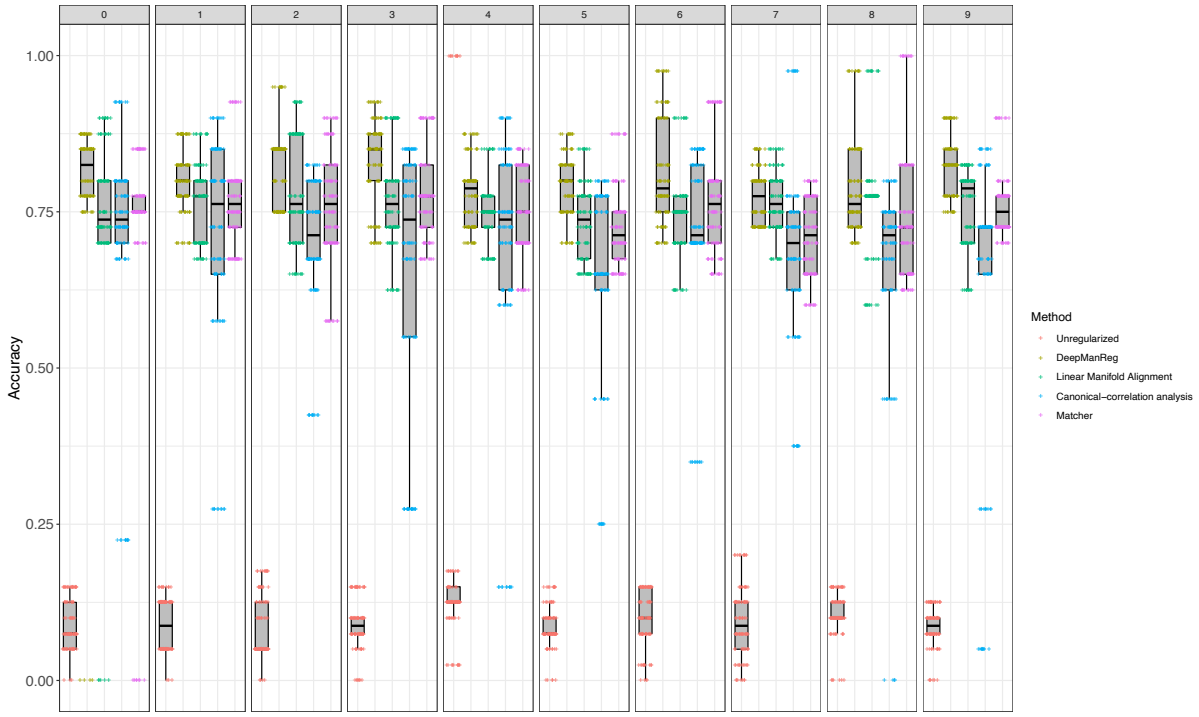
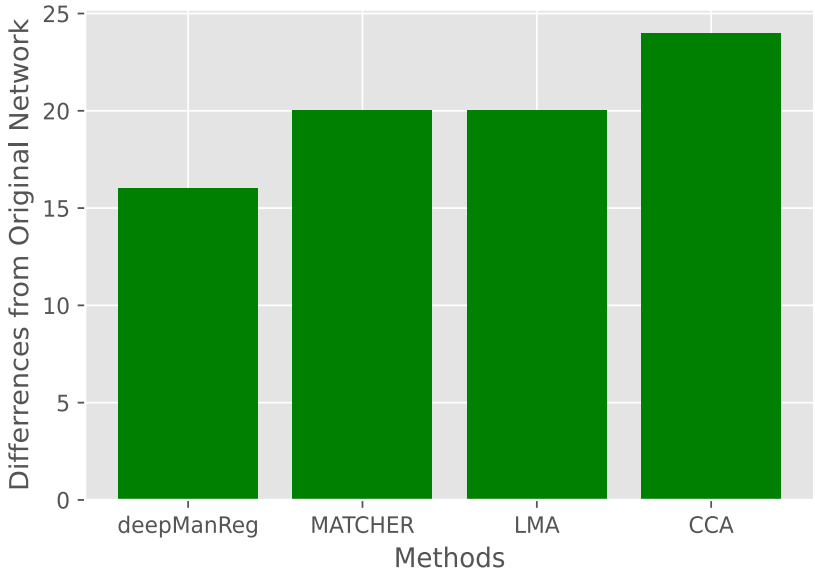
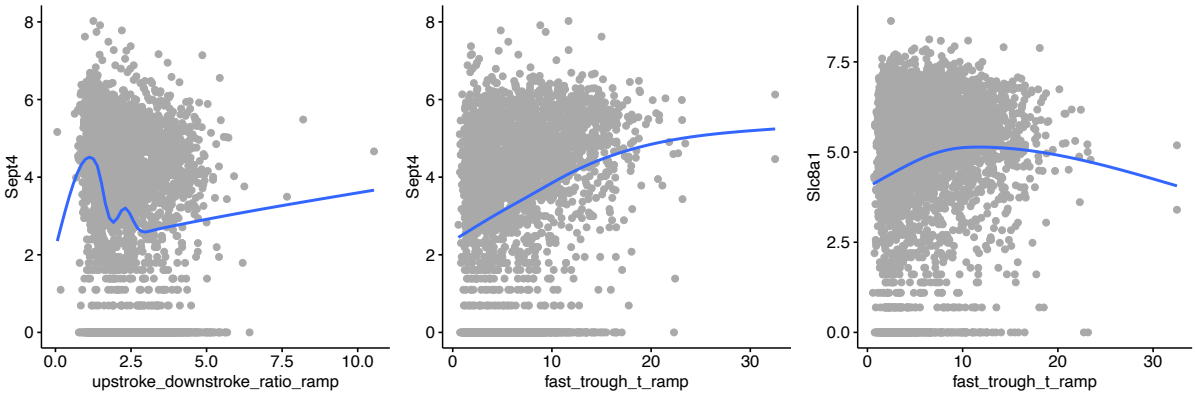


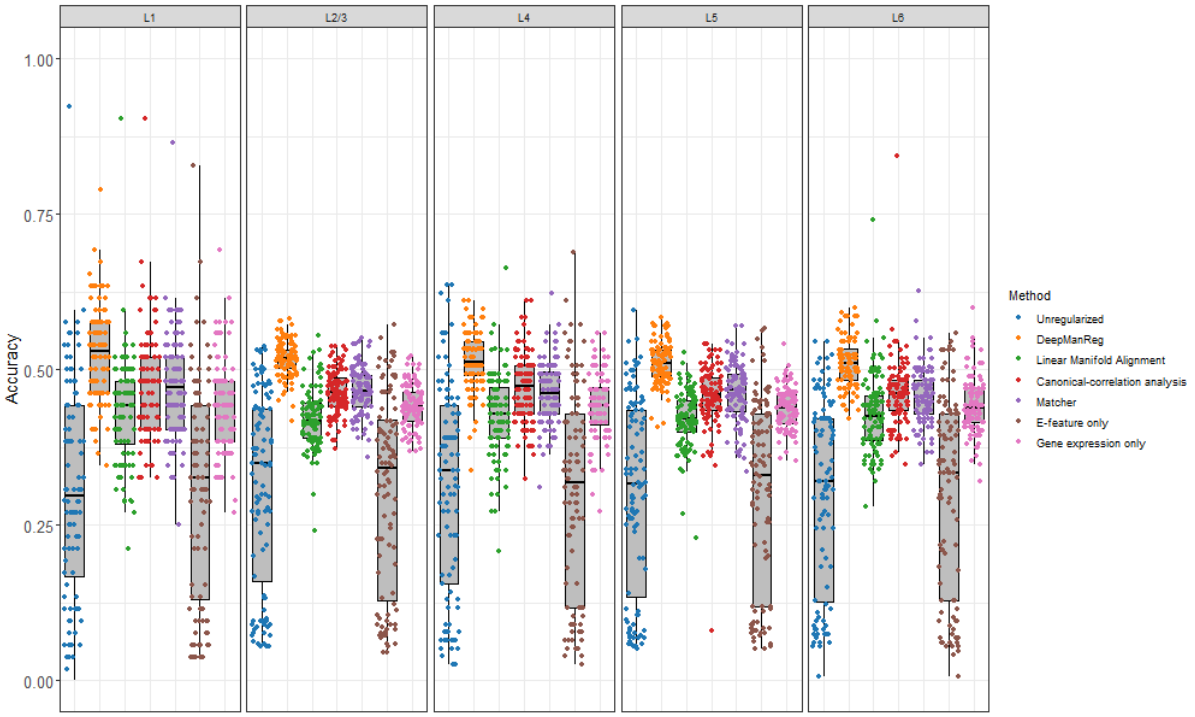
Supplementary Figure 1 Boxplots of testing accuracies for classifying digits in the mfeat dataset by deepManReg (Orange) vs. neural network classification without any regularization vs. Linear Manifold Alignment (Blue) vs. CCA (Red) vs. MATCHER (Purple). Points are shown on the plots.



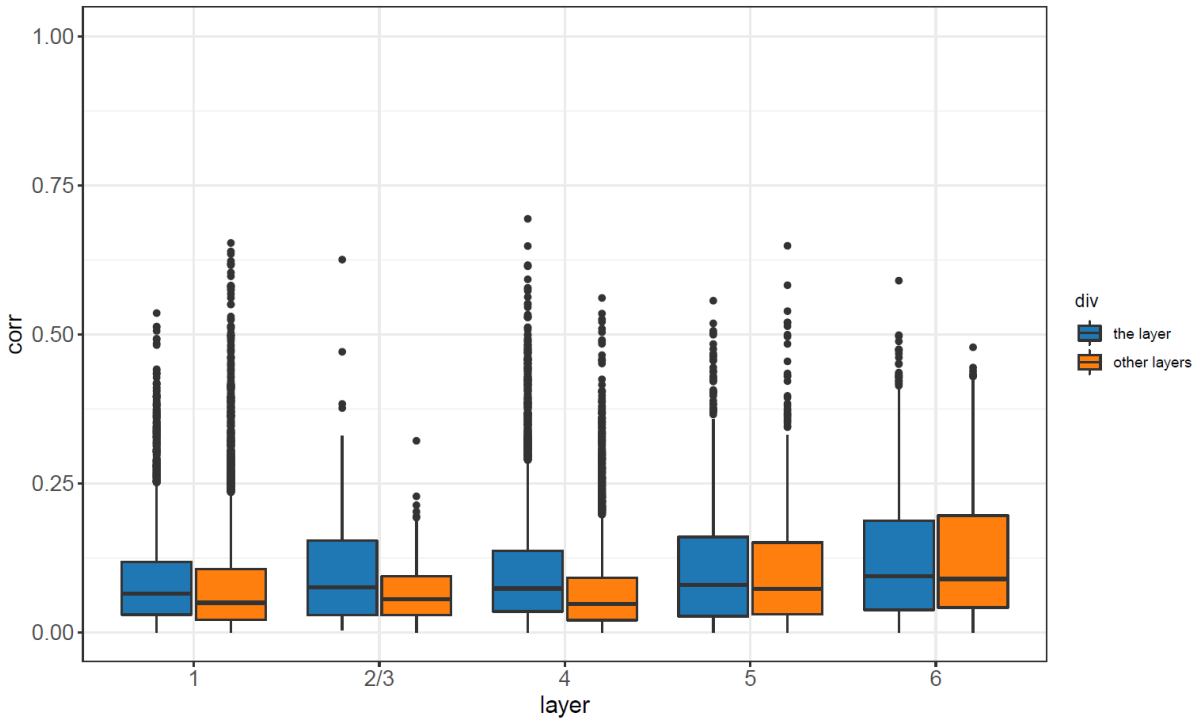
Supplementary Figure 2 Barplots showing the differences between the original model gene regulatory network [1] and the networks reconstructed by deepManReg, MATCHER, LMA, and CCA. The differences are characterized by the score: $S = |X' - \text{Original}| + |Y' - \text{Original}| + |X' - Y'|$, where X' and Y' are the adjacency matrices of reconstructed networks by RNA abundance and protein abundance (i.e., Modalities 1 and 2) respectively, Original is the original model network adjacency matrix.



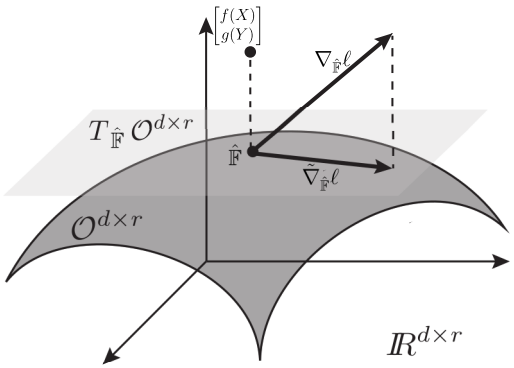
Supplementary Figure 3 Select connected genes vs. electrophysiological features in Figure 4 show significantly nonlinear relationships across cells. Dots: cells. Left: Sept4 vs. upstroke_downstroke_ratio_ramp (Spearman correlation ρ = -0.299, p-value < 2.2e-16). Middle: Sept4 vs. fast_trough_t_ramp (Spearman correlation ρ = 0.287, p-value < 2.2e-16). Right: Slc8a1 vs. fast_trough_t_ramp (Spearman correlation ρ = 0.127, p-value < 1.6e-14). The blue curves represent a fitted smoothing line by a generalized additive model to predict y ~ x.



Supplementary Figure 4 Boxplots of testing accuracies for classifying cell layers in the mouse visual cortex by deepManReg (Orange) vs. neural network classification without any regularization using both modalities (Blue), e-features only (Brown), and gene expression only (Pink) by Linear Manifold Alignment (Green), CCA (Red), and MATCHER (Purple). Points are shown on the plots.



Supplementary Figure 5 Boxplot showing spearman correlations between layer genes and e-features. For each layer, we calculate the correlations between layer genes and e-features on cells inside that layer (blue) and outside that layer (orange).



Supplementary Figure 6 Optimization on Stiefel Manifold. In forward pass, the concatenation of the outputs of the 2 neural networks, $\begin{bmatrix} f(X) \\ g(Y) \end{bmatrix}$, is projected onto Stiefel Manifold. In backward pass, the Euclidean gradient, $\nabla_{\hat{F}}\ell$, is projected onto the tangent space of the Stiefel Manifold to get the Riemannian gradient, $\tilde{\nabla}_{\hat{\mathbb{F}}}l$.

Algorithm 1: Deep Manifold Alignment

input : data for two modalities X & Y
params: training step T, learning rate η
output : parameters $\mathcal{W}_{T+1}$ & $\mathcal{Z}_{T+1}$

initialize $\mathcal{W}_0$ & $\mathcal{Z}_0$;
for $t = 0 : T$ **do**
  // forward pass
  $f_t \leftarrow f(X; \mathcal{W}_t)$;
  $g_t \leftarrow g(Y; \mathcal{Z}_t)$;
  $R_t \leftarrow \begin{bmatrix} f_t \\ g_t \end{bmatrix}$;
  Calculate L ; // L is the joint Laplacian
  // project the output onto Stiefel manifold
  $\hat{\mathbb{F}}_t \leftarrow U_t I V_t^T$ where $R_t = U_t \Sigma_t V_t^T$ is the SVD decomposition of R_t and $\mathbb{I}$ is the identity matrix;
  $\ell \leftarrow tr(\hat{\mathbb{F}}_t^T L \hat{\mathbb{F}}_t)$; // compute loss
  $e_t \leftarrow \nabla_{\hat{\mathbb{F}}_t} \ell$; // compute Euclidean gradient
  // project Euclidean gradient onto the tangent space of Stiefel manifold
  $p_t \leftarrow \hat{\mathbb{F}}_t skew(\hat{\mathbb{F}}_t^T e_t) + (I - \hat{\mathbb{F}}_t \hat{\mathbb{F}}_t^T) e_t$;
  // backpropagate the Riemannian gradient
  $\mathcal{W}_t, \mathcal{Z}_t \leftarrow \text{backprop}(p_t)$ $\mathcal{W}_{t+1} \leftarrow \mathcal{W}_t + g(\mathcal{W}_t, \eta, t)$;
  $\mathcal{Z}_{t+1} \leftarrow \mathcal{Z}_t + g(\mathcal{Z}_t, \eta, t)$ where g is an optimizer (e.g., SGD)
end

Supplementary Algorithm Optimization on Stiefel Manifold. In forward pass, the concatenation of the outputs of the 2 neural networks, $\begin{bmatrix} f(X) \\ g(Y) \end{bmatrix}$, is projected onto Stiefel Manifold. In backward pass, the Euclidean gradient, $\nabla_{\hat{F}} \ell$, is projected onto the tangent space of the Stiefel Manifold to get the Riemannian gradient, $\widetilde{\nabla}_{\hat{\mathbb{F}}} l$.

α	β	test accuracy
1	0	0.80275
0.995	0.005	0.80225
0.99	0.01	0.80125
0.8	0.2	0.79525
0.5	0.5	0.78875
0.3	0.7	0.78375
0.2	0.8	0.77925
0.1	0.9	0.75175
0	1	0.693

Supplementary Table 1 Ablation study on parameters α and β for feature-network regularized learning showing that feature-network regularization contributes most to improve the prediction outcomes.

References

[1] Robrecht Cannoodt, Wouter Saelens, Louise Deconinck, and Yvan Saeys. Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. Nature Communications, 12(1):1–9, 2021.