



Jul 12, 2019

GDSC data set and random forest model scripts [↗](#)

PLOS One

[Peter Anderson](#)<sup>1</sup><sup>1</sup>Physical Sciences Division, University of Washington Bothell *Works for me* [dx.doi.org/10.17504/protocols.io.3j9gkr6](https://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](https://doi.org/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](#)



This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited