Genome-wide polygenic scores to stratify risk for common diseases

Sekar Kathiresan, M.D.
Director, Center for Genomic Medicine, MGH
Director, Cardiovascular Disease Initiative, Broad

June 21, 2018
Key public health need: identify individuals at high risk for disease

Most diseases inherited component

Stratify individuals based on inherited DNA variation
Myocardial infarction (MI) or heart attack

Symptoms

EKG change

Elevation in cardiac biomarkers

- **Complex trait** with heritable & lifestyle components
- $\sim\frac{1}{2}$ all MIs, 1st presentation sudden death
Inherited component to early MI

MI at age < 55  Age onset at MI
Traditional approach: Genetic prediction focuses on rare, monogenic mutations

Familial hypercholesterolemia

↑ Cholesterol

Heart attack
3x increased risk

0.4% of the population

↑ Cholesterol
Question: can we identify additional patients with a polygenic risk model?
Concept: polygenic risk scores

- ‘N’ polymorphisms
- 0, 1, or 2 copies of the risk allele
- Score ranging from 0 to 2N for each person

Polygenic risk scores: move from top SNPs to a genome-wide set of 6.6M for prediction

Khera*, Chaffin*, bioRxiv 2017

Amit V. Khera
**Hypothesis:** a polygenic score including a genome-wide set of SNPs can identify individuals with risk equivalent to a monogenic mutation

---

**Step 1**
Training data set: effect sizes for 6.6 million variants from genome-wide association study

- **Cases** N = 60K
- **Controls** N = 120K

**Step 2**
Validation Dataset: ~125K

- **Cases** N = 4K
- **Controls** N = 120K

**Step 3**
Testing Dataset: ~300K

- **Cases** N = 8.7K
- **Controls** N = 288K

**Genotypes:** from arrays + imputation

---

A new quantitative metric of genetic liability to heart attack

Polygenic score of 6.6 million common variants

Genome-wide polygenic score: little correlation with currently measured MI risk factors

Correlation with ACC/AHA Pooled Cohorts Equation

Correlation coefficient $r = 0.02$

Using polygenic model, can we identify group with risk for MI equivalent to monogenic mutations?
>20-fold gradient in risk across bins of score

Top 5% of polygenic MI score: risk equivalent to monogenic mutations

Polygenic score of 6.6 million common variants

Remainder of distribution

Top 5%

High polygenic score definition

<table>
<thead>
<tr>
<th>High polygenic score definition</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 5%</td>
<td>3.3</td>
</tr>
<tr>
<td>Top 1%</td>
<td>4.7</td>
</tr>
</tbody>
</table>

## Monogenic, polygenic contributions to early MI

<table>
<thead>
<tr>
<th></th>
<th>Monogenic</th>
<th>Polygenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence in population</td>
<td>0.4%</td>
<td>5%</td>
</tr>
<tr>
<td>Odd ratio for MI</td>
<td>3.2</td>
<td>3.3</td>
</tr>
<tr>
<td>Mode of detection</td>
<td>↑ LDL cholesterol</td>
<td>Currently UNAWARE</td>
</tr>
<tr>
<td>Mechanism of risk</td>
<td>apoB lipoproteins</td>
<td>‘Gemish’</td>
</tr>
</tbody>
</table>
# Monogenic, polygenic contributions to early MI

<table>
<thead>
<tr>
<th></th>
<th>Monogenic</th>
<th>Polygenic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalence in population</strong></td>
<td>0.4%</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Odd ratio for MI</strong></td>
<td>3.2</td>
<td>3.3</td>
</tr>
<tr>
<td><strong>Mode of detection</strong></td>
<td>↑ LDL cholesterol</td>
<td>Currently UNAWARE</td>
</tr>
<tr>
<td><strong>Mechanism of risk</strong></td>
<td>apoB lipoproteins</td>
<td>‘Gamish’</td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td>Lifestyle, Medications</td>
<td>?</td>
</tr>
</tbody>
</table>
Is polygenic risk for MI modifiable? Yes

**Lifestyle**


↓48%

**Medicines**


↓44%
Monogenic vs polygenic contribution to early MI

100 patients with early MI

<table>
<thead>
<tr>
<th>%</th>
<th>↑ Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monogenic</td>
<td>3-4 fold</td>
</tr>
<tr>
<td>Polygenic</td>
<td>3-4 fold</td>
</tr>
<tr>
<td>Monogenic &amp; polygenic</td>
<td>6-8 fold</td>
</tr>
</tbody>
</table>
Polygenic risk for other common diseases... including those without monogenic risk factors

Potential for major impact on clinical practice

<table>
<thead>
<tr>
<th>Disease</th>
<th>% of population at &gt;3-fold risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td>6.2%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>3.6%</td>
</tr>
<tr>
<td>Inflammatory Bowel Disease</td>
<td>3.0%</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>2.1%</td>
</tr>
</tbody>
</table>
Conclusions

• Now possible to score polygenic component to any complex trait (from genotyping array data, simultaneous for many diseases, at birth)

• Those in extremes of score: at risk for disease approaching or exceeding monogenic mutations

• Use scores to guide prevention, treatment, or screening strategies