Northwestern Site Presentation

eMERGE II Steering Committee Meeting
June 29-30, 2015
Rex Chisholm, PhD & Maureen Smith, MS, LGC
Overview

• Phenotyping and Genomic Analysis
• Genomic Medicine Pilot & Return of Results
• EHR integration
• PGx
Lower GI* Case & Control Study Flow

*Diverticulosis & colon polyp phenotypes followed the same basic flowchart

Base population: pts w/ >=2 outpt visits

- Pts w/ abdominal CT scan
  - Pts w/ colonoscopy
    - Pts w/ “polyp” in pathology report(s): Colon Polyp Cases
    - Pts w/o pathology reports & not colon polyp case, Or not diverticulosis case: Controls
    - Pts w/ “diverticul*” in colonoscopy, pathology, or CT scan report(s): Diverticulosis Cases
NLP for Pathology and Procedure Notes

- Extracting polyp type & location from the EHR, & also diverticulosis, required NLP at most sites
- We developed an NLP pipeline using the Unstructured Information Management Architecture (UIMA), with a mixture of cTAKES and Northwestern-developed UIMA components
- Applied the NLP pipeline to generate clinical quality metrics such as physician adenoma detection rate [Gawron, Thompson, Keswani, Rasmussen, Kho (2014), American Journal of Gastroenterology]
We created executable phenotyping workflows for our GI phenotypes using the open-source KNIME (Konstanz Information Miner) tool. All sites have to do is plug in their data.

- NLP code that processes text is bundled into workflow as an executable node.
- Shared across different sites – at least 4 other sites used this
- At least 1 other site, Geisinger, then also used KNIME to create portable workflows for their phenotypes (AAA & T2DM remission after RYGB)
Validation Tool

Luke R. developed this note abstraction tool @Marshfield, to ease creation of training & validation data sets, then extended @NU to use for colon polyps

Links to a collection of clinical notes & enables fast annotation of relevant colon polyp features

Generates an output file which is read by a KNIME workflow that calculates PPV, Recall, & F1

2 other sites are using both the validation tool & the KNIME workflow

Northwestern validation (of histology & location)

Sensitivity = 0.95
PPV = 0.95
5,285 cases; 3,761 controls for diverticulosis
Diverticulosis GWAS—AA diverticulosis vs. diverticulitis

rs62382461
PRR16 / LOC102467226
p=4.62E-08
RAF=0.28
Colon Polyp Results

- **Validation:**
  - NU: 94% PPV for cases, 98% PPV for controls
  - Geisinger: 100% PPV for cases, 96% PPV for controls

- **10,012 colonoscopies across 6 adult sites:**
  - 4,739 cases & 5,272 controls
  - 58% Female & 9% African (all)
  - Cases: important types & loc. of polyps are in table below

<table>
<thead>
<tr>
<th>Side of colon:</th>
<th>Carcinoma</th>
<th>Serrated / sessile adenoma</th>
<th>Tubular adenoma</th>
<th>Hyperplastic</th>
<th>Inflammatory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>50</td>
<td>117</td>
<td>1,839</td>
<td>1,378</td>
<td>20</td>
</tr>
<tr>
<td>Right</td>
<td>80</td>
<td>220</td>
<td>2,463</td>
<td>575</td>
<td>27</td>
</tr>
<tr>
<td>NA (not assessed)</td>
<td>6</td>
<td>6</td>
<td>129</td>
<td>76</td>
<td>3</td>
</tr>
<tr>
<td><strong>Totals:</strong></td>
<td><strong>136</strong></td>
<td><strong>343</strong></td>
<td><strong>4,431</strong></td>
<td><strong>2,029</strong></td>
<td><strong>50</strong></td>
</tr>
</tbody>
</table>
Top Hits

• rs62100507
  - Intronic in *INO80C (Chromatin remodeling complex) chr 18*
  - RAF = 0.12
  - OR = 1.39
  - P = 2.2E-07

• rs78233968
  - Intronic in long non-coding RNA *LINC01520 chr 10*
  - RAF = 0.09
  - OR = 0.61
  - P = 1.4E-06

• rs174414
  - Intronic in *SLCO4C1*-organic anion transporter (OATP) family. OATPs are involved in the membrane transport of bile acids, conjugated steroids, thyroid hormone, eicosanoids, peptides, and numerous drugs in many tissues, chr 5
  - RAF = 0.62
  - OR = 1.43
  - P = 2.5E-06
**caMRSA** Community Associated Methicillin-resistant Staphylococcus aureus

- MRSA now represents the most common cause of skin and soft tissue infections (SSTIs) in the U.S.

- ca-MRSA has replaced healthcare associated strains in many communities & express increased virulence factors leading to increased tissue destruction & more severe infections.

- Genetic host factors are suspected as a risk factor for recurrent infection with ca-MRSA, with an increased prevalence in younger, healthier populations

- caMRSA eHR phenotype definition
  
  - Not easy to define using eHR: tried both “gold” & “silver” definitions, still hard to find cases, went thru multiple iterations w/ multiple sites
## caMRSA eHR phenotype definition

<table>
<thead>
<tr>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inclusion Criteria:</strong></td>
<td><strong>Inclusion Criteria:</strong></td>
</tr>
<tr>
<td>All of the following (&quot;gold&quot; definition):</td>
<td>• Patients, aged 0 to 89,</td>
</tr>
<tr>
<td>• Bacteria culture-confirmed MRSA</td>
<td>• who receive routine primary care</td>
</tr>
<tr>
<td>• Culture drawn in the outpatient (inc. ED) setting, or w/in 72 hrs. of admission</td>
<td>• within a minimum of a 3 year continuous period of enrollment</td>
</tr>
<tr>
<td>• SST as site of infection</td>
<td></td>
</tr>
<tr>
<td>OR (&quot;silver&quot; definition) using ICD-9 codes:</td>
<td></td>
</tr>
<tr>
<td>1. caMRSA diagnosis</td>
<td>• patients, aged 0 to 89,</td>
</tr>
<tr>
<td>2. within +/- 7 days of #1, SSTI diagnosis</td>
<td>• who receive routine primary care</td>
</tr>
<tr>
<td><strong>Exclusion Criteria:</strong></td>
<td><strong>Exclusion Criteria:</strong></td>
</tr>
<tr>
<td>None of the following:</td>
<td>No prior history of any of the following:</td>
</tr>
<tr>
<td>• in the last year before MRSA infection:</td>
<td>• positive MRSA screen</td>
</tr>
<tr>
<td>• hospitalization</td>
<td>• MRSA infection</td>
</tr>
<tr>
<td>• prior stay in a long term care facility/nursing home</td>
<td>• SSTI</td>
</tr>
<tr>
<td>• during present admission: catheterization or insertion of indwelling percutaneous devices</td>
<td></td>
</tr>
</tbody>
</table>
CA-MRSA genomics methods

- AA  72 Case / 713 Control
- EU  270 Case / 5064 Control

  - All analyses stratified by race as determined by PC

- Controlled for age, sex, chip, site, PCs*
  - Fewer PCs used for AA, as many covariates was causing many models not to converge due to small numbers
CA-MRSA EA
23 and Me performed GWAS on self-reported history of 23 infectious diseases in 350,000 customers of EA descent

10 of 23 had GWAS significant hits in the HLA region
- chicken pox, cold sores, plantar warts, childhood ear infection, mumps, pneumonia, TB, scarlet fever, shingles, tonsillectomy
- caMRSA was not included

Interestingly, previous GWAS of staph infection have not seen a similar signal

*Other HLA/infectious*: Zoster in EMERGE and many other examples
CA-MRSA AA

Not HLA—Intergenic

Intron $MSH3$
Fluomics Project

MULTISCALE ANALYSIS OF INFLUENZA HOST-PATHOGEN INTERACTIONS: FLUOMICS

- Novel cellular proteins required during virus replication, new host cell modifiers of infection, and their functional importance in restriction of infection that will give insight into the biological basis of host resilience
- Collaboration with eMERGE to obtain genomic DNA for patients who have tested positive for influenza A
- Waiting on approval from BioVU and DUA with CHOP

<table>
<thead>
<tr>
<th>Original date</th>
<th>Updated</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct. 1, 2014</td>
<td>July/Aug-2015</td>
<td>DNA samples sent to NU, w/ demographic &amp; treatment setting phenotype data</td>
</tr>
<tr>
<td>Dec. 1, 2014</td>
<td>July/Aug-2015</td>
<td>Remaining phenotype data sent to NU</td>
</tr>
<tr>
<td>June, 2015</td>
<td>Oct-2015</td>
<td>Extreme phenotype sequencing done</td>
</tr>
<tr>
<td>Winter, 2015</td>
<td>June 2017</td>
<td>1st draft of manuscript</td>
</tr>
</tbody>
</table>
Phenotyping Tools

- Accepted AMIA 2015, Interactive Panel: (Authoring) Rules, (Distributed Query) Tools, and Drools: The challenging new world of high throughput phenotyping. Moderator: Jennifer Pacheco

- eMERGE Data Dictionary / Data File Validation Tool
  https://github.com/lrasmus/emerge_data_validation

- Document Abstraction
  https://github.com/lrasmus/DocumentAbstraction

- Phenotype Execution & Modeling Architecture (PhEMA)
  Authoring tool
  https://github.com/PheMA/phema-author
## Phenotyping

<table>
<thead>
<tr>
<th>Activities at Northwestern</th>
<th>eMERGE I</th>
<th>eMERGE II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotypes led by NU</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Phenotypes validated</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Phenotypes Implemented</td>
<td>14</td>
<td>22</td>
</tr>
</tbody>
</table>
eMERGE I Phenotype Collaborations

• T2D
  - MEDIA
  - MAGIC
• Glycemic traits
  - AAGILE/MEDIA Fasting Glucose Meta Analysis
  - GIANT Consortium
• Lipids
• RBC traits
  - COGENT
Genomic Medicine Pilot: Structure

Consultation
- Physician and Patient Advisory Committees
- Qualitative Assessment

EHRI
- Structural design
- Content; fact sheets, BPAs, Infobutton

IRB
- Returning results to biobank participants
- How to handle incidental findings

Laboratory
- Logistics
- Integrate with CDS
Genomic Medicine Pilot

• Population - Biobank subjects in eMERGE
• FVL, PT, HH from GWAS results
• CLIA validation and EHR integration of test results
• Qualitative study of returning results-Physicians – interviews complete
• Quantitative study of patient reported data (baseline, 1 mo and 6 months after return)-ongoing
• Qualitative study of biobank participant perspectives on receiving personal results - 30 complete
Advisory Committee: Physicians

• Composed of an existing QuIC within General Internal Medicine Department

• Some general themes from the meetings:
  - Ensuring that the conditions we return genomic information on are actionable and reliable
  - For PGX, concern about incomplete and conflicting evidence
  - Burden on physicians and patients; downstream costs for patients, time
  - Updating tools
  - Wanted patients to get results directly from study
  - Wanted all results sent to their inbox, did not want alerts only

• After return of results
  - Some results were confusing; getting intermediate results for patients already on meds
  - Busy physicians wanted to receive fewer results per week
  - Results were time consuming for some physicians
  - Some medications were changed after alerts
Advisory Committee: Patients

- Composed of NUgene biobank participants who will not receive results from the study
- Discussed several disease and PGx scenarios for returning results
  - Also reviewed consent documents/recruitment materials
- Some emerging themes from meetings:
  - Want the option to receive results; some want it even if no immediate impact on health
  - Want both positive and negative results, i.e. results suggesting genetic risk as well as when no risk was found
  - Want knowledgeable people to discuss results with
  - Like information kept in MyChart™, easy to access
  - Concern about 3rd party access to genetic information in the EHR
  - Wanted really simple consent forms and brochures about the study
    - Led us to conduct a study evaluating short and long (standard) consent forms
System Architecture

CLIA Lab -> Secure Data Receiver -> Ancillary Genomic System

Secure Data Receiver -> Format and store -> Actionable Variant KB

Ancillary Genomic System -> Knowledge Engine

Knowledge Engine -> Computed observations

Knowledge Engine -> Epic

Epic -> Interface

Interface -> Lab Results

Lab Results -> BPAs

BPAs -> Inbox

Inbox -> Physicians

 Physicians

MyChart -> Patients
Ancillary Genomics System

• Provides the link between a phenotype in the EHR and original data
  - Complete trace of interpretation logic
• Support heterogeneity of CLIA lab results
  - Star variants (Mount Sinai)
  - SNPs (Johns Hopkins)
• Versioning/modifications to interpretations
  - Provides a linked history of results as new knowledge (interpretations) change
Ancillary Genomic System

<table>
<thead>
<tr>
<th>Creation Date</th>
<th>Patient Name</th>
<th>DOB</th>
<th>MRN</th>
<th>Test Name</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/31/2013 10:44:20 AM</td>
<td>Zztest, Hamish</td>
<td>12/26/1961</td>
<td>100000084</td>
<td>Simvastatin Metabolism</td>
<td>Normal Activity (Predicted)</td>
</tr>
<tr>
<td>1/26/2014 10:01:56 PM</td>
<td>Zztest, Hamish</td>
<td>12/26/1961</td>
<td>100000084</td>
<td>Warfarin Dosing</td>
<td></td>
</tr>
</tbody>
</table>

AGS Interpretation - Report Preview

**Warfarin Dosing**

- **Value:**
  - Created: 01/26/2014 22:01:56

- **Report:**
  - RESULT
  - CYP2C9 *1/*2
  - VKORC1 (rs9923231) G/G

Use specific information below to determine the appropriate starting dose at http://www.warfarindosing.org.

- VKORC1-1639/3673: GG (warfarin insensitive)
- CYP2C9*2: CT (heterozygous)
- CYP2C9*3: AA (wildtype)
- CYP2C9*5: CC (wildtype)
- CYP2C9*6: AA (wildtype)

**INTERPRETATION**

Patient carries one active and one reduced activity CYP2C9 allele and, therefore, is expected to be able to metabolize medications via CYP2C9 less effectively. Intermediate metabolizers may require non-conventional doses of medications whose major metabolic pathway is CYP2C9 or use of another drug that is not processed by CYP2C9.
EHR Implementation

• PGx
  - CLIA lab results received for 750 patients
  - Results being released to the EHR

• GMPP
  - CLIA lab results for 99 participants
  - Results released into EHR

• Infobuttons
  - Focus groups/interviews conducted
  - Evaluating usage statistics
  - Adjustments based on feedback

<table>
<thead>
<tr>
<th>Format of Lab Results</th>
<th>Storage of Results</th>
<th>Interface into Epic</th>
<th>Decision Support</th>
<th>Implementation Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGx</td>
<td>Excel documents</td>
<td>Ancillary Genomics System</td>
<td>HL7 interface</td>
<td>12 CDS rules implemented</td>
</tr>
<tr>
<td>GMPP</td>
<td>HL7 messages</td>
<td>Hospital Lab System (Cerner)</td>
<td>HL7 interface</td>
<td>8 CDS rules implemented</td>
</tr>
</tbody>
</table>
EHRI tools

• Joomla Infobutton Responder
  - Code available on request - to be made publicly available by the end of July

• Ancillary Genomics System (AGS)
  - Code available on request

• MyResearch – Genomic Results
  - Code available on request

• HL7 Infobutton.NET
  -https://github.com/lrasmus/HL7InfobuttonDotNet

• Clinical Knowledge Manager
  -https://github.com/lrasmus/ClinicalKnowledgeManager

• Patient Genomics Viewer
  -https://github.com/lrasmus/patient-genomics-viewer
Laboratory

• Hospital-based molecular diagnostics laboratory
• Plan to use model to fully integrate molecular results and CDS with existing hospital systems
  - Leadership fully support plan
  - Several operational issues to overcome
• Lessons learned
  - Changing operating procedures takes a lot of time
  - Clinical and research operations very different
  - Involvement of Epic and Powerchart teams at beginning is important
  - Time to develop CDS can be reduced with experience from past projects
  - IT teams are experienced with CDS implementation
• Laboratory assessing study; tentative plans to adopt reporting and fact sheets we developed
Return of Results

• 29 people have received results
  - 7 carriers (6 HH, 1 PT)
  - Results in EHR with BPA firing (all patients)

• Comparison of CLIA results with GWAS for 19 participants
  - Discordance for 2 participants who are heterozygous for PT mutation, both were predicted to be homozygous for the variant
PGx Study

- 750 participants enrolled
- Pre and post surveys sent to all participants
- Results returned to 450 participants
  - Remainder will be returned over next 5 weeks to keep in line with physician requests
  - Considered return of SCN5A and KCNH2 variants
    - Internal deliberation and consultation with specialists
Decision Tree: Gene on ACMG List

- **Variant classified as benign**
  - No further action

- **Variant classified as VUS**
  - Consult experts

- **Consider chart review**
  - No clinical phenotype present
    - No further action
  - Clinical phenotype/clinical issue present
    - Consult experts
      - No clinical phenotype present
        - Inform physician and patient
      - Clinical phenotype/clinical issue present
        - Inform physician and patient
Shortened Consent Study
Graduate Genetic Counseling Student Project

• Compared shortened consent vs. standard length consent for the first 150 PGx participants

<table>
<thead>
<tr>
<th></th>
<th>Word count</th>
<th>Paragraphs</th>
<th>Number pages</th>
<th>Flesch Reading Ease</th>
<th>Flesch-Kincaid Grade Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard consent</td>
<td>2524</td>
<td>86</td>
<td>6</td>
<td>48.5</td>
<td>11</td>
</tr>
<tr>
<td>Shortened consent</td>
<td>1888</td>
<td>64</td>
<td>4</td>
<td>49.2</td>
<td>10.6</td>
</tr>
</tbody>
</table>

• Participants showed no statistical differences in objective or subjective comprehension

• Finishing manuscript, sharing findings with IRB
Northwestern Team

• Rex Chisholm
• Maureen Smith
• Justin Starren
• Abel Kho
• Geoffrey Hayes
• Laura Rasmussen Torvik
• Luke Rasmussen
• Jennifer Pacheco
• Sharon Aufox
• Cathy Wicklund
• Vivian Pan

• Stephen Persell
• Tim Herr
• Siddhartha Jonnalagadda
• Will Thompson
• Kathryn Jackson
• Jess Behrens
• Loren Armstrong
• Michael Heathcote
• Angelica Espinoza
• Oana Popescu
Questions?