2014 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, 2014

A. TABLE OF CONTENTS

B. ACTIVITIES AND MISSION

DEPARTMENT OF BIOMEDICAL INFORMATICS MISSION STATEMENT

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

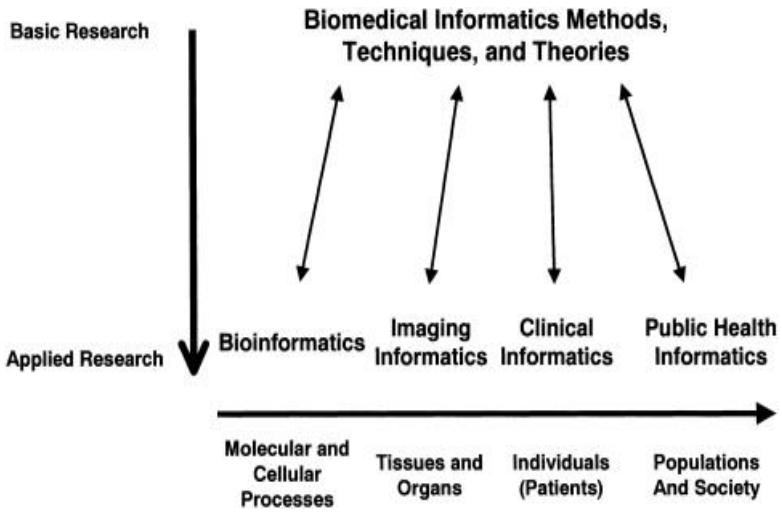
- To provide the **highest quality of training in Informatics** and to provide our students with a world-class education that prepares them to become outstanding leaders in biomedical informatics research, education, and practice.

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

2013-2014 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 July 1, 2013, the Department of Biomedical Informatics consisted of sixteen core faculty. No faculty additions or departures occurred, therefore the number of DBMI faculty remains at sixteen.

For FY2014, the Department had $8,444,278 in research revenue and $567,039 in other operating revenue from teaching and service contracts. With research funding remaining at a lower level due to the difficult financial environment nationally, the department did not meet the budgeted expectations for FY2014. Direct grant expenditures for FY2014 were $504,511 lower than FY2013, and the indirect revenue associated with those grants was $135,648 lower.

Under sequestration in FY 2013, NIH awarded 722 fewer grants than in FY 2012. The steadily declining budget allocations for NIH from FY 2010 through FY 2013

contributed, in part, to the steady decline in grant-application success rates, which decreased each year during this period and fell to an all-time low of 16.8 percent in FY 2013. The impact is there is an overall reduction of grants awarded, thus the downward trend of research dollars awarded to Biomedical Informatics continued during FY14. An indicator of this is presented by the total funding per faculty member being decreased by 6.52% from $549K per faculty member in FY2013 to $513K per faculty member in FY2014. Of note, at time of this report, DBMI is poised to dramatically rebound from the post-ARRA funding situation and has secured several new major grants which will be reflected in our next Annual Report to the SoM. Our current, projected FY15 budget is $10.5 M which is a greater than 20% increase in annual grant revenue from actual FY14.

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 FY14 are being launched including *NIH Big Data to Knowledge* (BD2K), (Cooper CO-PI) and *PaTH*, a PCORI project (Becich CO-I and Informatics/HIT faculty lead), making expectations for FY15 higher than in previous years. Additionally, an extreme emphasis on grant proposal efforts in FY14, resulting in nearly 60 submitted proposals, increases prospects as well. Of note, DBMI faculty had committed to submitting 46 grants and they actually submitted 59. The largest number of grants submitted by DBMI in past years was 26!!!

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/). 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 FY2014, the Department of Biomedical Informatics had $6,085,154 in direct research expenditures and $2,359,153 in indirect research revenue for a total of FY2014 research support of $8,444,277. Based on grant award notices received and estimates for pending awards, the approved FY2015 total research budget is $12,076,257, a 4.6 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:

MEASURES	FY10	FY11	FY12	FY13	FY14	FY 14 Change From FY10
Directs	**$7,305,726**	**$7,933,452**	**$6,235,831**	**$6,625,664**	**$6,085,154**	**(16.71)**
Indirects	**$2,654,219**	**$2,844,350**	**$2,305,331**	**$2,494,800**	**$2,359,153**	**(11.12)**
Total Grants	**$9,959,945**	**$10,777,802**	**$8,541,162**	**$9,120,465**	**$8,444,277**	**(15.22)**
Revenue/FTE	**$665,327**	**$698,497**	**$551,042**	**$549,527**	**$513,689**	**(22.79)**

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:

1) Amin W, Tsui F, Borromeo C, Chuang CH, Espino JU, Ford D, Hwang W, Kapoor W, Lehmann H, Martich GD, Morton S, Paranjape A, Shirey W, Sorensen A, **Becich MJ**, Hess, R. and the PaTH Network. PaTH: A Learning Health System for the Mid-Atlantic Region. JAMIA. 2014 Jul-Aug;21(4):633-6. doi: 10.1136/amiajnl-2014-002759. Epub 2014 May 12. PMID: 24821745
2) Amin W, Srinivasan M, Song SY, Parwani AV, **Becich MJ**. Use of automated image analysis in evaluation of Mesothelioma Tissue Microarray (TMA) from National Mesothelioma Virtual Bank. Pathol Res Pract. 2014 Feb;210(2):79-82 PMID:2435972 [PubMed - in process]
3) Amin W, Parwani A, Melamed J, Flores R, Pennathur A, Valdivieso F, Whelan N, Landreneau R, Luketich J, Feldman M, Pass H, and **Becich MJ**. National Mesothelioma Virtual Bank: A Platform for Collaborative Research and Mesothelioma Biobanking Resource to Support Translational Research. Lung Cancer International, Volume 2013, Article ID 765748.

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/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:

Amin W, Tsui FR, Borromeo C, Chuang CH, Espino JU, Ford D, et al. PaTH: towards a learning health system in the Mid-Atlantic region. J Am Med Inform Assoc. 2014 Aug;21(4):633–6.

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 CTSA 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
Recent results: Boyce. RD., Ryan. PB., Noren. N., et al., Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. Drug Safety. 2014. Volume 37, Issue 8 (2014), Page 557-567. PMCID: PMC4134480 and Schneider, J., Ciccarese, P., Clark, T., Boyce, RD. Using the Micropublications ontology and the Open Annotation Data Model to represent evidence within a drug-drug interaction knowledge base. The 4th Workshop on Linked Science 2014— Making Sense Out of Data (LISC2014). Collocated with the 13th International Semantic Web Conference (ISWC2014). October 19th and 20th, Riva del Garda, Trentino, Italy. [Proceedings in process.]
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.
Recent results: Boyce. RD., Ryan. PB., Noren. N., et al., Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. Drug Safety. 2014. Volume 37, Issue 8 (2014), Page 557-567. PMCID: PMC4134480 and Boyce, RD., Perera, S., Nace, Culley, C., Handler SM. Laboratory Monitoring Adverse Drug Event Alert Communication Preferences of Nursing Home Physicians. Recently accepted to Applied Clinical Informatics. 2014.
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 measureable 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.
Recent results: Boyce, RD., Freimuth, RR., Romagnoli, KM., Pummer, T., Hochheiser, H., Empey, PE. Toward semantic modeling of pharmacogenomic knowledge for clinical and translational decision support. Proceedings of the 2013 AMIA Summit on Translational Bioinformatics. San Francisco, March 2013. PMCID: PMC3814496 and Samwald, M., Freimuth, R., Luciano, JS., Lin, Simon., Powers, RL., Marshall, MS, Adlassniga, KP., Dumontier, M. Boyce, RD. An RDF/OWL Knowledge Base for Query Answering and Decision Support in Clinical Pharmacogenetics. Proceedings of the 14th World Congress on Medical and Health Informatics. Copenhagen, Denmark. August 2013. PMCID: PMC4028612

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: 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.

Future plans: Planned next steps are to improve the machine-learning methodology and evaluate the system prospectively on ICU cases.

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. A preliminary evaluation of the method indicates that it is able to perform useful detection and characterization of influenza disease outbreaks.

Gerald P. Douglas Ph.D.

Project Title: Optimizing the Value of EMR Systems in Low-Income Countries using Economic Models

Project Background: About two-thirds of the world's population lives in low-income countries (LICs). The World Bank defines LICs as those countries with a Gross Domestic Product per capita below $1,005 (less than $3 per day). LICs comprise 36 countries, of which 23 are on the African continent, and the combined populations of these countries represent 63% percent of the World's population. Thus finding ways to improve and optimize care delivery in these countries would improve quality of life for vast numbers of people.

Research objectives: This work will define, develop, and evaluate a novel, setting-driven approach that will enable LIC hospitals to make cost-effective choices about implementing EMRs. By using a Lean approach to identify the waste (opportunities) in clinical workflow processes and by identifying EMR components to address these specific opportunities, our tool will maximize the immediate impact and financial value of EMR technology in LIC hospital settings. Findings from this research may inform EMR adoption strategies in middle-income countries and low-resource settings in high-income countries, which face similar

challenges. Specific aims are: Aim 1: Populate our prototype Return on Investment (ROI) model with informatics interventions designed to address real-world problems experienced in LIC hospital settings, based on findings from partner hospitals. Aim 2: Validate and refine the predictive power of the ROI model. Aim 3: Create a web-based version of the ROI model that is preloaded with LIC health and economic data, and release it to the community.
Principal methods: Economic modeling, Lean Healthcare concepts, focus groups & contextual inquiry.
Recent results: We have modeled three informatics interventions and valued then in the context of a specific hospital setting in a Low-income country.
Conclusions: This is new research.
Future plans: We have submitted an R01 grant application to support this work in partner hospitals in Malawi and Haiti.

Project Title: Biosurveillance Using Routinely-Collected Outpatient Data In Malawi
Project Background: In December 2009, the government of Malawi introduced electronic registers to more than 16 out of 28 district hospitals to collect data for national planning of health resources and for biosurveillance. These electronic registers obviate the need for paper registers. With over four years of use, electronic registers present an opportunity to reduce the workload of health care workers and improve the quality of data for biosurveillance.
Research objectives: The objective of this research was to determine whether signal detection algorithms when applied to routinely collected data from electronic outpatient registers could be used to detect events of public health importance.
Principal methods: We compared three time series analysis algorithms for outbreak detection on simulated outbreaks, and then applied the optimal algorithm to the electronic outpatient data to detect previously reported measles and cholera outbreaks.
Recent results: We identified one measles and two cholera outbreaks and mapped 131 indicators to 11 syndromes. Of the three tested algorithms, simple moving averages produced the best results. This algorithm successfully raised an alert for the measles outbreak and cholera at one of the three districts within two weeks of an estimated start date.
Conclusions: This pilot study demonstrated that routinely collected data from electronic outpatient registers in Malawian clinics can be used for public health surveillance.
Future plans: Algorithms and approaches refined through this research will be piloted to detect outbreaks in real-time in district hospital clinics in Malawi.

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-inflammatory 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). Next, we constructed the interactome of schizophrenia genes with over 600 novel PPIs (this work is under review). 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 is under review at Nature). All 14 PPIs that we studied experimentally in wet lab (through collaborations at Pitt) have been validated to be true interactions.

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 – 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.

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: Drug Safety and Pharmacogenomics, Addressing Gaps in Clinically Useful Evidence on Drug-Drug Interactions *(NLM, 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.

User-Centered design and evaluation of PGX@Pitt: A pharmacogenomics information portal to support clinical decision making (K. Romagnoli, PhD project, under my supervision)

Clinicians need informatics resources that provide relevant and actionable information about the impact genetic variants have on medication response. We hypothesize that a user-centered, pharmacogenomics information resource will answer clinicians' pharmacogenomics questions more accurately and more efficiently than current practices.

Research objectives: To understand clinical information needs regarding drug-drug interactions and pharmacogenomics; build tools that will provide this information in a manner that will support effective decision-making and prescribing.

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; 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; preliminary literature review of drug-drug interaction information systems.

Conclusions: None yet

Future plans: Continued interviews; Iterative design and refinement of candidate information tools.

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: None yet
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: TIES is an existing software system for developing networked repositories of sharable de-identified surgical pathology reports. Previous funded projects extended the capabilities of the system by improving the portability of the system, extended the types of documents that can be processed, evaluated the system's NLP performance and usability, and built a user community to support this open-source application. Over the seven years of previous work, we have created a mature, hardened, and deployed software system that transforms siloed archives and repositories into investigator-friendly, easily accessible research repositories by using natural language processing (NLP) methods.
Research objectives: The focus of the current project is further enhancement of the TIES system. We expect 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 have focused on (1) the release of TIES 5.0, (2) establishing the TIES Cancer Research Network, including establish both regulatory, and (3) initiating pilot projects between centers.
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 expect to enhance the informatics technology to support inter-institutional "trust," paraffin registry development, tissue microarray (TMA) development, and nondestructive tissue use. During Y2 we expect to complete the initial set or requirements, design, and first UI implementation for regulatory aspects of protocol management across the network.

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 grant application brings together two Principal Investigators, Dr. Rebecca Jacobson, and Dr. Guergana Savova, with complementary skills and extensive expertise in biomedical informatics, cancer informatics, and natural language processing. They have equally shared in the development of this project and their combined skills are needed for the overall execution of this study. Dr. Jacobson serves in the role of contact principal investigator.
Research objectives: The proposed research is a truly collaborative and interdisciplinary endeavor that will rely heavily on the expertise brought by each of the PIs, and co-investigators. Because of the highly collaborative aspect of this project, its success will require a shared organizational model. 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.
Principal methods: 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. These labs will contribute phenotype variables for extraction, test utility and usability of the software, and provide the setting for an extrinsic evaluation.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: Since the start date of this project in May, 2014, we have implemented the communication mechanisms outlined in the Multi-PI Plan included in the submission. Team conference calls as well as PI phone meetings occur weekly. There has been suitable progress made on the Project Plan, which extensively outlines schedules and steps of the project.
Conclusions: This is new research beginning only in May of 2014**.**
Future plans: We are making good progress towards our stated goals for Y1 and expect to stay on track to progress as stated in our specific aims.

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. Dr. Jiang 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: 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.
Conclusions: 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.
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.

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

Project Title: Development of Personalized Predictive Models for Personalized Medicine
Project Background: The typical paradigm consists of learning a single predictive model from a database of individuals, which is then applied to predict outcomes for any future

individual. Such a model is called a population-wide model because it is intended to be applied to an entire population of future individuals. In contrast, personalized predictive modeling, the focus of this research theme, focuses on learning models that are tailored to the characteristics of the individual at hand. Personalized models that are optimized to perform well for a specific individual often have better predictive performance than the typical population-wide models that are optimized to have good predictive performance on average on all future individuals. Moreover, personalized models can identify features such as genomic factors that are specific for an individual.

Research objectives: Develop a new personalized predictive model that uses decision tree models and apply it to identify genomic factors that are specific for an individual with a disease of interest.

Principal methods: We developed a new Bayesian approach for evaluating personalized predictive decision tree models, and applied it to predict the risk of developing late-onset Alzheimer's disease using genomic variant data obtained from a genome-wide association study.

Recent results: The personalized predictive decision tree models had better predictive performance than the typically used population-wide decision trees.

Conclusions: The personalized predictive decision tree models can predict better the risk of developing late-onset Alzheimer's disease using genomic data.

Future plans: Apply the personalized predictive model approach to identify individual genomic factors in chronic pancreatitis using exome data.

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 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. Please see our newly designed website for recent developments: https://www.midas.pitt.edu/
Conclusions: Not applicable.
Future plans: We have begun implementing the Aims as stated above.

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.

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.

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:

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.

Standards implementation: Plans are to connect an additional two DTMs to the Apollo Web Services.

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:

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

A P2aTH Towards a Learning Health System in the Mid-Atlantic Region
PCORI CDRN
PI: Hess
2014 - 2015

Richard D. Boyce, Ph.D.

Improving Medication Safety For Nursing Home Residents Prescribed Psychotropic Drugs
NIH K01
PI: RD Boyce, Ph.D.
2013 - 2016

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

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

Cancer deep phenotype extraction from electronic medical records
NIH U24
PI: Rebecca S. Jacobson
2014 - 2019

Advanced Development of TIES – Enhancing Access to Tissue for Cancer Research
NIH U24
PI: Rebecca S. Jacobson, M.D., M.S.
2013 - 2018

Xia Jiang, PhD

A new generation clinical decision support system
NIH/NLM R01
PI: X Jiang, PhD
2014 - 2018

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

Probabilistic Disease Surveillance
NIH R01
PI: MM Wagner, M.D., Ph.D.
2013 - 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

SPORE in Skin Cancer
NIH
University of Pittsburgh Cancer Institute (UPCI)
PI/Program Director: M Becich, M.D., Ph.D.
2008 - 2018
Cancer Center Support (CCSG)
NIH/NCI
Core Director: M Becich, M.D., Ph.D.
1997 - 2015

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

Research to automate classification of medical reports in support of systematic reviewers.
Lister Hill National Center for Biomedical Communications, Communications Engineering Branch, US National Library of Medicine, NIH
Dr. Dina Demner-Fushman, MD, PhD, Staff Scientist

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

Roger Day, Sc.D.

Gynecologic Cancer Center of Excellence
Department of Defense, (Henry M. Jackson Foundation)
PI: L Maxwell, M.D., Walter Reed Army Medical Center, Washington, DC
2010 - 2013

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
PI: 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
PI: 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.

Windber Research Institute
K Kolli, Ph.D.

Inflammatory Bowel Disease
University of California Los Angeles
L Goodglick, 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

Steven Handler, M.D., Ph.D.

A Randomized Controlled Trial to Assess the Impact of a Telemedicine Medication Delivery Unit on Medication Adherence Following Hospitalization for Common Heart Conditions
INRange Systems Incorporated
PI: Steven Handler, M.D., Ph.D.
2012 - 2013

The Effect of a Computer-generated Rounding Report on Physician Workflow in the Nursing Home
American Medical Directors Association (AMDA)
PI: Steven Handler, M.D., Ph.D.
2012 - 2013

Harry Hochheiser, Ph.D.

FaceBase Management and Coordination Hub
NIH
University of Iowa
PI: M Marazita, Ph.D. Collaborators: University of Iowa - J Murray, M.D.; University of Washington - J Brinkley, M.D., Ph.D.; M Cunningham, M.D., Ph.D.; C Heike, M.D., M.S.; L Shapiro, Ph.D.; University of Southern California - Y Chai, D.D.S., Ph.D.; University of California Los Angeles - S Ruffins, Ph.D.; University of California – San Diego - W Chapman, Ph.D.
2009 – 2014

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.

Quantifying Electronic Medical Records Usability to Improve Clinical Workflow
AHRQ
San Diego VA;
PI: Z. Agha; San Diego VA: San Diego VA: S. Ashfaq, A. Calvitti, K. Bell, M. Gabuzda, UCSD – L. Liu, N. Farber, N. Weibel, J. Hollan; NYU – D. Zuest ; Texas A&M: R. Street; UC Irvine – Y. Chen; U. Michigan: K. Zheng; U. Utah: C. Weir
2012-2016.

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

Interactive Search and Review of Clinical Records with Multi-layered Semantic Annotation
NLM
UCSD
PI: W. Chapman (UCSD), R. Hwa, J. Wiebe
2011-2014.

Ontology-based integration of human studies data
NIDCR
U. Washington
PI: L Shapiro. U. Washington: J. Brinkley, C. Heike, M. Cunningham, T. Cox, R. Travillian, J. Mejino
2011-2014.

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

Advanced Development of TIES
University of Pennsylvania
MJ Feldman, M.D., Ph.D.

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

Pharmacogenomic Implementation in Remote Environments via EMRs
Vanderbilt University
Kevin B Johnson, M.D., M.S.
2013 - present

Advanced Development of Informatics Technology
Children's Hospital of Boston
Guergana Kirilova Savova, Ph.D.
2010 - present

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

Modeling and Analysis of Roles of Yeast Sphingolipids
Stony Brook University of New York
Y Hannun, M.D.

Facilitating Unintrusive Navigation in Online Health Communities
University of Illinois Urbana-Champaingn
Chengxiang Zhai, PhD

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

Pediatric Hospital Readmission Prediction Project
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
Christina Patterson, MD, Pediatric Neurologist
2013 - present

Adult Heart Failure Hospital Readmission Prediction Project
UPMC Heart and Vascular Institute
Mark Schmidhofer, M.S., M.D., Director, Coronary Intensive Care Unit UPMC Presbyterian
Ravi Ramani, M.D.
Michael Mathier, M.D.
2013 - present

Influenza Case Detection Project
Allegheny County Health Department
R Voorhees, M.D., Chief of Epidemiology and Biostatistics
2009 - present

Electronic Lab Reporting Project
William Pasculle, ScD, Director of the UPMC's Clinical Microbiology laboratories
Carlene Muto, MD, Medical Director of Infection Control and Hospital, UPMC
2004 - present

MedLEE Software
Columbia University
Carol Friedman, Ph.D.
2007 - present

RODS Deployment
Tarrant County Health Department, TX, Southwest Center for Advanced Public Health Practice
B Stephens, Manager
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

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
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

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

Dr. Bekhuis' research belongs to the emerging field of clinical 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. Recently, she studied barriers to implementing evidence-based guidelines *(Journal of Evidence-based Dental Practice 2010),* and investigated the information needs of dentists expressed in online messages to enable subsequent in-depth qualitative research *(Journal of Medical Internet Research 2011).* A follow-up study

considered dentists' interest in and knowledge of the connection between oral and systemic disease (*BMC Oral Health* under review). Dr. Bekhuis compared the representation of study designs in biomedical vocabularies and ontologies (*Journal of the Medical Library Association 2013*) and classified studies with nonrandomized designs for inclusion in systematic reviews *(Studies in Health Technology and Informatics 2010* and *Artificial Intelligence in Medicine 2012).* A follow-up project is underway to improve classification by considering various term sets derived statistically or extracted from external documents (*PLoS One* under review). For her dental informatics research, Dr. Bekhuis continues to collaborate with researchers in the School of Dental Medicine and the American Dental Association. For her primary research, she collaborates with Dr. Dina Demner-Fushman at Lister Hill, US National Library of Medicine. In sum, Dr. Bekhuis' main goal is to support human health and welfare by facilitating information retrieval and translation of medical research.

Richard D. Boyce, Ph.D.

Assistant Professor of Biomedical Informatics
Biomedical Informatics Training Program Core Faculty

Richard D. Boyce, Ph.D., is currently funded on 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.

Roger S. Day, Sc.D.

Associate Professor of Biomedical Informatics and Biostatistics

Over many years, Dr. Day has cared about the treatment of cancer, and how computational and modeling tools could help people create better biological understanding, then apply it to better individual treatment decisions. The technical developments worked on to support this goal constitute a collection of topics which all related to understanding cancer treatment better. These topics are knowledge representation, software architecture for comprehensive modeling and validation, multi-scale modeling in cancer, strategies for overcoming drug resistance in cancer, and how pharmaceutical and biological interactions should be statistically modeled and detected.

A major thrust of this effort is combining biomathematical models with research results and other knowledge, to better understand the natural history of cancer and develop individualized treatment strategies for cancer. The Oncology Thinking Cap (OncoTCap) provides a platform for identifying and solving problems along this path. Developed by the Educational Resource for Tumor Heterogeneity, OncoTCap integrates cancer knowledge engineering with a biomathematical modeling environment, used in education of cancer professionals. A sub-theme of great current interest is the connection with cancer stem cells. A hypothetic treatment strategy, the "worst drug rule", was discovered long ago using multitype branching process models. It fits neatly with new discoveries about cancer stem cells, with which it could be combined to yield new treatment plans of great effect.

A related effort is the study of biological, statistical and drug interaction, and the relationships among these related but distinct concepts. A generalized additive effects model has been developed for drug interactions, and computational methods are being developed and applied to data of the Combinatorial Chemistry for Cancer Drug Development project (J.Lazo, A.Vogt). A unique "weakest link" methodology for combining predictors has been developed and applied to microarray data for lung cancer

(T.Richards, N.Kaminski). This methodology was also applied to laser scanning cytometry data on lung cancer primaries in relation to recurrence (T.Luong, S.Shackney).

A distinct highly active research area is the development of statistical and computational methodologies and support software for ethically-oriented clinical trial design based on Bayesian decision theory. Methods were developed with M. Wang for Phase I clinical trials, to consider both adverse events and response endpoints, and to incorporate pharmacogenetic information. Later Dr. Day worked with a doctoral student, Y. Wang, to develop the Clinical Trials Design Evaluator, an evaluation platform based on simulation designed to foster cooperation and sharing of interoperable code modules for clinical trial designs, population models, outcome models, and evaluation criteria. This is currently under active development.

The most important recent work towards better cancer detection and treatment focuses on fostering more clinically relevant biomarker validation studies. Biomarker studies are almost uncountable, yet they have had little impact on clinical practice. The design of better studies called for a method of communicating clinical ethical comparative judgments and converting them into performance specifications and thence into sample size requirements. The right concept for this job is "number needed to treat", repurposed from its original role as a descriptor of treatment consequences. Along the way, an important result is a new "contra-Bayes" theorem facilitating the derivation of requirements for sensitivity and specificity.

A recent published paper extended our paradigm for using integration of bioinformatics datasets to create a measuring tool for comparing the quality of identifier mapping and identifier filtering methods. The extension utilizes a decision theory paradigm to create a richer framework for optimizing the decisions. Application to identifier filtering has been completed. In comparing nine different filtering methods in two very different, it showed superiority of the Jetset method over competitors. This work is backed by the publication of several R bioconductor packages.

Gerald P. Douglas Ph.D.

Assistant Professor of Biomedical Informatics
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 10 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.

Dr Ganapathiraju's most 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.

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 in collaboration with Drs. Jill Siegfried and Dr. William Bigbee.

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.

Steven Handler, M.D., Ph.D.

Assistant Professor in the Department of Biomedical Informatics, and Division of Geriatric Medicine, and Clinical and Translational Research, Director, Clinical Informatics, at the University of Pittsburgh, Medical Director, Long-term Care, Health Information Technology, UPMC
Core Faculty of the RAND-University of Pittsburgh Health Institute (RUPHI)

Dr. Handler is an Assistant Professor with a primary appointment in the Department of Biomedical Informatics and secondary appointments in Geriatric Medicine, and Clinical and Translational Research and Director of Clinical Informatics. Dr. Handler is also Core Faculty of the RAND-University of Pittsburgh Health Institute (RUPHI).. Dr. Handler is a practicing geriatrician with direct patient care and medical director responsibilities in the nursing home setting. Dr. Handler serves as the Medical Director for Long-term Care Health information technology at the University of Pittsburgh Medical Center (UPMC), as well as the Director of Clinical Informatics for the Department of Biomedical Informatics.

In these roles, Dr. Handler has been responsible for managing projects including the development of the first and only nursing home data repository in the world that contains laboratory, medication, vital-signs and Minimum Data Set Data that can now be accessed by the UPMC/University of Pittsburgh clinical and research community, adapting the first and only real-time active medication monitoring system to automate the detection and management of adverse drug events in the nursing home, adapting this real-time active medication monitoring system for use in the UPMC Presbyterian MICU, development of a computer-generated paper-based rounding report to provide structured clinical documentation in the nursing home, adapting a point-of-care charge capture program for use on mobile devices for the Division of Geriatric Medicine, and the development and distribution of the "lean medical record" to facilitate the distribution of critical clinical information when patients are transferred from a UPMC hospital to UPMC nursing home.

Dr. Handler's primary research area focuses on medication and patient safety primarily in the nursing home setting. As part of this research, he has developed the first nursing home specific, IHI-endorsed trigger tool for conducting rapid and targeted chart reviews for detecting potential adverse drug events. This tool is currently being evaluated for use in the VA Community Living Centers (i.e., nursing homes) for inclusion in the consultant pharmacist monthly medication regimen review process and has most recently been used by the Office of the Inspector General to conduct a national estimate of all-cause resident

harm. His most recent research at the intersection of biomedical informatics and geriatrics is funded by the Agency for Healthcare Research and Quality (AHRQ) and focuses on the use of active medication monitoring systems to detect and reduce the number of adverse drug events among older nursing home residents and hospitalized patients.

Finally, Dr. Handler Co-Directs the University of Pittsburgh Geriatric Pharmaceutical Outcomes and Gero-Informatics Research and Training Program with Dr. Hanlon. This program is a collaboration between the Division of Geriatric Medicine and Gerontology, and the Department of Biomedical Informatics in the School of Medicine. The overall goal of this interdisciplinary program is to serve as a hub for research and research training in Geriatric Pharmaceutical and Gero-Informatics research.

Active Funded Grants in FY 14:

Enhancing the Detection and Management of Adverse Drug Events in the Nursing Home
Agency for Healthcare Research and Quality (AHRQ): This study represents the first large, well-controlled, comprehensive examination of an active medication monitoring system in the NH. We developed and pilot-tested an active medication monitoring system for use in a single NH, where it was shown to detect Adverse Drug Events (ADEs) with a high degree of accuracy and at a rate of nearly 2.5 times that of usual care. We will conduct a cluster randomized controlled trial among 36 NH physicians working in one of 4 UPMC NHs in Southwestern PA for a period of 12 months. Our hypotheses are that NH patients managed by physicians who receive active medication monitoring alerts will have more ADEs detected, will have a faster ADE management response time, and will result in cost-savings from a societal perspective compared to usual care.

Transforming the Role of the Hospital Pharmacist to Improve Patients' Access, Adherence, and Self-management of Medication after Discharge: The goal of this quality improvement initiative is to enhance and evaluate a recently implemented pharmacy patient-care model at UPMC-Presbyterian Shadyside, using a tool that will enable the pharmacist to engage the patient in the process of shared decision-making for medication self-management.

Reduce Avoidable hospitalizations using Evidence-based interventions for Nursing facilities in Western Pennsylvania (RAVEN): UPMC will implement an intervention in 19 nursing facilities in the western region of Pennsylvania. UPMC Community Provider Services has created a program called "RAVEN" (Reduce AVoidable hospitalizations using Evidence-based interventions for Nursing facilities in western Pennsylvania). The goals of the program are to: 1) reduce the frequency of avoidable hospital admissions and readmissions among nursing facility residents, 2) Improve nursing facility resident health outcomes; 3) improve the bi-directional process of transitional care between hospitals and nursing facilities; 4) reduce overall health care spending without restricting access to care or choice of providers, and; 5) facilitate knowledge transfer and culture change within partner nursing facilities to sustain these best practices over the long-term. This program will include facility-based nurse practitioners to assist with determining resident care plan goals, and conduct acute change in condition assessments. It will also implement

evidence-based clinical communication tools such as INTERACT and others recommended by the American Medical Directors Association to assist in structuring and standardizing clinical assessments and recommendations. The intervention will also provide support from innovative telehealth and information technologies to connect participating nursing facilities into the Western PA Health Information Exchange.

Pilot Study to Determine the Feasibility of Using Telemedicine to Assist Nurse Practitioners with Managing Acute Change in Condition and Palliative Care Assessments of Nursing Home Residents: The primary goal of this quality improvement project is to determine the feasibility of using telemedicine to assist nurse practitioners with managing acute change in condition and palliative care assessments of UPMC nursing home patients. To assess feasibility, we intend to utilize web-based surveys to quantify the perception of these services in terms of quality of the medical care provided, quality of the equipment used, and perceived barriers to implementation, both before and following each telemedicine encounter.

Pittsburgh Older American Independence Center (Pepper Center): The specific aims are to develop interventions to improve function and independence in older adults with balance disorders, integrate studies of physiologic, biomechanical and psychosocial mechanisms affecting balance with clinical studies, and foster multidisciplinary research and research training.

A Multicenter Evaluation of Off-label Medication Use and Adverse Drug Events in Adult Intensive Care Unit Patients: The goal of this multi-center, comparative, prospective evaluation is to determine the incidence of adverse drug reactions (ADRs) associated with off-label drug use and elucidate if off-label drug use is an independent risk factor for the development of ADRs in ICU patients. These data will be used to develop a pharmacy practice model for prevention of ADRs from off-label drug use by incorporating clinical decision support into an alerting system.

The Effect of a Computer-generated Rounding Report on Physician Workflow in the Nursing Home: The goal of this quality improvement project is to determine the impact of a paper-based computer-generated rounding report (i.e., a summary document formatted like a SOAP note that contains the most relevant clinical and demographic data as determined by geriatric practitioners) on physician workflow in the nursing home. The findings of this project will likely result in the use of an information system that will yield improved physician efficiency by reducing the effort required to gather the data necessary to generate rounding reports, improve the completeness of available information, and reduce some of the commonly reported barriers to providing patient care in the nursing home setting.

Feasibility ofUusing a Telemedicine Medication Delivery Unit for Older Adults that Require Medication Assistance During Transition from Hospital to Home: Using Pennsylvania Department of Aging resources, this project's short-term goal is to assess the feasibility of using a telemedicine medication delivery unit for frail older adults that require medication assistance in their home immediately following an acute hospitalization. As part of this feasibility assessment, we will assess several methods and intervention-related components.

University of Pittsburgh Program for Pharmaceutical Outcomes Research in Aging: This National Institute on Aging project of the University of Pittsburgh Geriatric Pharmaceutical Outcomes and Gero-Informatics Research and Training Program is a collaboration between the Division of Geriatric Medicine and Gerontology, and the Department of Biomedical Informatics in the School of Medicine. The overall goal of this interdisciplinary program is to serve as a hub for research and research training in Geriatric Pharmaceutical and Gero-Informatics research.

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.
- Interactive collaboration search tools for research social network environments
- Design of diagnostic ontologies and support tools for genetic craniofacial disorders
- Examination of the usability of electronic medical records, particularly in low-resource settings.

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

Professor of Biomedical Informatics, Pathology, and Intelligent Systems
Director, Biomedical Informatics Training Program

Dr. Jacobson's research focuses on the development of knowledge based systems to facilitate medical research and education. Her research interests are manifested in two main areas: (1) cognitive systems and clinical decision-making, and (2) development of translational informatics systems using automated processing of clinical text. Dr. Jacobson currently has 67 accepted or published peer-reviewed publications, of which 16 are first authored and 35 are senior authored. Her H-index is 21. Two notable medical informatics systems are the direct result of her research efforts. Dr. Jacobson has also written two book chapters, and a number of white papers.

In the area of Cognitive Systems and Clinical Decision Making, Dr. Jacobson is a nationally recognized expert. Dr. Jacobson has created an advanced medical training system, and used the system to study the nature of diagnostic reasoning. The system is one of the only medical intelligent tutoring systems evaluated with results showing dramatic learning gains in diagnostic and reporting accuracy in formative evaluations. A summative evaluation demonstrated the system's ability to improve the quality and accuracy of melanoma reports among practicing community pathologists. Dr. Jacobson has described and tested methods for predictive modeling of skill acquisition in this complex domain, and has shown that the system can be used to trace known heuristics and biases. The project has also explored the use of written language interfaces, the role of immediate feedback, reification of domain knowledge, and the use of metacognitive scaffolding.

In the area of clinical and translational informatics research, Dr. Jacobson has achieved national recognition through her work on caTIES. Building on methods of Natural Language Processing (NLP), the open source caTIES system builds a repository of deidentified, highly processed and coded, free-text documents that can be searched within and across organizations. caTIES was one of the first developed systems for the Cancer Biomedical Informatics Grid (caBIG). The system is currently deployed at University of Pittsburgh and numerous other academic institutions across the country. The software provides a platform for the development of translational research networks, based on a foundation of research on data sharing. Since her promotion to Associate Professor, Dr. Jacobson has developed research programs in ontology enrichment and use of ontologies for natural language processing of clinical documents. A third system (ODIE) resulted from this work. Increasingly, Dr. Jacobson's work centers on use of information extraction techniques for deep phenotyping to support personalized medicine.

Since 2012, Dr. Jacobson has spent a significant part of her time in the development of a computing infrastructure for personalized medicine. Her role in this regard is two-fold. As Business Lead for the UPMC Enterprise Analytics (EA) Research Solutions Delivery Team, Dr. Jacobson provides direction for the research use of the UPMC Enterprise Data Warehouse, which incorporates the Oracle Translational Development Center and other

deployed informatics products to support translational research using data derived from next generation sequencing. In this role, she is also a member of two key Enterprise Analytics working groups at UPMC – the Data Governance Council and the Analytics Community Center. As newly named Chief Information Officer (CIO) for the Institute for Personalized Medicine (under the direction of Dr. Jeremy Berg), she provides oversight for infrastructure initiatives designed to advance personalized medicine. These include, for example, efforts to create a cohesive distributed health sciences computing platform for genomic 'big data', in collaboration with Dr. Mike Barmada at the Simulation and Modeling Center and Dr. Shawn Brown at the Pittsburgh Supercomputing Center.

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 Epistasisis, 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 scored 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.

Associate Director, RODS Laboratory
Research Assistant Professor of Biomedical Informatics and Intelligent Systems

Dr. Tsui is a world-renowned biomedical informatician **whose research publication impact ranking based on Google Index is top 65 in the world and only two other faculty members (chair and vice chair) in DBMI are in the top 100 list.** (http://allisonbmccoy.github.io/scholar-scraper/index-bmi.html). Dr. Tsui's research interest includes hospital readmission prediction and management, big data analysis, public health surveillance, time series analyses, machine learning, expert systems, grid computing, and artificial intelligence to biomedical informatics research problems.

SHARP System: In 2013, Dr. Tsui started working closely with UPMC Children's Hospital and Presbyterian University Hospital to develop predictive tools to automatically identify and manage hospitalized patients who may be in high risk of readmission within 30 days. Starting from 2013, the Centers for Medicare and Medicaid Services (CMS) imposes re-imbursement penalty for hospitals with excess readmissions; about 2,200 hospitals will be penalized this year, each of which could lose about $125,000 on average, with the total loss for the industry as $280 million. Moreover, Medicaid and private insurance companies like High Mark Blue Cross and Blue Shield started non-payment penalty for patients readmitted in 30 days. Currently hospitals primarily rely on manual process for readmission reduction, which is not sustainable and resource intensive. Dr. Tsui and his team have developed a **S**ystem for **H**ospital **A**daptive **R**eadmission **P**rediction and Management (SHARP) System based on big data approach by processing available structured and unstructured electronic health records (EHRs). SHARP can automatically identify critical features from EHRs and build predictive models using the features. The system is under review to be deployed in Pittsburgh Children's Hospital of UPMC.

ELR reporting project (Pennsylvania Department of Health): Dr. Tsui will be working with Pennsylvania Department of Health to facilitate electronic laboratory reporting using standard terminology funded by CDC. This initiative work will focus on HIV reporting and it aligns well with our existing case detection project under the Center of Excellence funding supported by CDC.

Biosurveillance (Public Health Surveillance) Grid: Dr. Tsui recently received a two-year CDC funding for building, deploying and evaluating the PA-OH Biosurveillance Grid (PA-OH BiG), which allows federated data sharing between five participating public health entities: the Pennsylvania Department of Health, the Ohio Department of Health, the Center for Disease Control and Prevention (CDC), the Allegheny County Health

Department, and the University of Pittsburgh. The goal of this project is to advance current public health practice by using an efficient secure grid network to exchange data instead of using fax or mails.

Bayesian Case Detection System: Dr. Tsui continues to build a Bayesian case detection system (CDS) using Bayesian networks and natural language processing tools for detection of influenza and three additional respiratory diseases. This project will move existing biosurveillance to the next level by notifying public health practitioners of suspected cases (e.g., suspected SARS or measles cases) before the confirmed lab reports become available, which will allow more time for public health officials to conduct an investigation and will provide more data to guide an investigation. The long-term goal of this project is to help hospital and public health officials identify cases of importance and to develop a technical and professional relationship between public health practice to local hospitals.

Influenza-like-Illness (ILI) Counts Reporting: With the imminent threat of H1N1 swine flu, this project is to automatically detect ILI cases and fever cases by processing UPMC Emergency Department (ED) freetext reports and provides daily counts for the last 7 days to the Allegheny County Health Department. This ILI detection system serves as an automated sentinel physician reporting system overseeing several UPMC hospitals.

Biomedical Research Database: Dr. Tsui is working on building a biomedical research database based on the framework of OpenMRS, a well-developed popular open source medical record database. The initial phase is to migrate existing Clinical Event Monitor (CLEM) database to the standard OpenMRS and adopts other open source tools such as Mirth to process various hospital data feeds such as Radiology reports, dictation reports, laboratory (including culture) reports, pharmacy orders and radiology/lab orders. Such database allows researchers perform efficient data queries for conducting various research projects. It currently serves MRSA case detection/reporting gthat Dr. Tsui recently initiated.

Algorithms for Biosurveillance: The goal is to develop spatial and temporal algorithms for the early detection of disease outbreaks from data that are commonly available electronically. For example, RODS Laboratory currently collects admission data in real-time from emergency departments in Pennsylvania and daily over-the-counter medicine sales from the entire country. Dr. Tsui focuses on the development of algorithms that use multiple data sets and spatial temporal information to achieve early detection of disease outbreaks.

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, patent-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

Source of Funding	Principal Investigator	Project Start	Project End	Direct Costs for current budget period	Indirect Cost for current budget	Description
National Institutes of Health	Michael J. Becich, M.D. Ph.D.	10-Sep-97	31-Jul-15	134,437	69,235	Cancer Center Support Grant
Centers for Disease Control and Prevention	Michael J. Becich, M.D. Ph.D.	1-Sep-06	31-Aug-16	395,549	213,596	Continuation of the National Mesothelioma Virtual Bank
National Institutes of Health	Michael J. Becich, M.D. Ph.D.	1-May-12	30-Apr-15	176,139	90,710	Sarcoidosis and A1AT Genomics & Informatics Center
National Institutes of Health	Michael J. Becich, M.D. Ph.D.	26-Aug-08	30-Jun-18	79,882	41,938	SPORE in Skin Cancer
National Institutes of Health	Michael J. Becich, M.D. Ph.D.	1-Jul-14	30-Jun-15	500,585	257,801	University of Pittsburgh Clinical and Translational Science Institute
National Institutes of Health	Michael J. Becich, M.D. Ph.D.	1-Sep-12	31-Aug-13	1,800	927	University of Pittsburgh Research Tissue, Data and Biological Specimen
John Hopkins University (NIH)	Michael J. Becich, M.D. Ph.D.	7-Sep-10	30-Jun-15	3,683	1,934	Refinement and Discovery of Nuclear Matrix Protein Markers for Colorectal Cancer
Patient-Centered Outcome Research Institute	Michael J. Becich, M.D. Ph.D.	14-Mar-14	30-Sep-15	390,442	146,129	A P2aTH Towards a Learning Health System in the Mid-Atlantic Region
National Institutes of Health	Tanja Bekhuis, Ph.D.	1-Jul-12	30-Jun-15	133,411	68,707	Screening Nonrandomized Studies for Inclusion in Systematic Reviews of Evidence
National Institutes of Health	Richard Boyce, Ph.D.	30-Nov-16	30-Nov-16	92,831	7,353	Improving Medication Safety for Nursing Home Residents Prescribed Antidepressants
National Institutes of Health	Richard Boyce, Ph.D.	15-Feb-14	14-Feb-14	284,232	115,677	Addressing gaps in clincally useful evidence on drug-drug interactions

National Institutes of Health	Gregory F. Cooper, M.D., Ph.D.	30-Sep-09	29-Sep-13	251,125	119,517	Using Medical Records Repositories to Improve the Alert System Design
Carnegie Mellon Univ (NSF)	Gregory F. Cooper, M.D., Ph.D.	1-Sep-09	31-Aug-13	75,896	39,087	III Large Discovering Complex Anomalous Patterns
National Institutes of Health	Roger Day, Sc.D.	10-Sep-97	31-Jul-15	7,595	3,914	Cancer Center Support Grant
National Institutes of Health	Roger Day, Sc.D.	28-Feb-16	31-Jul-15	1,497	120	Biostatistics with a Focus on Cardiovascular Health and Minority Populations
National Institutes of Health	Roger Day, Sc.D.	26-Aug-08	30-Jun-14	29,337	15,402	SPORE in Skin Cancer
Univerity of Michigan (NIH)	Roger Day, Sc.D.	8-Feb-10	31-Dec-13	4,944	2,596	Statistical Methods and Issues for Implementing Adaptive Phase I Trials
Centers for Disease Control and Prevention	Gerald Douglas. Ph.D.	1-Oct-12	30-Sep-13	104,766	38,134	Improving Quality of Care and Health Impact through Sustainable, Integrated, Inno
National Institutes of Health	Madhavi Ganapathiraju, Ph.D.	1-Aug-11	30-Apr-16	250,000	108,658	Discovery of Mental Health an Inflammation (MHAIN) Interactome
National Institutes of Health	Vanathi Gopalakrishnan, Ph.D.	1-May-06	31-Aug-16	1,518,152	781,848	Spore in Lung Cancer
National Institutes of Health	Vanathi Gopalakrishnan, Ph.D.	15-Aug-11	30-Jun-14	210,631	108,473	Bayesian Rule Learning Methods for Disease Prediction and Biomarker Discovery
National Institutes of Health	Vanathi Gopalakrishnan, Ph.D.	24-Sep-12	31-Jul-15	161,101	84,779	Transfer Learning Rule for Knowledge Based Biomarker Discovery and Predictive Bio
Agency for Healthcare Research & Quality	Steven Handler, M.D., M.S.	1-May-10	30-Apr-14	370,250	124,694	Enhancing the Detection and Management of Adverse Drug Events in the Nursing Home

National Institutes of Health	Steven Handler, M.D., M.S.	1-Sep-12	31-Aug-14	4,201	975	Pittsburgh Older Americans Independence Center
National Institutes of Health	Steven Handler, M.D., M.S.	30-Sep-09	31-Aug-14	4,506	360	UPITT Program for Pharmaceutical Outcomes Reseach in Aging
National Institutes of Health	Harry Hochheiser, Ph.D.	1-May-14	30-Apr-19	6,407	3,445	FaceBase 2: Management and Integration Hub
Univ of Washington (NIH)	Harry Hochheiser, Ph.D.	21-Sep-09	30-Apr-14	747,459	289,471	Technology Project: Shape-Based Retrieval of 3D Craniofacial Data
Univ of Utah (NIH)	Harry Hochheiser, Ph.D.	1-Oct-13	30-Sep-14	21,371	11,218	Interactive Search and Review of Clinical Records with Multilayered Semantic Annotation
University of Oregon (NIH)	Harry Hochheiser, Ph.D.	1-Feb-13	31-Jul-16	66,334	34,825	Semantic LAMHDI: Linking diseases to model organism res
University of Iowa (NIH)	Harry Hochheiser, Ph.D.	21-Sep-09	30-Apr-14	206,972	106,591	FaceBase Management and Coordination Hub
VA Research Foundation of Pittsburgh	Harry Hochheiser, Ph.D.	1-Dec-12	30-Jun-15	10,259	5,386	Quantifying Electronic Medical Records Usability to Imp
National Institutes of Health	Rebecca Jacobson, M.D., M.S.	21-Aug-03	31-Aug-13	300,255	18,426	The Cancer Training Web-A Multi-Institutional Tutoring System in Pathology
National Institutes of Health	Rebecca Jacobson, M.D., M.S.	1-Jan-09	31-Dec-13	100,638	51,829	Continued Development and Evaluation of caTIES
National Institutes of Health	Rebecca Jacobson, M.D., M.S.	1-Jul-87	30-Jun-17	881,840	52,926	Pittsburgh Biomedical Informatics Training Program
National Institutes of Health	Rebecca Jacobson, M.D., M.S.	29-Sep-09	30-Jun-14	11,472	918	Enhancing Research Informatics Capacity of Health Information in Columbia

National Institutes of Health	Rebecca Jacobson, M.D., M.S.	12-Jul-12	30-Jun-14	44,216	22,772	Implemenation of Automated Guideline Adherence Feedback in Malawi
National Institutes of Health	Rebecca Jacobson, M.D., M.S.	1-Oct-06	30-Jun-16	137,363	70,742	Contextual ASR to Support EHR Adoption
National Institutes of Health	Rebecca Jacobson, M.D., M.S.	1-Jul-87	30-Jun-17	1,026,245	63,660	Pittsburgh Biomedical Informatics Training Program
National Institutes of Health	Rebecca Jacobson, M.D., M.S.	25-Sep-13	31-Jul-18	533,958	153,032	Advanced Development of TIES-Enhancing Access to Tissue for Cancer Research
National Institutes of Health	Rebecca Jacobson, M.D., M.S.	1-Jul-87	30-Jun-17	29,781	2,382	Annual Biomedical Informatics Training Conference
National Institutes of Health	Rebecca Jacobson, M.D., M.S.	6-May-14	30-Apr-15	540,337	108,038	Cancer Deep Phenotype Extraction from Electronic Medica
National Institutes of Health	Xia Jiang, Ph.D.	1-Jun-14	31-May-18	323,290	93,471	A New Generation Clinical Decision Support System
National Institutes of Health	Xinghua Lu, M.D., Ph.D.	1-Sep-10	31-Aug-14	194,434	93,566	Statistical Methods for Integromics Discoveries
National Institutes of Health	Xinghua Lu, M.D., Ph.D.	1-Sep-11	31-Aug-15	192,561	99,169	Ontology-Driven Methods for Knowledge Acquisition and Knowledge Discovery
Stony Brook University (NIH)	Xinghua Lu, M.D., Ph.D.	1-May-12	30-Apr-14	10,192	5,248	Modeling and Analysis of Roles of Yeast Spingolipids
County of Sacramento (NIH)	Tsui, Fu-Chiang, Ph.D., M.S.	17-Nov-11	16-Nov-14	2,163	1,170	County of Sacramento CA National Retail Data Monitor Access
National Institutes of Health	Michael Wagner, M.D., Ph.D.	30-Sep-08	29-Sep-13	247,339	127,379	Decision Making in Biosurveillance

National Institutes of Health	Michael Wagner, M.D., Ph.D.	18-Apr-12	31-Mar-16	467,807	116,310	Apollo: Increasing Access and Use of Epidemic Models Through the Development and Adoption of a Standard Ontology
Centers for Disease Control and Prevention	Michael Wagner, M.D., Ph.D.	1-Sep-10	31-Aug-14	66,006	33,994	University of Pittsburgh Center for Advanced Study of Informatics in Public Health
National Institutes of Health	Michael Wagner, M.D., Ph.D.	1-Aug-13	30-Jun-16	461,712	128,311	Probabilistic Disease Surveillance

Entrepreneurial Activities

There are a number of entrepreneurial relationships that faculty in the Department have with outside enterprises, consisting of the following during FY2014:

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

Richard D. Boyce, Ph.D.

Dr. Boyce and his colleagues at Pitt disclosed an invention to the Office of Technology Management titled "Pharmacogenomics Information System" (#03041). The Technology Transfer Committee decided to pursue a copyright for this invention on May 28th 2013.

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

De-ID, Inc.
Co-inventor of software licensed to De-ID by the University of Pittsburgh
2005 - present

Roger Day, Sc.D.

Dr. Day engages with Alung Technologies through their data safety monitoring committee. He also maintains an active music performance and recording schedule.

Vanathi Gopalakrishnan, Ph.D.

Dr. Gopalakrishnan has acquired datasets from Metabolon, Inc., which is interested in joint publications and the use of rule learning technology, developed in her laboratory, towards the analysis of metabolomics data.

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

General Biodefense, LLC
Equity Ownership and Consultant
2004 - present

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 38 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

Frederick Adkins, B.A., M.S., Ph.D. B.A. (1990 Mathematics & Computer Science) Concordia College, M.S. (1992 Mathematics) University of Iowa, M.S. (1994 Mechanical Engineering) University of Iowa, Ph.D. (1996 Applied Mathematics & Computer Science) University of Iowa. His research interests include molecular modeling and the connection to biological information. [Biomedical Informatics Certificate Program]

Rafael Ceschin, B.S. (2009, Microbiology) University of Pittsburgh. His research involves analysis of neuroimaging data to study neonatal development following preterm birth. [Biomedical Informatics Master's Program (part-time)]

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]

Vicky Chen, B.S. M.S. B.S. (2008, Biomedical Engineering and Computer Science) University of Virginia, M.S. (2012, Biomedical Informatics) University of Pittsburgh. Her current research involves building a Python package of a graphic representation of the Gene Ontology that holds semantic, gene, and structural information. [Biomedical Informatics Doctoral Program]

Amie Draper, B.S. (2013 Biomedical Science) Ohio State. 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]

Joyeeta Dutta-Moscato, B.A., B.S.E., M.S. (2000, Biological Basis of Behavior) and B.S.E. (2000, Computer Science & Engineering) University of Pennsylvania. M.S. (2007, Neuroscience) University of Pittsburgh, M.S. (2013, Biomedical Informatics) University of Pittsburgh. Her research interests are in translational bioinformatics and systems biology. [Biomedical Informatics Doctoral Program]

Arielle Fisher, B.S. (2013 Biology) Bucknell University. 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., DDS. BS (1997 Mathematics) University of South Carolina, MSPH (2002 Biostatistics) University of South Carolina, DMD (2006 Dental Medicine) Medical University of South Carolina. His research interests are undetermined. [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. BS (2009 Electrical Engineering), Pontificia University, Colombia. His research interests include global health informatics.: [Biomedical Informatics Master's Program]

JoAnna Hillman, B.S., M.PH. BS (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]

Rick Jordan, B.S., M.S., M.S. B.S. (1996, Biology) University of Pittsburgh, Johnstown, PA. M.S. (2001, Molecular Biology & Biotechnology) East Carolina University. M.S. (2005; Biomedical Informatics) University of Pittsburgh. He is focused on his doctoral dissertation research on text mining for informative prior construction in omic data analysis. [Biomedical Informatics Doctoral Program (part-time)]

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

Zachary Landis Lewis, B.A., M.L.I.S., M.S. B.A. (2000, History) Goshen College. M.L.I.S. (2004, Library & Information Science) and M.S. (2010, Biomedical Informatics) University of Pittsburgh. His research involves the feasibility of automating audit and feedback for ART Guideline Adherence in Malawi. [Biomedical Informatics Doctoral Program]

Arturo Lopez Pineda, B.S., M.S., M.S. B.S.(2006, Computer Science), M.S. (2008, Intelligent Systems) Tecnologico de Monterrey, Mexico, and M.S. (2012 Biomedical Informatics) University of Pittsburgh. His research interests include monitoring influenza-like illness by mapping twitter messages contents. [Biomedical Informatics Doctoral Program]

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]

Danielle Mowery, B.S., M.S. B.S. (2006, Biological Sciences) and M.S. (2008, Health Information Systems; 2010, Biomedical Informatics) University of Pittsburgh. Her research involves automatic SOAP classification in emergency department reports. [Biomedical Informatics Doctoral Program]

Soyapi Mumba, B.S. BS (2011 Computer Science) University of Hertfordshire, United Kingdom. His research interests include biosurveillance for disease outbreak. [Biomedical Informatics Master's Program]

Henry Ogoe, B.S., M.S. (2007, Computer Science) Åbo Akademi University, Finland
His research involves data mining and pattern recognition in biomedicine.
[Biomedical Informatics Doctoral Program]

Richard Oldham, B.A., D.D.S. B.S.(2006, Economics) University of Virginia and D.D.S.(2011, Dental Surgery) Virginia Commonwealth University. Research involves the design and evaluation of tablet-based dental treatment-planning software in clinical use. [Biomedical Informatics Master's Program]

Katrina Romagnoli, B.S., M.L.I.S., M.S. B.S. (2008, Molecular Biology and English Literature), M.L.I.S. (2009, Library & Information Science), and M.S. (2012, Biomedical Informatics) University of Pittsburgh. Her research involves understanding post-hospitalization information needs of elderly patients: an evaluation of the experience of patients, their caregivers, and home care nurses, and the relationship to hospital readmissions. [Biomedical Informatics Doctoral Program]

Victor Ruiz Herrera, BS, BS (2011 Electrical Engineering) Pontificia University Javeriana, Colombia. His research interests are in natural language processing foundations.
[Biomedical Informatics Master's Program]

Reza Sadeghian, B.S., M.D., M.B.A. BSc (2003 Biology) Azad University, Iran,
MD (2010 Medicine) St. Georges University, MBA (2012 Business Administration) Auburn University, Master's Student, Department of Biomedical Informatics. His research interest utilizing telehealth and technologies in the reducing avoidable hospitalizations among nursing facilities residents.[Biomedical Informatics Master's Program]

Lucas Santana dos Santos, B.Sc, M.S. B.Sc. (2008, Biological Science) Univeridade Federal De Minas Gerais, 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]

Corey Stein, B.S. B.S. **(**2010, Biology) William Paterson University of New Jersey. His research involves utilizing mobile technology to enhance patient access to care and emergency management in dentistry. [Biomedical Informatics Master's Program]

Matthew E. Stokes, B.S., M.S. B.S. (2008, Systems and Control Engineering) Case Western Reserve University, M.S. (2011, Intelligent Systems Program) University of Pittsburgh. His research involves the development of machine learning algorithms to handle large-scale data problems, including searching the human genome for genetic variants associated with disease. [Intelligent Systems Program / Biomedical Informatics Doctoral Program]

Eric Strobl, B.A. (2009 Psychology and Biology) University of California at Berkeley. His research interests include kernel-based Markov blanket discovery.
[Biomedical Informatics Doctoral Program]

Wong, An-Kwok Ian, B.S., M.S. B.S. (2007, Engineering) Duke University, Durham, NC, MS (2009, Intelligent Systems) University of Pittsburgh, Pittsburgh, PA. His research interests 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 & Epidemiosixteenlogy) 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]

Certificate Awardees

None for FY14

Departing Fellows

Danielle Mowery, B.S., M.S., Ph.D. has accepted a postdoctoral research position at the Department of Biomedical Informatics at the University of Utah, Salt Lake City, Utah (August 2014)

Soyapi Mumba, B.S., M.S. is the Acting Director of Informatics at Baobab Health Trust in Malawi, Africa.
(April 2014)

Reza Sadeghian, B.S., M.D., M.B.A., M.S. began a residency program at the University of Southern Alabama in July 2014 (April 2014)

Corey Stein, B.S., M.S. was accepted into dental school at The College of Dental Medicine at the Western University of Health Sciences in Pomona, California (December 2013)

Matthew E. Stokes, B.S., M.S., Ph.D. (graduate of the Intelligent System Program with a concentration in Biomedical Informatics) has accepted a position as a Computational Biology Scientist for Celgene, at the Celgene Institute for Translational Research Europe (CITRE) in Sevilla, Spain. (August 2014)

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.

Ye Ye, B.S., M.S., M.S.P.H. B.S. (2006, Preventive Medicine) and M.S. (2009, Biostatistics & Epidemiosixteenlogy) 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. She is currently working with Dr. Michael Wagner.

Post-Docs

John Aronis, B.A., M.S., Ph.D. BS (Mathematics) State College of New York and and MS (Mathematics) Syracuse University; Ph.D. (Intelligent Systems) University of Pittsburgh. His research interests include application to data mining to the detection of anomalous cases and contagious outbreaks.

Viji Avali, Ph.D. BS (1987 Electrical Engineering) Madurai University, India; MS (1992 Process Monitoring and Control) University of Houston; ME (1996 Electrical and Computer Engineering) University of South Carolina; and PhD (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.

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

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 presently works for Dr. Xia Jiang.

Huichen Feng, Ph.D. BS (1995 Biology) Northeast Agriculture University, China; MS (2000 Medical Genetics) Harbin Medical University, China; and PhD (2005 Cancer Biology) University of Hong Kong. His research interests include transcriptomic and proteomic approaches to discovering viral causes for human hematologic malignant cancers.

Elizabeth Leslie, Ph.D. BS (2008 Chemistry) St. Olaf College, and PhD (2012 Genetics) University of Iowa. Her research interests include Identifying genetic risk factors for nonsyndromic orofacial clefts using massively parallel sequencing and subclinical phenotyping.

Songjian Lu, Ph.D.

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

Departing Post-Docs

Arthur Ward, Ph.D.

B.S., Managerial Economics (1985), Carnegie-Mellon University; MBA, Business Administration (1995),Carnegie-Mellon University; Ph.D., Intelligent Systems (2010), University of Pittsburgh. Dr. Ward worked with Dr. Rebecca Jacobson.

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:

Roger S. Day, Sc.D.
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.
Gerry Douglas, Ph.D.
Madhavi Ganapathiraju, Ph.D.
Steven Handler, M.D., Ph.D.
Harry Hochheiser, Ph.D.
Xia Jiang, Ph.D.
Shyam Visweswaran, M.D., Ph.D.

Research Assistant Professor:

Fu Chiang (Rich) Tsui, Ph.D.

Secondary Appointment Faculty

Professor:

Barton Branstetter, M.D., *Radiology*
Rajiv Dhir, M.D., *Pathology*
Louis Leff, M.D., *Medicine*
Daniel Martich, M.D., *Critical Care Medicine*
Anil Parwani, M.D., Ph.D., *Pathology*

Associate Professor:

Colleen Culley, Pharm D, *Pharmacy*
Liron Pantanowitz, M.D., *Pathology*

Clinical Associate Professor:

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

Assistant Professor:

Ashi Daftary, M.D., *OB/GYN & Reproductive Sciences*
Julia Driessen, Ph.D., *GSPH*
Hiroshi Ishikawa, M.D., *Ophthalmology*
Bruce Lee, M.D., M.B.A., *Medicine*
Ervin Sejdic, Ph.D., *Electrical Engineering*
Heiko Spallek, D.M.D., Ph.D., *Dental Medicine*

Clinical Assistant Professor:

Steven Hasley, M.D., *OB/GYN & Reproductive Sciences*

Adjunct Faculty

Clinical Professor:

Bruce Block, M.D., *Family Medicine*

Clinical Associate Professor:

Stephen Corey, M.D., *OB/GYN & Reproductive Sciences*

Associate Professor:

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

Assistant Professor:

Richard Ambrosino, M.D., *UPMC ISD*
Christa Bartos, Ph.D., *UPMC IS*
Anne Ben-Smith, Ph.D., *Malawi*
Brian Chapman, Ph.D., *University of San Diego*
Stephen Corey, M.D., *OB, UPMC Healthplan*
Denver Dash, Ph.D., *Intel Corporation*
Joel Diamond, M.D., *UPMC ISD*
Dan Drawbaugh, *UPMC ISD*
Artur Dubrawski, Ph.D., *Carnegie Mellon University*
William Fera, M.D., *UPMC ISD*
Cynthia Gadd, Ph.D., *Vanderbilt University*
William Hogan, M.D., M.S., *University of Arkansas*
John Houston, J.D. *UPMC ISD*
Valerie Monaco, Ph.D., M.H.C.I., *UPMC Health Plan*
Daniel Neill, Ph.D., *Carnegie Mellon University*
Rema Padman, Ph.D., *Carnegie Mellon University*
Adam Rothschild, M.D., *Private Practice*
Bertha Simwaka, Ph.D., *Malawi*

Thankam Thyvalikakath, M.D.S., Ph.D., *Indiana University*
Claudia Mello-Thoms, Ph.D., University of Australia
Darcy Waechter, R.N., M.S.N., *ISD*
Garrick Wallstrom, Ph.D., *University of Arizona*

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 Instittute, 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 it 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

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 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 Effect* and the *Journal of Evidence-based Dental Practice (JEBDP*). Her most recent publications appear in the *JEBDP, Journal of Medical Internet Research, Artificial Intelligence in Medicine, and the Journal of the Medical Library Association*. She is a member of the Phi Kappa Phi Honor Society and 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.* Recently, she was admitted to the Academy of Health Information Professionals (AHIP) at the senior level.

Students Participating in Research:

Dr. Bekhuis works with Mr. Corey Stein, a biomedical informatics master's student. She taught him how to re-create complex database queries appearing in published systematic reviews of medical evidence. She also taught him how to conduct various analyses for a Cochrane review, including a meta-analysis of reported outcomes for tumor recurrence. Dr. Bekhuis also mentors Dr. John Frazier, a doctoral student in DBMI who is an oral pathologist. They are developing a gold standard dataset of studies on prognostic biomarkers of oral squamous cell carcinoma in preparation for information retrieval studies.

Editorships, Honors, Awards, and Major Lectureships:

Editorial Board Member. *Journal of the Medical Library Association (JMLA)* 2008 to 2011
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

Richard D. Boyce, Ph.D. is an Assistant Professor of Biomedical Informatics in the University of Pittsburgh School of Medicine and Faculty in the Geriatric Pharmaceutical Outcomes and Gero-Informatics Research and Training Program. He received his Ph.D. in 2008 from the University of Washington program in Biomedical and Health Informatics completed a two-year postdoctoral fellowship in biomedical informatics at the University of Pittsburgh in 2010, and completed 3-years of career development training in Comparative Effectiveness Research in 2013. His research focuses on interventions that improve medication safety for nursing home residents, examines the effectiveness and safety of newer psychotropics in elderly nursing home residents, and creates new knowledge-based approaches to drug-drug interaction prediction and identification and drug safety decision support.

Study Sections and Advisory Committee Memberships:

Dr. Boyce was a Merit Reviewer the Patient Centered Outcomes Research Institute (Cycle III, Improving Health Systems). 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. Jeremy Jao – undergraduate in the Pitt Bioinformatics program
Mr. Yifan Ning – Masters student in the Pitt Information Science program

Editorships, Honors, Awards, and Major Lectureships:

Editorships:

Marshall MS., and Boyce RD, (Eds). Health Care and Life Science (HCLS) Linked Data Guide. World Wide Web Consortium (W3C). November 2012. http://www.w3.org/2001/sw/hcls/notes/hcls-rdf-guide/. Last Accessed 11/26/2012.

Rodríguez-González, A., Pathak, J., Wilkinson, MD., Shah, NH., Stevens, R., Boyce, RD., García-Crespo, Á., "Proceedings of the Joint Workshop on Semantic Technologies Applied to Biomedical Informatics and Individualized Medicine (SATBI+SWIM 2012)," Held at the 11th International Semantic Web Conference (ISWC 2012), http://ceur-ws.org/Vol-930/. November 12th, Boston, USA.

Honors:

Distinguished Paper Award
AMIA Summit on Translational Bioinformatics, April 2013

Major Lectureships:

University of Wisconsin – Milwaukee, Center for Biomedical Data and Language Processing (BioDLP) Seminar Series, Visiting lecture "Addressing Gaps in Clinically Useful Evidence on Drug-Drug Interactions", Milwaukee, WI, May 2nd 2013.

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.

Roger S. Day, Sc.D.

Associate Professor of Biomedical Informatics and Biostatistics

Roger Day, Sc.D., is an Associate Professor in the Department of Biomedical Informatics, University of Pittsburgh School of Medicine, with a secondary faculty appointment in the Graduate School of Public Health, Department of Biostatistics, and Affiliated Faculty at the University of Pittsburgh McGowan Institute and the University of Pittsburgh Molecular Medicine Institute. He received his Sc.D. degree in Biostatistics from Harvard University and carried out postgraduate work at the National Institute for Environmental Health Sciences and the Keystone Conference on the Histopathology of Cancer. He pursued research at Harvard University and the New England Journal of Medicine Project prior to joining the University of Pittsburgh Cancer Institute in 1986, as Director of the University of Pittsburgh Biostatistics Department. Currently, Dr. Day is Co-Investigator/Biostatistician on several grants including the UPCI Lung Spore. He leads the Educational Resource for Tumor Heterogeneity and continues the evolution of the Oncology Thinking Cap. He is author of over 60 publications in Biostatistics and Bioinformatics. His research has been funded by the National Institutes of Health.

Students Participating in Research:

Proteomic-to-Genomic ID Mapping
Kevin McDade, M.S. candidate, DBMI, University of Pittsburgh
2008 – 2014

Gerald P. Douglas Ph.D.

Assistant Professor of Biomedical Informatics
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. Dr. Douglas sits on the advisory council of Inveneo Inc., and is member of the inaugural group of Technology, Entertainment and Design (TED) Fellows (2009).

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:

Soyapi Mumba, Masters Student, DBMI, University of Pittsburgh
Biosurveillance using routinely-collected outpatient data in Malawi
2012 – 2014

Zachary Landis-Lewis, PhD Student, DBMI, University of Pittsburgh
Implementation of Automated Guideline Adherence Feedback in Malawi
2011 – 2014

JoAnna Hillman, PhD Student, DBMI, University of Pittsburgh
Challenges Facing Hospital Laboratories in Low-and Middle-Income Countries: a Protocol for a Scoping Review
2013 – present

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 Laboratory Turnaround Time at Kamuzu Central Hospital in Malawi: Challenges Encountered while Testing the Study Protocol
2014 – 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:

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

Students Participating in Research:

Graduate Students

- Lavanya Viswanathan, B.Engg., Advisor
- Suyoun Kim, B.S., M.S., Advisor
- Adam Handen, B.S., M.S., Advisor
- Adam Roth, B.S. Advisor
- Haohan Wang, B.S. Advisor

Undergraduate Students - Research Supervision:

2014 Spring	Daniel Hui	Undergrad	Pitt First Experiences in Research
	Maysum Chaudhri	Undergrad	Pitt First Experiences in Research
	Molly Johnson	Undergrad	Pitt First Experiences in Research

	Vasil Mico	Undergrad	Pitt First Experiences in Research
	Anis Adnani	Undergrad	Pitt First Experiences in Research
2014 Summer	Helen Li	Undergrad	Univ. of California at Santa Cruz

High School Students - Research Supervision:

2013 Summer	Sophia Cheng	High School	CoSBBI Summer Research
	Tonya Hammond	High School	CoSBBI Summer Research
	Ritwik Gupta	High School	CoSBBI Summer Research
2014 Summer	Thomas Nash	High School	CoSBBI Summer Research

Vanathi Gopalakrishnan, Ph.D.

Associate Professor of Biomedical Informatics, Intelligent Systems, and 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

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

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

Provisional patent application filing: Zaidi AH, Jobe BA, Zeng X, Balasubramanian JB, Gopalakrishnan V, Bigbee WL, inventors; SERUM BIOMARKER PANEL FOR THE DETECTION OF ESOPHAGEAL ADENOCARCINOMA. United States Provisional Patent Application No. 61/922,665. 2013, Dec 31.

Awards to mentored student:

Jessica Larusch – Best Student Paper, 3rd Place, DBMI Student Retreat, Pittsburgh PA 2012.

Lectureships:

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.
Stanford University, Palo Alto, CA: "Mining Big Biomedical Data for Actionable Knowledge", May 30, 2012.

Mining "Big" Biomedical Data for Actionable Knowledge, Department of Computer Science, April 10, 2012

"Classification Rules Modeling for Understanding Biomedical Data", Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, April 6, 2012.

"Data Mining for Predictive Biomedicine". University of California, San Diego, CA. March 26, 2012.

Steven Handler, M.D., Ph.D.

Assistant Professor in the Department of Biomedical Informatics, Division of Geriatric Medicine, and Clinical and Translational Research, Director of Clinical Informatics, at the University of Pittsburgh
Core Faculty of the RAND-University of Pittsburgh Health Institute (RUPHI)
Medical Director, Long-term Care, Health Information Technology, UPMC
Steven Handler, M.D., Ph.D., Dr. Handler's primary appointment is Assistant Professor in the School of Medicine, Department of Biomedical Informatics (DBMI), where he is the Director of Clinical Informatics. He has secondary appointments in Geriatric Medicine and Clinical and Translational Research. He is also Core Faculty of the RAND-University of Pittsburgh Health Institute (RUPHI) and Medical Director of Long-term Care, Health Information Technology, UPMC. Dr. Handler is a board-certified Family Physician who has completed his fellowship in Geriatric Medicine and continues to practice clinical medicine primarily in the long-term care setting.

Dr. Handler received his M.D. from New Jersey Medical School in 1997, completed a Family Medicine Residency at Tufts University in 2000, and completed a Geriatric Medicine fellowship in 2002 at the University of Pittsburgh. He completed a M.S. in Biomedical Informatics with concentration in health services research in 2004 and a Ph.D. in Biomedical Informatics in 2010. Following his fellowship in Geriatric Medicine, he remained at the University of Pittsburgh and accepted a position as a Research Instructor in 2002, transitioned to a Post-doctoral Scholar when a new T32 was awarded to Geriatric Medicine in 2004, and then accepted a position as an Assistant Professor upon receipt of a CTSA KL2 career development award in 2005. In 2008, he changed my primary appointment from Geriatric Medicine to Biomedical Informatics as his research focus evolved.

As an Assistant Professor in Biomedical Informatics, Dr. Handler is actively involved in research and teaching. His salary is primarily supported by research grants and contracts. He was awarded a four-year R01 from the Agency for Healthcare Research and Quality (AHRQ) to conduct a randomized controlled trial to assess the impact of an active medication monitoring system on the detection and management of adverse drug events in the nursing home setting. In addition to this funding source, He is also the Co-Director of a National Institute of Aging K07 (awarded to Joseph Hanlon, PharmD, MS) to fund the University of Pittsburgh Program for Pharmaceutical Outcomes Research in Aging.

In 2011, Dr. Handler was named Director of Clinical Informatics, Department of Biomedical Informatics. In this role, Dr. Handler will act as the point-person and interface between the Biomedical Informatics faculty and UPMC clinicians and researchers interested in conducting high-quality informatics-based research. His initial focus will be on translational informatics projects that can improve the quality of patient care in a variety of clinical settings.

Dr. Handler currently teaches Introduction to Patient Care and Clinical Environments. This is an elective course that he has been taught since 2008 and is designed for students who have no significant clinical experience with the U.S. healthcare system. In this course, students learn about the different clinical environments-such as the emergency room or hospital setting-by “shadowing” physicians during a typical work-day. This unique approach allows a first-hand experience of clinical settings that would otherwise be difficult to arrange. Students are also introduced to medical terminology through an interactive lecture format. Finally, students are provided with a set of tools and techniques derived from the Toyota Production System (TPS) to perform current condition assessments and develop countermeasures to problems noted during their clinical observations. This course has consistently received extremely high ratings and has been taken by students throughout the health sciences as well as from Carnegie Mellon University.

Dr. Handler has mentored pre-doctoral Biomedical Informatics students, medical students, M.S.T.P. students, post-doctoral Biomedical Informatics students, clinical fellows, and junior faculty since 2003: 9 graduate students in the Department of Biomedical Informatics (DBMI) Training Program, serving as primary research advisor for 3 of these students; 4 medical students, 3 internal medicine residents, 2 students from McMaster University, and 1 from the University of Massachusetts at Worcester. In addition, he has mentored 1 post-doctoral scholar and 4 junior faculty in biomedical informatics, geriatric medicine, and general internal medicine. Dr. Handler has a strong track record of publishing with trainees and junior faculty, which has led to 19 co-authored peer-reviewed papers. In addition, he has provided guidance and assistance leading to Dr. Richard Boyce (Assistant Professor in DBMI) receiving a K12 career development award, and Dr. Zachary Marcum (Assistant Professor in the Division of Geriatric Medicine) receiving two consecutive loan repayment awards (L30). With regard to student committee service, Dr. Handler has served on 2 doctoral preliminary examination committees and 6 masters committees, on 2 of which he was Chair of the Committee.

Dr. Handler has published over 68 peer-reviewed articles focusing on medication and patient safety primarily in the nursing home setting. He is frequently invited to speak to health professional groups locally and nationally. He is an active member of a number of professional organizations including AMIA, AMDA, AGS, ASCP, GSA, and Society for Clinical and Translational Science.

Study Sections and Advisory Committee Memberships:

Member of the Electronic Health Record Physician Advisory Council, UPMC Health System, 2004 – present.

Member of the Long-Term Care Research Network Steering Committee, American Medical Directors Association, 2006 - present

Health Level Seven (HL7) Long-term Care EHR-S Functional Profile Workgroup, 2007 – present

Health Level Seven (HL7) Long-term Care EHR-S Functional Profile Direct Care Task Group, 2007-present

Member of the Health Services Executive Committee, Gerontology Society of America, 2007 – present.

Member of the of the System Patient Safety and Quality Committee, UPMC Health System, 2007 – present.

Member of the Clinical Intake/Transfer of Care Team, UPMC Health System, 2008 – present.
Member of the Health Information Technology (HIT) Subcommittee, American Medical Directors Association, 2008 – present.

Physician Chair of the Unplanned Readmissions from Skilled Nursing Homes to Hospitals Work Group, UPMC Heath System and University of Pittsburgh Aging Institute, 2010 – present.

Member of the Health System Palliative Care Work Group, UPMC Heath System and University of Pittsburgh Aging Institute, 2011 – present.

Telehealth Advisory Committee, UPMC Health System, 2012 – present.

Beckwith Institute Clinical Transformation Program, UPMC Health System, 2012 – present.

Member of the Health Systems Innovation – Economics and Technology Committee, American Geriatrics Society, 2012-2015.

Member of the Advisory Committee for the Center for Assistance in Research using eRecord (CARe), University of Pittsburgh, UPMC, 2012 – present.

Member of the Enterprise Analytics Program Pharmacy Use Case, UPMC Health System, 2013 – present.

Physician and Patient Safety Expert Panelist, Centers for Medicare and Medicaid Services (CMS) F-329 (Unnecessary Drugs) Guidance for State Surveyors Redesign Committee
2012 – present

Key Informant for a systematic literature review being conducted on "Medication Therapy Management", Agency for Healthcare Research and Quality, 2013.

Students Participating in Research:

Biomedical Informatics:

Katrina Romagnoli, Doctoral Student
Content Mentor and Research Supervisor
2011 - 2014

Zachary Landis-Lewis, Doctoral Student
Content Mentor and Research Supervisor
2011 – 2014

Corey D. Stein, BS
Masters Committee Member
Masters Program, Dental School
2012 – 2013

Reza Sadeghian, MD, MBA, Masters Fellow, NLM Training Program
Research Advisor

Masters Fellow, NLM Training Program
2012 – 2014

Dio Kavalieratos, PhD,
Research Advisor
RAND-Pitt Postdoctoral Fellow
Health Services Research
2012 - present

External:

Pavel S. Roshanov
Mentor and external MSc Committee Examiner
McMasters University, Health Research Methodology, Ontario, Canada
2010 – 2013

Imaan Bayoumi, MD
Mentor and external MSc Committee Examiner
McMasters University, Health Research Methodology, Ontario, Canada
2011 – 2013

Alyssa McKenna
Research Advisor
Second Year
Lake Erie College of Osteopathic Medicine
2013 - Present

Editorships, Honors, Awards, and Major Lectureships:

Expert Reviewer of the Omnicare Geriatric Pharmaceutical Care Guidelines,
Representative from the American Geriatrics Society
2006 – present

Westat Research Corporation, Technical Expert for Developing a Patient Safety Culture Instrument for Nursing Home Use
2006 – present

Public Resource:

Developed and maintain publically the trigger tool for measuring adverse drug events in the nursing home (http://www.ihi.org/ihi/topics/patientsafety/medicationsystems/tools/triggertoolformeasuringadesinnursing+home.htm
Institute For Healthcare Improvement
2009 - Present

Lectureships:

1. Pennsylvania Medical Directors Association Annual Symposium, "The Nuts and Bolts of INTERACT II," Hershey, PA, October, 2012.
2. Clinical Update in Geriatric Medicine Annual Symposium, "Medication Adherence in the Elderly: Diagnosis & Management," Pittsburgh, PA, April 2013.
3. **Handler SM**. Evaluating the Impact of Computer-Generated Rounding Reports on Physician Workflow in the Nursing Home: A Feasibility Time-Motion Study. Symposium presented at the American Medical Directors Association Annual Symposium in National Harbor, MD, April, 2013.
4. Ouslander JG, **Handler SM**, Crecelius CA, Hays C, Nazir A, Phillips S. Improving Care by Reducing Avoidable Hospitalizations: Implementing the INTERACT Program in Your Facility. Symposium presented at the American Medical Directors Association Annual Symposium in National Harbor, MD, April, 2013.
5. Shulman E, Alexander GL, Agarwal K, **Handler SM**. Long-Term and Post-Acute Care Health IT Summit, "**R**educe **Av**oidable hospitalizations using **E**vidence-based interventions for **N**ursing facilities (**RAVEN**) in western Pennsylvania: a Focus on Telemedicine," Baltimore, MD, June, 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 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 an Associate 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 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 intelligent medical training 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 60 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

AMIA Annual Symposium Scientific Program Committee
2009 - 2010

International Artificial Intelligence in Education Society
Member
2000 - 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

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

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, Dr. Jiang received her 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, Dr. Jiang received her second M.S. from Northeastern Illinois University, where she first became involved in Bayesian network related research. Dr. Jiang received her Ph.D. from the University of Pittsburgh in 2008. At Pitt, she did her graduate work in the laboratory of Dr. Gregory Cooper. She then did a postdoctoral fellowship supported by NLM/ NIH for over three years. Dr. Jiang was a graduate level course instructor in the Computer Science Department at Northeastern Illinois University for three years before she came to Pitt. She is the author of 20 peer reviewed articles as well as the co-author of two books. Her first book titled Probabilistic Methods for Financial and Marketing Informatics was published by Morgan Kaufmann in April 2007. Her second book titled Contemporary Artificial Intelligence will be published by Chapman & Hall/CRC in August of this year. Dr. Jiang is 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 Dr. Jiang's 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. Dr. Jiang is also very interested in translational informatics. She will devote her efforts in developing advanced informatics tools that assist the translation of the findings in basic scientific research efficiently and effectively into patient medical care. Dr. Jiang's 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, Dr. Jiang has two goals: 1) she will devote her efforts to extend her collaborations with colleagues in other programs/departments at Pitt and other universities, in the hope of fostering interdisciplinary education and research, and 2) she will strive to build an amiable, collaborative, and productive microenvironment at her research lab which will focus on both research and education and training. She is willing to interact actively with students at all levels and across a variety of research domains.

Students Participating in Research:

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

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

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

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

Editorships, Honors, Awards, and Major Lectureships:

Editorships

Guest Editor, Cancer Informatics Supplement (2013-2014)

Awards

2010, National Library of Medicine postdoctoral training scholarship.

2010, Finalist Award, Student paper competition at American Medical Informatics Association (AMIA) 2010 Annual Symposium for paper titled: "A Fast Algorithm for Learning Epistatic Genomic Relationships", in proceedings of American Medical Informatics Association 2010.

Manuscript Reviews

2011 Reviewer, PLoS
2011 Reviewer, BMMC Bioinformatics
2012 Reviewer, Genetic Epidemiology
2013 Reviewer, SpringerPlus
2013 Reviewer, Bayesian Analysis
2013 Reviewer, Journal of Biomedical Informatics
2013 Reviewer, AMIA 2013 Annual Symposium
2013 Reviewer, IEEE Intelligent Systems

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
2008 - 2012

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 Sphingoloipids 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

Nuclear receptors and their co-regulators in cancers
Hatice Osmanbeyoglu, Ph.D. student, Department of Biomedical Informatics, University of Pittsburgh
2011 – 2012

Editorships, Honors, Awards, and Major Lectureships:

BMC Research Notes
Associate Editor
2007 - present

Open Systems Biology
Editorial Board
2009 – present

DNA Repair
Editorial Board
2013-present

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

Associate Director, RODS Laboratory
Research Assistant Professor of Biomedical Informatics and Intelligent Systems

Fu-Chiang (Rich) Tsui, Ph.D., is Research Assistant Professor of Biomedical Informatics and Intelligent Systems. He earned his Ph.D. in electrical engineering from the University of Pittsburgh and completed postdoctoral studies in biomedical informatics in Pitt's Center for Biomedical Informatics. Dr. Tsui is the Associate Director of the Real-time Outbreak and Disease Surveillance (RODS) Laboratory, which he co-founded, and was instrumental in the development of the widely used RODS public health surveillance open source system. Dr. Tsui serves as a PI on four grants and two sub-award grants, and serves as a Co-Investigator several grants; those grants are funded by the CDC, the NLM and state/local departments of health. He has authored 40 peer-reviewed articles on the subjects of medical informatics and biosurveillance.

Study Sections and Advisory Committee Memberships:

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

Students Participating in Research:

Wavelet-based Spatial Clustering Algorithm
Jialan Que, Ph.D. ISP, University of Pittsburgh
2006 – 2012

Evaluation of case detection system
Ye Ye, PhD student, ISP, University of Pittsburgh
2010 – present

Detection of patients with influenza syndrome using machine learning models learned from ED reports
Arturo López Pineda, PhD student, DBMI, University of Pittsburgh
2010 – present

Using a distributed search engine to identify optimal product sets for use in an outbreak
detection system
Ruhsary Rexit, MS student, CS, University of Pittsburgh
2011 – present

Editorships, Honors, Awards, and Major Lectureships:

Advances in Disease Surveillance
Associate Editor
2006 - present

American Medical Informatics Association
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

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.

Park S, Parwani AV, Aller RD, Banach L, **Becich MJ**, Borkenfeld S, Carter AB, Friedman BA, Rojo MG, Georgiou A, Kayser G, Kayser K, Legg M, Naugler C, Sawai T, Weiner H, Winsten D, Pantanowitz L. The history of pathology informatics: A global perspective. *Journal of Pathology Informatics*. 2013 May 30;4:7. PMCID: PMC3714902

Lee RE, McClintock DS, Balis UJ, Baron JM**, Becich MJ**, Beckwith BA, Brodsky VB, Carter AB, Dighe AS, Haghighi M, Hipp JD, Henricks WH, Kim JY, Klepseis VE, Kuo FC, Lane WJ, Levy BP, Onozato ML, Park SL, Sinard JH, Tuthill MJ, Gilbertson JR. Pathology informatics fellowship retreats: The use of interactive scenarios and case studies as pathology informatics teaching tools. *Journal of Pathology Informatics*. 2012;3:41. Epub 2012 Nov 28. PMCID: PMC3519095

Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics. *Journal of Pathology Informatics*. 2012;3:40. Epub 2012 Oct 31. PMCID: PMC3519097

Bartholow TL, **Becich MJ**, Chandran UR, Parwani AV. Immunohistochemical staining of slit2 in primary and metastatic prostatic adenocarcinoma. *Translational oncology*. 2011 Oct; 4 (5):314-20. PMCID: PMC3162306

Jiang X, Barmada MM, **Becich MJ**. Evaluating De Novo Locus-Disease Discoveries in GWAS Using the Signal-to-Noise Ratio. *AMIA Annual Symposium*; 2011 Oct 22. 617-24.

Jiang X, Barmada MM, Cooper GF, **Becich MJ**. A bayesian method for evaluating and discovering disease loci associations. *PLoS One*. 2011;6(8):e22075. Epub 2011 Aug 10 PMCID: PMC3154195

Bartholow TL, Chandran UR, **Becich MJ**, Parwani AV. Immunohistochemical profiles of claudin-3 in primary and mestastatic prostatic adenocarcinoma. *Diagnostic Pathology*, 2011 Jan 21;6(1):12. PMCID: PMC 3033791

Bartholow TL, Chandran UR, **Becich MJ**, Parwani AV. Immunohistochemical staining of radixin and moesin in prostatic adenocarcinoma. *BMC Clinical Pathology,* 2011 Jan 14;11(1):1. PMCID: PMC 3029218

Batholow TL, **Becich MJ**, Chandran UR, Parwani AV. Immunohistochemical Analysis of ezrin-radixin-moesin-binding phosphoprotein 50 in prostatic adenocarcinoma. *BMC Urology,* 2011 Jun 14;11(1):12. PMCID: PMC3132203

Maxwell GL, Hood BL, Day R, Chandran U, Kirchner D, Kolli VS, Bateman NW, Allard J, Miller C, Sun M, Flint MS, Zahn C, Oliver J, Banerjee S, Litzi T, Parwani A, Sandburg G, Rose S, **Becich MJ**, Berchuck A, Kohn E, Risinger JI, Conrads TP. Proteomic analysis of stage I endometrial cancer tissue: Identification of proteins associated with oxidative processes and inflammation. *Gynecologic Oncology*, 2011 Jun 1;121(3):586-94. [Epub 2011 Apr 1]. PMID: 21458040 PMC: not applicable

Tanja Bekhuis, PhD, MS, MLIS, AHIP

Bekhuis T, Demner-Fushman D. Screening nonrandomized studies for medical systematic reviews: A comparative study of classifiers. Artif Intell Med. 2012 Jul;55(3):197-207. Epub 2012 Jun 5. PMCID: PMC3393813. [Featured in the AHRQ SRC Methods Library Article Alert, November 2012]

Bekhuis T, Demner-Fushman D, Crowley RS. Comparative effectiveness research designs in MeSH and Emtree: an evaluation of coverage. JMLA: Journal of the Medical Library Association. 2013;101(2):92-100. PMCID: PMC3634392. [Featured in the AHRQ SRC Methods Library Article Alert, May 2013]

Bekhuis T, Demner-Fushman D, Crowley RS. Comparative effectiveness research designs in MeSH and Emtree: an evaluation of coverage. JMLA: Journal of the Medical Library Association. 2013;101(2):92-100. PMCID: PMC3634392. [Featured in the AHRQ Scientific Resource Center Methods Library, May 2013]. Impact factor 0.976.

Song M, O'Donnell JA, **Bekhuis T**, Spallek H. Are dentists interested in the oral-systemic disease connection? A qualitative study of an online community of 450 practitioners. BMC Oral Health. 2013 Nov 21;13(1):65. [Epub ahead of print] PMID: 24261423. PMCID: PMC3924341. [Highly accessed article; Also, featured in British Dental Journal. 2013; 215(11): 545, News section, online: 6 December 2013. doi:10.1038/sj.bdj.2013.1152]. Impact factor 1.339.

Bekhuis T, Tseytlin E, Mitchell K, Demner-Fushman D. Feature engineering and a proposed decision-support system for systematic reviewers of medical evidence. PLoS One. Epub 2014 Jan 27; 9(1):e86277. doi:10.1371/journal.pone.0086277. PMCID: PMC3903545. [Featured in the AHRQ Scientific Resource Center Methods Library, March 2014]. Impact factor 3.730.

Bekhuis T. Mentors and tormentors on the road to informatics [invited editorial]. J Med Lib Assoc 102(2), 2014 Apr;102(2):67-8. doi: 10.3163/1536-5050.102.2.001. PMCID: PMC3988775. Impact factor 0.976.

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

APA PsycEXTRA database]. American Psychological Association (APA). 2014. doi: 10.1037/e502572014-001.

Bekhuis T, Kim K, Valappil B, Spallek H. PRIME Fissure Sealant Questionnaire—US Version [questionnaire in APA PsycTESTS database]. American Psychological Association (APA). 2014. doi: 10.1037/t27847-000.

Richard D. Boyce, Ph.D.

Boyce RD, Collins C, Clayton M, Kloke J, Horn J. Inhibitory Metabolic Drug Interactions with Newer Psychotropic Drugs: Inclusion in Package Inserts and Influences of Concurrence in Drug Interaction Screening Software. Annals of Pharmacotherapy. 2012 Oct;46(10):1287-98. Epub 2012 Oct 2. DOI 10.1345/aph.1R150. PMID 23032655. (NIH Public Access Policy not applicable because the study was not NIH funded)

Marshall MS, **Boyce RD,** Deus H, Zhao J, Willighagen E, Samwald M, Pichler E, Hajagos J, Prud'hommeaux E, and Stephens, S. Emerging practices for mapping life sciences data to RDF - a case series. Journal of Web Semantics. Special Issue: Reasoning with Context in the Semantic Web. Volume 14, July 2012, Pages 2–13. DOI 10.1016/j.websem.2012.02.003. (NIH Public Access Policy not applicable because the study was not NIH funded)

Boyce RD, Horn JR, Hassanzadeh O, de Waard A, Schneider J, Luciano JS, Rastegar-Mojarad M, Liakata M. Dynamic enhancement of drug product labels to support drug safety, efficacy, and effectiveness. J Biomed Semantics. 2013 Jan 26;4(1):5. DOI 10.1186/2041-1480-4-5 PMID: 23351881. PMCID – PMC3698101

Handler SM, **Boyce RD**, Ligons FM, Perera S, Nace DA, Hochheiser H. Use and Perceived Benefits of Mobile Devices by Physicians in Preventing Adverse Drug Events in the Nursing Home. J Am Med Dir Assoc. 2013 Oct 2. doi:pii: S1525-8610(13)00472-6. 10.1016/j.jamda.2013.08.014. Epub 2013 Oct 2. PubMed PMID: 24094901. [PMCID – in process]

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

Boyce. RD., Ryan. PB., Noren. N., et al., Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. Drug Safety. 2014. Volume 37, Issue 8 (2014), Page 557-567. DOI: 10.1007/s40264-014-0189-0. PubMed PMID: 24985530. PMCID: PMC4134480

Refereed Conference Articles:

Boyce, RD., Freimuth, RR., Romagnoli, KM., Pummer, T., Hochheiser, H., Empey, PE. Toward semantic modeling of pharmacogenomic knowledge for clinical and translational decision support. Proceedings of the 2013 AMIA Summit on Translational Bioinformatics. San Francisco, March 2013. PubMed ID: 24303292. PMCID: PMC3814496

Hassanzadeh, O., Zhu, Qian., Freimuth, RR., **Boyce R.** Extending the "Web of Drug Identity" with Knowledge Extracted from United States Product Labels. Proceedings of the 2013 AMIA Summit on Translational Bioinformatics. San Francisco, March 2013. PubMed PMID: 24303301. PMCID: PMC3814463.

Rasteger-Mojarad, M., **Boyce RD.**, Prasad, R. UWM-TRIADS: Classifying Drug-Drug Interactions with Two-Stage SVM and Post-Processing. Proceedings of the 2013 International Workshop on Semantic Evaluation (SemEval), Task 9 - Extraction of Drug-drug Interactions from BioMedical Texts. Atlanta Georgia, June 2013. http://www.aclweb.org/anthology/S/S13/S13-2110.pdf. (NIH Public Access Policy not applicable because the study was not NIH funded).

Boyce RD, Horn JR, Hassanzadeh O, de Waard A, Schneider J, Luciano JS, Rastegar-Mojarad M, Liakata M. Dynamic enhancement of drug product labels to support drug safety, efficacy, and effectiveness. J Biomed Semantics. 2013 Jan 26;4(1):5. DOI 10.1186/2041-1480-4-5 PMCID: PMC3698101

Samwald, M., Freimuth, R., Luciano, JS., Lin, Simon., Powers, RL., Marshall, MS, Adlassniga, KP., Dumontier, M. **Boyce, RD.** An RDF/OWL Knowledge Base for Query Answering and Decision Support in Clinical Pharmacogenetics. Proceedings of the 14th World Congress on Medical and Health Informatics. Copenhagen, Denmark. August 2013. PubMed PMID: 23920613. PMCID: PMC4028612.

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

Sverchkov Y, Jiang X, **Cooper GF**. Spatial cluster detection using dynamic programming. BMC Medical Informatics & Decision Making 12:22 (2012). Doi:10.1186/1472-6947-12-22. PMID: 22443103 PMCID: PMC3403878

Batal I, Valizadegan H, **Cooper GF**, Hauskrecht M. A temporal pattern mining approach for classifying electronic health record data. ACM Transactions on Intelligent Systems and Technology (2012).

Batal I, **Cooper G**, Hauskrecht M. A Bayesian scoring technique for mining predictive and non-spurious rules. Machine Learning and Knowledge Discovery in Databases 7524 (2012) 260-276.

Hennings-Yeomans PH, **Cooper GF**. Improving the prediction of clinical outcomes from genomic data using multiresolution analysis. IEEE/ACM Transactions on Computational Biology and Bioinformatics 9 (2012) 1442 - 1450. PMID: 22641708

Sverchkov Y, Visweswaran S, Clermont G, Hauskrecht M, **Cooper GF**. A multivariate probabilistic method for comparing two clinical datasets. In: Proceedings of the ACM International Health Informatics Symposium (2012) 795-800.

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

Hauskrecht M, Batal I, Valko M, Visweswaran S, **Cooper GF**, Clermont G. Outlier-detection for patient monitoring and alerting. Journal of Biomedical Informatics 46 (2013) 47-55. doi: 10.1016/j.jbi.2012.08.004. PMID: 22944172 PMCID: PMC3567774.

Montefusco DJ, Chen L, Matmati N, Lu S, Newcomb B, **Cooper GF**, Hannun YA, Lu X. Distinct signaling roles of ceramide species in yeast revealed through systematic perturbation and systems biology analyses. Science Signaling 6 (2013). rs14. doi: 10.1126/scisignal.2004515. PMID: 24170935 PMC3974757

Ferreira A, **Cooper GF**, Visweswaran S. Decision path models for patient-specific modeling of patient outcomes. In: Proceedings of the Annual Symposium of the American Medical Informatics Association (2013) 413-21. PMID: 24551347. PMC3900188

Jiang X, Cai B, Xue D, Lu X, **Cooper GF**, Neapolitan RE. A comparative analysis of methods for predicting clinical outcomes using high-dimensional genomic datasets. Journal of the American Medical Informatics Association (2014) [Epub ahead of print]. doi:10.1136/amiajnl-2013-002358. PMID: 24737607 [PubMed in process]

Cooper GF, Villamarin R, Tsui FC, Millett N, Espino J, Wagner MM. A method for detecting and characterizing outbreaks of infectious disease from clinical reports. Journal of Biomedical Informatics (to appear).

Roger S. Day, Sc.D.

Lisovich A, **Day RS.** The IdMappingAnalysis package: critically comparing identifier maps retrieved from bioinformatics annotation resources. R package Version 1.1.1, in Bioconductor Release 2.11, October 3, 2012.

Lisovich A, **Day RS**. The IdMappingAnalysis package: critically comparing identifier maps retrieved from bioinformatics annotation resources. R package Version 1.1.1, in Bioconductor Release 2.11, March 2012.

Day RS, McDade KK. A decision theory paradigm for evaluating identifier mapping and filtering methods using data integration. BMC Bioinformatics. 2013 Jul 15;14(1):223. PMCID: PMC3734162

Osmanbeyoglu HU, Lu KN, Oesterreich S, **Day RS**, Benos PV, Coronnello C, Lu X. Estrogen represses gene expression through reconfiguring chromatin structures. Nucleic Acids Research. 2013 Jul 1. PMID:23821662

Risinger JI, Allard J, Chandran U, **Day R**, Chandramouli GV, Miller C, Zahn C, Oliver J, Litzi T, Marcus C, Dubil E, Byrd K, Cassablanca Y, Becich M, Berchuck A, Darcy KM, Hamilton CA, Conrads TP, Maxwell GL. Gene expression analysis of early stage endometrial cancers reveals unique transcripts associated with grade and histology but not depth of invasion. Frontiers in Oncology. 2013;3:139. PMCID: PMC3683664

Day RS, McDade KK. A decision theory paradigm for evaluating identifier mapping and filtering methods using data integration. BMC Bioinformatics. 2013 Jul 15;14:223. doi: 10.1186/1471-2105-14-223. PMCID: PMC3734162

Gerald P. Douglas Ph.D.

Driessen J, Cioffi M, Alide N, Landis-Lewis Z, Gamadzi G, Gadabu OJ, **Douglas G**. Modeling return on investment for an electronic medical record system in Lilongwe, Malawi. Journal of the American Medical Informatics Association : JAMIA. 2012 Nov 9. PMCID: PMC3721156.

Harries AD, Jahn A, Ben-Smith A, Gadabu OJ, **Douglas GP**, Seita A, Khader A, Zachariah R. Cohort analysis of antenatal care and delivery outcomes in pregnancy: a basis for improving maternal health. Public Health Action http://dx.doi.org/10.5588/pha.13.0103. 2014 Jun 24; 4 (2):75-78.

Gadabu OJ, Munthali CV, Zacharaiah R, Gudmund-Hinderaker S, Jahn A, Twea H, Gondwe A, Mumba S, Lungu M, Malisita K, Mhango E, Makombe SD, Tenthani L, Mwalwanda L, Moyo C, **Douglas G**P, Landis-Lewis Z, Chimbwandira F. Is transcription of data on antiretroviral treatment from electronic to paper-based registers reliable in Malawi? Public Health Action. 2011 Sep;1(1):10-12(3).

Madhavi Ganapathiraju, Ph.D.

Orii N, **Ganapathiraju MK**. Wiki-pi: a web-server of annotated human protein-protein interactions to aid in discovery of protein function. PloS one. 2012; 7 (11):e49029. PMCID: PMC3509123

Ganapathiraju MK, Mitchell AD, Thahir M, Motwani K, Ananthasubramanian S. Suite of tools for statistical N-gram language modeling for pattern mining in whole genome sequences. Journal of bioinformatics and computational biology. 2012 Dec; 10 (6):1250016. PMID: 22817111

Thahir M, Sharma T, **Ganapathiraju MK**. An efficient heuristic method for active feature acquisition and its application to protein-protein interaction prediction. BMC proceedings. 2012 Nov 13; 6:S2. PMCID: PMC3504800

Ananthasubramanian S, Metri R, Khetan A, Gupta A, Handen A, Chandra N, **Ganapathiraju MK**. Mycobacterium tuberculosis and Clostridium difficille interactomes: demonstration of rapid development of computational system for bacterial interactome prediction. *Microbial Informatics and Experimentation*. 2012; 2:4. PMCID: PMC3353838

Zhu J, Zhang Y, Ghosh A, Cuevas RA, Forero A, Dhar J, Ibsen MS, Schmid-Burgk JL, Schmidt T, **Ganapathiraju MK**, Fujita T, Hartmann R, Barik S, Hornung V, Coyne CB, Sarkar SN. Antiviral Activity of Human OASL Protein Is Mediated by Enhancing Signaling of the RIG-I RNA Sensor. Immunity. 2014 Jun 19; 40 (6):936-48. PMCID: PMC4101812. PMID: 24931123.

Kuppuswamy U, Ananthasubramanian S, Balakrishnan N, **Ganapathiraju MK**. Predicting gene ontology annotations of orphan GWAS genes using protein-protein interactions. Algorithms for Molecular Biology. 2014 Apr 3; 9 (1):10. PMID: 24708602.

Giorgio E, Rolyan H, Kropp L, Chakka AB, Yatsenko S, Di Gregorio E, Lacerenza D, Vaula G, Talarico F, Mandich P, Toro C, Pierre EE, Labauge P, Capellari S, Cortelli P, Vairo FP, Miguel D, Stubbolo D, Marques LC, Gahl W, Boespflug-Tanguy O, Melberg A, Hassin-Baer S, Cohen OS, Pjontek R, Grau A, Klopstock T, Fogel B, Meijer I, Rouleau G, Bouchard JP, **Ganapathiraju M**, Vanderver A, Dahl N, Hobson G, Brusco A, Brussino A, Padiath QS. Analysis of LMNB1 duplications in autosomal dominant leukodystrophy provides insights into duplication mechanisms and allele-specific expression. Hum Mutat. 2013 Aug;34(8):1160-71.

Ganapathiraju MK, Naoki O. Research prioritization through prediction of future impact on biomedical science: a position paper on inference-analytics. GigaScience. 2013 Aug 30; 2 (11). PMID: 24001106.

Zhu J, Zhang Y, Ghosh A, Cuevas RA, Forero A, Dhar J, Ibsen MS, Schmid-Burgk JL, Schmidt T, **Ganapathiraju MK**, Fujita T, Hartmann R, Barik S, Hornung V, Coyne CB, Sarkar SN. Antiviral Activity of Human OASL Protein Is Mediated by Enhancing Signaling of the RIG-I RNA Sensor. Immunity. 2014 Jun 19; 40 (6):936-48. PMCID: PMC4101812. PMID: 24931123.

Kuppuswamy U, Ananthasubramanian S, Balakrishnan N, **Ganapathiraju MK**. Predicting gene ontology annotations of orphan GWAS genes using protein-protein interactions. Algorithms for Molecular Biology. 2014 Apr 3; 9 (1):10. PMID: 24708602.

Vanathi Gopalakrishnan, Ph.D.

Liu G, Kong L, **Gopalakrishnan V**. A Partitioning Based Adaptive Method for Robust Removal of Irrelevant Features from High-dimensional Biomedical Datasets. *AMIA*

Summits on Translational Science proceedings AMIA Summit on Translational Science. 2012; 2012:52-61. PMCID: PMC3392052.

Bigbee, W. L*., **Gopalakrishnan, V***, Weissfeld J, L., Wilson, D. O., Dacic, S. Lokshin, A. E., Siegfried, J. M. A Multiplexed Serum Biomarker Immunoassay Panel Discriminates Clinical Lung Cancer Patients from High-Risk Individuals Found to be Cancer-Free by CT Screening. J Thorac Oncol. 2012 Apr;7(4):698-708. (*These authors contributed equally to the study). PMID: 22425918 PMCID: PMC3308353

Grover, H., **Gopalakrishnan, V**. Efficient Processing of Models for Large-scale Shotgun Proteomics Data. In Proceedings of the International Workshop on Collaborative Big Data (C-Big 2012), Pittsburgh, PA, October 14, 2012.

Grover H, Wallstrom G, Wu CC, **Gopalakrishnan V.** Context-Sensitive Markov Models for Peptide Scoring and Identification from Tandem Mass Spectrometry. *Omics : a journal of integrative biology*. 2013 Feb;17(2):94-105. doi: 10.1089/omi.2012.0073. Epub 2013 Jan 5. PMCID: PMC3567622.

Floudas, C. S., Balasubramanian, J, Romkes, M., **Gopalakrishnan, V.** An empirical workflow for genome-wide single nucleotide polymorphism-based predictive modeling. Proceedings of the AMIA Translational Bioinformatics Summit 2013, March 18-20, San Francisco, CA. PMID: 24303297 PMCID: PMC3814469.

Menon, P. G., Morris, L., Staines, M., Lima, J., Lee, D. C., Lee, C., **Gopalakrishnan, V**. Novel MRI-derived quantitative biomarker for cardiac function applied to classifying ischemic cardiomyopathy within a Bayesian rule learning framework. Proceedings of the SPIE Medical Imaging 2014. February 15-20, 2014. San Diego, CA.12.

Balasubramanian, J. B., Cooper, G.F., Visweswaran, S. **Gopalakrishnan, V**. Selective Model Averaging with Bayesian Rule Learning for Predictive Biomedicine. Proceedings of the AMIA 2014 Joint Summits in Translational Science, April, 2014, San Francisco, CA.

Dutta-Moscato J, **Gopalakrishnan V**, Lotze MT, Becich MJ. Creating a pipeline of talent for informatics: STEM initiative for high school students in computer science, biology, and biomedical informatics. Editorial in J Pathol Inform 2014;5:12. PMID: 24860688.

Gopalakrishnan V, Menon PG, Madan S. A novel framework to enhance scientific knowledge of cardiovascular MRI biomarkers and their application to pediatric cardiomyopathy classification. Proceedings of the Second International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2014); Granada, Spain. 2014. p. 798-809.

Shi, L., Grover, H. Balasubramanian, J. B., Kolli, K, Shriver, C., **Gopalakrishnan, V**. A Flexible Feature Selection Framework for Improving Breast Cancer Classification from

Sparse Spectral Count Proteomic Data. Proceedings of The 2014 International Conference on Data Mining, DMIN'14, July 21-24, 2014, Las Vegas, Nevada, USA.

Zaidi, A., **Gopalakrishnan, V**. et al. Evaluation of a four-protein serum biomarker panel – biglycan, annexin-A6, myeloperoxidase and protein S100-A9 (B-AMP©) – for detection of esophageal adenocarcinoma. Cancer. 2014 Aug 5. doi: 10.1002/cncr.2896. PMID: 2510029. (Impact Factor = 5.2)

Steven Handler, M.D., Ph.D.

Gellad WF, Aspinall SL, **Handler SM**, Stone RA, Castle N, Semla TP, Good CB, Fine MF, Dysken J, Hanlon JT. Use of antipsychotics among older residents in VA nursing homes. Medical Care. 2012 Nov;50(11):954-960 PMCID: PMC3746977

Nace DA, **Handler SM**, Hoffman EL, Perera S. Impact of the Rasing Immunizations Safely and Effectively (RISE) Program on Healthcare Worker Immunization Rates in Long-term Care Settings. Journal of the American Medical Directors Association.2012 Nov13;13(9):806-10. PMCID: PMC3650646

Marcum ZA, Pugh MV, Amuan ME, Aspinall SL, **Handler SM**, Ruby CM, Hanlon JT. Prevalence of Potentially Preventable Unplanned Hospitalizations Caused by Therapeutic Failures and Adverse Drug Withdrawal Events among Older Veterans. The Journal of Gerontology: Biological Sciences and Medical Sciences 2012;67(8 Aug):867-874. PMCID: PMC3403866.

Kane-Gill SL, Forsberg EA, Verrico MM, **Handler SM**. Comparison of three pharmacovigilance algorithms in the ICU setting: a retrospective and prospective evaluation of ADRs. Drug Safety. 2012, Aug 1;35(8):645-53 PMID: 22720659.

Schwarz EB, Parisi SM, **Handler SM**, Koren G, Shevchik GJ, Fischer GS. Clinical Decision Support to Promote Safe Prescribing to Women of Reproductive-age: A Cluster-Randomized Trial Journal of General Internal Medicine. 2012;27(7 July):831-838. PMCID: PMC3378745.

Romagnoli KM, **Handler SM**, Home Health Care: More than Just a Visiting Nurse [Invited Editorial]. BMJ: Quality and Safety in Healthcare. 2013 Aug 12 (Epub ahead of print). PMID: 23940375 PMC in process.

Schwarz EB, Parisi, SM, **Handler SM**, Koren G, Fischer GS. Patient-reported Counseling about Medication-induced Birth Defects with Clinical Decision Support in Primary Care. Journal of Women's Health (Epub ahead of print). 2013 Aug 9. PMID: 23930947 PMCID: PMC3837561 [Available on 2014/10/1]

Thorpe-Jamison PT, Cullet CM, Perera S, **Handler SM**. Evaluating the Impact of Computer-Generated Rounding Reports on Physician Workflow in the Nursing Home: A Feasibility Time-Motion Study. Journal of the American Medical Directors Association. 2013 May 14(5):358:62. PMID: 23318665.

Marcum ZA, Sevick MA, **Handler SM**. Medication Non-Adherence: A Diagnosable and Treatable Medical Condition. Journal of the American Medical Association (JAMA) 2013 May 22;309(20):2105-6. PMID: 23695479 PMCID: PMC3976600

Wagner LM, Castle NG, **Handler SM**. The use of Health Information Technology for Adverse Event Reporting in Nursing Homes: Barriers and Facilitators. Geriatric Nursing. 2013 Mar-Apr;34(2):112-5. Epub 2012 Dec 23 PMID:23266459

Roshanov PS, Fernandes N, Wilczynski J, **Handler SM**, Haynes RB. Features of effective computerized clinical decision support. British Medical Journal (BMJ) 2013;Feb 14;346:f657 PMID: 23412440, PMC: not applicable

Marcum ZA, Rovesti KL, Behrens MC, Logsdon MW, Myers J, Francis SD, Aspinall SL, Hanlon JT, **Handler SM**. Use of an adverse drug event trigger tool in Veterans Affairs nursing homes. Consultant Pharmacist. 2013 Feb;28(2):99-109. PMCID: PMC3677952

Romagnoli KM, **Handler SM**, Ligons FM, Hochheiser H. Home Health Care Nurses' Perceptions of the Post-hospitalization Information Needs of Geriatric Patients. BMJ: Quality and Safety in Healthcare 2013 Apr;22(4):324-32. doi: 10.1136/bmjqs-2012-001207. Epub 2013 Jan 29. PMID: 23362507 PMCID: PMC3694324

Handler SM, Boyce, RD, Ligons FM, Perera S, Nace DA, Hochhesier H. Use and Perceived Benefits of Mobile Devices by Physicians in Preventing Adverse Drug Events in the Nursing Home. J Am Med Dir Assoc. 2013 Dec;14(12):906-10. doi: 10.1016/j.jamda.2013.08.014. Epub 2013 Oct 2. PMID: 24094901.

Castle NG, **Handler SM**, Wagner LM. Potential Prescription Drug Misuse in Assisted Living. Res Gerontol Nurs. 2014 Jan-Feb;7(1):25-32. doi: 10.3928/19404921-20130910-01. Epub 2013 Sep 16. PMID:24044781.

Marcum ZA, Sevick MA, **Handler SM**. Medication Non-Adherence: A Diagnosable and Treatable Medical Condition. Journal of the American Medical Association (JAMA) 2013 May 22;309(20):2105-6. PMID: 23695479. [PubMed in process]

Thorpe-Jamison PT, Cullet CM, Perera S, **Handler SM**. Evaluating the Impact of Computer-Generated Rounding Reports on Physician Workflow in the Nursing Home: A Feasibility Time-Motion Study. Journal of the American Medical Directors Association. 2013 May 14(5):358:62. PMID: 23318665. [PubMed in process]

Castle NG, Wagner LM, Ferguson-Rome JC, Men A, **Handler SM**. Hand Hygiene Deficiency Citations in Nursing Homes. Appl Gerontol. 2014 Feb;33(1):24-50. doi: 10.1177/0733464812449903. Epub 2012 Aug 1. PMID: 24652942 [PubMed - in process]

Ligons FM, Mello-Thoms C, **Handler SM**, Romagnoli KM, Hochheiser H. Assessing the impact of cognitive impairment on the usability of an electronic medication delivery unit in an assisted living population. Int J Med Inform. 2014 Jul 27. pii: S1386-

5056(14)00139-7. doi: 10.1016/j.ijmedinf.2014.07.004. [Epub ahead of print] PMID: 25153770 PMC: In process.

Harry Hochheiser, Ph.D.

Kohle-Ersher A, Chatterjee P, Osmanbeyoglu HU, **Hochheiser H**, Bartos C. Evaluating the barriers to point-of-care documentation for nursing staff. Comput Inform Nurs doi: 10.1097/NCN.0b013e3182343f1. 2012 Mar;30(3):126-33. PMID:22024972.

Romagnoli KM, Handler SM, **Hochheiser H**. Home Care: More Than Just a Visiting Nurse. BMJ Qual Saf. 2013 Dec;22(12):972-4. doi0:10.1136/bmjqs-2013-002339 PMID:23940375

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

Brinkley JF, Borromeo C, Clarkson M, Cox TC, Cunningham MJ, Detwiler LT, Heike CL, **Hochheiser H**, Mejino L, Travillian RS, Shapiro LG. The Ontology of Craniofacial Development and Malformation for translational craniofacial research. Seminars in Medical Genetics DOI: 10.1002/ajmg.c.31377. 2013 Oct 4; 1-14. PMID: 24124010.

Handler SM, Boyce RD, Ligons FM, Perera S, Nace DA, **Hochheiser H**. Use and Perceived Benefits of Mobile Devices by Physicians in Preventing Adverse Drug Events in the Nursing Home. Journal of The American Medical Directors Association. 2013 Oct 2. PMID: 24094901.

Wentz B, **Hochheiser H**, Lazar J. A Survey of Blind Users on the Usability of Email Applications. In: Universal Access in the Information Society International Journal. Springer-Verlag, 2013 August. p.327-36.

Romagnoli KM, Handler SM, Ligons FM, **Hochheiser H**. Home-care nurses' perceptions of unmet information needs and communication difficulties of older patients in the immediate post-hospital discharge period. BMJ quality & safety. 2013 Apr; 22 (4):324-32. PMCID: PMC3694324. PMID: 23362507.

Ligons FM, Mello-Thoms C, Handler SM, Romagnoli KM, **Hochheiser H**. Assessing the impact of cognitive impairment on the usability of an electronic medication delivery unit in an assisted living population. International Journal of Medical Informatics. 2014 Jul 27. PMID:25153770.

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

Zheng J, Chapman WW, Miller TA, Chen L, **Crowley Jacobson RS**, Savova GK. A system for coreference resolution for the clinical narrative. Journal of the American Medical Informatics Association. 2012 Jul-Aug; 19(4):660-7. PMCID: PMC3384116

Mello-Thoms C, Mello CA, Medvedeva O, Castine M, Legowski E, Gardner G, Tseytlin E, **Crowley Jacobson RS**. Perceptual analysis of the reading of dermatopathology virtual slides by pathology residents. Arch Pathol Lab Med, 2012; 136, 551-62. PMCID: PMC3728442

Kuqi K, Eveleigh T, Holzer T, Sarkani S, Levin JE, **Crowley Jacobson RS**. Design of electronic medical record user interfaces: a matrix-based method for improving usability. *Journal of Healthcare Engineering*. 2013;4(3):427-52. PMID: 23965597 [PubMed - in process]

Bekhuis T, Demner-Fushman D, **Crowley Jacobson RS**. Comparative effectiveness research designs in MeSH and Emtree: an evaluation of coverage. JMLA: Journal of the Medical Library Association. 2013;101(2):92-100. PMCID: PMC3634392.

Liu K, Mitchell KJ, Chapman WW, Savova GK, Sioutos N, Rubin DL, **Crowley Jacobson RS**. Formative evaluation of ontology learning methods for entity discovery by using existing ontologies as reference standards. *Methods of Information in Medicine*. 2013 Aug 6;52(4):308-16. PMID: 23666409 [PubMed - in process]

Crowley Jacobson RS, Legowski E, Medvedeva O, Reitmeyer K, Tseytlin E, Castine M, Jukic D, Mello-Thoms C. Automated detection of heuristics and biases among pathologists in a computer-based system. *Advances in Health Sciences: Education, Theory and Practice*, 2013 Aug;18(3):343-63 PMCID: PMC3728442

Xia Jiang, Ph.D.

Jiang, X., and R.E. Neapolitan, Mining Strict Epistatic Interactions From High-Dimensional Datasets: Ameliorating the Curse of Dimensionality, *PLoS ONE*, 2012;7(10):e46771. PMCID: PMC3470561

Sverchkov Y, **Jiang X**, Cooper GF. Spatial Cluster Detection Using Dynamic Programming. *BMC Medical Informatics and Decision Making.* 2012 Mar 25;12(1):22 PMCID: PMC3403878

Jiang X, Visweswaran S, Neapolitan RE. Scoring and Searching Bayesian Network Models of Gene-Phenotype Association, to appear in Sinoquet C, and R. Mourad (Eds.): Probabilistic Graphical Models for Genetics, Oxford University Press, 2013

Jiang X, Neapolitan RE, The Bayesian Network Story, to appear in Hájek, A, and C. Hitchcock (Eds.): The Oxford Handbook of Probability and Philosophy, Oxford University Press, 2013

Jiang X, Cai B, Xue D, Lu X, Cooper GF, Neapolitan RE. A comparative analysis of methods for predicting clinical outcomes using high-dimensional genomic datasets.

Journal of the American Medical Informatics Association (2014) [Epub ahead of print]. doi:10.1136/amiajnl-2013-002358. PMID: 24737607 [PubMed in process]

Jiang X, Xue D, Brufsky A, Khan S, Neapolitan R. A new method for predicting patient survivorship using efficient bayesian network learning. Cancer Inform. 2014 Feb 13;13:47-57. doi: 10.4137/CIN.S13053. eCollection 2014. PMCID: PMC3928477.

Neapolitan R, Xue D**, Jiang X.** Modeling the altered expression levels of genes on signaling pathways in tumors as causal bayesian networks. Cancer Inform. 2014 May 25;13:77-84. doi: 10.4137/CIN.S13578. eCollection 2014. PMCID:PMC4051800.

Neapolitan RE, **Jiang X**. A Note of Caution on Maximizing Entropy. Entropy. 2014; 16 (7):4004-14. doi: 10.3390/e16074004

Cai B, **Jiang X**. Novel Artificial Neural Network Method for Biomedical Prediction based on Matrix Pseudo-Inversion. Journal of Biomedical Informatics. 2014 Apr; 48:114-21. doi: 10.1016/j.jbi.2013.12.009.

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

Feng H, Hu B, Liu KW, **Lu X**, Yiin JJ, Lu S, Keezer S, Fenton T, Furnari FB, Hamilton RL, Vuori, K, Nagane M, Nishikawa R, Cavenee WK, Cheng, SY. Aberrant Activation of Rac1 by Src-dependent Phosphorylation of Dock180Y1811 Mediates PDGFRα-stimulated Glioma Growth and Invasion. *Journal of Clinical Investigation.* 2012 121(12):4670–4684. PMCID: PMC3223070

Osmanbeyoglu H, Hartmaier R., Oesterreich S., **Lu X**. Improving ChIP-seq peak-calling for functional indirect co-regulator binding by integrating multiple sources of biological information. *BMC Genomics* 2012, 13(Suppl 1):S1. Epub 2012 Jan 17 PMCID: PMC3439677

Richards A, Schwacke J, Rohrer B, Cowart LA. **Lu X**. Revealing functionally coherent gene subset using spectral clustering and information integration approaches. *BMC Systems Biology*, 2012; 6 (Suppl 3) : S7. Epub 2012 Dec 17. PMCID: PMC3542577

Qin T, Tsoi LC, Sims KJ, **Lu X** and Zheng WJ. Signaling network prediction using the ontology fingerprint enhanced Bayesian networks. *BMC Systems Biology*, 2012; 6 (Suppl 3) : S3 (co-corresponding author) Epub 2012 Dec 17. PMCID: PMC3524013

Mowrey D, Cheng M, Liu L, Willenbring D, **Lu X**, Wymore T, Xu Y, Tang P. Asymmetric ligand binding facilitates conformational transitions in pentameric ligand-gated ion channels. *Journal of the American Chemical Society,* 2013 135(6):2172-80. PMCID: PMC3582375

Lu S, Jin B, Cowart LA, **Lu X**. From data towards knowledge: Revealing the architecture of signaling systems by unifying knowledge mining and data mining of systematic perturbation data. PLoS One. 2013 Apr 23;8(4):e61134 PMCID: PMC3634064

Osmanbeyoglu H, Lu K, Oesterreich S, Day RS, Benos PV, Coronnello C, **Lu X**. Estrogen represses gene expression through chromatin reconfiguration. *Nucleic Acid Research*. 2013 Jul 1. [Epub ahead of print] PMID: 23821662 [Pub-Med in process]

Lu S, **Lu X**. Using graph model to find transcription factor modules: the hitting set problem and an exact algorithm. *Algorithms for Molecular Biology,* 2013; 8:2. PMCID: PMC3622577

Montefusco, D., Chen, L, Matmati, N., Lu, S., Newcomb, B., Cooper, GF., Hannun, YA., **Lu, X**, (2013) Distinct signaling roles of ceramide species in yeast revealed through systematic perturbation and integromics analyses. Science Signaling 6:rs14 (reported as in press last year, published in October, 2013)

Zhao, Z, Shen, B., **Lu X,** and Vongsangnak, W (2013) Translational Biomedical Informatics and Computational Systems Medicine. BioMed Research International, vol. 2013, Article ID 237465, 2013. doi:10.1155/2013/237465

Jiang X, Cai B, Xue D, **Lu X**, Cooper GF, Neapolitan RE. A comparative analysis of methods for predicting clinical outcomes using high-dimensional genomic datasets. Journal of the American Medical Informatics Association (2014) [Epub ahead of print]. doi:10.1136/amiajnl-2013-002358. PMID: 24737607 [PubMed in process]

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

Que J. and **Tsui F.-C**, Spatial and temporal algorithm evaluation for detecting over-the-counter thermometer sale increasing during 2009 H1N1 pandemic, *Online Journal of Public Health Informatics* 2012;4(1). Epub 2012 May 17. PMCID: PMC3615801

Rexit R., **Tsui F.-C.**, et. al., Using a distributed search engine to identify optimal product sets for use in an outbreak detection system, International Workshop on Collaborative Big Data (C-Big 2012), Pittsburgh, PA, Oct. 14, 2012. DOI 10.4108/icst.collaboratecom.2012.250728

Villamarín R, Cooper G, Wagner M, **Tsui F.-C**, Espino JU. A Method for Estimating from Thermometer Sales the Incidence of Diseases that are Symptomatically Similar to Influenza, *Journal of Biomedical Informatics*. 2013 Jun;46(3):444-57. Epub 2013 Mar 14 PMID: 23501015 [PubMed - in process]

Que J. and **Tsui F.-C**., Modeling baseline shifts in multivariate disease outbreak detection, *Online Journal of Public Health Informatics*. Published online 2013 April 4, 2013 5(1): e9. PMCID: PMC3692939

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. 2013;8(3):e59273. Epub 2013 Mar 21. PMCID: PMC3605458

López PA, **Tsui F.-C.**, Visweswaran S, Cooper GF, Detection of Patients with Influenza Syndrome Using Machine-Learning Models Learned from Emergency Department Reports, *Online Journal of Public Health Informatics*. 2013; 5(1): e41. Published online 2013 April 4. PMCID: PMC3692886
Rexit R, **Tsui F.-C**., Espino J, et. al., An analytics appliance for identifying (near) optimal over-the-counter medicine products as health indicators for influenza surveillance, Journal Manager of Information Systems, 2014

Amin W, **Tsui FR**, et. al., PaTH: towards a learning health system in the Mid-Atlantic region. Journal of American Medical Informatics Association, 21(4), Jul-Aug, 2014, pp633-6, PMCID: PMC4078296

Ye Y, **Tsui F.-C**., Wagner M, Espino JU, Li Q, Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers, Journal of American Medical Informatics Association, Jan. 2014, PMID: 24406261

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

Refereed Journal Articles

Hauskrecht, M, Batal, I, Valko, M, **Visweswaran, S**, Cooper, GF, Clermont, G. Outlier detection for patient monitoring and alerting. *Journal of Biomedical Informatics*. 2013 Feb; 46(1):47-55. PMCID: PMC3567774.

Kalamangalam GP, Pestana Knight, EM, **Visweswaran, S**, Gupta, A. Noninvasive predictors of subdural grid seizure localization in children with nonlesional focal epilepsy. *Journal of Clinical Neurophysiology*. 2013 Feb;30(1):45-50 PMID: 23377441

Stokes, M, **Visweswaran, S**. Modular Relief for ranking genetic predictors of disease. *BioData Mining*. 2012 Dec 3;5(1):20 PMCID: PMC3554553

Strobl, EV, Eack, SM, Swaminathan, V, **Visweswaran, S**. Predicting the risk of psychosis onset: Advances and prospects. *Early Intervention in Psychiatry*. 2012 Jul 8. PMCID: PMC3470783 [Available on 2013/11/1]

Bhavnani SK, Bellala, G, Victor, S, Bassler, K, **Visweswaran, S**. The role of complementary bipartite visual analytical representations in the analysis of SNPs: A case study in ancestral informative markers. *Journal of the American Medical Informatics Association*. 2012 Jun 1;19(e1):e5-e12. PMCID: PMC3392853

Mowery, D, Weibe, J, **Visweswaran, S**, Harkema, H, Chapman, WW. Building an automated SOAP classifier for emergency department reports. *Journal of Biomedical Informatics*. 2012 Feb;45(1):71-81. PMCID: PMC3267853

Zaidi, AH, Gopalakrishnan, V, Kasi, PM, Malhotra, U, Balasubramanian, J, **Visweswaran, S**, Zeng, X, Sun, M, Bergman JJ, Bigbee, WL, Jobe, BA. Evaluation of a four-protein biomarker panel for detection of esophageal adenocarcinoma. Cancer. 2014 Aug 5. PMID: 25100294.

Aflakparast, M, Masoudi-Nejad, A, Bozorgmehr, JH, **Visweswaran, S**. Informative Bayesian Model Selection: A method for identifying interactions in genome-wide data. Molecular BioSystems, 2014 Aug;10(10):2654-2662. PMID: 25070634.

Aflakparast, M, Salimi, H, Gerami, A, Dubé, M-P, **Visweswaran, S**, Masoudi-Nejad, A. Cuckoo search epistasis: A new method for exploring significant genetic interactions. Heredity. 2014 Jun;112(6):666-74. PMID: 24549111 PMCID: PMC4023449.

Stokes, ME, Barmada, MM, Kamboh, MI, **Visweswaran, S**. The application of network label propagation to rank biomarkers in genome-wide Alzheimer's data. BMC Genomics. 2014 Apr;15(1):282. PMID: 24731236.

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

Lee BY, Tai JH, McGlone SM, Bailey RR, Wateska AR, Zimmer SM, Zimmerman RK, **Wagner MM**. The potential economic value of a 'universal' (multi-year) influenza vaccine. Influenza and other respiratory viruses. 2012 May;6(3):167-75. PMCID:PMC3253949

Wagner MM, Levander JD, Brown S, Hogan WR, Millett N, Hanna J. Apollo: giving application developers a single point of access to public health models using structured vocabularies and Web services. AMIA Annu Symp Proc. 2013 Nov 16;2013:1415-24. eCollection 2013

Cooper GF, Villamarin R, Rich Tsui FC, Millett N, Espino JU, **Wagner MM**. A method for detecting and characterizing outbreaks of infectious disease from clinical reports. J Biomed Inform. 2014 Aug 30. pii: S1532-0464(14)00192-0. doi: 10.1016/j.jbi.2014.08.011. [Epub ahead of print] PMID: 25181466.

Ye Y, Tsui FR, **Wagner M**, Espino JU, Li Q. Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers. J Am Med Inform Assoc. 2014 Jan 9; PMCID: PMC4147621 [Available on 2015/9/1].

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
5 Administrative Support Staff
2 Grants & Finance
3 Program Managers (Training Program, Operations, RODS Lab)
1 Clinical Data Scientist
2 Research Specialists
25 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

Technical Staff

Rebecca Boes
Software Design/Development

Charles Borromeo
Software Design/Development

Anish Bhaswanth Chakka
Software Design/Development

Girish Chavan
Software Design/Development

Jeremy Espino
Software Design/Development

Johnson Paul Kottakalil
Software Design/Development

Liz Legowski
Research Specialist

John Levander
Software Design/Development

Soumya Luthra
Software Design/Development

Michael Marks
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

FINANCIAL PLAN

FY2014 Financial Summary

The FY2014 actual performance for Hard Money (Dept. Operations) revenue was under budget. The actual net operating budget deficit was ($700,156) as the Office of the Senior Vice Chancellor Health Sciences (SVCHS) covered DBMI's operating cost deficit. The SVCHS will provide a 5 year period for DBMI to repay this deficit amount. The key measures of research support included indirect research revenue of $2.133 million, about $882,902 short of the projected budget. Direct research revenue of $6.085 million was $2.518 million less than budget, reflecting a less successful research portfolio. (See the Statement of Revenue and Expenses Fiscal Year 2014)

Department of Biomedical Informatics
Statement of Revenues and Expenses
Fiscal Year 2014

	FY14 Budget Operating Acct	FY14 Actual Operating Acct	FY14 Budget Service Accounts	FY14 Actual Service Accounts	FY14 Budget Discretionary	FY14 Actual Discretionary	FY14 Budget Research	FY14 Actual Research	FY14 Budget Pitt/UPMC	FY14 Actual Pitt/UPMC	FY14 Budget TOTAL	FY14 Actual TOTAL
Revenue												
Grant Revenue	$ 2,938,034	$ 2,133,871					$ 8,603,328	$ 6,085,154			$ 11,541,362	$ 8,219,025
Other Revenue	$ 600,000	$ 500,000	$ 258,529	$ 119,198	$ 47,277	$ 455,042			$ -		$ 905,806	$ 1,074,240
School of Medicine Start-Up		$ 1,350			$ -	$ -					$ -	$ 1,350
Education Support	$ 447,839	$ 447,839			$ 200,600	$ 26,572					$ 200,600	$ 474,411
Total Revenue	**$ 3,985,873**	**$ 3,083,060**	**$ 258,529**	**$ 119,198**	**$ 247,877**	**$ 481,614**	**$ 8,603,328**	**$ 6,085,154**	**$ -**	**$ -**	**$ 13,095,607**	**$ 9,769,026**
Expenses												
Personnel and Other	$ 3,580,744	$ 3,380,309	$ 219,743	$ 281,813	$ 47,277	$ 191,535	$ 8,603,328	$ 6,085,154	$ -	$ -	$ 12,451,092	$ 9,938,811
University Overhead	$ 443,915	$ 443,915	$ 38,786		$ -	$ -	$ -				$ 482,701	$ 443,915
University Overhead Transfer	$ (38,786)	$ (741,164)	$ -		$ (5,596)	$ -	$ -					
Total Expenses	**$ 3,985,873**	**$ 3,083,060**	**$ 258,529**	**$ 281,813**	**$ 41,681**	**$ 191,535**	**$ 8,603,328**	**$ 6,085,154**	**$ -**	**$ -**	**$ 12,889,411**	**$ 9,641,562**
Surplus / (Deficit)	$ -	$ (0)	$ -	$ (162,615)	$ 206,196	$ 290,079	$ -	$ -	$ -	$ -	$ 206,196	$ 127,464
Deficit Transfer				$ 200,464							$ -	$ 200,464
Net Surplus / (Deficit)	**$ -**	**$ (0)**	**$ -**	**$ 37,849**	**$ 206,196**	**$ 290,079**					**$ 206,196**	**$ 327,928**
Fund Balance as of 06/30/13				*$ 199,549*		*$ 17,410*					*$ -*	***$ 216,959***
Fund Balance as of 06/30/14				*$ 39,816*		*$ 287,631*					*$ -*	***$ 327,448***

Notes: **FY14**

* Operating Account (Indirect Grant revenue) was $804,163 below budget, based on unsuccessful grant application with the PACE, NMVB and AVON Foundation	$ 804,163	
* Operating Account expenses was $902,813 below budget. Dr. Levine loaned DBMI the amount of $700,156 in June14 to bring operating to a balance budget	$ 902,813	
* Service Account revenue was $139,331 lower than budget as a result of the FY13 NRDM revenue carryover, which reduced the cost per store that DBMI could charge	$ 139,331	
* Service Account expenses were above budget, as a result of more personnel spending add'l time on the NRDM project	$ (23,284)	
* Discretionary Account income exceeded budget projections, due to Personalized Medicine funds provided by Dr. Berg for Dr. Jacobson. Also, online registration income was generated.		$ 149,854
* Discretionary Account expenses were over budget as a result of PGRR project expenses.		
* Direct Grant revenue was significantly below budget, resulting from less research awards than budgeted.	$ 2,518,174	

FY15 budget anticipates:

1. Start Up costs for Xia Jiang (Post Doc) and Gerry Douglas (GSR) will stop by Oct14. Tanya Bekhuis (Staff) will hit the 02 Account in FY14.
2. Online course, under the direction of Rebecca Jacobson, will resume in the spring of 2015. No income was budgeted for this.
3. NRDM costs will flatline in FY15, as DBMI utilizes FY14 carryover funds in FY15 to project to a zero balance for June 30th, 2015.
4. Service Account revenue and expense projected will increase, based on requests for Uma/Soumya's time.
5. New projects for Cooper (BD2K) Jacobson (PGRR3) and Tsui (CHP-McCune Foundation) were not factored into the FY15 budget.

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 has placed it as one of the most well-funded departments 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 remains near the highest in total research funding.

The Department's historical five-year plan was to project a relatively flat growth in research funding. The FY2014 research revenue budget of $8,603,328 is less than 1% higher than FY2013 actual research revenue, mainly a result of the flat lined funding of the federal research support.

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. FY2014 represents the second 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

Due to the unexpected departure of Dr. Handler in July 2014 and multiple changes in UPMC Information Services Division leadership including loss of the CIO, Director of the TDC and the Director of the UPMC Enterprise Data Warehouse, DBMI is engaging in a Strategic Planning exercise and will "relaunch" its effort in the highest quality of support for the clinical practice of medicine in early 2015. The new plan will involve 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 going to be 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 is going to assume 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. More details on our activities in this important area will be in our next annual report.

FY2015 Financial Projections

The FY2015 budget projects an increase in research funding from the previous year. The FY2015 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 FY2015 reflects a surplus of $42,090 and will easily represent a major (20% increase or more) in grant revenue.

University of Pittsburgh School of Medicine
University of Pittsburgh Physicians
Department of Biomedical Informatics
Schedule of Revenue and Expenses Fiscal Year 2015 Budget

	University	UPP	Total Budget FY2015
Revenue	$ -	$ -	$ -
Patient Care			$ -
Grant:			$ -
Directs	$ 9,319,214		$ 9,319,214
Indirects	$ 2,757,043		$ 2,757,043
Hospital Contract			$ -
School of Medicine	$ 307,356		$ 307,356
VAMC			$ -
Other	$ 855,202		$ 855,202
Total Revenue	**$ 13,238,815**	**$**	**$ 13,238,815**
Expenses			
Salaries and Fringe Benefits:			
Faculty	**$ 7,607,213**		**$ 7,607,213**
Non-Faculty			**$ -**
Malpractice Insurance			**$ -**
Space Rental	$ 772,242		**$ 772,242**
UPP Overhead	$ -	-	**$ -**
University Overhead	$ 445,095		**$ 445,095**
Other Operating Expenses	$ 4,372,175		**$ 4,372,175**
Total Operating Expenses	**$ 13,196,725**	-	**$ 13,196,725**
Excess Revenue over Expenses	**$ 42,090**	**$**	**$ 42,090**
Captial Equipment/Improvements	**$ -**	**$**	**$ -**
Fund Balances			
University Restricted Accounts as of 6/30/14	**$ 290,079**	**$ -**	**$ 290,079**
University Endowments as of 6/30/14			
UPP Fund Balance as of 6/30/14			
UPMC Endowments as of 6/30/14			
UPMC SPF Accounts as of 6/30/14			
Total Fund Balances	**$ 290,079**	**$**	**$ 290,079**

END OF REPORT