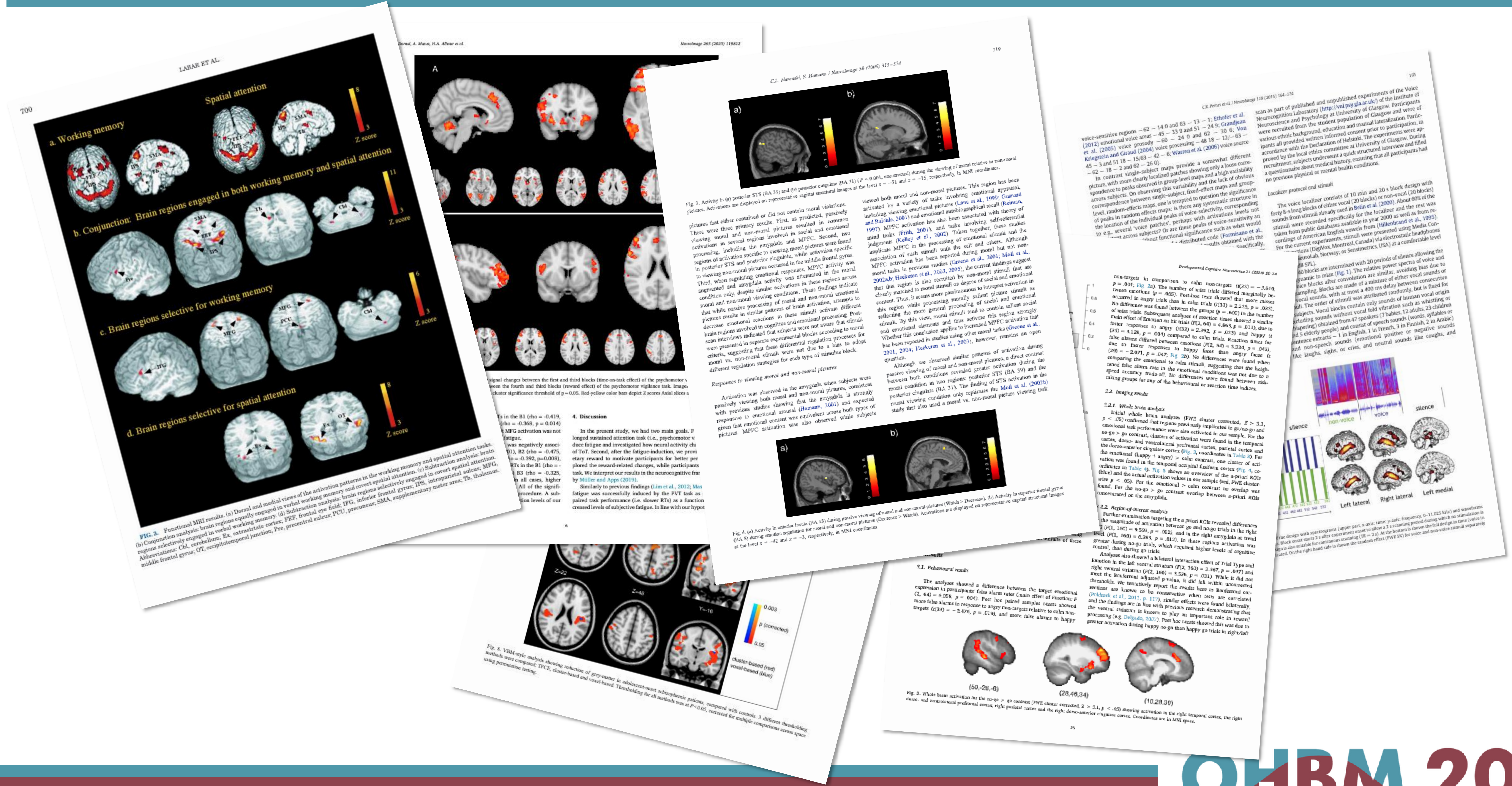
BRISBANE
OHBM 2025 JUNE 24-28

Cluster analysis revisited

Wouter Weeda

‘Classical’ cluster-extent analysis



‘Classical’ cluster-extent analysis

- For most functional MRI studies measured signal comes from distinct locations in the brain called voxels: a 3-dimensional grid of 3x3x3 mm cubes.
- Inference in functional MRI is done on each location (voxel) separately.
- The maps that you often see are the outcomes of this inference (usually in the form of a z or t-statistic indicating significance).

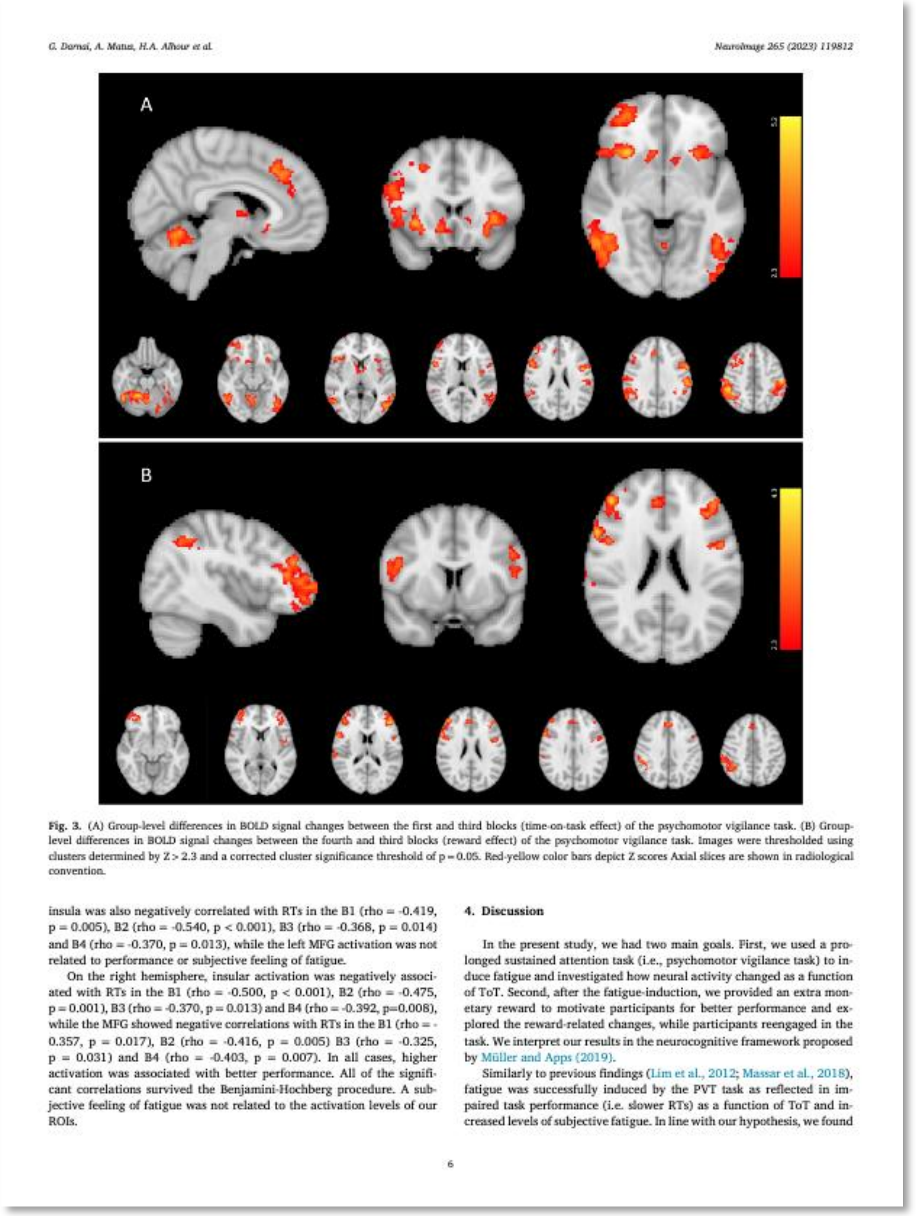


Fig. 3. (A) Group-level differences in BOLD signal changes between the first and third blocks (time-on-task effect) of the psychomotor vigilance task. (B) Group-level differences in BOLD signal changes between the fourth and third blocks (reward effect) of the psychomotor vigilance task. Images were thresholded using clusters determined by Z > 2.3 and a corrected cluster significance threshold of p = 0.05. Red-yellow color bars depict Z scores Axial slices are shown in radiological convention.

insula was also negatively correlated with RTs in the B1 (rho = -0.419, p = 0.005), B2 (rho = -0.540, p < 0.001), B3 (rho = -0.368, p = 0.014) and B4 (rho = -0.370, p = 0.013), while the left MFG activation was not related to performance or subjective feeling of fatigue.

On the right hemisphere, insular activation was negatively associated with RTs in the B1 (rho = -0.500, p < 0.001), B2 (rho = -0.475, p = 0.001), B3 (rho = -0.370, p = 0.013) and B4 (rho = -0.392, p=0.008), while the MFG showed negative correlations with RTs in the B1 (rho = -0.357, p = 0.017), B2 (rho = -0.416, p = 0.005) B3 (rho = -0.325, p = 0.031) and B4 (rho = -0.403, p = 0.007). In all cases, higher activation was associated with better performance. All of the significant correlations survived the Benjamini-Hochberg procedure. A subjective feeling of fatigue was not related to the activation levels of our ROIs.

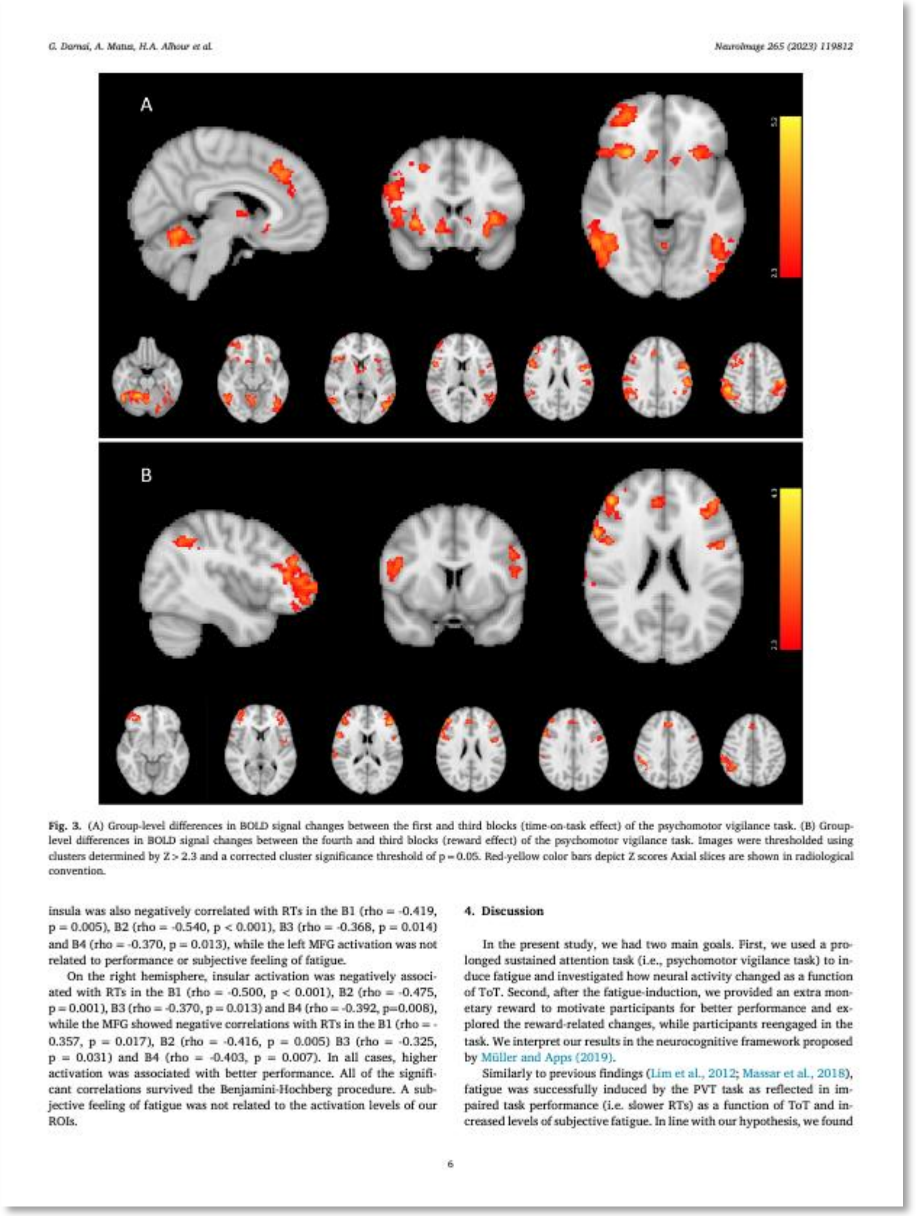
4. Discussion

In the present study, we had two main goals. First, we used a prolonged sustained attention task (i.e., psychomotor vigilance task) to induce fatigue and investigated how neural activity changed as a function of ToT. Second, after the fatigue-induction, we provided an extra monetary reward to motivate participants for better performance and explored the reward-related changes, while participants reengaged in the task. We interpret our results in the neurocognitive framework proposed by Müller and Apps (2019).

Similarly to previous findings (Lim et al., 2012; Massar et al., 2018), fatigue was successfully induced by the PVT task as reflected in impaired task performance (i.e. slower RTs) as a function of ToT and increased levels of subjective fatigue. In line with our hypothesis, we found

‘Classical’ cluster-extent analysis

- The goal of fMRI inference is to decide for each voxel whether it is active or not (using a hypothesis test).
- For each test we allow a little uncertainty of whether our decision is the right one.
- When doing multiple tests, the chances of making a wrong decision somewhere in our ‘family’ of tests increases dramatically.
- The family-wise error rate (FWER) of our family of tests is what we want ‘controlled’.



G. Darnai, A. Matuz, H.A. Alhour et al. NeuroImage 265 (2023) 119812

Fig. 3. (A) Group-level differences in BOLD signal changes between the first and third blocks (time-on-task effect) of the psychomotor vigilance task. (B) Group-level differences in BOLD signal changes between the fourth and third blocks (reward effect) of the psychomotor vigilance task. Images were thresholded using clusters determined by Z > 2.3 and a corrected cluster significance threshold of p = 0.05. Red-yellow color bars depict Z scores Axial slices are shown in radiological convention.

insula was also negatively correlated with RTs in the B1 (rho = -0.419, p = 0.005), B2 (rho = -0.540, p < 0.001), B3 (rho = -0.368, p = 0.014) and B4 (rho = -0.370, p = 0.013), while the left MFG activation was not related to performance or subjective feeling of fatigue.

On the right hemisphere, insular activation was negatively associated with RTs in the B1 (rho = -0.500, p < 0.001), B2 (rho = -0.475, p = 0.001), B3 (rho = -0.370, p = 0.013) and B4 (rho = -0.392, p=0.008), while the MFG showed negative correlations with RTs in the B1 (rho = -0.357, p = 0.017), B2 (rho = -0.416, p = 0.005) B3 (rho = -0.325, p = 0.031) and B4 (rho = -0.403, p = 0.007). In all cases, higher activation was associated with better performance. All of the significant correlations survived the Benjamini-Hochberg procedure. A subjective feeling of fatigue was not related to the activation levels of our ROIs.

4. Discussion

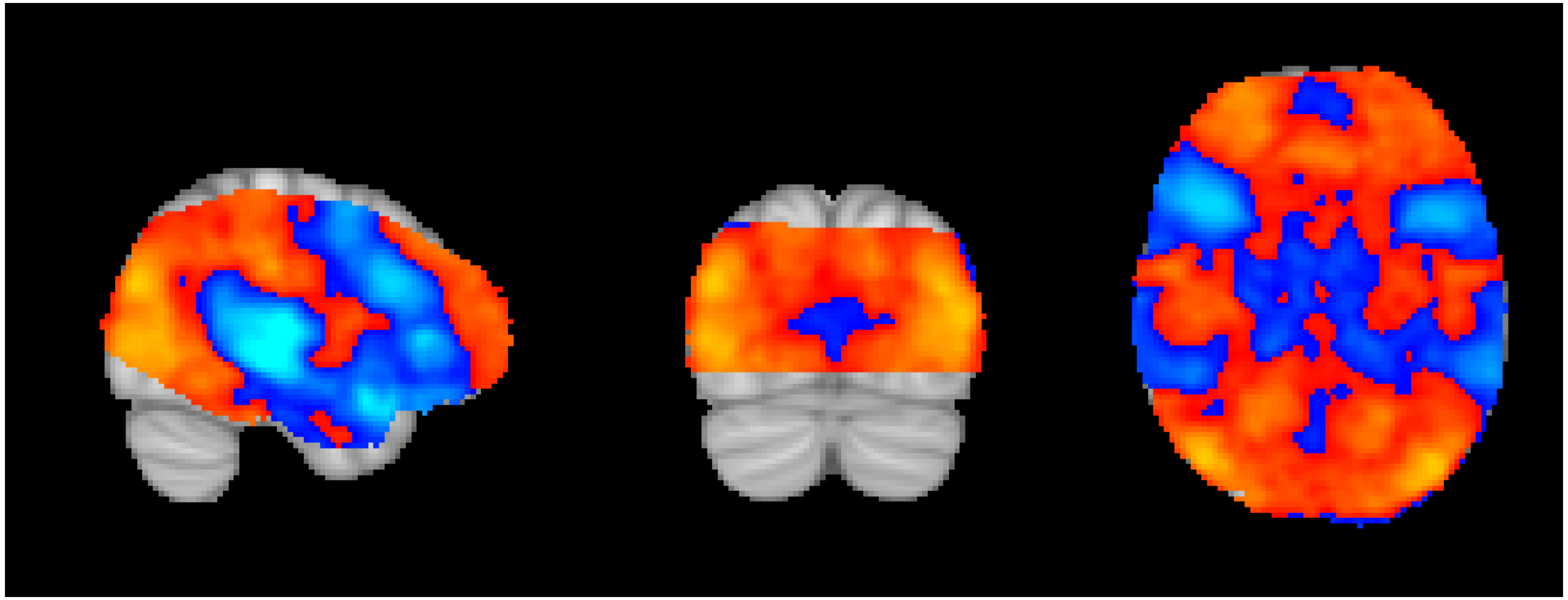
In the present study, we had two main goals. First, we used a prolonged sustained attention task (i.e., psychomotor vigilance task) to induce fatigue and investigated how neural activity changed as a function of ToT. Second, after the fatigue-induction, we provided an extra monetary reward to motivate participants for better performance and explored the reward-related changes, while participants reengaged in the task. We interpret our results in the neurocognitive framework proposed by Müller and Apps (2019).

Similarly to previous findings (Lim et al., 2012; Massar et al., 2018), fatigue was successfully induced by the PVT task as reflected in impaired task performance (i.e. slower RTs) as a function of ToT and increased levels of subjective fatigue. In line with our hypothesis, we found

6

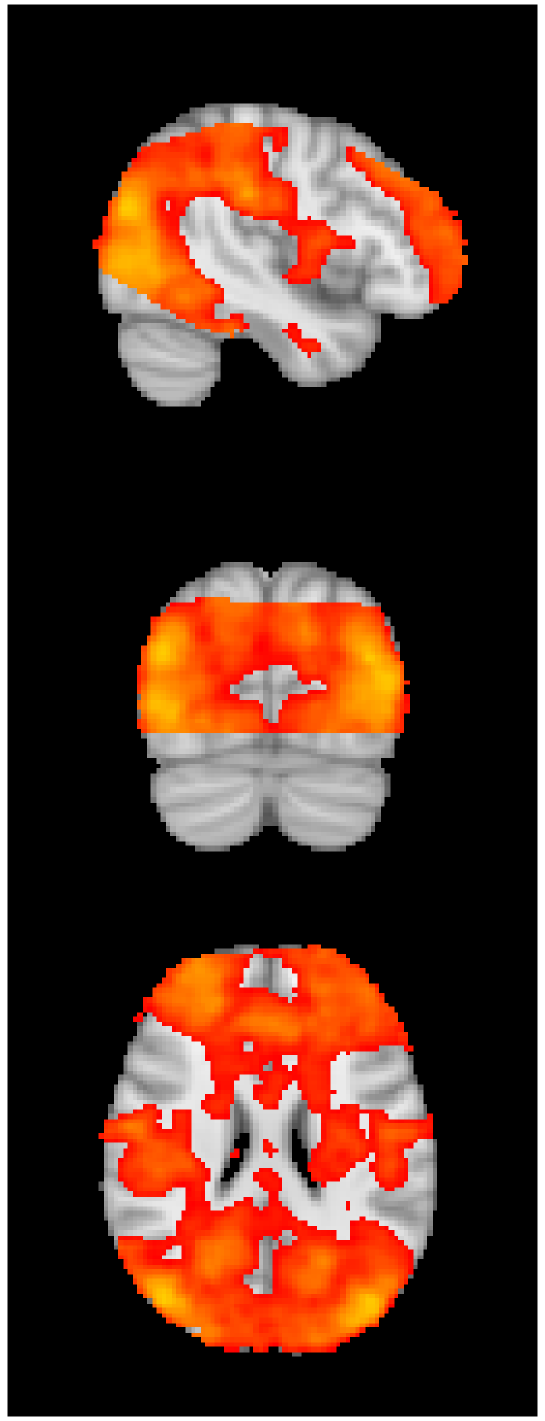
‘Classical’ cluster-extent analysis

Study on vocal and non-vocal sounds, Pernet et al., 2015



‘Classical’ cluster-extent analysis

- In total 166.407 in-mask voxels.
- Focus only on positive values for now.
- Z-statistics indicate whether a voxel is more active in the *non-vocal* condition than in the *vocal* condition.
    - H_0= not active (z-value = 0)
    - H_1= active (z-value > 0)

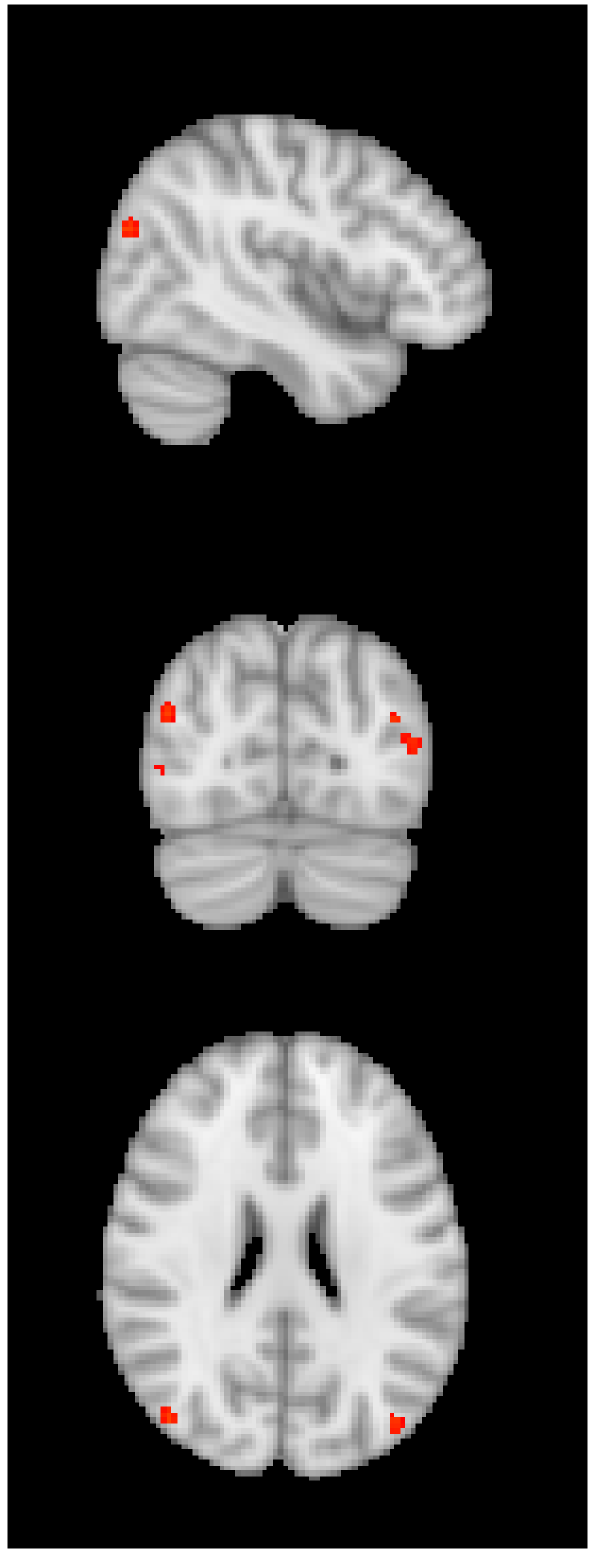


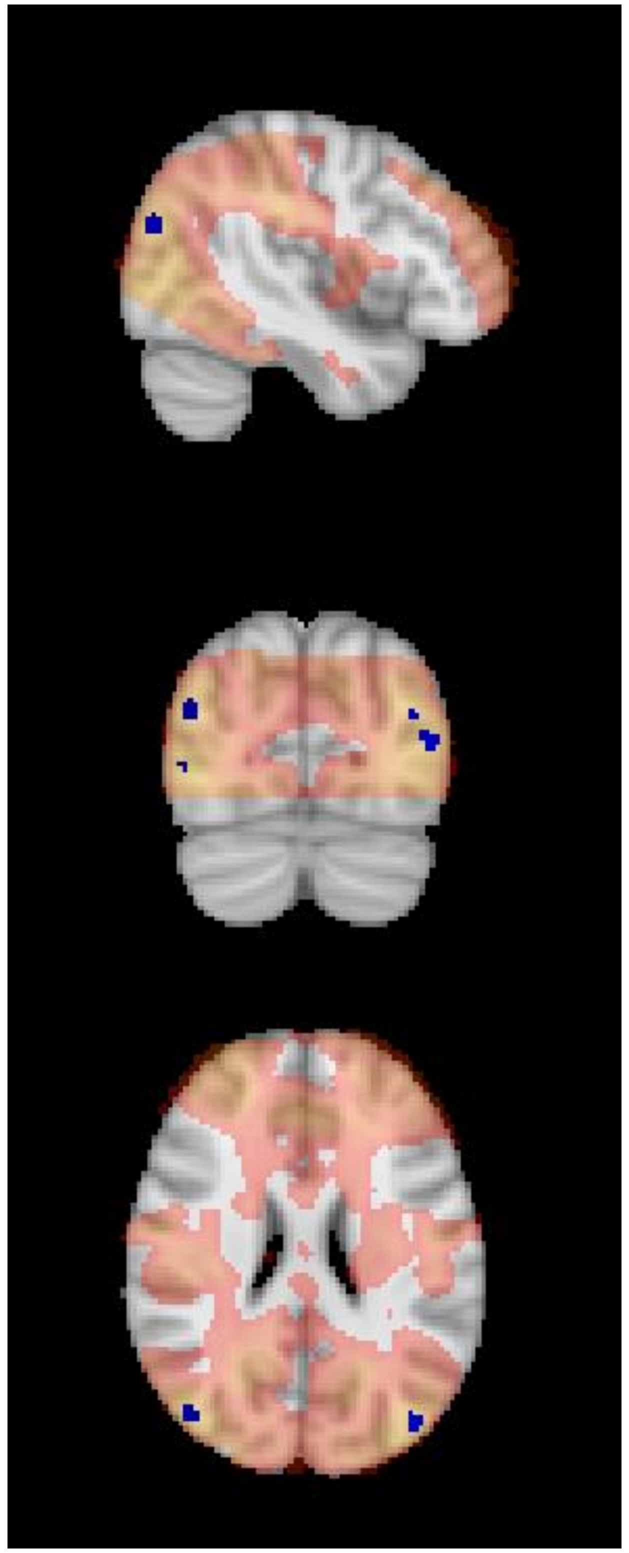
'Classical' cluster-extent analysis

- Controls the FWER over all voxels in the brain (mask). Family = all voxels.
- Easiest method to control the FWER is Bonferroni correction.
- Calculated by setting the per-voxel α to be α / #voxels

  .05 / 166407 = .0000003

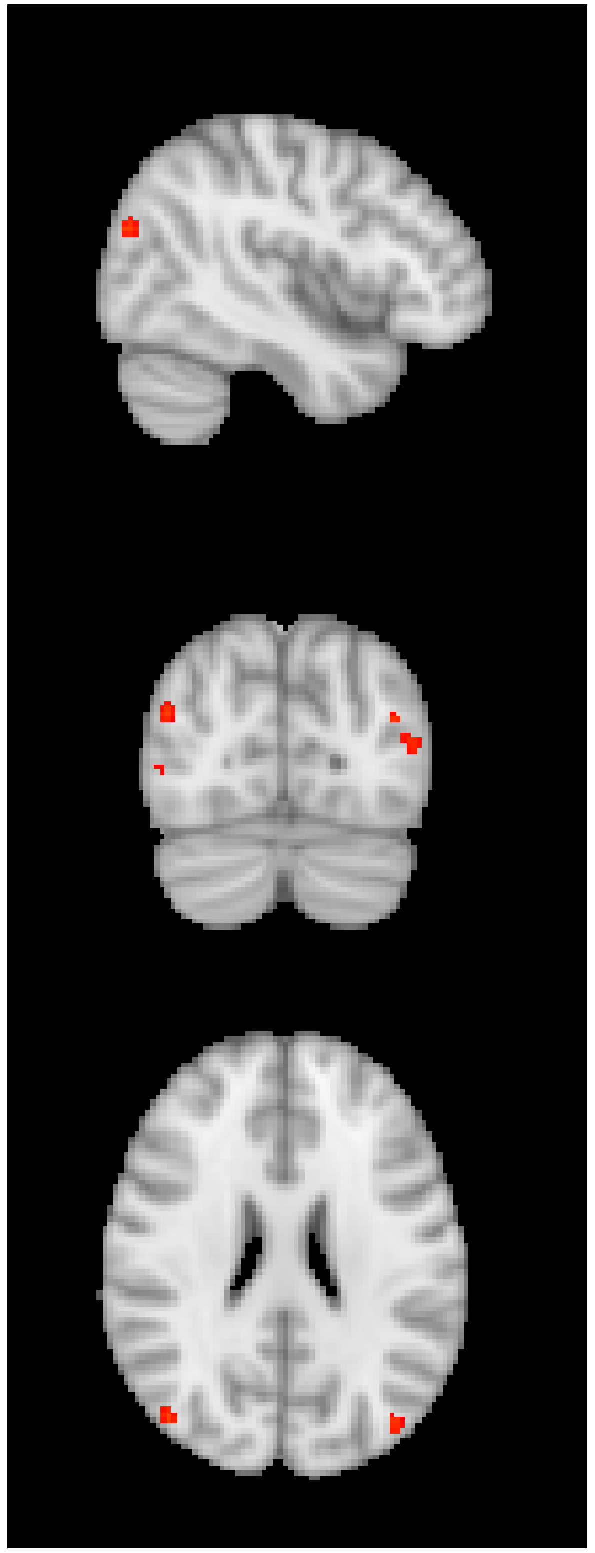
- Usually not very powerful.

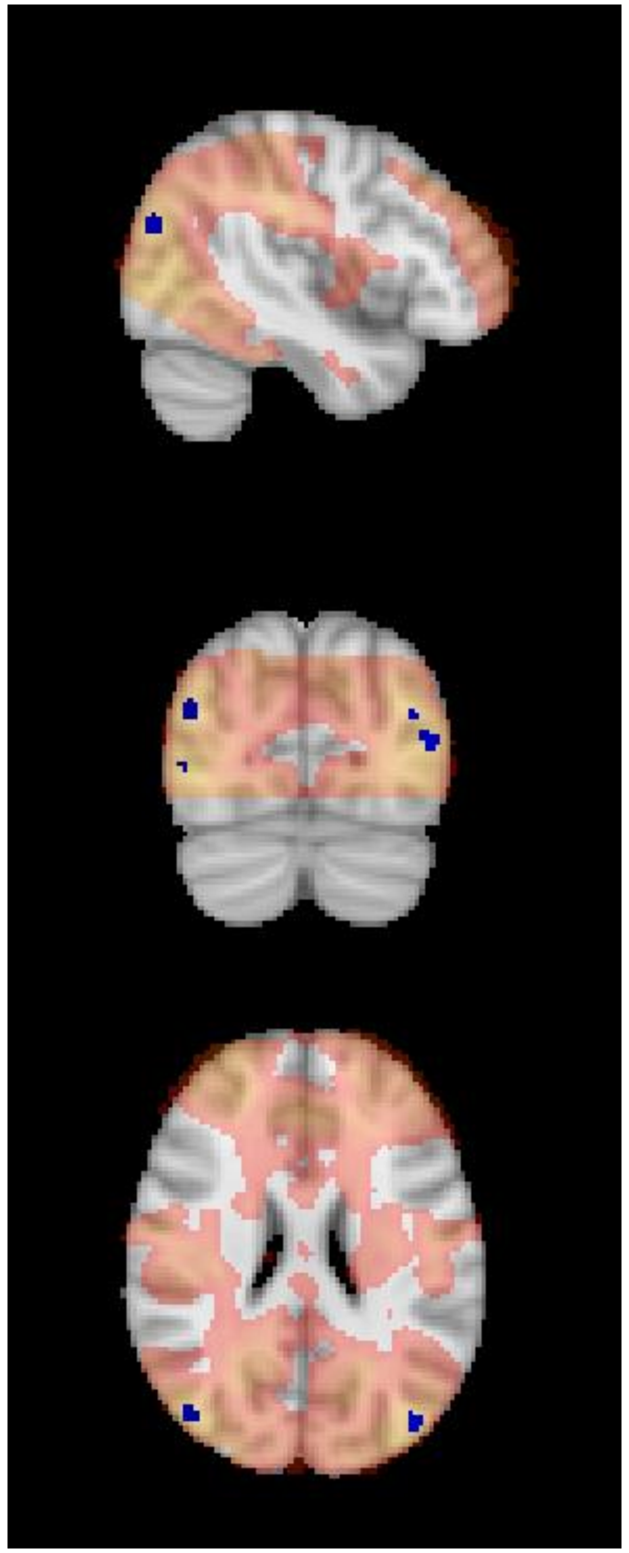




'Classical' cluster-extent analysis

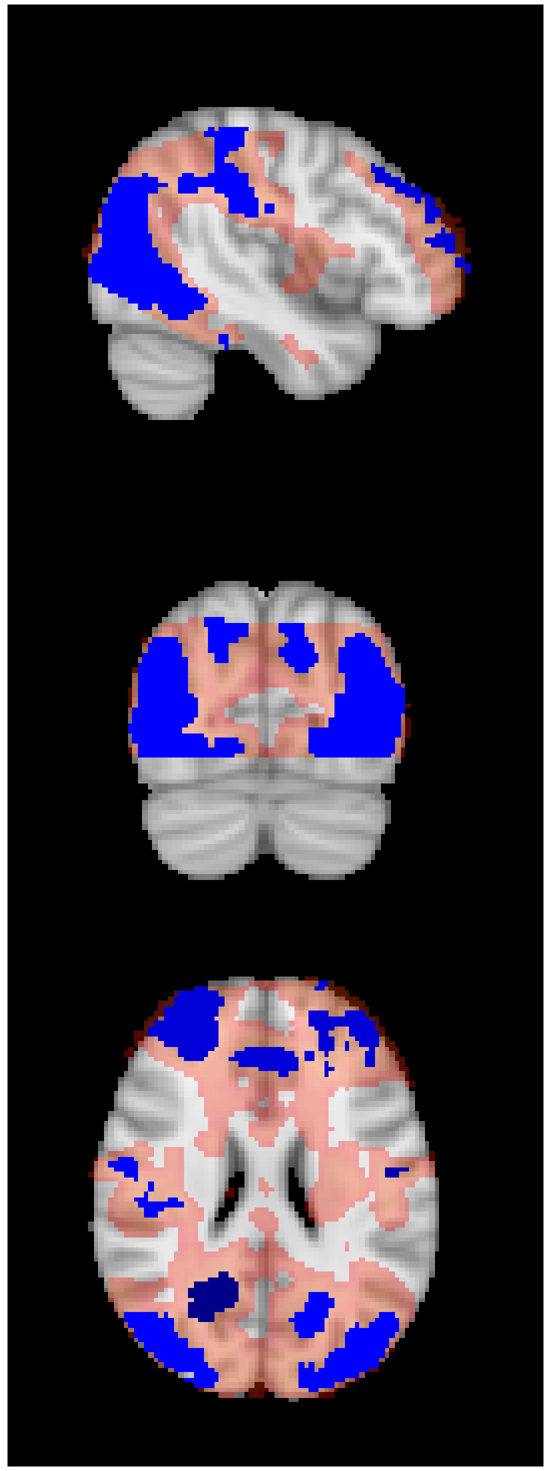
- But…
- Since our family is all voxels, we know exactly where the activation is!
- In other words: we have high spatial specificity.
- (because the chance of any of these voxels being a false-positive < 5%)





‘Classical’ cluster-extent analysis

- Usually, we are not interested in single-voxel activity per se. A more natural unit is a ‘cluster’ of voxels (which we will name ’blob’).
- A cluster or blob is defined as a contiguous/connected set of voxels.
- We control the number of false-positive blobs (our family in FWER is thus all possible blobs, not all voxels).



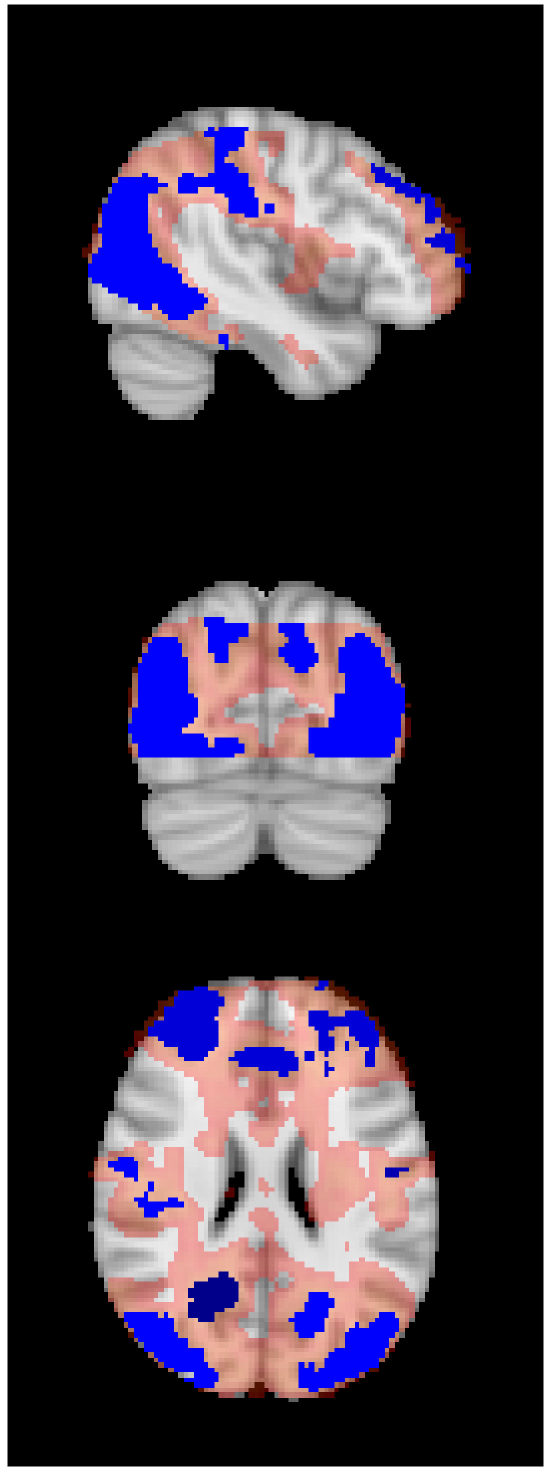
‘Classical’ cluster-extent analysis

- In practice, using a two-step approach:

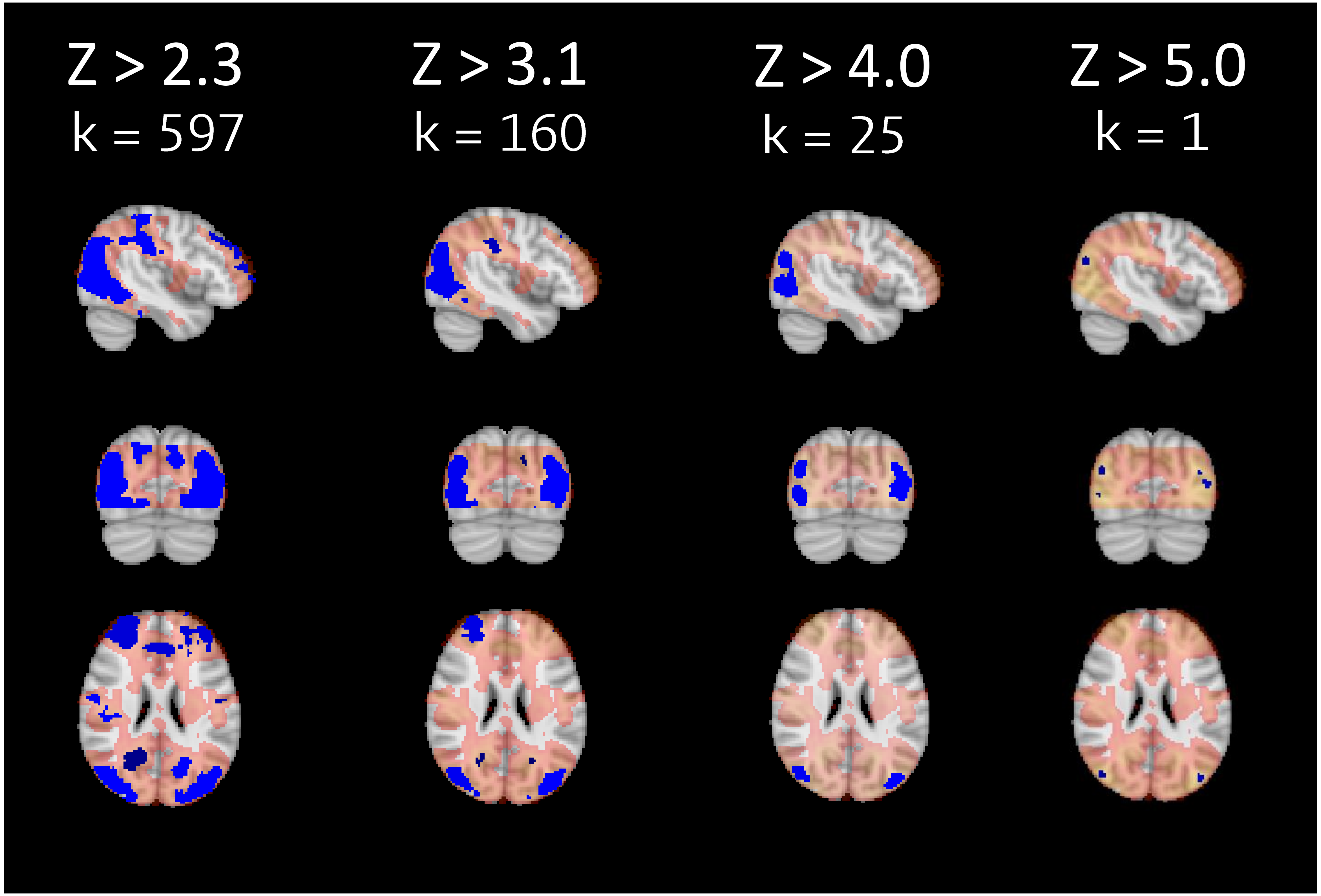
  1. Choose a ‘cluster-forming’ threshold *z* and estimate the size of all contiguous clusters above this threshold.

  Determine the minimum cluster size *k* that occurs by chance under the null (95%) given the smoothness of the data and the chosen threshold *z* (e.g., using RFT or permutations)

  2. Check which clusters are larger than *k* (all clusters that are larger are significant).



‘Classical’ cluster-extent analysis



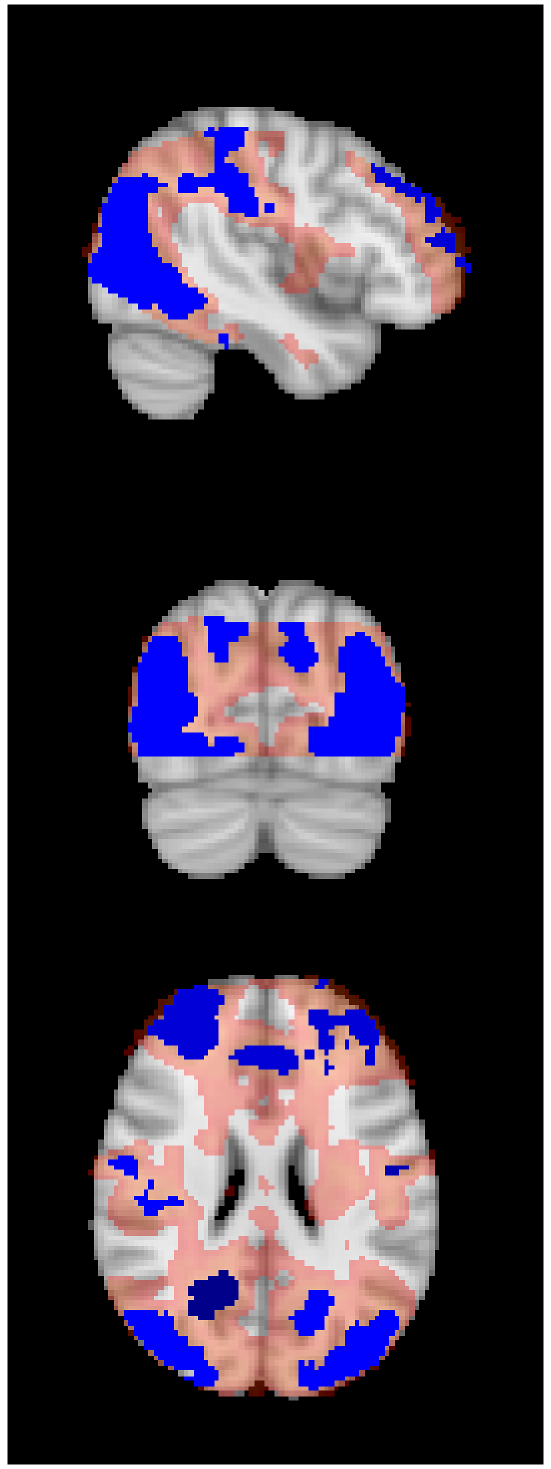
‘Classical’ cluster-extent analysis

- More powerful than voxel-wise approaches, but… more powerful in detecting activation, not in localizing it.

- Because of hypotheses being on the ‘cluster’ level:

  - Non-significant when cluster-extent is smaller than k
  - Significant when cluster-extent is larger than k

- No information about voxels within a cluster (clusters are large enough or not).



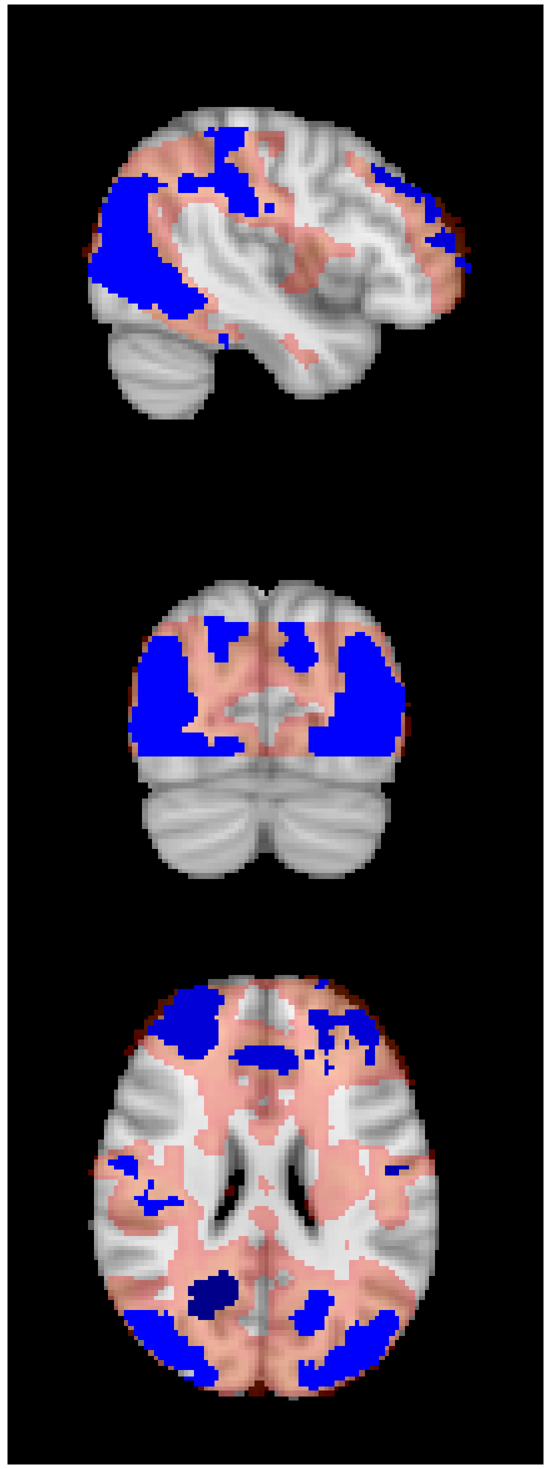
'Classical' cluster-extent analysis

- Formal way of stating this:

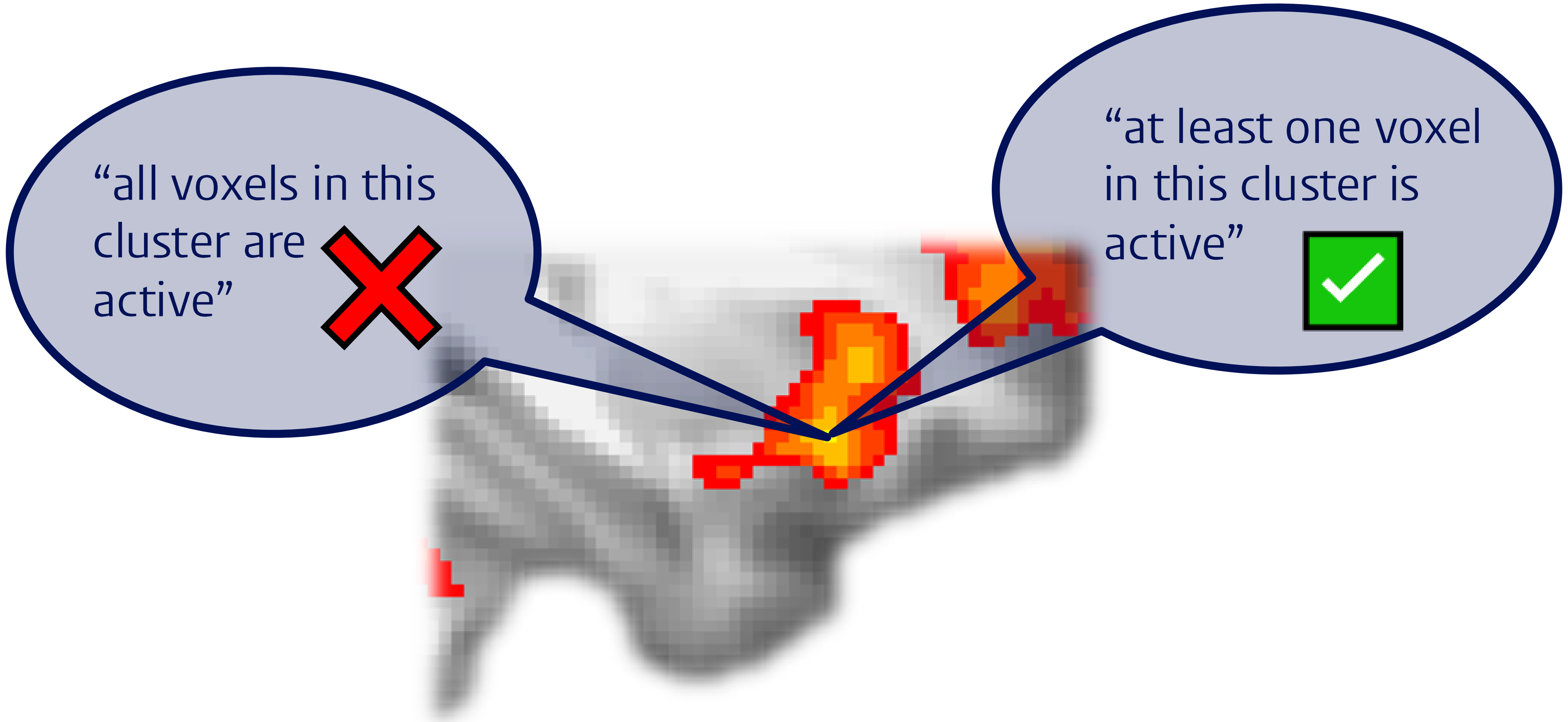
  H_0= no activation within a cluster
  H_1= at least one voxel active within a cluster

- So, the larger the cluster found, the less we know about activation within a cluster.

- This is called the Spatial specificity paradox (Woo et al., 2014, Lindquist & Mejia, 2015).



Spatial Specificity Paradox

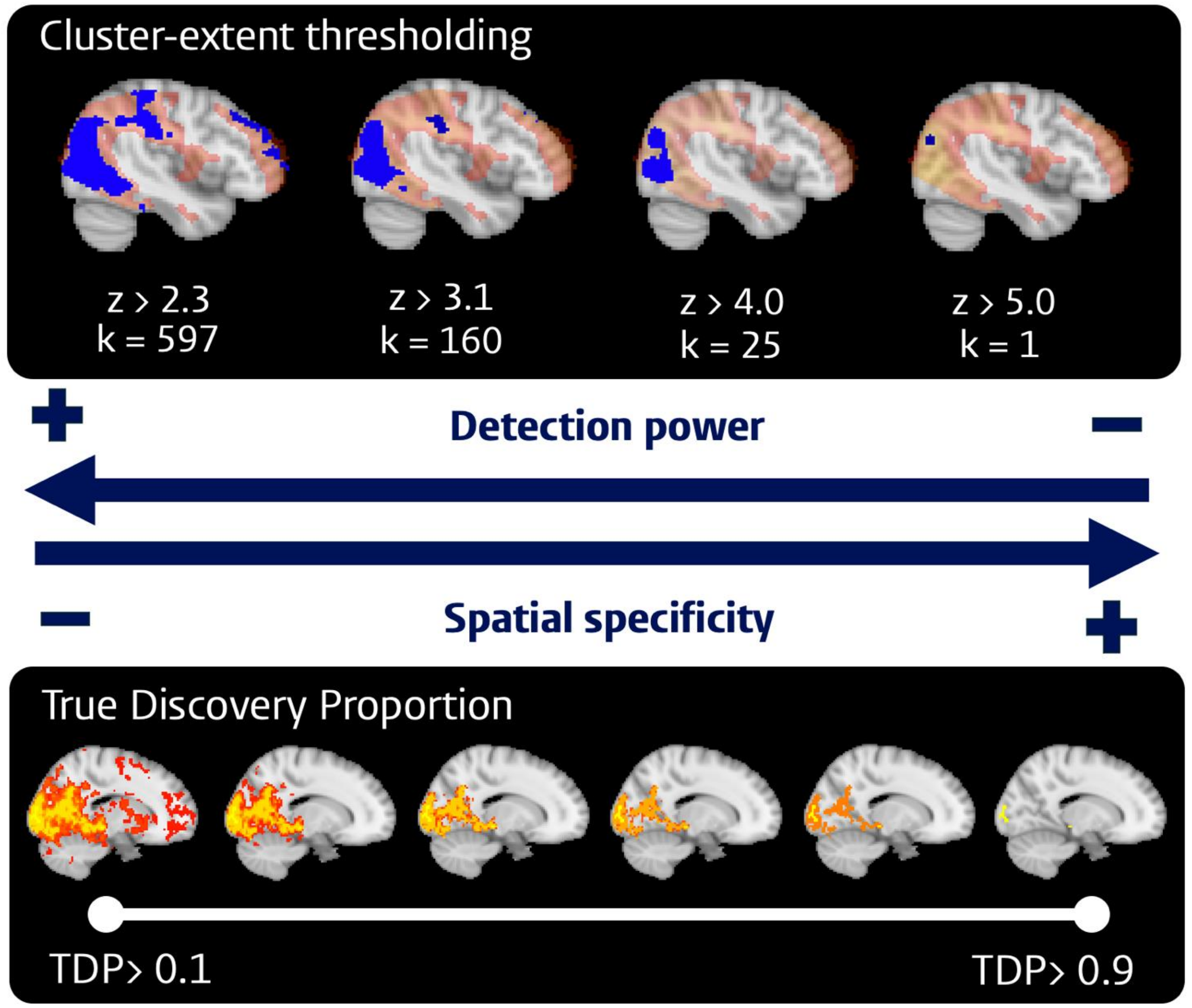


H_0= no activation within a cluster

H_1= at least one voxel active within a cluster

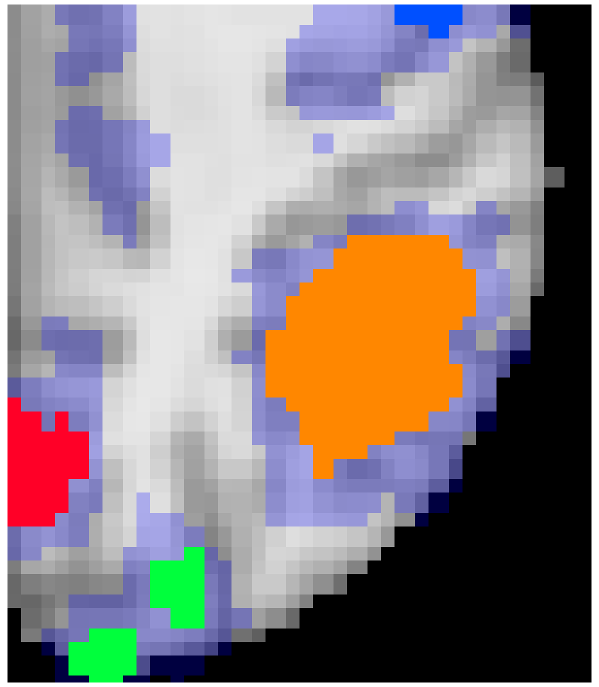
Spatial Specificity Paradox

- The statement “there is at least one voxel active” could mean that all voxels in a cluster are active. Or it could be one, we just don’t know.
- Intuition is that there is usually more than one voxel active. But cluster-extent statistics don’t allow us to test that.
- We can use TDP based methods to give us an in-depth analysis of what clusters are made of.

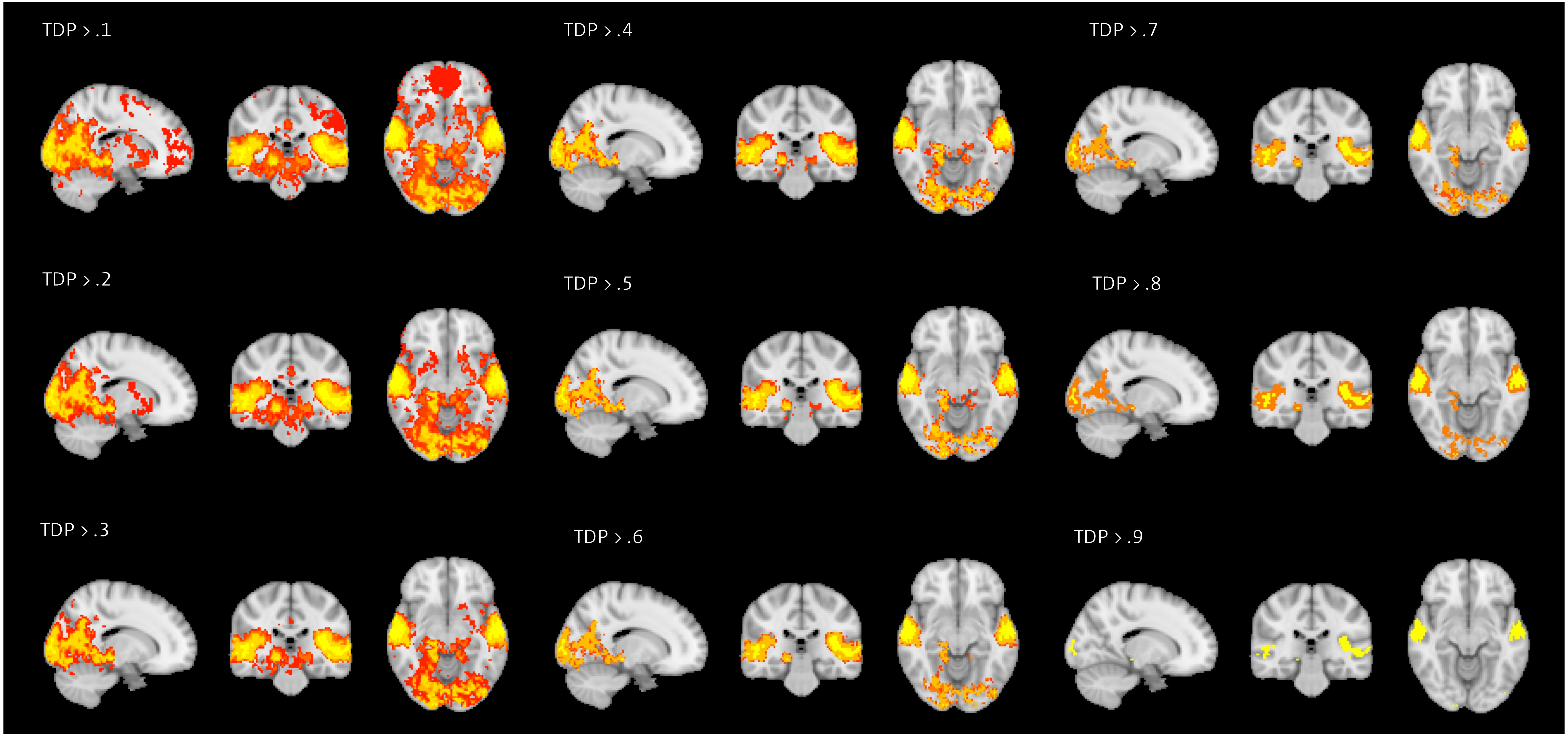


True Discovery Proportion (TDP) based methods
Rosenblatt, Finos, Weeda, Solari & Goeman (2018)

- Multiple testing correction method based on closed testing.
- True Discovery Proportion (TDP) based methods allow us to estimate the number of truly active voxels within a cluster, for all possible clusters, as many times a researcher wants, with full FWER control.
- We can also estimate clusters with at least a certain TDP. For example, “what is the largest cluster that contains at least 60% active voxels (TDP > .6)?”

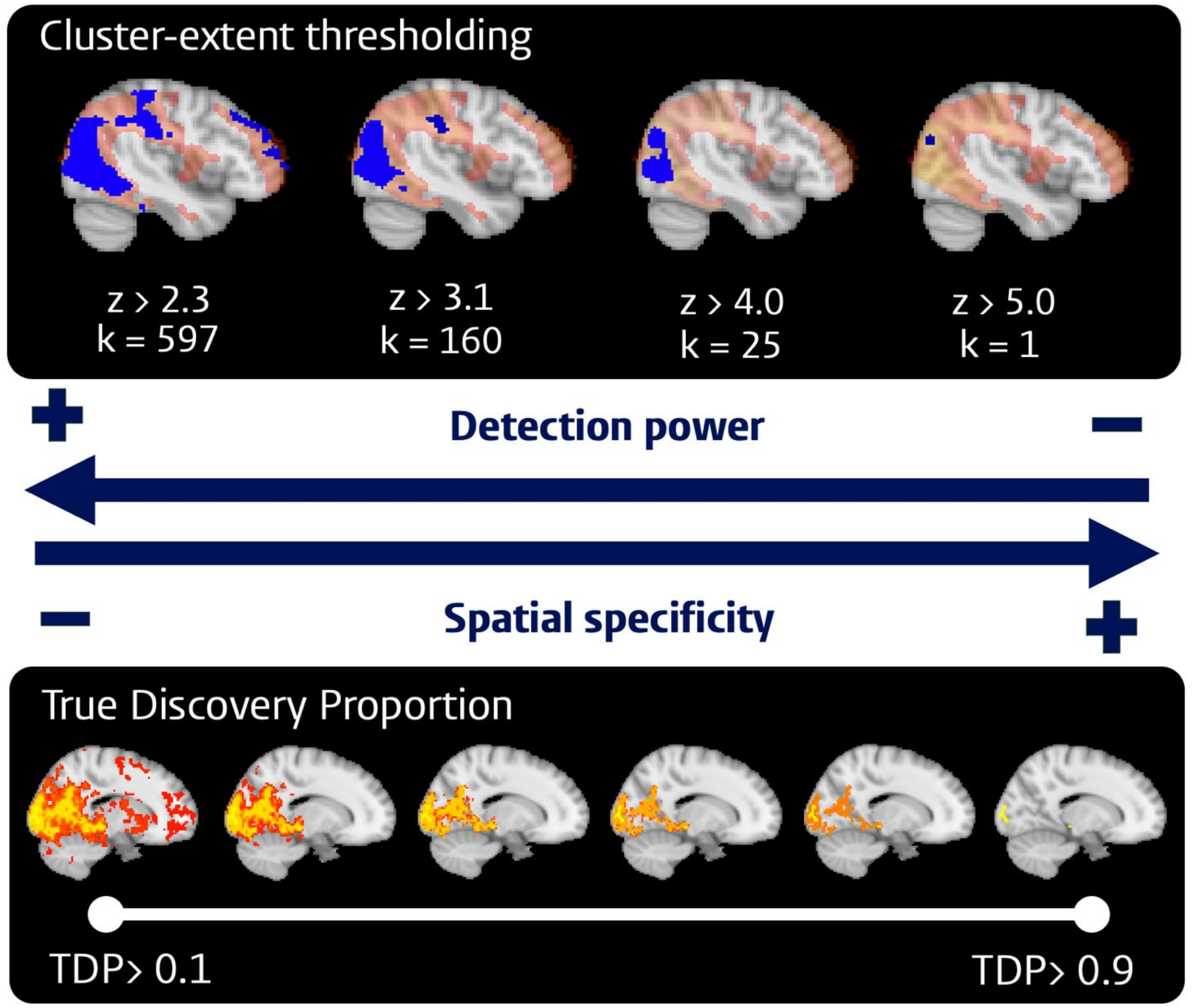


True Discovery Proportion (TDP) based methods



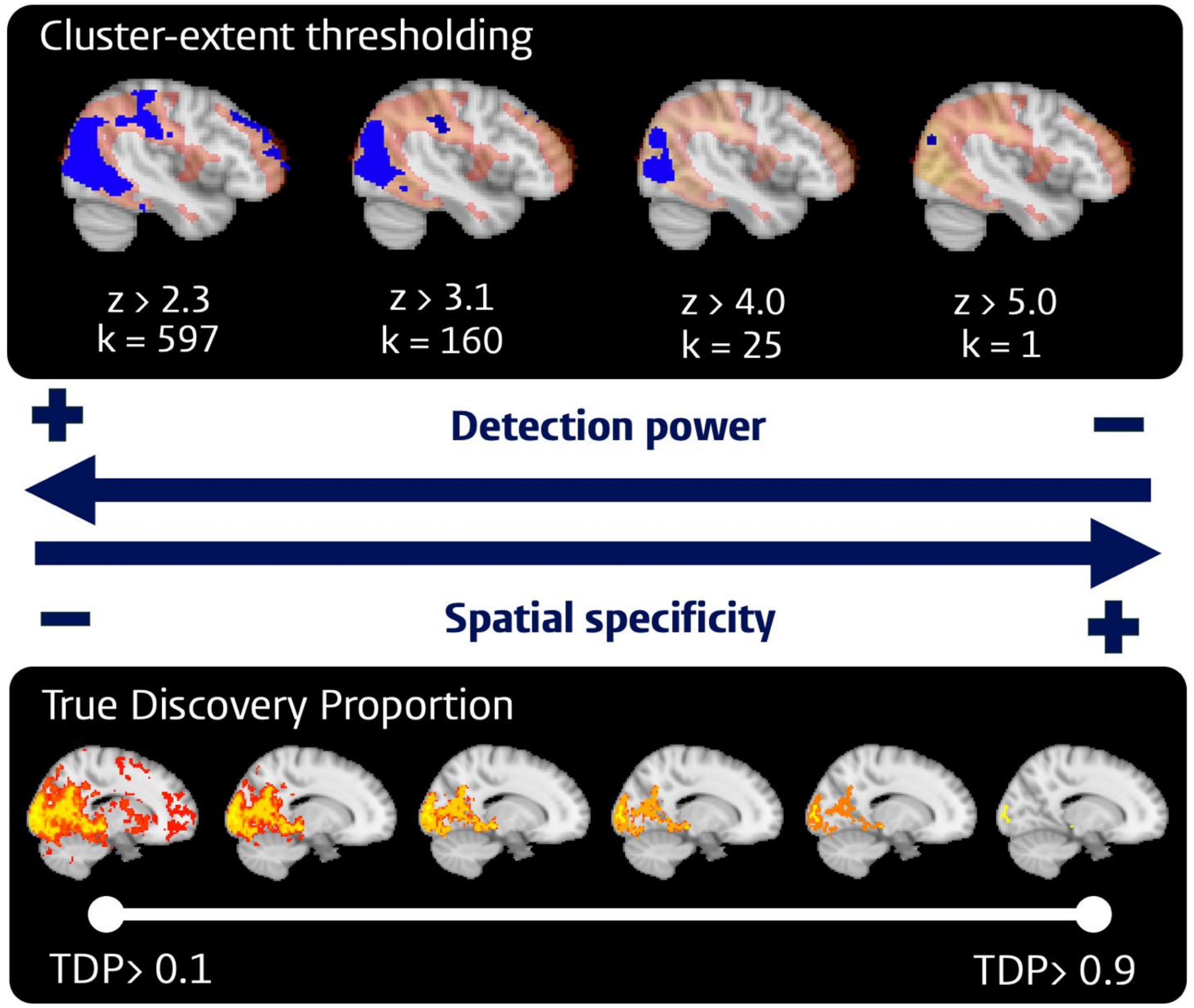
True Discovery Proportion (TDP) based methods

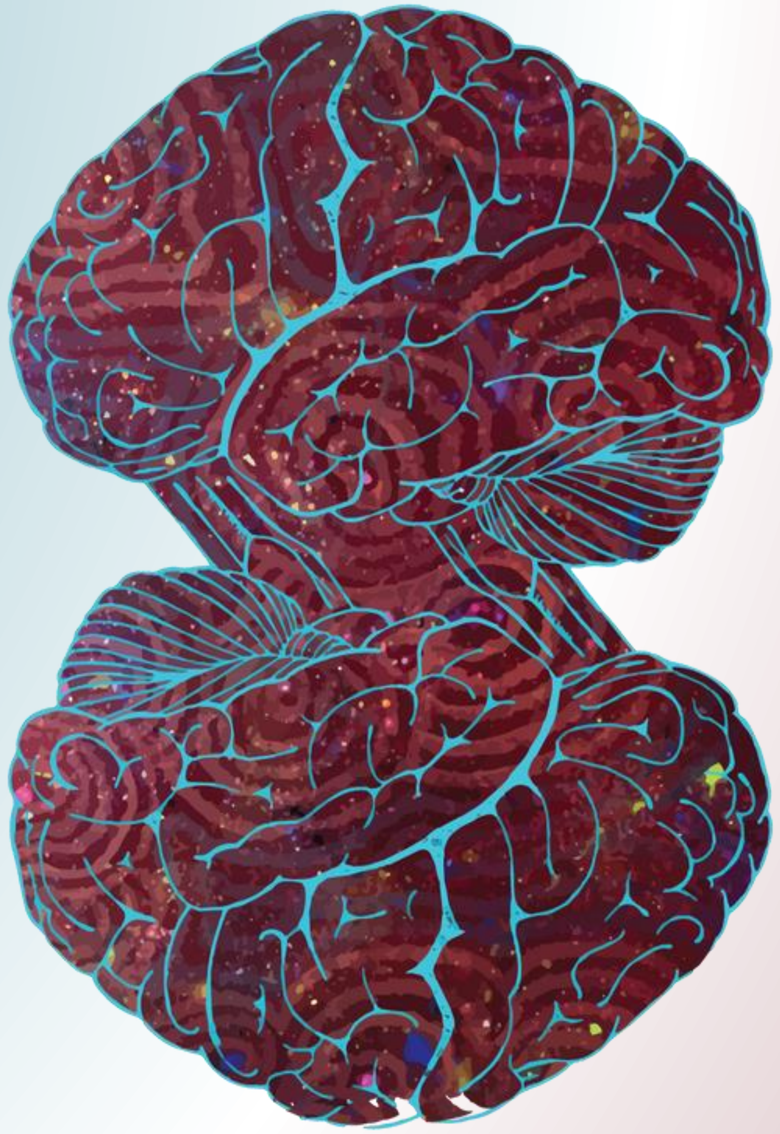
- Both methods are a trade-off between detection and localization, but:
- While cluster extent inference cannot go beyond the cluster-level
- TDP based methods can quantify this trade-off explicitly.



True Discovery Proportion (TDP) based methods

- TDP-based methods can mitigate the spatial shortcomings of cluster-extent thresholding by:
  - Quantifying the amount of spatial uncertainty/specificity.
  - Allowing researchers to drill-down into clusters until they are happy with a certain TDP (post-hoc).
  - Setting a desired TDP level and estimating all clusters with at least this TDP (i.e., choosing a spatial specificity level).





BRISBANE

OHBM 2025

JUNE
24–28