JUL 12, 2019

OPEN ACCESS

DOI:
[dx.doi.org/10.17504/protocols.io.3j9gkr6](https://doi.org/10.17504/protocols.io.3j9gkr6)

External link:
<https://doi.org/10.1371/journal.pone.0219774>

Protocol Citation: Peter Anderson 2019. GDSC data set and random forest model scripts . **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.3j9gkr6>

MANUSCRIPT CITATION:
 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)

License: 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

Protocol status: Working

Created: May 31, 2019

GDSC data set and random forest model scripts

PLOS One

Peter Anderson¹

¹Physical Sciences Division, University of Washington Bothell

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.

MATERIALS

Last Modified: Jul 12, 2019

PROTOCOL integer ID:
23905

Keywords: GDSC data set;
random forests; cancer drug
sensitivity; machine learning;
personalized medicine