

JnJCodelinkAPI - package vignette

October 4, 2012

1 About the package

1.1 Method

The connectivity map approach is used to find connection strengths between query compounds and the reference DB.

The following processing is done when the reference DB is searched using a specific query compound:

- 1). For the query compound, t-test against control is done to compute p-values. Probes are filtered based p-value and fold-change thresholds.
- 2). Probes that pass the filtering criteria are ranked based on their fold changes.
- 3). Sum-of-product is calculated for each Query-Reference pair. This gives "Connection Strength".
- 4). "Connection Strength" is divided by "Max. Connection Strength" to get "Connection Score"

2 How to use

2.1 Data API

```
> ##### Loading required libraries ##  
> library(JnJCodelinkAPI)  
> ##### Listing database compounds ##  
> comp <- listCompounds()  
> head(comp)
```

```
$Amitriptyline  
[1] "Dose:150; Dur:1d" "Dose:150; Dur:4d"
```

```
$Amiodarone  
[1] "Dose:300; Dur:4d" "Dose:1000; Dur:1d"
```

```
$Adrenocorticotropin  
[1] "Dose:0.5; Dur:1d"
```

```
$Aminoglutethimide  
[1] "Dose:200; Dur:1d"
```

```
$Cholecalciferol  
[1] "Dose:20; Dur:1d"
```

```
$Spironolactone  
[1] "Dose:800; Dur:1d"
```

```

> ##### Load reference data #####
> ### Gene Level Data ###
> data(geneLevelData)
> class(geneLevelData)

[1] "ExpressionSet"
attr(,"package")
[1] "Biobase"

> geneLevelData

ExpressionSet (storageMode: lockedEnvironment)
assayData: 9212 features, 605 samples
  element names: exprs
protocolData: none
phenoData
  sampleNames: AA|b1|AA13|Vehicle||4d|Oral AA|b1|AA14|Vehicle||4d|Oral
  ... DH|b1|DH28|Chloramphenicol|1500|1d|Oral (605 total)
  varLabels: BatchID LotID ... Duration (7 total)
  varMetadata: labelDescription
featureData
  featureNames: 309082 94201 ... 24383 (9212 total)
  fvarLabels: EntrezID Gene_Symbol FeatureID CODELINK_ID
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation:

> ### Probe Level Data ###
> data(probeLevelData)
> class(probeLevelData)

[1] "ExpressionSet"
attr(,"package")
[1] "Biobase"

> probeLevelData

ExpressionSet (storageMode: lockedEnvironment)
assayData: 35129 features, 605 samples
  element names: exprs
protocolData: none
phenoData
  sampleNames: AA|b1|AA13|Vehicle||4d|Oral AA|b1|AA14|Vehicle||4d|Oral
  ... DH|b1|DH28|Chloramphenicol|1500|1d|Oral (605 total)
  varLabels: SampleName BatchID ... Toxicity (10 total)
  varMetadata: labelDescription
featureData
  featureNames: 1001 1002 ... 328112 (35129 total)
  fvarLabels: CODELINK_ID NCBI_ACC ... ENTREZ_ID (5 total)
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation: rwgcod

> ##### Download expression data ###
> ### For Single compound ###
> expData <- getExpressionData(Compound="Phenytoin", Type="ProbeLevel")
> expData

```

```

ExpressionSet (storageMode: lockedEnvironment)
assayData: 35129 features, 9 samples
  element names: exprs
protocolData: none
phenoData
  sampleNames: AJ|b1|AJ04|Phenytoin|20|1d|Oral
               AJ|b1|AJ05|Phenytoin|20|1d|Oral ...
               AJ|b1|AJ12|Phenytoin|225|1d|Oral (9 total)
  varLabels: SampleName BatchID ... Toxicity (10 total)
  varMetadata: labelDescription
featureData
  featureNames: 1001 1002 ... 328112 (35129 total)
  fvarLabels: CODELINK_ID NCBI_ACC ... ENTREZ_ID (5 total)
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation: rwgcod

```

```

> expData <- getExpressionData(Compound="Gabapentin", Dose=c("1000", "2160"), Duration="1d", Type="G")
> expData

```

```

ExpressionSet (storageMode: lockedEnvironment)
assayData: 9212 features, 6 samples
  element names: exprs
protocolData: none
phenoData
  sampleNames: AD|b8|AD082|Gabapentin|2160|1d|Oral
               AD|b8|AD083|Gabapentin|2160|1d|Oral ...
               AJ|b3|AJ42|Gabapentin|1000|1d|Oral (6 total)
  varLabels: BatchID LotID ... Duration (7 total)
  varMetadata: labelDescription
featureData
  featureNames: 309082 94201 ... 24383 (9212 total)
  fvarLabels: EntrezID Gene_Symbol FeatureID CODELINK_ID
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation:

```

```

> ### For Multiple compounds ###
> expData <- getExpressionData(Compound=c("Gabapentin", "Flufenamic acid", "Metoprolol", "Digoxin"),
> expData

```

```

ExpressionSet (storageMode: lockedEnvironment)
assayData: 9212 features, 18 samples
  element names: exprs
protocolData: none
phenoData
  sampleNames: AD|b8|AD082|Gabapentin|2160|1d|Oral
               AD|b8|AD083|Gabapentin|2160|1d|Oral ...
               AJ|b3|AJ42|Gabapentin|1000|1d|Oral (18 total)
  varLabels: BatchID LotID ... Duration (7 total)
  varMetadata: labelDescription
featureData
  featureNames: 309082 94201 ... 24383 (9212 total)
  fvarLabels: EntrezID Gene_Symbol FeatureID CODELINK_ID
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation:

```

```

> ##### View annotations #####
> geneLevelAnnotation <- getAnnotation(dataType="GeneLevel")
> head(geneLevelAnnotation)

      EntrezId  Symbol FeatureId CodelinkId
309082  309082   Clrn3     67071  GE1100063
94201   94201    Cdk4     206040 GE1100112
83583   83583   Rabac1    188041 GE1100133
297383  297383    Rtkn     293017 GE1100158
312451  312451    Tcf3      29106  GE1100219
170566  170566  Slc16a10   234078  GE1100230

> probeLevelAnnotation <- getAnnotation(dataType="ProbeLevel")
> head(probeLevelAnnotation)

      CodelinkId Accession UnigeneId      Symbol EntrezId
1001  GE200017      <NA>      <NA>      <NA>      NA
1002  GE1275839  CB717085  Rn.24338 Large_predicted 361368
1004  GE1204628  CB613942  <NA>      <NA>      NA
1005  GE1289537  BF398129  Rn.7881    LOC681415  681415
1006  GE1250992  BG377995  Rn.107863 Surf4      619346
1007  GE200017      <NA>      <NA>      <NA>      NA

>

```

2.2 Connectivity Score API

```

> ##### Set parameters #####
> p.val <- 0.1
> fc <- 1.5
> ##### Load ranked reference database #####
> data(rankedRefDB)
> rankedRefDB[1:4,1:4]

      AA|b1|AA04|Amitriptyline|150|1d|Oral
309082      2765.0
94201     -4244.5
83583     -6486.5
297383      2294.5
      AA|b1|AA05|Amitriptyline|150|1d|Oral
309082     -5316.5
94201      6064.0
83583     -3276.0
297383      6050.0
      AA|b1|AA06|Amitriptyline|150|1d|Oral
309082      2141.0
94201     -4352.5
83583     -6776.0
297383      1411.0
      AA|b1|AA07|Amitriptyline|150|4d|Oral
309082      1572.5
94201      4570.0
83583     -2973.0
297383      2135.5

> #####Load query example #####
>

```

```
> data(sampleQuery)
> queryData <- sampleQuery # A custom query can also be created, as shown below
> head(queryData)
```

```
      logFC      pValue
24152 -0.02305170 0.7841735935
24153 -0.27953432 0.6027071296
24157  0.37126152 0.0047984685
24158 -0.02462885 0.7103978130
24159  0.98900578 0.0001592612
24162  0.08474472 0.2933326031
```

```
> ##### Read & process query data #####
> queryData <- read.table(file="WY24hrquery.txt", sep="\t", header=T) # Replace the file name with y
> queryData <- as.matrix(queryData)
> rownames(queryData) <- queryData[,1] # First column contains Entrez ids, set as rownames
> queryData <- queryData[,2:3]
> head(queryData)
```

```
      fc      p.val
171142 3.359879 2.33e-26
29740  2.394200 1.16e-25
50549  2.713894 2.18e-25
192272 2.946369 4.48e-25
64526  2.981228 1.53e-24
292486 4.280572 2.82e-24
```

```
> ##### Computing connection scores using the wrapper function #####
> cs <- connectivityCalculator(query=queryData, fc=1.5, p.val=0.1, limitProbes=NULL)
> head(cs$connectivityScores, 5) # 5 Most similar samples
```

```
      ConnectionScores
AG|b1|AG06|Benzafibrate|500|1d|Oral      0.7797052
AG|b1|AG04|Benzafibrate|500|1d|Oral      0.7765758
AG|b1|AG05|Benzafibrate|500|1d|Oral      0.7565615
AZ|b1|AZ04|Fenofibrate|500|1d|Oral      0.6834902
AE|b2|AE25|Perfluorooctanoate|150|1d|Oral 0.6800971
```

```
> tail(cs$connectivityScores, 5) # 5 Least similar samples
```

```
      ConnectionScores
AE|b3|AE54|Tacrine|50|1d|Oral      -0.5086006
AD|b2|AD022|Captopril|5000|1d|Oral  -0.5410676
AE|b3|AE53|Tacrine|50|1d|Oral      -0.5645552
AE|b3|AE52|Tacrine|50|1d|Oral      -0.5983422
AD|b2|AD023|Captopril|5000|1d|Oral  -0.6042862
```

```
> length(cs$absentProbes) # Query probes not found in Reference
```

```
[1] 283
```

```
> head(cs$absentProbes)
```

```
[1] "292486" "85249" "29651" "29726" "79131" "361637"
```

For a given query, the connection score with the reference samples show following trend.

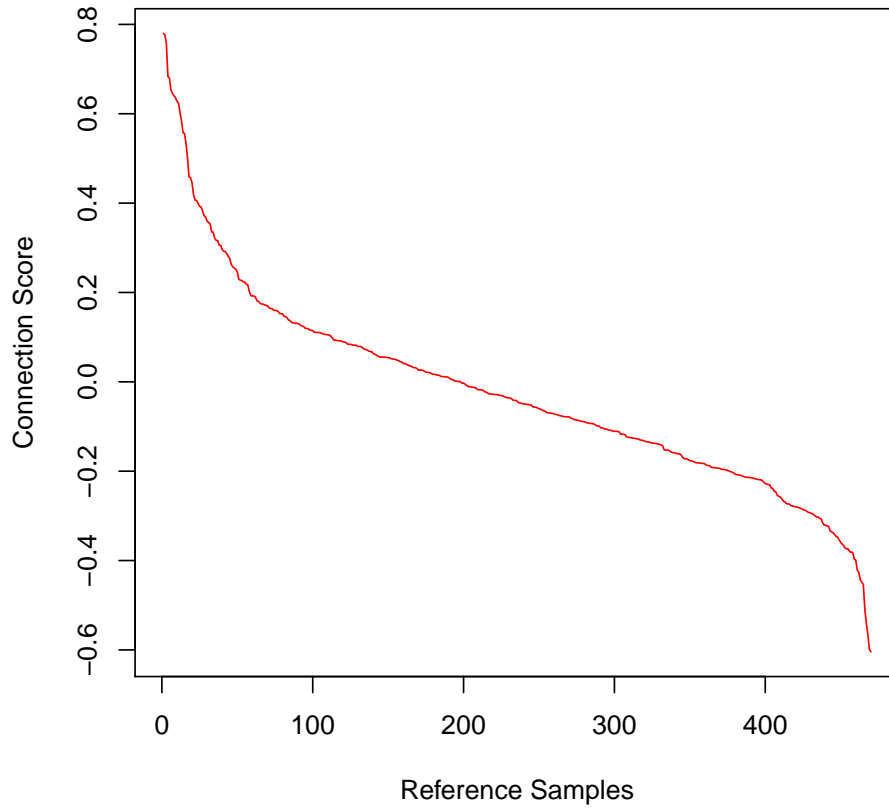


Figure 1: Plot showing Connectivity Score Curve