Methods and Applications in Gene-Environment Interactions

林苑 (Wan-Yu Lin)
2020.12.05  Taiwanese Association of Statistical Methodology
http://homepage.ntu.edu.tw/~linwy/
Institute of Epidemiology and Preventive Medicine,
College of Public Health,
National Taiwan University, Taipei, Taiwan
Gene-environment interactions

• Genetic effects are not constant for all subjects
• While genetic materials are inborn, environmental exposures can be changed
Single-nucleotide polymorphism (SNP)

Variation in DNA sequence
Changes in adenine (A), thymine (T), cytosine (C), or guanine (G)

https://isogg.org/wiki/Single-nucleotide_polymorphism
Three possible genotypes in a SNP

- For example, if a SNP has two alleles A and G
  - AA (0, 0 allele of G)
  - AG (1, 1 allele of G)
  - GG (2, 2 alleles of G)
Gene: a chromosomal region

https://medium.com/sanogenetics/snp-of-the-week-77753b4aea87
Phenotype

• A trait of interest
  ➢ Height
  ➢ Body mass index (BMI)
  ➢ Body fat percentage
  ➢ Blood pressure levels
  ➢ Disease status
Three scales of G x E interaction analysis

• SNP x E interaction analysis
  ➢ whether $p < 5 \times 10^{-8}$ (0.05/1,000,000)

• Gene x E interaction analysis
  ➢ whether $p < 2.5 \times 10^{-6}$ (0.05/20,000)

• GRS x E interaction analysis
  ➢ GRS: Genetic risk score
  ➢ whether $p < 0.05$ (0.05/1)
GRS: Genetic risk score

- A linear combination of effect alleles

\[ GRS_i = \sum_{j=1}^{L} \hat{\beta}_j G_{ij} \]

- Unweighted GRS (if all \( \hat{\beta}_j = 1 \))
- Weighted GRS (usually weighted by effect sizes)
The number of BMI-increasing alleles in 12 SNPs

European ancestry (78 %)
Asians (10 %)
Africans (2 %)
External genome-wide association studies (GWASs) may not be available, especially for non-European ethnicity.
97 BMI-associated SNPs ($p < 5 \times 10^{-8}$)


<table>
<thead>
<tr>
<th>In Taiwan Biobank ($n=18,424$)</th>
<th>BMI</th>
<th>Body fat %</th>
<th>Waist circumference</th>
<th>Hip circumference</th>
<th>Waist-to-hip ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of SNPs with $p &lt; 5 \times 10^{-8}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Number of SNPs with $p &lt; 0.05$</td>
<td>29</td>
<td>20</td>
<td>28</td>
<td>22</td>
<td>12</td>
</tr>
</tbody>
</table>
We need to build weights according to our data.
\[ g\{E(Y)\} = \beta_0 + \beta_{SNP,i}SNP_i + \beta_c Covariates \]

\[ g\{E(Y)\} = \gamma_0 + \gamma_{SNP,i}SNP_i + \gamma_c Covariates + \gamma_E E + \gamma_{Int,i}SNP_i \times E \]

Under \( H_0: \gamma_{Int,i} = 0 \), the maximum likelihood estimate \( \hat{\beta}_{SNP,i} \), is asymptotically independent to \( \hat{\gamma}_{Int,i} \)

Gene-based GxE interaction approach
Adjust for non-genetic covariates

\[ g\left[ E(Y_i) \right] = \alpha_0 + \alpha'X_i , \ i = 1, \ldots, n \]

Gender, age, smoking status, ancestry principal components

\[ \hat{\mu}_{0i} = \hat{\alpha}_0 + \hat{\alpha}'X_i \] (for continuous \( Y_i \)) or

\[ \hat{\mu}_{0i} = \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}'X_i)}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}'X_i)} \] (for binary \( Y_i \))
Covariate-adjusted phenotype

\[ \hat{\varepsilon}_i = Y_i - \hat{\mu}_{0i} \]

\[ g[E(\hat{\varepsilon}_i)] = \beta_0 + \sum_{j=1}^{L} \beta_j G_{ij} \]

0, 1, 2
Filtering stage
Ridge regression (RIDGE)

\[ \hat{\beta} = \underset{\beta}{\arg \min} \left[ \sum_{i=1}^{n} \left( \hat{e}_i - \beta_0 - \sum_{j=1}^{L} \beta_j G_{ij} \right)^2 + \lambda \sum_{j=1}^{L} \beta_j^2 \right] \]

LASSO (Least Absolute Shrinkage and Selection Operator)

\[ \hat{\beta} = \underset{\beta}{\arg \min} \left[ \sum_{i=1}^{n} \left( \hat{e}_i - \beta_0 - \sum_{j=1}^{L} \beta_j G_{ij} \right)^2 + \lambda \sum_{j=1}^{L} |\beta_j| \right] \]
**FIGURE 3.11.** Estimation picture for the lasso (left) and ridge regression (right). Shown are contours of the error and constraint functions. The solid blue areas are the constraint regions $|\beta_1| + |\beta_2| \leq t$ and $\beta_1^2 + \beta_2^2 \leq t^2$, respectively, while the red ellipses are the contours of the least squares error function.
ENET (Elastic net)

\[
\hat{\beta} = \argmin_{\beta} \left\{ \sum_{i=1}^{n} \left( \hat{e}_i - \beta_0 - \sum_{j=1}^{L} \beta_j G_{ij} \right)^2 + \lambda \sum_{j=1}^{L} \left[ \frac{1}{2} (1 - \alpha) \beta_j^2 + \alpha |\beta_j| \right] \right\}
\]

- \( \alpha = 0 \), RIDGE
- \( \alpha = 1 \), LASSO
- \( \alpha = 1/2 \), ENET

10-fold cross validation to select \( \lambda \) that leads to the minimum MSE (mean squared error)

\[
GRS_i = \sum_{j=1}^{L} \hat{\beta}_j G_{ij}
\]

0, 1, 2
Testing stage
\[ g[E(Y_i)] = \gamma_0 + \gamma_G GRS_i + \gamma_E E_i + \gamma_{Int} GRS_i \times E_i + \gamma_c' X_i, \ i = 1, \ldots, n \]

By testing \( H_0: \gamma_{Int} = 0 \) vs. \( H_1: \gamma_{Int} \neq 0 \), we evaluate whether GxE exists.

If \( \hat{\gamma}_{Int} > 0 \), E exacerbates the adverse influence of a candidate gene.

If \( \hat{\gamma}_{Int} < 0 \), E attenuates the adverse influence of a candidate gene.
Competing methods
SBERIA (Jiao et al. 2013, Genet. Epidemiol.)
(Set-Based gene-Environment Interaction test)

\[ g[E(Y_i)] = \beta_0 + \beta_j G_{ij} + \beta'_c X_i \]

\[ H_0: \beta_j = 0 \text{ vs. } H_1: \beta_j \neq 0 \text{ One SNP at a time} \]

\[ GRS_i = \sum_{j=1}^{L} [I(p_j < 0.1) \text{sign}(\hat{\beta}_j)]G_{ij} \]

\[ g[E(Y_i)] = \gamma_0 + \sum_{j=1}^{L} \gamma_{Gj} G_{ij} + \gamma_E E_i + \gamma_{\text{int}} GRS_i \times E_i + \gamma_c' X_i \]
iSKAT (Lin X.Y. et al. 2016, Biometrics)
interaction Sequence Kernel Association Test

\[ g[E(Y_i)] = \delta_0 + \sum_{j=1}^{L} \delta_{G_j}G_{ij} + \delta_{E}E_i + \sum_{j=1}^{L} \delta_{Int_j}G_{ij}E_i + \delta_C'X_i \]

Assuming \( \delta_{Int_j} \)s \((j = 1, \ldots, L)\) follow a distribution with a mean of 0 and a variance of \( \tau \)

\( H_0: \tau = 0 \) vs. \( H_1: \tau > 0 \)
Adaptive Combination of Bayes Factors (ADABF) Method  

\[ g[E(Y_i)] = \delta_0 + \delta_{G_j}G_{ij} + \delta_E E_i + \delta_{Int_j}G_{ij}E_i + \delta'_c X_i \]

\[ H_0: \delta_{Int_j} = 0 \text{ vs. } H_1: \delta_{Int_j} \neq 0 \]
Bayes factor

\[
BF = \frac{\Pr(Data \mid H_1)}{\Pr(Data \mid H_0)}
\]

- \(BF\) quantifies the ‘relative’ evidence in favor of \(H_1\).
Sort \( BF_{(1)} \geq BF_{(2)} \geq \cdots \geq BF_{(L)} \)

Significance score \( S_k = \sum_{l=1}^{k} \log \left( BF_{(l)} \right), \quad k = 1, \cdots, L \)

Summing the largest \( k \) \( \log(BF) \), \( k = 1, \cdots, L \)
• The significance scores will be compared with their counterparts from resampling replicates (under $H_0$)

• The R source code can be downloaded from [http://homepage.ntu.edu.tw/~linwy/ADABFEPoly.html](http://homepage.ntu.edu.tw/~linwy/ADABFEPoly.html)
Simulation Study
Taiwan Biobank

• 18,424 unrelated subjects (9,093 males and 9,331 females)

• Three genes were drawn for simulations:
  - $TNNT3$ (48 SNPs)
  - $RFX3$ (95 SNPs)
  - $FTO$ (242 SNPs)
Power evaluation

\[ Y_i = \sum_{d=1}^{4} \beta_{G_d} G_{id} + \beta_E E_i + \sum_{d=1}^{D} \beta_{Int_d} G_{id} E_i + \varepsilon_i \]

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$E$</th>
<th>$\beta_{G_1}$</th>
<th>$\beta_{G_2}$</th>
<th>$\beta_{G_3}$</th>
<th>$\beta_{G_4}$</th>
<th>$\beta_{Int_1}$</th>
<th>$\beta_{Int_2}$</th>
<th>$\beta_{Int_3}$</th>
<th>$\beta_{Int_4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Exacerbation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2 Attenuation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Power given a significance level of 0.05, for continuous traits and $P(E = 1) = 0.2$
In the filtering stage

\[ Y_i = \sum_{d=1}^{4} \beta_{G_d} G_{id} + \beta_E E_i + \sum_{d=1}^{D} \beta_{Int_d} G_{id} E_i + \varepsilon_i \]

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$E$</th>
<th>$\beta_{G_1}$</th>
<th>$\beta_{G_2}$</th>
<th>$\beta_{G_3}$</th>
<th>$\beta_{G_4}$</th>
<th>$\beta_{Int_1}$</th>
<th>$\beta_{Int_2}$</th>
<th>$\beta_{Int_3}$</th>
<th>$\beta_{Int_4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Exacerbation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2 Attenuation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Power evaluation

\[ Y_i = \sum_{d=1}^{4} \beta_{Gd} G_{id} + \beta_E E_i + \sum_{d=1}^{D} \beta_{Intd} G_{id} E_i + \epsilon_i \]

<table>
<thead>
<tr>
<th>Scenario</th>
<th>( E )</th>
<th>( \beta_{G1} )</th>
<th>( \beta_{G2} )</th>
<th>( \beta_{G3} )</th>
<th>( \beta_{G4} )</th>
<th>( \beta_{Int1} )</th>
<th>( \beta_{Int2} )</th>
<th>( \beta_{Int3} )</th>
<th>( \beta_{Int4} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Exacerbation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4 Attenuation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
3 Exacerbation

4 Attenuation

Sign-misspecifications (in %)

RIDGE
ENET
LASSO
SBERIA
Power evaluation

\[ Y_i = \sum_{d=1}^{4} \beta_{G_d} G_{id} + \beta_E E_i + \sum_{d=1}^{D} \beta_{\text{Int}_d} G_{id} E_i + \epsilon_i \]

<table>
<thead>
<tr>
<th>Scenario</th>
<th>( E )</th>
<th>( \beta_{G_1} )</th>
<th>( \beta_{G_2} )</th>
<th>( \beta_{G_3} )</th>
<th>( \beta_{G_4} )</th>
<th>( \beta_{\text{Int}_1} )</th>
<th>( \beta_{\text{Int}_2} )</th>
<th>( \beta_{\text{Int}_3} )</th>
<th>( \beta_{\text{Int}_4} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>cross-over</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

40
5 Cross-over
Application to the Taiwan Biobank (TWB)
Taiwan Biobank: Since October 2012
Taiwan residents aged 30 to 70 years

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total, n (%)</strong></td>
<td>18,424</td>
<td>9,093</td>
<td>9,331</td>
</tr>
<tr>
<td><strong>Age (years), mean (s.d.)</strong></td>
<td>48.9 (11.0)</td>
<td>49.0 (11.0)</td>
<td>48.9 (10.9)</td>
</tr>
<tr>
<td><strong>Smoking, n (%)</strong></td>
<td>2,134 (11.6)</td>
<td>1,882 (20.7)</td>
<td>252 (2.7)</td>
</tr>
<tr>
<td><strong>Drinking, n (%)</strong></td>
<td>1,345 (7.3)</td>
<td>1,178 (13.0)</td>
<td>167 (1.8)</td>
</tr>
<tr>
<td><strong>Regular exercise, n (%)</strong></td>
<td>7,652 (41.5)</td>
<td>3,896 (42.8)</td>
<td>3,756 (40.3)</td>
</tr>
</tbody>
</table>
**FTO x exercise interaction on obesity**

- The *fat mass and obesity-associated (FTO)* gene
- Chromosome 16 (53,737,875 - 54,148,379)
- 242 SNPs (minor allele frequency > 1%)
- Regular exercise: 30 minutes of exercise 3 times a week
- Covariates: sex, age, educational attainment, drinking status, smoking status, and the first 10 ancestry principal components (PCs).
Exercise attenuates the adverse influence of *FTO*.

<table>
<thead>
<tr>
<th>Trait</th>
<th>RIDGE</th>
<th>ENET</th>
<th>LASSO</th>
<th>SBERIA</th>
<th>iSKAT</th>
<th>ADABF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BMI (kg/m²)</strong></td>
<td>$\hat{\gamma}_{\text{Int}}$</td>
<td>-0.1743</td>
<td>-0.0821</td>
<td>-0.0964</td>
<td>-0.1482</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Int}}$</td>
<td><strong>0.0009</strong></td>
<td>0.1192</td>
<td>0.0671</td>
<td><strong>0.0067</strong></td>
<td>0.2043</td>
</tr>
<tr>
<td><strong>Body fat %</strong></td>
<td>$\hat{\gamma}_{\text{Int}}$</td>
<td>-0.2661</td>
<td>-0.2069</td>
<td>-0.2081</td>
<td>-0.2259</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Int}}$</td>
<td><strong>0.0031</strong></td>
<td>0.0212</td>
<td><strong>0.0205</strong></td>
<td><strong>0.0160</strong></td>
<td>0.2430</td>
</tr>
<tr>
<td><strong>Waist circumference</strong></td>
<td>$\hat{\gamma}_{\text{Int}}$</td>
<td>-0.3854</td>
<td>-0.3719</td>
<td>-0.3760</td>
<td>-0.2786</td>
<td></td>
</tr>
<tr>
<td>(cm)</td>
<td>$P_{\text{Int}}$</td>
<td><strong>0.0052</strong></td>
<td><strong>0.0069</strong></td>
<td><strong>0.0063</strong></td>
<td>0.0512</td>
<td>0.5369</td>
</tr>
<tr>
<td><strong>Hip circumference</strong></td>
<td>$\hat{\gamma}_{\text{Int}}$</td>
<td>-0.3868</td>
<td>-0.3286</td>
<td>-0.3291</td>
<td>-0.2902</td>
<td></td>
</tr>
<tr>
<td>(cm)</td>
<td>$P_{\text{Int}}$</td>
<td><strong>0.0001</strong></td>
<td><strong>0.0011</strong></td>
<td><strong>0.0011</strong></td>
<td><strong>0.0055</strong></td>
<td>0.5061</td>
</tr>
</tbody>
</table>
(A) BMI-GRS effect on BMI

$P_{Int}=0.0009$

- All subjects: $18,424$
- Non-exercisers: $10,764$
- Exercisers: $7,652$

8 subjects did not respond to this question
FGF5 x WHR interaction on blood pressure

• The fibroblast growth factor 5 (FGF5) gene
• Chromosome 4 (81,187,742 - 81,212,171)
• 38 SNPs (minor allele frequency > 1%)
• WHR: waist-hip ratio
• Covariates: sex, age, drinking status, smoking status, and the first 10 ancestry PCs.
<table>
<thead>
<tr>
<th>Trait</th>
<th>RIDGE</th>
<th>ENET</th>
<th>LASSO</th>
<th>SBERIA</th>
<th>iSKAT</th>
<th>ADABF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DBP (mmHg)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>舒張壓</td>
<td>(\hat{\gamma}_{\text{int}})</td>
<td>0.2419</td>
<td>0.1980</td>
<td>0.2141</td>
<td>0.2378</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(P_{\text{int}})</td>
<td>0.0013</td>
<td>0.0082</td>
<td>0.0042</td>
<td>0.0014</td>
<td>0.0154</td>
</tr>
<tr>
<td><strong>SBP (mmHg)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>收縮壓</td>
<td>(\hat{\gamma}_{\text{int}})</td>
<td>0.3396</td>
<td>0.3548</td>
<td>0.3551</td>
<td>0.3261</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(P_{\text{int}})</td>
<td>0.0027</td>
<td>0.0017</td>
<td>0.0017</td>
<td>0.0039</td>
<td>0.0482</td>
</tr>
</tbody>
</table>

The *FGF5* gene has a stronger effect on blood pressure in Han Chinese with a higher waist-hip ratio.
Summary

• Not only provides a \( p \)-value for a GxE test
• But also knows how E modifies the adverse effect of a gene
• We look forward to performing genome-wide GxE analyses on a larger TWB cohort
Thanks for your attention!

http://homepage.ntu.edu.tw/~linwy/