Tutorial for TransPRECISE

Rupam Bhattacharyya

1/31/2020

Setup

In this tutorial, we will be illustrating how to use the TransPRECISE pipeline to run Bayesial graphical regression on patient/cell lines proteomic data and collect the outputs for cross-cancer comparison of the pathway activities and personalized drug response predictions.

```
# You need to install the remotes package for this to work.
# Also, if you encounter errors while installing PRECISE due to some packages
# not being avaiable in CRAN, it is possibly because they are now
# available via Bioconductor!

remotes::install_github("MinJinHa/PRECISE")
```

Running Bayesian graphical regression

For a given cancer type and pathway, the typical proteomics data would look someething like this in R. (This is a subset of the ACC patient samples, the pathway exhibited is apoptosis.)

```
load("Input.rda")
head(Final[1,1][[1]])
```

```
##                       SampleID TumorType        BAD     BAK1      BAX
## 1 TCGA-OR-A5J2-01A-21-A39K-20       ACC 0.15764346 1.814608 1.056160
## 2 TCGA-PA-A5YG-01A-21-A39K-20       ACC 0.26195648 1.777795 1.124701
## 3 TCGA-OR-A5JV-01A-21-A39K-20       ACC 0.03683437 1.699956 1.499859
## 4 TCGA-OR-A5JT-01A-21-A39K-20       ACC 0.47255555 1.841973 1.073022
## 5 TCGA-OR-A5JR-01A-21-A39K-20       ACC 0.43722031 1.634123 1.352207
## 6 TCGA-OR-A5JP-01A-21-A39K-20       ACC 0.82496783 1.495358 1.220990
##         BCL2     BCL2L1       BID    BCL2L11      CASP7     BIRC2
## 1 -0.9289668 -0.1386618 0.5251967  0.7515081 -0.6100740 0.8210888
## 2 -0.5204113 -0.6395635 0.2103295 -0.1142627 -0.7660185 0.7378661
## 3 -0.6588039 -0.6342769 0.3343039  0.9816016 -0.8706080 1.1755916
## 4 -0.8085059 -0.6573536 0.5753181 -0.4115302 -0.4455411 1.1914605
## 5 -0.9098263 -0.5628068 0.3114742 -0.1270832 -0.9736350 1.3100153
## 6 -1.1663160 -0.6762326 0.4607540 -0.5356534 -0.9721746 0.7302129
```

To fit the Bayesian graphical regression model and save the outputs for future, we used the following code.

```
#Load Package#
library(PRECISE)

#Load Environment for Data#
load("Input.rda")
```

```
#Pathway Headers#
pw.array=Pathways

#Load Data : Tumor x Pathway#
RPPAdat=Final[1,1][[1]]

Details=RPPAdat[,1:2]
RPPAdat=as.matrix(RPPAdat[,-(1:2)])
p=ncol(RPPAdat)

#Uniform Prior#
Gmat=matrix(0.5,p,p)
diag(Gmat)=0

#Final Response Vector + Covariate Matrix#
dat=getregcovDat(Gmat=Gmat,RPPAdat=RPPAdat)

#Fit Model Now#
bmsfit=getBMS(dat,Gmat)

#Get Tumor-Specific Pathway Network#
nodes=names(dat$ylist)
netfit=getPosteriors(bmsfit$outlist,nodes)
Network=netfit$G

#Get Betas for Tumor-Specific Pathway Network#
Coeffs=Network

for(i in 1:nrow(Coeffs))
{
  Coeffs[i,]=append(bmsfit$outlist[[i]][[1]]$b1mo,0,after=i-1)
}

#Create Patient-Specific Networks#
delta=0.5
psNet=getPRECISE(bmsfit$outlist,bmsfit$pdlist,nodes,delta)
Scores=cbind(Details,psNet$score.mat)

#Calibrate PRECISE Scores#
Status=cbind(Details,apply(psNet$score.mat,1,which.max))
```

A typical set of outputs would look like this.

```
load("Results.rda")
head(Scores[1,1][[1]])
```

```
##                       SampleID TumorType Netscore.pos Netscore.neg
## 1 TCGA-OR-A5J2-01A-21-A39K-20       ACC    11.643569     3.903753
## 2 TCGA-PA-A5YG-01A-21-A39K-20       ACC     8.936196     6.305865
## 3 TCGA-OR-A5JV-01A-21-A39K-20       ACC     9.117722     5.687658
## 4 TCGA-OR-A5JT-01A-21-A39K-20       ACC    13.916109     3.217110
## 5 TCGA-OR-A5JR-01A-21-A39K-20       ACC     7.878172     8.555711
```

```
## 6 TCGA-OR-A5JP-01A-21-A39K-20       ACC     3.818500     9.421719
##   Netscore.neu
## 1     9.452679
## 2     9.757938
## 3    10.194621
## 4     7.866781
## 5     8.566118
## 6    11.759781
```

```
head(Status[1,1][[1]])
```

```
##                       SampleID TumorType
## 1 TCGA-OR-A5J2-01A-21-A39K-20       ACC
## 2 TCGA-PA-A5YG-01A-21-A39K-20       ACC
## 3 TCGA-OR-A5JV-01A-21-A39K-20       ACC
## 4 TCGA-OR-A5JT-01A-21-A39K-20       ACC
## 5 TCGA-OR-A5JR-01A-21-A39K-20       ACC
## 6 TCGA-OR-A5JP-01A-21-A39K-20       ACC
##   apply(psNet$score.mat, 1, which.max)
## 1                                    1
## 2                                    3
## 3                                    3
## 4                                    1
## 5                                    3
## 6                                    3
```

```
Graphs[1,1][[1]]
```

```
##               BAD      BAK1       BAX      BCL2    BCL2L1       BID
## BAD     0.0000000 0.1436017 0.3110661 0.1331649 0.2351913 0.1658809
## BAK1    0.2457277 0.0000000 0.2737218 0.1650390 0.1380636 0.1889904
## BAX     0.2498290 0.1623932 0.0000000 0.1593484 0.1327064 0.1404851
## BCL2    0.1908690 0.1418573 0.1402384 0.0000000 0.2917065 0.9957838
## BCL2L1  0.2097344 0.1558544 0.1305769 0.3003959 0.0000000 0.1825189
## BID     0.2559377 0.2249883 0.1830516 0.9926423 0.2979844 0.0000000
## BCL2L11 0.1561535 0.1387443 0.2669947 0.1507761 0.2156587 0.1526220
## CASP7   0.8231443 0.9743107 0.1588454 0.9993448 0.1656261 0.9999999
## BIRC2   0.4014934 0.9897746 0.2153235 0.9448860 0.3435591 0.1662488
##            BCL2L11     CASP7     BIRC2
## BAD     0.1455839 0.2483370 0.2354568
## BAK1    0.1431972 0.8548833 0.9752209
## BAX     0.2671333 0.1307038 0.2123031
## BCL2    0.1898248 0.9994499 0.9013517
## BCL2L1  0.1992838 0.1428022 0.6068297
## BID     0.1484929 0.9999999 0.2231740
## BCL2L11 0.0000000 0.1333129 0.1755751
## CASP7   0.1438772 0.0000000 0.9281494
## BIRC2   0.1335392 0.8455353 0.0000000
```

Now, using these sets of saved results, one would be able to compare the fitted pathway networks across cancer types and sample types, and identify any differential/conserved patterns in the network circuitry using different metrics to compare networks. We used the networks to compute connectivity scores and quantify

significance via permutation tests. Codes for those are available in the subdirectory **Figure 3** under the **Figures** folder.

One may also use the pathway scores to compare samples across cancers and correlate cancer types or samples types based on the scores. We took two different approaches in terms of this. For a cross-pathway comparison of the samples, we used Pearson correlations, codes and resutls for which are available in the **Figure 4** subdirectory under **Figures**. We also used hierarchical clustering across all patients and sample types based on the scores, codes for which are available under **Figure 5**.