May 20th, 9:00 AM

**Keynote Address: The Future of Cardiovascular Epidemiology: Current Trends?**

Vasan S. Ramachandran  
*Boston University School of Medicine*

---

**Repository Citation**  

---

**Creative Commons License**  
This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License. This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in UMass Center for Clinical and Translational Science Research Retreat by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
Future of Cardiovascular Epidemiology

Vasan S. Ramachandran MD

cHealth

sHealth

mHealth

eHealth

gHealth
Future of Cardiovascular Epidemiology

- **Background**
- **Role of**
  - cHealth (community)
  - sHealth (social)
  - mHealth (mobile)
  - eHealth (electronic)
  - gHealth (genomic)
- **A synthesis**
Time for a Creative Transformation of Epidemiology in the United States

Michael S. Lauer, MD

What has epidemiology done for medical science lately?

Answer: Much but not enough!

Dariush Mozaffarian et al. Circulation. 2016;133:e38-e360

Copyright © American Heart Association, Inc. All rights reserved.
1 in 3 adults


- Diabetes
- Obese
- Current smoking
- No leisure-time physical activity
- High blood pressure
- High cholesterol
- Low fruit consumption
- Low vegetable consumption

Percent
Eight Americas: Investigating Mortality Disparities across Races, Counties, and Race-Counties in the United States

Christopher J. L. Murray, Sandeep C. Kulkarni, Catherine Michaud, Niels Tomijima, Maria T. Bulzacchelli, Terrell J. Landiorio, Majid Ezzati

PLOS Medicine September 2006 | Volume 3 | Issue 9 | e260

**Gap in life expectancy in US of up to 14 years**

**Figure 3.** Life Expectancy at Birth in the Eight Americas (1982–2001)

Dariush Mozaffarian et al. Circulation. 2016;133:e38-e360

Copyright © American Heart Association, Inc. All rights reserved.
AHA Policy Statement

Forecasting the Future of Cardiovascular Disease in the United States

A Policy Statement From the American Heart Association

Paul A. Heidenreich, MD, MS, FAHA, Chair; Justin G. Trogdon, PhD; Olga A. Khavjou, MA;
Javed Butler, MD, MPH, FAHA; Kathleen Dracup, RN, DNSc;
Michael D. Ezekowitz, MBChB, DPhil, FRCP, FAHA; Eric Andrew Finkelstein, PhD, MHA;
Yuling Hong, MD, PhD, FAHA*; S. Claiborne Johnston, MD, PhD, FAHA; Amil Khera, MD, MSc;
Donald M. Lloyd-Jones, MD, MSc, FAHA; Sue A. Nelson, MPA;
Graham Nichol, MD, MPH, FRCP(C), FAHA; Diane Orenstein, PhD*;
Peter W.F. Wilson, MD, FAHA; Y. Joseph Woo, MD, FAHA; on behalf of the American Heart Association

_Circulation._ 2011;123:933-944

Table 1. Projections of Crude CVD Prevalence (%), 2010–2030 in the United States

<table>
<thead>
<tr>
<th>Year</th>
<th>All CVD*</th>
<th>Hypertension</th>
<th>CHD</th>
<th>HF</th>
<th>Stroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>36.9</td>
<td>33.9</td>
<td>8.0</td>
<td>2.8</td>
<td>3.2</td>
</tr>
<tr>
<td>2015</td>
<td>37.8</td>
<td>34.8</td>
<td>8.3</td>
<td>3.0</td>
<td>3.4</td>
</tr>
<tr>
<td>2020</td>
<td>38.7</td>
<td>35.7</td>
<td>8.6</td>
<td>3.1</td>
<td>3.6</td>
</tr>
<tr>
<td>2025</td>
<td>39.7</td>
<td>36.5</td>
<td>8.9</td>
<td>3.3</td>
<td>3.8</td>
</tr>
<tr>
<td>2030</td>
<td>40.5</td>
<td>37.3</td>
<td>9.3</td>
<td>3.5</td>
<td>4.0</td>
</tr>
<tr>
<td>% Change</td>
<td>9.9</td>
<td>9.9</td>
<td>16.6</td>
<td>25.0</td>
<td>24.9</td>
</tr>
</tbody>
</table>
Future of Cardiovascular Epidemiology

• Background

• Role of
  – cHealth (community)
  – sHealth (social)
  – mHealth (mobile)
  – eHealth (electronic)
  – gHealth (genomic)

• A synthesis
Aetiology confronts two distinct issues: the determinants of individual cases, and the determinants of incidence rate. If exposure to a necessary agent is homogeneous within a population, then case/control and cohort methods will fail to detect it: they will only identify markers of susceptibility. The corresponding strategies in control are the ‘high-risk’ approach, which seeks to protect susceptible individuals, and the population approach, which seeks to control the causes of incidence. The two approaches are not usually in competition, but the prior concern should always be to discover and control the causes of incidence.
At Risk            X              Rate of Disease    =     Community Burden

Overweight
Obese
Normal Weight


BMI  
≥ 35  4 %        X 32 % = 129 280 cases  12 %
30 à 34,9  13 %  X 21 % = 274 700 cases  26 %
Obese
25 à 29,9  41 %  X 10 % = 418 500 cases  40 %
Overweight
23 à 24,9  22 %  X 7 % = 157 800 cases  15 %
Normal Weight
< 23  20 %  X 3 % = 61 400 cases  6 %

A substantial proportion of cases arise from the middle of the distribution!

American Heart Association Guide for Improving Cardiovascular Health at the Community Level, 2013 Update
A Scientific Statement for Public Health Practitioners, Healthcare Providers, and Health Policy Makers

Thomas A. Pearson, MD, PhD, FAHA, Co-Chair; Latha P. Palaniappan, MD, MS, FAHA, Co-Chair;
Nancy T. Artinian, PhD, RN, FAHA; Mercedes R. Carnethon, PhD, FAHA;
Michael H. Criqui, MD, MPH, FAHA; Stephen R. Daniels, MD, PhD, FAHA;
Gregg C. Fonarow, MD, PhD, FAHA; Stephen P. Fortmann, MD; Barry A. Franklin, PhD, FAHA;
James M. Galloway, MD, FAHA; David C. Goff, Jr., MD, PhD, FAHA;
Gregory W. Heath, DHSc, MPH, FAHA; Ariel T. Holland Frank; Penny M. Kris-Etherton, PhD, RD;
Darwin R. Labarthe, MD, MPH, PhD, FAHA; Joanne M. Murabito, MD, ScM;
Ralph L. Sacco, MD, MS, FAHA; Comilla Sasson, MD, MS; Melanie B. Turner, MPH;

(Circulation. 2013;127:1730-1753.)
cHealth

Chronic diseases

High blood pressure, overweight/obesity, high glucose levels in the blood and high cholesterol levels

Smoking, unhealthy diet, physical inactivity, consumption of alcohol

Social health determinants and globalization/urbanization/aging of the population
Neighborhood Resources for Physical Activity and Healthy Foods and Incidence of Type 2 Diabetes Mellitus

The Multi-Ethnic Study of Atherosclerosis

Amy H. Auchincloss, PhD, MPH; Ana V. Diez Roux, MD, PhD; Mahasin S. Mujahid, PhD, MS; Mingwu Shen, MS; Alain G. Bertoni, MD, MPH; Mercedes R. Carnethon, PhD
Future of Cardiovascular Epidemiology

• Background

• Role of
  – cHealth (community)
  – sHealth (social)
  – mHealth (mobile)
  – eHealth (electronic)
  – gHealth (genomic)

• A synthesis
sHealth

Public Health Classics

Economic and social determinants of disease

Michael Marmot

Chetty R. JAMA 2016 April.

Mean household income in thousands, $^a$

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24</td>
<td>26</td>
<td>45</td>
<td>50</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>112</td>
<td>119</td>
<td>71</td>
<td>77</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>1.9 million</td>
<td>2.0 million</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Women by household income percentile
Bottom 1%: 78.8 (95% CI, 78.7-78.9)
Top 1%: 88.9 (95% CI, 88.7-89.1)

Men by household income percentile
Bottom 1%: 72.7 (95% CI, 72.6-72.9)
Top 1%: 87.3 (95% CI, 87.2-87.5)
Fat Friends Forever!

Ouchy, your Killin' Me?

Yeah Yeah Yeah

Eating Habits Are Contagious
How the people around us influence what we eat

Our friends influence the healthiness* of what we choose to eat by 34.5%
Future of Cardiovascular Epidemiology

• Background

• Role of
  – cHealth (community)
  – sHealth (social)
  – mHealth (mobile)
  – eHealth (elect)
  – gHealth (geno)

• A synthesis
Big data will become valuable to healthcare in what’s known as the internet of things (IoT). SAS describes the IoT as:

"a growing network of everyday objects from industrial machines to consumer goods that can share information and complete tasks while you are busy with other activities, like work, sleep, or exercise."
THE QUANTIFIED SELF:
Fundamental Disruption in Big Data Science and Biological Discovery
The FOUR V’s of Big Data

Volume
6 BILLION PEOPLE have cell phones

2020

6 BILLION PEOPLE have cell phones

World Population: 7 BILLION

40 ZETTATBYTES (43 TRILLION GIGABYTES) of data will be created by 2020, an increase of 300 times from 2005

2005

It’s estimated that 2.5 QUINTILLION BYTES (2.3 TRILLION GIGABYTES) of data are created each day

Volume SCALE OF DATA

100 TERABYTES (100,000 GIGABYTES) of data stored

Most companies in the U.S. have at least

Velocity

ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS -- almost 2.5 connections per person on earth

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION during each trading session

Modern cars have close to

100 SENSORS that monitor items such as fuel level and tire pressure

Velocity

Analysis of Streaming Data

Veracity

UNCERTAINTY OF DATA

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES (151 BILLION GIGABYTES)

By 2014, it’s anticipated there will be

420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month

30 BILLION PIECES OF CONTENT are shared on Facebook every month

400 MILLION TWEETS are sent per day by about 200 million monthly active users

Uncertainty of Data

Poors data quality costs the US economy around

$3.1 TRILLION A YEAR

1 in 3 business leaders don’t trust the information they use to make decisions

27% of respondents in one survey were unsure of how much of their data was inaccurate

Variety

DIFFERENT FORMS OF DATA

As a leader in the sector, IBM data scientists break big data into four dimensions: Volume, Velocity, Variety and Veracity

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015

4.4 MILLION IT JOBS will be created globally to support big data, with 1.9 million in the United States

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

https://www-01.ibm.com/software/data/bigdata/
mHealth: personally generated health data (PGHD)

Biosensing wearables allow continuous physiological monitoring in a wide range of form factors.

**Biosensors** are devices that convert a biological recognition element into a signal output.

**Biosensors** e.g. AliveCor, Scanadu

**Wearables** are on- or in-body accessories that enhance the user experience.

**Wearables** e.g. Google Glass, Oculus Rift

**Biosensing Wearables**
mHealth/quantitative sensor data

- wrist-based accelerometers in the Centers for Disease Control and Prevention National Health and Nutrition Examination Survey (NHANES) and the UK Biobank
- Health eHeart Study (a PCORnet Patient Powered Research Network)
- Apple’s ResearchKit, MyHeart Counts
- Extensive “physiome” data through wearable sensors are planned for a Baseline Study coordinated by Stanford, Duke University, and Google Inc
- mobile health data also planned for the NIH's Precision Medicine Initiative cohort
mHealth Advantages/Opportunities

- new knowledge about living with and managing health and illness.
- Increase compliance with meds
- ‘hovering’ to promote healthy behavior
- Use predictive analytics and behavioral economics
mHealth: Pitfalls & Challenges

- Few measurements from wearable sensors have been validated relative to existing metrics
- Continuous ambulatory data that do not directly match the tests done in the clinic
- Data quality can be dependent on individual participants and their level of engagement
- Accepting trade-offs in precision for more frequent, scalable measures
- Selection bias from the participants who “opt in” and who have sufficient technological knowledge and access
- Privacy and security of the data are critical
mHealth: Pitfalls and Challenges

• Technology necessary but not sufficient to induce health choice
• Adherence to use of mhealth technology unclear
• Must be integrated into clinical practice
• Applicability of approaches across diverse populations unknown
• Reach people when they are not patients
Future of Cardiovascular Epidemiology

• Background

• Role of
  – cHealth (community)
  – sHealth (social)
  – mHealth (mobile)
  – eHealth (electronic)
  – gHealth (genomic)

• A synthesis
How Big Data will change science

Here's how medical research traditionally works:

1. Come up with a question or hypothesis.
2. Design an experiment to test it. Wait for new data to come in.
3. Form your conclusion.

Big Data: EMR

• Enactment of the Patient Protection and Affordable Care Act of 2010 → hospitals and clinics received a mandate for electronic medical records (EMRs).
• Digitization of patients’ past histories & complaints, treatments, and outcomes → clinical research
• Lack of standardized data elements and definitions limits interoperability
• National standards have been promulgated, and EMRs are slowly mapping to these standards.
Big Data: EMR

- Infrastructure projects such as the National Institutes of Health (NIH) Collaboratory and the National Patient-Centered Clinical Research Network (PCORnet) facilitated linking of EMR data across multiple large health systems
- Large-scale post-market surveillance studies
- Recruit patients and collect information in practical clinical trials
- Incorporate quality improvement systems into the flow of clinical care.
PCORnet: clinical research and patient engagement on a large scale.

CDRNs indicates Clinical Data Research Networks; PCORI, Patient-Centered Outcomes Research Institute; PCORnet, National Patient-Centered Clinical Research Network; PPRNs, Patient Powered Research Networks.
Growth of Big Data in Health Care

- 1970s: Database machine
- 1980s: Megabyte to Gigabyte
- 1990s: Gigabyte to Terabyte
- 2011: Terabyte to Petabyte
- Petabyte to Exabyte

- 1990s: Google file system and MapReduce
- 2011: IDC report

- Kilobyte (KB) = \(2^{10}\) bytes
- Megabyte (MB) = \(2^{10} \times 2^{10}\) bytes
- Gigabyte (GB) = \(2^{10} \times 2^{10} \times 2^{10}\) bytes
- Terabyte (TB) = \(2^{10} \times 2^{10} \times 2^{10} \times 2^{10}\) bytes
- Petabyte (PB) = \(2^{10} \times 2^{10} \times 2^{10} \times 2^{10} \times 2^{10}\) bytes
- Exabyte (EB) = \(2^{10} \times 2^{10} \times 2^{10} \times 2^{10} \times 2^{10} \times 2^{10}\) bytes
Goals of Big Data Science in Medicine

<table>
<thead>
<tr>
<th>Goals of Big Data Science in Medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facilitating <strong>discovery science</strong>: avoiding duplication, ensuring reproducibility</td>
</tr>
<tr>
<td>Increasing <strong>understanding of human disease</strong></td>
</tr>
<tr>
<td>Improving the design, efficiency, and quality of <strong>clinical trials</strong></td>
</tr>
<tr>
<td>Improving the quality of <strong>care in clinical settings</strong></td>
</tr>
<tr>
<td>Increasing the effectiveness of <strong>prevention</strong></td>
</tr>
<tr>
<td>Translation to <strong>public</strong></td>
</tr>
<tr>
<td>Types of Data</td>
</tr>
<tr>
<td>--------------</td>
</tr>
<tr>
<td>Medication</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Demographics</td>
</tr>
<tr>
<td>Encounters</td>
</tr>
<tr>
<td>Diagnoses</td>
</tr>
<tr>
<td>Procedures</td>
</tr>
<tr>
<td>Diagnostics (ordered)</td>
</tr>
<tr>
<td>Diagnostics (results)</td>
</tr>
<tr>
<td>Genetics</td>
</tr>
<tr>
<td>Social History</td>
</tr>
<tr>
<td>Family History</td>
</tr>
<tr>
<td>Symptoms</td>
</tr>
<tr>
<td>Lifestyle</td>
</tr>
<tr>
<td>Socioeconomic</td>
</tr>
<tr>
<td>Social Network</td>
</tr>
<tr>
<td>Environment</td>
</tr>
</tbody>
</table>

**Examples of biomedical data**

- Pharmacy data
- Health care center (electronic health record) data
- Claims data
- Registry or clinical trial data
- Data outside of health care system

**Ability to link data to an individual**

- Easier to link to individuals
- Harder to link to individuals
- Only aggregate data exists

**Data quantity**

- More
- Less
Challenges

• integrating large data sets, but it is imperative that this is not uncoupled from biological investigation

• Longitudinal datasets: connect the large clinical data sets with an abundance of preclinical data,

• pharma companies externalizing and partnering on research
Future of Cardiovascular Epidemiology

- Background
- Role of
  - cHealth (community)
  - sHealth (social)
  - mHealth (mobile)
  - eHealth (electronic)
  - gHealth (genomic)
- A synthesis
Big data: The $1000 Genome

Carlson Curve

Cost per Genome

Moore’s Law
Big Data in Genomics Era

• deCODE Genetics: history records with genome data from 150,000 Icelandic people (including 15,000 whole-genome sequences).
• United Kingdom launched the 100,000 Genomes Project
• Geisinger-Regeneron collaboration launched 250,000 genomes
• PMI (US) and BGI (China): 1,000,000 genomes
**THE PRECISION MEDICINE INITIATIVE**

**WHAT IS IT?**

Precision medicine is an emerging approach for disease prevention and treatment that takes into account people’s individual variations in genes, environment, and lifestyle. The Precision Medicine Initiative will generate the scientific evidence needed to move the concept of precision medicine into clinical practice.

**WHY NOW?**

The time is right because of:
- Sequencing of the human genome
- Improved technologies for biomedical analysis
- New tools for using large datasets

**NEAR TERM GOALS**

- Intensity efforts to apply precision medicine to cancer.
- Innovative clinical trials of targeted drugs for adult, pediatric cancers
- Use of combination therapies
- Knowledge to overcome drug resistance

**LONGER TERM GOALS**

Create a research cohort of > 1 million American volunteers who will share genetic data, biological samples, and diet/lifestyle information, all linked to their electronic health records if they choose.

Pioneer a new model for doing science that emphasizes engaged participants, responsible data sharing, and privacy protection.

Research based upon the cohort data will:
- Advance pharmacogenomics, the right drug for the right patient at the right dose
- Identify new targets for treatment and prevention
- Test whether mobile devices can encourage healthy behaviors
- Lay scientific foundation for precision medicine for many diseases

[www.nih.gov/precisionmedicine](http://www.nih.gov/precisionmedicine)
Precision Medicine

- Better taxonomy of disease
- Better ontology of phenome
- Better predictive & prognostic biomarkers
- Multidimensional phenotypic/omic data
- Machine learning
- Better disease modeling, trajectory and time series
- Data lakes
Precision Medicine

- Requires an understanding of the **precise relationship between gene and phenotype, and the stratification of diseases into subtypes** according to their underlying biological mechanisms.
- Functions of most genes unknown, and what is known limited to a few cell types, tissues or physiological contexts.
- Descriptions of disease phenotypes often fail to capture the diverse manifestations of common diseases or to define subclasses of those diseases that predict the outcome or response to treatment.
  - Phenotype descriptions are typically “sloppy or imprecise”
The goal of predictive analytics in any field is to reliably predict the unknown.
Moving toward precision medicine. Ten challenges for achieving precision medicine are qualitatively ordered on the x axis by how much they are intrinsically technical versus sociopolitical challenges. The y axis qualitatively orders the difficulty each challenge currently presents if we are to attain the widely articulated goals for precision medicine.
Concept of Deep Phenotyping

- exhaustive examination of the discrete components of a phenotype that goes beyond what is typically recorded in medical charts
- There are a hundred ways to be “diabetic” involving different processes in the pancreas, liver, muscle, brain and fat
- Genetic studies lose statistical power by looking at a conglomereration of underlying causes.
Concept of Deep Phenotyping

• Different genes are responsible for particular subtypes of diabetes, so mixing them together obscures the reasons why people with the same genetic mutation respond differently to the same treatment.

• Studying ‘outbred’ mice better mirrors human diversity in diseases such as diabetes that have many genetic contributors.
Concept of Deep Phenotyping

• New human cell models of complex diseases.
• Induce skin cells to form stem cells, and can differentiate them into self-assembled clusters of cells called organoids, so they can study the connections between phenotypes, genomics and related biological data.
Genomic Big Data

• Harvesting genomes or even exomes at the population scale produces a vast amount of data, perhaps up to 40 petabytes (40 million gigabytes) each year
• Storage is not a problem
• Computational scales increase linearly
• Processing power is a limiting factor: no longer a desk top game!
• Cloud based architecture and hosting
Sharing Genomic Big Data

• A multinational coalition, the Global Alliance for Genomics and Health, developed the Framework for Responsible Sharing of Genomic and Health-Related Data.

• The Framework includes guidelines on privacy and consent, & on accountability and legal consequences for those who break the rules.

• Data-transfer agreements
Integrating genomics into electronic health records

• The NIH launched the Electronic Medical Records and Genomics (eMERGE) Network in 2007 to define best practices

• The issue there is, how do you take a practitioner who has 12 minutes per patient and about 45 seconds of time allocated for prescribing drugs, and influence their practice in a meaningful way?”

• Genome is only part of story...other omes!

• Each patient may become a big-data producer
Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data

Joshua C Denny¹,², Lisa Bastarache², Marylyn D Ritchie³, Robert J Carroll², Raquel Zink², Jonathan D Mosley¹, Julie R Field⁴, Jill M Pulley⁴,⁵, Andrea H Ramirez¹, Erica Bowton⁴, Melissa A Basford⁴, David S Carrell⁶, Peggy L Peissig⁷, Abel N Kho⁸, Jennifer A Pacheco⁹, Luke V Rasmussen¹⁰, David R Crosslin¹¹, Paul K Crane¹², Jyotishman Pathak¹³, Suzette J Bielinski¹⁴, Sarah A Pendergrass³, Hua Xu¹⁵, Lucia A Hindorff¹⁶, Rongling Li¹⁶, Teri A Manolio¹⁶, Christopher G Chute¹³, Rex L Chisholm¹⁷, Eric B Larson⁶, Gail P Jarvik¹¹,¹², Murray H Brilliant¹⁸, Catherine A McCarty¹⁹, Iftikhar J Kullo²⁰, Jonathan L Haines²¹, Dana C Crawford²¹, Daniel R Masy²²,²³, Rob M Parker²³

VOLUME 31  NUMBER 12  DECEMBER 2013  NATURE BIOTECHNOLOGY
<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Cases</th>
<th>Region</th>
<th>Nearest gene</th>
<th>SNP</th>
<th>Odds ratio (95% CI)</th>
<th>P-value</th>
<th>NHGRI Catalog disease(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoimmune</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psoriasis</td>
<td>327</td>
<td>6p21.33</td>
<td>HLA-C</td>
<td>rs10484554</td>
<td>1.71 (1.41, 2.08)</td>
<td>6.2E-08</td>
<td>Psoriasis</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>398</td>
<td>6p21.32</td>
<td>HCP5</td>
<td>rs2395029</td>
<td>2.38 (1.74, 3.26)</td>
<td>2.0E-08</td>
<td>Psoriasis</td>
</tr>
<tr>
<td>Hypothyroidism&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2,042</td>
<td>9q22.33</td>
<td>FOXE1</td>
<td>rs7850258</td>
<td>0.77 (0.71, 0.83)</td>
<td>1.1E-11</td>
<td>Hypothyroidism</td>
</tr>
<tr>
<td>Hematologic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron metabolism disorder</td>
<td>40</td>
<td>6p22.2</td>
<td>SLC17A1</td>
<td>rs17342717</td>
<td>6.84 (4.36, 10.7)</td>
<td>5.3E-17</td>
<td>Serum ferritin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HFE</td>
<td>rs1800562</td>
<td>12.3 (7.64, 19.7)</td>
<td>3.4E-25</td>
<td>Serum transferrin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIST1H2BJ</td>
<td>rs13194491</td>
<td>7.80 (4.76, 12.8)</td>
<td>3.8E-16</td>
<td>Serum transferrin</td>
</tr>
<tr>
<td>Neoplastic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanoma</td>
<td>268</td>
<td>16q24.3</td>
<td>MCIR</td>
<td>rs4785763</td>
<td>1.52 (1.27, 1.81)</td>
<td>2.8E-06</td>
<td>Melanoma</td>
</tr>
<tr>
<td>Nonmelanoma skin cancer</td>
<td>1,931</td>
<td>6p25.3</td>
<td>EXOC2</td>
<td>rs12210050</td>
<td>1.32 (1.20, 1.45)</td>
<td>6.0E-09</td>
<td>Basal cell carcinoma</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>848</td>
<td>8q24.21</td>
<td>Intergenic</td>
<td>rs1447295b</td>
<td>1.61 (1.34, 1.92)</td>
<td>2.8E-07</td>
<td>Prostate cancer</td>
</tr>
<tr>
<td>Circulatory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>1,382</td>
<td>9p21.3</td>
<td>CDKN2BAS</td>
<td>rs4977574</td>
<td>1.28 (1.17, 1.40)</td>
<td>4.0E-08</td>
<td>Myocardial infarction</td>
</tr>
<tr>
<td>Coronary atherosclerosis</td>
<td>3,499</td>
<td>9p21.3</td>
<td>CDKN2BAS</td>
<td>rs4977574b</td>
<td>1.26 (1.18, 1.34)</td>
<td>1.0E-12</td>
<td>Coronary heart disease</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>1,950</td>
<td>4q25</td>
<td>Intergenic</td>
<td>rs2200733</td>
<td>1.52 (1.34, 1.72)</td>
<td>1.5E-10</td>
<td>Atrial fibrillation</td>
</tr>
<tr>
<td>Endocrine/metabolic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>615</td>
<td>6p21.32</td>
<td>HLA-DQB1</td>
<td>rs2647044</td>
<td>1.42 (1.24, 1.61)</td>
<td>2.0E-07</td>
<td>Type 1 diabetes</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>3,122</td>
<td>10q25.2</td>
<td>TCF7L2</td>
<td>rs7903146b</td>
<td>1.31 (1.23, 1.40)</td>
<td>8.3E-16</td>
<td>Type 2 diabetes</td>
</tr>
<tr>
<td>Hypercholesterolemia</td>
<td>4,518</td>
<td>1p13.3</td>
<td>CELSR2</td>
<td>rs646776</td>
<td>0.77 (0.70, 0.85)</td>
<td>1.0E-07</td>
<td>LDL &amp; total cholesterol</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2p24.1</td>
<td>APOB</td>
<td>rs693</td>
<td>0.78 (0.73, 0.85)</td>
<td>7.4E-10</td>
<td>LDL &amp; total cholesterol</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19p13.2</td>
<td>LDLR</td>
<td>rs6511720</td>
<td>0.74 (0.65, 0.84)</td>
<td>2.5E-06</td>
<td>LDL cholesterol</td>
</tr>
<tr>
<td>Hyperglyceridemia</td>
<td>492</td>
<td>11q23.3</td>
<td>APOA5</td>
<td>rs12272004</td>
<td>2.24 (1.70, 2.95)</td>
<td>7.2E-09</td>
<td>Triglycerides</td>
</tr>
<tr>
<td>Gout</td>
<td>769</td>
<td>4p16.1</td>
<td>SLC2A9</td>
<td>rs16890979</td>
<td>0.67 (0.59, 0.78)</td>
<td>5.1E-08</td>
<td>Serum urate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4p16.1</td>
<td>Intergenic</td>
<td>rs13129697b</td>
<td>0.72 (0.63, 0.81)</td>
<td>2.4E-07</td>
<td>Gout, Serum urate</td>
</tr>
<tr>
<td>Hyperbilirubinemia</td>
<td>46</td>
<td>2q37.1</td>
<td>UGT1A1</td>
<td>rs887829a</td>
<td>33.8 (14.5, 78.5)</td>
<td>3.2E-16</td>
<td>Serum bilirubin</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alzheimer's disease</td>
<td>737</td>
<td>19q13.32</td>
<td>TOMM40</td>
<td>rs157580</td>
<td>0.70 (0.62, 0.80)</td>
<td>8.6E-08</td>
<td>Alzheimer's disease</td>
</tr>
<tr>
<td>Age-related macular degeneration</td>
<td>749</td>
<td>1q31.3</td>
<td>CFH</td>
<td>rs1329428</td>
<td>0.51 (0.45, 0.59)</td>
<td>5.2E-28</td>
<td>Alzheimer's disease</td>
</tr>
<tr>
<td>Fuchs' dystrophy</td>
<td>108</td>
<td>6p21.33</td>
<td>SKIV2L/C2/CFB</td>
<td>rs429608</td>
<td>0.57 (0.46, 0.70)</td>
<td>4.8E-08</td>
<td>Age-related macular degeneration</td>
</tr>
</tbody>
</table>
Genomics/’Omics over the Translational Stages of Cardiometabolic Therapy R&D

Target identification
Target assay construction and screening
Target validation, hits-to-leads medicinal chemistry safety
Phase I: safety
Phase II: efficacy, safety
Phase III: efficacy, safety
FDA review and approval

Harnessing Genomics/’Omics for Optimal Patient Care and Population Prevention

Target discovery & identification:
- Effect direction
- Effect size
- Correct tissue

Target validation and biomarkers:
- Patient subsets
- Risk prediction
- Genomic strata
- Biomarker strata

Drug indication selection & repositioning
- RCT patient stratification and enrichment

In era of WGS, optimal patient treatment guided by genome + adjunctive tests
### Instrumented health system study versus traditional trial or registry

<table>
<thead>
<tr>
<th></th>
<th>Traditional clinical trial or registry</th>
<th>Instrumented health system study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data source</strong></td>
<td>All data generated during and for the trial</td>
<td>Electronic health records, bio-specimen banks, laboratory information systems, payor claims, e-prescribing data, inpatient pharmacy data</td>
</tr>
<tr>
<td><strong>Data specifications</strong></td>
<td>Data formats fully specified but traditionally specific to the particular study rather than universal</td>
<td>Highly varied clinical data formats, with federal specification by the CMS and other agencies slowly increasing</td>
</tr>
<tr>
<td><strong>Data acquisition</strong></td>
<td>Data meticulously collected by trained personnel according to well-specified standard operating procedures</td>
<td>Data collected during the course of routine care by nonstandardized systems, including the <code>free text</code> dictation of physician notes</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td>Study design fully specified, including data types acquired</td>
<td>No preexisting nationwide standard of data from laboratory systems, or for annotations such as clinical notes</td>
</tr>
<tr>
<td><strong>Study hypotheses</strong></td>
<td>Small number of hypotheses tested—e.g., is drug A superior to drug B; often no secondary analysis is planned</td>
<td>Myriad questions to be asked and hypotheses to be tested in the future, not specified at the time of data acquisition</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>High cost for data standardization and collection</td>
<td>Low cost for acquisition, but variable cost for transformation and transmission</td>
</tr>
</tbody>
</table>
Principles of engagement in federated networks

• Transparency
• Representation
• Local benefit
• Right to reassort

• Cost neutrality
• Access
• Parsimony of data storage standards
Cloud Computing

- access a shared pool of data in an environment equipped with extensive and elastic computing resources and a sophisticated model for access control
- allows researchers to rent a data center under a pay-as-you-go model
- also a paradigm for writing algorithms to enable massive parallelization, allowing for scalable on-demand “supercomputers.”
- Because genomic computations are easily parallelized by genomic locus, they are ideally suited
Computational health care

• **60% of data are exogenous** (eg, behavioral, socioeconomic, environmental) and are rarely captured as part of EMR systems.

• data are generated in **uncontrolled environments** (ie, no hospital or supply-side control), which create highly fragmented value chains that need a neutral entity that can collect, store, manage, curate, and analyze data for insights

• To implement behavior modification in clinical care, it will be important to study the **biometrics, medication usage patterns, stress levels, sleep patterns, and social interactions** of individual patients
Future of Cardiovascular Epidemiology

• Background
• Role of
  – cHealth (community)
  – sHealth (social)
  – mHealth (mobile)
  – eHealth (electronic)
  – gHealth (genomic)
• A synthesis
Future of CV Epidemiology: Summing up

Ideal Health  Risk Factors  Disease
“It’s hard to tell who’s swimming naked until the tide goes out.”
Warren Buffet
What has epidemiology done for medical science lately?

Answer: much but not enough!

Suggests:
1. Refocused scientific questions
2. Centralized and integrated governance
3. Different types of exposures and outcome measures
4. Embedded clinical and policy trials
Disease Mx and Behavior Change?

• Opportunities to improve disease management and treatment may exist through context-aware data acquisition, medication/dosage and comorbidity management, and patient education and engagement

• behavior change and prevention can be addressed by using behavior models to develop recommendation services and by understanding habit-formation cycles to design new service models, incentives, and touch-point modifications
Personalized Medicine vs. Personalized Health Care

<table>
<thead>
<tr>
<th>PERSONALIZED MEDICINE</th>
<th>PERSONALIZED HEALTH CARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right</td>
<td>Best</td>
</tr>
<tr>
<td>MANTRA</td>
<td>Probabilistic</td>
</tr>
<tr>
<td>MODEL</td>
<td></td>
</tr>
<tr>
<td>FOCUS</td>
<td>Prevention, intervention, and treatment</td>
</tr>
<tr>
<td>DATA</td>
<td>Demographic, social, administrative, clinical, molecular, patient-generated/reported</td>
</tr>
</tbody>
</table>

“Figuring out how to get the right drug to the right person at the right dose at the right time.”

DR. FRANCIS COLLINS
DIRECTOR, NATIONAL INSTITUTES OF HEALTH

“If I wanted to be a doctor today I’d go to math school not to medical school.”

VINOD KHOSLA
VENTURE CAPITALIST
Thank You!

How's the big data project coming along, Hoskins?