Human Fibroblast IMR90 Hi-C Data (Dixon et al.)

Nicolas Servant

November 16, 2016

1 Introduction

The Hi-C technic was first introduced by Lieberman-Aiden et al. [2009]. In the continuity with 3C, 4C and 5C technics, the goal of the Hi-C is to simultaneoulsy detect all chromosomal contacts in a single experiment. All this technics aim at measuring the population-averaged frequency at which two genomic loci physically interact in three-dimensional space. In Hi-C, after a first crosslink and digestion, all genomic fragments are labeled with a biotinylated nucleotide before ligation. These junctions can then be purified efficiently by streptavidin-coated magnetic beads, and finally sequenced using a standard Illumina paired-end protocol.

The data available in this package were published by Dixon et al. [2012] and downloaded from the GEO website (GSE35156, sample GSM862724). This publication is one of the key paper in the field for two main reasons: i) it was the first time than Hi-C data were generated at such resolution (up to 20kb), ii) this resolution highlighted a new short range structure defined as topological domains (TADs), with high frequencies of intra-domain chromatin interactions but infrequent inter-domain chromatin interactions (Nora et al. [2012]).

If you use HiCDataHumanIMR90, please cite:


2 Hi-C Data

The hic_imr90_40 object is a HTClist object (see the HiTC package for more information (Servant et al. [2012])). It contains the complete genome-wide HiC data, with all inter and intrachromosomic contact maps at a resolution of 40kb.

> require(HiCDataHumanIMR90)
> require(HiTC)
> data(Dixon2012_IMR90)
> ## Show data
> show(hic_imr90_40)

HTClist object of length 325
25 intra / 300 inter-chromosomal maps

> ## Is my data complete (i.e. composed of intra + inter chromosomal maps)
> isComplete(hic_imr90_40)
Human Fibroblast IMR90 Hi-C Data (Dixon et al.)

[1] TRUE

> ## Note that a complete object is not necessarily pairwise
> ## (is both chr1-chr2 and chr2-chr1 stored ?)
> isPairwise(hic_imr90_40)

[1] FALSE

> ## Which chromosomes ?
> seqlevels(hic_imr90_40)

[1] "chr1" "chr2" "chr3" "chr4" "chr5" "chr6" "chr7" "chr8" "chr9" "chr10"
[11] "chr11" "chr12" "chr13" "chr14" "chr15" "chr16" "chr17" "chr18" "chr19" "chr20"
[21] "chr21" "chr22" "chrX" "chrY" "chrM"

> ## Details about a given map
> detail(hic_imr90_40$chrXchrX)

HTC object
Focus on genomic region [chrX:1-155270560]
CIS Interaction Map
Matrix of Interaction data: [3882-3882]
Binned data - window size = 40000
3882 genome intervals
Total Reads = 15349610
Number of Interactions = 3362484
Median Frequency = 1
Sparsity = 0.112

> ## Descriptive statistics
> head(summary(hic_imr90_40))

    seq1 seq2 nbreads nbinteraction averagefreq medfreq sparsity
chr1chr1 chr1 chr1 25914788 4524734 5.7274 1 0.8835
chr1chr2 chr1 chr2 504332 497291 1.0142 1 0.9869
chr1chr3 chr1 chr3 440865 434917 1.0137 1 0.9859
chr1chr4 chr1 chr4 456924 450005 1.0154 1 0.9849
chr1chr5 chr1 chr5 399067 393926 1.0131 1 0.986
chr1chr6 chr1 chr6 382580 377654 1.013 1 0.9858

3 Topological Domains

The tads_imr90 object is a GRanges object with a all TADs detected from this Hi-C data.

> show(tads_imr90)

GRanges object with 2338 ranges and 0 metadata columns:

<table>
<thead>
<tr>
<th>seqnames</th>
<th>ranges</th>
<th>strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;Rle&gt;</td>
<td>&lt;IRanges&gt;</td>
<td>&lt;Rle&gt;</td>
</tr>
<tr>
<td>TAD-1</td>
<td>chr1 [ 770138, 1290137]</td>
<td>*</td>
</tr>
<tr>
<td>TAD-2</td>
<td>chr1 [1290138, 1850140]</td>
<td>*</td>
</tr>
<tr>
<td>TAD-3</td>
<td>chr1 [1850141, 2330140]</td>
<td>*</td>
</tr>
<tr>
<td>TAD-4</td>
<td>chr1 [2330141, 3610140]</td>
<td>*</td>
</tr>
<tr>
<td>TAD-5</td>
<td>chr1 [3770141, 6077413]</td>
<td>*</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Human Fibroblast IMR90 Hi-C Data (Dixon et al.)

TAD-2334 chrX [146992309, 148552096] *
TAD-2335 chrX [148592096, 149929342] *
TAD-2336 chrX [149929343, 151969344] *
TAD-2337 chrX [152089345, 152746806] *
TAD-2338 chrX [152786807, 154946806] *

seqinfo: 23 sequences from an unspecified genome; no seqlengths

> ## Extract region
> regx <- extractRegion(hic_imr90_40$chrXchrX,
> +   chr="chrX", from=95000000, to=105000000)
> ## Plot Hi-C data with TADs
> plot(regx, tracks=list(tads_imr90), maxrange=20)

Package versions

This vignette was generated using the following package versions:

- R version 3.3.2 (2016-10-31), x86_64-pc-linux-gnu
- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, stats4, utils
- Other packages: BiocGenerics 0.20.0, GenomeInfoDb 1.10.1, GenomicRanges 1.26.1, HiCDataHumanIMR90 0.108.0, HiTC 1.18.0, IRanges 2.8.1, S4Vectors 0.12.0
- Loaded via a namespace (and not attached): Biobase 2.34.0, BiocParallel 1.8.1, BiocStyle 2.2.0, Biostrings 2.42.0, GenomicAlignments 1.10.0, Matrix 1.2-7.1, RColorBrewer 1.1-2, RCurl 1.95-4.8, Rsamtools 1.26.1, SummarizedExperiment 1.4.0, XML 3.98-1.5, XVector 0.14.0, bitops 1.0-6, grid 3.3.2, lattice 0.20-34, rtracklayer 1.34.1, tools 3.3.2, zlibbioc 1.20.0
References


