Testing gene-environment interactions without measuring the environment

Jiacheng Miao[1], Yuchang Wu[1,2], Kunling Huang[1], Zijie Zhao[2], Zheng Ni[1], Qiongshi Lu[1,2]

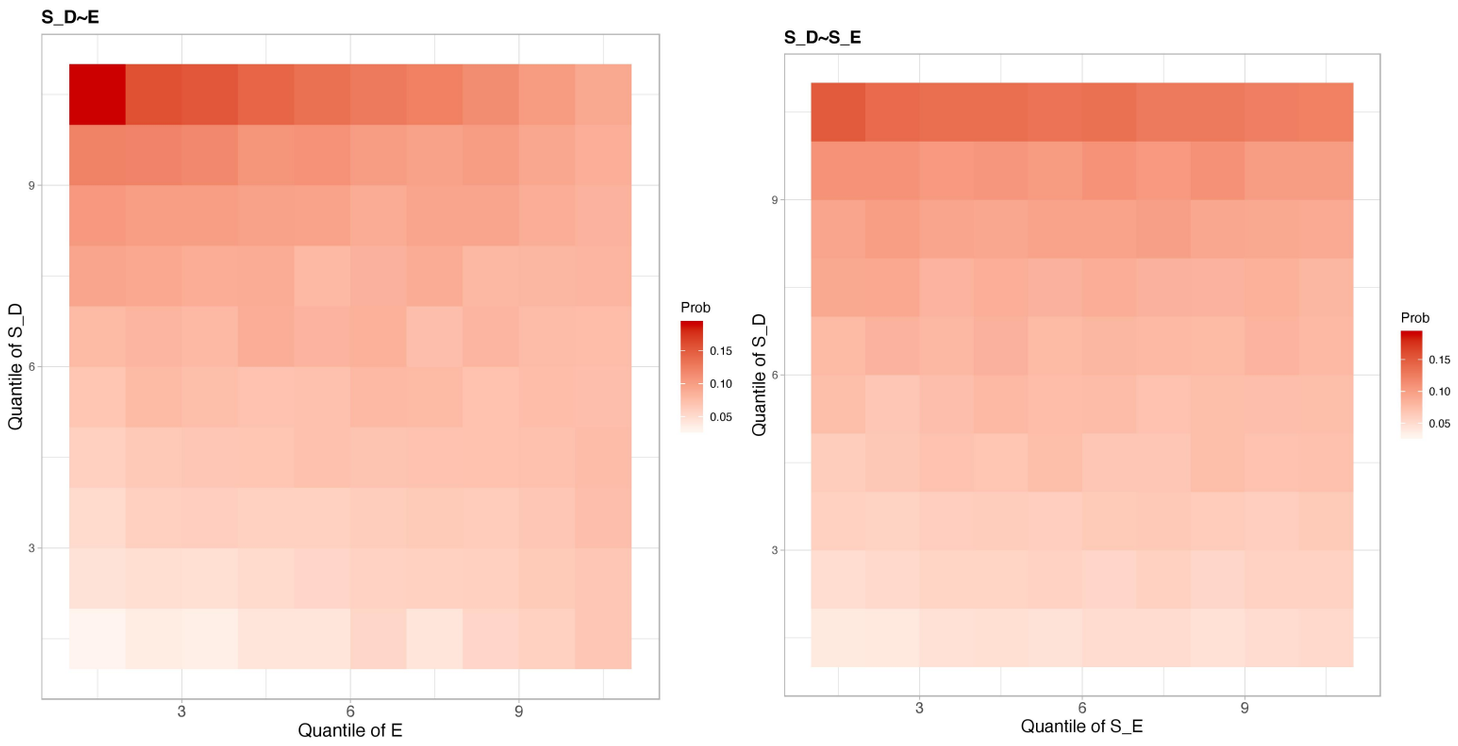
[1] Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA 53706
[2] Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA 53706

LU LABORATORY
statistical genetics & genome informatics

Motivation

- Gene-environment (GxE) interaction studies have only had limited success to date in part due to the lack of large datasets that provide all the required components in a successful GxE study :
    - Genome-wide genetic data
    - Robust measurements of environmental risk factors
    - A sufficient sample size for the disease

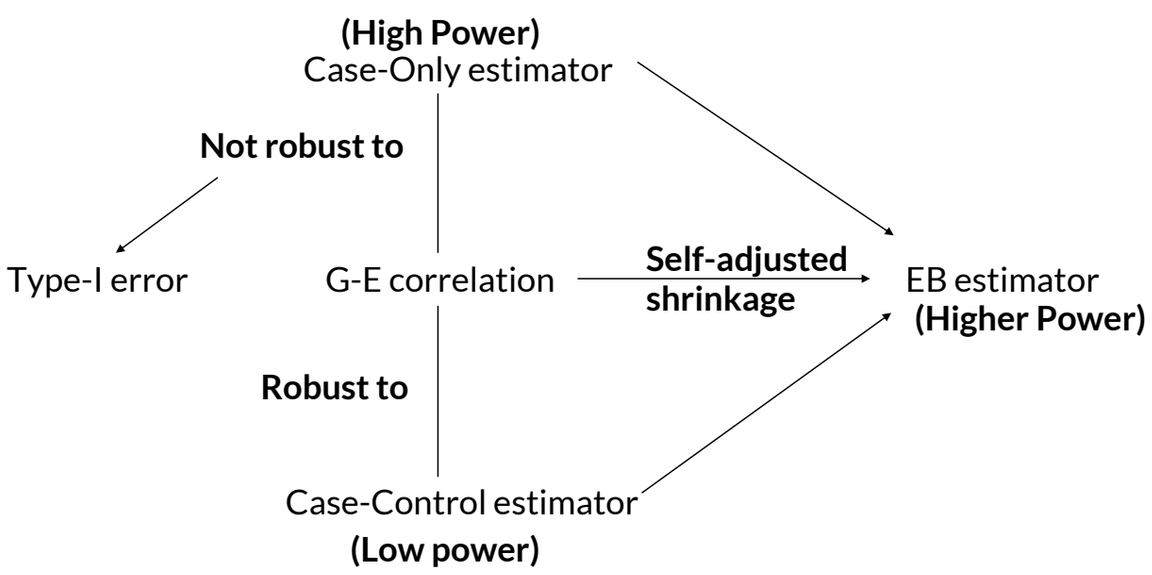
- In our work, we seek a solution for a seemingly impossible problem: can we test GxE interactions without measuring the environment ?

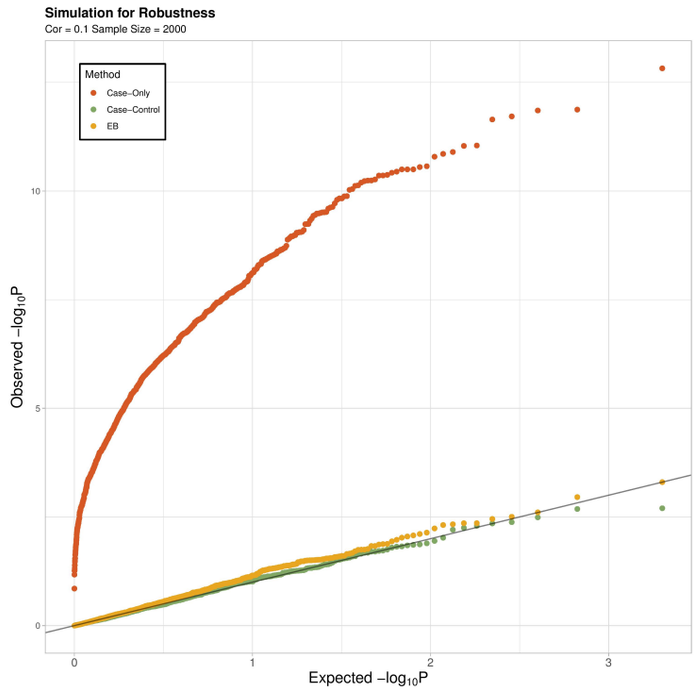


Methodology

- We use **polygenic score (PGS)** as a proxy for the E factor in GxE analysis, and this approach can be applied to diverse types of GxE study designs:
  - Quantitative trait GxE analysis
  - Binary trait GxE analysis
    I. Case-only analysis
    II. Case-Control analysis
    III. Cohort analysis

- **Empirical Bayes approach** can provide GxE effect estimates that are robust to G-E correlations





Dataset Description

- In order to investigate the possible GxE interactions for autism spectrum disorder (ASD), we obtain data from Autism Genome project (AGP), Simons Simplex Collection (SSC), and Simons Foundation Powering Autism Research for Knowledge (SPARK).

Probands				
Dataset		Overall	Male	Female
SPARK	trios	3823	3037	786
	duos	1369	1081	288
AGP		2188	1895	293
SSC	1Mv1	228	197	31
	1Mv3	840	728	112
	Omni2.5	726	627	99
Whole		9174	7565	1609

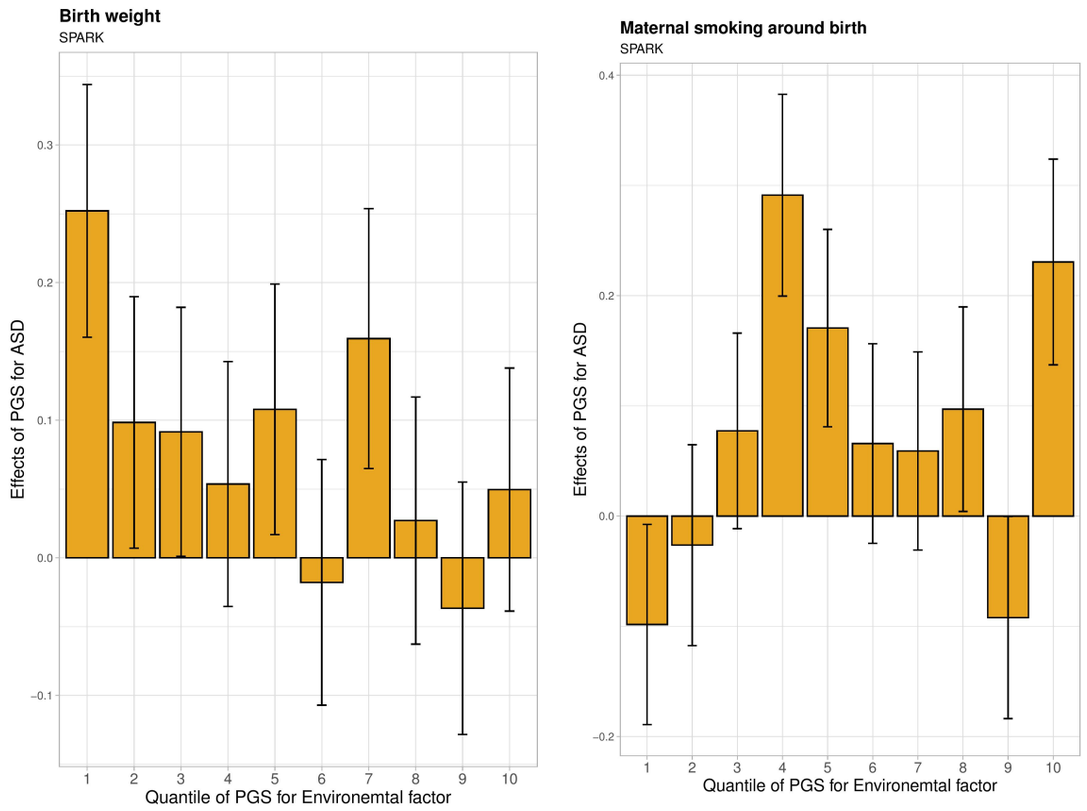
Siblings				
Dataset		Overall	Male	Female
SPARK	trios	1813	946	867
SSC	1Mv1	202	102	100
	1Mv3	634	288	346
	Omni2.5	596	283	313
Whole		3245	1619	1626

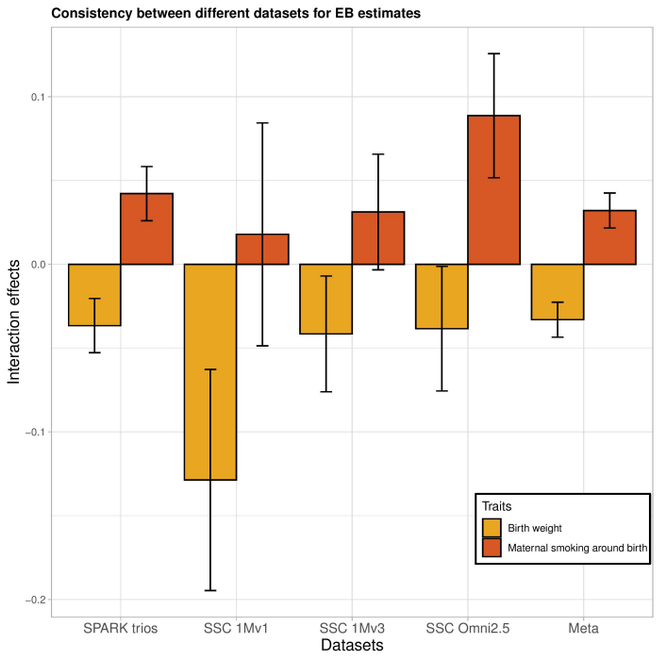
- We select four possible candidate environmental risk factor considering their temporal relationship with ASD :
  - Birth weight
  - Maternal smoking around birth
  - Adopted as a child
  - Breastfed as a baby
- The polygenic score are generated based on:
  - ASD: iPSYCH
  - Environmental: Ben's UK Biobank summary stats

Main findings

- **Higher birth weight** will buffer genetic risk for ASD while **maternal smoking around birth** will amplify the risk.

Traits	Effects	SE	P-value
Birth weight	-0.033	0.0104	1.56E-3
Breastfed as a baby	-0.020	0.0104	0.053
Maternal smoking around birth	0.032	0.0104	2.12E-3
Adopted as a child	-0.012	0.0104	0.243





Conclusion& Future direction

- We propose a novel and robust statistical framework to test GxE interactions by using the polygenic score as a proxy for the environmental factors.

- Our novel approach makes it possible to perform rigorous GxE inference on large GWAS datasets even if the environmental factor is not directly measured.

- We have performed some analysis in the maternal environmental factor, but has not been finished yet.

- We also consider using our framework to perform GxE interactions analysis on some other disease such as Alzheimer's disease

Thank you!

Jiacheng Miao

jmiao24@wisc.edu

Qiongshi Lu

qlu@biostat.wisc.edu

LU LABORATORY

statistical genetics & genome informatics