Enabling Oncology Research Through De-identification

Bradley Malin, Ph.D.
Assoc. Prof. & Vice Chair of Biomedical Informatics, School of Medicine
Assoc. Prof. of Computer Science, School of Engineering
Director, Health Information Privacy Laboratory
Vanderbilt University
2/24/2014
Disclosure

• Paid consultancies

  Celgene (2013 – present)
  Sanofi-Aventis (2013 – present)

* de-identification services for oncology clinical trials data


♣ development of de-identification guidance
Given Enough Effort
Given Enough Effort, Time, Incentive, Money…
Claim: De-identification Has Failed

High Profile Re-identification

Ethnicity  
Visit date  
Diagnosis  
Procedure  
Medication  
Total charge

ZIP Code  
Birthdate  
Gender

Name  
Address  
Date registered  
Party affiliation  
Date last voted

Hospital Discharge Data  
Voter List

Sweeney, JLME 1997
What is De-identification?

According to EU (Data Protection Directive):
“principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable”

According to HIPAA (Privacy Rule):
“information that does not identify an individual and ... no reasonable basis ... information can be used to identify an individual”

- Safe Harbor
  - Removal of 18 types of identifiers
  - No actual knowledge residual information can identify individual

- Expert Determination
  - Apply statistical or scientific principles
  - Very small risk that anticipated recipient could identify individual
**HIPAA “Cookbook” Standards**

<table>
<thead>
<tr>
<th>Field</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Names</td>
<td>Related to patient (not provider)</td>
</tr>
<tr>
<td>Unique Numbers</td>
<td>Phone, SSN, MRN, ...</td>
</tr>
<tr>
<td>Internet</td>
<td>Email, URL, IP addresses, ..</td>
</tr>
<tr>
<td>Biometrics</td>
<td>Finger, voice, ...</td>
</tr>
<tr>
<td>Dates</td>
<td>Less specific than year</td>
</tr>
<tr>
<td></td>
<td>Ages &gt; 89</td>
</tr>
<tr>
<td>Geocodes</td>
<td>Town, County, Less specific than Zip-3 (assuming &gt; 20,000 people in zone)</td>
</tr>
<tr>
<td>“Catch all”</td>
<td>“Any other unique identifying number, characteristic, or code”</td>
</tr>
</tbody>
</table>

*** Must have no *actual knowledge* the remaining data can be used to identify
Practice What You Preach
Vanderbilt’s BioVU
(~2 million patients records → over 100 TB of data)

Patient Identifier

One-Way Hashed Identifier

Clinical Notes

Orders (CPOE)

Clinical Messaging

Labs (Test Results)

Scrubbed Clinical Notes

Scrubbed Orders (CPOE)

Scrubbed Clinical Messaging

Scrubbed Labs (Test Results)
# Vanderbilt De-identified EMR + DNA

## Timeline

<table>
<thead>
<tr>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Community / Patient</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communications materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community Advisory Board established</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>On-going input</strong></td>
<td><strong>Poster study</strong></td>
<td><strong>Pre-launch awareness generation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Methods / Feasibility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistics/process mapping</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample acceptance validation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De-Identification effectiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proof of Concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form implementation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Protocol development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRB review and modifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethics review and modifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legal review and modifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final IRB approval</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OHRP confirmation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Live Operations Phase I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample accrual begins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demonstration proj.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient research, live setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Redaction in Natural Language

Original PHI

Smith, 61 yo ... daughter, Lynn, to ... oncologist Dr. White ... 5/13/10 to consider ... SWOG protocol 1811, ... was randomized 5/10 ... to call Mr. Smith on ... PLAN: Dr. White and I ...
Scrubbing Process

- Convert records to standard format
- Remove uninformative terms (e.g., “cc:”, “sincerely”)
- Add **PROTECTED[begin] & **PROTECTED[end] tags to retain necessary information
- Random Offset of **DATE
- Addition of hashed pseudonym

Recall = 0.999
Does Machine Learning Work?
(Vanderbilt EMR – *No Dictionaries*)

<table>
<thead>
<tr>
<th></th>
<th>Discharge</th>
<th>Laboratory</th>
<th>Letter</th>
<th>Order</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>200</td>
<td>400</td>
<td>200</td>
<td>400</td>
<td>1200</td>
</tr>
<tr>
<td>Test</td>
<td>50</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>Precision</td>
<td>0.946</td>
<td>0.905</td>
<td>0.931</td>
<td>0.993</td>
<td>0.943</td>
</tr>
<tr>
<td>Recall</td>
<td>0.986</td>
<td>0.966</td>
<td>0.956</td>
<td>0.999</td>
<td>0.978</td>
</tr>
</tbody>
</table>

Precision: 0.91 – 0.99
Recall: 0.95 – 0.99

Aberdeen et al. IJMI, 2010
Negligible Impact on Medication Extraction

- Conditional Random Field (@ Cincinnati Children’s Hospital)
- ~3500 clinical notes over 22 note types

<table>
<thead>
<tr>
<th></th>
<th>Original Notes</th>
<th>Scrubbed Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>96.3</td>
<td>96.3 – 96.5</td>
</tr>
<tr>
<td>Recall</td>
<td>89.3</td>
<td>88.9 – 89.5</td>
</tr>
<tr>
<td>F-measure</td>
<td>92.6</td>
<td>92.5 – 92.7</td>
</tr>
</tbody>
</table>

Deleger et al. JAMIA., 2013
Redaction Has its Limits

**Original PHI**

Smith, 61 yo ...
daughter, Lynn, to ...
oncologist Dr. White ...
5/13/10 to consider ...
SWOG protocol 1811, ...
was randomized 5/10 ...
to call Mr. Smith on ...
PLAN: Dr. White and I ...

**Redacted PHI & Leaked PHI**

**pt_name<A>,** age<60s> yo ...
daughter, Lynn, to ...
oncologist Dr. **MD_name<C>** ...
**date<5/28/10>** to consider ...
SWOG protocol **other_id,** ...
was randomized 5/10 ...
to call Mr. **pt_name<A>** on ...
PLAN: Dr. White and I ...
Redaction Has its Limits...
but it Isn’t the Only Option

Original PHI

Smith, 61 yo ...
daughter, Lynn, to ...
oncologist Dr. White ...
5/13/10 to consider ...
SWOG protocol 1811, ...
was randomized 5/10 ...
to call Mr. Smith on ...
PLAN: Dr White and I ...

**Redacted PHI & Leaked PHI

**pt_name<A>, **age<60s> yo ...
daughter, Lynn, to ...
oncologist Dr. **MD_name<C> ...
**date<5/28/10> to consider ...
SWOG protocol **other_id, ...
was randomized 5/10 ...
to call Mr. **pt_name<A> on ...
PLAN: Dr White and I ...

Surrogate PHI & Hidden PHI

Jones, a 64 yo ...
daughter, Lynn, for ...
oncologist Dr. Howe ...
5/28/10 to consider ...
SWOG protocol 1798, ...
was randomized 5/10 ...
to call Mr. Jones on ...
PLAN: Dr White and I ...

Idea: Inject surrogated information to hide the leaks!

Carrell et al., JAMIA 2013
Hiding in Plain Sight [HIPS]

• Added a surrogation component to MIST*
• ~130 oncology notes from Group Health Coop of Puget Sound

*MIST forced into a dumbed-down state for assessment

Can effectively raise de-identification performance from to >0.99

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HIPAA</td>
<td>PHI Residual Expected</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pat. name</td>
<td>186</td>
<td>147</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>130</td>
<td>91</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone #</td>
<td>23</td>
<td>16</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td>67</td>
<td>35</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>180</td>
<td>117</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRN</td>
<td>39</td>
<td>33</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acct. #</td>
<td>27</td>
<td>23</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other ID #s</td>
<td>102</td>
<td>67</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>323</td>
<td>47</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTHER</td>
<td>PHI Residual Expected</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prac name</td>
<td>82</td>
<td>9</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Org. name</td>
<td>27</td>
<td>20</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>109</td>
<td>29</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Carrell et al., JAMIA 2013
Even HIPS has Limits

Original PHI

Smith, 61 y/o, daughter, Lynn, to see oncologist Dr. White on 5/13/10
SWOG protocol 1811, was randomized 5/10
to call Mr. Smith on ...
PLAN: Dr. White and I ...

**Redacted PHI & Leaked PHI

**pt_name<A>, age<60s>, Smith, 64 y/o, daughter, Lynn, to see oncologist Dr. Howe on 5/28/10
SWOG protocol other_id, was randomized 5/10
to call Mr. **pt_name<A> on ...
PLAN: Dr. White and I ...

Surrogate PHI & Hidden PHI

Mr. Jones, 64 y/o, daughter, Lynn, to see oncologist Dr. Howe on 5/28/10
SWOG protocol 1798, was randomized 5/10
to call Mr. Jones on ...
PLAN: Dr. White and I ...

Unknown residual re-identification potential (e.g. “the Senator’s wife”)

Idea: Inject surrogated information to hide the leaks!

Policy:

Data Use Agreements

Carrell et al., JAMIA 2013
Certify via “generally accepted statistical and scientific principles & methods, that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by the anticipated recipient to identify the subject of the information.”
Towards a Risk-Based De-identification Model

Malin, Benitez, and Masys. JAMIA. 2011.
Xia, et al. ACM CODASPY. 2013
# Vandy ECG Case Study

<table>
<thead>
<tr>
<th>Who</th>
<th>State</th>
<th>State Population Size (2010 Census)</th>
<th>Cohort Size</th>
<th>Patients &gt;89 years old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vanderbilt</td>
<td>TN</td>
<td>~6 million</td>
<td>~3,000</td>
<td>12</td>
</tr>
</tbody>
</table>

## Policy Generalizations

<table>
<thead>
<tr>
<th>Policy</th>
<th>Gender</th>
<th>Race</th>
<th>Age</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safe Harbor</td>
<td>∅</td>
<td>∅</td>
<td>[90 - 120]</td>
<td>0.909</td>
</tr>
<tr>
<td>Alternative 1</td>
<td>[M or F]</td>
<td>∅</td>
<td>∅</td>
<td>0.476</td>
</tr>
<tr>
<td>Alternative 2</td>
<td>∅</td>
<td>[Asian or Other]</td>
<td>∅</td>
<td>0.857</td>
</tr>
<tr>
<td>Alternative 3</td>
<td>∅</td>
<td>∅</td>
<td>[52 - 53]</td>
<td>0.875</td>
</tr>
</tbody>
</table>

Evaluation in Multiple Populations

- Cohorts from the Electronic Medical Records and Genomics Consortia (http://www.gwas.net)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>$G_{Dem}$</td>
<td>GHC</td>
<td>WA</td>
<td>5,894,121</td>
<td>Dementia</td>
<td>3,616</td>
<td>1,483</td>
</tr>
<tr>
<td></td>
<td>$R_{Cat}$</td>
<td>Marshfield</td>
<td>WI</td>
<td>5,363,675</td>
<td>Cataracts</td>
<td>2,646</td>
<td>269</td>
</tr>
<tr>
<td></td>
<td>$Y_{PAD}$</td>
<td>Mayo</td>
<td>MN</td>
<td>4,919,479</td>
<td>Peripheral Arterial Disease</td>
<td>3,412</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>$N_{T2D}$</td>
<td>Northwestern</td>
<td>IL</td>
<td>1,2519,293</td>
<td>Type-II Diabetes</td>
<td>3,383</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>$V_{QRS}$</td>
<td>Vanderbilt</td>
<td>TN</td>
<td>5,689,283</td>
<td>QRS Duration</td>
<td>2,983</td>
<td>12</td>
</tr>
<tr>
<td>Quality Control</td>
<td>$N_{QRS}$</td>
<td>Northwestern</td>
<td>IL</td>
<td>1,2519,293</td>
<td>QRS Duration</td>
<td>149</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$V_{T2D}$</td>
<td>Vanderbilt</td>
<td>TN</td>
<td>5,689,283</td>
<td>Type-II Diabetes</td>
<td>2,015</td>
<td>18</td>
</tr>
</tbody>
</table>

Malin, Benitez, & Masys. JAMIA. 2011.
Risk Model: Uniques

Is the number of uniques expected to be greater than Safe Harbor?

<table>
<thead>
<tr>
<th>Disclosure Policy</th>
<th>Acceptable?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized Ethnicity (Black, White, Other)</td>
<td>Red</td>
</tr>
<tr>
<td>Age at 5 Year Bins</td>
<td></td>
</tr>
<tr>
<td>Generalized Ethnicity AND Age at 5 year bins</td>
<td>Green</td>
</tr>
<tr>
<td>Age at 10 Year Bins</td>
<td>Green</td>
</tr>
</tbody>
</table>

Red = more risk than Safe Harbor
Green = risk no worse than Safe Harbor

Malin, Benitez, & Masys. JAMIA. 2011.
Forthcoming Data from Sanofi

• Oncology clinical trial data for Project Data Sphere

• De-identification Decisions
  – Only field-structured data (no free text)
  – Suppression of contact information (e.g., phone #, medical record #)
  – Coarsen geographic area:
    • North America, South America, Western Europe, Eastern Europe, & Other
  – Age reported at year, but top-coded as 85+
  – Dates of trial-related events permitted, but
  – Death events limited to one-week interval

• Proof of Protection
  – Use population and dataset-specific distributions to show re-identification risk is no worse than Safe Harbor
  – Safe Harbor: 0.00029% of U.S. population estimated to be unique
  – Sanofi: ~0.000001%
Risk in a Multinational Setting

- Risk analysis initially performed using US population statistics
- Extrapolated analysis by simulating the diversity of various demographic distributions (e.g., age, race)
- Decision: no region less than 10M people
Prepping for Expert Determination

• Identifiability is proportional to
  
  Uniqueness (must be distinguishable)  x
  Replicability (must be reproducible)  x
  Availability (must be accessible)

• A drug dose may be unique, but may not be accessible to the public in any known resource

• “Adversaries” have incomplete knowledge
[Your Favorite Feature] Distinguishes You!!

- Demographics (Sweeney ‘97; Bacher ‘02; Golle ‘06; El Emam ‘08; Koot ‘10; Li ‘11)
- Diagnosis Codes (Loukides ‘10; Tamersoy ‘10, ‘12)
- Lab Tests (Atreya ‘13, Cimino ‘12)
- DNA (Lin ‘04; Malin ‘05; Homer ‘08; Wang ‘09; Gymrek ‘13)
- Health Survey Responses (Solomon ‘12)
- Hospital (Location) Visits (Malin ‘04; Golle ‘09; El Emam ‘11)
- Pedigree (Family) Structure (Malin ‘06)

- Movie Reviews (Narayanan ‘08)
- Social Network Structure (Backstrom ‘07; Narayanan ‘09; Yang ‘12)
- Search Queries (Barbaro ‘06)
- Internet Browsing (Malin ‘05; Eckersley ‘10; Banse ‘11; Herrmann ‘12, Olejnik ‘12)
- Smart Utility Meter Usage (Buchmann et al ‘12)
Diagnoses?

- ~50% of Vanderbilt patients with at least 1 diagnosis code are unique!


Loukides, Denny, and Malin. JAMIA. 2010.
Big Data ≠ End of Privacy
Simple Expert Model

$k$-Anonymity \(^{(\text{Sweeney, 2002})}\)

Ensure $k$ record for every set of identifiers
“Guaranteed” Privacy

- Privacy: No record links to $< k$ people using diagnoses
- Utility: Retain diagnoses codes for genome-phenome “validation”
- Cohort: 3000 Vanderbilt patients in a QRS study
- Results shown for $k = 5$

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Intelligent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma</td>
<td>✓</td>
</tr>
<tr>
<td>ADHD</td>
<td></td>
</tr>
<tr>
<td>Bipolar</td>
<td>✓</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td>✓</td>
</tr>
<tr>
<td>Coronary Disease</td>
<td>✓</td>
</tr>
<tr>
<td>Diabetes 1</td>
<td>✓</td>
</tr>
<tr>
<td>Diabetes 2</td>
<td>✓</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>✓</td>
</tr>
<tr>
<td>Pancreatic Cancer</td>
<td>✓</td>
</tr>
<tr>
<td>Platelet Related Phenotype</td>
<td></td>
</tr>
<tr>
<td>Preterm Birth</td>
<td>✓</td>
</tr>
<tr>
<td>Prostate Cancer</td>
<td>✓</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>✓</td>
</tr>
<tr>
<td>Renal Cancer</td>
<td>✓</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>✓</td>
</tr>
<tr>
<td>Sickle-Cell Disease</td>
<td>✓</td>
</tr>
</tbody>
</table>

Loukides, Gkoulalas-Divanis, & Malin. PNAS. 2010.
Phenome Wide Association Studies

(associated with longer QRS duration in normal hearts)

Ritchie et al., Circulation 2013
Big Data Can Mean Big Privacy

• Often use very strong adversary
• But almost perfect results can be achieved...
• ... in real world
• Validation of 192 SNP – phenotype associations

De-identification is NOT a Panacea

• There is *always* a risk of re-identification
• But risk exists in any security setting
• The challenges are
  – Determine an appropriate level of risk
  – Ensure accountability

• Combine with data use agreements

• Risk is proportional to anticipated recipient trustworthiness (public vs. vetted investigator)
De-identification Can Be Safe

• Reviewed all actual re-identification attempts and rates of success

• All attacks through 2010
  – 14 published re-identification attacks on any type of data
  – 11 were conducted by researchers as demo attacks
  – Only 2 datasets followed any standard
  – Only case with health data subject to Safe Harbor had a success likelihood of 0.00013

Challenges for De-identification

• 2014 recent report from NRC Committee on Revisions to the Common Rule for the Protection of Human Subjects in Research in the Behavioral and Social Sciences

• HIPAA calls for protection from identity disclosure... but does not address utility of the data

• No definitive standard for
  – Risk Assessment
  – De-identification Methodology (but the Office for Civil Rights issued HIPAA guidance in November 2012)

• Need for national clearinghouse of models, methods, and evaluations

• Protections should be proportional to harm, recipients, and generally the context

• Case studies are needed!
NRC Recommendations

• Data Protection Plans
  – Degree of identifiability
  – Computing environment where data is shared
  – Location & method of data storage
  – Controls to the data
  – Secure transmission of data
  – Methods of output (paper vs. electronic)
  – Mechanisms for audit and oversight

• Researchers should honor confidentiality agreements, but no further consent should be necessary for secondary use (including linkage to other resources, unless specified from the outset)
We Must be Reasonable & Practical

Models

How is identification achieved?
What are the Opportunities for Harm?

Measures

What is the likelihood of identification?
Assess the most risky record vs. the average risk?

Mitigation

Are legal, technical, or hybrid controls the most prudent?
What are the benefits vs. the risks?
Acknowledgements

- Edoardo Airoldi, Ph.D. (Harvard U.)
- Kathleen Benitez
- Mustafa Canim, Ph.D. (IBM)
- Ellen Wright Clayton, M.D., J.D.
- Josh Denny, M.D.
- Xiaofeng Ding, Ph.D. (U. So. Australia)
- Khaled El Emam, Ph.D. (U. Ottawa)
- Aris Gkoulalas-Divanis, Ph.D. (IBM)
- Jonathan Haines, Ph.D.
- Raymond Healtherly, Ph.D.
- Murat Kantarcioglu, Ph.D. (U. of Texas)
- Jiuyong Li, Ph.D. (U. So. Australia)
- Grigorios Loukides, Ph.D. (Cardiff U)

Funding

- NHGRI @ NIH
  - U01 HG006385 (eMERGE)
  - U01 HG006378 (VGER)
- NLM @ NIH
  - R01 LM009989
- Trustworthy Computing @ NSF
  - CCF-0424422 (TRUST)
Questions?

b.malin@vanderbilt.edu

Health Information Privacy Laboratory
http://www.hiplab.org/