

SOFTWARE

Open Access



caOmicsV: an R package for visualizing multidimensional cancer genomic data

Hongen Zhang, Paul S. Meltzer and Sean R. Davis*

Abstract

Background: Translational genomics research in cancers, e.g., International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA), has generated large multidimensional datasets from high-throughput technologies. Data analysis at multidimensional level will greatly benefit clinical applications of genomic information in diagnosis, prognosis and therapeutics of cancers. To help, tools to effectively visualize integrated multidimensional data are important for understanding and describing the relationship between genomic variations and cancers.

Results: We implemented the R package, caOmicsV, to provide methods under R environment to visualize multidimensional cancer genomic data in two layouts: matrix layout and combined biological network and circular layout. Both layouts support to display sample information, gene expression (e.g., RNA and miRNA), DNA methylation, DNA copy number variations, and summarized data. A set of supplemental functions are included in the caOmicsV package to help users in generation of plot data sets from multiple genomic datasets with given gene names and sample names. Default plot methods for both layouts for easy use are also implemented.

Conclusion: caOmicsV package provides an easy and flexible way to visualize integrated multidimensional cancer genomic data under R environment.

Keywords: Software, R package, Genomic data visualization, Multidimensional data visualization

Background

Featured with high-throughput technologies, current translational genomic research of cancers often generates multidimensional data such as mRNA/miRNA expression, DNA methylation, exome sequencing, and SNP/DNA copy number variations [1, 2]. Data analysis at multidimensional level will greatly benefit clinical applications of genomic information in diagnosis, prognosis and therapeutics of cancers. To help, tools to effectively visualize integrated multidimensional data are important for understanding and describing the relationship between genomic variation and cancers [3–5].

Visualizing multidimensional genomic data have been implemented in different ways: genomic coordinate based presentation, heatmaps, and networks views [4]. Genomic coordinate based tools such as UCSC genome browser

and Integrative Genomics Viewer are powerful in viewing of detailed sequence and various types of variations as well as epigenomic and transcriptomes profiles that tied to genomic loci [6, 7], and CIRCOS and its implementation under different environments [8–10] help in exploring relationships between genomic alterations or positions. One disadvantage of genomic coordinate based tools is the limitations on numbers of samples and genes displayed simultaneous and integration of genomic variations with network/pathway information. In contrast, heatmap and network views can integrate multiple types of genomic variations independent of genomic loci in multiple sample groups at gene set or pathway level and are commonly used in presenting relationship between genomics variations and sample features and relationship between different genomic alterations [11–13].

The R statistical programming environment, an important open source tool used in cancer research community for statistical analysis and visualization of cancer genomic data, has packages which implemented genomic coordinate

* Correspondence: sdavis2@mail.nih.gov
Genetics Branch, Center for Cancer Research, National Cancer Institute,
National Institutes of Health, Building 37, Room 6138, 37 Convent Drive,
Bethesda, MD 20892-4265, USA

based views [14–16] and complex heatmap views [17]. To facility the R with more flexible and easy way in presenting multidimensional genomic information, we developed the caOmicsV package for R, to provide a set of graphic functions for visualizing multidimensional genomic data with two different types of layout: matrix layout (bioMatrix) and circular layout on biological network (bioNetCircos).

Implementation

The caOmicsV package is implemented with R language only and provides two layouts for displaying multidimensional genomic dataset: bioMatrix and bioNetCircos layout. Both layouts support to display sample features, mRNA and miRNA expression, DNA copy number variations (CNV), DNA methylation data, and summarization data. On bioMatrix layout, clinical features of cancer samples are shown with different colored rectangles, gene expression (mRNA and miRNA) data are plotted as heatmap, DNA methylation status are presented as colored

rectangle outlines, and DNA copy number variations are displayed as colored points. Besides gene names and sample names, summarized data can also be presented on the layout as text or bars. On bioNetCircos layout, a biological network is built from given gene expression dataset and genes are presented as nodes on the network. Clinical features, mRNA and miRNA expression, DNA methylation, DNA copy number variation, and other summarized data for each sample are displayed in circular layout on each node (gene) as polygons, heatmap, bars, points, or lines. In the center of each node, link lines could be plotted to display the relationship between two samples. Both layouts are using low level plot functions of R graphics package. For bioNetCircos layout, installation of R igraph package is required.

Results

The presentation of multidimensional genomic information and sample feature with caOmicsV package is shown in Figs. 1 and 2. Two default plot methods,

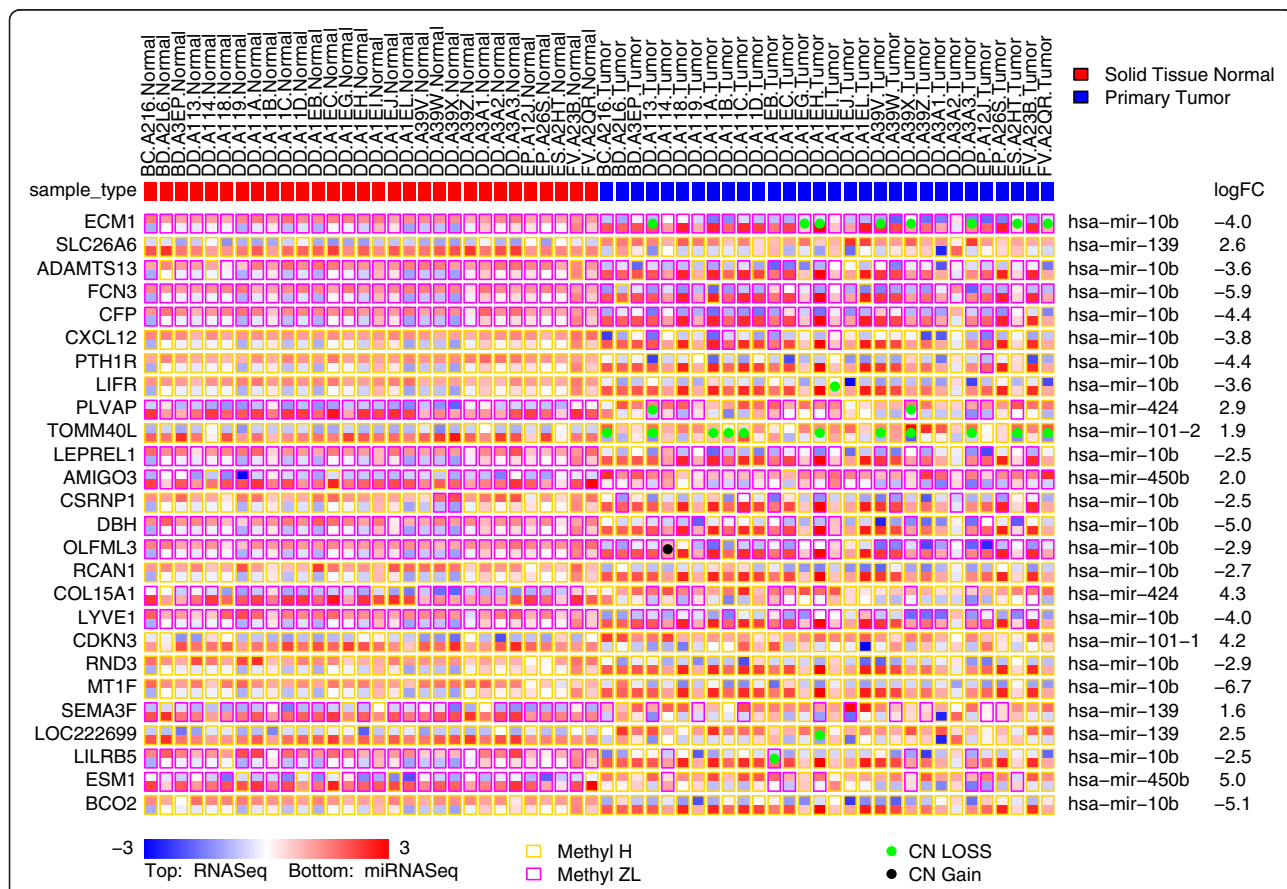
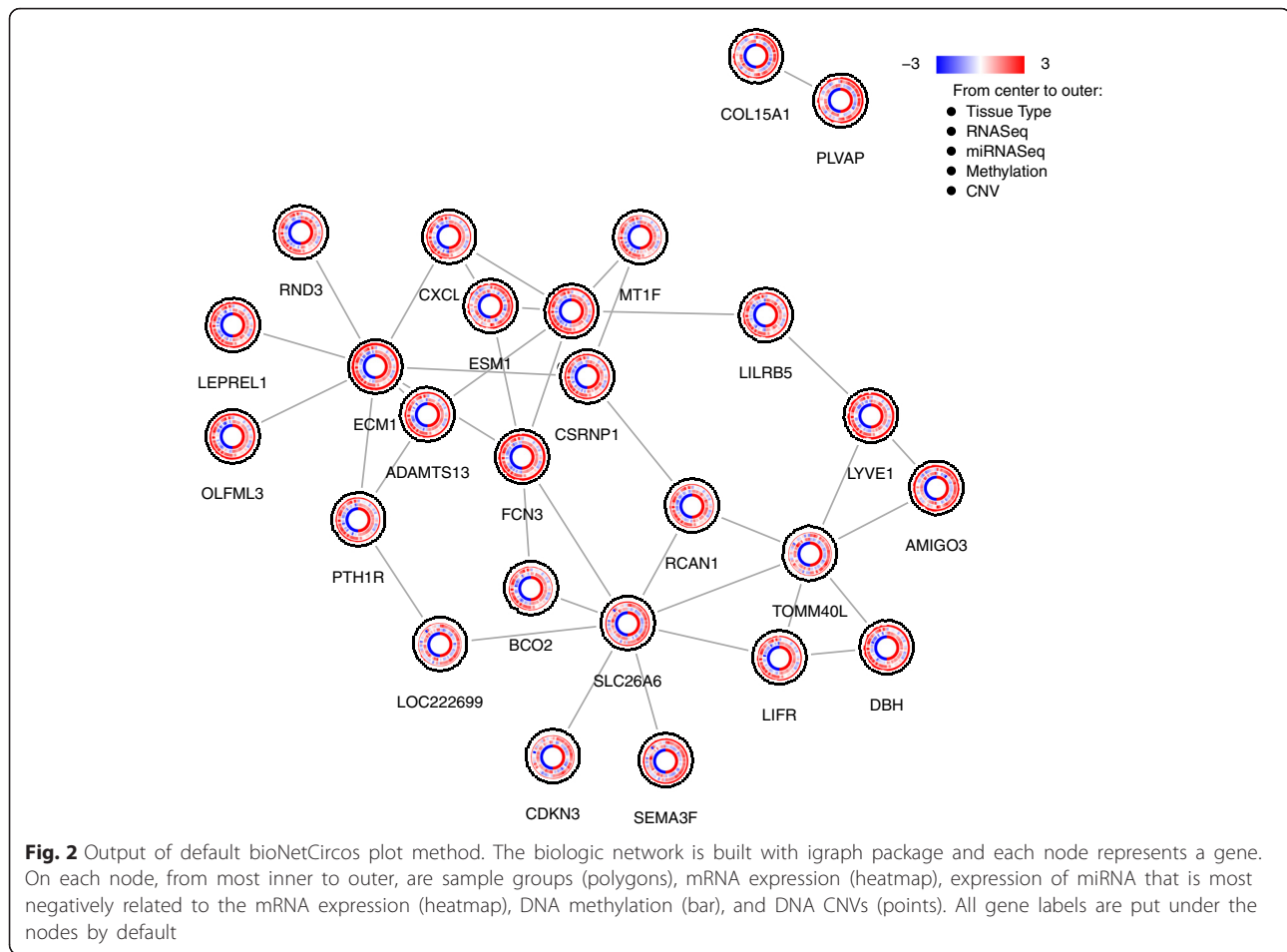


Fig. 1 Output of default bioMatrix plot method. Sample information is plot on the top of matrix and below that are data for each gene. In each gene row, each column represents a sample, and mRNA and miRNA expression are shown as heatmap, DNA methylation is represented by different colored outlines, DNA CNV are plotted as colored points. For each gene, top half heatmap show mRNA expression and bottom half are expression of miRNA that is most significant negatively related to the gene. The mean fold change for each gene is listed at the most right of plot area



plotBioMatrix() and plotBioNetCircos() are implemented for easy use. As shown below, simply pass the input data to relevant plot functions will generate the images with default parameter setting as Figs. 1 and 2.

Default plot method for bioMatrix plot

```
library(caOmicsV)
data(biomatrixPlotDemoData)
plotBioMatrix(biomatrixPlotDemoData,
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")
```

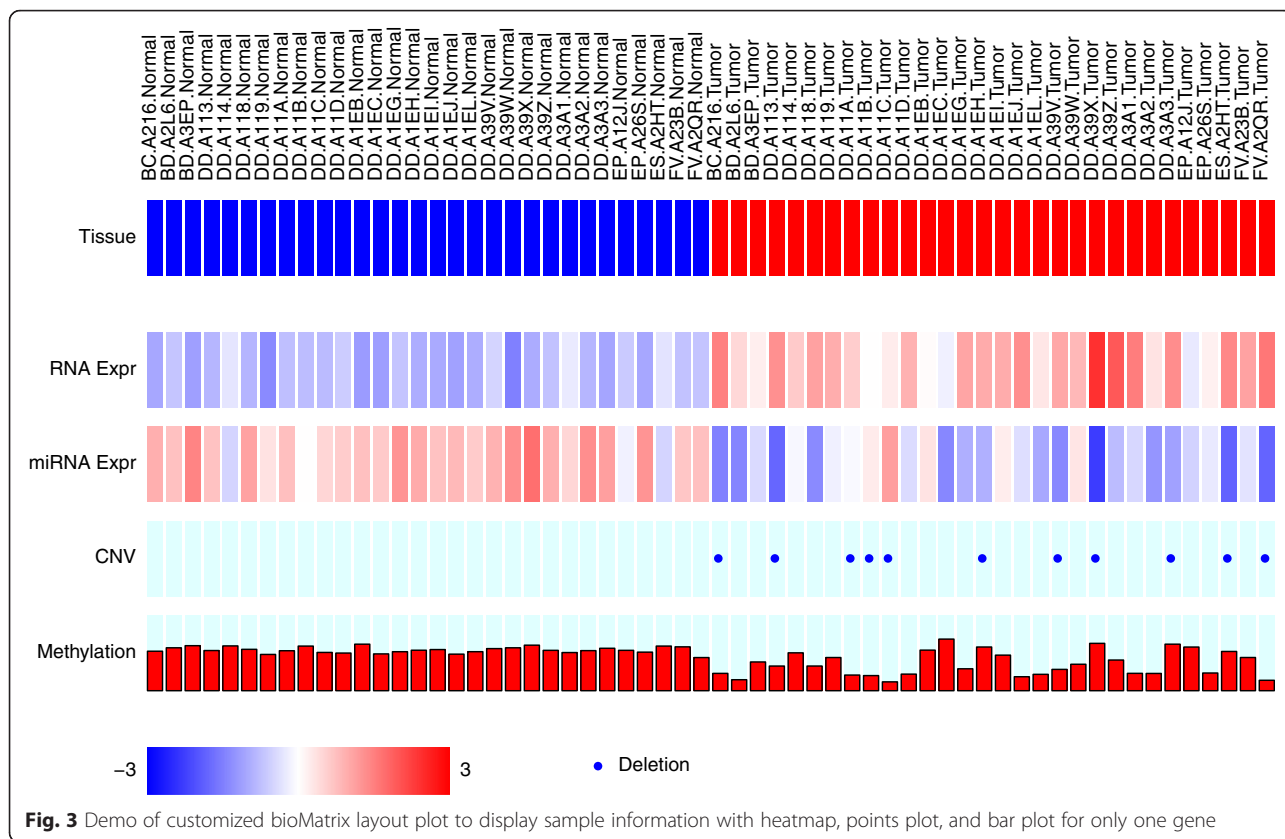
Default plot methods for bioNetCircos plot

```
library(caOmicsV)
data(bionetPlotDemoData)
plotBioNetCircos(bionetPlotDemoData)
```

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

The input data format for both bioMatrix and bioNetCircos layout plot is a list of data matrix and character vectors. A function getESet() was implemented in caOmicsV package to build the input data list, which take a given set of gene names, sample names, and plot data in data frame format as input and returns a list containing all plot data. To help in preparing input dataset for getESet() function, a set of supporting functions are also provided in the package including of methods of extracting subset data from big data set with given gene names and sample names as well as sorting datasets for desired orders required for the plot methods.

Beside of the default plot methods, caOmicsV package can also allow users to generate customized images with each specific plot function. Figure 3 is a demo of customized bioMatrix plot which displays sample information,



mRNA expression, miRNA expression, DNA CNV, and methylation for one gene. Also, caOmicsV plot are implemented with low level plot methods of R graphic package, more decoration items such as title and extra legend could be easily added onto the plot outputs. With the R graphic layout supporting, multiple caOmicsV plots could be generated on one image. In addition, caOmicsV plot functions use data matrix as input format, other data other than genomic data, when correctly formatted, may also be plotted with caOmicsV package.

Conclusions

caOmicsV package provides sample way to present integrated multidimensional genomic data under R environment with both matrix layout and circus layout on biological network.

Availability and requirements

- Project name:** caOmicsV
- Project home page:** <https://www.bioconductor.org/packages/caOmicsV>
- Operating systems:** any operating system supporting R
- Programming language:** R
- Other requirements:** working R installation
- Licence:** GPL

Any restriction to use by non-academics: none

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

HZ and SD designed and implemented the software package, and wrote manuscript. PM revised the manuscript critically. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by NCI intramural research funding in Center for Cancer Research, National Cancer Institute, National Institutes of Health, and the package was developed and tested on Helix system at the Center for Information Technology, National Institutes of Health, Bethesda, Maryland, USA. The data used for implementation and demo of caOmicsV package is based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

Received: 18 November 2015 Accepted: 14 March 2016

Published online: 22 March 2016

References

1. The Cancer Genome Atlas (TCGA). <http://cancergenome.nih.gov/>. Accessed 2 May 2015.
2. International Cancer Genome Consortium. <https://icgc.org/>. Accessed 2 May 2015.
3. Nielsen CB, Cantor M, Dubchak I, Gordon D, Wang T. Visualizing genomes: techniques and challenges. *Nat Methods*. 2010;7:55–S15.
4. Schroeder MP, Gonzalez-Perez A, Lopez-Bigas N. Visualizing multidimensional cancer genomics data. *Genome Med*. 2013;5:9.

5. Wang R, Perez-Riverol Y, Hermjakob H, Vizcaino JA. Open source libraries and frameworks for biological data visualisation: A guide for developers. *Proteomics*. 2015;15:1356–74.
6. The UCSC Cancer Genomics Browser. <https://genome-cancer.ucsc.edu>. Accessed 2 May 2015.
7. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–92.
8. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–45.
9. Zhang H, Davis S, Meltzer PS. RCircos: an R package for Circos 2D track plots. *BMC Bioinform*. 2013;14:244.
10. An J, Lai J, Sajjanhar A, Batra J, Wang C, Nelson CC. J-Circos: an interactive Circos plotte. *Bioinformatics*. 2015;31:1463–5.
11. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Disco*. 2012;2:401–4.
12. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics*. 2010;26:i237–45.
13. Wong CK, Vaske CJ, Ng S, Zachary Sanborn J, Ben S, Haussler D, Stuart JM. The UCSC Interaction Browser: multidimensional data views in pathway context. *Nucleic Acids Res*. 2013;41:W218–24.
14. Yin T, Cook D, Lawrence M. ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biology*. 2012;13:R77.
15. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30:2811–2.
16. Hu Y, Yan C, Hsu CH, Chen QR, Niu K, Komatsoulis GA, Meerzaman D. OmicCircos: A Simple-to-Use R Package for the Circular Visualization of Multidimensional Omics Data. *Cancer Inform*. 2014;13:13–20.
17. Gu Z. ComplexHeatmap: Making Complex Heatmaps. R package version 1.0.0, <https://github.com/jokergoo/ComplexHeatmap>. Accessed 2 May 2015.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

