



## Mapping effective connectivity by virtually perturbing a surrogate brain

Zixiang Luo<sup>1</sup>, Kaining Peng<sup>1</sup>, Zhichao Liang<sup>1</sup>, Shengyuan Cai<sup>1</sup>, Chenyu Xu<sup>2</sup>, Dan Li<sup>1</sup>,  
Yu Hu<sup>3</sup>, Changsong Zhou<sup>4</sup> and Quanying Liu<sup>1\*</sup><sup>1</sup>Southern University of Science and Technology, <sup>2</sup>Iowa State University, <sup>3</sup>The Hong Kong University of Science and Technology, <sup>4</sup>Hong Kong Baptist University

## Introduction

Effective connectivity (EC), indicative of the causal interactions between brain regions, is fundamental to understanding information processing in the brain. Traditional approaches, which infer EC from neural responses to stimulations, are not suited for mapping whole-brain EC in humans due to being invasive and limited spatial coverage of stimulations. To address this gap, we present Neural Perturbational Inference (NPI), a data-driven framework designed to map EC across the entire brain. NPI employs an artificial neural network trained to learn large-scale neural dynamics as a computational surrogate of the brain. NPI maps EC by perturbing each region of the surrogate brain and observing the resulting responses in the rest of regions. NPI captures the directionality, strength, and excitatory/inhibitory properties of EC on a brain-wide scale.

## Result

A surrogate brain, an artificial neural network (ANN), is trained to learn brain dynamics (Fig.1a,b). It can then replace the real brain to be perturbed. ANN is optimized for predicting the subsequent brain state  $x(t+\Delta t)$  given the current brain state  $x(t)$ . After training, ANN is systematically perturbed to infer EC (Fig.1c,d). The all-to-all EC can be inferred by perturbing the ANN region by region. This EC is a brain-wide map of causal influences that shows directionality, strength, and excitatory-inhibitory distinction. Recurrently feeding the result of one-step prediction back as input to ANN produces the generated neural signals. The model FC and empirical FC are respectively calculated from generated BOLD signals and empirical BOLD signals, respectively, and then averaged across 800 subjects. The model FC and the empirical FC are highly correlated ( $r=0.97$ ,  $p < 0.001$ ), indicating that the trained ANN as a surrogate brain captures the inter-regional relationship of the real brain (Fig.1e-h). NPI is validated on generative models with known ground-truth EC (Fig.1i). NPI is applied to rsfMRI

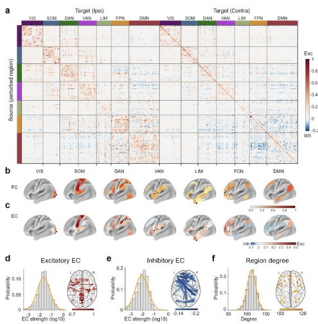


Fig 2. NPI-inferred Human Effective Brain Connectome (EBC).

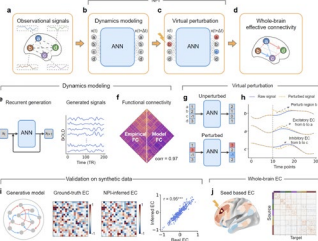


Fig 1. Neural Perturbational Inference (NPI) maps EC by virtually perturbing a surrogate brain.

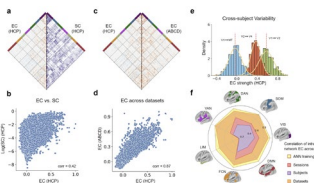


Fig 3. Reliability of EBC across ANN training and datasets.

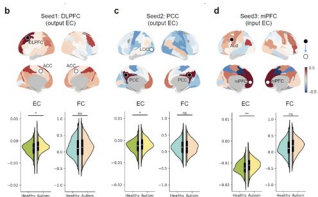


Fig 4. NPI reveals different EC in autism spectrum disorder.