National Library of Medicine Informatics Training Conference

June 23-24, 2015

National Institutes of Health
Natcher Conference Center, Building 45
Bethesda, Maryland

NATIONAL LIBRARY OF MEDICINE

TABLE OF CONTENTS

Wednesday, June 24, 2015

NLM Informatics Training Conference 2015
Natcher Conference Center, NIH Campus

Agenda

Tuesday, June 23, 2015

7:00 - 7:55 AM	Breakfast *Lower Level*
7:15 - 7:55 AM	Poster Setup *Atrium Lobby, Upper Level*
8:00 - 8:15 AM	Welcome and Opening Remarks Ms. Betsy Humphreys, Acting Director, NLM *Main Auditorium, Lower Level*
8:15 - 8:25 AM	Introduction of Training Directors and Trainees: Agenda Update Dr. Valerie Florance, NLM Extramural Programs *Main Auditorium, Lower Level*
8:25 – 8:30 AM	Welcome, Dr. Doug Fridsma, President and CEO, AMIA, *Main Auditorium*
8:30 - 9:45 AM	Plenary Session #1 Moderator: John Hurdle, Utah (1 hour 15 min, 5 papers) *Main Auditorium, Lower Level* 1. Predicting Discharge Date from the Neonatal ICU Using Progress Notes - Michael Temple/Vanderbilt 2. ActiveNotes: Designing an EMR Note with Free-Text Order Entry Authors - Adam Rule/UC San Diego 3. Optimizing Population Health Outcomes Utilizing Clinical and Administrative Data - Bobbie Kite/Ohio State 4. Reproducing a Prospective Clinical Study in MIMIC-II - Fabricio Kury/NLM 5. OrderRex: Data-Mining Clinical Decision Support from Electronic Medical Records - Jonathan Chen/VA

9:45 - 10:30 AM | Posters and Coffee Break – Attended: Posters 45 min Grouped by Topic
Atrium Lobby, Upper Level

Topic 1 - Healthcare Informatics: 10 posters

#101 A Dynamic Model of Usage of an Automated Physical Activity Intervention
- Daniel Schulman/VA

#102 Identifying Salient Concepts of Discussion in an Online ASD Community
- Michelle Chau/Columbia

#103 Modeling of Clinical Workflows in Ophthalmology Using EHR Data
- Michelle Hribar/OHSU

#104 Development and Preliminary Evaluation of a Prototype of a Learning Electronic Medical Record System
- Andrew King/Pittsburgh

#105 Developing a Local Terminology for Decision Support and Quality Measurement
- Yanhua Lin/Utah

#106 A Framework of Clinical Event Definitions and Usages
- Edna Shenvi/UC San Diego

#107 What is the Alignment of Discharge Readiness Perceptions Among Patients and Providers?
- Ross Lordon/Washington

#108 Self-Organizing Maps to Improve Risk Prediction in Hepatorenal Syndrome
- Jejo Koola/VA

#109 Determining Targets for Temporal Prediction Systems
- Eungyoung Han/Utah

#110 Computerizing Preclinical Interviews in Pediatric Diabetes Barrier Identification
- Yaa Kumah-Crystal/Vanderbilt

Topic 2 - Bioinformatics/Computational Biology: 7 posters

#201 Universally Conserved Spt5 Affects Antisense Transcription in *S. pombe*
- Scott Kallgren/Harvard

#202 Orthologous Retrotransposon Insertion Detection in Primate Genomes
- Thomas Meyer/OHSU

#203 Using Deep Learning to Find Low-Dimensional Representations of Gene Expression Data
- Jonathan Young/Pittsburgh

#204 Mitochondrial Heteroplasmy in 1000 Genomes
- Diego Calderon/Stanford

#205 Proteogenomics Discovery of Biomarkers of Lung Cancer Prognosis
- Michael Sharpnack/Ohio State

#206 A Pan-Cancer Modular Regulatory Network Analysis to Identify Common and Cancer-Specific Network Components
- Sara Knaack/Wisconsin

#207 Computational and Statistical Methods for Population Genomics
- Christopher Fragoso/Yale

Topic 3 - Clinical Research Translational Informatics: 4 posters

#301 Examination of Temporal Coding Bias Related to Acute Disease
- Mollie McKillop/Columbia

#302 Predicting Therapeutic Response and Prognosis in AML: A Crowdsourcing Approach
- David Noren/Rice U

#303 Ontological Content Auditing During Model Creation using the Foundational Model of Anatomy
- Lucy Wang/Washington

#304 Incorporation of Externally Generated Next-Generation Tumor Genotyping into Clinical Personalized Cancer Medicine Workflows
- Matthew Rioth/Vanderbilt

10:30 - 11:30 AM

Informatics Careers Panel: Career Transition Awardees (K01, K22, K99/R00)
Moderator: Robert El-Kareh, UC San Diego
(4 speakers, 12 minutes each plus questions)
Main Auditorium, Lower Level

1. Improving the Appropriateness of Clinical Decision Support Alerts and Clinician Responses
   - Allison McCoy, Tulane University
2. Protect Healthcare Data in the Cloud
   - Xiaquian Jiang, UC San Diego
3. High Throughput Phenotyping of Tumor Tissue
   - David Mayerich, U of Houston
4. Imaging Genomics as a “Big Data” Science: Mapping Genes on Brain Structure and Function
   - Kwangsik Nho, Indiana/Purdue

11:30 AM - 12:30 PM11:45 AM – 12:30 PM

Lunch
Natcher Grounds, Upper Level

- Birds of a Feather Lunch Tables
  - Tables with K awardees

- Topic tables: Clinical Informatics, Translational Bioinformatics, Clinical Research Informatics, Public Health Informatics, Dental Informatics

- Executive Session of Training Directors
(Dr. Florance chairs), 2 reps from each program, *E1/E2, Lower Level*
- Grant Program Session for Trainees and Faculty
(Ye/VanBiervliet) (K grants, ESI R01 grants), *F1/F2, Lower Level*

12:45 - 1:55 PM

Open Mic Session X1: Trainee Presentations in Translational Bioinformatics and Clinical Research Informatics
Moderator: Cindy Gadd, Vanderbilt
(10 speakers, 5 minutes per speaker including questions)
Main Auditorium, Lower Level

1. Bayesian Logistic Regression for Mining Biomedical Data
- Viji Avali/Pittsburgh
2. Evaluation of Current Technologies in Representing Concepts Found Within Genomic Reports
- Evelyn Rustia/Columbia
3. Drug Repositioning and Combination Therapy Discovery in Melanoma
- Kelly Regan/Ohio State
4. Search Optimization for Automated Clinical Trial Matching
- Tasneem Motiwala/Ohio State
5. Can Content-Based Image Retrieval Assist in Diagnosis? A Crowdsourcing Study
- Jessica Faruque/NLM
6. Ultrastructural Analysis of Sleep Deprived Brain Tissue using Automated Image Segmentation Techniques
- Kurt Weiss/Wisconsin
7. In-Depth Identification of Protein Sequence Variants and Postranslational Modification in 10 Human Cell Lines
- Anthony Cesnik/Wisconsin
8. Discovering Disease-Associated Molecular Interactions Using Discordant Correlation
- Charlotte Siska/Colorado
9. Visualization for Stability of Complex Phenotype and Biomarker Association
- Michael Hinterberg/Colorado
10. Integrating Literature and Experimental Data for Druggability Methods
- Emily Mallory/Stanford

2:00 - 3:00 PM

Parallel Paper Focus Session A
(3 papers at 10 minutes each plus 30 minutes for Q&A)

Focus Session A1
Moderator: Steven Bagley, Stanford
Main Auditorium, Lower Level

- Confidence and Information Access in Clinical Decision-Making: An

Examination of the Cognitive Processes that Affect the Information-Seeking Behavior of Physicians
- Raymonde Uy/NLM
- Clinical Information Needs in Acute Altered Mental Status
- Teresa Taft/Utah
- Systematic Review: Price Transparency in Clinical Care Lowers Cost and Quantity of Tests
- Katie Homann/UC San Diego

Focus Session A2
Moderator: Rebecca Jacobson, Pittsburgh
Balcony A, Upper Level

- Identifying Adverse Drug Events using Markov Networks and Temporal Dependence
- Aubrey Barnard/Wisconsin
- Use of Cytogenetic Data to Enable Drug Repurposing Studies
- Zachary Abrams/Ohio State
- Building Local Surgical Lexicons for Quality Improvement: The "Open Operative Report"
- Elizabeth Murphy/VA

Focus Session A3
Moderator: Philip Payne, Ohio State
Balcony B, Upper Level

- Concepts-Centric Approach to Research Registry Development
- Yauheni Solad/Yale
- Extending the Coverage of Phenotypes in SNOMED CT Through Post-Coordination
- Ferdinand Dhombres/NLM
- A Meta-Cluster Framework for Clustering Within and Across Datasets
- Katie Planey/Stanford

3:00 - 3:30 PM

Posters and Coffee Break – Attended: Posters 30 min Grouped by Topic
Atrium Lobby, Upper Level
Topic 1 - Healthcare Informatics: 10 posters

- Schulman/VA; Chau/Columbia; Hribar/Ohio State; King/Pittsburgh; Lin/Utah; Shenvi/UC San Diego; Lordon/Washington; Koola/VA; Han/Utah; Kumah-Crystal/Vanderbilt

Topic 2 - Bioinformatics/Computational Biology: 7 posters

- Kallgren/Harvard; Meyer/OHSU; Young/Pittsburgh; Calderon/Stanford; Sharpnak/Ohio State; Knaack/Wisconsin; Fragoso/Yale

Topic 3 - Clinical Research Translational Informatics: 4 posters

- McKillop/Columbia; Noren/Rice U; Wang/Washington; Rioth/Vanderbilt

3:30 - 4:45 PM	Plenary Session #2 Moderator: Lydia Kavraki, Rice (1 hour 15 min, 5 papers) *Main Auditorium, Lower Level* 1. Predicting Smoking Relapse through a Bayesian Model for fMRI Biomarker Identification - Sharon Chiang/Rice U 2. Ancestry Infused Variant Calling Pipeline - Latrice Landry/Harvard 3. Biological Network Visualizations: Exploratory versus Explanatory - Nikhil Gopal/Washington 4. Network-Regularization Improves Classification of Flu Vaccine Response - Stefan Avey/Yale 5. Genetic Association Testing in COPD Using Visual Assessment of Chest CT Images - Eitan Halper-Stromberg/Colorado
4:45 PM	Announcements, Dr. Florance
5:00 - 6:00 PM	Reception *NLM Lister Hill Center Lobby (Building 38A, Across the Street from Natcher)*
6:00 - 8:30 PM	Picnic *Natcher Grounds, Upper Level*

Wednesday, June 24, 2015

7:30 - 8:15 AM	Breakfast
8:15 AM	Announcements, Dr. Florance
8:20 - 9:15 AM	Open Mic Session X2: Trainee Presentations in Healthcare and Public Health Informatics Moderator: Bill Hersh, OHSU (11 speakers, 5 minutes per speaker including questions) *Main Auditorium, Lower Level* 1. RxMAGIC: A Dispensary Management Information System for Low-Resource Settings - Arielle Fisher/Pittsburgh 2. Situation Awareness for Labor and Delivery Suites in African Hospitals - JoAnna Hillman/Pittsburgh 3. Design of a Community-Engaged Health Informatics Platform - Mary Millery/Columbia 4. Evaluation of Data Visualizations from a Wearable Sensor System that Supports Elders Living in the Community - Uba Backonja/Washington 5. A Multivariate Timeseries Modeling Approach to Severity of Illness Assessment and Forecasting in ICU with Sparse, Heterogeneous Clinical Data Research Stage - Marzyeh Ghassemi/Harvard 6. Does Integrating Critical Radiology Alerts within the Electronic Health Record Impact Closed Loop Communication and Follow-Up? - Stacy O'Connor/Harvard 7. Identifying ECG Features in Congenital Heart Disease - Emily Hendryx/Rice U 8. MeTeOR: A Network with Biological and Translational Applications - Stephen Wilson/Baylor 9. Documentation of Race/Ethnicity using Unstructured Data in the Veteran Affairs Electronic Medical Record - April Mohanty/VA 10. Exploring Healthcare Mobility in the US to Improve Quality of Care Across Health Systems - Karen Wang/VA 11. The Quantified Brain: Mobile and Wearable Technologies for Measuring Brain Health - David Stark/Stanford
9:15 - 9:30 AM	Coffee Break
9:30 - 10:30 AM	Parallel Paper Focus Session B (3 papers at 10 minutes each plus 30 minutes for Q&A) Focus Session B1 Moderator: Cynthia Brandt, Yale

Main Auditorium, Lower Level

- Designing a Patient-Centered Informatics Tool for Brain Tumor Patients
  - Rebecca Hazen/Washington
- Growth of Secure Messaging Through a Patient Portal Across Clinical Specialties
  - Robert Cronin/Vanderbilt
- Measuring Patient-Perceived Quality of Care in US Hospitals from Twitter Data
  - Jared Hawkins/Harvard

Focus Session B2
Moderator: Peter Embi, Ohio State
Balcony A, Upper Level

- Predicting Negative Outcomes from Patient Surgical Vital Sign Quality
  - Risa Myers/Rice U
- Birth Month Affects Lifetime Disease Risk: A Retrospective Population Method
  - Mary Regina Boland/Columbia
- Recovering Causal Variables with Ridge Regularized Linear Models
  - Eric Strobl/Pittsburgh

Focus Session B3
Moderator: Mark Craven, Wisconsin
Balcony B, Upper Level

- Matrix Completion Methods and Imputation for EMR-Based Prediction
  - Erika Strandberg/Stanford
- Bayesian Tensor Factorization to Predict Drug Response in Cancer Cell Lines
  - Nathan Lazar/OHSU
- Alignment-free P-Clouds Extension and Detection of Ancient TE-Derived Fragments
  - Jaime Merlano/Colorado

10:30 - 11:30 AM

Open Mic Session X3: NLM Resources Available for Research Use
Moderator: Larry Hunter, Colorado
(10 speakers, 5 minutes per speaker)
Main Auditorium, Lower Level

1. RxNorm - Steve Emrick
2. Open-I - Dina Demner-Fushman
3. Training Opportunities at NLM - Paul Fontelo
4. Value Set Authority Center - Steve Emrick
5. NCBI Research Resources - Ben Busby
6. Semantic Medline - Tom Rindflesch
7. UMLS Resources - Steve Emrick
8. RxNav, RxClass, RxMix and APIs for NLM Drug Information Sources - Olivier Bodenreider
9. Metamap and MTI - Jim Mork
10. MeSH RDF - Steve Emrick

11:30 AM - 12:30 PM	Lunch *Natcher Grounds, Upper Level*
11:45 AM – 12:30 PM 11:45 AM – 12:30 PM	• Birds of a Feather Lunch Tables o Lister Hill Center NLM Summer Fellowships at NLM o NCI-VA Pilot Program o Topic tables: Clinical Informatics, Translational Bioinformatics, Clinical Research Informatics, Public Health Informatics, Dental Informatics • Grants Management/X-TRAIN Meeting (Mowery/McNair), *Room G1/G2 - Lower Level* • SciENcv & My Bibliography: Tools for Biosketches, Grant Reporting, & Compliance (Trawick), *Room F1/F2 - Lower Level*
12:30 - 1:45 PM	Plenary Session #3 Moderator: George Demiris, Washington (1 hour 15 min, 5 papers) *Main Auditorium, Lower Level* 1. Barriers to Physician Information-Gathering in the EHR: A Qualitative Study - Julie Doberne/OHSU 2. Machine Learning Analysis of Institutional Review Board Processing Times - Kimberly Shoenbill/Wisconsin 3. The Phenome Model: Probabilistic Phenotyping from Heterogeneous EHR Data - Rimma Pivovarov/Columbia 4. Novel Methods to Improve Household Food Environments Cheaply and at Scale - Philip Brewster/Utah 5. Using Laboratory Data for Prediction of 30-Day Hospital Readmissions - Amie Draper/Pittsburgh
1:45 – 2:00 PM	Closing Session *Main Auditorium, Lower Level*

TRAINING CONFERENCE PRESENTERS		
Program	**Presenter**	**E-Mail Address**
Columbia University		
George Hripcsak – Principal Investigator		hripcsak@columbia.edu
Plenary Presentation	Rimma Pivovarov	rp2521@cumc.columbia.edu
Focus Presentation	Mary Regina Boland	mb3402@columbia.edu
Poster Presentation	Michelle Chau	mmc2106@cumc.columbia.edu
Poster Presentation	Mollie McKillop	mm4234@cumc.columbia.edu
Open Mic	Mary Millery	mm994@cumc.columbia.edu
Open Mic	Evelyn Rustia	elr9078@nyp.org
Harvard University		
Alexa McCray – Principal Investigator		alexa_mccray@hms.harvard.edu
Plenary Presentation	Latrice Landry	latrice_landry@hms.harvard.edu
Focus Presentation	Jared Hawkins	jared_hawkins@hms.harvard.edu
Poster Presentation	Scott Kallgren	scottk@hms.harvard.edu
Open Mic	Marzyeh Ghassemi	mghassem@mit.edu
Open Mic	Stacy O'Connor	sdoconnor@partners.org
Ohio State University		
Peter Embi – Principal Investigator		peter.embi@osumc.edu
Philip Payne – Principal Investigator		Philip.payne@osumc.edu
Plenary Presentation	Bobbie Kite	bobbie.kite@osumc.edu
Focus Presentation	Zachary Abrams	zachary.abrams@osumc.edu
Poster Presentation	Michael Sharpnack	michael.sharpnack@osumc.edu
Open Mic	Tasneem Motiwala	tasneem.motiwala@osumc.edu
Open Mic	Kelly Regan	kelly.regan@osumc.edu
Oregon Health & Science University		
William Hersh – Principal Investigator		hersh@ohsu.edu
Plenary Presentation	Julie Doberne	doberne@ohsu.edu
Focus Presentation	Nathan Lazar	lazar@ohsu.edu
Poster Presentation	Michelle Hribar	hribarm@ohsu.edu
Poster Presentation	Josh Meyer	meyerjos@ohsu.edu
Rice University		
Tony Gorry – Principal Investigator		tony@rice.edu
Plenary Presentation	Sharon Chiang	sc4712@rice.edu
Focus Presentation	Risa Myers	rbm2@rice.edu
Poster Presentation	David Noren	david.noren@rice.edu
Open Mic	Emily Hendryx	emily.hendryx@rice.edu
Open Mic	Stephen Wilson	stephen.wilson2@bcm.edu

Stanford University		
Russ Altman – Principal Investigator		Russ.altman@stanford.edu
Focus Presentation	Katie Planey	katie.planey@stanford.edu
Focus Presentation	Erika Strandberg	estrandb@stanford.edu
Poster Presentation	Diego Calderon	dcal@stanford.edu
Open Mic	Emily Mallory	edoughty@stanford.edu
Open Mic	David Stark	dstark@stanford.edu
University of California, San Diego		
Lucila Ohno-Machado - Principal Investigator		lohnomachado@ucsd.edu
Plenary Presentation	Adam Rule	acrule@ucsd.edu
Focus Presentation	Katie Homann	khomann@ucsd.edu
Poster Presentation	Edna Shenvi	eshenvi@ucsd.edu
University of Colorado, HSC		
Lawrence Hunter - Principal Investigator		larry.hunter@ucdenver.edu
Plenary Presentation	Eitan Halper-Stromberg	eitan.halper-stromberg@ucdenver.edu
Focus Presentation	Jaime Merlano	jaime.merlano@ucdenver.edu
Open Mic	Michael Hinterberg	michael.hinterberg@ucdenver.edu
Open Mic	Charlotte Siska	charlotte.siska@ucdenver.edu
University of Pittsburgh		
Rebecca Jacobson - Principal Investigator		crowleyrs@upmc.edu
Plenary Presentation	Amie Draper	ajd109@pitt.edu
Focus Presentation	Eric Strobl	evs17@pitt.edu
Poster Presentation	Andrew King	ajk77@pitt.edu
Poster Presentation	Jonathan Young	jdy10@pitt.edu
Open Mic	Viji Avali	vra5@pitt.edu
Open Mic	Arielle Fisher	arf56@pitt.edu
Open Mic	JoAnna Hillman	jlh2422@pitt.edu
University of Utah		
John Hurdle - Principal Investigator		john.hurdle@utah.edu
Plenary Presentation	Philip Brewster	phil.brewster@utah.edu
Focus Presentation	Teresa Taft	teresa.taft@utah.edu
Poster Presentation	Eungyoung Han	e.han@utah.edu
Poster Presentation	Yanhua Lin	yanhua.lin@hsc.utah.edu

University of Washington		
George Demiris – Principal Investigator		gdemiris@u.washington.edu
Peter Tarczy-Hornoch - Principal Investigator		pth@u.washington.edu
Plenary Presentation	Nikhil Gopal	ngopal@u.washington.edu
Focus Presentation	Rebecca Hazen	hazenr@uw.edu
Poster Presentation	Ross Lordon	rlordon@uw.edu
Poster Presentation	Lucy Wang	lucylw@uw.edu
Open Mic	Uba Backonja	backonja@uw.edu
University of Wisconsin-Madison		
Mark Craven - Principal Investigator		craven@biostat.wisc.edu
Plenary Presentation	Kimberly Shoenbill	shoenbill@wisc.edu
Focus Presentation	Aubrey Barnard	barnard@cs.wisc.edu
Poster Presentation	Sara Knaack	saknaack@wisc.edu
Open Mic	Anthony Cesnik	cesnik@wisc.edu
Open Mic	Kurt Weiss	krweiss@wisc.edu
Vanderbilt University		
Cindy Gadd - Principal Investigator		cindy.gadd@vanderbilt.edu
Plenary Presentation	Michael Temple	michael.w.temple@vanderbilt.edu
Focus Presentation	Robert Cronin	robert.cronin@vanderbilt.edu
Poster Presentation	Yaa Kumah-Crystal	yaa.kumah@vanderbilt.edu
Poster Presentation	Matthew Rioth	matthew.j.rioth@vanderbilt.edu
Yale University		
Cynthia Brandt - Principal Investigator		cynthia.brandt@yale.edu
Michael Krauthammer - Principal Investigator		mkrauthammer@yale.edu
Plenary Presentation	Stefan Avey	stefan.avey@yale.edu
Focus Presentation	Yauheni Solad	yauheni.solad@yale.edu
Poster Presentation	Christopher Fragoso	christopher.fragoso@yale.edu
Veterans Administration		
Steven Brown - Training Director		steven.brown@va.gov
Plenary Presentation	Jonathan Chen	jonathan.chen@va.gov
Focus Presentation	Elizabeth Murphy	elizabeth.murphy5@va.gov
Poster Presentation	Jejo Koola	jejo.koola@va.gov
Poster Presentation	Daniel Schulman	daniel.schulman@va.gov
Open Mic	April Mohanty	april.mohanty@va.gov
Open Mic	Karen Wang	karen.wang@va.gov

National Library of Medicine		
Paul Fontelo - Training Director		pfontelo@mail.nih.gov
Plenary Presentation	Fabricio Kury	fabricio.kury@nih.gov

Focus Presentation	Ferdinand Dhombres	ferdinand.dhombres@nih.gov
Focus Presentation	Raymonde Charles Uy	raymondecharles.uy@nih.gov
Open Mic	Jessica Faruque	jessica.faruque@nih.gov
Informatics Careers Panel Open Mic Session		
Open Mic	Xiaquian Jiang	x1jiang@ucsd.edu
Open Mic	David Mayerich	dmayerich@gmail.com
Open Mic	Allison McCoy	amccoy1@tulane.edu
Open Mic	Kwangsik Nho	knho@iupui.edu
NLM Intramural Open Mic Session		
Open Mic	Steve Emrick	emricks@mail.nih.gov
Open Mic	Dina Demner-Fushman	ddemner@mail.nih.gov
Open Mic	Paul Fontelo	pfontelo@mail.nih.gov
Open Mic	Ben Busby	busbybr@mail.nih.gov
Open Mic	Tom Rindflesch	trindflesch@mail.nih.gov
Open Mic	Olivier Bodenreider	obodenreider@mail.nih.gov
Open Mic	Jim Mork	jmork@mail.nih.gov

PLENARY/FOCUS SESSION PRESENTATIONS

(Listed Alphabetically by Presenter)

Landry, Latrice	**Harvard University**	**Ancestry Infused Variant Calling Pipeline**	**56**
Lazar, Nathan	**Oregon Health & Science University**	**Bayesian Tensor Factorization to Predict Drug Response in Cancer Cell Lines**	**67**
Merlano, Jaime	**University of Colorado, Denver**	**Alignment-free P-Clouds Extension and Detection of Ancient TE-Derived Fragments**	**68**
Murphy, Elizabeth	**Veterans Administration**	**Building Local Surgical Lexicons for Quality Improvement: The "Open Operative Report"**	**51**
Myers, Risa	**Rice University**	**Predicting Negative Outcomes from Patient Surgical Vital Sign Quality**	**63**
Pivovarov, Rimma	**Columbia University**	**The Phenome Model: Probabilistic Phenotyping from Heterogeneous EHR Data**	**71**
Planey, Katie	**Stanford University**	**A Meta-Cluster Framework for Clustering Within and Across Datasets**	**54**
Rule, Adam	**University of California, San Diego**	**ActiveNotes: Designing an EMR Note with Free-Text Order Entry Authors**	**21**
Shoenbill, Kimberly	**University of Wisconsin-Madison**	**Machine Learning Analysis of Institutional Review Board Processing Times**	**70**
Solad, Yauheni	**Yale University**	**Concepts-Centric Approach to Research Registry Development**	**52**
Strandberg, Erika	**Stanford University**	**Matrix Completion Methods and Imputation for EMR-Based Prediction**	**66**
Strobl, Eric	**University of Pittsburgh**	**Recovering Causal Variables with Ridge Regularized Linear Models**	**65**
Taft, Teresa	**University of Utah**	**Clinical Information Needs in Acute Altered Mental Status**	**47**
Temple, Michael	**Vanderbilt University**	**Predicting Discharge Date from the Neonatal ICU Using Progress Notes**	**20**
Uy, Raymonde	**National Library of Medicine**	**Confidence and Information Access in Clinical Decision-Making: An Examination of the Cognitive Processes that Affect the Information-Seeking Behavior of Physicians**	**46**

POSTER PRESENTATIONS

(Listed Alphabetically By Presenter)

Rioth, Matthew	Vanderbilt University	#304 - Incorporation of Externally Generated Next-Generation Tumor Genotyping into Clinical Personalized Cancer Medicine Workflows	45
Schulman, Daniel	Veterans Administration	#101 – A Dynamic Model of Usage of an Automated Physical Activity Intervention	25
Sharpnack, Michael	Ohio State University	#205 – Proteogenomics Discovery of Biomarkers of Lung Cancer Prognosis	39
Shenvi, Edna	University of California, San Diego	#106 – A Framework of Clinical Event Definitions and Usages	30
Wang, Lucy	University of Washington	#303 - Ontological Content Auditing During Model Creation using the Foundational Model of Anatomy	44
Young, Jonathan	University of Pittsburgh	#203 – Using Deep Learning to Find Low-Dimensional Representations of Gene Expression Data	37

OPEN MIC PRESENTATIONS X1, X2, AND X3 AND THE INFORMATICS CAREER PANEL (Listed Alphabetically by Presenter)

Presenter	Institution	Title
Avali, Viji	University of Pittsburgh	Bayesian Logistic Regression for Mining Biomedical Data
Backonja, Uba	University of Washington	Evaluation of Data Visualizations from a Wearable Sensor System that Supports Elders Living in the Community
Bodenreider, Olivier	National Library of Medicine	RxNav, RxClass, RxMix and APIs for NLM Drug Information Sources
Busby, Ben	National Library of Medicine	NCBI Research Resources
Cesnik, Anthony	University of Wisconsin-Madison	In-Depth Identification of Protein Sequence Variance and Posttranslational Modification in 10 Human Cell Lines
Demner-Fushman, Dina	National Library of Medicine	Open-I
Emrick, Steve	National Library of Medicine	RxNorm/Value Set Authority Center/UMLS Resources/MeSH RDF
Faruque, Jessica	National Library of Medicine	Can Content-Based Image Retrieval Assist in Diagnosis? A Crowdsourcing Study
Fisher, Arielle	University of Pittsburgh	RxMAGIC: A Dispensary Management Information System for Low-Resource Settings
Fontelo, Paul	National Library of Medicine	Training Opportunities at NLM
Ghassemi, Marzhey	Harvard University	A Multivariate Timeseries Modeling Approach to Severity of Illness Assessment and Forecasting in ICU with Sparse, Heterogeneous Clinical Data Research Stage
Hendryx, Emily	Rice University	Identifying ECG Features in Congenital Heart Disease
Hillman, JoAnna	University of Pittsburgh	Situation Awareness for Labor and Delivery Suites in African Hospitals
Hinterberg, Michael	University of Colorado, Denver	Visualization for Stability of Complex Phenotype and Biomarker Association
Jiang, Xiaquian	University of California, San Diego	Protect Healthcare Data in the Cloud
Mayerich, David	University of Houston	Highthroughput Phenotyping of Tumor Tissue
McCoy, Allison	Tulane University	Improving the Appropriateness of Clinical Decision Support Alerts and Clinician Responses
Mallory, Emily	Stanford University	Integrating Literature and Experimental Data for Druggability Methods

Millery, Mary	Columbia University	Design of a Community-Engaged Health Informatics Platform
Mohanty, April	Veterans Administration	Documentation of Race/Ethnicity using Unstructured Data in the Veteran Affairs Electronic Medical Record
Mork, Jim	National Library of Medicine	Metamap and MTI
Motiwala, Tasneem	Ohio State University	Search Optimization for Automated Clinical Trial Matching
Nho, Kwangsik	Indiana/Purdue University	Imaging Genomics as a "Big Data" Science: Mapping Genes on Brain Structure and Function
O'Connor, Stacy	Harvard University	Does Integrating Critical Radiology Alerts within the Electronic Health Record Impact Closed Loop Communication and Follow-Up?
Regan, Kelly	Ohio State University	Drug Repositioning and Combination Therapy Discovery in Melanoma
Rindflesch, Tom	National Library of Medicine	Semantic Medline
Rustia, Evelyn	Columbia University	Evaluation of Current Technologies in Representing Concepts Found Within Genomic Reports
Siska, Charlotte	University of Colorado, Denver	Discovering Disease-Associated Molecular Interactions Using Discordant Correlation
Stark, David	Stanford University	The Quantified Brain: Mobile and Wearable Technologies for Measuring Brain Health
Wang, Karen	Veterans Administration	Exploring Healthcare Mobility in the US to Improve Quality of Care Across Health Systems
Weiss, Kurt	University of Wisconsin-Madison	Ultrastructural Analysis of Sleep Deprived Brain Tissue using Automated Image Segmentation Techniques
Wilson, Stephen	Baylor College of Medicine	MeTeOR: A Network with Biological and Translational Applications

Predicting Discharge Date from the Neonatal ICU Using Progress Notes

Authors:
Michael W Temple, Christoph U Lehmann, Daniel Fabbri, Vanderbilt University

Abstract:

Objectives

Discharging patients from the Neonatal Intensive Care Unit (NICU) can be delayed for non-medical reasons including the need for medical equipment, parental education, and children's services. We describe a method to identify patients who will be medically ready for discharge in the next 2-10 days – providing lead-time to resolve non-medical issues.

Methods

A matrix was constructed using 26 features (17 extracted, 9 engineered) and days to discharge (DTD) from daily progress notes of 4,693 patients (103,206 patient-days). ICD-9 codes classified patients as premature, cardiac, GI surgery, and/or neurosurgery. A supervised machine learning approach using a Random Forest defined important features and built a discharge prediction model.

Results

Three sub-populations (Premature, Cardiac, GI surgery) and all patients combined performed similarly at 2, 4, 7, and 10 DTD with AUC ranging from 0.854-0.865 at 2 DTD and 0.723-0.729 at 10 DTD. Neurosurgery performed worse scoring 0.749 at 2 DTD and 0.614 at 10 DTD. This model identified important features to construct simpler models. Using 2 features (oral percentage of feedings and weight) we constructed a model with an AUC of 0.843.

Conclusion

Using clinical features from daily progress notes provides an accurate method to predict when NICU patients are nearing discharge.

ActiveNotes: Designing an EMR Note with Free-Text Order Entry Authors

Authors:
Adam Rule, Steven Rick, Nadir Weibel, Zia Agha, University of California, San Diego

Abstract:
Notes are at the heart of medical records. They unify the record by describing a patient's history, interpreting test results, and justifying care plans. They are also where clinicians spend much of their time. Our prior research on primary care visits across several Veterans Affairs hospitals found that that clinicians spent nearly half (40%) of their EMR time in Notes and that most of their navigation between EMR sections involved going from Notes to Medications, Labs, or Orders and then returning to Notes. Despite their centrality, notes are rather static. In particular, executing the care plan described in a note usually requires leaving that note altogether. We attempt to unify the documenting and acting out of care plans with ActiveNotes, a prototype note editor that supports inline, free-text order entry for medications. Initial user tests with eight clinicians suggest that ActiveNotes' free-text order entry is quickly learned and could reduce documentation time by both unifying documentation and ordering and reducing the number of transitions clinicians need to make away from notes. Future research will develop methods for placing more complex orders, such as imaging or consult orders, and explore using dictation for free-text order entry.

Optimizing Population Health Outcomes Utilizing Clinical and Administrative Data

Authors:
Bobbie Kite, Tasneem Motiwala, Kelly Regan, Zachary Abrams, Ohio State University

Abstract:
Secondary data available from the payer side of the healthcare system necessitates population health informatics. These anonymized data provided from health plans include operational information (ex: claims, workflows), data from wearable devices and apps (ex: Fitbit, Garmin), as well as information regarding member interactions with their healthcare system (ex: labs, vital signs). The deidentified data of 61,000 members over the course of seven years from The Ohio State University Health Plan were used in conjunction with clinical measurement guidelines to characterize the member's status with regard to chronic conditions and develop population health program metrics. After member condition statuses were determined, this information was provided back to the health plan so the analyses could be used to influence health plan quality improvement operations as well as enlighten about the health plan population. These data were used in an iterative analysis/feedback loop to understand and improve the quality of both the patient/member and payer engagement while aiming to close the gaps in communication. The long-term objective is to make accessible data safely deidentified, and available for analyses on behalf of both research and business needs through an automated platform (data fusion), which allows for the portability and evaluation of tools.

Reproducing a Prospective Clinical Study in MIMIC-II

Authors:
Fabrício S P Kury, James Cimino, National Library of Medicine

Abstract:

Introduction: The widening scale of sharing of electronic health records (EHR) increases the interest in their possible secondary uses. In this paper, we approached one publicly available dataset of electronic health records – the MIMIC-II v. 2.6 database – and applied one possible study design, namely, a computational retrospective study.

Objective: To reproduce, as a computational retrospective study, a recent large prospective clinical study: the 2013 publication, by the Japanese Association for Acute Medicine (JAAM) Working Group, about disseminated intravascular coagulation (DIC), in the Critical Care Journal[1] (doi:10.1186/cc12783).

Methods: We designed an electronic phenotype and reproduced the prospective study's clinical data collection and statistical inference procedures in Java and R. All our source code and data are available online for free.
.

Results and Conclusion: Our electronic phenotype algorithm scanned 27,579 patients and identified 2,257 as eligible. DIC was diagnosed in 406 patients in the first day. Our results remarkably agreed with the prospective study and showed the prognostic power of the JAAM DIC score for predicting mortality. MIMIC-II lacked a minority of the data elements required by the reference study, and permitted statistical inferences with greater statistical power than the reference study in the majority of the cases.

References

Gando S. et al. A multicenter, prospective validation study of the Japanese Association for Acute Medicine disseminated intravascular coagulation scoring system in patients with severe sepsis. Critical Care 2013, 17:R111. Doi:10.1186/cc12783

OrderRex: Data-Mining Clinical Decision Support from Electronic Medical Records

Authors:
Jonathan H Chen[1,2], Mary K Goldstein[1,2], Steven M Asch[1,2], Russ B Altman[2]
[1]Department of Veterans Affairs, VA Palo Alto Health Care System, [2]Stanford University

Abstract:
Uncertainty and variability is pervasive in medical decision making with insufficient evidence-based medicine and inconsistent implementation where established knowledge exists. Clinical decision support constructs like order sets help distribute expertise, but are constrained by knowledge-based development. We produced a data-driven order recommender system to automatically generate clinical decision support content from structured electronic medical record data on >19K hospital patients. We present the first structured validation of such automatically generated content against the external standards-of-care established in clinical practice guidelines. For example scenarios of chest pain, gastrointestinal hemorrhage, and pneumonia in hospital patients, the automated method recommends orders that are referenced in practice guidelines with ROC AUCs (c-statistics) (0.89, 0.95, 0.83) that improve upon statistical prevalence benchmarks (0.76, 0.74, 0.73) and pre-existing human-expert authored order sets (0.81, 0.77, 0.73) ($P<10^{-30}$ in all cases). We demonstrate that automatically generated clinical decision support content can reproduce and optimize top-down constructs like order sets while largely avoiding inappropriate and irrelevant recommendations. This will be even more important when extrapolating to more typical clinical scenarios where well-defined external standards and decision support do not exist.

Poster #101
A Dynamic Model of Usage of an Automated Physical Activity Intervention

Authors:
Daniel Schulman, DeAnna Mori, Barbara Niles
Department of Veterans Affairs, VA Boston Healthcare System

Abstract:
Automated (eHealth) health behavior change interventions promise to improve the reach of behavioral medicine, but suffer from participant attrition and non-adherence with the intervention. Consequently, clients may not receive an adequate dose of intervention content. Data collection may be affected by missed assessment opportunities, which makes it more challenging to accurately evaluate an intervention, or to dynamically tailor an intervention to changing user behavior, attitudes, or beliefs.

We introduce a model of technology-assisted health behavior change which extends prior work to treat both usage and behavior change as dynamic processes: User characteristics, intervention characteristics, and environmental factors interact to influence intervention (non-)usage. Intervention usage leads to behavior change via various mechanisms of change. Behavior change modifies user characteristics, which influences future intervention usage.

We illustrate this model with a retrospective analysis of a telephone-based physical activity intervention for veterans with diabetes, jointly modeling self-reported physical activity and participation in weekly intervention sessions as coupled autoregressive processes. The results suggest a positive feedback loop, where greater self-reported physical activity predicts subsequent intervention usage, and usage predicts subsequent activity. This statistical model includes a principled approach to non-ignorable missing data due to non-usage, and allows for the prediction of future (non-)usage.

Poster #102
Identifying Salient Concepts of Discussion in an Online ASD Community

Authors:
Michelle M Chau, Sharon R Lipsky Gorman, Noémie Elhadad, Columbia University

Abstract:
Online health communities are emerging as a valuable source for identifying information needs of patients and care givers. Our study focuses on an online community for parents of children with autism spectrum disorders (ASD) and identifies specific health concepts that are salient to the community. We investigate a two-step approach to identifying such salient concepts. First, we establish a base lexicon of terms frequently used in the ASD community, by augmenting traditional health terminologies with automatically discovered ASD-relevant terms. Second, using a general parenting online community as a control, we determine which terms in our lexicon are statistically significantly more discussed in the ASD community than in the control (Chi-square association tests are carried out, controlled for multiple hypothesis false discovery rate). We find that in order to identify salient concepts representative of the community as a whole, its member-level frequency is more indicative than its raw frequency. Analysis of the salient concepts is ongoing. Current analysis indicates that (1) traditional terminologies lack coverage of concepts relevant to ASD discussions; and (2) the salient concepts in the ASD community span an impressive range of topics and semantic groups, suggesting of the high information burden of ASD parents.

Poster #103
Modeling of Clinical Workflows in Ophthalmology Using EHR Data

Authors:
Michelle R Hribar, Sarah Read-Brown, Leah Reznick, Lorinna Lombardi, Mansi Parikh, Thomas R Yackel, Michael F Chiang, Oregon Health & Science University

Abstract:
Patient flows within outpatient clinics are complex; patients see multiple providers in multiple rooms and have multiple procedures during a single encounter. Consequently, patient wait times can accumulate affecting patient satisfaction[1] and future care.[2] Being able to represent the time and variability of patient exams can greatly improve scheduling strategies through the use of simulation.[3,4] We investigate the accuracy of EHR timestamps for representing each provider's interaction length during the overall encounter; we present our model and the challenges for using EHR data to determine provider interaction lengths.

We performed a time-motion study of patient flows at 3 different outpatient ophthalmology clinics (112 encounters, 201 interactions). We then compared provider interaction lengths as determined from EHR timestamps to those from observations. Our initial analysis of one clinic (33 encounters, 82 interactions) showed that EHR timestamp data was ≤3 minutes from the observed timings for 55/82 (67%) interactions, and was ≥5 minutes from the observed timings for 14/82 (17%) interactions. Analysis showed that the greater time difference occurred when clinicians weren't using the EHR during exams. We conclude that the EHR timing data shows promise for representing clinical workflows for simulation studies of scheduling strategies.

Poster #104

Development and Preliminary Evaluation of a Prototype of a Learning Electronic Medical Record System

Authors:
Andrew J King, Gregory F Cooper, Harry Hochheiser, Shyam Visweswaran, University of Pittsburgh

Abstract:
Electronic medical records (EMRs) are capturing increasing amounts of data per patient. For clinicians to efficiently and accurately understand a patient's clinical state, we need a better way to determine when and how to display EMR data. We built a prototype system that records how physicians view EMR data, which we used to train models that predict which EMR data will be of interest in any given future patient. We call this approach a Learning EMR (LEMR). A physician used the prototype to review the laboratory test results of 59 de-identified ICU patient cases. We trained penalized logistic regression models that predict when each of 21 laboratory tests would be of interest. When judged relative to actual viewing behavior, the AUROC was as high as 0.92 and averaged 0.73, supporting that the approach is promising. To study the usability and acceptance of the approach, four ICU physicians used the interface to review five patient cases with data that had been manually highlighted. Overall, 3/4 physicians were enthusiastic about features of the prototype, and 4/4 of them thought a mature version of a LEMR would have a positive impact on quality of care.

Poster #105
Developing a Local Terminology for Decision Support and Quality Measurement

Authors:
Yanhua Lin, Catherine Staes, David Shields, Kensaku Kawamoto, Department of Biomedical Informatics, University of Utah, Salt Lake City, UT

Abstract:
Terminologies are a critical resource for enabling interoperable clinical decision support (CDS) and for performing quality measurement (QM). However, the enormous quantity of concepts available in standard terminologies makes it imperative to identify and maintain a much smaller subset of broader concepts with utility for both CDS and QM. Here, we report systematic methods to: 1) Identify concepts from standard terminologies and knowledge resources based on relevant attributes of the HL7 Virtual Medical Record (vMR) data model, and 2) Evaluate, 'clean', and maintain the local terminology. Candidate concepts were identified from various sources, including HITSP and the SNOMED CT CORE problem list. After review by a physician informaticist, approximately 3,100 CDS and QM-related concepts were uploaded into a terminology server (Apelon DTS) and mapped to vMR data attributes. This terminology was then 'cleaned' through de-duplication and naming standardization using UMLS MetaMap and Metathesaurus resources. The terminology is actively being applied to enterprise-level CDS and QM. Together, a CDS concept taxonomy based on the vMR can facilitate the use of standard terminologies to enable scalable and interoperable CDS and QM.

Poster #106
A Framework of Clinical Event Definitions and Usages

Author:
Edna Shenvi, University of California, San Diego

Abstract: The importance of temporality in clinical processes has prompted significant research and discussion about "events," although the definitions used for this term have varied widely. I review previous event models and definitions to show that they have generally been of three categories, for varied purposes: 1) temporally-associated data or concepts, for defining database storage or information extraction requirements; 2) system or entity interactions, for defining action triggers by information systems, decision support, or providers; 3) untoward occurrences, for quality and safety efforts or as illness or injury etiologies. Despite these formalized definitions and usages, the term "event" is commonly used differently in clinical settings. Using research on physician communication and information needs, informal literature on system design and practice commentary, and supplemented with a sample of medical texts, I demonstrate that key to this understanding of clinician usage of the term "event" is the centrality of change, in either patient status or clinical management, and propose a formal model for "clinical events" that reflects clinical usage of this term for filtering and processing of information. This has significant ramifications for information system design and a variety of informatics research efforts.

Poster #107
What is the Alignment of Discharge Readiness Perceptions Among Patients and Providers?

Authors:
Ross Lordon[1], Andrea Hartzler[2], Cheryl Armstrong[1], Heather Evans[1], William Lober[1]
1) University of Washington
2) Group Health Research Institute

Abstract:
Background: Postoperative patients with unmet information needs have increased risks for complications and readmission (1). Discharge teaching quality is a strong predictor of patients' readiness for discharge (2). However, little is known about patients' and providers' perception alignment concerning readiness. Methods: In a prospective observational study, 33 postoperative patients and 11 surgical providers were enrolled at an academic medical center and a county trauma center in Washington State. Patients were given discharge teaching and completed validated surveys (3) about their readiness perceptions and discharge teaching. Providers also completed the readiness survey. We assessed the kappa concordance of readiness perceptions between patients and providers, and the correlation of content patients needed versus received during discharge teaching. A convenience sample of 9 patients was used for concordance because not all providers completed the readiness survey for each patient. Results: The perceived readiness patient-provider concordance was 0.08, indicating little agreement pre-discharge. The provider readiness perception scores had little variation, with a mean score of 9. Patients reported a mean readiness perception score of 7.5. During discharge teaching, 12 patients (36%) received less information than needed. Indicating providers consistently overestimated their patients' readiness for discharge and some patients did not feel prepared for discharge.

Works Cited
1. Maloney L, Weiss M. Patients' Perceptions of Hospital Discharge Informational Content. Clinical Nursing Research. 2008;17(3):200-19.
2. Kneir S, Stichler J, Ferber L, Catterall K. Patients' Perceptions of the Quality of Discharge Teaching and Readiness for Discharge. Rehabilitation Nursing. 2014;0:1-10.
3. Weiss M, Piacentine L, Lokken L, Ancona J, Archer J, Gresser S, et al. Perceived Readiness for Hospital Discharge in Adult Medical-Surgical Patients. Clinical Nurse Specialist. 2007;21(1):31-42.

Poster #108

Self-Organizing Maps to Improve Risk Prediction in Hepatorenal Syndrome

Authors:
Jejo Koola[1,2], Samuel Ho[1,3,4], Michael E Matheny[1,2]
[1]Department of Veterans Affairs, VA Tennessee Valley Healthcare System, Nashville, [2]Vanderbilt University, [3]Department of Veterans Affairs, VA San Diego Healthcare System, [4]University of California at San Diego Medical Center

bstract: The Hepatorenal Syndrome (HRS), a form of cirrhosis specific kidney failure, has a high prevalence and mortality; the median survival for Type I HRS is two weeks. Current HRS mortality risk models suffer from poor performance and calibration. This study had two goals: (1) use unsupervised clustering to uncover variability in risk modeling performance among subgroups; (2) develop an effective visualization. We analyzed a retrospective cohort of 350 patients hospitalized at the Dept. of Veterans Affairs who had a discharge diagnosis of HRS. We modeled 14 input features: creatinine, bilirubin, albumin, vasopressor therapy, Model for End-Stage Liver Disease score, weight, oxygen requirement, temperature, respirations, oximetry, mean arterial pressure, sodium, International Normalized Ratio, and alcoholic cirrhosis. We constructed a 5x5 self-organizing map and predicted 90-day mortality using a logistic regression model for each resulting map-cluster unit. Median cluster size was 11 patients (range 2 – 33). Excluding the 6 units for which performance could not be calculated, the median AUC was 0.71 (range: 0.56 – 1.00). HRS mortality prediction demonstrated large variability for each subcohort. Unsupervised clustering and the resulting visualization may be used to inform model revision.

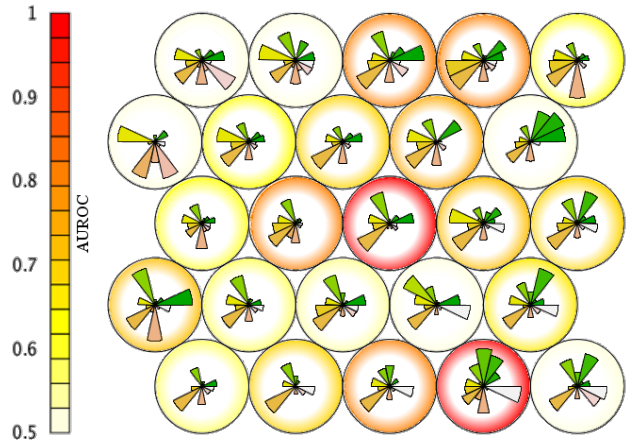


Figure 1: The twenty-five node Self-Organizing Map built with the 350 patient Hepatorenal Syndrome training set. The pie pieces correspond to each input feature referenced in the Methods section, starting from 12 o' clock and moving clockwise. Each unit's colored halo corresponds to its AUROC, which was set to clear for units with only one outcome class.

Poster #109
Determining Targets for Temporal Prediction Systems

Authors:
Eungyoung Han[1], Farrant Sakaguchi[2], Guilherme Del Fiol[1], Scott Narus[1,2], Bruce Bray[1], Peter J Haug[1,2]
[1]University of Utah; [2]Intermountain Healthcare, Salt Lake City, UT

Abstract:
Healthcare-Associated Sepsis (HAS) onset timing is troublesome to determine for various reasons: key pathophysiological variables are irregularly sampled; missing values are common; and the moment of onset is often unspecified. Timely automated identification is elusive because machine learning algorithms require these data. We propose a novel methodology to locate sepsis onset.
We aim to find temporal targets indicating HAS onset, which we validated with physicians' chart reviews. These targets help train Bayesian Network classifiers that can provide early HAS onset predictions from its clinical manifestations.
We developed an HAS onset identification toolkit comprised of kernelized support vector machines for classifying time intervals as septic given temporal features of predictor variables. We evaluate this approach's effectiveness through a three-physician chart review. The anticipated intra-class correlation coefficient determines the desired power's appropriate sample size.
We identified 301 subjects with antibiotics administration begun after a 72-hour incubation where sepsis as a non-primary discharge diagnosis.
Each patient's stay was partitioned into fixed-time-width intervals. We classified outermost intervals, attaining AUCs of over 0.95 for all kernel types tested. Classifier evaluation on intermediate intervals depends on our ongoing chart review results.

Poster #110
Computerized Preclinical Interviews in Pediatric Diabetes Barrier Identification

Authors:
Yaa A Kumah-Crystal, Laurie L Novak, Qingxia Chen, S Trent Rosenbloom, Vanderbilt University

Abstract:
Pediatric diabetes is a chronic disease that requires continuous daily self-maintenance. Identifying and addressing a child and family's barriers to diabetes self-management is important to improving care. During a clinic visit, factors such as time constraints and a family's inability to identify or communicate their barriers to adherence can result in missed opportunities for clinicians to help problem solve.

Computerized preclinical interview surveys (CPIS) can serve as key instruments to help families identify barriers to diabetes adherence, and to communicate those barriers to their clinicians. CPIS also offers the additional benefit of gathering data in a format that can be used to drive clinical documentation, thereby allowing the provider more time to discuss barriers with the family. In this study we will compare documentation resulting from CPIS facilitated clinical encounters to standard clinic visits. We will evaluate the utility of CPIS in facilitating the identification and documentation of barriers to diabetes management.

This study refines upon existing historic approaches for collecting patient information from computerized interviews, and develops a model for chronic diseases like diabetes. Demonstrating how patients can contribute to the documentation workflow in this manner should encourage reconsideration of our existing practices for clinical documentation.

Poster #201

Universally Conserved Spt5 Affects Antisense Transcription in *S. pombe*.

Authors:
Scott P Kallgren, Ameet Shetty, Burak H Alver, Peter J Park, Fred Winston, Harvard Medical School, Boston, MA

Abstract:
Spt5 is the only transcription elongation factor conserved in all three domains of life, but its molecular mechanisms are not yet thoroughly understood. Since it plays a global role in transcription, genomic approaches using an inducible depletion mutant should shed light on its general functions. We therefore sequenced nascent transcripts (NET-seq), mature mRNA (RNA-seq), and RNA polymerase II-associated chromatin (ChIP-seq) during Spt5 depletion. My analyses of RNA-seq show an increase in levels of transcripts antisense to 5' gene-coding regions upon Spt5 depletion. ChIP-seq corroborates this finding: I found a strong enrichment of RNA Pol II at the 5' ends of genes upon Spt5 depletion. I plan to next 1) compare RNA-seq and ChIP-seq results with NET-seq to determine whether this is the result of a transcript accumulation or nascent transcript increase, and 2) determine whether there are splicing defects associated with Spt5 depletion, and 3) and subset genes according to type of effect of Spt5: on mRNA 5' ends or splicing.

Poster #202
Orthologous Retrotransposon Insertion Detection in Primate Genomes

Authors:
Thomas J Meyer, Nathan H Lazar, and Lucia Carbone, Oregon Health & Science University

Abstract:
Over 50% of the human genome is comprised of transposable elements (Tes). Similar fractions have been observed in all of the sequenced non-human primate genomes. The most prolific TE class in these genomes is the retrotransposons, which mobilize via reverse transcription of an RNA-intermediate. This results in increasing copy numbers in host genomes over time. While studies often eliminate retrotransposon sequences from analysis due to the difficulty of working with these repetitive sequences, this can ignore important markers relevant to biomedical, population genetic, and evolutionary analyses. Retrotransposons have been shown to contribute to human disease through a variety of mechanisms, including insertion into genes and regulatory sequences as well as genomic rearrangements caused by non-allelic homologous recombination. They are also nearly homoplasy-free markers for population genetic and phylogenetic analyses. Investigations of these effects require the identification of orthologous retrotransposon insertions across samples, and our goal is to produce computational methods and workflows to meet these needs. Here, we present our initial strategies for automating the location and characterization of orthologous retrotransposon insertions across marmoset, rhesus macaque, gibbon, orangutan, gorilla, chimpanzee, and human draft genomes. Additionally, we present our initial analysis of polymorphic insertions identified in a population of rhesus genomes.

Poster #203

Using Deep Learning to Find Low-Dimensional Representations of Gene Expression Data

Authors:
Jonathan D Young, Xinghua Lu, University of Pittsburgh

Abstract:
Understanding the cellular signal transduction pathways that drive cells to become cancerous is fundamental to developing personalized cancer therapies that decrease the morbidity and mortality of cancer. The purpose of this study was to develop an unsupervised deep learning model for finding meaningful, low-dimensional representations of lung cancer gene expression data. Ultimately, we hope to use these low-dimensional representations to reveal hierarchical relationships (pathways) involved in cancer pathogenesis.
We downloaded 1,081 gene expression samples from The Cancer Genome Atlas. Each sample consisted of expression values for 20,501 genes. We modified publicly available MATLAB code to develop a Stacked Restricted Boltzmann Machines – Deep Autoencoder learning model for gene expression data. Hierarchical Clustering was performed on both the low-dimensional representations and the high-dimensional input data. Survival analysis of the clusters was performed using Kaplan-Meier plots.
More robust clustering was observed with the low-dimensional representations than with the high-dimensional representations. Statistically significant pairwise differences in cluster survival were observed in Kaplan-Meier plots of the low-dimensional representation clusters, but not with the high-dimensional representation Kaplan-Meier plots.
A deep learning model can be trained to represent meaningful abstractions of lung cancer gene expression data. Clustering the low-dimensional representations provided more insight into patient survival than clustering the high-dimensional input data.

Poster #204
Mitochondrial Heteroplasmy in 1000 Genomes

Authors:
Diego Calderon, Emily Glassberg, Arbel Harpak, Jonathan Pritchard, Stanford University

Abstract:
Mitochondrial heteroplasmy refers to the presence of multiple distinct mitochondrial genomes within a single individual. In humans, mutations in the mitochondrial genome are associated with a variety of bioenergetics defects and maternally inherited diseases. Here, we wanted to determine the prevalence of human mitochondrial heteroplasmy in the general population, and compare selective pressures acting on the mitochondria at the cellular and individual level. To identify heteroplasmic sites from next-generation sequencing data, we calculated log-likelihood ratios of simple models for homoplasmy and heteroplasmy at each position of the mitochondrial genome. We applied this method to 2535 individuals from the 1000 Genomes Project and found evidence of widespread heteroplasmy. Moreover, we compare observed patterns of heteroplasmy at sites that are likely to be neutral to those at sites that are likely to be influenced by selection, especially focusing on known disease-associated sites. These analyses form a basis for disentangling the contributions of various population genetic factors to observed patterns of mitochondrial genetic variation.

Poster #205
Proteogenomics Discovery of Biomarkers of Lung Cancer Prognosis

Authors:
Michael Sharpnack, Arunima Srivastava, Ferdinando Cerciello, David Carbone, Kun Huang, The Ohio State University

Abstract:
Cancer is a disease characterized by genetic alterations, many of which modulate RNA and protein expression. Mutations that affect nearly every level of biological regulation, including miRNA's, transcription factors, and splice factors, have been implicated in lung cancer. Although many biomarker studies have investigated the ability of RNA panels and low-throughput molecular screening of proteins to predict lung cancer prognosis, RNA-seq and shotgun proteomics have yet to be combined for this purpose. Interrogation of a single step in the creation of a functional protein, such as transcriptome profiling, is likely to be an incomplete snapshot of tumorigenesis. In addition to the increased sensitivity to differences between aggressive and non-aggressive tumors, integrative data can be used to discover regulation at the post-transcriptional level. With a network approach, we combine RNA-seq and proteomics datasets to discover crucial regulators of RNA and protein expression and coordination. Our results also have wider mechanistic implications for the role of dysregulation of the central dogma in cancer biology.

Poster #206

A Pan-Cancer Modular Regulatory Network Analysis to Identify Common and Cancer-Specific Network Components

Authors:
Sara Knaack, Alireza Fotuhi Siahpirani, Sushmita Roy, Wisconsin Institute for Discovery and University of Wisconsin, Madison

Abstract:
Human diseases like cancer are the result of changes to transcriptional regulatory networks of genes. Comparisons of regulatory networks across multiple cancer types are powerful for illuminating shared and unique network features of these disease states. Towards this goal, we recently undertook a pan-cancer analysis of clinical microarray expression data from The Cancer Genome Atlas (TCGA). We implemented a module- and network-based characterization of six cancers studied in TCGA: breast, colon, rectal, kidney, ovarian, and endometrial. We used the modular regulatory network learning with per-gene information algorithm (MERLIN) within a stability-selection framework to predict regulators of individual genes, and gene modules, for each cancer state. Our module-based analysis identifies a theme of immune system involvement. For each cancer, gene modules were inferred that were statistically enriched for immune response processes, as well as targets of key immune response regulators from the interferon regulatory factor (IRF) and signal transducer and activator of transcription (STAT) regulatory factor X (RFX) families of transcription factors. Comparing the regulatory networks inferred from each cancer type identified a core regulatory network that emphasized genes involved in chromatin remodeling, cell cycle, and immune response processes. Our integrated module and network analysis supported known themes in cancer biology, and emphasized regulatory hub genes with roles in immune response, cell cycle, and chromatin remodeling processes in all cancers.

Poster #207
Computational and Statistical Methods for Population Genomics

Authors:
Christopher Fragoso, Christopher Heffelfinger, Stephen Dellaporta, Hongyu Zhao (Yale University), and Mathias Lorieux (CIAT, IRD)

Abstract:
Our goal is to establish a population genomics pipeline for novel variant identification.
In the first phase of the project, genotyping information was obtained using a highly multiplexed, low-coverage, sequencing strategy called genotype-by-sequencing (GBS). GBS interrogates a genome at specific restriction sites to effectively reduce genome size in large population studies. A genomics pipeline was developed for read alignment, variant calling, and variant filtering. The next phase of the project was to develop a Hidden Markov Model (HMM) based method to impute missing markers. This step is critical because multiplexed sequencing results in missing markers and missing alleles in the form of erroneous homozygous calls. With the imputed datasets we are now characterizing the genetic structure of our study populations, including studies of recombination, segregation distortion, and residual heterozygosity. We will report on the genomic characterization of 2,000 rice Nested Associated Mapping (NAM) lines generated from ten diverse parental lines crossed to a common parent followed by single seed selfing for several generations.

Poster #301
Examination of Temporal Coding Bias Related to Acute Disease

Authors:
Mollie McKillop, Fernanda Polubriaginof, Chunhua Weng, Columbia University

Abstract:
Electronic Health Records (EHRs) hold great promise for secondary data reuse for clinical research but contain severe biases[1,2,3]. The temporal characteristics of coding biases remain unclear. This study used an empirical, data-driven approach to examine temporal bias trends for coding acute diabetic conditions. Specifically, 112 glucose-controlled patients initially ICD-9 coded for ketoacidosis between 2004 and 2014 were identified. Using survival analysis, it took an average of 9 months for 50% of glucose-controlled patients to have the incorrect code removed from their records. This effect was examined for different patient subgroups. We confirmed the bias trend in 17 hypoglycemic patients. We also discuss the implications for using ICD codes for acute conditions and provide recommendations for reducing temporal coding bias.

[1] Weng C, Li Y, Ryan P, *et al.* A distribution-based method for assessing the differences between clinicaltrial target populations and patient populations in electronic health records. *Appl Clin Inform* 2014;5:463–79.

[2] Pivovarov R, Albers DJ, Sepulveda JL, *et al.* Identifying and mitigating biases in EHR laboratory tests. *J Biomed Inform* 2014;51:24–34.

[3] Weiskopf NG, Rusanov A, Weng C. Sick Patients Have More Data: The Non-Random Completeness of Electronic Health Records. *AMIA Annu Symp Proc* 2013;2013:1472–7.

Poster #302

Predicting Therapeutic Response and Prognosis in AML: A Crowd Sourcing Approach

Authors:
David Noren[1], Raquel Norel[2], Byron Long[1], Gustavo Stolovitzky[2], Steven Kornblau[3], Amina Qutub[1]
1. Department of Bioengineering, Rice University
2. IBM Computational Biology Center
3. Department of Leukemia, University of Texas MD Anderson Cancer Center

Abstract:
Acute Myeloid Leukemia (AML) is a very potent form of cancer which currently has limited treatment options and a low 5 year survival rate (~25%). Although cytogenetic information has proven insightful in selecting personalized therapies for patients, the overall effectiveness of treatment remains low for AML. Clinical proteomics assays, like Reverse Phase Protein Array (RPPA), offer a more direct measure of the alterations of signaling pathways, which are the molecular targets of most cancer therapeutics. Unfortunately, incorporating clinical RPPA data into AML patient prognosis and therapy selection has been challenging due to the pronounced heterogeneity in patient characteristics and the noise inherent in the proteomics data.
Crowd sourcing experiments have proven effective in providing deeper insight into problems requiring complex data analysis. Using RPPA data and clinical information, we developed a number of baseline statistical models to predict AML patient response to therapy, relapse time, and survival time. We then designed the DREAM 9 AML Outcome Prediction Challenge as a crowd sourcing experiment, encouraging participants to develop more effective models. Through this effort, we were able to determine the patient characteristics which most impact prediction accuracy. In addition, we gained insight towards developing an improved method to determine patient prognosis.

Acknowledgement:
This research was funded in part by a training fellowship from the Gulf Coast Consortia, on the Training Program in Biomedical Informatics, NLM T15LM007093, PD - Tony Gorry.

Poster #303

Ontological Content Auditing During Model Creation using the Foundational Model of Anatomy

Authors:
Lucy L Wang, Eli Grunblatt, Mark Whipple, Ira J Kalet, University of Washington

Abstract:
Model construction using an established ontology is an opportune time to audit the content of that ontology. The Foundational Model of Anatomy (FMA) is one such ontology from which we can extract entities and relationships that constitute biological models. Because models generally focus on an organ system or part of the body, only a subset of FMA knowledge is used, allowing errors to be probed in depth. Then, by using the audited subset of knowledge immediately, the changes are also systematically tested and validated. Ontological inconsistencies can be either (1) computationally detectable and deducible, (2) computationally detectable but nondeducible, or (3) computationally undetectable (e.g. errors of content). We demonstrate the auditing of all three types of errors in a model of tumor dissemination through the head and neck lymphatics. Four classes of type 1 and two classes of type 2 errors were identified, involving 33 entities, leading to suggestions for 64 concepts to be added and 21 concepts to be removed. Additionally, 14 concepts were suggested for addition and 9 for removal due to type 3 errors. As these changes are propagated back into the FMA, they become available for future educational, research and modeling uses.

Poster #304

Incorporation of Externally Generated Next-Generation Tumor Genotyping into Clinical Personalized Cancer Medicine Workflows

Authors:
Matthew J Rioth, David Staggs, Lucy Wang, Jeremy L Warner, Vanderbilt University

Abstract:
Personalized cancer treatment is increasingly reliant on tumor genomic information for treatment decisions. Incorporation of genomic data into electronic health records (EHR) is needed; unfortunately, 3rd-party genotyping results are usually returned in formats (e.g. PDF) that are not computable for clinical decision support (CDS) or research purposes. We describe a solution for the discrete transfer of targeted exome tumor sequencing results from a 3rd-party CLIA-certified laboratory, in a standardized extensible markup language (XML) format. These data allow for displays of genotyping results in the Vanderbilt University Medical Center (VUMC) EHR and enable population-level aggregation for research use. The XML messages are received on a daily basis, automatically logged, parsed to match patients and displayed in their EHR with notifications sent to ordering providers. To date, VUMC has received 615 reports with 2,519 “actionable” genetic variants, sent to 44 ordering providers. Variants of unknown significance are included in a full-color PDF report that is transmitted along with the discrete data. To our knowledge, this is the first demonstration of the automated incorporation of 3rd-party tumor genotype information into a clinical EHR. Electronic transfer of genotype data enables rapid consumption of results by clinicians within their standard workflow, CDS, and secondary use.

Confidence and Information Access in Clinical Decision-Making: An Examination of the Cognitive Processes that Affect the Information-Seeking Behavior of Physicians

Authors:
Raymonde Charles Y Uy, Raymond Francis Sarmiento, Alex Gavino, Paul Fontelo, National Library of Medicine

Abstract:
Clinical decision-making involves the interplay between cognitive processes and physicians' perceptions of confidence in the context of their information-seeking behavior. The objectives of the study are: to examine how these concepts interact, to determine whether physician confidence, defined in relation to information need, affects clinical decision-making, and if information access improves decision accuracy.

We analyzed previously collected data about resident physicians' perceptions of information need from a study comparing abstracts and full-text articles in clinical decision accuracy. We found that there is a significant relation between confidence and accuracy ($\varphi=0.164$, $p<0.01$). We also found various differences in the alignment of confidence and accuracy, demonstrating the concepts of underconfidence and overconfidence across years of clinical experience. Access to online literature also has a significant effect on accuracy ($p<0.001$). These results highlight possible CDSS strategies to reduce medical errors.

Clinical Information Needs in Acute Altered Mental Status

Authors:
Teresa Taft, Charlene Weir, Stacey Slager, University of Utah

Abstract:
Delirium occurs frequently in patients over age 65, with up to 22% prevalence upon hospital admission and up to 30% occurrence during hospital stays. The annual economic burden of delirium in the United States has been estimated as high as $152 billion. Because delirium is sporadic and has differing presentations, it often goes undetected. However, when caught early and the cause is identified and treated, the duration of delirium can be reduced. Studies have shown that clinical decision support (CDS) tools can improve alignment with best practices; however, CDS that improves identification of the causes of delirium for diagnosis in real world settings has yet to be developed. Understanding physician's decision-making processes in addressing diagnostic dilemmas for acute mental status change would help in the design of future decision support. To help fill this gap, we conducted critical incident interviews with physicians reporting a complex case in diagnosing a patient with acute mental status changes. The interviews were audiotaped and transcribed. We conducted qualitative analysis to identify themes and developed a content coding taxonomy. These themes were then categorized to show decision processes, uncertainty factors, information that doctors consider before making decisions, sources of that information, and unmet information needs.

Systematic Review: Price Transparency in Clinical Care Lowers Cost & Quantity of Tests

Authors:
Katie Homann, Adam Rule, Skyler Kerzner, University of California, San Diego

Abstract:
Fewer than 20% of medical tests have readily available pricing (Bernstein, 2014). This significantly impacts patients, as they are paying a larger portion of their healthcare costs (Claxton, 2013) and the prices vary significantly between sites (Cauchi, 2013; Green, 2014). Price transparency is one proposed solution, however, it requires review of the scientific literature. Applying the PRISMA Systematic Review Criteria, we investigated PubMed and Web of Science databases for articles on price transparency. Over 2000 articles were identified, 69 articles met criteria for additional screening, and 11 met inclusion criteria as empirical studies. Price transparency lowered costs in the majority of articles (9 of 11, or 82%). In these studies, total costs decreased 3% to 37%, and number of tests decreased by 4.5% to 27%. Methodology varies between articles and literature in this field remains sparse. However, these findings suggest value in further exploring price transparency to control healthcare costs.

Identifying Adverse Drug Events Using Markov Networks and Temporal Dependence

Authors:
Aubrey Barnard, David Page, University of Wisconsin-Madison

Abstract:
Adverse drug events (ADEs) are a serious concern for public health. The scale of the problem has created a pressing need for automatic, computational approaches for ADE identification that will augment the existing spontaneous reporting, clinical research, and epidemiological studies. This talk/poster presents such a method for causal discovery in observational medical data that combines the probabilistic modeling of patient histories with scores that measure temporal dependence. Testing this Markov-network-assisted temporal scoring approach on a task and 5 datasets from the Observational Medical Outcomes Partnership shows that it can identify ADEs as well or better than existing epidemiological methods.

The Use of Cytogenetic Data to Enable Drug Repurposing Studies

Authors:
Zachary B Abrams, Philip RO Payne, Ohio State University

Abstract:
Background: Cytogenetic data in the form of karyotypes are commonly used in the diagnosis and treatment of many forms of cancer. Karyotype data are expressed in a text-based form that is not machine-readable. This limits the utility of these data for secondary use and research purposes. Utilizing the International System for Human Cytogenetic Nomenclature (ISCN), we developed a parsing and mapping system that allows karyotype data to be represented and analyzed in a computationally tractable manner. A Loss-Gain-Fusion model (LGF) was created that allowed us to represent each karyotype as a binary vector. Each cytogenetic region is represented three times (loss, gain, and fusion) in the model. We utilized the publicly available Mitelman database as a test-bed for analyses, focusing on problems related to drug repurposing.
Methods: Utilizing our computational model and the Mitelman database, we were able to successfully parse 98% of its karyotypes; of those parsed, 89.4% could be mapped into our binary Loss-Gain-Fusion model. We then classified karyotypes based on their disease labels and filtered out all diseases with less than 50 patients. We then selected genetic aberrations present in 20% or more of the population in which the cytogenetic event led to increased gene expression. Subsequently, we identified all genes in the affected region and found drugs that inhibited the function of the overexpressed gene using publicly available drug data in The Drug Gene Interaction Database (DGIdb). We performed a literature search on these results in Pub Med selecting diseases and drugs that did not co-occur and where the disease and the gene had co-occurred in at least one Pub Med abstract.
Results: We discovered 68,543 triplets containing (1) a disease, (2) an overexpressed gene, and (3) a drug that suppressed that specific gene. From this list, we discovered a total of 69 cancer disease-drug pairs that were not cited as co-occurring in the literature. Given this filtering process where the drug and gene are related, the drug suppressed the gene and the gene was implicated in the disease; it logically follows that the drug should be helpful in treating the disease.
Discussion: Our computational approach serves as a basis for new directions in drug repurposing, leveraging existing and commonly available bio-molecular phenotypic data. In order to validate our results, future laboratory-based testing will be conducted on a sub-set of our findings. The ability to link publicly available data sources is a central component of this work and emphasizes the importance of utilizing such data in conjunction with clinically-generated data sets so as to support in-silico hypothesis generation.

Building Local Surgical Lexicons for Quality Improvement: The "Open Operative Report"

Authors:
Elizabeth V Murphy[1, 2], James Cimino[3], Cynthia Brandt[4], Chris Lu[3]
[1]Department of Veterans Affairs, VA Portland Health Care System, [2]Oregon Health and Science University, [3]National Library of Medicine, [4]Department of Veterans Affairs, VA Connecticut Health Care System

Abstract: Operative reports are the primary source of clinical information about surgical diagnoses and procedures but they are almost exclusively in free text format, making evaluation of surgical outcomes and quality dependent on manual chart review, or ICD-9 and CPT codes that often are incomplete or inadequately categorized. We report the development of the "Open Operative Report" tool, using regular expression techniques to extract local operative report data including preoperative diagnosis, postoperative diagnosis, and operative procedures with greater than 99.5 % precision and recall into an operative report object and creating a local surgical lexicon. A categorization object accurately categorizes procedures by body part, including laterality, with a high degree of granularity including specific digits and spinal levels. The tool's output is easily transferred into a relational database for outcomes tracking, cohort retrieval, and surgical quality improvement. The local surgical lexicon is also used, with regular expression modifications, in a *MetaMap* strategy, mapping surgical terms to SNOMED ontology concepts with the goal of broadening the existing ontology to include specific surgical concepts that can then be used for cross institution outcomes research, and data-driven improvement efforts.

Concepts-Centric Approach to Research Registry Development

Authors:
Yauheni Solad[1], Allen Hsiao[2], Nitu Kashyap[2], James J Farrell[2,3], Richard Shiffman[1]
[1]Yale Center for Medical Informatics; [2]Yale New-Haven Hospital; [3]Yale Center for Pancreatic Diseases;

Abstract:
Identification of key elements for a research registry is a challenging task. Success depends on proper identification of key data elements and discovery of additional data elements that may be required for future studies.
Concepts-centric research registry design uses concepts extracted from narrative guidelines and relevant literature to identify key data elements during registry development.
We explored the feasibility of this approach in an attempt to improve and standardize the process of data elements selection for a research registry. We selected pancreatic cystic neoplasm (PCN) evaluation and management as a basis for a research registry.
We identified key concepts and extracted relevant data elements from a current PCN guideline to create an initial list of data elements. This list was subsequently expanded with relevant data elements identified in the literature review. Selected elements were used to create a PCN management ontology and were visualized in a concept map. We finalized a list of key data elements and used it to develop a research registry inside an Epic Electronic Medical Record (EMR) infrastructure.
Overall, we found this approach to be useful for researchers to ensure the inclusion of all necessary data elements in the research registry.

Extending the Coverage of Phenotypes in SNOMED CT Through Post-Coordination

Authors:
Ferdinand Dhombres, James T Case, Rainer Winnenburg, Olivier Bodenreider, National Library of Medicine

Abstract:
Our objective is to extend the coverage of phenotypes in SNOMED CT through post-coordination. The specific contribution of this work is to identify templates in SNOMED CT for the creation of post-coordinated expressions for phenotype concepts from the Human Phenotype Ontology (HPO).
We identified frequent modifiers in terms from HPO (e.g., "abnormality of", "congenital"), which we associate with templates for post-coordinated expressions in SNOMED CT (e.g., "abnormality of <ANATOMICAL_STRUCTURE>").
We identified 176 modifiers, created 12 templates, and generated 1,167 post-coordinated expressions. This approach significantly extends the mapping to pre-coordinated concepts used in most mapping studies, including our earlier work on the coverage of HPO terms in SNOMED CT.

In this preliminary study, we explored the automatic mapping of HPO terms to SNOMED CT through post-coordination. Through this novel approach, we were able to increase the current number of mappings by 50%.

A Meta-Cluster Framework for Clustering Within and Across Datasets

Authors:
Katie Planey, Olivier Gevaert, Stanford University

Abstract:
One way to prove subtypes (clusters) for a specific disease are robust is to show they exist across multiple datasets. But the problem of grouping similar clusters derived from clustering each dataset individually, for multiple datasets, poses methodological challenges. It can lead to an explosion of clusters, and it is not readily apparent how to compare them all across datasets. A common workaround is to normalize each dataset to remove study-specific signal, merge the datasets into one single matrix, and then cluster this matrix. However, this approach has several drawbacks, including: there is no consensus on the optimal normalization scheme, which highly affects the number of clusters selected and biological interpretation, feature sets may not completely intersect across all datasets, and outlier datasets can bias clustering results. Merging datasets also prevents the user from quantifying cohort similarities between each dataset-dataset pair, which may be helpful in quality control and data exploration.

We have developed a novel framework for discovering meta-clusters across datasets that does not require dataset merging and/or normalization to remove study-specific signal. We will present this framework and applications to breast cancer and ovarian gene expression databases.

Predicting Smoking Relapse through a Bayesian Model for fMRI Biomarker Identification

Authors:
Sharon Chiang[1], Michele Guindani[2], Francesco Versace[2], Marina Vannucci[1]
[1]Rice University [2] University of Texas MD Anderson Cancer Center

Abstract:
Despite the fact that cigarette smoking is the leading preventable cause of death in the United States, only 6% of smoking cessation attempts are successful after six months. Two major threats to relapse include smoking to avoid unpleasant withdrawal states, and cue-induced cravings in the presence of cigarette-related stimuli.

In this talk I will present a novel Bayesian statistical model that simultaneously (1) identifies neuroimaging biomarkers for smoking relapse, and (2) predicts the risk of smoking relapse on an individual patient level. The proposed model achieves high prediction accuracy and identifies several brain regions as potential fMRI regional biomarkers of smoking relapse.

Acknowledgement:
This research was funded in part by a training fellowship from the Gulf Coast Consortia, on the NLM Training Program in Biomedical Informatics, T15LM007093, PD - Tony Gorry.

Ancestry Infused Variant Calling Pipeline

Authors:
Latrice Landry, Sek Won Kong, Yassine Souilmi, Robert Green, Peter Tonellato, Harvard Medical School, Boston, MA

Abstract:
Clinical sequencing has the potential to revolutionize our understanding of human disease and the way in which we practice modern medicine. Genomic medicine has evolved primarily from research on populations with European Ancestry. The field of Clinical Molecular Genetics relies heavily on genetic databases and published research for the curation and classification of variants. Existing data deficits can translate into less meaningful results from genomic testing for non-European patients. The GATK 'best practices' pipeline does not include ancestry information. Inclusion of ancestry in variant calling, would allow us to assess the existence of any disparities between populations, as well as provide a foundation for addressing population based differences in human genome sequencing. Here we present our assessment of European bias in variant calling quality, accuracy and precision in the widely used GATK 'Best Practices' using 1000G sequencing data. Additionally, we present the design of our ancestry infused pipeline. The inclusion of genomics in medicine is expected to grow. As we transition to a more personalized paradigm in medicine, it is important to ensure the generalizability of the tools and advancements. This research is just a small part of that goal.

Biological Network Visualizations: Exploratory Versus Explanatory

Authors:
Nikhil Gopal, Neil F Abernethy, John H Gennari, University of Washington

Abstract:
Visualization can identify patterns otherwise undetected by traditional statistical analysis[1]. Although researchers often visualize biological networks intending to identify meaningful patterns, resulting networks appear as incomprehensible "hairballs" that obscure meaningful interactions in complexity[2]. We postulate that understanding how researchers use biological network visualizations can inform better designs.

We interviewed researchers about perspectives, challenges, and preferences in context of biological networks, pathway resources, and visualizations, and searched the transcripts for insights about objectives and barriers.

We carried out semi-structured interviews with 21 researchers. Through naïve thematic coding between three researchers, we generated a master codebook containing 10 sets of themes. Upon re-coding the interview transcripts using the master codebook, we generated a binary matrix representing membership between interview transcripts and codes. The themes (and code-sets therein) were systematically crossed with one another to yield hierarchically clustered dendrograms using *complete-linkage* clustering and *Binary* similarity.

Our results show (among others): a dichotomy between exploratory and explanatory use, certain layouts cluster with certain user interactions, clustering between challenges and training background, and that biology-specific researchers prefer to obtain context via text and others prefer cartographic methods. These results should be considered when designing biological networks visualizations to better support research.

Network-Regularization Improves Classification of Flu Vaccine Response

Authors:
Stefan Avey, Steven H Kleinstein, Yale University

Abstract:
Seasonal influenza viruses cause thousands of deaths annually worldwide and result in widespread disease and health care burden. While recommended for most individuals, the efficacy of the seasonal influenza vaccine is relatively low and host determinants of successful antibody response are poorly understood. Thus, identifying clinically relevant signatures of response is crucial to improve both vaccine delivery and design. Advances in high-throughput technologies over the last decade have resulted in large repositories of biological knowledge often summarized in gene interaction networks. Few others have attempted to utilize this prior biological knowledge for classification as it remains unclear how to best leverage this *a priori* knowledge. We apply a network-regularized machine learning algorithm to incorporate prior knowledge from gene networks into prediction of vaccination response from baseline gene expression data. Models trained using biological networks outperform random networks on an independent test set. Furthermore, the use of prior knowledge improves the interpretability of the gene signatures, as demonstrated by increased pathway enrichment. In this study, we propose a framework for incorporating prior biological knowledge into classification of influenza vaccination response and show that this improves both the accuracy and interpretability of the resulting model.

Genetic Association Testing in COPD Using Visual Assessment of Chest CT Images

Author:
Eitan Halper-Stromberg, University of Colorado, Denver

Abstract:
The utility of using standardized protocols for visually assessed computed tomography (CT) lung images to generate COPD phenotypes for use in genetic association studies is unknown. Automated analysis of CT lung images for this purpose is becoming common but likely has drawbacks relative to the performance of human analysts.
Methods: A standardized inspection protocol was used to visually assess chest CTs for 1264 non-Hispanic white (NHW) patients within the COPDGene cohort. Chest CTs from these individuals were assessed for the presence and severity of radiographic features related to both emphysema and small airway disease. For each visual CT phenotype, a genome-wide association study (GWAS) was performed, and two sets of SNPs with a higher prior likelihood of association were specified *a priori* for separate analysis. These sets were the set of SNPs previously associated with COPD susceptibility or emphysema at genome-wide significance (prior GWAS set) and the set of SNPs located within 100kb of a previously published set of COPD candidate genes (candidate gene set). For each visual CT feature, a corresponding semi-automated CT feature(s) was identified for comparison.
Results: Association for visual CT features in 1264 NHW subjects yielded significant results at all levels of inquiry; at the genome-wide level, at the candidate gene level, and at the prior GWAS lead SNP level.
Conclusion: Visually assessed emphysema was somewhat correlated with its semi-automated counterpart, but yielded better association results. Many of these associations were robust to adjustment for the semi-automated emphysema correlate. Visually assessed bronchial wall thickening was not correlated with its semi-automated counterparts and yielded different results from them. This study indicates that visual assessment of chest CT images holds new value in defining phenotypes for genetic association testing in COPD.

Designing a Patient-Centered Informatics Tool for Brain Tumor Patients

Authors:
Rebecca J Hazen, Mark H Phillips, John H Gennari, University of Washington

Abstract:
Patient involvement in care and decision-making has become an important issue in healthcare today. Understanding that health and disease extend beyond the walls of the clinic, researchers and clinicians have an opportunity, or even an obligation, to support and empower patients in understanding and managing their conditions outside of the clinic. Involving patients in the design and implementation of care activities and interventions is one way of achieving a more patient-centered environment. For patients with brain tumors, this can be a challenging task. These patients experience a wide range of symptoms and treatment effects that greatly affect their daily lives and abilities. They may also experience varying degrees of fatigue and cognitive impairment that impact their ability to participate in traditional stakeholder conversations and activities.

In this study, we employ Participatory Design techniques to engage patients and caregivers in designing an application to support patients in tracking, understanding, and communicating symptom information. Throughout this process, we work to understand the cognitive assumptions of design activities, data collection, and interpretation tasks in order to create a design process better reflecting the needs of this patient population. We present our approach and preliminary work towards developing a brain tumor-specific patient-centered application.

Growth of Secure Messaging Through a Patient Portal Across Clinical Specialties

Authors:
Robert M Cronin, Sharon E Davis, Jared A Shenson, S Trent Rosenbloom, Gretchen Purcell Jackson, Vanderbilt University

Abstract:
Patient portals are web-based applications that enable patients to interact with their healthcare providers. Secure messaging in portals is increasing as a form of outpatient interaction. Research about portals and messaging has been focused on primary care and medical specialties. We describe the adoption of secure messaging across a variety of clinical specialties. In our study we examined the use of patient-initiated secure messages and clinic visits in the three years following full deployment of a portal. We measured the proportion of secure messaging as a form of outpatient interactions (messages and clinic visits).

Over the study period 2,422,114 clinic visits occurred, and 82,159 unique portal users initiated 948,428 messages to 1,924 recipients. Medicine received the most messages (742,454), followed by surgery (84,001) and obstetrics/gynecology (53,424). The proportion of outpatient interactions through messaging increased from 12.9% in 2008 to 33.0% in 2009 and 39.8% in 2010. By 2010, this proportion was highest for obstetrics/gynecology (83.4% of outpatient interactions in obstetrics/gynecology), dermatology (71.6%), and medicine (56.7%). We demonstrated rapid growth of secure messaging across clinical specialties. With increased adoption of messaging across diverse clinical specialties, further research is necessary to determine implications for provider workloads and care delivered to patients.

Measuring Patient-Perceived Quality of Care in US Hospitals from Twitter Data

Authors:
Jared B Hawkins, Gaurav Tuli, Florence Bourgeois, John S Brownstein, Felix EC Greaves, Harvard Medical School, Boston, MA

Abstract:
Assessing the quality of patient care is essential to ensure that a high standard of care is maintained for all individuals across diverse healthcare institutions in the US. Traditionally, this has been measured using qualitative assessment of patients' perception of the quality of their own healthcare. The national standardized survey HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) has been considered the gold standard and allows valid comparisons to be made across hospitals. One major downside to HCAHPS is that there is a significant time-lag (+1 years) before official data release, which makes it difficult for patients to be informed about current opinions on the quality of healthcare at a given institution. Researchers have been experimenting with sentiment analysis on social media. Sentiment can be determined in several ways, with the principle being to classify the underlying emotional information as either positive or negative. Using this approach, we explored the use of Twitter as alternative, real-time, supplementary data to measure patient-perceived quality of care in US hospitals. Our aims are to: 1) capture hospital related posts from Twitter, 2) develop an automated approach to identify patient experience tweets, and 3) compare Twitter patient experience data to HCAHPS data.

Predicting Negative Outcomes from Patient Surgical Vital Sign Quality

Authors:
Risa B Myers[1], John C Frenzel[2,] Joseph R Ruiz[2,] Christopher M Jermaine[1]
[1] Rice University, [2] The University of Texas MD Anderson Cancer Center

Abstract:
Anesthesiologists carefully manage patient vital signs during surgery. Unfortunately, there is little empirical evidence that this management is correlated with patient outcomes. Using a database of over 90,000 cases, we seek to validate or repudiate current practice and determine whether those cases that anesthesiologists would subjectively decide are "low quality" are more likely to result in negative outcomes such as mortality, myocardial infarction or stroke. The problem reduces to one of multi-dimensional time series classification. To accomplish this goal, we have expert anesthesiologists independently label a small number of training cases, from which we train classifiers that label over 90,000 surgical cases.

Using these labels we can easily identify cases that anesthesiologists would consider low quality. We compare the prevalence of the negative outcomes in the best and worst 10% of cases and obtain strong, empirical evidence that current best practice is correlated with reduced negative patient outcomes. These labels can also be used as a teaching tool to identify and evaluate poor cases. Finally, our techniques can also be used to evaluate in-flight vital signs to determine if short-term negative outcomes such as elevated blood pressure values are likely.

Acknowledgement:
This research was funded in part by a training fellowship from the Gulf Coast Consortia, on the Training Program in Biomedical Informatics, NLM T15LM007093, PD – Tony Gorry.

Birth Month Affects Lifetime Disease Risk: A Retrospective Population Method

Authors:
Mary Regina Boland[1,5], Zachary Shahn[4], David Madigan[4-5], George Hripcsak[1,5], Nicholas P Tatonetti[1-3, 5]
[1]Department of Biomedical Informatics, [2]Department of Systems Biology, [3]Department of Medicine, [4]Department of Statistics, [5]Observational Health Data Sciences and Informatics (OHDSI), Columbia University, New York, NY, USA

Abstract:
An individual's birth month has a significant impact on their lifetime disease risk. Previous studies reveal relationships between birth month and several diseases including atherothrombosis, asthma, attention deficit hyperactivity disorder, and myopia, leaving most diseases completely unexplored. We developed a hypothesis-free method that systematically investigates disease-birth month patterns across all conditions. Our dataset includes 1,749,400 individuals with records at New York-Presbyterian/Columbia University Medical Center. We modeled associations between birth month and 1,688 diseases using logistic regression. Significance was assessed using a chi-squared test with multiplicity correction. We found 55 diseases with a significant birth month dependency. Of these 39 were reported in the literature, and a remaining 16 diseases were completely unreported. We found distinct incidence patterns across disease categories. Individuals born in birth months with higher cardiovascular disease incidence (February-June) were also associated with decreased life-expectancy in the literature corroborating our findings. Neurological diseases, pregnancy conditions and asthma associations revealed by our method were validated by European studies in the literature. Overall, we found that individuals born in May and July had the lowest overall disease risk. Lifetime disease risk is affected by birth month. Seasonally-dependent developmental mechanisms may help explain these associations.

Recovering Causal Variables with Ridge Regularized Linear Models

Authors:
Eric V Strobl, Shyam Visweswaran, University of Pittsburgh

Abstract:
Ridge regularized linear models (RRLMs), such as ridge regression and the SVM, are a popular group of methods which are used in conjunction with coefficient hypothesis testing to discover explanatory variables with a significant multivariate association to a response. However, many investigators are reluctant to draw causal interpretations of the selected variables due to the incomplete knowledge of the capabilities of RRLMs in causal inference. Under some reasonable assumptions, we show that a modified form of RRLMs can be used to get "very close" to detecting a subset of the Markov boundary by providing a worst-case bound on the space of possible solutions. The results hold for any convex loss, even when the underlying link function is highly nonlinear, and the solution is not unique. Our approach combines ideas in Markov boundary and sufficient dimension reduction theory. Experimental results suggest that the modified RRLMs are competitive against several other algorithms in discovering the Markov boundary from gene expression data.

Matrix Completion Methods and Imputation for EMR-Based Prediction

Authors:
Erika J Strandberg, Mohsen Bayati, Stanford University

Abstract:
The availability of electronic medical records (EMRs) has allowed for the use of a large amount of clinical data for predicting healthcare outcomes; however, the high proportion of missing entries, particularly those missing not at random, can make predictions and risk assessment difficult. Few studies have compared methodology for missing data completion with the objective of optimal predictive performance; even fewer studies have looked at imputing missing values in medical data with high amounts of incompleteness. We compare a novel multiple imputation method to matrix completion, imputation methods and complete case analysis. We assess performance on simulated data and on real EMR data. Results from this study indicate that the choice to impute is a good one; however, the best method to use may not be the same for all classification tasks. Multiple imputation and bagging hold much promise due to their good overall performance and much faster computation times.

Bayesian Tensor Factorization to Predict Drug Response in Cancer Cell Lines

Authors:
Nathan H Lazar, Mehmet Gonen, Shannon McWeeney, Kemal Sonmez, Oregon Health & Science University

Abstract:
Precision oncology aims to improve cancer patient outcomes by tailoring treatment to a given patient's tumor. Cell line screening panels give information on how specific tumors may respond to drugs by measuring the growth of tumor cells after exposure to compounds at varying doses in a high-throughput manner. By combining information from these panels with ‘omic profiles of cell lines as well as structural and target information on drugs we can build models to predict response and gain insight into response mechanisms.

Our method decomposes a central three-dimensional tensor encoding the responses of 70 breast cancer cell lines treated with 90 drugs at 10 doses into factors representing characteristics of the cell lines and compounds. By reconstructing the central tensor object from these factors, we predict full dose-response curves for missing cell line-drug combinations and can extrapolate to unseen cell lines and drugs. The use of Bayesian sparsity-inducing priors while training the model allows for the automatic determination of the most informative features. Lastly, by incorporating all of the data into one model, we are able to leverage information from each cell line and drug for the prediction of the others and examine relationships between cell line and drug features.

Alignment-free P-Clouds Extension and Detection of Ancient TE-Derived Fragments

Authors:
Jaime V Merlano, David Pollock, University of Colorado, Denver

Abstract:
We hypothesize that TE fragments degenerated and became part of what is known as the “black matter” of the genome. With no selections pressures, TE fragments accumulated mutations and diverged over millions of years. Over evolutionary time, the “leftovers” of TE expansion events diverged and created a background noise that, we believe, is the “fossil record” evidence of these expansion events.

A deeper exploration of these ancient genomic signatures is necessary to find the evidence to support or reject our hypothesis on the evolutionary origins of the remaining “black matter”. In order to meet the detection challenges of tracing back ancient TE fragments, we must balance the higher sensitivity of short annotation units by boosting the specificity of the current P-Clouds method.

We are achieving the detection specificity goal by taking advantage of relationships among clouds (i.e. cloud oligos, *k-mers* or simply cloud-mers) and exploiting the spatial information (orientation and positional) contained in pairs of adjacent non-overlapping clouds. The ability to enrich and filter the annotation units using spatial properties of cloud pairs (i.e. right order and right position) is a direct application of observed cloud organization and consistent patterns of cloud adjacency across several families of TEs.

Barriers to Physician Information-Gathering in the EHR: A Qualitative Study

Authors:
Julie W Doberne, Joan S Ash, Michael F Chiang, Oregon Health & Science University

Abstract:
Despite the purported benefits of electronic health records (EHRs) to improve quality of care and efficiency, questions remain about whether EHRs are meeting the information and workflow needs of physicians adequately. We conducted a qualitative study to describe what physicians perceive to be the current barriers to information gathering and overall workflow when using EHRs to evaluate new patients. Prominent themes pertaining to information gathering and overall workflow in the EHR were qualitatively identified from narrative survey responses using a grounded theory approach. Narrative responses from 327 physicians were obtained and analyzed. Major identified themes included: 1) Physicians struggle with unintuitive workflows and negative time impact; 2) EHR documentation was excessive and often of poor clinical value; 3) Provider-provider communication is negatively impacted by EHR challenges; and 4) Frustration with EHRs led to mistrust of vendors and clinical administration responsible for building and selecting the EHR software. Barriers such as inefficient workflows, increased time demands, and inconsistent documentation practices exist in EHRs that prevent ideal information gathering when evaluating a new patient. Results from this study could provide insights into new EHR interface redesign and development, and into new physician EHR training opportunities.

Machine Learning Analysis of Institutional Review Board Processing Times

Authors:
Kimberly Shoenbill, Yiqiang Song, Nichelle Cobb, Eneida A Mendonca
University of Wisconsin – Madison

Abstract:
Most human subjects research requires an Institutional Review Board (IRB) assessment of a protocol's adequacy of human subject protections before study initiation. Prior papers have discussed IRB inefficiencies, inconsistencies and lack of transparency impeding the conduct of research, but particular factors creating these barriers have not been adequately explored. In order to identify specific IRB and protocol features that correlate with delays or accelerations in the IRB review process, we evaluated two years of IRB data. These data contained IRB and protocol features (e.g., total IRB processing time, type of IRB review process completed, and protocol inclusion of vulnerable populations) from over 2,500 protocols. We used these data to develop predictive models to estimate total IRB processing time for newly submitted protocols.
Using descriptive statistics, we found several attributes having statistical significance including, Veterans' Administration study involvement, IRB member in charge of protocol review, and type of IRB review required (e.g., full, expedited). We also used machine learning algorithms including decision trees, support vector machines, neural nets and linear regression. Our most informative model was an M5P decision tree. Knowledge of these attributes and use of this model will assist efforts to improve IRB efficiency, identify areas where additional resources may be needed, better manage staff workload, and facilitate the conduct of human subjects research.

The Phenome Model: Probabilistic Phenotyping from Heterogeneous EHR Data

Authors:
Rimma Pivovarov, Adler J. Perotte, Edouard Grave, John Angiolillo, Chris Wiggins, Noémie Elhadad, Columbia University

Abstract:
Current patient phenotyping efforts, that include multiple expert panels and iterative design processes, produce specific and accurate disease models. However, because these expert-driven activities are often time-consuming, it is attractive to leverage unsupervised computational techniques for creating initial phenotypes, which can then be further refined by experts.

We developed the Phenome model to mine EHR data and learn models of disease. This initial version of the Phenome model is a fully unsupervised mixed-membership probabilistic model of longitudinal records and phenotypes. The model learns phenotypes from structured data (medication orders, laboratory tests, diagnosis codes) and clinical notes. The model output comprises computational disease models (the learned phenotypes) and an inference mechanism to assign phenotypes to unseen patient records.

The Phenome model was applied to two large EHR datasets and clinical settings for validity and generalizability. Quantitative evaluation on held-out data showed added-value over state-of-the-art mixture modeling techniques. There was a high correlation between inferred phenotypes and known disorders in gold-standard records annotated with SNOMED-CT, indicating accurate learned phenotypes. Qualitative evaluation showed that the learned phenotypes were more coherent than baseline models. Large-scale probabilistic phenotyping is a promising approach to learning accurate and interpretable computational disease models.

Novel Methods to Improve Household Food Environments Cheaply and at Scale

Authors:
Philip J Brewster, John F Hurdle, University of Utah

Abstract:
Households make decisions influencing dietary health by selecting specific sets of grocery food items while shopping. The goal of this research is to develop informatics application tools that measure and evaluate household grocery-purchasing patterns, as a basis for recommending food quality improvements at the household level.

Using data provided by a nationwide grocery chain, we analyze household grocery transaction sets longitudinally (~140,000 households for 15 months). We assess the quality of market baskets using their adherence to healthy patterns modeled in the USDA Food Plans, which meet the recommendations in the Dietary Guidelines for Americans at various budget levels. Our work proposes heuristics to evaluate the overall quality of any household's food purchases without relying on self-report, enabling inexpensive and scalable interventions.

We score food purchase quality using ratios of *actual* household expenditures to the *ideal* expenditures in the 29 food categories of the USDA Food Plans. We validate this scoring technique against a reference standard: the food-based component scores of the Healthy Eating Index 2010, using the CDC/USDA's NHANES survey. Further, we explore how geospatial factors from the retail data may refine hypotheses on how store neighborhoods impact the overall healthfulness of food purchases in specific at-risk populations.

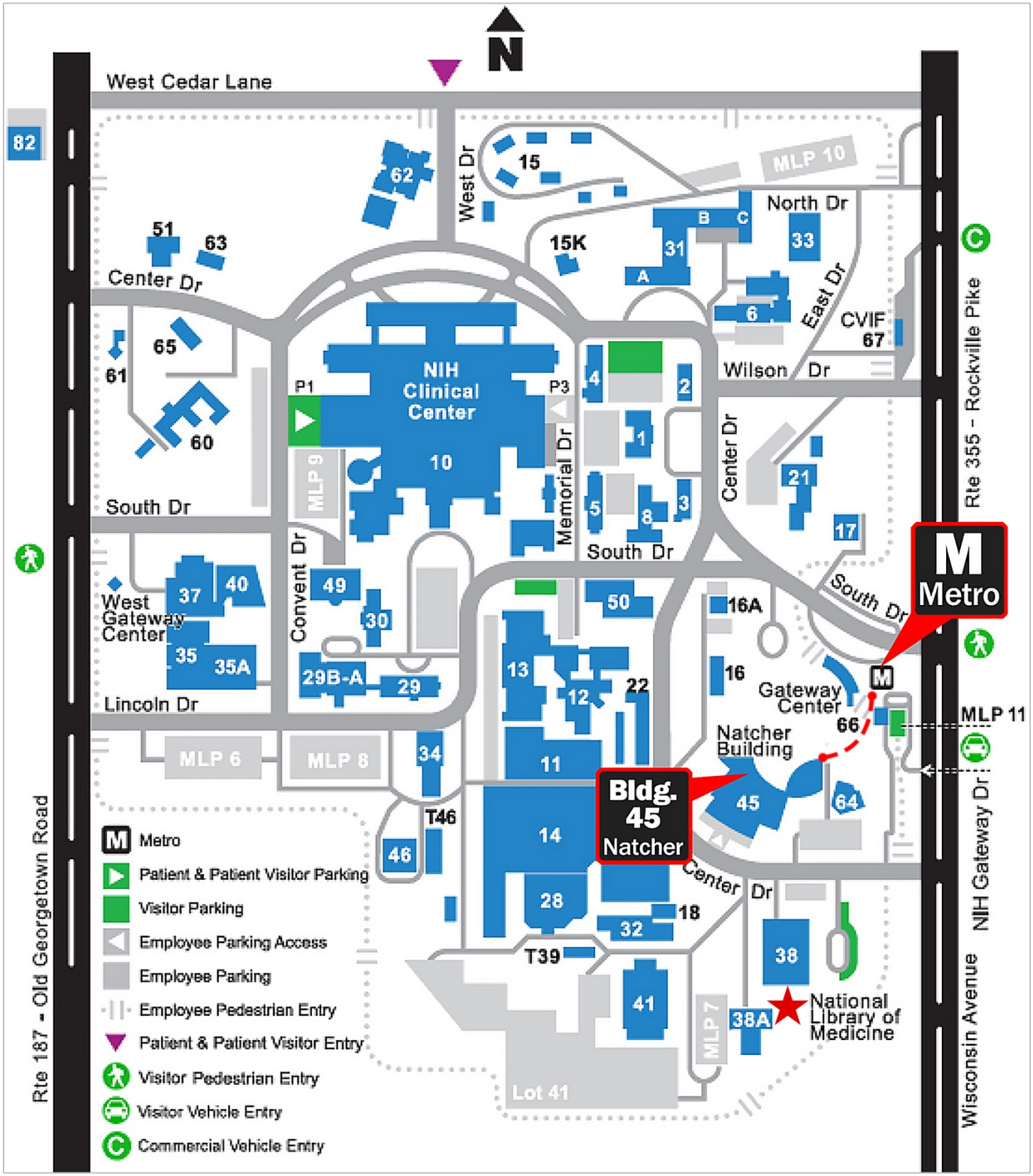
Using Laboratory Data for Prediction of 30-Day Hospital Readmissions

Authors:
Amie J Draper[1], Ye Ye[1], Victor M Ruiz[1], Christina Patterson[2], Andrew Urbach[2], Fereshteh Palmer[2], Christopher Myers[2], Fuchiang Tsui[1], [1]Department of Biomedical Informatics, University of Pittsburgh, [2]Children's Hospital of Pittsburgh of UPMC

Abstract:
Prior work in readmission risk prediction has under-utilized laboratory data, which may provide valuable information about a patient's condition. We aim to assess the contribution of laboratory data in predicting readmission risk. Preliminary work has focused on pediatric seizure, which has the highest volume of pediatric readmissions but no identified readmission risk factors.

We used discharge diagnosis ICD-9 codes to identify seizure-specific visits to Children's Hospital of Pittsburgh of UPMC during 2007-2012. Patients were considered readmitted if they returned to the hospital within 30-days post-discharge. We extracted features to summarize laboratory data for each patient. We used a training dataset (2007-2011) to rank features with information gain ratio and added features to a baseline model in order of rank. We kept features that improved prediction accuracy under 10-fold cross validation. A test (held-out) dataset (2012) was used to compare the AUROCs of the baseline model and the model with the added laboratory features. The addition of laboratory features significantly improved the prediction ability of the model, which suggests that laboratory data may be useful in identifying patients at risk of readmission. Ongoing work includes examining the contribution of laboratory data to readmission risk in adult heart failure patients.

NIH MAP


N
West Cedar Lane
82
Rte 187 - Old Georgetown Road
Rte 355 - Rockville Pike
Wisconsin Avenue
NIH Gateway Dr
Center Dr
West Dr
North Dr
East Dr
Wilson Dr
Memorial Dr
South Dr
Convent Dr
Lincoln Dr
MLP 10
MLP 9
MLP 6
MLP 8
MLP 7
MLP 11
NIH Clinical Center
P1
P3
West Gateway Center
Gateway Center
Natcher Building
Bldg. 45 Natcher
National Library of Medicine
Lot 41
CVIF
M Metro
Metro
Patient & Patient Visitor Parking
Visitor Parking
Employee Parking Access
Employee Parking
Employee Pedestrian Entry
Patient & Patient Visitor Entry
Visitor Pedestrian Entry
Visitor Vehicle Entry
Commercial Vehicle Entry

Agenda at a Glance

Tuesday, June 23, 2015

7:00 – 7:55	Breakfast **Lower Level**
7:15 – 7:55	Poster Setup **Atrium Lobby, Upper Level**
8:00 – 8:30	Welcome **Main Auditorium, Lower Level**
8:30 – 9:45	Plenary Paper Session #1 **Main Auditorium, Lower Level**
9:45 – 10:30	Poster and Coffee Break **Atrium Lobby, Upper Level**
10:30 – 11:30	Informatics Career Panel **Main Auditorium, Lower Level**
11:30 – 12:30	Lunch **Natcher Grounds, Upper Level**
11:30 – 12:30	Executive Session of Training Directors **Room E1/E2, Lower Level**
11:45 – 12:30	Grant Program Session for Trainees And Faculty **Room F1/F2, Lower Level**

Wednesday, June 24, 2015

7:30 – 8:15	Breakfast **Lower Level**
8:15 – 8:20	Announcements **Main Auditorium, Lower Level**
8:20 – 9:15	Open Mic Session X2 **Main Auditorium, Lower Level**
9:15 – 9:30	Coffee Break
9:30 – 10:30	Parallel Paper Focus Session B Focus Session B1 **Main Auditorium, Lower Level** Focus Session B2 **Balcony A, Upper Level** Focus Session B3 **Balcony B, Upper Level**
10:30 – 11:30	Open Mic Session X-3 **Main Auditorium, Lower Level**
11:30 – 12:30	Lunch **Natcher Grounds, Upper Level**

12:45 – 1:55	Open Mic Session X1 **Main Auditorium, Lower Level**
2:00 – 3:00	Parallel Paper Focus Session A Focus Session A1 **Main Auditorium, Lower Level** Focus Session A2 **Balcony A, Upper Level** Focus Session A3 **Balcony B, Upper Level**
3:00 – 3:30	Posters and Coffee Break **Atrium Lobby, Upper Level**
3:30 – 4:45	Plenary Paper Session #2 **Main Auditorium, Lower Level**
4:45 – 5:00	Announcements
5:00 – 6:00	Reception **NLM LHC Lobby, Bldg. 38A**
6:00 – 8:30	Picnic **Natcher Grounds, Upper Level**

11:45 – 12:30	Grants Management/X-TRAIN Meeting **Room G1/G2, Lower Level**
11:45 – 12:30 12:30 – 1:45	SciENcv and My Bibliography Tutorial **Room F1/F2, Lower Level** Plenary Paper Session #3 **Main Auditorium, Lower Level**
1:45 – 2:00	Closing Session **Main Auditorium, Lower Level**