Jul 12, 2019

GDSC data set and random forest model scripts

PLOS One

Peter Anderson[1]

[1]Physical Sciences Division, University of Washington Bothell

1 Works for me dx.doi.org/10.17504/protocols.io.3j9gkr6

Peter Anderson

ABSTRACT

This location contains the GDSC data set with 145 oncogene mutation statuses and ~1200 chemical descriptors. For instructions how to run binary classification (to predict compound activity vs inactivity) or regression (to predict values of log IC50), see the README.classification.txt and README.regression.txt files, respectively.

EXTERNAL LINK

https://doi.org/10.1371/journal.pone.0219774

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Lind AP, Anderson PC (2019) Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. PLoS ONE 14(7): e0219774. doi: 10.1371/journal.pone.0219774

MATERIALS TEXT

gdsc_oncogenes_descriptors.csv.gz
make_training_test_sets_classification.R
make_training_test_sets_regression.R
calc_stats_classification.R
run_all_classification.sh
run_all_regression.sh
calc_stats_regression.R
README.classification.txt
run_all_regression.sh
README.regression.txt