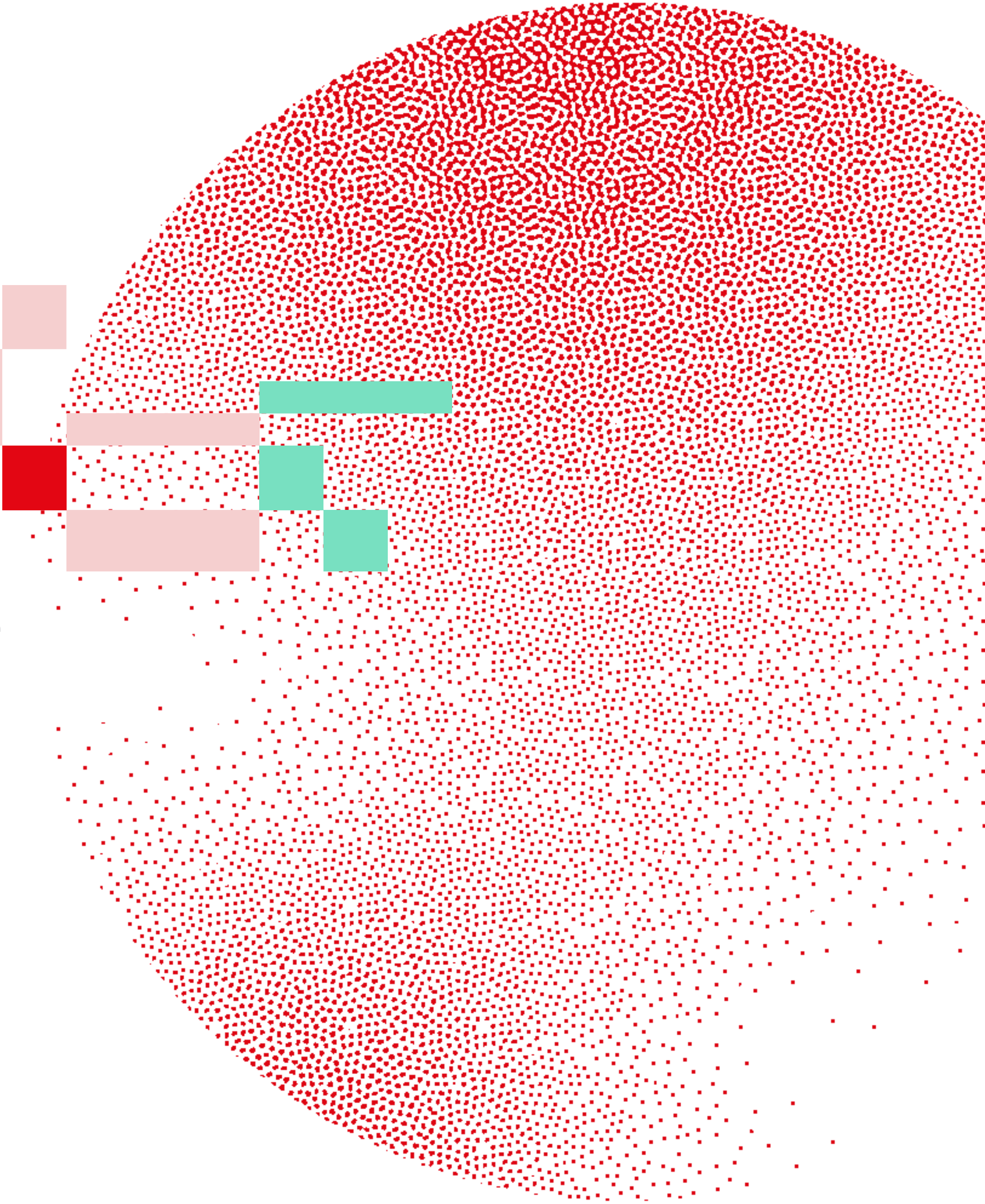
Swiss Institute of Bioinformatics



BIOLOGY-INFORMED MULTIOMICS DATA INTEGRATION AND VISUALIZATION

Dataset and data structure

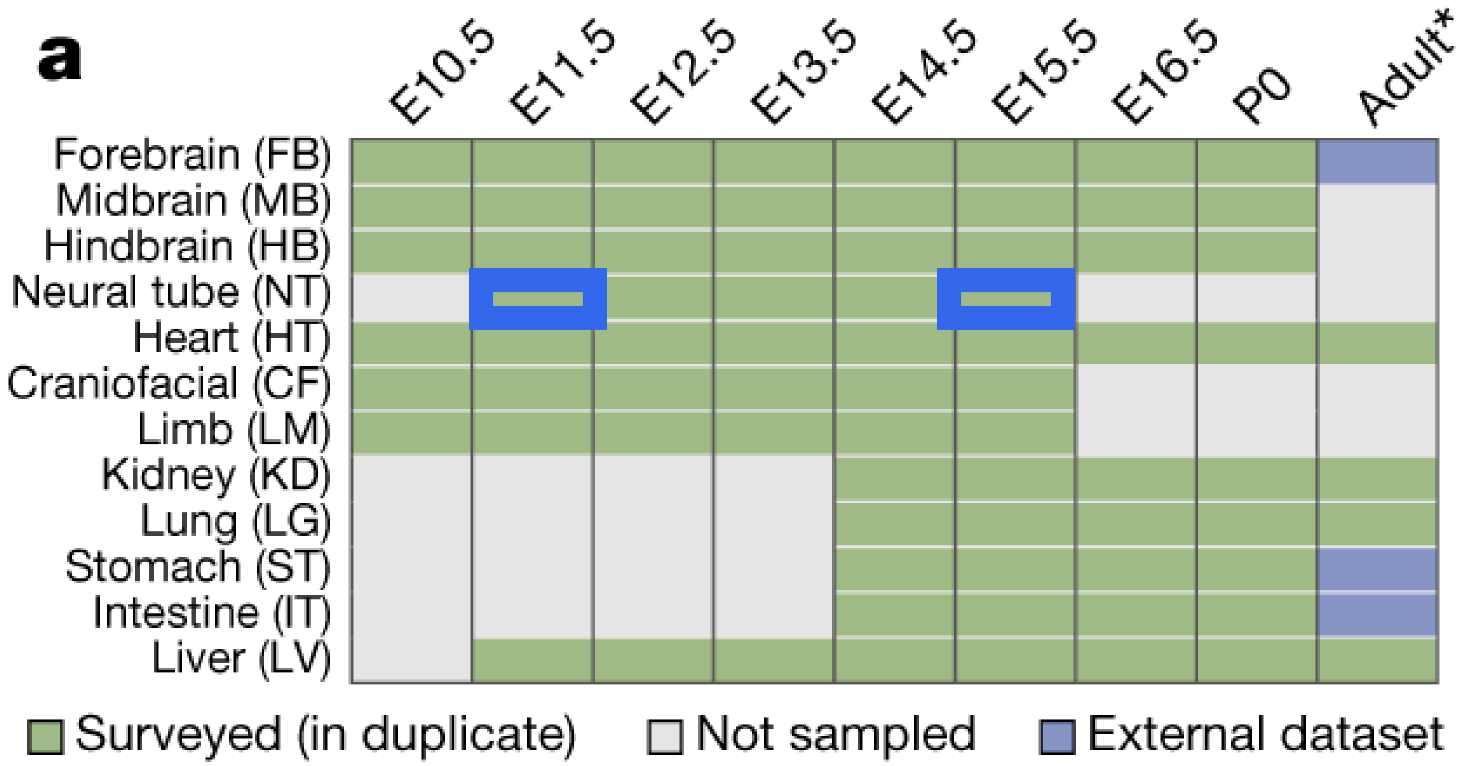
Deepak Tanwar
June 16-17, 2025

SWITZERLAND

Learning outcomes

- Know about the datasets used in this course
- Introduction to data structures:
    - SummarizedExperiment
    - GRanges

Dataset for the course



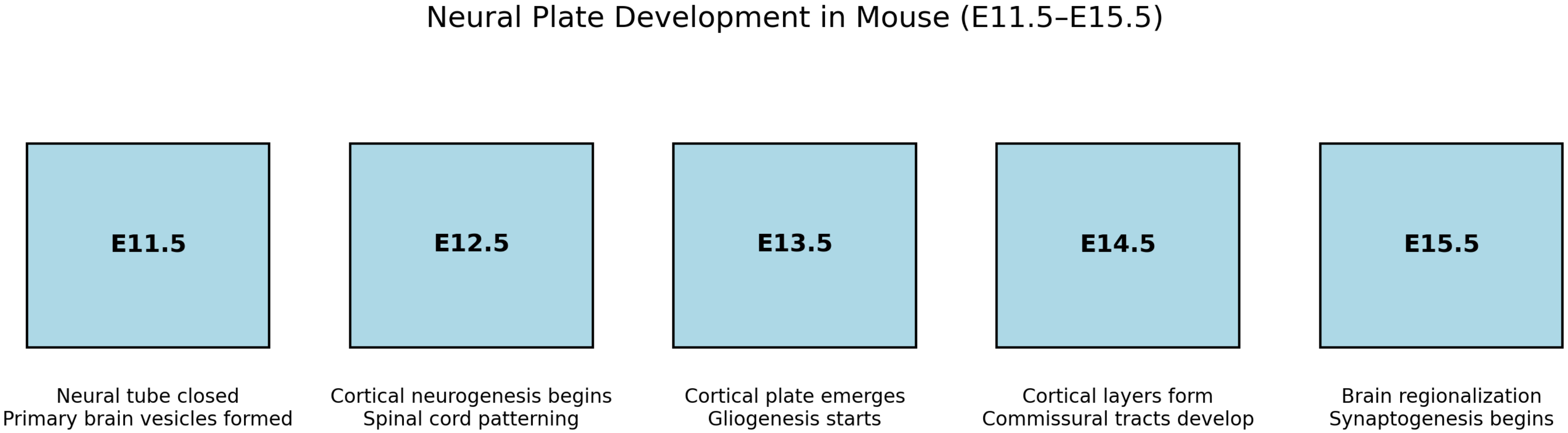
	E10.5	E11.5	E12.5	E13.5	E14.5	E15.5	E16.5	P0	Adult*
Forebrain (FB)	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟦
Midbrain (MB)	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	⬜
Hindbrain (HB)	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	⬜
Neural tube (NT)	⬜	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜
Heart (HT)	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Craniofacial (CF)	🟩	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜
Limb (LM)	🟩	🟩	🟩	🟩	🟩	🟩	⬜	⬜	⬜
Kidney (KD)	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟩	🟩
Lung (LG)	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟩	🟩
Stomach (ST)	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟩	🟦
Intestine (IT)	⬜	⬜	⬜	⬜	🟩	🟩	🟩	🟩	🟦
Liver (LV)	⬜	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩

🟩 Surveyed (in duplicate) ⬜ Not sampled 🟦 External dataset

https://www.nature.com/articles/s41586-020-2119-x

Neural tube development in mouse

In mouse embryonic development, the neural tube is a crucial structure that forms the central nervous system, encompassing the brain and spinal cord.

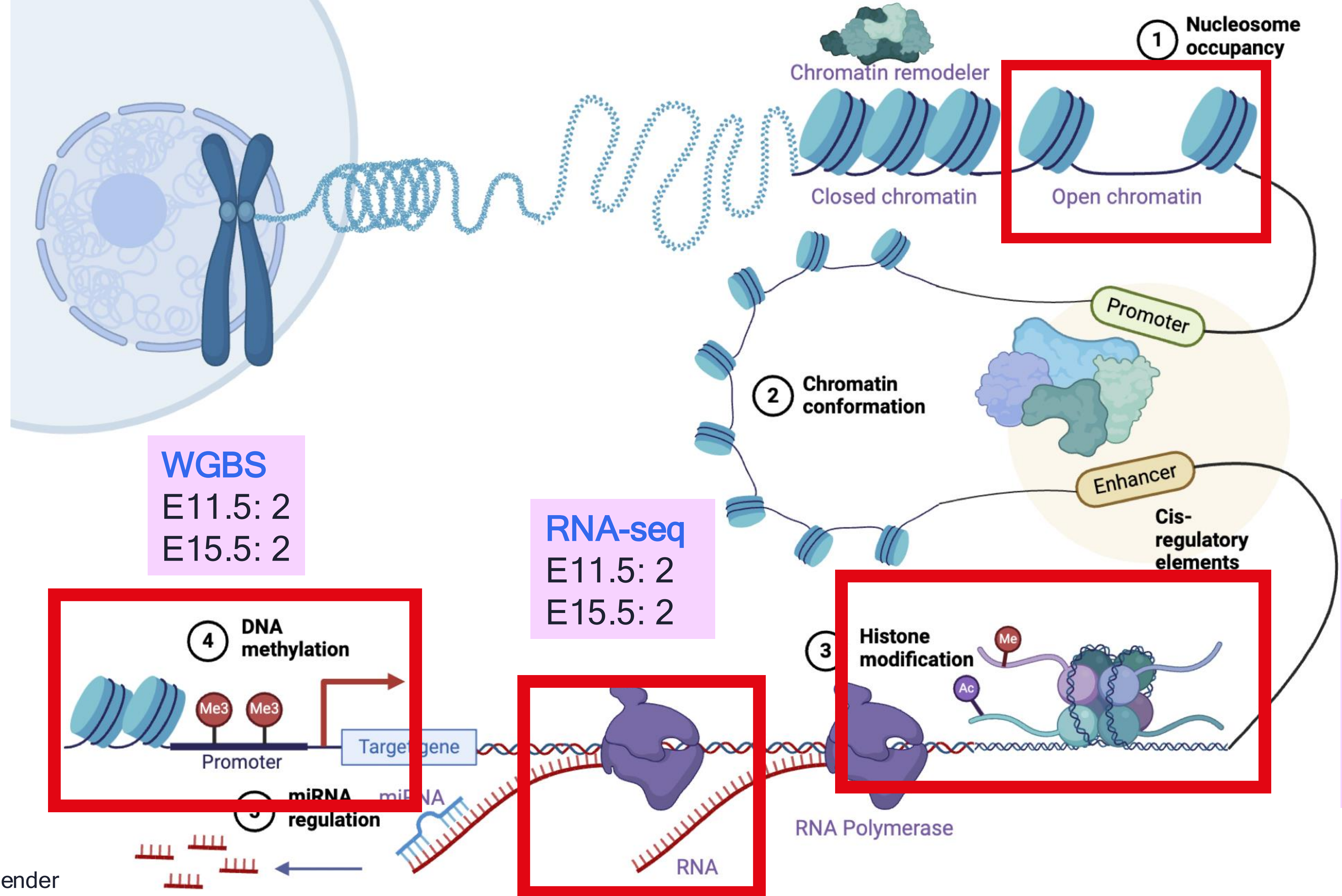


Neural Plate Development in Mouse (E11.5–E15.5)

E11.5	**E12.5**	**E13.5**	**E14.5**	**E15.5**
Neural tube closed Primary brain vesicles formed	Cortical neurogenesis begins Spinal cord patterning	Cortical plate emerges Gliogenesis starts	Cortical layers form Commissural tracts develop	Brain regionalization Synaptogenesis begins

https://www.nature.com/articles/s41586-020-2119-x

Multiomics datasets from mouse neural tube development



ATAC-seq
E11.5: 2
E15.5: 2

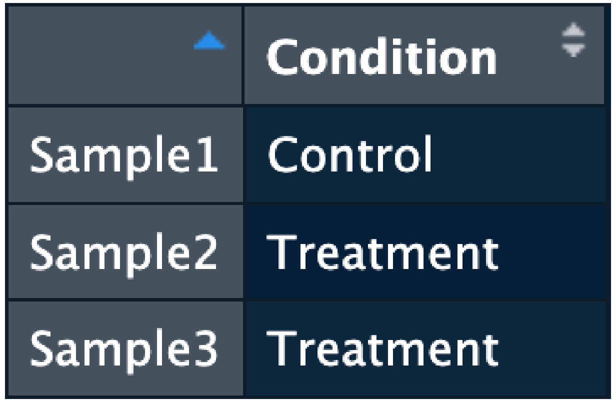
WGBS
E11.5: 2
E15.5: 2

RNA-seq
E11.5: 2
E15.5: 2

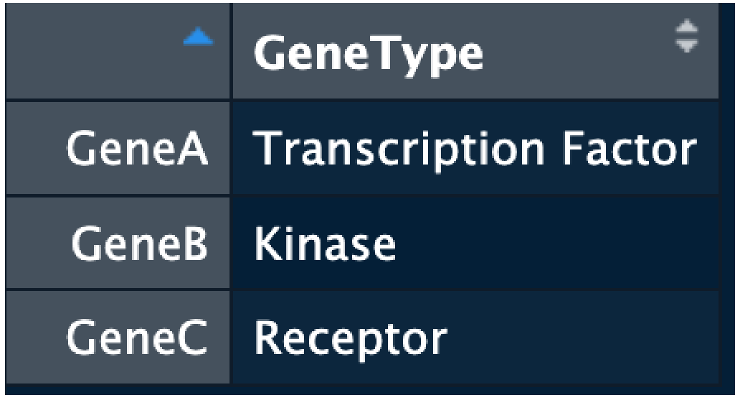
ChIP-seq
H3K4me3
H3K4me1
H3K27me3
H3K27ac
E11.5: 2
E15.5: 2

Generated using Biorender

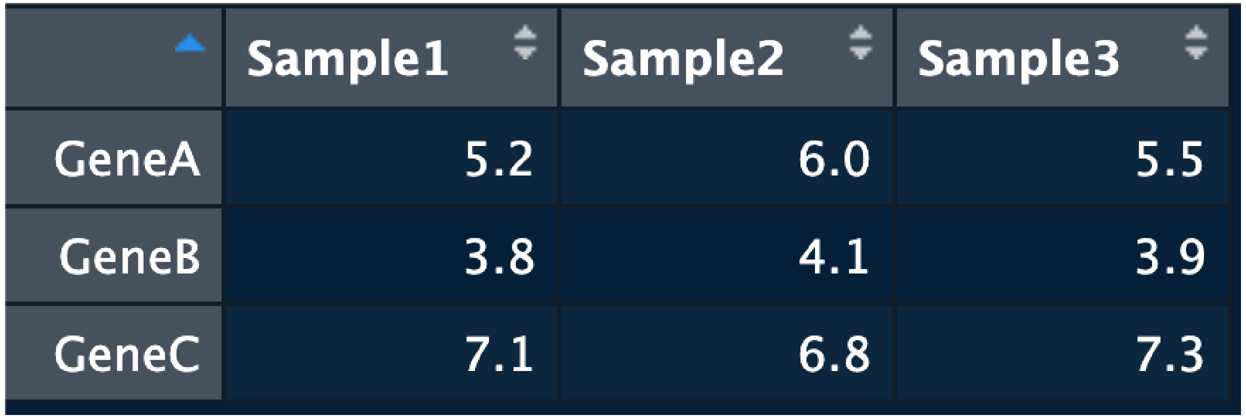
Gene expression data in a data.frame



	Condition
Sample1	Control
Sample2	Treatment
Sample3	Treatment

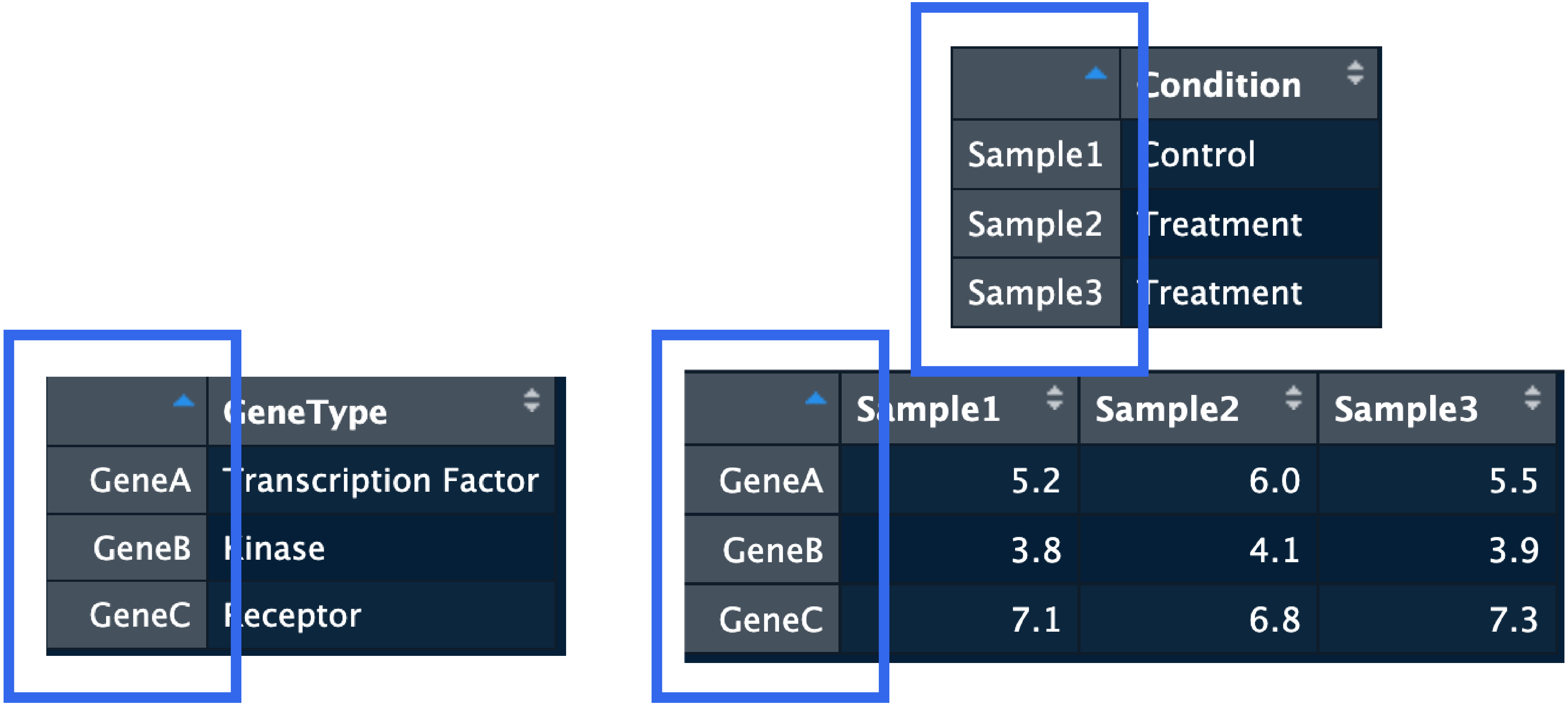


	GeneType
GeneA	Transcription Factor
GeneB	Kinase
GeneC	Receptor



	Sample1	Sample2	Sample3
GeneA	5.2	6.0	5.5
GeneB	3.8	4.1	3.9
GeneC	7.1	6.8	7.3

data.frame: rownames

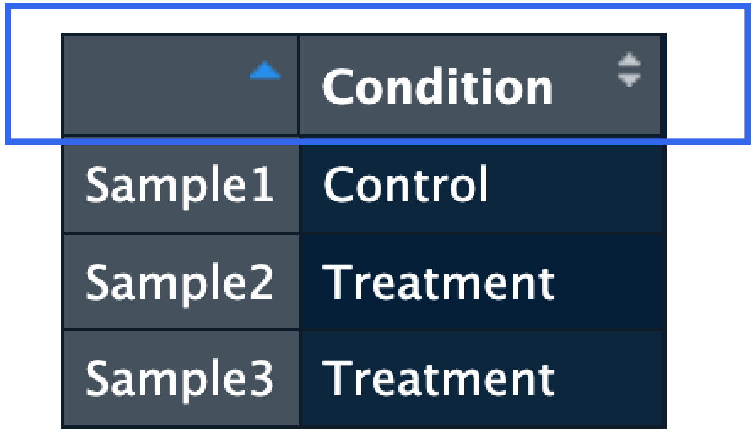


	Condition
Sample1	Control
Sample2	Treatment
Sample3	Treatment

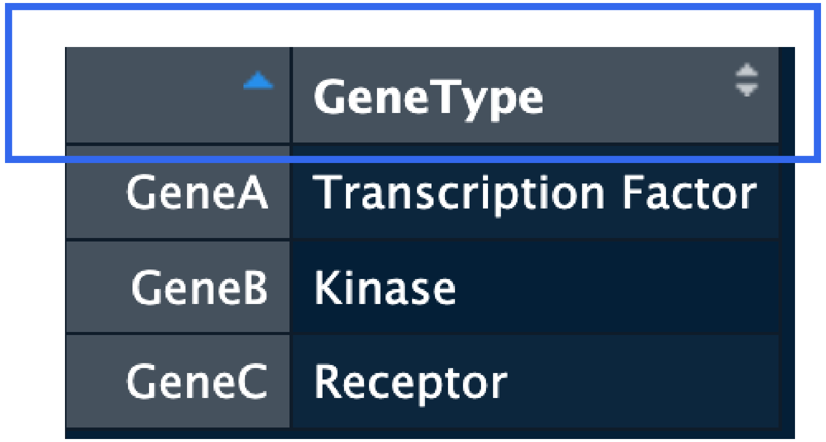
	GeneType
GeneA	Transcription Factor
GeneB	Kinase
GeneC	Receptor

	Sample1	Sample2	Sample3
GeneA	5.2	6.0	5.5
GeneB	3.8	4.1	3.9
GeneC	7.1	6.8	7.3

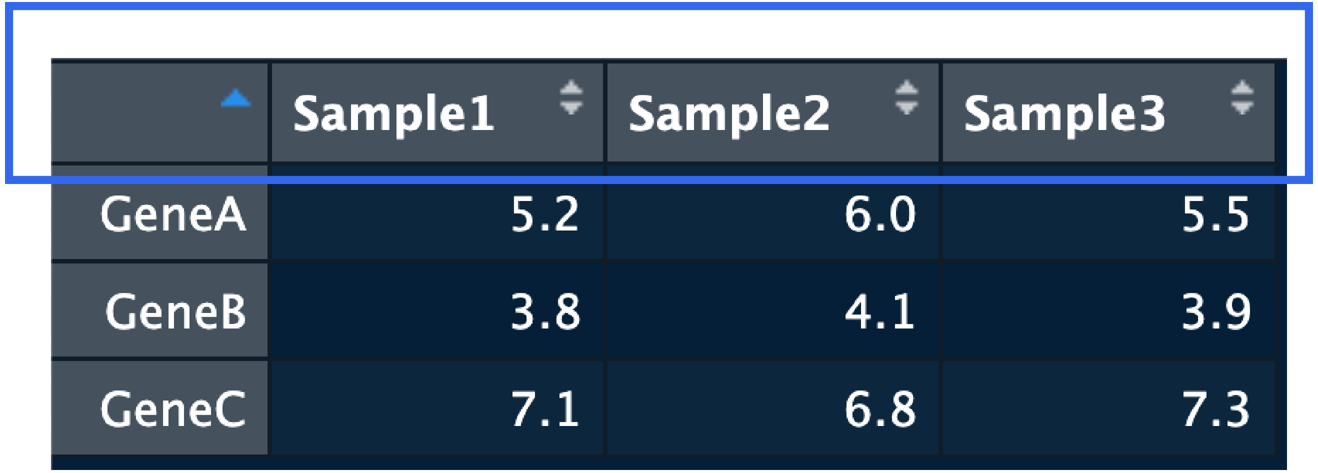
data.frame: colnames



	Condition
Sample1	Control
Sample2	Treatment
Sample3	Treatment

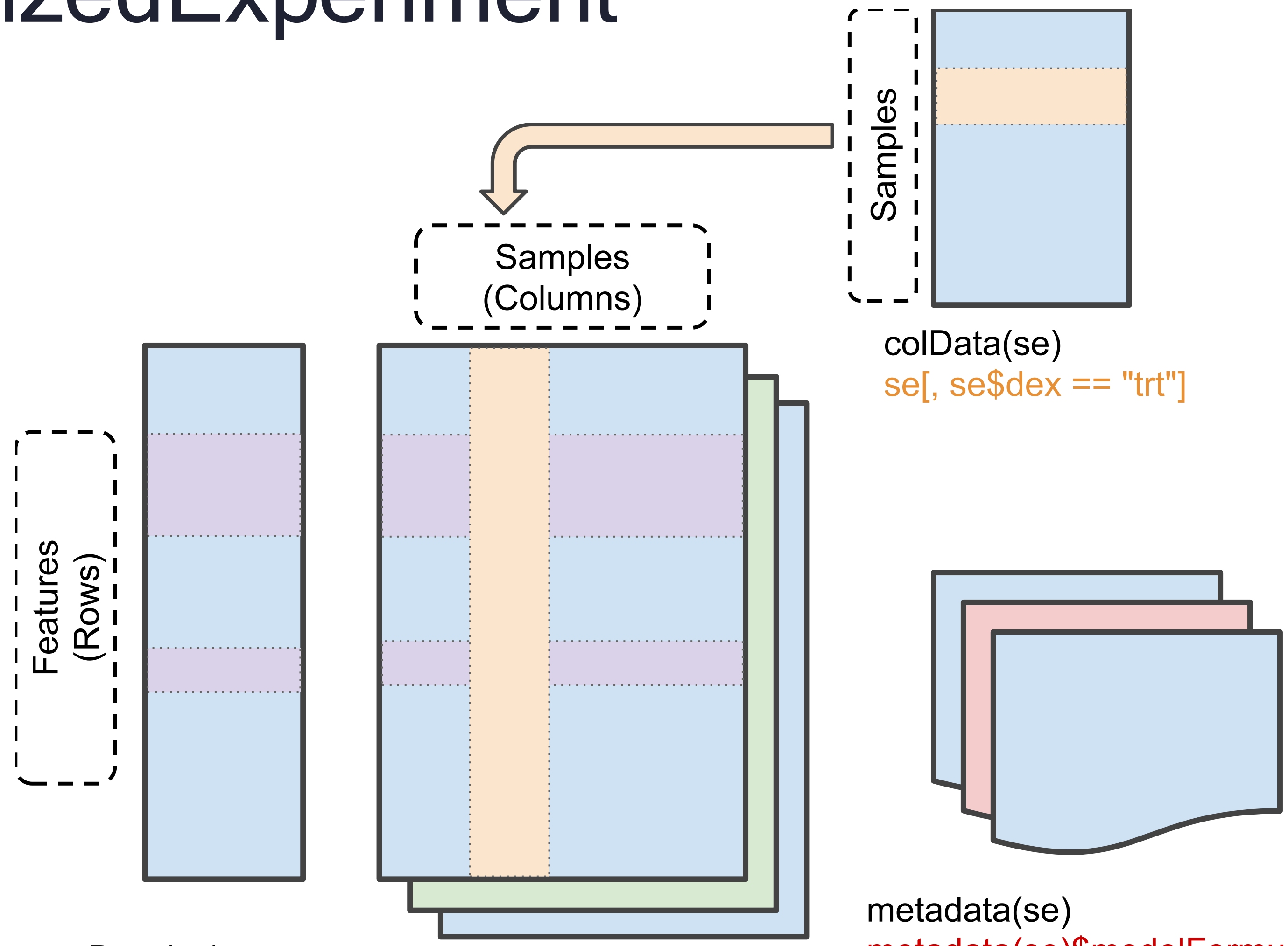


	GeneType
GeneA	Transcription Factor
GeneB	Kinase
GeneC	Receptor



	Sample1	Sample2	Sample3
GeneA	5.2	6.0	5.5
GeneB	3.8	4.1	3.9
GeneC	7.1	6.8	7.3

SummarizedExperiment



Samples

Samples
(Columns)

Features
(Rows)

colData(se)
se[, se$dex == "trt"]

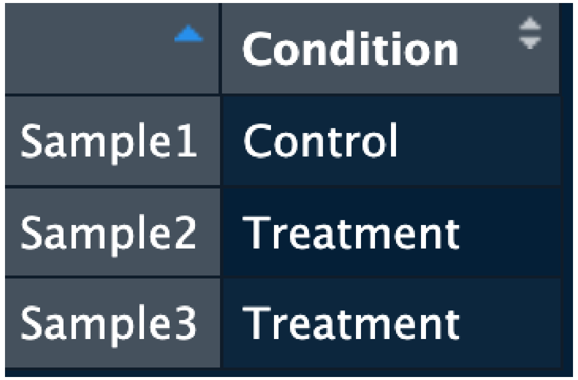
rowData(se)
subsetByOverlaps(se, roi)

assays(se)
assay(se, n = 2)
assay(subsetByOverlaps(se, roi))
assay(se[, se$dex == "trt"])

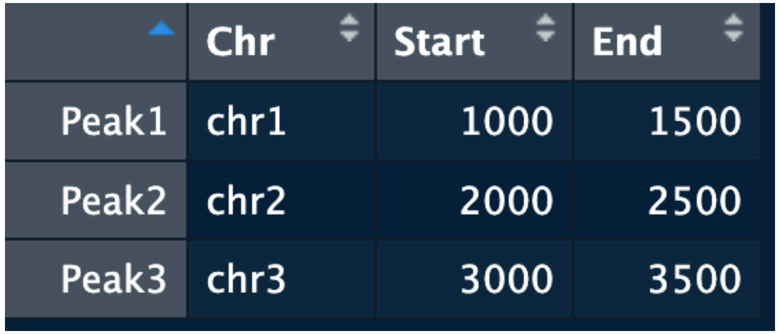
metadata(se)
metadata(se)$modelFormula

https://bioconductor.org/packages/devel/bioc/vignettes/SummarizedExperiment/inst/doc/SummarizedExperiment.html

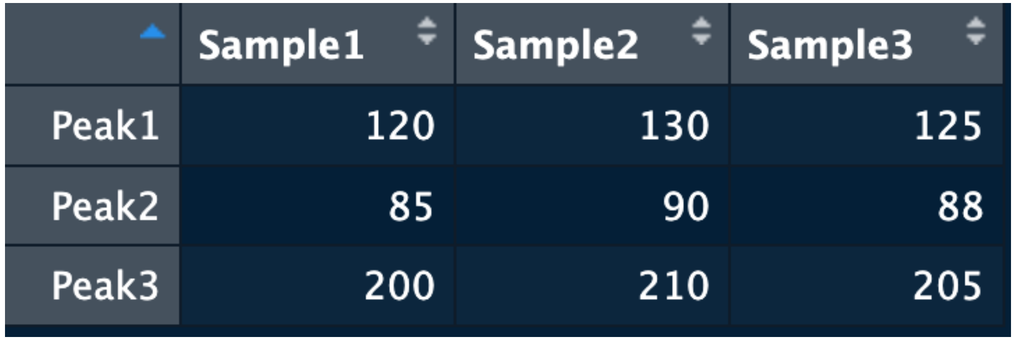
Data format for ATAC/ ChIP sequencing data



	Condition
Sample1	Control
Sample2	Treatment
Sample3	Treatment

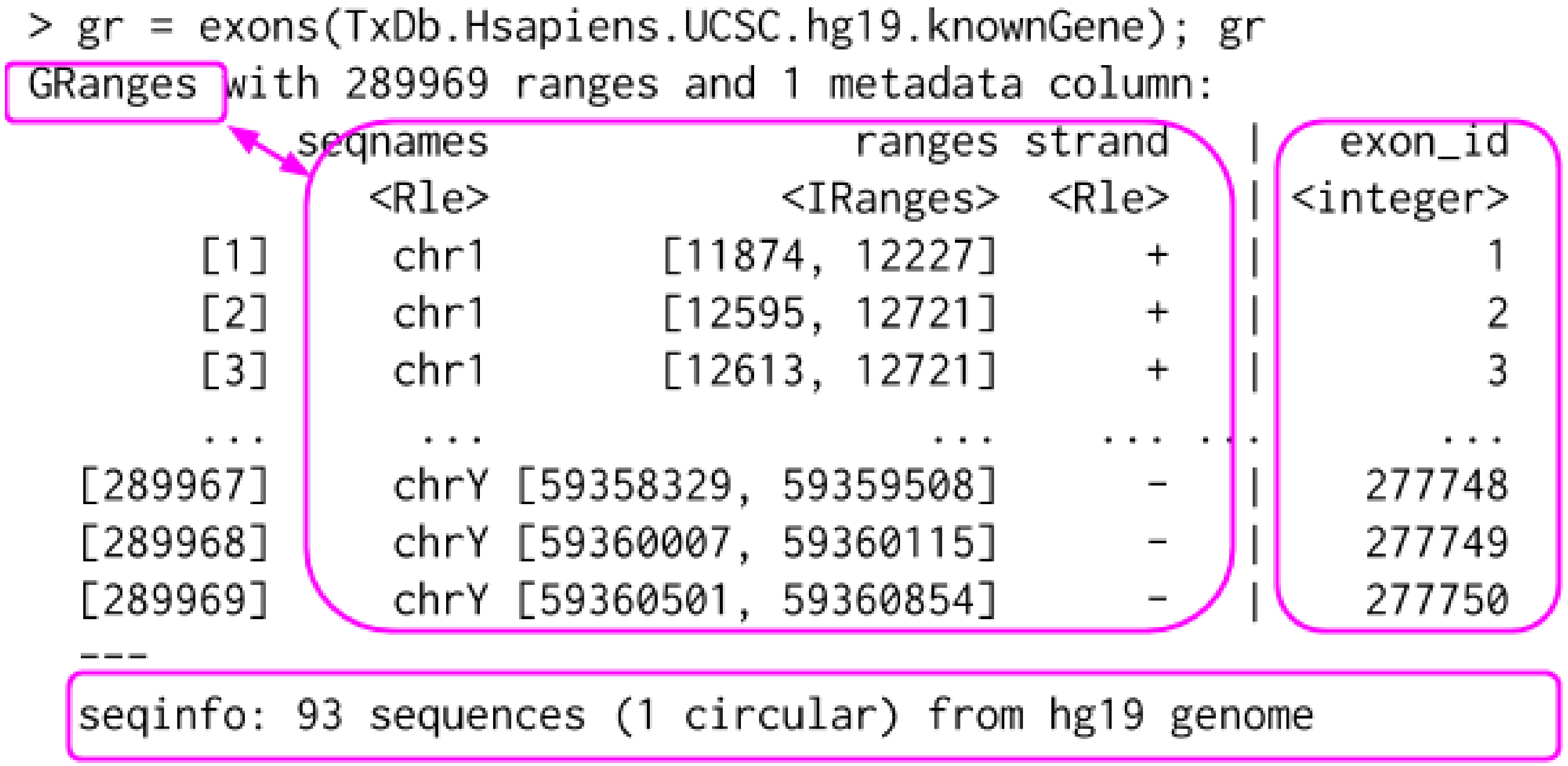


	Chr	Start	End
Peak1	chr1	1000	1500
Peak2	chr2	2000	2500
Peak3	chr3	3000	3500



	Sample1	Sample2	Sample3
Peak1	120	130	125
Peak2	85	90	88
Peak3	200	210	205

Genomic Ranges snapshot



```
> gr = exons(TxDb.Hsapiens.UCSC.hg19.knownGene); gr
GRanges with 289969 ranges and 1 metadata column:
           seqnames                 ranges strand   |   exon_id
              <Rle>              <IRanges>  <Rle>   | <integer>
       [1]     chr1       [11874, 12227]      +   |         1
       [2]     chr1       [12595, 12721]      +   |         2
       [3]     chr1       [12613, 12721]      +   |         3
       ...      ...                  ...    ... ...       ...
  [289967]     chrY [59358329, 59359508]      -   |    277748
  [289968]     chrY [59360007, 59360115]      -   |    277749
  [289969]     chrY [59360501, 59360854]      -   |    277750
  ---
  seqinfo: 93 sequences (1 circular) from hg19 genome
```

GRanges

```
length(gr); gr[1:5]
seqnames(gr)
start(gr)
end(gr)
width(gr)
strand(gr)
```

DataFrame

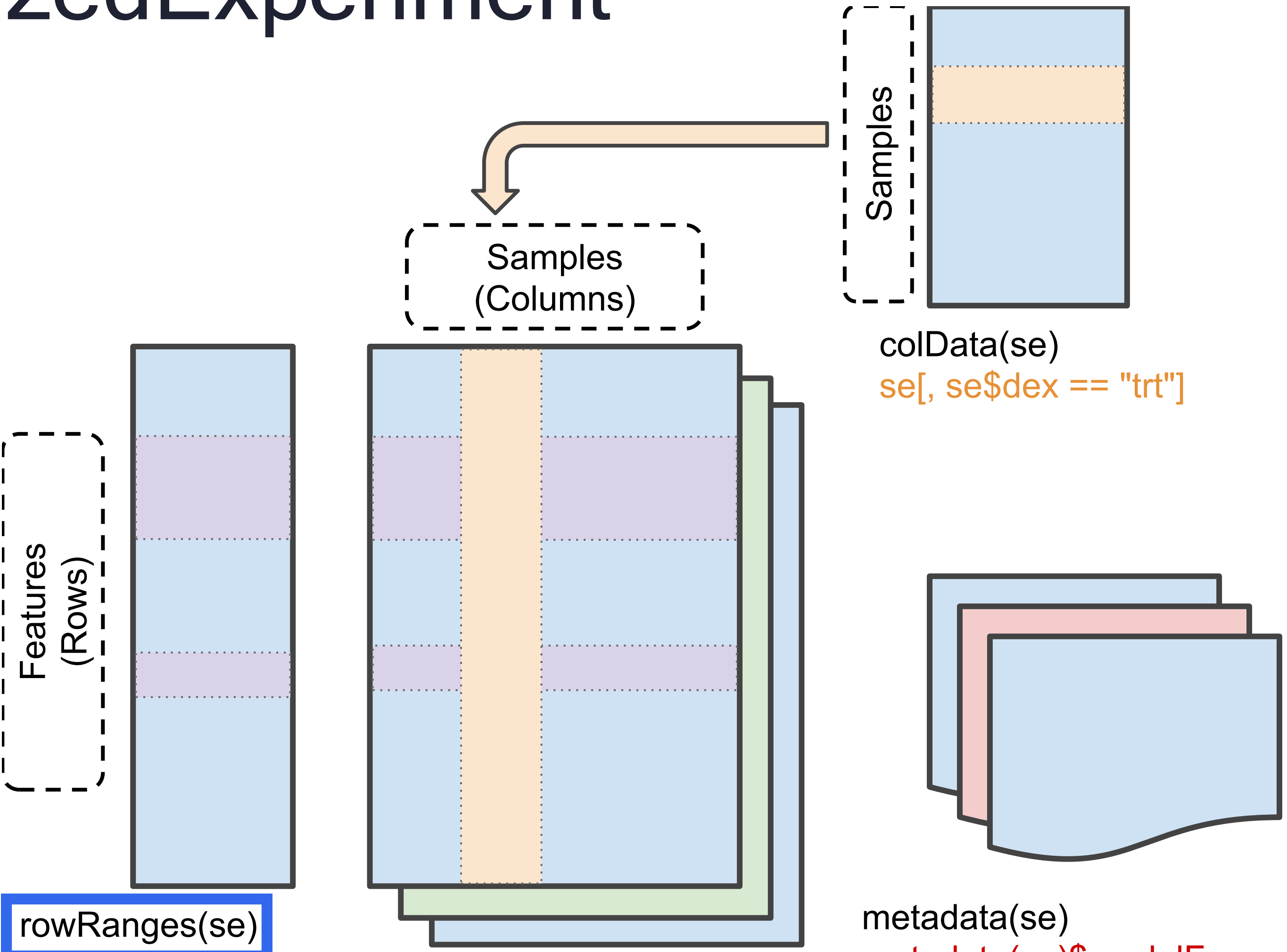
```
mcols(gr)
gr$exon_id
```

Seqinfo

```
seqlevels(gr)
seqlengths(gr)
genome(gr)
```

https://www.bioconductor.org/help/course-materials/2015/Uruguay2015/V3-GenomicRanges.html

SummarizedExperiment



colData(se)
se[, se$dex == "trt"]

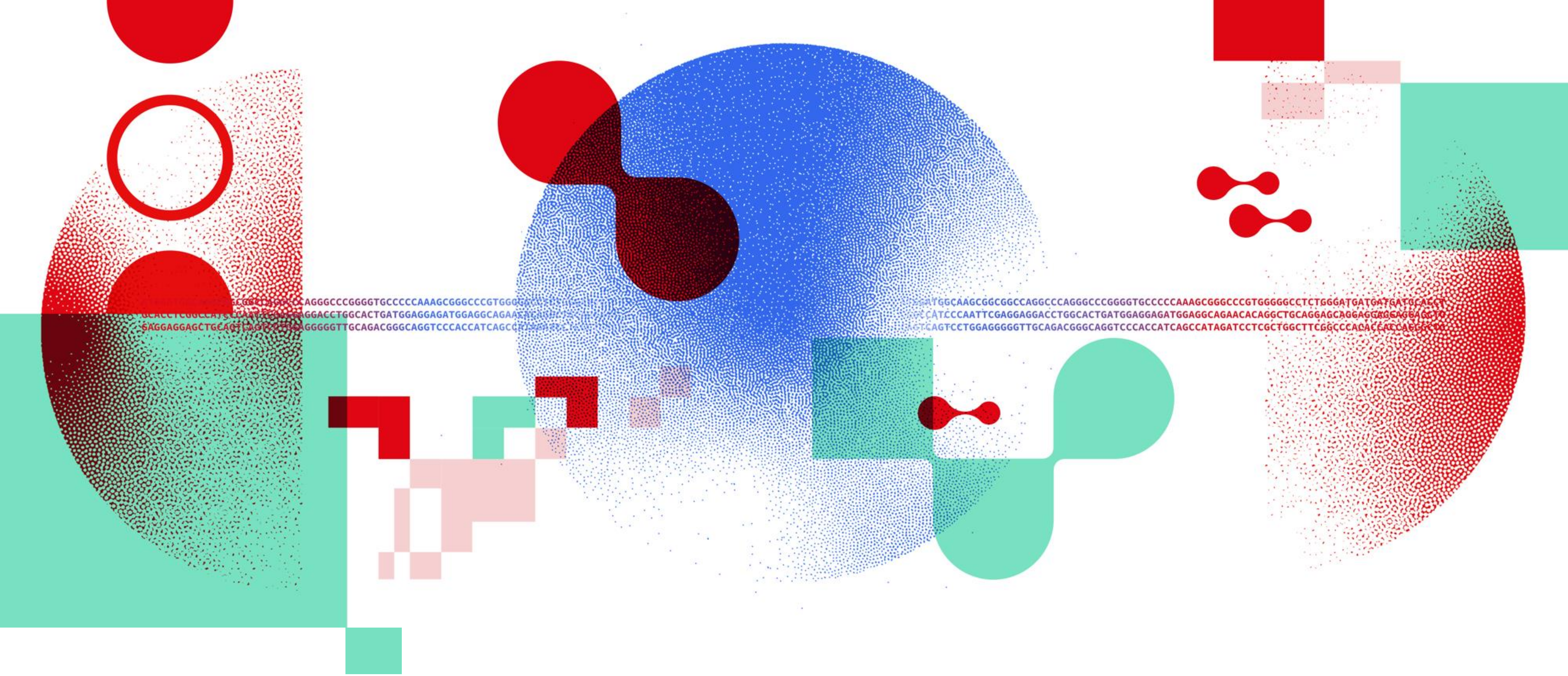
rowRanges(se)
rowData(se)
subsetByOverlaps(se, roi)

assays(se)
assay(se, n = 2)
assay(subsetByOverlaps(se, roi))
assay(se[, se$dex == "trt"])

metadata(se)
metadata(se)$modelFormula

https://bioconductor.org/packages/devel/bioc/vignettes/SummarizedExperiment/inst/doc/SummarizedExperiment.html

Exercise 1



Thank you

DATA SCIENTISTS FOR LIFE

sib.swiss