Precision Medicine, Big Data and Geisinger’s MyCode Community Health Initiative

W. Andrew (Andy) Faucett, MS, LGC
wafaucett@geisinger.edu
Professor / Geisinger Health System
May 24, 2016
“Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle in each person”.

“A longitudinal ‘cohort’ of 1 million or more Americans who have volunteered to participate in research. Participants will be asked to give consent for extensive characterization of biologic specimens . . . – including whole genome sequencing . . . , all linked to their electronic health records”

A New Initiative on Precision Medicine
Francis S. Collins, MD, PhD, and Harold Varmus, MD
NEJM, Jan 30 2015

$215 million for NIH in President’s 2016 budget
Geisinger and Precision Medicine Initiative

• Forum and online survey of MyCode members
  • 44.7% eager
  • 47.7% might be interested

• Year one recruitment goal
  • 10,000
  • 40% x 125,000 = 49,000

• Recruitment goal year 2 to 5
  • 35,000 per year
Geisinger Cohort Value for PMI

- Snapshot of biobank data
  - 11.96 Median years of EHR data
  - 47 Median number of clinical encounters
  - 455 Median number of lab test values
  - 94 Median vital measurements

- Geisinger Health Plan (GHP) – *sweet spot*
  - 2/3 GHP coverage at some point

- Large families
  - For every result, expect 6-10 1st degree relatives
PMI Recruitment

• Consent, evaluation and samples to “count”
• 5 models in proposal
  • 3 person teams – PA/NP, consenter, & support staff (clinic space)
  • Geisinger clinicians – IM & FP, consenter & staff
  • Careworks, consenter & staff
  • 2 mobile van clinics, each 2 exam rooms, 3 person teams
  • Annual physical from EHR – 2 person team – consenter & staff
Geisinger PMI Proposal

UG3

- MyCode consent protocols
- New Geisinger markets
- 21st Century Cohort
- Healthy Trios
- Senior Siblings

UG3 planning and development

- Well Connected
- Health Advances

Engagement

MyCode Participant Engagement

Participant Data

- Environmental
- Psychosocial
- Participant Reported
- Behavioral
- mHealth
- Sensor

UH3 implementation and expansion
Geisinger Regional Demographics

- ~4 million people in 41-county service area
  - 31 rural, 7 urban and 3 in NJ
- Older and poorer relative to national averages
  - More high school graduates, fewer non-HS and college
- Most counties federally designated as poor, underserved
- <1% annual out-migration rate in most counties
- Large, multi-generation families
- Stable population + EHR + integrated health system = prime for genomics research
### Census racial characteristics

<table>
<thead>
<tr>
<th>Region</th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td>92.3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Northeast</td>
<td>84.5</td>
<td>8.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>West</td>
<td>92.5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Harrisburg</td>
<td>83</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Atlantic City</td>
<td>68</td>
<td>5.1</td>
<td>4.5</td>
<td>5.2</td>
<td>1</td>
</tr>
<tr>
<td>US</td>
<td>62</td>
<td>12</td>
<td>7.5</td>
<td>5.5</td>
<td>3</td>
</tr>
</tbody>
</table>
Clinical Data Warehouse

• Near real-time, system-wide data warehouse

• Searchable data format
  • EHR data (back to 2004)
    • Demographic and lifestyle data (e.g., age, height, weight, smoking)
    • Clinical measures (e.g., blood pressure)
    • Orders (prescriptions, imaging)
    • Clinical laboratory data
  • Financial data (e.g., billing, payment)
  • Operational data (e.g., events, scheduling)
  • Claims data (e.g., ambulatory, hospital, pharmacy)

• Data are cleansed, normalized and stored at most granular levels to facilitate data mining and analytics
Research Data Broker and Clinical Decision Intelligence System (CDIS) Data Warehouse

- **EHR**
  - Orders
  - Diagnoses
  - Lab values
  - Meds
  - Procedures
  - Etc.

- **CDIS**
  - Searchable
  - Exportable

- **TPO**
  - Claims
  - Finance
  - Ops

- **Investigator**

- **MyCode Biobank**
  - >100,000 consented participants
  - >200,000 samples
  - Blood, serum, DNA, tissue

- **Governing Board**

- **PHI Firewall**

- **Research data**
  - Linkable
What is Big Data?

• Big data is a blanket term for any collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. – Wikipedia

• Data sets that are too large and complex to manipulate or interrogate with standard methods or tools. – Oxford Dictionary

• Computers, data sets, typically consisting of billions or trillions of records, that are so vast and complex that they require new and powerful computational resources to process. – Dictionary.com
Where do we see Big Data?

To put the data explosion in context, consider this. Every minute of every day we create:

• More than 204 million email messages
• Over 2 million Google search queries
• 48 hours of new YouTube videos
• 684,000 bits of content shared on Facebook
• More than 100,000 tweets

Where do we see Big Data?

90 PB (pedabytes)  
May, 2013

300 PB (pedabytes)  
April, 2014

Exabytes of data
Big Data: Is it all about size?

“I think you’ll find that mine is bigger”

Depends on your frame of reference
Cost per Genome

Moore's Law

National Human Genome Research Institute

gene.gov/sequencingcosts

Big Data for your health

Pharmacogenomics

How is Risk Calculated?

Simple Disease
Single Disease Gene

Risk is easy to calculate for rare disorders caused by a single gene.

Complex Disease
Many Disease Genes

But for complex diseases that are influenced by multiple genes, risk is much more difficult to calculate.

Risk must be estimated based on observation of data collected from large families affected by these diseases.

Disease Risk Prediction

Disease Prevention

CHRONIC DISEASE
AND WHAT YOU CAN DO TO PREVENT THEM
Pharmacogenomics
When medicine gets personal

The right medicine
The right patient
The right dose
The right time
Patient population

Standard approach

Treatment

Treatment A (effective in 20% of target population; 80% is waste)

Tailored approach

Treatment A
Treatment B
Treatment C
Treatment D

PREDICT helps match patient with proper drug

BY: KATHY WHITNEY

10/28/2010 - Had Scyble Van Cleve, a spry 83-year-old from Brentwood, had her heart procedure done a month ago instead of one week ago, she would have been prescribed the standard dose of clopidogrel, a blood thinner used to prevent blood clots from forming around her coronary stents.

Her doctors may not have known that, based on her genes, she needed a different blood-thinning regimen to safeguard her from possible fatal complications.

Thanks to a novel program implemented at Vanderbilt called PREDICT, Van Cleve
• 35 CPIC guidelines for gene-drug pairs
• Over 100 genetic variants
Genomics and Electronic Health Records:

A Dynamic Duo for Precision Medicine

EHR: Electronic Health Records
Data Flow Between Coordinating Center (CC) and Participant Sites

- Simpler queries matching core data handled by CC (fast, scalable)
- Over time, core data grows
- Select subset validation/data cleaning
- Core Data

Direct Volunteers

HPOs

Potential limitations

• Context is important
  • Tissue, time
  • Broad, unbiased sample collection probably not best strategy
• Data are incomplete
• Biological knowledge is incomplete
• Network models may be incomplete/incorrect

Do we need to know everything and have every data point to make inferences and learn new biology?
Geisinger Research Interests

Pharmacogenomics

Disease Risk Prediction

Disease Prevention
History:

- MyCode® began in 2006
- High consent rate – 85%+
- Rate limiting factor was number of consenters
MyCode Expansion By Year

Number of consented patients

- 0
- 500-999
- 1-49
- 1000-4999
- 50-99
- 5000+
- 100-499

animation and map created by Ryan Colonie and Craig Wood
Geisinger Research Strategic Vision- 2010

• Vision: Personalized Health Care (aka Precision)
• Emphasis on genomics
• Innovative clinical provider system and payor
• Both of whom see “value” in potential to change course of disease/improve health outcomes/reduce costs
• 10 year Vision - translational research is part of, not separate from, Clinical Medicine
• 2014 update added focus on participant engagement and Learning Healthcare System
“Geisinger is as close to Iceland as you’ll find in the United States” - Glenn Steele, 2003

Features for community engagement and longitudinal, genomic medicine research / implementation

- Large, stable population – >2.5 M people (>700K active patients) with many 3+ generation families (4 M current)
- Strong and trusting relationship between patients and Geisinger
- Integrated healthcare delivery system
- Longstanding EHR and comprehensive clinical data (EPIC – 1996)
- Innovative and supportive leadership
Why Change MyCode®

- National discussion about the duty to return research results
  - Engage research participants
  - Geisinger is a health system – patient benefit a priority
- Cost of Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) dropping rapidly
  - Results might actually be available for many participants
- Set the national biobank standard and become attractive to research partners
  - Bring benefits of WGS and WES to Geisinger Family
Focus Groups – community engagement

• 6 focus groups held
• Bariatric surgery group and Primary Care Clinics in Bloomsburg and Kulpmont
• 93 participants – 57% F, 43% M
  • 42% Age 61 – 70
  • 23% Age 51 – 60
  • 22% Age 71 & >
  • 8% Age 41 – 50
  • 5% Age 18 – 40
  • (Ages match biobank participants)
• 49.5% GHS patients for 20+ years
• 85% receive majority of care at GHS
Focus Group Discussion

• Pharmacogenomics (genetic variation that impacts medication dosage and choice)
• Recessive Carrier (risk for serious disease, information for children and grandchildren)
• Increased Risk for Preventable or Treatable condition
• Increased Risk for NON Preventable or Treatable condition
• Genetic changes that we currently do not understand
Results of Focus Groups

• **Wanted ALL results**

• Results should be returned to healthcare provider and participant (most preferred same time)

• Geisinger should develop educational materials and expert support system

• Results should be put in EHR
MyCode® Consent

• Critical Features
  • Re-contact
  • Longitudinal EHR data
  • Longitudinal samples
  • Return of results
  • Placement of results in EHR
  • Online consent
MyCode® Community Health Initiative: An Ethics Advisory Committee

- **Kevin T. FitzGerald, SJ, PhD, chair**
  Jesuit priest, PhDs in molecular genetics and bioethics, Georgetown faculty in oncology and bioethics, former member of the Secretary’s Advisory Committee on Genetics, Health & Society (SACGHS)

- **Kyle Brothers, MD, PhD**
  Pediatrician with research interests in childhood obesity, genomic research and clinical use of genomic data; trained at Vanderbilt, now on faculty at University of Louisville, lead publication of eMERGE paper on consent for children

- **Joan Scott, MS, CGC**
  Certified genetic counselor, former director of Genomics and Public Policy at Johns Hopkins and National Coalition for Health Professional Education in Genetics (NCHPEG)

- **Sylvia Mann Au, MS, CGC**
  Certified genetic counselor, State Genetics Coordinator for Hawaii Department of Health, and member of SACGHS

- And 4 Geisinger patients!
Aiming to Push Genomics Forward in New Study

By ANDREW POLLACK
**Objective:** Improve our ability to predict and prevent disease, optimize treatments based on genetic information for each individual

- Partnership between Geisinger Health System and Regeneron
- 5 – 10 Year Genomic Sequencing Study
- 250,000 Study Participants
- *(DiscovEHR)* – Data Integrating Sequencing, Communities, and Omics adding Value to the Electronic Health Record
Geisinger & Regeneron – scientific partners

**Geisinger Resources**
- Unique community partnership, trust
- Stable population (three generation families)
- High recruit rate for MyCode
- Strong research expertise in improving healthcare
- Research financial resources for return-of-results

**Regeneron Resources**
- Strong scientific team
- State-of-the-art DNA sequencing facility
- Strong financial resources for sequencing costs
- Focus on new drug development
Initiated 7/1/2014
9/1/2014 Fully-automated exome sample preparation
Averaged 1,000 exomes per week through the last third of 2014
Averaged >1200 exomes per week through 1st quarter of 2015
Scaled to >1,650 per week (7,000 per month)

Community Health Initiative

- Total consented: 104,659
- 100,000 (3/18)
- MyGeisinger (online)
  - 2,800
  - 343 reconsents (return-of-results)
- Almost 40,000 consented in 2015

As of April 27, 2016
MyCode Participation in PA

Map of MyCode Community Health Initiative Participation (as of March 21, 2016)
<table>
<thead>
<tr>
<th></th>
<th>Geisinger MyCode®</th>
<th>UK</th>
<th>Iceland</th>
<th>Estonia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opt in consent</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>EHR linked</td>
<td>Yes</td>
<td>Yes</td>
<td>Limited</td>
<td>Partial</td>
</tr>
<tr>
<td>Longitudinal data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Re-contact</td>
<td>Yes (broad)</td>
<td>Yes</td>
<td>No</td>
<td>Yes (targeted)</td>
</tr>
<tr>
<td>Individualized Return of Results</td>
<td>Yes (broad)</td>
<td>No</td>
<td>No</td>
<td>Allowed *</td>
</tr>
<tr>
<td>Current Enrollment</td>
<td>80,000</td>
<td>500,000</td>
<td>140,000</td>
<td>51,535</td>
</tr>
<tr>
<td>Planned Enrollment</td>
<td>250,000</td>
<td>500,000</td>
<td>140,000</td>
<td>1,050,000</td>
</tr>
</tbody>
</table>
Employee Campaign

Two Goals

• All employees know basics of campaign and where to refer
• Employees have opportunity to participate
What kind of results will be returned?

- 27 highly penetrant, medically actionable conditions
  - Pathogenic/likely pathogenic variants
  - 76 genes, inclusive of the ACMG 56
  - Cancer predisposition (e.g., BRCA1/2)
  - Cardiovascular disease (e.g., FH)
  - Malignant hyperthermia
  - Hereditary hemorrhagic telangietasia
  - Ornithine transcarbamylase (OTC) deficiency

- NOT carrier status, VUS, pharmacogenomics or GWAS
  - List will increase in future
Top Three Most Prevalent Conditions in the G76 Expected to Compose Half of those Returned

<table>
<thead>
<tr>
<th>GENOMIC CONDITION</th>
<th>POPULATION PREVALENCE</th>
<th>CLINICAL RISK</th>
<th>DISEASE-ALTERING INTERVENTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familial Hypercholesterolemia</td>
<td>1 in 175</td>
<td>Early-onset coronary artery disease &amp; stroke</td>
<td>Targeted screening &amp; aggressive medical management</td>
</tr>
<tr>
<td>Hereditary Breast and Ovarian Cancer Syndrome</td>
<td>1 in 400</td>
<td>Early-onset breast, ovarian &amp; prostate cancers</td>
<td>Targeted screening &amp; prophylactic surgical intervention</td>
</tr>
<tr>
<td>Lynch Syndrome</td>
<td>1 in 440</td>
<td>Early-onset colon &amp; uterine cancers</td>
<td>Targeted screening &amp; management of pre-cancerous changes</td>
</tr>
<tr>
<td>TOTAL</td>
<td>&gt; 1 in 100</td>
<td>Multiple cancers &amp; cardiovascular diseases</td>
<td>Life-saving screening &amp; intervention before development of disease</td>
</tr>
</tbody>
</table>

Associated genes to be screened: BRCA1, BRCA2, MLH1, MSH2, MSH6, PMS2, LDLR, APOB, PCSK9
Primary care provider (PCP) management

- CG sends EHR message/letters to patient
  - Result sent to PCP
  - PCP encounter scheduled
    - 5 days*
  - CG calls patient
    - Contact
      - 1. Disclosure phone script
      - 2. Family hx ID assigned
    - Result & support materials mailed to patient
    - Patient may follow up with PCP, CG or both

Clinical Genomics (CG) management

- CG assists w/referral from PCP
- Patient follows up with CG

Targeted follow-up w/PCP & condition-specific specialists

*Business; Follow-up includes genetic counseling & medical evaluation
Geisinger GenomeFIRST™ Case

Case Summary

• **May 2013** – Angelina Jolie NYT OP-ED “My Medical Choice”.

• **Summer 2013** – Prompted by story, 33 year old woman seeks advice at Geisinger for testing related to Breast Cancer in Family. Meets with Genetic Counselor, consultation reveals only a single case of Breast Cancer in family and she does not meet criteria to recommend *BRCA* gene test.

• **Fall 2015** - patient receives positive result for *BRCA1* through MyCode (GenomeFIRST).

Key Points

Current recommendations for *BRCA1/2* genetic testing are based on strong family history and miss an estimated 50% of cases where family history is not sufficient to prompt testing. (Gabai-Kapara E, et al. 2014)

GenomeFIRST allows for detection of both family history positive and family history negative cases of individuals with pathogenic variants in *BRCA1/2*. 
Type II, III, IV Diabetes?

Different treatment plans?

Different disease risk prediction?
Recalibration of frequency of “returnable” sequence results

1,415 WES

641 potentially pathogenic variants in G76

- Reported previously to be pathogenic
- Predicted loss-of-function

“clinical” curation

55 P, LP variants
63 individuals*

586 B, LB, VUS

*Geisinger now predicted to be 4.4%
48% of participants had ≥ 1 first or second degree relatives in the cohort
Genetically-inferred relationships among 50,730 Adults in MyCode

<table>
<thead>
<tr>
<th>Relation</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ Twins</td>
<td>20 pairs</td>
</tr>
<tr>
<td>Full Siblings</td>
<td>4957 pairs</td>
</tr>
<tr>
<td>Parent Child</td>
<td>6654 pairs</td>
</tr>
<tr>
<td>Second Degree Relatives</td>
<td>14,548 pairs</td>
</tr>
<tr>
<td>Third Degree Relatives</td>
<td>103,314 pairs</td>
</tr>
</tbody>
</table>

First degree relatives:
- grandparents, grandchildren, aunts, uncles, nephews, nieces or half-siblings

- first-cousins, great-grandparents or great grandchildren
Expansion

- Western Region
- Northeast
- Central
  - Bloomsburg Hospital
- AtlantiCare
  - July 2016
  - IRB
  - Staff
- Holy Spirit
  - September 2016
Learn More – geisinger.org/genomics

The Genomic Medicine Institute at Geisinger

Clinical Services
Genomic Sequencing
Family History
Simons VIP

ICCG
MyCode
eMerge
Mechanisms/Consequences

Choose from one of the links above to learn more.

MyCode™ Community Health Initiative

Breaking: Regeneron and Geisinger Health System announce major human genetics research collaboration.

Our mission: Partnering with patients, healthcare providers and researchers worldwide to enhance the quality of life through research, education and clinical care innovation in genomic medicine.

Our vision: To be the leader in the integration of genomics into integrated health care systems.

The Genomic Medicine Institute conducts innovative research in genetics, genomics and family history to enhance the quality of life and improve healthcare value for our patients.

Geisinger Health System, including the Geisinger Health Plan, upholds the Genetic Information Non-Discrimination Act (GINA) of 2008 in employment and in insurability and ensures that genetic information is not used in a discriminatory way.

Genomic Sequencing Study Overview

Geisinger Health System is partnering with Regeneron, a leading biopharmaceutical company, for a new, long-term, large research study.

Significance of Genomic Research

Genetic research holds great promise to increase understanding of the causes of diseases, disorders, and medical conditions - including conditions that today have limited or no treatments available. By comparing genetic information against medical histories, Geisinger and Regeneron hope to eventually develop new means of diagnosing, preventing, and/or treating medical conditions - before they cause significant harm. Some participants may also receive information that could be useful in their own medical care.

This study is aligned with Geisinger’s focus on innovative research to help patients in our communities - as well as improve healthcare for patients across the nation and worldwide.

Geisinger’s Goal and Role

Geisinger’s goal is to offer research opportunities and benefits to all members of the Geisinger community. To date the participation rate for MyCode has been very high and we anticipate continued interest, making it possible to reach our initial goal of 100,000 participants. We encourage all members of the Geisinger community to participate.

Geisinger will collect blood samples from consenting participants, along with relevant medical information. Regeneron will perform genomic analysis on the samples - in the hope of identifying new information on genetic variants that may be associated with specific diseases and health conditions.

All samples and records will be confidential - with personal information removed and all data scrupulously maintained through Geisinger’s secure MyCode Community Health Initiative repository.

Geisinger Health System, including the Geisinger Health Plan, upholds the Genetic Information Non-Discrimination Act (GINA) of 2008 in employment and in insurability and ensures that genetic information is not used in a discriminatory way.
Acknowledgements

• Thank you to the thousands of patients, volunteers, collaborators, investigators, and scientists who make this work possible.

RGC Leadership Team
- Aris Baras, M.D.
  Executive Director and Co-Head, RGC
- Rick Dewey, M.D.
  Director, Translational Genetics
- Omri Gottesman, M.D.
  Director and Head of Clinical Informatics
- John Overton, Ph.D.
  Director and Head of Sequencing and Lab Operations
- Jeffrey Reid, Ph.D.
  Senior Director and Head of Genome Informatics
- Alan Shuldiner, M.D.
  Vice President and Co-Head, RGC

RGC Scientific Advisory Board
- Richard Lifton, M.D., Ph.D. (Chair)
- Goncalo Abecasis, Ph.D.
- Wendy Chung, M.D., Ph.D.
- Peter Donnelly, Ph.D.
- Tim Hunkapiller, Ph.D.
- Sekar Kathiresan, M.D.
- James R. Lupski, M.D., Ph.D., D.Sc.(hon)
- Elaine Mardis, Ph.D.

Geisinger Leadership
- David Carey, Ph.D.
- Dan Davis, Ph.D.
- W. Andrew Faucett, M.S.
- Les Kirchner, Ph.D.
- David Ledbetter, Ph.D.
- Michael Murray, M.D.
- Marylyn Ritchie, Ph.D.

RGC Founders and Steering Committee
- George Yancopoulos, M.D., Ph.D., Aris Baras, M.D., Aris Economides, Ph.D., Scott Mellis, M.D., Ph.D., Andrew Murphy, Ph.D., Robert Phillips, Ph.D., Neil Stahl, Ph.D.
Geisinger Team

- Steering Committee
  - David Lederbetter*
  - David Carey*
  - Andy Faucett*
  - Les Kirchner
  - Mike Murray
  - Dan Davis
  - Marylyn Ritchie
  - Community Engagement
    - Carroll Flansburg
  - Biomedical/Translational Informatics
    - Sarah Pendergrass
  - John Wallace
  - Shefali Verma
  - Anurag Verma
  - Anna Okula

- Consenting
  - Sam Fetterolf
  - 18 consenters
  - Return of Results
    - Adam Buchanan
    - Monica Giovanni
    - Marci Barr
    - Steve Martin
    - Phenotype Core
      - Joe Leader
      - Brandon Geise
      - Lance Adams
      - Dustin Hartzel
      - Dan Lavage
      - Neil Manus
      - John Snyder

- *Project Principal Investigators
wafaucett@geisinger.edu
570-452-0043