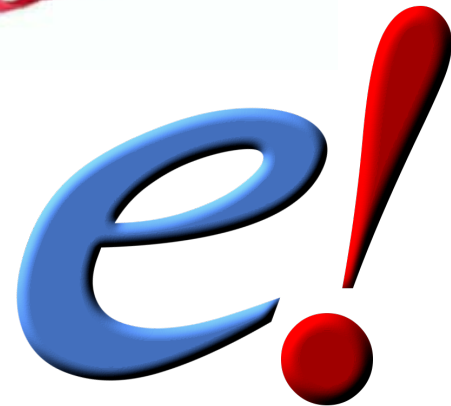


Ensembl Regulation



Ensembl Regulation

The goal of Ensembl Regulation team is to annotate the genome with features that may play a role in the transcriptional regulation of genes, ie **promoters**, **enhancers**, **insulators** etc, using:

- Predicted open/closed chromatin
 - DNase I sensitivity
- Transcription factor binding sites
- Epigenetic marks
 - Histone modifications
 - DNA methylation
- RNA Pol binding



We do not...

- ...link promoters/enhancers/insulators or any other regulatory features to genes. We allow you see what is where and make your own inferences.
- ...link regulatory features to gene expression. We have cell-line specific regulation data and tissue specific expression data – make of it what you will.

Regulatory data is incredibly complex and still in relative infancy. There is no comprehensive database of regulation data.

Data sources



A subset of cell types

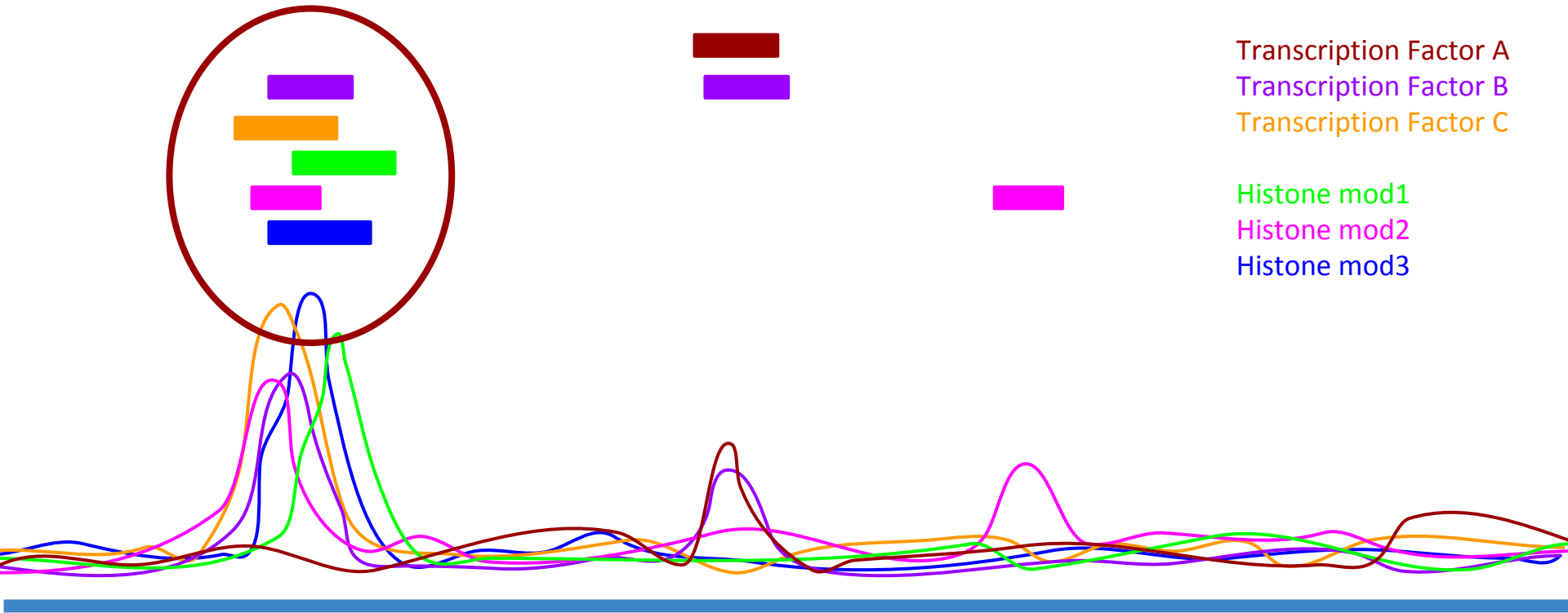
- Only a subset of available data is displayed in Ensembl.
- We display cell types that have, at a minimum:
 - CTCF binding
 - DNase or FAIRE data
 - H3K4me3, H3K27me3, H3K36me3 data
- We display all TFBS and histone modification data known in these cell types.
- We process these data to predict activity.
- Further data can be added using track hubs.

Processing the data

- The raw data is taken from the various sources.
- This is processed to predict the positions of regulatory features, such as **promoters**, **enhancers** and **insulators**.
- The activity of these features is predicted in the different cell types.

- All of this can be viewed in the genome browser.

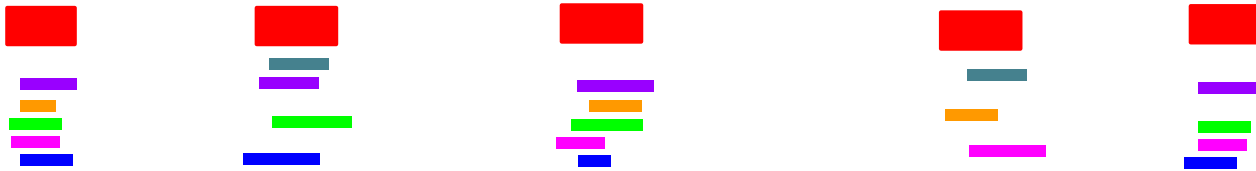
Raw data



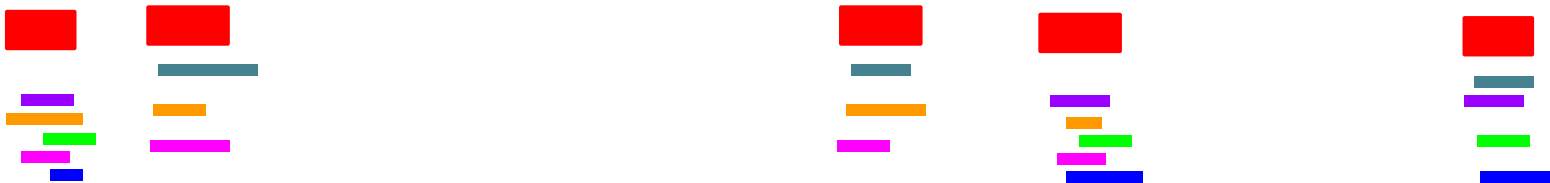
Searching for patterns



known promoter



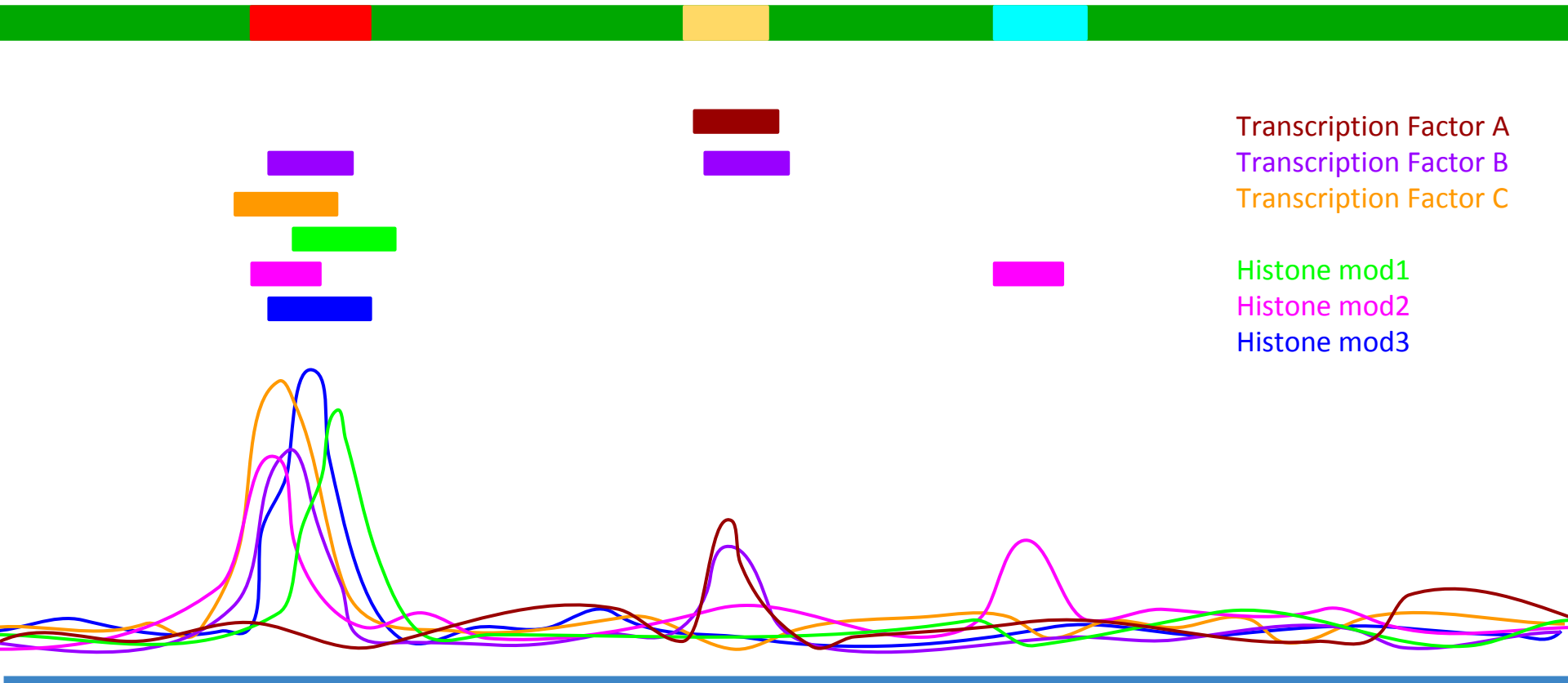
known promoter



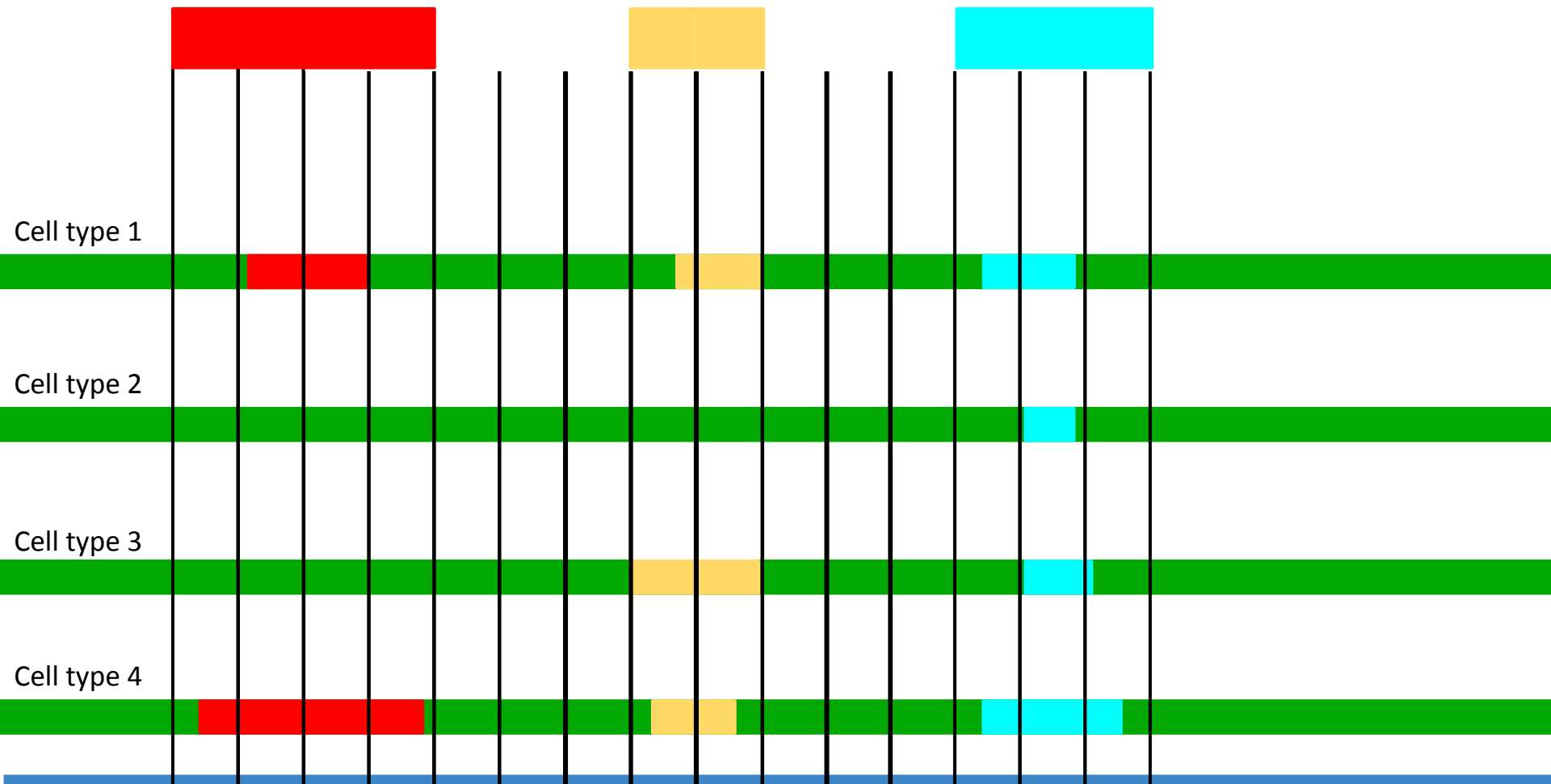
known promoter



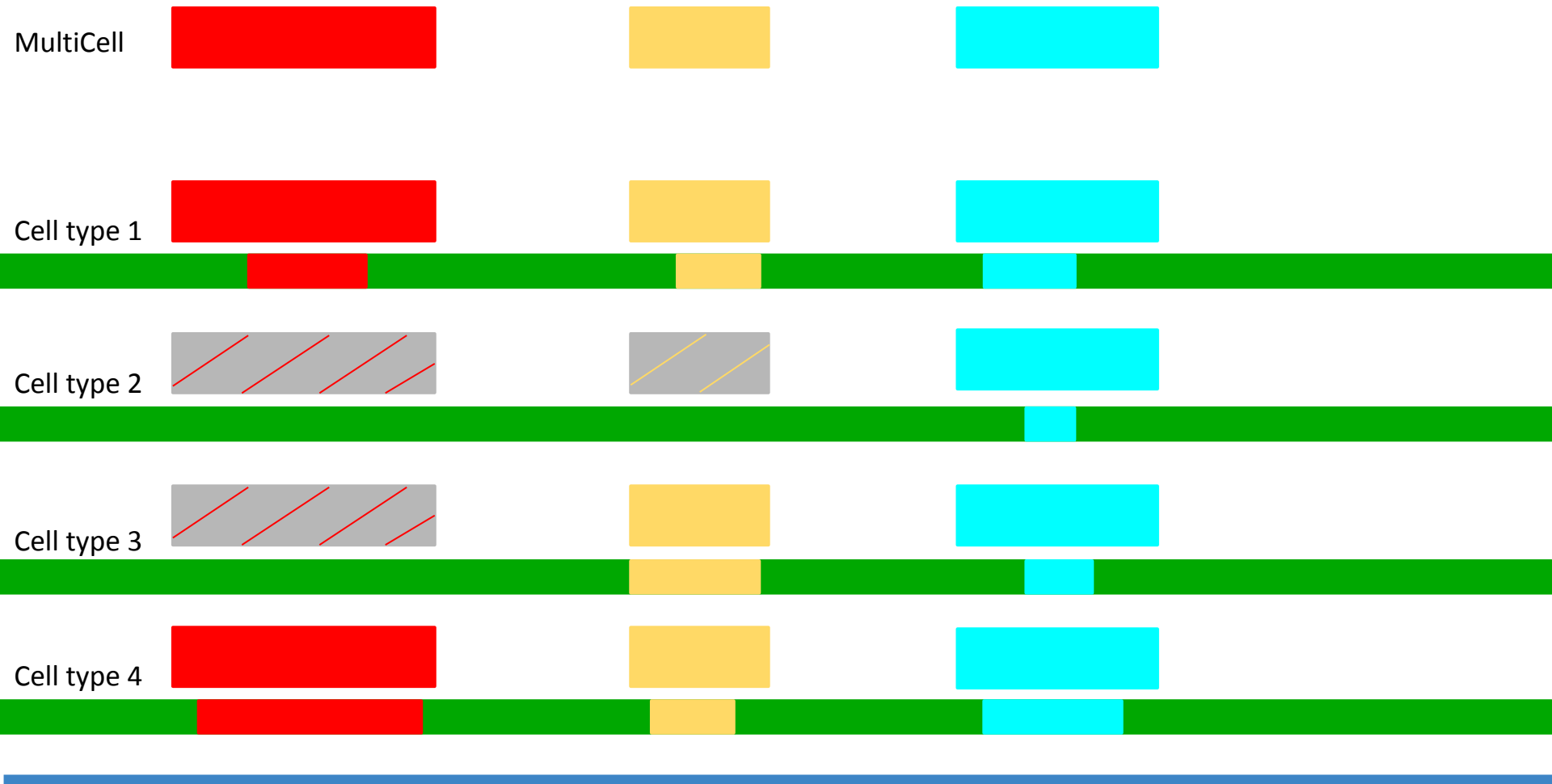
Segmentation



MultiCell features



Cell-specific features



Coverage

Label	Count	Mean length (bp)	Max length (bp)	Total length (Mbp)
TSS	40,249	973.2	11,400	39.2
Proximal Reg.	101,206	1005.5	15,000	101.8
Distal Reg.	209,081	526.1	8,400	110.0
CTCF	108,284	550.1	5,200	59.6
Unannotated TFBS	163,528	155.8	1,630	25.5
Union				299.2

Hands on

- We're going to look at the region of a gene *LIMD2* to find regulatory features and explore what cells types they are active in and what evidence there is to show this.
- We'll have a go at a quick BioMart query to get the loci and functions of some Regulatory Features:

ENSR00001601181

ENSR00001567543

ENSR00001601182

ENSR00000556855

ENSR00001601183

ENSR00000556857

ENSR00001601184

ENSR00000556858

ENSR00001601185

ENSR00000556859

ENSR00001567544

ENSR00000556863

ENSR00000556865

ENSR00000556867

ENSR00001567547