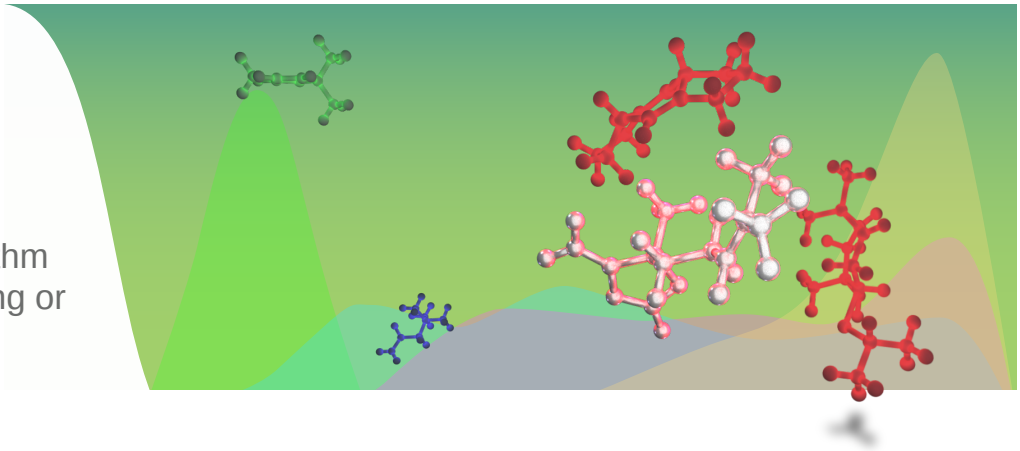
CYCLICA

Evaluating SWITCH$_X$ modulatory effect classification

SWITCH$_X$'s machine learning algorithm predicts ligand effects to be activating or inhibiting with 87% accuracy



PROBLEM

Few *in silico* technologies are capable of predicting the modulatory effect of a novel ligand-protein interaction

TECHNOLOGY

Ligand Express™ tools: SWITCH$_X$ Modulator Effect Prediction

SOLUTION

SWITCH$_X$ uses machine learning and data mining to predict the modulatory effect of a ligand on a novel protein target

INTRODUCTION

Many *in silico* drug development tools evaluate the likelihood of a small molecule ligand interacting with a target protein. However, once a probable interaction is identified, running *in vitro* experiments to determine its biological consequence can be costly and challenging. Insights into whether a predicted ligand-protein interaction is inhibitory or activating helps inform researchers about a compound's potential therapeutic benefit and helps identify compounds worth further evaluation.

Cyclica's SWITCH$_X$ is a cheminformatics tool that determines whether a small molecule ligand is likely to activate or inhibit a target, providing critical information for drug repurposing efforts, *de novo* drug discovery, or identifying a drug's mechanism of action. SWITCH$_X$ uses machine learning (ML) and natural language processing to combine structural data with published data to build a model specific to each protein. SWITCH$_X$ then uses this model to classify query ligands into the modulatory categories *activator*, *inhibitor*, and *unknown*.

METHODOLOGY

For comparison, we tested SWITCH$_X$ alongside three other classification methods:

1. Random selection with each of the four classes evenly weighted (**Even**)
2. Random selection with each of the four classes weighted based of the popularity of the classes from the aggregated data (**Weighted**)
3. Blind assignment to the most common class as determined by the aggregated data (**Common**)
4. SWITCH$_X$ machine learning algorithm

Each classification method was evaluated using a leave-one-out procedure[1]: for each protein, each ligand in the list of known interactions was removed and classified according to a model built from the remaining ligands. The percentage of correct classifications for each method was recorded, and the significance of SWITCH$_X$ results versus these methods was determined using the Mann-Whitney U test[2].

Table 1. A summary of the performance for the 4 different classification methods and the results from the Mann-Whitney U Test distinguishing SWITCH$_X$ from the other methods.

Method	Median(%)	IQR(%)	P-Value (compared to SWITCH$_X$)
Even	25.09	3.62	$6.53e^{-65}$
Weighted	59.57	35.31	$1.74e^{-26}$
Common	73.24	32.89	$1.25e^{-12}$
SWITCH$_X$	87.27	18.14	N/A

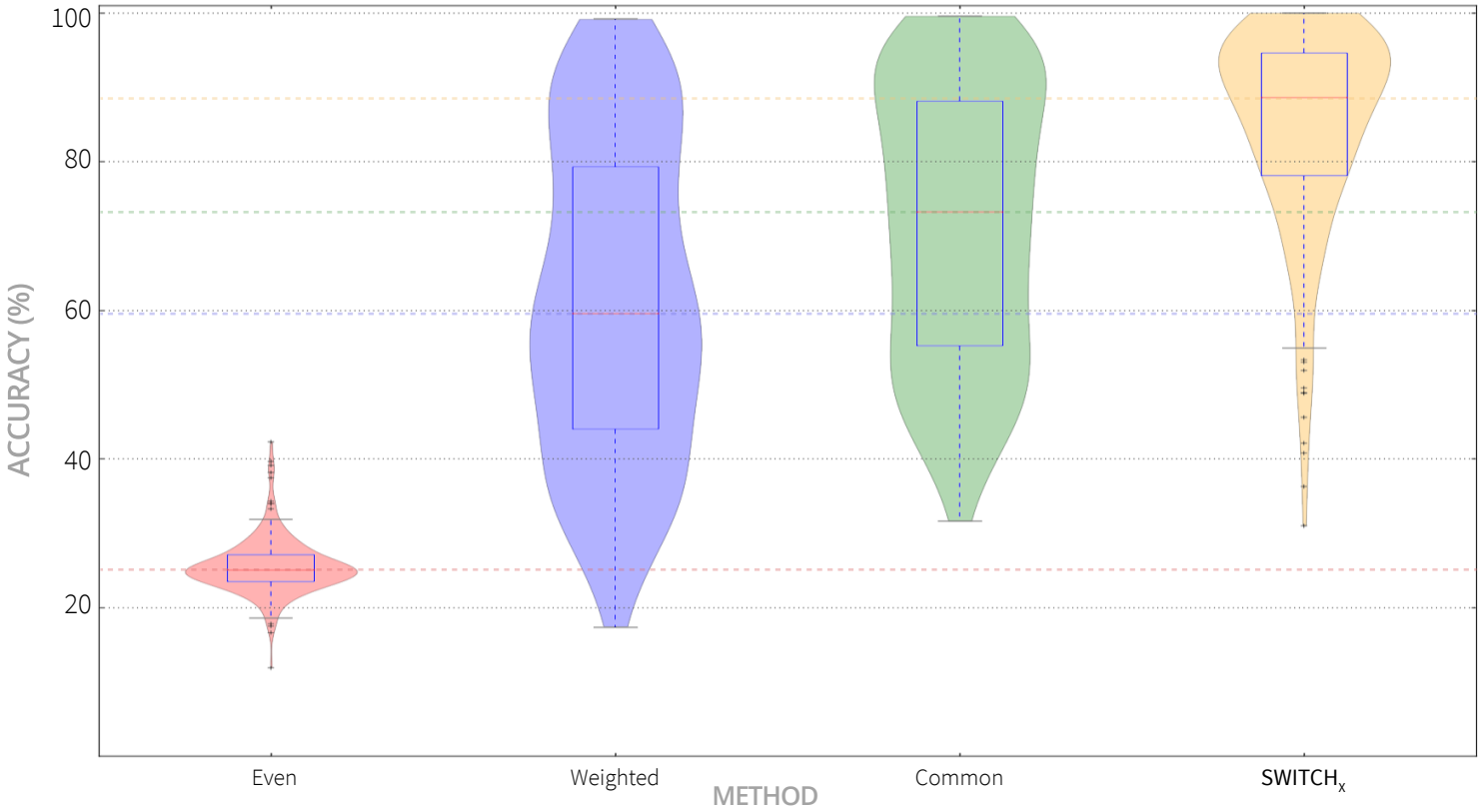


Figure 1. The accuracy and precision of different classifications methods visualized by a box-and-whisker plot overlaid on a violin plot.

To generate average results across a large number of cases, 201 proteins were selected with at least 20 known effectors and at least 7 effectors in 2 or more classes. The proteins selected represent a wide variety of functional protein types, demonstrating the application of $SWITCH_X$ across many domains.

RESULTS

Classification methods that used weighted averages (Weighted) and the most common-result (Common) performed better than completely random classification (Even), with medians of 59.57% and 73.24% respectively. $SWITCH_X$ classification showed a marked improvement over both of these approaches, with a median accuracy of 87.27%, and smaller interquartile ranges indicative of increased precision. **Table 1** contains a summary of these findings, along with the P-values indicating the significance of the SWITCHx improvement over each method. The P-values signifies that $SWITCH_X$ outperforms the other procedures. **Figure 1** visualizes the distributions of each classification method using a box-and-whisker plot overlaid on a violin plot.

SUMMARY

In predicting the modulatory effect of novel ligands, $SWITCH_X$ achieved a median accuracy of over 87% when tested using a leave-one-out method over ligands of known effect for 200 distinct proteins. Three other methods were also tested, and performed considerably worse. The results demonstrate that $SWITCH_X$ with a machine learning approach is capable of providing important pharmacological information that leads to improved insight into drug activity.

RESOURCES

1. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B* **36**, 111–147 (1974).

2. Mann, H. B. & Whitney, D. R. On a Test of Whether one of Tw o Random Variables is Stochastically Larger than the Other. *Ann. Math. Stat.* **18**, 50–60 (1947).

CYCLICA INC.

18 King St East, Suite 801
Toronto, Ontario, M5H 1A1, Canada
1-416-304-9201

101 Main Street, Floor 14
Cambridge, Massachusetts,
02014-1519, United States