OXFORD

Genome analysis

chromPlot: visualization of genomic data in chromosomal context

Karen Y. Oróstica and Ricardo A. Verdugo*

Programa de Genética Humana ICBM, Facultad de Medicina, Universidad de Chile, Independencia 1027, Santiago, Chile

*To whom correspondence should be addressed.
Associate Editor: John Hancock

Abstract

Summary: Visualizing genomic data in chromosomal context can help detecting errors in data processing and may suggest new hypotheses to be tested. Here, we report a new tool for displaying large and diverse genomic data along chromosomes. The software is implemented in R so that visualization can be easily integrated with its numerous packages for processing genomic data. It supports simultaneous visualization of multiple tracks of data. Large genomic regions such as QTLs or synteny tracts may be shown along histograms of number of genes, genetic variants, or any other type of genomic element. Tracks can also contain values for continuous or categorical variables and the user can choose among points, connected lines, colored segments, or histograms for representing data. chromPlot takes data from tables in data.frame in GRanges formats. The information necessary to draw chromosomes for mouse and human is included with the package. For other organisms, chromPlot can read Gap and cytoBandIdeo tables from the UCSC Genome Browser. We present common use cases here, and a full tutorial is included as the package's vignette.

Availability and Implementation: chromPlot is distributed under a GLP2 licence at http://www.bioconductor.org.

Contact: raverdugo@u.uchile.cl

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

When analyzing large amounts of genomic data, visualization in chromosomal context can reveal chromosome-specific or genome-wide patterns suggesting new hypotheses and their appropriate tests. In addition, it may show errors in data generation or in an analysis step. We distinguish between two types of genomic visualization: local and global. Local visualization displays data in a region within one chromosome. Global visualization displays data in whole chromosomes. Tools available for local visualization are very flexible, allowing highly customized plots with diverse data, e.g. GenomeGraphs (Durinck *et al.*, 2009) and GenoPlotR (Guy *et al.*, 2010) R packages and the Integrative Genomics Viewer (IGV) (Thorvaldsdóttir *et al.*, 2013).

Global Visualization has been incorporated in websites of cross-species repositories, e.g. UCSC, Ensembl and NCBI, and in species-specific websites, such as or the Chromosome Visualization Tool (Cvit)

(Cannon *et al.*, 2011) and Gviewer (gmod.org/wiki/Flash_GViewer). Standalone tools are task-sepecific, e.g. PhenoGram (Wolfe *et al.*, 2013) represents genotype-by-phenotype associations. Quantsmooth displays discrete and continuous data in the local view, and large segments in the global view. The IdeoViz R-package can plot continuous values as lines or bars for binned data (Pai and Ren, 2014) but has no support for categorical data. Thus, there is a need for a generic tool for global visualization to mach flexibility of local visualizers of genomic data.

2 Software

The chromPlot package was developed in R and uses builtin functions for data processing and plotting using points, lines, bars or colored segments. It automatically finds clusters of overlapping segments and stacks them for saving plotting space. When segments are too numerous, e.g. all genome's genes, it uses histograms.

3 Program overview

3.1 Input data

chromPlot takes data as R objects of class data.frame or GRanges. Data frames must include the columns 'Chrom', 'Start', 'End'. A 'Group' column can be added to classify elements in categories that will be differentiated by color.

3.2 Usage

The user interacts with a single main function called chromPlot with 50 arguments. The gaps argument uses the 'gap' table from the UCSC Table Browser (https://genome.ucsc.edu/) to set a system of genomic coordinates: number of chromosomes, chromosomes names and lengths and the position of centromeres (shown as solid circles). The tables for human and mouse are provided with the package. The following arguments take genomic data:

bands: data is shown as colored bands on chromosomal bodies. It detects cytoBandIdeo tables from the UCSC genome browser and shows them as an idiogram. Tables for human and mouse idiograms are included in the package.

annot1–4: they are plotted sequentially as histograms per genomic bins (1 Mb by default). This is ideal for genomic elements that are subsets of each other. For example, genes in the genome, in an array, detected and differentially expressed, respectively. A track of all genes in the genome can be directly obtained from the Ensembl repository by providing an organism name to the org argument.

segment and segment2: tracks of large segments plotted as colored lines on either side of chromosomes. Up to two categories for each segment are supported ('Group' and 'Group2' columns). Segment categories are distinguished by color and a shape in the middle. If segments are smaller than the size of bins or if they are too many (>200 by default), they are plotted as histograms, but this can be changed by setting the noHist argument to TRUE. Stacked histograms are used for categorized segments.

stats and stats2 display continuous data in XY plots along chromosomes. A column containing values must be indicated with the statCol argument. The user can choose between connected lines or points by setting statsTyp to 'l', 'p' or 'n' to omit. If an ID column is present, labels are printed, avoiding overlapping among labels.

Example command for creating the plot in Figure 1:

chromPlot(gaps=mm9_gap, bands=region, annot1=genes, annot2=array, annot3=filtered, annot4=selected, segment=qtl, chr=c(2, 11, 17))

4 Results

chromPlot creates karyotype-like graphs and the data can be plotted on three areas: to the left, to the right and on the chromosome. Multiple tracks may be overlaid or shown on different chromosomal sides. The user can generate a virtually infinite number of possible graphs. This is a list of common uses for the package:

Synteny plots: the genomic regions of homology are represented by colored segments within chromosomes. Data can be read in AXT format or as a table.

Differentially expressed genes: as histograms in chromosomal context may reveal clusters of differential expression. This may suggest effects from regional context such as chromatin state, epigenetics regulation or cis-eQTL as in Figure 1.

QTL mapping: QTLs can be represented as LOD curves or chromosomal segments. It is possible to integrate such information with results from *omics* experiments, such as differential expression (Fig. 1).
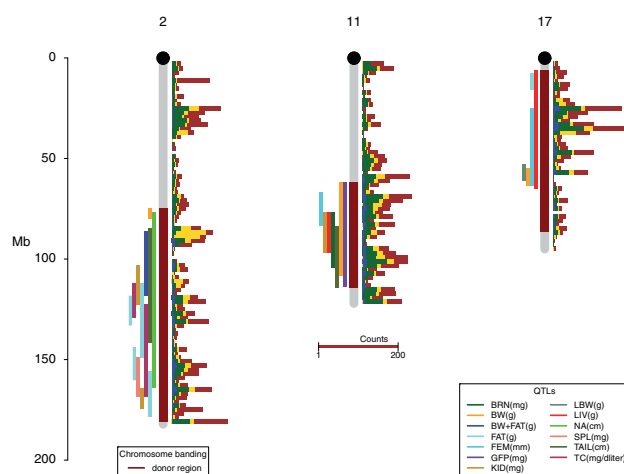
Fig. 1. **Differential expression and QTLs in three mouse congenic strains**. QTLs are colored segments, and number of genes in 2 Mbp bins as histograms: brown, total genes; yellow, genes in the microarray; green, expressed genes; blue, differentially expressed genes. Brown segments on chromosome bodies denote regions from a donor strain. Legends are placed under the shortest chromosomes by default. Source (Verdugo *et al.*, 2010)

GWAS: genomewide association studies in outbred populations can be summarized as, for instance, average $-\log 10(p)$ values by chromosomal bins. Similar plots may be created to summarize genomewide genetic divergence (e.g. *Fst*) or measures of positive selection like *iHS*.

Genome sequencing: chromPlot can be used for quality controlling or summarizing results from whole-genome sequencing experiments. For instance, it is possible to plot genome coverage as connected-lines or number of detected variants as histograms.

Local ancestry: chromosomes of admixed individuals can be painted by genetic ancestry on the body (bands) or on either side (segment and segment2).

Physical maps of genes or makers: use the stat argument, omitting the statCol argument and providing an 'ID' column.

5 Discussion and conclusion

chromPlot is a generic and highly flexible tool for global visualization of genomic data. It is able to plot large datasets for different scientific needs. It was developed in R to facilitate inclusion in genomic workflows of data analysis. This feature makes chromPlot a more versatile software compared to other global visualization tools currently available.

Acknowledgements

Funding

References

Cannon,E.K.S. *et al*. (2011) Chromosome visualization tool: a whole genome viewer. *Int J Plant Genomics*, **2011**. doi:10.1155/2011/373875.

Durinck,S. *et al.* (2009) GenomeGraphs: integrated genomic data visualization with R. *BMC Bioinformatics*, **10**(Jan.), 2.

Guy,L. *et al.* (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics*, **26**, 2334–2335.

Pai,S. and Ren,J. (2014) IdeoViz: plots data (continuous/discrete) along chromosomal ideogram. *R package version 1.4.0.*

Thorvaldsdóttir,H. *et al.* (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinf.*, **14**, 178–192.

Verdugo,R.A. *et al.* (2010) Serious limitations of the QTL/Microarray approach for QTL gene discovery. *BMC Biology*, **8**, 96.

Wolfe,D. *et al.* (2013) Visualizing genomic information across chromosomes with PhenoGram. *BioData Mining*, **6**, 18.