Statistical inference and multiple testing correction in MVPA

Johannes Stelzer, 1st of September 2015

MAX PLANCK INSTITUTE FOR HUMAN COGNITIVE AND BRAIN SCIENCES LEIPZIG

Max-Planck-Institut für biologische Kybernetik

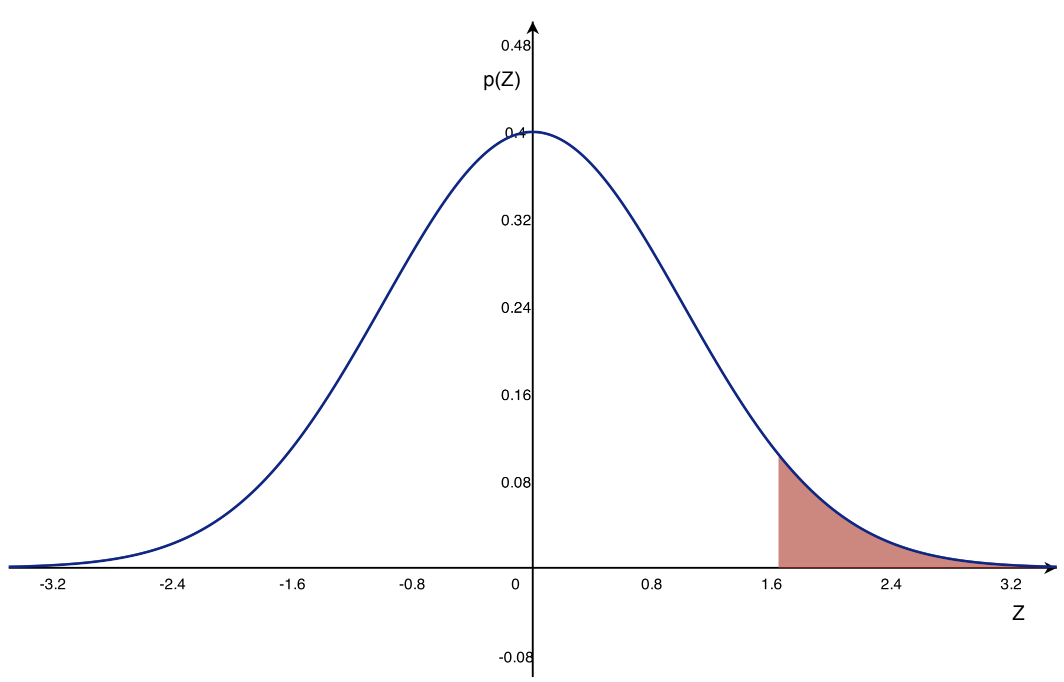
whole-brain map of classifier's estimates (e.g. decoding accuracies)

→

thresholded map indicating statistical significance

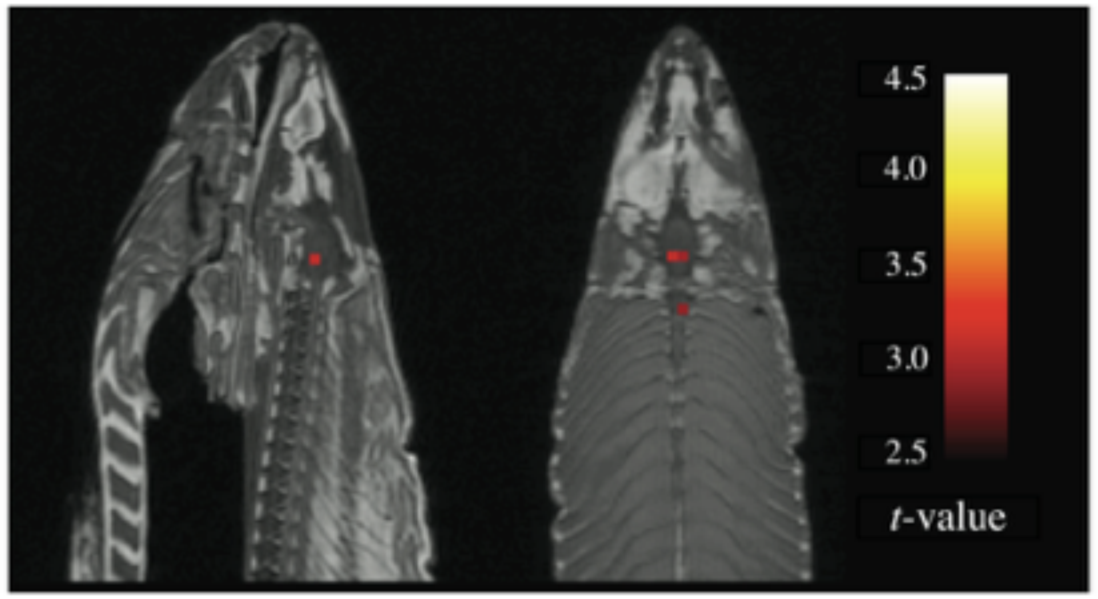
Assuming the classifier "guesses":
what is the likelihood to observe
a given decoding accuracy (or larger)?

how to get
null **distribution**



$$Pr\{Z \geq Z_0\} = \int_{Z_0}^{\infty} f(Z)dZ$$

how to correct
for **multiple** testing



how to get null distribution *parametrically* (common approaches in mvpa):

- binomial models
- t-based stats

real label	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
classifier’s estimate	b	a	a	b	b	a	a	b	a	a	b	b	a	b	a	b	a	b	b	a

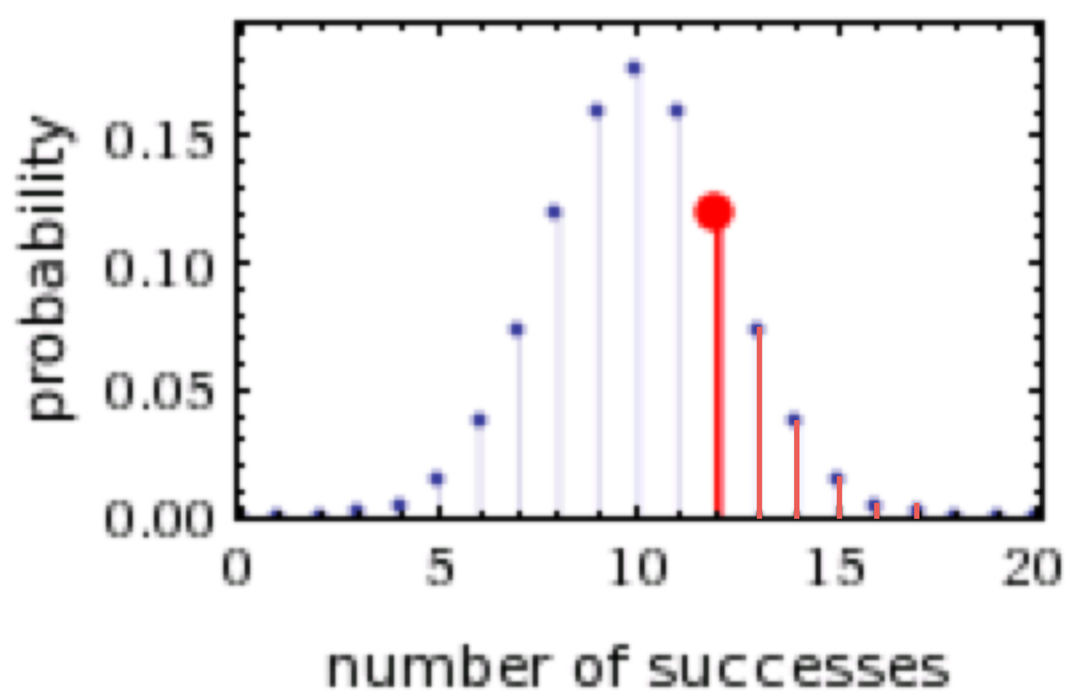
number of samples **N** = 20

correct samples **c** = 12

accuracy = 0.6

binomial model:

$$p_X(c) = \binom{N}{c} p^c (1-p)^{N-c}$$



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
real label	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
classifier’s estimate	b	a	a	b	b	a	a	b	a	a	b	b	a	b	a	b	a	b	b	a

test set T

2x CV											T	T	T	T	T	T	T	T	T	T
4x CV																T	T	T	T	T
5x CV																	T	T	T	T
10x CV																			T	T
20x CV																				T

single model (no CV): $$p_X(c) = \binom{N}{c} p^c (1-p)^{N-c} \qquad c = 0, 1, ..., N$$

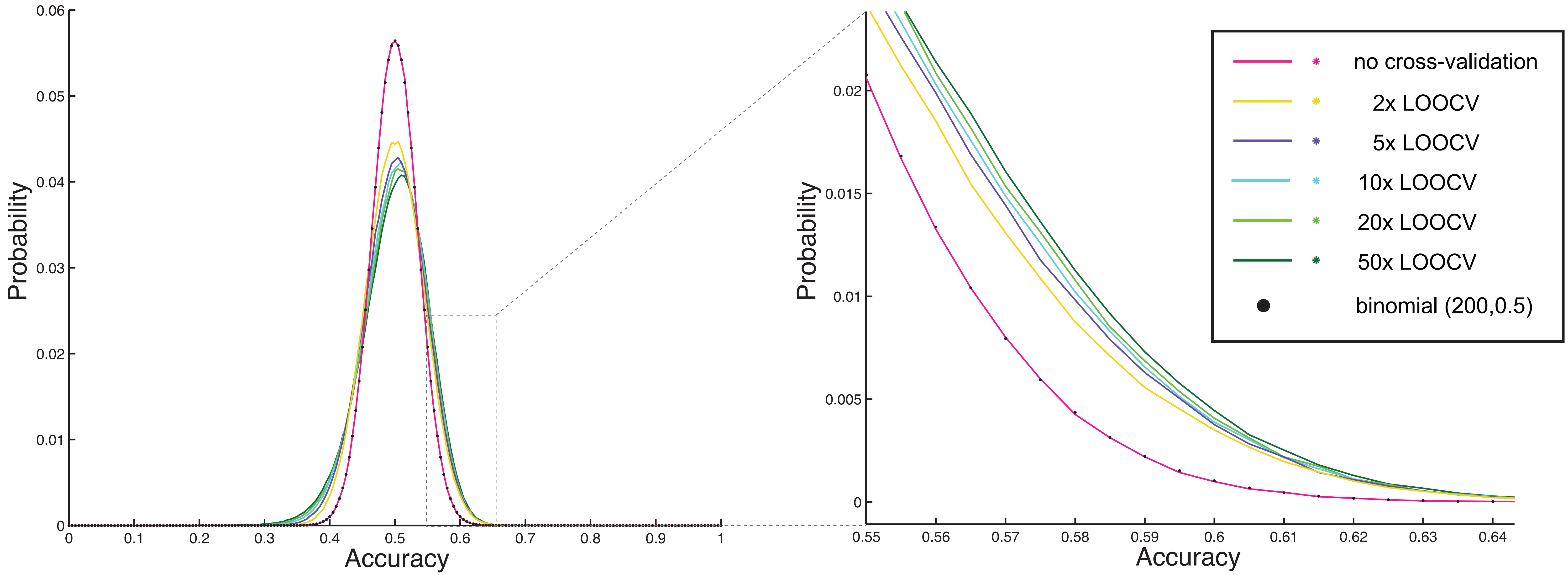
with **t** CV-folds:
(N trials per fold) $$p_X(c) = \binom{N \cdot t}{c} p^c (1-p)^{N \cdot t - c} \qquad c = 0, 1, ..., N \cdot t$$

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage, 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063

real label	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
classifier's estimate	b	a	a	b	b	a	a	b	a	a	b	b	a	b	a	b	a	b	b	a

test set T

2x CV											T	T	T	T	T	T	T	T	T	T
4x CV																T	T	T	T	T
5x CV																	T	T	T	T
10x CV																			T	T
20x CV																				T



Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage, 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063

problems regarding binomial model

- dependency between cross-validation folds undermines assumption of independency between trials
- (ignoring above problem): solution for one single voxel - multiple testing solution needed!

t-based methods and random fields procedure

- get voxel-wise accuracy map for each subject
- one sample t-test against 0.5 (2 classes)
- apply gaussian random-fields cluster level correction

problems regarding t-based framework

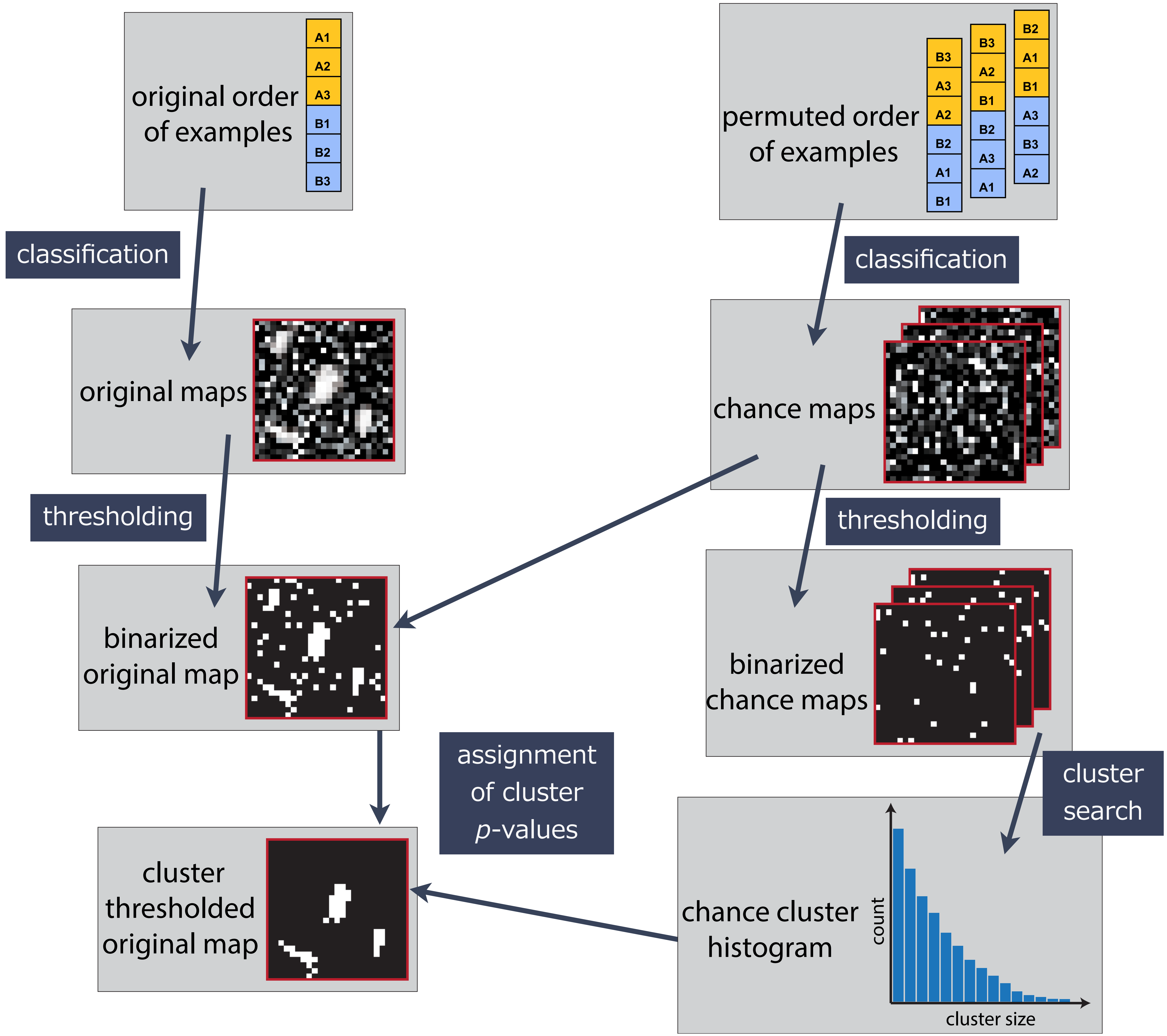
- decoding accuracies discrete (t-test assumes continuous)
- accuracies don’t follow normal distribution
- assumptions regarding random field theory & estimation of spatial smoothness not met

advantages of permutation testing

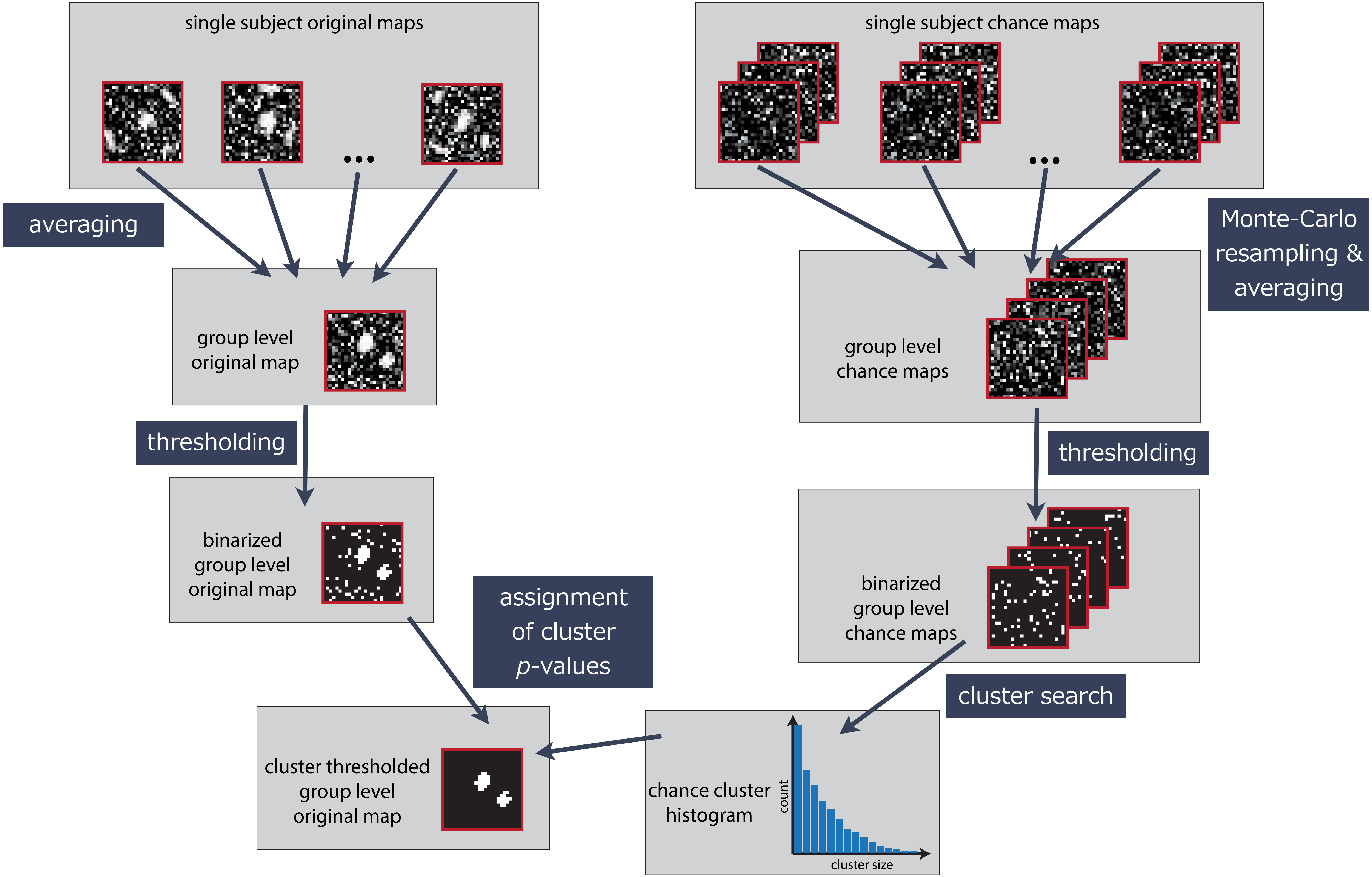
- (practically) assumption free
- yields voxel-wise null distribution
- spatial “chance maps” for free!
  - this provides a null distribution of cluster sizes!

Golland, P., Liang, F., Mukherjee, S., and Panchenko, D. (2005). Permutation Tests for Classification. 3559, 501–515. doi: 10.1007/11503415_34.

Hayasaka, S., and Nichols, T. E. (2003). Validating cluster size inference: random field and permutation methods. 20, 2343–2356. doi: 10.1016/j.neuroimage.2003.08.003.

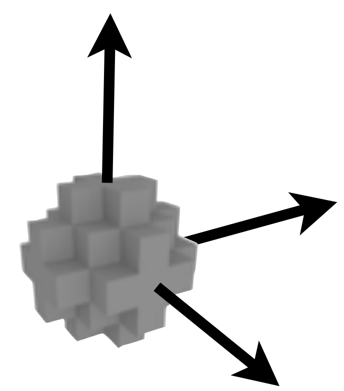


Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage, 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063

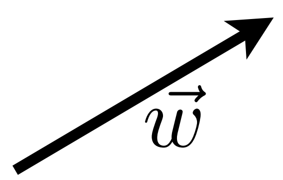


Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage, 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063

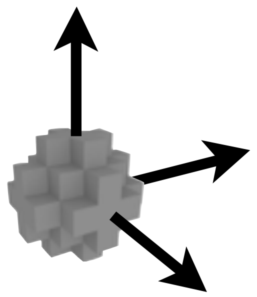
Two common MVPA mapping methods



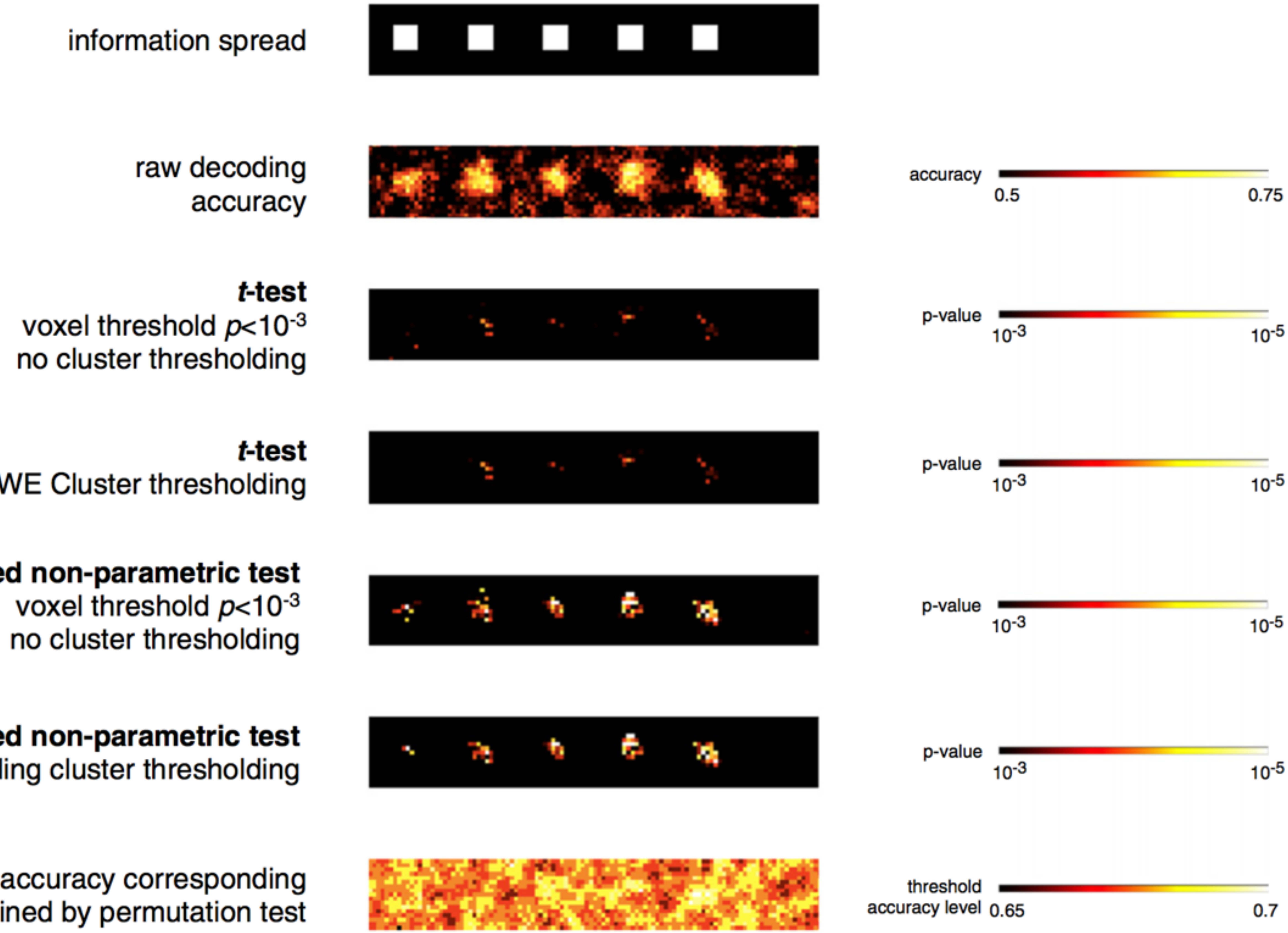
searchlight decoding

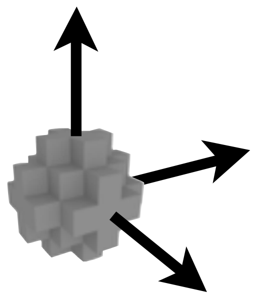


mapping classifier
weights

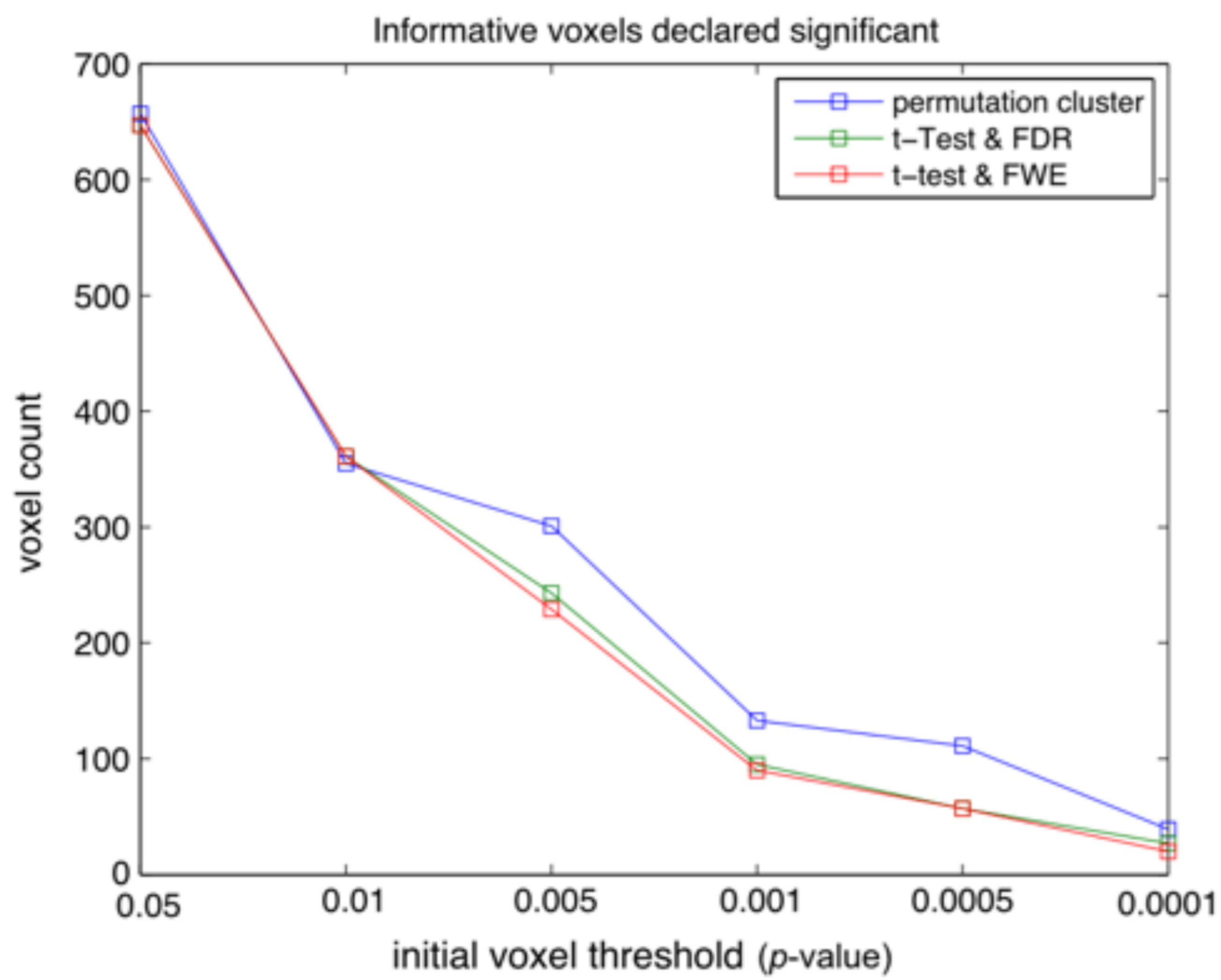


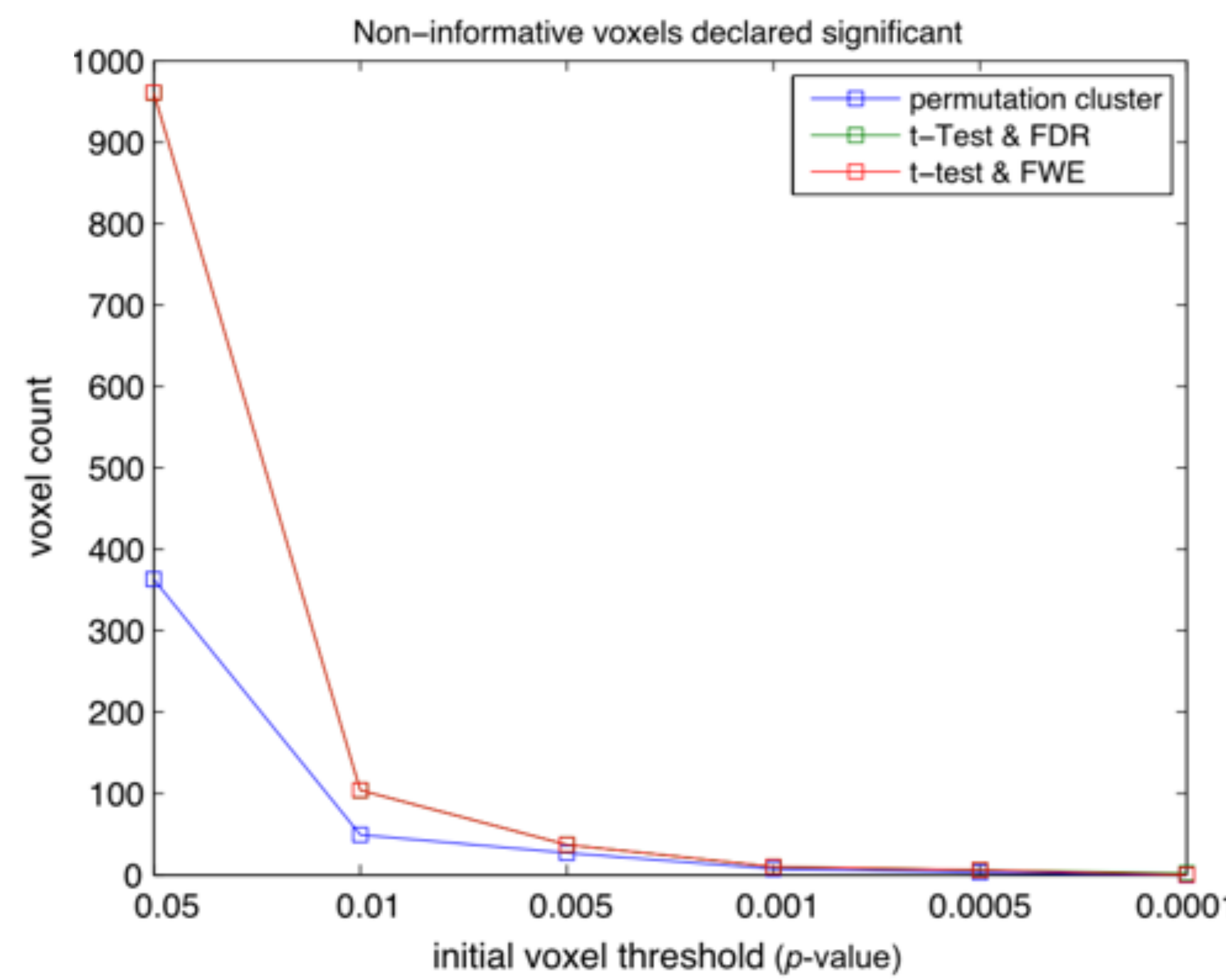
searchlight

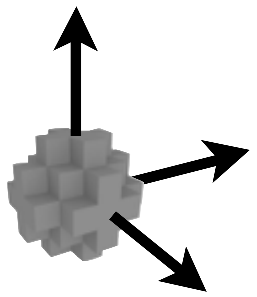




searchlight

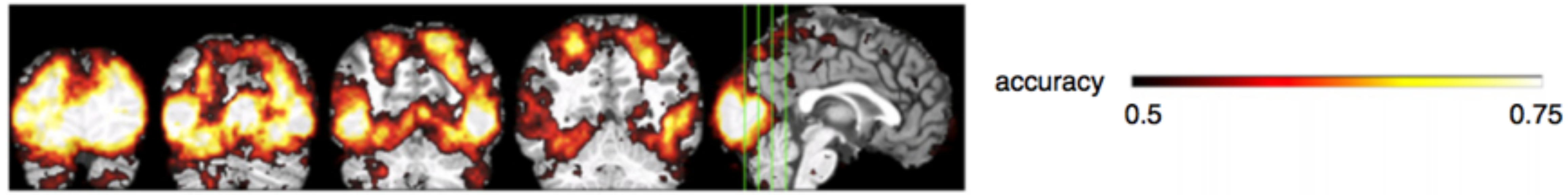




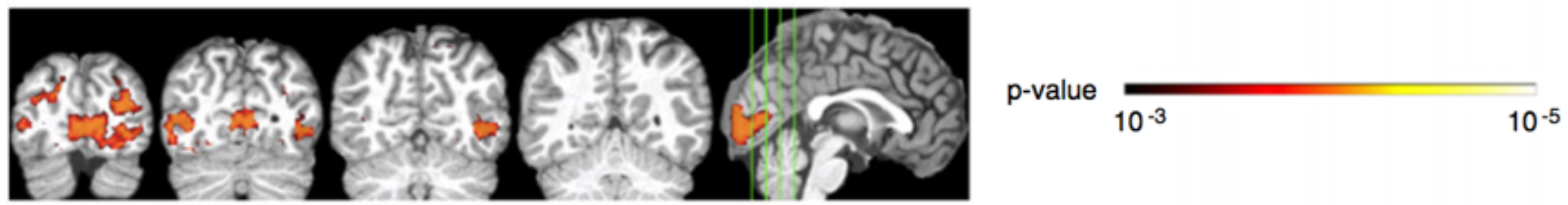


searchlight

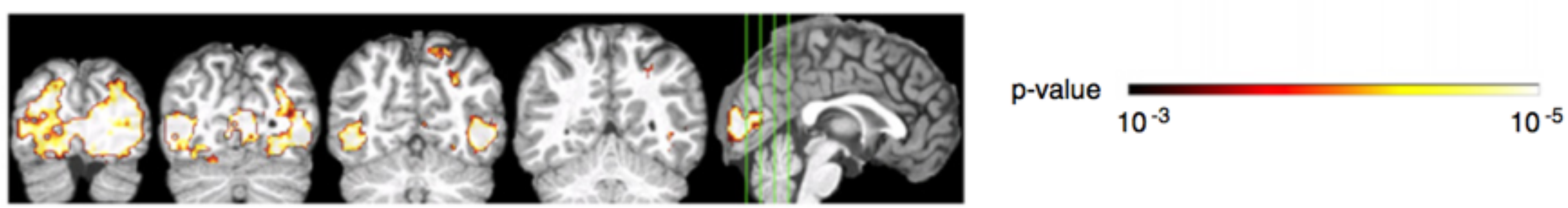
raw decoding accuracy
no cluster size thresholding

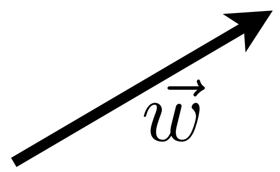


***t*-test**
FWE cluster thresholding

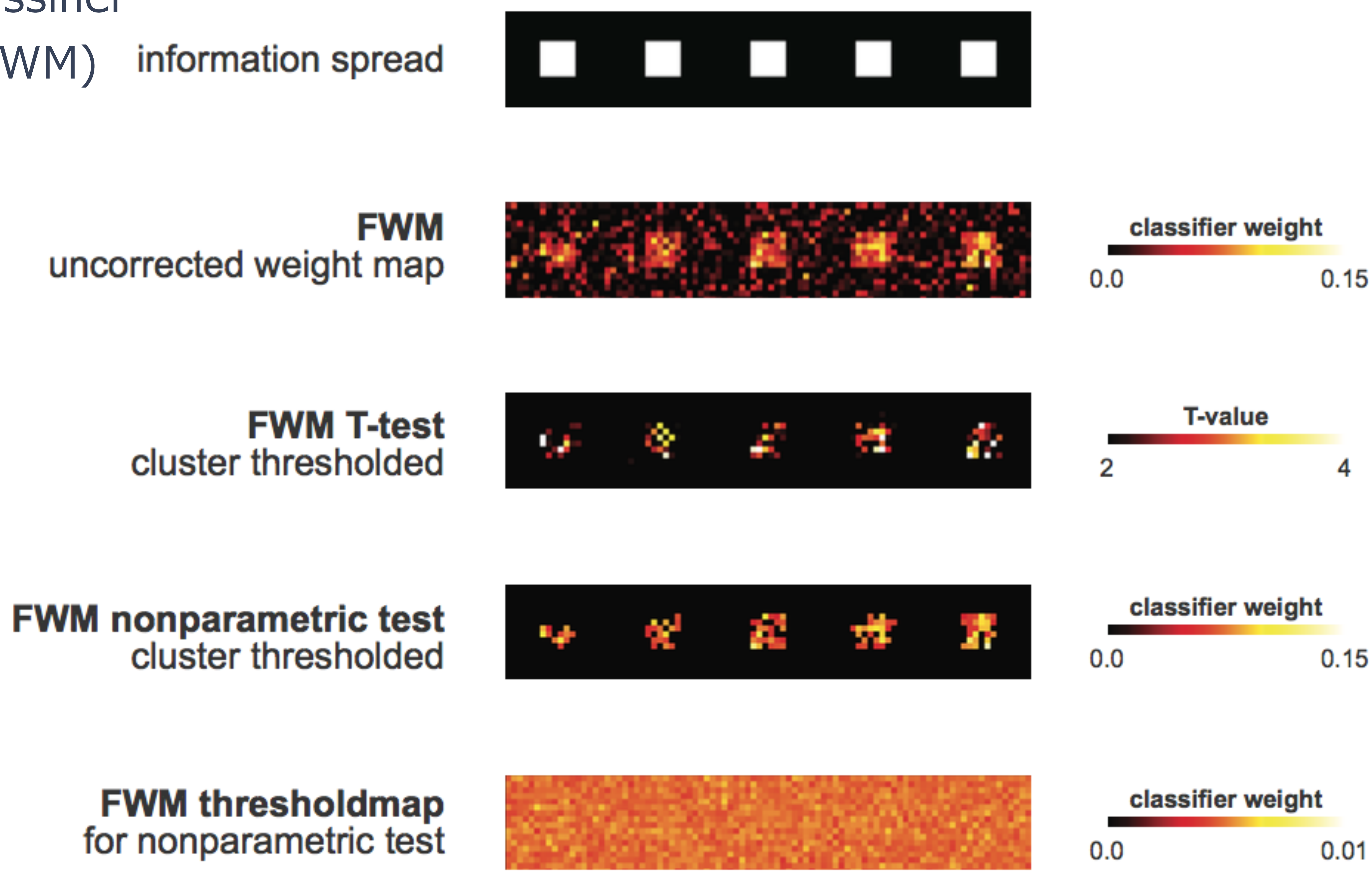


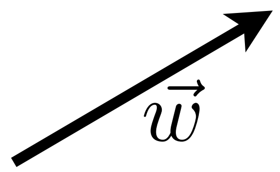
proposed non-parametric test
including cluster thresholding



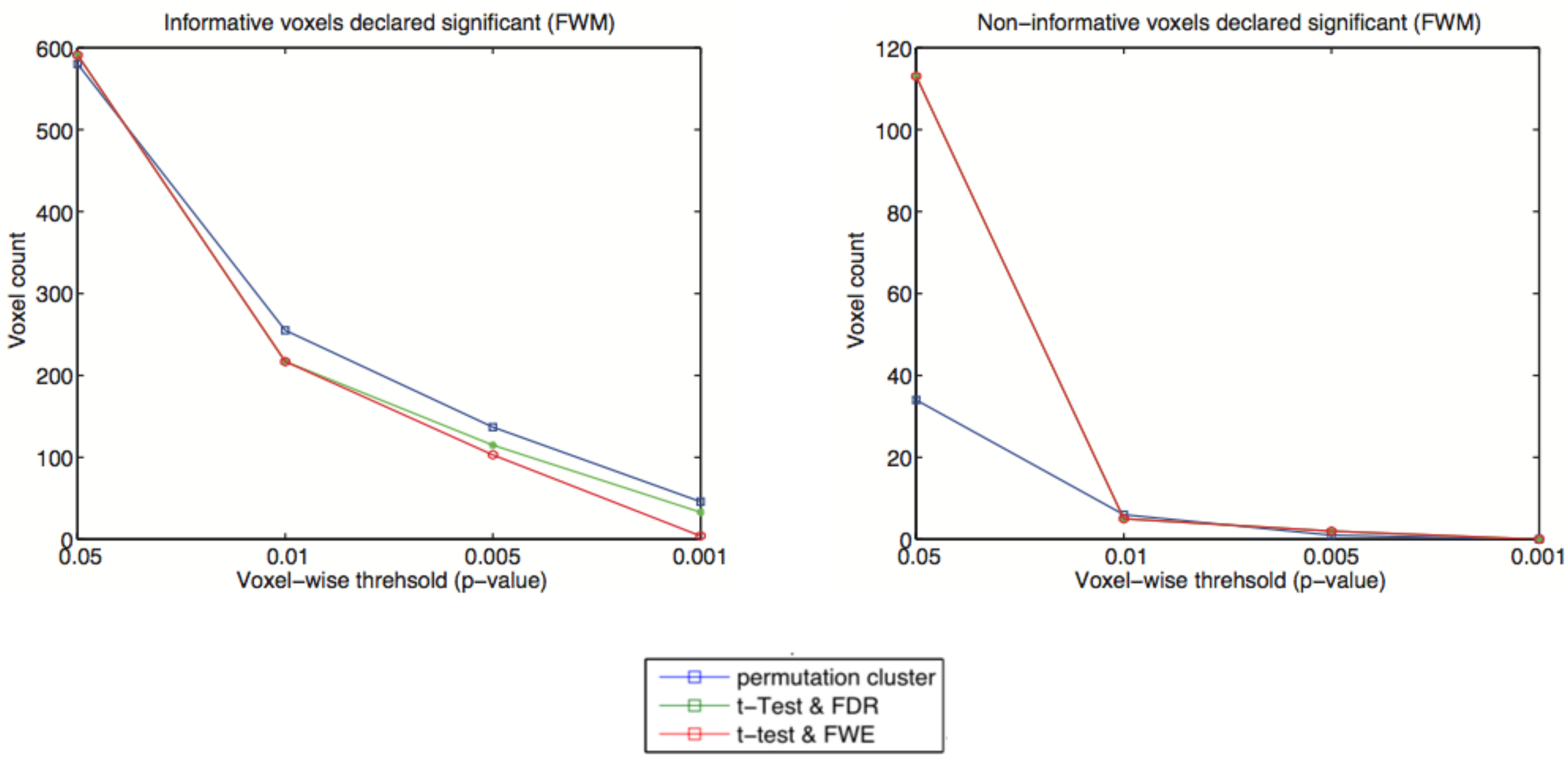


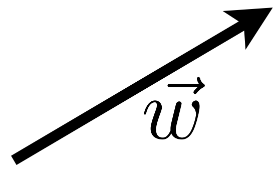
mapping classifier weights (FWM)



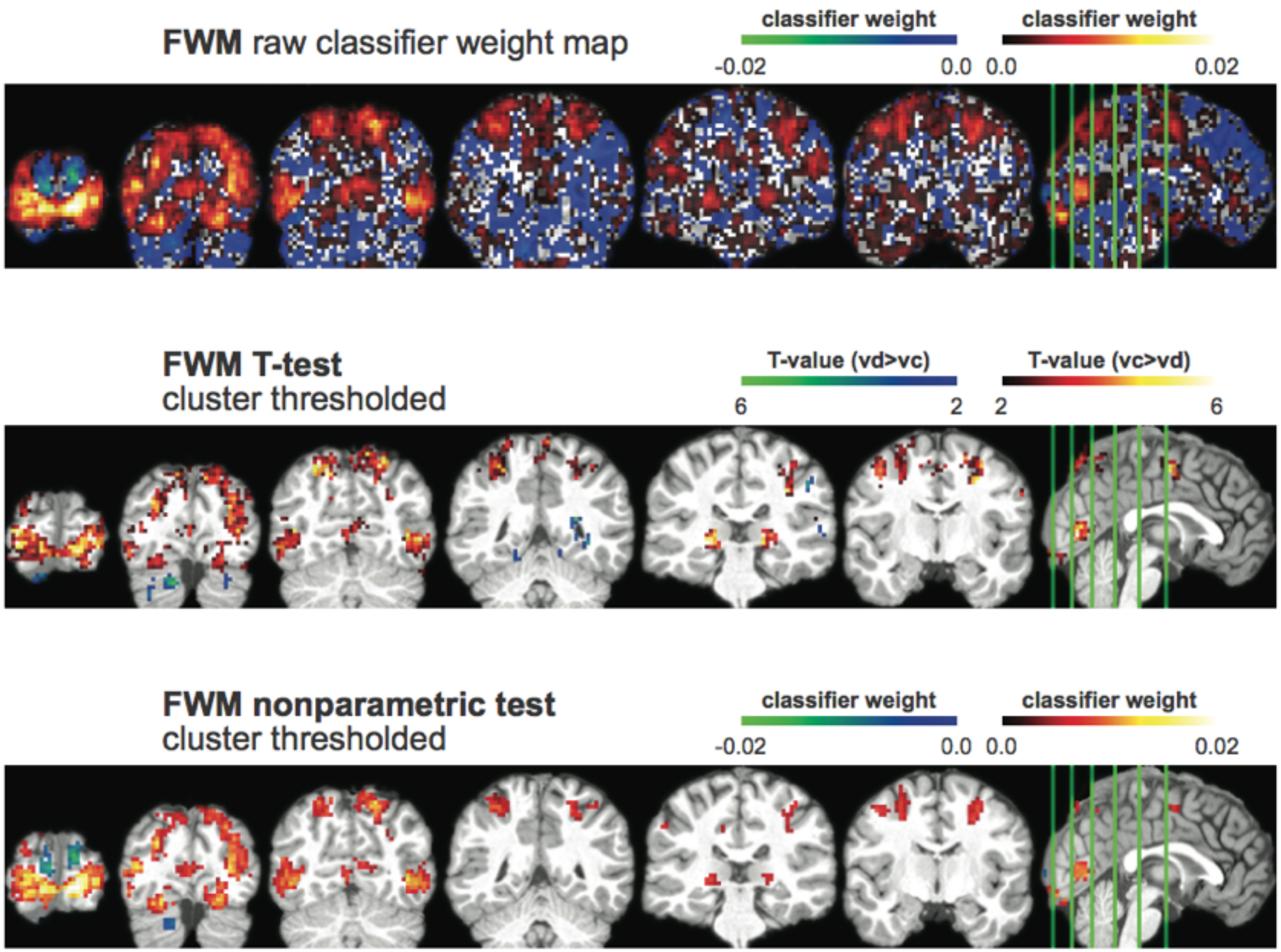


mapping classifier weights (FWM)

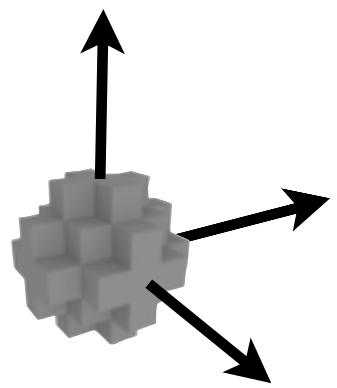




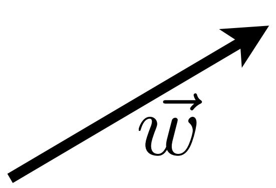
mapping classifier weights (FWM)



group level null simulations
100 simulations carried out



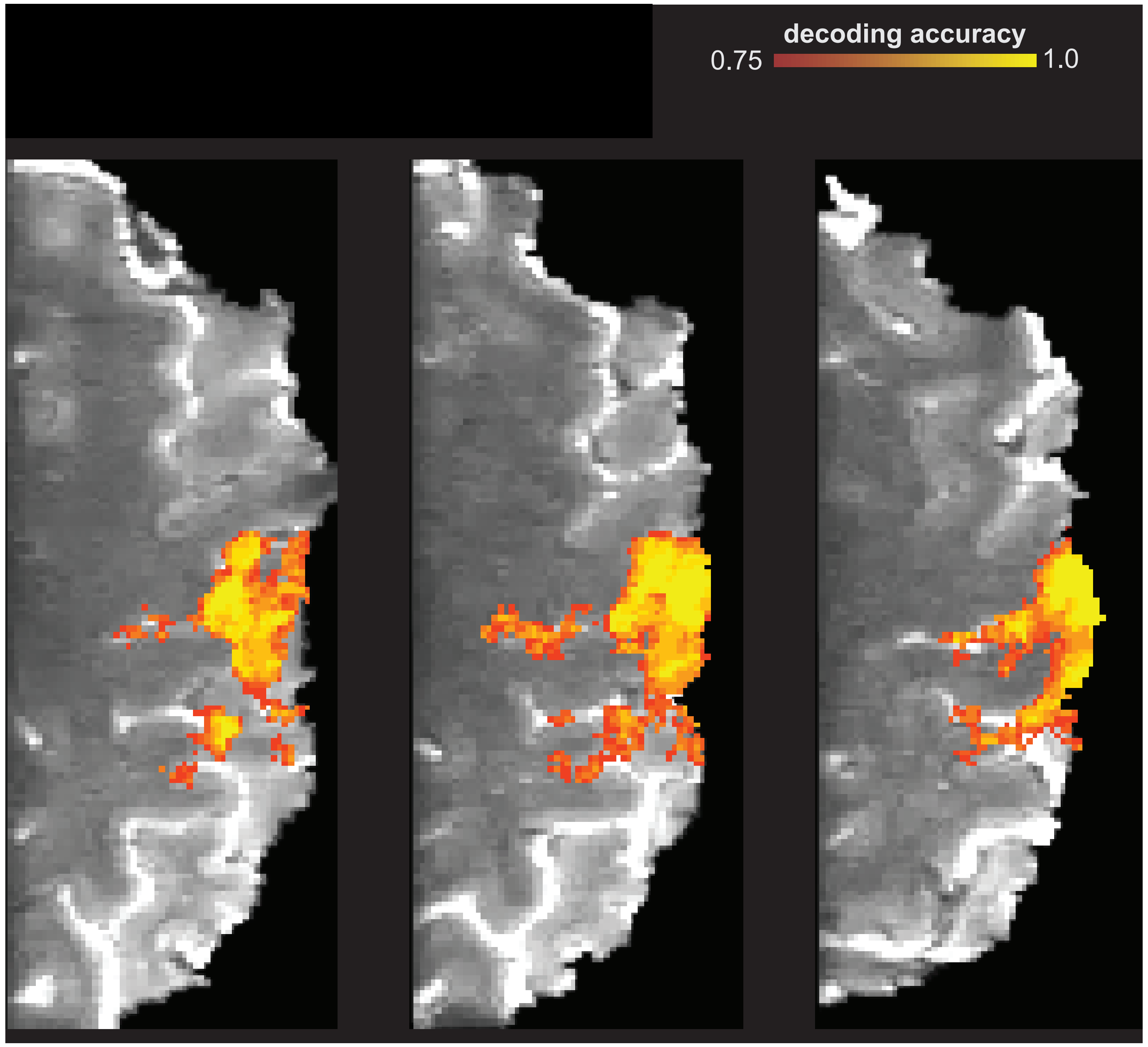
cluster threshold (p)	0,01	0,02	0,03	0,04	0,05
E(clusters)	1	2	3	4	5
nonparametric: clusters	0	1	1	1	1
T-based: clusters	4	7	11	13	17



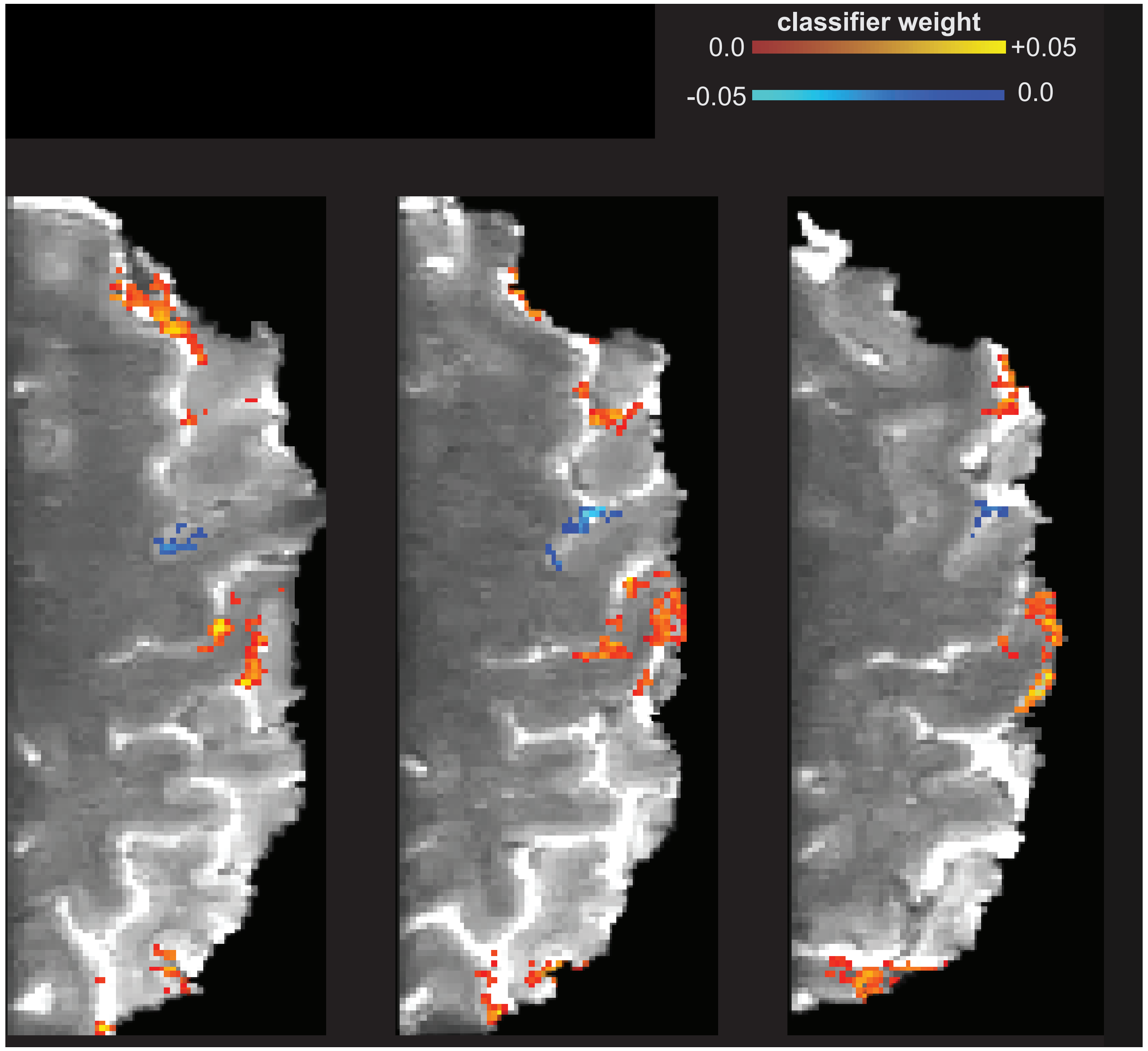
cluster threshold (p)	0,01	0,02	0,03	0,04	0,05
E(clusters)	1	2	3	4	5
nonparametric: (+)clusters	0	1	1	1	1
nonparametric: (-)clusters	0	2	2	2	2
T-based: (+)clusters	1	2670	2670	2670	2670
T-based: (-)clusters	0	2548	2548	2548	2548

MVPA @ 7T

searchlight decoding

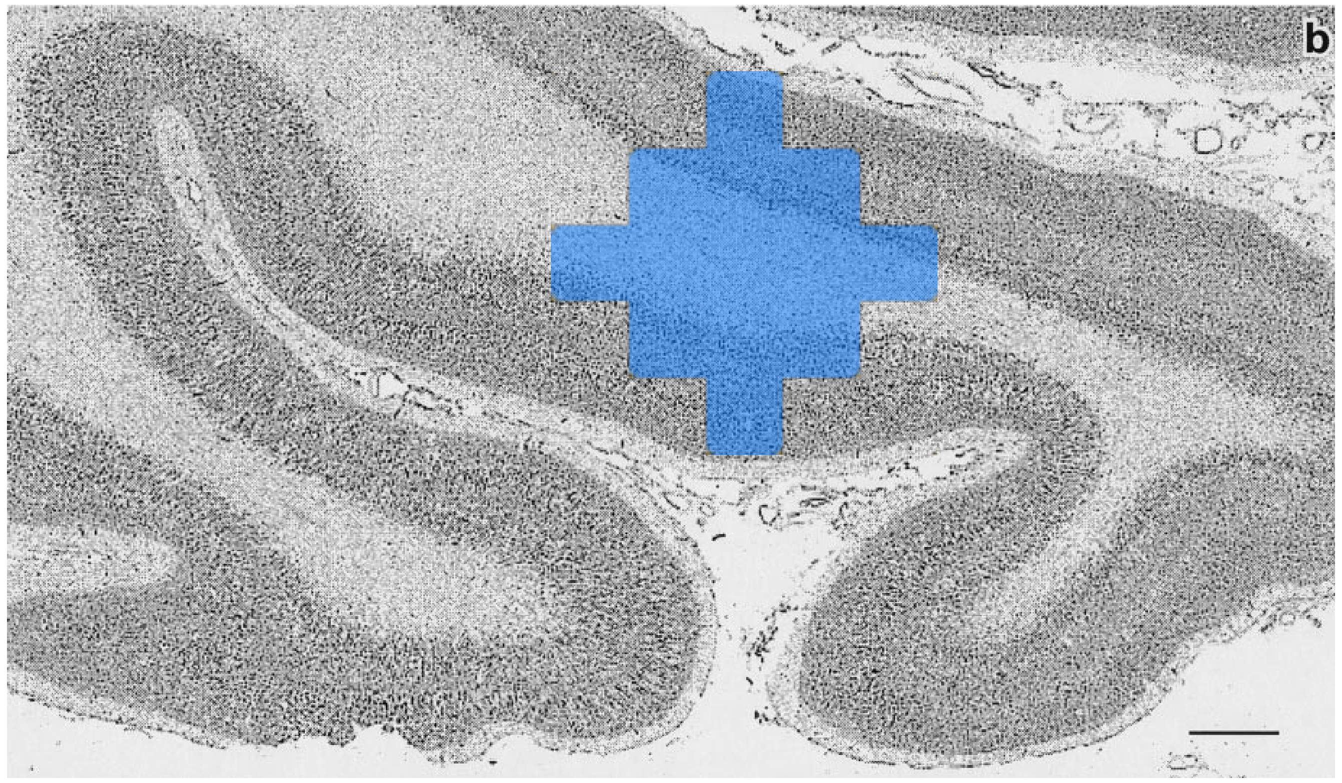


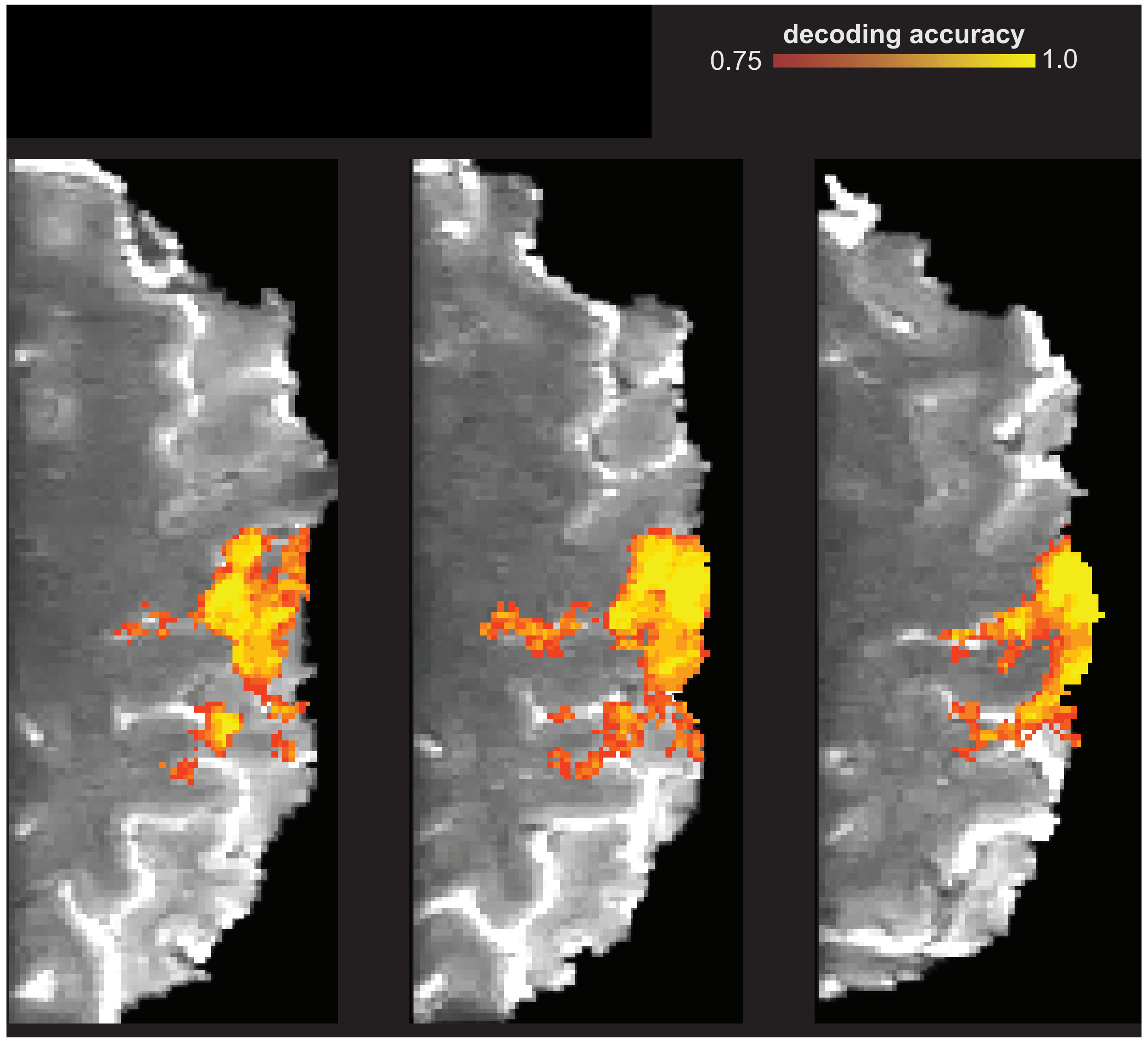
feature weight mapping

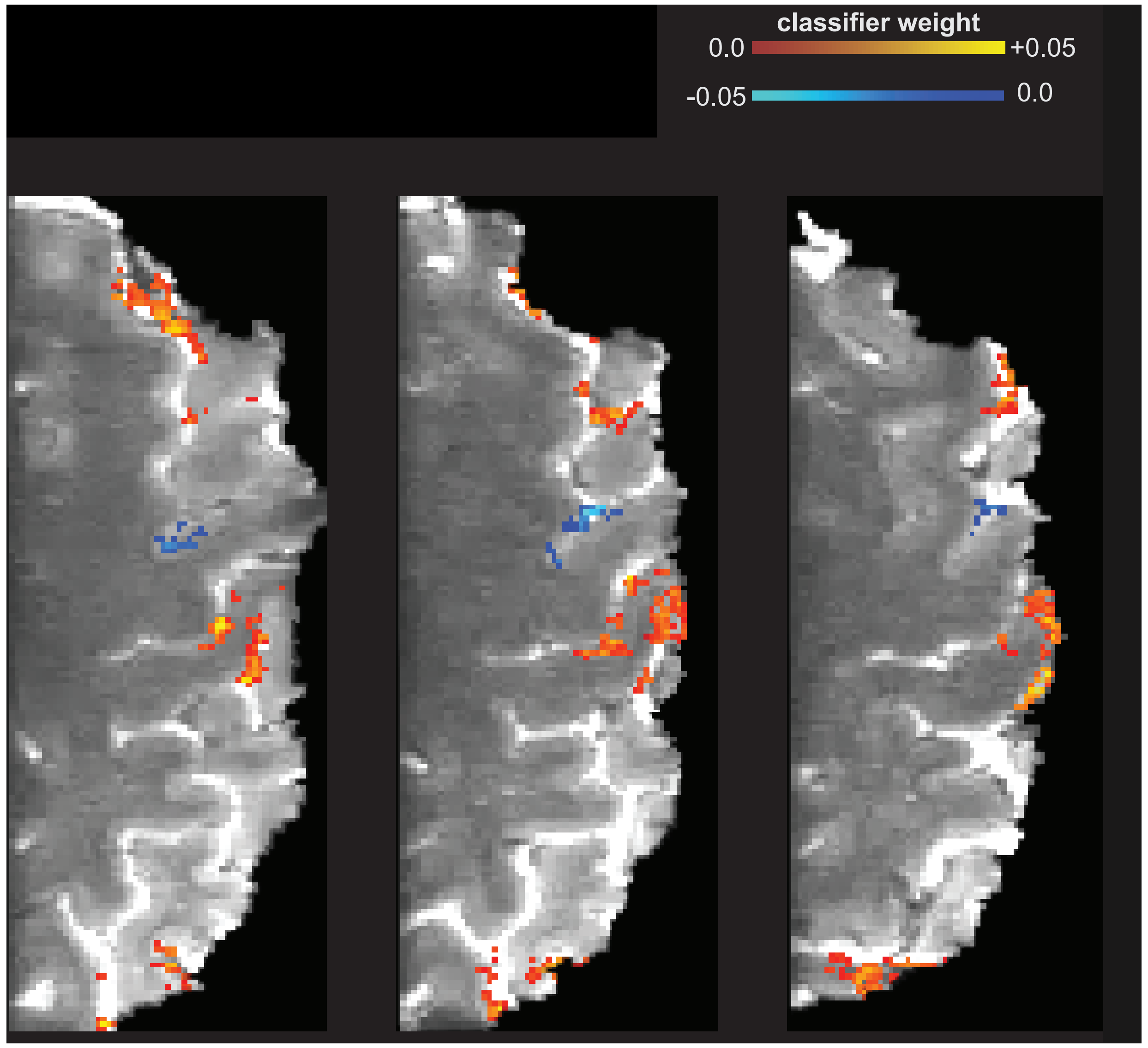


Stelzer, J., Buschmann, T., Lohmann, G., Margulies, D. S., Trampel, R., and Turner, R. (2014). Prioritizing spatial accuracy in high-resolution fMRI data using multivariate feature weight mapping. Brain Imaging Methods 8, 66. doi:10.3389/fnins.2014.00066.

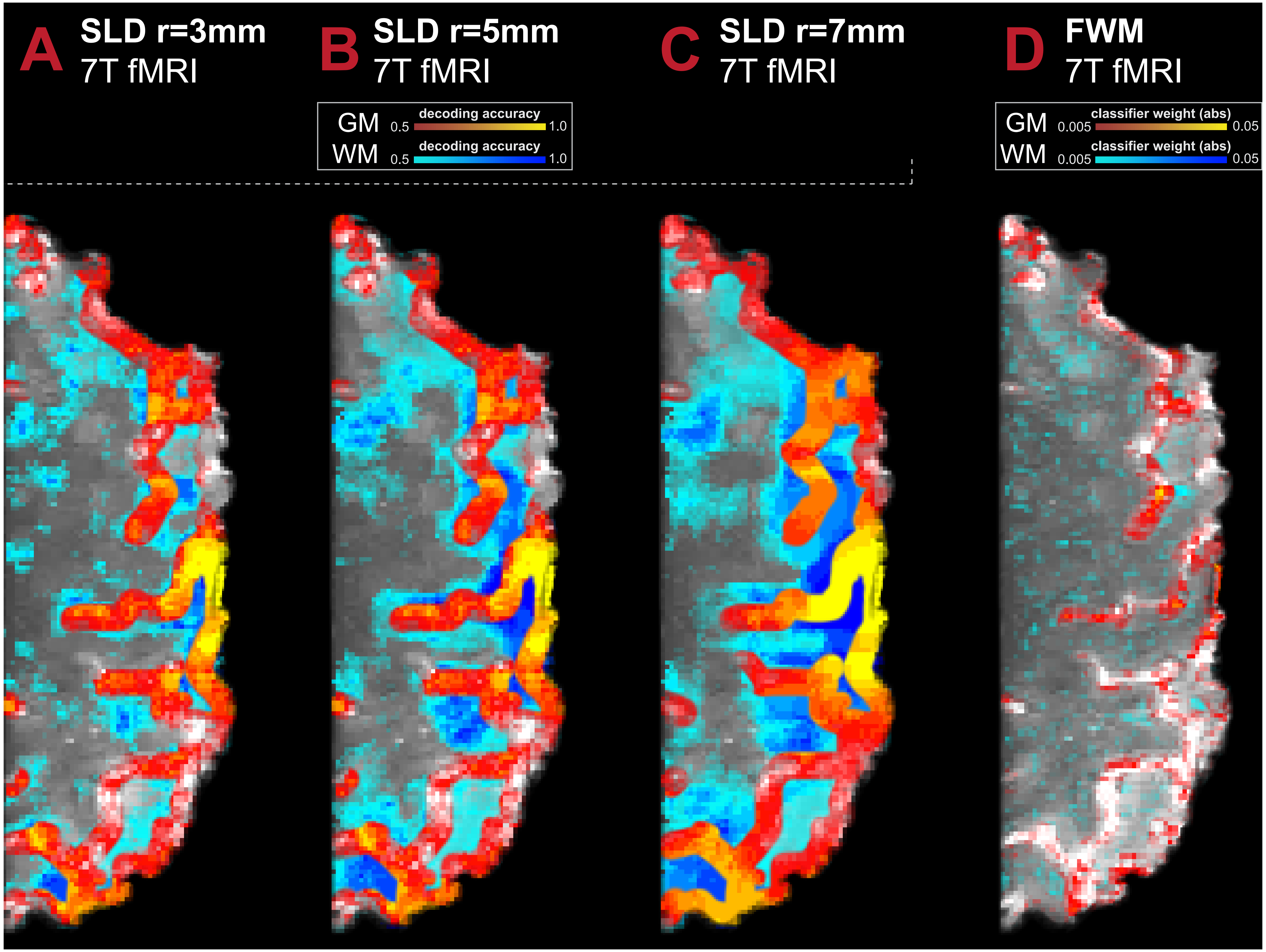
MVPA @ 7T







Stelzer, J., Buschmann, T., Lohmann, G., Margulies, D. S., Trampel, R., and Turner, R. (2014). Prioritizing spatial accuracy in high-resolution fMRI data using multivariate feature weight mapping. Brain Imaging Methods 8, 66. doi:10.3389/fnins.2014.00066.



Stelzer, J., Buschmann, T., Lohmann, G., Margulies, D. S., Trampel, R., and Turner, R. (2014). Prioritizing spatial accuracy in high-resolution fMRI data using multivariate feature weight mapping. Brain Imaging Methods 8, 66. doi:10.3389/fnins.2014.00066.

issues

- computational burden
- spatial smoothness homogeneous
- small-scale patterns invisible (cluster stats)
- group stat issues

Eklund, A., Dufort, P., Villani, M., and LaConte, S. (2014). BROCCOLI: Software for Fast fMRI Analysis on Many-Core CPUs and GPUs. Front Neuroinform 8. doi:10.3389/fninf.2014.00024.

Stelzer, J., Lohmann, G., Mueller, K., Buschmann, T., and Turner, R. (2014). Deficient approaches to human neuroimaging. Front. Hum. Neurosci. 8. doi:10.3389/fnhum.2014.00462.

implementations

- PyMVPA (http://www.pymvpa.org/)
- CoSMoMVPA (http://cosmomvpa.org/)
- Lipsia 3.0 (to be released soon)

thanks for your attention!

& thanks to

Robert Turner

Yi Chen

Gabriele Lohmann

Tilo Buschmann

MAX PLANCK INSTITUTE FOR HUMAN COGNITIVE AND BRAIN SCIENCES LEIPZIG

Max-Planck-Institut für biologische Kybernetik