Predictive Genomics for Prevention, Interception and Cure

Health Datapalooza

Emma Huang | Population Analytics | 27 April 2017
Looking to the future, our goal is to leverage the unprecedented advances in science to **treat, intercept and prevent diseases** in novel ways. Through a combination of internal expertise and external innovation, we hope to make a meaningful difference for patients all over the world.

– Bill Hait, Head of Janssen R&D
Accelerating learning across the health continuum

Population-level data

Clinical manifestation
No clinical manifestation

Onset of morbidity

HEALTHY: PREVENTION
DTC genotyping

INCUBATION: INTERCEPTION
Diagnostic Tests

DISEASE: CURE
Clinical WGS
PGx panels
Challenge 1: Integrating personalized data
Global Initiatives linking human genetics & clinical phenotypes

- deCODE genetics
  - UK Biobank: 500,000
- Amgen
  - 300,000
- Geisinger
  - Regeneron: 100,000
- Kaiser Permanente
  - > 200,000
- Geisinger + Regeneron
  - 100,000
- Human Longevity + AstraZeneca
  - 500,000
- Veterans Affairs
  - 1,000,000
- Estonian Precision Medicine Initiative
- Finland Founder Population
- Estonia Precision Medicine Initiative
- China PMI
  - 100,000,000 by 2030
- Kadoorie Biobank
  - ~500,000
- 23andMe
  - > 1.4M
- All of Us
  - 1,000,000
- Saudi Human Genome Program
  - 100,000
End goal for predictive genomics:

Using data collected from diverse instruments & sensors, we want to accurately map the probability for disease risk:

\[ P(D \mid \{G\}, \{E\}, \{L\}, \text{Gender, Age}) \]

taking into account family history, prior treatment, etc.
Data needs depend on stage of health

- **HEALTHY: PREVENTION**
  - Health Status:
    - Primary Care
    - Routine Screening
  - Health Data Contacts:
    - Employment
    - Taxation
    - Education
    - Social Media
    - Housing
    - Consumer Data

- **INCUBATION: INTERCEPTION**
  - Morbidity:
    - Primary Care
    - Outpatient Records
    - Clinical Test Results
    - Disease Registry
  - Volume of Data:
    - Hospital
    - Employment
    - Taxation
    - Education
    - Social Media
    - Housing
    - Consumer Data

- **DISEASE: CURE**
  - Multi-morbidity:
    - Primary Care
    - Emergency Care
    - Inpatient Hospital
    - Clinical Test Results
    - Disease Registry
  - Last Year Of Life:
    - Social Care
    - Social Security
    - Pension
    - Social Media
    - Housing
    - Consumer Data

Deeny and Steventon, BMJ Quality and Safety 2015
Challenge 2: Integrating stages of prediction

**Genomic prediction**
Predictive models to assess the risk of any disease, given genetic background

**PLATO: patient-level prediction**
Predictive models to assess the probability of a patient experiencing any outcome following initiation of any intervention, given personal medical history

**Dynamic prediction of individual prognoses**
Predictive models to assess the probability of a patient experiencing outcomes given personal history (genetic, environmental, behavioral)
Population Analytics

Biostatistics
Longitudinal studies
Statistical and population genomics
Data integration
Functional genomics

What patients responded?
Real world data
Descriptive

Why did they respond?
Disease biology, PGx, etc.
Diagnostic

Given genetics, lifestyle, and environment, what is your risk for disease?
Predictive

If your risk is high, how should we intercept (and when)?
Prescriptive

How should we maintain wellness?
Pre-emptive

Hindsight
Insight
Foresight

And growing!
Prevention
- Genomic prediction
- Spatiotemporal data integration
- Feature extraction from continuous streaming

Interception
- Molecular definitions of disease, biomarkers
- Dynamic prediction of individualized prognoses
- Tracking of risk and need for intervention

Cure
- Subgroup characterization
- Prediction of response to interventions
- Prediction of disease progression and best intervention
Challenge 3: Translation to patients

- Retrospective study of phase III clinical trial
- Goal: identify genetic predictors of response to anti-TNF drug golimumab
- Data: clinical phenotypes and WGS for 436 RA patients

Courtesy of Kris Standish
<table>
<thead>
<tr>
<th>CURRENT BASELINE</th>
<th>FUTURE BASELINE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Validation, infrastructure and analytics to demonstrate value of data generation</strong></td>
<td>WGS, continuous sensing, linked with EHR for millions of people</td>
</tr>
<tr>
<td><strong>Understanding of how to add value with analytics; of what needs to be measured; of disease</strong></td>
<td>Widespread data analytics adoption</td>
</tr>
<tr>
<td>Omics, SNPs for 1000s of people</td>
<td></td>
</tr>
<tr>
<td>Widespread data generation adoption</td>
<td></td>
</tr>
<tr>
<td>Quantification of self</td>
<td>Effortlessly, and can be interpreted meaningfully</td>
</tr>
<tr>
<td>Devices become unobtrusive; behavioral shift; evidence of value for health</td>
<td></td>
</tr>
<tr>
<td>Basis for regulation and payment is outcomes</td>
<td>Basis for regulation and payment is value</td>
</tr>
<tr>
<td>Definition of value; real-world evidence collection and analytics to quantify value</td>
<td></td>
</tr>
<tr>
<td>Vendors with proven success in healthcare; collaborations between technology and healthcare industries</td>
<td>Big data infrastructure for healthcare</td>
</tr>
<tr>
<td>Big data infrastructure for business</td>
<td></td>
</tr>
</tbody>
</table>
Thanks!