

Patterns, Volume 3

Supplemental information

**DeepND: Deep multitask learning of gene risk
for comorbid neurodevelopmental disorders**

Ilayda Beyreli, Oguzhan Karakahya, and A. Ercument Cicek

A Supplementary Figures

Supplementary Figure 1. Comparison of forecASD, DeepND and forecASD + DeepND performances.

(a) MCC between the rankings of each method and the ground truth genes are shown for varying rank percentage threshold values (x axis). This ranking threshold p sets the top $p\%$ of a ranking as positive predictions. and (b) precision-recall curves.

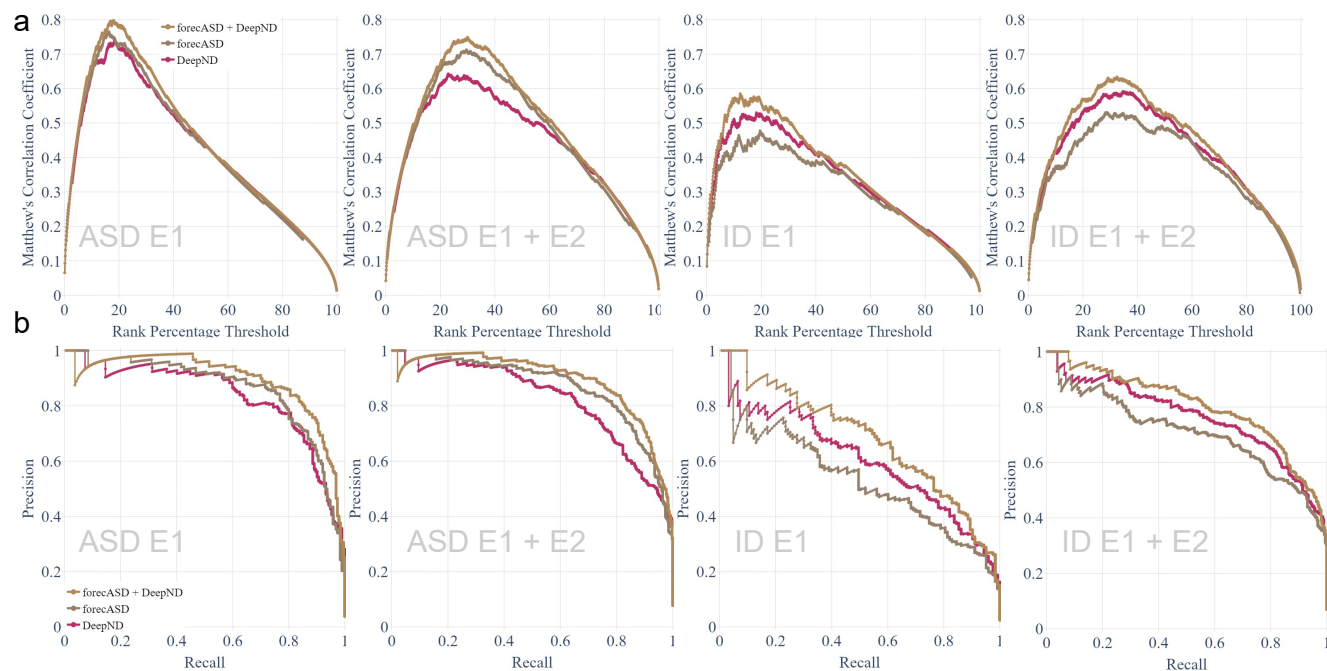


Figure 1: Comparison of forecASD, DeepND and forecASD + DeepND performances. (a) MCC between the rankings of each method and the ground truth genes are shown for varying rank percentage threshold values (x axis). This ranking threshold p sets the top $p\%$ of a ranking as positive predictions. and (b) precision-recall curves.

Supplementary Figure 2. Jarracd Index for DeepND ASD and ID Rankings

(a) The percentage overlap between corresponding deciles of the ASD and ID genome-wide risk rankings are shown. (b) The percentage overlap between corresponding percentiles of the ASD and ID genome-wide risk rankings are shown.

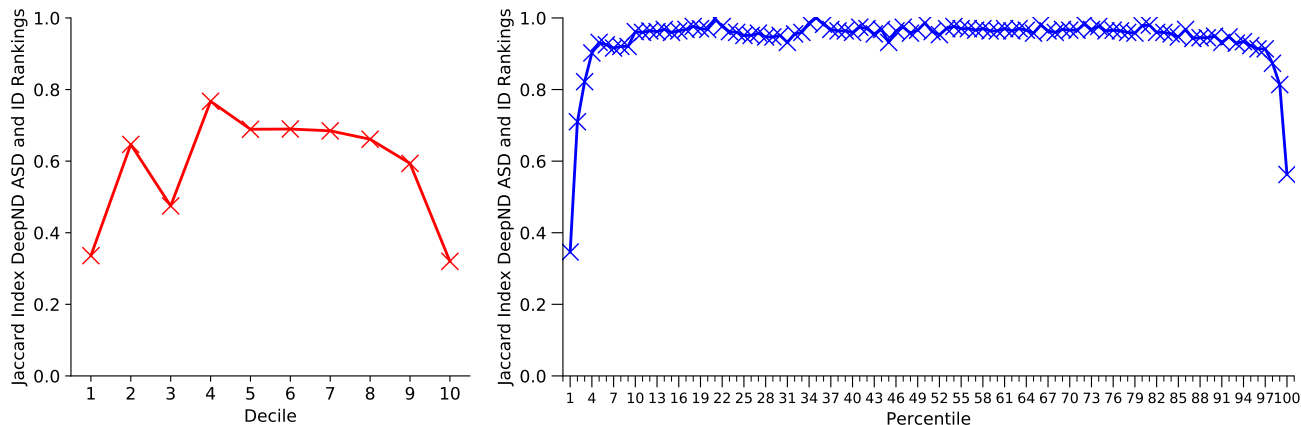


Figure 2: Jarracd Index for DeepND ASD and ID Rankings (a) The percentage overlap between corresponding deciles of the ASD and ID genome-wide risk rankings are shown. (b) The percentage overlap between corresponding percentiles of the ASD and ID genome-wide risk rankings are shown.

Supplementary Figure 3. Network Probability Heatmaps for ASD and ID

Heatmaps show which spatio-temporal windows lead to assignment of higher risk probabilities to the top percentile genes for ASD (left) and ID (right). The numbers in boxes for network probabilities are softmaxed outputs of each respective GCN, averaged for top percentile genes and then normalized. The categorization of brain regions and time windows are provided in Figure 3 in the main text.

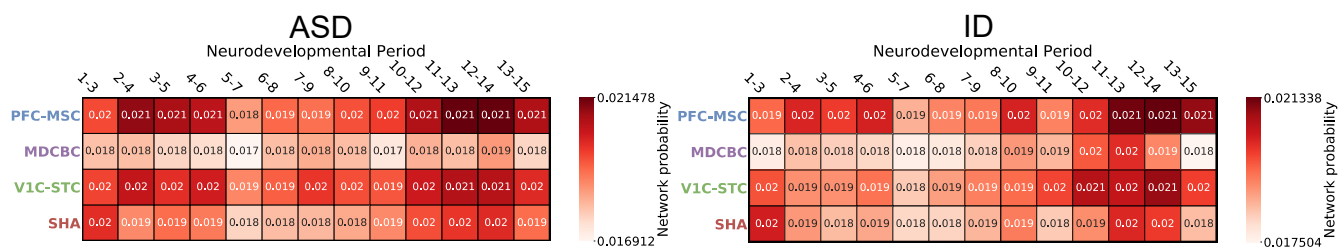


Figure 3: Network Probability Heatmaps for ASD and ID. Heatmaps show which spatio-temporal windows lead to assignment of higher risk probabilities to the top percentile genes for ASD (left) and ID (right). The numbers in boxes for network probabilities are softmaxed outputs of each respective GCN, averaged for top percentile genes and then normalized. The categorization of brain regions and time windows are provided in Figure 3 in the main text.

Supplementary Figure 4. Permutation Test Results for ASD

Results for the permutation test to assess the significance of the informativeness of each neurodevelopmental window for ASD. The box plots show the distribution of the average probabilities assigned by the corresponding GCN to top percentile genes, for each of the 100 randomized ground truth sets. The red cross indicates the result obtained with actual ground truth labels and match with the results presented in Figure 3b in the main text.

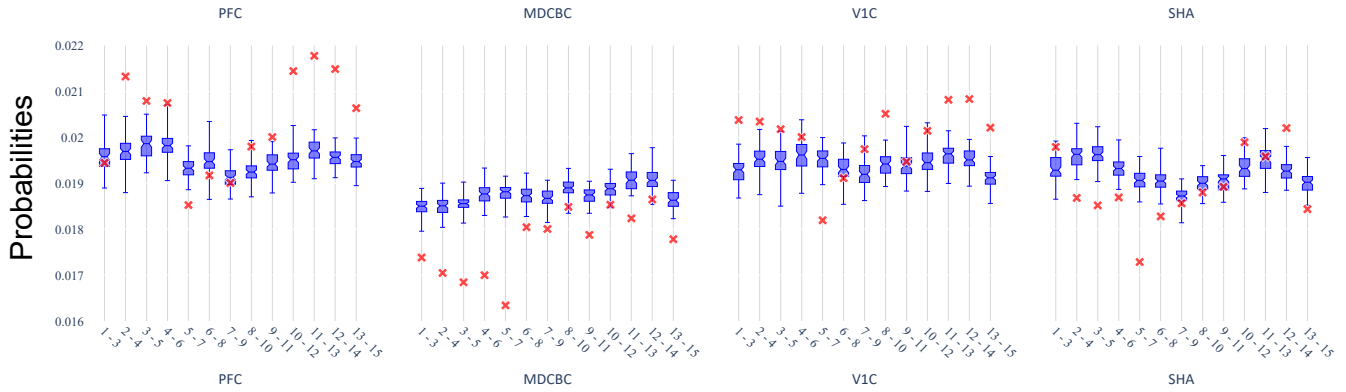


Figure 4: Permutation Test Results for ASD. Results for the permutation test to assess the significance of the informativeness of each neurodevelopmental window for ASD. The box plots show the distribution of the average probabilities assigned by the corresponding GCN to top percentile genes, for each of the 100 randomized ground truth sets. The red cross indicates the result obtained with actual ground truth labels and match with the results presented in Figure 3b in the main text.

Supplementary Figure 5. Permutation Test Results for ID

Results for the permutation test to assess the significance of the informativeness of each neurodevelopmental window for ID. The box plots show the distribution of the average probabilities assigned by the corresponding GCN to top percentile genes, for each of the 100 randomized ground truth sets. The red cross indicates the result obtained with actual ground truth labels and match with the results presented in Figure 3b in the main text.

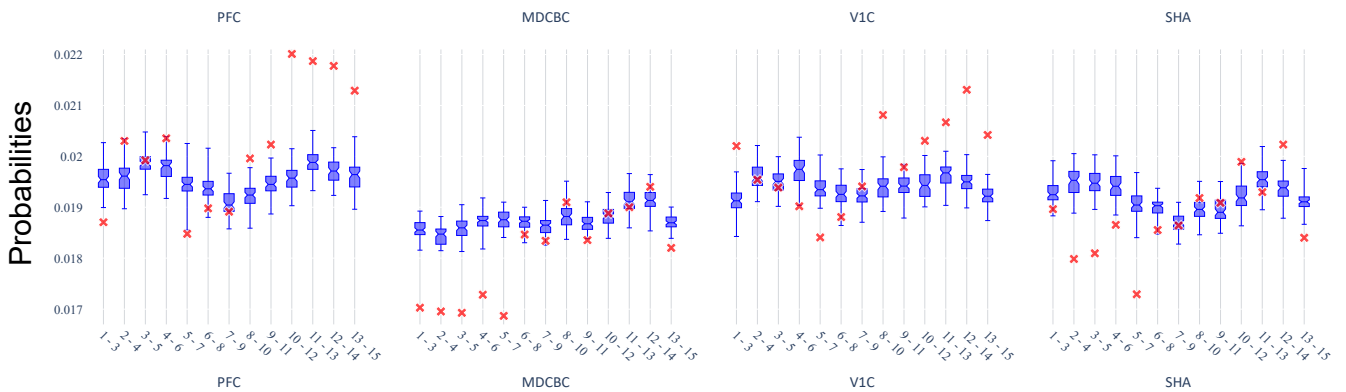


Figure 5: Permutation Test Results for ID. Results for the permutation test to assess the significance of the informativeness of each neurodevelopmental window for ID. The box plots show the distribution of the average probabilities assigned by the corresponding GCN to top percentile genes, for each of the 100 randomized ground truth sets. The red cross indicates the result obtained with actual ground truth labels and match with the results presented in Figure 3b in the main text.

Supplementary Figure 6. Evaluation of DeepND genome-wide risk assessment for breast and ovarian cancer.

(a) The area under ROC curve distributions of DeepND and DeepND-ST for breast and ovarian cancer genome-wide risk assessments. Every point corresponds to the performance on a test fold in the repeated cross validation setting. (b) The area under precision-recall curve distributions for the same methods as in (a).

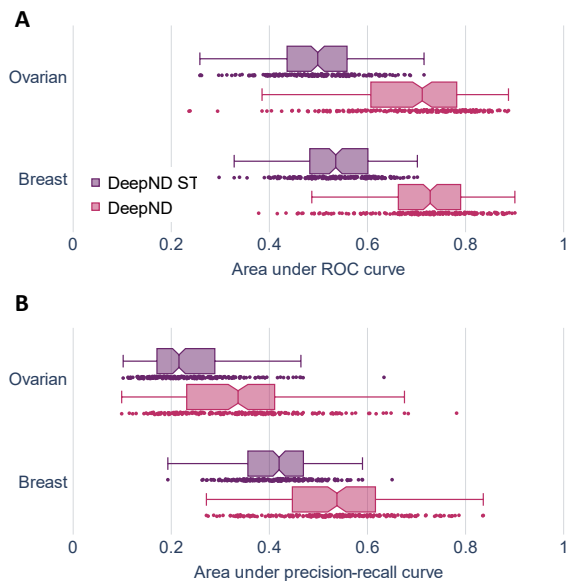


Figure 6: Evaluation of DeepND genome-wide risk assessment for breast and ovarian cancer.(a) The area under ROC curve distributions of DeepND and DeepND-ST for breast and ovarian cancer genome-wide risk assessments. Every point corresponds to the performance on a test fold in the repeated cross validation setting. (b) The area under precision-recall curve distributions for the same methods as in (a).

Supplementary Figure 7. Evaluation of DeepND genome-wide risk assessment for ADHD and ID

(a) The area under ROC curve distributions of DeepND and DeepND-ST for ADHD and ID. genome-wide risk assessments. Every point corresponds to the performance on a test fold in the repeated cross validation setting. (b) The area under precision-recall curve distributions for the same methods as in (a).

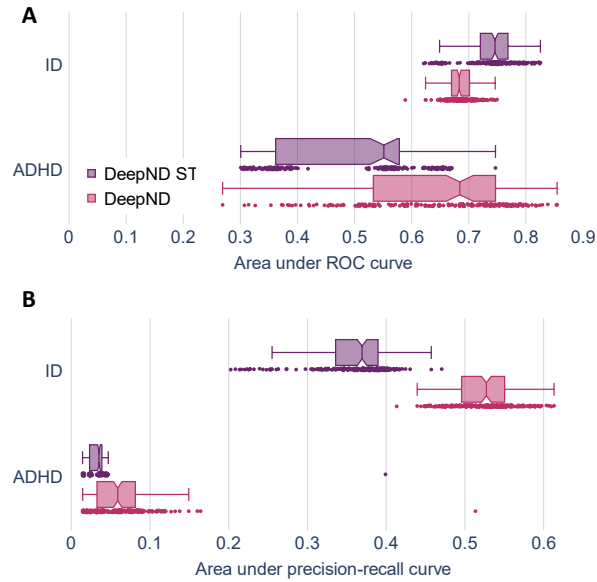


Figure 7: Evaluation of DeepND genome-wide risk assessment for ADHD and ID. (a) The area under ROC curve distributions of DeepND and DeepND-ST for ADHD and ID. genome-wide risk assessments. Every point corresponds to the performance on a test fold in the repeated cross validation setting. (b) The area under precision-recall curve distributions for the same methods as in (a).