Databasing Molecular Neuroimaging

Finn Årup Nielsen

Neurobiology Research Unit
Copenhagen University Hospital Rigshospitalet
and
Informatics and Mathematical Modelling
Technical University of Denmark

May 24, 2005

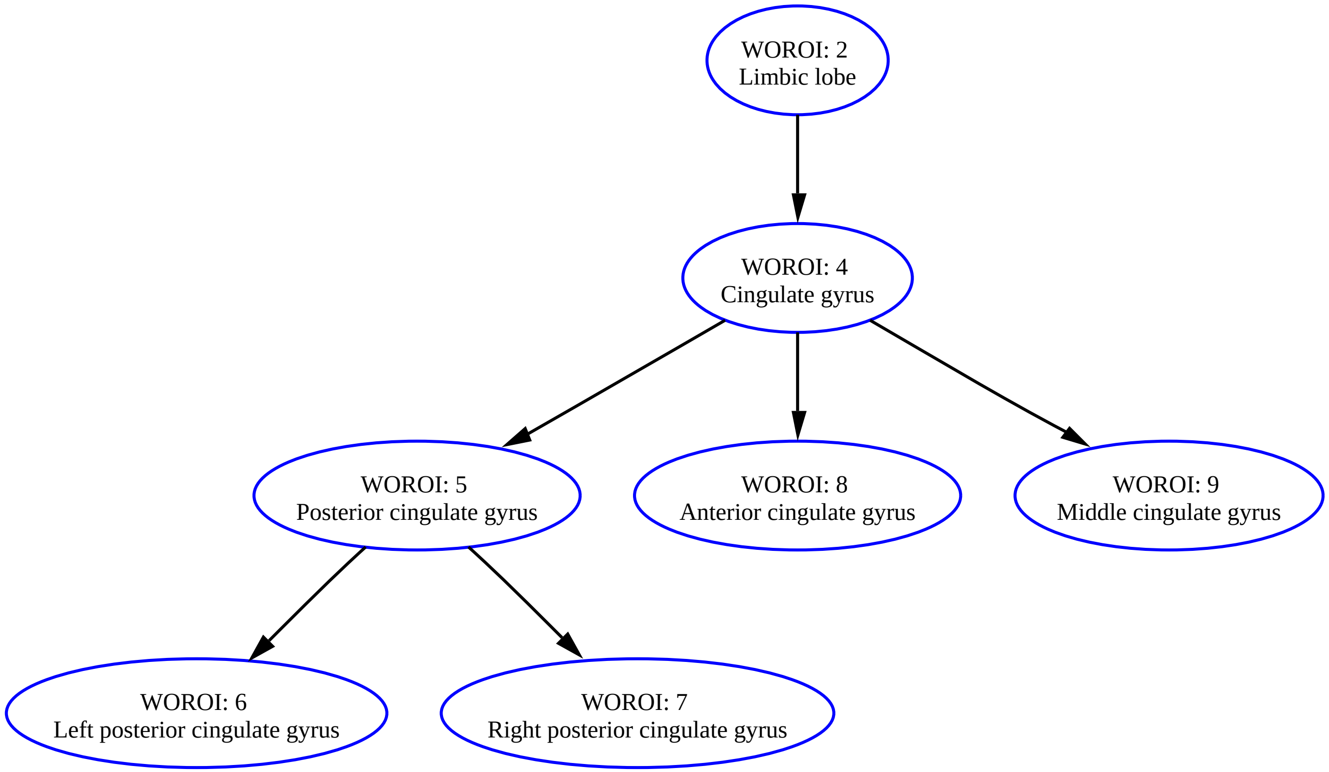
Molecular neuroimaging

Most molecular imaging studies relies on analysis of values from brain regions and report descriptive statistics for these values.

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.
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.

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]

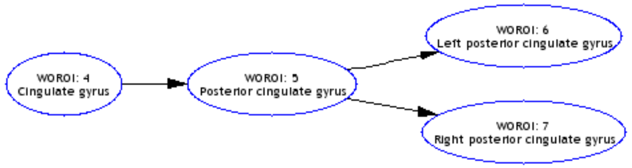
Roi -> [Alphabetic | Hammers | Tzourio-Mazoyer]

[Brede]

WOROI: 5 - Posterior cingulate gyrus

Abbreviation: **PCgG**

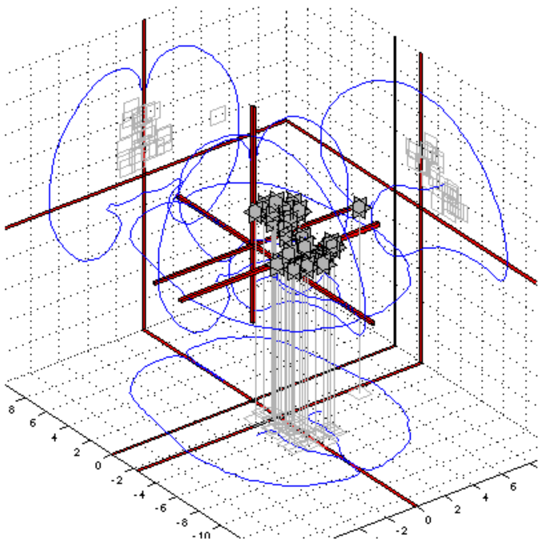
Variation: *posterior cingulate*
Variation: *posterior cingulate area*
Variation: *posterior gyrus cinguli*
BrainInfo: 144



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

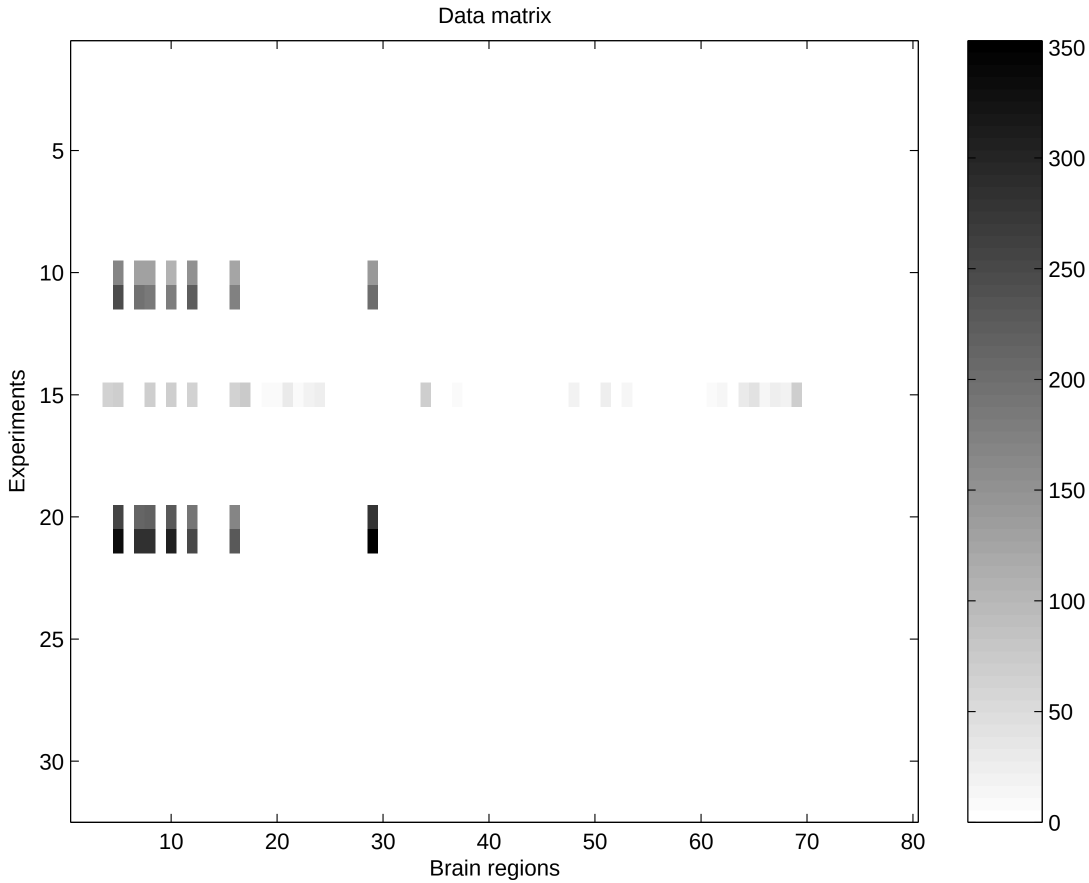


Brain region taxonomy included in the Brede Database.

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, Wikipedia.

Data matrix



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

For serotonin-2A part of the datamatrix $\mathbf{X}(32 \times 80)$:

Original matrix: $\approx$ 13% defined.

"One-back" imputation: $\approx$ 17% defined

Full forward/backward imputation: $\approx$ 64% defined

Handling different range among experiments

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

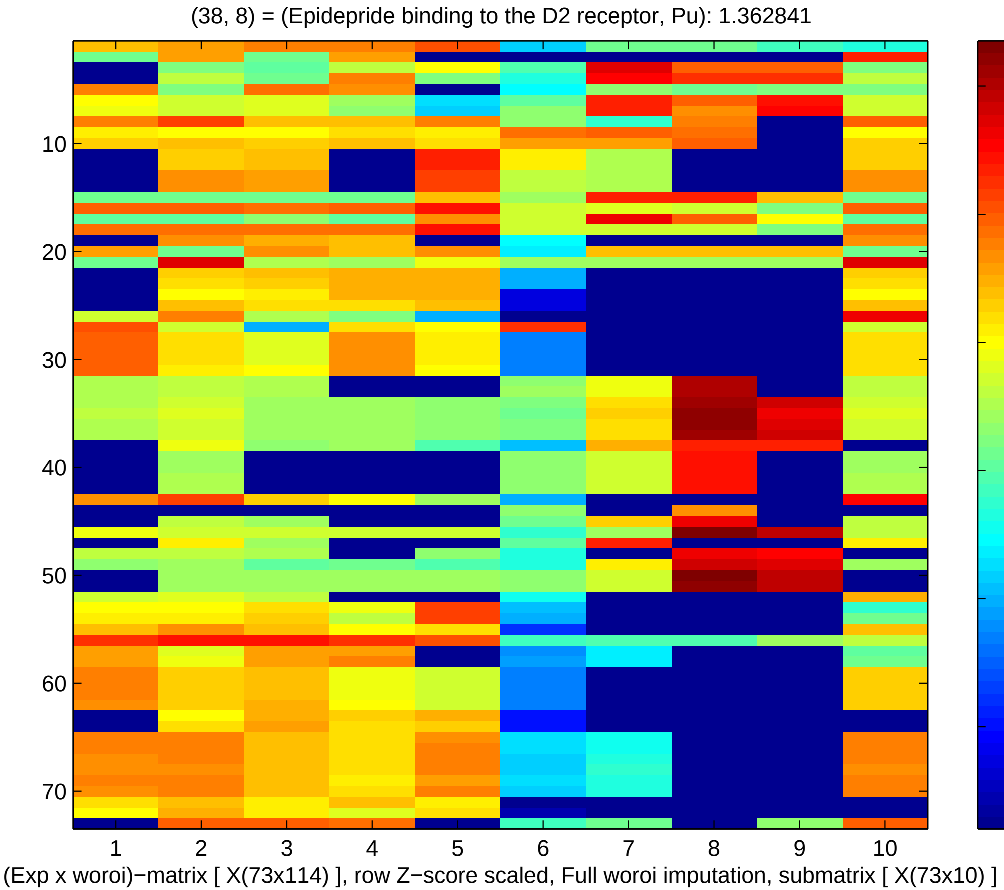
$$\bar{x}_n = \frac{1}{|\mathcal{P}_n|} \sum_{p \in \mathcal{P}_n} x_{np}, \qquad s_n = \sqrt{\frac{1}{|\mathcal{P}_n| - 1} \sum_{p \in \mathcal{P}_n} (x_{np} - \bar{x}_n)^2}. \tag{1}$$

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

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

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

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)

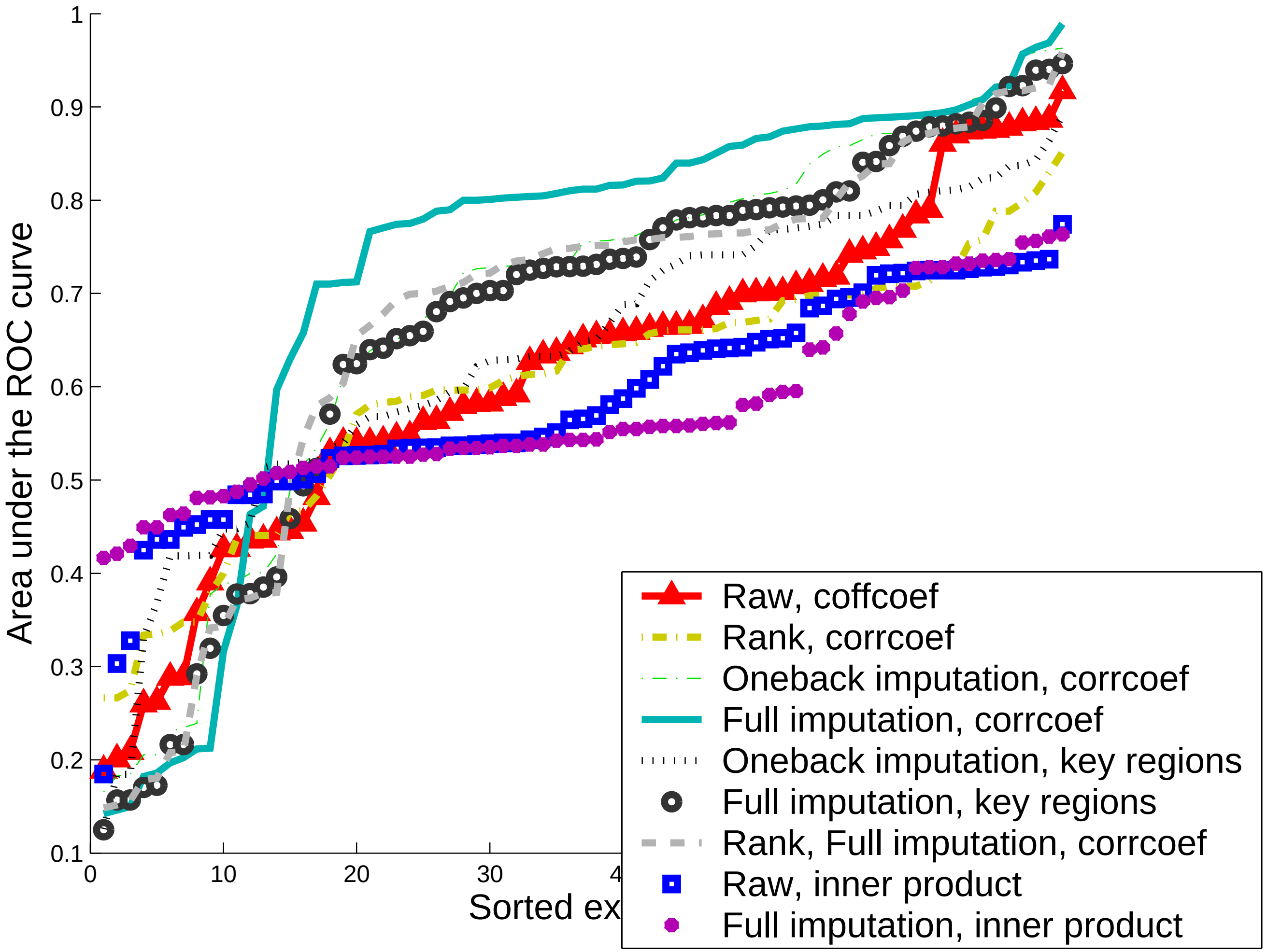
$$\tilde{r}_{nm} = \frac{\sum_{p \in \mathcal{P}_{nm}} \tilde{x}_{np} \tilde{x}_{mp}}{\sqrt{\sum_{p \in \mathcal{P}_{nm}} \tilde{x}_{np}^2} \sqrt{\sum_{p \in \mathcal{P}_{nm}} \tilde{x}_{mp}^2}}, \quad (3)$$

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

. . . Or just with an inner product

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

Information retrieval performance

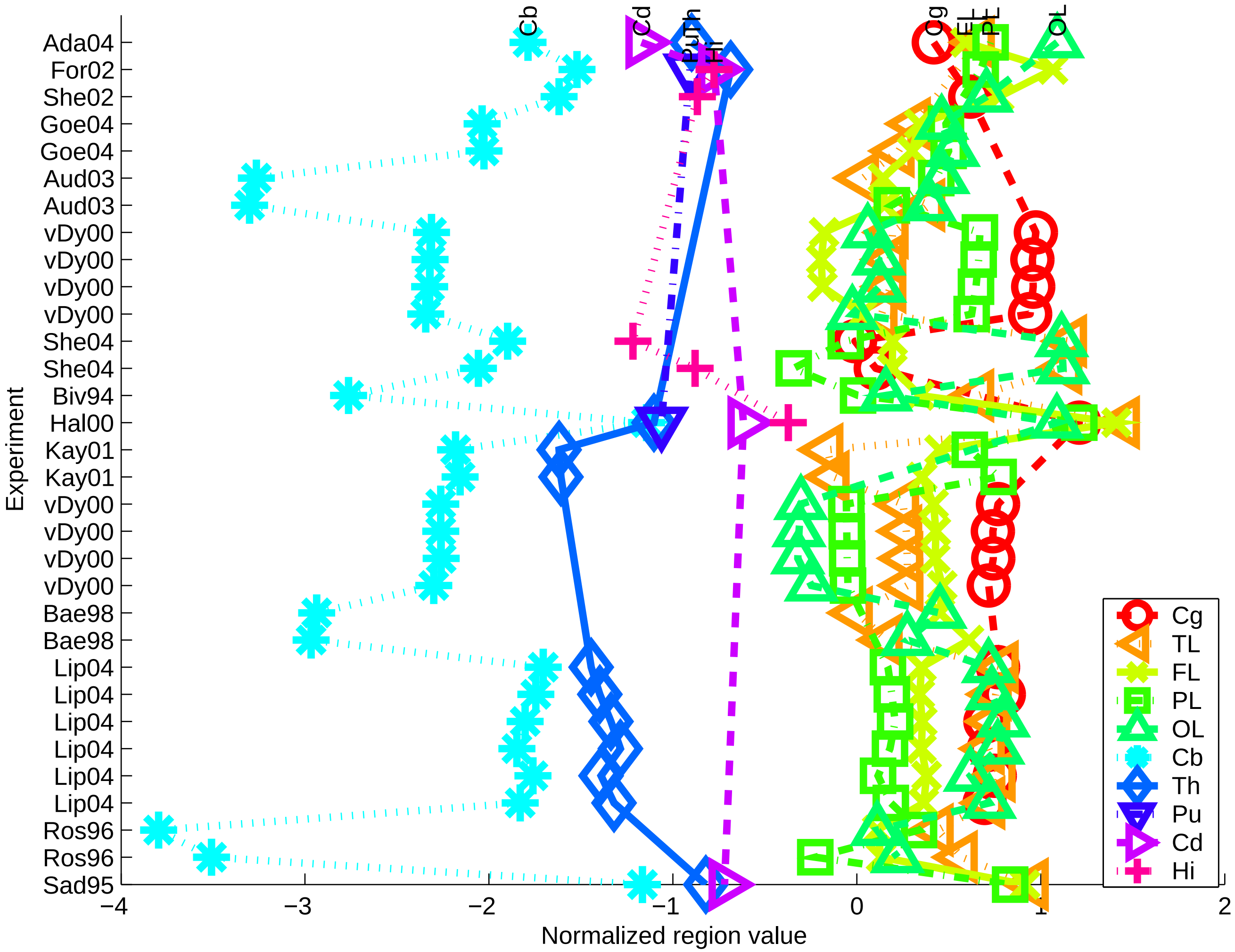


Area under the ROC curve as performance measure

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

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

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

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

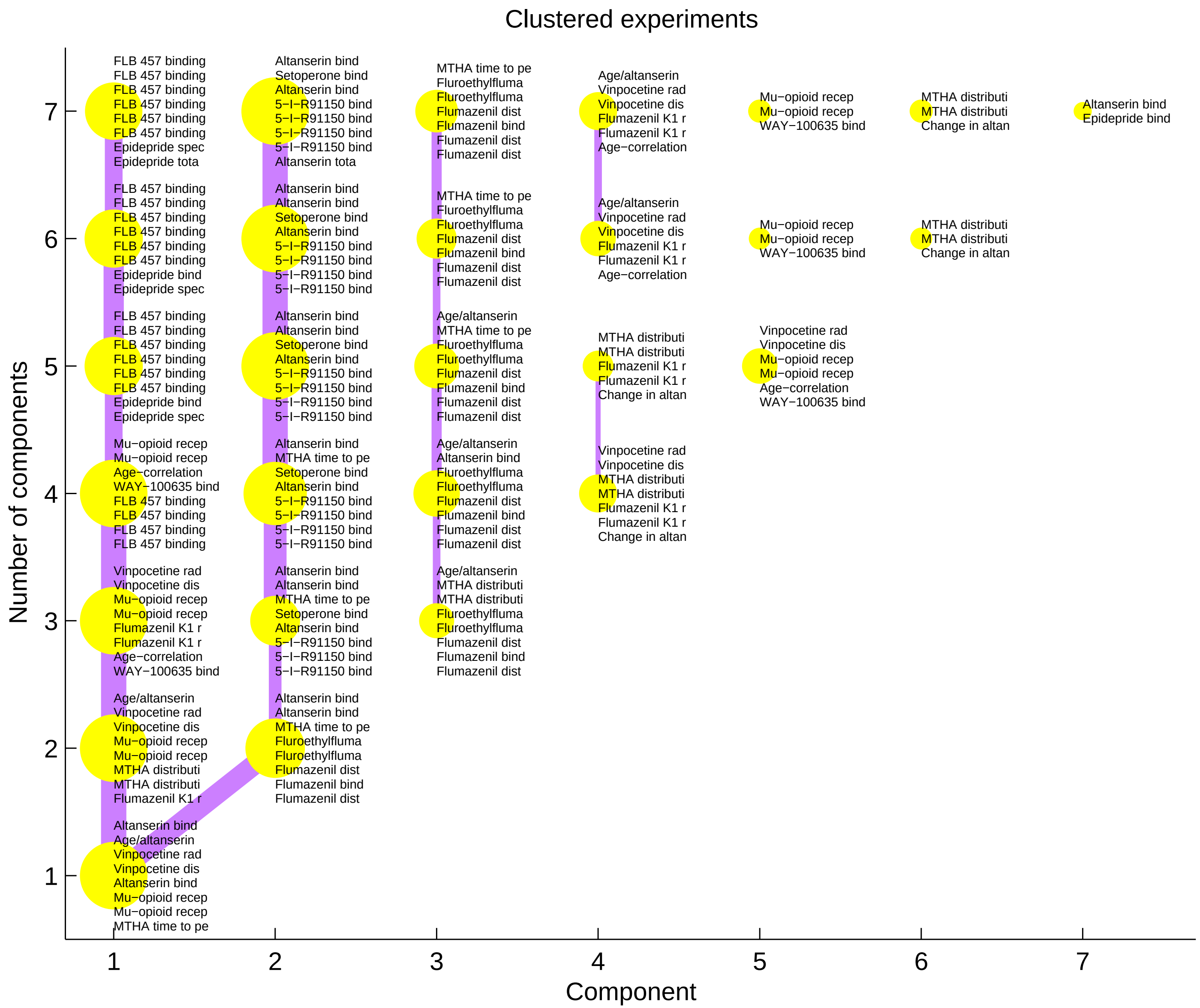
where $\mathbf{A}$ contains assignments for experiments, $\mathbf{C}$ the pattern across brain regions for prototypical tracers and $\mathbf{U}$ residuals.

...clustering brain regions

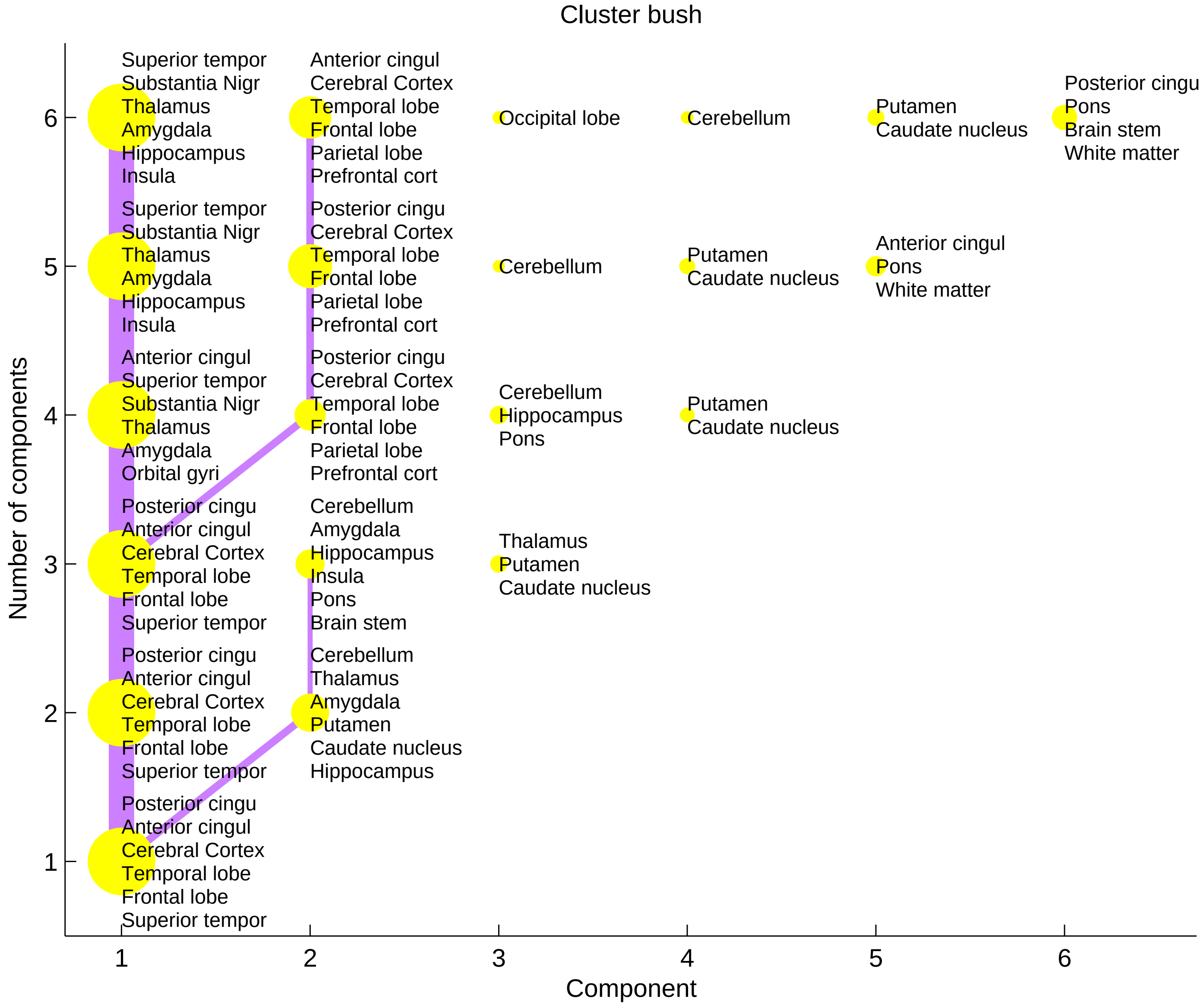
$$\mathbf{X} = \mathbf{C}\mathbf{A} + \mathbf{U} \tag{6}$$

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

Possible to make meta-analysis on brain region based molecular neuroimaging.

Information retrieval and clustering are dependent on key features of the tracer/receptor, e.g., altanserin has low/no binding in cerebellum.

References

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

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.

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.