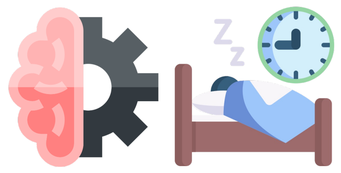
# Exploring causal relationships via gene expression for chronotype and neuropsychiatric disorders

Shane Crinion[1], Lorna M. Lopez[2], Derek W. Morris[1]

[1]Centre for Neuroimaging, Cognition and Genomics, Discipline of Biochemistry, University of Galway, University Road, Galway, Ireland;
[2]Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland

## 1. Background

- ❖ Neuropsychiatric disorders (NPD) are a complex group of disorders affecting cerebral function with psychiatric and neurodevelopmental features
- ❖ Chronotype (morning/evening person) is a behavioural indicator of underlying circadian rhythm. The morning chronotype has been linked to improved well-being while the evening chronotype has been linked to increased NPD risk
- ❖ A bidirectional relationship has been identified through Genome Wide Association Studies (GWAS) and post-GWAS analyses
- ❖ Mendelian randomisation (MR) is a post-GWAS analysis that is used to infer causal relationships between potential risk factors and outcomes
- ❖ MR using GWAS and expression quantitative trait loci (eQTL) data can identify gene expression alterations that influence chronotype and neuropsychiatric disorder risk
- ❖ Follow-up analysis using colocalization can validate the same variant is causal for gene expression and the phenotype



## 2. Objectives

This study aims to
1. Provide new evidence for genetic overlap and causal relationships between NPDs and chronotype.
2. Identify novel genes for these traits.
3. Provide supporting evidence for GWAS to map causal genes.

## 3. Mendelian Randomisation

MR is a post-GWAS analysis that can identify causal relationships between an exposure trait and risk for a disease outcome. Trait-associated SNPs, which are strongly associated with an exposure trait, are employed as instrumental variants (IVs) to represent the exposure and assess it as a potentially causal risk factor for the outcome. MR is similar to Randomised Controlled Trials (RCT), which are performed to study the effect of various therapies, exposures, or behaviours on disease risk. During an RCT, participants are randomly assigned to one of two study groups, such as an exposure (treatment) group and a control group. The results are compared between the two groups and any statistically significant difference is deemed to be due to the assigned exposure or treatment. The MR design and assumptions are in Figure 1 [1].

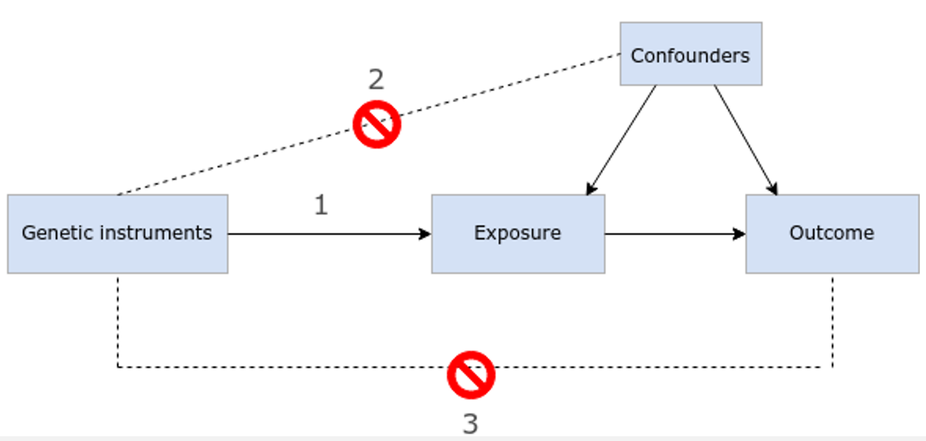
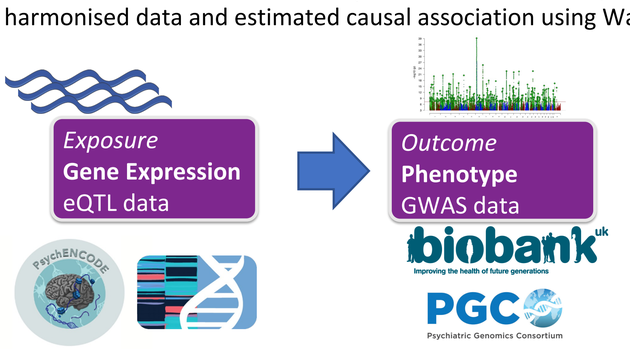



**Figure 1: Schematic diagram of MR study design and assumptions.** During an MR analysis, the effect estimates for genetic instruments are extracted from genomic data (e.g. GWAS, eQTL) and used to approximate the effect of the exposure on the outcome. The study design of MR is dependent on three core assumptions:
(1) the genetic instruments are robustly associated with the exposure
(2) genetic instruments have no association with outcome that is not mediated through the risk factor
(3) genetic instruments are not directly associated with the outcome

## 4. Study Design

*Step 1: Two-sample MR*

Two-sample MR extracts IV associations using different samples for the exposure and outcome.
We assessed the influence of gene expression alterations (exposure) on chronotype and neuropsychiatric disorder (outcome(s)) risk using a two sample MR framework. We selected significant IVs ($p < 5x10^{-8}$) from eQTL data (GTEx and PsychENCODE, Table 1)[2, 3] and extracted those instruments from GWAS data (UK Biobank and PGC, Table 2)[4, 5]. We clumped instruments, harmonised data and estimated causal association using Wald ratio estimate.




Wald ratio estimate

$$\beta_{IV} = \beta_{ZY} / \beta_{ZX}$$

*Step 2: Colocalization*

A method to identify whether genetic factors at a particular locus are shared between two traits. We use this method after MR analysis to confirm that two distinct locus are not causal in a region. We used the coloc R package and calculated posterior probability [6].

## 5. Data

**Table 1: eQTL data for exposure data**

| | Blood | Cortex | Hippocampus | Hypothalamus |
|---|---|---|---|---|
| **n** | 675 | 1592 | 165 | 170 |

**Table 2: GWAS data for outcome data**

| | Morning Person | ADHD | ASD | BD | Insomnia | MDD | SZ |
|---|---|---|---|---|---|---|---|
| **Cases** | 252,287 | 19,099 | 18,382 | 41,917 | 386,533 | 170,756 | 53,386 |
| **Cont.** | 150,908 | 34,194 | 27,969 | 371549 | - | 329,443 | 77,258 |

## 6. Results

- ❖ MR uncovered 175 eQTL-gene pairs that mediate change through gene expression, of which 105 influenced schizophrenia risk.
- ❖ Of these 175 eQTL-gene pairs, colocalization found that 11 were caused by the same variant (Figure 2).
- ❖ Results identify new genes associated to the phenotypes and identify new genes linking neuropsychiatric disorders and chronotype e.g. *CYB561* linked to chronotype via gene expression alterations was previously linked to neuropsychiatric traits.

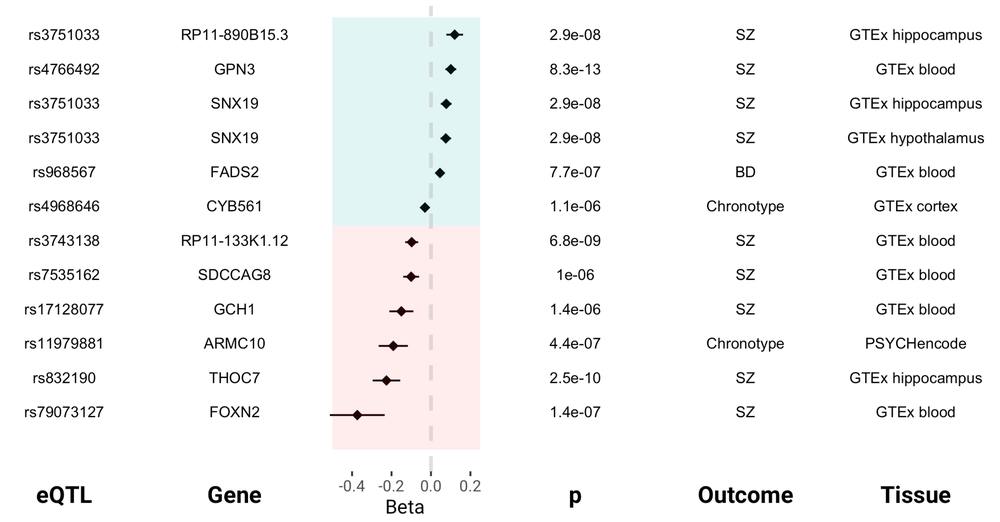



**Figure 2: Forest plot for significant associations found from MR analysis and colocalization of the causal effect of gene expression on chronotype and neuropsychiatric disorders.** For each SNP, the beta represents the effect size of gene expression alterations in the respective gene on the corresponding outcome and tissue. Blue indicates increased expression is linked to the phenotype, while red indicates decreased expression is linked to the phenotype. The p-value indicates the strength of the effect on the outcome and the tissue column indicates the source and the sample. Each result passed MR test for causal effect estimate and colocalization for same causal variant.

## 7. Discussion

MR found evidence for causal associations that link gene expression changes to neuropsychiatric disorders and chronotype. Colocalization analysis validates that the same variant influences gene expression and phenotype. The results include new genes of interest (*ARMC10*, *GCH1*), new SNPs in known genes (*SDCCAG8*, *CYB561*, *FADS2*, *SNX19*, *GPN3*) and new links between neuropsychiatric disorders (*GCH1*, *CYB561*). Colocalization assumes that only one causal variant exists in a region and therefore may limit the number of final eQTL-gene pairs identified as significant. Future work may use a multi-instrument approach and additional genomic datasets.

## 8. References

1. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. *Int J Epidemiol*. 2003;32(1):1-22. doi:10.1093/ije/dyg070
2. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653
3. Wang D, Liu S, Warrell J, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362(6420):eaat8464. doi:10.1126/science.aat8464
4. Sullivan PF, Agrawal A, Bulik CM, et al. Psychiatric Genomics: An Update and an Agenda. Am J Psychiatry. 2018;175(1):15-27. doi:10.1176/appi.ajp.2017.17030283
5. Jones SE, Lane JM, Wood AR, et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat Commun*. 2019;10(1):343. Published 2019 Jan 29. doi:10.1038/s41467-018-08259-7
6. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383. Published 2014 May 15. doi:10.1371/journal.pgen.1004383

HOST INSTITUTION

PARTNER INSTITUTIONS

This publication has emanated from research supported by a research grant from Science Foundation Ireland (SFI) under Grant number 18/CRT/6214 and is co-funded under the European Regional Development Fund and by Genuity Science