Databasing molecular neuroimaging studies

Finn Årup Nielsen

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

July 4, 2006

Databasing molecular neuroimaging studies

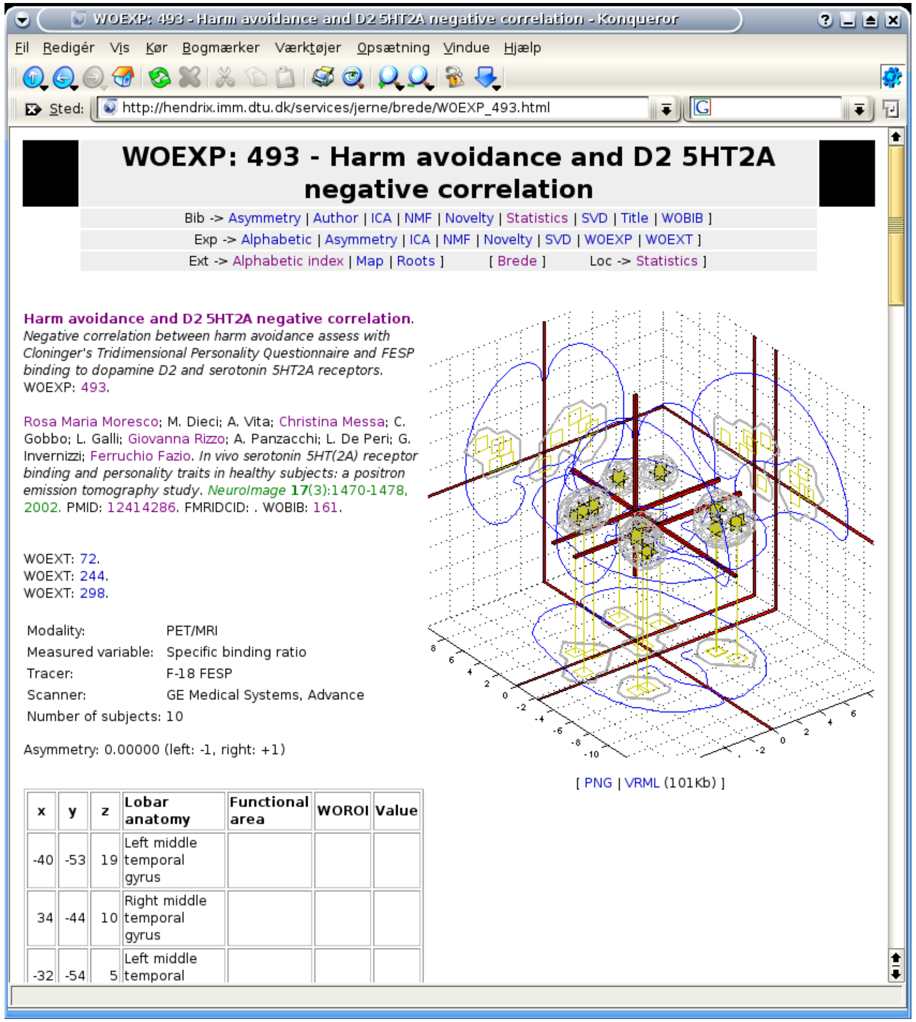
Issues in databasing information from the human molecular neuroimaging literature are:

1. Find a suitable representation for the data: How should the results from a typical molecular neuroimaging study be represented?

2. Construct information retrieval tools: Is it possible, e.g., to find "Related Articles" based on the quantitative neuroimaging results presented in papers?

3. Construct analysis tools: Is it possible to develop methods to extract information across studies?

Brede Database



Contains stereotaxic Talairach coordinates, cf. BrainMap (Fox and Lancaster, 1994).

Search on “Related Articles” based on stereotaxic coordinates possible (Nielsen and Hansen, 2004).

Multivariate analysis of abtract text and voxelized stereotaxic coordinates (Nielsen et al., 2004; Nielsen et al., 2005)

Available on the Internet in HTML and XML (Nielsen, 2003).

Handling molecular neuroimaging

Most molecular imaging studies relies on analysis of values from brain regions and report descriptive statistics for these values. There are not many voxel-based studies reporting in stereotaxic space, thus the standard Brede Database framework cannot be used.

There are two significant difficulties when comparing molecular neuroimaging studies:

1. Regions differ between studies: E.g., some include values for "temporal cortex" others do not. Solution: Build a taxonomy for brain regions and use imputation.

2. Measured and reported values differ between studies and they are not comparable: Tracers and receptors; transport rates (e.g., K_1), distribution volume, binding potentials; different methods to compute the values. Solution: Use some kind of normalization.

Entry of data

Information from molecular neuroimaging studies are entered in much the same way as for stereotaxic coordinate-based studies using the Brede Toolbox (Nielsen and Hansen, 2000).

Instead of stereotaxic coordinate the reported regional value is entered as well as an identifier for the associated item in a brain region taxonomy.

Extra information include bibliographic information and description of the experiment, such as tracer and scanner. Furthermore the experiment is linked to items in a taxonomy for so-called "external components", which, e.g., relate the individual neuroreceptors in a directed acyclic graph (Nielsen, 2005).

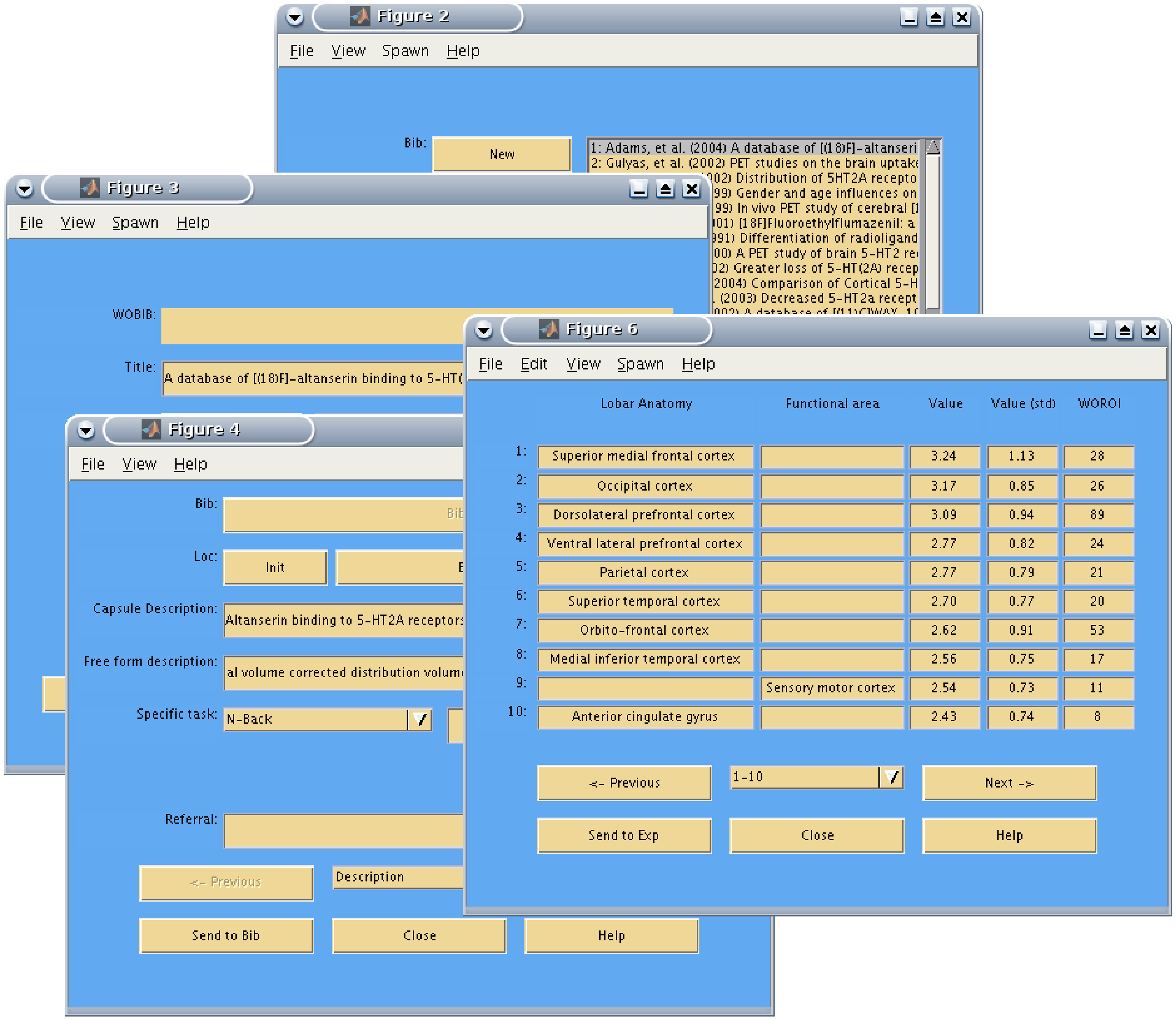
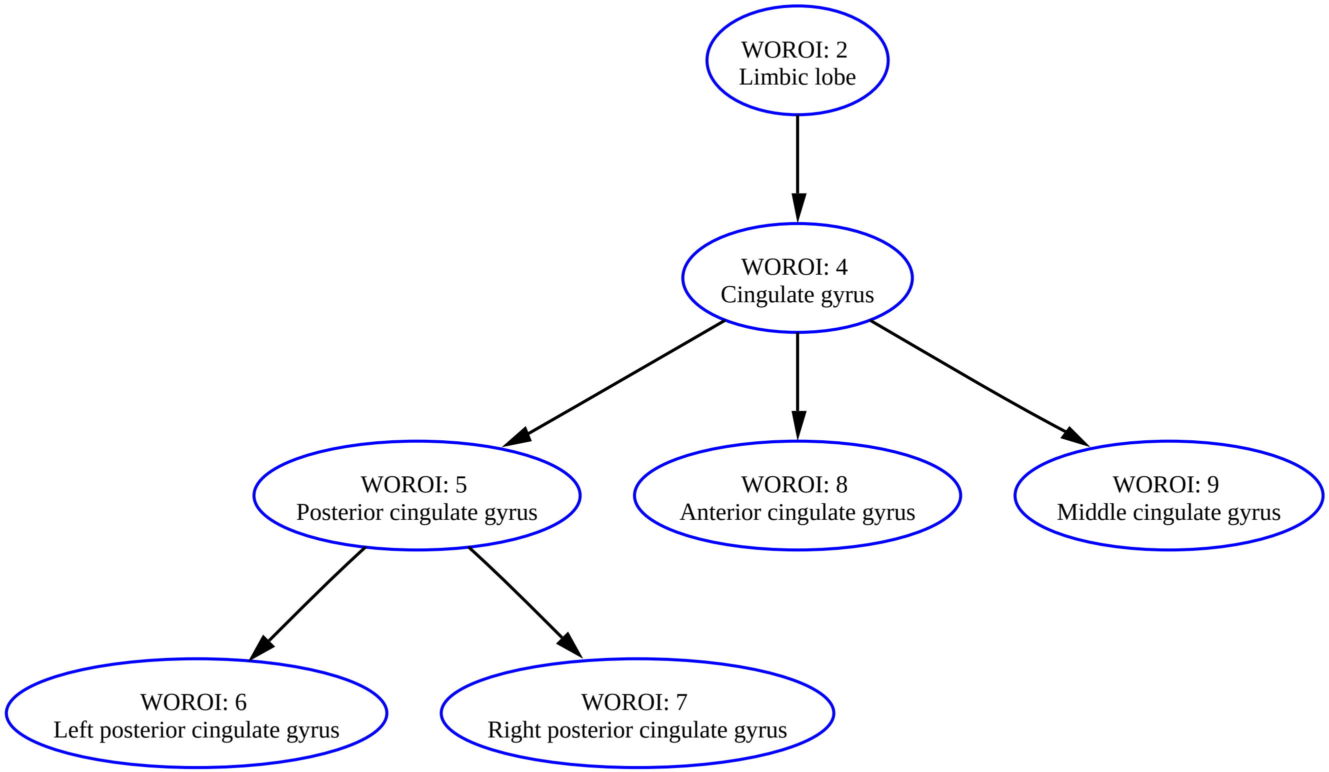


Figure 2
File View Spawn Help
Bib:
New
1: Adams, et al. (2004) A database of [(18)F]-altanseri
2: Gulyas, et al. (2002) PET studies on the brain uptake
Figure 3
File View Spawn Help
WOBIB:
Title:
A database of [(18)F]-altanserin binding to 5-HT(
Figure 4
File View Help
Bib:
Loc:
Init
Capsule Description:
Altanserin binding to 5-HT2A receptors
Free form description:
al volume corrected distribution volum
Specific task:
N-Back
Referral:
<- Previous
Description
Send to Bib
Close
Help
Figure 6
File Edit View Spawn Help
Lobar Anatomy
Functional area
Value
Value (std)
WOROI
1: Superior medial frontal cortex 3.24 1.13 28
2: Occipital cortex 3.17 0.85 26
3: Dorsolateral prefrontal cortex 3.09 0.94 89
4: Ventral lateral prefrontal cortex 2.77 0.82 24
5: Parietal cortex 2.77 0.79 21
6: Superior temporal cortex 2.70 0.77 20
7: Orbito-frontal cortex 2.62 0.91 53
8: Medial inferior temporal cortex 2.56 0.75 17
9: Sensory motor cortex 2.54 0.73 11
10: Anterior cingulate gyrus 2.43 0.74 8
<- Previous
1-10
Next ->
Send to Exp
Close
Help

Brain region taxonomy



Hierarchical taxonomy of brain regions records which brain areas are a part of other brain areas.

Imputation: If "left posterior cingulate" and "right posterior cingulate" values are available in a specific study these are used to define a value for "limbic lobe" — if this is not available.

Brain region taxonomy in the Brede database

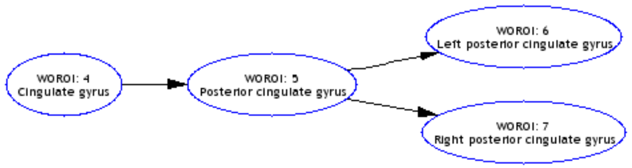
WOROI: 5 - Posterior cingulate gyrus

Bib -> Asymmetry | Author | ICA | NMF | Novelty | Statistics | SVD | Title | WOBIB]
Roi -> Alphabetic | Hammers | Tzourio-Mazoyer]
[Brede]

WOROI: 5 - Posterior cingulate gyrus

Abbreviation: **PCgG**

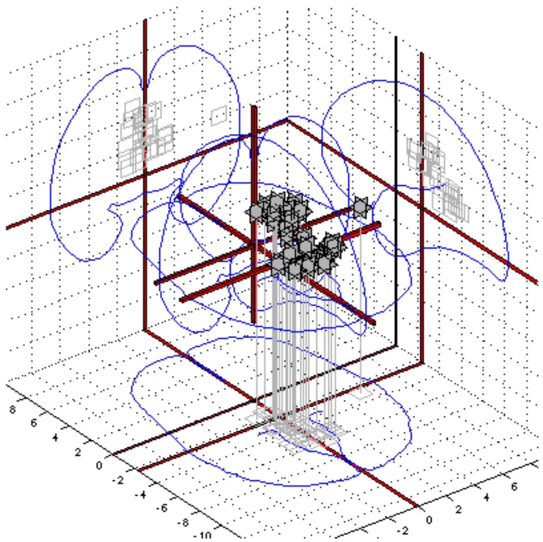
Variation: *posterior cingulate*
Variation: *posterior cingulate area*
Variation: *posterior gyrus cinguli*
BrainInfo: 144



Parents	Siblings	Children
Cingulate gyrus		Left posterior cingulate gyrus Right posterior cingulate gyrus

Talairach coordinates

x	y	z	Lobar anatomy	WOBIB	WOEXP
6	-29	38	Right posterior cingulate gyrus and precuneus	21	66
9	-53	14	Right posterior cingulate gyrus	32	109
4	-53	14	Right posterior cingulate gyrus	32	110
2	-40	40	Posterior cingulate gyrus	35	117
52	-30	20	Right postcentral gyrus/posterior cingulate gyrus	35	119
-4	-36	24	Left posterior cingulate gyrus	41	135
-4	-35	29	Left posterior cingulate gyrus	41	137
-4	-35	40	Left posterior cingulate gyrus	41	138
0	-26	29	Posterior cingulate gyrus	41	140
-2	-48	20	Left posterior cingulate gyrus	49	164
-9	-33	46	Posterior cingulate gyrus	57	183
0	-17	28	Right posterior cingulate gyrus	60	186
3	-53	15	Right posterior cingulate gyrus	71	223



Brain region taxonomy included in the Brede Database.

Talairach coordinates extracted where their anatomical label corresponds to the item in the taxonomy.

Links to NIH MeSH, BrainInfo (NeuroNames) (Bowden and Martin, 1995), segmented volumes, IBVD, Wikipedia.

Data matrix

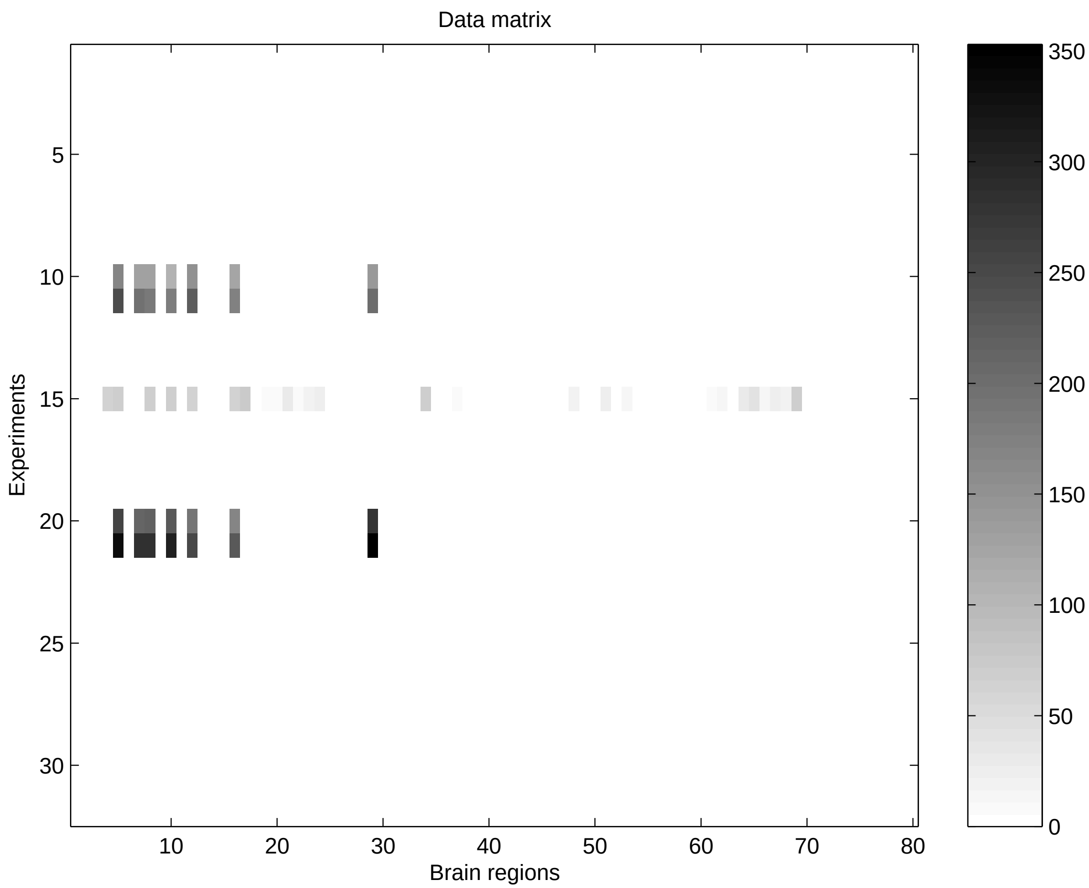


Figure 1: For serotonin-2A part of the datamatrix $\mathbf{X}(32\times80)$: Original matrix: $\approx$ 13% defined. "One-back" imputation: $\approx$ 17% defined. Full forward/backward imputation: $\approx$ 64% defined

When data from a number of experiments are entered a data matrix can be constructed:

$\mathbf{X}$(experiments $\times$ regions),

where each row corresponds to an "experiment" (e.g., computed values of binding across brain regions) and where each column represents a specific item in the brain region taxonomy.

Data matrix

Further processing works on the data matrix

- By normalizing the values in a row, i.e., within an experiment.
- By imputation of the missing value in the matrix setting an element based on the values of its taxonomic children, e.g., setting the value of "cingulate gyrus" to the value of "posterior cingulate gyrus" if that is defined.
- By excluding columns of the matrix (i.e. brain regions) so only key brain regions, such as the lobes, remain.

Normalization

Studentize values across $P_n = |\mathcal{P}_n|$ regions with the n'th experiment: $\tilde{x} = (x - \bar{x}_n)/s_n$ with missing values

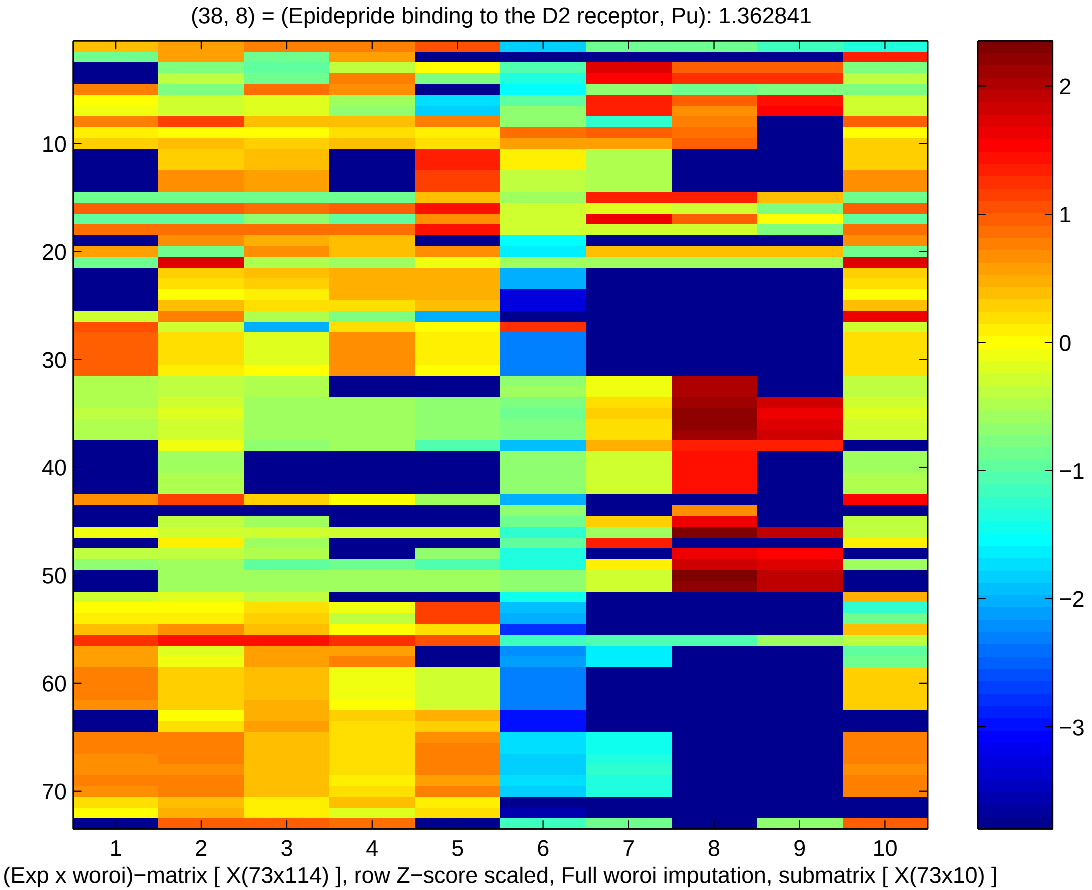
$$\bar{x}_n = \frac{1}{|\mathcal{P}_n|} \sum_{p \in \mathcal{P}_n} x_{np}, \qquad s_n = \sqrt{\frac{1}{|\mathcal{P}_n| - 1} \sum_{p \in \mathcal{P}_n} (x_{np} - \bar{x}_n)^2}. \tag{1}$$

Conversion of data matrix to a "rank order data matrix": $\mathbf{X}(N \times P) \rightarrow \tilde{\mathbf{X}}\left(N \times \frac{P!}{2(P-2)!}\right)$ (Jajuga et al., 2003)

$$\tilde{x}_{n\tilde{p}} = \begin{cases} 1 & \text{if } x_{np} > x_{np'} \\ -1 & \text{if } x_{np} < x_{np'} \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where "otherwise" is with $x_{np} = x_{np'}$ or if any of the values for the two regions p or p' is not defined.

The transformed data matrix



Restriction to key regions: The 5 lobes, cerebellum, caudate, putamen, thalamus and hippocampus: $\mathbf{X}(73 \times 10)$

After full imputation and restriction to key regions: $\approx$ 74% defined values

Outlying brain regions (columns in the data matrix) are: Cerebellum (blue), Caudate and Putamen (red).

Measuring difference between experiments

Comparison of two experiments represented in vectors $\mathbf{x}_n$ and $\mathbf{x}_m$ with the cross-correlation for missing values (pairwise complete version)

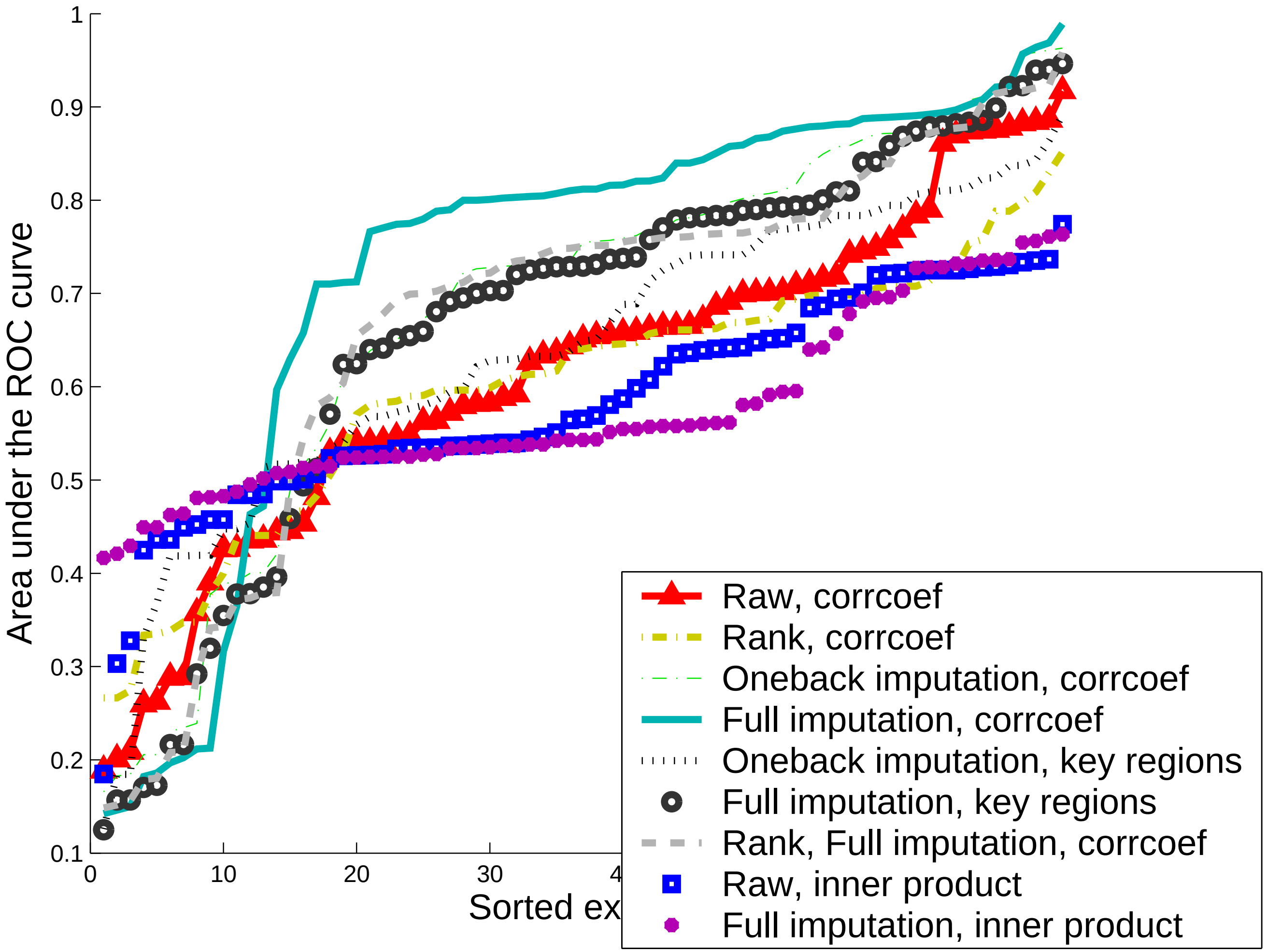
$$\tilde{r}_{nm} = \frac{\sum_{p \in \mathcal{P}_{nm}} \tilde{x}_{np} \tilde{x}_{mp}}{\sqrt{\sum_{p \in \mathcal{P}_{nm}} \tilde{x}_{np}^2} \sqrt{\sum_{p \in \mathcal{P}_{nm}} \tilde{x}_{mp}^2}}, \qquad (3)$$

where $\mathcal{P}_{nm} = \mathcal{P}_n \cap \mathcal{P}_m$ with centered data.

. . . Or just with an inner product

$$t_{nm} = \sum_{p \in \mathcal{P}_{nm}} x_{np} x_{mp} \qquad (4)$$

Information retrieval performance



Area under the ROC curve as performance measure

Task: Segregate between serotonin-2A and non-serotonin-2A studies.

Full imputation with cross-correlation coefficient is the best method.

Example on "Related Experiments"

Related - positive correlated volumes

+1: 29.90915 **Altanserin binding in normal women**. *Altanserin binding to the 5-HT2A receptor in normal women.*
W. H. Kaye; G. K. Frank; C. C. Meltzer; J. C. Price; C. W. McConaha; P. J. Crossan; K. L. Klump; L. Rhodes. *Altered serotonin 2A receptor activity in women who have recovered from bulimia nervosa.*. *Am J Psychiatry* **158**(7):1152-1155, 2001. PMID: 11431241. FMRIDCID: .

+2: 29.51611 **Altanserin binding in former bulimic women**. *Altanserin binding to the 5-HT2A receptor in women recovered from bulimia nervosa.*
W. H. Kaye; G. K. Frank; C. C. Meltzer; J. C. Price; C. W. McConaha; P. J. Crossan; K. L. Klump; L. Rhodes. *Altered serotonin 2A receptor activity in women who have recovered from bulimia nervosa.*. *Am J Psychiatry* **158**(7):1152-1155, 2001. PMID: 11431241. FMRIDCID: .

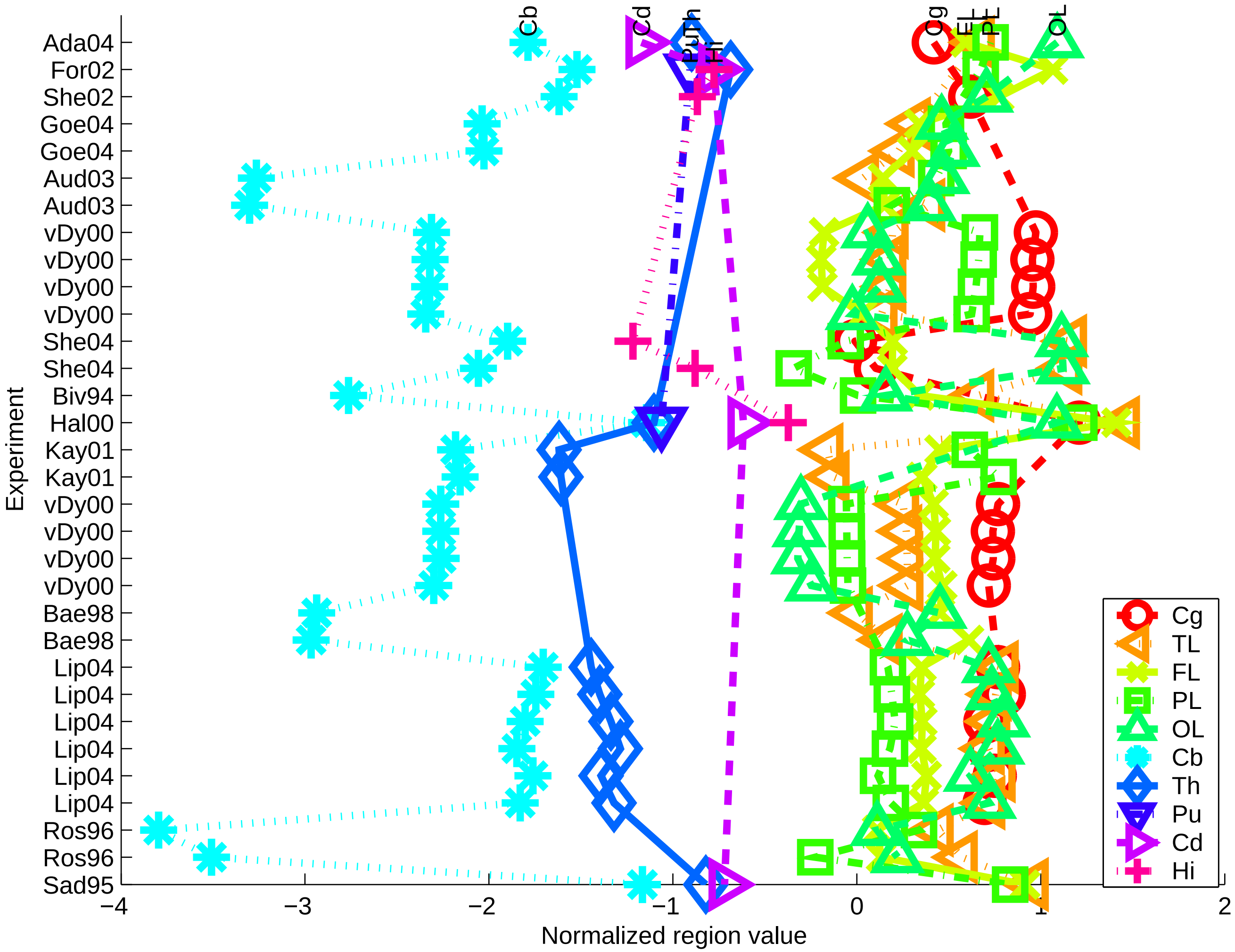
+3: 12.58705 **Altanserin binding in depressed patients**. *Altanserin binding to the serotonin-2A receptor in older major depression disorder patients.*
Yvette I. Sheline; Mark A. Mintun; Deanna M. Barch; Consuelo Wilkins; Abraham Z. Snyder; Stephen M. Moerlein. *Decreased hippocampal 5-HT(2A) receptor binding in older depressed patients using [18F]altanserin positron emission tomography.*. *Neuropsychopharmacology* **29**(12):2235-2241, 2004. PMID: 15367923. DOI: 10.1038/sj.npp.1300555. FMRIDCID: .

+4: 12.27997 **Altanserin binding in normals**. *Altanserin binding to the serotonin-2A receptor in normal elders.*
Yvette I. Sheline; Mark A. Mintun; Deanna M. Barch; Consuelo Wilkins; Abraham Z. Snyder; Stephen M. Moerlein. *Decreased hippocampal 5-HT(2A) receptor binding in older depressed patients using [18F]altanserin positron emission tomography.*. *Neuropsychopharmacology* **29**(12):2235-2241, 2004. PMID: 15367923. DOI: 10.1038/sj.npp.1300555. FMRIDCID: .

+5: 5.31778 **Altanserin binding to 5-HT2A receptors**. *Altanserin binding to 5-HT2A receptors as partial volume corrected distribution volume (DV'3) with cerebellum as reference region.*
Karen H. Adams; Lars H. Pinborg; Claus Svarer; Steen G. Hasselbalch; Soren Holm; Steven Haugbol; Karine Madsen; Vibe Frokjaer; Lars Martiny; Olaf B. Paulson; Gitte M. Knudsen. *A database of [(18)F]-altanserin binding to 5-HT(2A) receptors in normal volunteers: normative data and relationship to physiological and demographic variables.*. *Neuroimage* **21**(3):1105-13, 2004. PMID: 15006678. DOI: 10.1016/j.neuroimage.2003.10.046.

+6: 4.96921 **Altanserin specific binding to non-specific binding ratio**. *Ratio between specific binding and non-specific binding with cerebellum as the reference region for altanserin to the 5-HT2A receptor.*
Annemie Rosier; Patrick Dupont; Jozef Peuskens; Guy Bormans; Rik Vandenberghe; Michael Maes; Tjibbe de Groot; Christian Schiepers; Alfons Verbruggen; Luc Mortelmans. *Visualisation of loss of 5-HT2A receptors with age in healthy volunteers using [18F]altanserin and positron emission tomographic imaging*. *Psychiatry Research* **68**(1):11-22, 1996. PMID: 9027929. FMRIDCID: .

Comparisons on serotonin-2A studies



3 "clusters": Cerebellum (reference), Low binding (caudate, putamen, thalamus, hippocampus), high binding (cerebral cortex).

Little coherence among serotonin studies in the cerebral cortex, i.e., the ordering change between regions change.

Clustering

K-means clustering capable of handling missing values in the data matrix $\mathbf{X}$(experiments $\times$ regions) (Wishart, 2003).

Clustering experiments

$$\mathbf{X} = \mathbf{AC} + \mathbf{U}, \tag{5}$$

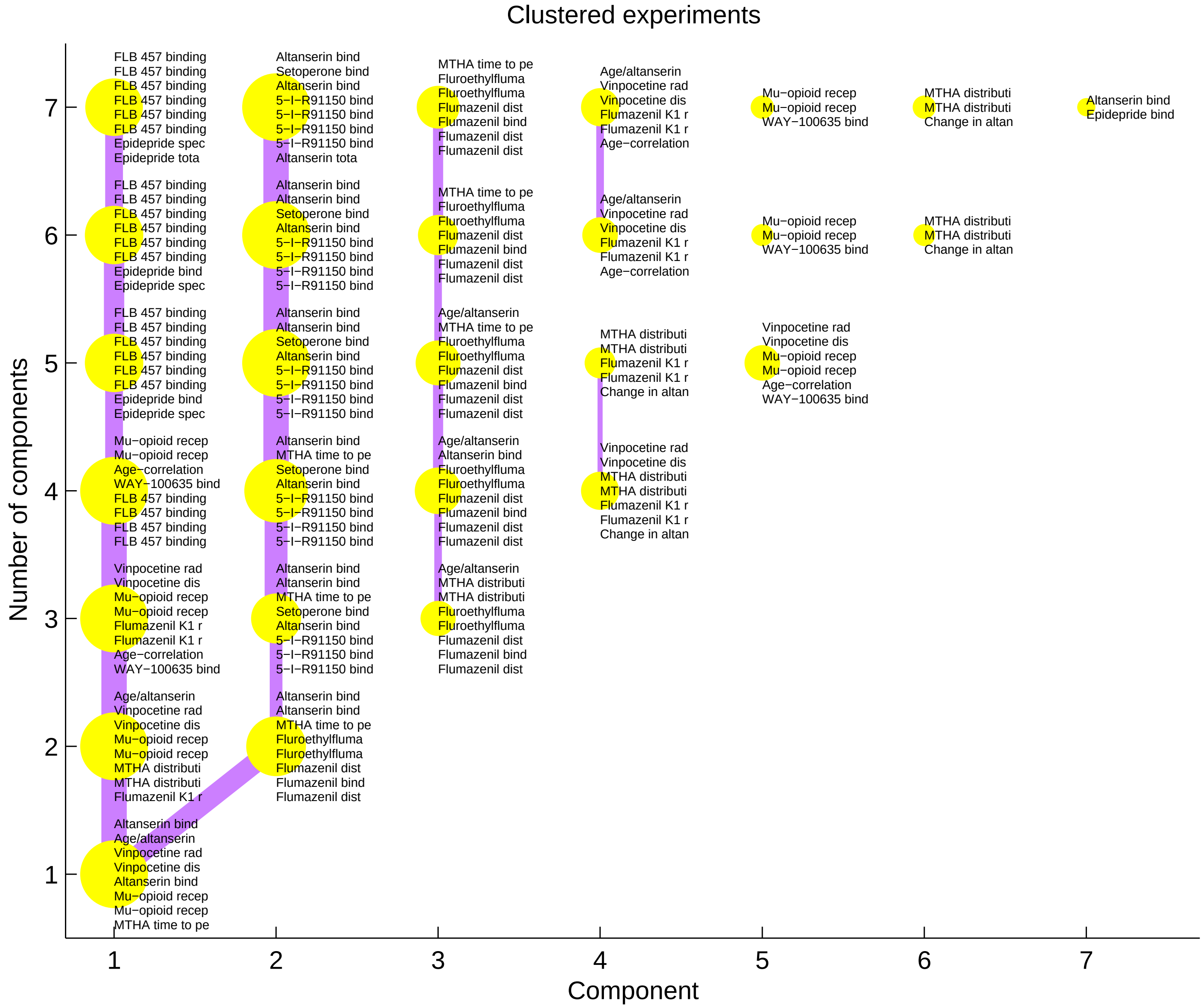
where $\mathbf{A}$ contains assignments for experiments, $\mathbf{C}$ the pattern across brain regions for prototypical tracers and $\mathbf{U}$ residuals.

. . . clustering brain regions

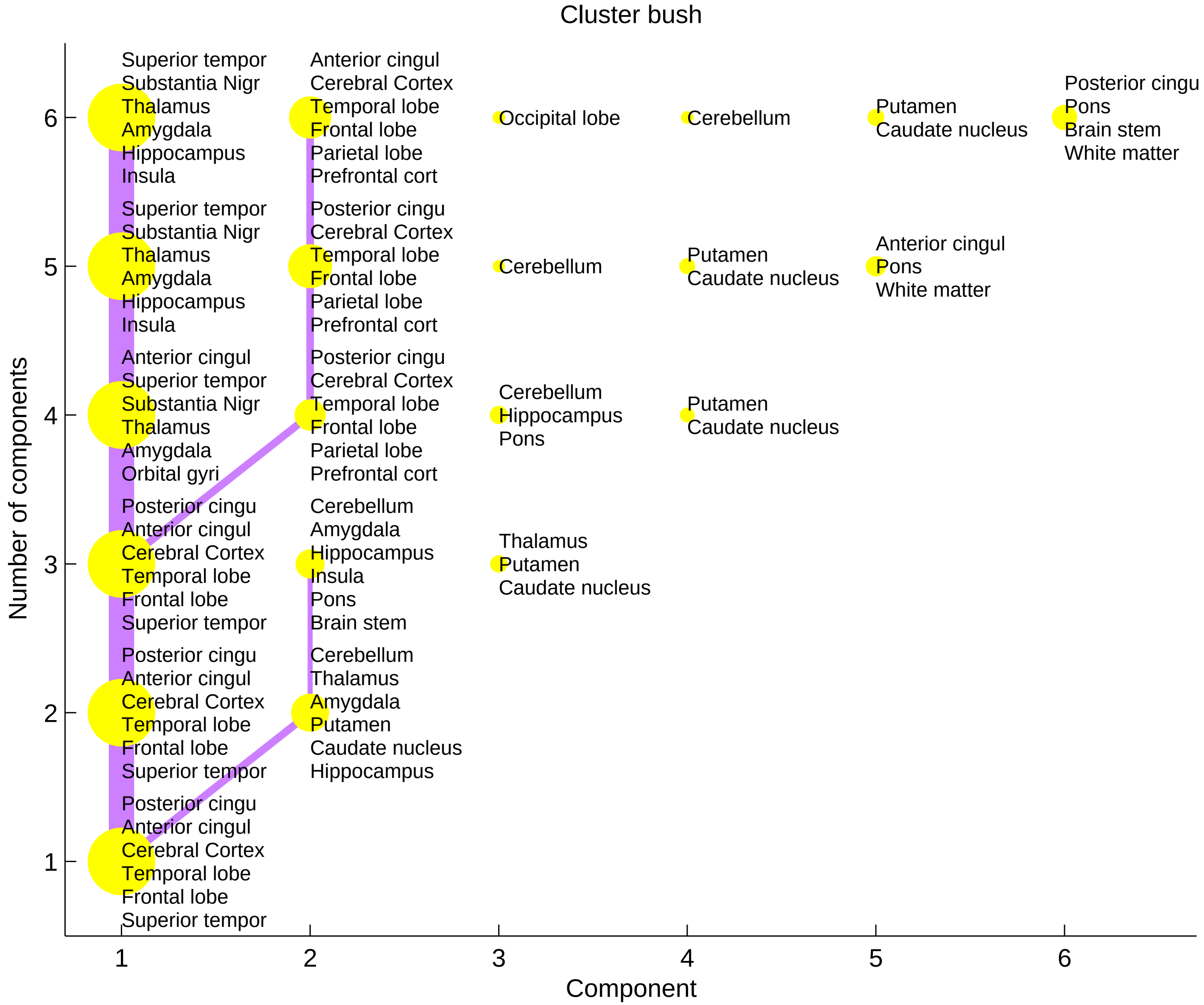
$$\mathbf{X} = \mathbf{CA} + \mathbf{U} \tag{6}$$

These kind of analyses have been made in humans and macaque with autoradiography, e.g., (Kötter et al., 2001).

Clustering of experiments



Clustering of brain regions



Summary

Framework established for databasing molecular neuroimaging studies.

Possible to make information retrieval methods that work on the quantitative region-based neuroimaging results.

Possible to make meta-analysis, e.g., by K-means clustering where experiments or brain regions are clustered.

References

Bowden, D. M. and Martin, R. F. (1995). NeuroNames brain hierarchy. *NeuroImage*, 2(1):63–84. PMID: 9410576. ISSN 1053-8119.

Fox, P. T. and Lancaster, J. L. (1994). Neuroscience on the net. *Science*, 266(5187):994–996. PMID: 7973682.

Jajuga, K., Walesiak, M., and Bak, A. (2003). On the general distance measure. In Schwaiger, M. and Opitz, O., editors, *Exploratory Data Analysis in Empirical Research. Proceedings of the 25th Annual Conference of the Gesellschaft für Klassifikation e.V., University of Munich, March 14-16, 2001*, volume 16 of *Studies in Classification, Data Analysis, and Knowledge Organization*, pages 104–109. Springer, Berlin, Germany. ISBN 3540441832.

Kötter, R., Stephan, K. E., Palomero-Gallager, N., Geyer, S., Schleicher, A., and Zilles, K. (2001). Multimodal characterisation of cortical areas by multivariate analyses of receptor binding and connectivity. *Anatomy and Embryology*, 204(4):333–349. PMID: 11720237. DOI: 10.1007/s004290100199. ISSN 0340-2061. A study on macaque brain regions using binding characteristics from 9 different ligands as well as using anatomical connectivity information. Multidimensional scaling and hierarchincal clustering are used to two receptor-times-brain-regions data matrices.

Nielsen, F. Å. (2003). The Brede database: a small database for functional neuroimaging. *NeuroImage*, 19(2). http://208.164.121.55/hbm2003/abstract/abstract906.htm. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, 2003, New York, NY. Available on CD-Rom.

Nielsen, F. Å. (2005). Mass meta-analysis in Talairach space. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*, pages 985–992, Cambridge, MA. MIT Press. http://books.nips.cc/papers/files/nips17/NIPS2004_0511.pdf.

Nielsen, F. Å., Balslev, D., and Hansen, L. K. (2005). Mining the posterior cingulate: Segregation between memory and pain component. *NeuroImage*, 27(3):520–532. DOI: 10.1016/j.neuroimage.2005.04.034.

Nielsen, F. Å. and Hansen, L. K. (2000). Experiences with Matlab and VRML in functional neuroimaging visualizations. In Klasky, S. and Thorpe, S., editors, *VDE2000 - Visualization Development Environments, Workshop Proceedings, Princeton, New Jersey, USA, April 27–28, 2000*, pages 76–81, Princeton, New Jersey. Princeton Plasma Physics Laboratory. http://www.imm.dtu.dk/pubdb/views-/edoc_download.php/1231/pdf/imm1231.pdf. CiteSeer: http://citeseer.ist.psu.edu/309470.html.

Nielsen, F. Å. and Hansen, L. K. (2004). Finding related functional neuroimaging volumes. *Artificial Intelligence in Medicine*, 30(2):141–151. PMID: 14992762. http://www.imm.dtu.dk/~fn/Nielsen2002Finding/.

Nielsen, F. Å., Hansen, L. K., and Balslev, D. (2004). Mining for associations between text and brain activation in a functional neuroimaging database. *Neuroinformatics*, 2(4):369–380. http://www2.imm.dtu.dk/~fn/ps/Nielsen2004Mining_submitted.pdf.

Wishart, D. (2003). k-means clustering with outlier detection, mixed variables and missing values. In Schwaiger, M. and Opitz, O., editors, *Exploratory Data Analysis in Empirical Research. Proceedings of the 25th Annual Conference of the Gesellschaft für Klassifikation e.V., University of Munich, March 14-16, 2001*, volume 16 of *Studies in Classification, Data Analysis, and Knowledge Organization*, pages 216–226. Springer, Berlin, Germany. ISBN 3540441832.