GGSB PRELIM QUESTION # 9

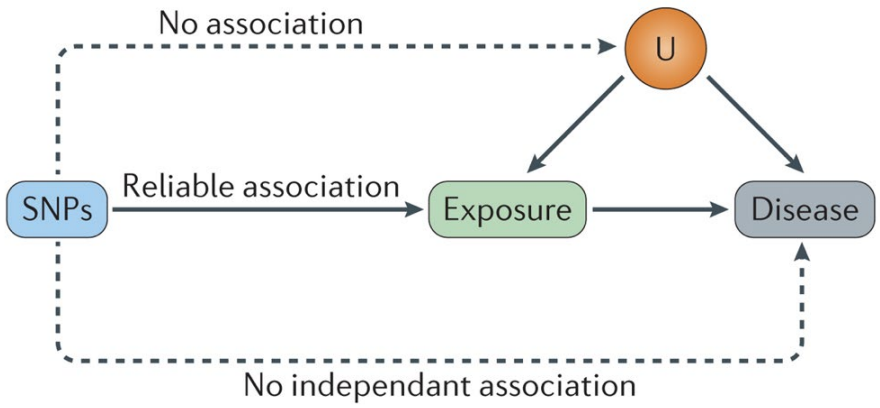
Identifying causal genes for complex traits

Genome-wide association studies have been performed on a large number of complex traits. These have discovered thousands of loci robustly associated with the traits. These loci fall in non-coding regions of the genome, which makes mapping out the underlying mechanisms challenging. The general consensus is that GWAS variants alter gene expression traits and through this mechanism affect the phenotype. PrediXcan [Gamazon et al 2015] and subsequent methods such as TWAS [Gusev et al 2016] were proposed to test this hypothesis by correlating the genetically regulated components of gene expression with the phenotype.

1. How can the genetic architecture of gene expression traits guide the optimal choice of prediction models used in these gene mapping methods. What kind of models would work best if the architecture is highly polygenic (many variants with small effect sizes affects gene expression) or very sparse (a few causal eQTLs)?

Mendelian randomization (MR) methods use genetic variations of known effects as instruments to test the mediating role of intermediate variables (exposure, molecular traits) on phenotypes. For example, Voight et al [https://www.ncbi.nlm.nih.gov/pubmed/22607825] found strong evidence for a non-causal role of HDL cholesterol levels on cardiovascular events, contrary to the accepted wisdom of the benefits of “good cholesterol" for cardiovascular health. This negative result is consistent with the failure of clinical trials of drugs that target HDL cholesterol exclusively.

3. Briefly describe the mendelian randomization (MR) method and its key assumptions.



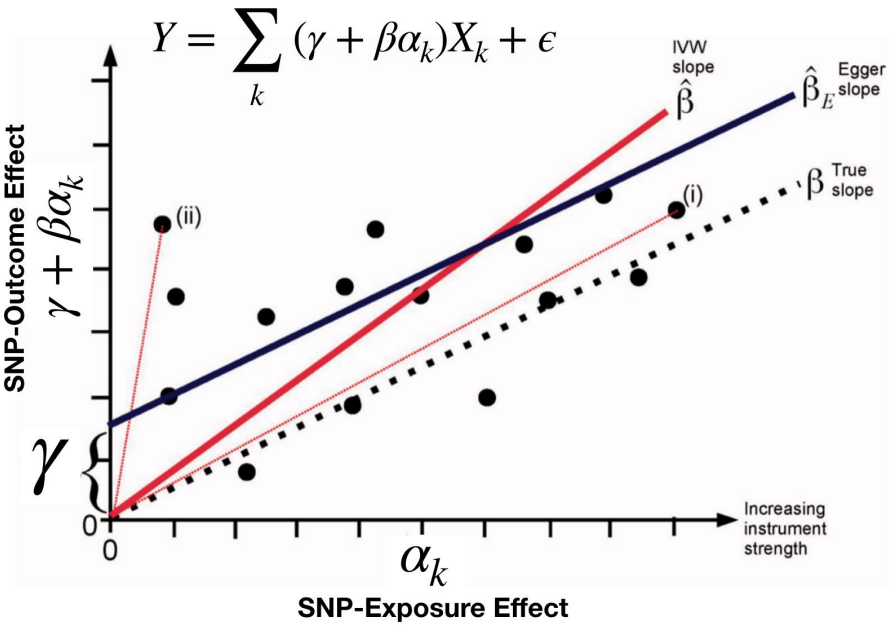
Nature Reviews | Cardiology

4. How can Predixcan be interpreted as a MR approach?

5. What assumptions need to be satisfied to conclude that the association is causal?

MR-Egger [Bowden et al] extends the applicability of MR by relaxing the requirement that there is no direct genetic effects and works well when there are multiple independent instruments (k=1:K). The phenotype here is modeled with a direct genetic effect γ (common for all k) and an indirect effect ($\beta\ \alpha_k$) as follows (with X_k representing the allelic dosage of SNP k)

$$Y = \sum_k (\gamma + \beta\alpha_k) X_k + \epsilon$$



6. Write down the PrediXcan model with this notation and compare to MR Egger.

7 What are the biological assumptions under which PrediXcan's simplifying assumptions would be justified?

8. How would this approach work if only some of the eQTL affected the phenotype. How would you interpret biologically if this were the case?

9. How do the following papers extend the PrediXcan framework.

Barfield et al [https://www.biorxiv.org/content/early/2018/04/27/223263]
Mancuso et al [https://www.biorxiv.org/content/early/2017/12/20/236869]
Park et al [https://www.biorxiv.org/content/early/2017/12/01/219428]

What are the assumptions that are dropped in each paper and under which biological assumptions would these make sense?

References

[Bowden et al]
J. Bowden, G. Davey Smith, and S. Burgess, “Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression,” *International Journal of Epidemiology*, vol. 44, no. 2, pp. 512–525, Jun. 2015.

[Gamazon et al 2015]
E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, GTEx Consortium, D. L. Nicolae, N. J. Cox, and H. K. Im, “A gene-based association method for mapping traits using reference transcriptome data.,” *Nat Genet*, vol. 47, no. 9, pp. 1091–1098, Sep. 2015.

[Gusev et al 2016]
A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. J. H. Penninx, R. Jansen, E. J. C. de Geus, D. I. Boomsma, F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. Lehtimäki, E. Raitoharju, M. Kähönen, I. Seppälä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, and B. Pasaniuc, “Integrative approaches for large-scale transcriptome-wide association studies.,” *Nat Genet*, vol. 48, no. 3, pp. 245–252, Mar. 2016.

[Voight et al]
B. F. Voight et al “Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study,” *The Lancet*, vol. 380, no. 9841, pp. 572–580, Aug. 2012.
[https://www.ncbi.nlm.nih.gov/pubmed/22607825]

[Holmes et al]
M. V. Holmes, M. Ala-Korpela, and G. D. Smith, “Mendelian randomization in cardiometabolic disease: challenges in evaluating causality.,” *Nat Rev Cardiol*, vol. 14, no. 10, pp. 577–590, Oct. 2017.
[https://www.nature.com/articles/nrcardio.2017.78]

[Barfield et al]
Richard Barfield, Helian Feng, Alexander Gusev, Lang Wu, Wei Zheng, Bogdan Pasaniuc, Peter Kraft, “Transcriptome-wide association studies accounting for colocalization using Egger regression”
bioRxiv 223263; doi: https://doi.org/10.1101/223263
Now published in *Genetic Epidemiology* doi: 10.1002/gepi.22131

[Mancuso et al]
Probabilistic fine-mapping of transcriptome-wide association studies
Nicholas Mancuso, Gleb Kichaev, Huwenbo Shi, Malika Freund, Alexander Gusev, Bogdan Pasaniuc
bioRxiv 236869; doi: https://doi.org/10.1101/236869

[Park et al]
Yongjin Park, Abhishek Sarkar, Liang He, Jose Davila-Velderrain, Philip L. De Jager, Manolis Kellis, “A Bayesian approach to mediation analysis predicts 206 causal target genes in Alzheimer's disease”
bioRxiv 219428; doi: https://doi.org/10.1101/219428
https://www.biorxiv.org/content/early/2017/12/01/219428