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

Nicolas Servant

October 27, 2024

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 finaly 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:

- Servant N (2014). HiCDataHumanIMR90: Human Fibroblast IMR90 HiC data from Dixon et al. 2012. R package version 1.1.0.
- Dixon JR, Selvaraj S, Yue F, Kim A et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398):376-80.

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)
```

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

```
> ## 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)

[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"
[10] "chr10" "chr11" "chr12" "chr13" "chr14" "chr15" "chr16" "chr17" "chr18"
[19] "chr19" "chr20" "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

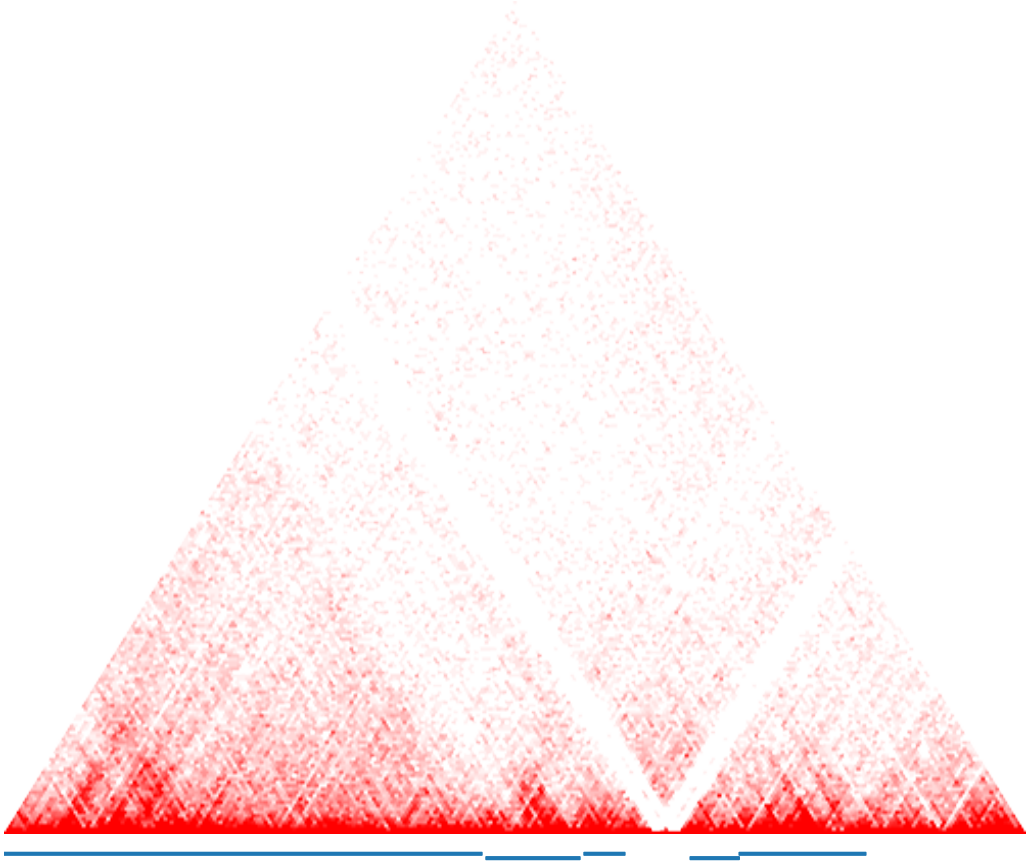
The `tads_imr90` object is a *GRanges* object with a all TADs detected from this Hi-C data.

```
> show(tads_imr90)
```

```
GRanges object with 2338 ranges and 0 metadata columns:
            seqnames              ranges strand
               <Rle>           <IRanges>  <Rle>
     TAD-1     chr1      770138-1290137      *
     TAD-2     chr1     1290138-1850140      *
     TAD-3     chr1     1850141-2330140      *
     TAD-4     chr1     2330141-3610140      *
     TAD-5     chr1     3770141-6077413      *
       ...      ...                 ...    ...
  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)
```

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



Package versions

This vignette was generated using the following package versions:

- R Under development (unstable) (2024-10-21 r87258), `x86_64-pc-linux-gnu`
- Running under: `Ubuntu 24.04.1 LTS`
- Matrix products: default
- BLAS: `/home/biocbuild/bbs-3.21-bioc/R/lib/libRblas.so`
- LAPACK: `/usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.12.0`
- Base packages: base, datasets, grDevices, graphics, methods, stats, stats4, utils
- Other packages: BiocGenerics 0.51.3, GenomeInfoDb 1.41.2, GenomicRanges 1.57.2, HiCDataHumanIMR90 1.25.0, HiTC 1.49.0, IRanges 2.39.2, S4Vectors 0.43.2
- Loaded via a namespace (and not attached): Biobase 2.65.1, BiocIO 1.15.2, BiocManager 1.30.25, BiocParallel 1.39.0, BiocStyle 2.33.1, Biostrings 2.73.2, DelayedArray 0.31.14, GenomeInfoDbData 1.2.13, GenomicAlignments 1.41.0, Matrix 1.7-1, MatrixGenerics 1.17.1, R6 2.5.1, RColorBrewer 1.1-3, RCurl 1.98-1.16, Rsamtools 2.21.2, S4Arrays 1.5.11, SparseArray 1.5.45, SummarizedExperiment 1.35.5, UCSC.utils 1.1.0, XML 3.99-0.17, XVector 0.45.0,

abind 1.4-8, bitops 1.0-9, cli 3.6.3, codetools 0.2-20, compiler 4.5.0, crayon 1.5.3, curl 5.2.3, digest 0.6.37, evaluate 1.0.1, fastmap 1.2.0, grid 4.5.0, htmltools 0.5.8.1, httr 1.4.7, jsonlite 1.8.9, knitr 1.48, lattice 0.22-6, matrixStats 1.4.1, parallel 4.5.0, restfulr 0.0.15, rjson 0.2.23, rlang 1.1.4, rmarkdown 2.28, rtracklayer 1.65.0, tools 4.5.0, xfun 0.48, yaml 2.3.10, zlibbioc 1.51.2

References

J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, Apr 2012. doi: 10.1038/nature11082. URL http://dx.doi.org/10.1038/nature11082.

E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, Oct 2009. doi: 10.1126/science.1181369. URL http://dx.doi.org/10.1126/science.1181369.

E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker, and E. Heard. Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, Apr 2012. doi: 10.1038/nature11049. URL http://dx.doi.org/10.1038/nature11049.

N. Servant, B. R. Lajoie, E. P. Nora, L. Giorgetti, C. Chen, E. Heard, J. Dekker, and E. Barillot. Hitc : Exploration of high-throughput 'c' experiments. *Bioinformatics*, Aug 2012. doi: 10.1093/bioinformatics/bts521. URL http://dx.doi.org/10.1093/bioinformatics/bts521.