ToxicoGx: An R Platform for Integrated Toxicogenomics Data Analysis

Sisira Kadambat Nair [1] Esther Yoo [2] Christopher Eeles [3] Amy Tang [1] Nehme El-Hachem [1] Petr Smirnov [1, 2, 5] Benjamin Haibe-Kains [1, 2, 3, 4, 5]

[1] *Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada*

[2] *Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada*

[3] *Department of Computer Science, University of Toronto, Toronto, ON, Canada*

[4] *Ontario Institute of Cancer Research, Toronto, ON, Canada*

[5] *Vector Institute, Toronto, ON, Canada*

Plotting gene expression changes from TG-GATEs dataset

Our first case study is based on a plot from a paper by Rangel-Escareño et al., in which gene expression changes of CYP1A1 (gene associated with xenobiotic metabolism) has been plotted at all concentrations and time points. The plot shows clear differential expression at time 8(hr) suggesting that higher the dose, larger the impact of CCL4 on this gene. For plotting the gene expression under same conditions using the package, the first step is to load the datasets from disk or download them using the downloadTSet function above. In the following example, we use the toy dataset provided with the package to illustrate the process. To plot, the function drugGeneResponseCurve has been used wherein mandatory inputs such as dataset, drug name, cell-line, molecular type, gene name, dose and time points should be specified.

```
devtools::install_github("bhklab/ToxicoGx")
```

Plot time dependent dose response of Carbon tetra chloride on CYP1A1 gene

```
library(PharmacoGx)
library(ToxicoGx)
library(ggplot2)
library(tibble)
# Load the tset
data(TGGATESsmall)
ToxicoGx::drugGeneResponseCurve(TGGATESsmall,
                    duration = c("2", "8", "24"),
                    cell_line = "Hepatocyte", mDataTypes = "rna",
                    features = "ENSG00000140465_at",
                    dose = c("Control", "Low", "Middle", "High"),
                    drug = "Carbon tetrachloride",
                    ggplot_args = list(labs(title="Effect of Carbon tetra chloride on CYP1A1")),
                    summarize_replicates = FALSE
                    )
```

Effect of Carbon tetra chloride on CYP1A1
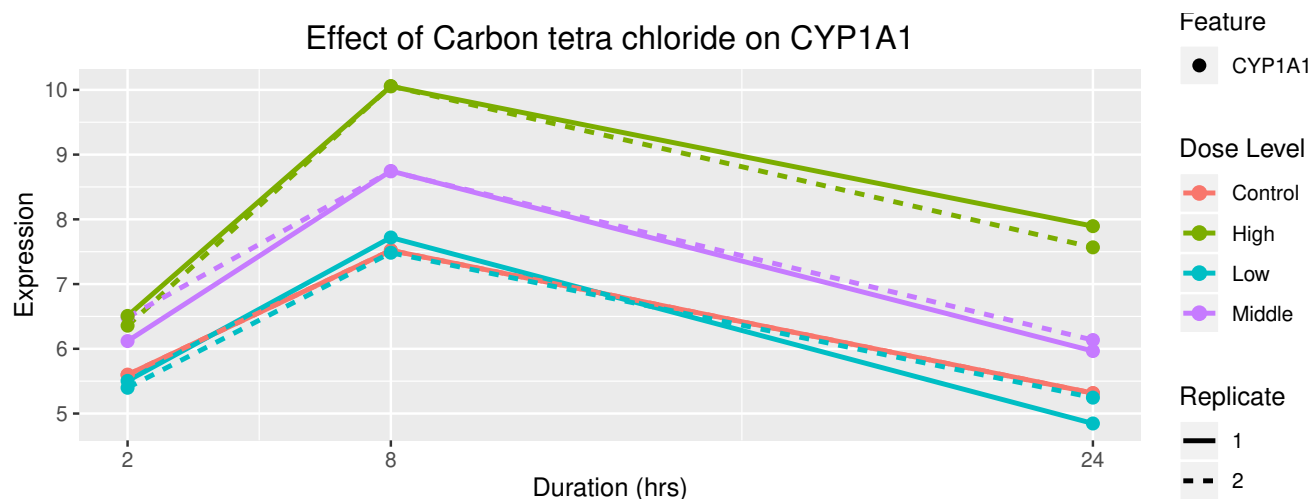
cytochrome P450, family 1, subfamily A, polypeptide 1 MB = 14.3 rank= 1
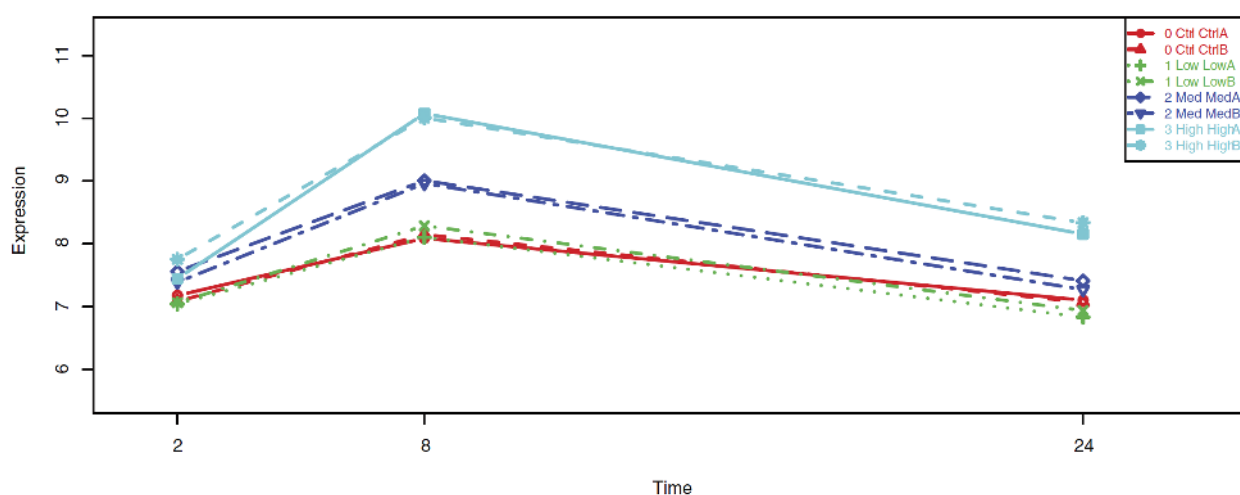
Connectivity map analysis on TG-GATEs and human hepatocarcinoma signatures

For the second case study, we will recreate an analysis from the paper by Jos Kleinjans et al., wherein connectivity mapping has been used to predict compound carcinogenicity by linking in vivo human hepatocarcinoma (HCC) signature geneset with in vitro TG-GATEs primary human hepatocyte data. In this example, we are using the toy dataset. The full dataset has to be downloaded to carry out the whole analysis done in the paper. The HCC signature, already mapped to the gene level, has been included in this package and it can be loaded by calling data(HCC_sig). Once the dataset is loaded, recreate drug signatures for each drug using the function drugPerturbationSig to peform statistical modelling of the transcriptomic response to the application of each drug. We then compare the observed up-regulated and down-regulated genes to HCC signature published in the paper. The output will be the GSEA connectivity score with FDR values that can be used to determine the correlation between the two signatures.

```
library(xtable)
data("TGGATESsmall")
# To compute the effect of drug concentration on the molecular profile of the cell
drug.perturbation <- ToxicoGx::drugPerturbationSig(tSet = TGGATESsmall,
                                    mDataType = "rna",
                                    cell_lines = "Hepatocyte",
                                    duration = "24",
```

```
                                        dose = c("Control", "Low"),
                                        drugs = c("Omeprazole", "Isoniazid"),
                                        returnValues=c("estimate","tstat", "pvalue", "fdr"),
                                        verbose = FALSE)
```

```
## Warning in .checkParamsForMissing(funName = funName, tSet = tSet, mDataType =
## mDataType, : features parameter not specified, defaults to all features in the
## given tSet for the specified mDataType!
```

```
data(HCC_sig)
res <- apply(drug.perturbation[,,c("tstat", "fdr")],
             2, function(x, HCC){
               return(PharmacoGx::connectivityScore(x = x,
                                         y = HCC[,2,drop = FALSE],
                                         method = "fgsea", nperm = 100))
             }, HCC = HCC_sig)
```

```
## Warning in PharmacoGx::connectivityScore(x = x, y = HCC[, 2, drop = FALSE], :
## GSEA method: query gene set (y) larger than signature (x)
```

```
## Warning in PharmacoGx::connectivityScore(x = x, y = HCC[, 2, drop = FALSE], :
## GSEA method: query gene set (y) larger than signature (x)
```

```
rownames(res) <- c("Connectivity", "P Value")
res <- t(res)
res <-  cbind(res,"FDR" = p.adjust(res[,2], method = "fdr"))
res <- res[order(res[,3]),]
xtable::xtable(res,
       caption = 'Connectivity Score results for HCC and TG-GATEs PHH gene
       signature')
```

% latex table generated in R 3.6.2 by xtable 1.8-4 package % Tue Jan 14 16:55:29 2020

	Connectivity	P Value	FDR
Omeprazole	0.00	1.00	1.00
Isoniazid	0.00	1.00	1.00

Table 1: Connectivity Score results for HCC and TG-GATEs PHH gene signature

In the above table, omeprazole showed a positive connectivity score in contrast to isoniazid. This observation aligns with the trends reported in the paper. The above example is to demonstrate the ease with which drug perturbation analysis can be done using ToxicoGx package.