NLM Informatics Training Conference 2012

University of Wisconsin-Madison, Madison, WI
June 26-27, 2012

Table of Contents

Table of Contents (Continued)

Table of Contents (Continued)

Table of Contents (Continued)

NLM Informatics Training Conference 2012

University of Wisconsin-Madison, Madison, WI
June 26-27, 2012

Agenda

Tuesday, June 26, 2012

7:00 AM – 7:45 AM	Breakfast **Microbial Sciences Building 1st Floor Atrium, 2nd Floor Café Area** Poster Setup – Day 1 Group **Biochemical Sciences Building Atrium**
7:45 AM – 8:00 AM	Welcome Remarks Computation and Informatics in Biology and Medicine Training Program George Phillips, Director Mark Craven, Incoming Director **Ebling Symposium Center, Microbial Sciences Building Room 1220**
8:00 AM – 8:20 AM	NLM Director's Remarks (Dr. Donald A.B. Lindberg)
8:20 AM – 8:30 AM	Introduction of Training Directors and Trainees (Valerie Florance)
8:30 AM – 9:30 AM	Plenary Paper Session #1 – 4 papers (1 hour) **Ebling Symposium Center – Microbial Sciences Building, Room 1220** **Moderator: George Hripcsak, Columbia University** • Adverse Drug Events Caused by Serious Medication Administration Errors (Kale - Harvard) • Providing Explanations for Unexpected Treatment Decisions (Wong - Pittsburgh) • Supervised Machine Learning Analysis of Drug Induced Gene Expression (Fernald - Stanford) • Pharmacovigilance and Drug Repurposing Model and Application: A Case Study (Haerian - Columbia)
9:30 AM –10:30 AM	Poster Session Day 1 Group **Biochemical Sciences Building Atrium** • **Health Care and Public Health**: Nwanze/Harvard; Ballard/Indiana; Carr/VA; Zirkle/VA; Misquitta/Harvard; Bahr/OHSU; Romagnoli/Pittsburgh; Alexander/Vanderbilt; Dunn/JHU • **Clinical/Translational**: McNamara/UCLA; Black/Washington; Gimenez/Stanford; Hinterberg/Colorado; Fearn/Washington; Jing/NLM • **Translational and Bioinformatics**: Vingara/OHSU; Collins/Wisconsin; Rance/NLM; Nabavi/Harvard; Gadala-Maria/Yale; Gordon/UC Irvine; Hooker-Rice/Baylor
10:30 AM –11:30 AM	Parallel Paper Focus Session A Focus Session A1 **Ebling Symposium Center, Microbial Sciences Building Room 1220**

Agenda (Continued)

Moderator: Larry Hunter, University of Colorado

- Coupling Semantic Predications and Microarray Data to Build Gene Expression Networks
  (Chen – NLM)
- Greedy Total Correlation Detects Functional Gene Sets in Cancer Copy Number Data
  (Melamed – Columbia)
- Using Natural Language Processing Methods to Improve EEG Classification in Brain-Computer Interfaces
  (Speier – UCLA)

Focus Session A2
Microbial Sciences Building Room 1520
Moderator: Pierre Baldi, University of California – Irvine

- Determining Protein Conformational Variability Using Robotic Motion Planning
  (Gipson – Rice)
- Improving Knowledge-Based Scoring Functions for Protein-Ligand Interactions by Accounting for Sparse Data in the Training Set
  (Grinter – Missouri)
- Mapping Protein Complex Interaction Networks by Quantitative Mass Spectrometry
  (Kaake – UC Irvine)

11:30 AM – 12:45 PM Lunch – with Birds of a Feather Tables
Wisconsin Institutes for Discovery Town Center

11:30 AM – 12:30 PM Executive Session of Training Directors
Microbial Sciences Building Room 6201, Faculty Conference Room
(Session Chair: Dr. Donald A.B. Lindberg)

11:45 AM – 12:30 PM Introduction to Career Development Awards and New Investigator R01 Grants (For Trainees In or Nearing Their Final Year)
Wisconsin Institutes for Discovery Room 3280

12:45 PM - 2:00 PM Plenary Paper Session #2 – 5 papers (1 hour, 15 minutes)
Ebling Symposium Center, Microbial Sciences Building Room 1220
Moderator: Alexa McCray, Harvard University

- Implementation of the ATHENA-HTN Clinical Decision Support System
  (Shluzas – Veterans Administration)
- Ontology-Guided Feature Engineering for Clinical Text Classification
  (Garla – Yale)
- Development and Application of ReO: a Biomedical Research Reagent Ontology
  (Brush – OHSU)
- Large-Scale Image Retrieval Using Text-Based and Content-Based Features
  (Simpson – NLM)
- Learning Context-Sensitive Causal Models
  (Garcia-Gathright – UCLA)

2:00 PM – 3:00 PM Open Mic Session - 13 talks (4.5 minutes each)
Ebling Symposium Center, Microbial Sciences Building Room 1220
Moderator: Perry Miller, Yale University

- Participatory Simulation for Health Care Policy Analysis
  (Engelhard – Virginia)

Agenda (Continued)

- Information Technology Needs of Remote Family Caregivers
  (Williamson – OHSU)
- Application Ontology for Medically Unexplained Syndromes
  (Doing-Harris – Utah)
- Prediction of Kidney Disease Outcomes Using Structured and Narrative Longitudinal Data
  (Pivovarov – Columbia)
- Weaving Clinical expertise into Online Peer-Patient Conversations
  (Huh – Washington)
- Evaluating Standard Terminologies for Encoding Allergy and an NLP based approach to documenting free-text Allergies
  (Goss – Harvard)
- Every Needle in a Haystack: Finding Fingerprints in a Safe Harbor Dataset
  (Fort – Columbia)
- Randomized Pilot Study to Improve the Quality of Patient Visits and the Quality of Documentation in Autism Clinic by Collecting Pre-Visit Data Using a Patient Portal
  (O'Rourke – Harvard)
- Population Genomics of the Mouse and its Functional Implications
  (Liu – Rice)
- Genetic Feedback Provides Robustness to Mutation
  (Marciano – Rice/Baylor)
- Building a Probeset Quality Measure and Development of a Validation Method on an Integrated Cancer Database
  (McDade – Pittsburgh)
- SAXS Restraints to Promote de Novo Protein Folding
  (Putnam – Vanderbilt)
- Pathway Correlation Profile of Gene-Gene Co-Expression for Identifying Pathway Perturbation
  (Tegge – Missouri)

3:00 PM – 4:00 PM Poster Session Day 1 Group
Biochemical Sciences Building Atrium

- **Health Care and Public Health**:
  Nwanze/Harvard; Ballard/Indiana; Carr/VA; Zirkle/VA; Misquitta/Harvard; Bahr/OHSU; Romagnoli/Pittsburgh; Alexander/Vanderbilt; Dunn/JHU

- **Clinical/Translational**:
  McNamara/UCLA; Black/Washington; Gimenez/Stanford; Hinterberg/Colorado; Fearn/Washington; Jing/NLM

- **Translational and Bioinformatics**:
  Vingara/OHSU; Collins/Wisconsin; Rance/NLM; Nabavi/Harvard; Gadala-Maria/Yale; Gordon/UC Irvine; Hooker-Rice/Baylor

4:00 PM – 5:00 PM Parallel Paper Focus Session B
Focus Session B1
Ebling Symposium Center, Microbial Sciences Building Room 1220
Moderator: Karen Eden, OHSU

- Using a Web-based Tool to Help Women Make Informed Choices About Breast Cancer Screening
  (Scariati – OHSU)
- Leveraging an Automated Cumulative Antibiogram for Clinical Decision Support
  (Fidahussein – Indiana)
- Towards a Clinically Efficient Decision Support Model Linked to the VA EMR
  (Phipps – Veterans Administration)

Agenda (Continued)

Focus Session B2
Microbial Sciences Building Room 1520
Moderator: Jude Shavlik, University of Wisconsin

- Integrating NLP and Sequence Features for More Accurate Function Prediction
  (Funk – Colorado)
- Predicting Dendrimer Cytotoxicity via Molecular Descriptors and Data Mining
  (Jones – Utah)
- Automation of In Situ Gene Expression at Single-Cell Resolution in C. Elegans to Study Development and Aging
  (Aerni – Stanford)

5:00 PM — Buses to Hotel and Dinner

6:00 PM - 9:00 PM — Dinner and Reception
Tripp Commons and Tripp Deck, 2nd Floor Memorial Union

8:00 p.m. – Open Mic Music Session (Dmitry Kondrashov, MC)

Wednesday, June 27, 2012

7:15 AM - 8:00 AM — Breakfast
Microbial Sciences Building 1st Floor Atrium, 2nd Floor Café Area

Poster Setup – Day 2 Group
Biochemical Sciences Building Atrium

8:00 AM – 9:15 AM — Plenary Paper Session #3; 5 papers (1 hour, 15 minutes)
Ebling Symposium Center, Microbial Sciences Building Room 1220
Moderator: Lydia Kavraki, Rice University

- Multi-Marker Tests and Proxy Association for Functional Rare Variants
  (Schuyler – Colorado)
- Inferring Metabolic Networks Using the Bayesian Graphical Lasso
  (Peterson – Rice)
- A Lineage-Hidden Markov Model for Genomic Annotation from Epigenetic Modifications
  (Biesinger – UC Irvine)
- Simulating Glycemic Variability in Critically Ill Burn Patients
  (Ortiz – Virginia)
- Pathway Index Models for Construction of Patient-Specific Risk Profiles
  (Eng – Wisconsin)

9:15 AM – 10:15 AM — Poster Session Day 2 Group
Biochemical Sciences Building Atrium

- **Health Care and Public Health:**
  Duval-Arnould/JHU; Natarajan/Columbia; Sakaguchi/Utah; Smith/Vanderbilt; Jacobs/Utah; Buell/Missouri; Hugine/Virginia; Ronquillo/Harvard; Morea/Indiana; Anderson/Missouri

- **Bioinformatics**:
  Overby/Columbia; Dvorkin/Colorado; Chen/Pittsburgh; Sun-Baylor/Rice; Poznik/Stanford; Muganda-Rippchen/Wisconsin; Evans/Yale; Eickholt/Missouri; Williams/Virginia; Kao/UC-Irvine

Agenda (Continued)

10:15AM –11:15 AM

University of Wisconsin-Madison Showcase
Ebling Symposium Center, Microbial Sciences Building Room 1220
Introduction: Mark Craven

Speaker: Richard Davidson
Departments of Psychology and Psychiatry and Waisman Center
Laboratory for Brain Imaging & Behavior, University of Wisconsin

11:15 AM -12:30 PM

Lunch
Wisconsin Institutes for Discovery Town Center

11:45 AM –12:30 PM

NLM – 2012 Biomedical Informatics Training Program Overview
(For Training Program Administrators)
Wisconsin Institutes for Discovery Room 3280

12:30 PM - 1:30 PM

Parallel Paper Focus Session C
Focus Session C1; (4 10-minute papers + 20 minutes for discussion)
Ebling Symposium Center, Microbial Sciences Building Room 1220
Moderator: John Hurdle, University of Utah

- Development and Evaluation of a Web-Based Electronic Medical Record Without Borders
  (Saavedra – Washington)
- HIE Utilization to Coordinate and Track Seasonal and H1N1 Flu Vaccination
  (Kudesia – Harvard)
- Investigating the Second Pandemic of the Black Plague through Metabolic Modeling
  (Baumler – Wisconsin)
- EHR Implementation Planning Processes in Critical Access Hospitals
  (Craven – Missouri)

Focus Session C2; (4 10-minute papers + 20 minutes for discussion)
Microbial Sciences Building Room 1520
Moderator: Ellen Bass, University of Virginia

- The Impact of Resident Shift Schedule on Handoff of Care
  (Price – Virginia)
- Information Flow as a Metric for Evaluating Clinical Documentation Systems
  (Khan – Vanderbilt)
- Which NLP Annotations Contribute to Accurate Automatic Problem List Generation?
  (Mowery – Pittsburgh)
- Representing Clinical Guideline Recommendations Using the Quality Data Model
  (Michel – Yale)

1:30 PM – 2:30 PM

Poster Session Day 2 Group
Biochemical Sciences Building Atrium

- **Health Care and Public Health**:
  Duval-Arnould/JHU; Natarajan/Columbia; Sakaguchi/Utah; Smith/Vanderbilt; Jacobs/Utah; Buell/Missouri; Hugine/Virginia; Ronquillo/Harvard; Morea/Indiana; Anderson/Missouri

- **Bioinformatics**:
  Overby/Columbia; Dvorkin/Colorado; Chen/Pittsburgh; Sun-Baylor/Rice; Poznik/Stanford; Muganda-Rippchen/Wisconsin; Evans/Yale; Eickholt/Missouri; Williams/Virginia; Kao/UC-Irvine

Agenda (Continued)

2:30 PM – 3:30 PM

Plenary Paper Session #4; 4 papers (1 hour)
Ebling Symposium Center, Microbial Sciences Building Room 1220
Moderator: Rebecca Crowley, University of Pittsburgh

- Building the Next Generation of Medical Education: Case Based Online Interactive
  Learning from Drawing Board to the Conference Room
  (Imler – Indiana)
- Exposing Benefits of Real-Time Tracking During Cancer Care with a Technology Probe
  (Patel – Washington)
- Feasibility of Collecting CAM Data Through a Computerized Patient Interview
  (Scarton – Utah)
- Structured Family Health Information using www.MyFamilyatVanderbilt.com
  (Holt – Vanderbilt)

3:30 PM – 3:45 PM

Closing Session and Poster Awards, Mark Craven and Valerie Florance
Ebling Symposium Center, Microbial Sciences Building Room 1220

PODIUM PRESENTATIONS (Listed Alphabetically by Author)			
Presenter	**Institution**	**Title**	**Page**
Aerni, Sarah	Stanford University	Automation of In Situ Gene Expression at Single-Cell Resolution in *C. elegans* to Study Development and Aging	58
Baumler, David J.	University of Wisconsin-Madison	Investigating the Second Pandemic of the Black Plague through Metabolic Modeling	92
Biesinger, Jacob	University of California, Irvine	A Lineage-Hidden Markov Model for Genomic Annotation from Epigenetic Modifications	67
Brush, Matthew H.	Oregon Health and Science University	Development and Application of ReO: a Biomedical Research Reagent Ontology	50
Chen, Jose G.	National Library of Medicine	Coupling Semantic Predications and Microarray Data to Build Gene Expression Networks	42
Craven, Catherine	University of Missouri	EHR Implementation Planning Processes in Critical Access Hospitals	93
Eng, Kevin H.	University of Wisconsin-Madison	Pathway Index Models for Construction of Patient-Specific Risk Profiles	69
Fernald, Guy	Stanford University	Supervised Machine Learning Analysis of Drug Induced Gene Expression	18
Fidahussein, Mustafa	Regenstrief Institute	Leveraging an Automated Cumulative Antibiogram for Clinical Decision Support	54
Funk, Christopher	University of Colorado	Integrating NLP and Sequence Features for More Accurate Function Prediction	56
Garcia-Gathright, Jean	University of California, Los Angeles	Learning Context-Sensitive Causal Models	52
Garla, Vijay N.	Yale University	Ontology-Guided Feature Engineering for Clinical Text Classification	49
Gipson, Bryant	Rice University	Determining Protein Conformational Variability Using Robotic Motion Planning	45
Grinter, Sam Z.	University of Missouri	Improving Knowledge-Based Scoring Functions for Protein-Ligand Interactions by Accounting for Sparse Data in the Training Set	46
Haerian, Krystl	Columbia University	Pharmacovigilance and Drug Repurposing Model and Application: A Case Study	19
Holt, Jonathan	Vanderbilt University	Structured Family Health Information Using www.MyFamilyatVanderbilt.com	101
Imler, Timothy D.	Regenstrief Institute Indiana University	Building the Next Generation of Medical Education: Case Based Online Interactive Learning from Drawing Board to the Conference Room	98
Jones, David E.	University of Utah	Predicting Dendrimer Cytotoxicity via Molecular Descriptors and Data Mining	57

PODIUM PRESENTATIONS (Continued)			
Presenter	**Institution**	**Title**	
Kaake, Robyn M.	University of California, Irvine	Mapping Protein Complex Interaction Networks by Quantitative Mass Spectrometry	47
Kale, Abhivyakti	Harvard Medical School	Adverse Drug Events Caused by Serious Medication Administration Errors	16
Khan, Naqi A.	Vanderbilt University	Information Flow as a Metric for Evaluating Clinical Documentation Systems	95
Kudesia, Valmeek	Harvard Medical School	HIE Utilization to Coordinate and Track Seasonal and H1N1 Flu Vaccination	91
Melamed, Rachel	Columbia University	Greedy Total Correlation Detects Functional Gene Sets in Cancer Copy Number Data	43
Michel, Jeremy J.	Yale University	Representing Clinical Guideline Recommendations Using the Quality Data Model	97
Mowery, Danielle	University of Pittsburgh	Which NLP Annotations Contribute to Accurate Automatic Problem List Generation?	96
Ortiz, Edward A.	University of Virginia	Simulating Glycemic Variability in Critically Ill Burn Patients	68
Patel, Rupa A.	University of Washington	Exposing Benefits of Real-Time Tracking During Cancer Care with a Technology Probe	99
Peterson, Christine B.	Rice University	Inferring Metabolic Networks Using the Bayesian Graphical Lasso	66
Phipps, Michael S.	Department of Veterans Affairs, VA Connecticut Healthcare System	Towards a Clinically Efficient Decision Support Model Linked to the VA EMR	55
Price, Taryn	University of Virginia	The Impact of Resident Shift Schedule on Handoff of Care	94
Saavedra, Francisco	University of Washington	Development and Evaluation of a Web-Based Electronic Medical Record Without Borders	90
Scariati, Paula	Oregon Health and Science University	Using a Web-based Tool to Help Women Make Informed Choices About Breast Cancer Screening	53
Scarton, Lou Ann A.	University of Utah	Feasibility of Collecting CAM Data Through a Computerized Patient Interview	100
Schuyler, Ronald	University of Colorado	Multi-Marker Tests and Proxy Association for Functional Rare Variants	65

PODIUM PRESENTATIONS (Continued)			
Presenter	**Institution**	**Title**	**Page**
Shluzas, Lauren A.	Department of Veterans Affairs, VA Palo Alto Health Care System	Implementation of the ATHENA-HTN Clinical Decision Support System	48
Simpson, Matthew S.	National Library of Medicine	Large-Scale Image Retrieval Using Text-Based and Content-Based Features	51
Speier, William	University of California, Los Angeles	Using Natural Language Processing Methods to Improve EEG Classification in Brain-Computer Interfaces	44
Wong, An-Kwok Ian	University of Pittsburgh	Providing Explanations for Unexpected Treatment Decisions	17

POSTER PRESENTATIONS (Listed Alphabetically By Author)			
Presenter	**Institution**	**Title**	**Page**
Alexander, Pauline T.	Vanderbilt University	A Rapid Assessment of Learning Style Preference	27
Anderson, Blake	University of Missouri	Mining for Common Behaviors in Image Analysis Using Eye-tracking Fixation Sequences	79
Bahr, Nathan	Oregon Health Science University	Alternative Design Considerations for Collaborative Medication Management	25
Ballard, Jeanne E.	Regenstrief Institute	Harnessing the Power of a Health Information Exchange for Post-Market Surveillance of Surgical Devices to Treat Pelvic Organ prolapse or Stress Urinary Incontinence	21
Black, Wynona	University of Washington-Seattle	Predicting Patient Risk; Using EMR Data to Build Models	30
Buell, Joseph	University of Missouri	Determining Credibility of Online Posts about Health	75
Carr, Thomas A.	VA Medical Center Indianapolis	Functionality of a Colorectal Cancer Personal Health Record: A User-Centered Study	22
Chen, Vicky	University of Pittsburgh	Identifying Functionally-Coherent Gene Subsets through Ontology Merging Algorithms	82
Collins, Maxwell D.	University of Wisconsin-Madison	Random Walker Cosegmentation for Medical Imaging	36
Dunn, Kyle S.	Johns Hopkins University	Threshold Analysis of Cancer Risk Acceptability Across U.S. Health Agencies	28
Duval-Arnould, Jordan	Johns Hopkins University	A Practical Tool to Improve Engagement in Quality and Safety Interventions	70
Dvorkin, Daniel	University of Colorado	Mixture Models vs. Supervised Learning for Integrative Genomic Analysis	81

POSTER PRESENTATIONS (Continued)			
Presenter	**Institution**	**Title**	**Page**
Eickholt, Jesse	University of Missouri-Columbia	Deep Networks and CUDA for Contact Prediction, Disorder Prediction and Gene Interaction	87
Evans, Perry	Yale University	Evaluating Melanoma Whole Exome Sequences Suggests New Driver Genes	86
Fearn, Paul A.	University of Washington	Enhancement and Integration of Caisis and REDCap for Longitudinal Follow-Up	33
Gadala-Maria, Daniel F.	Yale University	Pipeline for Analysis of Immunoglobulin Sequences on the Repertoire Scale	39
Gimenez, Francisco	Stanford University	Prediction of Radiologist Observations Using Computational Image Features: Method and Preliminary Results	31
Gordon, Elizabeth A.	University of California-Irvine	Computational Prediction of MAPK Substrates and Integrations with Phosphoproteomes	40
Hinterberg, Michael A.	University of Colorado-Denver	Selecting Important Features for Beta-Blocker Response in Heart Failure	32
Hooker, Stanley	Rice/Baylor College of Medicine	Keloid Susceptibility loci 1q41, 3q22, and 15q21 Replicated in a Nigerian Population	41
Hugine, Akilah L.	University of Virginia	Rectangular Area Judgments as a Function of Size Difference and Placement	76
Jacobs, Jason R.	University of Utah	Evaluation of Methods for Translating the CCD to the vMR	74
Jing, Xia	National Library of Medicine	Application of a Graphical Method for Analyzing Large Data Sets to Aid Knowledge Discovery	34
Kao, Athit	University of California-Irvine	Mapping the Topology of the 19S Proteasome Using Cross-Linking Mass Spectrometry and Probabilistic Modeling	89
McNamara, Mary	University of California-Los Angeles	A Patient Portal for Radiology with a Consumer-Oriented Terminology	29
Misquitta, Donald	Harvard Medical School-Boston	Predictors of Use of an IVR System using Multivariate Adaptive Regression Splines	24
Morea, Justin	Regenstrief Institute	Adding Social Networking to the EHR	78
Muganda-Rippchen, Deborah	University of Wisconsin – Madison	Computing Clustered Alignments of Gene-Expression Time Series	85
Nabavi, Sheida	Harvard Medical School-Boston	Comparative Analysis of Sequence-Based Copy Number Variation Detection Methods	38
Natarajan, Karthik	Columbia University	Creation of a Gold Standard for EHR-Based Information Retrieval	71

POSTER PRESENTATIONS (Continued)			
Presenter	**Institution**	**Title**	**Page**
Nwanze, Chukwuemeka	Massachusetts General Hospital-Boston	Bevacizumab vs. Ranibizumab for Macular Degeneration: A Cost-Effectiveness Analysis	20
Overby, Casey	Columbia University	Estimating Heritability Captured by Common Variants in Pharmacological Traits	80
Poznik, David	Stanford University	Demographic Insight from Sequencing Y Chromosomes in Diverse Populations	84
Rance, Bastien	National Library of Medicine	Identifying Missed Synonymy in the UMLS	37
Romagnoli, Katrina M.	University of Pittsburgh	Visiting Nurses' Perceptions of Geriatric Post-hospitalization Information Needs	26
Ronquillo, Jeremiah G.	Massachusetts General Hospital, Boston	Social Network Analysis of Genetic Testing Relationships	77
Sakaguchi, Farrant	University of Utah	The Quality of Discharge Summaries for Facilitating Transitions of Care	72
Smith, Joshua C.	Vanderbilt University	Exploring Adverse Drug Effect Discovery from Data Mining of Clinical Notes	73
Sun, Jiayi M.	Rice/Baylor College of Medicine	Epigenomic Profiling of the Osteosarcoma Genome	83
Vingara, Lisa K.	Oregon Health and Science University	Visualization of Quality for Magnetic Resonance Spectroscopic Imaging	35
Williams, Aaron L.	University of Virginia	Modeling the Elastic Change of Depolarization in a Sensory Neuron	88
Zirkle, Maryan	Oregon Health Sciences University	Toward the Post Traumatic Stress Disorder Ontology: Understanding Symptoms and Treatments	23

Open Mic Session Presentations

Health Care & Public Health Informatics		
Presenter	**Institution**	**Title**
Engelhard, Matthew	University of Virginia	A Participatory Simulation for Health Care Policy Analysis
Goss, Foster	Harvard Medical School	Evaluating Standard Terminologies for Encoding Allergy and an NLP-Based Approach to Documenting Free-Text Allergies
O'Rourke, Julia	Harvard Medical School	A Randomized Pilot Study to Improve the Quality of Patient Visits and the Quality of Documentation in Autism Clinic by Collecting Pre-Visit Data Using a Patient Portal
Pivovarov, Rimma	Columbia University	Prediction of Kidney Disease Outcomes Using Structured and Narrative Longitudinal Data

OPEN MIC SESSION PRESENTATIONS (Continued)		
Williamson, Steven	Oregon Health Science University	Information Technology Needs of Remote Family Caregivers

Translational Bioinformatics and Clinical Research Informatics		
Presenter	**Institution**	**Title**
Doing-Harris, Kristina	University of Utah	Application Ontology for Medically Unexplained Syndromes
Fort, Daniel	Columbia University	Every Needle in a Haystack: Finding Fingerprints in a Safe Harbor Dataset
Huh, Jina	University of Washington	Weaving Clinical Expertise Into Online Peer-Patient Conversations
Liu, Kevin	Rice University	Population Genomics of the Mouse and its Functional Implications
Marciano, David C.	Baylor College of Medicine	Genetic Feedback Provides Robustness to Mutation
McDade, Kevin	University of Pittsburgh	Building a Probeset Quality Measure and Development of a Validation Method on an Integrated Cancer Dataset
Putnam, Daniel	Vanderbilt University	SAXS Restraints to Promote de Novo Protein Folding
Tegge, Allison N.	University of Missouri	Pathway Correlation Profile of Gene-Gene Co-Expression for Identifying Pathway Perturbation

Tuesday
June 26, 2012

Tuesday – Table of Contents

Adverse Drug Events Caused by Serious Medication Administration Errors

Authors:
Abhivyakti Kale, Carol A Keohane, Saverio Maviglia, Tejal K Gandhi, Eric G Poon
Brigham & Women's Hospital, Boston, MA
Harvard Medical School, Boston MA

Abstract:
Medication errors occur at every stage, but those that occur at the medication administration stage are important because they are rarely intercepted and have a high potential to cause harm. A study performed in 36 hospitals showed 19% medication administrations contained an error and seven percent administration errors had the potential to cause patient harm and were judged potential ADEs. With so many medication doses administered, the potential for harm cannot be underestimated. Past studies have estimated error rates but very few have established how many lead to patient harm.

We performed a retrospective chart review for cases of known medication administration errors established in a previous study; in order to assess the relationship between serious medication administration errors and actual adverse drug events (ADE) in a hospital setting. We found that serious medication administration errors with a potential to cause harm (potential ADE) can cause serious patient harm if not intercepted. 7.5% serious and life threatening potential ADE result in ADE. In a hospital where 6 million medication doses are administered per year, about 4000 preventable ADEs are attributable to medication administration errors annually. These findings will have a significant influence on hospital leaders and policy makers.

Email of First Author: abhivyakti_kale@hms.harvard.edu

Providing Explanations for Unexpected Treatment Decisions

Authors:
An-Kwok Ian Wong, MS, Shyam Visweswaran, MD, PhD, University of Pittsburgh

Abstract:
Methods that identify unexpected events during medical care can be useful in the development of automated clinical alerting systems to alert clinicians to treatment choices that warrant additional consideration. We have developed a method for identifying unexpected medication administration in the intensive care unit that is based on learning naive Bayes models from past patient data that when applied to current patient data identifies statistically unusual treatment decisions. However, explaining why an unusual treatment decision is unexpected to a clinician is equally important, and the provision of such explanations will enhance the utility of automated clinical alerting systems. The preliminary results reported here describe the statistical properties of identified unexpected treatment decisions and the types of explanations that can be automatically generated. Future work will identify the most effective methods for offering explanations that are useful in automated clinical alerting systems.

Email of First Author: aiw5@case.edu

Supervised Machine Learning Analysis of Drug Induced Gene Expression

Authors:
Guy Fernald, Russ Altman, Stanford University

Abstract:
Understanding the mechanism of action for a drug is essential for drug development and optimization. Drug failure rates have increased in recent years and most often the underlying cause for failure remains unknown. By studying properties of small molecules and analyzing drug induced gene expression researchers have identified molecular targets and gained insight into the mode of action for some drugs. However the effect a drug will have on a given tissue is difficult to predict. The accumulation of large repositories of studies of drug induced RNA expression changes presents an opportunity to relate molecular properties of small molecules to gene expression changes. We apply supervised machine learning techniques to relate properties of small molecules to significant gene expression changes in a large publicly available gene expression database of cancer tissues. The method successfully identifies features of small molecules that result in significantly increased expression of a subset of genes. Further analysis of these features identifies properties of small molecules that can be applied to rational drug development.

Email of First Author: guyhf@stanford.edu

Pharmacovigilance and Drug Repurposing Model and Application: A Case Study

Authors:
Krystl Haerian, Carol Friedman, Columbia University

Abstract:
An important component of drug safety, pharmacovigilance, entails detecting the risk of unknown serious adverse drug reactions. The primary goal of our study was to use the information in free text EHRs to develop a model for drug induced valvular heart disease (VHD) that can translate retrospective health data into knowledge. Our VHD model was based on 123,177 inpatients. We used natural language processing to obtain information from discharge summaries and linear regression to control for potential confounders. The model was applied to screen known drugs, published drug suspects (identified though binding assays and quantitative structure-activity relationship (QSAR) models), and other drugs of interest. Our model identified a positive signal with the known valvulopathogens, fenfluamine and pergolide. We also found that two atypical antipsychotic medications resulted in a lower than expected prevalence of VHD.
NLP and statistical modeling of EHR data is useful for pharmacovigilance studies. Furthermore, existing EHR data can be utilized to perform translational studies, examining molecular based theories in observational databases. In addition, the unexpected finding of drugs that lowered the risk of VHD has preliminary implications that the EHR may also be a valuable source of information for drug repurposing studies.

Email of First Author: krh7003@dbmi.columbia.edu

Bevacizumab vs. Ranibizumab for Macular Degeneration: A Cost-Effectiveness Analysis

Authors:
Chukwuemeka Nwanze[1,2], Abumere Akinwale[3], Ron Adelman[4]
1. Massachusetts General Hospital, Boston
2. Harvard Medical School, Boston
3. Joslin Diabetes Center, Boston
4. Yale Department of Ophthalmology and Visual Science, New Haven

Abstract:
-OBJECTIVE: To evaluate the cost-effectiveness of monthly and as-needed dosing protocols using ranibizumab or bevacizumab for the treatment of neovascular age-related macular degeneration (AMD), when the treatment costs of severe ocular and systemic adverse events are considered.
-METHODS: A Markov model was developed to assess the cost effectiveness of each of the following protocols: monthly ranibizumab, monthly bevacizumab, as-needed ranibizumab and as-needed bevacizumab. Direct costs and utilities were assessed from the perspective of a third-party payer or an insurance company. Cost effectiveness was evaluated in 2011 US dollars per quality-adjusted life year (QALY).
-RESULTS: Considering the treatment costs of severe medical and ocular adverse events, the cost effectiveness of each protocol is as follows: monthly ranibizumab $63,333 /QALY, ranibizumab as needed $18,571/QALY, bevacizumab monthly $2,676/QALY and bevacizumab as needed $3,333/QALY. Sensitivity analysis of the treatment costs of medical and ocular adverse events demonstrated minimal impact on relative cost-effectiveness.
-CONCLUSION: At current prices, monthly bevacizumab is the most cost-effective anti-VEGF AMD treatment protocol. Ranibizumab is as cost effective as bevacizumab at a maximum price of $158 per dose.

Email of First Author: mnwanze@yahoo.com

Harnessing the Power of a Health Information Exchange for Post-Market Surveillance of Surgical Devices to Treat Pelvic Organ Prolapse or Stress Urinary Incontinence

Authors:
Jeanne E Ballard, Marc B Rosenman, Regenstrief Institute

Abstract:
Surgical devices to treat pelvic organ prolapse (POP) or stress urinary incontinence (SUI) occasionally fail or are associated with complications. On January 3, 2012, the FDA issued a 522 Post-Market Surveillance order to 33 manufacturers of 88 devices that treat POP, and 7 manufacturers of 11 devices that treat SUI. The 522 order requires tracking systems and reporting of device malfunctions, serious injuries, or deaths. The FDA also has a safety surveillance strategy that relies upon voluntary reporting of medical device failures and complications by healthcare providers, institutions, device manufacturers, and patients. However, less than 0.5% of device failures are captured this way.

Using data from the Indiana Network for Patient Care, we are creating a computer-based database extraction method including automated searches of tree-text operative notes to identify surgical cases where POP or SUI devices are used. After the cases have been identified and the method validated, we can perform regular surveillance across multiple hospital systems of the frequency of use and of complications associated with these devices. Detectable complications include those identified during the hospitalization when the device is implanted, or those that resulted in an admission. This method could also be used to populate device registries longitudinally.

Email of First Author: jballard@regenstrief.org

Functionality of a Colorectal Cancer Personal Health Record: A User-Centered Study

Authors:
Thomas A Carr, Matthew Walsh, Michael Weiner, David A Haggstrom, Department of Veterans Affairs, Richard L. Roudebush VA Medical Center Indianapolis.

Abstract:
Cancer Personal Health Records (PHRs), like other chronic disease PHRs (CD PHRs), have unique information-handling demands; disease management function needs, and interface usability issues. In this study, we endeavored to define the requirements of a colorectal cancer (CRC) PHR, and differentiate it from other CD PHRs. Information was gathered in the context of a Human-Computer Interaction evaluation of a CRC PHR, developed for the Veterans Health Administration at the Center of Excellence on Implementing Evidence-based Practice/Health Services Research and Development, Roudebush VA Medical Center, Indianapolis, Indiana. Video/audio tapes of focus groups, structured interviews, and performance-based testing were reviewed. Observations were transcribed onto “sticky” notes and underwent grounded theory affinity analysis to identify minor and major themes. The CRC PHR and other cancer PHRs were very similar, but they differed from other CD PHRs in access controls, journaling function, and the emphasis on cancer surveillance tracking. With the degree of similarity among the electronic cancer PHRs, it should be possible to create a single cancer PHR.

Email of First Author: carrt@iupui.edu

Toward the Post Traumatic Stress Disorder Ontology: Understanding Symptoms and Treatments

Authors:
Maryan Zirkle[1,2], David Hickam[1,2], Bryan T Gamble[2]
[1]Department of Veterans Affairs, Portland VA Medical Center, [2]Oregon Health Sciences University

Abstract:
Post Traumatic Stress Disorder (PTSD) is common among VA patients and is an important problem among OEF/OIF veterans. Given the heterogeneous nature of the disorder, there is a need for a better understanding of this domain. However, a formal explicit description of concepts has not been previously developed. We are developing ontology to capture this knowledge specific to PTSD. In order to understand PTSD symptoms and treatments, data will be collected from several resources: focus groups, cognitive interviews, clinical guidelines, SNOMED-CT, text mining, chart annotations, and natural language processing. Data have been gathered from several of these resources thus far, yielding 43 terms from focus groups, 158 terms from guidelines, 172 terms from SNOMED-CT, 20 terms from text mining, and 985 terms from annotations. Each concept in our ontology will be clearly defined and uniquely identified to capture the distinct, direct and indirect relationships of these PTSD concepts. This ontology will help to answer questions about the existence of patterns in constellations of symptoms and treatment delivered. Completeness of the ontology will be assessed by a second round of annotations of progress notes from VA mental health clinics.

Email of First Author: maryan.zirkle@va.gov

Predictors of Use of an IVR System using Multivariate Adaptive Regression Splines

Authors:
Donald Misquitta[1,2], Robert Friedman[2]
1. Harvard Medical School, Boston
2. Boston University, Boston

Abstract:
Understanding the relationship between the factors that mediate use of health intervention systems and the type of intervention are important for the design of multimodal systems and for their impact on patient outcomes. While information exists on predictors of use of various novel health interventions, there is a small amount of data on demographic factors that mediate the use of interactive voice response systems, with prior studies showing no significant predictors. Telephone-linked care (TLC), an IVR system, has been shown in randomized controlled trials to improve patient outcomes. The system carries out fully automated, interactive conversations with patients using both patient-specific data and patient responses to questions to determine how to navigate down a large network of possible conversations.

To evaluate predictors of use, a recent TLC study of 69 intervention patients was analyzed using multivariate adaptive regression splines to model the relationship between predictors and the outcome of total number of completed calls during the study period. This technique has the ability to look for nonlinearities and complex interactions among variables. Of the 13 predictors, none were found to be individually important, but the results suggested that the combination of middle age and COPD predicted increased use.

Email of First Author: donald_misquitta@hms.harvard.edu

Alternative Design Considerations for Collaborative Medication Management

Authors:
Nathan Bahr, Ravi Teja Bhupatiraju, Pamela Lam, Paul N Gorman, Oregon Health & Science University

Abstract:
A “medication list” is a moving target: antihypertensives are adjusted to achieve a goal; antibiotics added for infection; analgesics adjusted, stopped, resumed to control pain. Physicians, consultants, pharmacists, and patients track medications in systems tailored to individual tasks and organizational requirements. Rarely are systems integrated so everyone is “on the same page.” The result is medication lists which do not match, jeopardizing patient safety and increasing workload as clinicians re-enter data and resolve discrepancies among lists. We aim to support collaborative medication management by enabling interaction among systems and users so that medication lists are maintained in synchrony. We assume asynchronous, distributed activity by multiple users of diverse applications in separate organizations. Modeled after the Common Framework, we developed a prototype to support distributed, collaborative medication management. Here we review different successful approaches used for collaboration in other domains and describe our use of a distributed versioning system as a platform to implement a shared medication list management system.

Email of first author: bahrn@ohsu.edu

Visiting Nurses' Perceptions of Geriatric Post-hospitalization Information Needs

Authors:
Katrina M Romagnoli, Steven M Handler, Frank M Ligons, Harry Hochheiser, University of Pittsburgh

Abstract:
Geriatric patients experience information needs at discharge that might contribute to undesirable outcomes, including readmission. As the only clinicians who provide medical care in the home, home-care nurses have unique insight into these needs. We conducted a series of qualitative investigations aimed at understanding home care nurses' perceptions of information challenges faced by these patients. Two Nominal Group Technique (NGT) sessions with seventeen home care nurses produced a list of 28 information needs grouped into six themes: medications, disease/condition, discharge, non-medication care, communication, and functional limitations. The list was used, along with prior research, to develop a survey administered to home care nurses' to elicit perceptions of patients' post-hospitalization information needs. The survey was sent to 220 home-care nurses, with a 54.1% (119/220) response rate. Respondents identified several information needs that have not been previously discussed in the readmission literature, including information about medication regimens; the severity of their condition; hospital discharge process; non-medication care regimens; the extent of care needed, and which providers are best suited to provide that care.
These results might be used to develop interventions that may improve information sharing among clinicians, patients and caregivers during care transitions, potentially preventing unplanned hospital readmissions.

E-mail of First Author: kak59@pitt.edu

A Rapid Assessment of Learning Style Preference

Authors:
Pauline T Alexander, Nunzia B Giuse, Taneya Y Koonce, Vanderbilt University

Abstract:
Background: Learning style, a cognitive preference for gathering, organizing and thinking about information, represents an important aspect of patient communication. Study investigators developed the Fast Learning Preference Screen (FLPS) in response to the need for a rapid, clinically relevant and appropriate learning style preference assessment.

Objective: To compare the results of the newly developed FLPS and the VARK™ learning style preference assessment tool.

Methods: This cross-sectional study compared data that resulted from use of the FLPS and VARK™ between May - June 2011 in English speaking, adult, emergency department (ED) patients. After administering the VARK and the FLPS to all patients, study investigators established concordance by asking participants whether they agreed or disagreed with the results for each measure.

Results: Of 216 patients screened, 101 participants met inclusion/exclusion criteria and were included in the analysis. Patients' agreement rate was 90% (91/101) for the VARK™ and 92% (93/101) for the FLPS. McNemar's Test found no statistical difference between agreement rates (p=0.77).

Conclusions: Investigators note the possible advantages of using a rapid assessment of learning style preferences in healthcare settings given the uniqueness of the environment. Future research should focus on developing tools designed for this setting and work towards validating their effectiveness.

Email of First Author: pauline.alexander@vanderbilt.edu

'
'

Threshold Analysis of Cancer Risk Acceptability Across U.S. Health Agencies

Author:
Kyle S Dunn, Johns Hopkins University

Abstract:
United States federal agencies enforce distinct limits on human exposure to cancer-causing agents. The limits relate to either an acceptable dose received by an individual, or to an acceptable number of resulting cancer cases in a population. The acceptability of cancer risk is defined differently by each agency and determined by factors outside toxicological or epidemiological research. This study aims to identify the major political, technological, economic and social determinants that contribute to acceptable risk, and to construct agency-specific decision thresholds based on the respective weight of each determinant. Exposure to ionizing radiation has been chosen as a data-rich case study with application across a number of federal agencies, including the Occupational Safety and Health Administration (OSHA), the Environmental Protection Agency (EPA), the Food and Drug Administration (FDA), the Transportation Security Administration (TSA) and the Consumer Product Safety Commission (CPSC). Divergent radiation limits among these agencies reflects the lack of a unified federal framework for protecting the public's health. Threshold analysis is presented as means of acknowledging the separate influences met by each agency, and characterizing a best practices analysis to condense cancer risk decision-making into a single framework.

Email of First Author: kdunn@jhsph.edu

A Patient Portal for Radiology with a Consumer-Oriented Terminology

Authors:
Mary McNamara, Corey Arnold, Suzie El-Saden, Ricky Taira; University of California, Los Angeles

Abstract:
The development of a personal health record (PHR) cannot rely on the transferring of electronic health record (EHR) content verbatim. While PHR content should be modeled on EHR content, consumer information visualization should be specific to the needs of the patient. Visualization models for PHRs help to structure content in a way that is relevant to the patient's mental model of their condition. In addition, there is a significant difference in the vocabulary used by practitioners and patients. Simply making the same information available to patients that is used by practitioners does not address the issue of health literacy. The Open Access Collaborative Consumer Health Vocabulary (OAC CHV) ontology attempts to collect all terminology and definitions relevant to consumers across the spectrum of medicine. This method is in contrast to other medical ontologies, which tend to focus on a specific subdivision such as laboratory results (e.g., LOINC) or clinical drug names (e.g., RxNorm). Thus, while the OAC CHV includes a wide variety of terms, explanations for terms are occasionally missing or are too general to be insightful. This project extends the OAC CHV, focusing on the area of neuro-radiology. We created a patient portal that provided access to neuro-radiology records by displaying the conclusion section of radiology reports. Terms deemed too difficult for the average consumer were defined using mouse roll-overs that, when possible, were taken from the OAC CHV. However, when the OAC CHV lacked a definition for the term, or the term was unspecific, we manually generated definitions with the help of a physician.

Email of First Author: mmcnamara@ucla.edu

Predicting Patient Risk; Using EMR Data to Build Models

Authors:
Wynona Black[1], Eric Horvitz[1,2], Meliha Yetisgen-Yildiz[1] & John H Gennari[1]
[1]Biomedical & Health Informatics, University of Washington, Seattle;
[2]Microsoft Research, Redmond

Abstract:
Acute Coronary Syndrome (ACS), which encompasses unstable angina, non-ST elevated (NSTEMI) and ST-elevated (STEMI) myocardial infarction, is the primary cause of an estimated 1.4 million hospitalizations in 2005 and costs over $300 million annually. As EMR use increases in the US (especially with the passage of the HITECH Act), there is an opportunity to analyze EMR patient data to build better risk models of patient outcomes. The Acute Coronary Syndrome Patient Database (ACSPD) is a unique, very large, 9-year patient database derived from EMR data from 128 health care institutions across the United States. In patient care, information about outcome risk is beneficial, and depends on the discriminant ability of the risk model. We explored patient mortality and found the following: 1) To predict patient risk, one needs a broad range of features, including patient demographics, the type of care facility, and adherence to clinical practice guidelines. 2) The specific features that influence patient outcome vary by syndrome. For example, the teaching status of a healthcare facility influences mortality risk for NSTEMI patients (teaching: 9.2% mortality versus non-teaching 11.6%, p-value <0.0001), but is not a factor for STEMI patients (teaching 8.1% versus non-teaching 8.7%, p-value = 0.17).

Email of First Author: blackw@uw.edu

Prediction of Radiologist Observations Using Computational Image Features: Method and Preliminary Results

Authors:
Francisco Gimenez, Jiajing Xu, Christopher Beaulieu, Daniel Rubin, Sandy Napel, Stanford University

Abstract:
We aim to predict radiological observations using computationally-derived imaging features extracted from computed tomography (CT) images. We created a dataset of 79 CT images containing liver lesions identified and annotated by a radiologist using a controlled vocabulary of 76 semantic terms. Computationally-derived features were extracted describing intensity, texture, shape, and edge sharpness. L_1-regularized logistic regression (LASSO) was used in order to predict the radiological observations using computational features. The approach was evaluated by leave one out cross-validation. Informative radiological observations such as lesion enhancement, hypervascular attenuation, and homogeneous retention were predicted well by computational features. By exploiting relationships between computational and semantic features, this approach could lead to more accurate and efficient radiology reporting.

Email of First Author: fgimenez@stanford.edu

Selecting Important Features for Beta-Blocker Response in Heart Failure

Author:
Michael A Hinterberg, University of Colorado-Denver

Abstract:
Congestive heart failure (CHF) is a debilitating and costly disease, affecting more than 5 million people in the United States, and responsible for 3.4 million hospital visits annually. Beta-blocker therapy has been shown to be effective for some but not all patients, while efficacy of beta-blocker treatment is not evident for months, in which case more invasive surgical options may be suggested. Therefore, predicting which patients are likely to be responsive to beta-blocker therapy is important for ensuring appropriate treatment as soon as possible.

In order to characterize the phenotype of responsive patients, we obtained various sources of patient data from electronic medical records, including demographic data, lab tests, echocardiogram measurements and coded observations, nuclear medicine, and catheterization, in a cohort of 45 patients with congestive heart failure, across two time points. We developed a script to combine these data sources, along with mRNA and miRNA expression data, and compared several linear and non-linear models for predicting outcomes, using techniques such as principal components analysis, random forests, and support vector machines. Different sets of explanatory phenotype features were found between the linear and non-linear models, which may provide insights on heart remodeling.

Email of First Author: michael.hinterberg@ucdenver.edu

Enhancement and Integration of Caisis and REDCap for Longitudinal Follow-Up

Authors:
Paul A Fearn, Daniel R Masys, Peter Tarczy-Hornoch, University of Washington

Abstract:
At high volume centers of excellence or tertiary care centers, many patients return to their primary or secondary healthcare providers for long-term follow-up. Systematically assessing the long-term effectiveness of major treatments can become a complex and expensive process to manage. Several major diseases and procedures are tracked and reported through registries (e.g. tumor registries). However, the data collected by registries may be inadequate for secondary use in clinical or translational research. This project aims to improve the coverage, effectiveness and efficiency of long-term follow-up for specified patient populations.

Given the economic constraints on biomedical research and informatics infrastructure, increasing demand for interoperability, the difficulties in achieving widespread adoption of new systems, we based our solution to this problem on the integration and enhancement of two popular open source platforms, Caisis and REDCap. In collaboration with the software developers, care providers, and research stakeholders, we developed a detailed functional specification. For our pilot implementation, required enhancements to Caisis have been released in version 6.0, and enhancements to REDCap are in progress.

We will use criteria developed by the caBIG Oversight Ad Hoc Subcommittee to evaluate the specifications, progress and outcomes of this pilot project.

Email of First Author: fearnp@uw.edu

Application of a Graphical Method for Analyzing Large Data Sets to Aid Knowledge Discovery

Authors:
Xia Jing, James J Cimino, Lister Hill NCBC, National Library of Medicine, NIH

Abstract:
Objective: To demonstrate an application of a graphical method for reducing and analyzing large data sets in which the data are coded with a hierarchical terminology; to facilitate visualization in clinical research and guide knowledge discovery.

Methods: Data in clinical data repositories are often coded using hierarchical terminologies. These data can be aggregated based on their classification. We use a combination of aggregated class counts and differences between data sets to reduce the numbers of nodes in hierarchical graphs, allowing important areas of the hierarchies to be highlighted. Details of the technique, including guidelines for thresholds setting, are presented.

Results: We queried a clinical data repository (BTRIS) in NIH on the ICD9-CM diagnoses of patients who had been exposed to either of two similar drugs (pioglitazone and rosiglitazone). After reducing the size of the data sets, the graphs of the ICD9-CM hierarchy highlight important areas that differ between the two groups, confirming previously known adverse effects.

Conclusions: This is a demonstration of the new graphic method that can be applied to the kinds of data sets typically obtained from clinical data repositories. Our method can aid knowledge validation by highlighting expected results and knowledge discovery by revealing unexpected results.

Email of First Author: xia.jing@nih.gov

Visualization of Quality for Magnetic Resonance Spectroscopic Imaging

Authors:
Lisa K Vingara, Manoj Sammi, Bill Rooney, Eilis Boudreau, Oregon Health and Science University

Abstract:
Neurological disorders are challenging to study given the inaccessibility of the brain. A means to investigate metabolic information noninvasively from the brain is to use Magnetic Resonance Spectroscopic Imaging (MRSI). MRSI allows for numerous spectra to be obtained from small volumes throughout the brain, yielding a valuable source of metabolic information. A typical dataset contains hundreds of spectra per subject. Due to the complexity of data acquisition, some spectra may be of poor quality due to excess noise, motion, or other technical artifacts. These poor quality spectra may skew the results of further analysis. Herein, we visualize various quality measures of each individual spectrum, such as signal to noise ratio and spectral linewidths, in order to quickly and efficiently evaluate individual spectral quality as well as the quality of the overall dataset. We demonstrate its usability on phosphorous MRSI datasets.

Email of First Author: vingara@ohsu.edu

Random Walker Cosegmentation for Medical Imaging

Authors:
Maxwell D Collins, Jia Xu, Leo Grady, Vikas Singh, University of Wisconsin-Madison

Abstract:
Cosegmentation is the task of selecting similar regions from a set of related images. We recast the cosegmentation problem using Random Walker (RW) segmentation as the core segmentation algorithm, rather than the traditional MRF approach adopted in the literature so far. We show that the resulting optimization can be expressed in terms of linear algebra operations on sparse matrices which are easily mapped to a GPU architecture. Our model further allows an optimization scheme exploiting quasiconvexity for model-based segmentation with no dependence on the scale of the segmented foreground, enabling robust template-based labeling of known image regions. We provide a highly specialized CUDA library for cosegmentation exploiting this special structure. Our implementation scales to large images more efficiently than the previous state of the art, making possible effective use on biomedical images. We use this method in an interactive setting to perform region of interest (ROI) extraction, a key step in research using medical imaging.

Email of First Author: mcollins@cs.wisc.edu

Identifying Missed Synonymy in the UMLS

Authors:
Bastien Rance, Olivier Bodenreider, National Library of Medicine, NIH

Abstract:
The Unified Medical Language System (UMLS) is a terminology integration system developed by the National Library of Medicine. The 2011AB version of the UMLS integrates concepts from 161 biomedical vocabularies and contains more than 2.6M concepts and 8M synonyms. The UMLS Metathesaurus groups into a single concept all synonymous names found in the source vocabularies for a given biomedical entity. The integration of the source vocabularies associates automatic lexical processing and human review.

Despite the quality assurance procedures built in the UMLS development process, errors have been reported, which it is not surprising given the magnitude of the Metathesaurus. Such errors are either present in the source terminologies or generated during the integration process.

In this study, we focus on the detection of missed synonyms, i.e. distinct UMLS concepts that have the same meaning and should be merged. We use a variety of features to identify such errors, namely lexical, semantic and structural information in the UMLS, as well as contextual information in the biomedical literature and internet sources. Using a machine learning approach, we identified 545 potential missed synonyms among 8451 candidates. Some of these errors have already been reported to and corrected by the developers of source terminologies.

Email of First Author: bastien.rance@nih.gov

Comparative Analysis of Sequence-Based Copy Number Variation Detection Methods

Authors:
Sheida Nabavi, Zhengqiu Cai, and Peter J Tonellato, Harvard Medical School, Boston

Abstract:
Copy number variations (CNVs) are an important type of genetic variation associated with a diverse range of disease. Identification of disease associated CNVs provides insight into disease progression and treatment. Recently, next generation sequencing (NGS) technologies have created an opportunity for detecting CNVs with higher accuracy and resolution. In this work, we compare the performance of six publically available sequence-based CNV detection tools, using eight breast cancer cell line and six synthesized NGS datasets. We compare sensitivity, precision and breakpoint accuracy of the tools. We also analyze statistical characteristics of detected CNVs and the computational requirements of the tools. The results indicate tools that employ more advanced algorithms and incorporate information imbedded in paired-end data perform well, while their computational cost is higher. However, the tool's precisions are not very high which results in false positive CNVs. Also we observed that, there are few consensus CNVs across the tools in the breast cancer datasets. These indicate more efficient and advanced algorithms are required for detecting true CNVs. The results of this work can facilitate selection of an appropriate CNV detection method and can serve as a guide to develop new sequence-based CNV detection algorithms to address the current limitations.

Email of First Author: sheida_nabavi@hms.harvard.edu

Pipeline for Analysis of Immunoglobulin Sequences on the Repertoire Scale

Authors:
Daniel F Gadala-Maria[1], Mohamed Uduman[1], Jason Kasvin-Felton[1], Gur Yaari[1], Uri Laserson[2,3,4], Francois Vigneault[2,5], George M Church[2], Steven H Kleinstein[1]
[1]*Yale University;* [2]*Harvard University;* [3]*Harvard-MIT Division of Health Sciences and Technology;* [4]*Massachusetts Institute of Technology;* [5]*Ragon Institute of MGH, MIT, and Harvard*

Abstract:
During an immune response, B cells encoding immunoglobulin (Ig) able to bind antigen are selected to expand clonally and undergo subsequent rounds of somatic mutation and selection. An individual's Ig repertoire can be represented by the millions of unique Ig sequences that result from this process, as well as the relative proportions of these sequences. Next-generation sequencing (NGS) now allows for large-scale characterization of the repertoire in humans, as well as model systems. Analyzing these data presents several challenges. Here we describe our efforts to: (1) develop an automated analysis pipeline that is robust to the high error rates associated with NGS data, (2) identify somatic mutations and separate sequences into clonally-related groups, (3) quantify selection and analyze mutation properties. We demonstrate our approach using human NGS data collected from three individuals at various times following influenza vaccination.

Email of First Author: daniel.gadala-maria@yale.edu

Computational Prediction of MAPK Substrates and Integrations with Phosphoproteomes

Authors:
Elizabeth A Gordon, [1,2,3] Vishal R Patel[2,3], Thomas C Whisenant[1,2,3] Robyn M Kaake [1,] Lan Huang [1] Pierre Baldi [1,2,3]and Lee Bardwell[1,2,3]
1.Department of Developmental and Cell Biology , 2. *Institute for Genomics and Bioinformatics,* 3 *Center for Complex Biological Systems, University of California, Irvine, CA 92697*

Abstract:
To understand signaling networks, new methods are needed to identify novel kinase substrates. MAP kinases make extensive use of docking and scaffolding interactions to bind their regulators and substrates. We have developed a hybrid computational search algorithm that combines machine learning and expert knowledge to identify novel MAP kinase docking sites (D-sites), and used this algorithm to search the human genome. Predictions were tested by peptide array followed by rigorous biochemical verification with in vitro binding and kinase assays. We identified several new D-site-dependent MAPK substrates and incorporated them into the previous training set. With the expanded training set and improved pattern matching we were able to create more high confidence hits. Gene Ontology analysis of genes revealed known MAPK targets and other important biological processes. Using new set we created a web interface that would map putative phosphosites and score likelihood of which kinase could phosphorylate the site and real phosphorylation sites near the D-site.

Email of First Author: eagordon@uci.edu

Keloid Susceptibility loci 1q41, 3q22, and 15q21 Replicated in a Nigerian Population

Authors:
Stanley Hooker[1], Peter B Olaitan[2], Suzanne M Leal[3], Ernst J Reichenberger[4]
[1, 3]Baylor College of Medicine
[2]Ladoke Akintola University of Technology Teaching Hospital, Osogbo Nigeria
[3]University of Connecticut Health Center

Abstract:
Keloids are overgrown scars that stem from abnormal wound healing and their pathogenesis is poorly understood. The highest incidence occurs in darker-skinned individuals, with the greatest prevalence found in those with African ancestry. Recently a genome-wide association (GWA) study was performed, using a Japanese population, to identify linked genes or loci (Nakashima et al., Nat Genet 2010). Four variants were associated at GWAS significance levels on chromosomes 1q41, 3q22.3 and 15q21.3. Only one SNP lies within a known gene, NEDD4 on chromosome 15. We genotyped 96 tagSNPs within these regions in 286 case and 267 control Yorubans that were ascertained in Nigeria. Analysis was performed within a logistic regression framework using an additive allelic effect model controlling for both age and sex. Significant findings for each region included: 1q41 - rs873549 OR = 1.54 (p = 0.012); 3q22 - rs13064974, chr3:140.1Mb, OR = 1.32 (p = 0.026) and rs940187, chr3:140.3Mb, OR = 1.33 (p = 0.027); and 15q21 - rs8031043 OR = 0.68 (p = 0.002). The minor alleles of rs873549 and rs940187 increased the risk of keloid development in both studies. This is the first study to characterize the genetic susceptibility of keloids in a sub-Saharan African population.

This project is supported in part by a training fellowship from the Keck Center NLM Training Program in Biomedical Informatics of the Gulf Coast Consortia (NLM Grant No. T15LM007093).

Email of First Author: hooker@bcm.edu

Coupling Semantic Predications and Microarray Data to Build Gene Expression Networks

Authors:
Jose G Chen, Michael J Cairelli, Thomas C Rindflesch, National Library of Medicine

Abstract:
We are investigating the impact of literature-derived knowledge on the interpretation of microarray data. Incorporation of knowledge into the analysis of microarray data has generally been limited to referencing genes to probe sets. Existing knowledge related to gene expression can be extracted by semantic interpretation of the literature to provide more focused and more salient analysis. Our goal is to develop a directed and weighted gene expression network, providing the individual contribution of each gene to its target. We use invasive ductile carcinoma of the breast as a target for literature search and a microarray set that demonstrates the effect of estrogen on gene expression in these cells. We extract relevant relationships from citations containing genes or proteins as arguments and INHIBITS or STIMULATES as predicates to create initial gene interaction networks. Classification of the microarray data embellishes the initial networks with relevant genes not included from the literature. A genetic algorithm was used to optimize the contribution weight with a fitness function based on either a linear or neural network model. Several data normalization models were explored: intensity normalization, log ratio normalization and variance stabilization method. Resulting networks contain both known relationships and novel relationships for further exploration.

Email of First Author: jose.chen@nih.gov

Greedy Total Correlation Detects Functional Gene Sets in Cancer Copy Number Data

Authors:
Rachel Melamed, Raul Rabadan, Columbia University

Abstract:
Compendiums that profile many tumor samples have increasingly enabled detection of the highly recurrent genetic lesions that are likely to drive cancer development. However, ranking by only by recurrence ignores the effects of epistasis. Tumor survival can be promoted by damage to only one of a set of alternate genes in a pathway (mutual exclusivity of aberration), while other genetic changes only provide a selective advantage to a cancer in a given mutational context (co-occurrence of aberration). We develop Greedy Total Correlation, an information theoretic method to find modules of genes with mutually informative patterns of aberration in cancer. Our method greedily adds genes to create the module with the lowest relative entropy. Additionally, it corrects for the effect of genes that may be co-aberrant in copy number data due to genomic co-location. When the method is applied to the TCGA glioblastoma copy number data, the top module contains known patterns, such as a mutual exclusive relationship between cell cycle regulators CDKN2A and CDK4. In addition, using recurrence and mutual exclusivity uncovers evidence for cell cycle and pro-proliferation as the dominant selective forces in glioblastoma. Thus, utilizing entropy of mutational patterns identifies sets of genes with relationships in cancer development.

Email of First Author: rdm2114@columbia.edu

Using Natural Language Processing Methods to Improve EEG Classification in Brain-Computer Interfaces

Authors:
William Speier, Corey Arnold, Jessica Lu, Ricky K Taira, and Nader Pouratian; University of California, Los Angeles

Abstract:
Victims of brain stem injuries and neuromuscular disorders such as amyotrophic lateral sclerosis (ALS) can lose the ability to move and communicate, a condition called locked-in syndrome. By detecting patterns in these patients' electroencephalography (EEG) signals, brain-computer interface (BCI) systems can be developed to restore some functionality. The P300 speller is one such device that presents visual stimuli and types characters based on the detection of an evoked response in the associated EEG signal. Although the most common application of this system has been communicating language, the properties and constraints of the linguistic domain have not to date been exploited when decoding brain signals. We created a probability model for the P300 speller's output using language information learned from the Brown corpus. We then used a naïve Bayes classifier to combine the standard stepwise linear discriminant analysis (LDA) algorithm with prior probabilities based on our language model. With integration of natural language processing (NLP), we observed significant improvements in system accuracy, and 40-60% increases in bit rate for all six subjects in a pilot study. This investigation suggests that integrating information about the linguistic domain can greatly improve signal classification.

Email of First Author: speier@jhu.edu

Determining Protein Conformational Variability Using Robotic Motion Planning

Authors:
Bryant Gipson[1], Mark Moll[1], Steven Ludtke[2], Lydia Kavraki[1]
[1]Department of Computer Science, Rice University, Houston Texas
[2]Department of Biochemistry, Baylor College of Medicine, Houston, Texas

Abstract:
The ability of a protein to switch among different conformations often gives rise to its ability to function. Dramatic conformational changes however, especially in the case of very large proteins or protein complexes, can be difficult to model with traditional modeling methods. We present a scalable, fully parallel method inspired from the principles of robotic motion planning that allows rapid exploration of conformational variability with minimal storage requirements. Our method, which we call the "Hierarchical Fragment Library" employs the Rosetta library (http://www.rosettacommons.org/) for protein manipulation and representation and is connected to the open source motion planning library OMPL (http://ompl.kavrakilab.org/), allowing seamless and customizable integration of features from both libraries. The output of operation is a single sqlite3 database that can be accessed, in real time, simultaneously by a host of programs, each of which can independently generate, analyze and reduce data. Additionally we have developed a robust toolset for visualization, data-access and statistical evaluation, designed to aid structural biologists and chemists in the interpretation and refinement of results. We present a number of results for conformational exploration and discovery on protein systems that would be otherwise too large to model with alternative methods.

This project is supported in part by a training fellowship from the Keck Center NLM Training Program in Biomedical Informatics of the Gulf Coast Consortia (NLM Grant No. T15LM007093).

Email of First Author: bryant.gipson@rice.edu

Improving Knowledge-Based Scoring Functions for Protein-Ligand Interactions by Accounting for Sparse Data in the Training Set

Authors:
Sam Z Grinter [1,3,] and Xiaoqin Zou [1,2,3]
[1] MU Informatics Institute, [2] Department of Physics & Astronomy, and Department of Biochemistry, and [3] Dalton Cardiovascular Research Center, University of Missouri, Columbia, MO

Abstract:
Even with large training sets, knowledge-based scoring functions face the inevitable problem of sparse data. In this work, we use a statistical approach to estimate the inaccuracy caused by sparse count data in a potential of mean force (PMF) for protein-ligand docking. This new scoring function, STScore, uses a consensus approach to combine a PMF and a simple force-field-based potential (FFP), where the relative weight given to the PMF and FFP is a function of their estimated inaccuracies. This weighting scheme implies that less weight will be given to the PMF for any pairs or distances that occur rarely in the training data, thus providing a natural way to deal with the sparse data problem. We show that STScore effectively combines the PMF and FFP, exceeding the performance of either potential alone. Its binding mode prediction success rate is 80% using the set of 100 complexes from Wang *et al*. and ligand decoys prepared by Huang *et al*. Its binding affinity predictions have a Pearson correlation of 0.671 with the experimentally-determined affinities in Wang's test set, and 0.489 when comparing with the large PDBBind test set.

Email of First Author: sam.grinter@gmail.com

Mapping Protein Complex Interaction Networks by Quantitative Mass Spectrometry

Authors:
Robyn M Kaake, Xiaorong Wang, Yingying Yang, Lei Fang, Vishal Patel, Pierre Baldi, Lan Huang, University of California, Irvine

Abstract:
Essential cellular functions responsible for maintaining cell homeostasis and proper signaling are carried out by a vast array of protein complexes. Protein-protein interactions (PPIs) are vital for protein complex assembly, stability, and function. Abnormal PPI signaling can have severe biological consequences and typically results in human disease or cancer. Recently, protein interaction interfaces have emerged as new targets for drug development. Therefore, global mapping of PPI networks of protein complexes is crucial for understanding how these “protein machines” work in cells and how their disruption can lead to cellular deficiencies. We utilize affinity purification coupled with label-free quantitative mass spectrometry (MS) to capture, identify, and characterize PPIs of protein complexes, specifically human Cop9 Signalosomes (CSN) and 26S proteasome complexes, both macromolecular protein complexes involved in the essential ubiquitin-proteasome degradation pathway. To determine the specificity and significance of putative interacting proteins it is essential to compare the purified protein complexes to mock/tag-only control purifications. Thus, we have integrated several computational strategies for quantitative comparison and statistical analysis of experimental and control populations. In this manner, we have generated comprehensive lists of high confidence CSN and 26S proteasome interactors, and subsequent PPI network analysis has characterized their extensive and dynamic interaction networks.

Email of First Author: rkaake@uci.edu

Implementation of the ATHENA-HTN Clinical Decision Support System

Authors:
Lauren A Shluzas, Ruth C Cronkite, Dallas Chambers, Douglas K Owens, Mary K Goldstein, Department of Veterans Affairs, VA Palo Alto Health Care System, and Stanford University

Abstract:
This research documents the implementation of the ATHENA-Hypertension (ATHENA-HTN) clinical decision support (CDS) system at five medical sites within the Department of Veterans Affairs. ATHENA-HTN provides clinicians with patient-specific recommendations regarding blood pressure control and therapy at point-of-care. We describe organizational factors that influenced the ATHENA-HTN implementation process during the VISN Collaborative Study from 2006-2010. This study occurred during a system-wide reorganization of the VA's Office of Information & Technology (OI&T), from a locally autonomous IT system to a centralized system with heightened security measures.

The research design involved a retrospective analysis of emails and meeting minutes pertaining to the implementation process, followed by stakeholder interviews. The data captures implementation barriers with respect to changes in rules, roles, and routines within the VA's OI&T, and strategies used to overcome these barriers. Implementation barriers included limited autonomy within a centralized reporting structure, permissions issues for remote computer access, and uncertainty regarding standardized procedures. Strategies to overcome these barriers included engaging high-level leaders to garner study support, validating data with local PIs, and establishing IT communication channels.

The findings from the VISN Collaborative Study provide investigators with insight for the implementation of future quality improvement initiatives within integrated health care networks.

Email of First Author: rkaake@uci.edu

Ontology-Guided Feature Engineering for Clinical Text Classification

Authors:
Vijay N Garla, Cynthia Brandt, Yale University

Abstract:
We present novel feature engineering techniques that leverage the biomedical domain knowledge encoded in the Unified Medical Language System (UMLS) to improve machine-learning based clinical text classification. Critical steps in clinical text classification include identification of features and passages relevant to the classification task, and representation of clinical text to enable discrimination between documents of different classes. We developed novel information-theoretic techniques that utilize the taxonomical structure of the Unified Medical Language System (UMLS) to improve feature ranking, and we developed a semantic similarity measure that projects clinical text into a feature space that improves classification. We evaluated these methods on the 2008 Integrating Informatics with Biology and the Bedside (I2B2) obesity challenge. The methods we developed improve upon the results of this challenge's top machine-learning based system, and may improve the performance of other machine-learning based clinical text classification systems. We have released all tools developed as part of this study as open source, available at http://code.google.com/p/ytex.

Email of First Author: rkaake@uci.edu

Development and Application of ReO: a Biomedical Research Reagent Ontology

Authors:
Matthew H Brush, Carlo Torniai, Melissa Haendel, Oregon Health and Science University

Abstract:
The volume of information available to biomedical researchers has far surpassed levels amenable to human access and comprehension. Accordingly, we are becoming increasingly reliant on informatics tools to discover and exploit such information. These tools must bridge diverse research domains that rely on specialized vocabularies and data models. In recent years, ontologies have emerged as a valuable approach for dealing with such semantic and syntactic data heterogeneity, with community-consortiums such as the Open Biomedical Ontologies (OBO) Foundry leading the way. Here, we describe development of the Reagent Ontology (ReO), and its potential to support linking and inferencing across data sources that describe the people, resources, and publications comprising the research network.

ReO models the domain of biomedical research reagents, considered broadly to include materials applied 'chemically' in scientific techniques to facilitate generation of data (examples include antibodies, constructs, cell lines, biochemicals). ReO was developed to be deeply integrated into the biomedical landscape, importing classes from a diverse set of ontologies and encoding their relation to reagents through rich semantic axioms. By providing a shared, computationally-tractable resource for modeling reagents, ReO can support applications aimed at research resource discovery, knowledge extraction from experimental data, and evaluation of scientific expertise and productivity.

Email of First Author: brushm@ohsu.edu

Large-Scale Image Retrieval Using Text-Based and Content-Based Features

Authors:
Matthew S Simpson, Dina Demner-Fushman, Sameer K Antani, George R Thoma,
Lister Hill National Center for Biomedical Communications, National Library of Medicine

Abstract:
Images are sources of essential information within biomedical texts. However, given the rapid growth of biomedical literature, it is increasingly important to provide a means for quickly accessing the most relevant images for a given need. Whereas text-based retrieval systems utilize text describing an image, such as its caption, as a surrogate for its content, content-based and multimodal retrieval systems extract visual descriptors of an image's content, such as its texture, and use these features to retrieve images that are visually similar to some given examples. Unfortunately, the use of content-based features becomes problematic at large scales when both the relevance of retrieved images and the time needed to obtain results are equally important.

This presentation describes a multimodal image indexing and retrieval approach suitable for use with large-scale image collections. The method operates by first clustering visual features extracted from a collection of images and then mapping the resulting clusters to artificial "words." The method combines these words with other image-related text to create multimodal documents that it then indexes with a traditional text-based information retrieval system. Experimental results demonstrate that this approach allows for low latency retrieval and improves upon the precision of existing content-based and multimodal methods.

Email of First Author: simpsonmatt@mail.nih.gov

Learning Context-Sensitive Causal Models

Authors:
Jean Garcia-Gathright, Ricky Taira, Denise Aberle, University of California, Los Angeles

Abstract:
In this project, we investigate natural language processing (NLP) methods for the creation and representation of causal association networks from medical literature in the domain of non-small cell lung cancer (NSCLC). This research aims to inform and improve patient care and clinical research by providing structured and tractable knowledge on NSCLC, aggregating scientific knowledge to enable novel hypotheses; and illuminating causal pathways between biological processes, targeted therapies, and patient outcomes. Current paradigms of causal relation extraction represent relations in a context-free manner. In contrast, we seek to provide a comprehensive view of the knowledge domain by extracting both the causal relations and the context in which these relations are found. A semi-supervised method for causal relation extraction is described, in which features such as lexico-syntactic patterns, Unified Medical Language System (UMLS) semantic types, and cue-phrase probability are used to train a support vector machine (SVM) classifier. This presentation will briefly discuss possible methods for extracting contextual elements at the sentence level (e.g., negation and other modifiers), the study level (e.g., population group, *p*-values), and the inter-study level (e.g., corroboration between multiple papers).

Email of First Author: jigarcia@ucla.edu

Using a Web-based Tool to Help Women Make Informed Choices About Breast Cancer Screening

Authors:
Paula Scariati, Lisa N Nelson, Jayashree Kalpathy-Cramer, Karen B Eden, Oregon Health and Science University

Abstract:
In November 2009 the US Preventive Services Task Force advised against routine screening mammography for women age 40-49. They recommended, instead, an individual choice be made that considered patient values regarding specific benefits and harms. To help patients make decisions, we built and tested a web-based, interactive breast cancer screening decision aid tool.

-The tool was reviewed by policy makers, providers, and patients. Pilot testing was conducted in a convenience sample of 51 age, risk-appropriate women to provide a preliminary assessment of the impact of the decision aid on screening choice and decisional conflict.

-A Wilcoxon signed-rank test was used to compare a woman's plans for screening mammography before and after using the decision aid tool. No significant change was seen ($Z = -1.5$, $p = 0.132$). Pre-post tool analysis of decisional conflict scores was undertaken using the same approach. A significant reduction in overall decisional conflict scores was observed ($Z = -5.3$, $p < 0.001$).

-The women studied here did not change their intention to have a screening mammogram in the next 1 to 2 years. However, using the web-based breast cancer screening decision aid tool significantly decreased the amount of decisional conflict they experienced in making that choice.

Email of First Author: scariati@ohsu.edu

Leveraging an Automated Cumulative Antibiogram for Clinical Decision Support

Authors:
Mustafa Fidahussein, Jason Cadwallader, Paul Dexter, Regenstrief Institute

Abstract:
Clinicians rely on their hospital's microbiology "antibiogram" to intelligently guide initial empiric antibiotic choice. The antibiogram is generated from cumulative antimicrobial susceptibility data derived from individual patients' microbial isolates. It is most commonly prepared annually, manually and in the form of a static report. We have automated the process of creating an antibiogram based on standard guidelines by applying information extraction techniques to individual patients' microbiology susceptibility data retrieved from our hospital's lab system. This has enabled us to generate an antibiogram on-demand using the most recent microbiology data with detailed customizable reports. Moreover, the encoded susceptibility data can now be applied for clinical decision support at the point-of-care to recommend optimal single or multiple antibiotic regimens based on individual patient characteristics. We also demonstrate additional benefits of this data including analyzing minimum inhibitory concentrations trends and the use of patient location information to generate epidemiologic reports for resistant isolates.

Email of First Author: mfidahussein@regenstrief.org

Towards a Clinically Efficient Decision Support Model Linked to the VA EMR

Authors:
Michael S Phipps, Hajime Tokuno, Nallakkandi Rajeevan, Cynthia Brandt, Perry Miller, Department of Veterans Affairs, VA Connecticut Healthcare System, and Yale University

Abstract:
We have built a prototype clinical decision support tool (Neuropathic/CDS) that uses patient data automatically extracted from the Veterans Administration (VA) patient record to assist the primary care clinician in decisions regarding the pharmacologic management of neuropathic pain (NP). NP is a chronic condition that requires monitoring/adjustment of treatment, and many primary providers feel uncomfortable managing it. Providers require accurate, easily accessible, and evidence-based information to care for these patients. Neuropathic/CDS has features designed for efficiency in the busy clinical environment:

Clinical Function:

- The goal is to provide as much focused, useful information as possible in one place that is easily accessed and assimilated.
- It is an optional adjunct to care, providing guideline-based options for management tailored to the patient's comorbidities and current treatment regimen.

Technical Design:

- It is driven from the patient record, retrieving patient demographics, co-morbidity information, and current and past drug prescription information in a few seconds.
- The information is almost entirely accessible from a single screen; it is a visual outline of pharmacologic management, with “hover boxes” containing focused information about management options.

We believe that this CDS model can ultimately be applied to a variety of focused clinical patient management problems.

Email of First Author: michael.phipps2@va.gov

Integrating NLP and Sequence Features for More Accurate Function Prediction

Authors:
Christopher Funk, Karin Verspoor, University of Colorado, Asa Ben-Hur, Colorado State University

Abstract:
Finding the function of unknown proteins is an important area of research in molecular biology. Due to the large amounts of biological data available (sequence, protein-protein interaction, expression, experimental, eg.), many computational approaches have been developed for automatic function prediction. The biomedical literature is another source of data that is not widely used but provides a wealth of information not available in public databases. This work utilizes GOstruct, a machine learning framework created specifically for function prediction. We performed natural language processing on all MEDLINE to extract two different types of co-occurrence: protein-protein (PPC) and protein-gene ontology term (PGC). Both types of co-occurrences were calculated as counts within a certain span; we examined spans of sentence and document level. PPC act as an approximation to protein-protein interactions and suggest similar function while PGC act as an approximation of function. These counts were the input features for GOstruct. Compared to a baseline of only using protein-protein interactions from databases and sequence similarity, BLAST, we found integrating literature co-occurrence features improved the prediction in a mouse dataset. (0.787 vs 0.814 AUC)

Email of First Author: christopher.funk@ucdenver.edu

Predicting Dendrimer Cytotoxicity via Molecular Descriptors and Data Mining

Authors:
David E Jones, Julio C Facelli, University of Utah

Abstract:
Dendrimers are highly branched, polymeric nanoparticles that can easily be modified to well-defined specifications, making them highly useful in the field of medicine as delivery vectors. However, there is a significant setback for their use in biomedicine due to their toxicological effects. Manufacturing methods are being used to counteract their inherent toxicity, however, this is very time consuming and resource intensive. *In silico* approaches have been very successful in medicinal chemistry and are commonly used to guide the design of small pharmaceutical compounds, but little is known about how to guide the experimental efforts leading to biocompatible dendrimers using *in silico* approaches.

This project aims to develop an *in silico* environment to model and predict with confidence a variety of dendrimer properties determining their biocompatibility and therefore reducing the search space for promising biological compatible dendrimers that have to be synthesized. The methods pipeline is working, on a subset of 40 dendrimer molecules for which we have found cytotoxicity data in the literature, and we are in the process of finding which molecular descriptors and machine learning techniques provide the best combination to establish a practical protocol that can be widely applied to the computer aided design of biocompatible dendrimers.

Email of First Author: tenace.jones@utah.edu

Automation of In Situ Gene Expression at Single-Cell Resolution in C. Elegans to Study Development and Aging

Authors:
Sarah Aerni, Xiao Liu, Andy Nguyen, Chuong Do, Samuel Gross, Stanford University

Abstract:
Gene expression at the sub-tissue or cell-specific level provides scientists with a deeper understanding of how organisms develop and insight into how they change with age. In particular, the transcriptome of each cell is tightly regulated and controlled during development, and individual cells may play different roles in the decline of organisms.

In C. elegans, microscopy permits high-resolution in-situ studies using reporter gene assays that allow us to observe differences at the level of single-cells. As an adult, C. elegans has 959 cells, each of which can be assigned a unique label by expert scientists identifying its exact lineage. Few experts exist who can perform this annotation, and the process is extremely time-consuming. Automated labeling could significantly increase progress in research by making these high-resolution studies more widely accessible and faster.

We present the first automated pipeline for single-cell annotation of adult C. elegans using a model trained on human-annotated adult worms. Our model integrates spatial information with characteristics of individual cells based on morphology and tissues. In addition, we show results of the single-cell analysis of the partial single-cell transcriptome in larvae, young and old adult C elegans.

Email of First Author: saerni@cs.stanford.edu

Wednesday
June 27, 2012

Wednesday – Table of Contents

Multi-Marker Tests and Proxy Association for Functional Rare Variants

Authors:
Ronald Schuyler, Lawrence E Hunter, University of Colorado

Abstract:
Genotyping chips used in genome wide association studies are designed to measure variants which are common in the population, but rare variants are also thought to play a significant role in disease susceptibility. Using a multi-marker genome wide association method we identify genomic regions strongly associated with resistance to the autoimmune condition celiac disease which are not detectable using standard single-maker tests. Further, due to the high specificity of these two-marker associations, we may conduct a proxy association analysis using a fully sequenced reference panel without the need for genotype imputation. Imputation with standard methods is a probabilistic process, and is not accurate for variants which are rare in the population. As an alternative, we identify individuals from the reference panel having the highly penetrant two-marker genotypes and determine what other possibly functional variants are over-represented in this group. We identify rare variants which are predicted to disrupt function within two genes of the same gene family. The genes are specifically expressed in the relevant organ and have overlapping functions. These rare and likely functional variants which may confer resistance to celiac disease suggest potential treatment strategies, and were completely undetectable using standard association tests and imputation methods.

Email of First Author: ron.schuyler@gmail.com

Inferring Metabolic Networks Using the Bayesian Graphical Lasso

Authors:
Christine B Peterson, Mirjana Maletic-Savatic, Marina Vannucci, Rice University

Abstract:
Chronic neuroinflammation is a hallmark of many neurodegenerative diseases such as Alzheimer's and Huntington's. Despite its significance, the mechanisms by which neuroinflammation leads to disruptions in cellular function are not well understood. One of the less explored cellular functions affected by neuroinflammation is metabolism. Metabolic changes underlying neuroinflammation can be exploited for biomarker discoveries and future drug developments. To understand these changes, we have developed statistical methods that allow inference of the cellular metabolic networks from a set of critical metabolites identified and quantified in inflamed cells. Our approach builds on the Bayesian graphical lasso, a technique for learning a sparse graph structure from the covariance of the data. To encourage edges between metabolites with known relationships, we allow each entry in the precision matrix to have a unique shrinkage parameter with a gamma prior shifted towards zero for molecules closer together in a reference network. By incorporating informative priors, we improve the reliability of network inference given the small sample sizes and high degree of noise. The resulting identification of cellular reactions under neuroinflammatory conditions not only increases scientific understanding of the mechanisms of neuroinflammatory diseases, but also provides potential targets for future treatments.

This project is supported in part by a training fellowship from the Keck Center NLM Training Program in Biomedical Informatics of the Gulf Coast Consortia (NLM Grant No. T15LM007093).

Email of First Author: christine.b.peterson@rice.edu

A Lineage-Hidden Markov Model for Genomic Annotation from Epigenetic Modifications

Authors:
Jacob Biesinger, Yuanfen Wang, Xiaohui Xie, University of California, Irvine

Abstract:
More than a decade after the draft sequence of the human genome was released, many biologists continue to focus on the descriptive annotation of the genome. Technological advances in high throughput sequencing coupled with new biological techniques are paving the way for systematic, comprehensive annotation of the genome and the differences between cell types as well as disease/normal states. Epigenetic modifications are a rich source of diversity amongst cell types and correlate tightly with cell-type specific gene expression and certain DNA sequence features. Previous attempts to understand the combinations of epigenetic modifications that give rise to chromatin state have not considered the lineage of the cells in question.

We present a bayesian network that uses epigenetic modifications to simultaneously capture underlying chromatin states and the transitions between cell types through differentiation or disease progression. We apply our model to a recent histone modifications dataset, covering nine human cell types each annotated with nine epigenetic modifications. Since exact inference in this model is intractable for all but the smallest datasets, we develop several variational approximations. Our results show improved recovery of cell-type specific transitions compared to previous methods while allowing missing data to be inferred accurately.

Email of First Author: jake.biesinger@uci.edu

Simulating Glycemic Variability in Critically Ill Burn Patients

Authors:
Edward A Ortiz, Stephen D Patek, Leon Farhi, Marc Breton, Boris P Kovatchev,
University of Virginia

Abstract:
Tight glycemic control in the intensive care unit (ICU) requires insulin therapy, and comes with the risk of hypoglycemia. Computer simulation can be an essential tool in evaluating protocols for insulin delivery in this setting. To this end, it is necessary to have mathematical models that explain BG variability within this patient population.

Current models account for glycemic variability with the concept of time varying insulin sensitivity, particularly in peripheral tissues, but non-insulin dependent hepatic contribution to glycemic variability is significant. An adapted model based on 300 *in silico* patients derived from non-ICU, real patients incorporates this contribution.

Hourly blood glucose, insulin, and feeding data from 186 burn-unit patients were fitted to the model. Peripheral glucose uptake was accounted for by finding the closest *in silico* patient, using least squares with weighting against negative error. Using this *in silico* patient, hepatic glucose production (HGP) was fitted with a time-varying coefficient ("SA", stress action). HGP was limited to a literature-derived maximum of 4.25 mg/kg/min. 147 SA vectors of at least 24 hours each and 82 unique *in silico* patients were identified.

This approach forms the basis for a simulator that would allow testing insulin protocols *in silico*, before use in patients.

Email of First Author: eao7r@virginia.edu

Pathway Index Models for Construction of Patient-Specific Risk Profiles

Authors:
Kevin H Eng, Sijian Wang, Christina Kendziorski, University of Wisconsin-Madison

Abstract:
Statistical methods for variable selection, prediction, and/or classification have proven extremely useful in moving personalized genomics medicine forward, in particular leading to a number of genomic based assays now in clinical use for predicting recurrence of breast, colon, and prostate cancer, as well as transplant rejection. Although invaluable in individual cases, the information provided by these assays is limited. In particular, a patient is classified into one of very few groups (e.g. recur or not), limiting the potential for truly personalized treatment. Furthermore, although these assays provide information on whether or not to treat (e.g. if recurrence is predicted), they provide no information on how to treat. This talk reviews an approach based on modeling a time-to-event outcome as a function of known biological pathways, identifying important genomic aberrations, and providing pathway-based patient-specific assessments of risk. As we discuss, applications of the approach to the Cancer Genome Atlas (TCGA) ovary project provide results that may significantly impact ovarian cancer research as well as patient treatment.

Email of First Author: eng@stat.wisc.edu

A Practical Tool to Improve Engagement in Quality and Safety Interventions

Authors:
Jordan Duval-Arnould[1,2], Simon Mathews[1,2], Kristina Weeks[1,2], Peter Pronovost[1,2,3], Sean Berenholtz[1,2]
Departments of [1]Anesthesiology and Critical Care Medicine, [2]Armstrong Institute for Patient Safety and Quality Johns Hopkins Medicine; Departments of [3]Health Policy and Management

Abstract:
-Introduction:
The Opportunity Estimator tool was developed for the national On the CUSP: Stop BSI (Stop BSI) program to reduce central line-associated bloodstream infections (CLABSI). This application is used to calculate evidence-based estimates of the impact of CLABSI presenting them in terms of excess mortality, cost, and length of stay.

-Description:
The Opportunity Estimator includes an implementation guide and a web-based interface. The estimator has input, output, and reference display functions along with an export feature providing results in a customizable Microsoft Word report. References include article citations and abstracts providing the evidence used to estimate attributable mortality and costs associated with CLABSI. Participating teams use the tool to share data with providers and monitor performance with a goal toward maintaining engagement and encouraging improvement in the initiative.

-Conclusion:
Users have reported providing data in terms of deaths, dollars, and days compared to traditional CLABSI rates has made preventable harm and economic consequences of their current performance more visible and meaningful. Between October 2009 and October 2011 there were 5,072 unique visitors to the Opportunity Estimator. After the report export option was made available approximately 45% of visitors generate a report.

Email of First Author: jordan.duval@gmail.com

Creation of a Gold Standard for EHR-Based Information Retrieval

Authors:
Karthik Natarajan, Noémie Elhadad, Columbia University

Abstract:
Searching within the electronic health record (EHR) is an area of promising research. Today, search functionality within EHRs is based on exact string matching, which can be improved upon in many ways; however, there needs to exist a platform to evaluate these different approaches. This work describes the development of a gold standard that can be used to evaluate an EHR-based search engine. In creating the gold standard, cohorts of real patients were selected, information needs for the cohorts were developed, and clinical experts identified relevant paragraphs within notes of each medical record. The relevant paragraphs were determined through consensus building. The gold standard contained a set of information needs and the relevant paragraphs and documents associated to them. There were 30 information needs that were used to determine relevancy. 2,114 paragraphs were identified as relevant out of a total of 20,643 paragraphs from 638 clinical notes. To our knowledge, this is one of the first attempts to develop a gold standard for within-patient search in the EHR. The different types of information needs, along with their relevance judgments, are shown to belong to a representative set of note types, providing a realistic and rich dataset.

Email of First Author: karthik.natarajan@dbmi.columbia.edu

The Quality of Discharge Summaries for Facilitating Transitions of Care

Authors:
Farrant Sakaguchi, Leslie Lenert, Michael Strong, Charlene Weir, University of Utah

Abstract:
-Introduction
Effective communication and coordination of care between inpatient and outpatient medical providers may decrease risks for readmission and medical errors. Prior work suggests discharge summaries typically lack critical information for effective handoffs. We examined discharge summaries at the University Hospital to develop specifications for an information sharing tool to facilitate transitions of care.

-Design
Retrospective review of dictated discharge summaries completed by Internal Medicine residents, laboratory orders and results, and discharge medications retrieved from an enterprise data warehouse.

Population: 60 patients discharged between August 1, 2010 and August 20, 2010

-Results
The patients were an average age of 60 years, 50% were male, and the average length of stay was 5 days. Discharge summaries had significant gaps in information needed for effective transfers of care. The receiving primary care provider was unambiguously identified (name, and address, clinic, or fax) in only 48%. Lab results pending at discharge were frequently not mentioned: only 17% were documented. Medical reasoning regarding changes in medications was described in only 40% of discharge summaries.

-Conclusions
Discharge summaries frequently lack critical data useful for safe and effective transitions of care. Informatics tools to facilitate inclusion of critical information may improve the quality of discharge summaries.

Email of First Author: farrant.sakaguchi@hsc.utah.edu

Exploring Adverse Drug Effect Discovery from Data Mining of Clinical Notes

Authors:
Joshua C Smith, Joshua C Denny, S Trent Rosenbloom, Anderson Spickard, III, and Randolph A Miller, Vanderbilt University

Abstract:
After 80 million people worldwide received prescriptions for the drug rofecoxib (Vioxx), its manufacturer withdrew it from the marketplace in 2004. Epidemiological data showed that it increases risk of heart attack and stroke. Recently, the FDA warned that the commonly prescribed statin drug class (e.g., Lipitor, Zocor, Crestor) may increase risk of memory loss and Type 2 diabetes. These incidents illustrate the difficulty of identifying adverse effects of prescription medications during premarketing trials. Only post-marketing surveillance can detecting some types of adverse effects (e.g., those requiring years of exposure). Through an NLM-sponsored R01 project, we explored use of data mining on clinical notes to detect adverse drug effects. We constructed a knowledge base using UMLS and other data sources that could classify drug-finding pairs as “currently known adverse effects” (drug causes finding), “known indications” (drug treats/prevents finding), or “unknown relationship”. With IRB approval, we used natural language processing (NLP) to extract current medications and clinical findings (including diseases) from 360,000 de-identified history and physical examination (H&P) notes. We identified 35,000 “interesting” co-occurrences of medication-finding concepts that exceeded threshold probabilities of appearance. These involved ~400 drugs and ~200 findings. Among the identified pairs are several that the FDA identified as harmful in postmarketing surveillance. Our preliminary results illustrate the problems and potential of using data mining for adverse drug effect discovery.

Email of First Author: joshua.smith@vanderbilt.edu

Evaluation of Methods for Translating the CCD to the vMR

Authors:
Jason R Jacobs, Kensaku Kawamoto, David E Shields, University of Utah

Abstract:
Effectively implemented and utilized, clinical decision support (CDS) provides a means to improve healthcare quality and decrease costs. As the phases of Meaningful Use progress, implementation of CDS will become increasingly significant.
OpenCDS is an effort to develop CDS tools and resources that can be widely adopted to enable CDS at scale. OpenCDS uses a standards-based, service-oriented, open-source approach to CDS. The Virtual Medical Record (vMR) is an XML-based specification for data input in OpenCDS.

The Continuity of Care Document (CCD) is an electronic document standard that specifies encoding of patient summary clinical data. The CCD is becoming popular as a means for exchanging patient healthcare data to facilitate meeting Meaningful Use objectives.

Automated translation of the CCD schema to the vMR schema would remove the barrier of data input for those organizations that also seek to leverage the capabilities of OpenCDS. We evaluated alternative methods for automating this translation with the intent to build this capability into the OpenCDS framework.

Email of First Author: jason.jacobs@utah.edu

Determining Credibility of Online Posts about Health

Author:
Joseph Buell, University of Missouri

Abstract:
There are currently many communities online where people can post health-related question for other users to answer. Currently, there is no uniform way for users to assess the reliability of answers posted in this way. Therefore, we will evaluate different methods of determining users' reputation for sites like this, and estimate their efficacy in filtering out accurate responses from inaccurate responses to recommend the better responses.

Email of First Author: jlbkwd@mail.missouri.edu

Rectangular Area Judgments as a Function of Size Difference and Placement

Authors:
Akilah L Hugine, Ellen J Bass, and Stephanie A Guerlain, University of Virginia

Abstract:
Comparing the size of displaced rectangles is important in several types of visualizations such as bar charts and treemaps. This study investigated the effect of size difference (aspect ratio and area difference) and relative placement (distance between rectangles and offset angle) on the speed and accuracy of rectangular area judgments. It also considered the effect of combinations of aspect ratio and area difference where rectangles are contained within each other as compared to cases where rectangles do not fit. In a controlled experiment participants were presented with two rectangular stimuli. They identified which was smaller and then estimated the percent difference in area by making a “quick visual judgment”. The results showed that the size (area difference) and location (offset angle) of rectangular stimuli affect judgment accuracy and speed. The lowest error and shortest judgment time occurred for comparing pairs of rectangles when the smaller rectangle fits inside of the larger. Relative displacement of rectangular stimuli resulted in low accuracy of relative area judgments and do not help to support fast judgments.

Email of First Author: alh8p@virginia.edu

Social Network Analysis of Genetic Testing Relationships

Authors:
Jeremiah G Ronquillo, William T Lester
Massachusetts General Hospital, Boston
Harvard Medical School, Boston

Abstract:
-BACKGROUND: Electronic medical records have begun storing diverse information about genetic test ordering behavior and relationships, but the large volume of results makes practical application using traditional data mining difficult.
-METHODS: A social network of doctor-patient genetic testing relationships was constructed using the Research Patient Data Registry to identify the physician and pediatric patient population involved with genetic testing at Massachusetts General Hospital from 2008-2010. Network analysis was performed using Pajek software (version 2.05) and visualized using the Kamada-Kawai algorithm.
-RESULTS: The genetic testing social network consisted of a total of 1,519 nodes. The center of the network was characterized by dense clusters of patient nodes surrounding a small number of physician nodes, along with a large number of sparsely connected physician-patient dyads at the periphery of the network. The four most connected physician nodes each had ties to over 50 patients, while approximately 96.5% of nodes involved orders for 10 or fewer genetic tests.
-CONCLUSION: Social networking analysis of genetic testing relationships can provide a potentially efficient way of visualizing important clusters and patterns of behavior for the population involved with testing. Further studies are needed in order to develop effective genome-enabled clinical decision support tools.

Email of First Author: jgr9@hms.harvard.edu

Adding Social Networking to the EHR

Authors:
Justin Morea, Jason Cadwallader, Jon Duke, Burke Mamlin, Regenstrief Institute

Abstract:
Since online social networking burst onto the scene with once popular sites such as Friendster and MySpace, social networking has become an integral part of daily living in most of the developed world. Facebook is one of the most accessed websites in the US and has more than 800 million active members, equaling about 11.5 percent of the world's population.

The passing of the health care reform law has made prominent a push for patient centered homes and accountable care organizations. As the delivery of health care has become increasingly fragmented, the burden upon health care providers to maintain communication and stay up to date on a patient's plan of care has grown tremendously. To meet the challenges of improving patient health, increasing communication between health care providers, and facilitating the participation of the patient in the communication, we propose the development of an innovative social networking health care platform. By putting to use ubiquitous technology and exploiting the notion of asynchronous communication, we feel that we can improve efficiency and outcomes. The net result will be a more highly connected team of primary care practitioners, consultants, nurses, therapists, case workers and other care providers for the patient.

Email of First Author: jmorea@iupui.edu

Mining for Common Behaviors in Image Analysis Using Eye-tracking Fixation Sequences

Authors:
Blake Anderson, Chi-Ren Shyu, University of Missouri

Abstract:
Medical experts spend many years studying medical images and developing a set of tacit strategies for uncovering relevant information quickly and accurately. The unconscious nature and the difficulty for experts to verbalize these strategies suggest that studying behaviors requires innovative capture methods. This research uses eye-tracking data to generate a sequence of eye movements and their associated image, gaze, and semantic features. In contrast to qualitative or statistical analysis of the eye movements, we apply temporal mining techniques to the fixation data from novices and experts to produce rules for frequent sequences. The resulting rules represent feature-oriented behaviors, which are tolerant of distracted eye movements or noisy data collection. Moreover, these behaviors can be studied quantitatively and allow direct contrast of subjects. This research could potentially identify a set of best-practice expert behaviors which could be used in training and detection of image-based medical errors.

Email of First Author: beakv6@mizzou.edu

Estimating Heritability Captured by Common Variants in Pharmacological Traits

Authors:
Casey Overby, George Hripcsak, Yufeng Shen, Columbia University

Abstract:
The heritability of complex conditions is traditionally estimated through twin and family studies. However, in the context of low prevalent pharmacological traits, it is difficult to recruit and obtain clinical outcome data in families. Therefore, we investigate the ability of the Genome-wide Complex Trait Analysis (GCTA) algorithm to estimate the heritability of low prevalent traits of moderate sample size from genome-wide single nucleotide polymorphism (SNP) data. Phenotypes were simulated for case-control datasets from real genome-wide association study data such that known causal variants explained all the genetic variance. In simulation studies, variables of disease prevalence, sample size, and heritability (h^2) were varied. For each scenario, we calculated h^2 based on the genetic relationship estimated from all SNPs. Estimates were averaged over 30 simulations. Results indicated that h^2 estimates appear to be accurate for a broader range of true h^2 values with larger sample sizes. Also, results indicated that the GCTA algorithm provides relatively accurate h^2 estimates for conditions with true h^2 values ≥ 0.1 and with modest sample sizes. Therefore, we believe the GCTA algorithm would be robust to provide estimates for the proportion of h^2 captured by common SNPs in low prevalent/high heritable pharmacological traits.

Email of First Author: casey.overby@dbmi.columbia.edu

Mixture Models vs. Supervised Learning for Integrative Genomic Analysis

Authors:
Daniel Dvorkin, Katerina J Kechris, University of Colorado

Abstract:
Gene classification using multiple data sources such as expression and transcription factor binding is an important problem in bioinformatics. Given high-quality training data, supervised machine learning methods such as logistic regression, naive Bayes classifiers, and random forests are generally more powerful than unsupervised methods. However, we show that when training data sets are small or contain labeling errors, unsupervised or semi-supervised mixture models are more robust. We present a comparison of mixture models to fully supervised machine learning methods, and discuss the circumstances under which the former may be expected to outperform the latter, using identification of critical genes in Drosophila development as an example of a case in which training data may not be of sufficient size, quality, or completeness to justify the use of fully supervised methods.

Email of First Author: daniel.dvorkin@gmail.com

Identifying Functionally-Coherent Gene Subsets through Ontology Merging Algorithms

Authors:
Vicky Chen, Xinghua Lu, University of Pittsburgh

Abstract:
Motivation: Contemporary "omics" studies often result in a long list of genes of potential interest. In this study, we report a computational approach based on the Gene Ontology (GO) to identify non-disjoint subsets of genes performing coherently related functions from such a gene list and summarize the functions of the subsets using GO terms. To achieve this goal, we utilized GO ontology structure, semantic, and protein information to identify gene subsets, and we further used information-theory-based semantic distances between GO terms to assess functional coherences of a set of genes.

Results: We developed a software package that can represent multiple types of information relevant to our study: the ontology structure of the GO, the semantic context of GO terms, and protein annotation information. We designed and evaluated three different algorithms
for identifying functionally coherent gene subsets and summarizing their functions. We found that utilizing semantic information enhanced our capability of identifying functionally coherent gene sets and applying the methods to analyze real-world examples led to biological insights.

Email of First Author: vic14@pitt.edu

Epigenomic Profiling of the Osteosarcoma Genome

Authors:
Jiayi M Sun[1,2], Joseph Luan[2,3], Alex Yu[2,3], Horatiu Voicu[3,4], Chris Man[2,3,4], Rudy Guerra[5], Ching C Lau[1,2,3,4,]
[1]Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine
[2]Texas Children's Cancer Center
[3]Department of Pediatrics, Baylor College of Medicine
[4]Dan L. Duncan Cancer Center
[5]Department of Statistics, Rice University

Abstract:
Osteosarcoma is the most common primary bone tumor diagnosed in adolescents and young adults. Due to the high degree of genomic heterogeneity, the tumor biology and pathogenesis of osteosarcoma is still elusive. Overall survival for patients with metastatic osteosarcoma remains dismal at 20-30%. To comprehensively evaluate the osteosarcoma genome, we have generated transcriptomic, DNA copy, microRNA and DNA methylation profiling data on 44 cases of pre-treatment biopsy. We hypothesize that epigenomic aberrations leading to repression of tumor suppressor genes or overexpression of oncogenes are involved in the pathogenesis of osteosarcoma. To identify regions or genes with aberrant DNA methylation, we performed whole genome high resolution DNA methylation profiling using Illumina's 450K array, which interrogates the methylation status at over 450,000 sites within the genome, including CpG sites outside of promoter regions. Comparison of methylation data with expression data reveals a set of genes differentially methylated in tumors which show high negative correlation with gene expression. Our current results suggest that methylation plays a fundamental role in altering the transcription of tumor suppressor genes. Further network based studies involving pathway and gene set enrichment analysis will clarify the role of these genes in the initiation and progression of these tumors.

This project is supported in part by a training fellowship from the Keck Center NLM Training Program in Biomedical Informatics of the Gulf Coast Consortia (NLM Grant No. T15LM007093).

Email of First Author: jmsun@bcm.edu

Demographic Insight from Sequencing Y Chromosomes in Diverse Populations

Authors:
David Poznik, Peter Underhill, Brenna Henn, Carlos Bustamante, Stanford University, Jeffrey Kidd, University of Michigan

Abstract:
The Y chromosome harbors the longest stretch of non-recombining DNA in the human genome. It is therefore a unique tool that enables the tracking of migrations and inference of demographic history. However, due to inefficient selection, a relatively high mutation rate, and a small effective population size, the chromosome is particularly subject to drift. It has accumulated large expanses of highly repetitive sequence, which pose considerable challenge within a short-read sequencing paradigm. Thus, the first objective of this work was to construct a pipeline to reliably call Y chromosome alleles from shotgun sequence data.

We have sequenced 69 samples from nine globally diverse populations, including three African hunter-gatherer groups. Haplogroups were assigned and a tree recapitulating the extant Y Chromosome phylogeny was constructed. We resolve a major polytomy by identifying a variant for which haplogroup G retains the ancestral allele, whereas haplogroups H and IJK share a derived allele, thus indicating common ancestry and uniting the two branches. This result has been validated by genotyping a larger panel. Finally, we employ the program BEAST, which uses MCMC to average over tree space, in order to co-estimate the phylogenetic tree and demographic parameters, including ancestral population size and divergence times.

Email of First Author: dpoznik@stanford.edu

Computing Clustered Alignments of Gene-Expression Time Series

Authors:
Deborah Muganda-Rippchen*, Mark Craven, University of Wisconsin – Madison

Abstract:
Identifying similarities and differences in expression patterns across multiple time series can provide a better understanding of the relationships among various chemical treatments or the effects induced by a gene knockout. We consider the task of identifying sets of genes that have a high degree of similarity both in their (i) expression profiles within each treatment, and (ii) changes in expression responses across treatments. Previously, we developed an approach for aligning time series that computes *clustered alignments*. In this approach, an alignment represents the correspondences between two gene expression time series. Portions of one of the time series may be compressed or stretched to maximize the similarities between the two series. A clustered alignment groups genes such that the genes within a cluster share a common alignment, but each cluster is aligned independently of the others. Unlike standard gene-expression clustering, which groups genes according to the similarity of their expression profiles, the clustered-alignment approach clusters together genes that have similar *changes* in expression responses across treatments. We have now extended the clustered alignment approach to produce multi-level clusterings that identify subsets of genes that have a high degree of similarity both in their (i) expression profiles within each treatment, and (ii) changes in expression responses across treatments. We examine the validity of this multi-level clustering method by performing a GO-term enrichment analysis of the clusters. Additionally, we use permutation testing to determine if our clusters that have alignment scores that are unlikely to occur by chance.

Email of First Author: deborah@cs.wisc.edu

Evaluating Melanoma Whole Exome Sequences Suggests New Driver Genes

Authors:
Perry Evans, Yong Kong, Richard Lifton, Ruth Halaban, Michael Krauthammer, Yale University

Abstract:
The identification of mutations in receptor and non-receptor protein kinases that drive malignant transformation of cells into melanoma, the most deadly form of skin cancer, has revolutionized the care of patients by providing patient-tailored targeted therapies that improve survival. The most promising example is the activating V600E/K mutation in the protein kinase BRAF. This mutation is present in ~50% of cutaneous melanomas, and drugs have been developed to inhibit this mutant kinase. There is a great interest in identifying additional 'drivers' that contribute to melanoma pathogenesis because large numbers of patients do not have BRAF mutations, and melanomas with BRAF mutations tend to develop resistance to the inhibitors with time. In this study, we propose additional driver genes by examining the mutational landscape of melanoma determined by sequencing the exomes of 148 tumors. We develop models of background silent mutation rates across the exome that account for gene expression, tumor type, and sequence context. Using our background silent mutation rates, we identify genes with a significant amount of nonsynonymous mutations across the 148 tumors. Our study successfully identifies known driver genes like BRAF and NRAS, and suggests important new drivers.

Email of First Author: perry.evans@yale.edu

Deep Networks and CUDA for Contact Prediction, Disorder Prediction and Gene Interaction

Authors:
Jesse Eickholt, Jianlin Cheng, University of Missouri, Columbia

Abstract:
Deep belief networks and CUDA are emerging technologies which allow complex models to be trained on large datasets in a reasonable amount of time. This reduction in time for both training and classification allows boosted ensembles of classifiers to be trained to predict both protein residue-residue contacts and disordered residues. In this project we have developed classifiers based on deep belief networks using CUDA for contact prediction and disorder prediction and studied the performance of bagged and boosted ensembles of such classifiers. We are also using CUDA and various clustering algorithms to attempt to cluster gene-gene contact data from different cancers and identify some distinguishing gene-gene interactions.

Email of First Author: jlec95@mail.mizzou.edu

Modeling the Elastic Change of Depolarization in a Sensory Neuron

Authors:
Aaron L Williams, Gregory J Gerling, University of Virginia

Abstract:
A major goal for next generation neural prosthetics is to provide the user with realistic sensory feedback. Previous research has focused on low-level transduction models that output action potential timings or higher-level psychophysics experiments demonstrating stimulation of varying sensations. Between these areas is a gap in the research, concerning the effects of stimulating signals on the peripheral nervous system. We seek to better understand how to reliably stimulate sensory neurons without causing damage to the nerve from large stimuli by finding the threshold for depolarization. A model will characterize the elastic shift in the sensory nerve's strength-duration curve as a function of the time since the previous action potential. The model is informed by finding changes in the rheobase and chronaxie of a sensory neuron with a computational simulation, based upon the Hodgkin-Huxley equations and instantiated in Matlab. These results will be validated by delivering electrical stimulations distally to the sural nerve of the rat and making *in vivo* electrophysiological recordings proximally on the nerve. The results of this model will improve our understanding of both the sensation of touch and methods for stimulating the peripheral nervous system.

Email of First Author: alw3fn@virginia.edu

Mapping the Topology of the 19S Proteasome Using Cross-Linking Mass Spectrometry and Probabilistic Modeling

Authors:
Athit Kao, Arlo Randall, Scott Rychnovsky, Pierre Baldi, Lan Huang, University of California, Irvine

Abstract:
The 26S proteasome maintains cellular homeostasis by controlling degradation of ubiquitinated proteins. While the 20S subcomplex structure has been resolved, the topology of the 19S subcomplex has remained elusive. Chemical cross-linking coupled with mass spectrometry (XL-MS) is recognized as a promising technology for structural characterization of protein complexes. Although XL-MS is hindered by complex analysis of cross-linked peptides, new methodologies are being developed.

A practical XL-MS workflow requires rapid and accurate identification of cross-linked peptides for structural elucidation by computational methods. To facilitate analysis, we recently developed a new integrated workflow combining our novel cross-linking reagent disuccinimidyl sulfoxide with multi-stage tandem mass spectrometry and bioinformatics tools (DOI:10.1074/mcp.M110.002212). This methodology was used to successfully identify cross-linked peptides from the 20S and recently from the 19S. Over a 170 binary interactions have been generated from the cross-linking data that provides useful distance constraints for computational modeling.

To generate the structure topology, we have developed a probabilistic modeling strategy to assess our cross-linking data. The predicted model of 19S substructures based on our cross-linking data is fully supported by current 19S structural knowledge. This is the first example of a rigorous probabilistic assessment of the 19S organization solely using XL-MS data.

Email of First Author: atet.kao@uci.edu

Development and Evaluation of a Web-Based Electronic Medical Record Without Borders

Authors:
Francisco Saavedra, Fredric M Wolf, University of Washington

Abstract:
Despite implementation of electronic medical record (EMR) systems in the United States and other countries, EMRs often lack open global access, standardization, efficient interface, and effective knowledge-based tools at the point of care. Consequently, the information needs of patients, practitioners, administrators, researchers, and policymakers often go unmet. To address this multifaceted problem, we are creating a novel EMR design to ensure that the most vital pieces of patient EMRs are available to make health care decisions. We have developed and are evaluating a standardized family medicine clinical history model as an EMR clinical core that integrates state-of-the-art terminology and peer-reviewed evidence-based protocols with real-time access to diagnostic decision support systems and the biomedical literature. This project aims to facilitate structured clinical documentation, interoperability, global access, and decision-making processes to better address not only local, clinical, and psychosocial primary care problems in targeted underserved communities in the State of Guanajuato, Mexico, but also transnational migration health issues based on information exchange among primary care settings. We hypothesize that its use will have a measurable, positive effect on provider satisfaction, and that it will make a contribution to the field of biomedical informatics.

Email of First Author: saavej@uw.edu

HIE Utilization to Coordinate and Track Seasonal and H1N1 Flu Vaccination

Authors:
Valmeek Kudesia[1,2], Dan Newman, Boston University[2]
1. Harvard Medical School, Boston, MA
2. Boston University, Boston, MA

Abstract:
Annual influenza outbreaks and vaccination are major national public health concerns. The H1N1 influenza pandemic in 2009-2010 superimposed upon annual influenza season posed new difficulties. These challenges included evolving vaccination recommendations, a limited supply of H1N1 vaccine and some groups requiring two H1N1 vaccine doses. Vaccination histories dispersed across multiple EHRs hindered efforts to address those challenges by a safety-net system serving an urban, disadvantaged population with low health-literacy. Seasonal and H1N1 influenza and pneumonia vaccination histories from surrounding safety-net clinics were available at the main safety-net hospital by implementing a Health Information Exchange (HIE). Vaccination histories were made available to care providers at the hospital's outpatient clinics, emergency department and inpatient service. Vaccine sharing was part of an overall software architecture that monitored local influenza incidence and hospital resource utilization for influenza related needs. We hypothesized that sharing vaccine histories would increase the percentage of high-risk individuals vaccinated and decrease over-vaccination. A retrospective study is underway to assess the hypothesized benefits of utilizing an HIE to share vaccination histories. This is particularly relevant given recently proposed Stage 2 Meaningful Use criteria which identified sharing of immunization histories and electronic surveillance of syndromic data as specific measures.

Email of First Author: valmeek.kudesia@bmc.org

Investigating the Second Pandemic of the Black Plague through Metabolic Modeling

Authors:
David J Baumler, Jennifer L Reed, and Nicole T Perna**,** University of Wisconsin-Madison

Abstract:
Yersinia pestis is believed to have been the causative agent responsible for the 2nd pandemic of the Black Plague that killed 30-50% of the European population in ~1300 A.D. Recently, the genome of an ancient *Yersinia pestis* strain was sequenced from the dental pulp of human corpses believed to be victims of the 2nd pandemic of the black plague. Using a paleo systems biology approach, a genome-scale metabolic model (GEM) of this ancient microorganism has been constructed and compared with GEMs of seven modern *Yersinia* strains. Using these seven metabolic models, the metabolic capacity for carbon, nitrogen, phosphorous, sulfur, and iron utilization in aerobic and anaerobic conditions was conducted and found that *in silico* predictions accurately predict substrate utilization phenotypes to >80% accuracy when compared to experimental data. Overall, we identified virulence factor genes and differentiating catabolic phenotypes that are unique to the ancient plague strain and one modern *Y. pestis* strain, and not found in two *Y. pestis* strains that no longer cause human disease. This work reveals new insight about this human pathogen and has identified new targets for treatment methods if a similar plague was to reoccur.

Email of First Author: dbaumler@wisc.edu

EHR Implementation Planning Processes in Critical Access Hospitals

Authors:
Catherine K Craven, Pre-Doctoral Fellow[1], Lanis L Hicks[1], Gregory L Alexander[1], Leonard B Hearne[2], John H Holmes[3], MaryEllen C Sievert[1], [1]University of Missouri Informatics Institute, Columbia MO, [2]University of Missouri Bond Life Sciences Center and Statistics Department, Columbia, MO, [3]University of Pennsylvania, Perleman School of Medicine, Philadelphia, PA

Abstract:
Improved, high-quality, safer patient care, and our ability to pool clinical data for important secondary uses, depends on national adoption and effective use of integrated EHRs. The 2009 American Recovery and Reinvestment Act provided $30 billion in incentives for hospitals and provider practices to implement certified, integrated electronic health records (EHRs). Many small, rural hospitals, including most of those designated as Critical Access Hospitals (CAHs) will implement integrated EHRs for the first time. The study aim is to determine the EHR implementation planning processes in CAHs. Such processes are linked to whether or not EHRs are successfully adopted and used in larger hospitals, where human factors are far different. Grounded theory method will be used in this evaluation study. Data collection will comprise interviews and focus groups in four CAHs in Arkansas (1), Kansas (2), Tennessee (1), and two similarly small, rural hospitals in Chile for an international comparison. Atlas.ti coding software will be used to perform first and second cycle Process and Focused Coding. Results will aid in the development of a framework and improvements to guidelines for rural HIT implementation, inform additional EHR implementation waves here and internationally, and inform possible future directions in federal HIT policy.

Email of First Author: catherine.craven@mail.missouri.edu

The Impact of Resident Shift Schedule on Handoff of Care

Authors:
Taryn Price, Ellen J Bass, Ted Perez, Sharon Meth and Margaret Plews-Ogan, University of Virginia

Abstract:
Some residency programs employ a cross-cover system where, in the afternoon, several day teams handover patients to one day team called the cross-cover team. This cross-cover team supports all patients until handing over to a night float team that cares for the patients overnight. This work investigated the impact of such a shift schedule.
-Verbal handovers of 100 patients discussed at the cross-cover sign-out (when day teams handed over patients to the cross cover team) were compared to the discussion of the same patients at the night float sign-out (when cross-cover teams handed over patients to the night float team). Day team residents discussed their patients longer than the cross-cover teams did at the nightfloat handover. Day teams discussed background, problems, and anticipatory guidance significantly longer.
-To compare the night float handovers of one's own patients to the handover of patients handed over at the cross-cover, 33 verbal handovers of the cross-cover team's day patients were compared to their handovers of 153 patients for which they had taken responsibility at the cross-cover. Cross-cover residents discussed their day patients longer. The cross-cover teams discussed labs significantly longer for their own patients.

Email of First Author: thp4p@virginia.edu

Information Flow as a Metric for Evaluating Clinical Documentation Systems

Author:
Naqi A Khan, Vanderbilt University

Abstract:
Healthcare providers distill information-dense clinical encounters into focused patient histories, and then document their observations as clinical notes. This study tested the hypothesis that "information flow" is a valid metric for tracing information from patient, through a healthcare provider, and then to a resultant clinic note. The documentation systems of interest include a computer-based documentation (CBD) system and a dictation-based documentation (DBD) system. This study measured information flow as clinical concepts transferred from standardized case scenario descriptions to resultant notes. Two standardized clinical simulation studies provided the primary dataset for this study. The dataset included scenario descriptions, notes generated by physician subjects, and the simulation encounters' audio transcripts. Two independent physician reviewers identified clinical concepts present in the scenario descriptions and in the resultant notes. The reviewers analyzed four case descriptions and 64 resultant clinic notes, with a mean of 100 concepts/scenario and a mean of 79.9 concepts/note. Notes created via CBD yielded a mean of 95.1concepts/note, while those created via DBD contained a mean of 62.9 concepts/note. A significant ($p = 0.007$) distinction in concept permeability is observed for the studied tools. These preliminary findings suggest that information flow is a viable objective measure for evaluating documentation systems.

Email of First Author: naqi.a.khan@vanderbilt.edu

Which NLP Annotations Contribute to Accurate Automatic Problem List Generation?

Authors:
Danielle Mowery, Jan Wiebe, Shyam Visweswaran, Titus Schleyer, University of Pittsburgh and Wendy W Chapman, University of California, San Diego

Abstract:
Problems identified from clinical reports (emergency department and dental) may be used to support transition of patient care. Our objective is to determine the types of NLP annotations that can best contribute to accurate generation of a problem list. As a first step, we recruited medical (n=6) and non-medical (n=4) students to annotate 283 problems from 30 de-identified emergency department reports. Annotated information i.e., the annotation schema, included *attributes* indicating a problem's *experiencer, negation*, *uncertainty*, *intermittency*, *change*, and *generalized or conditional* status*.*
-Using Cohen's kappa, we observed the following inter annotator agreement against reference standard annotations: *experiencer* (1.0+-0), *negation* (0.8+-0.12), *certainty* (0.52+-0.11), *change* (0.63+-0.09), *intermittency (*0.39+-0.14*),* and *generalized/conditional* (0.46+-0.34)*.* Using a generalizability coefficient above 0.70, we determined we would need the following number of annotators to obtain reliable annotations for each category: 1 (*experiencer* and *negation*), 2 (*certainty* and *change*), 4 (*intermittency*) and 8 (*generalized or conditional*). Results suggest that some attributes are very difficult to annotate and may be difficult to obtain automatically.
We are currently generating a reference standard for the dental domain, which includes attributes such as *condition type, tooth location*, *negation, procedure codes, etc*.

-This research is being supported by the NLM Fellowship 5T15LM007059.

Email of First Author: dlm31@pitt.edu

Representing Clinical Guideline Recommendations Using the Quality Data Model

Authors:
Jeremy J Michel, Richard N Shiffman, Yale University

Abstract:
The National Quality Forum developed the Quality Data Model (QDM) as an information model for performance measures. Performance measures and clinical guidelines contain similar information, but this information is expressed differently. To date, the QDM has only been used with performance measures. Phrases built in the QDM are not organization dependent and may be implementable at multiple facilities. Using the QDM to represent the knowledge in guideline recommendations could allow for the increased adoption of guideline-based decision support.
We sought to investigate the QDM as a vehicle for representing clinical recommendations. First, we selected a recent guideline from the American Academy of Pediatrics.[1] We abstracted this guideline, parsed all individual recommendations into the Guideline Elements Model, and developed logic statements from each recommendation. We then developed QDM elements using the knowledge contained within selected recommendations. We next investigated how the QDM elements representing these recommendations could be linked together with other QDM structures. By making use of the Measure Authoring Tool (maintained by the National Quality Forum), we developed an active QDM representation of the selected clinical recommendations.

1. American Academy of Pediatrics. Urinary Tract Infection: Clinical Practice Guideline for the Diagnosis and Management of the Initial UTI in Febrile Infants and Children 2 to 24 Months. Pediatrics. 2011;128(3):595-610 Epub August 28, 2011.

Email of First Author: jeremy.michel@yale.edu

Building the Next Generation of Medical Education: Case Based Online Interactive Learning from Drawing Board to the Conference Room

Authors:
Timothy D Imler (Regenstrief Institute & Indiana University School of Medicine Department of Gastroenterology) and Jason Cadwallader (Regenstrief Institute)

Abstract:
Medical education has remained unchanged since the “modern” movement began with Osler. With information for practicing medicine increasing at an exponential rate the manner education is imparted to all levels of learners must be readdressed. Education time has also been getting compressed with increasing work hour restrictions and shift towards service requirements. In 2008 FromtheCaseFiles.com was designed by the authors at a single teaching institution to provide case based medical education to resident physicians. Multiple iterations of the system have languished without a structured mechanism for expansion with more than 380 physicians with 575 cases. Research on the tool has yielded significance ($p < 0.005$) in regards to allowing more information to be presented in a shorter period. The submission for publications also increased in a group of residents using the system for 9 months.

A major redesign of the structure is currently underway. The emphasis is on improving mobile interaction during group presentations. Online and interactive case based medical education is an important tool moving forward. The lessons learned in early construction of the tool have been instrumental in the redesign process and ongoing research will help to assess the impact that can be made.

Email of First Author: timler@iupui.edu

Exposing Benefits of Real-Time Tracking During Cancer Care with a Technology Probe

Authors:
Rupa A Patel, Predrag Klasnja, Andrea Civan Hartzler, Kenton T Unruh, Wanda Pratt, University of Washington

Abstract:
People with cancer experience many unanticipated symptoms and have difficulties communicating the full extent of symptoms to clinicians. Researchers have developed patient-reported outcomes (PRO) tools to address this issue. However, these tools capture retrospective data at appointments that are intended for clinicians to review. In contrast, real-time tracking tools could potentially affect patient outcomes and communication with clinicians, while also improving patients' symptom management strategies. We need to understand how real-time tracking tools can help patients.

We studied the tracking behaviors of 25 women with breast cancer. We provided 10 of these participants with a real-time tracking tool that served as a "technology probe" to uncover behaviors and benefits that emerged out of actual voluntary use. Our findings demonstrated that, compared to the fragmented and sporadic nature of tracking without a tool, participants with access to the real-time tracking tool used it more than was required. They experienced benefits that allowed them to see patterns among symptoms, feel psychosocial comfort from consistent tracking, and support two-way communication with clinicians. We conclude with recommendations, such as pre-populating the tool with personalized diagnosis- and treatment-related symptom metrics, to inform the design of future real-time tracking tools.

Email of First Author: rupatel@uw.edu

Feasibility of Collecting CAM Data Through a Computerized Patient Interview

Authors:
Lou Ann A Scarton, Qing T Zeng, Bruce E Bray, Laura Shane-McWhorter, University of Utah

Abstract:
Objective: To determine the feasibility of gathering Complementary and Alternative Medicine (CAM) data directly from the patient via a computerized patient interview.
Design: A quantitative descriptive study was utilized to determine whether patients would be willing to self-report their CAM usage and whether the self-reported data complements clinicians perceptions and medical records.
Measurements: 40 patients were recruited to test the computerized patient interview application. Clinicians and staff (n=15) were also surveyed to determine their perceived CAM usage. In addition, a retrospective chart review (n=100) was done to estimate the documented CAM usage rate.
Results: In this study, we had a 85% participation rate, suggesting patients are willing to share their CAM use through the computer application. The self-reported usage rate was 85%, as compared to the chart documented usage rate of 9.5% and the average clinician/staff estimated usage rate of 43%.

Email of First Author: louann.scarton@utah.edu

Structured Family Health Information Using www.MyFamilyatVanderbilt.com

Authors:
Jonathan Holt and Dario A Giuse, Vanderbilt University

Abstract:
Family Health Information (FHI) allows clinicians to identify patients that are at increased risk for adverse health conditions based on genetic predisposition. Patient-reported FHI is by its nature hearsay and is limited by each family member's health literacy. Clinicians will often clarify patient-reported FHI, updating the medical record and then will use this information for clinical decision-making, such as stratified preventative strategies. However, appropriate stratification of familial risk relies on the clinician's ability to ascertain and the patient's ability to report complete and accurate FHI. Complicating this factor, the collection of detailed FHI often requires more time than is available in the typical patient encounter in the primary care setting. As a result, FHI is often inconsistently and ineffectively communicated during clinical encounters. Yet, FHI epitomizes a cost effective strategy, and is critical to the emerging practice of genome-informed and personalized medicine. This gap creates an opportunity for the development of web-enabled tools to facilitate the ascertainment of FHI directly from the patient and their family members. At Vanderbilt University Medical Center's Department of Biomedical Informatics, we have developed w*ww.MyFamilyatVanderbilt.com (MyFaV),* which aims to enable the systematic collection of structured FHI directly from patients via a web-based portal.

Email of First Author: jonathan.holt@vanderbilt.edu

Richard J. Davidson

Wisconsin Showcase Speaker

Richard J. Davidson is the William James and Vilas Research Professor of Psychology and Psychiatry, Director of the Waisman Laboratory for Brain Imaging and Behavior, the Laboratory for Affective Neuroscience and the Center for Investigating Healthy Minds, Waisman Center at the University of Wisconsin-Madison. He received his Ph.D. from Harvard University in Psychology and has been at Wisconsin since 1984. He has published more than 250 articles, many chapters and reviews and edited 13 books. He has been a member of the Mind and Life Institute's Board of Directors since 1991. He is the recipient of numerous awards for his research including a National Institute of Mental Health Research Scientist Award, a MERIT Award from NIMH, an Established Investigator Award from the National Alliance for Research in Schizophrenia and Affective Disorders (NARSAD), a Distinguished Investigator Award from NARSAD, the William James Fellow Award from the American Psychological Society, and the Hilldale Award from the University of Wisconsin-Madison. He was the Founding Co-Editor of the new American Psychological Association journal EMOTION and is Past-President of the Society for Research in Psychopathology and of the Society for Psychophysiological Research. He was the year 2000 recipient of the most distinguished award for science given by the American Psychological Association –the Distinguished Scientific Contribution Award. In 2003 he was elected to the American Academy of Arts and Sciences and in 2004 he was elected to the Wisconsin Academy of Sciences, Arts and Letters. He was named one of the 100 most influential people in the world by Time Magazine in 2006. In 2006 he was also awarded the first Mani Bhaumik Award by UCLA for advancing the understanding of the brain and conscious mind in healing. Madison Magazine named him Person of the Year in 2007. In 2011, he was given the Paul D. MacLean Award for Outstanding Neuroscience Research in Psychosomatic Medicine.