[3] *Washington University School of Medicine*

Background: Polygenic risk score represents weighted genetic disposition for some disease and is based on individual effects of multiple markers in the DNA. In the standard approach the score is the sum of GWAS weights for the risk allele. Selecting a p-value threshold, a polygenic score for each individual is calculated by summing over the number of alleles for each SNP-the most common type of genetic variation that involves a single nucleotide substitution in a DNA sequence weighted by the effect size drawn from a GWAS. The score then represents the composite additive effect of these multiple variants. Palindromic SNPs are removed in most cases to match exact allele in discovery and target samples. Clumping based on some known reference panel (for example, 1000G phase 3 panel based on 2502 individuals) is used to select independent or less correlated SNPs. Desirable reference panel and also discovery sample must have the same or close ancestry with target sample to ensure appropriate clumping and scoring.
Methods: Ancestries other than European American and African American have very limited number of discovery samples having the same ancestry and most of reference panels have small number of individuals. For example, ancestry subsets of 1000G phase 3 other than EUR and AFR are too small. Thus, creating PRS for other ancestries requires special approaches involving minor ancestries. One of recent methods uses combination of EUR (or AFR) based PRS with minor ancestry PRS. In the case when there is no discovery sample having the same minor ancestry it is possible to construct PRS using part of the sample as target and create weights based on other part of the sample as suggested in (Ref: Márquez-Luna C. (2016). Multiethnic polygenic risk scores improve risk prediction in diverse populations. bioRxiv preprint first posted online May. 2, 2016; doi: http://dx.doi.org/10.1101/051458). 10 fold cross-validation method is used to create minor ancestry PRS. The proposed method also incorporates ancestry principal components to improve prediction accuracy. PRS based on superpopulation weights, PRS with 10 fold cross-validation and two ancestry PC are optimized to get highly predictable overall PRS.
Results: Spit for Science sample has two main superpopulations (EUR and AFR) with satisfactory numbers of individuals. PRS for minor ancestry groups (say, for AFR sample) is calculated as an optimal linear combination of PRS based on EUR discovery sample, AFR discovery sample, 10 fold cross-validated method using AFR sample itself and two ancestral PCs. Optimal combination is determined as a fitted model value of the regression between disease and mentioned variables.
Discussion: Our analyses show that risk prediction significantly improves when incorporating risk scores based on multiple approaches.

Disclosure: Nothing to disclose.

doi: 10.1016/j.euroneuro.2018.08.105

F26
"GWAS QUALITY SCORE" (GQS) FOR ASSOCIATED REGIONS IN GWAS META-ANALYSES

Swapnil Awasthi[1], C. Anthony Altar[2], Stephan Ripke[3]

[1] *Charite Universitatsmedizien*
[2] *Assurex Health, Inc.*
[3] *MGH*

Background: In the last decade, genome-wide association studies (GWAS) have helped identify 18,257 distinct SNPs associated with various traits.2 The GIANT consortium alone recently identified 3,290 independent SNPs associated with height via a GWAS meta-analysis of 700,000 individuals.1 Due to the growing number of such loci, visual inspection of associated regions becomes cumbersome, provides only a subjective evaluation of GWAS significance, and is increasingly neglected. So, it becomes important to systematically separate spurious from true signals. Here we propose a novel method that evaluates the relationship between test statistics and linkage disequilibria (LD) within a genomic region with an associated independent SNP. To these, our method systematically assigns a general quality score and flags regions for secondary inspections.
Methods: Our method is based on the assumption that SNPs within an associated region have test statistics proportional to their LD with the region-index-SNP. To evaluate this criterion, we fit a linear model between the negative log of p-value and LD-r2 of each SNP. Further, we apply regression diagnostic methods to identify those SNPs that do not follow this linear trend. Specifically, we use studentized residuals and leverages to classify SNPs that have a combination of extreme p values and extreme LD with respect to the regression line. We used hundreds of examples and compared this method to the commonly used visual inspection of associated regions.
Results: We generated hundreds of RP-plots of significantly associated regions for various traits. Visually unreliable regions were identified by their high number of extreme outlier SNPs. These showed high leverages and high standardized residuals, thus receiving low-quality scores. Simultaneously our method identified index SNPs with suspiciously high significance as well.
Discussion: Many GWAS analyses have such a high number of associated regions that researchers may not have the capacity or a reliable method to perform careful visual inspections. The objective statistical tool described here will automatically identify problematic regions that warrant a secondary inspection.

Disclosure: Nothing to disclose.

doi: 10.1016/j.euroneuro.2018.08.106

F27
GENE EXPRESSION DEPENDENCY AND INTERACTION OF FUNCTIONAL ANNOTATION IN GWAS

Christiaan de Leeuw[1], Danielle Posthuma[1], Tom Heskes[2]

[1] *Vrije University*
[2] *Radboud University*

Background: The swift growth in sample sizes has seen a tremendous increase in the discovery of genetic associations but understanding how associated variants actually affect their phenotype remains a challenge. Although gene annotation analysis methods such as gene-set analysis and network analysis have been successfully applied to gain insight into underlying biology, much remains unknown. In part this is due to the nature of such methods, which are often limited in both the types of annotation that can be analysed and the hypotheses that can be tested.

To address this issue, we have extended and generalized our gene-set analysis framework. In the present research we have applied it to the relation of functional annotation associations in GWAS with tissue-specific gene expression and gene co-expression. Gene associations are often found to correlate with patterns of gene expression in specific phenotype-relevant tissues, as well as correlating across co-expression structures. Integrating such information therefore allows the effects of functional annotations to be analysed and understood in that broader context.
Methods: We generalized the regression framework of our gene-set analysis tool MAGMA to allow for simultaneous analysis of functional gene sets and gene expression levels, using conditional and interaction models to evaluate how issue specificity affects functional associations. We further extended it to also allow inclusion of random effect components, which can be used to model co-expression between genes. This can be used to investigate the possibility of indirect functional involvement of associated genes via co-expression.
Results: Analysis of a range of different phenotypes revealed numerous dependencies and interactions between functional annotation and gene expression. For many phenotypes the expression in specific tissues was found to be predictive of genetic association and confounded the associations of some of the functional gene sets. Moreover, interaction analysis showed that many gene sets affected the phenotype only in conjunction with tissue-specific expression. For such gene sets, only or primarily genes that were both in that functional category and expressed in phenotype-relevant tissues showed genetic association. Often these gene sets showed little or no marginal association, making then very difficult to detect in a standard gene-set analysis.
Discussion: The paths from genetic mutation to phenotype are long and winding, making them difficult to discern. Combining genetic data with information from different levels of biology can help us better understand those paths, but sophisticated statistical models are needed to do justice to the complexity of that biology. The methods presented here are a step in that direction, and our analyses show that they can successfully be used to combine different sources of biological data to obtain a more refined insight into the genetic etiology of the phenotype.

Disclosure: Nothing to disclose.

doi: 10.1016/j.euroneuro.2018.08.107

F28
SINGLE-CELL ENRICHMENT ANALYSIS FOR THE IDENTIFICATION OF TRAIT-RELEVANT GENES AND BRAIN CELL TYPES IN PSYCHIATRIC DISORDERS

Diogo Marinho Almeida [1], Samuel Demharter [2], iPSYCH-BROAD Working Group, Thomas Werge [1], Konstantin Khodosevich [2], Alfonso Demur [1]

[1] *Institut for Biologisk Psykiatri*
[2] *University of Copenhagen*

Background: Genome-Wide Association Studies allowed the characterization of several genetic traits which translates to several SNPs involved in the development of complexes diseases such as diabetes, autoimmune diseases and schizophrenia. Alongside with all GWAS data generated in the past years, neuronal single cell gene expression data has been also produced and plays an important role to unravel the genetic nature of psychiatric and neurological disorders when integrated with GWAS summary statistics data. In order to fill up the gap between GWAS studies and single cell gene expression analysis we present a simple and promising enrichment strategy to identify trait-relevant genes and cell types underlying psychiatric disorders.
Methods: The enrichment method consists in a pipeline of four basic steps: (1) the integration of the software MAGMA to perform gene-based analysis in the summary statistics level; (2) a Wilcoxon rank-sum test for identification of exclusive cell-type differentially expressed genes; (3) a Bonferroni correction of the p-values; (4) fisher exact test statistics to check the level of significance in the overlapping genes from (1 and 2).
Results: The pipeline was primary tested using the GWAS summary statistics from the League against epilepsy along with two normalized single-cell gene expression data-sets from the visual cortex and the frontal cortex, with 19110 and 10420 single cells respectively. The enrichment highlighted one epilepsy relevant gene SCN1A in the visual cortex area in one inhibitory neuron ($P=0.02244$) and two excitatory neurons ($P=0.03527$ and $P=0.015410$).
Discussion: Despite the simplicity, this pipeline was able to identify one epilepsy-relevant gene, already known to be involved with generalized epilepsy, febrile seizure and epileptic encephalopathy. The dichotomy of this pipeline is very dependent on the p-value correction. The stringency of the Bonferroni method influences a lot the distribution of the number of overlapping genes from the MAGMA output and the differentially expressed gene analysis. Moreover, new tests of this pipeline using GWAS summary statistics of epilepsy comorbid with several other psychiatric disorders (schizophrenia, autism, depression, anorexia and bipolar disorder) are on the way.

Disclosure: Nothing to disclose.

doi: 10.1016/j.euroneuro.2018.08.108