

Intrudocion to caOmicsV

Hongen Zhang, Ph.D.
Genetics Branch, Center for Cancer Research,
National Cancer Institute, NIH

March 10, 2015

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | An Quick Demo | 2 |
| 3 | Making eSet | 4 |
| 4 | Plot bioMatrix Layout Manually | 5 |
| 5 | Plot bioNetCircos Layout Manually | 8 |
| 6 | sessionInfo | 13 |

1 Introduction

Translational genomics research often generates large multidimensional data from high-throughput technologies. Effectively visualization and presentation of genomics data at multidimensional level will be great benefit for understanding and describing the relationship between genomics variation and cancers. The caOmicsV package provides a set of graphic functions to present multi-dimensional genomics data with two different types of layout: matrix layout (bioMatrix) and circos-like layout on biological network (bioNetCircos).

The data that could be plotted with each layout is listed below:

- Clinical (phenotypes) data such as gender, tissue type, and diagnosis, plotted as colored rectangles
- Expression data such as RNASeq and miRNASeq, plotted as heatmap
- Category data such as DNA methylation status, plotted as colored box outlines on bioMatrix layout and bars on bioNetCircos layout

- Binary data such as mutation status and DNA copy number variations, plotted as colored points
- Text labeling, for gene names, sample names, summary in text format

In addition, link lines can also be plotted on bioNetCircos layout to show the relationship between two samples

For bioNetCircos layout, igraph and bc3net packages must be installed first.

2 An Quick Demo

Following code will generate an bioMatrix layout image with the build-in demo data.

```
> library(caOmicsV)
> data(caOmicsV.biomatrix.eset)
> png("caOmicsV.bioMatrix.Layout.Plot.png", height=8, width=11,
+     units="in", res=300, type="cairo")
> plotBioMatrix(caOmicsV.biomatrix.eset, summaryType="text")
> bioMatrixLegend(heatmapNames=c("RNASeq", "miRNASeq"),
+     categoryNames=c("Methyl H", "Methyl L"),
+     binaryNames=c("CN LOSS", "CN Gain"),
+     heatmapMin=-3, heatmapMax=3, colorType="BlueWhiteRed")
> dev.off()
```

null device

1

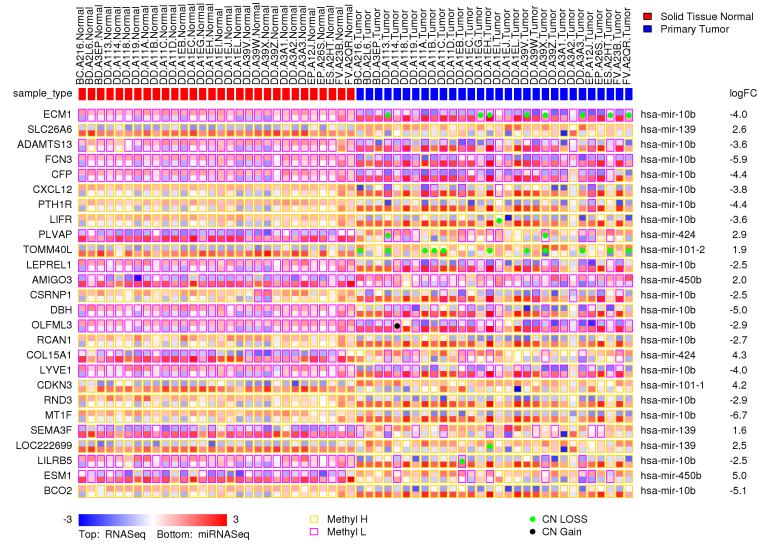


Figure 1. caOmicsV bioMatrix layout plot

Run the code below will get an image in bioNetCircos layout with the build-in demo data.

```
> data(caOmicsV.bionet.eset)
> png("caOmicsV.bioNetCircos.Layout.Plot.png", height=8, width=11,
+     units="in", res=300, type="cairo")
> plotBioNetCircos(caOmicsV.bionet.eset)

plot polygon
plot heatmap
plot bar
plot points

> dataNames <- c("Tissue Type", "RNASeq", "miRNASeq", "Methylation", "CNV")
> bioNetLegend(dataNames, heatmapMin=-3, heatmapMax=3)
> dev.off()

null device
      1
```

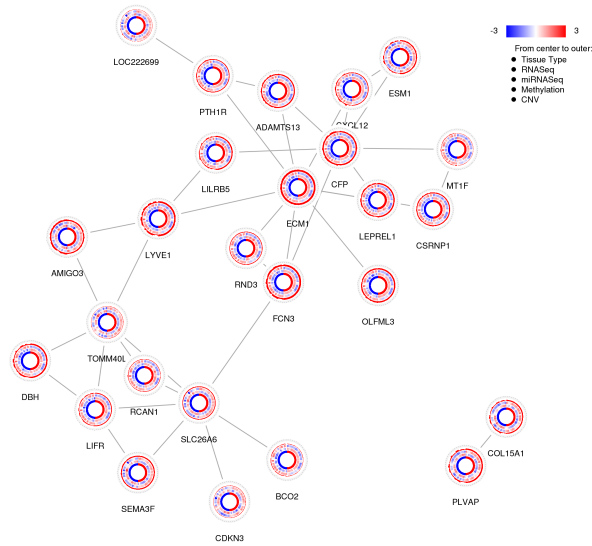


Figure 2. caOmicsV bioNetCircos layout plot

3 Making eSet

To use the default plot function shown as above, the first step is making an eset to hold all plot dataset in a list object. Two demo esets are included in the package installation and could be explored with:

```
> data(caOmicsV.biomatrix.eset)
> names(caOmicsV.biomatrix.eset)

[1] "sampleNames"      "geneNames"         "secondGeneNames"  "sampleInfo"
[5] "heatmapData"      "categoryData"      "binaryData"       "summaryInfo"

> data(caOmicsV.bionet.eset)
> names(caOmicsV.bionet.eset)

[1] "sampleNames"      "geneNames"         "secondGeneNames"  "sampleInfo"
[5] "heatmapData"      "categoryData"      "binaryData"       "summaryInfo"
```

caOmicsV package has one function `getESet()` to make eset object as above. The input data to pass to the function are as below:

- `sampleNames`: required, character vector with names of samples to plot

- `geneNames`: required, character vector with names of genes to plot
- `sampleData`: required, data frame with rows for samples and columns for features. The first column must be sample names same as the `sampleNames` above in same order
- `heatmapData`: list of data frames (maximum 2), continue numeric data such as gene expression data.
- `categoryData`: list of data frames(maximum 2), categorical data such as methylation High, Low, NO.
- `binaryData`: list of data frames(maximum 3), binary data such as 0/1.
- `summaryData`: list of data frames (maximum 2), summarization for genes or for samples.
- `secondGeneNames`: names of second set of genes for labeling on matrix layout plot such as miRNA names

All genomics data must be held with data frame in the format of rows for genes and columns for samples. The first column of each data frame must be gene names same as the `geneNames` as above in same order. The column names of each data frame must be sample names same as `sampleNames` as above in same order. `caOmicsV` package contains functions to sort data frame for given order and several supplement functions are also included in the package to help extract required data set from big datasets by supplying required gene names and sample names in a given order.

To make `caOmicsV` plot, the `eset` must contains at least one genomics data (one of heatmap data, category data, or binary data)

4 Plot bioMatrix Layout Manually

The default plot method, `plotBiomatrix()`, is a convenient and efficient way to make bioMatrix plot. In case of necessary, users can follow procedures below to generate bioMatrix layout plot manually.

1. Demo data

```
> library(caOmicsV)
> data(caOmicsV.biomatrix.eset)
> eset <- caOmicsV.biomatrix.eset
> names(eset)
```

| | | | |
|-------------------|----------------|-------------------|---------------|
| [1] "sampleNames" | "geneNames" | "secondGeneNames" | "sampleInfo" |
| [5] "heatmapData" | "categoryData" | "binaryData" | "summaryInfo" |

2. Initialize bioMatrix Layout

```
> numOfGenes <- length(eset$geneNames);
> numOfSamples <- length(eset$sampleNames);
> numOfPhenotypes <- nrow(eset$sampleInfo)-1;
> numOfHeatmap <- length(eset$heatmapData);
> numOfSummary <- length(eset$summaryData);
> phenotypes <- rownames(eset$sampleInfo)[-1];
> sampleHeight <- 0.4;
> sampleWidth <- 0.1;
> samplePadding <- 0.025;
> geneNameWidth <- 1;
> sampleNameHeight <- 2.5;
> remarkWidth <- numOfHeatmap;
> rowPadding <- 0.1;
> initializeBioMatrixPlot(numOfGenes, numOfSamples, numOfPhenotypes,
+   sampleHeight, sampleWidth, samplePadding, rowPadding, geneNameWidth,
+   remarkWidth, sampleNameHeight)
> caOmicsVColors <- getCaOmicsVColors()
> png("caOmicsVbioMatrixLayoutDemo.png", height=8, width=12,
+   unit="in", res=300, type="cairo")
> par(cex=0.75)
> showBioMatrixPlotLayout(eset$geneNames, eset$sampleNames, phenotypes)
```

The layoue will look like below.

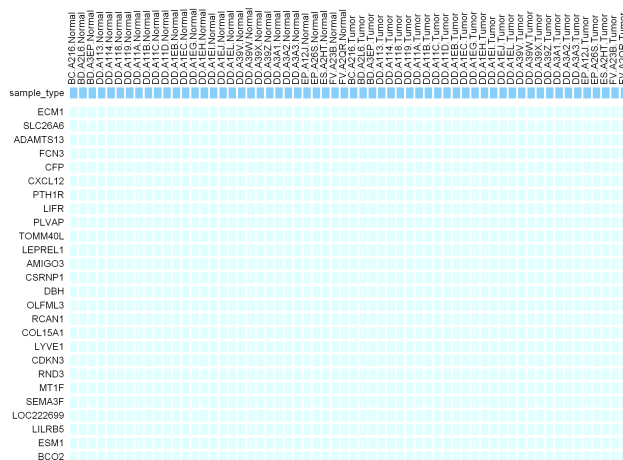


Figure 3. caOmicsV bioMatrix layout

3. Plot tissue types on phenotype area

```
> head(eset$sampleInfo)[,1:3]

                TCGA.BC.A216.Normal  TCGA.BD.A2L6.Normal  TCGA.BD.A3EP.Normal
sampleID      "TCGA.BC.A216.Normal" "TCGA.BD.A2L6.Normal" "TCGA.BD.A3EP.Normal"
sample_type    "Solid Tissue Normal" "Solid Tissue Normal" "Solid Tissue Normal"

> rowIndex <- 2;
> sampleGroup <- as.character(eset$sampleInfo[rowIndex,])
> sampleTypes <- unique(sampleGroup)
> sampleColors <- rep("blue", length(sampleGroup));
> sampleColors[grep("Tumor", sampleGroup)] <- "red"
> rowNumber <- 1
> areaName <- "phenotype"
> plotBioMatrixSampleData(rowNumber, areaName, sampleColors);
> geneLabelX <- getBioMatrixGeneLabelWidth()
> maxAreaX <- getBioMatrixDataAreaWidth()
> legendH <- getBioMatrixLegendHeight()
> plotAreaH <- getBioMatrixPlotAreaHeight()
> sampleH <- getBioMatrixSampleHeight()
> sampleLegendX <- geneLabelX + maxAreaX
> sampleLegendY <- plotAreaH + legendH - length(sampleTypes)*sampleH
> colors <- c("blue", "red")
> legend(sampleLegendX, sampleLegendY, legend=sampleTypes,
+       fill=colors, bty="n", xjust=0)
```

4. Heatmap plot

```
> heatmapData <- as.matrix(eset$heatmapData[[1]][,]);
> plotBioMatrixHeatmap(heatmapData, maxValue=3, minValue=-3)
> heatmapData <- as.matrix(eset$heatmapData[[2]][,])
> plotBioMatrixHeatmap(heatmapData, topAdjust=sampleH/2,
+   maxValue=3, minValue=-3);
> secondNames <- as.character(eset$secondGeneNames)
> textColors <- rep(caOmicsVColors[3], length(secondNames));
> plotBioMatrixRowNames(secondNames, "omicsData", textColors,
+   side="right", skipPlotColumns=0);
```

5. Draw outline for each samples to show methylation status.

```
> categoryData <- eset$categoryData[[1]]
> totalCategory <- length(unique(as.numeric(eset$categoryData[[1]])))
> plotColors <- rev(getCaOmicsVColors())
> plotBioMatrixCategoryData(categoryData, areaName="omicsData",
+   sampleColors=plotColors[1:totalCategory])
```

6. Binary data plot

```
> binaryData <- eset$binaryData[[1]];
> plotBioMatrixBinaryData(binaryData, sampleColor=caOmicsVColors[4]);
> binaryData <- eset$binaryData[[2]];
> plotBioMatrixBinaryData(binaryData, sampleColor=caOmicsVColors[3])
```

7. Plot summary data on right side of plot area

```
> summaryData <- eset$summaryInfo[[1]][, 2];
> summaryTitle <- colnames(eset$summaryInfo[[1]])[2];
> remarkWidth <- getBioMatrixRemarkWidth();
> sampleWidth <- getBioMatrixSampleWidth();
> col2skip <- remarkWidth/2/sampleWidth + 2;
> plotBioMatrixRowNames(summaryTitle, areaName="phenotype",
+   colors="black", side="right", skipPlotColumns=col2skip);
> plotBioMatrixRowNames(summaryData, "omicsData",
+   colors=caOmicsVColors[3], side="right",
+   skipPlotColumns=col2skip)
```

8. Add legend

```
> bioMatrixLegend(heatmapNames=c("RNASeq", "miRNASeq"),
+   categoryNames=c("Methyl H", "Methyl L"),
+   binaryNames=c("CN LOSS", "CN Gain"),
+   heatmapMin=-3, heatmapMax=3,
+   colorType="BlueWhiteRed")
> dev.off()
```

null device

1

Run code above should generate an image same as Figure 1.

5 Plot bioNetCircos Layout Manually

With the default bioNetCircos layout plot method, the node layout and labeling relies on the igraph package and sometimes the node layout and labeling may not be in desired location. In that case, it is recommended to manually check out the layout first then plot each item.

Following are basic procedures to make a bioNetCircos plot:

1. Demo data

```
> library(caOmicsV)
> data(caOmicsV.bionet.eset)
```



```

> eset <- caOmicsV.bionet.eset
> sampleNames <- eset$sampleNames
> geneNames <- eset$geneNames
> numOfSamples <- length(sampleNames)
> numOfSampleInfo <- nrow(eset$sampleInfo) - 1
> numOfSummary <- ifelse(eset$summaryByRow, 0, col(eset$summaryInfo)-1)
> numOfHeatmap <- length(eset$heatmapData)
> numOfCategory <- length(eset$categoryData)
> numOfBinary <- length(eset$binaryData)
> expr <- eset$heatmapData[[1]]
> bioNet <- bc3net(expr)

```

2. Initialize bioNetCircos layout

```

> widthOfSample <- 100
> widthBetweenNode <- 3
> lengthOfRadius <- 10
> dataNum <- sum(numOfSampleInfo, numOfSummary, numOfHeatmap,
+   numOfCategory, numOfBinary)
> trackheight <- 1.5
> widthOfPlotArea <- dataNum*2*trackheight
> initializeBioNetCircos(bioNet, numOfSamples, widthOfSample,
+   lengthOfRadius, widthBetweenNode, widthOfPlotArea)
> caOmicsVColors <- getCaOmicsVColors()
> supportedType <- getCaOmicsVPlotTypes()
> par(cex=0.75)
> showBioNetNodesLayout()

```

3. Manually label each node

At this point, each node has its index labeled. Manually check out the desired location for node name (gene) labeling then label each node (the node index here may be different from your graph).

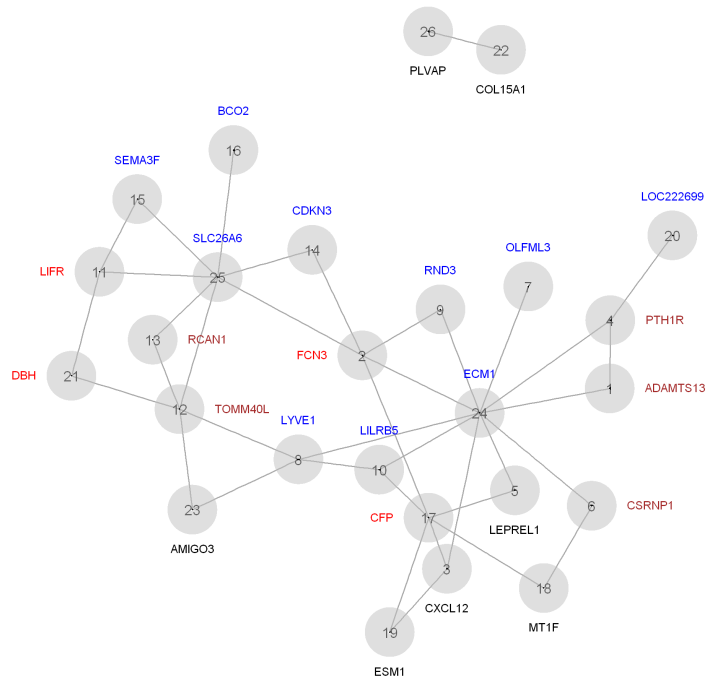


Figure 4. caOmicsV bionetcircos node layout

```
> par(cex=0.6)
> onTop <- c(14, 15, 16, 9, 7, 20, 8, 24, 10, 25)
> labelBioNetNodeNames(nodeList=onTop,labelColor="blue",
+   labelLocation="top", labelOffset = 0.7)
> onBottom <- c(26, 22, 23, 18, 19, 3, 5)
> labelBioNetNodeNames(nodeList=onBottom,labelColor="black",
+   labelLocation="bottom", labelOffset = 0.7)
> onLeft <- c(2, 11, 21, 17)
```

```

> labelBioNetNodeNames(nodeList=onLeft,labelColor="red",
+   labelLocation="left", labelOffset = 0.7)
> onRight <- c(13, 12, 4, 1, 6)
> labelBioNetNodeNames(nodeList=onRight,labelColor="brown",
+   labelLocation="right", labelOffset = 0.7)

```

Once all node names are labeled correctly, plot area of each node could be erased for plotting.

```

> eraseBioNetNode()

```

4. Plot each data set

```

> inner <- lengthOfRadius/2
> outer <- inner + trackheight

```

Plot tissue type for each node. Repeat this step if there are more than one clinical features.

```

> groupInfo <- as.character(eset$sampleInfo[2, ])
> sampleColors <- rep("blue", numOfSamples);
> sampleColors[grep("Tumor", groupInfo)] <- "red"
> plotType=supportedType[1]
> groupInfo <- matrix(groupInfo, nrow=1)
> bioNetCircosPlot(dataValues=groupInfo,
+   plotType, outer, inner, sampleColors)
> inner <- outer + 0.5
> outer <- inner + trackheight

```

Heatmap plot for each node. Repeat this step if there are more heatmap data.

```

> exprData <- eset$heatmapData[[1]]
> plotType <- supportedType[4]
> bioNetCircosPlot(exprData, plotType, outer, inner,
+   plotColors="BlueWhiteRed", maxValue=3, minValue=-3)
> inner <- outer + 0.5
> outer <- inner + trackheight

```

Category data plot for each node. Repeat this step if there are more category datasets.

```

> categoryData <- eset$categoryData[[1]]
> plotType <- supportedType[2];
> bioNetCircosPlot(categoryData, plotType,
+   outer, inner, plotColors="red")

```

length of data values and colors are different.
First color will be used for all samples.

```
> inner <- outer + 0.5
> outer <- inner + trackheight
```

Binary data plot for each node. Repeat this step if there are more binary datasets

```
> binaryData <- eset$binaryData[[1]]
> plotType <- supportedType[3]
> plotColors <- rep(caOmicsVColors[1], ncol(binaryData))
> bioNetCircosPlot(binaryData, plotType,
+   outer, inner, plotColors)
> inner <- outer + 0.5
> outer <- inner + trackheight
```

Link samples on a node. Repeat this step for each node when needed.

```
> outer <- 2.5
> bioNetGraph <- getBioNetGraph()
> nodeIndex <- which(V(bioNetGraph)$name=="PLVAP")
> fromSample <- 10
> toSample <- 50
> plotColors <- "red"
> linkBioNetSamples(nodeIndex, fromSample,
+   toSample, outer, plotColors)
> fromSample <- 40
> toSample <- 20
> plotColors <- "blue"
> linkBioNetSamples(nodeIndex, fromSample,
+   toSample, outer, plotColors)
```

5. Add legend

```
> dataNames <- c("Tissue Type", "RNASeq", "Methylation", "CNV")
> bioNetLegend(dataNames, heatmapMin=-3, heatmapMax=3)
```

The output is shown below.

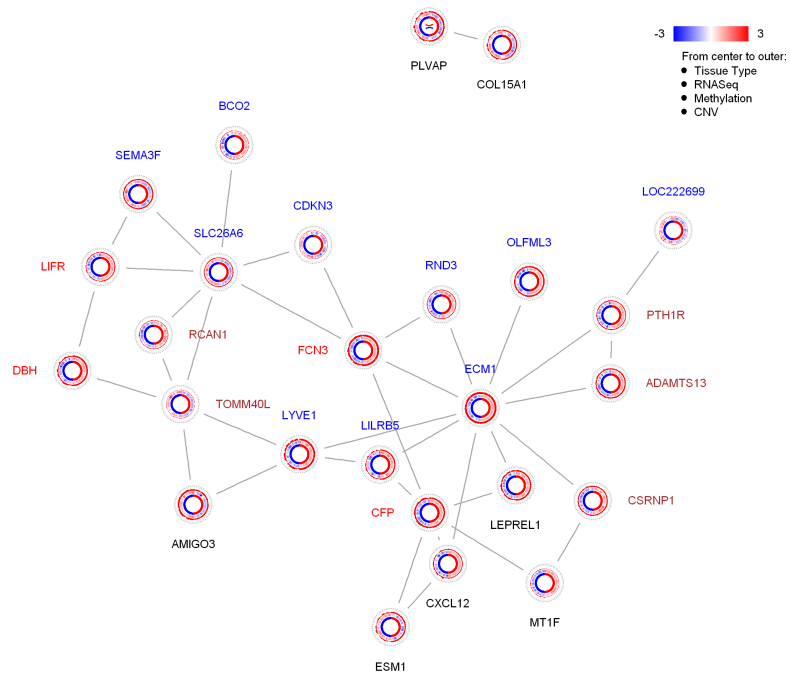


Figure 5. caOmicsV bionetcircos plot

6 sessionInfo

```
> sessionInfo()
```

R version 3.1.2 (2014-10-31)

Platform: x86_64-unknown-linux-gnu (64-bit)

```
locale:
[1] C

attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods    base

other attached packages:
[1] caOmicsV_0.99.0 bc3net_1.0.2    lattice_0.20-29 Matrix_1.1-5
[5] infotheo_1.2.0  c3net_1.1.1      igraph_0.7.1

loaded via a namespace (and not attached):
[1] grid_3.1.2  tools_3.1.2
```