Supplemental Data Items

Supplemental Figures

Supplemental Tables

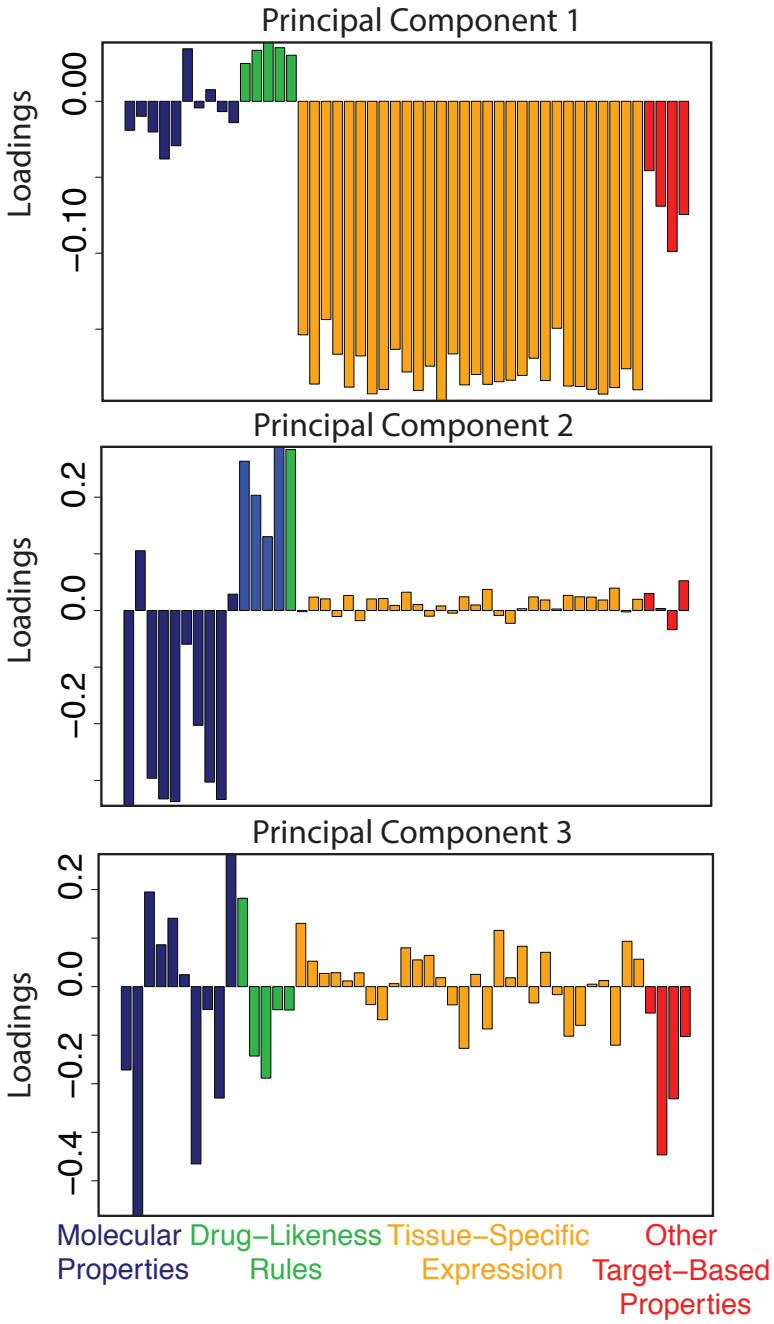


Figure S1, Related to Figure 3. Feature correlations identified by principal component analysis. Bar plots of the loadings for the first (top), second (middle), and third (bottom) principal components.

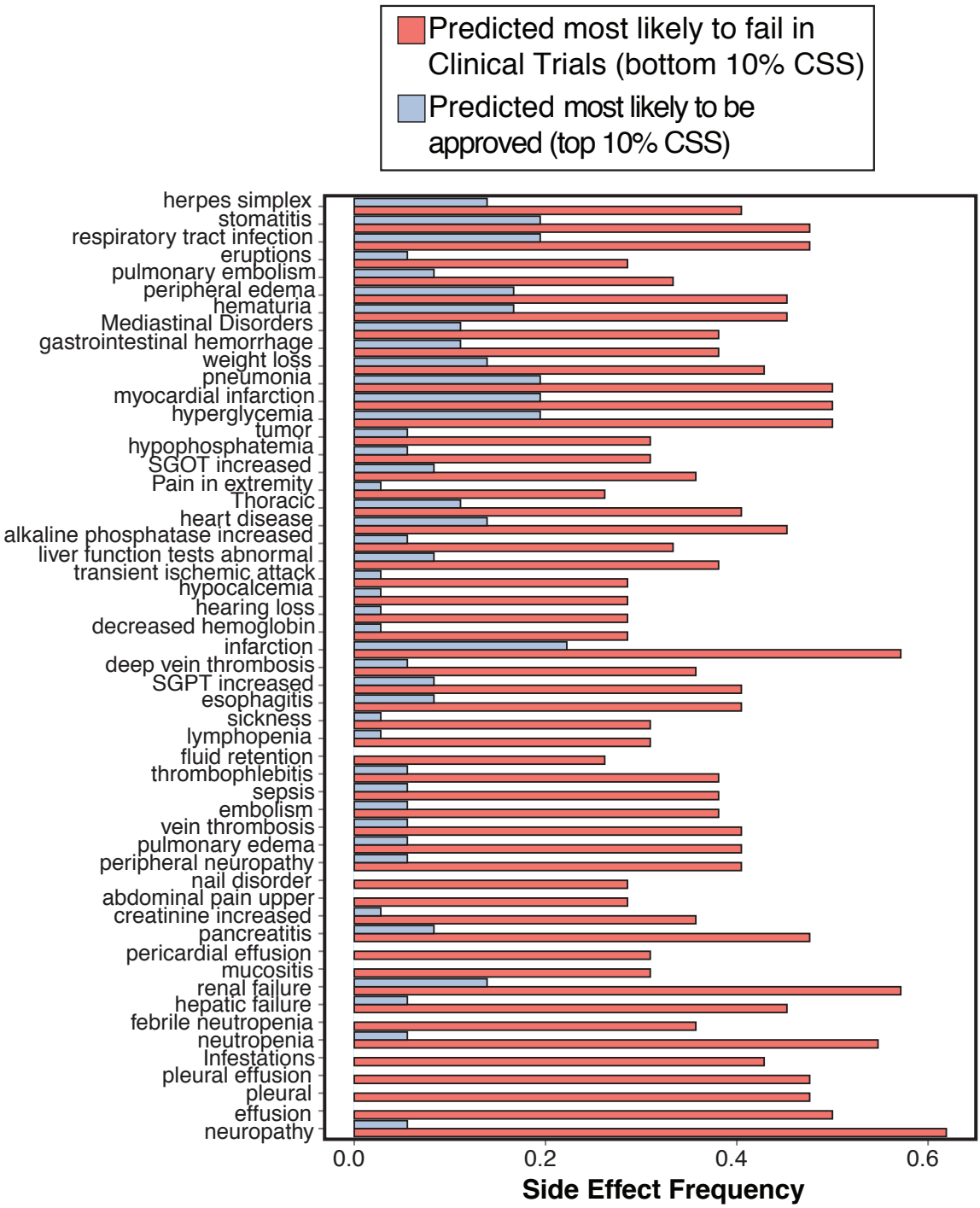


Figure S2, Related to Figure 4. Side Effects Frequency. The frequency of adverse for drugs within the bottom ten percentile compared to the top ten percentile of PrOCTOR scores.

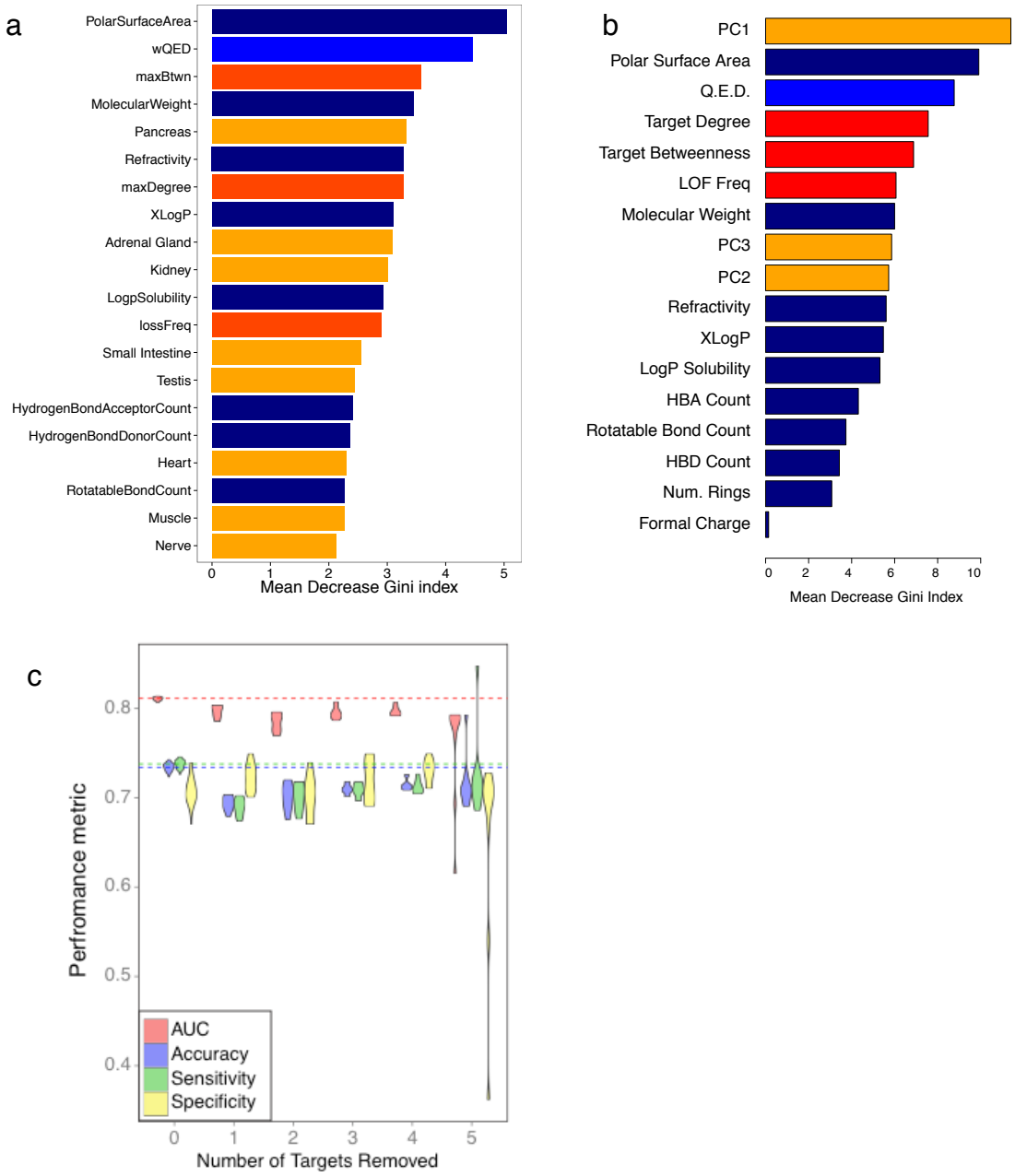
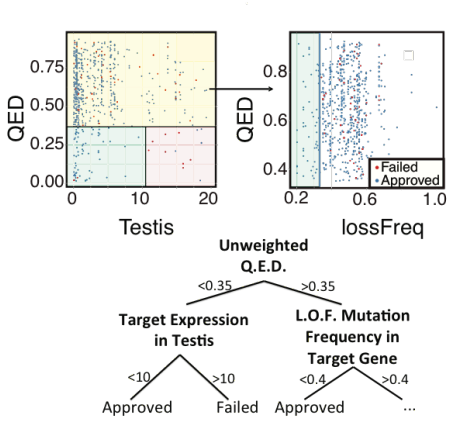


Figure S3, Related to Figure 3. Feature Importance and Model Robustness. Mean decrease Gini coefficient observed upon feature removal for the top 20 features **(a)** with all individual expression features and **(b)** with top 3 expression principle components instead of individual expression features. **(c)** Violin plots showing the range of AUC, accuracy, sensitivity and specificity for 0-5 targets removed.

a



b

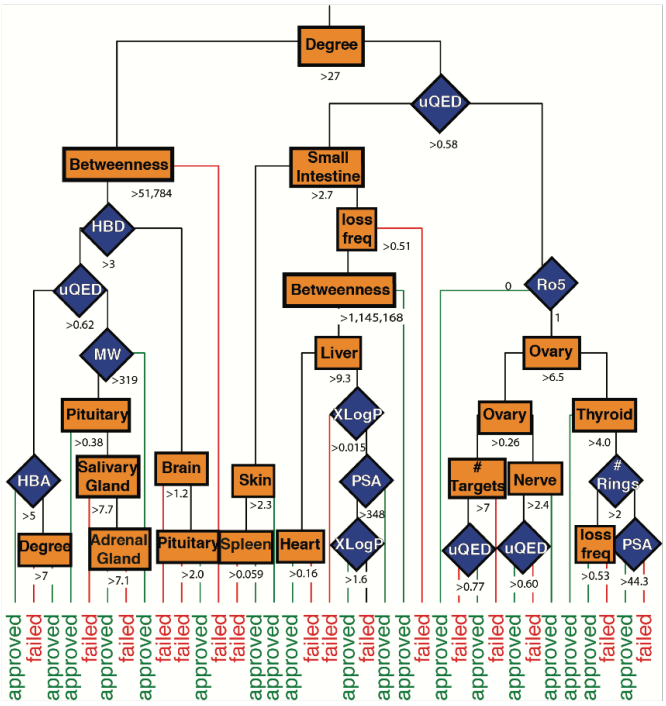


Figure S4, Related to Figure 3. Model Interpretation. **(a)** (top) unweighted Q.E.D. metric vs testis expression of drug targets. (middle) unweighted Q.E.D. metric vs frequency of loss of function mutations in target gene. (bottom) Sample simplified decision tree outputted by the method. **(b)** Consensus decision tree identified by the PrOCTOR model.

Supplementary Methods

Feature Importance Analysis

We measured the importance of each feature in the model by removing each feature and recording the AUC, accuracy, sensitivity and specificity for 20 replicates. The impact on performance was statistically quantified using the Wilcoxon signed rank test. Additionally, the R *randomForest* package (Liaw and Wiener, 2002) was used to measure the mean decrease in the Gini coefficient to further quantify the importance of each feature.

We found that many of the features within broad categories (eg. chemical properties, drug-likeness rules, expression in specific tissues, other target-based properties) were correlated by measuring pearson correlation coefficients and performing principal component analysis. We next removed these entire categories and statistically tested the impact on performance.

Target Removal Robustness Analysis

To investigate the effect of annotated targets on method performance, we measured the change in accuracy, sensitivity, specificity and area under the ROC curve (AUC). For n=1..5, we removed *n* randomly chosen targets from drugs with at least (n+1) targets. We repeated this ten times to get a range of values for each performance metric and used the Wilcoxon signed rank test to statistically quantify the impact on performance.

Supplementary Tool

In order to facilitate use and interpretation of our method, we have developed a tool for investigating model features and predictions. It is currently maintained on GitHub at **https://github.com/kgayvert/PrOCTOR.**

Supplementary Tool Visuals

Feature Quantile Plot

The feature quantile plot is a barplot that displays the quantile value of each feature for a selected molecule, defaulting to median values of each feature across the training set. Quantiles were chosen in order to visualize all variables on the same scale.

Structure/Target Feature Value Tables

Exact feature values, along with their corresponding quantile value, can be viewed in the **Structure and Feature Value** Table tabs.

Prediction Text and Plots

For the current molecule being displayed, the PrOCTOR model predictions, as well as target-only and structure-only model predictions. Additionally, these predictions are visualized with an arrow in comparison to the distributions of failed toxic trials and FDA approved drugs in the **Predictions tab**.

Loading features of an existing drug

While the visual defaults to median values for each feature, the features for a known drug can instead be loaded using the dropdown box on the bottom-right of the main screen

Changing feature values

The impact on how modifying a given molecule's features on PrOCTOR's predictions can be investigated by altering specific feature values. Feature values for the currently loaded drug can be changed in two ways:

1) Using a manually entered new value

2) Clicking directly on the barplot to set a new quantile value for a feature

In both cases, the input is used to update all plots, tables and prediction scores.

Update features using correlations (*default OFF).*

Since many of these features are not independent, an option is available for updating any correlated features along with inputted changes. Whenever a given feature *F* is changed, the values of all correlated features are updated as follows:

1) Correlated features are identified as those in the training set with Pearson correlation coefficient r>0.5.
2) A subset of the training set T_N is created by extracting all drugs that have *F* within 10% quantile range of the new value.
3) The new value for each correlated feature is set to be the median value of that feature within T_N.

After all feature values have been updated, all plots, tables and prediction values are updated accordingly.

References for Supplementary Material

Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. R News *2*, 18-22.