**Supplementary Material**

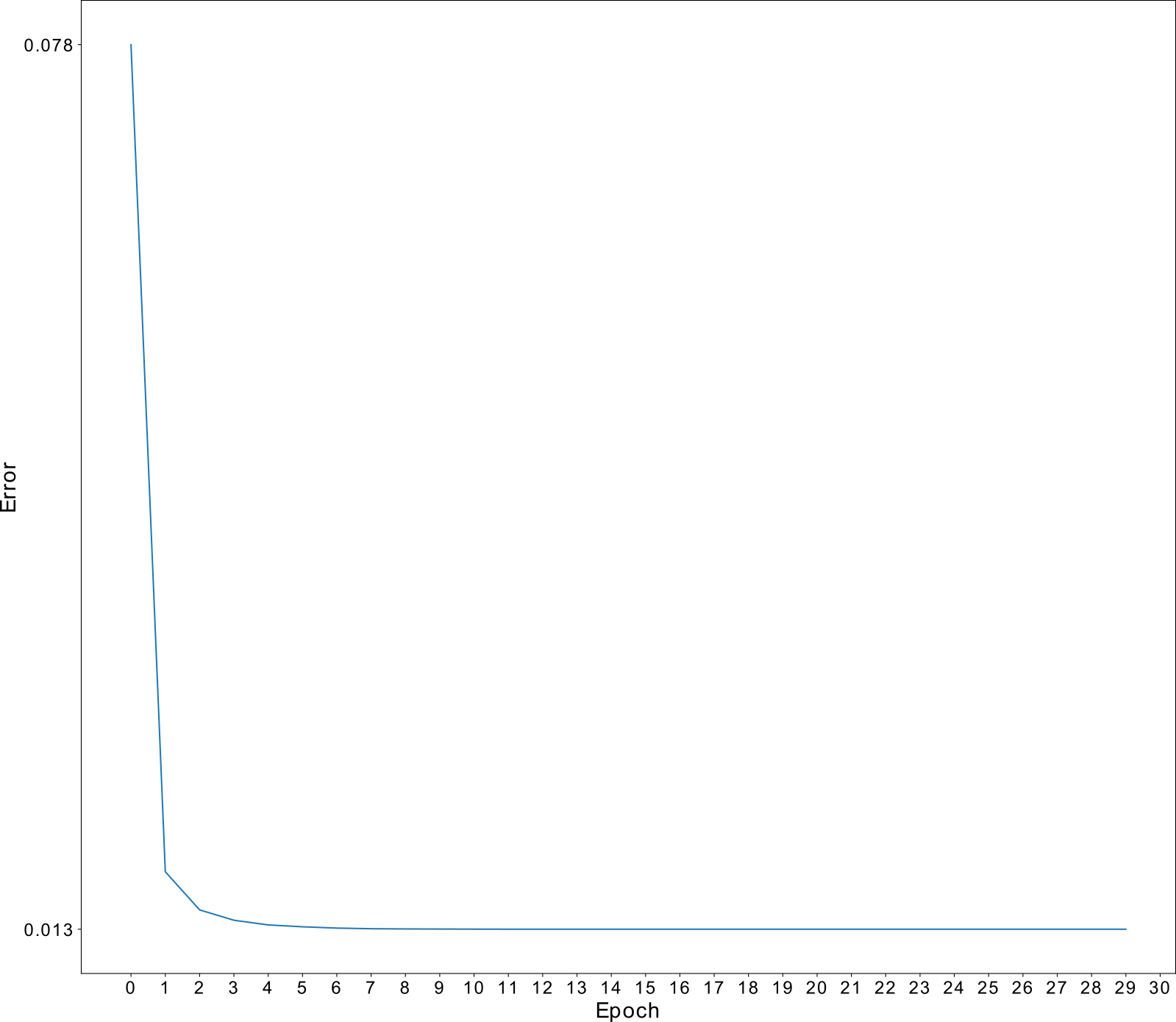


Fig. S1: Cross-entropy loss of the network evaluated on the validation set before each epoch of training. The network contained motifs of consecutive length starting from 5 and ending with 24. The network was trained for 30 epochs and quickly reached a stable local minimum after approximately seven epochs. This loss value did not change for the remainder of training, even after training for over 100 epochs, indicating that the network was able to generalize to the full dataset without overfitting.

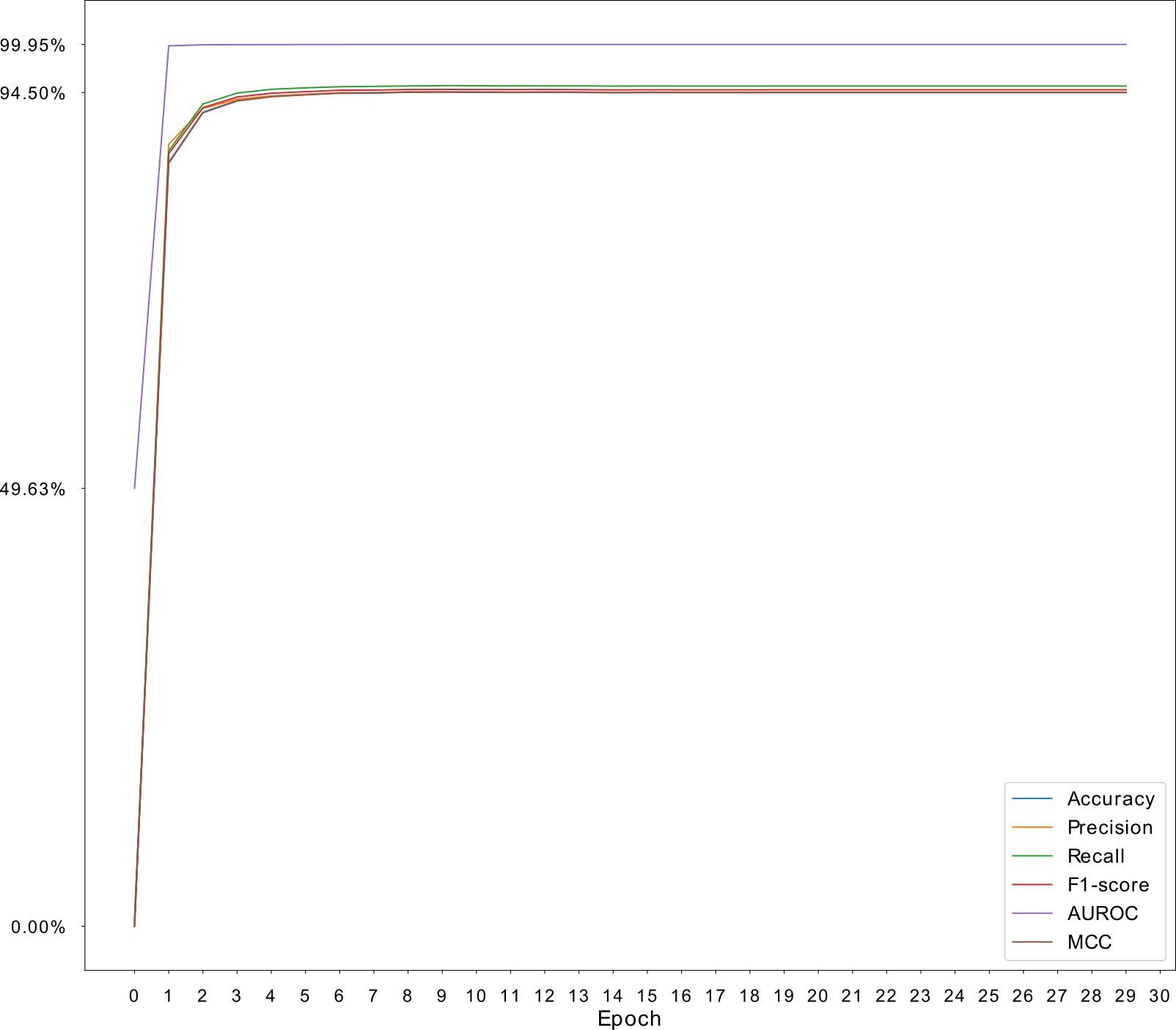


Fig. S2: Plot of accuracy, precision, recall, F1-Score, AUROC and MCC of the model evaluated on the validation set before each epoch. The model quickly found a stable local minimum with AUROC reaching 97% at the end of the first epoch. For this reason, the network was tested on the validation set once before training. The results of the untrained network are shown here as epoch 0. Many of the metrics, including MCC, reached approximately the same values, indicating that the steps taken to avoid the effects of class imbalance were effective.

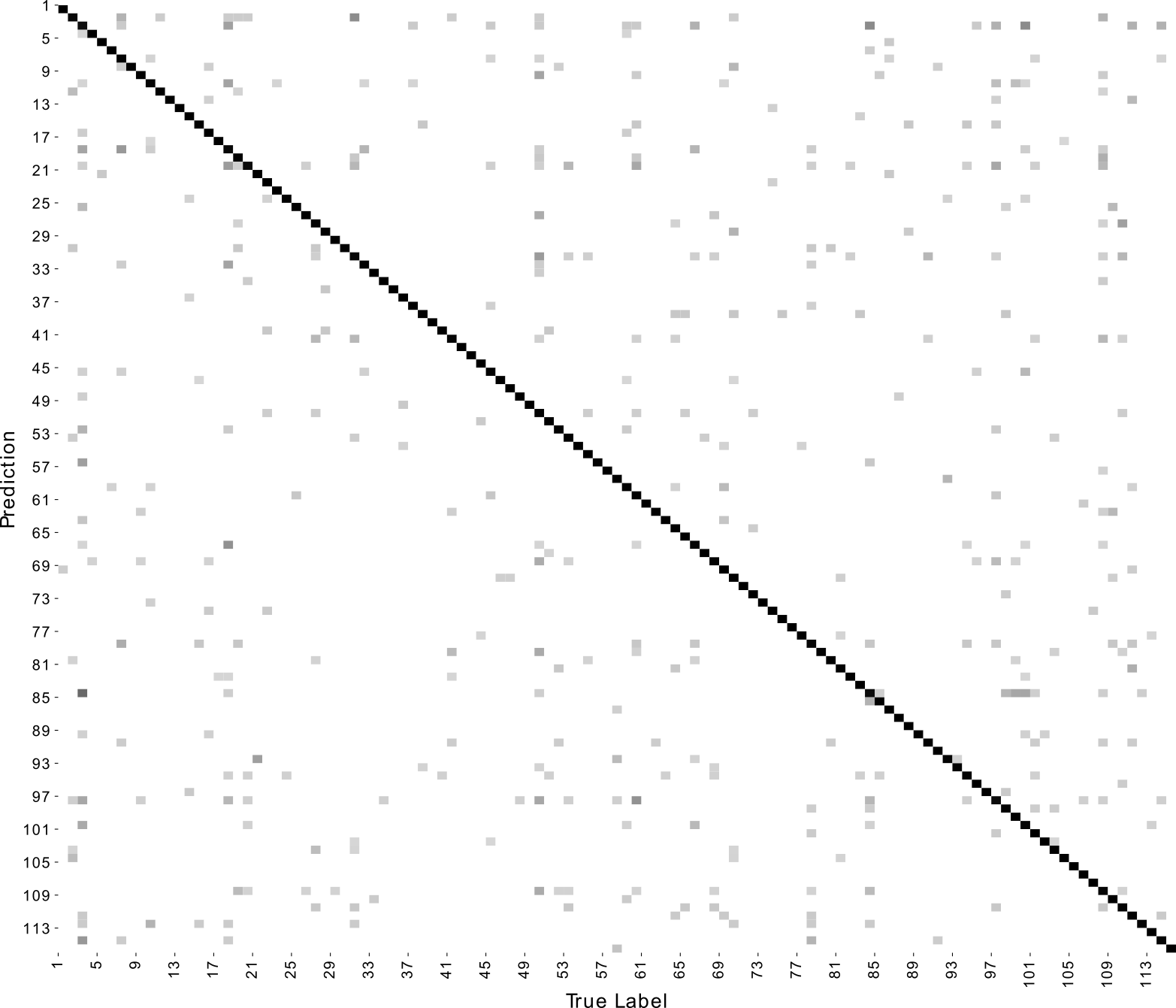


Fig. S3: Confusion matrix of the trained neural network. Darker cells represent a higher count than lighter cells. The cells are drawn on a logarithmic scale to make the lightest colored erroneous (off-diagonal) values more visible.

Heatmaps

Heatmaps for each of the resulting networks from the grid search are listed below. Each heatmap represents a different metric; these being loss, accuracy, precision, recall, F1-Score, are under the ROC curve and Matthews correlation coefficient. Each box represents a network with motif lengths within the range [min, max] where min is the corresponding length on the x-axis and max is on the y-axis. The diagonal represents networks containing motifs of the same length. Blue boxes represent the lowest values in the figure and red boxes represent the highest values.

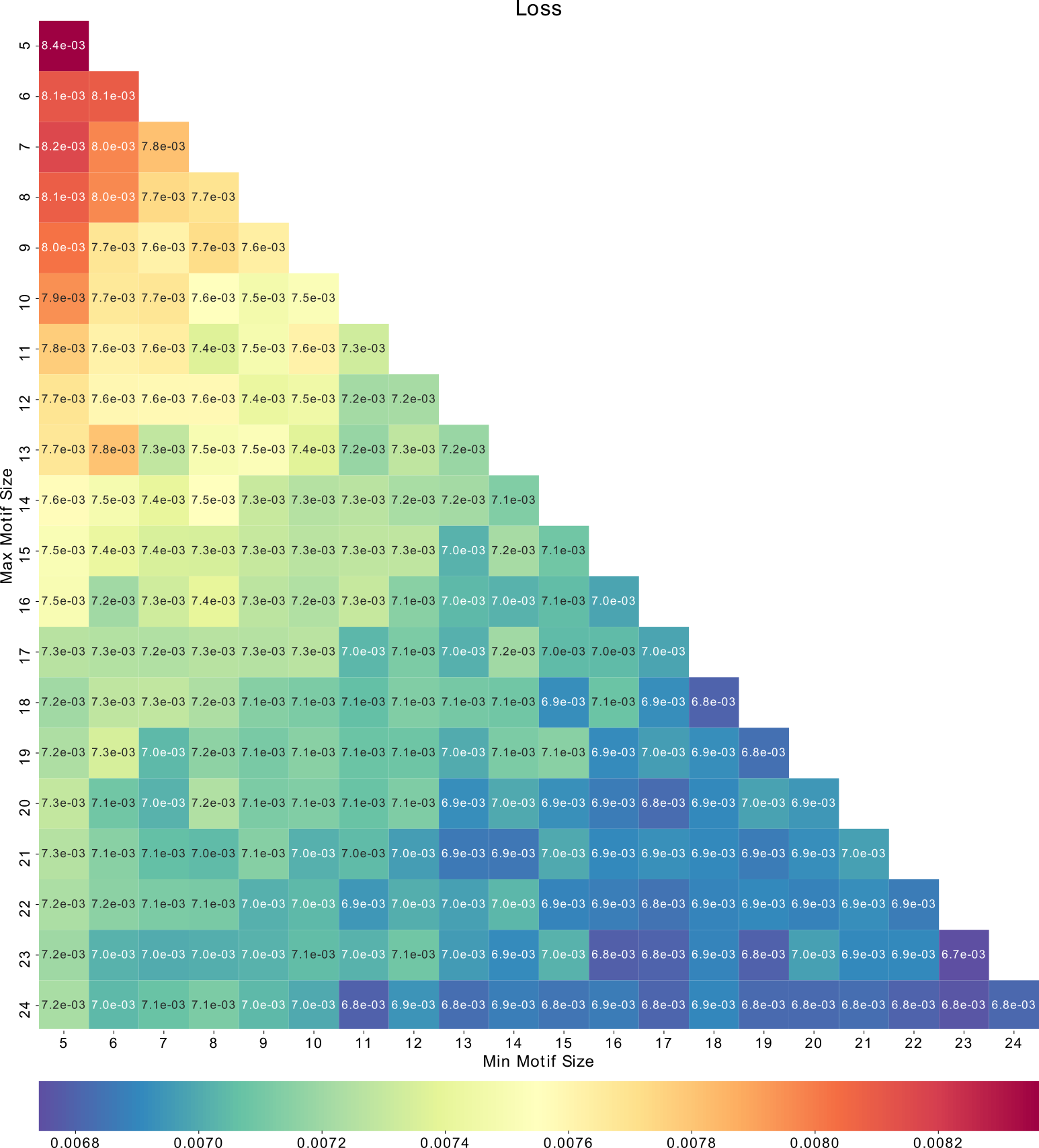


Fig. S4: Heatmap of the cross-entropy loss for each of the networks in the grid search.

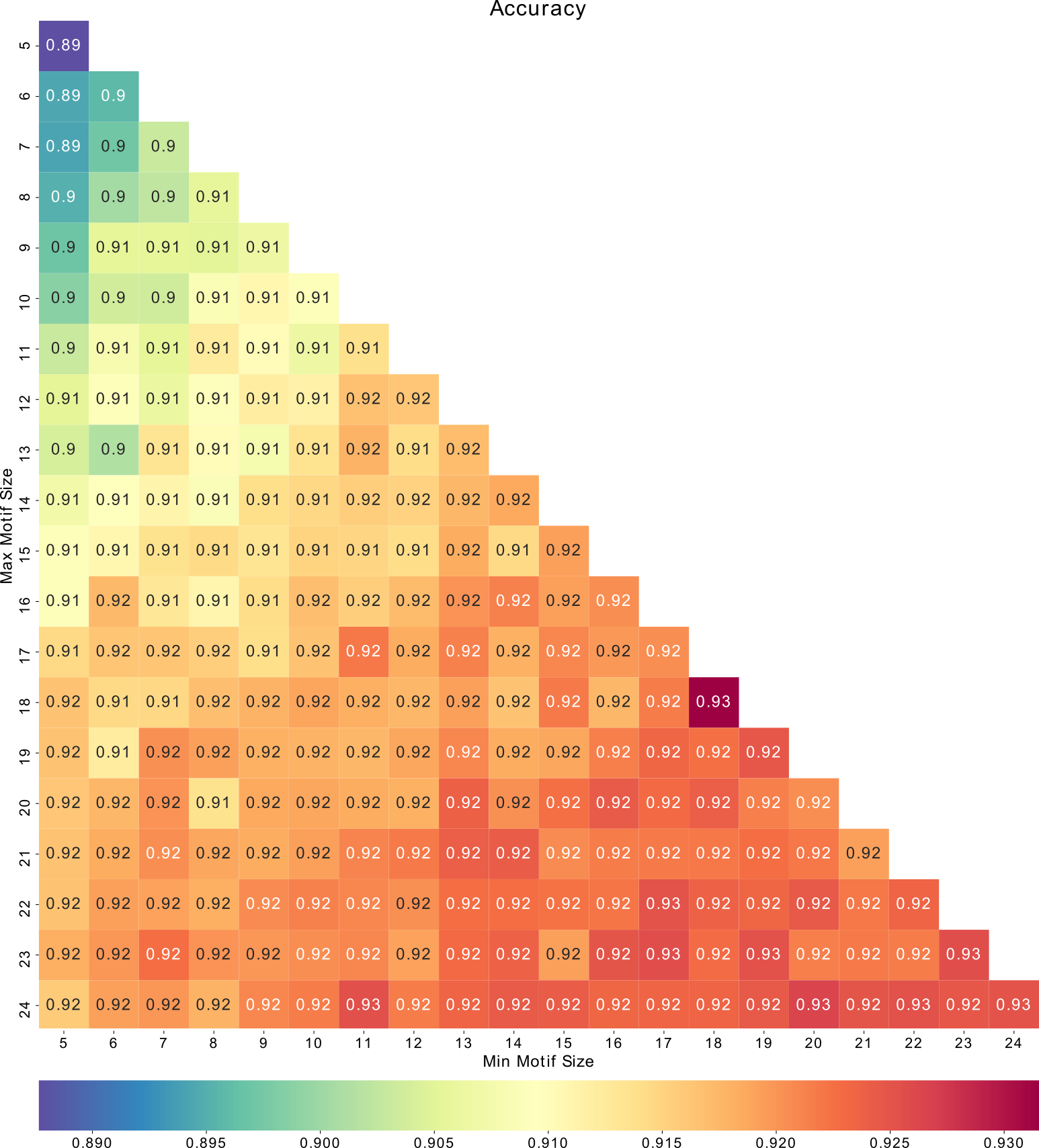


Fig. S5: Heatmap of the accuracy for each of the networks in the grid search.

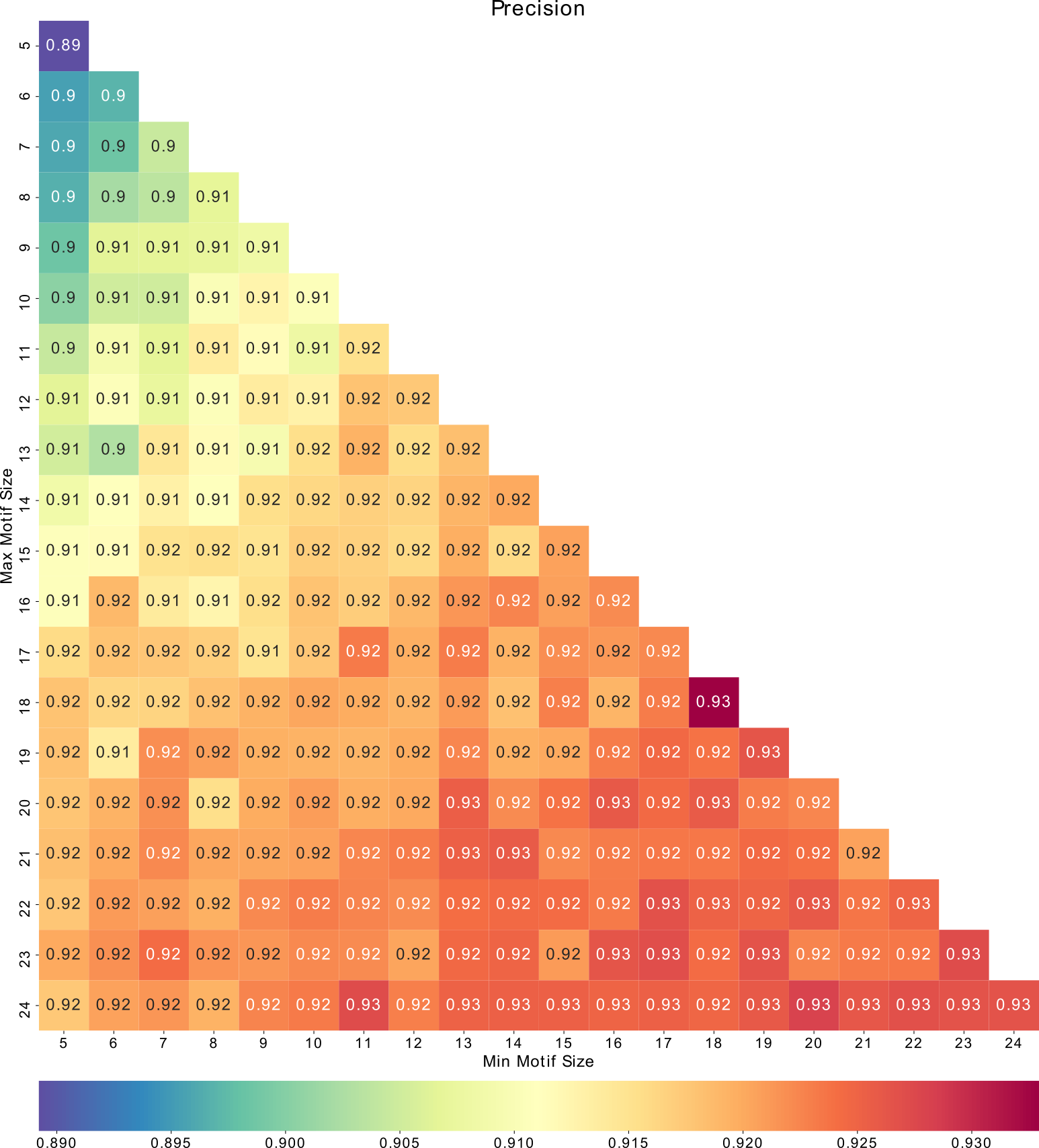


Fig. S6: Heatmap of the precision for each of the networks in the grid search.

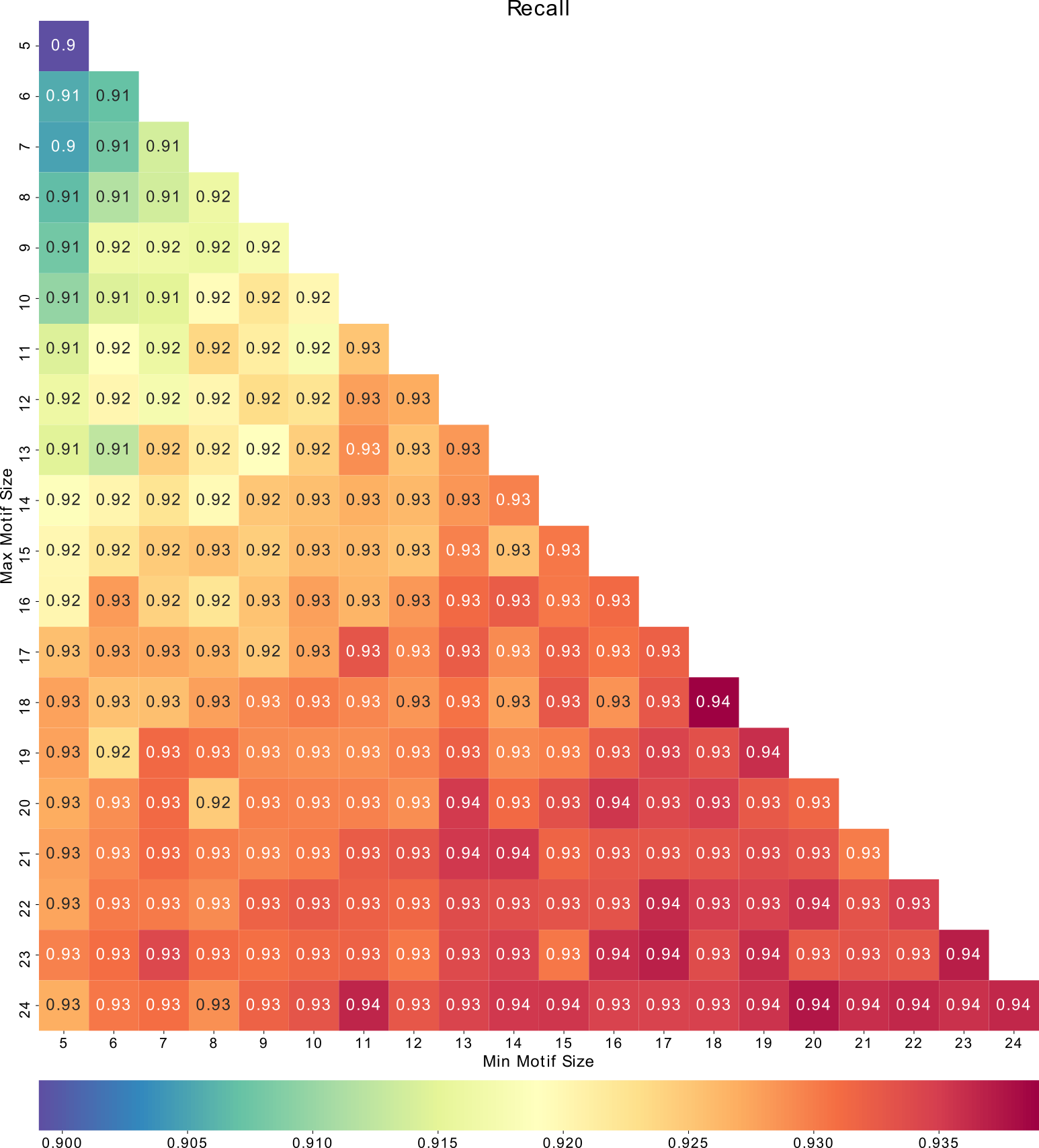


Fig. S7: Heatmap of the recall for each of the networks in the grid search.

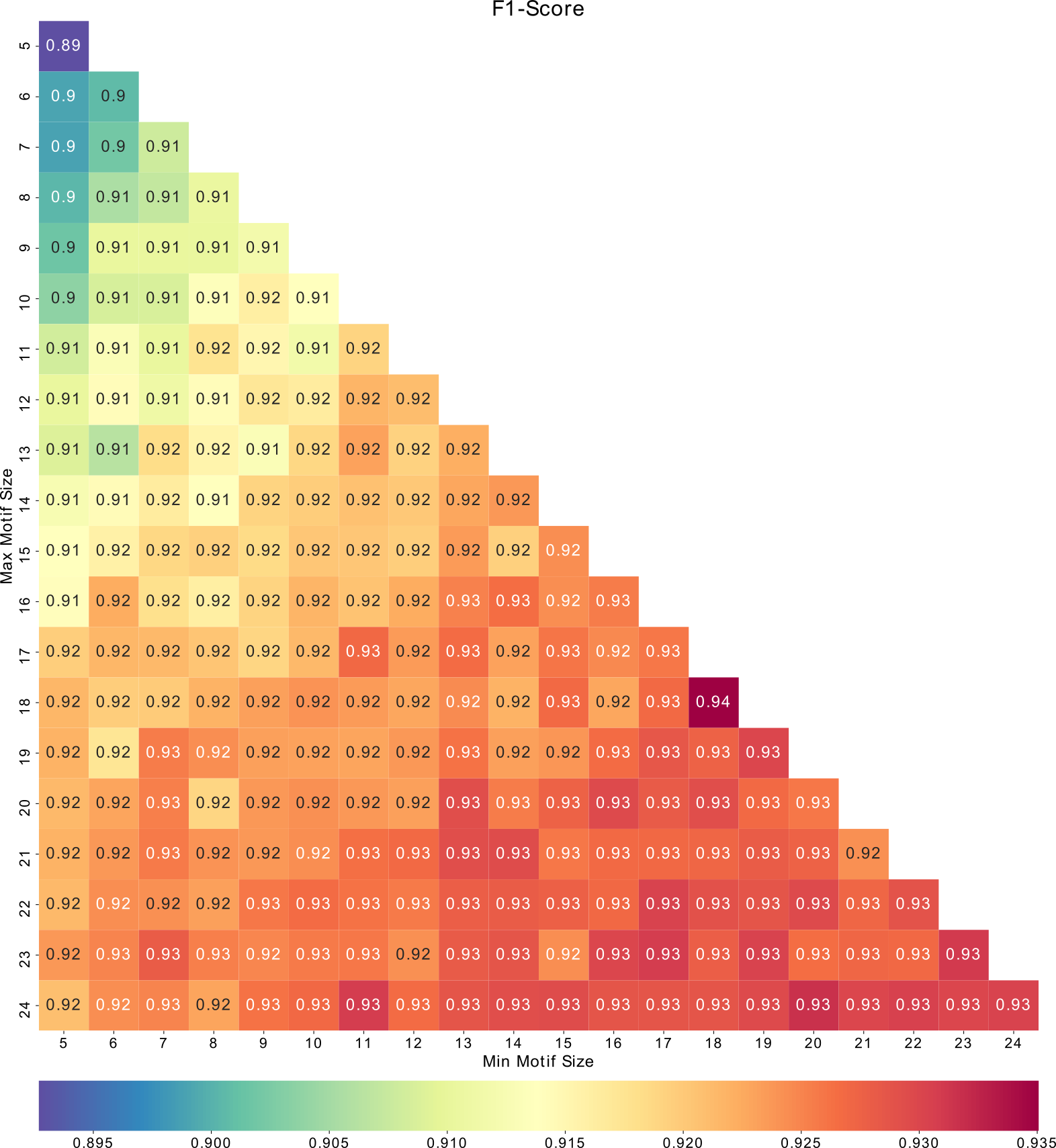


Fig. S8: Heatmap of the F1-Score for each of the networks in the grid search.

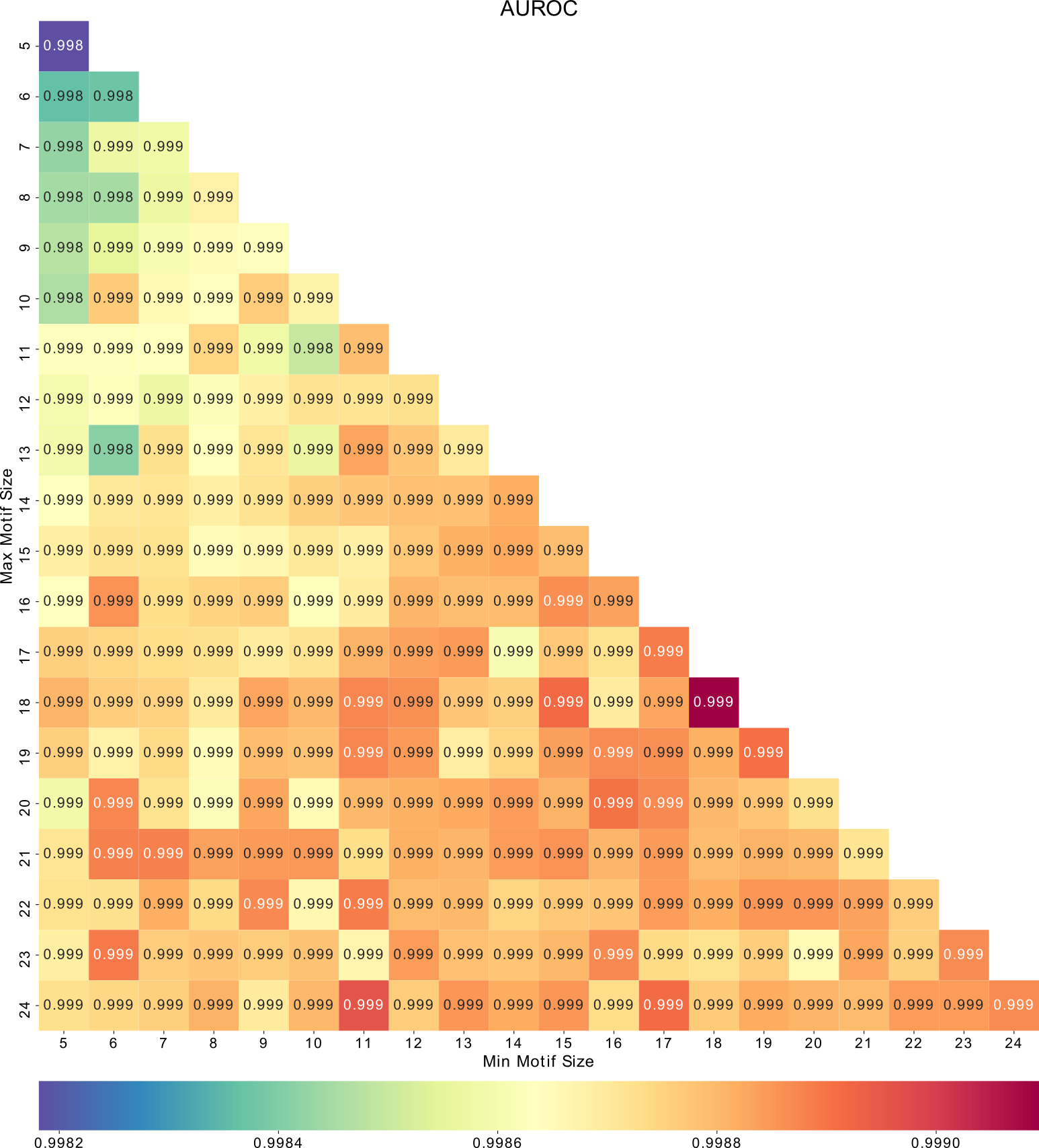


Fig. S9: Heatmap of the area under the receiver operator characteristic curve for each of the networks in the grid search.

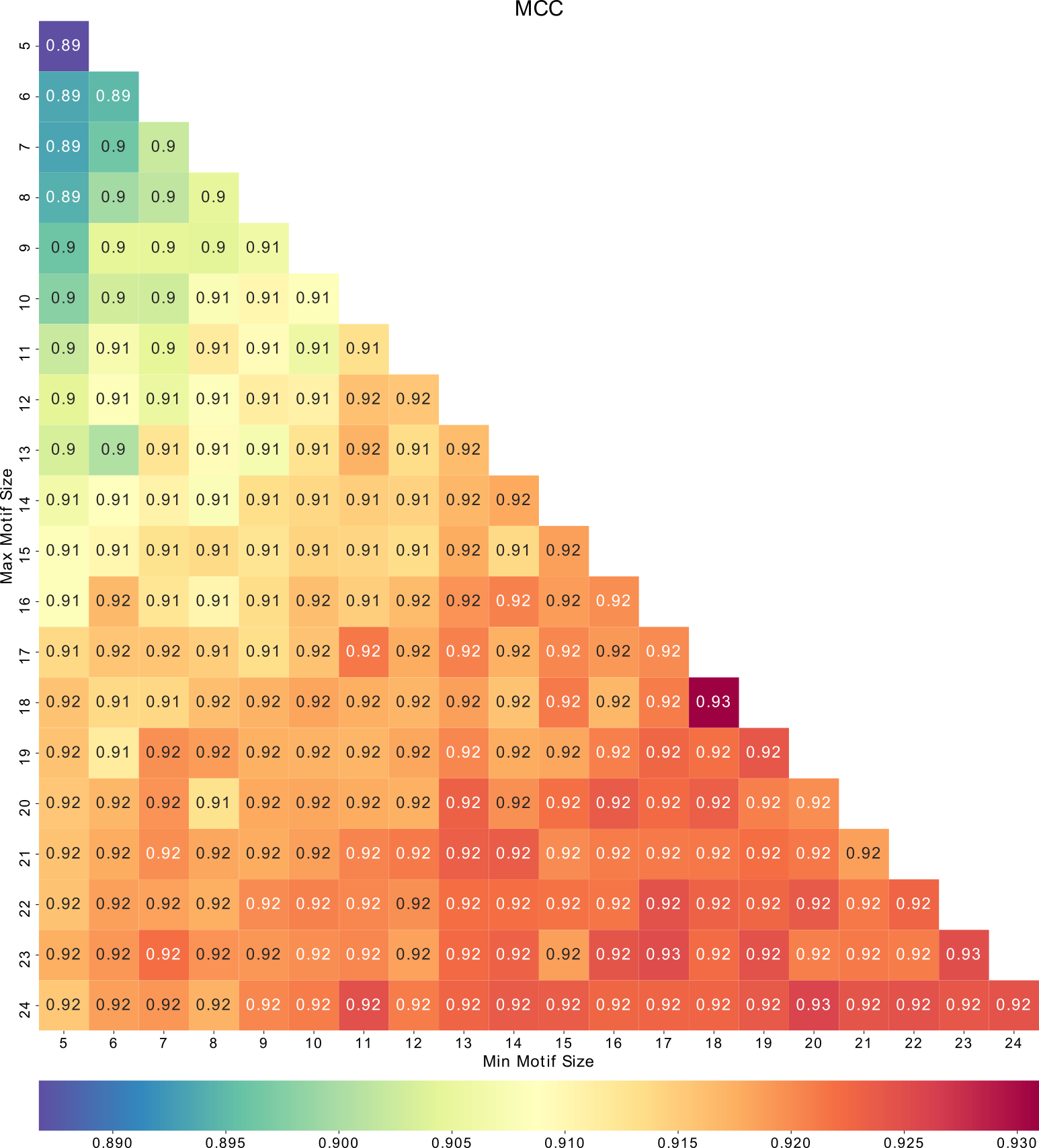


Fig. S10: Heatmap of the Matthews correlation coefficient for each of the networks in the grid search.

Logos

All SLiMs were extracted from the convolutional layer of the network using the method described by the DeepBind paper (Alipanahi et al., 2015). We then took the SLiMs, in PFM format, and used TomTom (Gupta et al., 2007) to compare them to known protein SLiMs. TomTom then calculated the e-value of each SLiM and created logos for those with the lowest e-value. These logos are shown below along with their associated e-values.

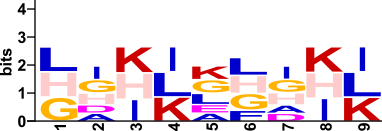


Fig. S11: Logo of a SLiM of length 9 discovered by the network. e-value = 0.0212009

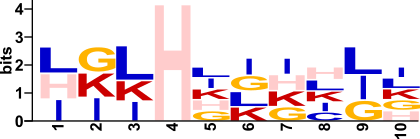


Fig. S12: Logo of a SLiM of length 10 discovered by the network. e-value = 0.000443802



Fig. S13: Logo of a SLiM of length 12 discovered by the network. e-value = 0.0360164



Fig. S14: Logo of a SLiM of length 14 discovered by the network. e-value = 0.127821



Fig. S15: Logo of a SLiM of length 16 discovered by the network. e-value = 0.282643



Fig. S16: Logo of a SLiM of length 16 discovered by the network. e-value = 0.0453992



Fig. S17: Logo of a SLiM of length 19 discovered by the network. e-value = 0.278652



Fig. S18: Logo of a SLiM of length 22 discovered by the network. e-value = 0.237427. It can be seen here that it is possible for a shorter motif to be detected within a longer filter. The pattern here is a repeating pattern of [ILKD][LHG][HKI][KLGE][GHI] with slight variations throughout.