A Small Wiki for Personality Genetics

Finn Årup Nielsen

The Lundbeck Foundation Center for Integrated Molecular Brain Imaging;
Department of Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark;
Neurobiology Research Unit, Rigshospitalet, Copenhagen, Denmark.

Summary

A fielded wiki for genetic association studies with personality traits is described that features easy entry, on-the-fly meta-analysis of effect sizes and forest and funnel plotting with export of data in different formats.

Background

Our experience with a meta-analytic neuroinformatics database, the *Brede Database*,[1] shows that data entry is a bottleneck. Web-based collaborative entry with a wiki could be a solution to speed up the process. To explore this idea a small wiki was constructed for a well-defined field: Personality genetics.

The wiki idea has inspired a number of other projects within bioinformatics, e.g., *WikiProteins*,[2] *WikiGenes*,[3] *SNPedia* and the automatic setup of gene articles in *Wikipedia*.[4] Non-wiki web-based systems with genetic association studies and meta-analysis are also in operation, e.g., *AlzGene*.[5]

Many genetic association studies on personality traits exist, and several meta-analyses have already been performed: On a serotonin transporter gene polymorphism (5-HTTLPR) and anxiety-related traits.[6–10] But also other polymorphisms and traits have been examined in meta-analyses.[6, 11, 12] These meta-analyses typically focus on a single personality trait and a single polymorphism. The wiki system presented here handles multiple polymorphisms across multiple personality traits.

## Genetic association data in the wiki web interface

Rather than a free-form text-based wiki a fielded wiki framework was setup where data was structured in a table with fixed columns: Gene symbol, polymorphism, genotype, personality test, personality trait, personality score mean and standard deviation, recruitment group and PubMed identifier. Each row represents one personality trait measured for one group of humans with a specific genotype reported in one specific published paper.

So far data from 22 papers have been entered amounting to 26 different personality traits from 7 different test batteries, 9 different polymorphisms from 7 different genes, 20 different subject groups (patient groups as well as healthy). It adds up to a total of 523 trait/genotype personality score values.

Technical details

The genetic information is identified by the standard gene symbol and a short string for the polymorphisms and genotypes with the variant nucleotides alphabetically sorted.

Editors of the wiki are able to easily add and edit the table and see revisions. Presently, the wiki is "fully open"[13] which means that everyone is able to add and edit. However, a simple cookie-based authentication scheme is in place enabling the switch to a "lockable" wiki type.
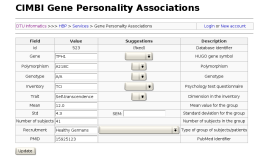
On the fly plots of the mean and standard deviations of both raw data and effect sizes are constructed in the image format *Scalable Vector Graphics* (SVG) that allows hyperlinks.

Data export to comma-separated values file allows more flexible and advanced off-line analyses, and export to *MediaWiki* templates make it possible to include the data in *Wikipedia* or any other *MediaWiki*-based wiki.

Where possible the individual items have a link to *PubMed*, and the gene symbols and polymorphisms link to corresponding pages on *Wikipedia*. One single script in the programming language *Python* with a *SQLite* database backend implements the web-service.

Effect size

An *effect size* d is defined as a standardized mean difference

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}, \qquad (1)$$

where $\bar{x}_1$ is the mean personality score for one group of subjects, $\bar{x}_2$ is the mean for a comparison group and s is the standard deviation within the groups. Slight variations exist for this computation and the present application uses Hedges' unbiased d for the computation.[14]

Presently effect sizes are computed automatically for all homozygotes-homozygotes comparisons for each polymorphism and personality trait examined in each paper.

Meta-analysis computation

Given several studies of a single polymorphism/trait pair a meta-analytic estimate of a common effect size is possible.

Hedges and Olkin suggest a large sample approximation for the variance of d of the kth study as

$$\hat{\sigma}^2(d_k) = \frac{n_{1,k} + n_{2,k}}{n_{1,k} n_{2,k}} + \frac{d_k^2}{2(n_{1,k} + n_{2,k})}. \qquad (2)$$

A combined effect size d_+ among K studies may be found as

$$d_+ = \hat{\sigma}^2(d_+) \sum_{k=1}^{K} \frac{d_k}{\hat{\sigma}^2(d_k)}, \qquad (3)$$

where the variance of the combined effect, $\hat{\sigma}^2(d_+)$, is estimated as

$$\hat{\sigma}^2(d_+) = \left(\sum_{k=1}^{K} \frac{1}{\hat{\sigma}^2(d_k)} \right)^{-1}. \qquad (4)$$

With a large sample Gaussian approximation the combined effect size and its variance may be used in a comparison against the normal distribution to obtain P-value and confidence intervals.

Effect sizes and meta analysis in the wiki
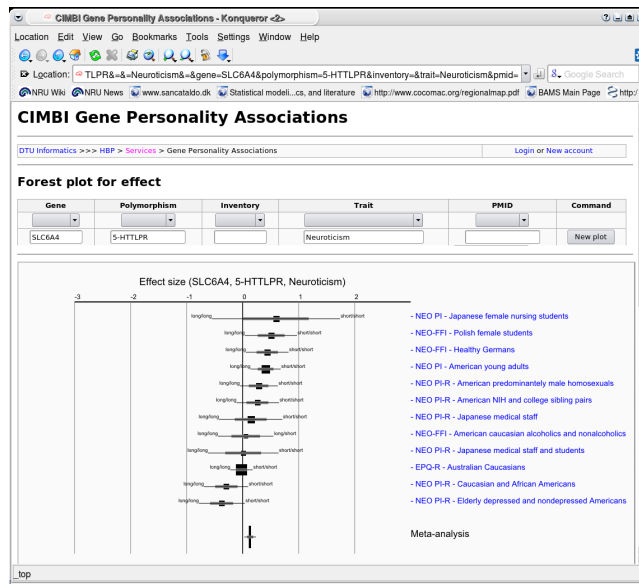
The effect size and meta-analysis computation allows for the generation of *forest plots* with the standard deviation and 95% confidence interval for the estimate for both the individual effect sizes (a total of 178) and the combined meta-analytic effect size. Also *funnel plots* can be made. For the well-studied polymorphism (5-HTTLPR) in the serotonin transporter gene (*SLC6A4*) the wiki database can presently compute 12 effect sizes against the 'neuroticism' trait.

The fixed effect meta-analysis shows that the polymorphism has a small and significant association. However, the Cochran's test for inhomogeneity[15] shows that the studies are confounded by between-study variation. The variation can be estimated with a random effect model, but it does not point to its origin. Some variation may be due to subject group differences and possible personality test differences. Variations in the geographic distribution of personality traits have been observed,[16, 17] but should these variations be carried over to gene/trait interactions?

Observations

A first attempt on a meta-analytic bioinformatics fielded wiki was made. The data of personality genetics has relatively well-defined entities, is easy to represent, analyze and visualize compared to typical neuroinformatics data.

In principle the personality genetics wiki allows for mass meta-analysis across multiple personality traits and multiple polymorphisms.

The present structure does not make it possible to record or analyze combined effects of polymorphisms, and the present wiki lacks discussion pages ("Talk pages") or any other means for supporting social interaction among users and editors.

References

[1] Nielsen FÅ. The Brede database: a small database for functional neuroimaging. *NeuroImage*, 2003;19. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, 2003, New York, NY. Available on CD-Rom.

[2] Mons B, et al. Calling on a million minds for community annotation in wikiproteins. *Genome Biology*, 2008;9:R89.

[3] Hoffman R. A wiki for the life sciences where authorship matters. *Nature Genetics*, 2008;40:1047–1051.

[4] Huss, III JW, et al. A gene wiki for community annotation of gene function. *PLoS Biology*, 2008;6:e175.

[5] Bertram L, et al. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature Genetics*, 2007;39:17–23.

[6] Munafò MR, et al. Genetic polymorphisms and personality in healthy adults: A systematic review and meta-analysis. *Molecular Psychiatry*, 2003;8:471–484.

[7] Schinka JA, Busch RM, and Robichaux-Keene N. A meta-analysis of the association between the serotonin transporter gene polymorphism (5-HTTLPR) and trait anxiety. *Molecular Psychiatry*, 2004;9:197–202.

[8] Sen S, Burmeister M, and Ghosh D. Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)*, 2004;127B:85–89.

[9] Munafò MR, Clark TG, and Flint J. Are there sex differences in the association between the 5HTT gene and neuroticism? a meta-analysis. *Personality and Individual Differences*, 2004;37:621–626.

[10] Munafò MR, Clark T, and Flint J. Does measurement instrument moderate the association between the serotonin transporter gene and anxiety-related personality traits? a meta-analysis. *Molecular Psychiatry*, 2005;10:415–419.

[11] Schinka JA, Letsch EA, and Crawford FC. DRD4 and novelty seeking: Results of meta-analyses. *American Journal of Medical genetics (Neuropsychiatric Genetics)*, 2002;114:643–648.

[12] Munafò MR, et al. Association of the dopamine D4 receptor (drd4) gene and approach-related personality traits: meta-analysis and new data. *Biological Psychiatry*, 2008;63:197–206.

[13] Leuf B and Cunningham W. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley, Boston, 2001.

[14] Hedges LV and Olkin I. *Statistical Methods for Meta-Analysis*. Academic Press, Orlando, Florida, 1985.

[15] Hartung J, Knapp G, and Sinha BK. *Statistical Meta-Analysis with Applications*. Wiley Series in Probability and Statistics. Wiley, Hoboken, New Jersey, 2008.

[16] Schmitt DP, et al. The geographic distribution of big five personality traits, patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, 2007;38:173–212.

[17] Rentfrow PJ, Gosling SD, and Potter J. A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science*, 2008;3:339–369.