

# Environment vs Genetics: Predicting Common Disease Risk in the UK

Yixuan He

*PhD student in Bioinformatics and Integrative Genomics*

*2019 UK Biobank Meeting*

*June 19, 2019*



**HARVARD**

**MEDICAL SCHOOL**

*Phenotype*

**P**

**=**

*Genotype*

**G**

**+**

*Exposures*

**X**

Type 2 Diabetes

Lung Function

Coronary Artery Disease

Atrial Fibrillation

Genetic Variants

Pollutants

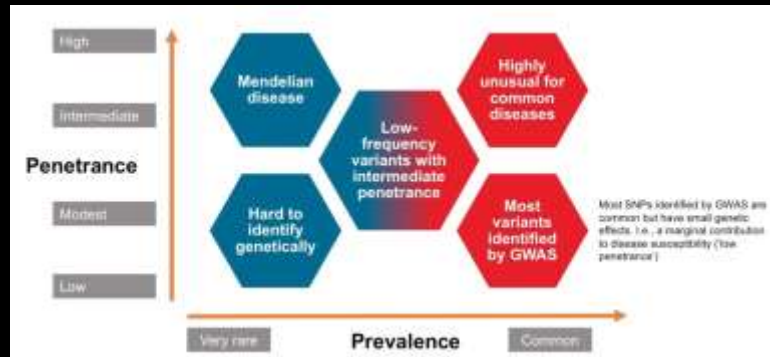
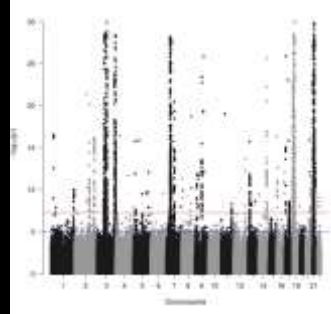
Physical Activity

Nutrients + Diet

Education Attainment

# Genotype

# G



# Traditional Approaches

Genome Wide Association Studies (GWAS)  
Identify variants associated with phenotype  
Genome wide significance of  $p=5 \times 10^{-8}$

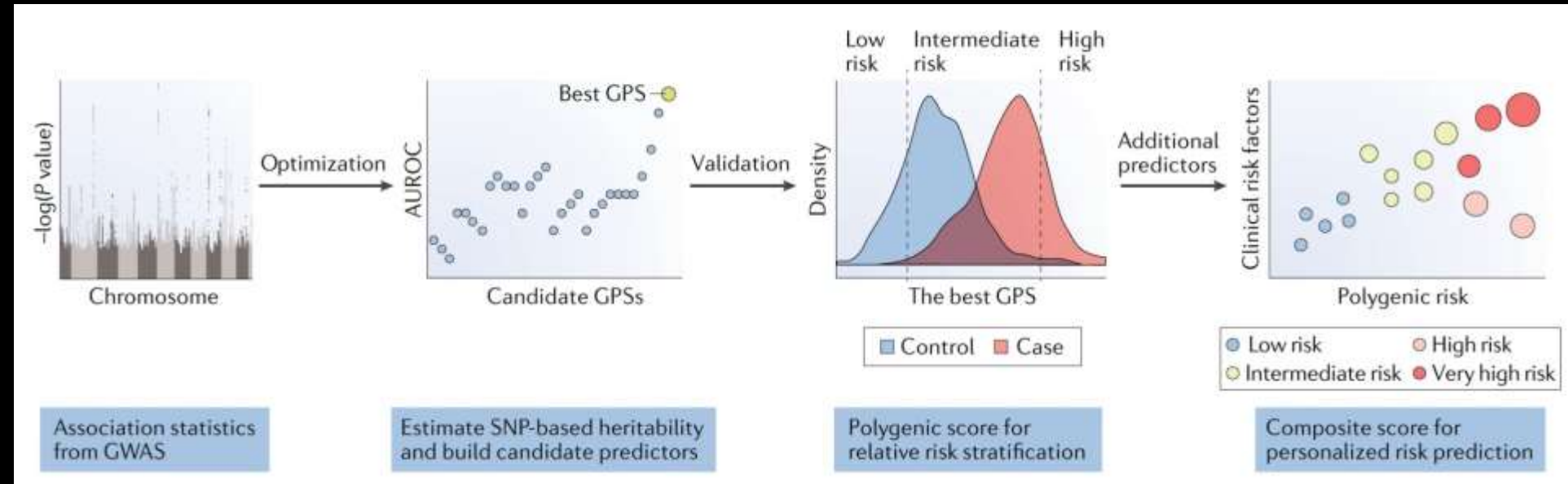
GWAS Catalog  
~4000 publications and >138,000 associations (2019 May)  
<https://www.ebi.ac.uk/gwas/>

Common disease, Common variant model  
most common disorders are due to a large number of alleles

Genotype

# Polygenic Risk Scores (PRS)

# G



Combines individual effects of all GWAS variants across the genome to predict risk

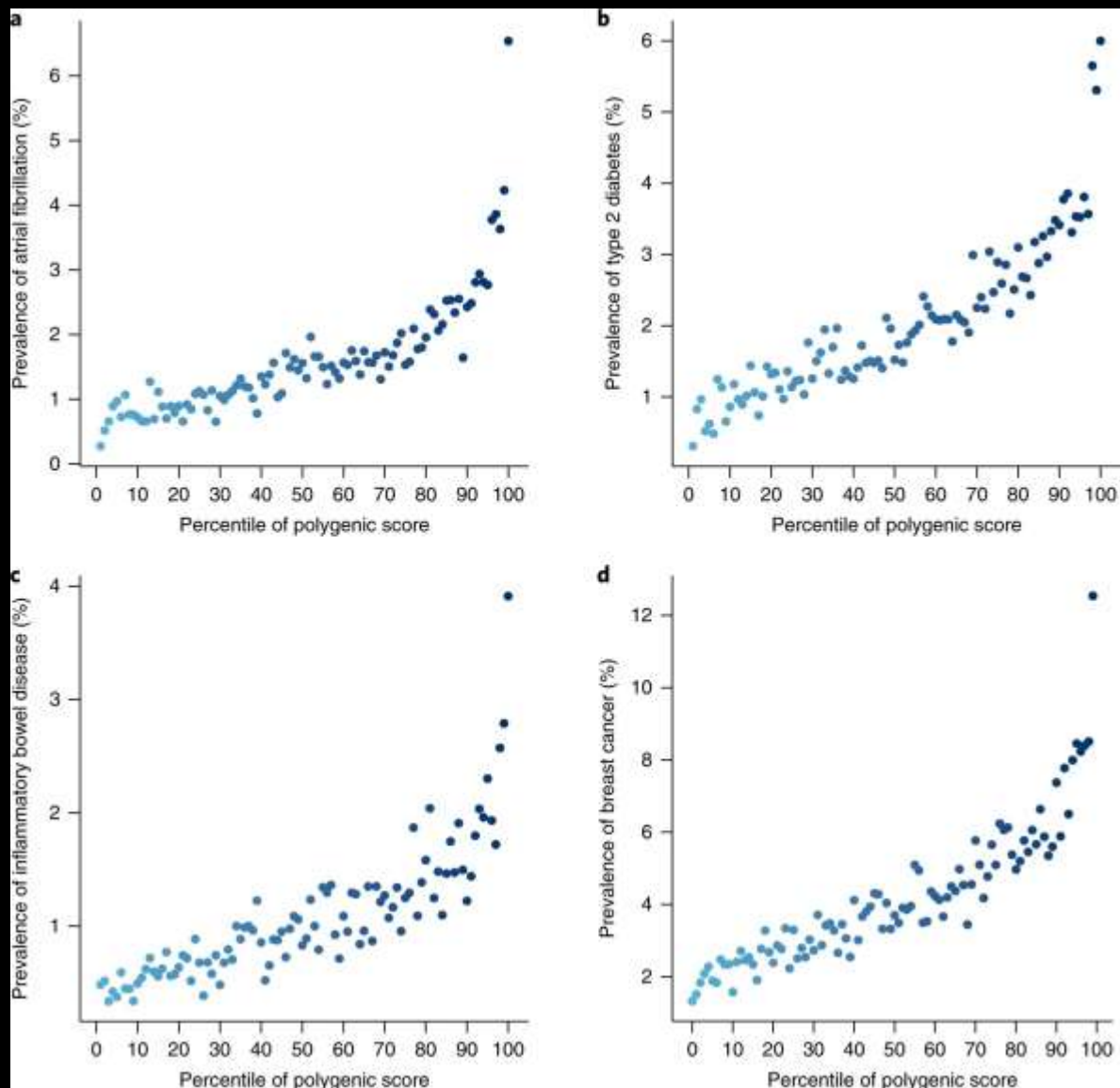
Accounts for correlation between variants (linkage disequilibrium)

Theoretically maximizes power of GWAS

Genotype

# PRS in Common Diseases

# G



Khera et al. 2018

Atrial Fibrillation, Type 2 Diabetes, Inflammatory Bowel Disease, Breast Cancer, and Coronary Heart Disease.

Identified individuals with greater than threefold increased risk.

*Exposures*

**X**

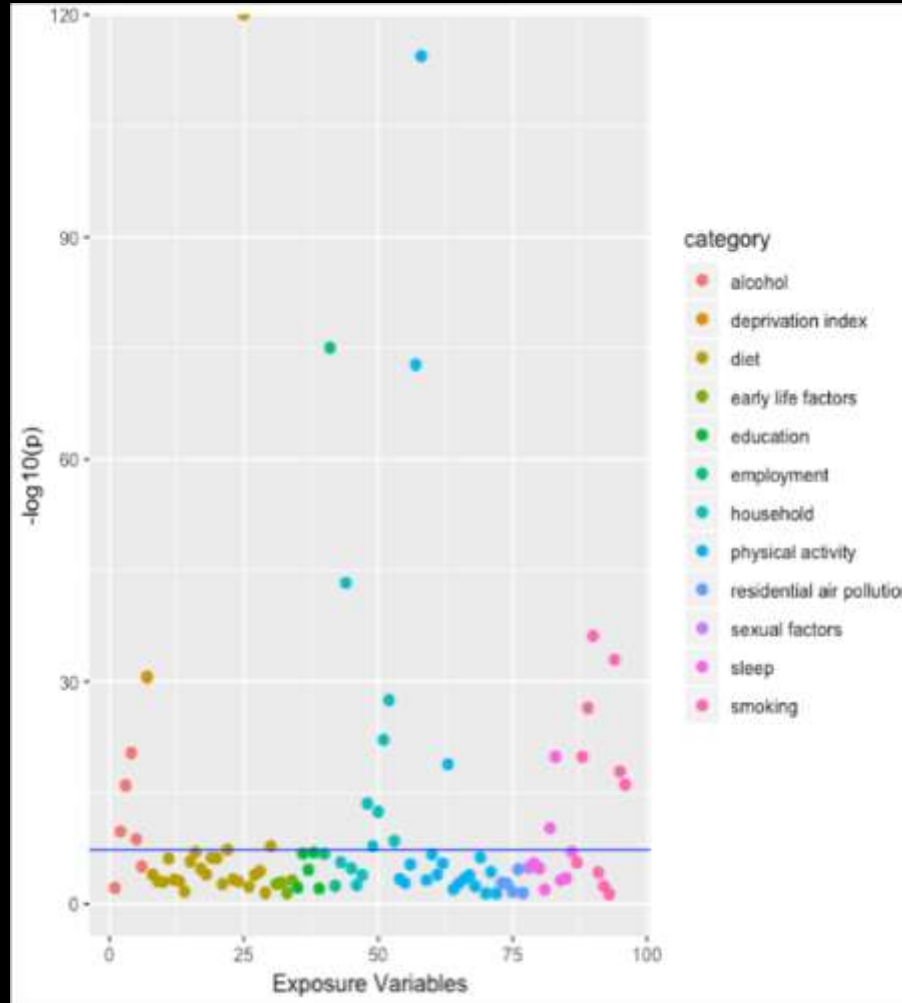
???

Exposures

# eXposure Wide Association Study (XWAS)

# X

## Type 2 Diabetes (T2D)



Environmental exposure variables can be extracted based on category name.

Only those with <10% missing entries were retained.

Deemed Benjamini-Hochberg False Discovery Rate (FDR) of <0.05 as significant.

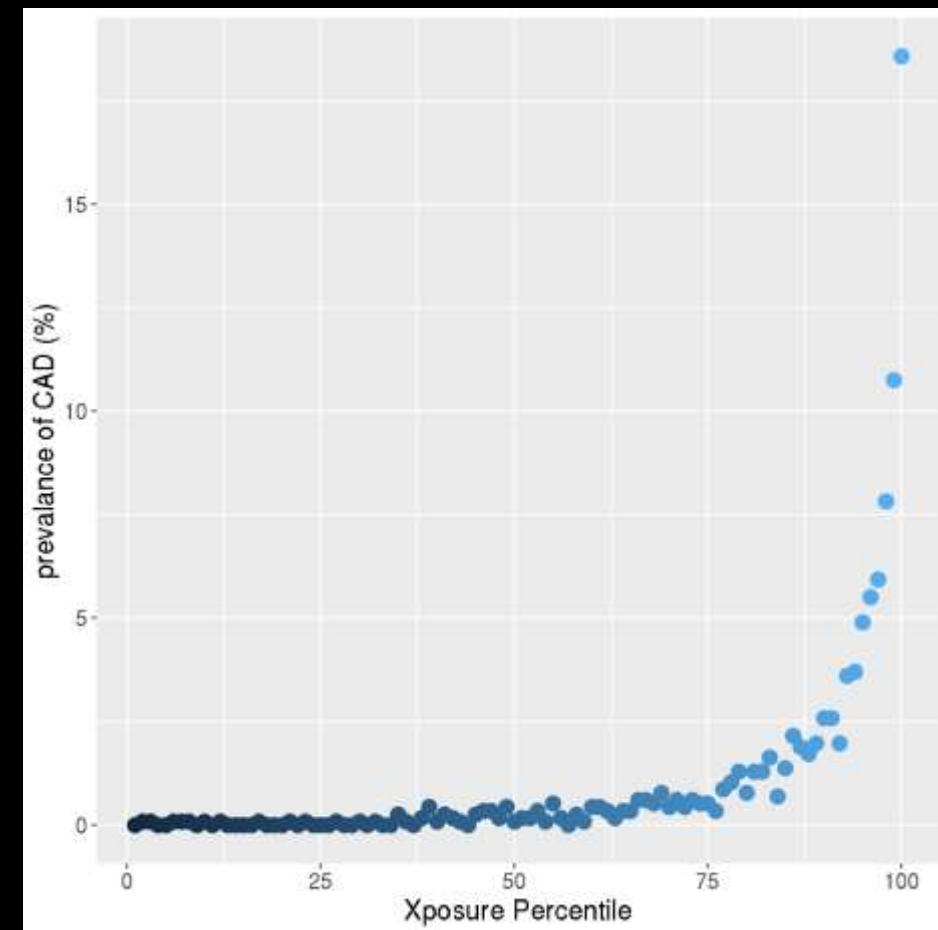
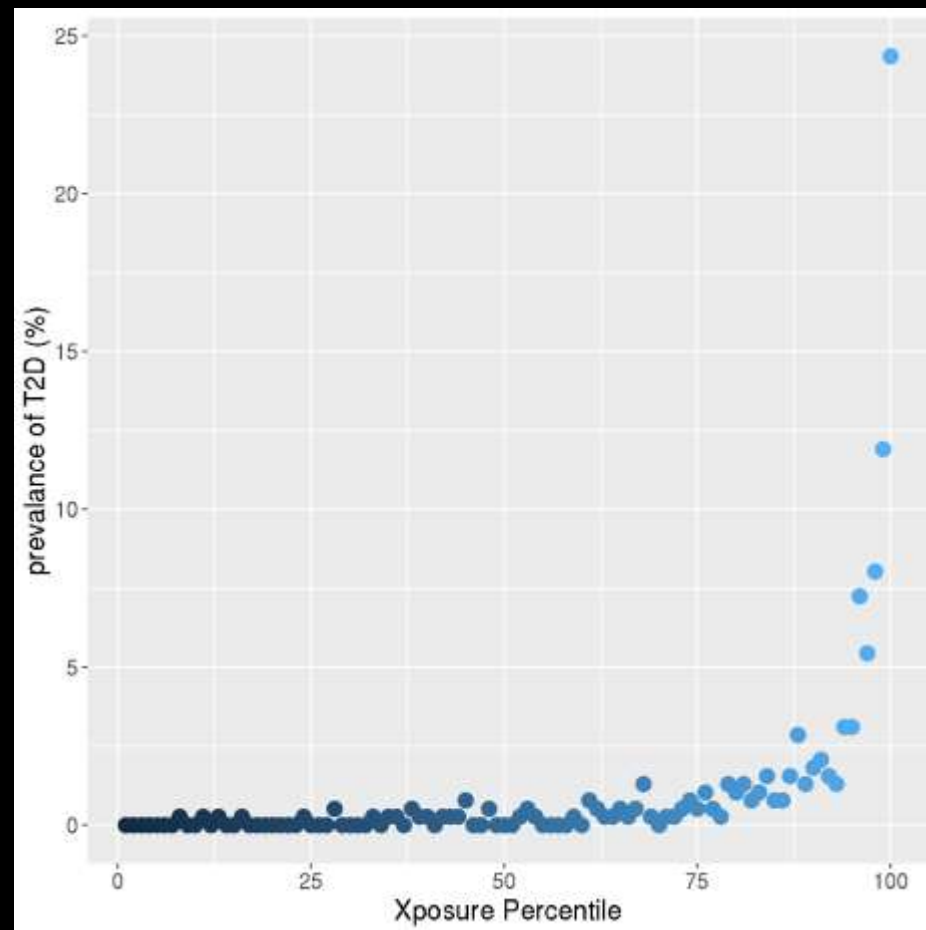
# Exposures

# X

**T2D**: type 2 diabetes

**CAD**: coronary heart disease

# eXposure Score



Prevalence of disease in each of the 100 groups binned according to the percentile of significant eXposures.



$$\begin{array}{ccccc} \textit{Phenotype} & & \textit{Genotype} & & \textit{Exposures} \\ \mathbf{P} & = & \mathbf{G} & + & \mathbf{X} \end{array}$$

- Information in large databases like the UKBiobank allows us to study associations between environment and phenotype.
- In addition to genetics, environmental exposures also provide critical information that can help improve clinical care in complex common diseases.

# Acknowledgements

Rag Group:



Chirag Patel



Chirag Lakhani



Arjun K Manrai



BLAVATNIK INSTITUTE  
BIOMEDICAL INFORMATICS



[yixuan\\_he@hms.harvard.edu](mailto:yixuan_he@hms.harvard.edu)



@he\_yix