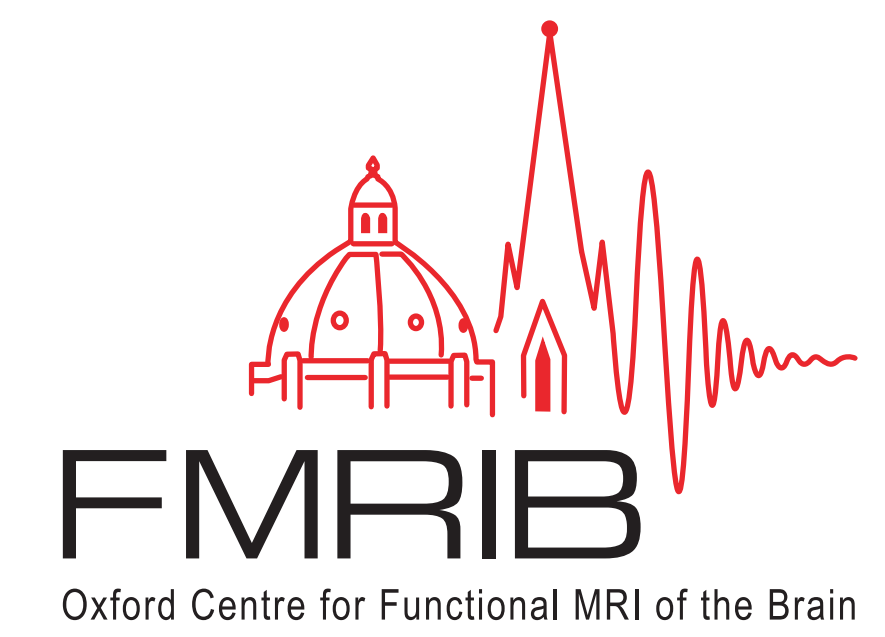


# Multi-level block permutation

Anderson M. Winkler<sup>1</sup>, Diego Vidaurre<sup>2</sup>, Matthew A. Webster<sup>1</sup>, Mark W. Woolrich<sup>2</sup>, Thomas E. Nichols<sup>1,3</sup>, Stephen M. Smith<sup>1</sup>

1. Oxford Centre for Functional MRI of the Brain (FMRIB), University of Oxford, UK. 2. Oxford Centre for Human Brain Activity (OHBA), University of Oxford, UK. 3. Department of Statistics & Warwick Manufacturing Group, University of Warwick, Coventry, UK.



## 1 Introduction

Under weak and reasonable assumptions, mainly that the data are exchangeable under the null hypothesis, permutation tests can provide exact control of false positives and allow the use of various non-standard statistics. There are, however, various common examples where simple exchangeability can be violated, including paired tests, tests that involve repeated measurements, when subjects are relatives (members of pedigrees), and any dataset with known dependence among subjects or observations. In these cases, certain permutations would create realisations that would be incompatible with the original data even under the null hypothesis, and thus, that cannot be used to construct the reference distribution.

To allow permutation inference in such cases, we propose to test the null hypothesis using only a subset of all otherwise possible permutations, i.e., using only the rearrangements of the data that respect the exchangeability, thus retaining the original joint distribution unaltered. Instead of defining exchangeability at the level of each datum, we assert exchangeability for blocks of data, either within block (observations are shuffled inside each block only) or between blocks (without permuting the observations inside, but the blocks as whole). Importantly, we also allow nested blocks to be defined, in a hierarchical, multi-level fashion. Our proposal does not require modelling explicitly the degree of dependence between observations; this dependence is implicitly accounted for by the permutation scheme.

## 2 Method

To evaluate our method, we simulated two simple datasets that reproduce the structured dependence in the data of the Human Connectome Project (HCP). These two sets consisted of:

**Dataset A:** Three sets of three siblings, each comprising a pair of monozygotic twins and a non-twin.

**Dataset B:** Three sets of three siblings, each with a different structure: one with a pair of monozygotic twins and a non-twin, one with a pair of dizygotic twins and a non-twin, and one with three non-twin siblings.

The structure in each of these datasets and the possibilities for rearrangements can be represented graphically as dependence trees, as shown in Figure 1. The complexity of the trees contrast with the case where shuffling is allowed freely between all observations, as shown in Figure 2.

Figure 1: The dependence structure between observations (here, simulated sib-pairs) can be represented as a table, in which each column indicates a level, or as a tree.

Dataset A				Dataset B			
1	-1	1	1	-1	-1	1	1
1	-1	1	2	-1	-1	1	2
1	-1	2	3	-1	-1	2	3
1	-2	1	4	-1	-2	1	4
1	-2	1	5	-1	-2	1	5
1	-2	2	6	-1	-2	2	6
1	-3	1	7	-1	3	1	7
1	-3	1	8	-1	3	2	8
1	-3	2	9	-1	3	3	9

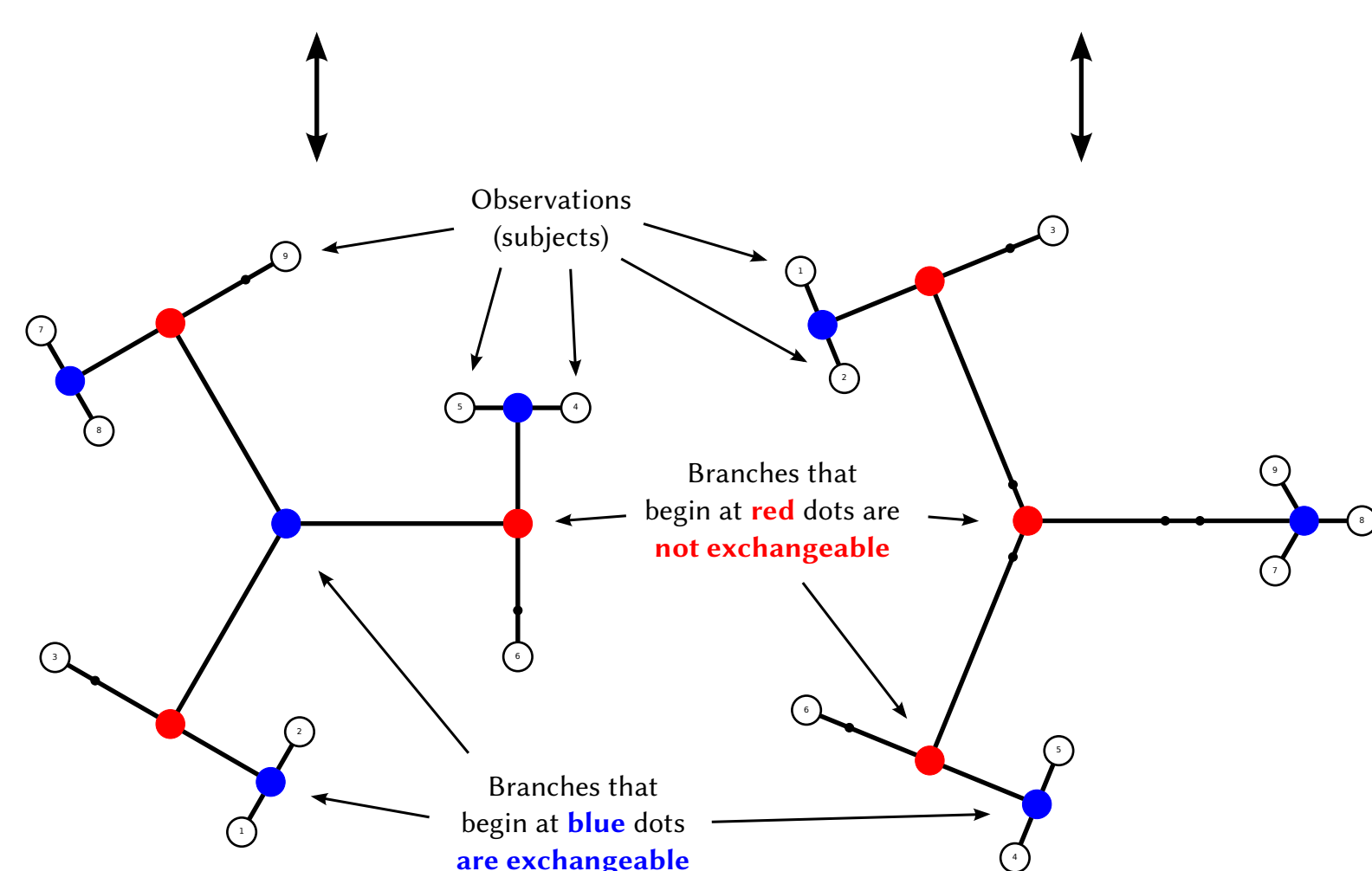
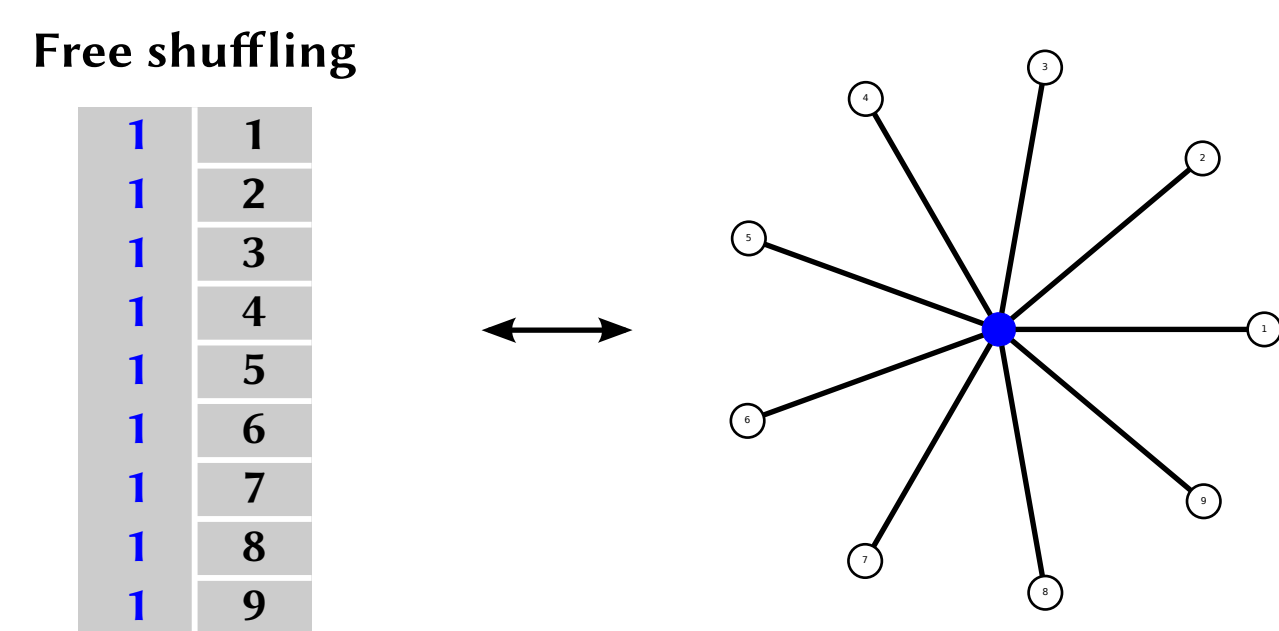
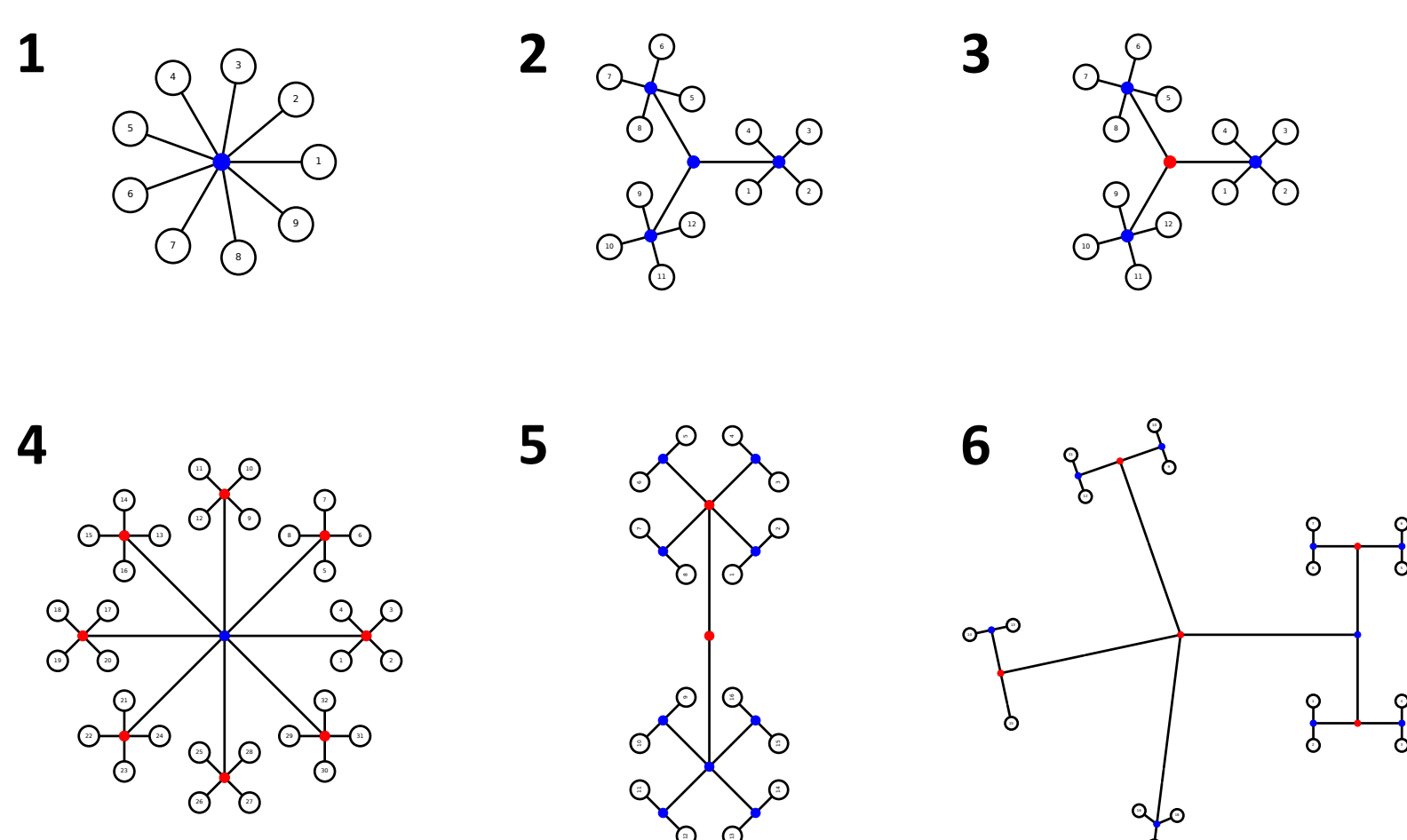


Figure 2: If the data is freely exchangeable, the structure is much simpler.



In addition to the two family configurations above, used to examine error type I rates and power, we also constructed various other cases of dependency to assess in more detail the influence of the permutation scheme over power. These are shown in Figure 3.

Figure 3: The various dependence structures used to examine power.



## 3 Results

If each observation is a subject, and if the dependency reflects kinship, then this dependence is equivalent to the heritability, i.e., the fraction of the observed variance that is attributable to kinship. The simulations shown here used varied degrees of heritability ( $h^2 = 0, 0.4$  and  $0.8$ ), as well as the introduction of signal.

Figure 4 shows the results. For both datasets (A and B), shuffling that ignores the structure between observations caused the error rate not to be controlled as the dependence became stronger. Shuffling respecting the tree-like structure between the observations controlled the error rate at the nominal level (here, 0.05).

Although power cannot be considered when the error rate is not controlled, it is clear that, even in the absence of true dependence, shuffling within block is less powerful than shuffling freely. This reduction in power stems from the reduced amount of perturbations caused on the data by the shuffling process. This "amount of perturbations" can be measured by the average Hamming distance, i.e., the average number of observations that change their position at each permutation. When the variation in power is contrasted with the average Hamming distance, as shown in Figure 5, it becomes clear that the less perturbation, less power.

Figure 5: Relationship between power gain (compared to free shuffling when there is no heritability) and the amount of perturbation (average Hamming distance, normalised to unity using the theoretical upper limit as the reference) of the observations at each shuffling. The plots 1-6 are the structures shown in Figure 3; the plots 7-8 in Figure 6.

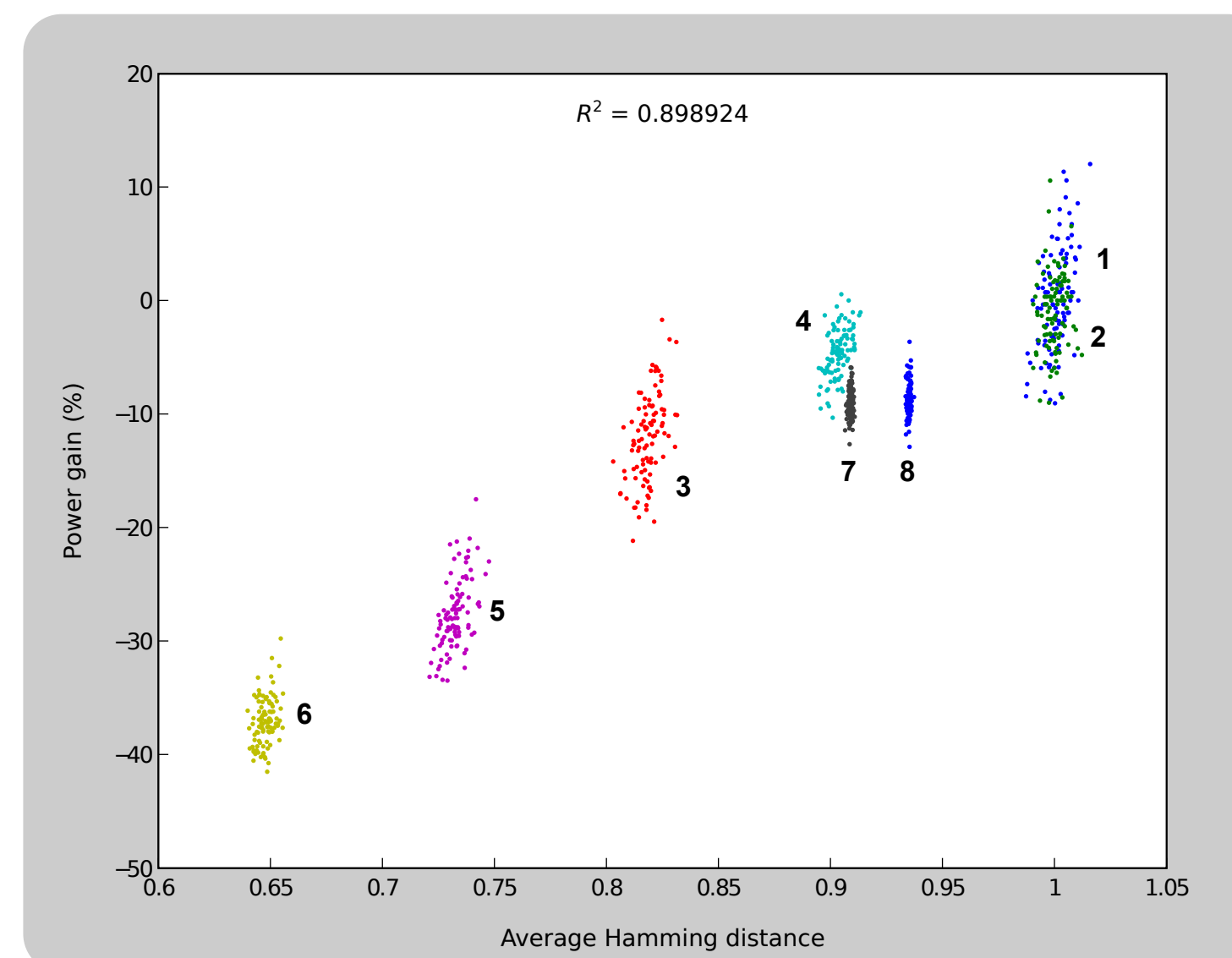
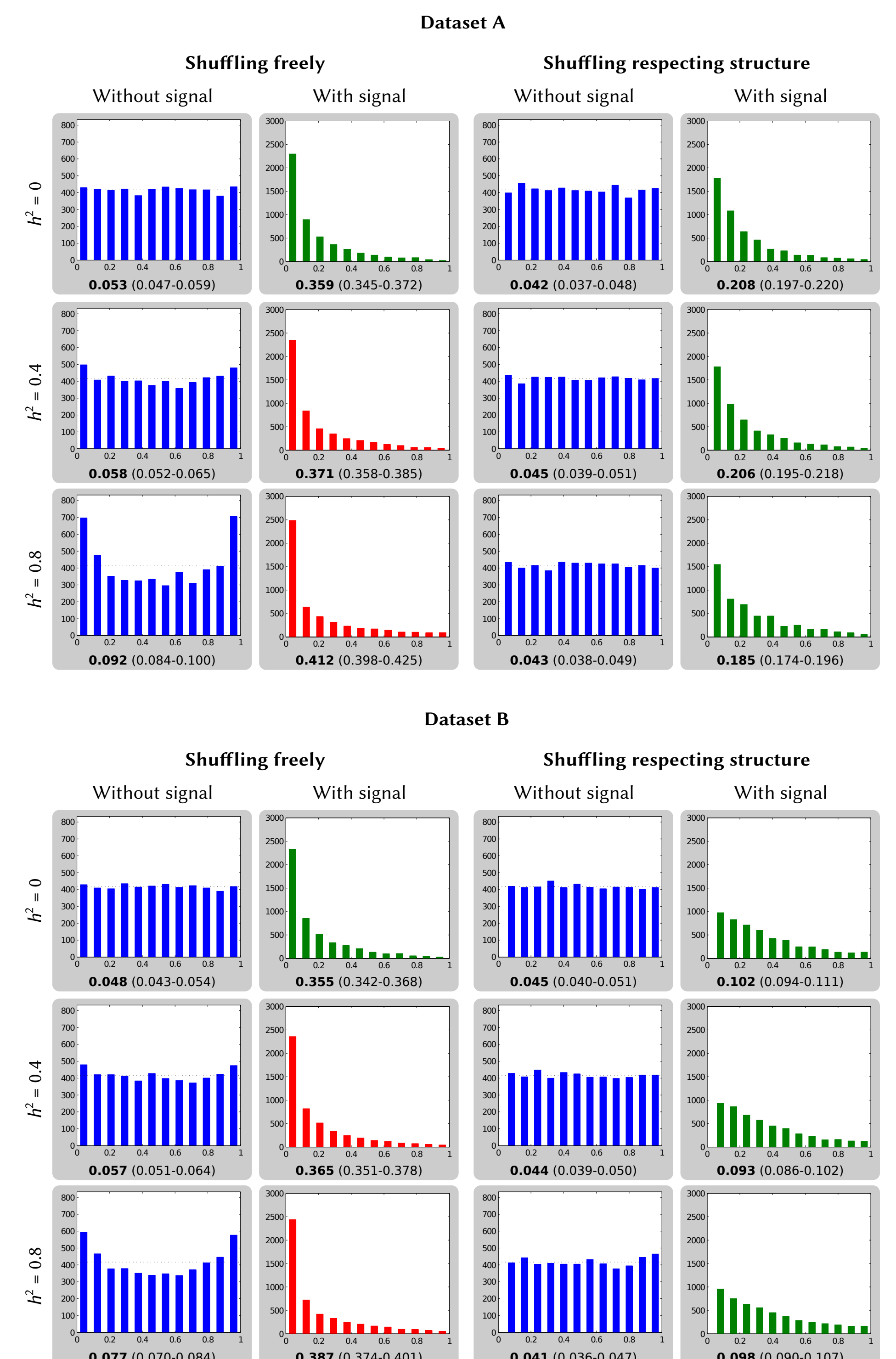


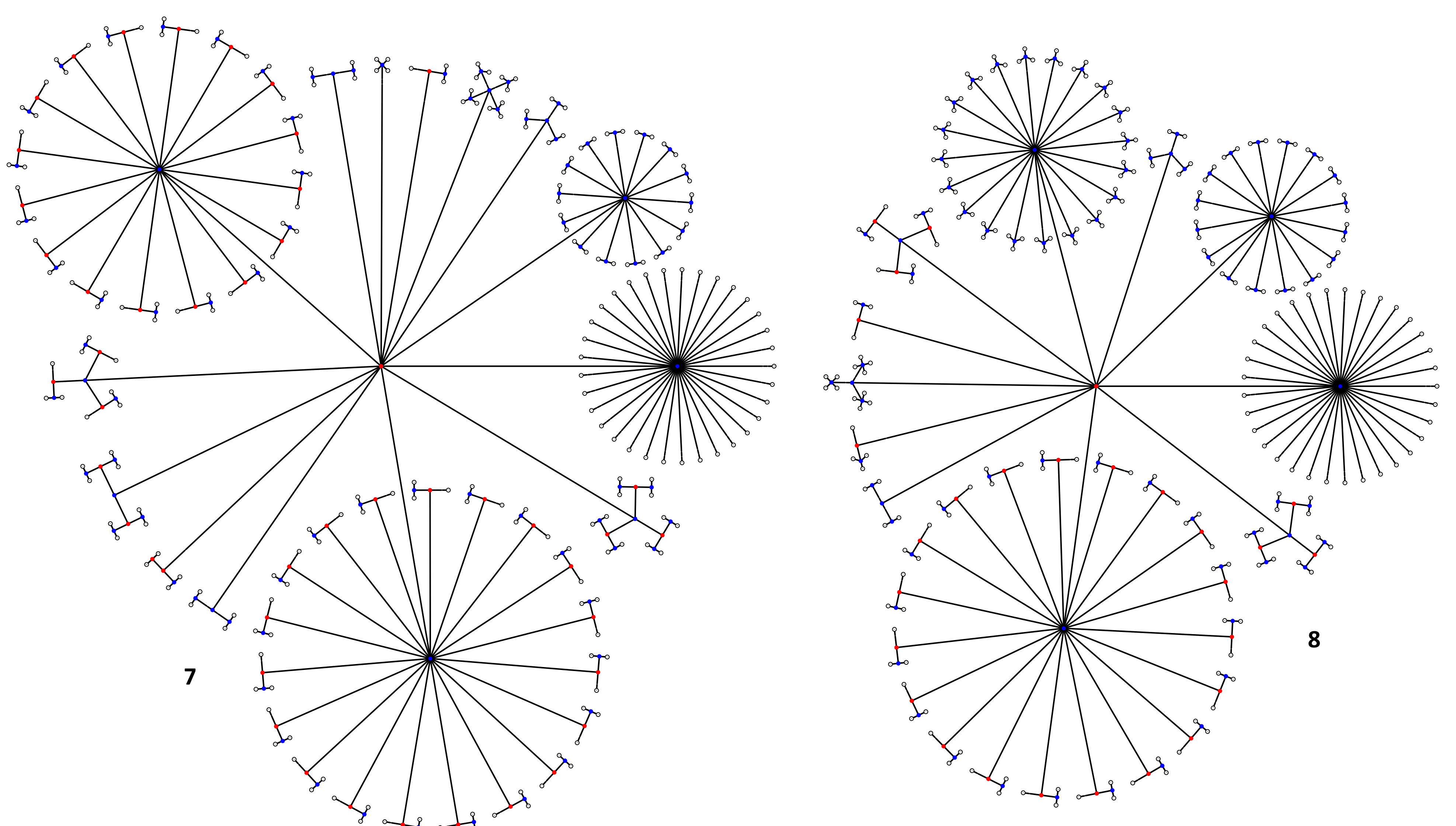
Figure 4: Error type I (in blue) and power (red or green) when shuffling respecting or not the data structure. Red bars are for power when the error rate is not controlled (flagged by red bars), thus when the test is not valid. The proportion of discoveries is in bold, and the 95% confidence intervals between parenthesis.



## 4 Data structure in the Human Connectome Project

The HCP recruits subjects together with their siblings, with emphasis on twins (monozygotic and dizygotic). As before, this structure can be represented as a tree that indicates which pieces of data can be shuffled for inference, and the permutation methods described this far can therefore be applied. Depending on whether there is interest in considering or not common effects in DZ twins, these can be treated as a category on their own (Figure 6 (left)), or be allowed to be exchanged with ordinary siblings (right).

Figure 6: Dependence structure of the HCP data. On the left, DZ twins are treated as a category on its own; on the right, DZ twins are treated as ordinary siblings (non twins).



## 5 Conclusion

Multi-level block permutation effectively controls the rate of type I errors, even in the presence of strong dependence between observations, and can be used as a general inference tool when the dependence structure can be organised in blocks.

There is an unavoidable loss of power due to insufficient shuffling, although in large datasets, with relatively complex dependence structure as the HCP, this loss might not be substantial.

## 6 References

- Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp.* 2002 Jan;15(1):1-25.
- Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, WU-Minn HCP Consortium. The WU-Minn Human Connectome Project: an overview. *Neuroimage.* 2013 Oct 15;80:62-79.
- Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. *Neuroimage.* 2014 May 15;92:381-97.