Inference

how surprising is your statistic? (thresholding)

But ... can I trust it?

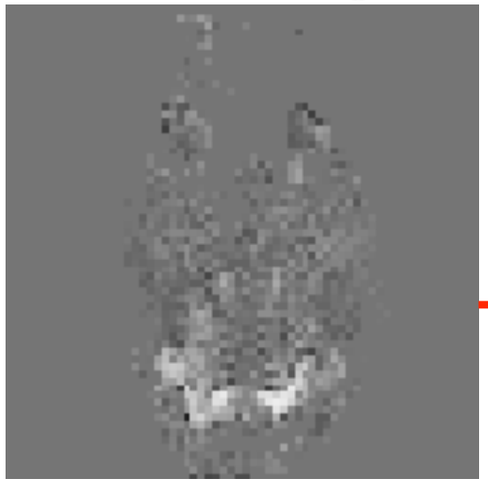
Outline

- Null-hypothesis and Null-distribution
- Multiple comparisons and Family-wise error
- Different ways of being surprised
  - Voxel-wise inference (Maximum z)
  - Cluster-wise inference (Maximum size)
- Parametric vs non-parametric tests
- Enhanced clusters (TFCE)
- FDR - False Discovery Rate

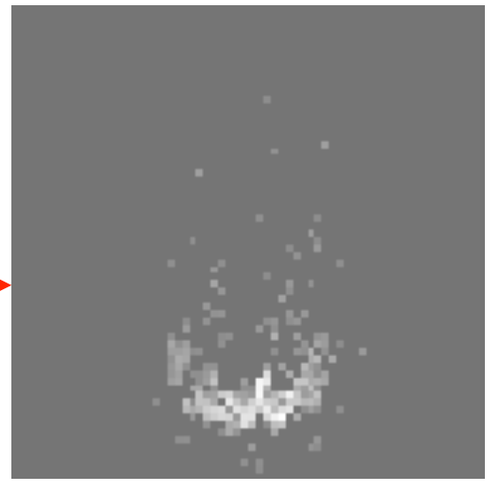
Clustering cookbook

Instead of resel-based correction, we can do clustering:

z stat image



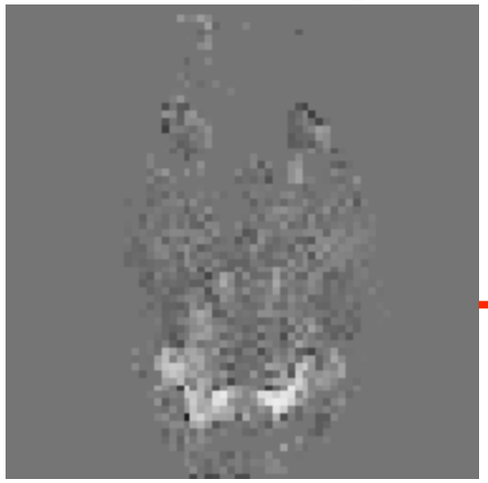
Threshold at (arbitrary) z level



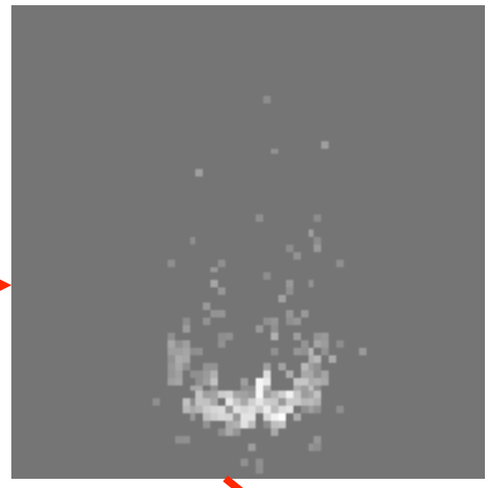
Clustering cookbook

Instead of resel-based correction, we can do clustering

z stat image



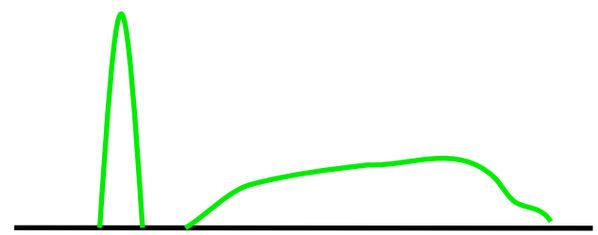
Threshold at
(arbitrary) z level



Form clusters from surviving voxels.
Calculate the size threshold $u(R,z)$.
Any cluster larger than u "survives" and we reject the null-hypothesis for that.

How do we choose the z-threshold?

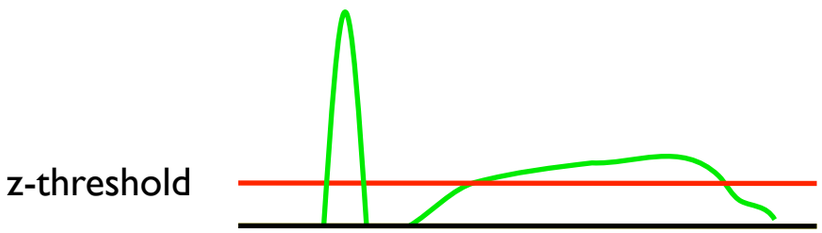
This is arbitrary (within reason) and a trade-off



How do we choose the z-threshold?

This is arbitrary (within reason) and a trade-off

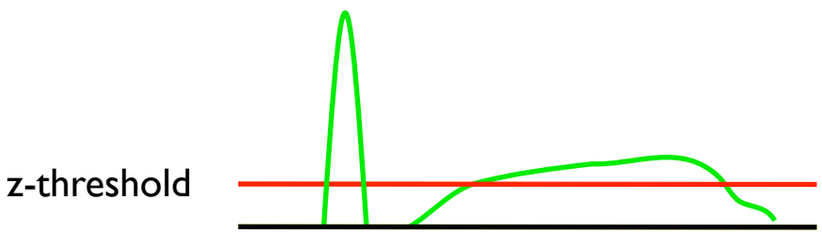
1. **Low threshold** - can violate RFT assumptions, but can detect clusters with large spatial extent and low z



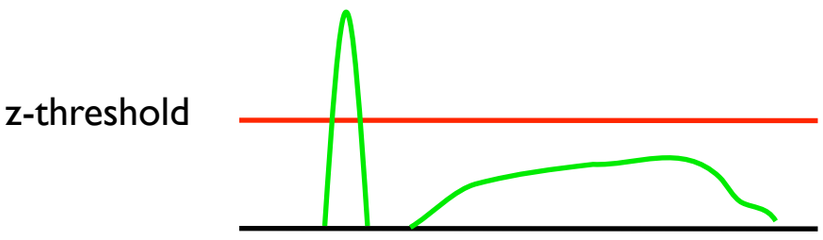
How do we choose the z-threshold?

This is arbitrary (within reason) and a trade-off

1. **Low threshold -** can violate RFT assumptions, but can detect clusters with large spatial extent and low z



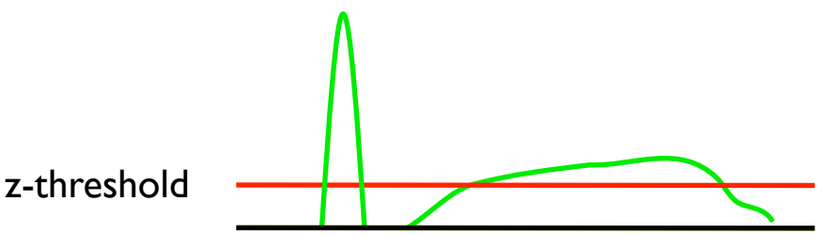
2. **High threshold -** gives more power to clusters with small spatial extent and high z



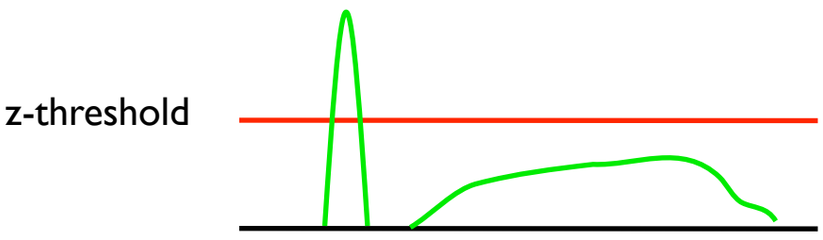
How do we choose the z-threshold?

This is arbitrary (within reason) and a trade-off

1. **Low threshold -** can violate RFT assumptions, but can detect clusters with large spatial extent and low z



2. **High threshold -** gives more power to clusters with small spatial extent and high z



Tends to be more sensitive than voxel-wise corrected testing

Results depend on extent of spatial smoothing in pre-processing

TFCE

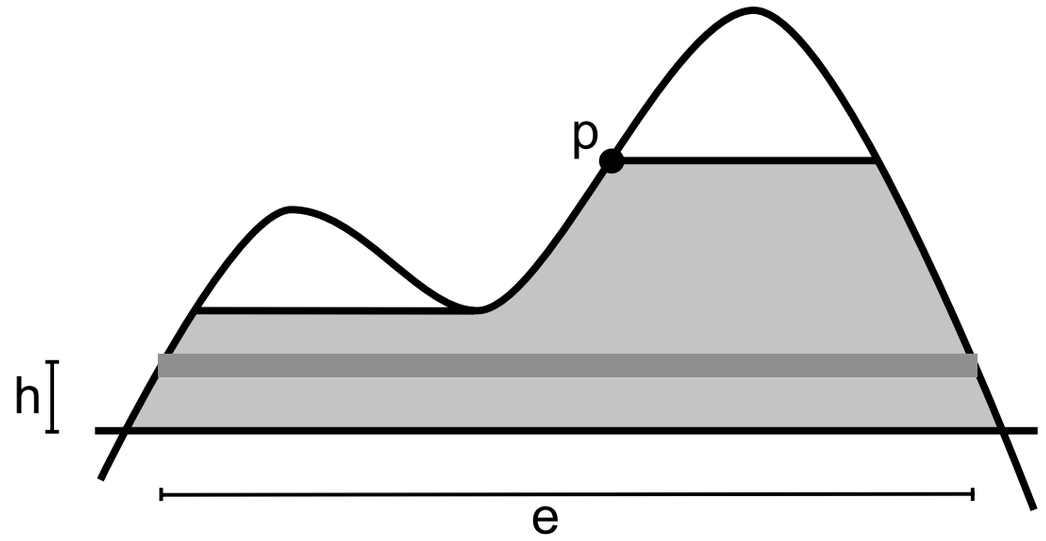
Threshold-Free Cluster Enhancement

[Smith & Nichols, NeuroImage 2009]

- Cluster thresholding:
  - popular because it's sensitive, due to its use of spatial extent
  - but the pre-smoothing extent is arbitrary
  - and so is the cluster-forming threshold

    ➡ unstable and arbitrary

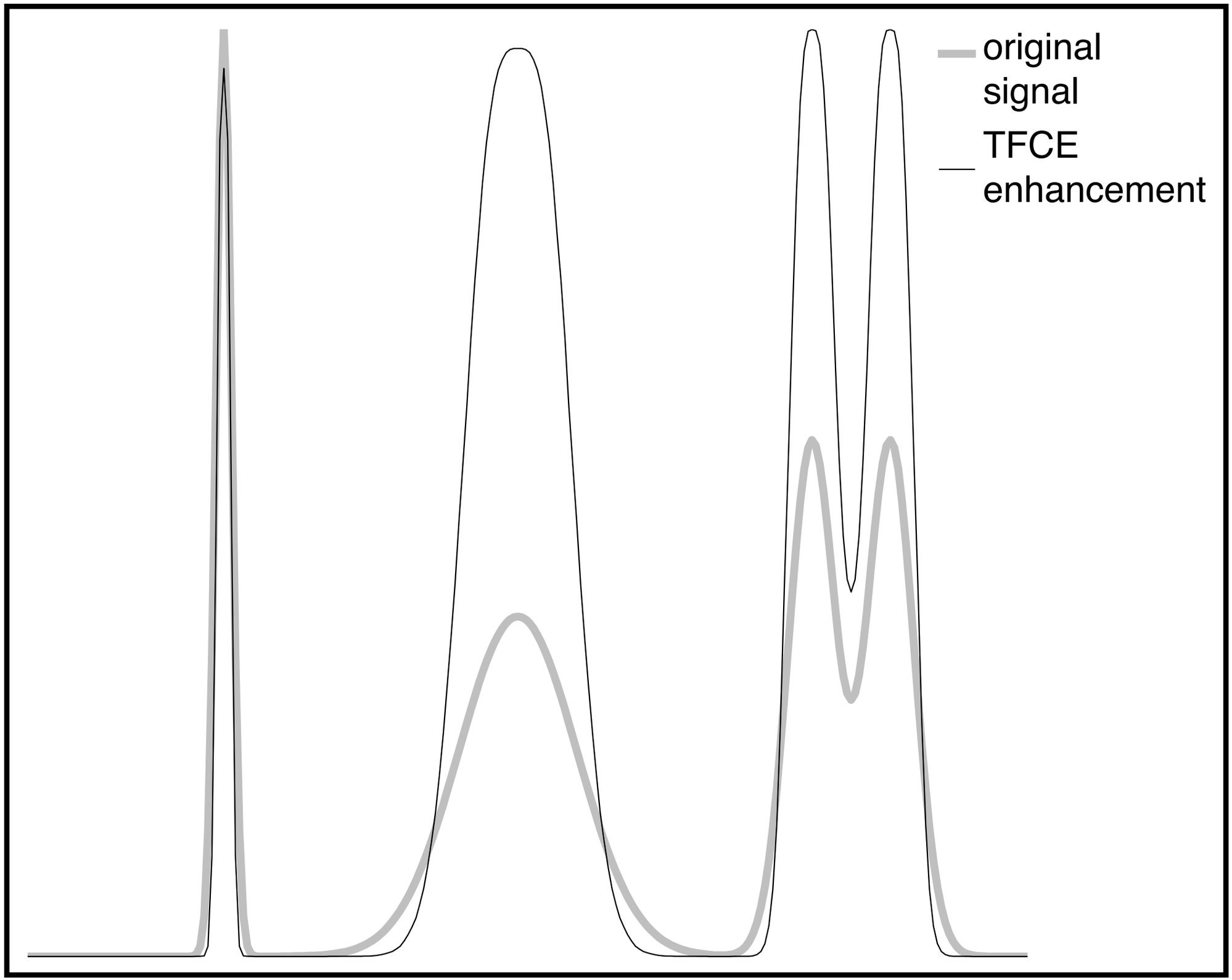
- TFCE
  - integrates cluster "scores" over all possible thresholds
  - output at each voxel is measure of local cluster-like support
  - similar sensitivity to optimal cluster-thresholding, but stable and non-arbitrary



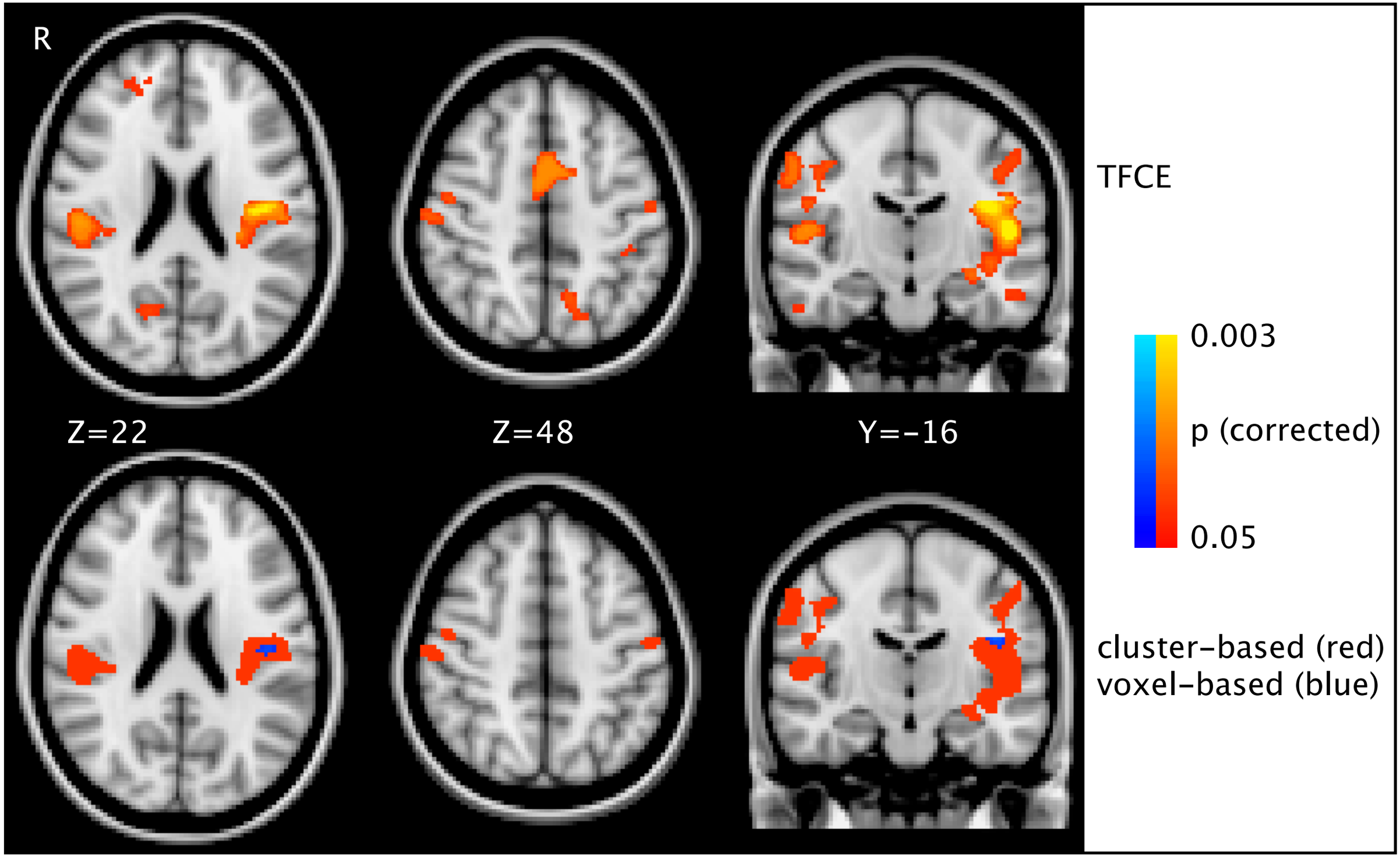
The TFCE value at point p is given by the sum, over the shaded area, of the score from each contributing incremental section:

$$TFCE(p) = \sum_h e(h)^E \cdot h^H$$

Qualitative example



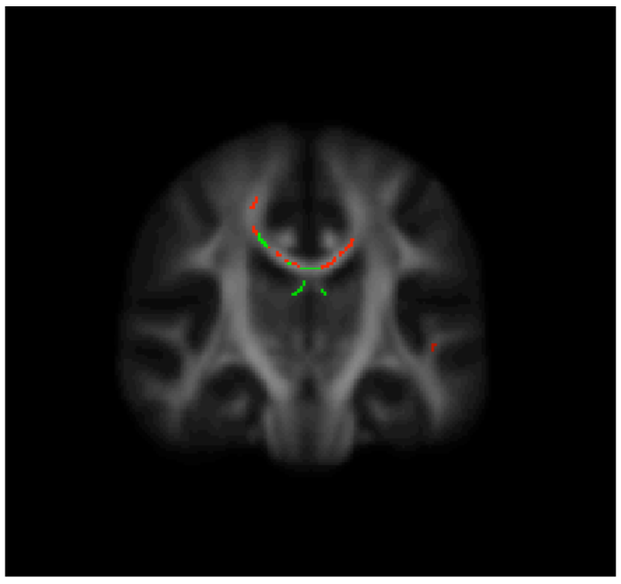
TFCE for FSL-VBM

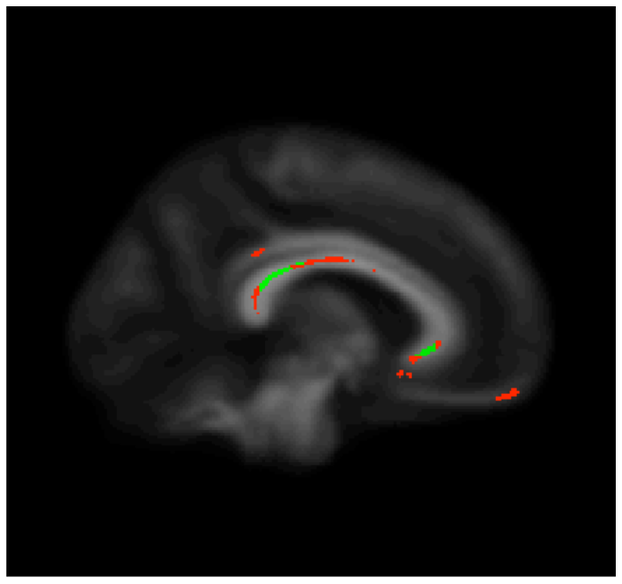


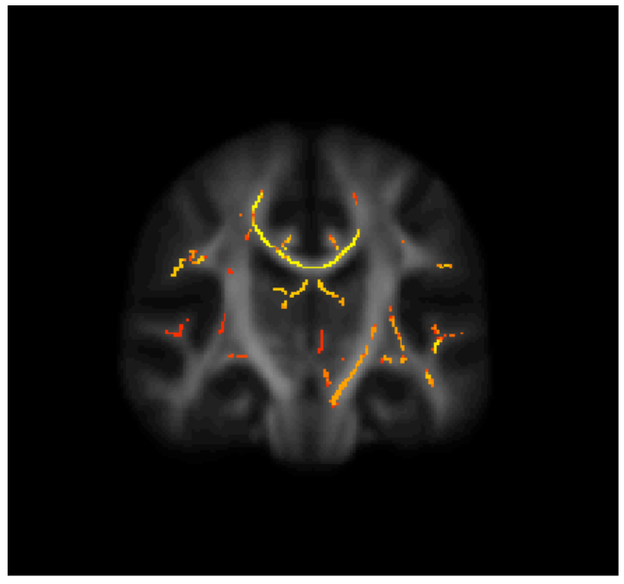
TFCE for TBSS

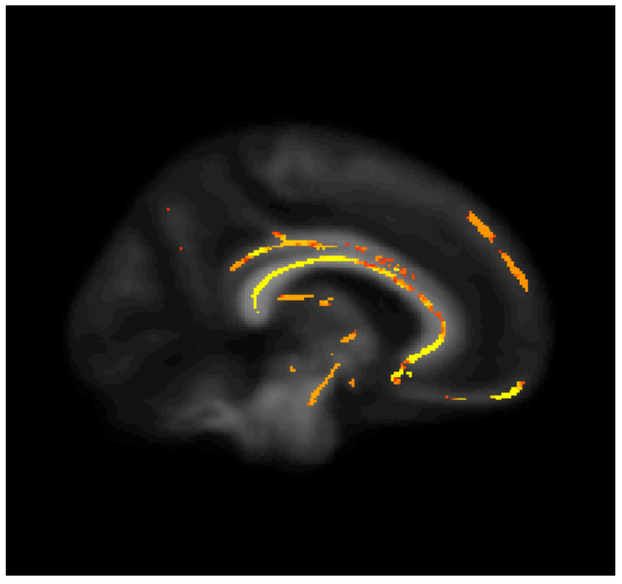
controls > schizophrenics

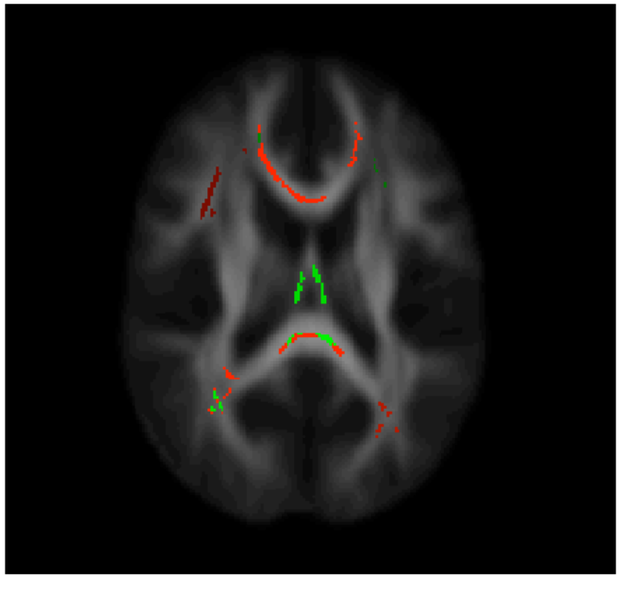
p<0.05 corrected for multiple comparisons across space, using randomise



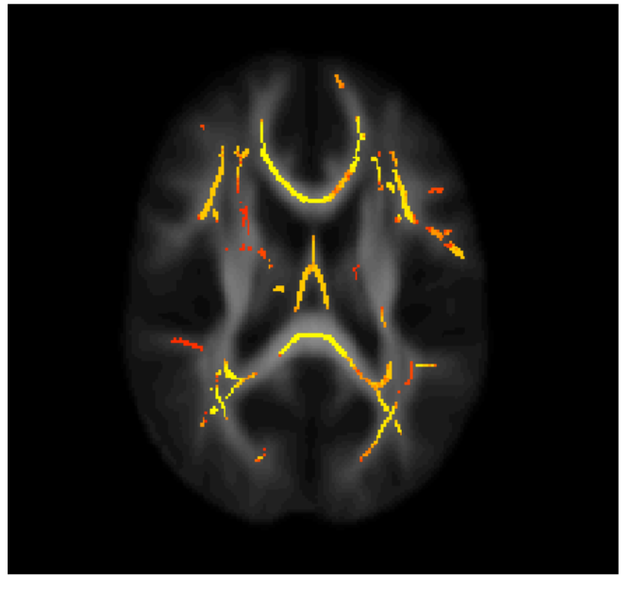








cluster-based:
cluster-forming
threshold =
2 or 3



TFCE

Outline

- Null-hypothesis and Null-distribution
- Multiple comparisons and Family-wise error
- Different ways of being surprised
  - Voxel-wise inference (Maximum z)
  - Cluster-wise inference (Maximum size)
- Parametric vs non-parametric tests
- Enhanced clusters
- FDR - False Discovery Rate

False Discovery Rate

- FDR: False Discovery Rate
  A "new" way to look at inference.

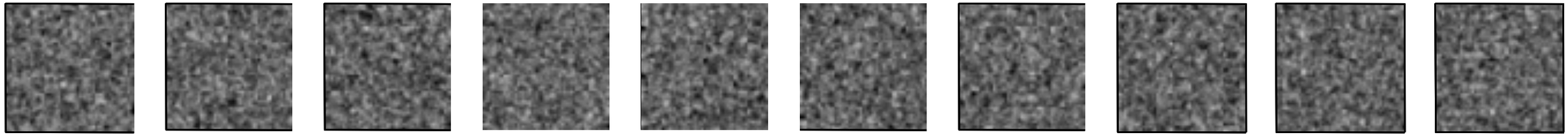
- Uncorrected (for multiple-comparisons):
  - On average, **5% of all voxels** are false positives

- Family-Wise Error (FWE):
  - On average, **5% of all experiments** have one or more false positive voxels

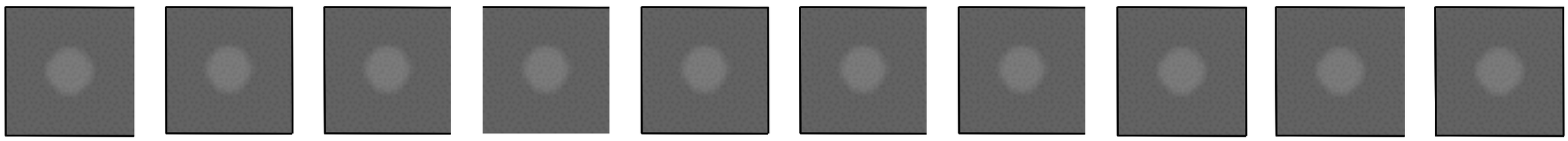
- False Discovery Rate
  - On average, **5% of significant voxels** are false positives

Little imaging demonstration.

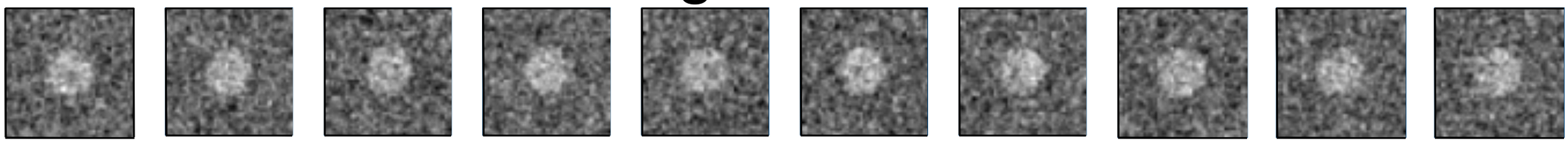
Noise



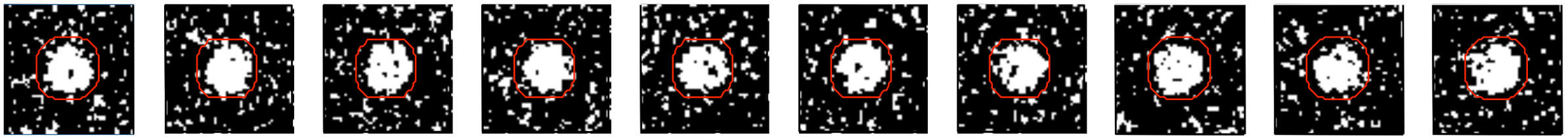
Signal



Signal+Noise

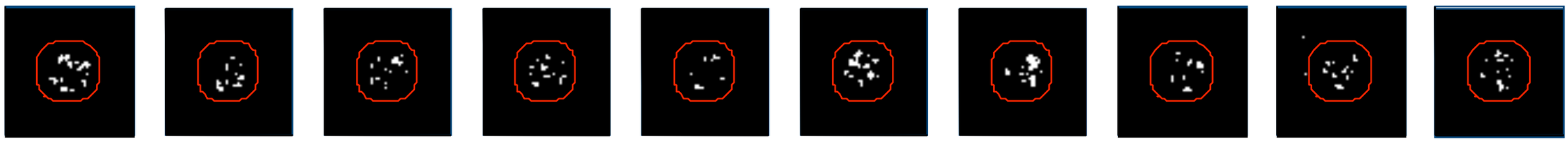


uncorrected voxelwise control of FP rate at 10%



percentage of all null pixels that are False Positives

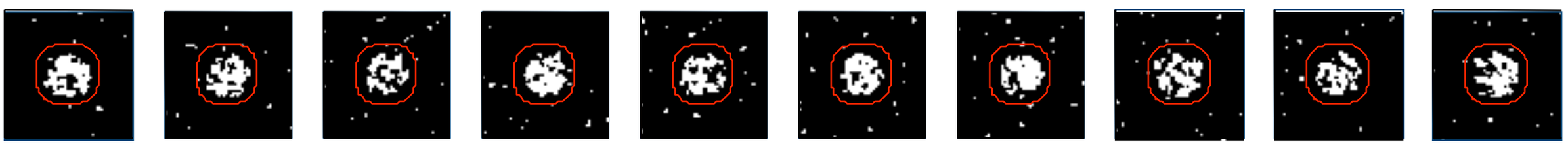
control of FamilyWise Error rate at 10%



occurrence of FamilyWise Error

FWE

control of False Discovery Rate at 10%



percentage of activated (reported) pixels that are False Positives

FDR for dummies

- Makes assumptions about how errors are distributed (like GRT).
- Used to calculate a threshold.
- Threshold such that X% of super-threshold (reported) voxels are false positives.
- Threshold depends on the data. May for example be very different for [1 0] and [0 1] in the same study.