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

Murat Günel, MD

Nixdorff-German Professor and Chairman, Dept. of Neuosurgery
Professor of Neurobiology and Genetics
Executive Director, Genomic Medicine, Yale Center for Genome Analysis
Yale School of Medicine

Yale SCHOOL OF MEDICINE

Yale NewHaven Health

Precision Medicine Initiative (PMI)

The Precision Medicine Initiative Cohort

- 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 1.1: Proposed PMI Budget Allocations for FY 2016 Department of Health and Human Services		
Investment	Agency	Purpose
$130 million	National Institutes of Health	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.
$70 million	NIH National Cancer Institute	To scale up efforts to identify genomic drivers in cancer and develop more effective approaches to cancer treatment.
$10 million	Food and Drug Administration	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.
$5 million	Office of the National Coordinator	To support the development of interoperability standards and requirements that address privacy and enable secure exchange of data across systems.

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

PMI Timeline

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.

		Time in years			
		0-2	3-5	5-10	>10
Cohort Capabilities	1. Discovery of disease risk factors	+	+++	+++	++++
	2. Pharmacogenomics	+	+++	+++	+++
	3. Discovery of disease biomarkers	+	++	+++	+++
	4. mHealth connections with disease outcomes		+	++	++++
	5. Impact of loss-of-function mutations		+	+++	+++
	6. New classifications of diseases		+	+++	++++
	7. Empowering participants	+++	+++	+++	+++
	8. Clinical trials of targeted therapies		+	+++	+++

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

National biobanks

Marshfield Clinic Personalized Medicine Research Program	>2 million	20,000	In-person, recruited via phone and mailers	16 months at 4 sites of Marshfield clinic
Mayo Clinic	2 million	>60,000	In-person consent at clinic	7 years, current rate about 8000-9000/year
Children's Hospital of Philadelphia	2.5 million	110,000	In-person consent at clinics	9 years
Cincinnati Children's Hospital Medical Center	670 thousand	>56,000	Hospital-wide consent by registrars at registration	4 years

Data set

- includes data from:
  - EHRs,
  - health insurance organizations,
  - participant surveys,
  - mHealth technologies, and
  - biologic investigations

Data collection

Table 5.1: Categories, Sources, and Uses of Data				
Category	**Examples**	**Source(s)**	**Example Uses**	**Core/ Subgroup**
Individual demographics and contact information	Date and place of birth, sex and gender, detailed and multiple races/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	Study participant, healthcare provider organizations	Participant-specific communications, analytics, risk stratification, assessment of covariates and confounds, study appointment reminders, invitations to participate in sub-studies	C
Terms of consent and personal preferences for participation in the project	Fine-grained consent for options to participate e.g., receive research results	Study participant	"Precision Participant Engagement"	C
Self-reported measures	Pain scales, disease-specific symptoms, functional capabilities, quality of life and well-being, gender identity, structured family health history	Study participant	Many	C/S
Behavioral and lifestyle measures	Diet, physical activity, alternative therapies, smoking, alcohol, assessment of known risk factors (e.g., guns, Illicit drug use)	Study participant (retrospective and prospective) and healthcare provider organizations	Correlation with clinical events, drug response, and health outcomes	C/S
Sensor-based observations through phones, wearables, home-based devices	Location, activity monitors, cardiac rate and rhythm monitoring, respiratory rate	Smartphone sensors, commercial and research-grade physiologic monitors	Functional ability and impairment assessment	C/S
Structured clinical data derived from Electronic Health Records (EHRs)	ICD/CPT billing codes, clinical lab values, medications, problem lists	Multiple provider organizations per study participant, via institutionally managed channels or direct from	Correlation of clinical events with other categories of data	C

Data collection

Unstructured and specialized types of clinical data derived from EHRs	Narrative documents, images, EKG and EEG waveform data	Multiple providers, via federated queries rather than inclusion in core dataset	Correlation of clinical events with other categories of data	S
PMI baseline health exam	Vital signs, medication assessment, past medical history	Study participant interacting with healthcare provider organization	Provides baseline measures on all participants	C
Healthcare claims data	Periods of coverage, charges and associated billing codes as received by public and private payers, outpatient pharmacy dispensing (product, dose, amount)	CMS and other federal sources, private insurers, pharmacy benefits management organizations	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	C
Research specific observations	Research questionnaires, ecological momentary assessments, performance measures (six minute walk test), disease specific monitors (e.g. glucometers, spirometers)	Study participants, research organizations	Many	S
Biospecimen-derived laboratory data	Genomics, proteomics, metabolites, cell-free DNA, single cell studies, infectious exposures, standard clinical chemistries, histopathology	Study participants, provider organizations, outsourced laboratories	Correlation of tissue findings and high throughput biomolecular data with other categories of data	C
Geospatial and environmental data	Weather, air quality, environmental pollutant levels, food deserts, walkability, population density, climate change	Public and private sources not directly part of PMI	Epidemiology, epidemic surveillance	C/S
Other data	Social networking e.g., Twitter feeds, social contacts from cell phone text and voice, OTC medication purchases	Public and private sources not directly part of PMI	Predictive analytics	S

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

1. **Phenotyping and sample collection**
2. **Bio-bank creation**
3. **Genomic profiling**
4. **Computer infrastructure**
5. **Analysis**
6. **Patient interface**
7. **Medical applications**

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

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

GHC/CT-PMI

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

GHC/CT-PMI

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

GHC/CT-PMI

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