Nothing to disclose

Distinct explanations underlie gene-environment interactions in the UKBiobank

Arun Durvasula, PhD
Assistant Professor
Center for Genetic Epidemiology
University of Southern California

Keck School of Medicine of USC
Center for Genetic Epidemiology

Outline

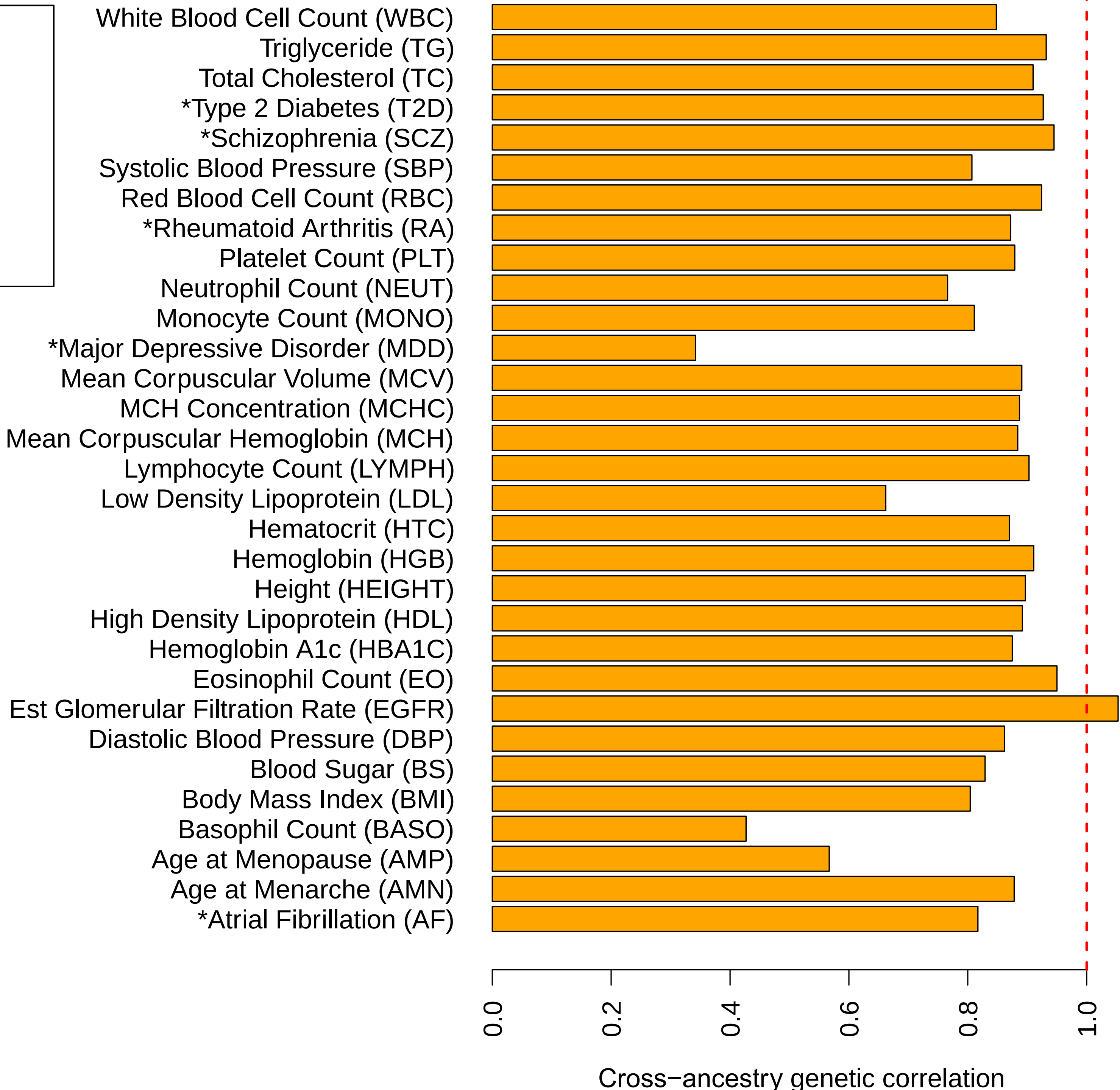
1. Background on gene-environment interactions
2. Methods
3. Results
4. Summary

Outline

1. **Background on gene-environment interactions**
2. Methods
3. Results
4. Summary

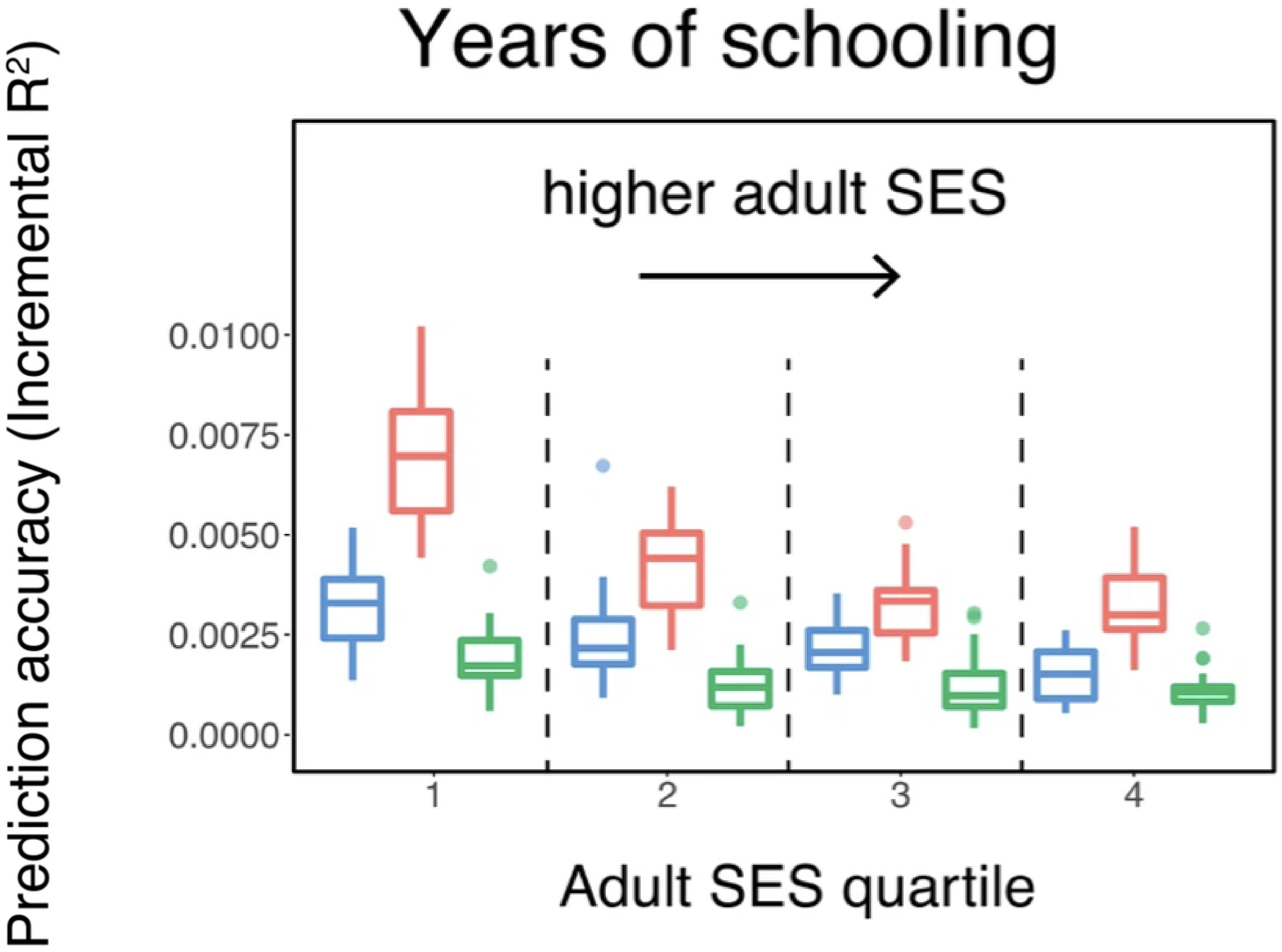
Genetic effects across the genome may exhibit context dependence

- European-East Asian cohort genetic correlation is less than 1 (average: **85%** SE: 1%) across a wide range of traits
- Polygenic GxE is one possible explanation



Data replotted from Shi et al 2021 Nat Commun

Polygenic GxE may explain variable prediction accuracy of polygenic scores within an ancestry



- Incremental R^2 changes as a function of E variables
- Could be due to interactions, among other explanations

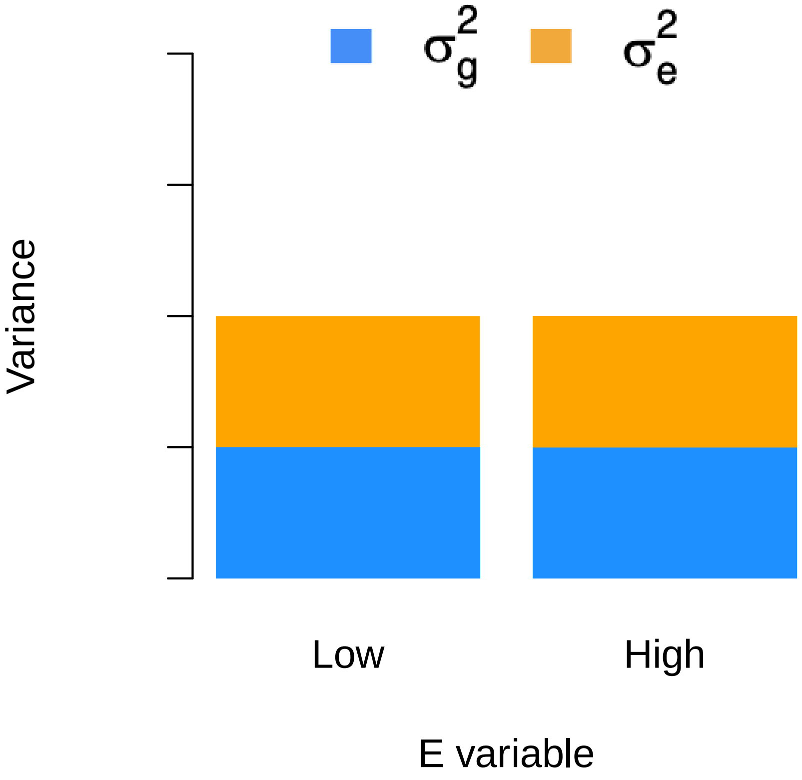
Mostafavi et al 2020 eLife

Major questions remain about the contribution of GxE to disease heritability

- How widespread are context dependent genetic effects?
- What is the contribution of polygenic GxE to disease heritability?

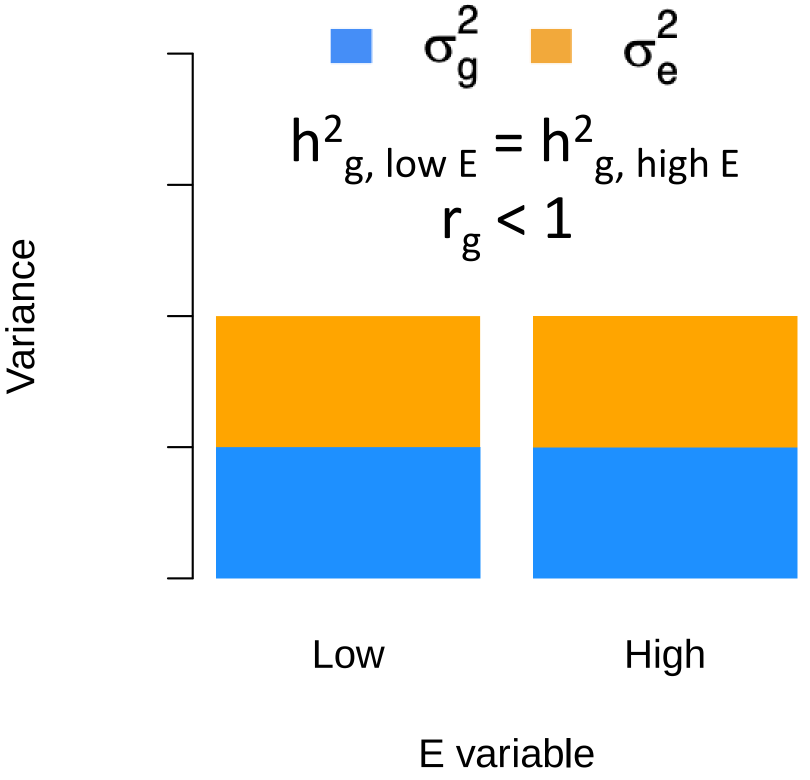
GxE can arise through different scenarios

Imperfect genetic correlation



GxE can arise through different scenarios

Imperfect genetic correlation

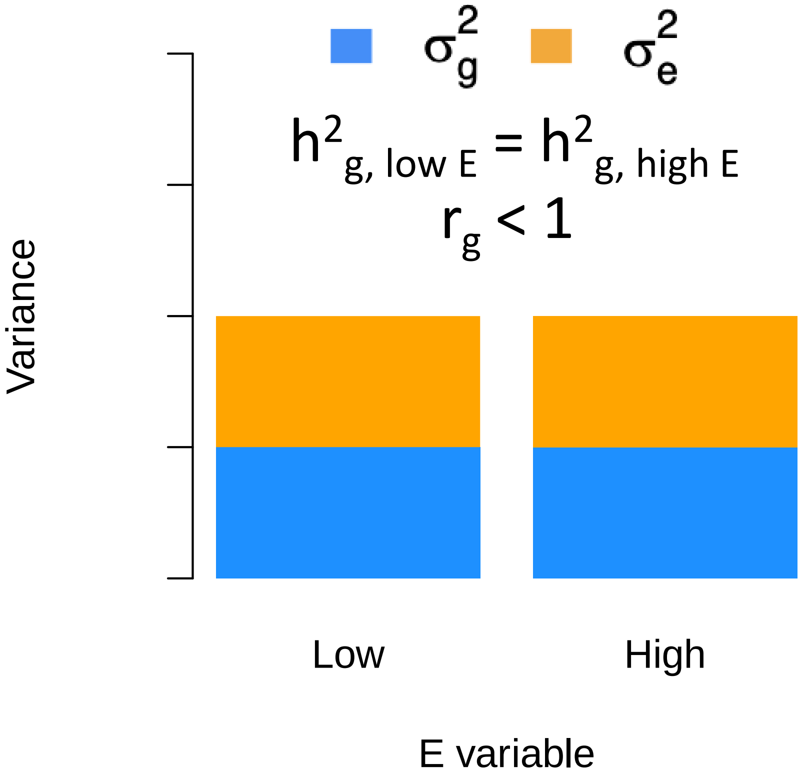


$h^2_{g,\ \text{low E}} = h^2_{g,\ \text{high E}}$

$r_g < 1$

GxE can arise through different scenarios

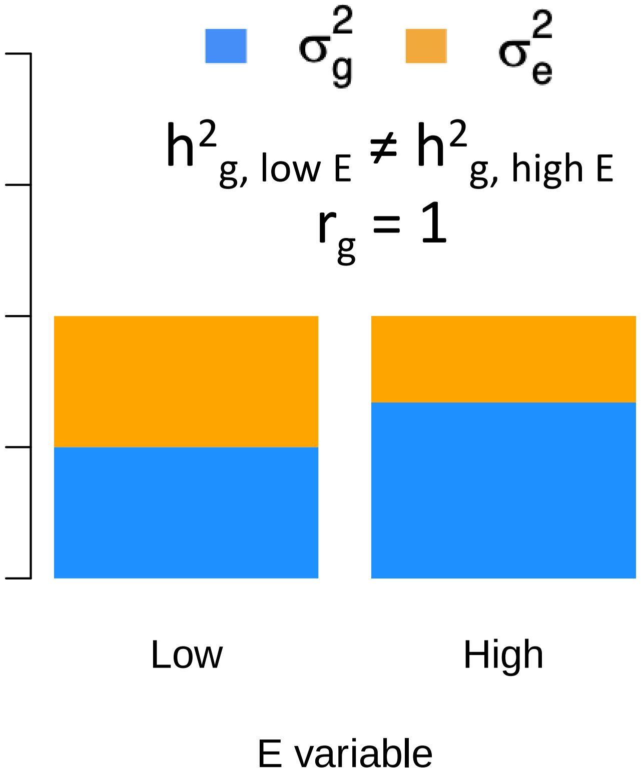
Imperfect genetic correlation



$h^2_{g,\ \text{low E}} = h^2_{g,\ \text{high E}}$

$r_g < 1$

Varying genetic variance

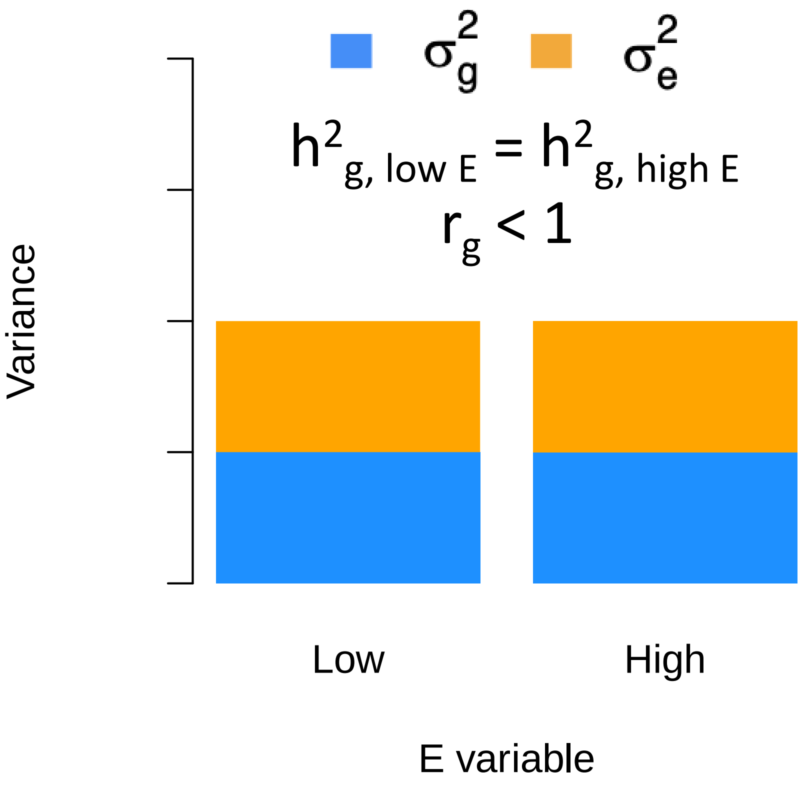


$h^2_{g,\ \text{low E}} \neq h^2_{g,\ \text{high E}}$

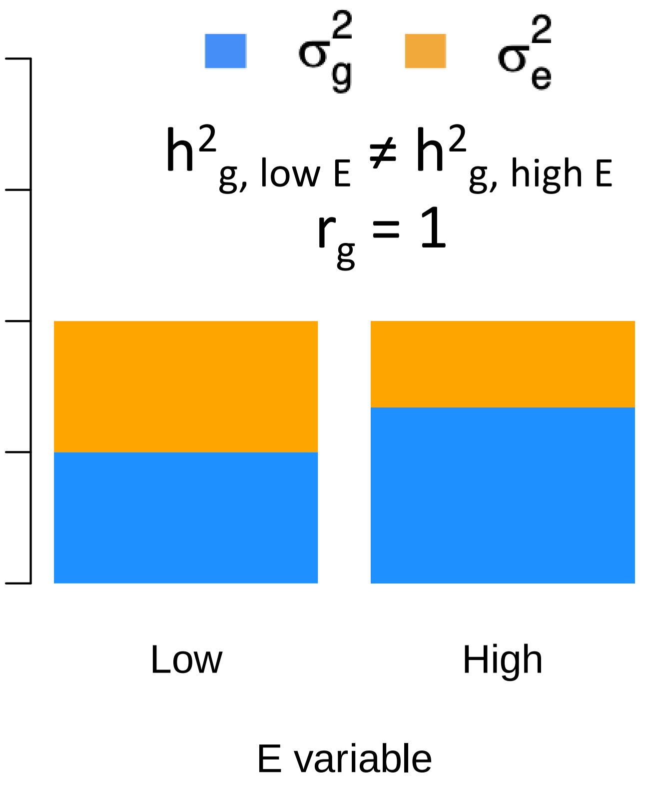
$r_g = 1$

GxE can arise through different scenarios

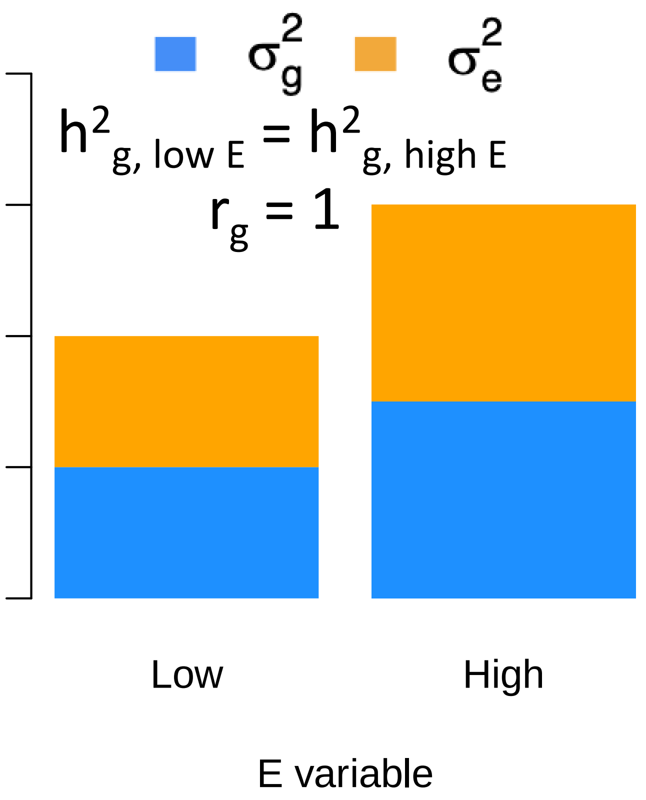
Imperfect genetic correlation



Varying genetic variance

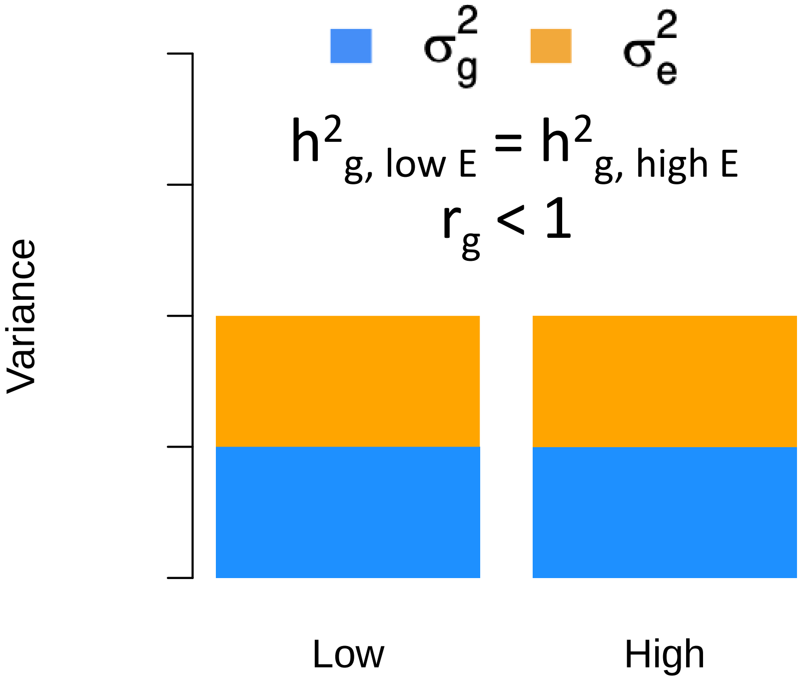


Proportional amplification



GxE can arise through different scenarios

Imperfect genetic correlation

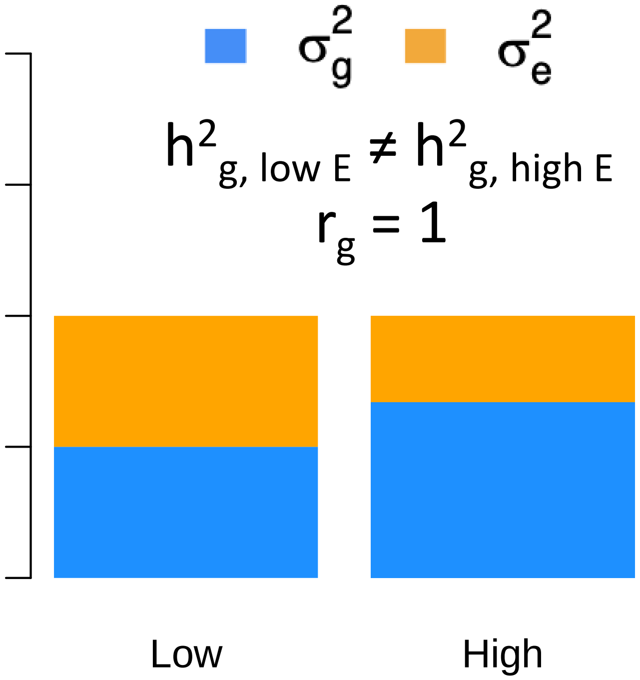


$h^2_{g,\ \text{low E}} = h^2_{g,\ \text{high E}}$

$r_g < 1$

Locus dependent GxE

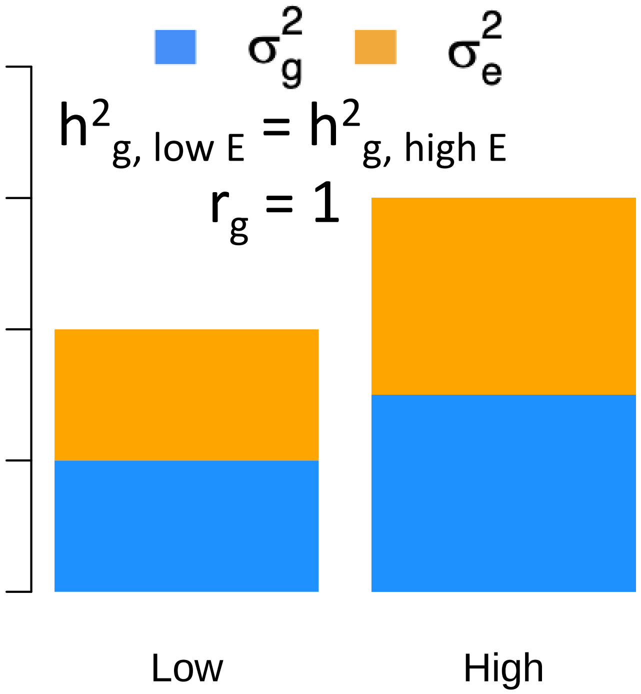
Varying genetic variance



$h^2_{g,\ \text{low E}} \neq h^2_{g,\ \text{high E}}$

$r_g = 1$

Proportional amplification



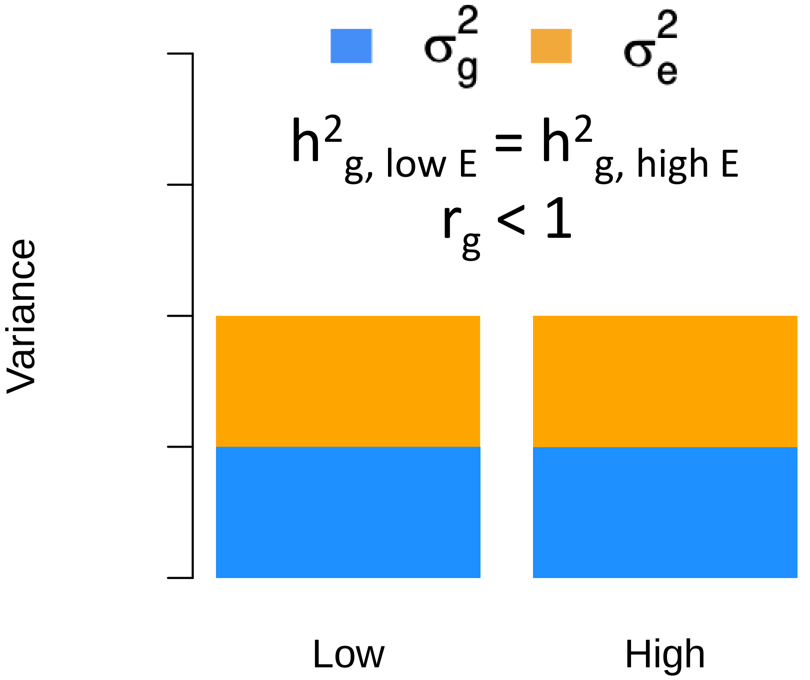
$h^2_{g,\ \text{low E}} = h^2_{g,\ \text{high E}}$

$r_g = 1$

Non-locus dependent GxE

GxE can arise through different scenarios

Imperfect genetic correlation

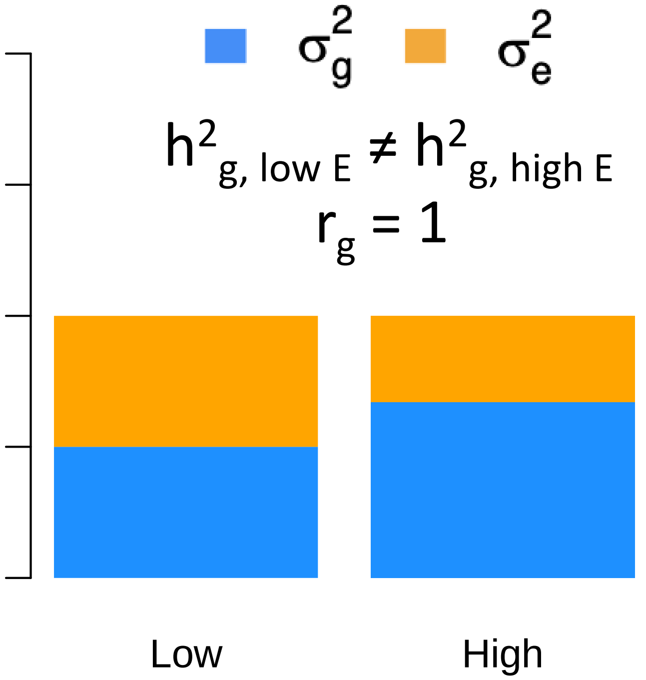


$h^2_{g,\text{ low E}} = h^2_{g,\text{ high E}}$

$r_g < 1$

Locus dependent GxE:
Effect sizes are multiplied by a SNP-specific factor

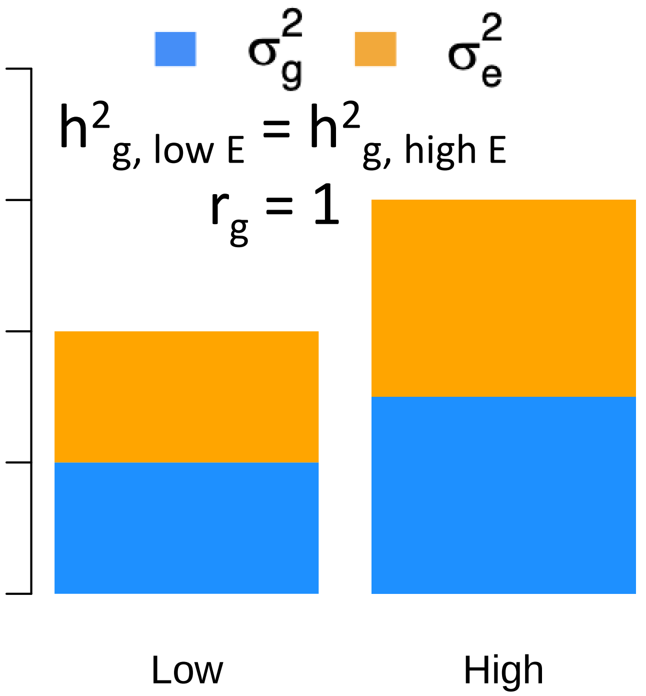
Varying genetic variance



$h^2_{g,\text{ low E}} \neq h^2_{g,\text{ high E}}$

$r_g = 1$

Proportional amplification



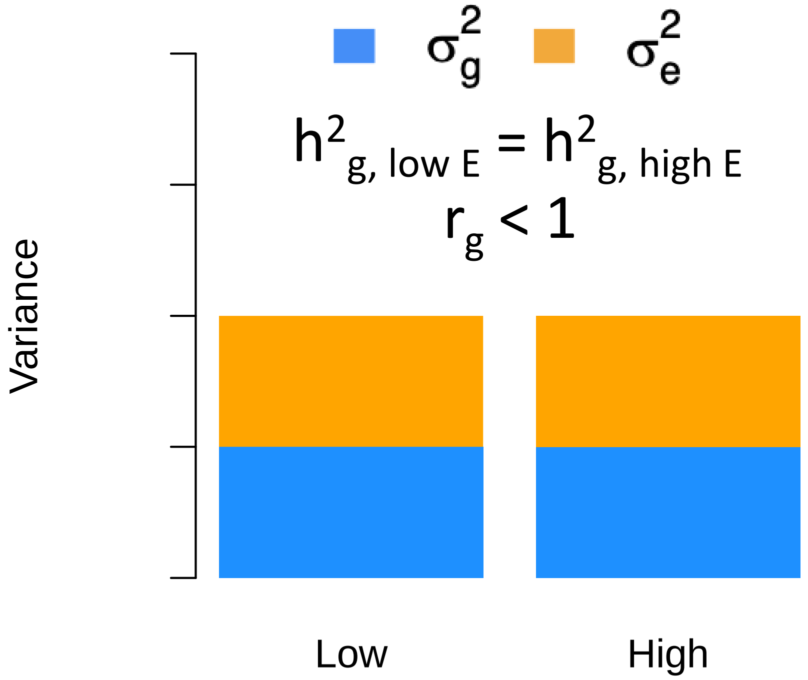
$h^2_{g,\text{ low E}} = h^2_{g,\text{ high E}}$

$r_g = 1$

Non-locus dependent GxE

GxE can arise through different scenarios

Imperfect genetic correlation

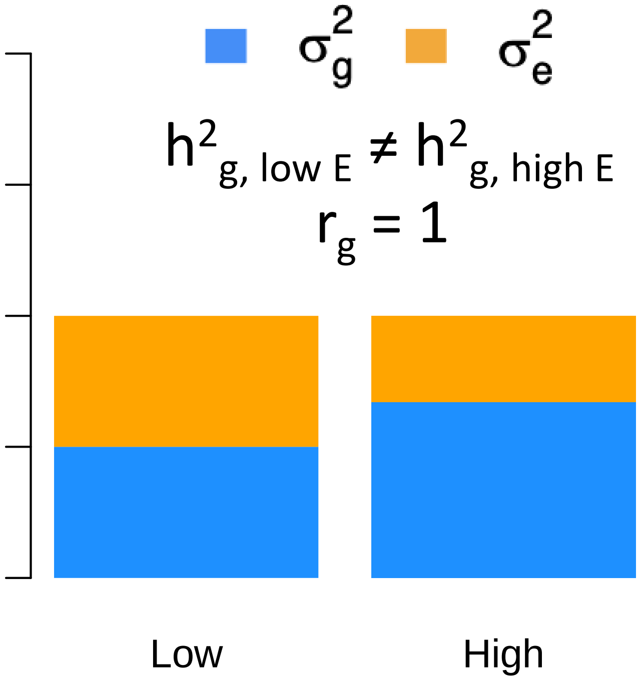


$h^2_{g,\text{ low E}} = h^2_{g,\text{ high E}}$

$r_g < 1$

Locus dependent GxE:
Effect sizes are multiplied by a SNP-specific factor

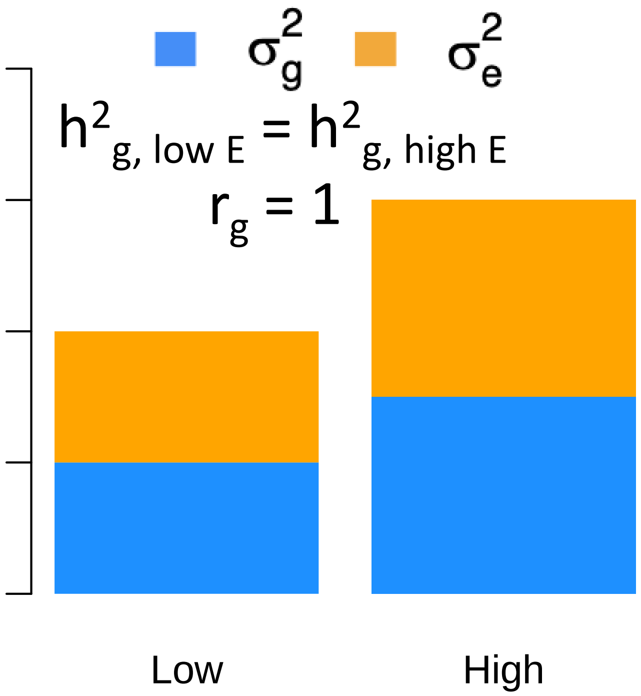
Varying genetic variance



$h^2_{g,\text{ low E}} \neq h^2_{g,\text{ high E}}$

$r_g = 1$

Proportional amplification



$h^2_{g,\text{ low E}} = h^2_{g,\text{ high E}}$

$r_g = 1$

Non-locus dependent GxE:
Effect sizes or phenotypic variance multiplied by a constant factor across all SNPs

Outline

1. Background on gene-environment interactions
2. **Methods**
3. Results
4. Summary

We can distinguish between polygenic GxE scenarios with three metrics

Scenarios / Metrics	Imperfect genetic correlation	Varying genetic variance	Proportional amplification
Genetic correlation	✓	✗	✗

We can distinguish between polygenic GxE scenarios with three metrics

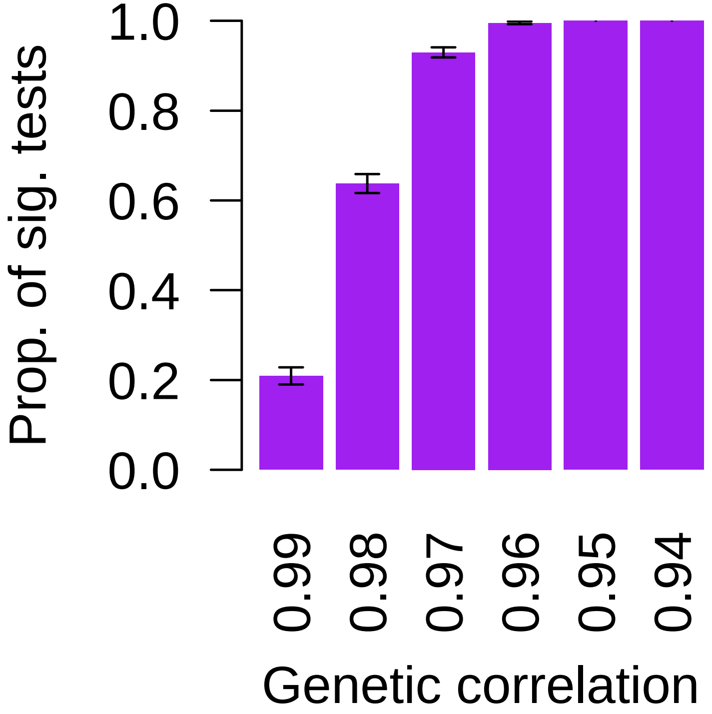
Scenarios Metrics	Imperfect genetic correlation	Varying genetic variance	Proportional amplification
Genetic correlation	✓	✗	✗
PRS x E regression	✗	✓	✓

We can distinguish between polygenic GxE scenarios with three metrics

Scenarios / Metrics	<u>Imperfect genetic correlation</u>	<u>Varying genetic variance</u>	<u>Proportional amplification</u>
Genetic correlation	✅	❌	❌
PRS x E regression	❌	✅	✅
h^2 by E	❌	✅	❌

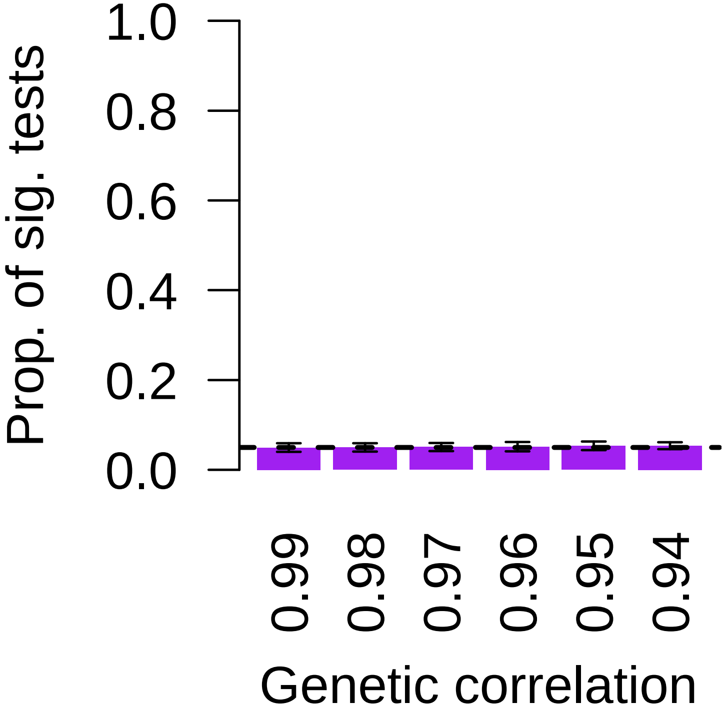
LDSC genetic correlation is well-powered to detect $r_g < 97\%$ across E bins in Scenario 1

Genetic correlation across E bins



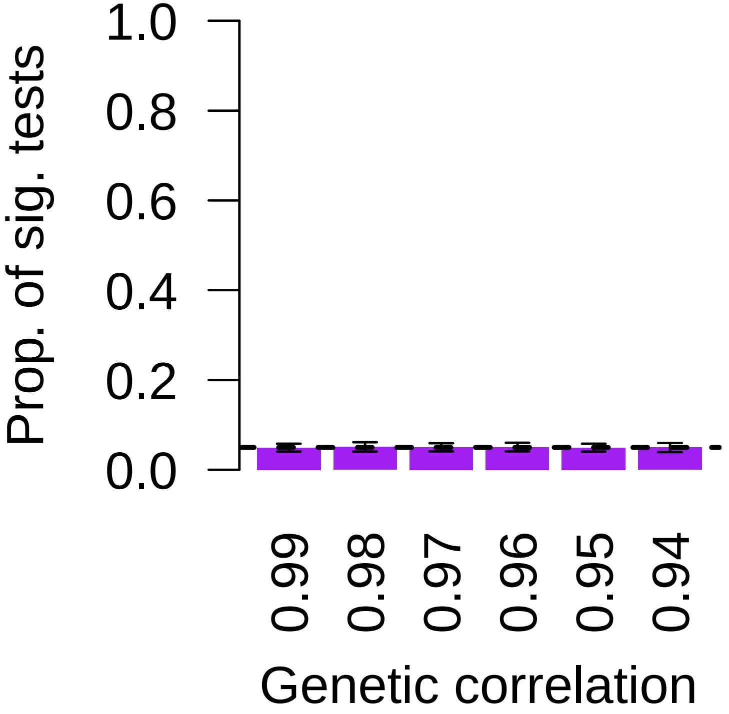
N=67K per bin

PRSxE regression



N=337K training
N=47K test

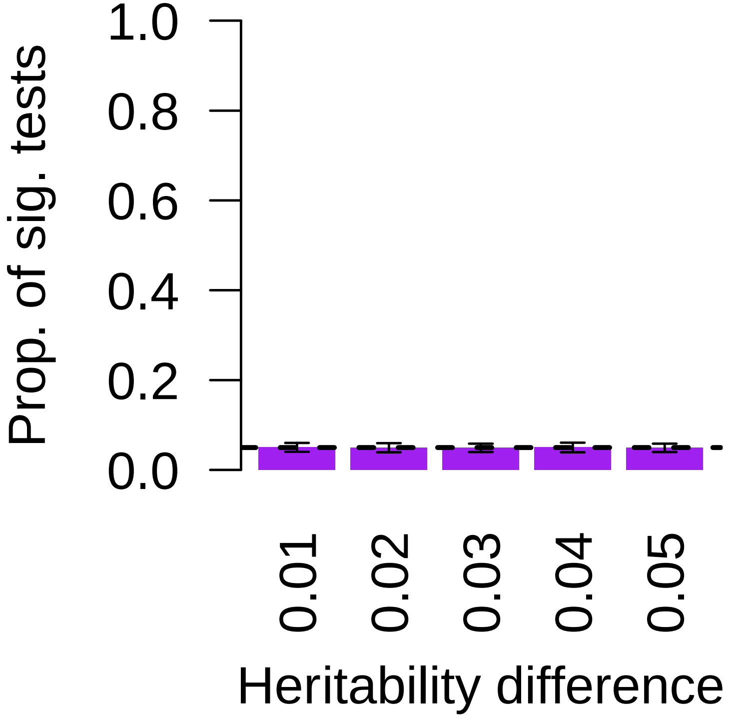
SNP-heritability across E bins



N=67K per bin

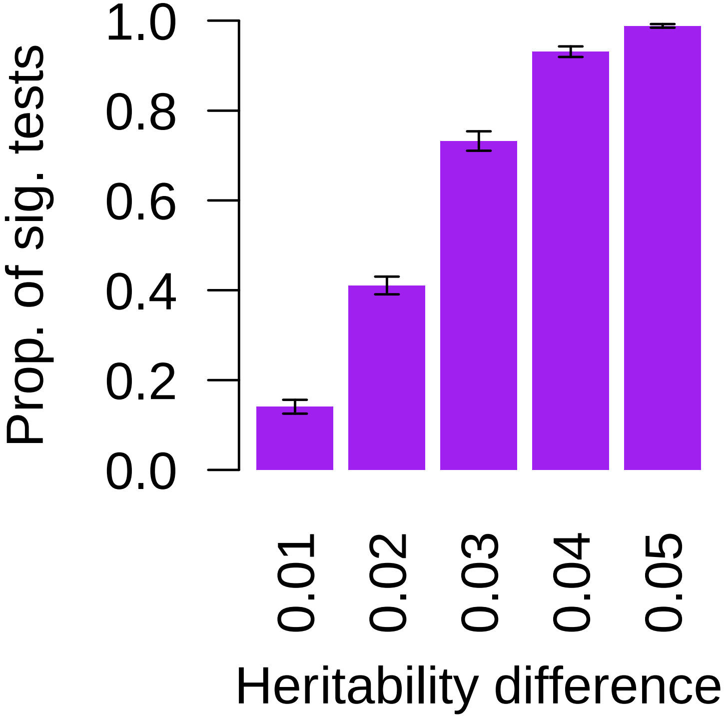
h^2 estimation across E bins and PRSxE regression are well-powered to detect heritability differences between E bins > 2% in Scenario 2

Genetic correlation
across E bins



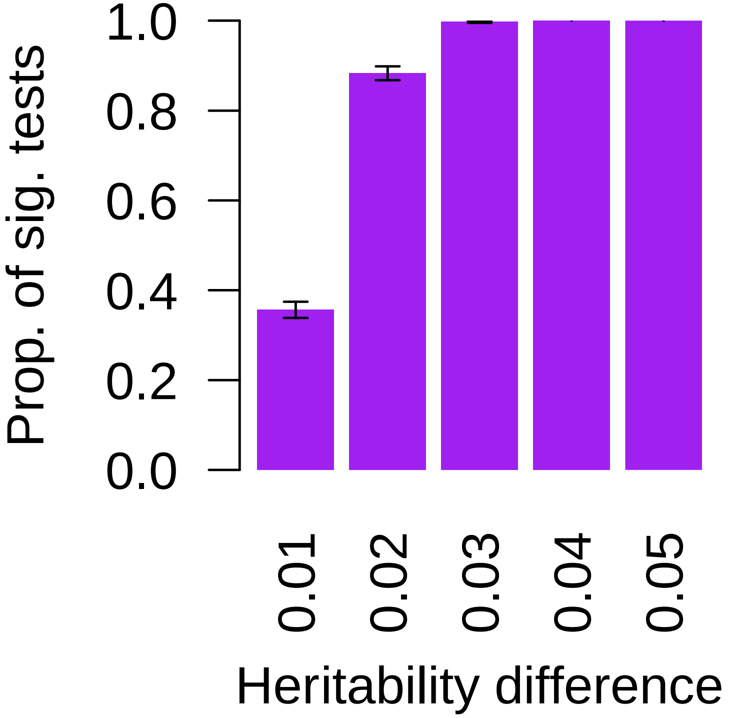
N=67K per bin

PRSxE
regression



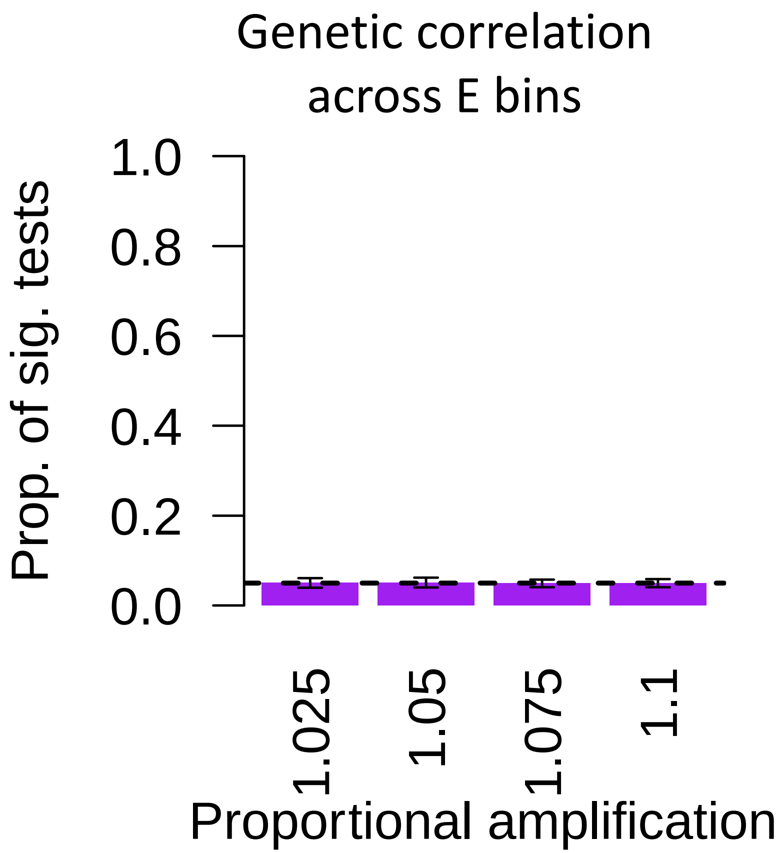
N=337K training
N=47K test

SNP-heritability
across E bins

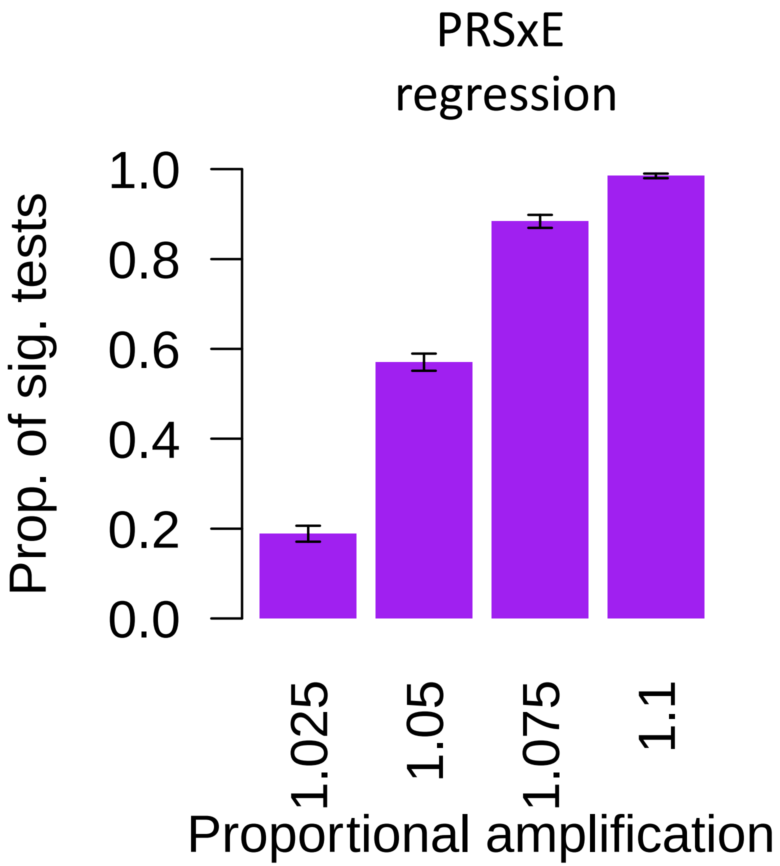


N=67K per bin

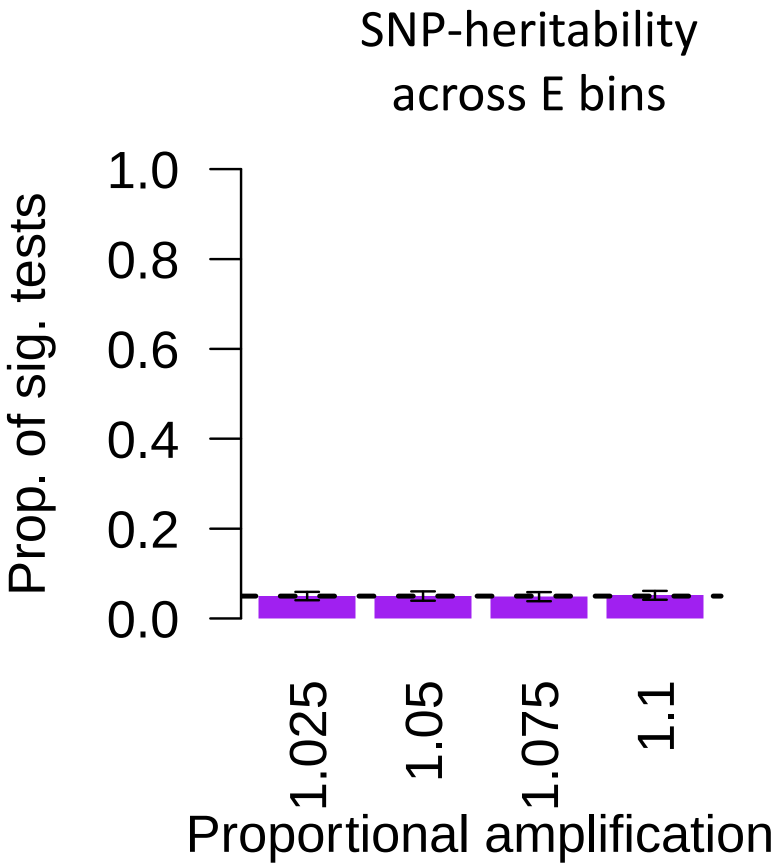
PRSxE regression is well-powered to detect proportional phenotype amplification >5% in Scenario 3



N=67K per bin



N=337K training
N=47K test



N=67K per bin

Inferring the contribution of GxE in UKBiobank data

1. We analyzed 33 complex diseases and traits across 10 E variables

Inferring the contribution of GxE in UKBiobank data

1. We analyzed 33 complex diseases and traits across 10 E variables
    - Diet
    - Sleep
    - Alcohol
    - Smoking
    - Physical activity
    - Sex

Inferring the contribution of GxE in UKBiobank data

1. We analyzed 33 complex diseases and traits across 10 E variables
    - Diet
    - Sleep
    - Alcohol
    - Smoking
    - Physical activity
    - Sex

Average E variable h^2 = 6%
Max = 15% (Smoking)

Inferring the contribution of GxE in UKBiobank data

1. We analyzed 33 complex diseases and traits across 10 E variables
    - Diet
    - Sleep
    - Alcohol
    - Smoking
    - Physical activity
    - Sex

2. For h^2 and r_g we binned individuals into 5 bins (or 2 bins for binary E variables) according to their E variables:
    - r_g and h^2 : N=67K per bin
    - PRSxE: N=47K (PRS trained on 337K)

Outline

1. Background on gene-environment interactions
2. Methods
3. Simulations
4. **Results**
5. Summary

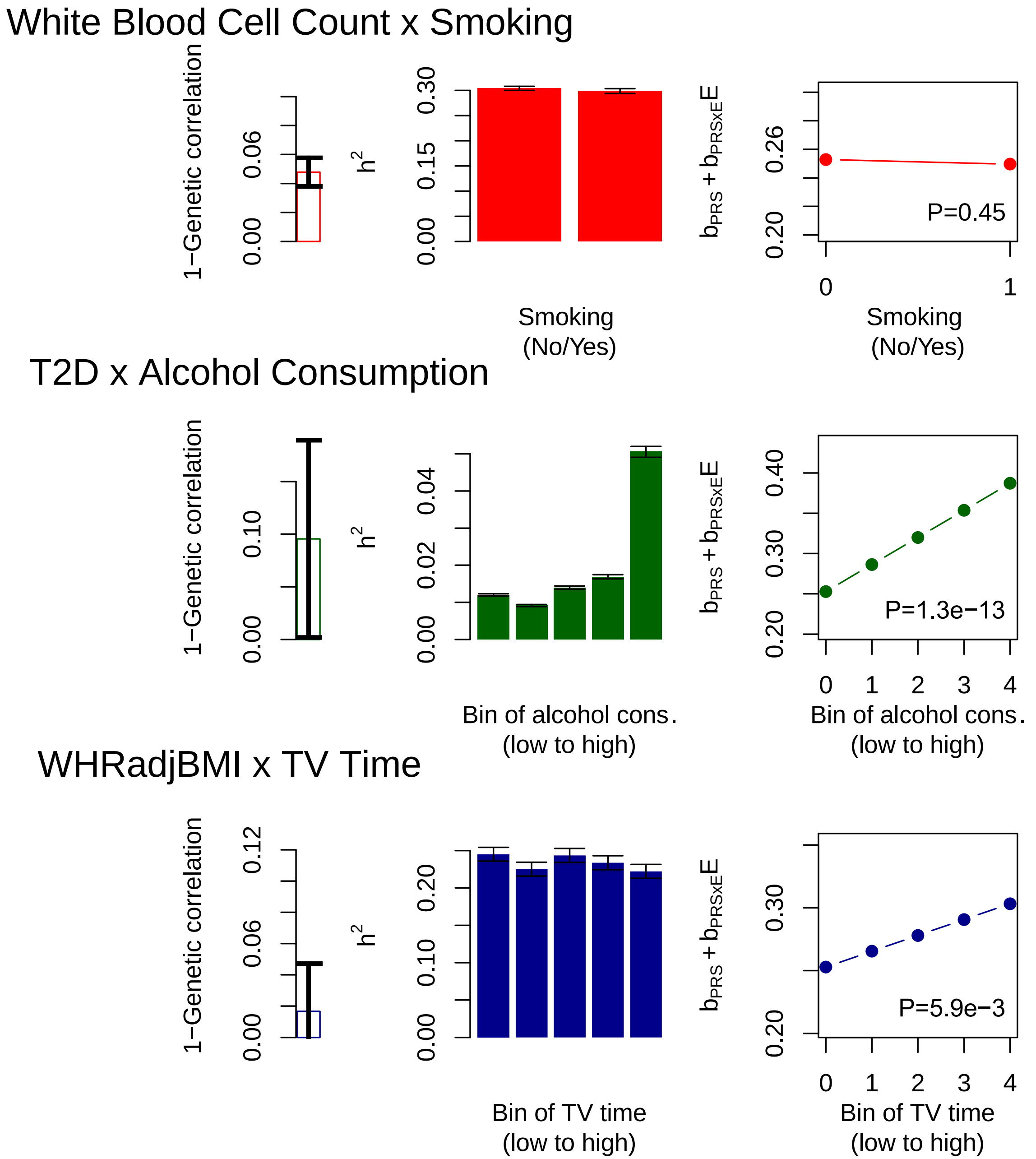
We find examples of all three Scenarios of GxE

Scenario 1:
Imperfect Genetic correlation

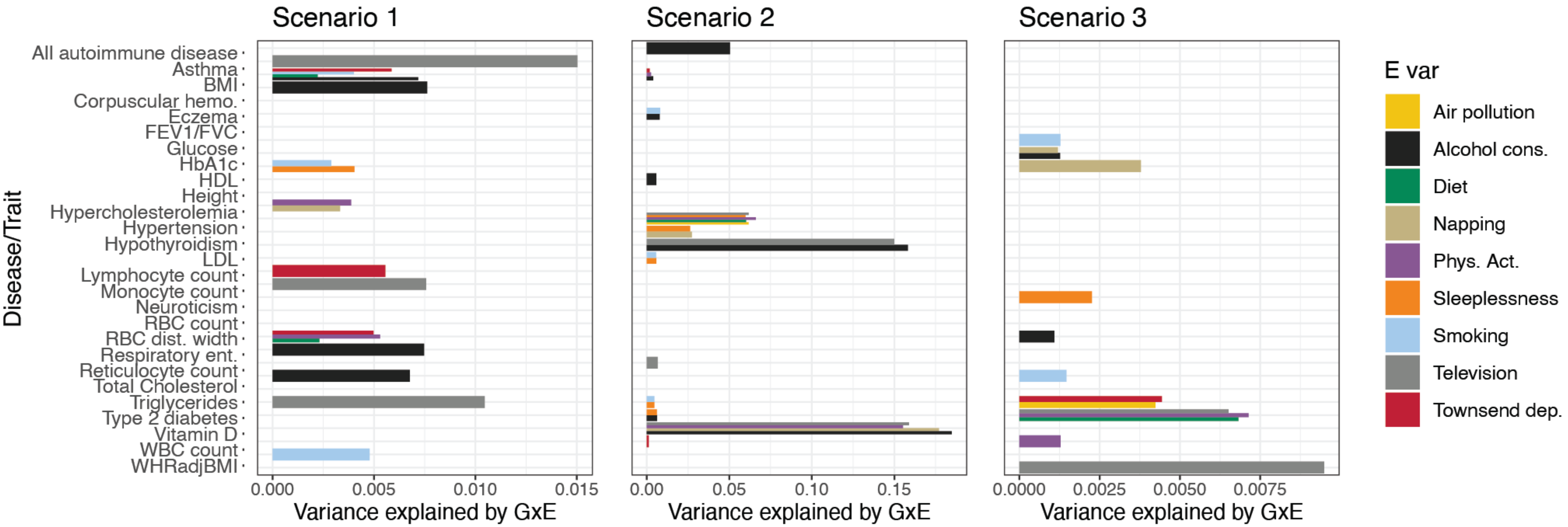
Scenario 2:
Varying Genetic variance

Scenario 3:
Proportional amplification

Genetic correlation | Heritability by E | PRSxE regression



GxE explains 3% of trait variance on average across traits



We find examples of all three Scenarios of GxSex

Genetic correlation

Heritability by E

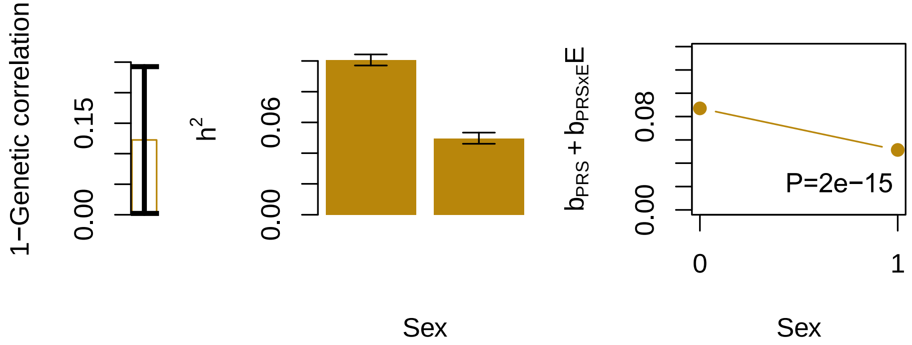
PRSxE regression

Scenario 1:
Imperfect Genetic correlation

A. Neuroticism x Sex

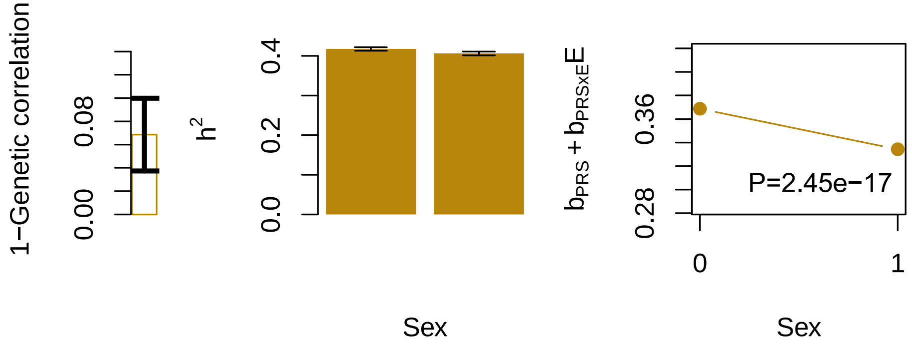
Scenario 2:
Varying Genetic variance

B. Autoimmune Disease x Sex

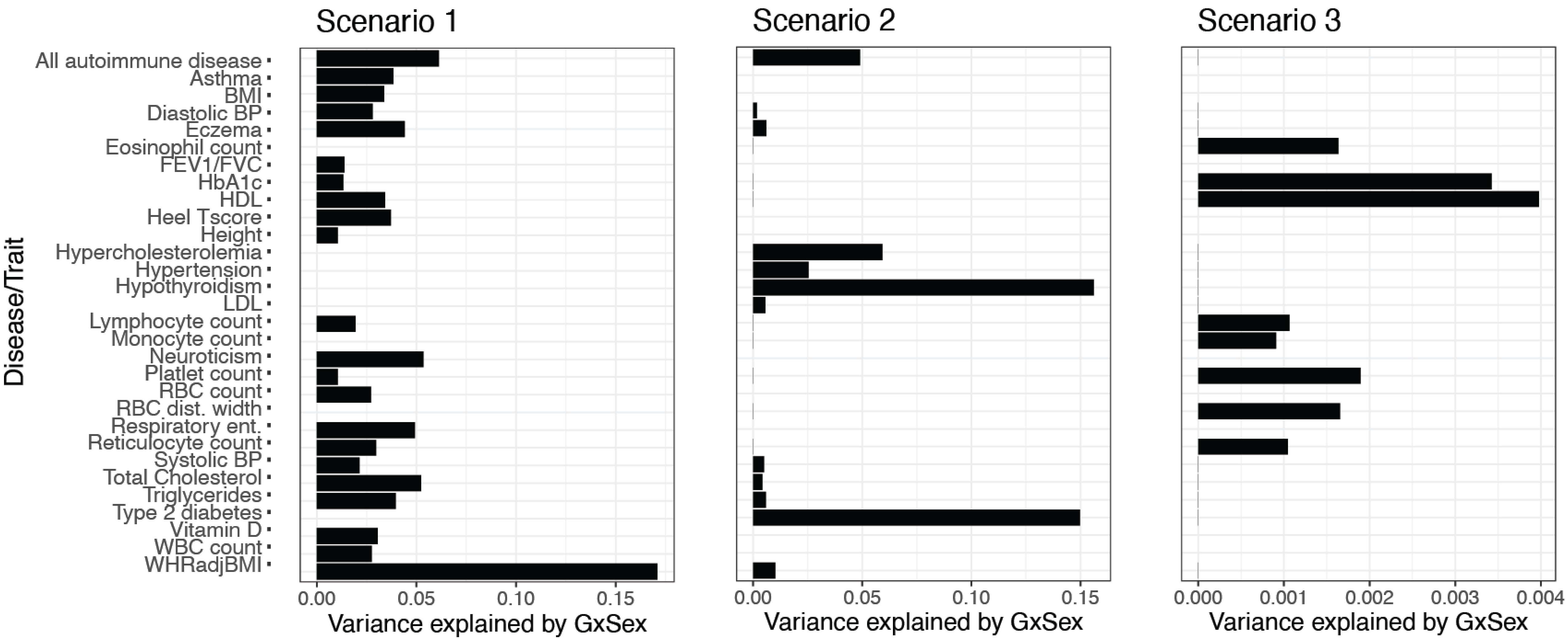


Scenario 3:
Proportional amplification

C. HDL Cholesterol x Sex



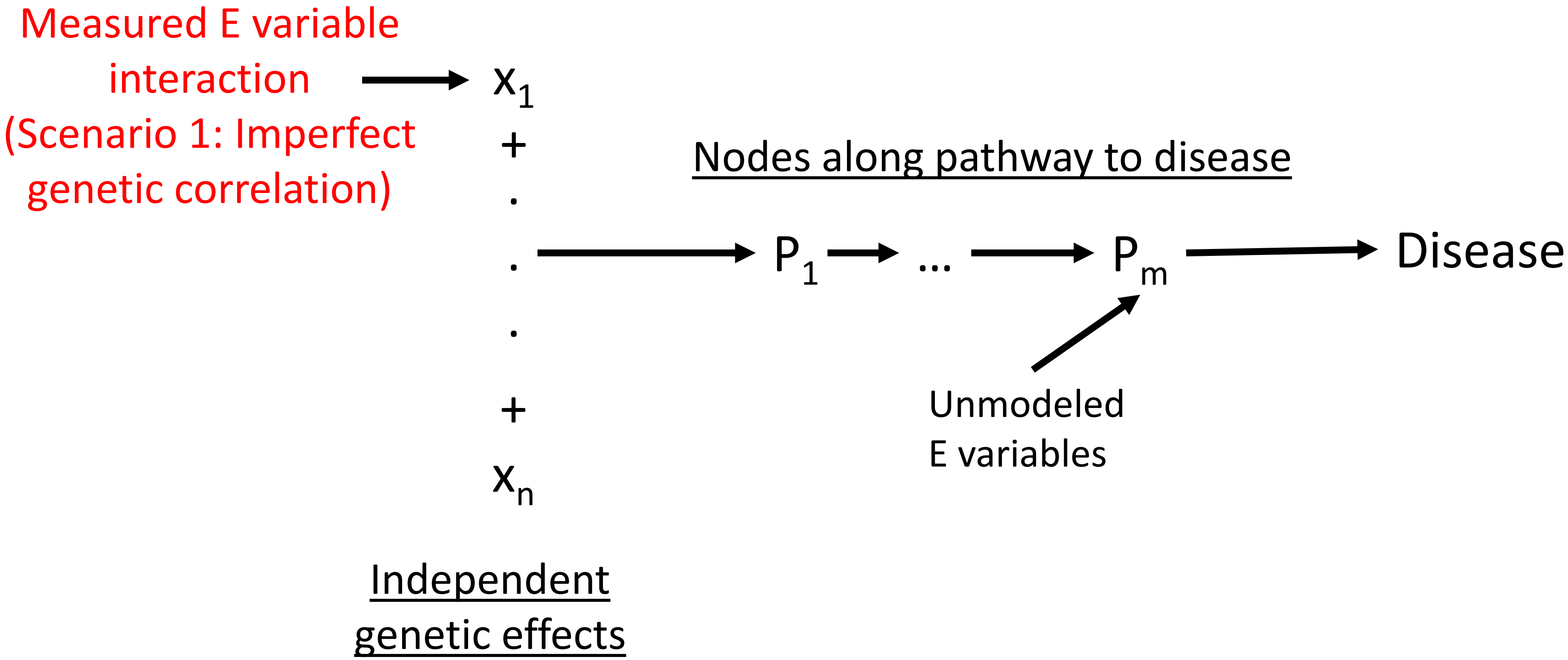
GxSex explains 4% of trait variance on average across traits



Outline

1. Background on gene-environment interactions
2. Methods
3. Results
4. **Summary**

Our results point to a model where E variables interact with genetics at several levels



Measured E variable interaction (Scenario 1: Imperfect genetic correlation)

x_1
+
.
.
.
+
x_n

Independent genetic effects

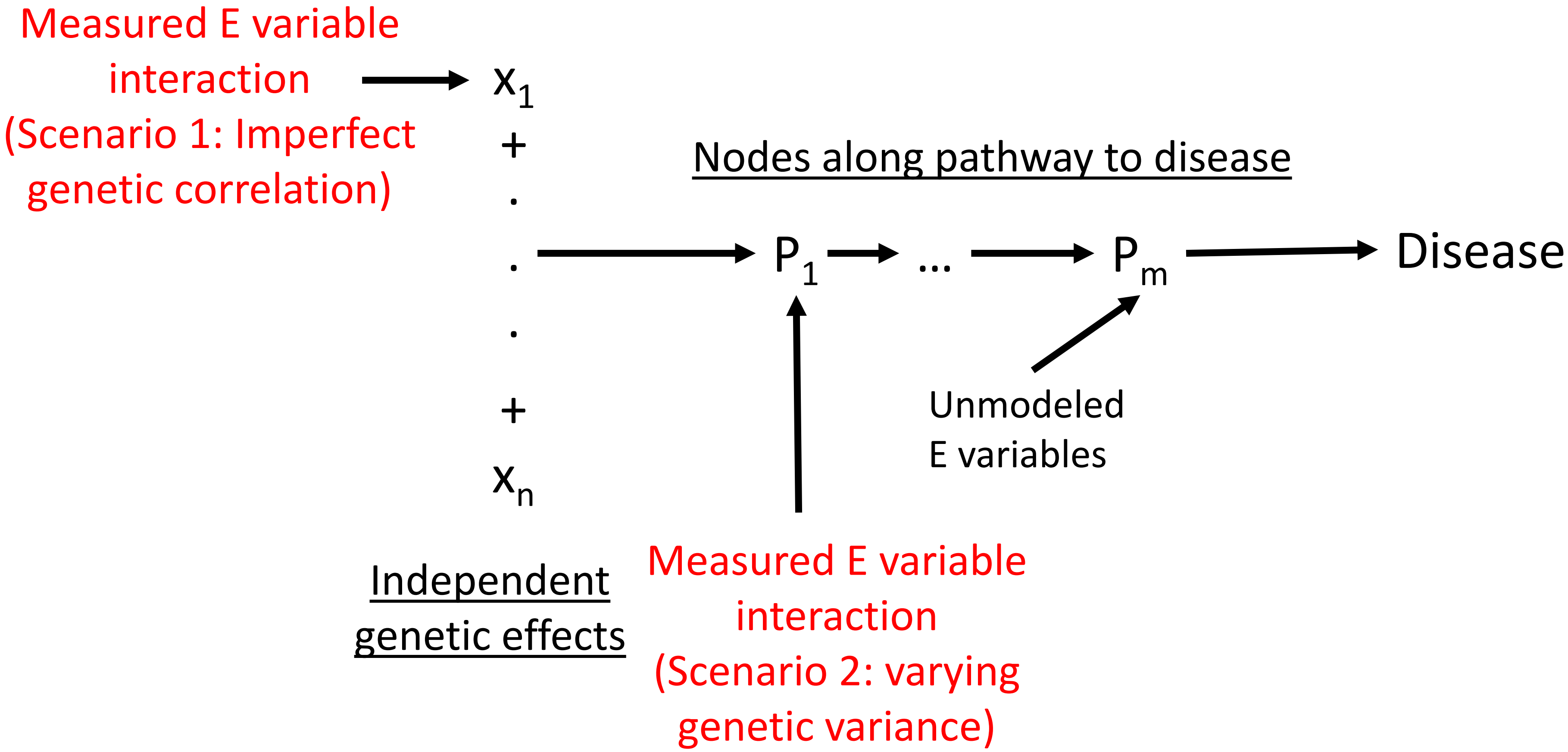
Nodes along pathway to disease

P_1 … P_m

Disease

Unmodeled
E variables

Our results point to a model where E variables interact with genetics at several levels



Measured E variable interaction (Scenario 1: Imperfect genetic correlation)

x_1 + . . . + x_n

Independent genetic effects

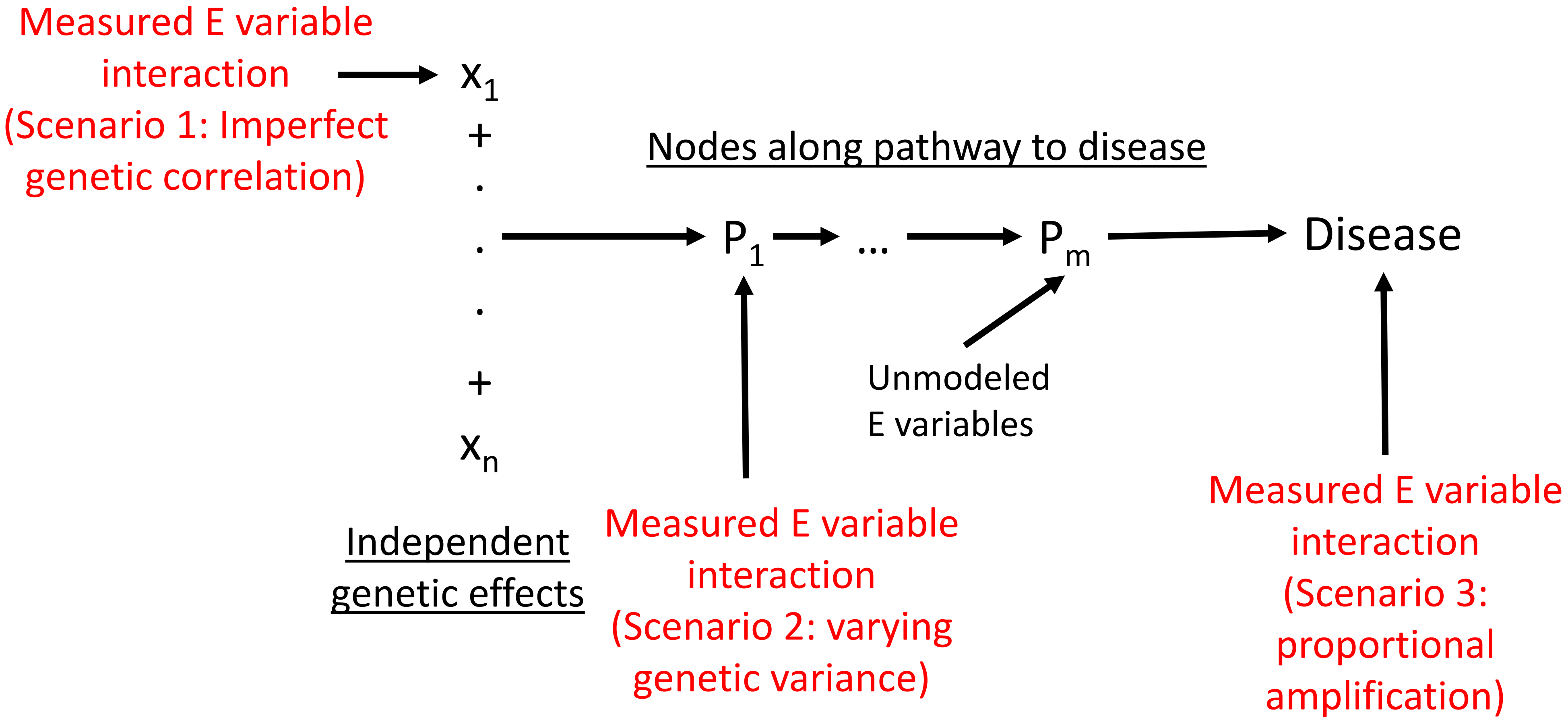
Nodes along pathway to disease

P_1 → … → P_m → Disease

Unmodeled E variables

Measured E variable interaction (Scenario 2: varying genetic variance)

Our results point to a model where E variables interact with genetics at several levels



Measured E variable interaction (Scenario 1: Imperfect genetic correlation)

x_1 + . . . + x_n

Independent genetic effects

Nodes along pathway to disease

P_1 → … → P_m → Disease

Unmodeled E variables

Measured E variable interaction (Scenario 2: varying genetic variance)

Measured E variable interaction (Scenario 3: proportional amplification)

Summary and conclusions

- For many diseases, the genetic effects are context-dependent

Summary and conclusions

- For many diseases, the genetic effects are context-dependent

- We find evidence for polygenic GxE arising from locus dependent interactions

Summary and conclusions

- For many diseases, the genetic effects are context-dependent
- We find evidence for polygenic GxE arising from locus dependent interactions
- We also find evidence for polygenic GxE arising from non-locus dependent interactions

Acknowledgements

- UKBiobank participants
  - Application #16549
- Alkes Price
- Xilin Jiang
- Ben Strober
- Martin Zhang
- Price Lab

The Durvasula Lab @ USC is hiring

- We are hiring postdoctoral researchers and graduate students
- We have projects related to gene-environment interactions, statistical genetics, and complex trait evolution
- Get in touch if you are interested!

Email: durvasul@usc.edu
Website: sites.usc.edu/durvasula

Keck School of Medicine of USC
Center for Genetic Epidemiology

Acknowledgements

- UKBiobank participants
  - Application #16549
- Alkes Price
- Xilin Jiang
- Ben Strober
- Martin Zhang
- Price Lab

Preprint available on medRxiv:

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES