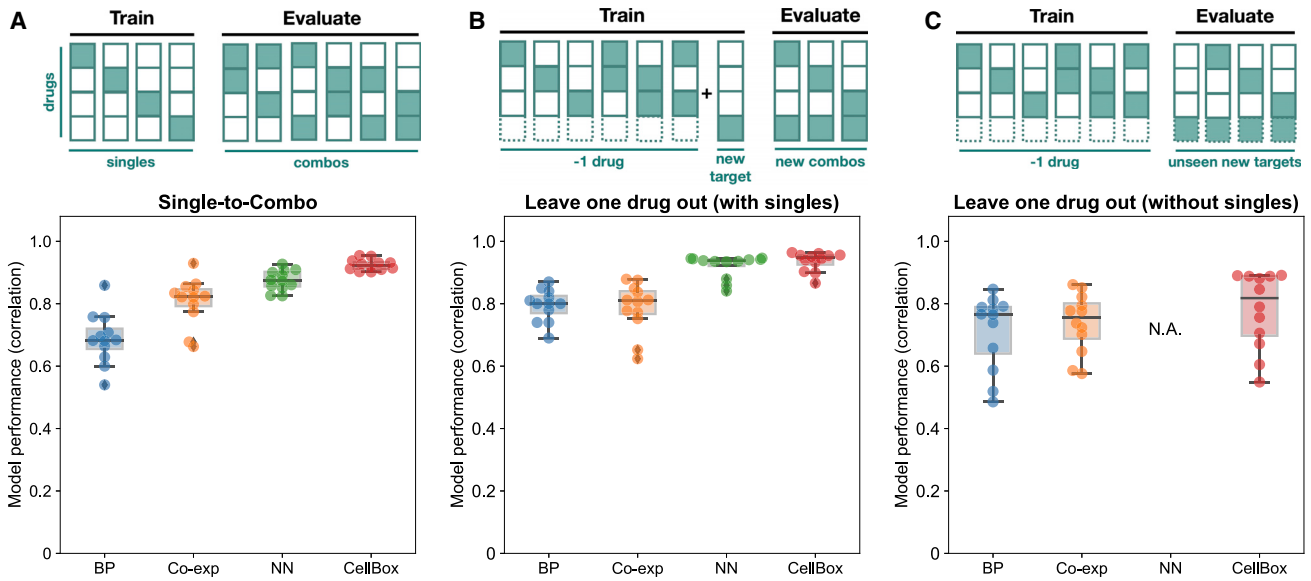

Figure 3. CellBox Can Accurately Predict Cell Response for Single-to-Combo and Leave-One-Drug-out Cross-Validations

(A) When only single conditions were used for training (single-to-combo), the CellBox models predict the effects of combinatorial conditions with high accuracy and outperform the dynamic network model inferred by using BP, the static co-expression network model (Co-exp), and a neural network regression model (NN) trained on the same data.

(B) When combinatorial conditions associated with one drug were withheld from training, the CellBox models retain high accuracy for predicting the effects of unseen drug pairs.

(C) When all conditions associated with one drug were withheld from training, the ODE network models predict the effects of the withheld drug with reduced accuracy, but direct-regression models such as NN cannot generalize to unseen targets at all. For each model type, the performance was evaluated by Pearson's correlation between predicted cell response and experimental cell response. The box charts indicate the group means and standard deviations.

been seen by the model, is a nontrivial challenge in the context of making accurate predictions of experimentally untested drug combinations.

In order to address these points, rather than training the model with random data partitioning, we instead designed more rigorous tasks: single-to-combo (Figures 3A and S7) and leave-one-drug-out cross-validation (Figures 3B and 3C) for each drug. In the single-to-combo analysis, all single-drug-treatment conditions were used for training, and predictions were made on all combinatorial drug conditions. In leave-one-drug-out cross-validation, all the combination conditions containing the treatment of a particular drug with or without the corresponding single-drug conditions were withheld and the rest of the conditions were used for training. In these more stringent tests, we found that the predicted values for withheld data were still highly correlated with the experimental observations (average Pearson's correlation: 0.93 for single-to-combo; 0.94 for leave-one-drug-out with single conditions, similar to that of the training with random partition; 0.79 for complete leave-one-drug-out). Under all three scenarios, on this dataset, CellBox outperforms the BP dynamic model approach previously used in perturbation biology (Korkut et al., 2015) in terms of predictive accuracy. These results indicate that the CellBox model can be trained with a relatively small set of perturbation data and that its predictions can be generalized to unseen combinatorial perturbations. In particular, CellBox models predict more accurately than linear models in the single-to-combo scenario (Figure S7), suggesting that CellBox can capture the nonadditive (synergistic or antagonistic) effects, which is particularly useful in nominating therapeutic drug combinations.

CellBox models are dynamic network models of a cell-biological system. To test whether such interpretable network models of molecular interactions help increase model predictive power, we compared the results to those of a static biological network model and a deep neural network model. The static network model was constructed by learning co-expression correlation for each pair of protein nodes (Co-exp) while the deep neural network model was trained to directly regress phenotypic changes against parameterized perturbations (NN) (STAR Methods). In all three tasks, the static network models had lower accuracy relative to the dynamic CellBox. The NN had comparable performance to CellBox in the cross-validation for individual drugs, but its performance dropped significantly in the single-to-combo analysis (Figure 3A). Furthermore, the NN was unable to generalize to unseen targets whose information is completely excluded from training (Figures 3C and S6). Altogether, because of the lack of mechanistic and dynamic information, static network or direct-regression models appear to be less suitable for facilitating the search for combinatorial targets.

Model Performance Is Robust against Noise and Reduced Training-Set Size

To examine model robustness of the CellBox models against a reduction in training data, we tested the stability of model performance when either the data quality or quantity is compromised. To test the former, we introduced different levels of multiplicative Gaussian noise (STAR Methods) into the input molecular and cellular response data and trained models on the resultant noisy datasets. The assumption behind such multiplicative noise is