Plotting discrete or continuous datasets in the context of chromosomal location has several useful applications in genomic analysis. Examples of possible metrics include RNA expression levels, densities of epigenetic marks or genomic variation, while applications could range from the analysis of a single variable in a single context, to multiple measurements in several biological contexts (e.g. age/sex/tissue/disease context). Visualization of metrics superimposed on the chromosomal ideogram could provide varied insights into the metric of interest:

1. It could identify distinctive spatial distribution that could further hypotheses about the functional role of the metric (e.g. telocentric or pericentromeric enrichment)
2. It could highlight distribution differences between different groups of samples, suggesting different regulatory mechanisms; in extreme cases, visualization may identify large genomic foci of differences
3. It could confirm that a quantitative difference measured between groups of interest is consistent throughout the genome (i.e. that there are no foci, and that the change is global).

This package provides a method to plot one or several datasets against the chromosomal ideogram. It provides some simple options (vertical/horizontal orientation, display in bars or line graphs). Data are expected to be binned; IdeoViz provides a function for user-specified bin widths. Ideograms for the genome of choice can also be automatically downloaded from UCSC using the getIdeo() function.

1 Setup

> require(IdeoViz)
> require(RColorBrewer) ### nice colours
> data(binned_multiSeries)
Example 1: Plotting several trendlines along one ideogram

The ideogram table containing cytogenetic band information is used to render chromosomes. This table corresponds directly to the cytoBandIdeo table from the UCSC genome browser. There are two ways to supply an ideogram table to plotOnIdeo():

1. First, it can be automatically downloaded from UCSC for your genome of choice, using the getIdeo() function.

2. Alternately, a pre-downloaded cytoBandIdeo table can be provided to downstream functions such as plotOnIdeo().

In this case, the table must be provided as a data.frame object with a header row and the column order matching that of the cytoBandIdeo() table at UCSC.

```r
> ideo <- getIdeo("hg18")
> head(ideo)

<table>
<thead>
<tr>
<th>chrom</th>
<th>chromStart</th>
<th>chromEnd</th>
<th>name</th>
<th>gieStain</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1</td>
<td>0</td>
<td>2300000</td>
<td>p36.33</td>
<td>gneg</td>
</tr>
<tr>
<td>chr1</td>
<td>2300000</td>
<td>5300000</td>
<td>p36.32</td>
<td>gpos25</td>
</tr>
<tr>
<td>chr1</td>
<td>5300000</td>
<td>7100000</td>
<td>p36.31</td>
<td>gneg</td>
</tr>
<tr>
<td>chr1</td>
<td>7100000</td>
<td>9200000</td>
<td>p36.23</td>
<td>gpos25</td>
</tr>
<tr>
<td>chr1</td>
<td>9200000</td>
<td>12600000</td>
<td>p36.22</td>
<td>gneg</td>
</tr>
<tr>
<td>chr1</td>
<td>12600000</td>
<td>16100000</td>
<td>p36.21</td>
<td>gpos50</td>
</tr>
</tbody>
</table>
```

```r
> plotOnIdeo(chrom=seqlevels(binned_multiSeries), # which chrom to plot?
>             ideogram=ideo, # ideogram name
>             values_GR=binned_multiSeries, # data goes here
>             value_cols=colnames(mcols(binned_multiSeries)), # col to plot
>             col=brewer.pal(n=5, 'Spectral'), # colours
>             val_range=c(0,10), # set y-axis range
>             ylab="array intensities",
>             plot_title="Trendline example")
```

Trendline example

![Trendline example graph](image-url)
Example 2: Plotting a single series in bar format

For this example, we specify a local file to obtain the chromosome ideograms, rather than having IdeoViz download it from UCSC.

```r
> data(binned_singleSeries)
> data(hg18_ideo) # cytoBandIdeo table downloaded previously and stored as a data.frame.
> plotOnIdeo(chrom=seqlevels(binned_singleSeries),
+     ideo=hg18_ideo,
+     values_GR=binned_singleSeries,
+     value_cols=colnames(mcols(binned_singleSeries)),
+     plotType='rect', # plot as bars
+     col='blue', vertical=T,
+     val_range=c(-1,1), ylab="dummy score",
+     plot_title="Discretized example")
```

Discretized example
Example 3: Plotting a single series in bar format along entire genome

```r
> data(binned_fullGenome)
> plotOnIdeo(chrom=seqlevels(binned_fullGenome),
+     ideo=ideo,
+     values_GR=binned_fullGenome,
+     value_cols=colnames(mcols(binned_fullGenome)),
+     plotType='rect',
+     col='orange', addScale=F, # hide scale to remove visual clutter
+     plot_title="Whole genome view",
+     val_range=c(-1,1),cex.axis=0.5,chromName_cex=0.6)
```
3 Example 4: Binning data using IdeoViz functions

In this example, we do everything in IdeoViz: download the ideogram from UCSC, bin the data, and finally, plot along chromosomes. For the example, we use histone H3K9me3 peak intensities mapped in the human lymphoblastoid cell line.
GM12878 (GEO accession GSM733664, only 3 chromosomes shown for simplicity). Here, average peak signal is plotted in 500Kb bins along the chromosome. The ideogram plots show high signal in pericentromeric and telomeric regions, consistent with the association of this histone mark with heterochromatin.


```r
ideo_hg19 <- getIdeo("hg19")
chroms <- c("chr1","chr2","chrX")
data(GSM733664_broadPeaks)
head(GSM733664_broadPeaks)

  chrom chromStart chromEnd name score strand signalValue pValue qValue
1  chr1     10141   10374 .  993 .  10.796883 10.3 -1
2  chr1     567457  567702 . 1000 .  16.590333 100.0 -1
3  chr1     569826  570047 . 1000 .  15.757614 100.0 -1
4  chr1     723167  727602 .  808 .  8.389733  14.2 -1
5  chr1     816959  817136 .  793 .  8.188648  1.7 -1
6  chr1     821181  821421 .  753 .  7.660859  3.4 -1

chrom_bins <- getBins(chroms, ideo_hg19, stepSize=5*100*1000)
avg_peak <- avgByBin(data.frame(value=GSM733664_broadPeaks[,7]),
                       GSM733664_broadPeaks[,1:3], chrom_bins)
plotOnIdeo(chrom=seqlevels(chrom_bins),
           ideoTable=ideo_hg19,
           values_GR=avg_peak, value_cols='value',
           val_range=c(0,50),
           plotType='rect',
           col='blue', vertical=T)
```

![Image of ideogram plots showing high signal in pericentromeric and telomeric regions.](image.png)
Session info

> sessionInfo()

R Under development (unstable) (2014-10-07 r66723)
Platform: x86_64-unknown-linux-gnu (64-bit)

locale:
[1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C
[3] LC_TIME=en_US.UTF-8 LC_COLLATE=C
[5] LC_MONETARY=en_US.UTF-8 LC_MESSAGES=en_US.UTF-8
[7] LC_PAPER=en_US.UTF-8 LC_NAME=C
[9] LC_ADDRESS=C LC_TELEPHONE=C

attached base packages:

other attached packages:
[1] IdeoViz_1.1.0 rtracklayer_1.27.0 RColorBrewer_1.0-5
[4] GenomicRanges_1.19.0 GenomeInfoDb_1.3.0 IRanges_2.1.0
[7] S4Vectors_0.5.0 Biobase_2.27.0 BiocGenerics_0.13.0

loaded via a namespace (and not attached):
[1] BBmisc_1.7 BatchJobs_1.4 BiocParallel_1.1.0
[4] Biostrings_2.35.0 DBI_0.3.1 GenomicAlignments_1.3.0
[7] RCurl_1.95-4.3 RSQLite_0.11.4 Rsamtools_1.19.0
[10] XML_3.98-1.1 XVector_0.7.0 base64enc_0.1-2
[13] bitops_1.0-6 brew_1.0-6 checkmate_1.4
[16] codetools_0.2-9 digest_0.6.4 fail_1.2
[19] foreach_1.4.2 iterators_1.0.7 sendmailR_1.2-1
[22] stringr_0.6.2 tools_3.2.0 zlibbioc_1.13.0