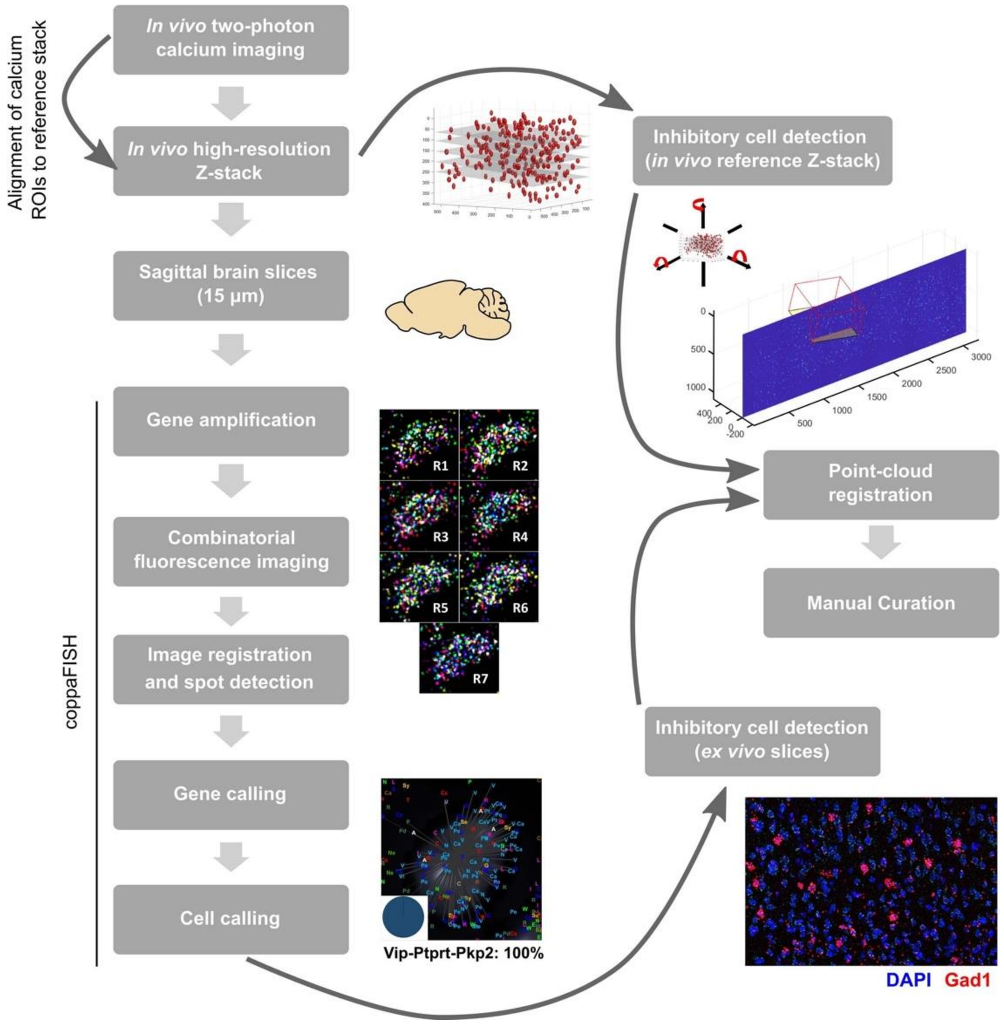
Neural Representational Consistency Emerges from Probabilistic Neural-Behavioral Representation Alignment

Yu Zhu, Chunfeng Song, Wanli Ouyang ,
Shan Yu , Tiejun Huang

Jun 16, 2025

1. Neuroscience relies on multimodal data analysis

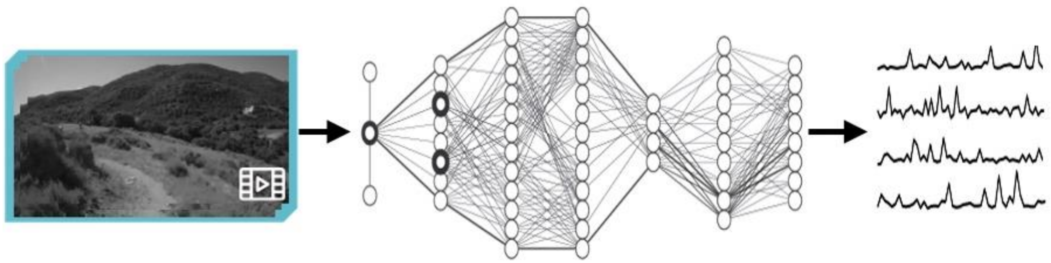
Multimodal data for decoding brain function: neural activity, behaviors, stimulus, gene…



Bugeon et al. (Nature, 2022)

2. Modeling paradigms vary across cortical regions

Sensory cortex (e.g., Visual cortex)



Walker et al. (Nat. Neurosci., 2019)

Predictive modeling (digital twin)

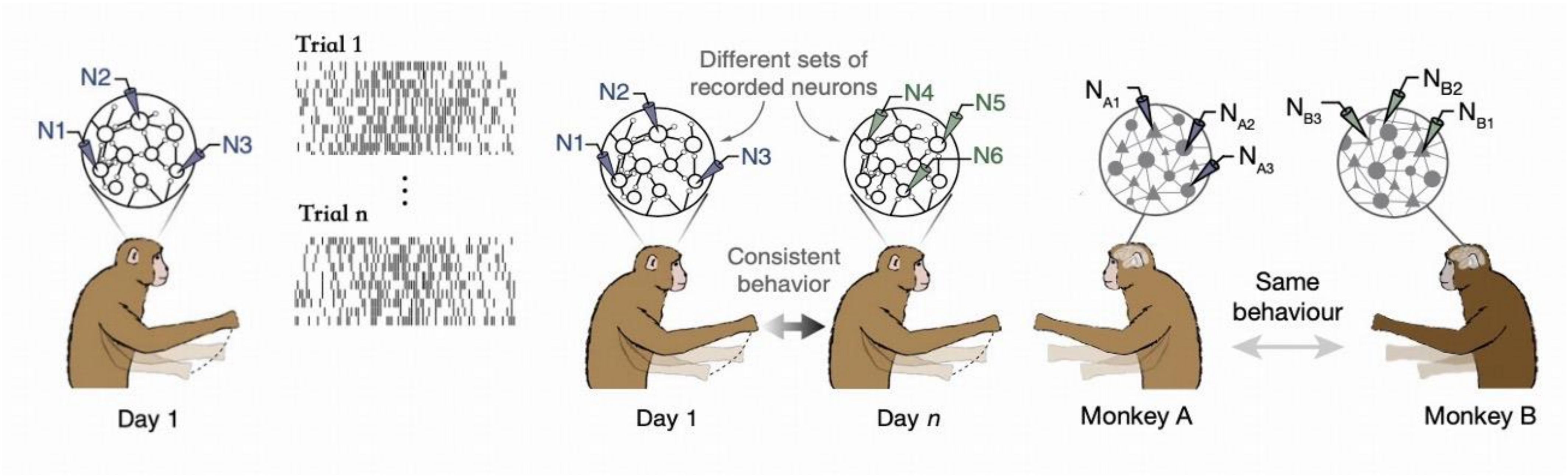
Motor cortex (e.g., PMd)

Gallego et al. (Nat. Neurosci., 2020)

Behavioral decoding

➡ Computational modeling fundamentally involves aligning neural activity patterns with their corresponding behaviors.

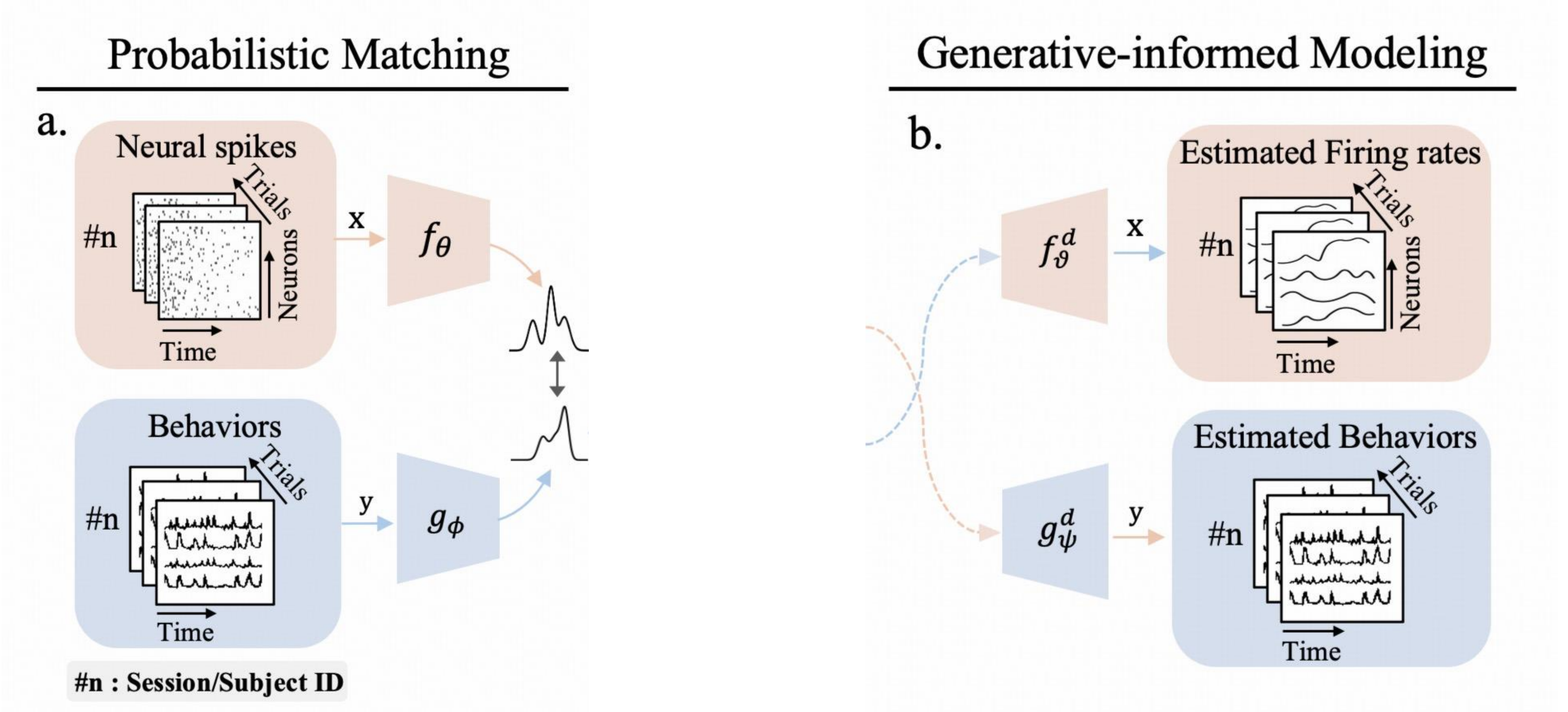
3. Inherent challenges in unified modeling



Cross-trials variability   Cross-session variability   Cross-subject variability

The inherent one-to-many correspondence between behavior and neural activity at the raw data level presents a theoretical barrier to establishing deterministic mathematical relationships.

4. Our probabilistic representation modeling paradigm



1. Adopt probabilistic representation modeling instead of seeking deterministic bijective mappings.
2. Strengthen representation learning through reconstruction constraints to prevent representational collapse.

5. Theoretical advantages of the alignment modeling

Theorem 3.1. *Let $f_\theta : \mathcal{X} \to \mathcal{Z}$ and $g_\phi : \mathcal{Y} \to \mathcal{Z}$ denote the neural and behavioral encoders. The following properties hold:*

(i) [Non-degeneracy] The optimization objective $\mathcal{L}_{total}$ prevents representation degeneration by ensuring:

$$\lim_{f_\theta(\mathbf{x}) \to \mathbf{z}_{const}} \mathcal{L}_{total} = +\infty, \quad \forall \mathbf{z}_{const} \in \mathcal{Z} \tag{10}$$

(ii) [Information Preservation] The generative constraints ensure enhanced mutual information between input and representation spaces:

$$I(f_{\mathcal{L}_{total}}(\mathbf{x}); \mathbf{x}) \geq I(f_{\mathcal{L}_{ProbMatch}}(\mathbf{x}); \mathbf{x}) + \eta, \quad \eta > 0 \tag{11}$$

(iii) [Representation Stability] For neural responses $\mathbf{x}_i, \mathbf{x}_j$ corresponding to the same behavioral variable $\mathbf{y}$, the learned representations maintain bounded distances:

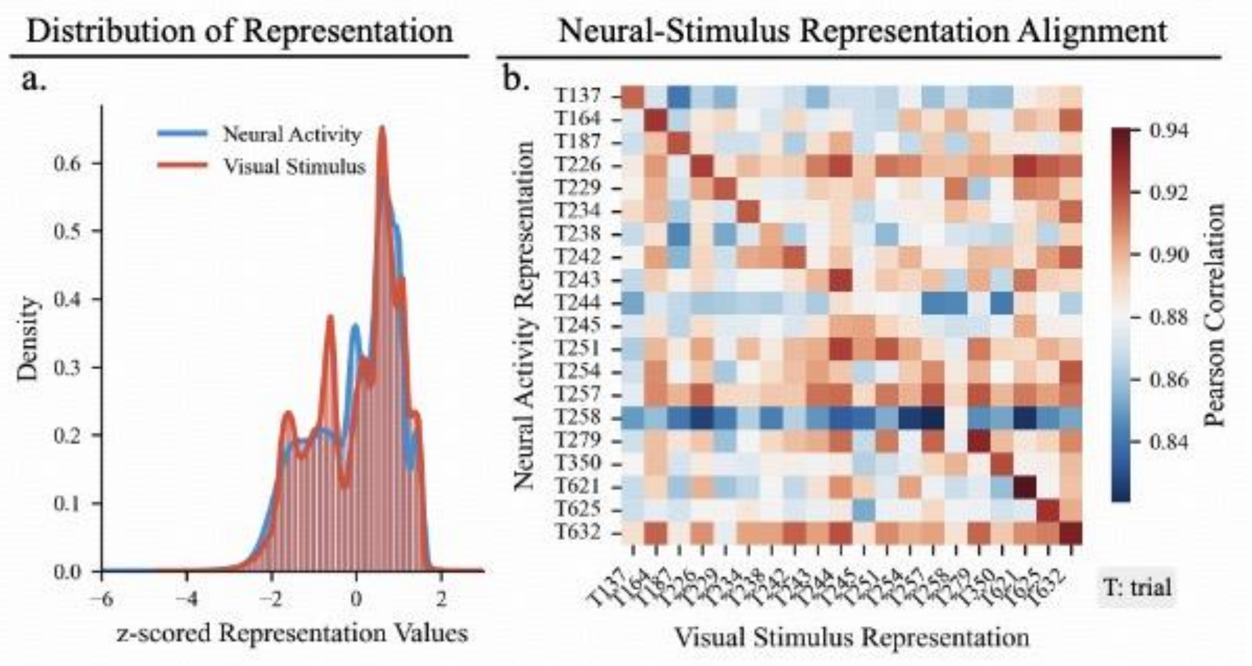
$$0 < \alpha \leq \|f_\theta(\mathbf{x}_i) - f_\theta(\mathbf{x}_j)\|_2 \leq \beta \tag{12}$$

Note: property (iii) addresses the inherent neural variability associated with identical behavioral outputs by maximizing representational similarity while preserving biological diversity.

Theoretical results show that the representation alignment paradigm is optimal for computational modeling.

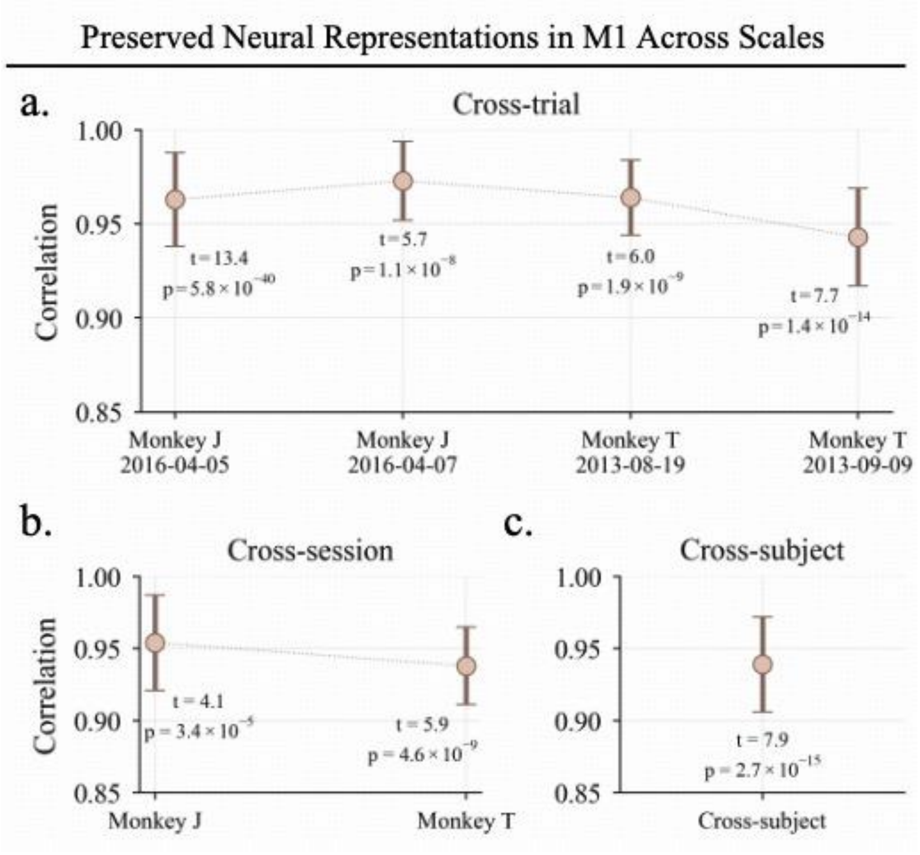
6. Alignment results of latent representation modeling

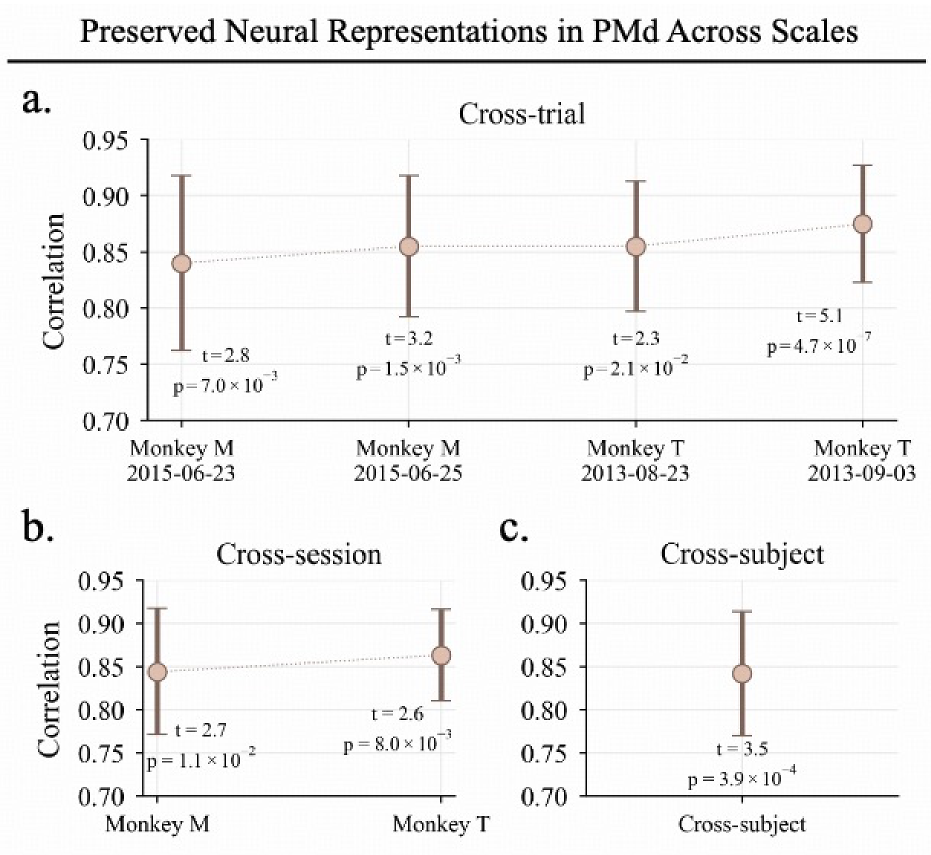
Cortical Area	Method	Correlation (R) Training Subjects	Correlation (R) New Subjects
Motor Cortex (M1)	VAE§	0.0197	0.0016
	FA+Procrustes†	0.3334	0.2009
	PCA+CCA†	0.3520	0.2160
	FA+amLDS†	0.5807	0.3627
	Neuroformer*	0.5214	–
	MEME	0.7756	0.7060
	PNBA (Ours)	**0.9465**	**0.9302**
Motor Cortex (PMd)	VAE§	0.0063	0.0028
	FA+Procrustes†	0.3605	0.2877
	PCA+CCA†	0.3916	0.3397
	FA+amLDS†	0.4733	0.4366
	Neuroformer*	0.3283	–
	MEME	0.5279	0.5255
	PNBA (Ours)	**0.9248**	**0.9176**
Visual Cortex (V1)	VAE§	0.0029	-0.0009
	FA+Procrustes†	0.1221	0.1207
	PCA+CCA†	0.1210	0.1209
	FA+amLDS†	0.1509	0.1501
	Neuroformer*	0.4116	–
	MEME	0.6357	0.5980
	PNBA (Ours)	**0.8830**	**0.8705**



Our probabilistic representation alignment framework achieves optimal alignment performance.

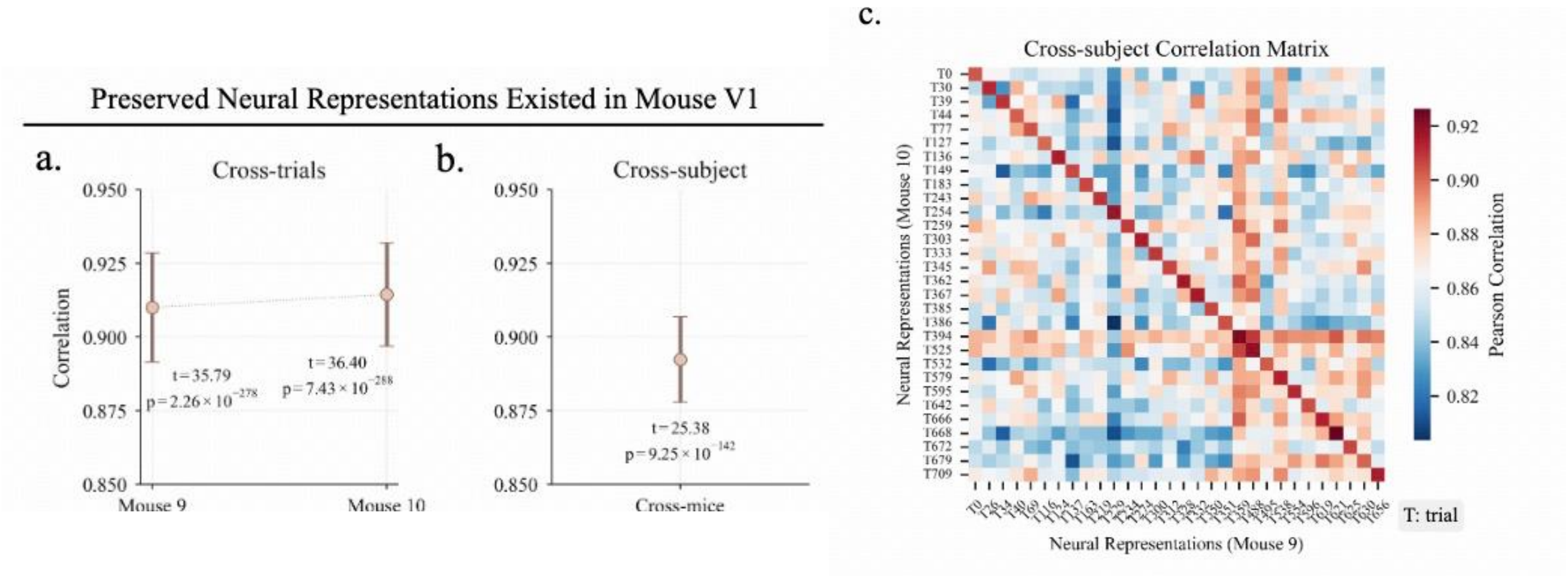
7. Preserved neural representations emerge





Using a data-driven paradigm, we provide evidence that preserved neural representations persist across novel subjects, thereby strengthening the findings of Safaie et al. (2023, Nature).

8. Preserved Neural Representations broadly exist



We further identified preserved neural representations in mouse somatosensory cortex, extending the conclusions of Safaie et al. (2023, Nature).

9. Conclusion

- **Unified Framework**: We introduce a probabilistic representation alignment paradigm applicable across cortical regions.
- **Theoretical Optimization**: Our framework ensures maximum biological plausibility through theoretical constraints.
- **Clinical Impact**: We extend Safaie et al. findings and demonstrate zero-shot BCI potential using preserved V1 representations.