Estimation of a polygenic risk score for ambulatory care sensitive conditions

Arindam Basu
School of Health Sciences,
University of Canterbury
arindam.basu@canterbury.ac.nz
Outline of the presentation

- Ambulatory sensitive hospitalisations
- Common genetic variants that can explain ASH
- What do GWAS studies tell us about ASH
- Estimation of PRS from GWAS studies
- Next steps
Ambulatory sensitive hospitalisation

- Asthma
- COPD
- TYPEII DM
- CHF
- HTN

Can be managed in the primary care setting

Avoid hospitalised care
Hospitalisation due to ASH in NZ

Source: NZHiS data, 21 DHBs, Year 2005

Note: Avoidable hospitalisation as percent from total hospitalisation
One sample of GWAS results due to Asthma
## List of GWAS studies on Asthma

**Table 2: Asthma Genetic Loci Identified by Genome-Wide Association Studies (GWAS)**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Primary Cohort Size</th>
<th>Replication Sample Size</th>
<th>Gene/Region</th>
<th>Novel Gene/Pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moffatt et al. (14)</td>
<td>994 subjects with asthma, 1,243 subjects without asthma</td>
<td>5,621 subjects</td>
<td>17q21 (ORMDL3)(^3)</td>
<td>Yes</td>
</tr>
<tr>
<td>Himes et al. (9)</td>
<td>359 subjects with asthma, 846 control subjects</td>
<td>18,891 subjects</td>
<td>PDE4D</td>
<td>No</td>
</tr>
<tr>
<td>Li et al. (11)</td>
<td>473 subjects with asthma, 1,892 control subjects</td>
<td>6q21 (HLA-DR, HLA-DQ), 5q31 (IL13, RAD50)</td>
<td>Yes (RAD50)</td>
<td></td>
</tr>
<tr>
<td>Sleiman et al. (16)</td>
<td>793 subjects with asthma, 1,988 control subjects</td>
<td>917 subjects with asthma, 1,546 control subjects</td>
<td>17q21(^2), DENND1B(^2)</td>
<td>Yes (DENND1B)</td>
</tr>
<tr>
<td>Hancock et al. (17)</td>
<td>492 Mexican trios</td>
<td>177 Mexican trios</td>
<td>TLE4</td>
<td>Yes</td>
</tr>
<tr>
<td>Choudhry S et al. (18)</td>
<td>96 cases, 88 controls (Puerto Rican)</td>
<td>284 Puerto Rican trios</td>
<td>5q23.3</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Large consortium meta-analysis of GWAS studies on Hypertension (partial list)

Table 1

<table>
<thead>
<tr>
<th>SNP identifier</th>
<th>Chr</th>
<th>Position</th>
<th>Gene</th>
<th>MAF</th>
<th>CHARGE Meta-analysis SBP</th>
<th>CHARGE Meta-analysis DBP</th>
<th>CHARGE Meta-analysis Hypertension</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2681492</td>
<td>12</td>
<td>88537220</td>
<td>ATP2B1</td>
<td>0.20</td>
<td>-1.26 0.19 3.0E-11</td>
<td>-0.62 0.11 4.6E-08</td>
<td>-0.14 0.03 8.4E-08</td>
</tr>
<tr>
<td>rs2681472</td>
<td>12</td>
<td>88533990</td>
<td>ATP2B1</td>
<td>0.18</td>
<td>-1.29 0.19 3.5E-11</td>
<td>-0.64 0.11 3.7E-08</td>
<td>-0.16 0.03 1.7E-08</td>
</tr>
<tr>
<td>rs11105354</td>
<td>12</td>
<td>88556654</td>
<td>ATP2B1</td>
<td>0.18</td>
<td>-1.30 0.20 3.7E-11</td>
<td>-0.63 0.11 5.8E-08</td>
<td>-0.16 0.03 1.8E-08</td>
</tr>
<tr>
<td>rs11105364</td>
<td>12</td>
<td>88593407</td>
<td>4.8E-11</td>
<td>0.18</td>
<td>-1.30 0.20 4.8E-11</td>
<td>-0.63 0.12 1.2E-07</td>
<td>-0.16 0.03 2.1E-08</td>
</tr>
<tr>
<td>rs17249754</td>
<td>12</td>
<td>88584717</td>
<td>5.2E-11</td>
<td>0.18</td>
<td>-1.30 0.20 5.2E-11</td>
<td>-0.63 0.12 1.0E-07</td>
<td>-0.16 0.03 2.2E-08</td>
</tr>
<tr>
<td>rs11105368</td>
<td>12</td>
<td>88598572</td>
<td>5.3E-11</td>
<td>0.18</td>
<td>-1.30 0.20 5.3E-11</td>
<td>-0.63 0.12 1.3E-07</td>
<td>-0.16 0.03 2.2E-08</td>
</tr>
<tr>
<td>rs12579302</td>
<td>12</td>
<td>88574634</td>
<td>ATP2B1</td>
<td>0.18</td>
<td>-1.29 0.20 6.2E-11</td>
<td>-0.62 0.12 1.3E-07</td>
<td>-0.16 0.03 2.2E-08</td>
</tr>
<tr>
<td>rs1220074</td>
<td>12</td>
<td>88614998</td>
<td>9.1E-11</td>
<td>0.17</td>
<td>-1.31 0.20 9.1E-11</td>
<td>-0.62 0.12 3.4E-07</td>
<td>-0.17 0.03 2.9E-08</td>
</tr>
<tr>
<td>rs11105378</td>
<td>12</td>
<td>88614872</td>
<td>9.1E-11</td>
<td>0.17</td>
<td>-1.31 0.20 9.1E-11</td>
<td>-0.62 0.12 3.1E-07</td>
<td>-0.17 0.03 2.8E-08</td>
</tr>
<tr>
<td>rs4842666</td>
<td>12</td>
<td>88456580</td>
<td>6.5E-09</td>
<td>0.17</td>
<td>-1.29 0.21 6.5E-09</td>
<td>-0.62 0.12 4.5E-07</td>
<td>-0.15 0.03 3.4E-07</td>
</tr>
<tr>
<td>rs8096897</td>
<td>18</td>
<td>13428905</td>
<td>CASZ1</td>
<td>0.01</td>
<td>-12.87 2.33 3.2E-08</td>
<td>-4.07 1.33 2.9E-03</td>
<td>-0.73 0.35 0.04</td>
</tr>
<tr>
<td>rs11105328</td>
<td>12</td>
<td>88466521</td>
<td>4.2E-08</td>
<td>0.18</td>
<td>-1.11 0.20 4.2E-08</td>
<td>-0.61 0.12 5.1E-07</td>
<td>-0.15 0.03 7.1E-07</td>
</tr>
<tr>
<td>rs880115</td>
<td>1</td>
<td>10719453</td>
<td>CASZ1</td>
<td>0.35</td>
<td>0.09 0.17 2.1E-07</td>
<td>0.20 0.10 2.9E-03</td>
<td>0.09 0.02 6.2E-05</td>
</tr>
<tr>
<td>rs3184504</td>
<td>12</td>
<td>11036899</td>
<td>SNX25</td>
<td>0.48</td>
<td>0.75 0.15 5.7E-07</td>
<td>0.50 0.09 1.7E-08</td>
<td>0.07 0.02 7.4E-04</td>
</tr>
<tr>
<td>rs581815</td>
<td>11</td>
<td>16858844</td>
<td>PLEKH7</td>
<td>0.26</td>
<td>0.84 0.17 5.8E-07</td>
<td>0.51 0.10 4.3E-07</td>
<td>0.09 0.02 1.7E-04</td>
</tr>
</tbody>
</table>
It is possible to obtain a list of studies on the five conditions from GWAS central
Study question

genetics

ASH Outcomes

All Other Variables
Polygenic Risk Score

- Single weighted summed score of SNPs from GWAS studies
- Weighted by their respective beta coefficients for continuous outcomes such as blood pressure scores or outcome scores
- (Alternative), weighted by their Odds Ratios for binary outcomes
- Single PRS is then used in a regression model to predict or study association between genotypic contribution to the phenotype, as in
- Phenotype ~ PRS score + Other variables
Steps in our case

- Identify candidate SNPs through an exploratory meta-analysis
- Quality control of the genotype data
- The candidate SNPs form a base population
- Identify a target population for whom genotype and phenotype data are available (phenotype == “access to primary care” for ASH)
- Construct the PRS in the base population
- Apply to the target population, run models
Example meta-analysis script file from the metal meta-analysis helper page

```plaintext
# VERBOSE ON

# Describe and process the DGI input files
MARKER SNP
WEIGHT N
ALLELE EFFECT_ALLELE NON_EFFECT_ALLELE
FREQ FREQ_EFFECT
EFFECT BETA
STDERR SE
PVAL P_VALUE

PROCESS DGI_three_regions.txt

# Describe and process the FUSION input files
MARKER SNP
ALLELE EFFECT_ALLELE NON_EFFECT_ALLELE
FREQ FREQ_EFFECT
WEIGHT N
EFFECT BETA
STDERR SE
PVAL P_VALUE

PROCESS MAGIC_FUSION_Results.txt.gz

# Describe and process the SardiNIA input files
MARKER SNP
DEFAULT 4106
ALLELE AL1 AL2
FREQ FREQ1
EFFECT EFFECT
STDERR SE
PVAL P_VALUE

PROCESS magic_SARDINIA.tbl

# Execute meta-analysis
ANALYZE
```
### Top 10 Meta-Analysis Results

<table>
<thead>
<tr>
<th>MarkerName</th>
<th>Allele1</th>
<th>Allele2</th>
<th>Weight</th>
<th>Zscore</th>
<th>P-value</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs560887</td>
<td>t</td>
<td>c</td>
<td>6806</td>
<td>-7.075</td>
<td>1.491E-12</td>
<td>- - -</td>
</tr>
<tr>
<td>rs853787</td>
<td>t</td>
<td>g</td>
<td>6806</td>
<td>6.691</td>
<td>2.221E-11</td>
<td>+ + +</td>
</tr>
<tr>
<td>rs853789</td>
<td>a</td>
<td>g</td>
<td>5339</td>
<td>-6.597</td>
<td>4.189E-11</td>
<td>? - ?</td>
</tr>
<tr>
<td>rs853773</td>
<td>a</td>
<td>g</td>
<td>6806</td>
<td>-6.132</td>
<td>8.662E-10</td>
<td>- - -</td>
</tr>
<tr>
<td>rs537183</td>
<td>t</td>
<td>c</td>
<td>6806</td>
<td>6.007</td>
<td>1.887E-09</td>
<td>+ + +</td>
</tr>
<tr>
<td>rs557462</td>
<td>t</td>
<td>c</td>
<td>6806</td>
<td>6.005</td>
<td>1.917E-09</td>
<td>+ + +</td>
</tr>
<tr>
<td>rs502570</td>
<td>a</td>
<td>g</td>
<td>6806</td>
<td>-6.001</td>
<td>1.955E-09</td>
<td>- - -</td>
</tr>
<tr>
<td>rs563694</td>
<td>a</td>
<td>c</td>
<td>6806</td>
<td>5.975</td>
<td>2.300E-09</td>
<td>+ + +</td>
</tr>
<tr>
<td>rs475612</td>
<td>t</td>
<td>c</td>
<td>6806</td>
<td>-5.867</td>
<td>4.423E-09</td>
<td>- - -</td>
</tr>
<tr>
<td>rs853781</td>
<td>a</td>
<td>g</td>
<td>6806</td>
<td>-5.844</td>
<td>5.092E-09</td>
<td>- - -</td>
</tr>
</tbody>
</table>
Using PRSice to compute Polygenic Risk Score

Rscript PRSice.R --dir . \\
  --prsice ./PRSice \\
  --base BASE_GWAS.assoc \\
  --target TARGET_DATA \\
  --thread 1 \\
  --stat OR \\
  --binary-target T \\
  (Script to run a PRS scoring algorithm based on GWAS)
Interpretation of the PRS output for the phenotype (here access to care)
Next steps for this project

- Quality control of the original GWAS data and then conduct a meta-analysis of GWAS studies on a set of defined population for the five conditions together
- Pool together the results of the five conditions and identify a set of candidate genes
- On that basis identify a target population (hardest hurdle to overcome)
- Construct the PRS and fit the PRS to the target population and identify the Odds Ratios
Outcomes and benefits

- A common set of variants for the common conditions will indicate a genetic component for access to care
- It’d be possible to study gene*gene and gene*environment interactions
- This study will extend the scope of genome wide association studies to preventive health
- This is an example of Precision public health as we can now cluster and quantify which population groups based on their genetic profile can benefit most from targeted preventive interventions