecosystem-size models. The project will also provide non-computational services in support of software engineering, model validation, dataset acquisition, and meeting logistical support for the MIDAS network.

**Recent results:** This is a new project with a start date of August 1, 2014. This is a new project with a start date of August 1, 2014.

The Software development service (SDS) of the group has developed the following systems and software:

- a revised website for the MIDAS network built on the Drupal content management system
- a self-serve bibliographic database to the MIDAS portal
- Ontology Based Catalog of Infectious Disease Epidemiology (OBC.ide)
- Apollo Location Service – a database and web-based API for representing temporally and spatially unambiguous locations on the earth
- Apollo Library Service – a database and web-based API for machine interpretable outbreak data
- Apollo Library Viewer – a web-based application that accesses the library service

The MISSION group of ISG which focuses on training of developers and customers on topics relating to productization of software and best-practices in software development held two online training sessions, one in-person training session and held three online seminars.

The High Performance Computing Services of the ISG has built and made available the Olympus compute cluster. It comprises 1536 compute cores and 7.5TB of RAM. 48 users of 14 different institutions have used Olympus for work in the domain of disease and outbreak simulation totaling three million CPU hours. The HPC services are also hosting the Apollo Web Service, which currently supports five different outbreak simulation models.

The Information Management Group of the ISG cataloged over 900 publications, datasets, reports, models, and software programs into the OBC.ide system. They have reengineered the user interface of the OBC.ide system to make it more usable and show more details about each indexed item.

The Synthetic Ecosystems Group of the ISG has developed software that can generate synthetic ecosystems for the entire world. To date, they have created synthetic ecosystems for eight West African countries and the entire United States.

The Apollo Library group has created libraries of all known Ebola outbreaks and selected influenza outbreaks.

We have provided logistical support for two MIDAS meetings one in New Orleans and one in Atlanta. This support is through the efforts of Ms. Szczepaniak (DBMI staff) and Ms. Bocaner (Cosmic Events, LLC).
Conclusions: Not applicable.

Future plans: The Software Development Service is collaborating with Los Alamos and Tulane University to develop the MIDAS Visualization service which is a web-based tool for modelers to create and share interactive maps and time series that illustrate disease data and/or the output of their models.

The HPC services group is developing a virtual machine based architecture for the HPC Web Service. It will allow users to easily connect to multiple compute resources including Amazon EC2, XSEDE and Olympus.

The ISG is collaborating with Willem Van Panhuis from the Pittsburgh MIDAS Center of Excellence. Mentored by Drs. Wagner and Cooper, he was awarded a new BD2k K award. He will be creating a library of known Chikungunya and Dengue outbreaks, develop logic-based algorithms to combine data with simulators and develop value of information methods to prioritize data collection. He also intends on taking Project Tycho data and migrating it to the Apollo standard.

The Synthetic Ecosystems group will distributing their software as an R package that anyone can use and modify.

We are developing the Peak-date-prediction Challenge which permits contestants to submit a disease transmission model of influenza outbreaks or submit weekly predictions of the peak date of influenza season. We built this using a system developed by Gregory Cooper’s group—a system that detects and characterizes outbreaks of disease from clinical reports.

Project Title: Probabilistic Disease Surveillance

Project Background: This project is furthering development and evaluation of a probabilistic approach to disease surveillance. In this approach, a probabilistic case detection system (CDS) uses Bayesian diagnostic networks to compute the likelihoods of patient findings for each of a set of infectious diseases for every patient in a monitored population. This is a novel, integrated, Bayesian approach for the early and accurate detection of cases of diseases that threaten health and for detection of outbreaks of diseases that threaten public health. Our approach has significant potential to improve the information available to public health officials and physicians, which can be expected to improve clinical and public health decision making, and ultimately to improve population health.

Research objectives: The specific aims of the proposed research are to (1) Significantly advance the development and integration of the NLP, case detection (CDS) and outbreak detection and characterization (ODCS) system components. We will add structured data types to CDS and advanced capabilities to ODCS, including the ability to detect concurrent outbreaks and outbreaks of unknown diseases. We also propose to further develop the knowledge bases of the Topaz NLP framework that extracts patient findings from ED reports. (2) Expand the disease models. We propose to extend the approach from influenza to three other respiratory diseases. (3) Evaluate the components and the system. We will measure the performance of the NLP, CDS, and ODCS components individually, including their portability, and as a system. (4) Disseminate the projects results, through publications, presentations, and computer code.

Principal methods: ODCS also utilizes a Bayesian approach to compute the probability that an outbreak is ongoing for each of a set of infectious diseases of interest, given information from CDS. ODCS also computes probability distributions over the expected size of a detected outbreak, its expected time course, and other characteristics required by public
health officials to respond effectively to an outbreak. The research will extend the approach, which we have already developed and evaluated for the disease influenza in one region, to a second region and eight additional respiratory infectious diseases. The research will also extend the capabilities of ODCS to utilize non-EMR data, detect an unknown disease, and detect and characterize concurrent outbreaks. The planned evaluations will measure the accuracy of both CDS and ODCS using historical surveillance data from the two regions and simulated outbreak data, created by adding outbreak cases generated by an agent-based epidemic simulator to the historical surveillance data.

**Recent results:** In year 1, we performed a new empirical evaluation of the outbreak detection and characterization system (ODS) using simulated influenza outbreaks. The results indicate that an outbreak is typically detected as highly likely at 52 days after it started, at which point about 6% of the outbreak cases have occurred. At 56 days into the outbreak, which corresponds to the first 10% of outbreak cases, the total number of cases (past, present, and future) is estimated with an error rate of about 9%, and the peak is estimated within an error of about 3 days. Since the peak day occurred on average at 74 days into the outbreak, these results provide support that influenza outbreaks can be detected and characterized well before the peak day is reached. The overall ODS performed well on this simulated data, which provides some support for its utility.

**Conclusions:** We have made good progress towards our stated goals for Y1 and expect to stay on track to progress as stated in our specific aims.

**Future plans:** We plan to complete the design of a method for modularizing how diseases are represented in ODS. We also plan to finish developing a version of ODS that runs on an HPC cluster. In Year 2, we will investigate using five-digit zip codes for spatial modeling of infectious diseases in the project. We will perform empirical evaluations using simulated data and using real data from Salt Lake City and Allegheny County. We also plan to implement and evaluate a method for detecting outbreaks of unmodeled diseases.

**Project Title:** Apollo: Increasing Access and Use of Epidemic Models through the Development and Adoption of a Standard Ontology

**Project Background:** The broad aim of the proposed work is to make epidemic models as readily available to health agencies and as easy-to-use as are weather models for local TV stations. The project will standardize the vocabulary and syntax used in epidemic modeling. By standardizing the many parameters that epidemic models require to represent disease transmission, the health status of a population, disease transmission, and disease control measures, the project will decrease the time and effort required to develop epidemic models and will make them more available to scientific and clinical users of these models.

**Research objectives:** The specific aims are to (1) develop a standard vocabulary for the field of epidemic modeling using a tool called Protégé; (2) create two extensions to Protégé that are needed by the project; (3) develop a standard syntax using the vocabulary for representing the inputs (e.g., disease control measures) and outputs of epidemic models and to use this syntax in an existing system called the Apollo Web Service that makes it possible for other computer programs to access epidemic models; and (4) to increase the capacity to run epidemic models on supercomputers so as to demonstrate the value of the work of the first three aims.
**Principal methods:** The project works with epidemic modelers from the Models of Infectious Disease Agent Study (MIDAS) research network to develop the standard vocabulary and syntax. It uses the extended Protégé software to develop the standard vocabulary as an ontology, which is a formal representation of objects and their interrelations. It represents syntax using the XML language. An analysis of the input and output data currently used by epidemic models will drive the design of the syntax.

**Recent results:** In year 2, we created version 2.0 of the Apollo standard syntax and terminology for representing infectious disease scenarios in DTM configuration files. The standard syntax is defined by Apollo XSD 2.0 and the standard terminology by Apollo-SV ontology 2.0. Apollo-SV 2.0 and Apollo XSD 2.0 are free and open under version 2.0 of the Apache license at our Google Code site [http://code.google.com/p/apollo](http://code.google.com/p/apollo). Apollo-SV is available in a standard Web Ontology Language (OWL) format, and is viewable at [http://www.ontobee.org/browser/index.php?o=APOLLO_SV](http://www.ontobee.org/browser/index.php?o=APOLLO_SV).

**Conclusions:** We have made significant progress towards our stated goals and are on track to complete our aims as listed above.

**Future plans:** In year 3, our plans for developing, implementing, and promoting the adoption of standards include the following examples:

1. **Standards development:** Includes adding standard identifiers for vaccines, infectious diseases, drugs, and entities in abiotic ecosystem censuses (e.g., school, household, group quarters, and weather) to Apollo and recommend their adoption to the MIDAS research network.
2. **Standards implementation:** Plans are to connect an additional two DTMs to the Apollo Web Services.
3. **Standards Adoption:** We plan to propose adoption of standard identifiers to the MIDAS Executive Committee.
New Research Initiatives

Faculty in the Department of Biomedical Informatics have a number of research collaborations that have been initiated during FY2014 consisting of:

Richard D. Boyce, Ph.D.

Addressing Gaps In Clinically Useful Evidence On Drug-Drug Interactions
NIH R01
PI: R.D. Boyce, Ph.D.
2014 – 2018

Gregory F. Cooper, M.D., Ph.D.

Center for Causal Modeling and Discovery of Biomedical Knowledge from Big Data
NIH/U54
PI: G. Cooper, MD, PhD
2014-2018

Xia Jiang, PhD

A New Generation Decision Support System
NIH/NLM
PI: X. Jiang, Ph.D.
2014-2018

Xinghua Lu, M.D., Ph.D.

Deciphering Cellular Signaling System
NIH/R01
PI: X. Lu, MD, PhD
2015-2019

Fu-Chiang (Rich) Tsui, Ph.D.

Children's Hospital of Pittsburgh SHARP
CHP SHARP
PI: Rich Tsui, Ph.D
2015-2016
Major Collaborations (Outside of the University)

Faculty in the Department of Biomedical Informatics maintain a wide range of significant collaborations with other faculty at other institutions. The following listing provides some of the major and active collaborations:

**Michael J. Becich, M.D., Ph.D.**

Clinical and Translational Science Award (CTSA) Biomedical Informatics Core for the Clinical and Translational Science Institute (CTSI) at Pittsburgh
60 funded CTSA sites
Co-Director, CTSI: M Becich, M.D., Ph.D.
Member, CTSA Steering Committee and Operations Committee: M Becich
2006 - 2011

National Mesothelioma Virtual Bank for Translational Research
NMVB/CDC NIOSH
University of Pennsylvania, Roswell Park Cancer Institute, New York University
PI/Program Director: M. Becich, M.D., Ph.D.
2006 – 2016

Sarcoidosis and A1AT Genomics & Informatics Center
NHLBI
Arizona Health Sciences Center, Johns Hopkins University, Medical University of South Carolina, National Jewish Health, UCSF, University of Pennsylvania, Vanderbilt University and Yale University
PI/PDs: N. Kaminski (Yale), M. Becich, S. Wisniewski
2012-2015

PaTH: A Learning Health System for the Mid-Atlantic Region
PCORI CDRN
Penn State University, Temple University and Johns Hopkins University
Co-I and Informatics Coordinating Site PI
2014-2016

**Tanja Bekhuis, PhD, MS, MLIS, A.H.I.P.**

Research to support systematic reviewers.
Dr. Dina Demner-Fushman, MD, PhD, Staff Scientist
Lister Hill National Center for Biomedical Communications, Communications Engineering Branch, US National Library of Medicine, NIH
Promotion of interoperable dental terminologies.
Dr. Elsbeth Kalenderian, DDS, MPH, PhD, Chair, Oral Health Policy and Epidemiology; Chief of Quality; Associate Professor in Oral Health Policy and Epidemiology
Harvard School of Dental Medicine

Clinical decision-support system to improve delivery of tobacco interventions and quitline referrals by dentists.
Dr. Donald Rindal, DDS, Senior Research Investigator
Dental Group, HealthPartners Institute for Education and Research

Promotion of social interactions to change clinical practice in dentistry.
Sharon Tracy, PhD, Assistant Director
American Dental Association Center for Evidence-based Dentistry

Commercialization of Evidence in Documents, Discovery, and Analysis (Edda) System
Don Taylor, MS, MBA, PhD, Executive-In-Residence
Pittsburgh Life Sciences Greenhouse

Richard D. Boyce, Ph.D.

Individualized Drug Interaction Alerts
University of Arizona
PI: Daniel C. Malone, Ph.D.
Role: Co-Investigator
2015 - present

Gregory F. Cooper, M.D., Ph.D.

Complex Pattern Analysis in Data
Carnegie Mellon University
PI: A Dubrawski, Ph.D.
Role: Co-Investigator
2009 - present

Women’s Cancer Research Center
Developmental Project
PIs: A Lee and R Edwards, Magee Women’s Research Institute
2010 - 2014
Gerald P. Douglas Ph.D.

Non Communicable Diseases using Chronic Care Clinic Model
World Diabetes Foundation
Baobab Health Trust
Pl: B Simwaka, Ph.D.
Jul 2012 – Jun 2014

Strengthening Management of HIV/AIDS and Non Communicable Diseases Through Appropriate Medical Informatics
International Union Against Tuberculosis and Lung Disease
Baobab Health Trust
Pl: Multiple
Jan 2012 – Dec 2013

Madhavi Ganapathiraju, Ph.D.

Study of Computationally Discovered Human Protein-protein Interactions
Indian Institute of Science, Bangalore, India
V Ravindranath

Vanathi Gopalakrishnan, Ph.D.

Inflammatory Bowel Disease
University of California Los Angeles
Graeber Lab, Ph.D.

Pediatric Cardiomyopathy
Children’s Hospital of Los Angeles
John Wood, M.D., PhD

Proteomics for Lung Cancer
Vanderbilt University
Pierre Massion, M.D.

Biomarkers for Classification of Esophageal Adenocarcinoma
Institute for the Treatment of Esophageal and Thoracic Disease, Allegheny Health Network, Pittsburgh, Pennsylvania
Jobe Blair, MD and Zaidi Ali, MD

Transfer Rule Learning - Functional Mapping for Integrative Modeling of Panomics Data
McDonnell Genome Institute
Washington University School of Medicine
Makedonka Mitreva, PhD
Harry Hochheiser, Ph.D.

Semantic LAMHDI: Linking diseases to model organism resources
NIH
Oregon Health Sciences University
PI: M. Haendel, PhD. Collaborators: Lawrence Berkeley National Labs – C. Mungall, S. Lewis, N. Washington; University of California San Diego: - M. Martone, J. Grethe, A. Bandrowski, T. Whetzel, A. Gupta
2012-2014.

Building Phenotype Comparison Tools for the Undiagnosed Disease Program
NIH/NHGRI Contract
NIH Undiagnosed Disease Program
2013-2014

Quantifying Electronic Medical Records Usability to Improve Clinical Workflow
AHRQ
San Diego VA;
2012-2016.

Sarcoidosis and A1AT Genomics & Informatics Center
NHLBI
Yale
PI: N. Kaminski (Yale), M. Becich, S. Wisniewski
2012-2015

Interactive Search and Review of Clinical Records with Multi-layered Semantic Annotation
NLM
UCSD
PI: W. Chapman (UCSD), R. Hwa, J. Wiebe
2011-2014.
Rebecca Jacobson, M.D., M.S.
Formerly Rebecca Crowley MD, MS

Advanced Development of TIES

University of Pennsylvania
MJ Feldman, M.D., Ph.D.
2004 - present

Roswell Park Cancer Institute
Carmelo Gaudioso, MD, MBA, PhD
2013 - present

Roni Bollag, MD, PhD
Georgia Regents University
2013-present

Cancer Deep Phenotyping from Electronic Medical Records

Children’s Hospital of Boston
Guergana Kirilova Savova, Ph.D.
2010 - present

Guoxiang Jian, MD, PhD
Mayo Clinic
2015 - present

Xinghua Lu, M.D., Ph.D.

Cancer therapeutic biomarker identification
Stony Brook University of New York
Yusuf A Hannun, M.D.

Cancer metabolics and metastasis of glioblastoma
Shiyuan Cheng PhD,
Northwestern University, Chicago, IL

Ichiro Nakano, MD
University of Alabama at Birmingham, Birmingham, AL

Fu-Chiang (Rich) Tsui, Ph.D.

Pediatric Hospital Readmission Prediction
Children’s Hospital of Pittsburgh of UPMC
Steve Docimo, MD, Chief Medical Officer
Andrew Urbach, MD, Associate Chief Medical Officer
Harun Rashid, Chief Information Officer
2013 – present

**Adult Heart Failure Hospital Readmission Prediction**
UPMC Heart and Vascular Institute
Mark Schmidhofer, MS, MD, Director, Cardiology Consult Service
Ravi Ramani, MD, Director, UPMC Integrated Heart Failure Program
Michael Mathier, MD, Director, Section of Heart Failure and Pulmonary Hypertension
2013 – present

**Post-discharge Monitoring and Evaluation**
UPMC Heath Plan
Pamela Peele, PhD, Chief Analytics Officer
2015 – present

**Influenza and ILI Case Detection Project**
Allegheny County Health Department
Ron Voorhees, MD, Chief of Epidemiology and Biostatistics
Kristen Mertz, MD, MPH, Medical Epidemiologist
2009 – present

**Freetext Clinician Reports Processing**
Columbia University, Carol Friedman, PhD, Professor, 2007-present
Nuance Communications, Henk Harkema, PhD, Principal NLP Engineer, 2014-present

**RODS Deployment**
Tarrant County Department of Health, TX
Jeremy Espino, MD, MS
2005 – present

**Shyam Visweswaran, M.D., Ph.D.**

**Complex Pattern Analysis in Data**
Carnegie Mellon University, Pittsburgh, PA, USA
A Dubrawski, Ph.D., D Neill, Ph.D., and J Schneider, Ph.D.
2009 - present

**Visual Analytics of Genomic Data**
The University of Texas Medical Branch, Galveston, TX, USA
SK Bhavnani, Ph.D.
2011 – present
Predictive Modeling in Cardiovascular Disease in Subjects of Indian Origin
Jawaharlal Institute of Postgraduate Medical Education & Research (JIPMER),
Puducherry, India
S Satheesh, M.D., D.M.
2011 – present

Michael Wagner, M.D., Ph.D.

Probabilistic Disease Surveillance
University of Utah and Intermountain Healthcare, Salt Lake City, UT
2013- present

Apollo: Increasing Access and Use of Epidemic Models Through the Development and Adoption of a Standard Ontology
Co-PI with William R. Hogan, Division of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR
2012- present

Trisano-CASIPH Project Collaboration
Allegheny County Department of Health and Tarrant County Department of Health, TX
AK Dey; JU Espino; RE Voorhees; and, M Wagner

MIDAS – Models of Infectious Disease Agent Study
MM Wagner, JU Espino
2014-present
Research Faculty: Summary of Interests & Initiatives

Michael J. Becich, M.D., Ph.D.
Chairman, Department of Biomedical Informatics
Professor of Biomedical Informatics, Pathology, Information Sciences, and Clinical and Translational Science
Director, Center for Commercial Applications (CCA) of Healthcare Data
Associate Director for Cancer Informatics, University of Pittsburgh Cancer Institute
Associate Director, Clinical and Translational Science Institute, University of Pittsburgh

Dr. Becich’s current research focuses on developing translational informatics tools and services, including data warehouses and data mining strategies for genomic, proteomic and microbiomic data. His interests also include clinical research informatics, particularly tissue bank information systems, clinical trials information systems and imaging repositories that currently are serving the clinical and translational research needs at University of Pittsburgh. Dr. Becich leads nationally recognized programs in Translational Informatics including the Cancer Clinical and Translational Science Awards (CTSA) Biomedical Informatics Core (see http://www.ctsi.pitt.edu/ ) and is founder of both the Association for Pathology Informatics (see http://www.pathologyinformatics.org) and Advancing Practice, Instruction and Innovation through Informatics (APIII) which in 2010 became Pathology Informatics 201X (see http://www.pathologyinformatics.com ). Pathology Informatics is a national meeting transforming translational research through training and continuing medical education in pathology informatics.

Dr. Becich’s laboratories are funded by grants from the National Institutes of Health, National Center for Research Resources, National Heart, Lung and Blood Institute, National Library of Medicine, National Cancer Institute, Centers for Disease Control and the Patient Centered Outcomes Research Institute as well as multiple corporate sponsored research programs and unrestricted educational grants which support Pathology Informatics and API.

Tanja Bekhuis, Ph.D., M.S., M.L.I.S., A.H.I.P.
Assistant Professor of Biomedical Informatics
Biomedical Informatics Training Program Core Faculty
Assistant Professor of Dental Public Health
Assistant Professor, School of Information Sciences

Dr. Bekhuis’ focus is on translational research informatics. She is particularly interested in developing methods to support comparative effectiveness researchers, including systematic reviewers who synthesize medical evidence, and biomedical librarians and trials search coordinators who develop complex search strategies. She also studies the language scientists use to describe their own research in published literature.
Regarding her work to facilitate translational research, Dr. Bekhuis published a paper in *Plos One* (2014) in which she proposed a decision-support system for systematic reviewers of medical evidence. This paper was featured in the AHRQ Scientific Resource Center Methods Library (March 2014). More recently, Dr. Bekhuis improved upon her methods reported in *Plos One* by developing a hybrid system. She described this novel strategy in a paper recently submitted to the International Workshop on Biomedical and Health Informatics, IEEE International Conference on Bioinformatics and Biomedicine. This new system combines logical rules as a preprocessor in combination with a machine learning component. Together, these components can reduce screening burden, as well as offer informative feedback for quality assurance and control. The EDDA Group released some of their code as open source; it is freely available in the RapidMiner Marketplace.

For several years, Dr. Bekhuis and her research team have been developing the EDDA Study Designs and Publications terminology. The early work for the terminology grew out of content analyses of classification tools developed by organizations such as the Agency for Healthcare Research and Quality, and the Cochrane Collaboration. The terminology is useful for automated classification and knowledge representation, and for information professionals who develop complex search strategies for systematic review teams. Her foundational work was published by the *Journal of the Medical Library Association (JMLA)*; 2013). A companion crosswalk for study designs was also featured in the AHRQ Scientific Resource Center Methods Library, May 2013. The EDDA terminology, version 1.3, was released in April 2015. It is freely available and can be downloaded in several file formats via the NCBO BioPortal (http://bioportal.bioontology.org/ontologies/EDDA). A major revision will be released upon completion of the current enrichment study.

Dr. Bekhuis continues to collaborate with researchers in the School of Dental Medicine, the American Dental Association, the Harvard School of Dental Medicine, and the HealthPartners Institute for Education and Research. Dr. Heiko Spallek, Associate Dean and Executive Director, Dr. Bernard Costello, Director of Translational Research, and Dr. Tanja Bekhuis, Director of Translational Research Methods, lead a new Center for Informatics in Oral Health Translational Research (CIOHTR) in the School of Dental Medicine. The overarching mission is to support research and education aimed at improving delivery of dental care and patient outcomes, as well as treatment of oral and maxillofacial conditions, particularly those related to systemic health. This Center significantly expands the mission of the old Center for Dental Informatics. Subsequent to opening the CIOHTR, the US National Institute of Dental and Craniofacial Research agreed to renew funding for trainee positions in the Department of Biomedical Informatics.

Regarding her dental informatics research, Dr. Bekhuis published a paper about the connection between oral health and systemic disease in *BMC Oral Health* (2013), which has “one of the highest ever scores” computed by Altmetrics for this journal. The *British Dental Journal* featured this paper in their news section (December 2013). The mixed methods Dr. Bekhuis developed and reported earlier in a paper published by the *Journal
of Medical Internet Research enabled this research. Several NIH proposals are under review relating to her dental informatics research.

**Richard D. Boyce, Ph.D.**  
*Assistant Professor of Biomedical Informatics  
Biomedical Informatics Training Program Core Faculty*

Richard D. Boyce, Ph.D., is currently the Principal Investigator of two NIH grants, a K01 from the NIA titled “Improving medication safety for nursing home residents prescribed psychotropic drugs” (K01 AG044433-01) and an R01 from the NLM titled “Addressing gaps in clinically useful evidence on drug-drug interactions” (R01 LM011838-01). Research projects include:

- Drug safety and decision support for older adults
- Knowledge-based approaches to drug-drug interaction and adverse drug event prediction, and identification
- Pharmacogenomics decision support

**Gregory F. Cooper, M.D., Ph.D.**  
*Vice Chairman, Department of Biomedical Informatics and Professor of Biomedical Informatics, Intelligent Systems, and Computational and Systems Biology.*

Dr. Cooper’s research involves the application of decision theory, probability theory, machine learning, Bayesian statistics, and artificial intelligence to biomedical informatics research problems. He has been investigating these topic areas for the past 25+ years and has published over 140 peer-reviewed papers. He is currently involved in the following research projects:

**Discovering Complex Anomalous Patterns in Data** (PI of a Pitt subcontract of an NSF grant to CMU) The goals of this project are to develop, implement, and evaluate a general and widely applicable framework for detecting potentially complex statistical patterns from data about entities in some set of interest, such as patterns of maintenance of jet aircraft in a fleet.

**Probabilistic Disease Surveillance** (Co-Investigator of an R01 from NLM) The aims of this project are to improve the ability of public health officials and physicians to estimate the current incidence of influenza and other infectious diseases and to predict the future course of epidemics of those diseases. The improved information will better support decisions made by health departments to control epidemics, which is expected to reduce morbidity and mortality from epidemic diseases.

**Transfer Rule Learning for Knowledge Based Biomarker Discovery and Predictive Biomedicine** (Co-Investigator of an R01 from NLM) The major goal of this project is develop, apply and evaluate novel Transfer Rule Learning (TRL) methods for integrative biomarker discovery from related biomedical data sets.
Center for Causal Modeling and Discovery of Biomedical Knowledge from Big Data (contact PI). The Center will develop and provide a powerful set of concepts, tools, training, and consortium activities that will accelerate the discovery and sharing of causal knowledge derived from very large and complex biomedical datasets.

Gerald P. Douglas Ph.D.
Assistant Professor of Biomedical Informatics
Assistant Professor of Health Policy & Management
Director, Center for Health Informatics for the Underserved

Dr. Douglas’s research focuses on applying the principles of medical informatics to improve healthcare in low-resource settings, both within the United States as well as internationally. He has particular interest in user-interface design and user experience. His research builds on techniques developed through 14 years of experience building point-of-care electronic medical record systems in Malawi. These techniques are captured in the curriculum of the graduate-level Principles of Global Health Informatics course, and Global Health Informatics Summer Internship in Malawi, created and taught by Dr. Douglas.

Madhavi Ganapathiraju, Ph.D.
Assistant Professor of Biomedical Informatics and Intelligent Systems

Dr. Ganapathiraju’s primary area of research is in Systems Biology, protein-protein interaction prediction at the system level and in translational application of the predicted interactions. A second core area is in Structural Biology, with membrane protein structure prediction from primary sequences. The third core area is in Sequence Analysis, pattern mining in whole-genome and whole-proteome sequences, with application of suffix array data structures for preprocessing the genome sequences. For all of these biomedical domains, Dr. Ganapathiraju develops novel algorithms with research basis in machine learning and network analysis and other areas of computer science.

One of Dr Ganapathiraju’s active research currently is in computationally discovering protein-protein interactions in the human interactome, specifically for proteins involved in mental health and in the systemic process of inflammation. This research is funded by the R01 grant of the BRAINS Award from National Institute of Mental Health. She is also working on developing strategies to identify protein-protein interactions that may be of high biological impact (impact prediction) and on strategies to disseminate the discovered interactions effectively to biologists.

Next, Dr. Ganapathiraju is working on personal genomes sequence analyses to identify the variation of pattern of sequences induced by single nucleotide variants.
Vanathi Gopalakrishnan, Ph.D.
Associate Professor of Biomedical Informatics, Intelligent Systems, and Computational Biology

Dr. Gopalakrishnan is interested in the design and development of computational methods for solving clinically relevant biological problems. She is fundamentally interested in technologies for data mining and discovery that allow incorporation of prior knowledge. For the last decade, she has developed and applied novel rule learning methods to biomarker discovery and verification from proteomic profiling studies. Her current research projects involve the development and application of novel variants of rule learning techniques to biomarker discovery and disease prediction for early detection and better understanding of mechanisms that cause neurodegenerative diseases, lung and breast cancers. Methods for incorporating prior knowledge that are being researched in her laboratory include text mining and ontology construction.

Current research projects include:

Bayesian Rule Learning for Disease Prediction and Biomarker Discovery: The major goals of this project are to develop, evaluate and refine novel Bayesian Rule Learning (BRL) methods that are algorithmically efficient, result in parsimonious models and accurately estimate predictive uncertainty from sparse biomedical datasets.

SPORE in Lung Cancer (Co-Director of Bioinformatics and Biostatistics CORE): The objectives of the UPCI Lung Cancer SPORE are to improve detection and treatment of lung cancer and to understand the mechanisms of increased susceptibility of women to lung cancer.

Transfer Rule Learning for Integrative Biomarker Discovery and Predictive Biomedicine: The major goals of this project are to develop, evaluate and refine novel Transfer Rule Learning (TRL) for learning predictive models from two or more biomedical data sets that are from related classification tasks.

Mining Biomedical Image Data for Actionable Knowledge: The major goals of this project are to develop, evaluate and refine a novel clinical workflow framework called Cardiovascular Magnetic Resonance Imaging Biomarker Extraction and Discovery (CMRI-BED) for classification of pediatric congenital heart disease.
Harry Hochheiser, Ph.D.
Assistant Professor of Biomedical Informatics

Dr. Hochheiser’s research activities stem from the application of techniques from the study of human-computer interaction to a variety of problems in biomedical informatics, in both basic research and clinical domains.

Bioinformatics Data Portals in Support of Collaborative Research:

For basic research, Dr. Hochheiser is interested in the development and evaluation of novel data exploration tool for bioinformatics, clinical, and translational data. For the Monarch/Semantic LAMHDI initiative and the related phenotype comparison tool project, Dr. Hochheiser has worked with basic researchers in the development of interactive information visualization widgets for exploring ontologically-inferred similarities between phenotype profiles, matching human phenotype profiles to comparable descriptions of model organisms and system.

For the Genomics Information Center for Genomic Research in Alpha-1 Antitrypsin Deficiency Syndrome and Sarcoidosis, we are building a data portal and cohort selection browser that will allow researchers from seven collaborating groups (and eventually from the broader research community) to identify subsets of interest for further analysis. Efforts include development of interactive tools that will support browsing, dynamic querying, and basic analyses for comparison across cohorts.

Clinical and Research Informatics: A variety of projects address the development and/or evaluation of usable information systems for patients, clinicians, and researchers. Currently active projects include:

- Development and evaluation of tools for the interactive view of structured data extracted from free-text in electronic medical records.
- Analysis of the interactions of physicians with electronic medical records in primary care settings, via video, eye-capture, and screen recording.
- Development of visual analytics tools in support of analysis of temporal phenotypes of cancer data, as extracted from clinical records via natural language processing.
- Examination of the usability of electronic medical records, particularly in low-resource settings.
- Study of user needs and design of information tools for the interpretation of pharmacogenomics and drug-drug interaction information
- Analytic tools for large-scale causal modeling
Rebecca S. Jacobson, M.D., M.S  
Professor of Biomedical Informatics, Pathology, and Intelligent Systems  
Director, Biomedical Informatics Training Program

Dr. Jacobson’s research focuses on (1) the development of methods deep phenotyping from electronic medical records, and the use of these methods in (2) large scale federated data and biospecimen sharing networks, and (3) learning health systems.

**Methods for Deep Phenotyping:** Dr. Jacobson has developed a unique system for information extraction from free text clinical reports. The TIES system was one of the first NLP pipeline systems to be used for clinical information extraction uses sophisticated natural language processing methods to process clinical text, and extract the meaning from the clinical records. The TIES system is deployed at University of Pittsburgh where it processes all pathology and radiology records, providing de-identified access to more than 24 million records on over 4 million patients, using an advanced interface for flexible search. The system is integrated with the HSTB and Radiology Honest Broker Systems, enabling rapid access to FFPE, frozen tissue, TMAs and radiology images. The TIES system is also in use at numerous institutions across the country including University of Pennsylvania, Roswell Park Cancer Center, Georgia Regents University, Thomas Jefferson Kimmel Cancer Center, MD Anderson Cancer Center, Memorial Sloan Kettering Cancer Center and a variety of other sites.

Funded by an NCI U24 (Jacobson, MPI), Dr. Jacobson is also collaborating with Dr. Guergana Savova to develop the Cancer Deep Phenotyping Pipeline (DeepPhe) which is designed to use NLP methods and other informatics methods to extract finely granular information from clinical records of patients with cancer and to combine these data with other information such as Cancer Registry. The project provides an entirely novel reframing of existing ideas about phenotyping (such as those described in the eMerge project). Results of this project will provide the methodologic underpinnings for large scale deep phenotyping required for national efforts such as the President’s Precision Medicine Initiative.

**Large Scale Federated Networks:** Through a second NCI U24 (Jacobson, PI), Dr. Jacobson is developing the TIES Cancer Research Network - a model for a national federated data and biospecimen sharing network. This network currently encompasses four cancer research networks and is expected to grow over the next three years. Efforts to develop the underlying trust agreements, policies and processes for the TCRN has established innovative new frameworks that are being used by the PCORI and CTSA networks. A publication describing TCRN was recently provisionally accepted for publication in Cancer Research – a leading AACR Journal.

**Learning Health Systems:** The development of methods for deep phenotyping is also supporting efforts for developing Learning Health System models to improve patient care. As an example, Dr Jacobson is currently working on a pipeline to automatically extract BIRADS scores from free text mammogram, ultrasound and Breast MR and correlate them with subsequent radiology and pathology results. The feedback will be
used for providing feedback to radiologists. Dr. Jacobson has previous experience in developing advanced educational systems.

Xia Jiang, Ph.D.
Assistant Professor of Biomedical Informatics, Intelligent Systems Program, and Carnegie Mellon – University of Pittsburgh Ph.D. Program in Computational Biology
Biomedical Informatics Training Program Core Faculty

Dr. Jiang has over 15 years of teaching and research experience in Bayesian Network modeling, machine learning, and algorithm design. She is currently focusing on developing novel algorithms/systems that improve the computational efficiency of high-dimensional data analysis, and network modeling of cancer genome data.

Dr. Jiang’s funded work, Detecting Genome-Wide Epistasis with Efficient Bayesian Network Learning, focuses on Epistasis, which is the interaction between two or more genes to affect phenotype. It is now widely accepted that epistasis plays an important role in susceptibility to many common and complex diseases. The crucial challenge to analyzing epistasis is finding a way to efficiently handle high-dimensional genomic data. This career award is to develop a succinct Bayesian network model representing epistasis and efficient algorithms, which are tailored to investigating such models, integrate the algorithms into methods for learning epistasis, and use simulated datasets to test the effectiveness of the methods and compare their performance to other methods.

Learning genetic epistasis using Bayesian network scoring criteria
Gene-gene epistatic interactions likely play an important role in the genetic basis of many common diseases. Recently, machine-learning and data mining methods have been developed for learning epistatic relationships from data. A well-known combinatorial method that has been successfully applied for detecting epistasis is Multifactor Dimensionality Reduction (MDR). Dr. Jiang created a combinatorial epistasis learning method called BNMBL to learn Bayesian network (BN) epistatic models. She compared BNMBL to MDR using simulated data sets. Each of these data sets was generated from a model that associates two SNPs with a disease and includes 18 unrelated SNPs. For each data set, BNMBL and MDR were used to score all 2-SNP models, and BNMBL learned significantly more correct models. In real data sets, we ordinarily do not know the number of SNPs that influence phenotype. BNMBL may not perform as well if we also score models containing more than two SNPs. Furthermore, a number of other BN scoring criteria have been developed. They may detect epistatic interactions even better than BNMBL.

Using Posterior Probability For Evaluating and Discovering Disease Loci Associations
Dr. Jiang introduced the Bayesian network posterior probability (BNPP) method which addresses the difficulties. The method represents the relationship between a disease and SNPs using a directed acyclic graph (DAG) model, and computes the likelihood of such models using a Bayesian network scoring criterion. The posterior probability of a hypothesis is computed based on the likelihoods of all competing hypotheses. The BNPP can not only be used to evaluate a hypothesis that has previously been discovered or suspected, but also to discover new disease loci associations. The results of her
experiments using simulated and real data sets are presented. Her results concerning simulated data sets indicate that the BNPP exhibits both better evaluation and discovery performance than does a p-value based method. For the real data sets, previous findings in the literature are confirmed and additional findings are found. The BNPP resolves a pressing problem by providing a way to compute the posterior probability of complex multi-locus hypotheses. A researcher can use the BNPP to determine the expected utility of investigating a hypothesis further. Furthermore, the BNPP is a promising method for discovering disease loci associations.

Evaluating De Novo Locus-Disease Discoveries in GWAS Using the Signal-to-Noise Ratio

A genome-wide association study (GWAS) involves examining representative SNPs obtained using high throughput technologies. A GWAS data set can entail a million SNPs and may soon entail many millions. In a GWAS researchers often investigate the correlation of each SNP with a disease. With so many hypotheses, it is not straightforward how to interpret the results. Strategies include using the Bonferroni correction to determine the significance of a model and Bayesian methods. However, when we are discovering new locus-disease associations, i.e., so called de novo discoveries, we should not just endeavor to determine the significance of particular models, but also concern ourselves with determining whether it is likely that we have any true discoveries, and if so how many of the highest ranking models we should investigate further. Dr. Jiang develop a method based on a signal-to-noise ratio that targets this issue. She applies the method to a GWAS Alzheimer's data set.

A new generation clinical decision support system

This is newly funded R01 grant is pertaining to big data to decision support in breast cancer clinic.

Critical clinical activities involve decision making. For both individual patients and for society at large, making good healthcare decisions is a paramount task. The objective of this research is to develop a novel decision support system that utilizes both the clinical features and the genomic profile of a breast cancer patient to assist the physician in integrating information about a specific patient (diagnostic subtype, tumor stage and grade, age, comorbidities) to make therapeutic plans for the patient.

Traditional clinical data are becoming increasingly available in electronic form. Unprecedentedly abundant genomic data are available to researchers as the results of advanced sequencing technologies such as next generation sequencing. Patient-specific genomic data are likely to become available for most patients in the foreseeable future. These sources of data provide significant opportunities for developing new generation clinical decision support systems that can achieve substantial progress over what is currently possible. However, the sheer magnitude of the number of variables in these data (often in the millions) presents formidable computational and modeling challenges. Also, integrating the heterogeneous information in multiple clinical datasets and genomic datasets presents an arduous challenge.
Breast cancer is the commonest cancer among women. Various breast cancer subtypes have been defined which, along with tumor stage, predict response to therapy and survival, albeit imperfectly. For example, HER2-amplified breast cancer is a subtype with poor prognosis, and therapy with an antibody to HER2 (Herceptin) has vastly improved the survival of such patients. Although Herceptin is used in the therapy of all patients with HER2-amplified tumors, only some respond. Also, it is expensive and can cause cardiac toxicity. So, it is important to give it only to patients benefiting from it. Studies show thousands of genes are associated with subtype and prognosis of breast cancer, and particular allele combinations may usefully guide the selection of effective treatment. The proposed system will amass all this genomic information and combine it with clinical information and therefore holds promise to provide accurate classification and treatment choices. It will build on previous results of the investigators in using Bayesian Network to learn from high-dimensional data sets.

As the principal investigator for this project, I will be responsible for managing the project at the University of Pittsburgh site. This includes setting milestones, ensuring achievement of the proposed research activities, coordinating timely publication and presentation of research results, and overseeing the fiscal aspects and compliance and regulatory concerns of the project. I have a long-standing history of successful research collaboration with the Co-PI, Dr. Neapolitan, at Northwestern University (NU), who will be responsible for managing the project at the NU site. I will work closely with Dr. Neapolitan to oversee successful communication and collaboration between the two sites. I will participate directly in the development and evaluation of the Bayesian network based decision support models and system. I will lead the tasks in Aim 1 involving acquiring datasets, processing data, and determining relevant features, and I will ensure the fulfillment of the tasks in Aim 2 involving the determination of the decision set, the utility set, and target/features relations. I will also take major responsibility for the design and development of algorithms involved in all the aims.

Xinghua Lu, M.D., Ph.D.
Associate Professor of Biomedical Informatics and Biomedical Informatics Training Program Core Faculty

Dr. Lu’s research focuses on the computational methods for identifying signaling pathways underlying biological processes and diseases as well as statistical methods for acquiring knowledge from biomedical literature. He was trained in Pharmacology and works in the field of bioinformatics after NLM sponsored postdoctoral training in Biomedical Informatics. His research interest concentrates on applying latent variable models to simulate biological signaling system and text mining. Currently, Dr. Lu is working on developing his research in translational bioinformatics and systems/computational biology and its application to specific domains relevant to human disease. He is pursuing collaboration in the area of natural language processing and text mining with the eventual goal of establishing a Center or Institute in Translational Bioinformatics.
Fu-Chiang (Rich) Tsui, Ph.D.
Assistant Professor of Biomedical Informatics, Intelligent Systems, Bioengineering, and Clinical and Translational Science
Biomedical Informatics Training Program Core Faculty

Dr. Tsui’s current research interest focuses on (big) clinical data science, mobile healthcare (m-health), medical natural language processing (NLP), machine learning, clinical informatics. His interests also include precision medicine, signal processing, data warehouse, large real-time production systems, biosurveillance, and public health informatics. Dr. Tsui has published over 90 peer-reviewed publications and his publications have been actively cited (Google h-index: 38).

Dr. Tsui has extensive knowledge and experience of building large clinical informatics production systems. He has built four large real-time clinical production systems: (1) Clinical Event Monitor System (CLEM) that serves as a decision support system and reports notifiable diseases at University of Pittsburgh Medical Center (UPMC) in real-time, (2) Realtime Outbreak and Disease Surveillance System (RODS) that collects and monitors emergency department visits at UPMC, and (3) National Retail Data Monitor (NRDM) that collects and monitors over-the-counter medication sales from over 32,000 retail stores across the nation.

Dr. Tsui leads several innovative projects in clinical informatics as new initiatives in 2015. He receives funding from McCune Foundation, Coulter Foundation, and Pennsylvania Innovation Works. His patent on readmission prediction and management is published in March 2015. One of his large new project is to build a production system (4) SHARP, a realtime decision support system embedded in Cerner®, one of the largest commercial EHR systems, which helps clinicians and patient care teams identify and manage patients at high-risk of 30-day hospital readmissions at the Children’s Hospital of Pittsburgh of UPMC.

Shyam Visweswaran, M.D., Ph.D.
Assistant Professor of Biomedical Informatics, Intelligent Systems, Clinical and Translational Science, and Computational Biology
Associate Director, Biomedical Informatics Training Program

Dr. Visweswaran’s research interests lie in the application of artificial intelligence and machine learning to personalized and genomic medicine, patient-specific predictive
modeling and computerized clinical decision support. Artificial intelligence and machine learning methods focus on the development of algorithms and methods for the learning of computational models from data.

**Personalized and Genomic Medicine:** Personalized medicine calls for the use of clinical, genomic and environmental data to more precisely evaluate risk, diagnose, assess prognosis, and tailor therapies to the individual. Genomic medicine is driving personalized medicine and focuses on the use of information obtained from sequences such as whole exomes and whole genomes. Genomic information, in combination with other clinical data, will lead to increased understanding of the biology of human health and disease, improved prediction of disease and effect of therapy, and ultimately the realization of precision medicine. Dr. Visweswaran’s work focuses on single nucleotide variants (SNVs) data obtained from genome-wide studies (GAWSs), and whole exome and whole genome data. In particular, Dr. Visweswaran focusses on 1) discovery of interacting SNVs in high-dimensional genomic data using Bayesian and information-theoretic methods, 2) development of efficient multivariate methods to rank SNVs in high-dimensional genomic data, and 3) development of computationally efficient population-wide and patient-specific predictive models from high-dimensional genomic data. This research is in collaboration with Gregory F Cooper, M.D., Ph.D. (Department of Biomedical Informatics) and Vanathi Gopalakrishnan, Ph.D. (Department of Biomedical Informatics)

**Patient-Specific Modeling:** In predictive modeling, the typical paradigm consists of learning a single model from a database of patient cases, which is then applied to predict outcomes for any future patient. Such a model is called a population-wide model because it is intended to be applied to an entire population of future cases. In contrast, patient-specific modeling focuses on learning models consists of learning models that are personalized to the characteristics of the patient at hand. Patient-specific models that are optimized to perform well on a specific patient are likely to be more precise than the typical population-wide models that are optimized to have good predictive performance on average on all future patients. Dr. Visweswaran focusses on 1) developing Bayesian and information-theoretic methods for learning patient-specific models from clinical and genomic data, and 2) applying patient-specific modeling to risk assessment, diagnosis, prognosis and selection of therapy. This research is in collaboration with Gregory F Cooper, M.D., Ph.D. (Department of Biomedical Informatics).

**Computerized Clinical Decision Support:** Dr. Visweswaran is involved in anomaly detection in clinical care including developing and implementing machine-learning methods that predict anomalies or deviations in therapy and clinical management of patients. This research is in collaboration with Milos Hauskrecht, Ph.D. (Department of Computer Science) Gilles Clermont, M.D., M.Sc. (Department of Critical Care Medicine), and Gregory F Cooper, M.D., Ph.D. (Department of Biomedical Informatics).
Michael M. Wagner, M.D., Ph.D.
Director, Corporate Relations; Director, RODS Laboratory
Associate Professor of Biomedical Informatics and Intelligent Systems

Dr. Wagner has been building public health information systems and conducting basic research in public health informatics for the past 14 years, having served as principal investigator on multi-institutional, interdisciplinary projects in this area funded by NLM, AHRQ, DARPA, DHS, CDC, the Pennsylvania Department of Health, and the Sloan Foundation. He authored The Handbook of Biosurveillance (Academic Press, 2006), which covers important aspects disease surveillance, model validation and data-use agreements. Dr. Wagner also served on two Defense Science Board Task Forces (Needs for Homeland Defense, Bio Warfare Panel, 2001; Defense Against Terrorist Use of Biological Weapons, 2002-2003) and was a planning member for the 2011 IOM meeting that studied the federal role in biosurveillance data integration (Information-sharing and collaboration: Applications to integrated biosurveillance, 2012, National Academies Press). He is currently the principal investigator for two NIH RO1 awards in closely related areas—the Probabilistic Disease Surveillance and Apollo projects.

Dr. Wagner has an abiding commitment to translation of research to practice. The Real-time Outbreak and Disease Surveillance (RODS) System is one of a handful of medical informatics projects that have completed the journey from research prototype to capitalized commercial products. The Apollo Web Services (NIGMS) and the Probabilistic Disease Surveillance System funded by the NLM are additional examples of systems that we have been aggressively translating to practice. New projects underway will integrate and operationalize all the systems, techniques, and theoretical concepts imagined possible when he began work in this area in 1999.
## Summary of Research Grants by Faculty

<table>
<thead>
<tr>
<th>Source of Funding</th>
<th>Principal Investigator</th>
<th>Project Start</th>
<th>Project End</th>
<th>Direct Costs for current budget period</th>
<th>Indirect Cost for current budget</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Institutes of Health</td>
<td>Michael J. Becich, M.D. Ph.D.</td>
<td>10-Sep-97</td>
<td>31-Jul-15</td>
<td>437,516</td>
<td>219,285</td>
<td>Cancer Center Support Grant</td>
</tr>
<tr>
<td>Centers for Disease Control and Prevention</td>
<td>Michael J. Becich, M.D. Ph.D.</td>
<td>1-Sep-06</td>
<td>31-Aug-16</td>
<td>718,550</td>
<td>307,014</td>
<td>Continuation of the National Mesothelioma Virtual Bank</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Michael J. Becich, M.D. Ph.D.</td>
<td>1-May-12</td>
<td>30-Nov-15</td>
<td>156,696</td>
<td>80,699</td>
<td>Sarcoidosis and A1AT Genomics &amp; Informatics Center</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Michael J. Becich, M.D. Ph.D.</td>
<td>26-Aug-08</td>
<td>30-Jun-18</td>
<td>80,925</td>
<td>43,700</td>
<td>SPORE in Skin Cancer</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Michael J. Becich, M.D. Ph.D.</td>
<td>1-Jul-14</td>
<td>30-Jun-15</td>
<td>523,767</td>
<td>262,772</td>
<td>University of Pittsburgh Clinical and Translational Science Institute</td>
</tr>
<tr>
<td>Patient-Centered Outcome Research Institute</td>
<td>Michael J. Becich, M.D. Ph.D.</td>
<td>14-Mar-14</td>
<td>30-Sep-15</td>
<td>127,584</td>
<td>51,034</td>
<td>A PaTH Towards a Learning Health System in the Mid-Atlantic Region</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Tanja Bekhuis, Ph.D.</td>
<td>1-Jul-12</td>
<td>30-Jun-15</td>
<td>133,411</td>
<td>68,707</td>
<td>Screening Nonrandomized Studies for Inclusion in Systematic Reviews of Evidence</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Richard Boyce, Ph.D.</td>
<td>30-Nov-16</td>
<td>30-Nov-16</td>
<td>177,333</td>
<td>14,113</td>
<td>Improving Medication Safety for Nursing Home Residents Prescribed Antidepressants</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Richard Boyce, Ph.D.</td>
<td>15-Feb-14</td>
<td>14-Feb-14</td>
<td>284,232</td>
<td>115,677</td>
<td>Addressing gaps in clinically useful evidence on drug-drug interactions</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Gregory F. Cooper, M.D., Ph.D.</td>
<td>15-Aug-14</td>
<td>31-May-18</td>
<td>49,729</td>
<td>26,854</td>
<td>Real-time detection of deviations in clinical care in ICU data streams</td>
</tr>
<tr>
<td>Project Title</td>
<td>Investigator</td>
<td>Start Date</td>
<td>End Date</td>
<td>Award Amount</td>
<td>Project Duration</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>------------</td>
<td>------------</td>
<td>---------------</td>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Gregory F. Cooper, M.D., Ph.D.</td>
<td>29-Sep-14</td>
<td>31-Aug-18</td>
<td>1,282,386</td>
<td>368,335</td>
<td>Center for Causal Modeling and Discovery of Biomedical</td>
</tr>
<tr>
<td>Centers for Disease Control and Prevention</td>
<td>Gerald Douglas, Ph.D.</td>
<td>1-Oct-12</td>
<td>28-Feb-15</td>
<td>104,766</td>
<td>38,134</td>
<td>Improving Quality of Care and Health Impact through Sustainable, Integrated, Inno</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Madhavi Ganapathiraju, Ph.D.</td>
<td>1-Aug-11</td>
<td>30-Apr-16</td>
<td>250,000</td>
<td>108,658</td>
<td>Discovery of Mental Health an Inflammation (MHAIN) Interactome</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Vanathi Gopalakrishnan, Ph.D.</td>
<td>1-May-06</td>
<td>31-Aug-16</td>
<td>46,250</td>
<td>23,819</td>
<td>Spore in Lung Cancer</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Vanathi Gopalakrishnan, Ph.D.</td>
<td>24-Sep-12</td>
<td>31-Jul-15</td>
<td>161,101</td>
<td>84,779</td>
<td>Transfer Learning Rule for Knowledge Based Biomarker Discovery and Predictive Bio</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Harry Hochheiser, Ph.D.</td>
<td>1-May-14</td>
<td>30-Apr-19</td>
<td>12,957</td>
<td>6,982</td>
<td>FaceBase 2: Management and Integration Hub</td>
</tr>
<tr>
<td>Univ of Utah (NIH)</td>
<td>Harry Hochheiser, Ph.D.</td>
<td>1-Oct-13</td>
<td>30-Sep-14</td>
<td>19,049</td>
<td>10,286</td>
<td>Interactive Search and Review of Clinical Records with Multilayered Semantic Annotation</td>
</tr>
<tr>
<td>University of Oregon (NIH)</td>
<td>Harry Hochheiser, Ph.D.</td>
<td>1-Feb-13</td>
<td>31-Jul-16</td>
<td>71,299</td>
<td>38,501</td>
<td>Semantic LAMHD: Linking diseases to model organism res</td>
</tr>
<tr>
<td>VA Research Foundation of Pittsburgh</td>
<td>Harry Hochheiser, Ph.D.</td>
<td>1-Dec-12</td>
<td>30-Jun-15</td>
<td>10,473</td>
<td>5,655</td>
<td>Quantifying Electronic Medical Records Usability to Imp</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Rebecca Jacobson, M.D., M.S.</td>
<td>1-Jul-87</td>
<td>30-Jun-17</td>
<td>1,089,010</td>
<td>68,418</td>
<td>Pittsburgh Biomedical Informatics Training Program</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Rebecca Jacobson, M.D., M.S.</td>
<td>6-May-14</td>
<td>30-Apr-15</td>
<td>540,337</td>
<td>108,038</td>
<td>Cancer Deep Phenotype Extraction from Electronic Media</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Xia Jiang, Ph.D.</td>
<td>1-Mar-12</td>
<td>28-Feb-15</td>
<td>135,226</td>
<td>63,621</td>
<td>Detecting Genome-Wide Epistasis with Efficient Bayesian Network Learning</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Xia Jiang, Ph.D.</td>
<td>1-Jun-14</td>
<td>31-May-18</td>
<td>323,290</td>
<td>93,471</td>
<td>A New Generation Clinical Decision Support System</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Xinghua Lu, M.D., Ph.D.</td>
<td>1-Sep-10</td>
<td>31-Aug-14</td>
<td>194,434</td>
<td>93,566</td>
<td>Statistical Methods for Integromics Discoveries</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Xinghua Lu, M.D., Ph.D.</td>
<td>1-Sep-11</td>
<td>31-Aug-15</td>
<td>192,561</td>
<td>99,169</td>
<td>Ontology-Driven Methods for Knowledge Acquisition and Knowledge Discovery</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Xinghua Lu, M.D., Ph.D.</td>
<td>1-Apr-15</td>
<td>31-Mar-19</td>
<td>227,221</td>
<td>116,330</td>
<td>Deciphering Cellular Signaling system by deep mining a comprehensive genomic compendium</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Shyam Visweswaran, M.D., Ph.D.</td>
<td>1-Jul-14</td>
<td>30-Jun-15</td>
<td>167,074</td>
<td>86,042</td>
<td>University of Pittsburgh Clinical and Translational Science Institute</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Michael Wagner, M.D., Ph.D.</td>
<td>18-Apr-12</td>
<td>31-Mar-16</td>
<td>467,807</td>
<td>116,310</td>
<td>Apollo: Increasing Access and Use of Epidemic Models Through the Development and Adoption of a Standard Ontology</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Michael Wagner, M.D., Ph.D.</td>
<td>1-Aug-13</td>
<td>30-Jun-16</td>
<td>461,712</td>
<td>128,311</td>
<td>Probabilistic Disease Surveillance</td>
</tr>
<tr>
<td>National Institutes of Health</td>
<td>Michael Wagner, M.D., Ph.D.</td>
<td>5-Aug-14</td>
<td>30-Apr-19</td>
<td>1,714,387</td>
<td>507,415</td>
<td>MIDAS - Modeling of Infectious Disease Agent Study Information</td>
</tr>
</tbody>
</table>
Entrepreneurial Activities

There are a number of entrepreneurial relationships that faculty in the Department have with outside enterprises, consisting of the following during FY2015:

**Michael J. Becich, M.D., Ph.D.**

Empire Genomics  
Scientific Advisory Board  
2012 - present

NinePoint Medical  
Scientific Advisory Board  
2011 - present

Omnyx, LLC  
Scientific Advisory Board and Sponsored Research Agreement  
2008 – present

**Tanja Bekhuis, PhD, MS, MLIS, AHIP**

*Invention Disclosure*

Dr. Bekhuis filed an invention disclosure with the University of Pittsburgh, May 26, 2015, for copyrightable software known as the *Evidence in Documents, Discovery, and Analysis (EDDA) System* (Invention Disclosure No. 03608).

The EDDA System can streamline reviews of medical reports to reduce costs and speed translation of evidence to providers. It can support information specialists, teams of scientists, solo investigators, and organizations. In turn, subsequent production of quality reviews will support clinical practice guideline developers and policymakers who can improve the quality of health care by knowing which regimens, drugs, or devices to recommend. By standardizing health care in line with evidence in the research literature, the financial impact could be considerable around the world.

**Richard D. Boyce, Ph.D.**

“An automated multifactorial patient-specific fall risk intervention designed for the nursing home setting that includes consideration of potential drug-drug interactions”,  
University of Pittsburgh Invention Disclosure Number: 03647. 2015

**Gregory F. Cooper, M.D., Ph.D.**

De-ID, Inc.  
Co-inventor of software licensed to De-ID by the University of Pittsburgh
Vanathi Gopalakrishnan, Ph.D.

Investigating the transfer of technology (intellectual property) in the form of biomarkers to industry – still in the planning or early stage.

Harry Hochheiser, Ph.D.

Invention disclosures filed with the University of Pittsburgh Innovation Institute:

- The Phenogrid phenotype-model similarity visualization (No. 03540, April 2015),
  - Co-innovators: C. Borromeo (Pitt), M. Haendel (OHSU), N. Washington (LBNL), Chris Mungall (LBNL)
  - Status: Copyrighted
  - Licensing discussions currently under negotiation with Appistry, Inc.
- An Interactive Tool for Natural Language Processing on Clinical Text (No. 03589, June 2015)
  - Co-Innovators: G. Trivedi (Pitt), P. Pham (Pitt), R. Hwa (Pitt), J. Wiebe (Pitt), W. Chapman (U. Utah)
  - Status: Copyrighted
- Learning Electronic Medical Record System (Filed July 2015)
  - Co-Innovators: S. Visweswaran, G. Cooper, A. King (all Pitt)
  - Status: Under review

Fu-Chiang (Rich) Tsui, Ph.D.

General Biodefense, LLC
Equity Ownership and Consultant
2004 – present

Invention Disclosure
- Surveysmith: An efficient web-based tool for intelligent survey design, deployment, collection, and summary, 3/19/2015
- IRIS: Intelligent Realtime Interactive SMS, 2/20/2015
- Software code for Automatic Adaptive Readmission Prediction System, 9/30/2014
- TOPAZ, 9/9/2013
- Automatic Adaptive Readmission Prediction System (AARPS), 8/21/2013

Patent Publication
System for hospital adaptive readmission prediction and management (SHARP), pub. No.: us 2015/0081328 a1, March 19, 2015
Shyam Visweswaran, M.D., Ph.D.

Applied for a patent titled “Computer-Implementable Algorithm for Biomarker Discovery Using Bipartite Networks” whose inventors are SK Bhavnani, KE Bassler and S Visweswaran.

Chief Medical Advisor, RoboClinics, Inc., Fernley, NV. 4/2014 - present

Michael Wagner, M.D., Ph.D.

General Biodefense, LLC
Equity Ownership and Consultant
2004 - present
TEACHING ACTIVITIES

Training Program / Graduate Student Education

The Pittsburgh Biomedical Informatics Training Program, established in 1987, prepares individuals for research and development careers emphasizing the application of modern information technology to health care, basic biological and clinical research, and education of health professionals. The program welcomes talented individuals who may be health professionals seeking formal training in informational and computational methods, as well as others who may be scientifically trained and seek to prepare themselves for careers emphasizing biomedical applications of information technology. The program currently has 31 students and offers a range of training experiences to accommodate the diverse backgrounds and aspirations of its students. Active participation in research and development projects is a key element of the training experience in Pittsburgh. The program is based in the Department of Biomedical Informatics, which works closely with all six University schools of the health sciences, and with the University of Pittsburgh Medical Center (UPMC), one of the most prestigious academic health systems in the nation. The University ranked seventh last year in research funding from the National Institutes of Health (NIH).

The mission of the Training Program is to provide our students with a world-class education that prepares them to become outstanding leaders in biomedical informatics research, education, and practice.

Degree Programs

Almost all students in the program undertake a formal course of study leading to a master’s or doctoral degree or to a biomedical informatics certificate. Admission to the program is highly competitive. Students can choose between two programs as the home for their studies:

The Biomedical Informatics Program

This program offers both master’s and doctoral degrees. Most students follow a general course of study in biomedical informatics; then, elect a specialized concentration through elective credits in one of the following areas: bioinformatics, dental informatics, health services research, or infectious disease and public health (biosurveillance) informatics. Training is also available for those interested in the components of medical informatics and bioinformatics that relate to clinical oncology and cancer research (pathology informatics).

Trainees may seek a 15-credit certificate in lieu of an academic degree. The biomedical informatics certificate can augment professional training in fields related to informatics and/or fulfill educational needs associated with a professional position.
The degree and certificate programs in biomedical informatics are administered through the School of Medicine, but welcome trainees from all health professions.

*The Intelligent Systems Program*

Also offered are master’s and doctoral degrees in intelligent systems with specialization in biomedical informatics. The degree program in intelligent systems enjoys a long-standing relationship with the Biomedical Informatics Training Program and may be the degree option of choice for students with specific interests in artificial intelligence applications to health care.

*MD/PhD Program*

Through the University of Pittsburgh School of Medicine, students accepted into the MD/PhD Program may choose to complete their PhD training in the Biomedical Informatics program. Completion of the PhD degree in the Biomedical Informatics program will typically take four years. Program requirements for the PhD and the MD/PhD are identical except that the MD/PhD students receive credit for medical school coursework completed prior to entering the Biomedical Informatics program.

*Summer Short-Term Trainee Program*

The Pittsburgh Biomedical Informatics Training Program offers a short-term training opportunities during the summer to appropriately qualified undergraduate, graduate, or medical students from underrepresented racial or ethnic groups, have a disability or from disadvantaged backgrounds. The goals of this training opportunity are to permit individuals who have a nascent interest in informatics to gain research experience and to encourage minorities to consider informatics as a career field. We envision that a student will work closely with a single faculty member on a project that can be completed during the summer. The nature of this project would depend on the skill set and interest of the student. Also, the student would have the opportunity to meet one-on-one with other faculty members to learn about their research areas.

*PhD, Masters and Certificate Graduate Students*

**Jhon Jairo Camacho Sanchez, B.A., M.S.** B.S. (2003, Electronic Engineering) Pontificia University, Colombia. M.S. (2012, Biomedical Informatics) University of Pittsburgh. His research involves informatics needs assessment of Columbian low birth-weight infant care programs. He is currently on a two-year leave of absence and will return in Fall 2015. [Biomedical Informatics Doctoral Program]

**Sergio Castro Diaz, M.D.** (2009, Medicine) Pontificia University, Colombia. His research interests include applied medical informatics, automatic relapse prediction models for mental health and cognitive informatics for health care. [Biomedical Informatics Certificate Program]
Rafael Ceschin, B.S., M.S. B.S. (2009, Microbiology) University of Pittsburgh, M.S. (2014, Biomedical Informatics) University of Pittsburgh. His research involves analysis of neuroimaging data to study neonatal development following preterm birth. [Biomedical Informatics Doctoral Program]

Lujia Chen, B.S., M.S. B.S. (2008, Biological Technology) University of Science & Technology, Beijing, China, M.S. (2012, Biomedical Informatics) University of Pittsburgh. Her research involves the collection of lipidomic data and gene expression data. [Biomedical Informatics Doctoral Program]


Michael Ding, B.A. (2012, Molecular and Cellular Biology) Harvard University. His research interest include generating predictive models for cancer drug sensitivity using large pharmacogenomics datasets. [Biomedical Informatics Doctoral Student]

Amie Draper, B.S., M.S. B.S. (2013 Biomedical Science) Ohio State, M.S. (2015, Biomedical Informatics) University of Pittsburgh. Her research interest includes utilizing laboratory information to aid in predicting patient risk of hospital readmission for pediatric seizure patients and adult heart failure patients. [Biomedical Informatics Doctoral Program]


Arielle Fisher, B.S., M.S. (2013 Biology) Bucknell University and M.S. (2015 Biomedical Informatics) University of Pittsburgh. Her research interest includes using contextual inquiry to document the dispensary workflow at the Birmingham Free Clinic. This qualitative, social method, in addition to a formal interview of three pharmacists working in this setting, will be used to identify workflow inefficiencies and other challenges encountered by pharmacists working at the clinic. A series of informatics interventions will be proposed and defined; a cost benefit analysis will be done on these proposed interventions. [Biomedical Informatics Doctoral Program]

John Frazier, B.S., M.S.P.H., D.M.D, M.S. BS (1997, Mathematics) University of South Carolina, MSPH (2002 Biostatistics) University of South Carolina, DMD (2006 Dental Medicine) Medical University of South Carolina, and M.S. (2014 Biomedical Informatics) University of Pittsburgh. His research interests includes computer assisted diagnosis of microorganisms in histopathological whole slide images. [Biomedical Informatics Doctoral Program]
Adam Handen, B.S., M.S. (2013, Bioinformatics) Rochester Institute for Technology. His current interests are interactomics, genetics, and disordered protein function as well as data mining, data curation/database development and artificial intelligence. [Biomedical Informatics Doctoral Program]

Andres Hernandez Camacho, B.S., M.S. B.S. (2009 Electrical Engineering) Pontificia University, Columbia, M.S. (2014, Biomedical Informatics) University of Pittsburgh. His research interests include global health informatics [Biomedical Informatics Masters Program]

JoAnna Hillman, B.S., M.PH. B.S. (2002 Human Science and Services) University of Rhode Island, and MPH (2006 Epidemiology and Biostatistics) Boston University. Her research interest includes modeling return on investment and evaluating impact of informatics solutions on process and workflow in resource-constrained hospital laboratories. [Biomedical Informatics Doctoral Program]


Andrew King, B.S., M.S. B.S. (2013 Computer Science) and M.S. (2015, Biomedical Informatics) University of Pittsburgh. His research interests include preliminary modeling, evaluation and usability study of a Learning EMR. [Biomedical Informatics Doctoral Program]


Yuzhe Brian Liu, B.S., M.S. B.S. (2010, Computer Science), M.S. (2011, Computer Science) Cornell University. His research interests include image marker discovery in cardiac MRI. [Biomedical Informatics Doctoral Student]

Kevin McDade, B.S., M.S. B.S. (2000, Microbiology) University of Pittsburgh. M.S. (2007, Biology) Chatham College. His research focus is the reformation of
transcriptome-proteome correlation using probe level quality. [Biomedical Informatics Doctoral Program]


Jose Posada Aguilar, B.S., M.S. B.S. (2007, Electrical Engineering), M.S. (2009, Engineering) Universidad del Norte, Colombia. His research interests include readmission prediction for psychiatry patients using Natural Language Processing. [Biomedical Informatics Doctoral Program]


Adam Roth, B.S., M.S. B.S. (2012, Biology), M.S. (2015, Biomedical Informatics) University of Pittsburgh. His research interests involve text mining for validating protein interactions. [Biomedical Informatics Master’s Program]

Victor Ruiz Herrera, B.S., M.S. BS (2011 Electrical Engineering) Pontificia University Javeriana, Colombia, M.S. (2014, Biomedical Informatics) University of Pittsburgh. His research involves prediction of readmission risks. [Biomedical Informatics Doctoral Program]

Lucas Santana dos Santos, B.Sc, M.S. B.Sc. (2008, Biological Science) Univeridade Federal, Brazil, M.S. (2012, Biomedical Informatics) University of Pittsburgh. His research interests include predicting oxidative phosphorylation levels of breast cancer cell lines from gene expression data. [Biomedical Informatics Doctoral Program]


include detection of gene-gene interaction to identify the genetic basis of phenotypic traits. [Intelligent Systems Program Doctoral Program/Biomedical Informatics Track]

Ye Ye, B.S., M.S., M.S.P.H. B.S. (2006, Preventive Medicine) and M.S. (2009, Biostatistics & Epidemiology) Peking University, China, and M.S.P.H. (2011, Public Health Informatics) Emory University. Her research involves evaluation of Bayesian case detection system and application of artificial intelligence technologies and statistics methods to topics in Public Health Informatics and Medical Informatics. [Intelligent Systems Program Doctoral Program/Biomedical Informatics Track]

Jonathan Young, B.A., M.S., M.D. BA (2002 Biology) University of Virginia, MS (2005 Physiology and Biophysics) Georgetown University, and MD (2010 Medicine) Eastern Virginia Medical School. His research interests include using deep learning to explore gene expression data. [Biomedical Informatics Master’s Program]

She Zhang, B.S. (2014, Biopharmaceuticals) Nanjing University, China. His research interest include simulation of dynamics computational tool development. [Biomedical Informatics Master’s Program]

Certificate Awardees

None for FY15

Departing Fellows

Zachary Landis Lewis, B.A., M.L.I.S., M.S., PhD received his Ph.D. in Biomedical Informatics (December 2014). He was a postdoctoral scholar in the Department of Biomedical Informatics, University of Pittsburgh from October 2014 to July 2015. He accepted a faculty position at Department of Learning Health Sciences, University of Michigan beginning August 1, 2015.

Andres Hernandez Camacho, B.S., M.S. received his M.S. in Biomedical Informatics (December 2014). He is currently a researcher at the Center for Biomedical Informatics and Computational Biology in Manizales Colombia.

Adam Roth, B.S., M.S. received his M.S. in Biomedical Informatics (April 2015). He is currently an enterprise architect at the US Department of Veterans Affairs.

Richard Wilson, B.A., M.S. received his M.S. in Biomedical Informatics (April 2010) University of Pittsburgh. After being on a two-year leave of absence, Lt. Col. Wilson was unable to return to the program due to his military duties.

Jonathan Young, B.A., M.S., M.D. continues to work on his M.S. in Biomedical Informatics (expected December 2015). He has been accepted into the doctoral program...
in the Intelligent Systems Program in the School of Arts and Sciences, University of Pittsburgh.

**She Zhang, B.S.** transferred to the doctoral program in Computational and Systems Biology, University of Pittsburgh.

### Graduate Student Researchers (from other programs)

**Jeya Balasubramanian, B.Tech. (Bachelor of Technology), M.S.** B.Tech. (2009, Bioinformatics) S.R.M. (Sri Ramaswamy Memorial) University, India, and M.S. (2010, Computational Biology) Carnegie Mellon University. His research involves design, evaluation, and interpretation of ensemble methods in data mining for biomarker discovery. He is presently working with Dr. Vanathi Gopalakrishnan.

**Diyang Xue, B.S., M.S.** B.S. (2006, Preventative Medicine) Shan Dong University, China and M.S. (2009, Health Statistics and Epidemiology) Peking University Health Science Center, China. His research interests include feature selection methods, study theory and compare their performance in high-dimensional Genome data and is currently working with Dr. Xia Jiang.
Post-Docs

John Aronis, B.A., M.S., Ph.D.  
B.A. (1979 Mathematics) State College of New York; M.S. (1981 Mathematics) Syracuse University; Ph.D. (1993 Intelligent Systems) University of Pittsburgh. His research interests include application to data mining to the detection of anomalous cases and contagious outbreaks. Dr. Aronis works with Dr. Gregory Cooper.

Viji Avali, Ph.D.  
B.S. (1987 Electrical Engineering) Madurai University, India; M.S. (1992 Process Monitoring and Control) University of Houston; M.E. (1996 Electrical and Computer Engineering) University of South Carolina; and Ph.D. (2010 Computer Science and Engineering) University of South Carolina. Her interests include Building a classifier incorporating domain knowledge in Bayesian Logistic Regression to use it in biomedical data analysis. Dr. Avali is works with Dr. Vanathi Gopalakrishnan and Dr. Madhavi Ganapathiraju.

Chunhui Cai, Ph.D.  
B.S. (2005 Physics) Hong Kong Baptist University, Ph.D (2010 Physics) Hong Kong Baptist University. Dr. Cai works with Dr. Xinghua Lu.

Virginia Dato, M.S., MPH  

Songjian Lu, Ph.D.  
B.S., (1988 Mathematics), Guangxi University, China; M.S., (1996 Mathematics Biology), Xi’an Jiaotong University, China; M.S., (2001 Applied Mathematics), University of Houston – Clear Lake, Houston, TX; M.S., (2003 Computer Science), University of Houston – Clear Lake; Ph.D., (2009 Computer Science), Texas A&M University. Dr. Lu works with Dr. Xinghua Lu on translational bioinformatics.
Departing Post-Docs

Zachary Landis Lewis, M.L.I.S., Ph.D.
Has accepted an Assistant Professor position working with Dr. Charles Friedman in the Department of Learning Health Sciences, University of Michigan. Ann Arbor, MI.

Binghaung Cai, Ph.D.
B.S. Electronic Information Engineering (2004), Department of Electronic Engineering, Shantou University, China; M.S. Signal and Information Processing (2007), Department of Electronic Engineering, Shantou University, China, Ph.D. Communication and Information Systems (2010), School of Information Science and Technology, Sun Yat-sen University, China. Dr. Cai worked for Dr. Xia Jiang.
## Faculty Data

### Faculty Roster

#### Primary Faculty

**Professor:**
- Michael J. Becich, M.D., Ph.D.
- Gregory F. Cooper, M.D., Ph.D.
- Rebecca S. Jacobson, M.D., M.S.I.S.

**Associate Professor:**
- Vanathi Gopalakrishnan, Ph.D
- Xinghua Lu, M.D., Ph.D.
- Michael M. Wagner, M.D., Ph.D.

**Assistant Professor:**
- Tanja Bekhuis, Ph.D., M.S., M.L.I.S., A.H.I.P.
- Richard Boyce, Ph.D.
- Gerald Douglas, Ph.D.
- Madhavi Ganapathiraju, Ph.D.
- Harry Hochheiser, Ph.D.
- Xia Jiang, Ph.D.
- Shyam Visweswaran, M.D., Ph.D.

**Research Assistant Professor**
- Tsui, Fu-Chiang, Ph.D., M.S.

#### Secondary Appointment Faculty

**Professor:**
- Barton Branstetter, M.D., *Radiology*
- Rajiv Dhir, M.D., *Pathology*
- Daniel Martich, M.D., *Critical Care Medicine*
- Anil Parwani, M.D., Ph.D., *Pathology*
- Bruce Rollman, MD, M.P.H, *Internal Medicine*

**Associate Professor:**
- Takis Benos, PhD, *Computational & Systems Biology*
- Liron Pantanowitz, M.D., *Pathology*
Clinical Associate Professor:

Anthony Fiorillo, M.D., UPMC ISD
Michael Dunn, M.D., Medicine

Assistant Professor:

Julia Driessen, Ph.D., GSPH
Bruce Lee, M.D., M.B.A., Medicine
Ervin Sejdic, Ph.D., Electrical Engineering
Heiko Spallek, D.M.D., Ph.D., Dental Medicine

Adjunct Faculty

Associate Professor:

Titus Schleyer, D.M.D, Ph.D., Indiana University

Assistant Professor:

Christa Bartos, Ph.D., UPMC IS
Anne Ben-Smith, Ph.D., Malawi
William Hogan, M.D., M.S., University of Arkansas
Claudia Mello-Thoms, Ph.D., University of Australia
Valerie Monaco, Ph.D., M.H.C.I., UPMC Health Plan
Rema Padman, Ph.D., Carnegie Mellon University
Brandon Reines, D.V.M., St. Francis Animal Hospital
Adam Rothschild, M.D., Private Practice
Bertha Simwaka, Ph.D., Malawi
Thankam Thyvalikakath, M.D.S., Ph.D., Indiana University
Faculty Biographies

Michael J. Becich, M.D., Ph.D.
Chairman, Department of Biomedical Informatics
Professor of Biomedical Informatics, Pathology, Information Sciences, and Clinical and Translational Science
Associate Director for Cancer Informatics, University of Pittsburgh Cancer Institute
Associate Director, Clinical and Translational Science Institute, University of Pittsburgh
Co-Director, Pathology Informatics 2010

Michael J. Becich, M.D., Ph.D., is the Chairman of the Department of Biomedical Informatics (www.dbmi.pitt.edu) at the University of Pittsburgh School of Medicine. He is a Professor of Biomedical Informatics, Pathology, Information Sciences/Telecommunications, and Clinical and Translational Science. He is currently director for Biomedical Informatics Core(BIC, http://www.ctsi.pitt.edu/ ). BIC was formed as a result of the successful funding of the Clinical and Translational Science Institute (www.ctsi.pitt.edu) at the University of Pittsburgh as part of the NIH Roadmap Clinical and Translational Science Awards.

Dr. Becich obtained his M.D. and Ph.D. in Experimental Pathology from Northwestern University and served as a staff anatomic pathologist for Washington University (St. Louis) after completing his pathology residency.

His laboratories are funded by grants from the National Institutes of Health, National Center for Advancing Translational Science, National Heart, Lung and Blood Institute, National Library of Medicine, National Cancer Institute, CDC and the Patient Centered Outcomes Research Institute as well as multiple corporate sponsored research programs. He is a member of 14 professional societies and has contributed to over 150 papers as well as several on-line presentations. He is a member of the American College of Medical Informatics.

While at the University of Pittsburgh, Dr. Becich founded the nation's first Pathology Informatics fellowship program and is founder of both the Association for Pathology Informatics (see http://www.pathologyinformatics.org) and Advancing Practice, Instruction and Innovation through Informatics (APIII) which in 2010 became Pathology Informatics 201X (see http://www.pathologyinformatics.com ). Pathology Informatics is a national meeting transforming translational research through training and continuing medical education in pathology informatics and is currently in its 18th year as the academic home for this transformative program.

Study Sections and Advisory Committee Memberships:

American Institute of Biological Sciences
Ad Hoc Member Reviewer
1999 - present
National Cancer Institute (NCI) Cancer Center Support Grants (CCSG), Parent Committee
Ad Hoc Reviewer
2002 - present

NCI Cancer Center Support Grants Program (CCSG)
Ad Hoc Reviewer
2001 – present

NCI Molecular Epidemiology
Transitioning Member
2000 – present

NCI Program Project Study Section
Ad Hoc Member Reviewer
1998 - present

NCI Sponsored Programs of Research Excellence (SPORE) Grants Program
Ad Hoc Reviewer
2001 - present

NCI, STTR & SBIR Program - Informatics Technology and Genomics
Ad Hoc Reviewer
2001 - present

Roswell Park Cancer Institute
External Consultant
2006 - present

Cancer Center Consulting – External Advisory Boards:

MD Anderson Cancer Center, Electronic Medical Records Program
2006 - present

Massey Cancer Center, Virginia Commonwealth University
2010 - present

Moffitt Cancer Center
2009 - present

National Functional Genomics Center (NFGC)
2005 - present

University of Buffalo
2010 - present
University of Colorado Cancer Center
2007 - present

University of Medicine and Dentistry of New Jersey Cancer Center
2007 – present

CTSA - External Advisory Boards:

Duke University Translational Medicine Institute
2011 - present

Medical College of Wisconsin Clinical and Translational Science Institute
2010 - present

Northwestern University Clinical and Translation Science Institute
Also Serve as External Advisory Board Chair
2009 - present

University of Arkansas for Clinical and Translational Research
2010 - present

University of California Irvine Institute for Clinical and Translational Science
2011 - present

University of California Davis Clinical and Translational Science Center
2009 - present

University of California Los Angeles Clinical and Translational Science Center
Also Serve as Consultant
2010 - present

University of Chicago Institute for Translational Medicine
2011 - present

University of Indiana Clinical and Translational Sciences Institute
2011 - present

University of Kentucky Center for Clinical and Translational Science and Biomedical Informatics Program
Also Serve as External Advisory Board Chair
2009 - present

University of Michigan Institute of Clinical and Translational Sciences
2010 - present

University of Wisconsin, Marshfield Clinic, Wisconsin Genome Institute
2010 - present
Washington University (St. Louis) Institute of Clinical and Translational Sciences
2007 - present

Editorships, Honors, Awards, and Major Lectureships:

Editorial Boards:

Journal of the American Medical Informatics Association
2010 - present

Journal of Pathology Informatics
2010 - present

Clinical Prostate Cancer
2003 - present

Clinical Proteomics
2003 - present

Archives of Pathology & Laboratory Medicine
2000 - present

Advances in Anatomic Pathology
1994 - present

Honors and Awards:

Elected Fellow, American College of Medical Informatics (ACMI) - 2006
Distinguished Visiting Professor, Mount Sinai Hospital, 38th Harold G Pritzker
Memorial Lecture, Toronto, Ontario, Canada - 2002
Visiting Professor of Pathology, Emory University School of Medicine, Atlanta, GA - 2001
Visiting Professor of Pathology, Montefiore Medical Center, Bronx, NY - 2001
Visiting Professor of Pathology, Ohio State University, Columbus, OH - 2001
Visiting Professor of Pathology, Mayo Clinic, Rochester, MN - 2001
Quest Distinguished Visiting Professor of Pathology, Harvard University, Boston, MA - 2000
Visiting Professor of Pathology, University of Pennsylvania, Philadelphia, PA - 2000
Distinguished Visiting Professorship of Pathology, The Leo Kaplan, MD Lectureship,
UCLA, Cedar Sinai Medical Center, Los Angeles, CA - 2000
Distinguished Visiting Professorship of Pathology, 17th Annual Francis P. Boland, MD
Memorial Surgical Symposium, University of Scranton, Scranton, PA - 2000
Visiting Professor of Pathology, MD Anderson Medical Center, Houston, TX - 1999
Intel Internet Health Hero - 1999
Guest Distinguished Visiting Professorship of Pathology, Johns Hopkins University,
Baltimore, MD - 1999
Pathology Teaching Award for Anatomic Pathology, University of Pittsburgh - 1992
National Cancer Institute (NCI) Cancer Biology Fellowship Program - 1986 - 89
Medical Student Research Fellowship Award, University of Nebraska - 1983
Graduate Research Assistant Scholarship - 1981 - 83
National Student Forum Participant - 1980 - 83
Graduate Fellowship, Dept. of Pathology, Northwestern University Medical School - 1979 - 82
Summer Fellow, Dept. of Pathology, Northwestern University Medical School - 1976

**Tanja Bekhuis, Ph.D., M.S., M.L.I.S., A.H.I.P.**
Assistant Professor of Biomedical Informatics
Biomedical Informatics Training Program Core Faculty
Assistant Professor of Dental Public Health
Assistant Professor, School of Information Sciences

**Tanja Bekhuis, PhD, MS, MLIS, AHIP**, is an Assistant Professor in the Department of Biomedical Informatics (DBMI) at the University of Pittsburgh School of Medicine. She has secondary appointments in the Department of Dental Public Health, School of Dental Medicine, and the School of Information Sciences. She attained her PhD in quantitative psychology from the highly regarded Thurstone Psychometric Laboratory at the University of North Carolina-Chapel Hill. She also has master's degrees in educational research from the University of Miami, and library and information science from the University of Pittsburgh. Dr. Bekhuis completed a National Library of Medicine (NLM) and National Institute of Dental and Craniofacial Research (NIDCR)-funded postdoctoral scholar position in the Center for Dental Informatics at the University of Pittsburgh School of Dental Medicine. During that time, she went to Lister Hill to participate in the summer internship program for NLM fellows. Lister Hill is the research center for NLM. She subsequently accepted a position as a postdoctoral associate in DBMI to continue her research in machine learning and natural language processing of medical text. At DBMI, Dr. Bekhuis developed a successful NIH K99/R00 career development grant for which she is the principal investigator (PI). NLM funds her K99/R00 award. Additionally, she was a co-investigator on an R21 grant funded by NIDCR (PI: Dr. Heiko Spallek).

Prior to retraining in biomedical informatics, Dr. Bekhuis received several awards for innovative research, as well as fellowships from the National Institute of Mental Health and the National Institute on Aging. She worked as a statistical consultant for both the University of North Carolina Medical School and the Penn State Statistical Department. Dr. Bekhuis has published extensively in peer-reviewed journals, and written numerous essays for reference books on research methods and measurement in education, psychology, and biomedicine. Additionally, she has evaluated many systematic reviews for the Database of Abstracts of Reviews of Effects and the Journal of Evidence-based Dental Practice (JEBDP). Her most recent publications appear in *Artificial Intelligence in Medicine, Journal of the Medical Library Association, BMC Oral Health, PLoS One, APA PsycEXTRA, APA PsycTESTS, Proceedings of the 14th European Association for Health Information and Libraries*, and the *Journal of Dental Education*. She is a member of the Phi Kappa Phi Honor Society, as well as Beta Phi Mu, the International Library
and Information Studies Honor Society. Dr. Bekhuis completed a three-year position on the editorial board of the *Journal of the Medical Library Association*. She was admitted to the Academy of Health Information Professionals (AHIP) at the senior level.

**Study Sections and Advisory Committee Memberships:**

**Advisory Committee Memberships**

Dr. Bekhuis is Chair of the Eugene Garfield Research Fellowship Committee for the Medical Library Association. She is also on the Program Committee for the 2015 International Workshop on Biomedical and Health Informatics in conjunction with the IEEE International Conference on Bioinformatics and Biomedicine.

**Students Participating in Research:**

**Zachary Landis-Lewis**, PhD, MLIS, post-doctoral scholar, serving as primary mentor (September 2014-July 2015). Obtained a position as an Assistant Professor, tenure track, University of Michigan, August 2015.

**John Frazier**, BS, MSPH, MS (obtained 2014), DMD, a doctoral student in DBMI who is an oral pathologist. They developed a gold standard dataset of studies on prognostic biomarkers of oral squamous cell carcinoma in preparation for information retrieval studies.

**Other Mentoring:**

**Ashleigh Faith**, BA, MA, PhD candidate (expected 2017), School of Information Sciences, University of Pittsburgh. Projects: Enriching the EDDA Study Designs and Publications terminology; Topics and Study Designs in Health Technology Assessment.

**Editorships, Honors, Awards, and Major Lectureships:**

Editorial Board Member. *Journal of the Medical Library Association (JMLA)*, 2008-11.

Senior Member. Academy of Health Information Professionals, 2013 to present.

**Richard D. Boyce, Ph.D.**

*Assistant Professor of Biomedical Informatics*

*Biomedical Informatics Training Program Core Faculty*

The use of informatics to support safe and effective medication therapy has been Dr. Boyce’s primary interest since the early stages of his research career. He now has more than a decade of training and experience in pharmaco-informatics research. For his PhD dissertation, he worked closely with pharmacy, pharmaceutics, and informatics experts to develop a novel method for predicting potential pharmacokinetic drug-drug interactions. During his postdoctoral training, he began to explore novel methods for monitoring
nursing home patients exposed to drug-drug interactions. His training since becoming faculty includes three years of clinical research and pharmacoepidemiology training (as a scholar in the University of Pittsburgh Comparative Effectiveness Research program). He is currently the Principal Investigator of an R01 project that is studying a novel approach to addressing gaps in drug-drug interaction evidence, and a K01 project that is investigating how to actively monitor nursing home patients for fall risk while they are exposed to drug-drug interactions. His publication record includes several peer reviewed research at the intersection of medication safety, knowledge representation, and decision support. He co-leads the Knowledge Base Workgroup in the Observational Health Data Sciences and Informatics consortium.

**Study Sections and Advisory Committee Memberships:**

Dr. Boyce was a reviewer for the University of Pittsburgh Office of Research Competitive Medical Research Fund, a reviewer for ZonMw (the Netherlands organisation for health research and development Rational Pharmacotherapy program). He is currently an invited member of the Evidence Appraisal Working Group for the AHRQ Drug-Drug Interaction Clinical Decision Support Conference Series (1R13HS021826-01).

**Students Participating in Research:**

Mrs. Katrina Romagnoli – PhD student in the Department of Biomedical Informatics  
Mr Samuel Rosko – incoming PhD student in the Pitt Information Science program  
Mr. Eric Chou – Undergraduate in the Pitt Bioinformatics program  
Mr. Andres Hernandez-Camacho – Masters student in the Department of Biomedical Informatics (completed December 2014)

**Major Lectureships:**

3. Addressing Gaps in Clinically Useful Evidence on Drug-drug Interactions – A Knowledge-Based Foundation. Invited talk at the University of Washington Department of Biomedical Informatics and Medical Education. August 11th, 2015. Seattle, WA, USA.
Gregory F. Cooper, M.D., Ph.D.
Vice Chairman, Department of Biomedical Informatics
Vice Chairman, Department of Biomedical Informatics and Professor of Biomedical Informatics, Intelligent Systems, and Computational and Systems Biology.

Gregory F. Cooper, M.D., Ph.D., is an Associate Professor of Biomedical Informatics and Vice Chair of the Department of Biomedical Informatics. He holds joint appointments in Intelligent Systems and in Computational Biology at the University of Pittsburgh. He received a Ph.D. degree in Medical Information Sciences and an M.D. from Stanford University. He was a Senior Research Scientist in biomedical informatics at Stanford University before joining the University of Pittsburgh in 1990. His primary research interests involve the application of decision theory, probability theory, Bayesian statistics, and artificial intelligence to biomedical informatics research problems, with a current focus on causal modeling and discovery from biomedical data, computer-aided medical diagnosis and prediction, machine-learning approaches to improving patient safety, and biosurveillance of disease outbreaks.

Study Sections and Advisory Committee Memberships:

Standing Study Section of the National Library of Medicine (NLM)
Member (Chair from November 2011 through June 2013)
2009 - 2013

External Advisory Board of the NYU Clinical and Translational Science Institute
Member
2009 - present

Students Participating in Research:

Yuriy Sverchkov, Ph.D. student, Intelligent Systems Program, University of Pittsburgh
2009 - 2014 (Ph.D. completed in summer, 2014)
Title of doctoral dissertation: Detection and Explanation of Statistical Differences Across a Pair of Groups

Andrew King, Ph.D. student, Biomedical Informatics Training Program, University of Pittsburgh.
Research topic: A learning electronic medical record.
2014 - present

Editorships, Honors, Awards, and Major Lectureships:

Artificial Intelligence in Medicine
Editorial Board
1990 - present

Journal of Biomedical Informatics
Editorial Board
2002 - present

Awards

1984 Martin Epstein award for best paper in the student paper competition at the Ninth Annual Symposium on Computer Applications in Medical Care

2005 Distinguished paper award, Symposium of the American Medical Informatics Association.

2010 Homer R. Warner clinical-informatics research paper award at the Annual Symposium of the American Medical Informatics Association (AMIA). First author: Milos Hauskrecht. Other co-authors: Michal Valko, Iyad Batal, Gilles Clermont, Shyam Visweswaran, and Gregory Cooper.

2011 Marco Ramoni distinguished paper award at the 2011 AMIA Summit on Translational Bioinformatics. First author: Wei Wei. Other co-authors: Shyam Visweswaran and Gregory Cooper.

Gerald P. Douglas Ph.D.
Assistant Professor of Biomedical Informatics
Assistant Professor of Health Policy & Management
Director, Center for Health Informatics for the Underserved

Gerald P. Douglas, Ph.D., is an Assistant professor of Biomedical Informatics in the Department of Biomedical Informatics. He received a B.Sc. (Honors) in Computer Science from the University of Victoria, and a Master’s degree in Information Science and a Ph.D. in Biomedical Informatics from the University of Pittsburgh. In 2000 Dr. Douglas founded a non-profit organization based in Malawi, Africa working in collaboration with the Malawi Ministry of Health to improve healthcare delivery through medical informatics. In 2011 he received the University of Victoria Faculty of Engineering Distinguished Alumnus Award in recognition of his work in global health informatics in Malawi.

Study Sections and Advisory Committee Memberships:

Inveneo Inc.
Advisory Council Member
2011 - present

Professional Organization Membership:

American Medical Informatics Association (AMIA)
Member
1998 - present

**Students Participating in Research:**

**JoAnna Hillman, PhD Student, DBMI, University of Pittsburgh**
Improving Situational Awareness in a High-Burden Labor and Delivery Suite in Malawi
2013 – 2015

**Arielle Fisher, PhD Student, DBMI, University of Pittsburgh**
Understanding the Dispensary Workflow at the Birmingham Free Clinic: Responding to Challenges with Informatics Interventions
2013 – present

**Lia Petrose, BPhil Student, University Honors College, University of Pittsburgh**
Assessing Barriers to Laboratory Testing at Kamuzu Central Hospital in Malawi:
Challenges Encountered while Testing the Study Protocol
2014 – present

**Timothy Mtonga, Certificate Student, DBMI, University of Pittsburgh**
Improving Antiretroviral supply chain in Malawi using eDispensing
2015 – present

**Jhon Camacho, PhD Student, DBMI, University of Pittsburgh**
Addressing Medication Management in a Low-resource Hospital Setting
2015 – present

**Editorships, Honors, Awards, and Major Lectureships:**

Technology, Entertainment and Design (TED) Fellow
2009 - present

**Madhavi Ganapathiraju, Ph.D.**
*Assistant Professor of Biomedical Informatics and Intelligent Systems*

Madhavi Ganapathiraju, Ph.D., is an Assistant Professor in the Department of Biomedical Informatics, and Intelligent Systems Program. She received a Masters in Engineering degree in Electrical and Communications Engineering from the Indian Institute of Science and a Ph.D. in Language and Information Technologies from the School of Computer Science, Carnegie Mellon University. The focus of her Ph.D. thesis was on the application of signal processing and language processing methods to the study of protein and proteome sequences and development of novel algorithms that have since been used widely in the analysis of biological sequences. Her research interests include machine learning and development of multidisciplinary approaches to computational biology and translational bioinformatics, specifically in protein-protein interaction prediction, membrane protein structure prediction and whole-genome sequence pattern mining.
Other appointments include:

Faculty, Joint Carnegie Mellon University – University of Pittsburgh Ph.D. program in Computational Biology

Member, Molecular and Cellular Cancer Biology Program, University of Pittsburgh Cancer Institute

Faculty, Language Technologies Institute, Carnegie Mellon University

Advisory Board Member, Biotechnology Innovation and Computation Program, Carnegie Mellon University

Study Sections and Advisory Committee Memberships:

2013  NIMH Special Emphasis Panel ZMH1 ERB-L(04) (for BRAINS awards)
2013  NIMH Special Emphasis Panel ZMH1 ERB-M(07) (for EUREKA awards)
2014  NIMH Special Emphasis Panel ZMH1-ERB-M(06) (for BRAINS awards)
2014  NIMH Special Emphasis Panel ZMH1 ERB-R(02) (for EUREKA awards)

Editorships, Honors, Awards, and Major Lectureships:

International Symposium on Nanoelectronic and Information Systems, Indore, India
Track Chair *Cyber Physical Systems and Social Networks Track* (2015)


Great Lakes Bioinformatics Conference (Ann Arbor, Michigan) Scientific Program Committee, 2013

Students Participating in Research:

**Graduate Students**
Adam Handen, B.S., M.S., Advisor
Haohan Wang, B.S., Advisor
Aman Gupta, B.Tech., Advisor
Sandeep Subramanian, B.Tech, Advisor
Undergraduate Students - Research Supervision:

2014 Summer  Helen Li   Undergrad  University of California at Santa Cruz

High School Students - Research Supervision:

2014 Summer  Thomas Nash   High School  CoSBBI Summer Research

Vanathi Gopalakrishnan, Ph.D.
Associate Professor of Biomedical Informatics,
Associate Professor, Intelligent Systems,
Associate Professor, Computational Biology

Vanathi Gopalakrishnan, Ph.D., is an Associate Professor of Biomedical Informatics, Intelligent Systems, and Computational Biology. She received her degree in Computer Science from the University of Pittsburgh in 1999. Dr. Gopalakrishnan is interested in the design and development of computational methods for solving clinically relevant biological problems. Her research encompasses the development and application of symbolic, probabilistic and hybrid machine learning techniques to the mining of structural and genomic databases in order to learn useful, robust models and associations. Her current collaborative projects include modeling of protein sequence-structure-function relationships and identification of disease-specific proteomic biomarkers for neurodegenerative diseases, lung, breast and esophageal cancers and pediatric heart disease. Dr. Gopalakrishnan is a member of the International Society for Computational Biology, the American Association for Artificial Intelligence, the Association for Computing Machinery (ACM), and the American Association for the Advancement of Science. She is the recipient of major grants from the National Institutes of Health including the NIGMS, NLM, and NCI.

Study Sections and Advisory Committee Memberships:

NIH/NCRR SBIR Grants and Contract Review
Study Section Member

NSF
Review Panelist

2012       NSF BIG Data Review Panelist
2013       NIH P41 Special Review Panelist
2013       NIH P51 Review Panelist
2014       NIH – Several types of Peer Review Panels
2014       Expert Committee Review for the Canada Foundation for Innovation: Integrating ‘Big Data’ for Health
2015       NIH Winter BDMA Study Section Review Panelist
2015 NIH Special emphasis panelist: Imaging and Biomarkers for Early Cancer Detection
2015 NIH Academic Industrial Partnership (AIP) Review Meeting

Internal Advisory Board, Genomics, Proteomics and Bioinformatics Core Laboratories, UPSOM

Students Participating in Research:

Text Mining for Informative Prior Construction in Omic Data Analysis
Rick Jordan, B.S., M.S. Ph.D. candidate, DBMI, University of Pittsburgh
2010 – present

Ensemble Bayesian Rule Learning
Jeya Balasubramanian, Ph.D. candidate, Intelligent Systems, University of Pittsburgh
2013 - present

Transfer Learning of Classification Rules for Biomarker Discovery through Functional Models
Henry Ogoe, B.S., M.S. Ph.D. candidate, DBMI, University of Pittsburgh
2011 – present

Lung Cancer Methylation Biomarker Discovery from Multiple Platforms
Arturo Lopez Pineda, B.S., M.S. Ph.D. candidate, DBMI, University of Pittsburgh
2010-present

Functional Magnetic Resonance Imaging Data Mining
Rafael Ceschin, B.S. Master’s candidate, DBMI, University of Pittsburgh
2012-present

Cardiac MRI data mining for Early Detection of Pediatric Heart Disease
Yuzhe (Brian) Liu, Medical Student, MSTP, University of Pittsburgh
2014-present

Postdoctoral Trainee

Building a classifier incorporating domain knowledge in Bayesian Logistic Regression
Viji Avali, BS, MS, ME, PhD, Post-Doctoral Scholar, DBMI, University of Pittsburgh
2014-2015

Editorships, Honors, Awards, and Major Lectureships:

Bioinformatics
Reviewer
2003 - present
BMC Bioinformatics  
Reviewer  
2009 - present

Very Large Databases  
Reviewer  
2003 - present

Dental Informatics Journal  
Reviewer  
2003 - present

Data and Knowledge Engineering  
Reviewer  
2003 - present

Algorithms for Molecular Biology - BMC Journal  
2009 - present

International Patent Application Filing  
Zaidi AH, Jobe BA, Zeng X, Balasubramanian JB, Gopalakrishnan V, Bigbee WL, inventors; Methods for the detection of Esophageal Adenocarcinoma. PCT Application Filed by University of Pittsburgh (Dec 2014).

Awards to mentored student:

2015 Marco Ramoni Distinguished Paper Award at AMIA:  

Lectureships:

“Novel Biomarker Discovery Data to Knowledge Pipeline for Precision Medicine,”  
Genome Institute at Washington University, St. Louis, MO. May 4, 2015.


University of Southern California, Keck School of Medicine, Cardiology, Children’s Hospital of Los Angeles, Angeles, Los Angeles, CA. “Informatics of Large Datasets.” November 25, 2013.
Harry Hochheiser, Ph.D.
Assistant Professor of Biomedical Informatics

Harry Hochheiser, Ph.D., is an assistant professor in the department of Biomedical Informatics. He received a B.S. and M.S. in Computer Science from the Massachusetts Institute of Technology and a Ph.D. in Computer Science from the University of Maryland. His postdoctoral work at the National Institute on Aging's Gerontology Research Center focused on the development of user tools for an open-source microscopy informatics platform. He was an assistant professor of Computer and Information Science at Towson University from 2006-2009.

Dr. Hochheiser's research has covered a range of topics, including human-computer interaction, information visualization, bioinformatics, universal usability, security, privacy, and public policy implications of computing systems. At Towson University, he was an investigator on NSF-funded projects in computer security in introductory computer science classes and computational thinking. He is currently working on the development of highly-interactive, user-centered systems for finding and exploring biomedical datasets, with specific applications ranging from basic research data to electronic health records.

Dr. Hochheiser is a reviewer for several journals, including Information Visualization, ACM Transactions on Human Computer Interaction, Interacting with Computers, Risk Analysis, and Advances in Bioinformatics. He has also served on program committees for several conferences, including the IEEE Information Visualization Symposium (2007-2009), the ACM Symposium on Usable Privacy and Security (2009-2010), and the Security and Privacy in Medical and Home-Care Systems Workshop (2009). Dr. Hochheiser has been a member of the Executive Committee of the Association of Computing Machinery's US Public Policy Committee (USACM) since 2004, and he is a co-author of Research Methods in Human-Computer Interaction (Wiley, 2010).

Professional Organization Membership:

Association of Computing Machinery
Member, U.S. Public Policy Committee
2004 - present

American Medical Informatics Association
Member
2009 – present

Students Participating in Research:

K. Romagnoli, DBMI Ph.D. student (expected 2015); MS defended April 2013.
G. Trivedi, ISP PhD Student

Editorships, Honors, Awards, and Major Lectureships:

Academic Editor PeerJ
Rebecca S. Jacobson, M.D., M.S.I.S.
Professor of Biomedical Informatics, Pathology, and Intelligent Systems
Director, Biomedical Training Program

Rebecca Jacobson, M.D., M.S.I.S., is a Professor in the Department of Biomedical Informatics, University of Pittsburgh School of Medicine, with secondary faculty appointments in the Intelligent Systems Program and in the Department of Pathology. She is also the Chief Information officer for the Institute for Personalized Medicine and Director of the Pittsburgh Biomedical Informatics Training Program. Dr. Jacobson is an elected Fellow of the American College of Medical Informatics (ACMI). She received her M.D. from the University of Pittsburgh Medical School and her M.S.I.S. from the University of Pittsburgh School of Information Sciences. She pursued an internship and fellowship at Stanford University, followed by a residency in Pathology at the University of Pittsburgh, and a fellowship with the Department of Biomedical Informatics at the University of Pittsburgh. Her research interests include development and evaluation of natural language processing systems, medical knowledge representation and decision support, information extraction from medical free-text, and cognitive studies of diagnostic expertise. She is a member of four professional societies and is the author of over 75 papers. Her work has been funded by the National Library of Medicine, National Cancer Institute, and Agency for Health Care Research and Quality.

Professional Organization Memberships:

American College of Medical Informatics
Elected Fellow
2010 - present

American Society for Clinical Pathology
Member and Informatics Committee
2007 - present

Alpha Omega Alpha Medical Honor Society
Member
1993 - present

American Medical Informatics Association
Member
1999 - present

Study Sections and Advisory Committee Memberships:

Biodata Management and Analysis (BDMA) Study Section, National Institutes of Health, Center for Scientific Review
Standing Member
2011 – 2017
Information Technology for Cancer Research Study Section, National Cancer Institute
Member, 2014
Chair (X2), 2015

Students Participating in Research:

Using Electronic Medical Records to Measure Guideline Adherence in Low-Resource Settings
Zach Landis-Lewis, M.S.I.S., M.S. (obtained 2010), Ph.D. candidate (obtained 2014),
DBMI, University of Pittsburgh
2007 – 2014

BIRADS Information Extraction
Sergio Castro Diaz (current MS student)

Editorships, Honors, Awards, and Major Lectureships:

American College of Medical Informatics
Elected as a Fellow
2010 - present

Xia Jiang, Ph.D.
Assistant Professor of Biomedical Informatics, Intelligent Systems Program, and
Carnegie Mellon – University of Pittsburgh Ph.D. Program in Computational Biology

Dr. Xia Jiang, Ph.D., is an assistant professor in the Department of Biomedical Informatics at the University of Pittsburgh School of Medicine. In 1997, I received my first M.S. from Rose-Hulman Institute of Technology, a highly rated engineering school with a #1 ranking by the U.S. News in the category of Best Undergraduate Engineering Programs. In 1999, I received my second M.S. from Northeastern Illinois University, where I first became involved in Bayesian network related research. I received my Ph.D. from the University of Pittsburgh in 2008. At Pitt, I did my graduate work in the laboratory of Dr. Gregory Cooper. I then did a postdoctoral fellowship supported by NLM/ NIH for over three years. I was a graduate level course instructor in the Computer Science Department at Northeastern Illinois University for three years before I came to Pitt. I am the author of 20 peer reviewed articles as well as the co-author of two books. My first book titled Probabilistic Methods for Financial and Marketing Informatics was published by Morgan Kaufmann in April 2007. My second book titled Contemporary Artificial Intelligence will be published by Chapman & Hall/CRC in August of this year. I am the principal investigator of an NIH R00 project titled “Detecting Genome Wide Epistasis with Efficient Bayesian Network Learning,” which runs from 2012 to 2015.

One of my specific areas of interests is developing advanced computational methods for high-dimensional data analysis. The abundance of high-dimensional datasets, such as the
next-generation sequencing data, offers us unprecedented opportunities to make major breakthroughs in biomedical and translational science and technology. To harness the potential in these datasets we need analytical methods specialized to overcoming the difficulties inherent in analyzing these data. I am also very interested in translational informatics. I will devote my efforts in developing advanced informatics tools that assist the translation of the findings in basic scientific research efficiently and effectively into patient medical care. My research collaborators include mathematicians, computer scientists, statisticians, physicians, pathologists, biologists, geneticist, and peer informaticians from University of Pittsburgh, Carnegie Mellon University, Northeastern Illinois University, and Northwestern University. To promote the state-of-the-art scientific research and training, I have two goals: 1) I will devote my efforts to extend my collaborations with colleagues in other programs/departments at Pitt and other universities, in the hope of fostering interdisciplinary education and research, and 2) I will strive to build an amiable, collaborative, and productive microenvironment at my research lab which will focus on both research and education and training. I am willing to interact actively with students at all levels and across a variety of research domains.

Students Participating in Research:

I helped Yuriy Sverchkov, an ISP student, in developing a spatial cluster detection algorithm using dynamic programming.

From October 1, 2012 to present, I served as the mentor for Binghuang, Cai, a postdoc associate in DBMI.

From May 1, 2012 to July 31, 2012, I served as the research advisor for Diyang Xue, who worked as a research volunteer.

From August 1, 2013 to present, I served as the research advisor for Diyang Xue, a GSR in ISP at Pitt.

From Aug 1, 2015, I served as the research advisor for Sanghoon Lee, a PhD student of DBMI.

From August 1, 2015 to present, I served as the research advisor for Chandra Rathnam, a MS student in DBMI.

Editorships, Honors, Awards, and Major Lectureships:

Editorships

Guest Editor, Cancer Informatics Supplement (2013-2014)

Awards

2010, National Library of Medicine postdoctoral training scholarship.

**Manuscript Reviews**
Reviewer, PLoS
Reviewer, BMC Bioinformatics
Reviewer, Genetic Epidemiology
Reviewer, SpringerPlus
Reviewer, Bayesian Analysis
Reviewer, Journal of Biomedical Informatics
Reviewer, IEEE Intelligent Systems
Reviewer, AMIA 2013 Annual Symposium

**Xinghua Lu, M.D., Ph.D.**

*Associate Professor of Biomedical Informatics and Biomedical Informatics Training Program Core Faculty*

Xinghua Lu, M.D., Ph.D., is an Associate Professor of Biomedical Informatics and Biomedical Informatics Training Program Core Faculty. He received a M.S. in Cardiology and a M.D. in medicine at Shandong Medical University. He received a Ph.D. degree in Pharmacology from the University of Connecticut Health Center and a certificate in Biomedical Informatics at the University of Pittsburgh. He was an Associate Professor in the Department of Biostatistics, Bioinformatics, and Epidemiology/Director of the Bioinformatics Division and the Director of the NLM Training Program at the University of South Carolina before joining the University of Pittsburgh in 2010.

Dr. Lu is a biomedical informatics researcher with broad experiences in clinical, basic and computational biosciences. His current research efforts span two complementary domains: 1) developing statistical text mining approaches to acquire biomedical knowledge from biomedical literature; and, 2) developing statistical data mining approaches to study cellular signaling systems. He has extensive experience in developing and implementing tools for semantic analysis of free texts, using advanced computational statistics techniques, graph theory, and information theory. Dr. Lu’s basic science training and research experience in systems biology enables him to combine knowledge mining and data mining to gain insight of biological systems and disease mechanisms.

Dr. Lu also has extensive experience in biomedical informatics education. He served as the Director for NLM training program the Medical University of South Carolina for 4 years. He also served a co-director of two NIH-funded T32 training program in biostatistics and a Department of Education-supported training program in systems biology. He has supervised over 20 trainees across the spectrum from undergraduate, graduate to postdoctoral fellows. As a core faculty of the training program, he also offers
didactic courses in machine learning, artificial intelligence and translational bioinformatics and serves as an advisor for trainees in the program.

**Study Sections and Advisory Committee Memberships:**

**Study Section of the National Library of Medicine (NLM)**  
Member, NIH Special Panel on Pharmacogenomics  
Member, NLM Special Panel on Biomedical Informatics

**Students Participating in Research:**

**Identifying Signal Transduction Pathways Underlying Cancers**  
Songjian Lu, Ph.D., Postdoc Associate, DBMI, University of Pittsburgh  
2009 – present

**Ontology-guided knowledge mining and knowledge acquisition**  
Chunhui Cai, PhD, Postdoc Associate, DBMI, University of Pittsburgh

**Modeling the Signaling Roles of Sphingolipids in Yeast**  
Lujia Chen, Ph.D. student, DBMI, University of Pittsburgh  
2009 - present

**Ontology-guided knowledge mining and knowledge acquisition**  
Vicky Chen, Ph.D. student, Department of Biomedical Informatics, University of Pittsburgh  
2009 - present

**Revealing transcriptional regulation mechanism through deep learning**  
Joyeeta Dutta-Moscato, Ph.D. student, Department of Biomedical Informatics, University of Pittsburgh  
2012 – present

**Revealing cellular signaling mechanisms for personalized cancer therapy**  
Jonathon Young, MD, MS student, Department of Biomedical Informatics  
2013 – present

**Editorships, Honors, Awards, and Major Lectureships:**

**BMC Research Notes**  
Associate Editor  
2007 - present

**DNA Repair**  
Editorial Board  
2013-present
Fu-Chiang (Rich) Tsui, Ph.D.
Assistant Professor of Biomedical Informatics, Intelligent Systems, Bioengineering, and Clinical and Translational Science
Biomedical Informatics Training Program Core Faculty

Fu-Chiang (Rich) Tsui, Ph.D., is Assistant Professor of Biomedical Informatics, Intelligent Systems, Bioengineering, and Clinical and Translational Science. He is also Associate Director of RODS Laboratory.

Dr. Tsui received his PhD in Electrical Engineering, premed training, and postdoctoral training in medical informatics at the University of Pittsburgh.

Dr. Tsui’s laboratory is funded by grants from the National Institutes of Health, National Library of Medicine, the Patient Centered Outcomes Research Institute (PCORI) as well as non-traditional (i.e., non-NIH) sources of funding including McCune Foundation, Coulter Foundation, and Pennsylvania Innovation Works. Dr. Tsui serves as a PI of two grants and a co-Investigator of several grants. He is a member of several professional societies and has contributed to over 90 papers and his publications have been actively cited (Google h-index: 38).

Study Sections and Advisory Committee Memberships:

Office of Naval Research, Department of the Navy
Ad-hoc Grant Reviewer
2007 - present

Students Participating in Research:

Adult Heart Failure Readmission Prediction
Diyang Xue, PhD student, ISP, University of Pittsburgh
2015 - present

Laboratory Tests for Readmission Prediction
Amie Draper, PhD student, DBMI, University of Pittsburgh
2014 - present

Single vs. Multiple Report per Emergency Department Visits for Influenza Detection
Victor Ruiz, PhD student, DBMI, University of Pittsburgh
2014 - present

Use of Machine Learning Algorithms for Building Influenza Detection Models
Arturo Lopez Pineda, PhD student, DBMI, University of Pittsburgh
2014 - present

Evaluation of NLP and Bayesian classifiers
Ye Ye, Ph.D. student, ISP, University of Pittsburgh
2011 – present
Search of (near) optimal OTC medication products as health indicators for influenza surveillance
Ruhsary Rexit, M.S. student, CS, University of Pittsburgh
2012 – 2015

Editorships, Honors, Awards, and Major Lectureships:

Advances in Disease Surveillance
Associate Editor
2006 - present

American Medical Informatics Association
Ad-hoc Reviewer
2006 - present

Artificial Intelligence in Medicine
Ad-hoc Reviewer
2008 - present

The Open Medical Informatics Journal
Editorial Board Member and Ad-hoc Reviewer
2008 - present

Scientific Journals International
Editorial Board Member
2007 - present

BMC Medicine
Ad-hoc Reviewer
2008 – present

Shyam Visweswaran, M.D., Ph.D.
Assistant Professor of Biomedical Informatics, Intelligent Systems, Clinical and Translational Science, and Computational Biology
Associate Director, Biomedical Informatics Training Program

Shyam Visweswaran, M.D., Ph.D., is an Assistant Professor of Biomedical Informatics. His medical education began in the Jawaharlal Institute of Post-Graduate Medical Education and Research, Pondicherry, India, where he received the M.B.B.S. in 1989. He received a M.S. in Physiology and Biophysics at the University of Illinois at Urbana-Champaign, IL, in 1996. He subsequently did a residency in Neurology at Boston University, Boston, MA, that he completed in 2000. He received a Ph.D. in Intelligent Systems (Biomedical Informatics track) at the University of Pittsburgh, PA, in 2007. His research interests lie in the application of artificial intelligence and machine learning to personalized and genomic medicine, patient-specific predictive modeling and computerized clinical decision support.
Study Sections and Advisory Committee Memberships:

National Science Foundation (NSF) review panel: NSF CISE, 2012.

Students Participating in Research:

**Biomarker Discovery in Exome Data**  
Ian Wong, M.S., Doctoral candidate, Intelligent System Program (Biomedical Informatics track), University of Pittsburgh  
2006 – present

**Ranking of Predictive Biomarkers in High Dimensional Genomic Data**  
Matt Stokes, Doctoral candidate, Intelligent System Program (Biomedical Informatics track), University of Pittsburgh  
2009 – present

**Deep Learning**  
Eric V. Strobl, Masters and Doctoral candidate, Biomedical Informatics, University of Pittsburgh  
2013 – present

Editorships, Honors, Awards, and Major Lectureships:

**International Journal of Medical Engineering and Informatics (IJMEI)**  
Editorial board  
2007 - present

Distinguished paper award at the 2013 AMIA Summit on Translational Bioinformatics for a co-authored paper.

A computer guy’s take on personalized medicine. An interview published in PITTMED, University of Pittsburgh School of Medicine, Summer 2013, Vol. 15, Issue 2.

**Michael M. Wagner, M.D., Ph.D.**  
Director, Corporate Relations  
Director, RODS Laboratory  
Associate Professor of Biomedical Informatics and Intelligent Systems

Michael M. Wagner, M.D., Ph.D., is Associate Professor of Biomedical Informatics and Intelligent Systems. He earned a M.D. from the New York University School of Medicine and a Ph.D. in intelligent systems from the University of Pittsburgh. Dr. Wagner is the director of the Real-time Outbreak and Disease Surveillance (RODS) Laboratory, a unit within the Department of Biomedical Informatics. In 1999, Dr. Wagner founded the RODS Laboratory with a mission to investigate methods for the real-time detection and assessment of disease outbreaks using information technology. RODS developed the first real-time public health surveillance system (the RODS system), which examines anonymous clinical data from hospitals, as well as the National
Retail Data Monitor (NRDM), a surveillance tool that monitors sales of over-the-counter (OTC) drugs looking for patterns that indicate a disease outbreak. Dr. Wagner has served as the Principal Investigator of multi-institutional, interdisciplinary projects supported by the Department of Homeland Security, CDC, AHRQ, DARPA, Pennsylvania Department of Health, Alfred P. Sloan Foundation, and the National Institutes of Health and Library of Medicine. He is the author of more than 60 peer-reviewed publications and is the editor of a full-length book entitled: *Handbook of Biosurveillance—a decision analytic tool for use by analysts working in health departments*.

**Editorships, Honors, Awards, and Major Lectureships:**

**Journal of the American Medical Informatics Association**  
Reviewer  
1999 - present

**PLoS Med (Public Library of Science Medicine - Journal)**  
Reviewer 2010
BIBLIOGRAPHY

Michael J. Becich, M.D., Ph.D.


**Tanja Bekhuis, PhD, MS, MLIS, AHIP**

**Bekhuis T, Demner-Fushman D, Crowley RS.** Comparative effectiveness research designs in MeSH and Emtree: an evaluation of coverage. JMLA: J Med Libr Assoc 2013;101(2):92-100. PMCID: PMC3634392. [Featured in the AHRQ Scientific Resource Center Methods Library, May 2013].


**Bekhuis T, Kim K, Valappil B, Spallek H.** Adapting the Scottish PRIME Fissure Sealant Questionnaire for implementation research in the United States: lessons learned. APA


Richard D. Boyce, Ph.D.


Giménez, JAM., Blagec, K., Boyce, RD., Adlassnig, KP., Samwald, M.. An Ontology-Based, Mobile-Optimized System for Pharmacogenomic Decision Support at the Point-of-Care. PLOS ONE. 2014 May 9(5). DOI: 10.1371/journal.pone.0093769. PMCID: PMC4008421


Refereed conference articles:


**Gregory F. Cooper, M.D., Ph.D.**

Villamarin R, Cooper G, Wagner M, Tsui FC, Espino JU. A method for estimating from thermometer sales the incidence of diseases that are symptomatically similar to influenza. Journal of Biomedical Informatics 46 (2013) 444-457. PMID: 23501015


**Gerald P. Douglas Ph.D.**


**Madhavi Ganapathiraju, Ph.D.**


Vanathi Gopalakrishnan, Ph.D.


Harry Hochheiser, Ph.D.


Lazar J, Hochheiser H. Legal Aspects of Interface Accessibility in the U.S. Communications of the ACM. 2013 Dec; 56 (12):74-80.


Trivedi, G., Pham, P., Chapman, W., Hwa, R., Wiebe, J., **Hochheiser, H.** An Interactive Tool for Natural Language Processing on Clinical Text. Paper presented at the 2015 TextVis workshop.


Rebecca Jacobson, M.D., M.S.I.S.


Book Chapter:


Xia Jiang, Ph.D.


Cai B, Jiang X. Revealing Biological Pathways Implicated in Lung Cancer from TCGA Gene Expression Data using Gene Set Enrichment Analysis. Supplementary Issue: Computational Advances in Cancer Informatics (A), 2014: 113-121.


Xinghua Lu, M.D., Ph.D.


systematic perturbation and integromics analyses. Science Signaling 6:rs14 (reported as in press last year, published in October, 2013)


Fu-Chiang (Rich) Tsui, Ph.D.


Liu TY, Sanders JL, Tsui F-C, Espino JU, Dato VM, and Suyama J, Association of Over-The-Counter pharmaceutical sales with Influenza-Like-Illnesses to patient volume in an urgent care setting, PLOS One, 8(3), March 2013


Peer-Reviewed Abstract, Conference Proceedings


Shyam Visweswaran, M.D., Ph.D.


**Michael M. Wagner, M.D., Ph.D.**


STAFF LISTING

The Department of Biomedical Informatics operates from two floor locations. The primary administrative offices and graduate training program are located on the 4th and 5th floors at the offices at 5607 Baum Blvd in Shadyside. The faculty and the group of software design and development staff are split between the 4th and 5th floors at the offices at 5607 Baum Blvd in Shadyside. Although the department is divided into two floor locations, regular meetings and working groups bring faculty and staff together.

1 Executive Administrator
1 Operations Manager
5 Administrative Support Staff
2 Grants & Finance
2 Program Managers (Training Program, RODS Lab)
3 Senior Researchers
1 Clinical Data Scientist
2 Research Specialists
20 Software Design and Development

Administrative Staff

Genine Bartolotta
Administrative Support Staff

Lucy Cafeo
Administrative Support Staff

Robert Cecchetti
Executive Administrator

Pamela Farneth
Administrative Support Staff

Celeste Flaherty-Thomas
Grants and Finance Administrator

Barbara Karnbauer
Operations Manager

Jesse Kummer
Sponsored Project Accountant

Linda Mignogna
Administrative Support Staff

**Nova Smith**  
Administrative Support Staff

**Toni Porterfield**  
Training Program Manager

**Cleat Szczepaniak**  
RODS Laboratory Program Manager

---

**Senior Research Staff**

**Waqas Amin, M.D.**  
Senior Research Scientist

**Uma Chandran, Ph.D.**  
Research Associate

**Jeremy Espino, M.D., M.S.**  
Senior Research Scientist

---

**Technical Staff**

**Rebecca Boes**  
Software Design/Development

**Charles Borromeo**  
Software Design/Development

**Anish Bhaswanth Chakka**  
Software Design/Development

**Girish Chavan**  
Software Design/Development

**Johnson Paul Kottakalil**  
Software Design/Development

**Liz Legowski**  
Research Specialist

**John Levander**  
Software Design/Development
Soumya Luthra  
Software Design/Development

Olga Medvedeva  
Software Design/Development

Nicholas Millett  
Software Design/Development

John Milnes  
Software Design/Development

Kevin Mitchell  
Software Design/Development

Michelle Morris  
Software Design/Development

Yifan Ning  
Software Design/Development

Melissa Saul  
Clinical Data Scientist

Shiyi Shen  
Software Design/Development

William Shirey  
Software Design/Development

Harpreet Singh  
Software Design/Development

Hoah-Der Su  
Software Design/Development

Kerry Trent  
Research Specialist

Eugene Tseytlin  
Software Design/Development

Sahawut Wasaratchakit  
Software Design/Development

Joyce Zelnis  
Software Design/Development
FY2015 Financial Summary

The FY2015 actual performance for Hard Money (Dept. Operations) revenue exceeded the budget. The actual net operating budget surplus was $107,899. A transfer was made to move surplus to the Office of the Senior Vice Chancellor Health Sciences (SVCHS), as they covered DBMI’s operating cost deficit from FY14. The SVCHS has provided a 3 year period for DBMI to repay this remaining deficit amount. The key measures of research support included indirect research revenue of $2.821 million, about $64k above the projected budget. Direct research revenue of $9.497 million was $178k more than budget, reflecting a more successful research portfolio. (See the Statement of Revenue and Expenses Fiscal Year 2015)
### Department of Biomedical Informatics
#### Statement of Revenues and Expenses
##### Fiscal Year 2015

<table>
<thead>
<tr>
<th></th>
<th>FY15 Budget</th>
<th>FY15 Actual</th>
<th>FY15 Budget</th>
<th>FY15 Actual</th>
<th>FY15 Budget</th>
<th>FY15 Actual</th>
<th>FY15 Budget</th>
<th>FY15 Actual</th>
<th>FY15 Budget</th>
<th>FY15 Actual</th>
<th>FY15 Budget</th>
<th>FY15 Actual</th>
<th>FY15 Budget</th>
<th>FY15 Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Revenue</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grant Revenue</td>
<td>$2,707,043</td>
<td>$2,821,326</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$9,319,214</td>
<td>$9,497,643</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Revenue</td>
<td>$640,403</td>
<td>$649,463</td>
<td>$188,903</td>
<td>$119,457</td>
<td>$26,716</td>
<td>$198,266</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School of Medicine</td>
<td></td>
<td></td>
<td>$1,300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education Support</td>
<td>$307,356</td>
<td>$307,356</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Revenue</td>
<td>$3,654,802</td>
<td>$3,774,665</td>
<td>$188,903</td>
<td>$119,457</td>
<td>$26,716</td>
<td>$198,266</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Expenses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personnel and Other</td>
<td>$3,267,474</td>
<td>$3,355,694</td>
<td>$188,903</td>
<td>$167,406</td>
<td>$11,276</td>
<td>$462,970</td>
<td>$9,319,214</td>
<td>$9,497,643</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University Overhead</td>
<td>$445,096</td>
<td>$445,096</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$26,550</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Expenses</td>
<td>$3,712,570</td>
<td>$3,750,784</td>
<td>$188,903</td>
<td>$167,406</td>
<td>$127,826</td>
<td>$489,520</td>
<td>$9,319,214</td>
<td>$9,497,643</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surplus / (Deficit)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficit Transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net Surplus / (Deficit)</td>
<td>$-</td>
<td>$-</td>
<td>$-</td>
<td>$-</td>
<td>$-</td>
<td>$-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fund Balance as of 06/30/16</td>
<td>$39,916</td>
<td>$287,631</td>
<td>$127,928</td>
<td>$1,964,928</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- Operating Account (Indirect Grant revenue) was $64,163 above budget, based on successful grant applications with the EDJ and MIDAS
- Operating Account expenses was $55,552 above budget. DBMI was able to start repaying the loan from DoD (payment was for $107,094)
- Service Revenue accounted for $69,546 under budget as a result of the FY14 NRDAM revenue carryover, which reduced the cost per store that DBMI could charge
- Service Account expenses were under budget, as a result of less personnel devoting effort on the NRDAM project
- Discretionary Account income exceeded budget projections, due to Personalized Medicine funds provided by Dr. Berg for Dr. Jacobson. Also, recruitment funds were added. $470,344
- Discretionary Account expenses were over budget as a result of PGRF and Bioinformatics project expenses.
- Direct Grant revenue was exceeded budget, resulting from more research awards than budgeted.

**FY16 budget anticipates:**
1. Start Up costs for Dr. Jiang (Staff) and Eddy Douglas (GRF) will start again in FY16.
2. Large spike in IDC due to the Center for Commercial Applications (CCA)
3. NRDAM costs will slightly increase in FY16, as DBMI utilizes FY15 deficit balance in FY16 to project to a zero balance for June 30th, 2016.
4. Service Account revenue and expense projected will increase, based on requests for Umangonya's time for the pending SAG Core.
5. PGRF under the direction of Dr. David Boone should increase for the coolBill program, as program expands from 2 months to 12 months.
Five-Year Plan

The key five-year objectives for the Department of Biomedical Informatics (DBMI), as stated in the FY2007 annual report, included: 1) Providing national and regional leadership in research in informatics; 2) Providing the highest quality instruction in informatics; and, 3) Providing the highest quality of support for the clinical practice of medicine.

1. Leadership in Research in Informatics

Historically, the research portfolio of the Department now places it as the top funded department of biomedical informatics in the nation. There are no official national reports for biomedical informatics funding, but based on continual informal discussions among biomedical informatics department chairs, the University of Pittsburgh, Department of Biomedical Informatics Pittsburgh is now agreed to be the best funded medical school department.

The Department’s funding FY2015 research revenue budget of $12.3M is nearly 40% higher than FY2014 actual research revenue, mainly a result of the successful funding of several new sources of federal research support. We are projecting $15.9M in funding for FY16 which will represent another 30% increase in research funding.

2. Highest Quality of Instruction in Informatics

Since 1987, the National Library of Medicine (NLM) has supported a fellowship training program in biomedical informatics at the University of Pittsburgh. FY2015 represents the third year in the current five-year approved Training Grant award. A competitive renewal application submitted in FY12 for another five years of support was awarded in FY13. Additionally, there are approximately twenty additional full and part-time graduate students beyond the NLM funded fellows.

Since there are no official quality measures of biomedical informatics training programs, it is difficult to ascertain where DBMI is positioned relative to other programs. However, DBMI has committed to a five-year goal of increasing the number of students enrolled in the graduate training program and to increase the number of trainee publications and graduates going into academic practice. Student publications continue to represent improvement in the academic quality of the Training Program.

An on-line certificate program was implemented in FY2014, which has added a new dimension to the program, both in terms of student recruitment, but also financial support, as the tuition from the on-line program will remain with the program.

Following the establishment of the Center for Health Informatics for the Underserved, DBMI has created an additional focus of concentration in the area of global health informatics, offering educational opportunities in the classroom, on-line, and “in vivo” as part of a Summer Internship in Global Health Informatics on-site in Malawi, Africa.
3. Highest Quality of Support for the Clinical Practice of Medicine

The current plan involves a major initiative in Personalized Medicine led by Dr. Rebecca Jacobson, CIO for the Institute for Personalized Medicine (IPM). It will also include new leadership in clinical informatics and focus on DBMI’s efforts in Big Data to Knowledge. Since the “computable phenotype” is critical in this effort, Shyam Visweswaran who is leading a new NCAT funded effort to link electronic medical records at 20 CTSA sites to enhance accrual to the highest priority clinical trials has assumed a new leadership role for DBMI in Clinical Informatics. In addition, Dr. Becich PCORI CDRN program will also play an instrumental role in creating a Learning Health System in partnership with UPMC. A major new UPMCE led initiative which involved DBMI is called the Pittsburgh Health Data Alliance was launched July 2015 and Dr. Becich was named the Director for Center for Commercial Application (CCA) of Healthcare Data. UPMCE funded CCA with $12M ($2M/yr for six years) of Early Stage Research Projects and $22M for Directed Major Projects. The purpose of CCA is to improve health through the commercialization of healthcare big data and analytics technologies. DBMI will be funded to commercialize Fall Sentinel (PI = Richard Boyce). Finally, Rich Tsui has successfully commercialized his SHARP technology which uses analytics to predict and prevent readmission for pediatric patients at risk for hospital readmission and is implementing SHARP it in partnership with UPMC at Children’s Hospital of Pittsburgh. This is exceptional progress in building strong and lasting relationships with UPMC.

FY2016 Financial Projections

The FY2016 budget projects an increase in research funding from the previous year. The FY2016 budget also includes a decrease in service account revenue, largely from the National Retail Drug Monitoring (NRDM) program, and a budgetary surplus in the discretionary account balance. Overall, the net projected revenue for FY2016 reflects a significant surplus of $4,771,775, which is mainly attributable to recruitment funds received from the SoM and a 31.8% increase in grant revenue, largely in part from the Center of Commercial Application funding from UPMC.
### Schedule of Revenue and Expenses Fiscal Year 2016 Budget

<table>
<thead>
<tr>
<th></th>
<th>University</th>
<th>UPP</th>
<th>Total Budget FY2016</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Revenue</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient Care</td>
<td>$ -</td>
<td>$ -</td>
<td>$ -</td>
</tr>
<tr>
<td>Grant:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directs</td>
<td>$ 12,821,173</td>
<td></td>
<td>$ 12,821,173</td>
</tr>
<tr>
<td>Indirects</td>
<td>$ 3,776,741</td>
<td></td>
<td>$ 3,776,741</td>
</tr>
<tr>
<td>Hospital Contract</td>
<td>$ -</td>
<td></td>
<td>$ -</td>
</tr>
<tr>
<td>School of Medicine</td>
<td>$ 376,710</td>
<td></td>
<td>$ 376,710</td>
</tr>
<tr>
<td>VAMC</td>
<td>$ -</td>
<td></td>
<td>$ -</td>
</tr>
<tr>
<td>Other</td>
<td>$ 5,945,625</td>
<td></td>
<td>$ 5,945,625</td>
</tr>
<tr>
<td><strong>Total Revenue</strong></td>
<td>$ 22,920,249</td>
<td>$ -</td>
<td>$ 22,920,249</td>
</tr>
<tr>
<td><strong>Expenses</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salaries and Fringe Benefits:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty</td>
<td>$ 10,650,668</td>
<td></td>
<td>$ 10,650,668</td>
</tr>
<tr>
<td>Non-Faculty</td>
<td>$ -</td>
<td></td>
<td>$ -</td>
</tr>
<tr>
<td>Malpractice Insurance</td>
<td>$ -</td>
<td></td>
<td>$ -</td>
</tr>
<tr>
<td>Space Rental</td>
<td>$ 797,715</td>
<td></td>
<td>$ 797,715</td>
</tr>
<tr>
<td>UPP Overhead</td>
<td>$ -</td>
<td></td>
<td>$ -</td>
</tr>
<tr>
<td>University Overhead</td>
<td>$ 471,563</td>
<td></td>
<td>$ 471,563</td>
</tr>
<tr>
<td>Other Operating Expenses</td>
<td>$ 6,228,230</td>
<td></td>
<td>$ 6,228,230</td>
</tr>
<tr>
<td><strong>Total Operating Expenses</strong></td>
<td>$ 18,148,474</td>
<td>$ -</td>
<td>$ 18,148,474</td>
</tr>
<tr>
<td><strong>Excess Revenue over Expenses</strong></td>
<td>$ 4,771,775</td>
<td>$ -</td>
<td>$ 4,771,775</td>
</tr>
<tr>
<td><strong>Capital Equipment/Improvements</strong></td>
<td>$ -</td>
<td>$ -</td>
<td>$ -</td>
</tr>
<tr>
<td><strong>Fund Balances</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University Restricted Accounts as of 6/30/15</td>
<td>$ -</td>
<td>$ -</td>
<td>$ -</td>
</tr>
<tr>
<td>University Endowments as of 6/30/15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPP Fund Balance as of 6/30/15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPMC Endowments as of 6/30/15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPMC SPF Accounts as of 6/30/15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Fund Balances</strong></td>
<td>$ -</td>
<td>$ -</td>
<td>$ -</td>
</tr>
</tbody>
</table>
END OF REPORT