Genome-wide gene-environment interactions

Statistical methods and their application in asthma research

Markus Ege, MD MA MPH Dr. von Hauner Children's Hospital Ludwig Maximilian University Munich

## What are G\*E interactions?



genotype (A = interacting allele)

How to model genome wide interaction analyses?

- 1. Full interaction model
- 2. Case-only approach
- 3. Weighted combination of the above (Mukherjee & Chatterjee 2008)
- 4. Two-step approach (modified following Murcray et al. 2009)

### The Interaction Odds Ratio



## $\beta_{G^*E} = \beta_{GD|E^+} - \beta_{GD|E^-} = \beta_{GE|D^+} - \beta_{GE|D^-}$

### 1. Full interaction model

# $\beta_{G^*E} = \beta_{GE|D^+} - \beta_{GE|D^-}$ Model: D = G + E + G\*E

logistic regression with interaction term

### 2. Case-only approach

# $\beta_{G^*E} = \beta_{GE|D^+} - \beta_{GE|D^-} \operatorname{since} \beta_{GE|D^-} = 0$ Modell: G = E for D+

- limited to informative subjects, but:
- severe bias possible, if Mendelian randomisation does not hold

### 3. Weighted combination

$$\beta_{G^*E} = \frac{\sigma^2_{G^*E} \cdot \beta_{GE|D^+} + \beta^2_{GE|D^-} \cdot \beta_{G^*E}}{\sigma^2_{G^*E} + \beta^2_{GE|D^-}}$$

following Mukherjee & Chatterjee. Biometrics 2008; 64:685

if  $\sigma^2_{G^*E}$  ↑↑ → case-only if  $\beta^2_{GE|D}$  >>0 → full interaction Disadvantage of previous approaches: strong correction for multiple testing necessary

remedy: perform fewer tests



Ege et al. J Allergy Clin Immunol 2011



Ege et al. J Allergy Clin Immunol 2011

### Advantage of two steps

correction for multiple testing less extreme because:

- <u>fewer</u> tests performed at step 2
- tests at step 1 irrelevant for step 2
- steps statistically <u>independent</u>

### Simulation study

Parameter	Range (steps)
prevalence of disease (P <sub>D</sub> )	0.05, 0.1 (0.1) 0.6
number of control subjects (N*[1-D])	1,000 (1,000) 9,000
prevalence of exposure (P <sub>E</sub> )	0.1 (0.1) 0.6
prevalence of genotype (P <sub>G</sub> )	0.1 (0.1) 0.6
population attributable risk fraction of exposure (PARF <sub>E</sub> )	0.1 (0.1) 0.5
population attributable risk fraction of exposure (PARF <sub>G</sub> )	0 (0.05) 0.25
interaction odds ratio (OR <sub>G*E</sub> )	1.05, 1.1 (0.1) 2.0
OR in controls (OR <sub>EG D</sub> )	$0.8^{\pm 1}$ , $0.9^{\pm 1}$ , $0.95^{\pm 1}$ 1.0
proportion of corrected alpha level allocated to step 1 ( $\psi$ )	0 (0.1) 1

### Logistic regression models

(i)	$D = G + E + G^*E$	full interaction model
(ii)	E = G in D+	gene-environment association in cases
(iii)	E = G in D–	gene-environment association in controls or non-cases
(iv)	D = G in E+	gene-disease association in exposed subjects
(v)	D = G in E–	gene-disease association in unexposed subjects
(vi)	E = G	gene-environment association in all subjects combined
(vii)	D = G	gene-disease association in all subjects combined

### Number of regression models

### Alternative hypothesis

- = interaction present
- 7 regression models
- 486 scenarios
- 1,000 iterations
- 3,402,000 models

### Null hypothesis

- = interaction absent
- 7 regression models
- 252 scenarios
- 1,000,000 iterations
- 1,764,000,000 models
  - → HLRB2, Linux-Cluster, SuperMUC

### Parallelization

- Performed in R
- Package 'multicore' (256 cores) in combination with package 'survey'
- Splitting data in chunks, calling many R instances
- Package 'pbdMPI'

### False Positive Rate

**Population setting** 

**Case-control setting** 



Ege & Strachan, Eur J Epidemiol 2013

### **True Positive Rate**



### Real world data

- Childhood onset asthma (prevalence=10%)
- Major Public Health concern
- ≈ 50% risk reduction in children growing up on farms ("experiment by nature")
- Gene-environment interaction previously described for innate immunity receptors
- Gene-environment interaction at genomewide scale?

### GWIS for childhood-onset asthma



Ege et al. J Allergy Clin Immunol 2011

## **Biologic function?**



#### Pacheco et al. J Neuroimmunol 2007

### Conclusions

- Genome-wide interaction analyses valid
- High statistical power achieved
- Few interesting hits, but at low allele frequencies
- Mostly negative findings
- Gene-environment interaction on different level, e.g. gene-expression?







European Research Council Established by the European Commission

## Thank you!

