Biomedical Informatics

Division Photo

First Row: J. Ma, M. Medvedovic, A. Jegga, J. Lu; Second Row: E. Hall, M. Kouril, J. Hutton; Third Row: K. Marsolo, M. Wagner; Fourth Row: A. Spooner

Division Data Summary
Research and Training Details

<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Faculty</td>
<td>11</td>
</tr>
<tr>
<td>Number of Joint Appointment</td>
<td>7</td>
</tr>
<tr>
<td>Number of Research Fellows</td>
<td>3</td>
</tr>
<tr>
<td>Number of Research Students</td>
<td>26</td>
</tr>
<tr>
<td>Number of Support Personnel</td>
<td>28</td>
</tr>
<tr>
<td>Direct Annual Grant Support</td>
<td>$773,182</td>
</tr>
<tr>
<td>Peer Reviewed Publications</td>
<td>45</td>
</tr>
</tbody>
</table>

Faculty Members

John Hutton, MD, Professor; Director, Division Chief

Bruce Aronow, PhD, Professor; Co-Director, Computational Medicine Center
  Research Interests: Gene Expression Analysis, Gene Regulation, Clinical Genomics, Functional Genomics of Development and Disease

Anil Jegga, MS, DVM, Research Assistant Professor
  Research Interests: Gene Regulatory Networks, Biomedical Ontologies, Integrative Genomics

Michal Kouril, PhD, Research Instructor; Director, Research IT
  Research Interests: Computational Support, High-performance computing, Parallel Programming, High-end Data Storage

Long Jason Lu, PhD, Assistant Professor
  Research Interests: Bioinformatics, Machine Learning, Integrative Genomics, Biological Networks, Computational Modeling, Software Development

Jun Ma, PhD, Associate Professor
Research Interests: Development, Transcription, Morphogen Gradient, Embryo, Robustness, Quantitative Studies

Keith Marsolo, PhD, Research Instructor; Director, Software Development and Data Warehouse
Research Interests: i2b2, Data Integration, Data Warehousing and Data Management

John Pestian, PhD, MBA, Associate Professor; Director, Computational Medicine Center
Research Interests: Natural Language Processing, Clinical Decision Support, Suicide Research, Pathology Research, Psychiatric Research

Jerry Phillips, MS, Field Service Assistant Professor
Research Interests: High-end SUN Solaris Computational Support

S. Andrew Spooner, MD, FAAP, Associate Professor Clinical; Chief Medical Information Officer
Research Interests: Decision Support, Pharmacy Information Systems

Michael Wagner, PhD, Research Assistant Professor; Faculty Liaison
Research Interests: Machine Learning, Proteomics, Genome-wide Association, Parallel Computing, Computational Infrastructure, Bioinformatics

Joint Appointment Faculty Members

Karen Davis, PhD, Associate Professor
  UC ECECS
  Electrical & Computer Engineering and Computer Science

Eric Hall, PhD, Research Instructor
  Neonatology & Pulmonary Biology
  Clinical Informatics, Knowledge Discovery Tools, Data Mining and Warehousing

Mark Halsted, MD, Associate Professor
  Radiology

Michael Jansen, PhD, Research Instructor
  Experimental Hematology

Mario Medvedovic, PhD, Associate Professor
  UC Environmental Health

Jarek Meller, PhD, Associate Professor
  UC Environmental Health

Yan Xu, PhD, Associate Professor
  Pulmonary Medicine

Trainees

- Jacek Biesiada, PhD, University of Silesia, Poland
- Walter Jessen, PhD, Texas A&M University
- Miroslaw Kordos, PhD, Silesian University of Technology, Poland

Significant Accomplishments in FY08

Developing computational approaches to understand the regulatory networks of tumor suppressor protein p53 (Anil Jegga, MS, DVM)

Dr. Jegga’s current research is focused on deciphering the micro and macro regulation of pathways and networks underlying cancer. Specifically, he is developing computational approaches to help understand the regulatory networks – transcriptional and posttranscriptional – of the tumor suppressor protein p53, the master regulator and the guardian of the genome. Mutations in p53 are implicated in over 50% of human cancers. Following toxic exposure, p53 binds to specific sites on our DNA sequence called p53 response elements and brings about numerous, coordinated changes in target genes. The way p53 binds to its target genes is governed by several rules, which Jegga and his team is identifying. Understanding these rules and the consequences of breaking them (because of DNA mutations) will provide a better understanding of what makes some of us more prone to cancer than others.

Establishment of an i2b2 Research Data Warehouse (Keith Marsolo, PhD)

One of the biggest challenges faced by medical investigators today is determining whether a group of patients is available for a research study. Trying to identify patients based on criteria like their demographics or diagnostic history typically requires the analysis of individual medical records, which is a difficult, time consuming task constrained by
privacy concerns. To circumvent this process, Dr. Marsolo has implemented a research data warehouse that employs the i2b2 framework (Informatics for Integrating Biology and the Bedside, https://www.i2b2.org/). The i2b2 framework allows users to perform a CCHMC-wide search on a de-identified dataset to determine the existence of a patient cohort. Marsolo and his team have loaded information on almost 500,000 patients, including roughly 5 years worth of basic clinical information and patient data (demographics, diagnoses, laboratory, medications). The warehouse incorporates data from the hospital's new Epic electronic health record, as well as information from research groups, including genetic variations and associated biomarkers. At the same time, Marsolo's group is ensuring that the appropriate regulatory and security safeguards are in place to protect the privacy of patients. By bringing these disparate resources together, the group can present CCHMC investigators with a comprehensive view of patients who are potential subjects for clinical research. Marsolo believes that the research data warehouse and associated software applications will be a key resource for investigators addressing medicine's most complex puzzles.

**Division Highlights**

**Bruce Aronow, PhD**

The overall goal for Bruce Aronow's group is to enable, through the emerging field of integrative genomics, the advancement of basic and translational research across many different areas of disease research that can lead to improved human health. His research is based on the application of modeling and data analysis techniques both to existing knowledge and to new data from high-throughput genomic and genetic analyses. His research group provides infrastructure and helps to develop models of how biological systems assemble, adapt, are genetically modified, and become impaired in disease.

Publications this year detail how integrative biological analyses can be accomplished in cardiac gene regulatory programs, brain injury, prostate cancer, septic shock, and colon cancer. In the case of colon cancer, the group showed that oncogenesis leads to the subversion of normal embryonic colon development.

**Eric Hall, PhD**

Dr. Hall's research has focused on the development of knowledge discovery tools that enable the identification of clinical concepts and care practice patterns in clinical documentation. Specifically, he has focused on the extraction of associations between nursing care patterns and patient outcomes in hospital Labor and Delivery.

**Michal Kouril, PhD**

Cincinnati Children's is implementing a new Research Data Storage (RDS) system. The goals of the RDS are: 1) to make centrally managed, secure data storage easily and transparently available to all researchers at Cincinnati Children's; 2) to manage growth of the storage system so that it is predictable for the institution, and 3) to provide basic amounts of storage for free and make larger allocations available for a very competitive and attractive price. This new system will accommodate data that are currently stored on external hard drives and will facilitate compliance with federal rules governing clinical research, as well as protect investigators from loss of data.

**Long (Jason) Lu, PhD**

Dr. Lu is a computational biologist with several ongoing projects:

1. Integrating diverse genomic features (microarray data, GO function annotation, proteomics data, etc) to predict protein-protein interaction networks as well as transcription regulatory network in yeast and mouse, and further prioritize protein pairs or TF-TG pairs for experimental validation.

2. Using advanced machine-learning and statistical techniques to identify key therapeutic targets in human microbial pathogens.

3. Computational modeling of the dynamics of morphogen diffusion in the fly embryo using imaging data.

4. Development of software to facilitate multi-source data integration and visualization.

**Jun Ma, PhD**

A remarkable feature of development is its reproducibility, the ability to correct embryo-to-embryo variations and instruct precise patterning. In *Drosophila*, embryonic patterning along the anterior-posterior axis is controlled by the morphogen gradient Bicoid, by mechanisms that remain controversial. Jun Ma recently performed quantitative studies to evaluate the behaviors of the native Bicoid gradient and its direct target Hunchback in early embryos. The results support a simple model of a robust Bicoid gradient sufficient to achieve scaled and precise activation of its target genes. These results will be published in the journal, Developmental Cell. Dr. Ma's group is now conducting
interdisciplinary studies to model the formation and action of the Bicoid gradient mathematically and to test the model directly by measurements in the developing embryo.

Jarek Meller, PhD

Jarek Meller's recent efforts are primarily focused on developing methods and tools for functional and structural annotation of proteins and their complexes. Several web servers and portals, such as SABLE and POLYVIEW-3D, have been developed to enhance such annotations, and are being widely used (over 250,000 submissions from over 80 countries to this time). Meller is also actively participating in a number of collaborative projects that involve applications of computational approaches to map, characterize and predict protein interaction sites, and other functional hot spots, that can be subsequently targeted for therapeutics. In collaboration with basic research and clinical colleagues, Meller is currently using docking simulations (utilizing BMI computational cluster for these CPU intensive tasks) to identify compounds that could attenuate protein signaling and other crucial events in the context of gastroenteritis, leukemia, and prostate cancer.

John Pestian, PhD, MBA

John Pestian's research group focuses on use of natural language processing (NLP) to analyze clinical free-text, with the goal of enhancing clinical processes and outcomes. This free-text can be discharge summaries, radiology reports or text books. While free-text is data, it is different from structured data, which have forced choices -- for example, zip codes, states, specific diagnostics or therapeutics. There are many methods for natural language processing, but Pestian's group is focusing on rule-based inference and neurocognitive computing. Together with collaborators, they are applying these methods to anonymize clinical free-text, cluster text into categories, develop corpora, develop visual languages and, most recently, develop artificial experts. The specific clinical problem they are addressing is the estimation of risk of recurrent suicide attempts based upon analyses of suicide notes using natural language processing.

Andrew Spooner, MD

Dr. Spooner is engaged in a collaborative project with pediatric investigators at Vanderbilt University. The aim is to fill two gaps in drug-related decision support in child health care: automated rounding of oral liquid drug doses to convenient and safe quantities, and representing formulas for pediatric drugs which are commonly compounded by compounding pharmacies.

Michael Wagner, PhD

Michael Wagner is a computational scientist who focuses on machine learning techniques to interpret proteomics data. With graduate student Rachana Jain he has been using large-scale linear programming techniques to improve database search engines commonly used for peptide mass fingerprinting. Dr. Wagner is also involved in collaborations to perform genome-wide association studies, where his work has concentrated on providing adequate, rapid data flow infrastructure, including the parallelization of genotype calling algorithms.

Division Publications


<table>
<thead>
<tr>
<th>PI</th>
<th>Title of Project</th>
<th>Agency</th>
<th>Grant Number</th>
<th>Start Date - End Date</th>
<th>Direct Costs</th>
<th>Total Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lu, L</td>
<td>Breast Cancer Biomarker Discovery Using Protein Network</td>
<td>American Cancer Society - National (University of Cincinnati)</td>
<td>01/01/08 - 12/31/08</td>
<td>$20,000 / $20,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ma, J</td>
<td>Activities Of The Bicoid Gradient In Drosophila Embryos</td>
<td>National Institutes of Health</td>
<td>R01 GM 072812</td>
<td>07/01/05 - 06/30/09</td>
<td>$180,154</td>
<td>$760,000</td>
</tr>
<tr>
<td>Pestian, J</td>
<td>Gene Expression in Pediatric Arthritis - Core B</td>
<td>National Institutes of Health</td>
<td>P01 AR 048929</td>
<td>08/22/03 - 07/31/08</td>
<td>$100,521</td>
<td>$304,734</td>
</tr>
<tr>
<td>Wagner, M.</td>
<td>Biomarkers for Amyotrophic Lateral Sclerosis in Active Duty Military</td>
<td>Department of Defense - Army (University of Cincinnati)</td>
<td>W81XWH-06-2-0016</td>
<td>01/23/06 - 01/22/09</td>
<td>$29,534</td>
<td>$85,238</td>
</tr>
<tr>
<td></td>
<td>Bayesian Mixture Modeling of Functional Genomics Data</td>
<td>National Institutes of Health (University of Cincinnati)</td>
<td>R01 HG 003749</td>
<td>07/01/06 - 06/30/10</td>
<td>$13,536</td>
<td>$53,679</td>
</tr>
<tr>
<td></td>
<td>Optimization of Structures and Networks of Proteins</td>
<td>National Institutes of Health (The University of Texas at Austin)</td>
<td>R01 GH 067823</td>
<td>06/01/08 - 05/31/11</td>
<td>$18,613</td>
<td>$57,535</td>
</tr>
<tr>
<td></td>
<td>Cincinnati Rheumatic Disease Core Center - Core 4</td>
<td>National Institutes of Health</td>
<td>P30 AR 047363</td>
<td>09/01/06 - 06/30/11</td>
<td>$53,834</td>
<td>$264,829</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>$618,827</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Current Year Direct** $773,182