A fielded 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 (a highly structured 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

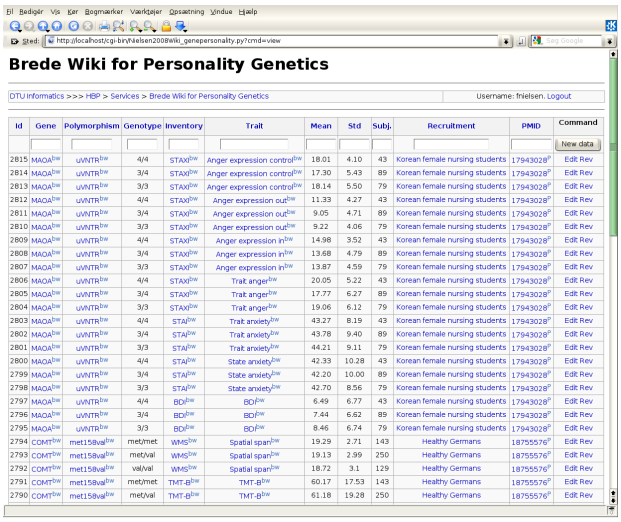
Our experience with a meta-analytic neuroinformatics database, the *Brede Database*,[1] shows that data entry is a bottleneck. Web-based collaborative entry with a wiki could be a solution to speed up the process. To explore this idea a small wiki was constructed for a well-defined field: Personality 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 researchers perform meta-analyses across polymorphisms and personality traits.[6]

Semantic wikis may represent the data. However, these wikis will often lack the ability to do advanced numerical computations or the specialized visualizations that meta-analysts want. The fielded wiki represents the data as well as implements the specialized computational and visualization functions.

Genetic association data in the wiki web interface



Id	Gene	Polymorphism	Genotype	Inventory	Trait	Mean	Std	Subj.	Recruitment	PMID	Command
											New data
2815	MAOA^{bw}	uVNTR^{bw}	4/4	STAXI^{bw}	Anger expression control^{bw}	18.01	4.10	43	Korean female nursing students	17943028^{bw}	Edit Rev
2814	MAOA^{bw}	uVNTR^{bw}	3/4	STAXI^{bw}	Anger expression control^{bw}	17.30	5.43	89	Korean female nursing students	17943028^{bw}	Edit Rev
2813	MAOA^{bw}	uVNTR^{bw}	3/3	STAXI^{bw}	Anger expression control^{bw}	18.14	5.50	79	Korean female nursing students	17943028^{bw}	Edit Rev
2812	MAOA^{bw}	uVNTR^{bw}	4/4	STAXI^{bw}	Anger expression out^{bw}	11.33	4.27	43	Korean female nursing students	17943028^{bw}	Edit Rev
2811	MAOA^{bw}	uVNTR^{bw}	3/4	STAXI^{bw}	Anger expression out^{bw}	9.05	4.71	89	Korean female nursing students	17943028^{bw}	Edit Rev
2810	MAOA^{bw}	uVNTR^{bw}	3/3	STAXI^{bw}	Anger expression out^{bw}	9.22	4.06	79	Korean female nursing students	17943028^{bw}	Edit Rev
2809	MAOA^{bw}	uVNTR^{bw}	4/4	STAXI^{bw}	Anger expression in^{bw}	14.98	3.52	43	Korean female nursing students	17943028^{bw}	Edit Rev
2808	MAOA^{bw}	uVNTR^{bw}	3/4	STAXI^{bw}	Anger expression in^{bw}	13.68	4.79	89	Korean female nursing students	17943028^{bw}	Edit Rev
2807	MAOA^{bw}	uVNTR^{bw}	3/3	STAXI^{bw}	Anger expression in^{bw}	13.87	4.59	79	Korean female nursing students	17943028^{bw}	Edit Rev
2806	MAOA^{bw}	uVNTR^{bw}	4/4	STAXI^{bw}	Trait anger^{bw}	20.05	5.22	43	Korean female nursing students	17943028^{bw}	Edit Rev
2805	MAOA^{bw}	uVNTR^{bw}	3/4	STAXI^{bw}	Trait anger^{bw}	17.77	6.27	89	Korean female nursing students	17943028^{bw}	Edit Rev
2804	MAOA^{bw}	uVNTR^{bw}	3/3	STAXI^{bw}	Trait anger^{bw}	19.06	6.12	79	Korean female nursing students	17943028^{bw}	Edit Rev
2803	MAOA^{bw}	uVNTR^{bw}	4/4	STAI^{bw}	Trait anxiety^{bw}	43.27	8.19	43	Korean female nursing students	17943028^{bw}	Edit Rev
2802	MAOA^{bw}	uVNTR^{bw}	3/4	STAI^{bw}	Trait anxiety^{bw}	43.78	9.40	89	Korean female nursing students	17943028^{bw}	Edit Rev
2801	MAOA^{bw}	uVNTR^{bw}	3/3	STAI^{bw}	Trait anxiety^{bw}	44.21	9.11	79	Korean female nursing students	17943028^{bw}	Edit Rev
2800	MAOA^{bw}	uVNTR^{bw}	4/4	STAI^{bw}	State anxiety^{bw}	42.33	10.28	43	Korean female nursing students	17943028^{bw}	Edit Rev
2799	MAOA^{bw}	uVNTR^{bw}	3/4	STAI^{bw}	State anxiety^{bw}	42.20	10.00	89	Korean female nursing students	17943028^{bw}	Edit Rev
2798	MAOA^{bw}	uVNTR^{bw}	3/3	STAI^{bw}	State anxiety^{bw}	42.70	8.56	79	Korean female nursing students	17943028^{bw}	Edit Rev
2797	MAOA^{bw}	uVNTR^{bw}	4/4	BDI^{bw}	BDI^{bw}	7.44	6.62	43	Korean female nursing students	17943028^{bw}	Edit Rev
2796	MAOA^{bw}	uVNTR^{bw}	3/4	BDI^{bw}	BDI^{bw}	7.44	6.62	89	Korean female nursing students	17943028^{bw}	Edit Rev
2795	MAOA^{bw}	uVNTR^{bw}	3/3	BDI^{bw}	BDI^{bw}	8.46	6.74	79	Korean female nursing students	17943028^{bw}	Edit Rev
2794	COMT^{bw}	met158val^{bw}	met/met	WMS^{bw}	Spatial span^{bw}	19.29	2.71	143	Healthy Germans	18755576^{b}	Edit Rev
2793	COMT^{bw}	met158val^{bw}	met/val	WMS^{bw}	Spatial span^{bw}	19.13	2.99	250	Healthy Germans	18755576^{b}	Edit Rev
2792	COMT^{bw}	met158val^{bw}	val/val	WMS^{bw}	Spatial span^{bw}	18.72	3.1	129	Healthy Germans	18755576^{b}	Edit Rev
2791	COMT^{bw}	met158val^{bw}	met/met	TMT-B^{bw}	TMT-B^{bw}	60.17	17.53	143	Healthy Germans	18755576^{b}	Edit Rev
2790	COMT^{bw}	met158val^{bw}	met/val	TMT-B^{bw}	TMT-B^{bw}	61.18	19.28	250	Healthy Germans	18755576^{b}	Edit Rev

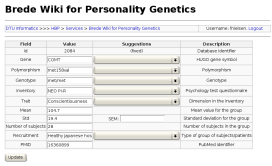
Rather than a free-form text-based wiki a fielded wiki framework was setup where data was structured in a table with fixed columns: Gene symbol, polymorphism, genotype, personality test, personality trait, personality score mean and standard deviation, recruitment group and PubMed identifier. Each row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

So far data from 87 papers have been entered amounting to 113 different traits from 25 different test batteries, 39 different polymorphisms from 24 different genes, 104 different subject groups (patient groups as well as healthy). It adds up to a total of 2815 trait/genotype personality score values.

Technical details

New information is easily entered in a form displayed as an extra row in the table. An autocomplete feature copies items from the previous row into the form fields, and autoreplace changes, e.g., 'n' to 'Neuroticism'.

Editors of the wiki are also able to edit the table and see revisions. A simple cookie-based authentication scheme is in place and the revisions are tracked with editor and time.



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 JSON or comma-separated values files allows more flexible and advanced offline analyses, and export to *MediaWiki* templates make it possible to include the data in *Wikipedia* or any other *MediaWiki*-based wiki, such as the *Brede Wiki*.

Where possible the individual items have automatically a deep link to *PubMed*, *Wikipedia* and the *Brede Wiki*.

One single script in the programming language *Python* with a *SQLite* database backend implements the web-service.

Effect size and meta-analysis

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.[7] An approximation of the combined effect size d_+ among K studies may be found as

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

These equations, together with the estimates of variance, are hard-coded into the wiki and the results from running through the entire data set of the wiki are added to a table that the user can view but cannot edit.

Furthermore, all meta-analytically combined effect sizes are added to another table, so that mass meta-analysis can be displayed across all traits and all polymorphims. The user initiates this computation.

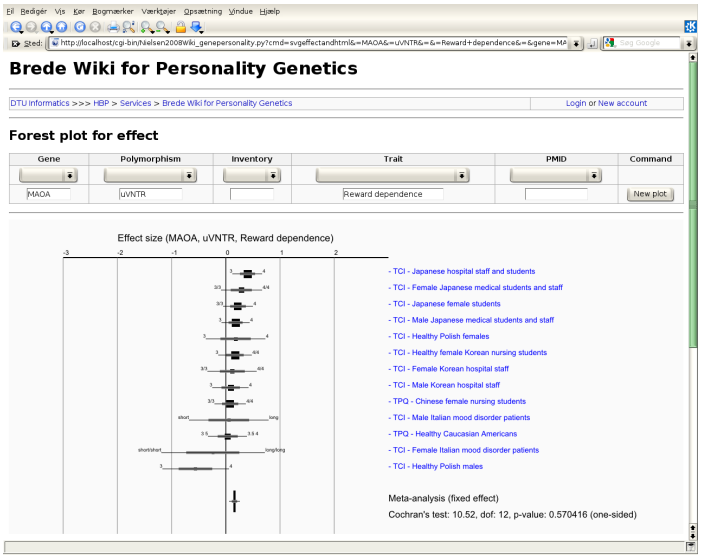
Display of data

Each of the four database tables (data, revisions, effects, meta-analysis) are displayed in HTML tables. By selecting a column the user can sorted items based on items in that column, and the user may also page through the rows of the table.

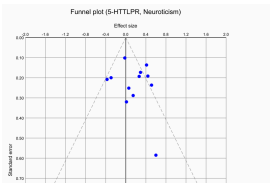
The displayed items can be filtered by following the hyperlink of the items, e.g., following a 'Neuroticism' link in the 'trait' column will select only the rows that have Neuroticism as trait.

When the data is exported to the Brede Wiki then template definitions formats the data as an HTML table. Items are hyperlinked, and sorting based on a column is possible, but filtering is not. No numerical computations or visulization are possible unless a MediaWiki-extension was developed.

Effect sizes and meta analysis in the wiki

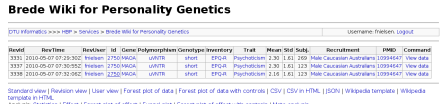


The effect size and meta-analysis computation allows for the generation of *forest plots* with the standard deviation and 95% confidence interval for the estimate for both the individual effect sizes (a present total of 992) and the combined meta-analytic effect size (presently 431).



Funnel plot.

Blue graphical elements in the SVG plots are hyperlinked to items in the tables. Via a form interface the user controls which data are shown.



Revision table.

Observations

A meta-analytic fielded wiki was construted, where users may enter new data, edit, view, compute and plot.

The present wiki lacks discussion pages ("Talk pages") or any other means for supporting social interaction among users and editors. It neither support additional fields, but extra information may be added on corresponding pages in the *Brede Wiki*.

While data may be represented in a MediaWiki the specialized fielded wiki allows for convenient and fast data entry helped by autocomplete features. Semantic MediaWiki[8] with form-based input provides an interface, but data entry would probably be more cumbersome, and the numerical and visualization functionality needs to be implemented with extra MediaWiki extensions.

Another alternative for a data set such as the one presented here would be online sharable spreadsheets, such as Google Docs. Ordinary spreadsheets have been used for meta-analysis and such spreadsheets have been distributed on the Web.[9] The fielded wiki may export its data in comma-separated files to spreadsheets.

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

[8] Völkel M, et al. Semantic wikipedia. In *Proceedings of the 15th international conference on World Wide Web*. ACM, New York, NY, USA, 2006; pages 585–594.

[9] Kempton MJ, et al. Meta-analysis, database, and meta-regression of 98 structural imaging studies in bipolar disorder. *Archives of General Psychiatry*, 2008; 65:1017–1032.