YNHHS/Yale Med Genome Health Center
Connecticut Precision Medicine Initiative (CT-PMI)

Murat Günel, MD

Nixdorff-German Professor and Chairman, Dept. of Neurosurgery
Professor of Neurobiology and Genetics
Executive Director, Genomic Medicine, Yale Center for Genome Analysis
Yale School of Medicine
Precision Medicine Initiative (PMI)

- PMI Cohort Program (PMI-CP)
  - Aimed at building a large research cohort
    - 1 million or more Americans
  - All participants must agree:
    - to share their health data,
    - provide a biospecimen, and
    - be recontacted for future research
    - must reflect the diversity of the U.S.

- Collaborate with healthcare provider organizations (HPOs) to recruit participants
  - enabling any individual living in America to volunteer
  - use a standardized consent protocol
  - return to each participant their own results and aggregated results from its studies to all participants
Precision medicine imitative (PMI)

- A unique approach to disease treatment and prevention that seeks to maximize effectiveness by taking into account individual variability in genes, environment, and lifestyle
- Individuals can have markedly variable responses to therapy, ranging from highly efficacious outcome, to no effect, to deleterious outcome
- Taking advantage of significant advances in:
  - data collection and storage,
  - mobile health applications,
  - genomic technologies, and
  - computational analysis
PM-Approach

• The ability to recognize individuals at high risk of developing specific disorders and the development of new interventions that can prevent subsequent development of overt disease
  – ideal approaches of preventing disease in the first place
  – no specific predictive biomarkers in Alzheimer’s disease and type II diabetes mellitus

• Endeavors to redefine our understanding of:
  – disease onset and progression,
  – treatment response, and health outcomes
  – through the more precise measurement of potential contributors:

• Molecular measurements as captured through:
  – DNA sequencing technologies
  – environmental exposures
  – other information captured through increasingly ubiquitous mobile devices
Successes of Precision Medicine

- Newborn diseases
- Prenatal diagnosis
- Chronic diseases
  - cystic fibrosis
- Pharmacogenomics
  - Avoiding drugs likely to cause serious adverse effects
  - optimizing therapies based on how different polymorphisms predict therapeutic response
- Targeted treatments for cancer
National Investment in PMI

<table>
<thead>
<tr>
<th>Investment</th>
<th>Agency</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>$130 million</td>
<td>National Institutes of Health</td>
<td>To develop a voluntary national research cohort to propel our understanding of health and disease and set the foundation for a new way of doing research.</td>
</tr>
<tr>
<td>$70 million</td>
<td>NIH National Cancer Institute</td>
<td>To scale up efforts to identify genomic drivers in cancer and develop more effective approaches to cancer treatment.</td>
</tr>
<tr>
<td>$10 million</td>
<td>Food and Drug Administration</td>
<td>To acquire additional expertise and advance the development of high quality, curated databases to support the regulatory structure needed to advance innovation in precision medicine.</td>
</tr>
<tr>
<td>$5 million</td>
<td>Office of the National Coordinator</td>
<td>To support the development of interoperability standards and requirements that address privacy and enable secure exchange of data across systems.</td>
</tr>
</tbody>
</table>
Specific opportunities

1. Development of quantitative estimates of risk for a range of diseases by integrating environmental exposures, genetic factors, and gene-environment interactions;
2. Identification of determinants of individual variation in efficacy and safety of commonly used therapeutics;
3. Discovery of biomarkers that identify people with increased or decreased risk of developing common diseases;
4. Use of mobile health (mHealth) technologies to correlate activity, physiologic measures and environmental exposures with health outcomes;
5. Determination of the health impact of heterozygous loss of function mutations;
6. Development new disease classifications and relationships;
7. Empowerment of participants with data and information to improve their own health; and
8. Creation of a platform to enable trials of targeted therapy.
2. Pharmocogenomics

- Wide variation in response for many commonly used drugs, including some individuals who have no response, contributing to unnecessary expense and worse health outcomes.

- Currently, more than 150 FDA-approved drugs include genomic information in their labeling*
  - to guide their prescription
  - use based on observed associations between genotypes and treatment outcomes

- Identification of the predictors of individual response for most commonly used therapeutics

*http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm
3. Biomarker discovery

- Lacking the ability to identify individuals with a high risk for future development of a wide range of common diseases
  - thwarting prevention efforts
- Dense genotyping or genome sequencing can reveal the inherited contribution to traits
  - other biomarkers can integrate the effects of both inherited and environmental influences
**Table 2.1: Timeline when expected PMI cohort capabilities will be realized.** The estimated timeline for focused research for each type of investigation is indicated by the number of “+” characters in each cell.

<table>
<thead>
<tr>
<th>Cohort Capabilities</th>
<th>Time in years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-2</td>
</tr>
<tr>
<td>1. Discovery of disease risk factors</td>
<td>+</td>
</tr>
<tr>
<td>2. Pharmacogenomics</td>
<td>+</td>
</tr>
<tr>
<td>3. Discovery of disease biomarkers</td>
<td>+</td>
</tr>
<tr>
<td>4. mHealth connections with disease outcomes</td>
<td>+</td>
</tr>
<tr>
<td>5. Impact of loss-of-function mutations</td>
<td>+</td>
</tr>
<tr>
<td>6. New classifications of diseases</td>
<td>+</td>
</tr>
<tr>
<td>7. Empowering participants</td>
<td>+++</td>
</tr>
<tr>
<td>8. Clinical trials of targeted therapies</td>
<td>+</td>
</tr>
</tbody>
</table>
Biobank

- highest priority for a national collection would be for blood collection
- CLIA compliant procedure
- should be in place before the start of recruitment
# National biobanks

<table>
<thead>
<tr>
<th>Biobank</th>
<th>HPO system size</th>
<th>Current Biobank Size</th>
<th>Recruitment method</th>
<th>Time to achieve size (during active enrollment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Million Veteran Project</td>
<td>6 million</td>
<td>400,000</td>
<td>Mailed veterans info about MVP and enrolled at visit</td>
<td>4 years in 54 sites</td>
</tr>
<tr>
<td>Kaiser Permanente</td>
<td>10.1 million</td>
<td>245,000 (goal of 500,000)</td>
<td>Mailed consent and mailed saliva sample (N = 189,500); electronic or in-person consent and blood samples (N= 50,000)</td>
<td>3.5 years using direct mail to 2 million</td>
</tr>
<tr>
<td>Partners Healthcare Biobank</td>
<td>6 million</td>
<td>&gt;30,000</td>
<td>In-person at outpatient visits and inpatient floors; Electronic consent via emails using patient portal</td>
<td>5 years since launch: 2 year pilot study; 3 years via in person recruitment; eConsent for past 1 year; current rate is 1100/month</td>
</tr>
<tr>
<td>Geisinger MyCode</td>
<td>1.3 million with an EHR encounter in last 10 years</td>
<td>&gt;86,000</td>
<td>In-person during routine outpatient visit; Electronic consenting pending</td>
<td>10 years; however, current rate is 1000/week</td>
</tr>
</tbody>
</table>
# National biobanks

<table>
<thead>
<tr>
<th>Biobank</th>
<th>Subjects</th>
<th>Participants</th>
<th>Recruitment Method</th>
<th>Timeframe Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marshfield Clinic Personalized Medicine</td>
<td>&gt;2 million</td>
<td>20,000</td>
<td>In-person, recruited via phone and mailers</td>
<td>16 months at 4 sites of Marshfield clinic</td>
</tr>
<tr>
<td>Research Program</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayo Clinic</td>
<td>2 million</td>
<td>&gt;60,000</td>
<td>In-person consent at clinic</td>
<td>7 years, current rate about 8000-9000/year</td>
</tr>
<tr>
<td>Children’s Hospital of Philadelphia</td>
<td>2.5 million</td>
<td>110,000</td>
<td>In-person consent at clinics</td>
<td>9 years</td>
</tr>
<tr>
<td>Cincinnati Children's Hospital Medical Center</td>
<td>670 thousand</td>
<td>&gt;56,000</td>
<td>Hospital-wide consent by registrars at registration</td>
<td>4 years</td>
</tr>
</tbody>
</table>
Data set

- includes data from:
  - EHRs,
  - health insurance organizations,
  - participant surveys,
  - mHealth technologies, and
  - biologic investigations
## Table 5.1: Categories, Sources, and Uses of Data

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Source(s)</th>
<th>Example Uses</th>
<th>Core/Subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual demographics and contact information</td>
<td>Date and place of birth, sex and gender, detailed and multiple race/ethnicities (e.g., Asian of Indian descent, Asian of Chinese descent), name, mailing address, phone number, cell phone number, email address, marital status, educational status, occupation/income</td>
<td>Study participant, healthcare provider organizations</td>
<td>Participant-specific communications, analytics, risk stratification, assessment of covariates and confounds, study appointment reminders, invitations to participate in sub-studies</td>
<td>C</td>
</tr>
<tr>
<td>Terms of consent and personal preferences for participation in the project</td>
<td>Fine-grained consent for options to participate e.g., receive research results</td>
<td>Study participant</td>
<td>“Precision Participant Engagement”</td>
<td>C</td>
</tr>
<tr>
<td>Self-reported measures</td>
<td>Pain scales, disease-specific symptoms, functional capabilities, quality of life and well-being, gender identity, structured family health history</td>
<td>Study participant</td>
<td>Many</td>
<td>C/S</td>
</tr>
<tr>
<td>Behavioral and lifestyle measures</td>
<td>Diet, physical activity, alternative therapies, smoking, alcohol, assessment of known risk factors (e.g., guns, illicit drug use)</td>
<td>Study participant (retrospective and prospective) and healthcare provider organizations</td>
<td>Correlation with clinical events, drug response, and health outcomes</td>
<td>C/S</td>
</tr>
<tr>
<td>Sensor-based observations through phones, wearables, home-based devices</td>
<td>Location, activity monitors, cardiac rate and rhythm monitoring, respiratory rate</td>
<td>Smartphone sensors, commercial and research-grade physiologic monitors</td>
<td>Functional ability and impairment assessment</td>
<td>C/S</td>
</tr>
<tr>
<td>Structured clinical data derived from Electronic Health Records (EHRs)</td>
<td>ICD/CPT billing codes, clinical lab values, medications, problem lists</td>
<td>Multiple provider organizations per study participant, via institutionally managed channels or direct from</td>
<td>Correlation of clinical events with other categories of data</td>
<td>C</td>
</tr>
</tbody>
</table>
## Data collection

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Sources/Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured and specialized types of clinical data derived from EHRs</td>
<td>Narrative documents, images, EKG and EEG waveform data</td>
<td>Correlation of clinical events with other categories of data</td>
</tr>
<tr>
<td>PMI baseline health exam</td>
<td>Vital signs, medication assessment, past medical history</td>
<td>Provides baseline measures on all participants</td>
</tr>
<tr>
<td>Healthcare claims data</td>
<td>Periods of coverage, charges and associated billing codes as received by public and private payers, outpatient pharmacy dispensing (product, dose, amount)</td>
<td>Assessments requiring complete longitudinal record of exposures/outcomes during specific periods, e.g., within X years of a diagnosis or medication exposure; health services research, exposure and outcomes assessment</td>
</tr>
<tr>
<td>Research specific observations</td>
<td>Research questionnaires, ecological momentary assessments, performance measures (six minute walk test), disease specific monitors (e.g. glucometers, spirometers)</td>
<td>Many</td>
</tr>
<tr>
<td>Biospecimen-derived laboratory data</td>
<td>Genomics, proteomics, metabolites, cell-free DNA, single cell studies, infectious exposures, standard clinical chemistries, histopathology</td>
<td>Study participants, provider organizations, outsourced laboratories</td>
</tr>
<tr>
<td>Geospatial and environmental data</td>
<td>Weather, air quality, environmental pollutant levels, food deserts, walkability, population density, climate change</td>
<td>Correlation of tissue findings and high throughput biomolecular data with other categories of data</td>
</tr>
<tr>
<td>Other data</td>
<td>Social networking e.g., Twitter feeds, social contacts from cell phone text and voice, OTC medication purchases</td>
<td>Public and private sources not directly part of PMI</td>
</tr>
<tr>
<td></td>
<td>Public and private sources not directly part of PMI</td>
<td>Predictive analytics</td>
</tr>
</tbody>
</table>
1. Phenotyping and sample collection
2. Bio-bank creation
3. Genomic profiling
4. Computer infrastructure
5. Analysis
6. Patient interface
7. Medical applications
1. **Phenotyping and sample collection**
   - Diverse CT population
   - Patient identification and sample collection
     - Uniform Consent
   - Phenotyping
     - Data collection
   - EHR (EPIC) interface

2. **Bio-bank creation**
   - Robotic storage
   - Database management
   - New collection
     - the highest priority is to obtain blood specimens
3. **Genomic profiling**
   - Exome sequencing (for PAF rare variants) and whole genome genotyping for common variants
   - Aiming for genome sequencing

4. **Computer infrastructure**
   - HPC cluster
   - HIPAA compliant data storage
   - Maintenance
5. Analysis
   - Variant calling
     • Confirmation
     • Core reports
   - Pipeline production
   - Reports
     • Ancestry
     • Clinically relevant variants
     • Pharmacogenomics
     • Genetic risk scores
       - Rare LOF variants
       - Common variant scores
     • Scientific discoveries
6. Patient interface
   - Patients as partners
     • Reporting data back
   - EPIC data load
   - Genetic counselors
   - Domain expert physicians

7. Medical applications
Opportunities

• Collaborations across institutions
  – Throughout CT and nationally
• New scientific discoveries
  – Increased NIH funding
• Insurance collaborations
• Industry partnerships
  – Start-ups
  – Big pharma
• Opportunities in education
• Population and personal health