

Knitr / Rmarkdown example using *GenomicInteractions*

Liz Ing-Simmons

Contents

Introduction	1
Making a GenomicInteractions object	1
Annotation	3
References	4
Session details	5

Introduction

This vignette shows you how `GenomicInteractions` can be used to investigate significant interactions in HiC data that has been analysed using [HOMER](#) software [1]. `GenomicInteractions` can take a HOMER [interaction file](#) as input.

HiC data comes from [chromosome conformation capture](#) followed by high-throughput sequencing. Unlike 3C, 4C or 5C, which target specific regions, it can provide genome-wide information about the spatial proximity of regions of the genome. The raw data takes the form of paired-end reads connecting restriction fragments. The resolution of a HiC experiment is limited by the number of paired-end sequencing reads produced and by the sizes of restriction fragments. To increase the power to distinguish real interactions from random noise, HiC data is commonly analysed in bins from 20kb - 1Mb. There are a variety of tools available for binning the data, controlling for noise (e.g. self-ligations of restriction fragments), and finding significant interactions.

The data we are using comes from [this paper](#) [2] and can be downloaded from [GEO](#). It is HiC data from wild type mouse double positive thymocytes. The experiment was carried out using the HindIII restriction enzyme. The paired-end reads were aligned to the mm9 mouse genome assembly and HOMER software was used to filter reads and detect significant interactions at a resolution of 100kb. For the purposes of this vignette, we will consider only data from chromosomes 14 and 15.

Making a GenomicInteractions object

Load the data by specifying the file location and experiment type. You can also include an experiment name and description.

The `GenomicInteractions` object consists of two linked `GenomicRanges` objects containing the anchors of each interaction, and the number of reads supporting each interaction. Metadata for each interaction (in this case, p-values and FDRs) is stored as a `DataFrame` accessed by `mcols()` or `elementMetadata()`, similar to the metadata of a simple `GRanges`. You can also access single metadata columns using `$`.

```
## GenomicInteractions object with 8171 interactions and 15 metadata columns:
## Name: HiC 100kb
## Description: HiC 100kb resolution
## Sum of interactions: 284294
## Annotated: no
```

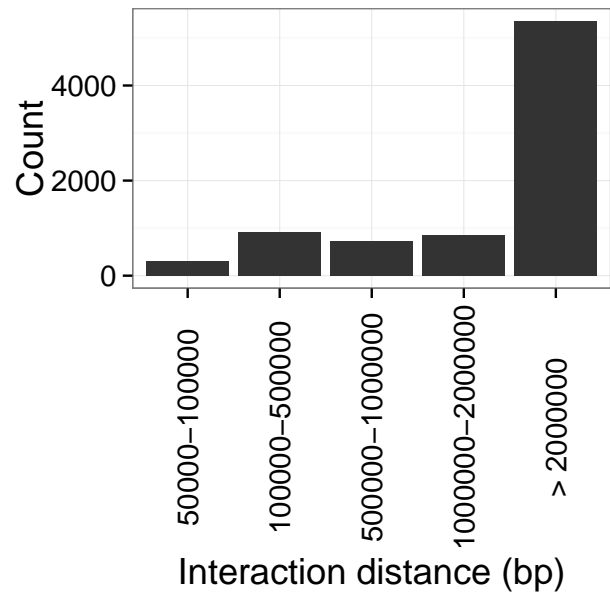
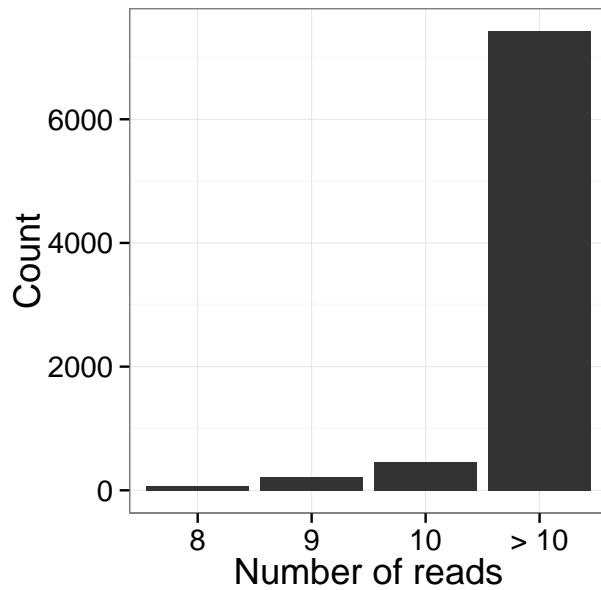
```

## Interactions:
##
##           Anchor One                      Anchor Two Counts
##   [1] chr15:97600000..97699999 --- chr15:97500000..97599999 344
##   [2] chr15:74800000..74899999 --- chr15:74700000..74799999 373
##   [3] chr14:55000000..55099999 --- chr14:54800000..54899999 258
##   [4] chr15:80400000..80499999 --- chr15:80300000..80399999 397
##   [5] chr14:55100000..55199999 --- chr14:54800000..54899999 213
##   ...
## [8167] chr14:64400000..64499999 --- chr14:56600000..56699999 12
## [8168] chr15:93200000..93299999 --- chr15:51700000..51799999 11
## [8169] chr15:103400000..103490927 --- chr15:95300000..95399999 12
## [8170] chr15:62600000..62699999 --- chr15:60900000..60999999 24
## [8171] chr14:80200000..80299999 --- chr14:74800000..74899999 13
##
##           | InteractionID      PeakID.1. strand.1. Total.Reads.1.
##   [1] | interaction66 chr15-97600000      +          7144
##   [2] | interaction94 chr15-74800000      +          8002
##   [3] | interaction103 chr14-55000000     +          7617
##   [4] | interaction118 chr15-80400000     +          9403
##   [5] | interaction122 chr14-55100000     +          6775
##   ...
## [8167] | interaction93141 chr14-64400000     +         11104
## [8168] | interaction93171 chr15-93200000     +         12050
## [8169] | interaction93173 chr15-103400000    +          8926
## [8170] | interaction93174 chr15-62600000     +         11155
## [8171] | interaction93186 chr14-80200000     +          9904
##
##           PeakID.2. strand.2. Total.Reads.2. Distance Expected.Reads
##   [1] chr15-97500000      +          8598    80527    59.66314
##   [2] chr15-74700000      +         11112    93528    79.84362
##   [3] chr14-54800000      +         11577   198082    37.47250
##   [4] chr15-80300000      +         11387    80980    94.90884
##   [5] chr14-54800000      +         11577   298783    25.45554
##   ...
## [8167] chr14-56600000      +          9067   7795028    2.687874
## [8168] chr15-51700000      +         11113  41526951    2.261388
## [8169] chr15-95300000      +         10042   8094968    2.688256
## [8170] chr15-60900000      +         12466  1695366    9.023079
## [8171] chr14-74800000      +         10361   5386967    3.136097
##
##           Z.score      LogP FDR.Benjamini..based.on.3.68e.08.total.tests.
##   [1] 4.887464 -327.7435                                0
##   [2] 4.303700 -290.9446                                0
##   [3] 3.648479 -284.0124                                0
##   [4] 3.996785 -274.6071                                0
##   [5] 3.558349 -271.0352                                0
##   ...
## [8167] 1.536492 -10.58734                                0.099790
## [8168] 1.617159 -10.58611                                0.099881
## [8169] 1.536628 -10.58605                                0.099885
## [8170] 1.174996 -10.58604                                0.099885
## [8171] 1.487317 -10.58560                                0.099916
##
##           Circos.Thickness      p.value      fdr
##   [1] 30 4.600386e-143          0
##   [2] 26 4.409070e-127          0
##   [3] 26 4.518538e-124          0
##   [4] 24 5.491174e-120          0

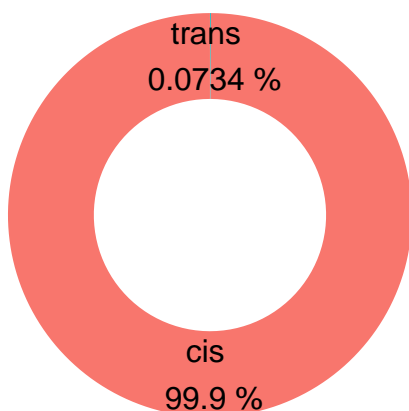
```

```
##      [5]          24 1.953899e-118          0
##      ...          ...          ...
##      [8167]          2 2.523340e-05 0.099790
##      [8168]          2 2.526451e-05 0.099881
##      [8169]          2 2.526592e-05 0.099885
##      [8170]          2 2.526620e-05 0.099885
##      [8171]          2 2.527740e-05 0.099916
##
##      -----
##      seqinfo: 2 sequences from an unspecified genome
```

HiC 100kb



Cis / Trans Percentages



Annotation

One of the most powerful features of GenomicInteractions is that it allows you to annotate interactions by whether the anchors overlap genomic features of interest, such as promoters or enhancers. However this

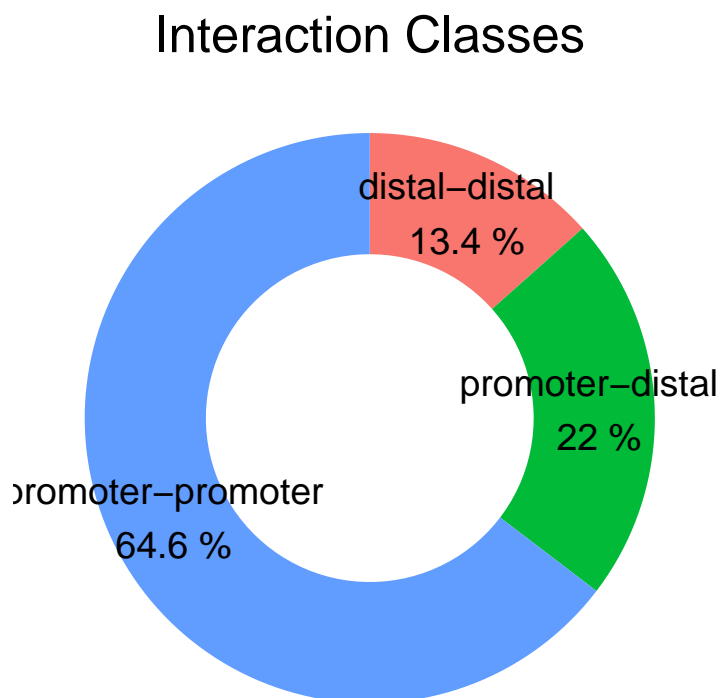
process can be slow for large datasets, so here we annotate with just promoters.

Genome annotation data can be obtained from, for example, UCSC databases using the GenomicFeatures package. We will use promoters of Refseq genes extended to a width of 5kb. Downloading all the data can be a slow process, so the data for promoters for chromosomes 14 and 15 is provided with this package.

`annotateInteractions` takes a list of features in GRanges or GRangesList format and annotates the interaction anchors based on overlap with these features. The list of annotation features should have descriptive names, as these names are stored in the annotated GenomicInteractions object and used to assign anchor (node) classes.

Interaction types

Interaction types are determined by the classes of the interacting nodes. As we only have two node classes, we have three possible interaction classes, summarised in the plot below. Most of the interactions are between promoters. We can subset the data to look at interaction types that are of particular interest.



You can also summarise interaction types in a table.

category	count
promoter-promoter	5281
promoter-distal	1794
distal-distal	1096

References

1. Heinz S, Benner C, Spann N, Bertolino E et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell 2010 May 28;38(4):576-589.
2. Seitan, V. C. et al. Cohesin-based chromatin interactions enable regulated gene expression within

pre-existing architectural compartments. Genome Res. 23, 2066-77 (2013).

Session details

This report was generated on Thu Oct 01 2015 at 19:02:26.

```
## R version 3.2.2 (2015-08-14)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: OS X 10.10.5 (Yosemite)
##
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] stats4      parallel    stats      graphics  grDevices  utils      datasets
## [8] methods     base
##
## other attached packages:
## [1] GenomicRanges_1.21.28      GenomeInfoDb_1.5.15
## [3] IRanges_2.3.21             S4Vectors_0.7.18
## [5] BiocGenerics_0.15.6        GenomicInteractions_1.3.6
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.1                biovizBase_1.17.2
## [3] lattice_0.20-33            Rsamtools_1.21.17
## [5] Biostrings_2.37.8          assertthat_0.1
## [7] digest_0.6.8              R6_2.1.1
## [9] plyr_1.8.3                 chron_2.3-47
## [11] futile.options_1.0.0       acepack_1.3-3.3
## [13] RSQLite_1.0.0              evaluate_0.8
## [15] highr_0.5.1                ggplot2_1.0.1
## [17] zlibbioc_1.15.0            GenomicFeatures_1.21.26
## [19] data.table_1.9.6           rpart_4.1-10
## [21] rmarkdown_0.8              proto_0.3-10
## [23] labeling_0.3               splines_3.2.2
## [25] BiocParallel_1.3.52        stringr_1.0.0
## [27] foreign_0.8-66             igraph_1.0.1
## [29] RCurl_1.95-4.7             biomaRt_2.25.3
## [31] munsell_0.4.2              rtracklayer_1.29.27
## [33] Gviz_1.13.7                htmltools_0.2.6
## [35] nnet_7.3-11                SummarizedExperiment_0.3.9
## [37] gridExtra_2.0.0            Hmisc_3.17-0
## [39] codetools_0.2-14           matrixStats_0.14.2
## [41] XML_3.98-1.3               dplyr_0.4.3
## [43] GenomicAlignments_1.5.17   MASS_7.3-44
## [45] bitops_1.0-6              grid_3.2.2
## [47] gtable_0.1.2              DBI_0.3.1
## [49] magrittr_1.5               formatR_1.2.1
## [51] scales_0.3.0              stringi_0.5-5
## [53] XVector_0.9.4             reshape2_1.4.1
## [55] latticeExtra_0.6-26        futile.logger_1.4.1
## [57] Formula_1.2-1             lambda.r_1.1.7
## [59] RColorBrewer_1.1-2         tools_3.2.2
```

## [61]	dichromat_2.0-0	BSgenome_1.37.5
## [63]	Biobase_2.29.1	survival_2.38-3
## [65]	yaml_2.1.13	AnnotationDbi_1.31.18
## [67]	colorspace_1.2-6	cluster_2.0.3
## [69]	knitr_1.11	VariantAnnotation_1.15.29