

The OmicCircos usages by examples

Ying Hu and Chunhua Yan

April 16, 2015

Contents

1	Introduction	2
2	Input file formats	2
2.1	segment data	2
2.2	mapping data	2
2.3	link data	3
2.4	link polygon data	3
3	The package functions	3
3.1	sim.circos	3
3.2	segAnglePo	5
3.3	circos	6
4	Plotting parameters	6
4.1	basic plotting	6
4.2	annotation	12
4.3	label	13
4.4	heatmap	16
4.5	traditional plotting and OmicCircos	17
4.6	zoom	22

1 Introduction

The OmicCircos package generates high-quality circular plots for visualizing variations in omics data. The data can be gene or chromosome position-based values from mutation, copy number, expression, and methylation analyses. This package is capable of displaying variations in scatterplots, lines, and text labels. The relationships between genomic features can be presented in the forms of polygons and curves. By utilizing the statistical and graphic functions in R/Bioconductor environment, OmicCircos is also able to draw boxplots, histograms, and heatmaps from multiple sample data. Each track is drawn independently, which allows the use to optimize the track quickly and easily.

In this vignette, we will introduce the package plotting functions using simulation data and TCGA gene expression and copy number variation (cnv) data (<http://www.cancergenome.nih.gov/>).

A quick way to load the vignette examples is:

```
1 vignette("OmicCircos")
```

2 Input file formats

Four input data files are used in the package: segment data, mapping data, link data and link polygon data. Segment data are required to draw the anchor circular track. The remaining three data sets are used to draw additional tracks or connections.

2.1 segment data

The **segment data** lay out the foundation of a circular graph and typically are used to draw the outmost anchor track. In the segment data, column 1 should be the segment or chromosome names. Columns 2 and 3 are the start and end positions of the segment. Columns 4 and 5 are optional which can contain additional description of the segment. The package comes with the segment data for human (hg18 and hg19) and mouse (mm9 and mm10). Let's start by loading the package

```
1 options(stringsAsFactors = FALSE);
2 library(OmicCircos);
3 ## input hg19 cytogenetic band data
4 data(UCSC.hg19.chr);
5 head(UCSC.hg19.chr);
```

	chrom	chromStart	chromEnd	name	gieStain
1	chr1	0	2300000	p36.33	gneg
2	chr1	2300000	5300000	p36.32	gpos25
3	chr1	5300000	7100000	p36.31	gneg
4	chr1	7100000	9200000	p36.23	gpos25
5	chr1	9200000	12600000	p36.22	gneg
6	chr1	12600000	16100000	p36.21	gpos50

2.2 mapping data

The **mapping data** are an R data frame which includes values to be drawn in the graph. In the mapping data, columns 1 and 2 are segment name and position respectively. Column 3 and beyond is optional which can be the value or name. In the following example, the third column is the gene symbol. Column 4 and 5 are the gene expression values for each sample.

```
1 options(stringsAsFactors = FALSE);
2 # load the OmicCircos-package
```

```

3 library(OmicCircos);
4 ## TCGA gene expression data
5 data(TCGA.BC.gene.exp.2k.60);
6 head(TCGA.BC.gene.exp.2k.60[,c(1:5)]);

```

	chr	po	NAME	TCGA.A1.AOSK.01A	TCGA.A1.AOSO.01A
1	1	13799938	PDPN	-2.295	-2.217
2	1	16204544	C1orf64	-6.666	-5.936
3	1	17292188	PADI2	0.994	2.514
4	1	19484755	AKR7A3	-1.418	-1.083
5	1	25113972	RUNX3	-1.536	-1.898
6	1	34994967	GJB5	3.349	-0.809

2.3 link data

The **link data** are for drawing curves between two anchor points. In the link data, columns 1, 2, 3 are the segment name, position, label of the first anchor point; columns 4, 5, 6 are segment name, position, label of the second anchor point. Column 7 is optional and could be used for the link type description.

```

1 options(stringsAsFactors = FALSE);
2 # load the OmicCircos-package
3 library(OmicCircos);
4 ## TCGA fusion gene data
5 data(TCGA.BC.fus);
6 head(TCGA.BC.fus[,c(1:6)]);

```

	chr1	po1	gene1	chr2	po2	gene2
1	2	63456333	WDPCP	10	37493749	ANKRD30A
2	18	14563374	PARD6G	21	14995400	POTED
3	10	37521495	ANKRD30A	3	49282645	CCDC36
4	10	37521495	ANKRD30A	7	100177212	LRCH4
5	18	18539803	ROCK1	18	112551	PARD6G
6	12	4618159	C12orf4	18	1514414	PARD6G

2.4 link polygon data

The **link polygon data** are for connecting two segments with a polygon graph. In the link polygon data, columns 1, 2 and 3 are the name, start and end points for the first segment and columns 4, 5 and 6 are the name, start and end points for the second segment.

3 The package functions

There are three main functions in the package: **sim.circos**, **segAnglePo** and **circos**. **sim.circos** generates simulation data for drawing circular plots. **segAnglePo** converts the genomic (linear) coordinates (chromosome base pair positions) to the angle based coordinates along circumference. **circos** enables users to superimpose graphics on the circle track.

3.1 sim.circos

The **sim.circos** function generates four simulated input data files, which allows users to preview the graph quickly with different parameters and design an optimal presentation with desired features. In the following example, there are 10 segments, 10 individuals, 10 links, and 10 link polygons. Each segment has the value

ranging from 20 to 50. The values will be generated by $\text{rnorm}(1) + i$. The i is the ordinal number of the segments. The values are increased by the segment order.

```

1 options(stringsAsFactors = FALSE);
2 # load the OmicCircos-package
3 library(OmicCircos);
4 # set up the initial parameters
5 seg.num      <- 10;
6 ind.num      <- 20;
7 seg.po       <- c(20:50);
8 link.num     <- 10;
9 link.pg.num  <- 10;
10 # run sim.circos function
11 sim.out      <- sim.circos(seg=seg.num, po=seg.po, ind=ind.num, link=link.num, link.pg
    =link.pg.num);
12 # display the data set names
13 names(sim.out)
14 # display the segment data
15 head(sim.out$seg.frame[,c(1:3)])

```

	seg.name	seg.Start	seg.End
1	chr1	0	1
2	chr1	1	2
3	chr1	2	3
4	chr1	3	4
5	chr1	4	5
6	chr1	5	6

```

1 # display the mapping data
2 head(sim.out$seg.mapping[,c(1:5)])

```

	seg.name	seg.po	name1	name2	name3
1	chr1	1	1.452	0.446	1.131
2	chr1	2	2.205	-0.152	-0.293
3	chr1	3	0.251	1.614	-0.187
4	chr1	4	2.897	1.861	2.837
5	chr1	5	0.305	0.734	0.453
6	chr1	6	1.539	-0.709	0.058

```

1 # display the linking data
2 head(sim.out$seg.link)

```

	seg1	po1	name1	seg2	po2	name2	name3
1	chr8	19	n1	chr2	18	n1	n1
2	chr1	31	n2	chr3	13	n2	n2
3	chr3	15	n3	chr6	40	n3	n3
4	chr5	19	n4	chr9	19	n4	n4
5	chr9	4	n5	chr9	0	n5	n5
6	chr5	23	n6	chr4	44	n6	n6

```

1 # display the linking polygon data
2 head(sim.out$seg.link.pg)

```

	seg1	start1	end1	seg2	start2	end2
1	chr4	1	29	chr6	4	16
2	chr3	1	13	chr7	9	15
3	chr4	23	34	chr9	27	19
4	chr7	19	2	chr9	12	20
5	chr3	2	12	chr7	17	27
6	chr3	4	14	chr8	0	10

3.2 segAnglePo

The `segAnglePo` function converts the segment pointer positions (linear coordinates) into angle values (the angle based coordinates along circumference) and returns a data frame. It specifies the circle size, number of segments, and segment length.

```

1 library(OmicCircos);
2 options(stringsAsFactors = FALSE);
3 set.seed(1234);
4 ## initial values for simulation data
5 seg.num ← 10;
6 ind.num ← 20;
7 seg.po ← c(20:50);
8 link.num ← 10;
9 link.pg.num ← 4;
10 ## output simulation data
11 sim.out ← sim.circos(seg=seg.num, po=seg.po, ind=ind.num, link=link.num,
12   link.pg=link.pg.num);
13 seg.f ← sim.out$seg.frame;
14 seg.v ← sim.out$seg.mapping;
15 link.v ← sim.out$seg.link
16 link.pg.v ← sim.out$seg.link.pg
17 seg.num ← length(unique(seg.f[,1]));
18 ## select segments
19 seg.name ← paste("chr", 1:seg.num, sep="");
20 db ← segAnglePo(seg.f, seg=seg.name);

```

	seg.name	angle.start	angle.end	seg.sum.start	seg.sum.end	seg.start
[1,]	"chr1"	"270"	"294.984"	"0"	"23"	"0"
[2,]	"chr2"	"296.984"	"333.917"	"23"	"57"	"0"
[3,]	"chr3"	"335.917"	"385.885"	"57"	"103"	"0"
[4,]	"chr4"	"387.885"	"423.732"	"103"	"136"	"0"
[5,]	"chr5"	"425.732"	"462.665"	"136"	"170"	"0"
[6,]	"chr6"	"464.665"	"499.425"	"170"	"202"	"0"
[7,]	"chr7"	"501.425"	"524.236"	"202"	"223"	"0"
[8,]	"chr8"	"526.236"	"568.601"	"223"	"262"	"0"
[9,]	"chr9"	"570.601"	"603.188"	"262"	"292"	"0"
[10,]	"chr10"	"605.188"	"628"	"292"	"313"	"0"

	seg.end
[1,]	"23"
[2,]	"34"
[3,]	"46"
[4,]	"33"
[5,]	"34"
[6,]	"32"
[7,]	"21"
[8,]	"39"

```
[9,] "30"
[10,] "21"
```

In the above example, there are 10 segments in a circle. Column 1 is segment name. Columns 2, 3 are the start and end angles of the segment. Column 4 and 5 are the accumulative start and end positions. Column 6 and 7 are the start and end position for the segment. The plotting is clockwise starting at 12 o'clock (270 degree).

3.3 circos

The **circos** is the main function to draw different shapes of the circle. For example, expression and CNV data can be viewed using basic shapes like scatterplots and lines while structural variations such as translocations and fusion proteins can be viewed using curves and polygons to connect different segments. Additionally, multiple sample expression and CNV data sets can be displayed as boxplots, histograms, or heatmaps using standard R functions such as `apply`. The usage of this function is illustrated in the next section.

4 Plotting parameters

4.1 basic plotting

The input data sets were generated by `textttsim.circos` function.

```
1 options(stringsAsFactors = FALSE);
2 library(OmicCircos);
3 options(stringsAsFactors = FALSE);
4 set.seed(1234);
5
6 # initial
7 seg.num      <- 10;
8 ind.num      <- 20;
9 seg.po       <- c(20:50);
10 link.num     <- 10;
11 link.pg.num  <- 4;
12
13 sim.out <- sim.circos(seg=seg.num, po=seg.po, ind=ind.num, link=link.num,
14   link.pg=link.pg.num);
15
16 seg.f      <- sim.out$seg.frame;
17 seg.v      <- sim.out$seg.mapping;
18 link.v     <- sim.out$seg.link
19 link.pg.v  <- sim.out$seg.link.pg
20 seg.num    <- length(unique(seg.f[,1]));
21
22 # name segment (option)
23 seg.name <- paste("chr", 1:seg.num, sep="");
24 db      <- segAnglePo(seg.f, seg=seg.name);
25 # set transparent colors
26 colors  <- rainbow(seg.num, alpha=0.5);
```

To get perfect circle, the output figure should be in square. The output file is the same width and height. The same line values are in the margin of the graphical parameters.

```
1 par(mar=c(2, 2, 2, 2));
2 plot(c(1,800), c(1,800), type="n", axes=FALSE, xlab="", ylab="", main="");
3
```

```

4 | circos(R=400, cir=db, type="chr", col=colors, print.chr.lab=TRUE, W=4, scale=TRUE);
5 | circos(R=360, cir=db, W=40, mapping=seg.v, col.v=3, type="l", B=TRUE, col=colors
  | [1], lwd=2, scale=TRUE);
6 | circos(R=320, cir=db, W=40, mapping=seg.v, col.v=3, type="ls", B=FALSE, col=colors
  | [9], lwd=2, scale=TRUE);
7 | circos(R=280, cir=db, W=40, mapping=seg.v, col.v=3, type="lh", B=TRUE, col=colors
  | [7], lwd=2, scale=TRUE);
8 | circos(R=240, cir=db, W=40, mapping=seg.v, col.v=19, type="ml", B=FALSE, col=colors
  | , lwd=2, scale=TRUE);
9 | circos(R=200, cir=db, W=40, mapping=seg.v, col.v=19, type="ml2", B=TRUE, col=colors,
  | lwd=2);
10 | circos(R=160, cir=db, W=40, mapping=seg.v, col.v=19, type="ml3", B=FALSE, cutoff=5,
  | lwd=2);
11 | circos(R=150, cir=db, W=40, mapping=link.v, type="link", lwd=2, col=colors[c(1,7)]);
12 | circos(R=150, cir=db, W=40, mapping=link.pg.v, type="link.pg", lwd=2, col=sample(
  | colors, link.pg.num));

```

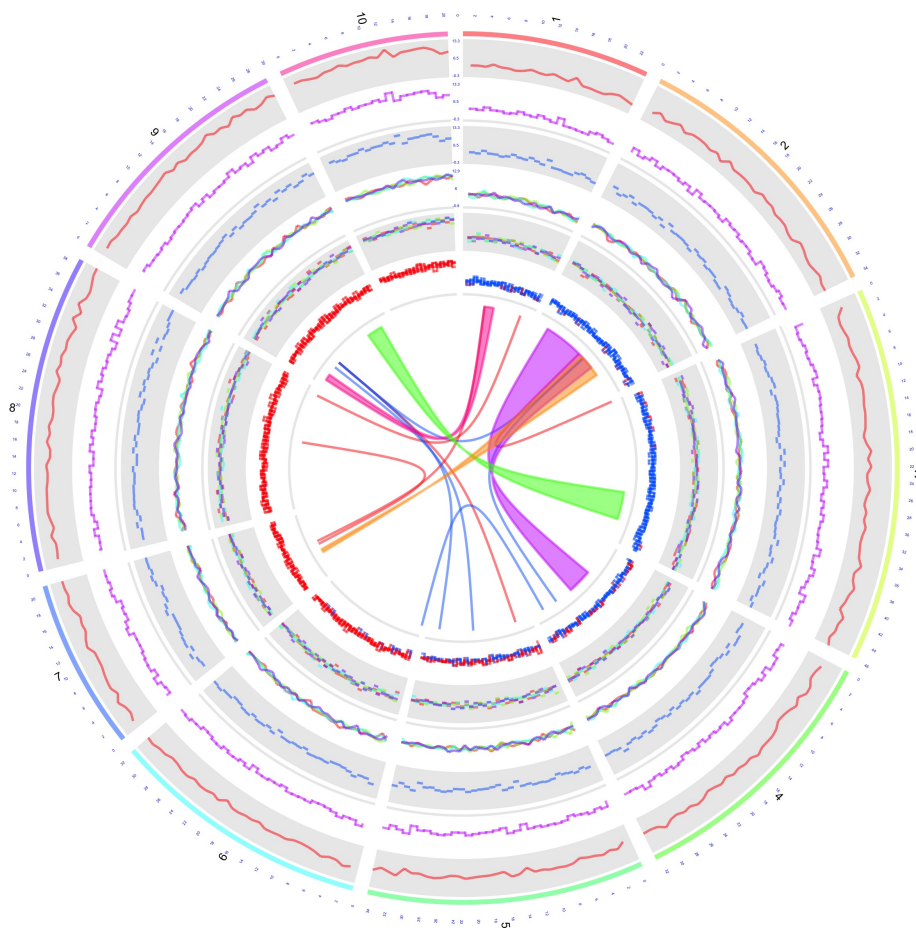


Figure 1

Figure 1 from outside to inside: Track 1 is lines; Track 2 is the stair steps; Track 3 is the horizontal lines; Tracks 4, 5 and 6 are the multiple lines, stair steps and horizontal lines for multiple the samples.

```

1 options(stringsAsFactors = FALSE);
2 library(OmicCircos);
3 set.seed(1234);
4
5 ## initial values for simulation data
6 seg.num      <- 10;
7 ind.num      <- 20;
8 seg.po       <- c(20:50);
9 link.num     <- 10;
10 link.pg.num  <- 4;
11 ## output simulation data
12 sim.out <- sim.circos(seg=seg.num, po=seg.po, ind=ind.num, link=link.num,
13   link.pg=link.pg.num);
14
15 seg.f      <- sim.out$seg.frame;
16 seg.v      <- sim.out$seg.mapping;
17 link.v     <- sim.out$seg.link
18 link.pg.v  <- sim.out$seg.link.pg
19 seg.num    <- length(unique(seg.f[,1]));
20
21 ## select segments
22 seg.name <- paste("chr", 1:seg.num, sep="");
23 db      <- segAnglePo(seg.f, seg=seg.name);
24
25 colors  <- rainbow(seg.num, alpha=0.5);

```

```

1 par(mar=c(2, 2, 2, 2));
2 plot(c(1,800), c(1,800), type="n", axes=FALSE, xlab="", ylab="", main="");
3
4 circos(R=400, type="chr", cir=db, col=colors, print.chr.lab=TRUE, W=4, scale=TRUE);
5 circos(R=360, cir=db, W=40, mapping=seg.v, col.v=8, type="box", B=TRUE, col=colors
6   [1], lwd=0.1, scale=TRUE);
7 circos(R=320, cir=db, W=40, mapping=seg.v, col.v=8, type="hist", B=TRUE, col=colors
8   [3], lwd=0.1, scale=TRUE);
9 circos(R=280, cir=db, W=40, mapping=seg.v, col.v=8, type="ms", B=TRUE, col=colors
10  [7], lwd=0.1, scale=TRUE);
11 circos(R=240, cir=db, W=40, mapping=seg.v, col.v=3, type="h", B=FALSE, col=colors
12  [2], lwd=0.1);
13 circos(R=200, cir=db, W=40, mapping=seg.v, col.v=3, type="s", B=TRUE, col=colors,
14   lwd=0.1);
15 circos(R=160, cir=db, W=40, mapping=seg.v, col.v=3, type="b", B=FALSE, col=colors,
16   lwd=0.1);
17 circos(R=150, cir=db, W=40, mapping=link.v, type="link", lwd=2, col=colors[c(1,7)]);
18 circos(R=150, cir=db, W=40, mapping=link.pg.v, type="link.pg", lwd=2, col=sample(
19   colors, link.pg.num));

```

Figure 2 from outside to inside: Track 1 is the boxplot for the samples from column 8 (col.v=8) to the last column in the data frame seg.v with the scale; Track 2 and track 3 are the histograms (in horizontal) and the scatter plots for multiple samples as track 1. Tracks 4, 5 and 6 are the histogram (in vertical), scatter plot and vertical line for just one sample (column 3 in the data frame seg.v).

```

1 options(stringsAsFactors = FALSE);
2 library(OmicCircos);
3 set.seed(1234);
4

```

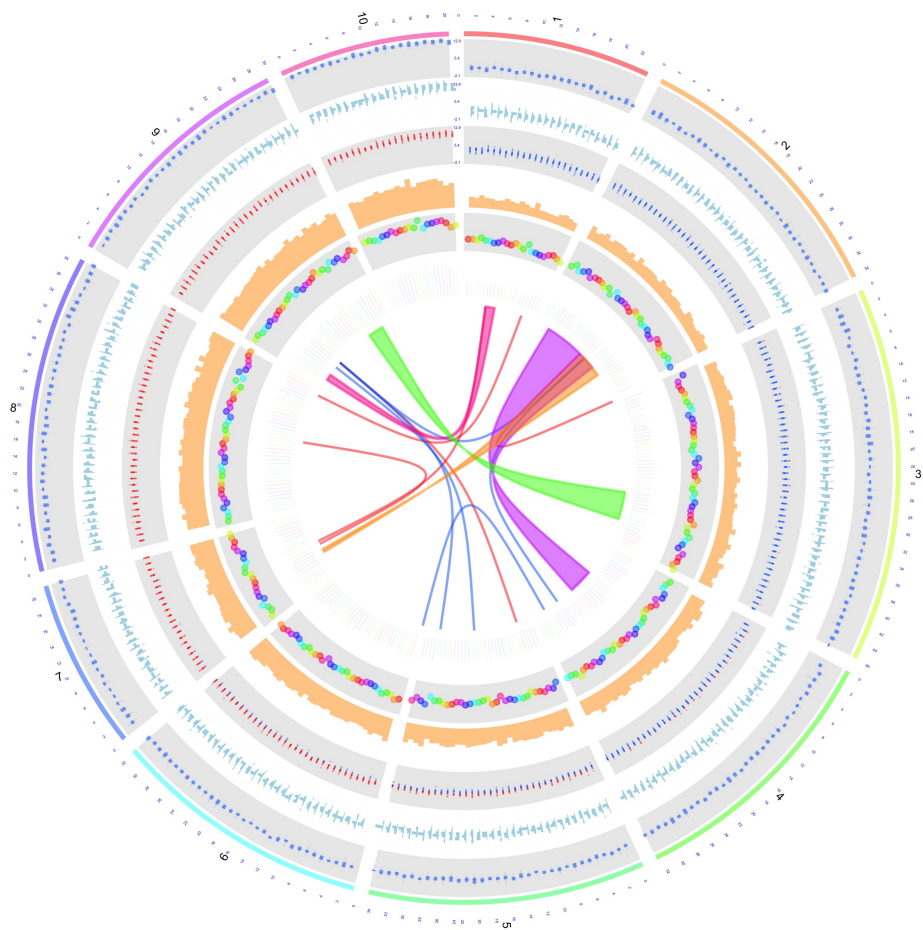



Figure 2

```

5 ## initial values for simulation data
6 seg.num      ← 10;
7 ind.num      ← 20;
8 seg.po       ← c(20:50);
9 link.num     ← 10;
10 link.pg.num  ← 4;
11 ## output simulation data
12 sim.out ← sim.circos(seg=seg.num, po=seg.po, ind=ind.num, link=link.num,
13   link.pg=link.pg.num);
14
15 seg.f      ← sim.out$seg.frame;
16 seg.v      ← sim.out$seg.mapping;
17 link.v     ← sim.out$seg.link
18 link.pg.v  ← sim.out$seg.link.pg
19 seg.num    ← length(unique(seg.f[,1]));
20
21 ##
22 seg.name   ← paste("chr", 1:seg.num, sep="");
23 db         ← segAnglePo(seg.f, seg=seg.name);
24
25 colors     ← rainbow(seg.num, alpha=0.5);

1 par(mar=c(2, 2, 2, 2));
2 plot(c(1,800), c(1,800), type="n", axes=FALSE, xlab="", ylab="", main="");
3
4 circos(R=400, type="chr", cir=db, col=colors, print.chr.lab=TRUE, W=4, scale=TRUE);
5 circos(R=360, cir=db, W=40, mapping=seg.v, col.v=8, type="quant90", B=FALSE, col=
  colors, lwd=2, scale=TRUE);
6 circos(R=320, cir=db, W=40, mapping=seg.v, col.v=3, type="sv", B=TRUE, col=colors
  [7], scale=TRUE);
7 circos(R=280, cir=db, W=40, mapping=seg.v, col.v=3, type="ss", B=FALSE, col=colors
  [3], scale=TRUE);
8 circos(R=240, cir=db, W=40, mapping=seg.v, col.v=8, type="heatmap", lwd=3);
9 circos(R=200, cir=db, W=40, mapping=seg.v, col.v=3, type="s.sd", B=FALSE, col=colors
  [4]);
10 circos(R=160, cir=db, W=40, mapping=seg.v, col.v=3, type="ci95", B=TRUE, col=colors
  [4], lwd=2);
11 circos(R=150, cir=db, W=40, mapping=link.v, type="link", lwd=2, col=colors[c(1,7)]);
12 circos(R=150, cir=db, W=40, mapping=link.pg.v, type="link.pg", lwd=2, col=sample(
  colors, link.pg.num));
13
14 the.col1=rainbow(10, alpha=0.5)[3];
15 highlight ← c(160, 410, 6, 2, 6, 10, the.col1, the.col1);
16 circos(R=110, cir=db, W=40, mapping=highlight, type="hl", lwd=1);
17
18 the.col1=rainbow(10, alpha=0.1)[3];
19 the.col2=rainbow(10, alpha=0.5)[1];
20 highlight ← c(160, 410, 3, 12, 3, 20, the.col1, the.col2);
21 circos(R=110, cir=db, W=40, mapping=highlight, type="hl", lwd=2);

```

Figure 3 from outside to inside: Track 1 is the three lines for quantile values for the samples from column 8 (col.v=8) to the last column in the data frame seg.v with the scale. The middle line is for the median, the outside line and the inside line are for 90% and the 10%, respectively; Track 2 is the circle points with the center=median and radius=variance; Track 3 is the circle plot with the center equal to the mean and scaled value (for example, the range from 0 to 3); Tracks 4 is the heatmap for the samples from column 8

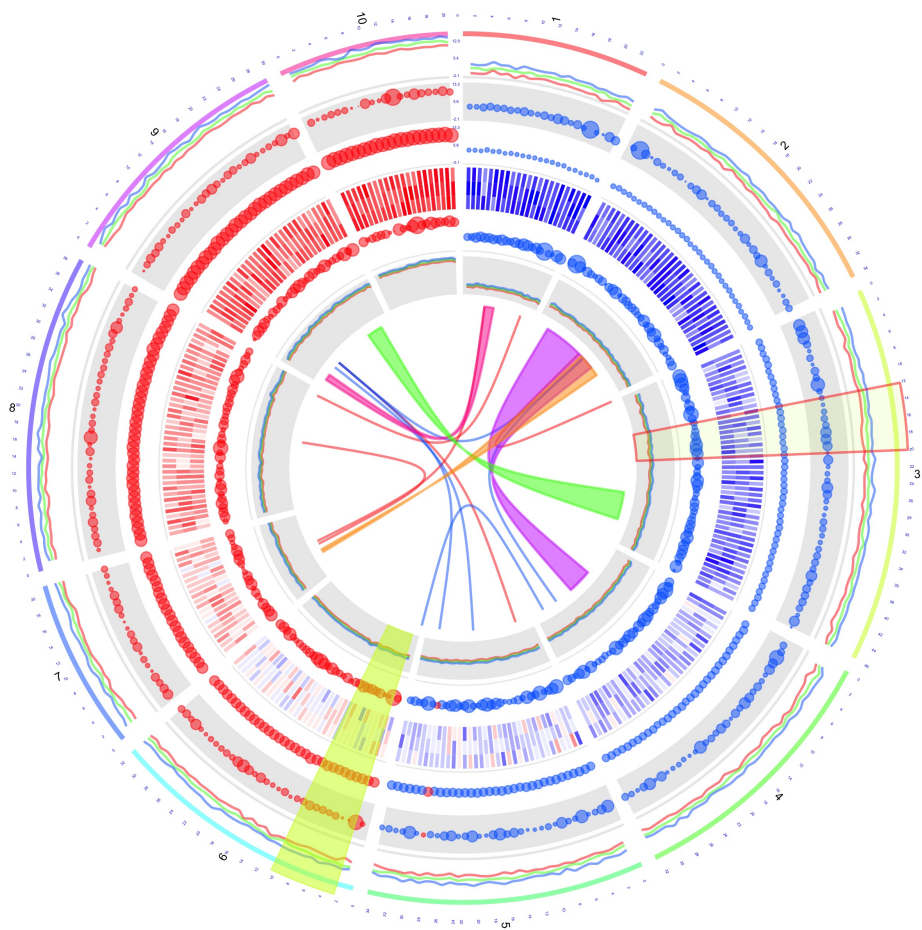


Figure 3

(col.v=8) to the last column in the data frame seg.v; Track 5 is the circle plot with the center=median and radius=standard deviation; Track 6 is the 95% confidence interval of the samples.

4.2 annotation

```

1 options(stringsAsFactors = FALSE);
2 library(OmicCircos);
3 set.seed(1234);
4
5 ## load mm cytogenetic band data
6 data("UCSC.mm10.chr", package="OmicCircos");
7 ref <- UCSC.mm10.chr;
8 ref[,1] <- gsub("chr", "", ref[,1]);
9 ## initial values for simulation data
10 colors <- rainbow(10, alpha=0.8);
11 lab.n <- 50;
12 cnv.n <- 200;
13 arc.n <- 30;
14 fus.n <- 10;
15
16 ## make arc data
17 arc.d <- c();
18 for (i in 1:arc.n){
19   chr <- sample(1:19, 1);
20   chr.i <- which(ref[,1]==chr);
21   chr.arc <- ref[chr.i,];
22   arc.i <- sample(1:nrow(chr.arc), 2);
23   arc.d <- rbind(arc.d, c(chr.arc[arc.i[1],c(1,2)], chr.arc[arc.i[2],c(2,4)]));
24 }
25 colnames(arc.d) <- c("chr", "start", "end", "value");
26
27 ## make fusion
28 fus.d <- c();
29 for (i in 1:fus.n){
30   chr1 <- sample(1:19, 1);
31   chr2 <- sample(1:19, 1);
32   chr1.i <- which(ref[,1]==chr1);
33   chr2.i <- which(ref[,1]==chr2);
34   chr1.f <- ref[chr1.i,];
35   chr2.f <- ref[chr2.i,];
36   fus1.i <- sample(1:nrow(chr1.f), 1);
37   fus2.i <- sample(1:nrow(chr2.f), 1);
38   n1 <- paste0("geneA", i);
39   n2 <- paste0("geneB", i);
40   fus.d <- rbind(fus.d, c(chr1.f[fus1.i,c(1,2)], n1, chr2.f[fus2.i,c(1,2)], n2));
41 }
42 colnames(fus.d) <- c("chr1", "po1", "gene1", "chr2", "po2", "gene2");
43
44 cnv.i <- sample(1:nrow(ref), cnv.n);
45 vale <- rnorm(cnv.n);
46 cnv.d <- data.frame(ref[cnv.i,c(1,2)], value=vale);

```

```

1 par(mar=c(2, 2, 2, 2));
2 plot(c(1,800), c(1,800), type="n", axes=FALSE, xlab="", ylab="");

```

```

3
4 circos(R=400, type="chr", cir="mm10", print.chr.lab=TRUE, W=4, scale=TRUE);
5 circos(R=340, cir="mm10", W=60, mapping=cnv.d, type="b3", B=TRUE, col=colors[7]);
6 circos(R=340, cir="mm10", W=60, mapping=cnv.d, type="s2", B=FALSE, col=colors[1],
  cex=0.5);
7 circos(R=280, cir="mm10", W=60, mapping=arc.d, type="arc2", B=FALSE, col=colors, lwd
  =10, cutoff=0);
8 circos(R=220, cir="mm10", W=60, mapping=cnv.d, col.v=3, type="b2", B=TRUE, cutoff=
  -0.2, col=colors[c(7,9)], lwd=2);
9 circos(R=160, cir="mm10", W=60, mapping=arc.d, col.v=4, type="arc", B=FALSE, col=
  colors[c(1,7)], lwd=4, scale=TRUE);
10 circos(R=150, cir="mm10", W=10, mapping=fus.d, type="link", lwd=2, col=colors[c
  (1,7,9)]);

```

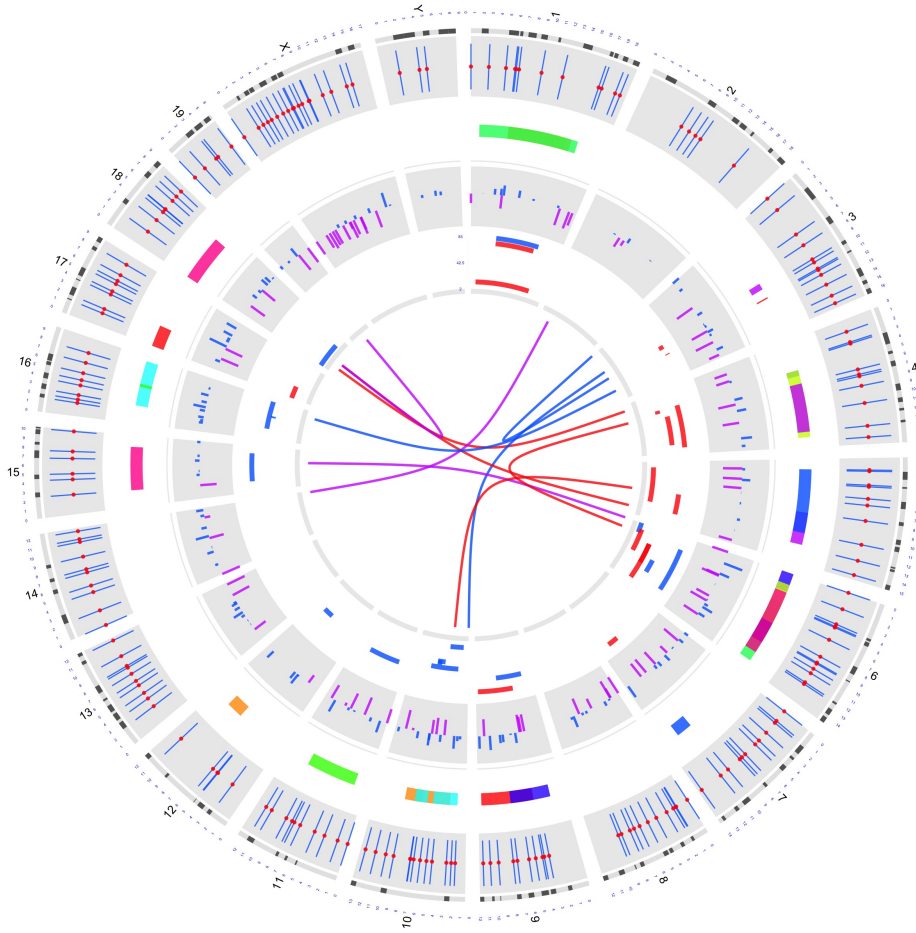


Figure 4

4.3 label

Figure 4 from outside to inside: Track 1 is the vertical lines with the same length and radius which can be used for the annotation of SNP positions; Track 2 is the arcs with the same radius which can be used for the

segment annotation, e.g. `cnv` (copy number variation); Track 3 is the barplot with positive and negative values; Track 4 is the arcs in the different radius.

```

1 options(stringsAsFactors = FALSE);
2 library(OmicCircos);
3
4 data("TCGA.PAM50_genefu_hg18");
5 data("TCGA.BC.fus");
6 data("TCGA.BC.cnv.2k.60");
7 data("TCGA.BC.gene.exp.2k.60");
8 data("TCGA.BC.sample60");
9 data("TCGA.BC_Her2_cnv_exp");
10
11 pvalue ← -1 * log10(TCGA.BC_Her2_cnv_exp[,5]);
12 pvalue ← cbind(TCGA.BC_Her2_cnv_exp[,c(1:3)], pvalue);
13
14 Her2.i ← which(TCGA.BC.sample60[,2] == "Her2");
15 Her2.n ← TCGA.BC.sample60[Her2.i,1];
16
17 Her2.j ← which(colnames(TCGA.BC.cnv.2k.60) %in% Her2.n);
18 cnv ← TCGA.BC.cnv.2k.60[,c(1:3, Her2.j)];
19 cnv.m ← cnv[,c(4:ncol(cnv))];
20 cnv.m[cnv.m > 2] ← 2;
21 cnv.m[cnv.m < -2] ← -2;
22 cnv ← cbind(cnv[,1:3], cnv.m);
23
24 Her2.j ← which(colnames(TCGA.BC.gene.exp.2k.60) %in% Her2.n);
25 gene.exp ← TCGA.BC.gene.exp.2k.60[,c(1:3, Her2.j)];
26 colors ← rainbow(10, alpha=0.5);

1 par(mar=c(2, 2, 2, 2));
2 plot(c(1,800), c(1,800), type="n", axes=FALSE, xlab="", ylab="");
3
4 circos(R=300, type="chr", cir="hg18", print.chr.lab=FALSE, W=4);
5 circos(R=310, cir="hg18", W=20, mapping=TCGA.PAM50_genefu_hg18, type="label",
6       side="out", col=c("black", "blue", "red"), cex=0.4);
7 circos(R=250, cir="hg18", W=50, mapping=cnv, col.v=4, type="ml3", B=FALSE, col=
8       colors[7], cutoff=0, scale=TRUE);
9 circos(R=200, cir="hg18", W=50, mapping=gene.exp, col.v=4, type="ml3", B=TRUE, col=
10      colors[3], cutoff=0, scale=TRUE);
11 circos(R=140, cir="hg18", W=50, mapping=pvalue, col.v=4, type="l", B=FALSE, col=
12      colors[1], scale=TRUE);
13 ## set fusion gene colors
14 cols ← rep(colors[7], nrow(TCGA.BC.fus));
15 col.i ← which(TCGA.BC.fus[,1]==TCGA.BC.fus[,4]);
16 cols[col.i] ← colors[1];
17 circos(R=132, cir="hg18", W=50, mapping=TCGA.BC.fus, type="link", col=cols, lwd=2);

```

Figure 5 is an example of adding outside labels.

```

1 par(mar=c(2, 2, 2, 2));
2 plot(c(1,800), c(1,800), type="n", axes=FALSE, xlab="", ylab="", main="");
3 circos(R=300, type="chr", cir="hg18", col=TRUE, print.chr.lab=FALSE, W=4);
4 circos(R=290, cir="hg18", W=20, mapping=TCGA.PAM50_genefu_hg18, type="label", side="
5       in", col=c("black", "blue"), cex=0.4);
6 circos(R=310, cir="hg18", W=50, mapping=cnv, col.v=4, type="ml3", B=TRUE, col=colors
7       [7], cutoff=0, scale=TRUE);

```



```

6 | circos(R=150, cir="hg18", W=50, mapping=gene.exp, col.v=4, type="ml3", B=TRUE, col=
  | colors[3], cutoff=0, scale=TRUE);
7 | circos(R=90, cir="hg18", W=50, mapping=pvalue, col.v=4, type="l", B=FALSE, col=
  | colors[1], scale=TRUE);
8 | circos(R=82, cir="hg18", W=50, mapping=TCGA.BC.fus, type="link", col=cols, lwd=2);

```

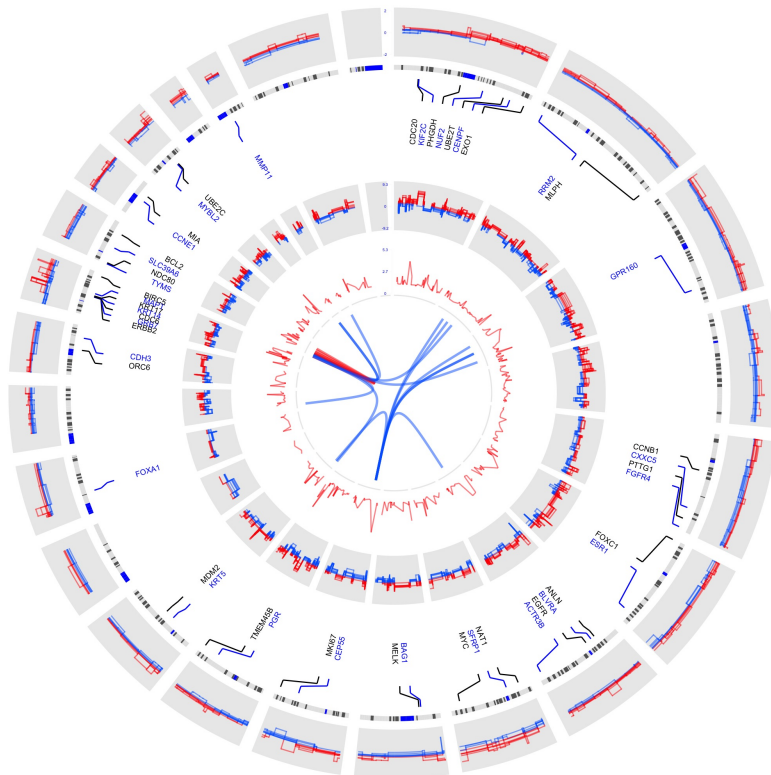


Figure 6

Figure 6 is an example of the inside labels.

4.4 heatmap

```
1 options(stringsAsFactors = FALSE);
2 library(OmicCircos);
3
4 data("TCGA.PAM50_genefu_hg18");
5 data("TCGA.BC.fus");
6 data("TCGA.BC.cnv.2k.60");
7 data("TCGA.BC.gene.exp.2k.60");
8 data("TCGA.BC.sample60");
```



```

9 data("TCGA.BC_Her2_cnv_exp");
10
11 pvalue <- -1 * log10(TCGA.BC_Her2_cnv_exp[,5]);
12 pvalue <- cbind(TCGA.BC_Her2_cnv_exp[,c(1:3)], pvalue);
13
14 Her2.i <- which(TCGA.BC.sample60[,2] == "Her2");
15 Her2.n <- TCGA.BC.sample60[Her2.i,1];
16
17 Her2.j <- which(colnames(TCGA.BC.cnv.2k.60) %in% Her2.n);
18 cnv <- TCGA.BC.cnv.2k.60[,c(1:3, Her2.j)];
19 cnv.m <- cnv[,c(4:ncol(cnv))];
20 cnv.m[cnv.m > 2] <- 2;
21 cnv.m[cnv.m < -2] <- -2;
22 cnv <- cbind(cnv[,1:3], cnv.m);
23
24 Her2.j <- which(colnames(TCGA.BC.gene.exp.2k.60) %in% Her2.n);
25 gene.exp <- TCGA.BC.gene.exp.2k.60[,c(1:3, Her2.j)];
26
27 colors <- rainbow(10, alpha=0.5);

1 par(mar=c(2, 2, 2, 2));
2
3 plot(c(1,800), c(1,800), type="n", axes=FALSE, xlab="", ylab="", main="");
4
5 circos(R=400, cir="hg18", W=4, type="chr", print.chr.lab=TRUE, scale=TRUE);
6 circos(R=300, cir="hg18", W=100, mapping=gene.exp, col.v=4, type="heatmap2",
7       cluster=TRUE, col.bar=TRUE, lwd=0.1, col="blue");
8 circos(R=220, cir="hg18", W=80, mapping=cnv, col.v=4, type="ml3", B=FALSE, lwd
9       =1, cutoff=0);
10 circos(R=140, cir="hg18", W=80, mapping=pvalue, col.v=4, type="l", B=TRUE,
11       lwd=1, col=colors[1]);
12
13 cols <- rep(colors[7], nrow(TCGA.BC.fus));
14 col.i <- which(TCGA.BC.fus[,1]==TCGA.BC.fus[,4]);
15 cols[col.i] <- colors[1];
16 circos(R=130, cir="hg18", W=10, mapping=TCGA.BC.fus, type="link2", lwd=2, col=cols)
17 ;

```

Figure 7: An example of a circular plots generated by OmicCircos showing the expression, CNV and fusion protein in 15 Her2 subtype samples from TCGA Breast Cancer data. Circular tracks from outside to inside: genome positions by chromosomes (black lines are cytobands); expression heatmap (red: up-regulated; blue: down-regulated); CNVs (red: gain; blue: loss); correlation p values between expression and CNVs; fusion genes.

4.5 traditional plotting and OmicCircos

```

1 options(stringsAsFactors = FALSE);
2 library(OmicCircos);
3
4 data("TCGA.BC.fus");
5 data("TCGA.BC.cnv.2k.60");
6 data("TCGA.BC.gene.exp.2k.60");
7 data("TCGA.BC.sample60");
8

```

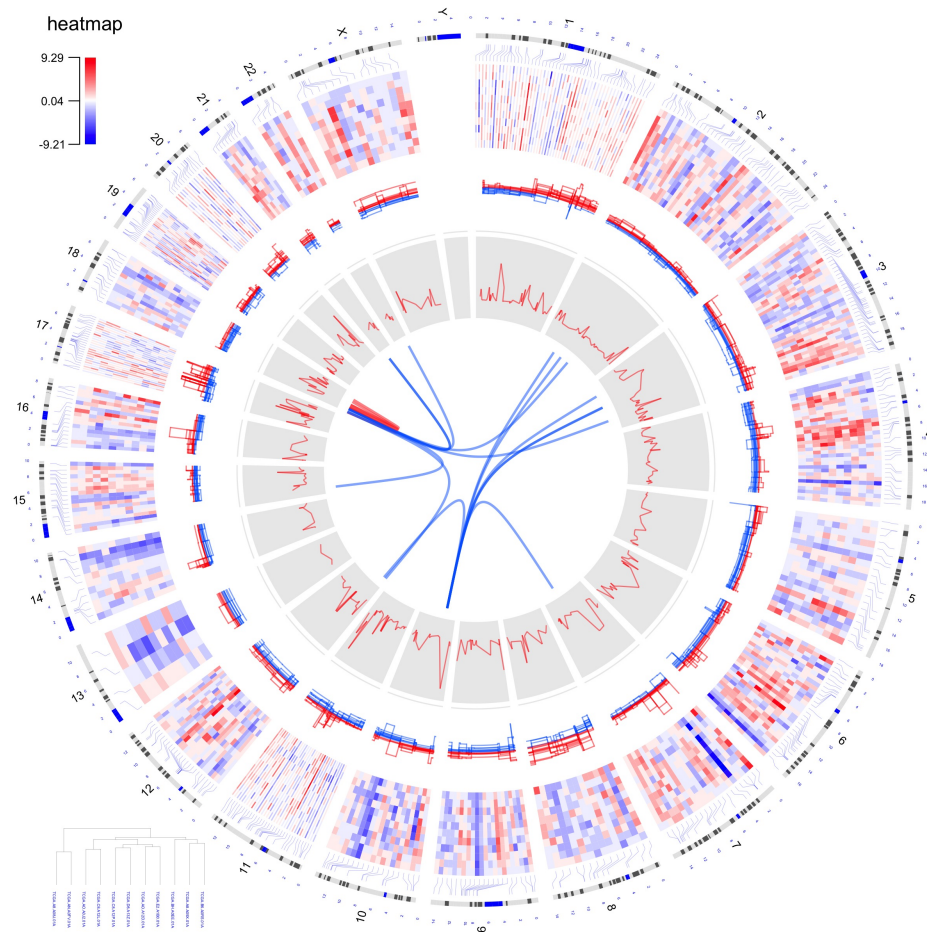


Figure 7

```

9  ## gene expression data for PCA
10 exp.m  <- TCGA.BC.gene.exp.2k.60[,c(4:ncol(TCGA.BC.gene.exp.2k.60))];
11 cnv     <- TCGA.BC.cnv.2k.60;
12 type.n  <- unique(TCGA.BC.sample60[,2]);
13 colors  <- rainbow(length(type.n), alpha=0.5);
14
15 ## sub-type colors
16 pca.col <- rep(NA, nrow(TCGA.BC.sample60));
17 for (i in 1:length(type.n)){
18   n  <- type.n[i];
19   n.i <- which(TCGA.BC.sample60[,2] == n);
20   n.n <- TCGA.BC.sample60[n.i,1];
21   g.i <- which(colnames(exp.m) %in% n.n);
22   pca.col[g.i] <- colors[i];
23 }
24
25 ## run PCA
26 exp.m  <- na.omit(exp.m);
27 pca.out <- prcomp(t(exp.m), scale = TRUE);
28
29 ## subtype cnv
30 cnv.i <- c();
31 for (i in 1:length(type.n)){
32   n  <- type.n[i];
33   n.i <- which(TCGA.BC.sample60[,2] == n);
34   n.n <- TCGA.BC.sample60[n.i,1];
35   cnv.i <- which(colnames(cnv) %in% n.n);
36 }

1  ## PCA is plotting.
2  plot(pca.out$x[,1]*5, pca.out$x[,2]*5, pch=19, col=pca.col, main="",
3       cex=2, xlab="PC1", ylab="PC2", ylim=c(-200, 460), xlim=c(-200,460));
4  legend(200,0, c("Basal","Her2","LumA","LumB"), pch=19, col=colors[c(2,4,1,3)], cex
5         =1,
6         title ="Gene Expression (PCA)", box.col="white");
7
8  ## It is going to plot the circos.
9  circos(xc=280, yc=280, R=168, cir="hg18", W=4,      type="chr", print.chr.lab=TRUE);
10 R.v <- 135;
11 for (i in 1:length(type.n)){
12   n  <- type.n[i];
13   n.i <- which(TCGA.BC.sample60[,2] == n);
14   n.n <- TCGA.BC.sample60[n.i,1];
15   cnv.i <- which(colnames(cnv) %in% n.n);
16   cnv.v <- cnv[,cnv.i];
17   cnv.v[cnv.v > 2] <- 2;
18   cnv.v[cnv.v < -2] <- -2;
19   cnv.m <- cbind(cnv[,c(1:3)], cnv.v);
20   circos(xc=280, yc=280, R=R.v, cir="hg18", W=34, mapping=cnv.m, col.v=4, type="ml3
21         ", B=FALSE, lwd=0.5, cutoff=0);
22   R.v <- R.v - 25;
23 }
24
25 legend(-80,460, c("1 Basal", "2 Her2", "3 LumA", "4 LumB", "(center)"), cex=1,
26        title ="CNV (OmicCircos)", box.col="white");

```

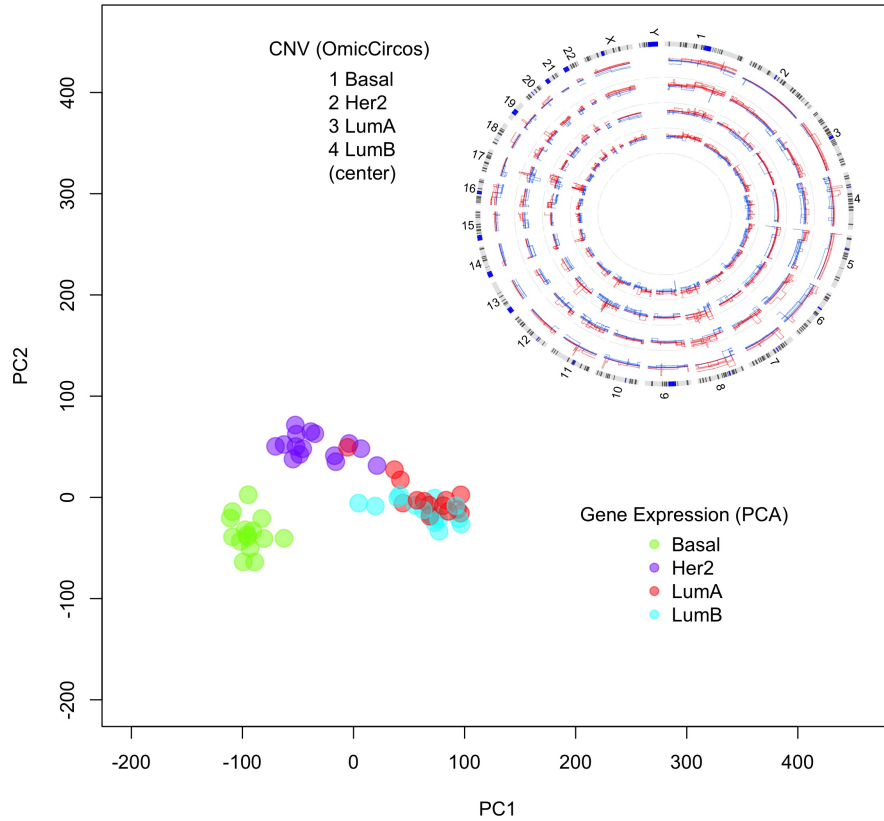


Figure 8

Figure 8: Integration of circular plots by OmicCircos and PCA from R for TCGA breast cancer data. Circular tracks from outside to inside: genome positions by chromosomes (black lines are cytobands); CNVs in Basal, Her2, Luminal A (LumA) and Luminal B (LumB) subtypes with 15 samples in each track (red: gain; blue: loss); At the bottom left corner, four clusters were generated by the from principle component analysis (PCA) of the gene expression data using R packages in the bioconductor.

```

1 plot(c(1,800), c(1,800), type="n", axes=FALSE, xlab="", ylab="", main="");
2
3 legend(680,800, c("Basal","Her2","LumA","LumB"), pch=19, col=colors[c(2,4,1,3)], cex
4     =0.5,
5     title ="Gene Expression (PCA)", box.col="white");
6
7 legend(5,800, c("1 Basal", "2 Her2", "3 LumA", "4 LumB", "(center)"), cex=0.5,
8     title ="CNV (OmicCircos)", box.col="white");
9
10 circos(xc=400, yc=400, R=390, cir="hg18", W=4, type="chr", print.chr.lab=TRUE,
11     scale=TRUE);
12 R.v <- 330;
13 for (i in 1:length(type.n)){
14     n <- type.n[i];

```

```

13   n.i   <- which(TCGA.BC.sample60[,2] == n);
14   n.n   <- TCGA.BC.sample60[n.i,1];
15   cnv.i <- which(colnames(cnv) %in% n.n);
16   cnv.v <- cnv[,cnv.i];
17   cnv.v[cnv.v > 2] <- 2;
18   cnv.v[cnv.v < -2] <- -2;
19   cnv.m <- cbind(cnv[,c(1:3)], cnv.v);
20   circos(xc=400, yc=400, R=R.v, cir="hg18", W=60, mapping=cnv.m, col.v=4, type="ml3",
21         B=FALSE, lwd=1, cutoff=0, scale=TRUE);
21   R.v <- R.v - 60;
22 }
23
24 points(pca.out$x[,1]*3.6+400, pca.out$x[,2]*3.6+400, pch=19, col=pca.col, cex=2);

```

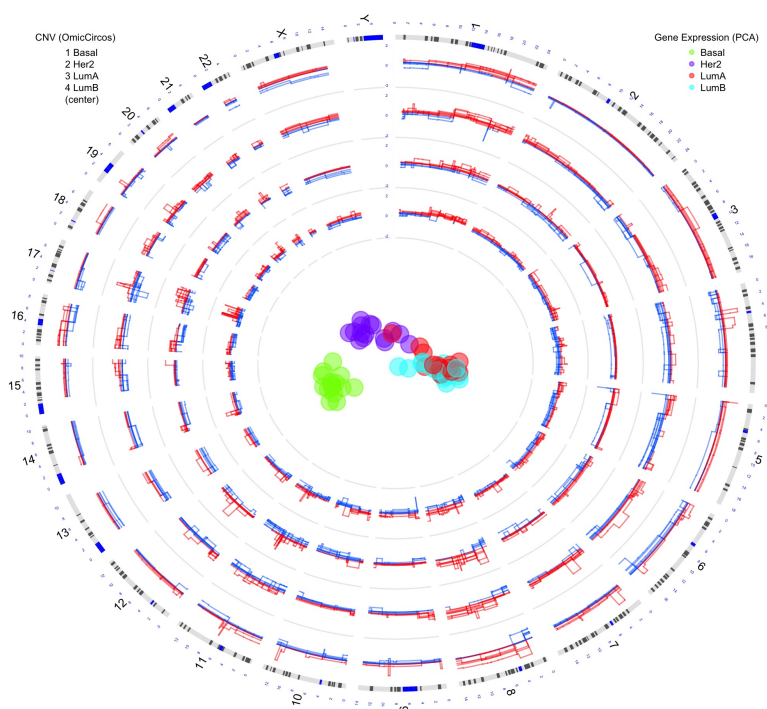


Figure 9

Figure 9 is an example showing the PCA plot is located at the center of the circos plot. Integration of circular plots by OmicCircos and PCA from R of TCGA breast cancer data. Circular tracks from outside to inside: genome positions by chromosomes (black lines are cytobands); CNVs in Basal, Her2, Luminal A (LumA) and Luminal B (LumB) subtypes with 15 samples in each track (red: gain; blue: loss); In the center, four clusters from principle component analysis (PCA) of the gene expression data with R packages

in the bioconductor.

4.6 zoom

```
1 options(stringsAsFactors = FALSE);
2 library(OmicCircos);
3
4 data("TCGA.PAM50_genefu_hg18");
5 data("TCGA.BC.fus");
6 data("TCGA.BC.cnv.2k.60");
7 data("TCGA.BC.gene.exp.2k.60");
8 data("TCGA.BC.sample60");
9 data("TCGA.BC_Her2_cnv_exp");
10 data("TCGA.PAM50_genefu_hg18");
11
12 pvalue <- -1 * log10(TCGA.BC_Her2_cnv_exp[,5]);
13 pvalue <- cbind(TCGA.BC_Her2_cnv_exp[,c(1:3)], pvalue);
14
15 Her2.i <- which(TCGA.BC.sample60[,2] == "Her2");
16 Her2.n <- TCGA.BC.sample60[Her2.i,1];
17
18 Her2.j <- which(colnames(TCGA.BC.cnv.2k.60) %in% Her2.n);
19 cnv <- TCGA.BC.cnv.2k.60[,c(1:3, Her2.j)];
20 cnv.m <- cnv[,c(4:ncol(cnv))];
21 cnv.m[cnv.m > 2] <- 2;
22 cnv.m[cnv.m < -2] <- -2;
23 cnv <- cbind(cnv[,1:3], cnv.m);
24
25 gene.exp <- TCGA.BC.gene.exp.2k.60[,c(1:3, Her2.j)];
26
27 colors <- rainbow(10, alpha=0.5);

1 par(mar=c(2, 2, 2, 2));
2
3 plot(c(1,800), c(1,800), type="n", axes=FALSE, xlab="", ylab="", main="");
4 # In figure 7, the chromosome 1 to chromosome 22 are going to be plotted from the
5 # angle 0 (12 O'clock)
6 # to 180 degree (6 O'clock).
7 zoom <- c(1, 22, 939245.5, 154143883, 0, 180);
8 circos(R=400, cir="hg18", W=4, type="chr", print.chr.lab=TRUE, scale=TRUE, zoom=
9 zoom);
10 circos(R=300, cir="hg18", W=100, mapping=gene.exp, col.v=4, type="heatmap2",
11 cluster=TRUE, col.bar=TRUE, col.bar.po = "bottomright", lwd=0.01, zoom=zoom);
12 circos(R=220, cir="hg18", W=80, mapping=cnv, col.v=4, type="ml3", B=FALSE,
13 lwd=1, cutoff=0, zoom=zoom);
14 circos(R=140, cir="hg18", W=80, mapping=pvalue, col.v=4, type="l", B=TRUE, lwd
15 =1, col=colors[1], zoom=zoom);
16 circos(R=130, cir="hg18", W=10, mapping=TCGA.BC.fus, type="link", lwd=2, zoom=zoom)
17 ;
18
19 # zoom in links by using the highlight functions
20 # highlight
21 the.col1=rainbow(10, alpha=0.5)[1];
22
```

```

17 # The highline region is radium from 140 to 400 and from position 282412.5 to
    133770314.5 in chromosome 11.
18 highlight <- c(140, 400, 11, 282412.5, 11, 133770314.5, the.col1, the.col1);
19 circos(R=110, cir="hg18", W=40, mapping=highlight, type="hl", lwd=2, zoom=zoom);
20 the.col2=rainbow(10, alpha=0.5)[6];
21 highlight <- c(140, 400, 17, 739525, 17, 78385909, the.col2, the.col2);
22 circos(R=110, cir="hg18", W=40, mapping=highlight, type="hl", lwd=2, zoom=zoom);
23 ## highlight link
24 highlight.link1 <- c(400, 400, 140, 376.8544, 384.0021, 450, 540.5);
25 circos(cir="hg18", mapping=highlight.link1, type="highlight.link", col=the.col1, lwd
    =1);
26 highlight.link2 <- c(400, 400, 140, 419.1154, 423.3032, 543, 627);
27 circos(cir="hg18", mapping=highlight.link2, type="highlight.link", col=the.col2, lwd
    =1);
28
29 # The chromosome 11 region is going plotting from 180 (6 Oclock) to 270 degree (9
    Oclock).
30 zoom <- c(11, 11, 282412.5, 133770314.5, 180, 270);
31 circos(R=400, cir="hg18", W=4, type="chr", print.chr.lab=TRUE, scale=FALSE, zoom=
    zoom);
32 circos(R=300, cir="hg18", W=100, mapping=gene.exp, col.v=4, type="heatmap2",
    cluster=TRUE, lwd=0.01, zoom=zoom);
33 circos(R=220, cir="hg18", W=80, mapping=cnv, col.v=4, type="ml3", B=FALSE,
    lwd=1, cutoff=0, zoom=zoom);
34 circos(R=140, cir="hg18", W=80, mapping=pvalue, col.v=4, type="l", B=TRUE, lwd
    =1, col=colors[1], zoom=zoom);
35
36 # The chromosome 17 region is going plotting from 180 (6 Oclock) to 270 degree (9
    Oclock).
37
38 gene.names <- c("ERBB2", "CDC6");
39 PAM50.17 <- which(TCGA.PAM50_genefu_hg18[,3]==gene.names);
40 TCGA.PAM50 <- TCGA.PAM50_genefu_hg18[PAM50.17,];
41
42 # zoom in chromosome 17
43 zoom <- c(17, 17, 739525, 78385909, 274, 356);
44 circos(R=400, cir="hg18", W=4, type="chr", print.chr.lab=TRUE, scale=TRUE, zoom=
    zoom);
45 circos(R=300, cir="hg18", W=100, mapping=gene.exp, col.v=4, type="heatmap2",
    cluster=TRUE, lwd=0.01, zoom=zoom);
46 circos(R=220, cir="hg18", W=80, mapping=cnv, col.v=4, type="ml3", B=FALSE,
    lwd=1, cutoff=0, zoom=zoom);
47 circos(R=140, cir="hg18", W=80, mapping=pvalue, col.v=4, type="l", B=TRUE, lwd
    =1, col=colors[1], zoom=zoom);
48 circos(R=410, cir="hg18", W=40, mapping=TCGA.PAM50, type="label", side="out", col="
    blue", zoom=zoom);

```

Figure 10: Circular plots by OmicCircos of expression, CNV, and fusion proteins in 15 Her2 subtype samples from TCGA breast cancer data. Circular tracks from outside to inside: genome positions by chromosomes (black lines are cytobands), expression heatmap of 500 most variable genes whose genomic coordinates are indicated by the blue connectors, CNVs (red: gain, blue: loss), correlation p-values between expression and CNV, fusion genes (intra/inter chromosomes: red/blue). Chromosomes 1–22 are shown in the right half of the circle. Zoomed chromosomes 11 and 17 are displayed in the left half. The circle arc was drawn using trigonometric functions. The fusion protein links were plotted with the Bézier curve algorithm.

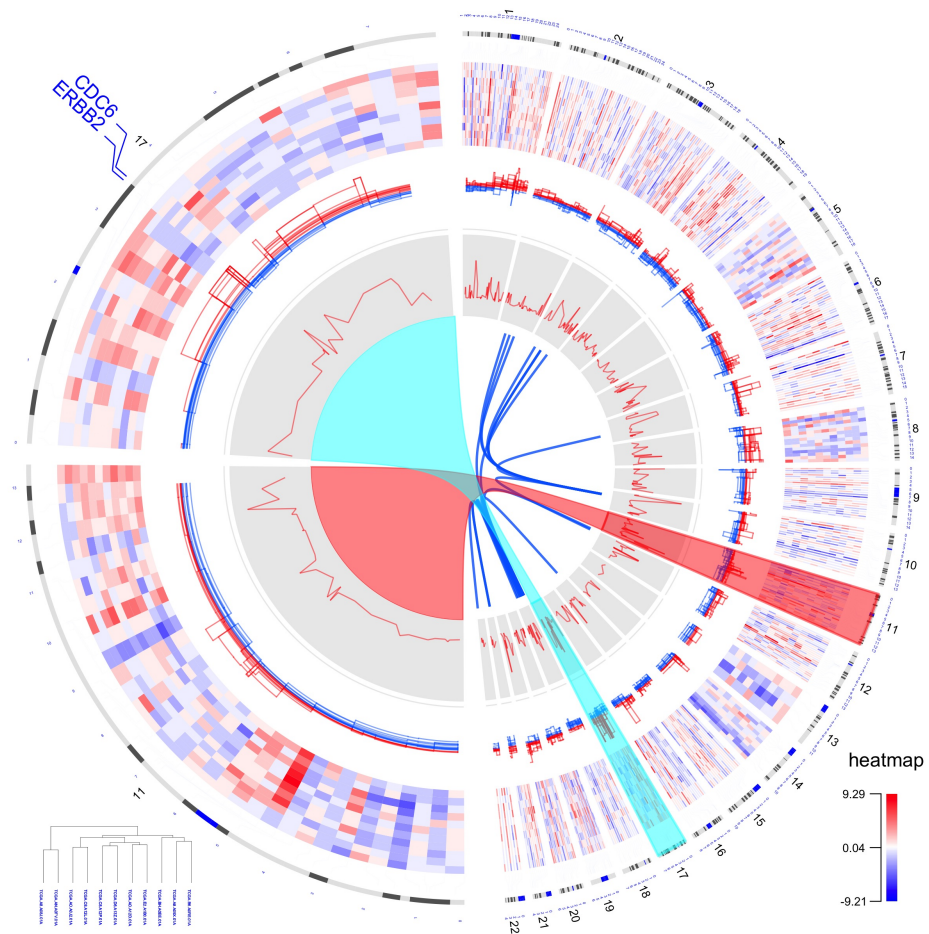


Figure 10