Patterns, Volume *4*

Supplemental information

Connectome-based machine learning models are vulnerable to subtle data manipulations

Matthew Rosenblatt, Raimundo X. Rodriguez, Margaret L. Westwater, Wei Dai, Corey Horien, Abigail S. Greene, R. Todd Constable, Stephanie Noble, and Dustin Scheinost

	IQ prediction		**Self-reported sex classification**					
	rCPM		SVM (Linear kernel)			Logistic regression		
Dataset	*r*	*q*^{2}	Acc	Sens	Spec	Acc	Sens	Spec
ABCD	-0.025 (0.010)	-0.031 (0.003)	0.860 (0.003)	0.850 (0.006)	0.869 (0.007)	0.805 (0.017)	0.800 (0.048)	0.809 (0.049)
HCP	0.177 (0.016)	0.031 (0.006)	0.883 (0.009)	0.859 (0.017)	0.903 (0.011)	0.767 (0.035)	0.736 (0.082)	0.794 (0.072)
PNC	0.243 (0.012)	0.058 (0.005)	0.807 (0.010)	0.743 (0.021)	0.855 (0.013)	0.711 (0.025)	0.629 (0.083)	0.774 (0.069)

Table S1. Baseline accuracies for regression models of IQ and classification models of self-reported sex in ABCD, HCP, and PNC. Prediction performance is evaluated with 10-fold cross-validation, with nested cross-validation to select L_2 regularization. The numbers in parentheses reflect the standard deviation of the metrics across 100 iterations of different random seeds.

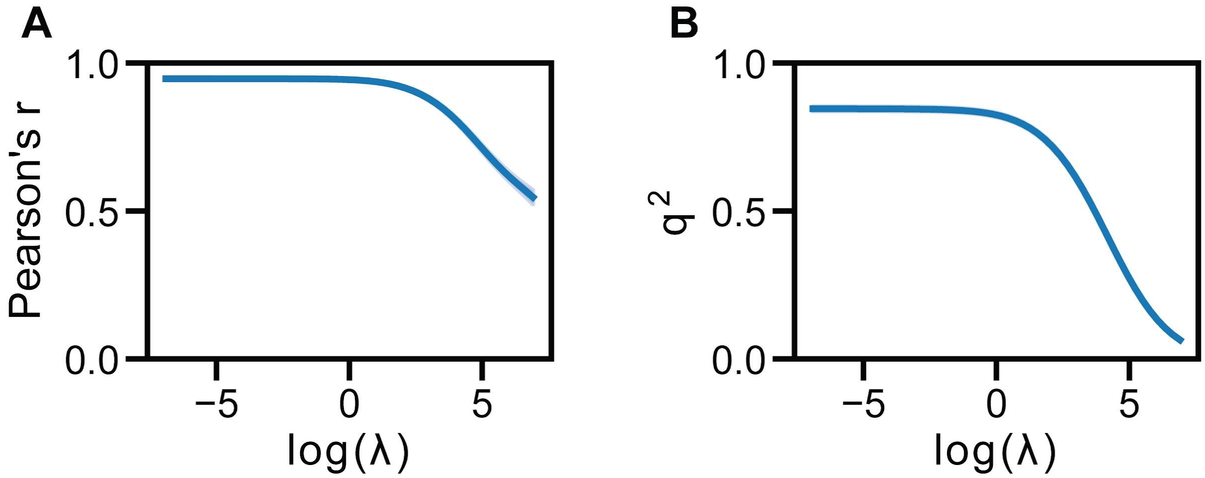

Figure S1. Enhancement performance for an attack scale, or mean absolute value of the enhancement pattern, of 0.01 and a variety of λ. The plots show the enhanced **a)** Pearson's r and **b)** q^2 as a function of the regularization parameter.

	Ridge regression		**Neural network**	
Scale	*r*	q^2	*r*	q^2
0	0.245 (0.013)	0.060 (0.006)	0.229 (0.012)	-0.088 (0.014)
0.01	0.441 (0.010)	0.159 (0.005)	0.575 (0.011)	0.330 (0.012)
0.02	0.770 (0.005)	0.379 (0.004)	0.898 (0.003)	0.761 (0.005)
0.03	0.921 (0.002)	0.592 (0.004)	0.967 (0.001)	0.910 (0.003)

Table S2. Enhancement attacks in HCP resting-state functional connectomes to predict IQ with ridge regression and with neural networks. Experiments were repeated ten times for different cross-validation splits.

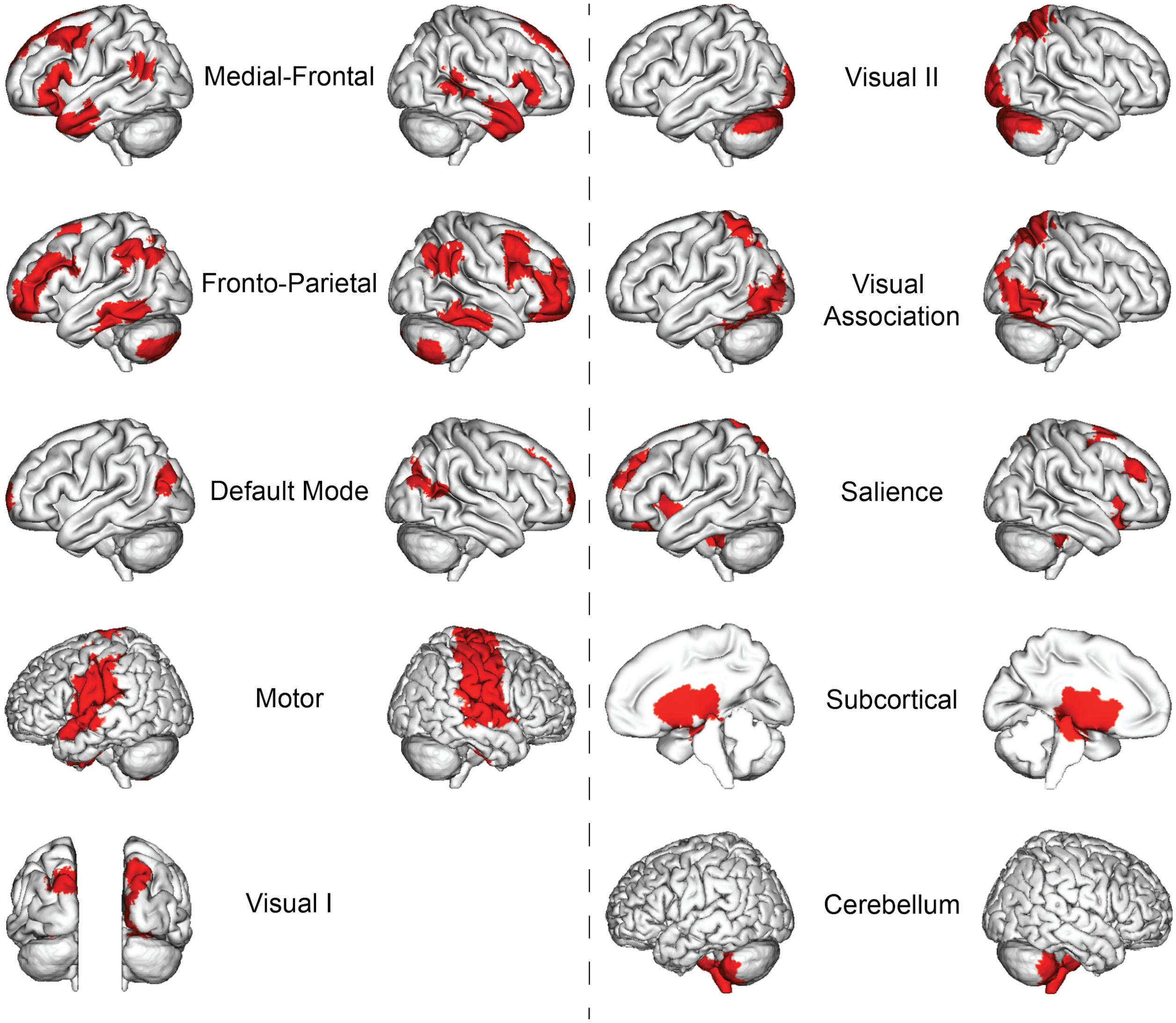

Figure S2. Network definitions with the Shen 268 atlas, related to Figures 3 and 7.

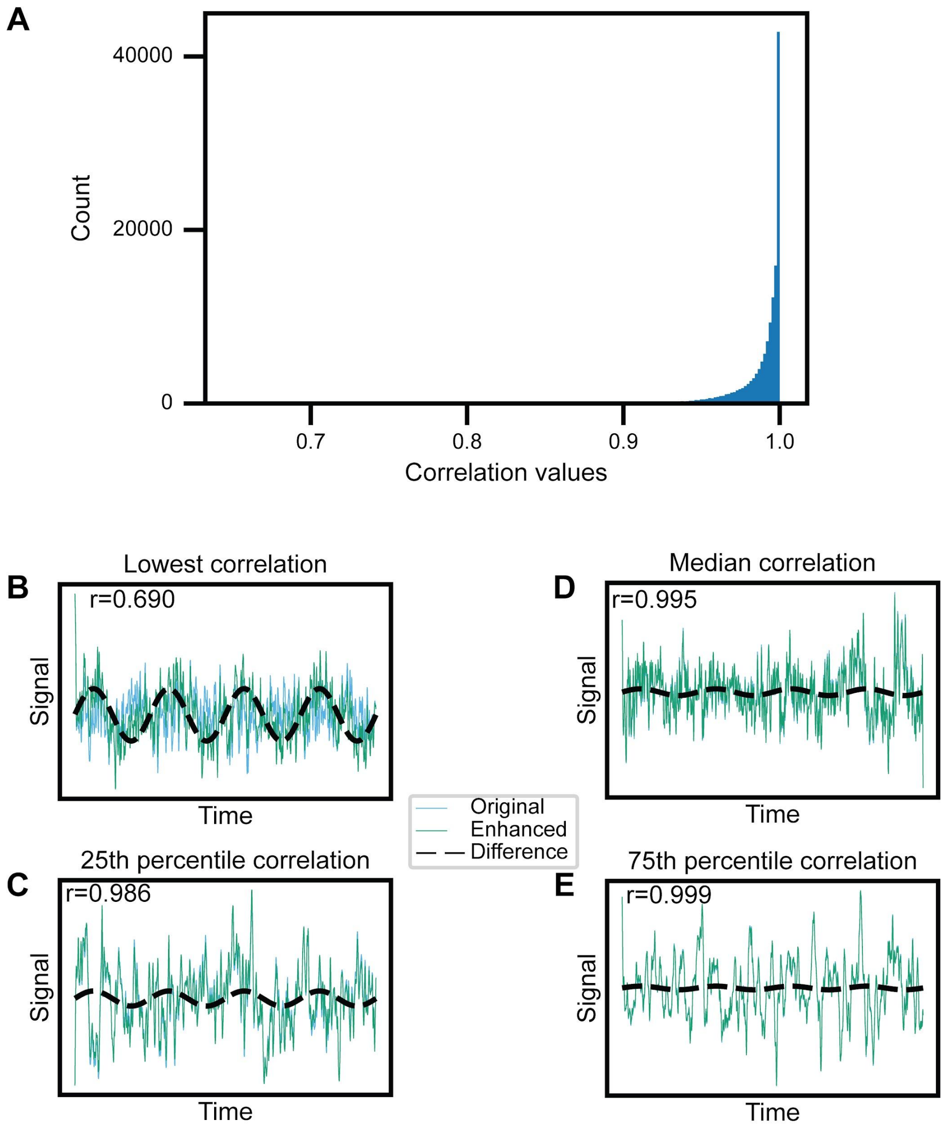

Figure S3. Examples of original and enhanced node time-series data, related to Figure 5. **a)** Histogram of correlation values between original and enhanced time-series data across all nodes (268) and participants (506), the vast majority of which are *r*>0.9. **b)** Original and enhanced time-series data with the lowest correlation across all nodes and participants (*r=0.690*). **c)** Data with the 25th percentile of correlations (*r=0.986*). **d)** Data with the median correlation (*r=0.995*). **e)** Data with the 75th percentile of correlations (*r=0.999*).

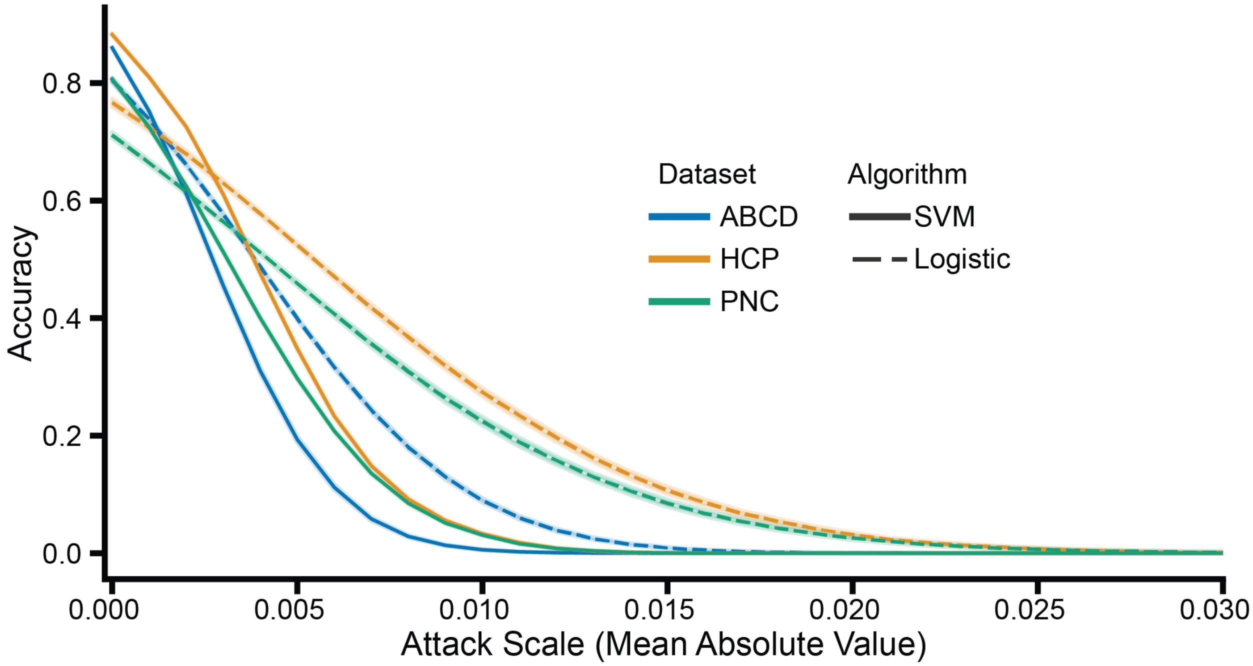

Figure S4. Comparison of adversarial robustness of SVM and logistic regression, related to Figure 6. The logistic regression models had higher robustness to manipulations for these particular predictions, meaning that a larger attack scale was required to decrease the accuracy. However, the baseline accuracy of logistic regression models was lower than that of SVM.