Synopsis no.: *S2.38*

Preliminary title: The EU-GEI genome-wide association study: the relationship between the genetic risk of schizophrenia, cognitive performance and schizophrenia symptoms

Contact info for the person(s) proposing the synopsis

Name: Dr Alex Richards
Partner no: P3 (CU)
e-mail address: wmdalr@cf.ac.uk

Publication category: 1

Working and writing group: A Richards, L Ferrero, S Goluksuz, D Quattrone, Uli Reininghaus, C Morgan, C Gayer-Anderson, C Iyegbe, A Pocklington, P Holmans, J Walters, G Leonenko, R Murray, J van Os, B Rutten, M Owen, MC O'Donovan

Work Packages involved: WP2, WP3, WP6

Partners involved from whom candidate co-authors (*additional to working and writing group*) should be nominated: 1-3, 5-9, 15, 22, 26-27 (all for sample access)

Objectives (scientific background, hypothesis, methods, and expected results):

Scientific background

Large genome-wide association studies, such as the meta-analysis conducted by the Psychiatric Genomics Consortium, have identified many novel genetic loci associated with schizophrenia. However, large sample size is achieved in these studies at the expense of phenotypic detail, with presence of schizophrenia reduced to a simple dichotomous variable. The EU-GEI dataset addresses these concerns by including common variant genotype data and a range of cognitive, symptom and environmental data on a large, pan-European cohort of controls, psychosis cases and their first-degree relatives. This allows the comparison of aggregate measures of genetic risk, such as risk profile scores or principal components derived from risk variants, with these cognitive and symptom data types.

Aims

We will perform a genome-wide association study of single variants against affection status and major cognitive and symptom variables. We also seek to discover whether risk profile scores based on major psychiatric disorder case/control training datasets (SZ, BD, MDD) have a relationship with cognitive and symptom variables and derived dimensions. We also wish to determine if any significant results from this analysis can be related to biological functional categories that have shown prior relevance to schizophrenia. Lastly, we will conduct factor analysis and clustering analysis on the EU-GEI genotypic data, and again examine the resulting clusters and factors for relationships with symptom and cognitive data. Simultaneous clustering of genotype and phenotype variables will also be attempted using canonical correlations.

Methods

The samples have been genotyped at Cardiff on the IPMCN chip. Genotype data were called using the GenomeStudio package and transferred into PLINK format. After quality control, has been imputed

using the software package Minimac3 and the Haplotype Reference Consortium reference panel. The single variant association analysis will be performed using linear mixed model analysis, to take into account sample relatedness and the variety of populations in the dataset. Risk profile scores will use the most powerful GWAS studies currently available from the PGC as training sets.

We will evaluate the performance of several factor analysis and clustering techniques on the genomic EU-GEI data, including principal component analysis and canonical correlations, the latter simultaneously correlating genomic and phenotypic data in an unsupervised way with respect to either. Specific focus will be given to ensuring that neither population stratification within samples nor linkage disequilibrium acts as a confounding factor in our analysis. Replication of any relationship between clusters or factors across the subpopulations of the EU-GEI and in appropriate external datasets will be a priority.

Hypotheses

We hypothesise that schizophrenia risk profile scores should be associated with cognitive performance. We hypothesise that different symptom domains will correlate best with different GWAS. We hypothesise that scores in functional categories (eg ARC/NMDAR) that have shown previous evidence for a link with schizophrenia and also with intellectual disability will predict cognitive impairment in schizophrenia.

Data needed for the study: Genotype data from WP2, WP6 and GROUP cases, controls and relatives. Schizophrenia case/control, cognitive and symptom data from WP2, WP6 and GROUP cases, controls and relatives. We will also link with Synopsis 2.34 which seeks to derive symptom dimensions in order to use the above methods to test the validity of those in the EU-GEI and then expanded non EU-GEI data.

Plan for statistical analysis (overall strategy): Risk profile scores will be compared to cognitive and symptom variables using linear regression, using covariate data such as age, sex, collection centre and population principle components to correct for confounding influences.
Mixed linear model analysis will be used for the single variant GWAS tests, in order to most effectively correct for sample relatedness and the variety of populations in the dataset.
GCTA and LD regression will be used to examine the overlap in genetic architecture between schizophrenia affection status, schizophrenia symptoms and cognition.
The clustering and factor analysis methods employed will include principal component analysis and canonical correlations.

Other analyses/methods:

Involvement of external Parties (non EU-GEI): Allied GROUP data to be used where overlap between cognitive variables permits. Professor S Zammit to advise on the use of cognitive and symptom variables.

IPR check:

Timeframe: Final imputed genotype and phenotype data to be produced by end of September 2016. Analysis and writing to occur by March 2017.

Additional comments:

Synopsis for EU-GEI Publication