TBX20 RNA-Seq data subset

Daniel Bindreither

April 21, 2026

1 Introduction

The TBX20 data set [4] provides ChIP-Seq and RNA-Seq data. In here only the RNA-Seq part of the data is utilized. The raw data where downloaded from Gene Expression Omnibus (GEO) [1], accession number GSM767225-GSM767230. TBX20 (T-box 20) in general is a transcriptional regulator essential for cardiac development and maintenance of mouse heart tissue. In this study TXB20 was knocked-out by using a Tamoxifen mediated conditional knock-out system. Transcriptional changes caused by the ablation of the second exon of TBX20 result in rapid onset of heart failures and the subsequent death of the mice. TBX20 knock-out adult heart tissue was compared to wild type adult heart tissue. This package provides a subset of the RNA-Seq data (chromosome 19) for demonstrating the capabilities of the *SpliceGraph* package. The vignette describes how to access the phenotypic data and the raw reads aligned with Bowtie [3] to the mm9 assembly of *Mus musculus* from UCSC Genome Browser [2].

Accessing the experimental design ...

```
> library("TBX20BamSubset")
> fn <- system.file("extdata", "phenoData.txt",
+                   package="TBX20BamSubset")
> pd <- read.table(fn, header=TRUE,
+                  stringsAsFactors=FALSE)
```

Accessing the raw reads ...

```
> library("Rsamtools")
> fls <- getBamFileList()
> bfs <- BamFileList(fls)
```

	SRX	SRR	GSM	condition	replicate
1	SRX085099	SRR316184	GSM767225	normal	1
2	SRX085100	SRR316185	GSM767226	normal	2
3	SRX085101	SRR316186	GSM767227	normal	3
4	SRX085102	SRR316187	GSM767228	Tbx20 knockout	1
5	SRX085103	SRR316188	GSM767229	Tbx20 knockout	2
6	SRX085104	SRR316189	GSM767230	Tbx20 knockout	3

Table 1: Design of the TBX20 experiment

References

[1] Tanya Barrett, Dennis B. Troup, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Rolf N. Muertter, Michelle Holko, Oluwabukunmi Ayanbule, Andrey Yefanov, and Alexandra Soboleva. Ncbi geo: archive for functional genomics data sets - 10 years on. *Nucleic Acids Research*, 39(suppl 1):D1005–D1010, 2011.

[2] Pauline A. Fujita, Brooke Rhead, Ann S. Zweig, Angie S. Hinrichs, Donna Karolchik, Melissa S. Cline, Mary Goldman, Galt P. Barber, Hiram Clawson, Antonio Coelho, Mark Diekhans, Timothy R. Dreszer, Belinda M. Giardine, Rachel A. Harte, Jennifer Hillman-Jackson, Fan Hsu, Vanessa Kirkup, Robert M. Kuhn, Katrina Learned, Chin H. Li, Laurence R. Meyer, Andy Pohl, Brian J. Raney, Kate R. Rosenbloom, Kayla E. Smith, David Haussler, and W. James Kent. The ucsc genome browser database: update 2011. *Nucleic Acids Research*, 2010.

[3] Ben Langmead, Cole Trapnell, Mihai Pop, and Steven Salzberg. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biology*, 10(3):R25, 2009.

[4] Noboru J. Sakabe, Ivy Aneas, Tao Shen, Leila Shokri, Soo-Young Park, Martha L. Bulyk, Sylvia M. Evans, and Marcelo A. Nobrega. Dual transcriptional activator and repressor roles of tbx20 regulate adult cardiac structure and function. *Human Molecular Genetics*, 2012.

2 Session Information

```
R version 4.6.0 RC (2026-04-17 r89917)
Platform: x86_64-pc-linux-gnu
```

```
Running under: Ubuntu 24.04.4 LTS

Matrix products: default
BLAS:   /home/biocbuild/bbs-3.24-bioc/R/lib/libRblas.so
LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.12.0  LAPACK version 3.12.0

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_GB              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
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

time zone: America/New_York
tzcode source: system (glibc)

attached base packages:
[1] stats4    stats     graphics  grDevices utils     datasets  methods
[8] base

other attached packages:
 [1] xtable_1.8-8          TBX20BamSubset_1.47.0 Rsamtools_2.27.2
 [4] Biostrings_2.79.5     XVector_0.51.0        GenomicRanges_1.63.2
 [7] IRanges_2.45.0        S4Vectors_0.49.2      Seqinfo_1.1.0
[10] BiocGenerics_0.57.1   generics_0.1.4

loaded via a namespace (and not attached):
[1] codetools_0.2-20    parallel_4.6.0      bitops_1.0-9
[4] compiler_4.6.0      tools_4.6.0         crayon_1.5.3
[7] BiocParallel_1.45.0
```