Neurodag 2004
Poster: #33

Meta-analytic clustering of molecular neuroimaging studies

Finn Årup Nielsen[1,2], Claus Svarer[1], Lars Kai Hansen[2], Gitte Moos Knudsen[1]

[1]Neurobiology Research Unit, Rigshospitalet, Copenhagen, Denmark.
[2]Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark

Aim

The aim of this study is to show that it is possible to extract relevant information about molecular imaging from published articles.

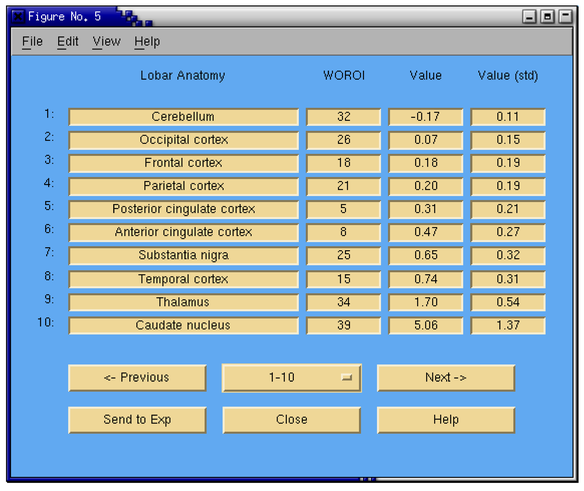
Introduction: Databasing molecular neuroimaging

The Brede Database [1] is a database that so far has recorded data from neuroimaging studies with the stereotaxic "Talairach" coordinates [2]. Such a representation allows for a variety of information extraction methods [3, 4, 5]. These coordinates will typically originate from cognitive neuroimaging studies with positron emission tomography or functional magnetic resonance imaging. Results from molecular neuroimaging studies with single-photon or positron emission tomography (PET) such as receptor imaging data are often reported in the form of values of interest across a set of brain regions, with only a few molecular neuroimaging studies using Talairach coordinates

We here report the first method for information extraction across multiple published molecular neuroimaging studies:

We used K-means to cluster experiments based on the normalized values of interest across brain regions. Molecular neuroimaging studies was entered in a database from published peer-reviewed articles via the Brede Neuroinformatics Toolbox, the brain regions were linked to an ontology, and an experiments-by-brain-regions data matrix was constructed. This matrix was subsequently analyzed with a hierarchical K-means algorithm.

Data: Brain region values



The example above is the graphical user interface for data entry with information from one of the four experiments reported in a dopamine receptor study [6]. This study reports estimated binding potentials across 11 brain regions with four different analysis techniques.

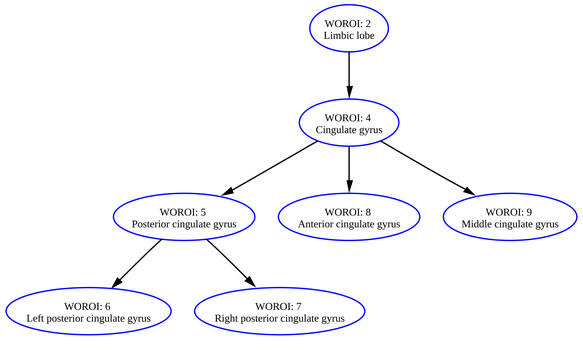
The values that are entered are primarily neuroreceptor specific, such as binding potentials and distribution volumes for neuroligands. A few other types of values were also entered, e.g., K_1 rate constants and distribution of compounds such as vinpocetine.

To account for the variability in the values of interest these values were normalized with "studentization", i.e., for each experiment the mean is extracted and the centered values are normalized with the standard deviation.

Brain anatomy ontology

The brain regions included will vary, e.g., some studies only report values for cerebral cortex while others include values for thalamus, putamen, caudate, pons, ... The detail of the segmentation also vary, e.g., with studies providing values from 4 to 21 regions [7, 8].

The segmentation used for the brain regions differs between studies. Typically the brain regions are only referred to by name, e.g., "frontal cortex", and the exact delineation is not given. It is difficult to know exactly what subregion the same designation refers to when used in different studies. We have presently chosen to ignore this problem and link the brain regions in the individual studies to the "nearest" items in a taxonomy (a simple ontology).



The diagram above shows part of the taxomony for cingulate regions.

Only regions where more than two experiments report values were included, but even then the experiment-by-brain-region data matrix contains many "wholes", i.e., "missing values".

Method: K-means clustering

K-means is a simple standard cluster analysis algorithm where the intra-component variance is minimized, see, e.g., [9]:

Initialization

1. Select K objects randomly and use these as component centroids

Iterative scheme

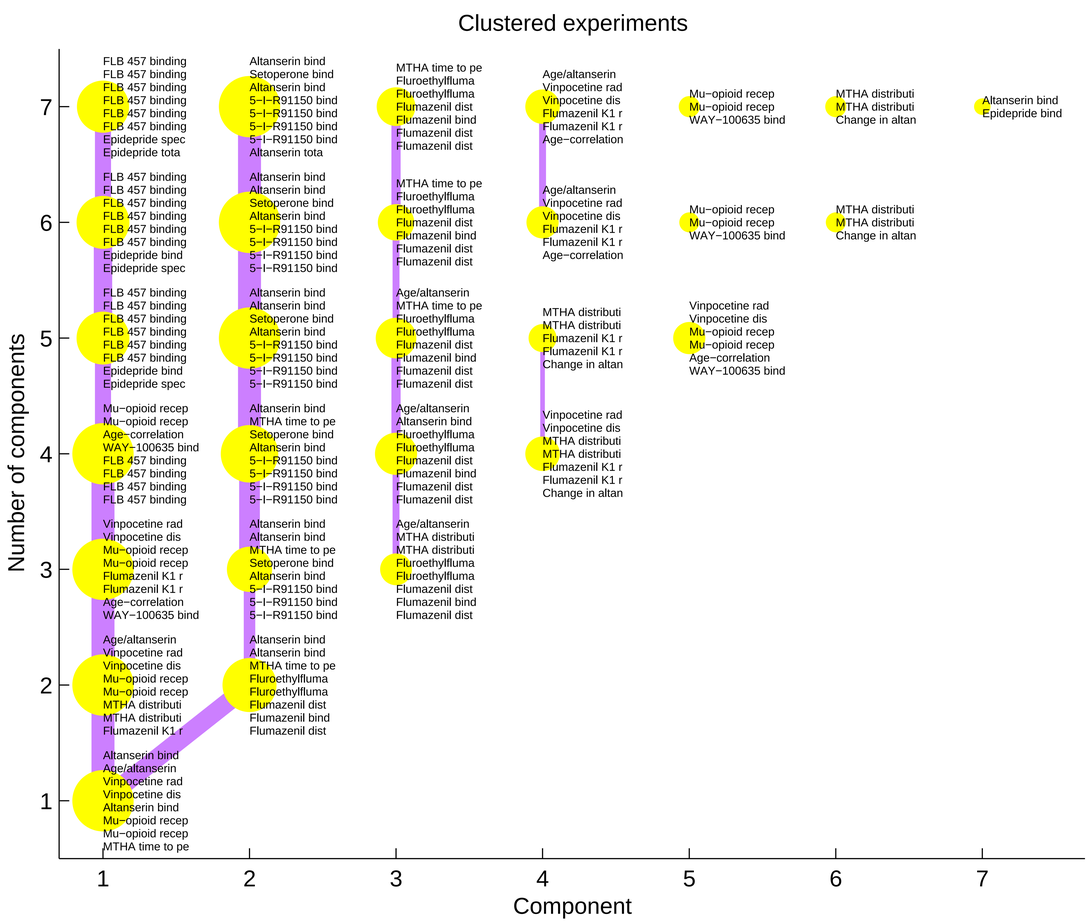
1. Calculate distances between all objects and all centroids and reassign the objects to the closest centroid.
2. Calculate the centroid for each component using all of their assigned objects.

The K-means algorithm was modified to handle "missing values" in the data matrix. The missing value K-means typically converges to unfavorable local minima and the algorithm is run multiple times with different initialization assignments. The run with the lowest intra-component variance is selected.

The number of components in the clustering was varied and a graphical technique was used to get an overview of the components [10].

Results: Hierarchical K-means clustering of experiments

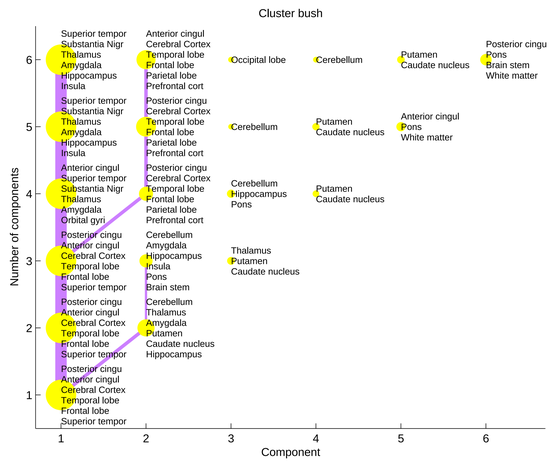
We have presently extended the database of region-based molecular neuroimaging studies to 43 experiments contained in 18 articles. Combined they report values in 68 different brain regions, and it results in a 43-by-68 matrix. After excluding regions where less than 3 experiments report values the matrix is sized 43-by-30, i.e., excluding 38 regions.



Results of the clustering of experiments appear above. The yellow circles each denotes a component in the cluster analysis and their sizes indicate how many experiments are assigned to the component. The x-axis is each an individual component and the y-axis is the number of components K in the clustering. The lines between the yellow circles show how related the components are.

The present database contains predominantly studies which target the receptor types serotonin-2A (5-HT_{2A}, with altanserin, setoperone and 5-I-R91150 radiotracers), dopamine D_2 (with FLB 457 and epidepride) and benzodiazepine (with flumazenil and fluoroethylflumazenil). These three types of experiments group in three separate components.

Results: Brain regions



Instead of clustering experiments it is also possible to cluster brain regions. As shown above, this type will tend to cluster cerebral cortex regions together.

By examining the centroid from the clustering of experiments it is possible to determine which brain regions have the highest values within a component of clustered experiments, e.g., for the dopamine studies with the FLB 457 and epidepride tracers the highest values appear (not surprisingly) in putamen and caudate nucleus. For the component with (5-HT_{2A}) receptor experiments the regions are anterior cingulate gyrus, parietal, prefrontal and occipital cortex.

Summary

Without knowledge of the target under study our method is able to group molecular neuroimaging studies according to the neuroreceptor under study.

Our approach is completely automated once the data has been entered in the database.

Acknowledgment & Availability

Finn Årup Nielsen is funded by the Villum Kann Rasmussen Foundation.

The tools for the analysis are available in the Brede neuroinformatics toolbox [11] presently available from http://hendrix.imm.dtu.dk/software-/brede/.

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] Talairach J and Tournoux P. *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme Medical Publisher Inc, New York, 1988.

[3] Fox PT, et al. Functional volumes modeling: Theory and preliminary assessment. *Human Brain Mapping*, 1997;5:306–311.

[4] Nielsen FÅ and Hansen LK. Modeling of activation data in the BrainMap™ database: Detection of outliers. *Human Brain Mapping*, 2002;15:146–156.

[5] Nielsen FÅ and Hansen LK. Finding related functional neuroimaging volumes. *Artificial Intelligence in Medicine*, 2004;30:141–151.

[6] Cselényi Z, et al. Wavelet-aided parametric mapping of cerebral dopamine D_2 receptors using the high affinity PET radioligand [^{11}C]FLB 457. *NeuroImage*, 2002; pages 47–60.

[7] Fujita M, et al. Effect of scatter correction on the compartmental measurement of striatal and extrastriatal dopamine D_2 receptors using [^{123}I]epidepride SPET. *European Journal of Nuclear Medicine and Molecular Imaging*, 2004;31:644–654.

[8] Rabiner EA, et al. A database of [(11)C]WAY-100635 binding to 5-HT(1A) receptors in normal male volunteers: normative data and relationship to methodological, demographic, physiological, and behavioral variables. *NeuroImage*, 2002;15:620–632.

[9] Goutte C, et al. On clustering fMRI time series. *NeuroImage*, 1999;9:298–310.

[10] Nielsen FÅ, Balslev D, and Hansen LK. Mining posterior cingulate. *NeuroImage*, 2004;22. Presented at the 10th Annual Meeting of the Organization for Human Brain Mapping, June 14–17, 2004, Budapest, Hungary. Available on CD-ROM.

[11] Nielsen FÅ and Hansen LK. Experiences with Matlab and VRML in functional neuroimaging visualizations. In Klasky S and Thorpe S, eds., *VDE2000 - Visualization Development Environments, Workshop Proceedings, Princeton, New Jersey, USA, April 27–28, 2000*. Princeton Plasma Physics Laboratory, Princeton, New Jersey, 2000; pages 76–81.