Predicting Drug Sensitivity

Your Name
Date

One of the ongoing challenges for the fields of personalized medicine is determining which genomic features will confer sensitivity or resistance to a drug. One way to study this is to use cancer cell lines from human tumors that have been genomically characterized. There are several publically available datasets that include gene expression, copy number, and mutation data for large numbers of cancer cell lines. Finding biomarkers that predict response to therapeutics would benefit many. One of the difficulties in finding these biomarkers is the complexity of biological pathways. Interactions among players in those pathways are often not captured in prediction algorithms.

I plan include three analyses:

1. The first is to validate comparisons of gene expression, copy number, and mutation data from two publicly available datasets: Cancer Cell Line Encyclopedia (CCLE) and Genomics of Drug Sensitivity in Cancer (GDSC).
2. The second is to evaluate existing methods of analyses. This will include validation of elastic net on GDSC.
3. The third is to improve upon current methods by including pathways biology in these analyses. One way to do this is to use gene sets derived from gene set enrichment analysis (GSEA).

Background

It has proven challenging to translate genomic data into therapeutic and biological knowledge. There were several initiatives to characterize large numbers of cancer cell lines, as well as initiatives to screen libraries of small molecules against hundreds of cancer cell lines to aid in the ability to predict sensitivity from genomic features. The Cancer Cell Line Encyclopedia (CCLE), which has characterized over 947 human cancer cell lines, was one such effort [1]. Cancer Target Discovery And Development (CTD^2) used a subset of CCLE cell lines to screen close to 500 compounds [2]. Genomics of Drug Sensitivity in Cancer (GDSC) was another large screening effort, including 138 anticancer drugs against almost 700 cell lines [3].

These datasets, as well as others in the field have used many different machine-learning algorithms, aimed at predicting drug response from genomic features. One effort to compare methods was the National Cancer Institute's (NCI) Dialogue on Reverse Engineering Assessment and Methods (DREAM) project, which analyzed 44 drug sensitivity prediction algorithms[4]. One of their findings was that application of biological pathways knowledge improved predictions. One way to incorporate biological pathways knowledge is to include gene sets and perform

gene set enrichment analysis[5]. Molecular signatures database (MSigDB) is a source of gene-sets that can be used[6].

Data Sources

I plan to use the following data sources:

CCLE
http://www.broadinstitute.org/ccle

Genomics of Drug Sensitivity in Cancer
http://www.cancerrxgene.org/downloads/

CTD^2 Data Portal
https://ctd2.nci.nih.gov/dataPortal/

Algorithms

Building predictive models:

Cforest
https://cran.r-project.org/web/packages/party/party.pdf

Elastic Net
https://cran.r-project.org/web/packages/elasticnet/elasticnet.pdf
or
https://cran.r-project.org/web/packages/glmnet/index.html

Gene set enrichment analysis
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239896/

References

[1]Barretina, Caponigro, Stransky et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012 Mar 28;483(7391):603-7. doi: 10.1038/nature11003.

[2]Rees MG, Seashore-Ludlow B, Cheah JH, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action.*Nature Chemical Biology Nat Chem Biol* 2015;12(2):109–116.

[3]Yang W, Soares J, Greninger P, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*. 2013;41(Database issue):D955-D961. doi:10.1093/nar/gks1111.

[4]Costello JC, Heiser LM, Georgii E, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nature biotechnology*. 2014;32(12):1202-1212. doi:10.1038/nbt.2877.

[5]Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.*Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102.

[6]Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260.