Model-based clustering using generative embedding

Kay H Brodersen[1,2,3] · Zhihao Lin[2] · Lorenz Deserno[4,5] · Ajita Gupta[2] · Will D Penny[6] · Alexander P Leff[6]
Morteza H Chehreghani[2] · Alberto-Giovanni Busetto[2,7] · Florian Schlagenhauf[4,5] · Joachim M Buhmann[2] · Klaas E Stephan[1,3,6]
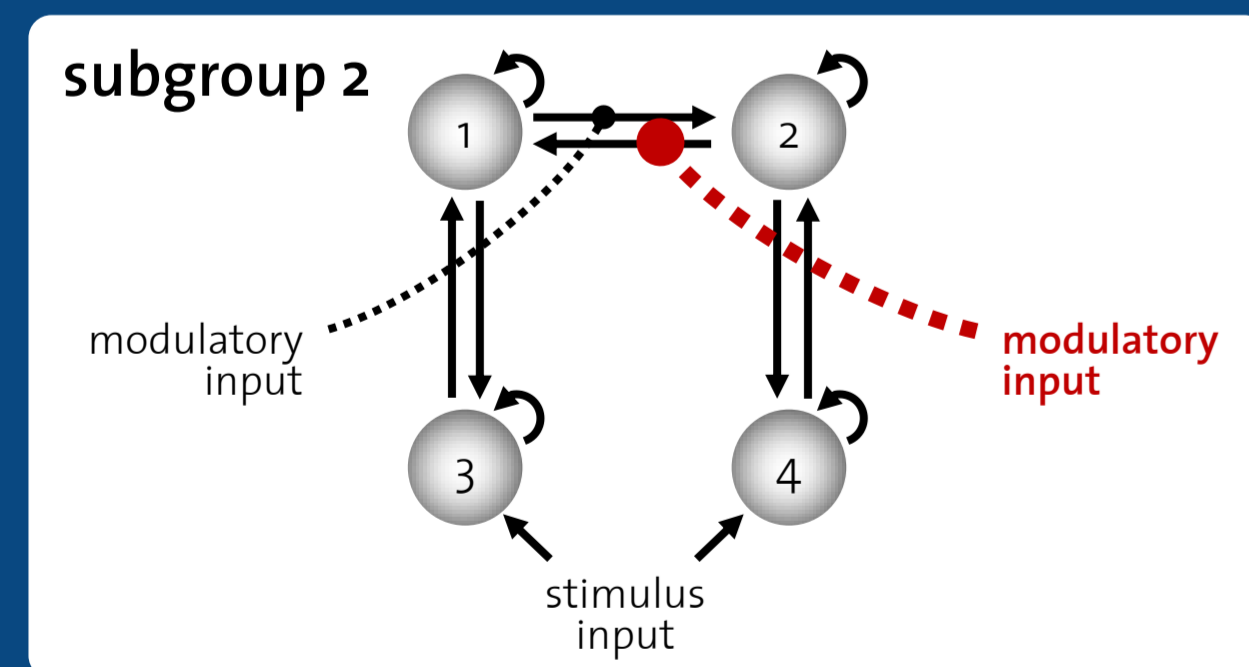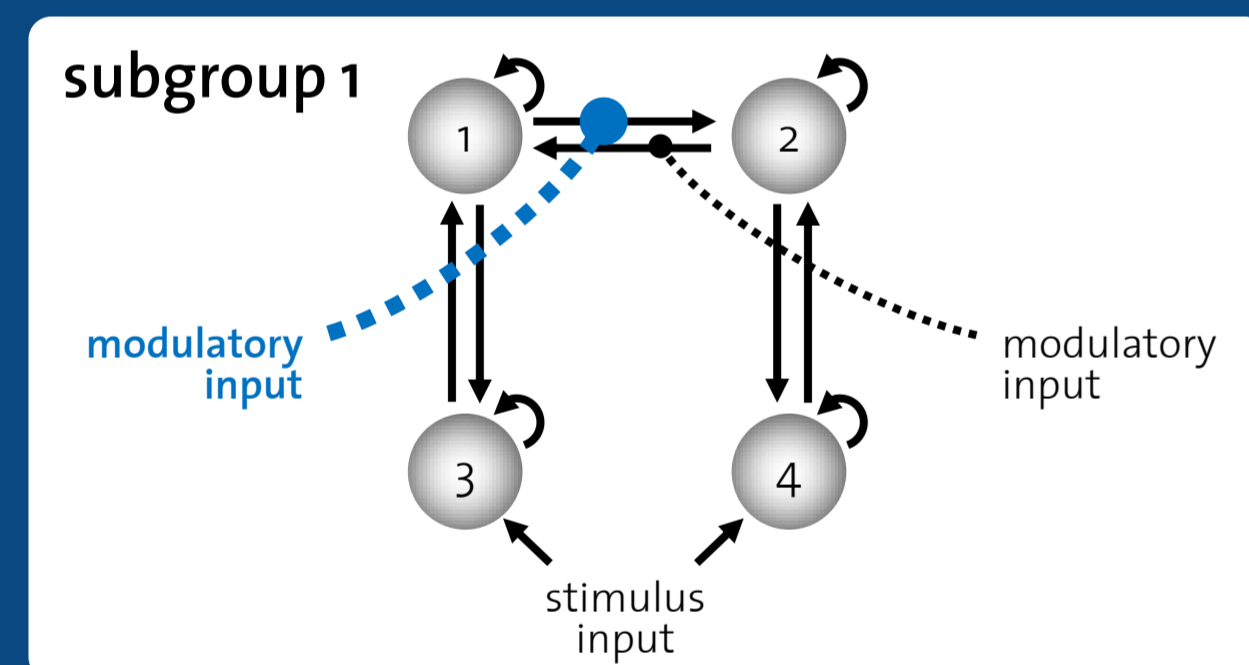
[1] Translational Neuromodeling Unit (TNU), University of Zurich & ETH Zurich, Switzerland [2] Department of Computer Science, ETH Zurich, Switzerland [3] Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, Switzerland [4] Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin Berlin, Germany
[5] Max Planck Institute for Cognitive and Brain Sciences, Leipzig, Germany [6] Wellcome Trust Centre for Neuroimaging, University College London, United Kingdom [7] Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland

1 Introduction

- An important problem in psychiatry is the lack of diagnostic classifications that are based on pathophysiological mechanisms rather than symptoms.

- It is conceivable that one could solve this problem by constructing generative models of brain function that enable inference on the computational and neuronal processes that underlie an observed collection of symptoms.

- We recently showed that generative embedding based on such models can yield highly accurate predictions of a symptom-based diagnostic state from fMRI data [1,2].

- In this study, we are beginning to address the open question of whether generative embedding might allow us to *discover* clinically relevant conditions when such conditions are not known *a priori*.
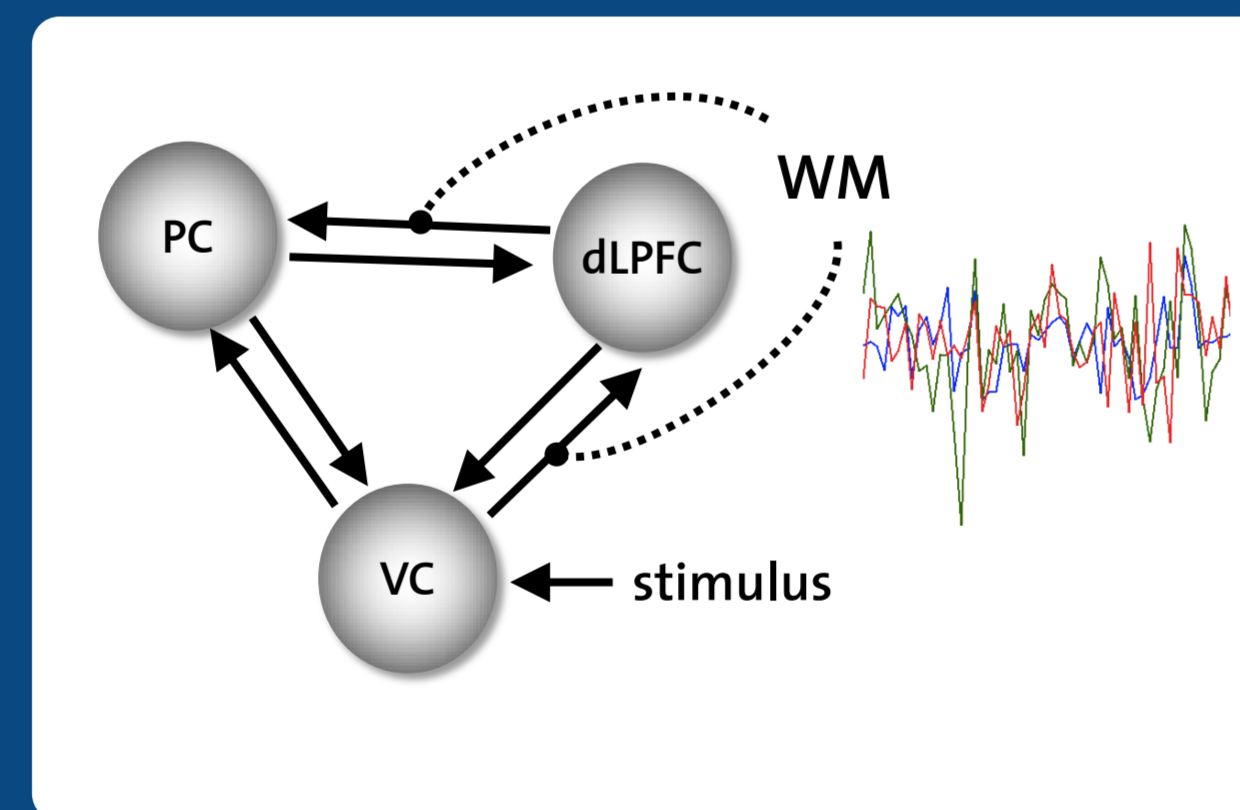
2 Datasets

Synthetic fMRI data (*n* = 80)

subgroup 1
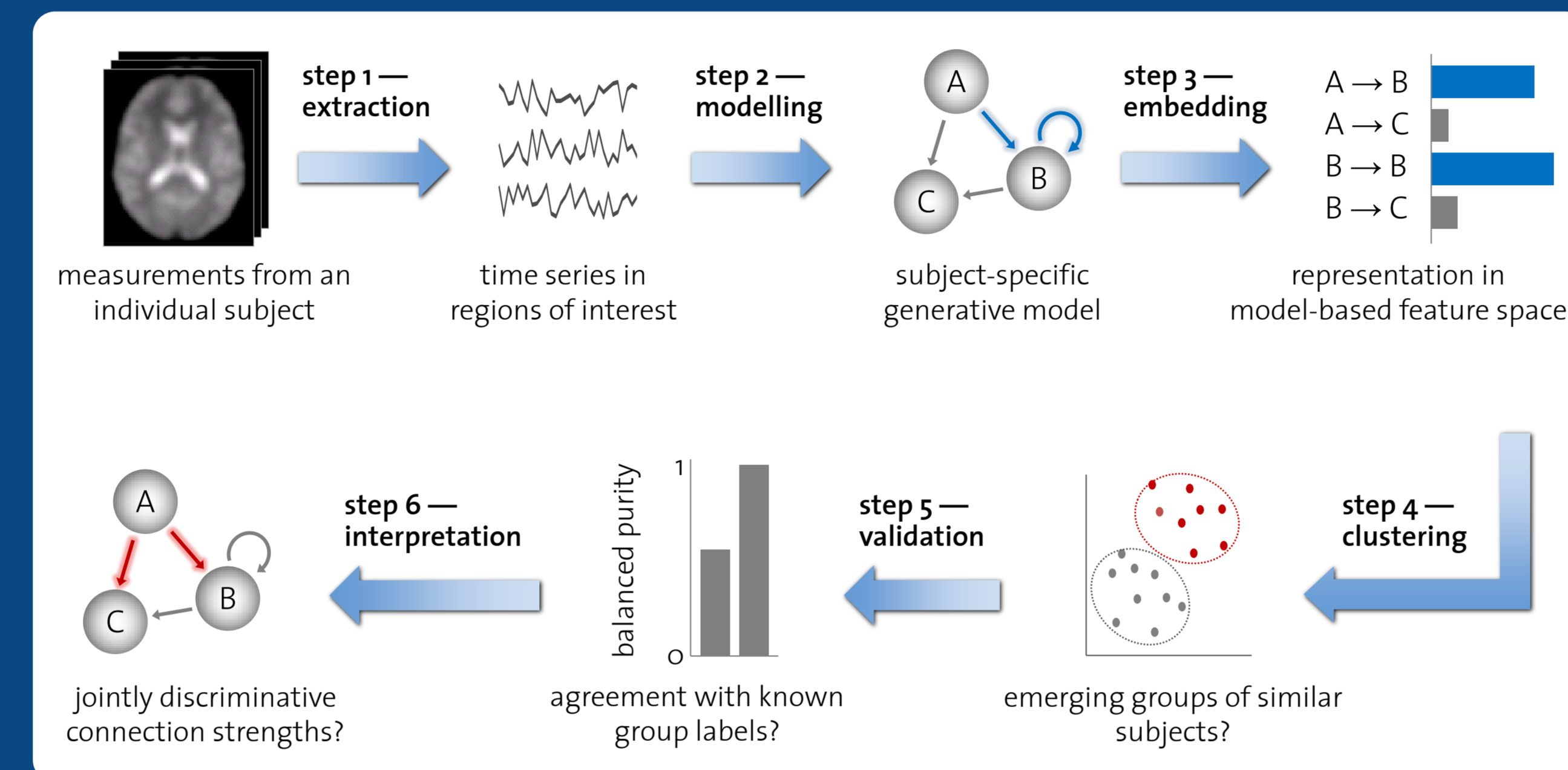
subgroup 2

Empirical fMRI data (*n* = 83)

To assess the empirical validity of our approach, we analysed fMRI data from schizophrenia patients and healthy controls engaged in a working-memory task [4]. Using a DCM of prefrontal and parietal activity, we asked whether we could discover the diagnostic category 'schizophrenia' from patterns of connectivity.

To investigate the theoretical properties of our approach, we generated fMRI data for two synthetic subject groups using a simple four-region DCM [3], as shown above. For each subject, the true DCM parameters were drawn from a Gaussian with a group-specific mean. The two subgroups differed in terms of modulatory effects on their intrinsic connectivity (μ_B). We then generated fMRI data from these DCMs, estimated the model parameters, and submitted these estimates to clustering.

References
1. Brodersen, K.H., Haiss, F., Ong, C.S., Jung, F., Tittgemeyer, M., Buhmann, J.M., Weber, B., Stephan, K.E. (2011). Model-based feature construction for multivariate decoding. *NeuroImage*, 56, 601-615.
2. Brodersen, K.H., Schofield, T.M., Leff, A.P., Ong, C.S., Lomakina, E.I., Buhmann, J.M., Stephan, K.E. (2011). Generative embedding for model-based classification of fMRI data. *PLoS Comput Biol*, 7(6): e1002079.
3. Friston, K.J., Harrison, L., & Penny, W., 2003. Dynamic causal modelling. *NeuroImage*, 19(4), 1273-1302.
4. Deserno, L., Sterzer, P., Wüstenberg, T., Heinz, A., & Schlagenhauf, F. (2012). Reduced prefrontal-parietal effective connectivity and working memory deficits in schizophrenia. *Journal of Neuroscience*, 31(1), 12-20.
5. Penny, W. (in preparation). Variational Bayes for d-dimensional Gaussian mixture models.
6. Stephan, K. E., Friston, K. J., & Frith, C. D. (2009). Dysconnection in schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia Bulletin*, 35(3), 509-527.
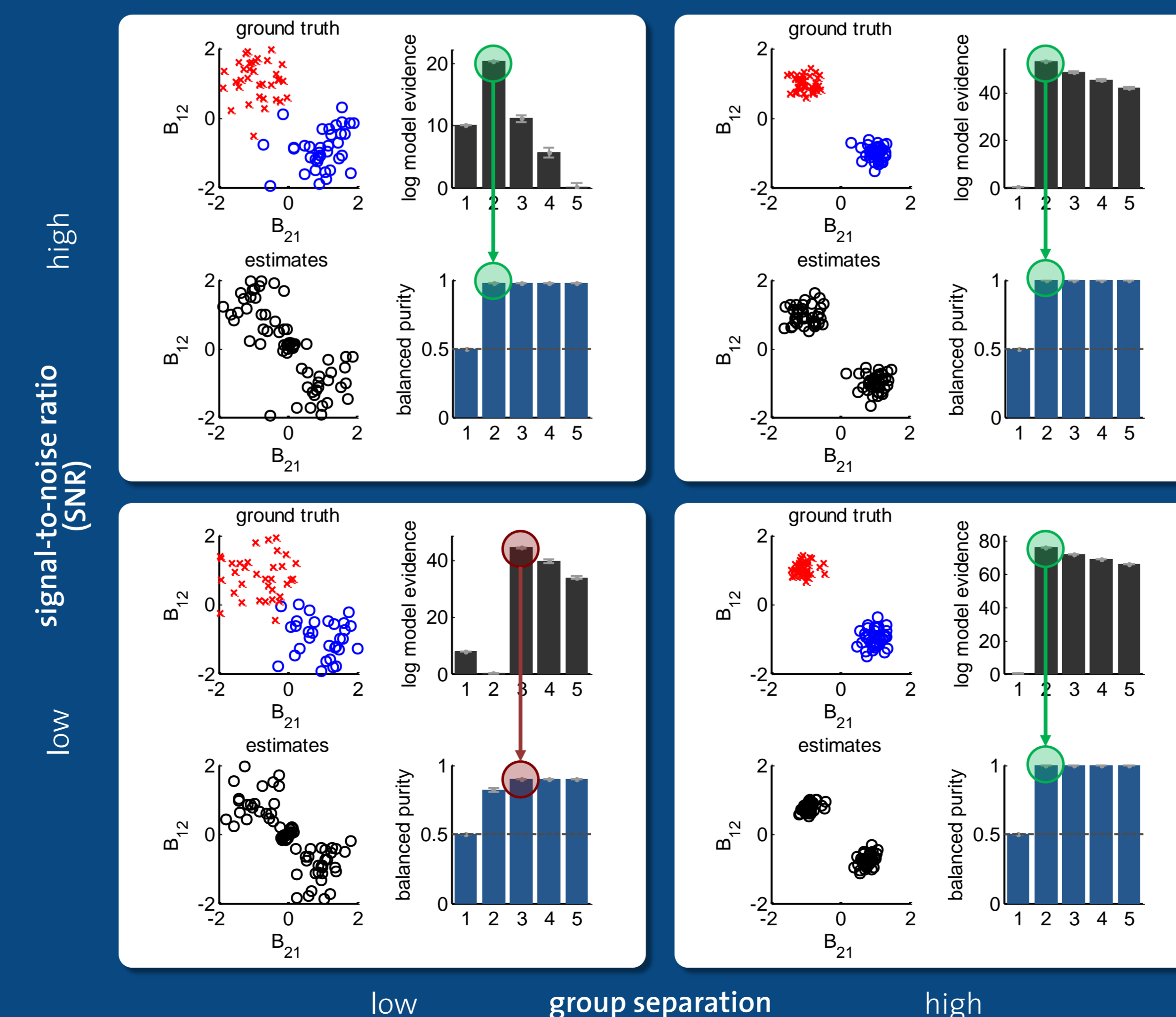
3 Model-based clustering

We introduce generative embedding for model-based clustering using a combination of dynamic causal models (DCM) and variational Gaussian Mixture Models (GMM) clustering [5].

step 1 — extraction — measurements from an individual subject

step 2 — modelling — time series in regions of interest

step 3 — embedding — subject-specific generative model — representation in model-based feature space

step 4 — clustering — emerging groups of similar subjects?

step 5 — validation — agreement with known group labels?

step 6 — interpretation — jointly discriminative connection strengths?
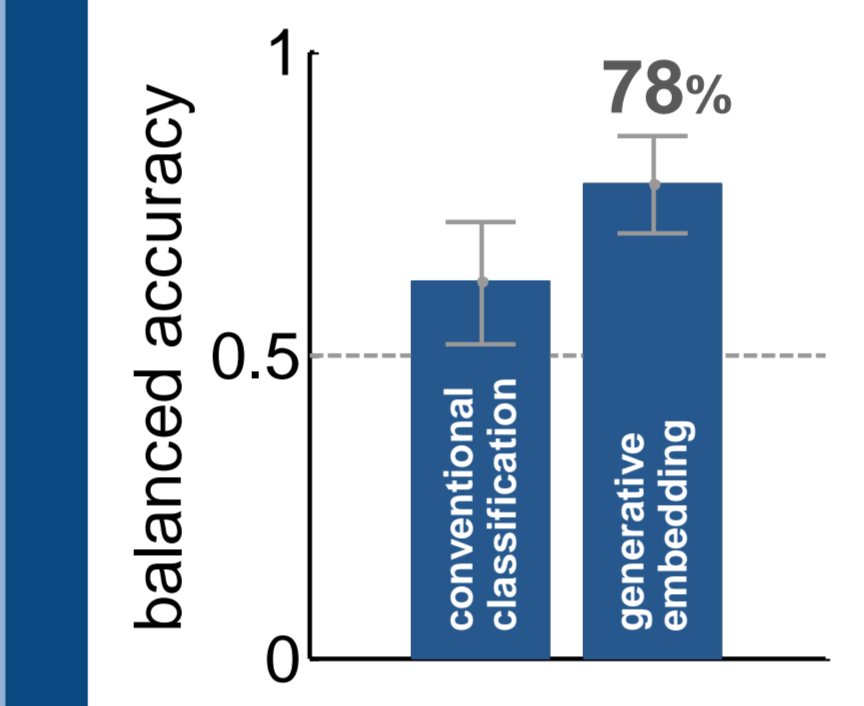
4 Results on synthetic fMRI data

Our analysis discovered the correct number of clusters (two) when the groups were well separated or there was a sufficiently high signal-to-noise ratio.

signal-to-noise ratio (SNR) — high / low

group separation — low / high

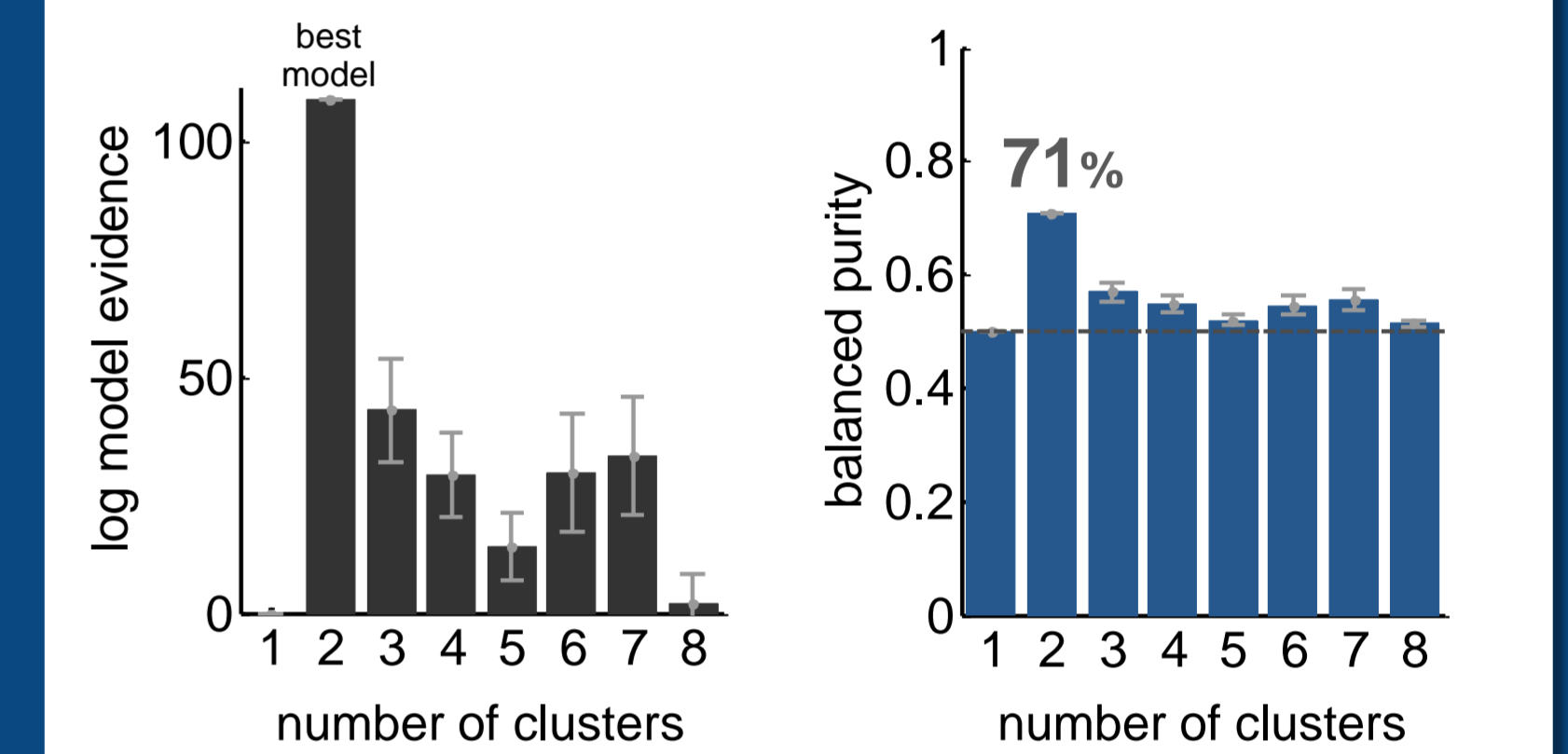ground truth · estimates · log model evidence · balanced purity

5 Results on empirical fMRI data

Using a linear support vector machine (SVM), we were able to predict a subject's diagnostic status with an accuracy of 78% (left). We then adopted an unsupervised exploratory approach: generative embedding inferred that the data comprised two subgroups. These subgroups showed a 71% correspondence with schizophrenic patiens and healthy controls (right).

Supervised setting: support vector classification — balanced accuracy — conventional classification / generative embedding — **78%**

New unsupervised setting: GMM clustering — log model evidence — best model — number of clusters — balanced purity — **71%** — number of clusters

Model-based solutions can be interpreted in terms of the underlying generative model. In the model underlying cluster 1, which contained almost exclusively healthy controls, working memory had a significantly stronger modulatory effect than in cluster 2, which was mostly composed of patients.

Clustering solution — cluster 1 — cluster 2 — + SZ patient — ○ healthy control
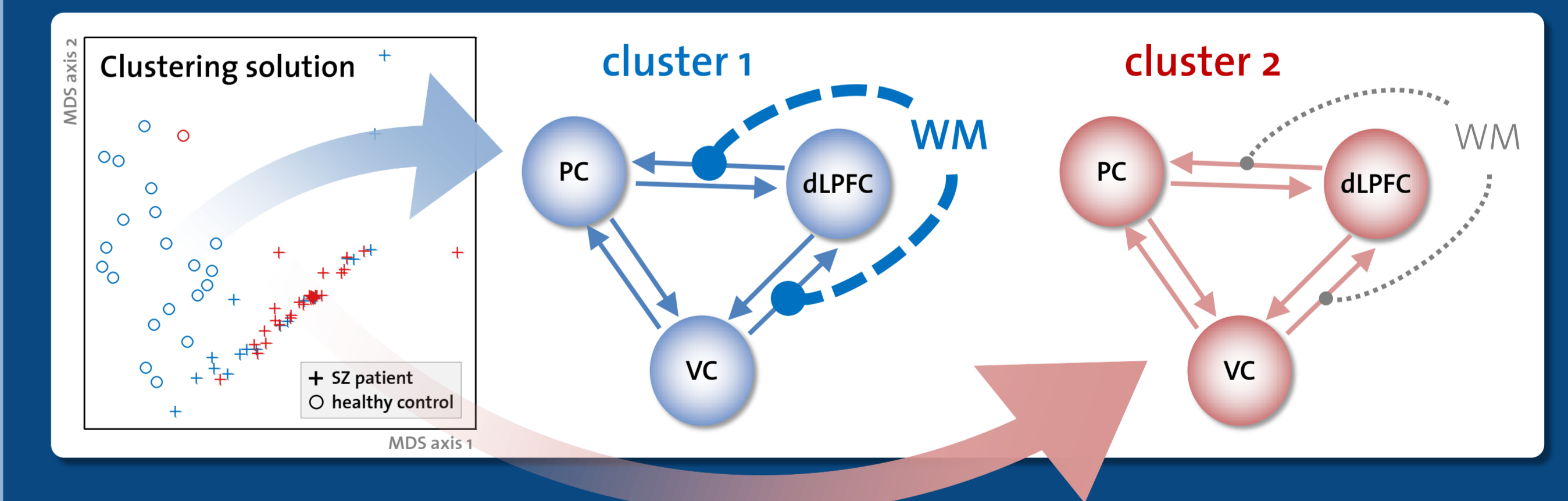
6 Conclusions

- Clustering using generative embedding may enable us to decompose groups of patients with similar symptoms into pathophysiologically distinct subtypes.

- In contrast to conventional activation-based, correlation-based, or symptom-based clustering schemes, our approach exploits discriminative information encoded in 'hidden' physiological quantities such as synaptic connection strengths.

- Critically, generative embedding enables a mechanistic interpretation of the discovered structures.