

Studying the transcriptome using RNA-seq

Cecilia Coimbra Klein



Outline

Outline

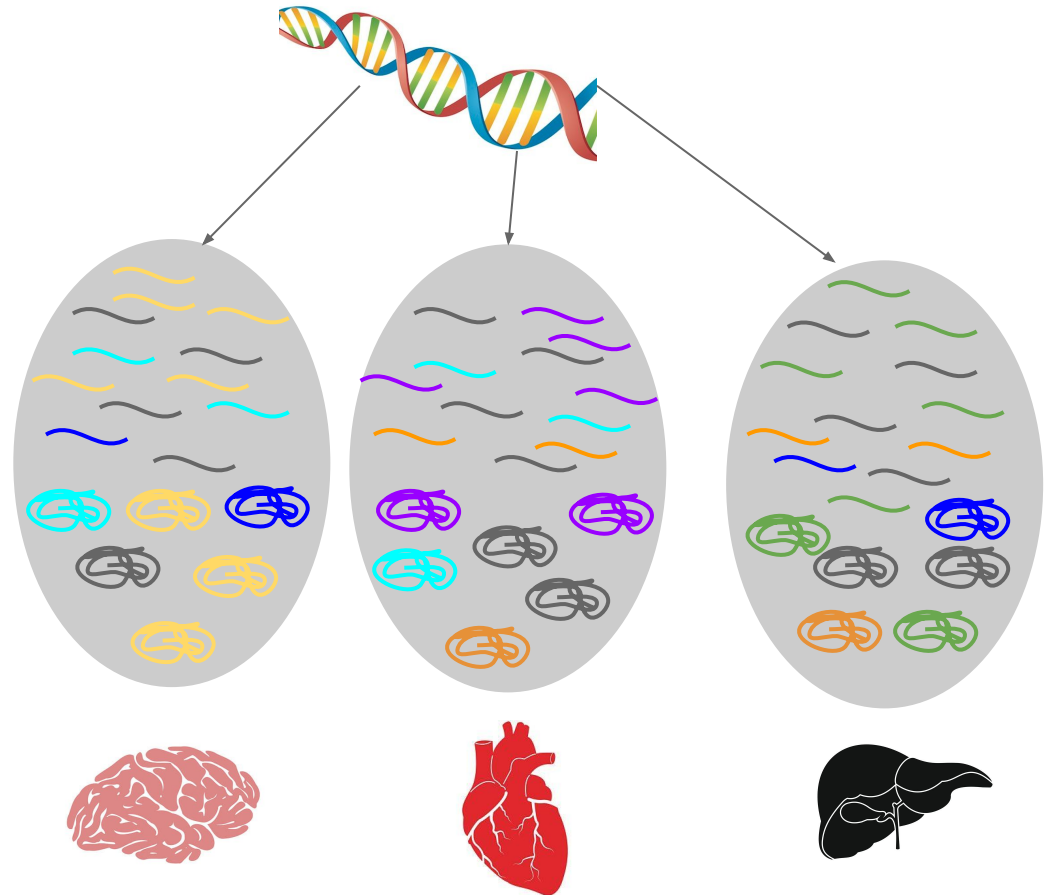
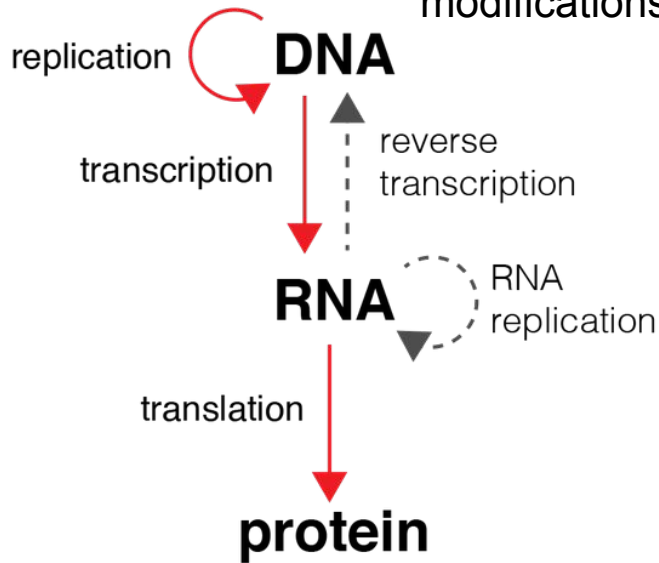
- **Summary of the course**
 - RNA-seq introduction and processing
 - RNA-seq analysis (clustering, differential gene expression, GO enrichment)
 - RNA-seq analysis (splicing)
- **Multiomics**
 - ChIP-seq (processing and data analysis)
 - ATAC-seq (visualization)
- **Hands-on Multiomics**
 - ChIP-seq and ATAC-seq signal in the UCSC genome browser
 - promoter regions of differentially expressed genes
 - promoter regions of differentially spliced genes
 - omics portals

RNA-seq

introduction and processing

Molecular biology dogma

epigenetic
modifications



- The **genome** is identical in all cell types, however not all cell types have the same function. That's why the **transcriptome** (and the **epigenome**) becomes also relevant.

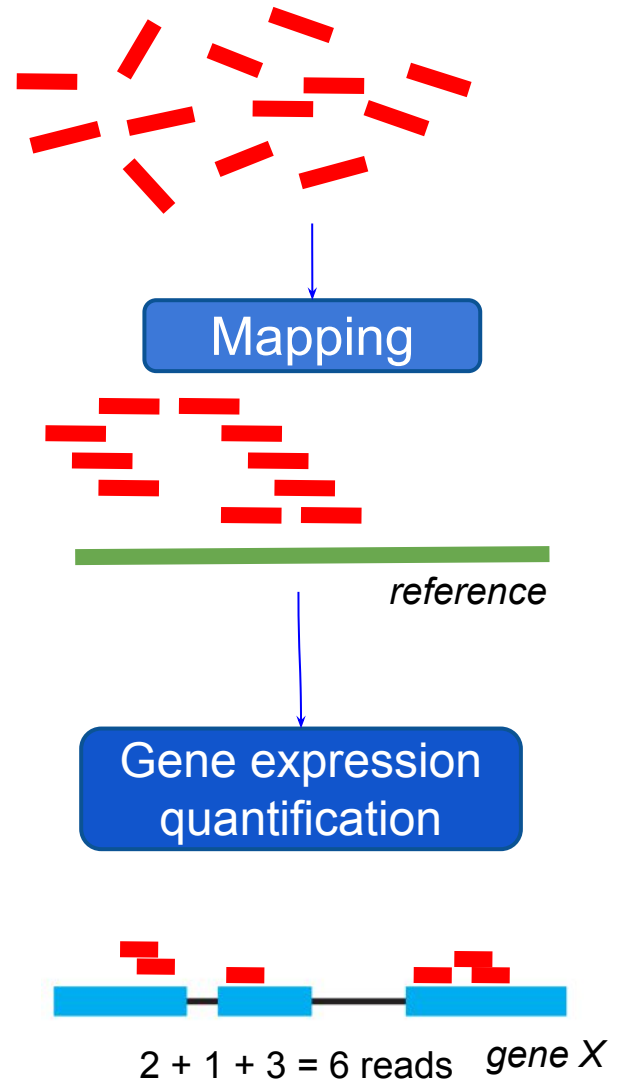
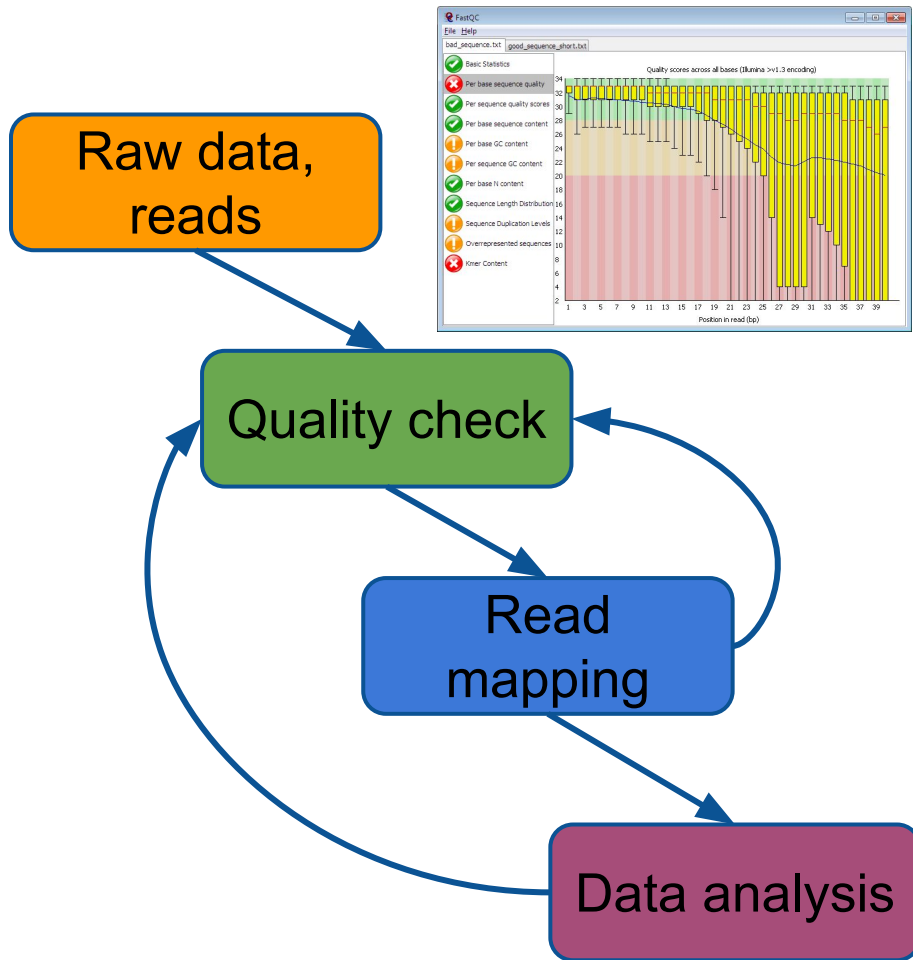
<https://sdtimes.com/23andme/sd-times-blog-using-dna-for-access-control/>

<https://www.pinterest.com/pin/702139398127827000/>

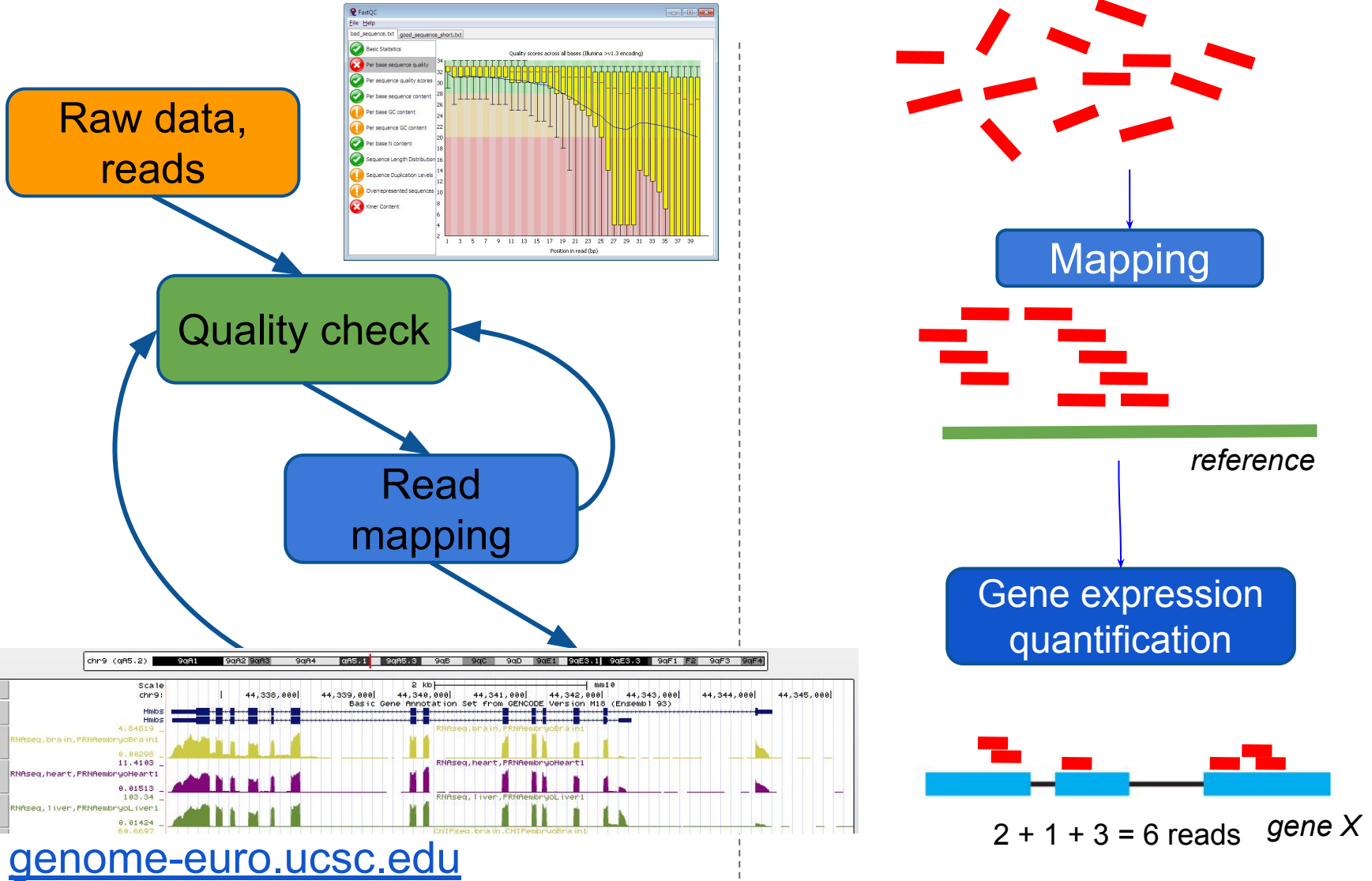
https://www.etsy.com/market/human_heart_svg

<https://www.vectorstock.com/royalty-free-vector/human-liver-black-icon-vector-7458173>

Typical pipeline



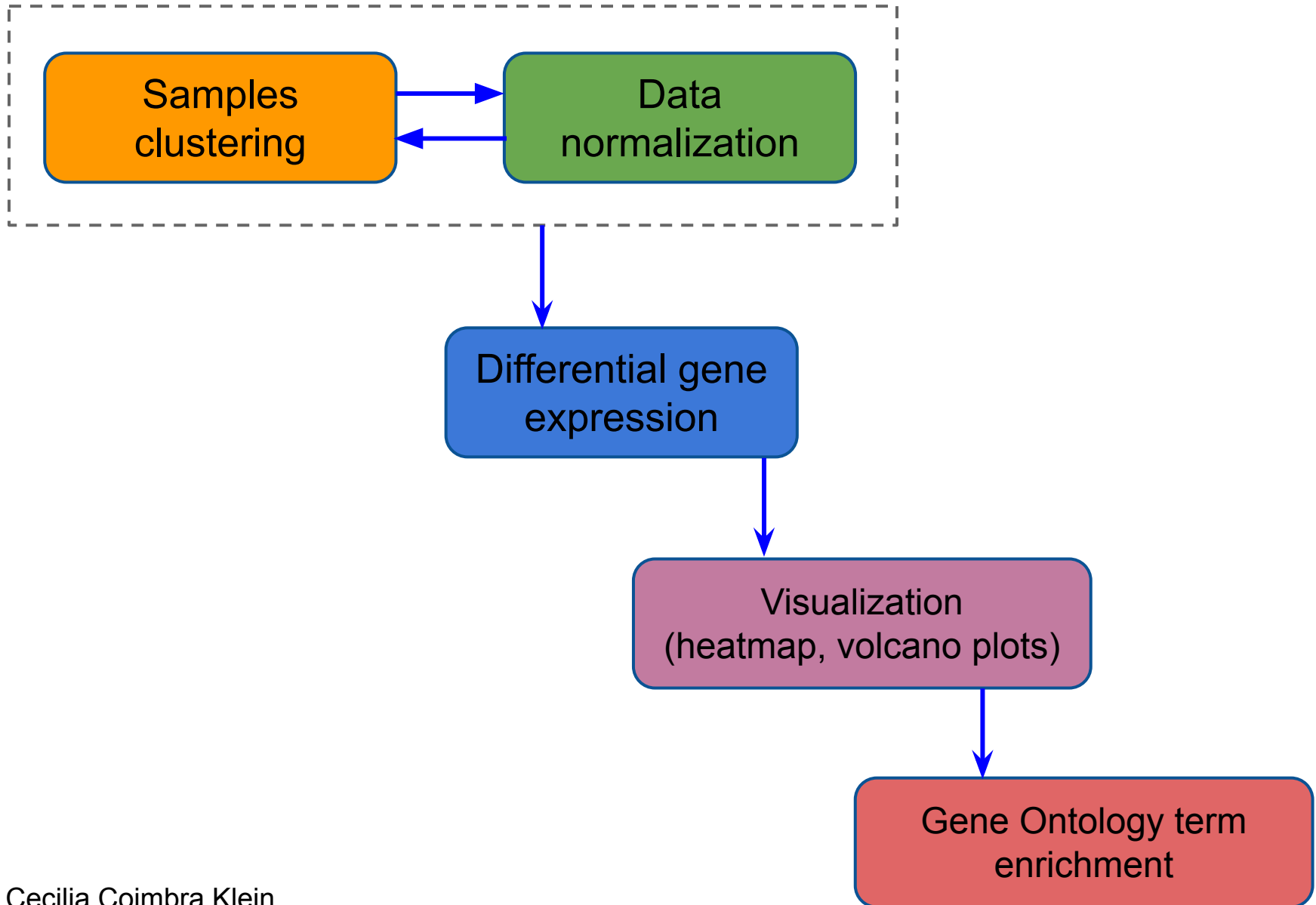
Typical pipeline



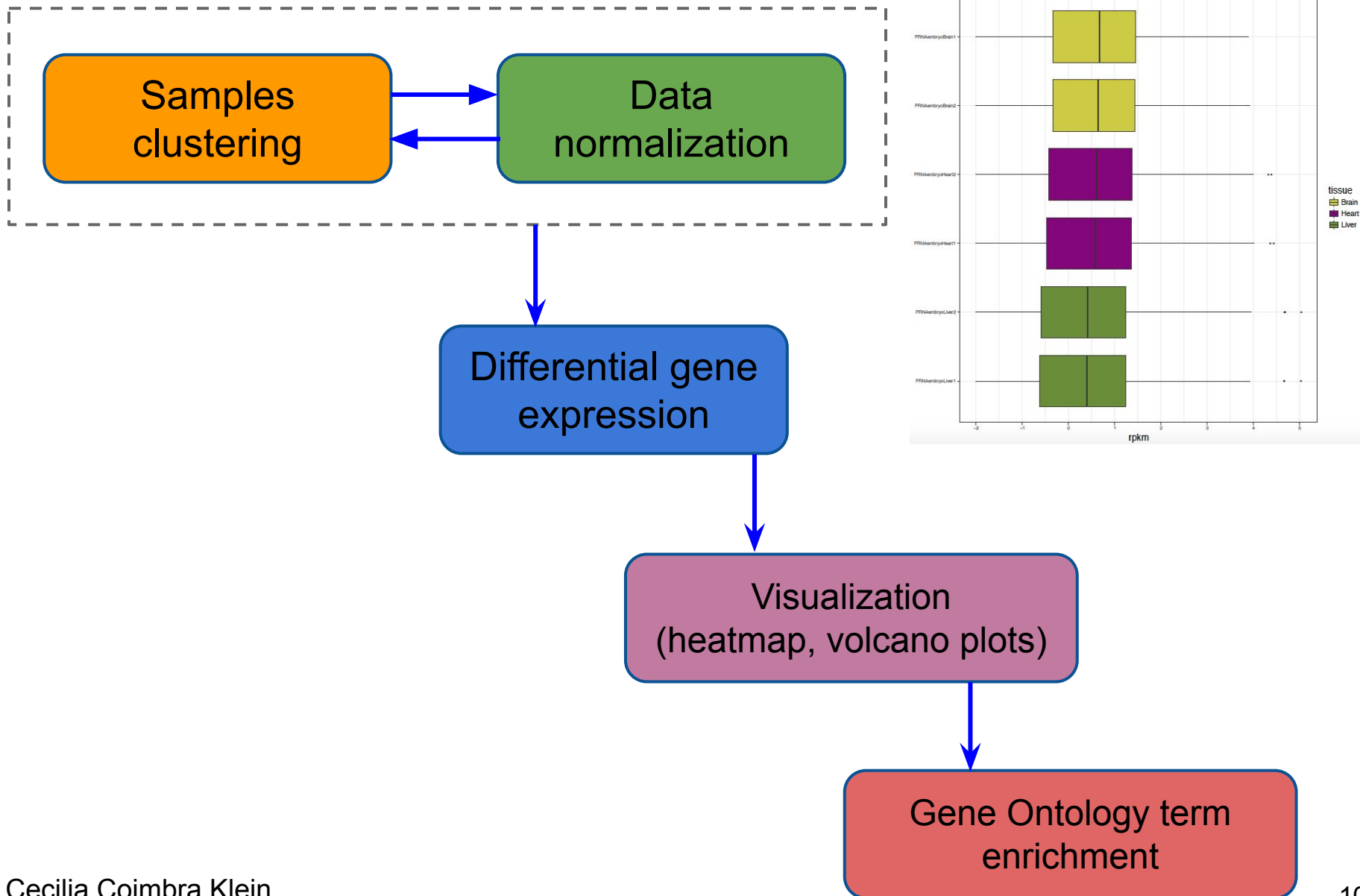
RNA-seq analysis

(clustering, differential gene expression, GO enrichment)

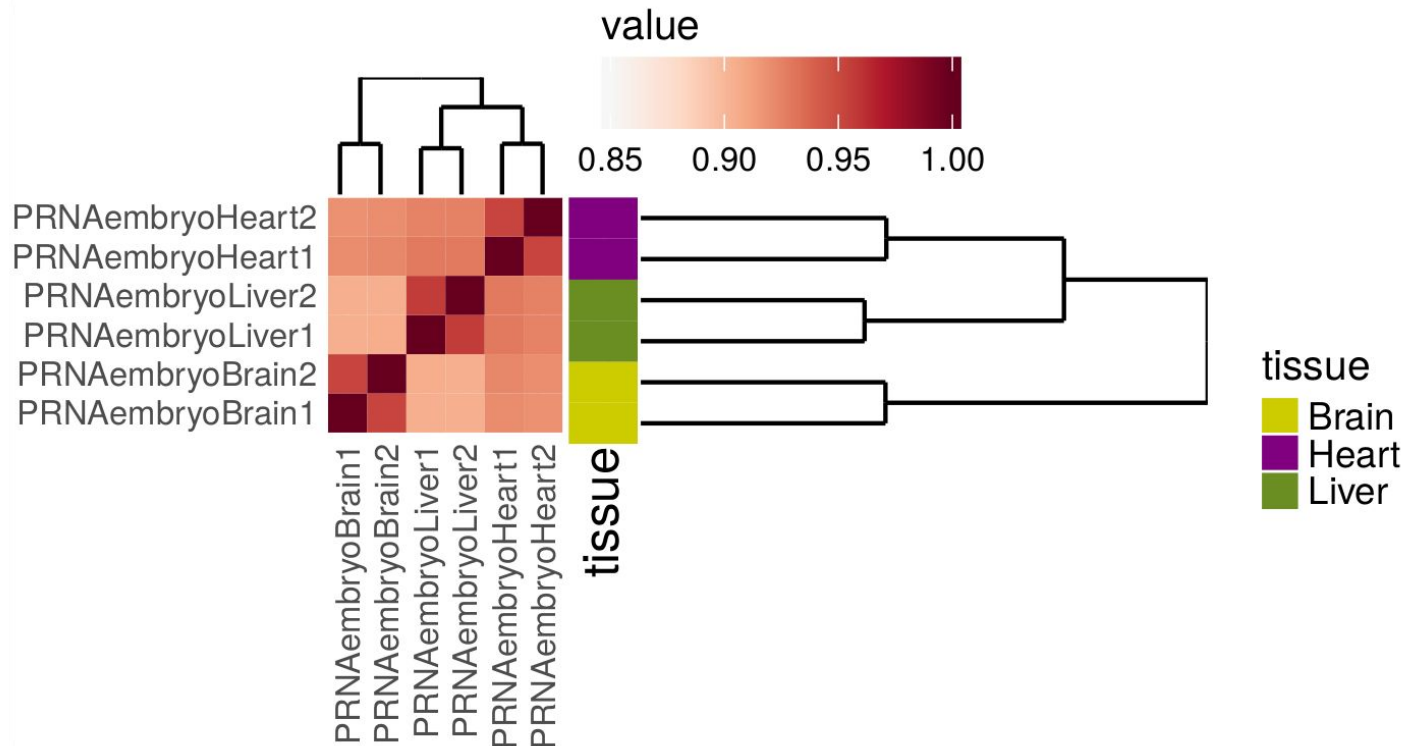
Analysis pipeline



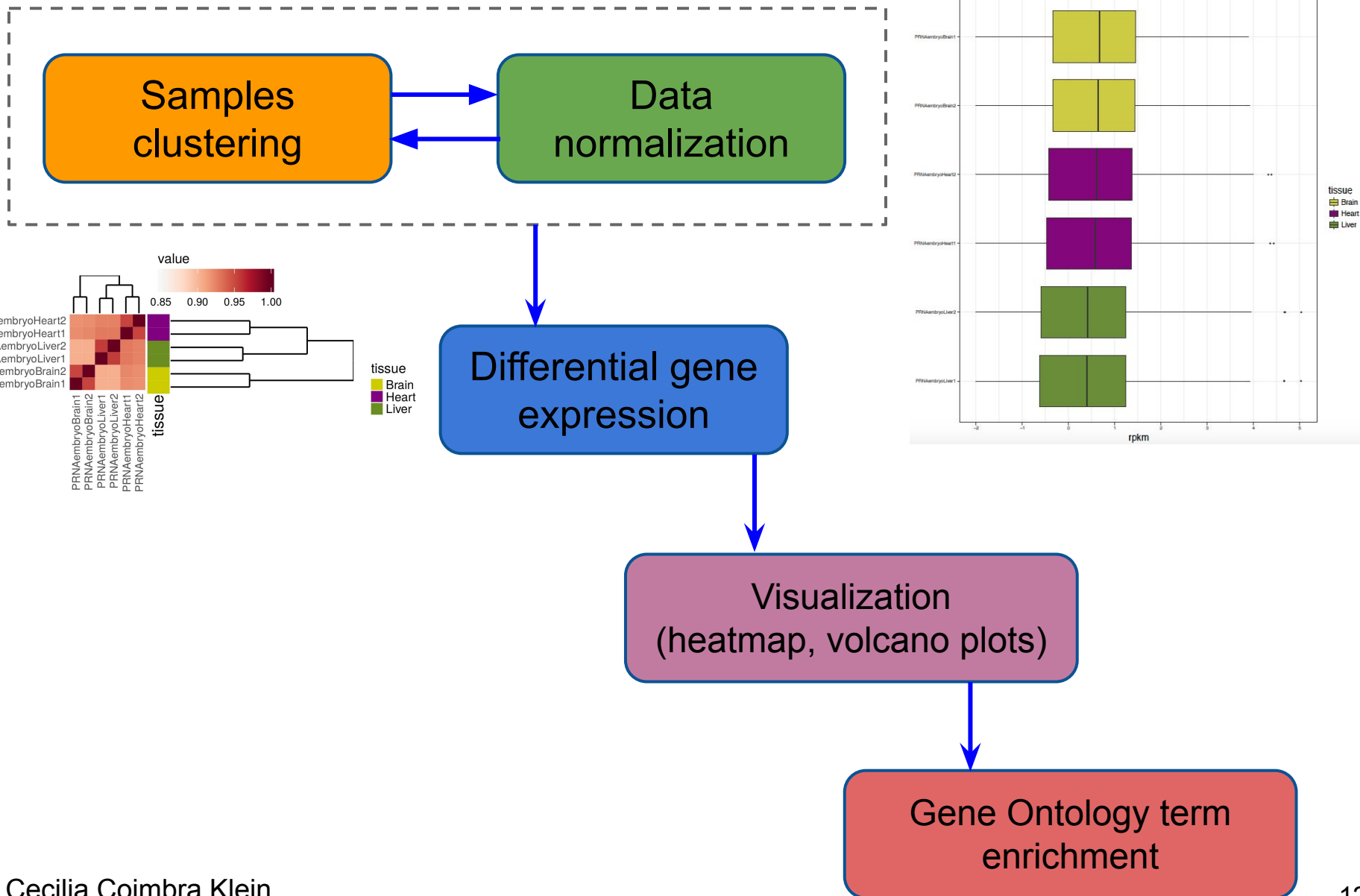
Analysis pipeline



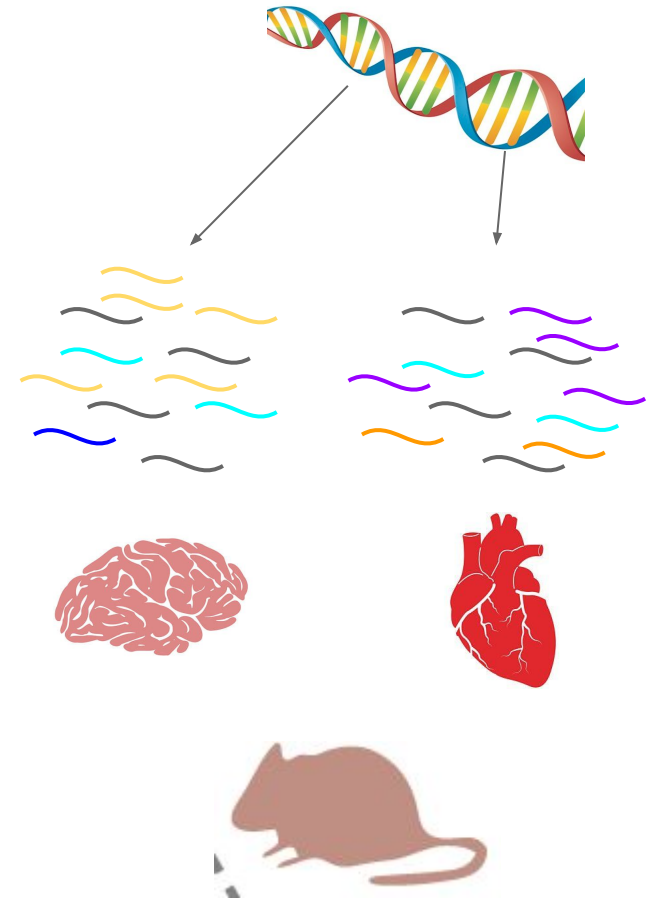
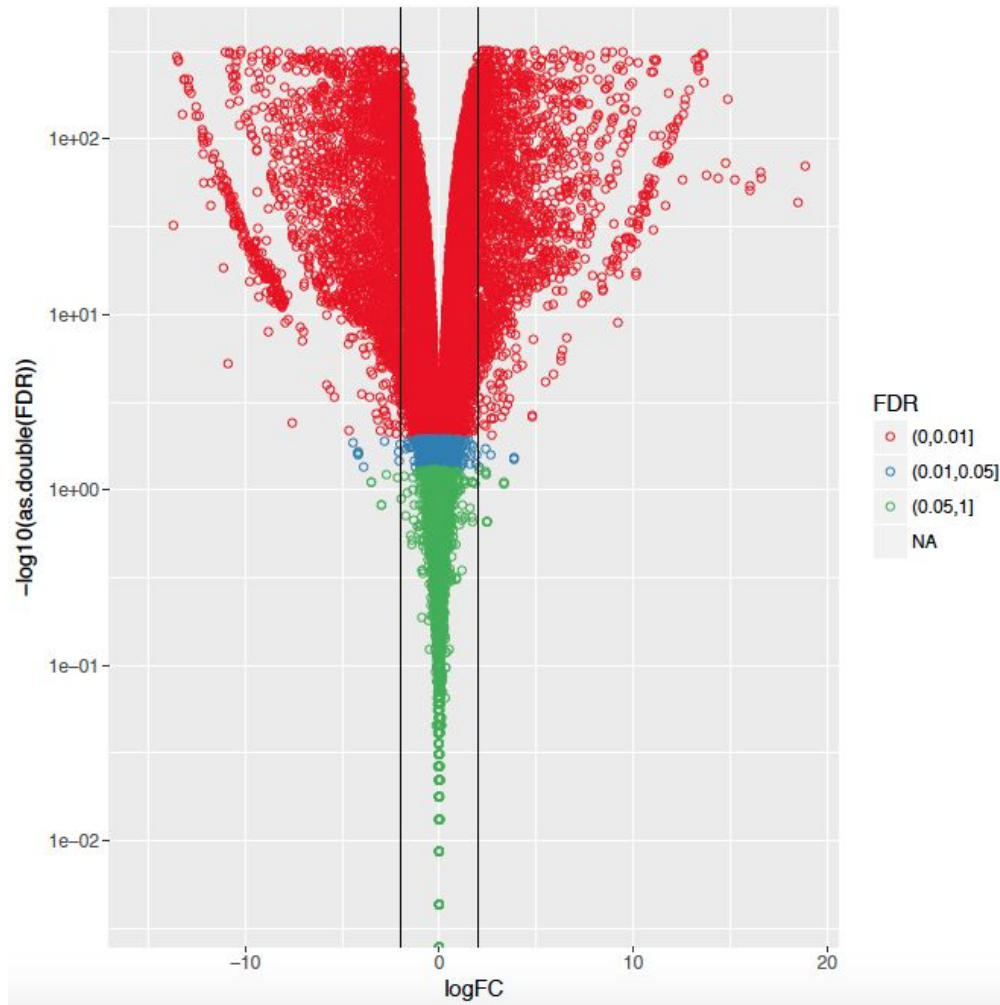
Samples clustering



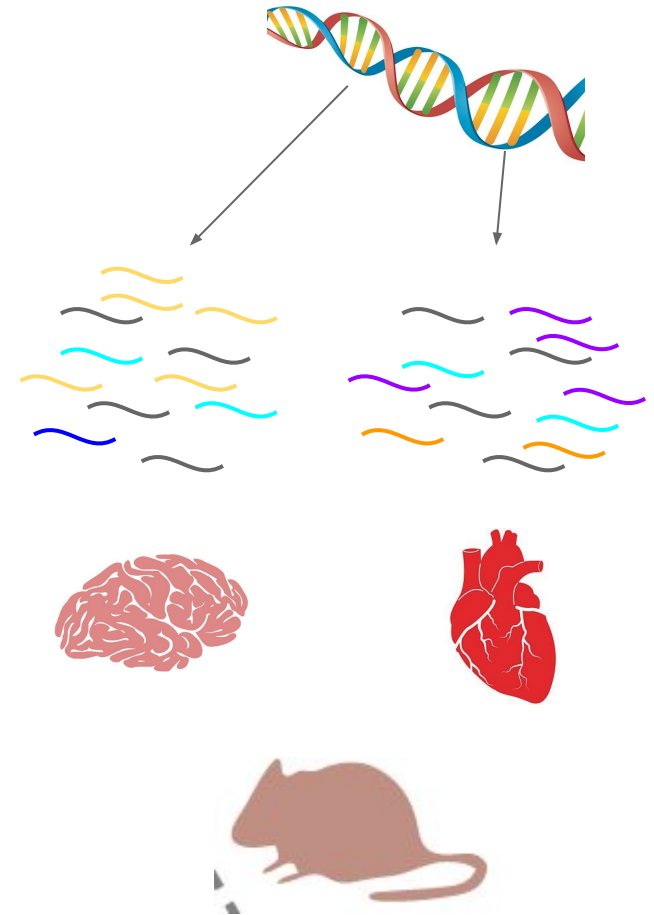
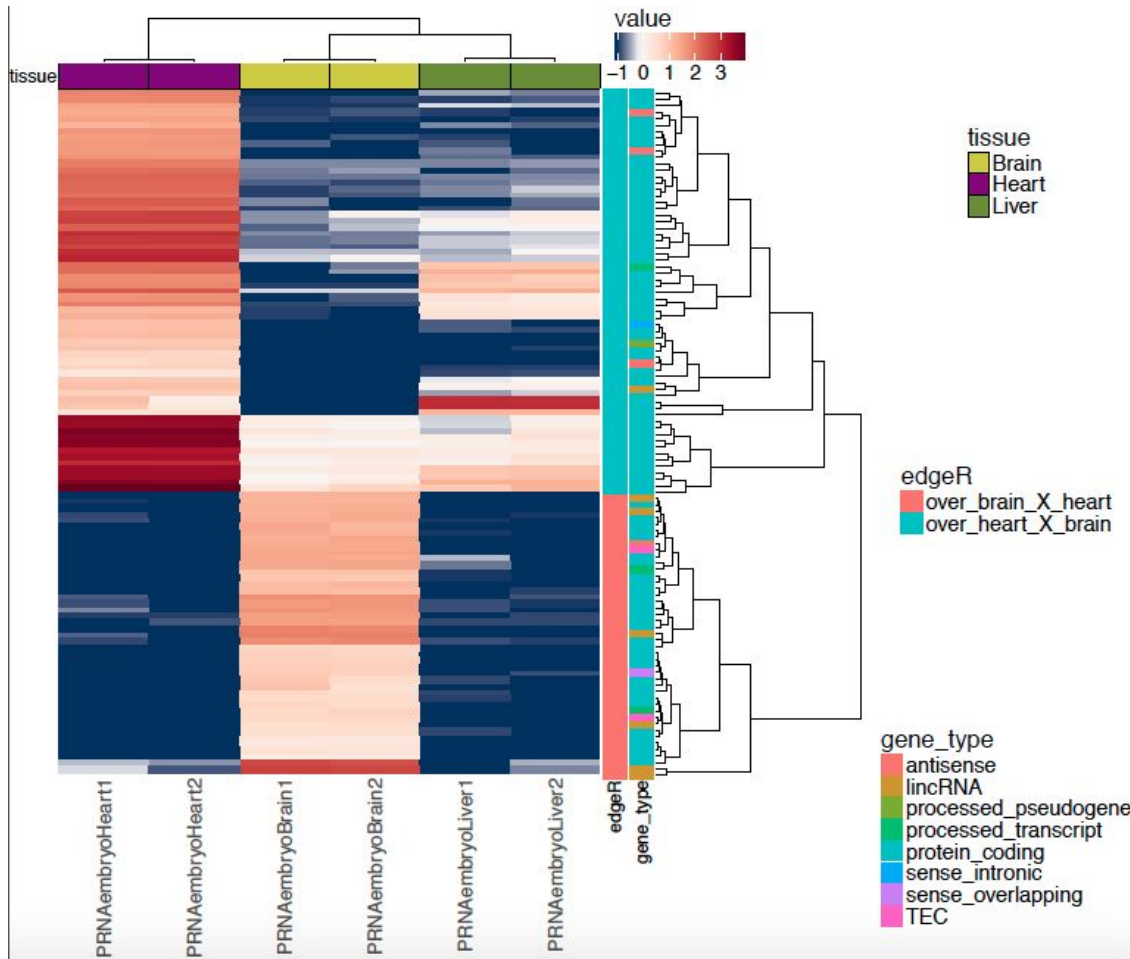
Analysis pipeline



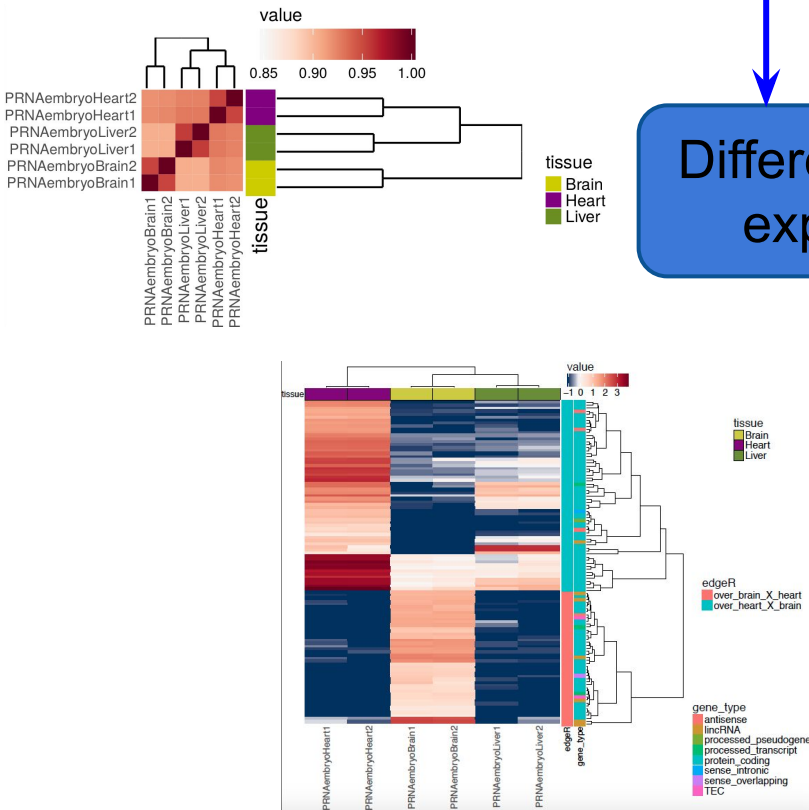
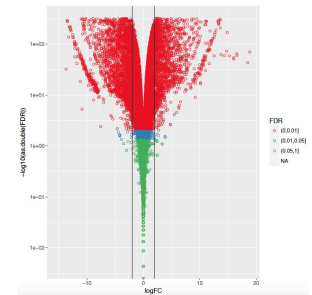
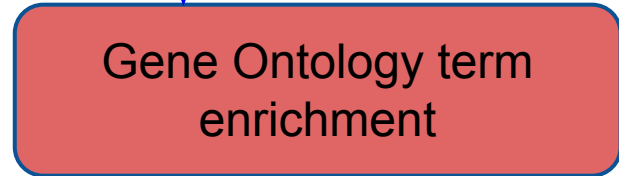
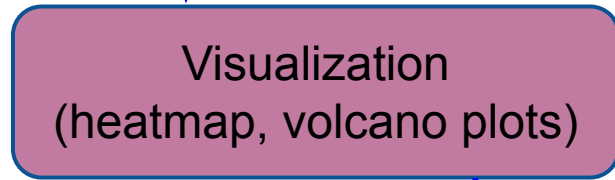
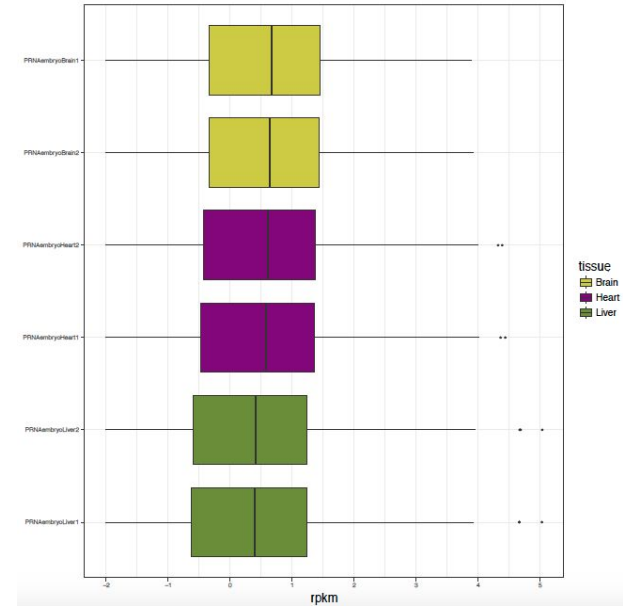
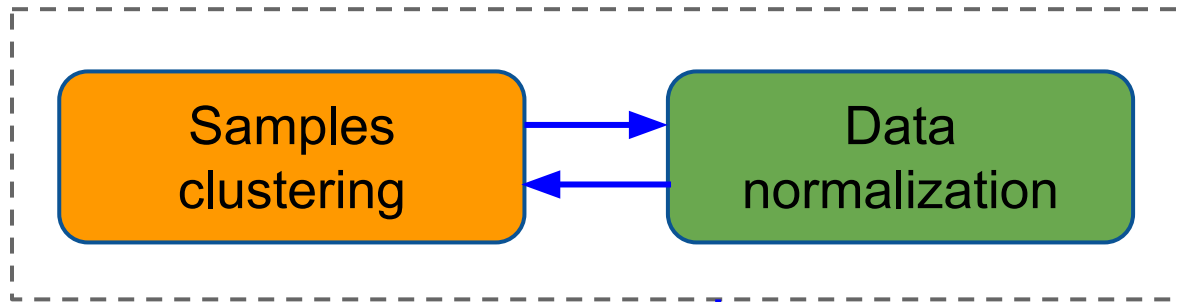
Differential Gene Expression



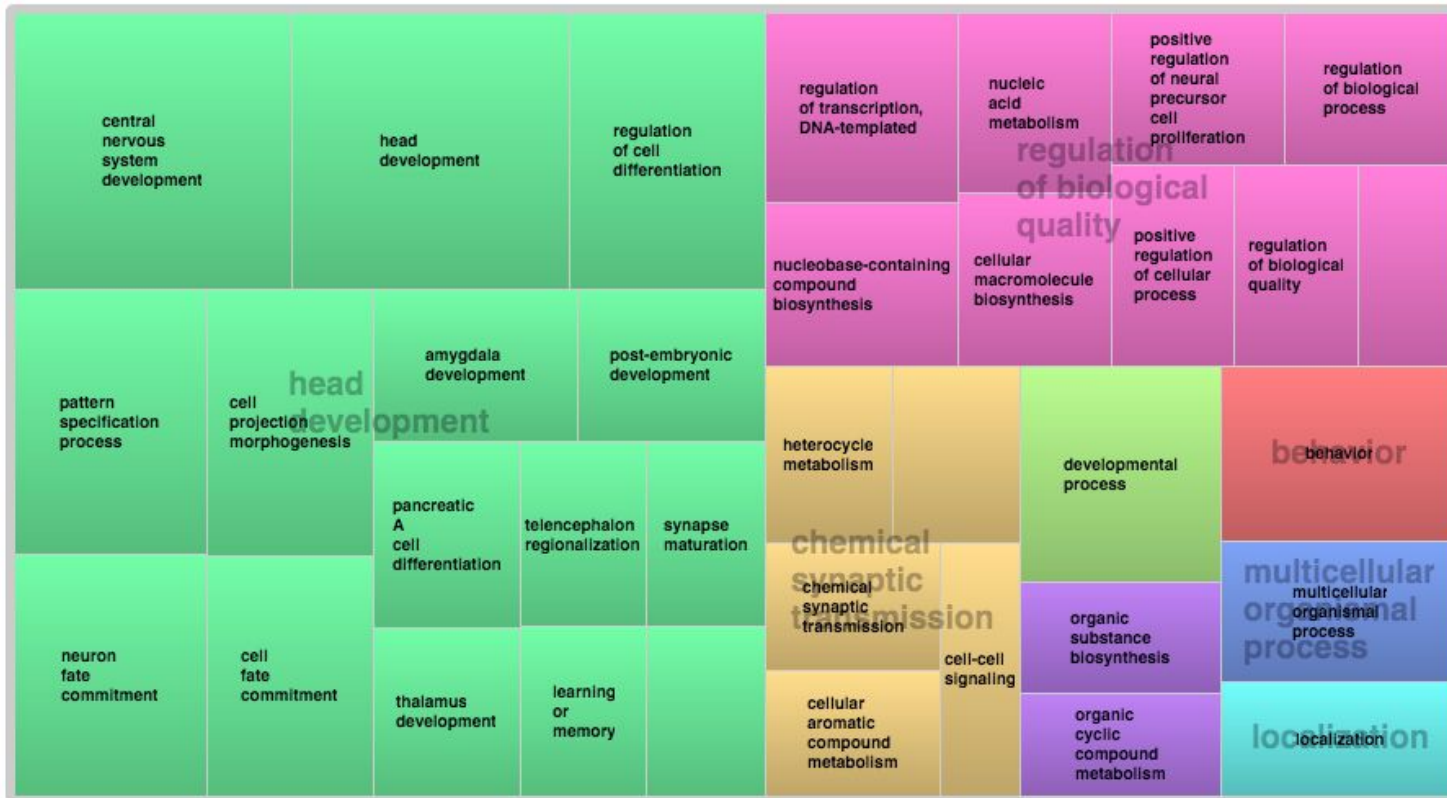
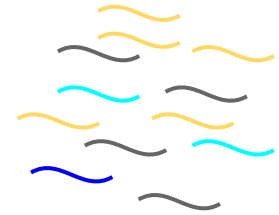
Differential Gene Expression



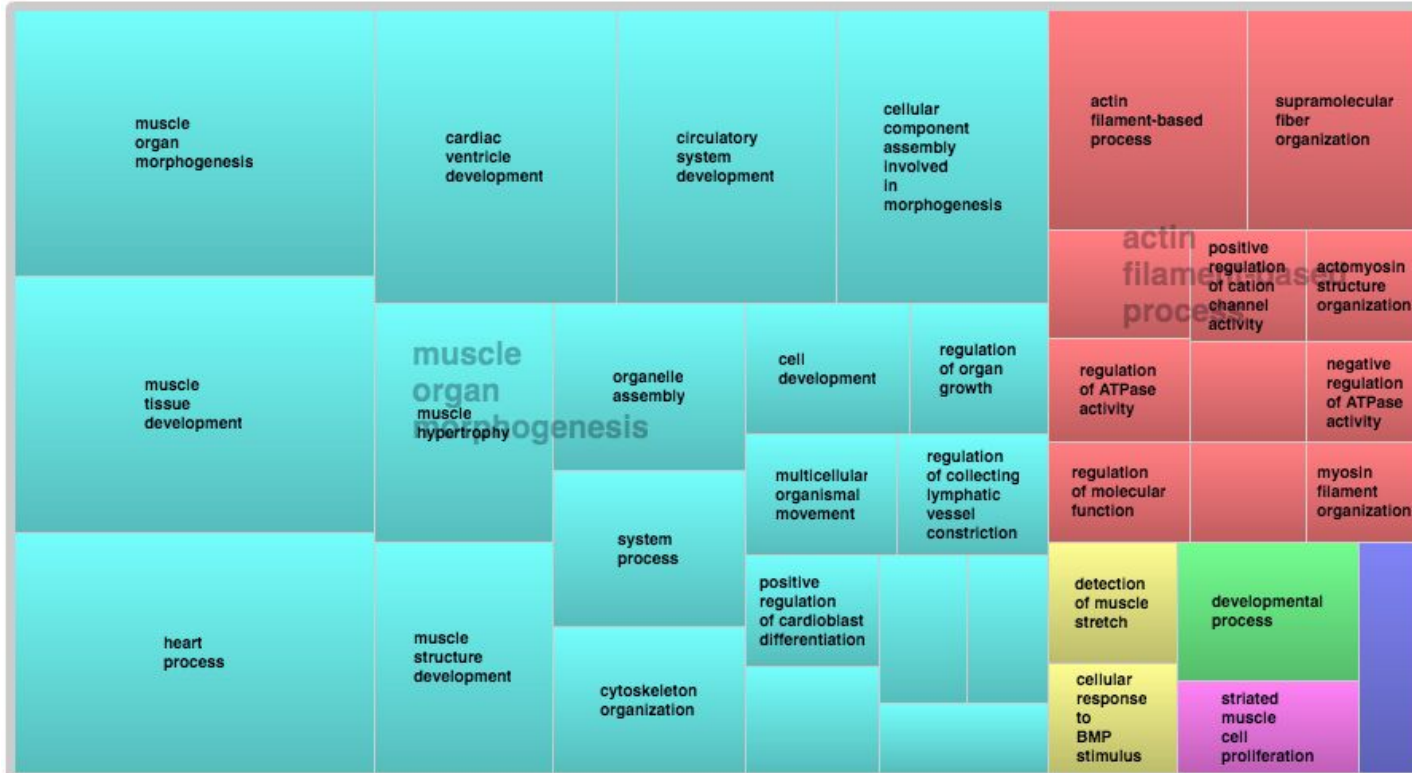
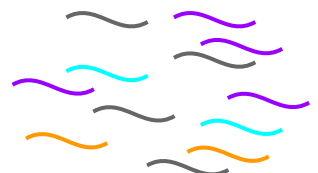
Analysis pipeline



GO term enrichment



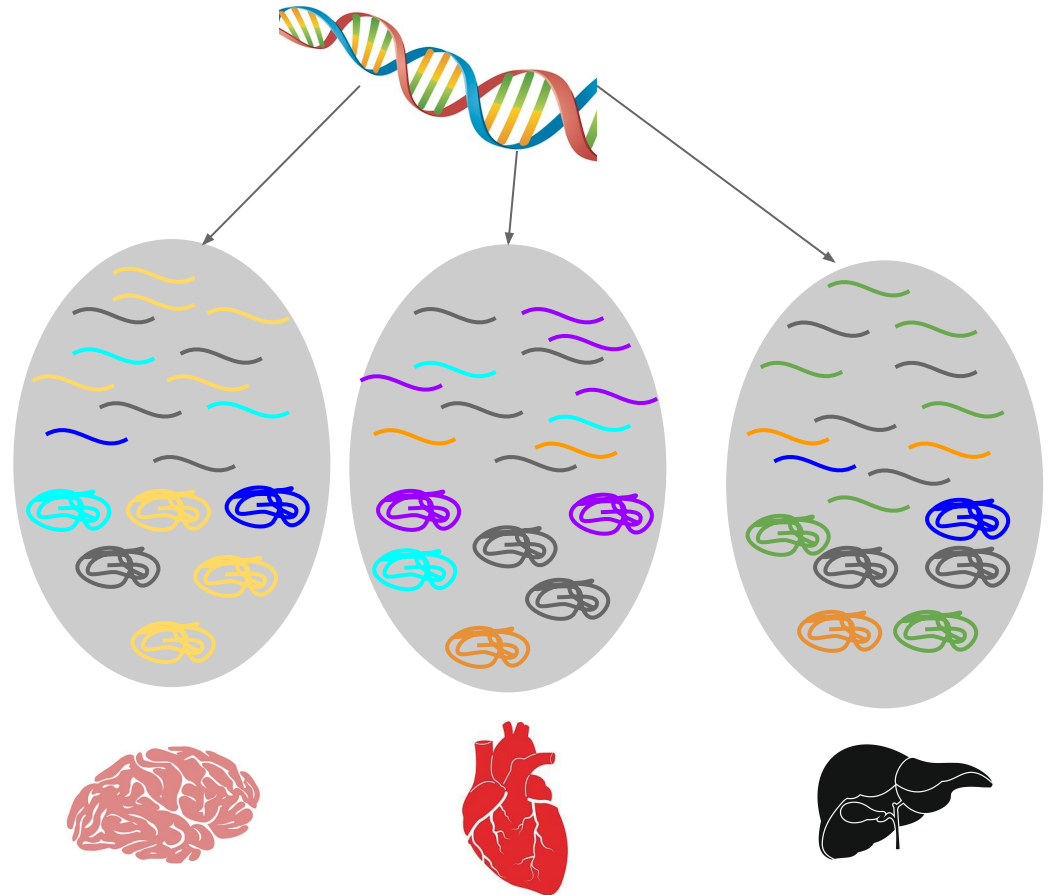
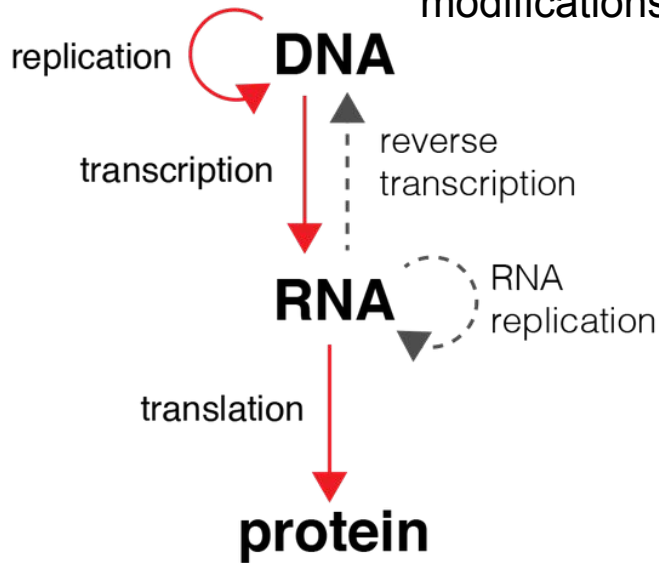
GO term enrichment



RNA-seq analysis (splicing)

Molecular biology dogma

epigenetic
modifications



- The **genome** is identical in all cell types, however not all cell types have the same function. That's why the **transcriptome** (and the **epigenome**) becomes also relevant.

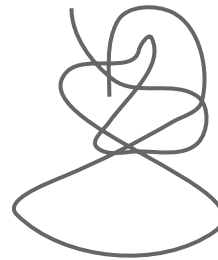
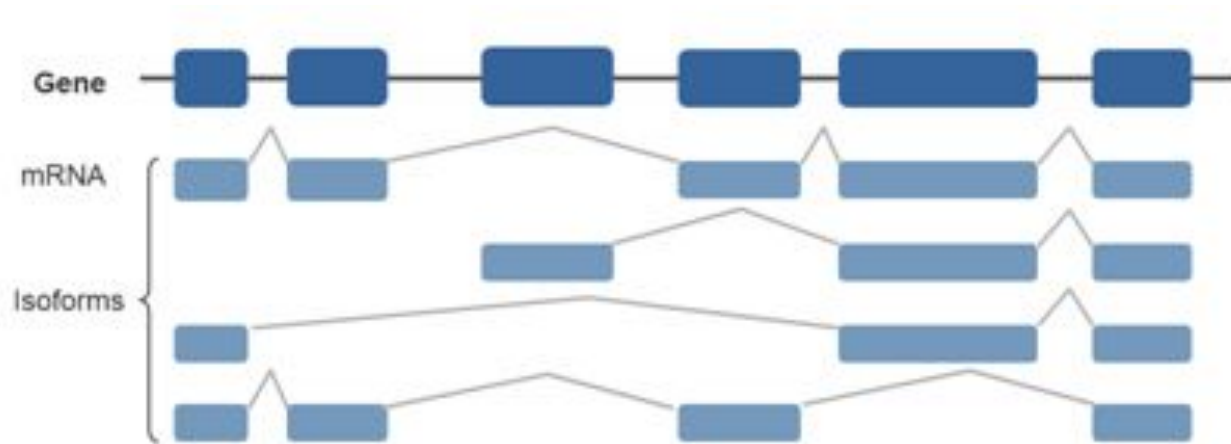
<https://sdtimes.com/23andme/sd-times-blog-using-dna-for-access-control/>

<https://www.pinterest.com/pin/702139398127827000/>

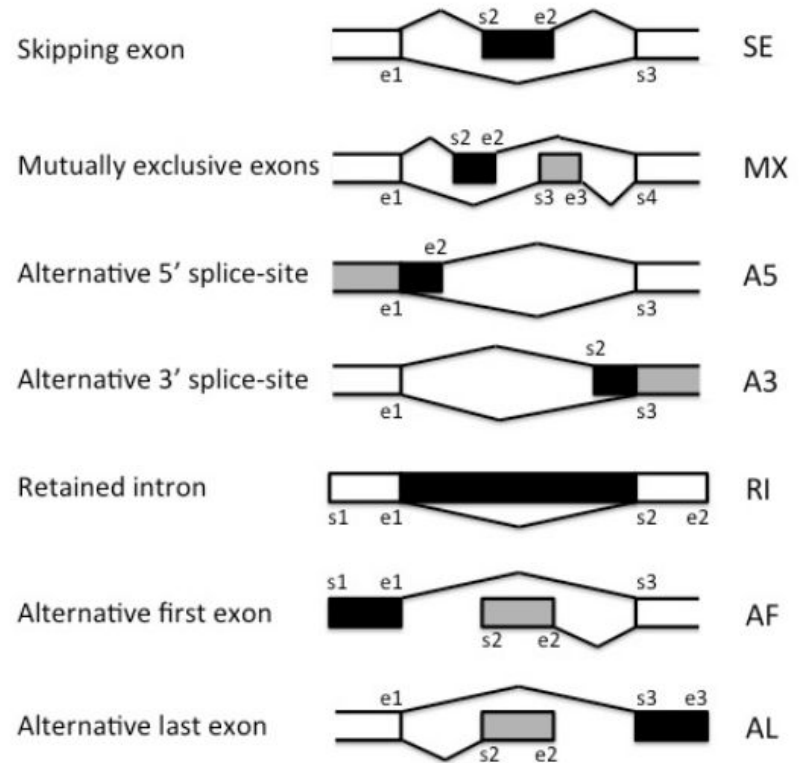
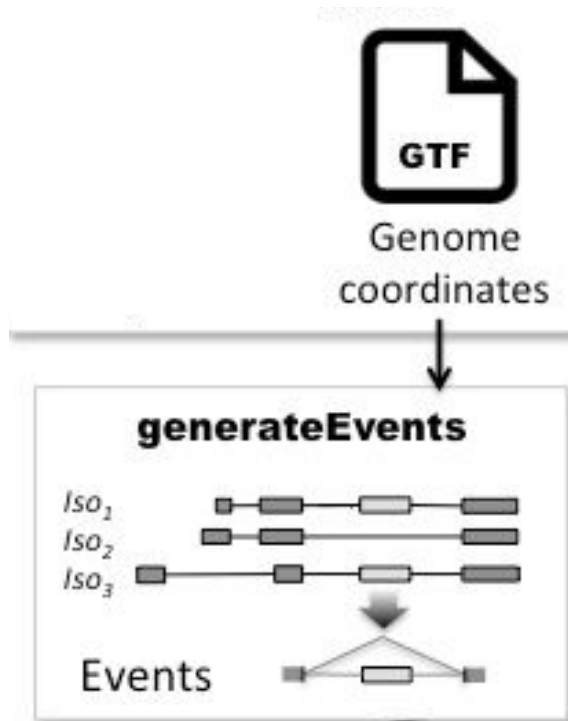
https://www.etsy.com/market/human_heart_svg

<https://www.vectorstock.com/royalty-free-vector/human-liver-black-icon-vector-7458173>

Isoform usage

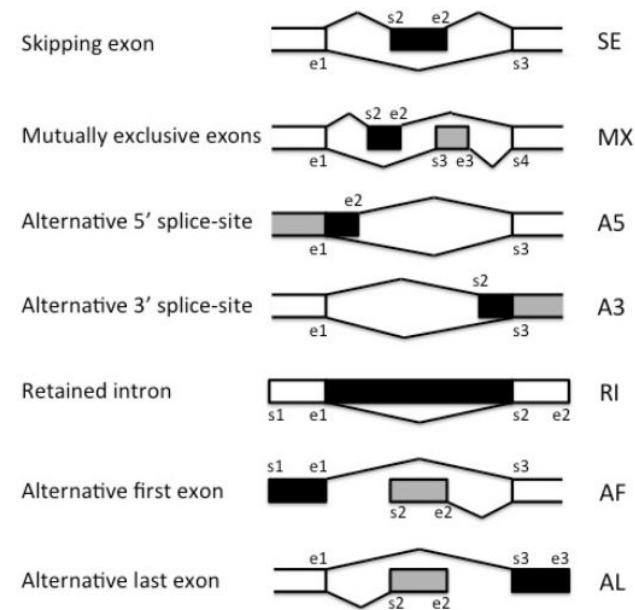
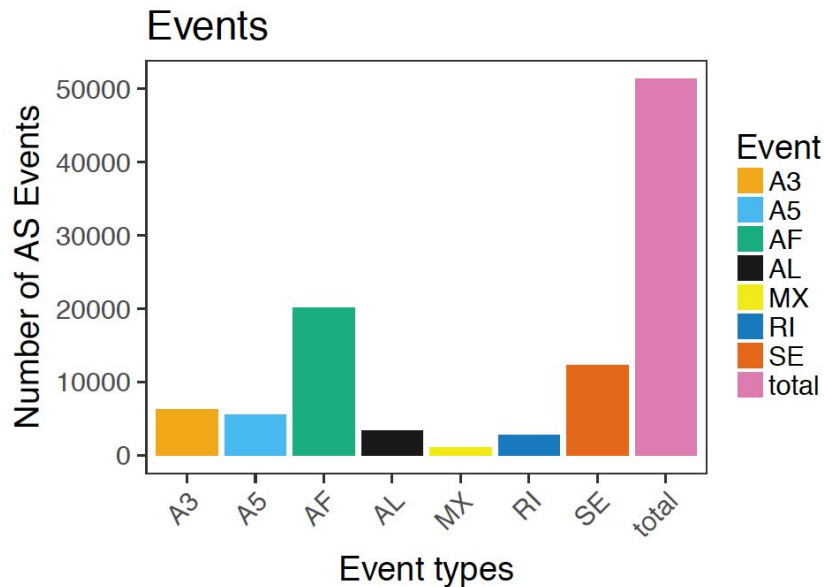


SUPPA: generate events based on gene annotation



<https://bitbucket.org/regulatorygenomicsupf/suppa>

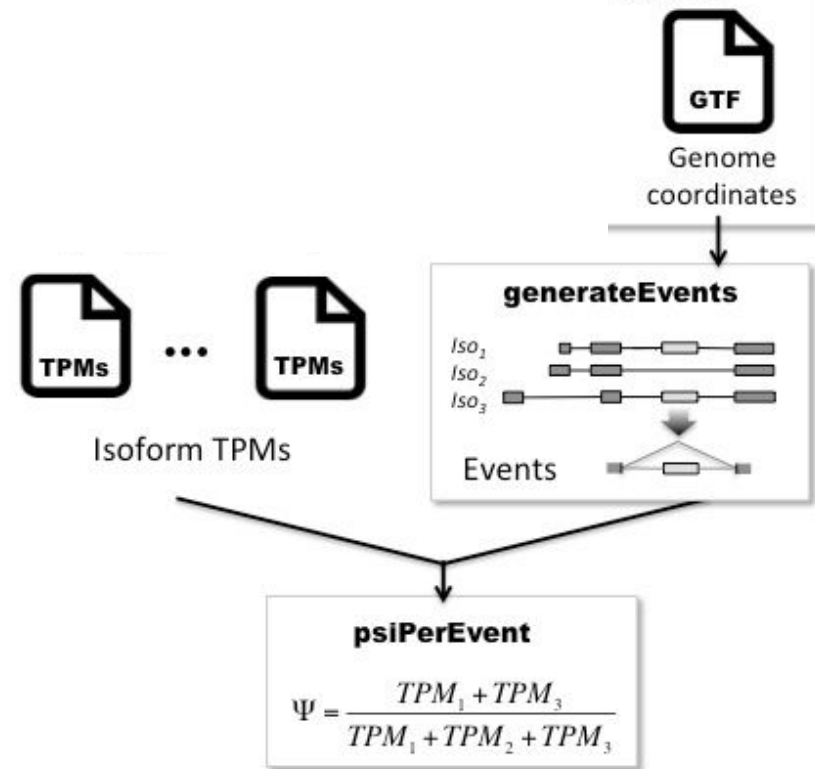
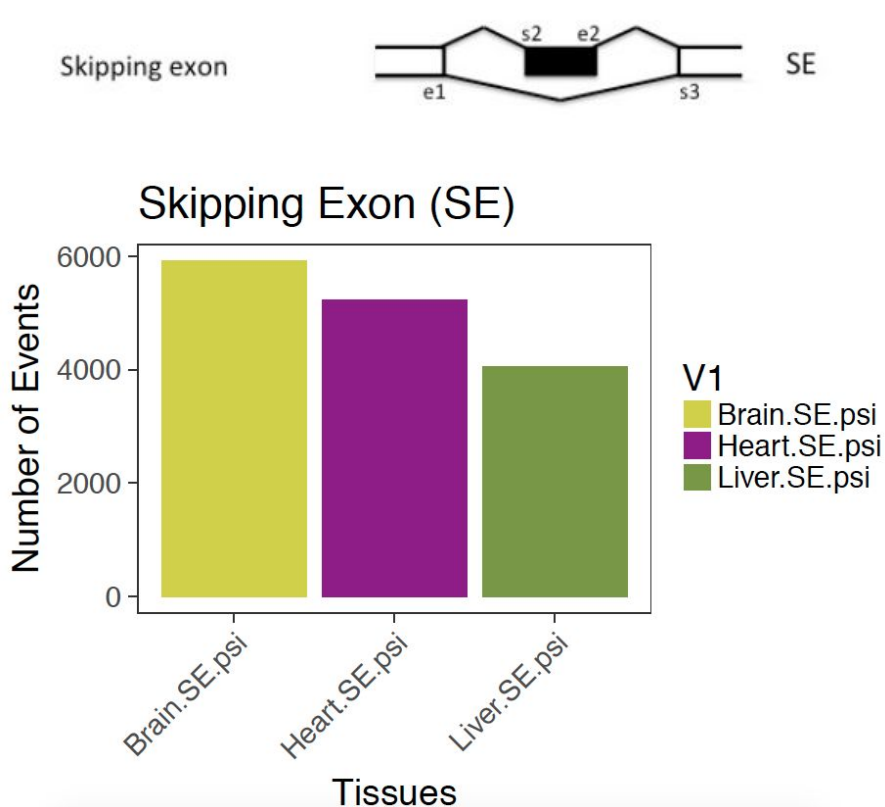
SUPPA: generate events based on gene annotation



```
# prepare input for plot
wc -l localEvents*ioe|awk '$2=="total"{print $2"\t"$1}$2!="total"{
split($2,a,"_"); print a[2]"\t"$1}' |sed 'liEvent\tNumber'> input.tsv

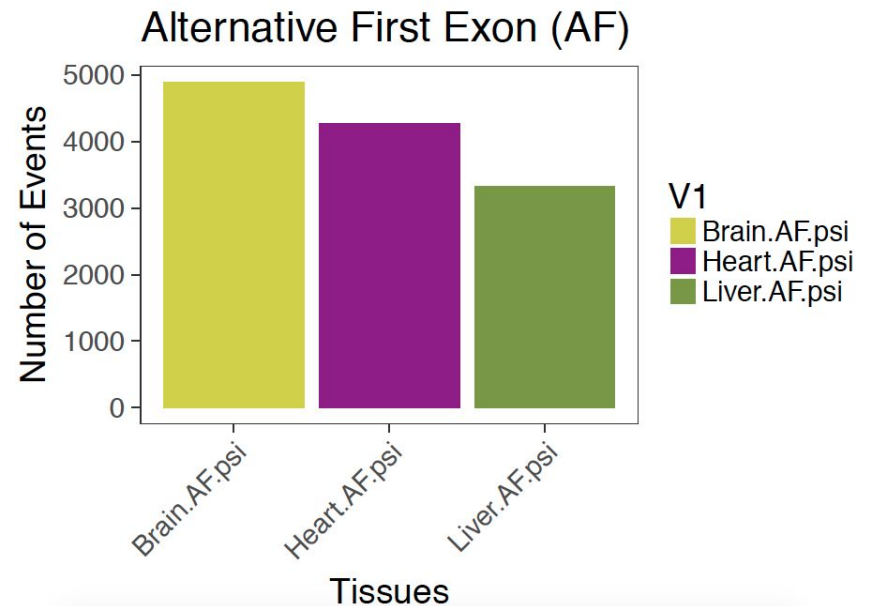
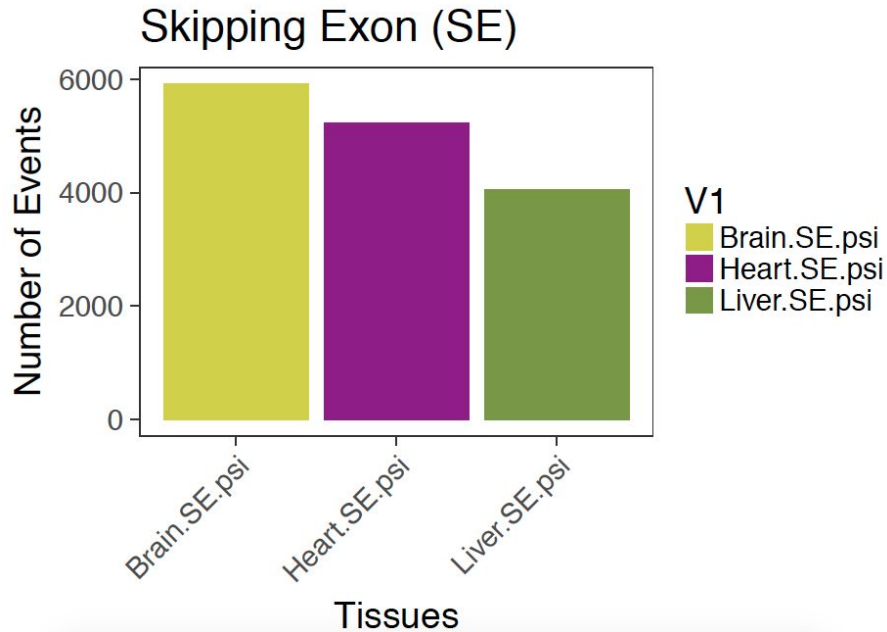
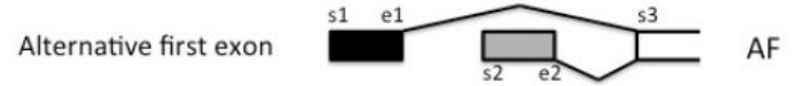
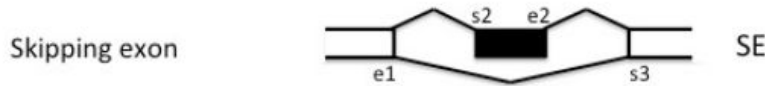
# number of alternative events with PSI values
ggbargplot.R -i input.tsv -o number_of_events.pdf --title "Events"
--y_title "Number of AS Events" --x_title "Event types" --palette_fill
/tutorial/palettes/cbbPalette.8.txt --fill_by 1 --header
```

SUPPA: Quantify event inclusion levels (PSIs)



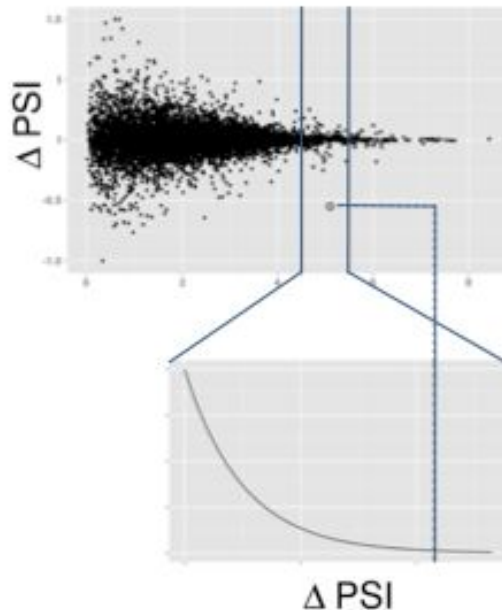
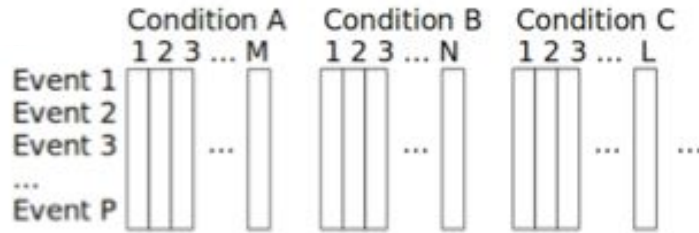
```
# number of alternative single exon skipping
ls *.SE.psi|while read f;do echo -e $f"\t"$(grep -v nan $f |wc -l);done |
ggbartplot.R -i stdin -o number_of_SE.pdf -f 1 --palette_fill
/tutorial/palettes/palTissue.txt --title "Skipping Exon (SE)" --y_title
"Number of Events" --x_title "Tissues"
```

SUPPA: Quantify event inclusion levels (PSIs)



```
# number of alternative first exons
ls *.AF.psi|while read f;do echo -e $f"\t"$(grep -v nan $f |wc -l);done |
ggbartplot.R -i stdin -o number_of_AF.pdf -f 1 --palette_fill
/tutorial/palettes/palTissue.txt --title "Alternative First Exon (AF)"
--y_title "Number of Events" --x_title "Tissues"
```

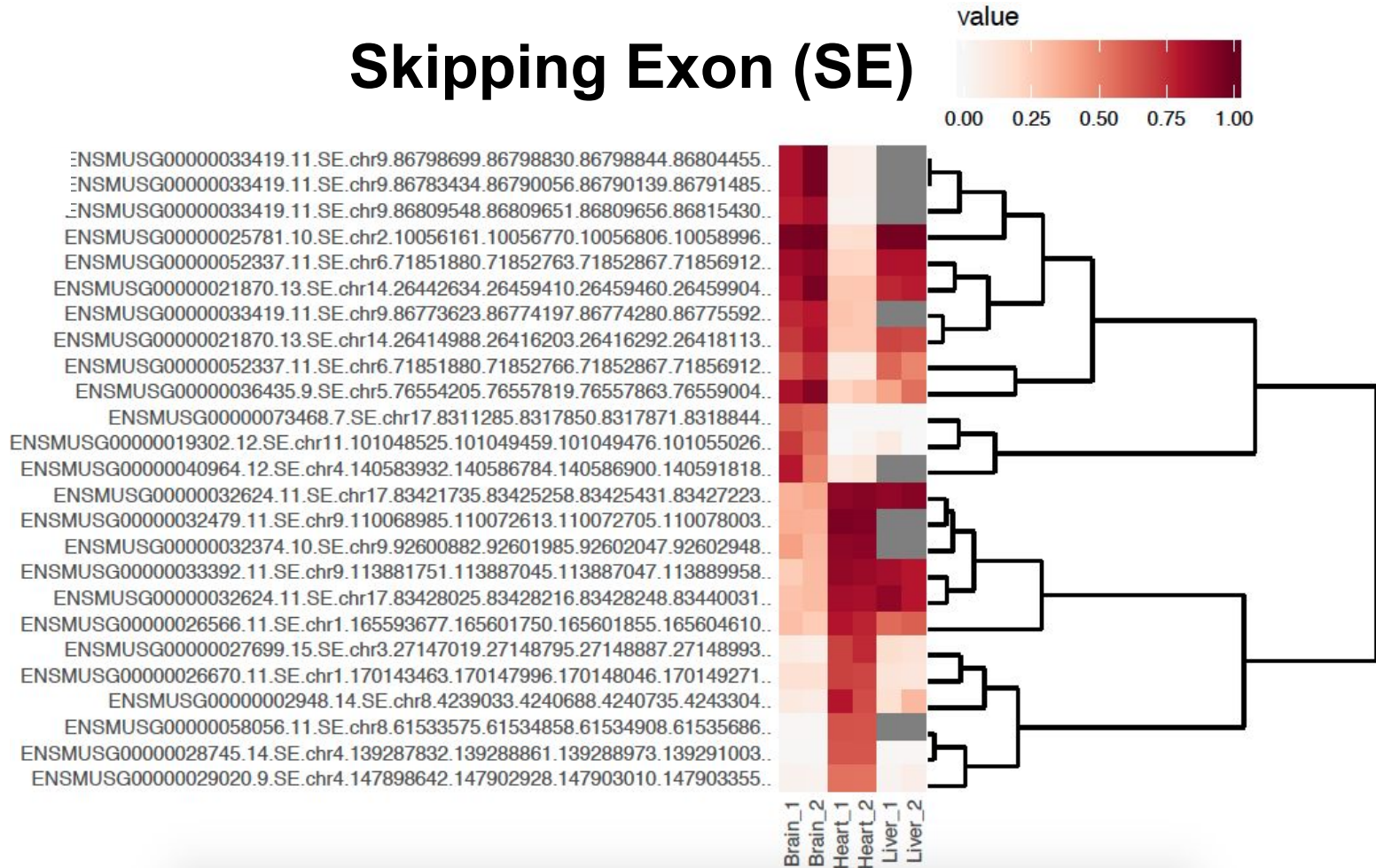

SUPPA: compare conditions



- SUPPA calculates the magnitude of splicing change (Δ PSI) and their significance across multiple biological conditions, using two or more replicates per condition.
- Statistical significance is calculated by comparing the observed Δ PSI between conditions with the distribution of the Δ PSI between replicates as a function of the gene expression (measured as the expression of the transcripts defining the events).

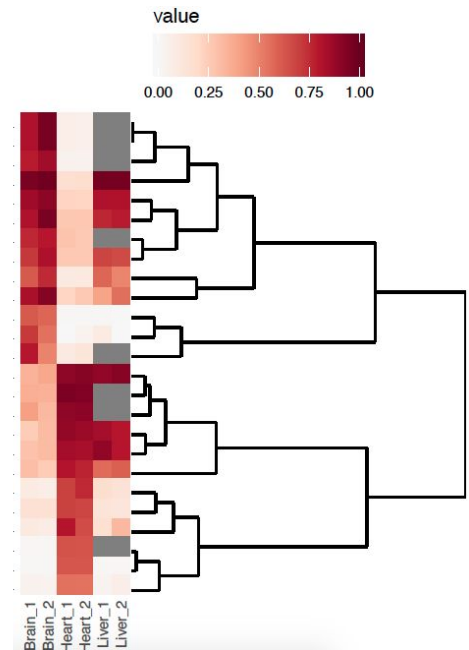
<https://bitbucket.org/regulatorygenomicsupf/suppa>

Skipping Exon (SE)



Skipping Exon (SE)

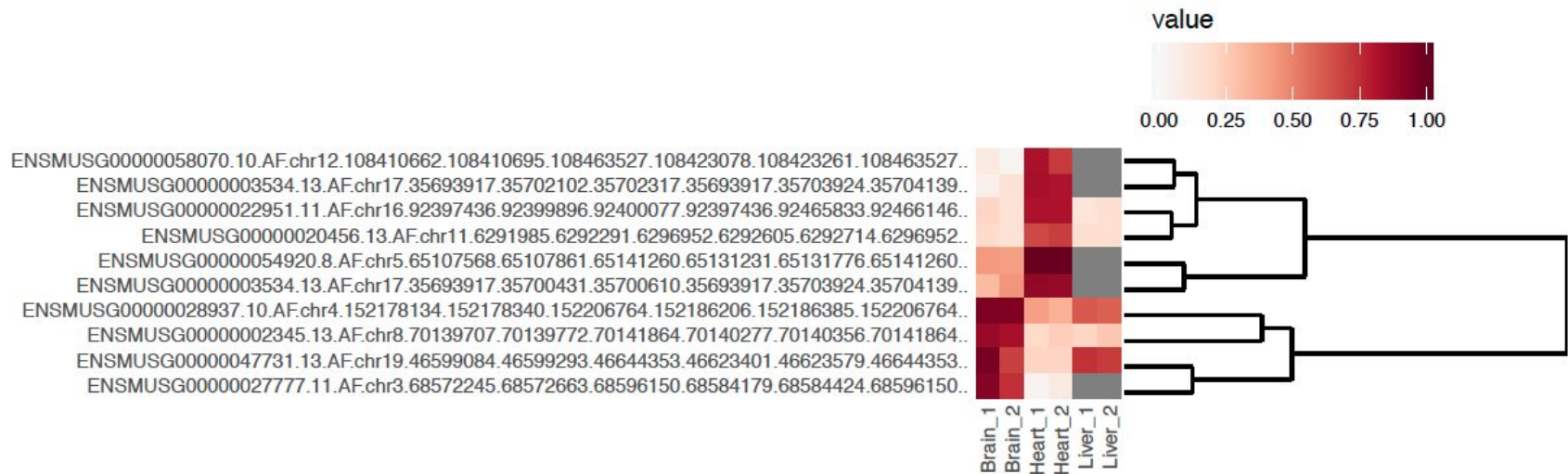
- Select top events from pairwise comparison of Brain and heart
- p-value < 0.05
- $\Delta\text{PSI} > 0.5$
 - PSI = 1 inclusion of exon
 - PSI = 0 exclusion of exon



```
# prepare input for heatmap
event=SE; awk 'BEGIN{FS=OFS="\t"}NR>1 && $2!="nan" && ($2>0.5 || $2<-0.5)
&& $3<0.05{print}' DS.${event}.dpsi|cut -f1 > top-examples-SE.txt
selectMatrixRows.sh top-examples-SE.txt DS.SE.psvvec >
matrix.top-examples-SE.tsv
```

```
# heatmap SE
ggheatmap.R -i matrix.top-examples-SE.tsv -o heatmap_top-examples-SE.pdf
--matrix_palette /tutorial/palettes/palSequential.txt --row_dendro
--matrix_fill_limits "0,1" -B 8
```

Alternative First exon (AF)

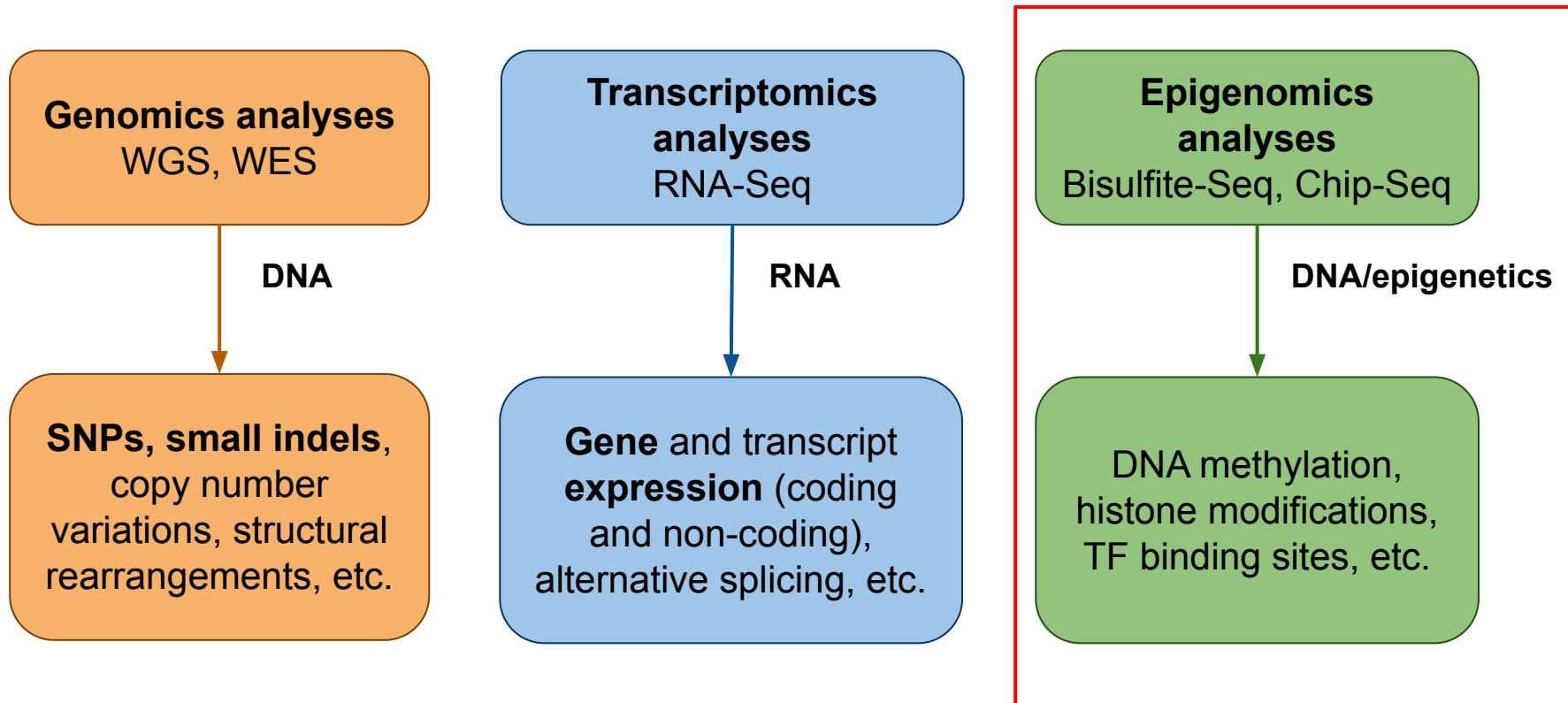


```
# prepare input for heatmap
event=AF; awk 'BEGIN{FS=OFS="\t"}NR>1 && $2!="nan" && ($2>0.5 || $2<-0.5)
&& $3<0.05{print}' DS.${event}.dpsi|cut -f1 > top-examples-AF.txt
selectMatrixRows.sh top-examples-AF.txt DS.AF.psivec >
matrix.top-examples-AF.tsv
```

```
# heatmap alternative first exons top examples
ggheatmap.R -i matrix.top-examples-AF.tsv -o heatmap_top-examples-AF.pdf
--matrix_palette /tutorial/palettes/palSequential.txt --row_dendro
--matrix_fill_limits "0,1" -B 8
```

Epigenomics

Which *-Seq do I need?



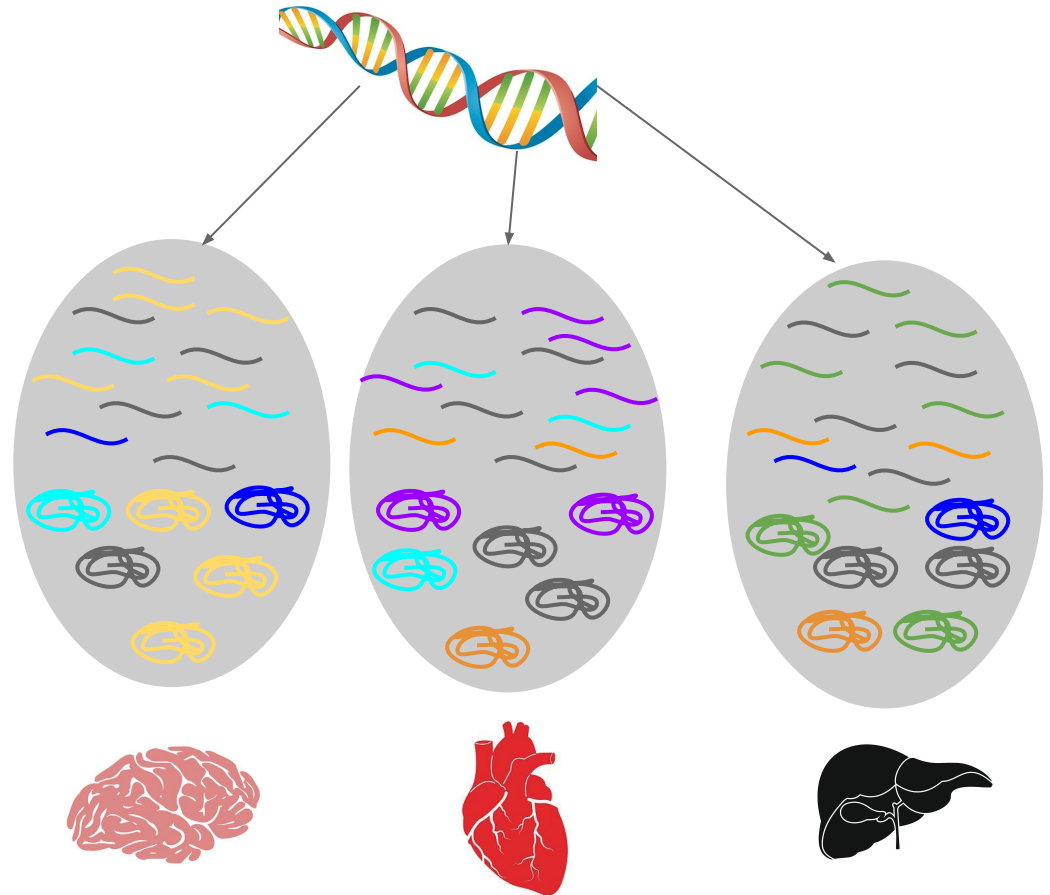
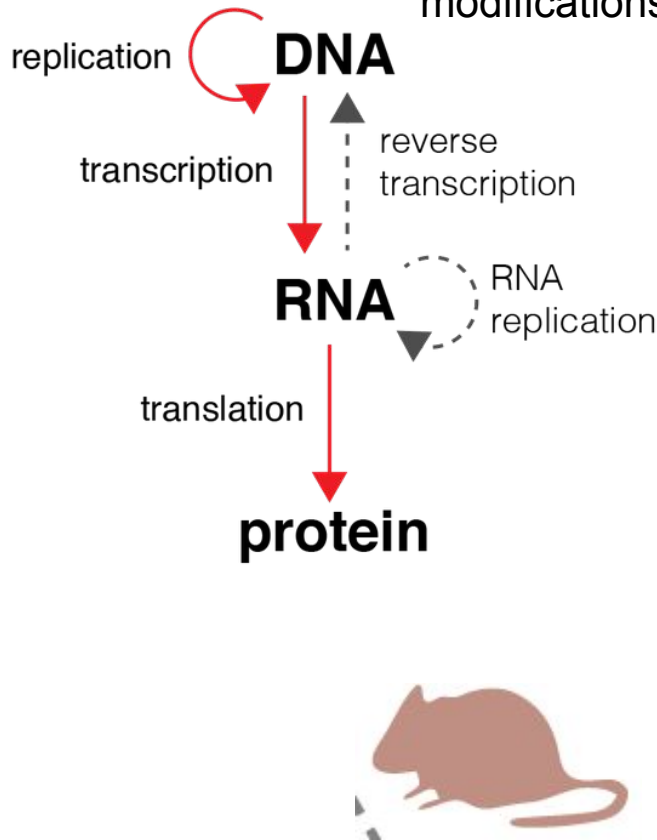
- Learn more about your favourite *-Seq [here!](#)
- Note that we are always talking about *re-sequencing*, which is something different from *de novo sequencing* (what is done for a new genome assembly)

ChIP-seq

Basic concepts

Molecular biology dogma

epigenetic
modifications



- The **genome** is identical in all cell types, however not all cell types have the same function. That's why the **transcriptome** (and the **epigenome**) becomes also relevant.

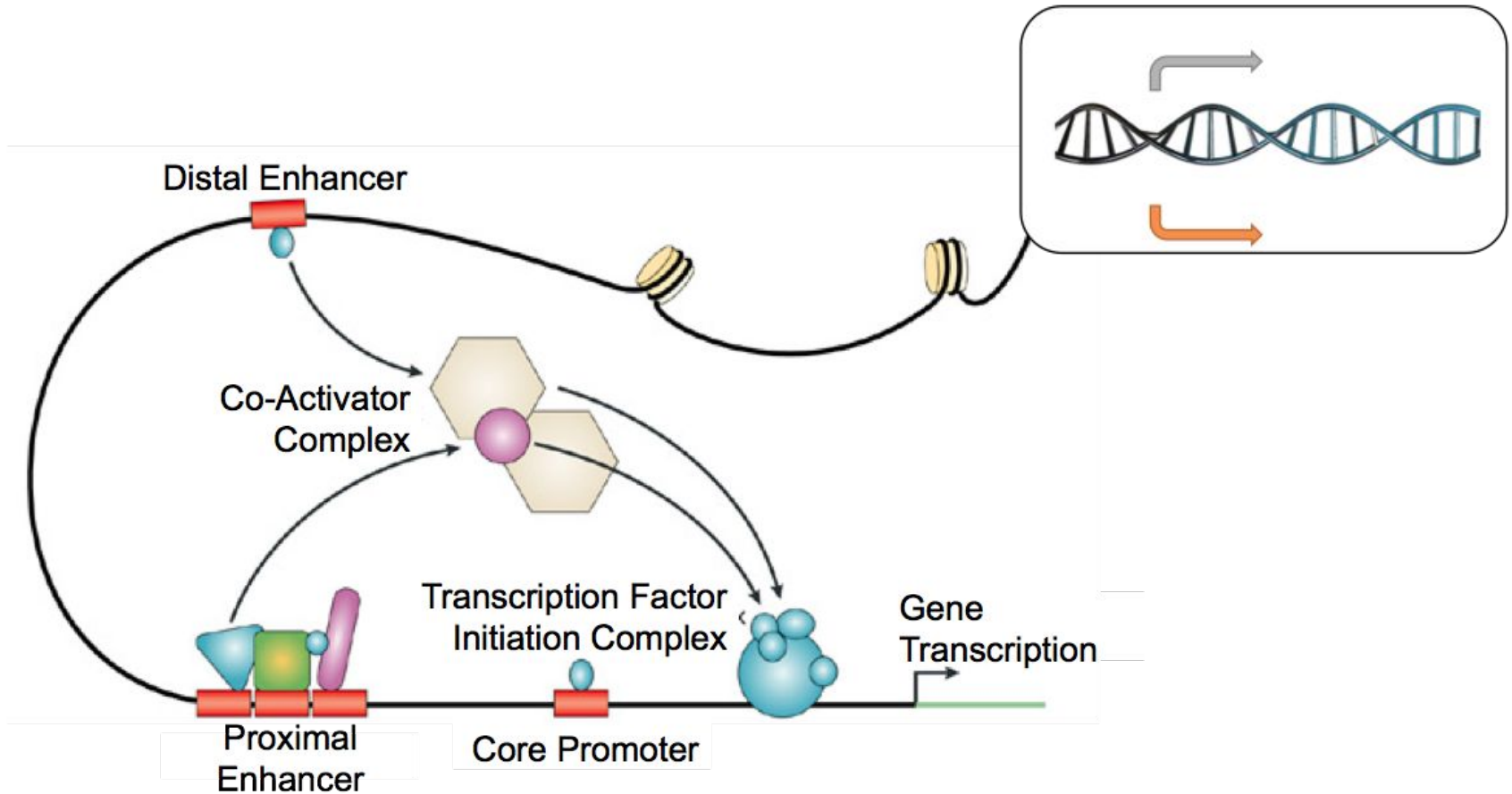
<https://sdtimes.com/23andme/sd-times-blog-using-dna-for-access-control/>

<https://www.pinterest.com/pin/702139398127827000/>

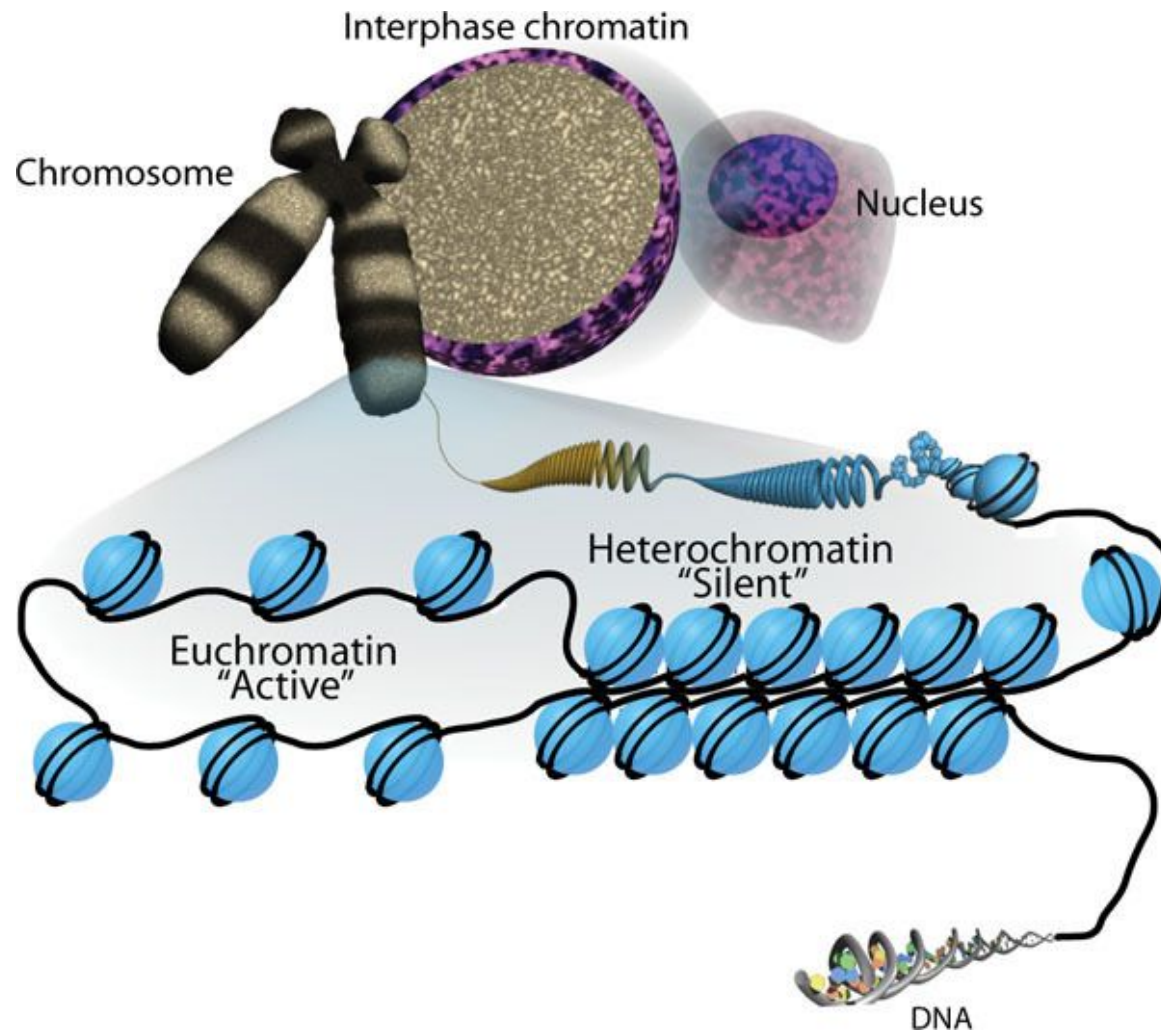
https://www.etsy.com/market/human_heart_svg

<https://www.vectorstock.com/royalty-free-vector/human-liver-black-icon-vector-7458173>

Dynamics of gene regulation

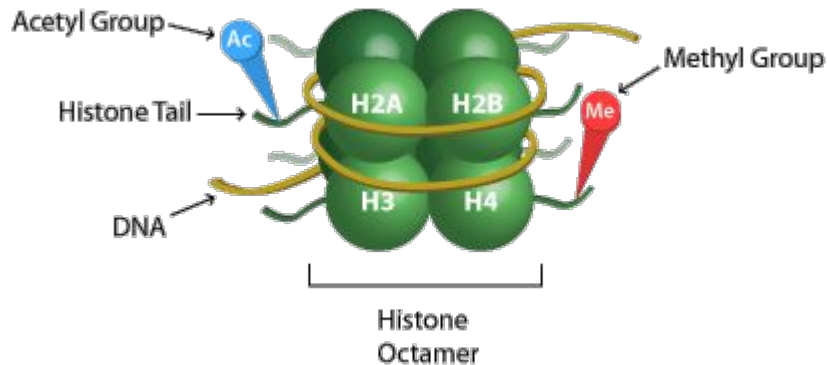
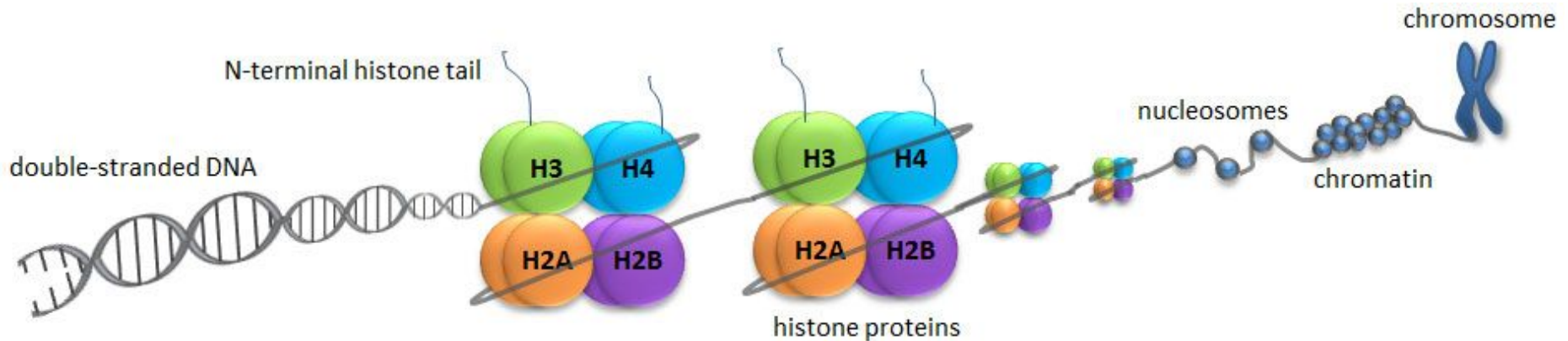


Chromatin organization



The chromatin signature of pluripotent cells, StemBook, NCBI
<https://www.ncbi.nlm.nih.gov/books/NBK27041/figure/thechromatinsignature.F1/>

Histone modifications

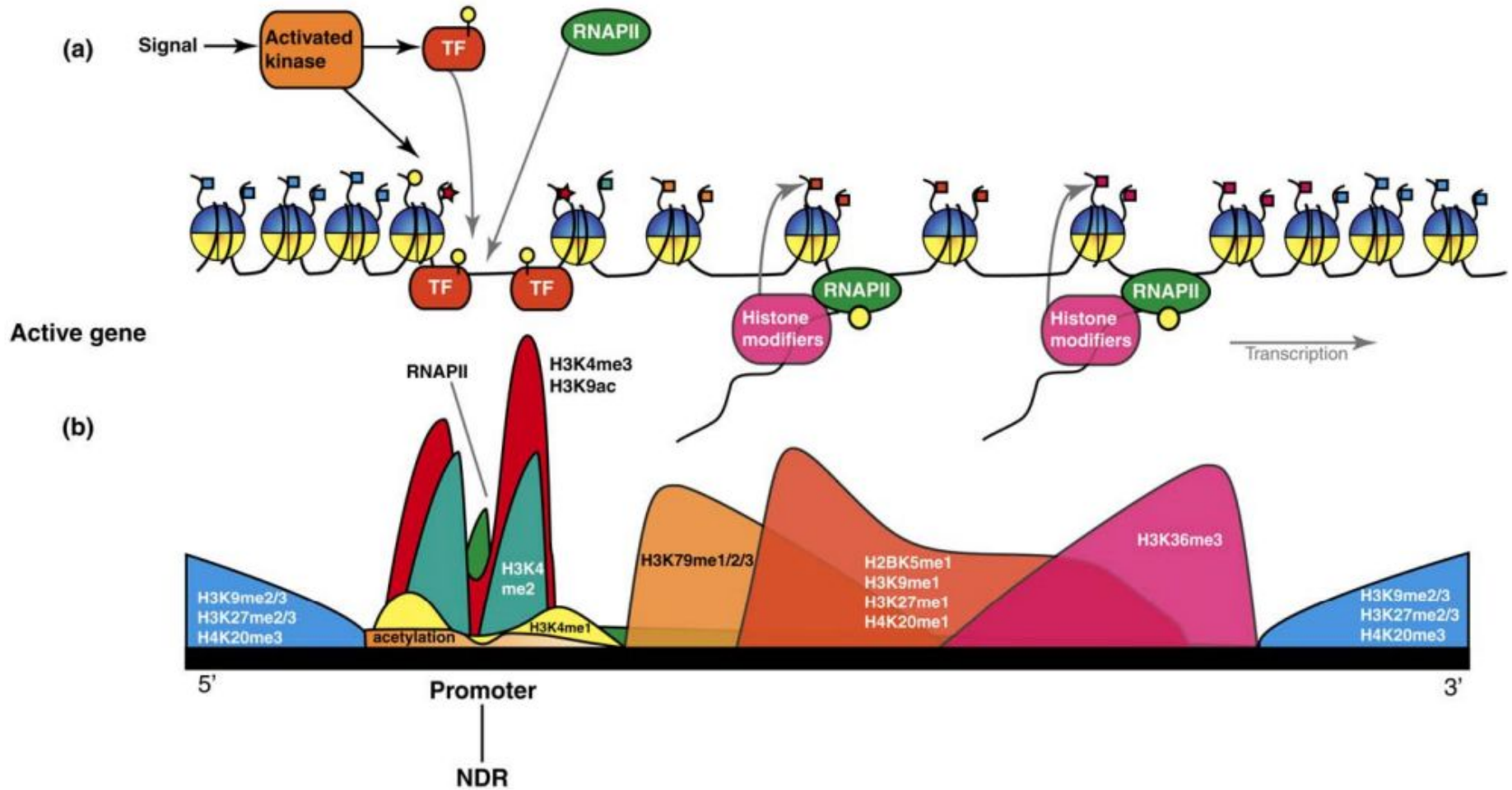


- methylation, acetylation and phosphorylation
- involve covalent post-translational modifications mostly to the residues at the positively charged N-terminal tails of histones.

<https://www.whatisepigenetics.com/histone-modifications/>

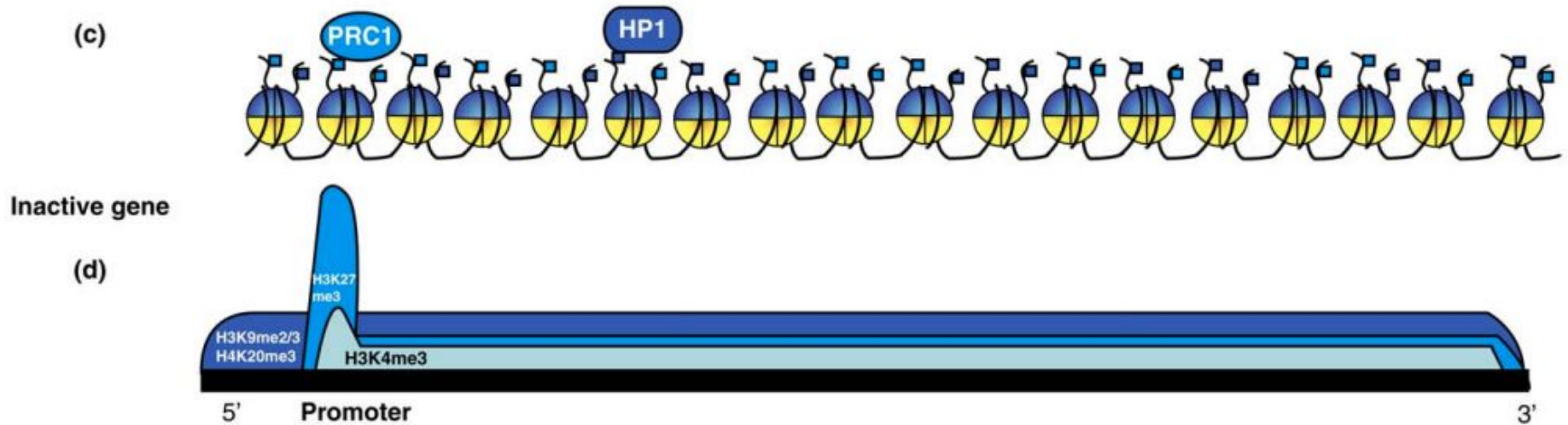
<https://www.epigentek.com/catalog/advanced-epigenetic-overview-of-histone-modifications-n-5.html?currency=es>

Histone modifications



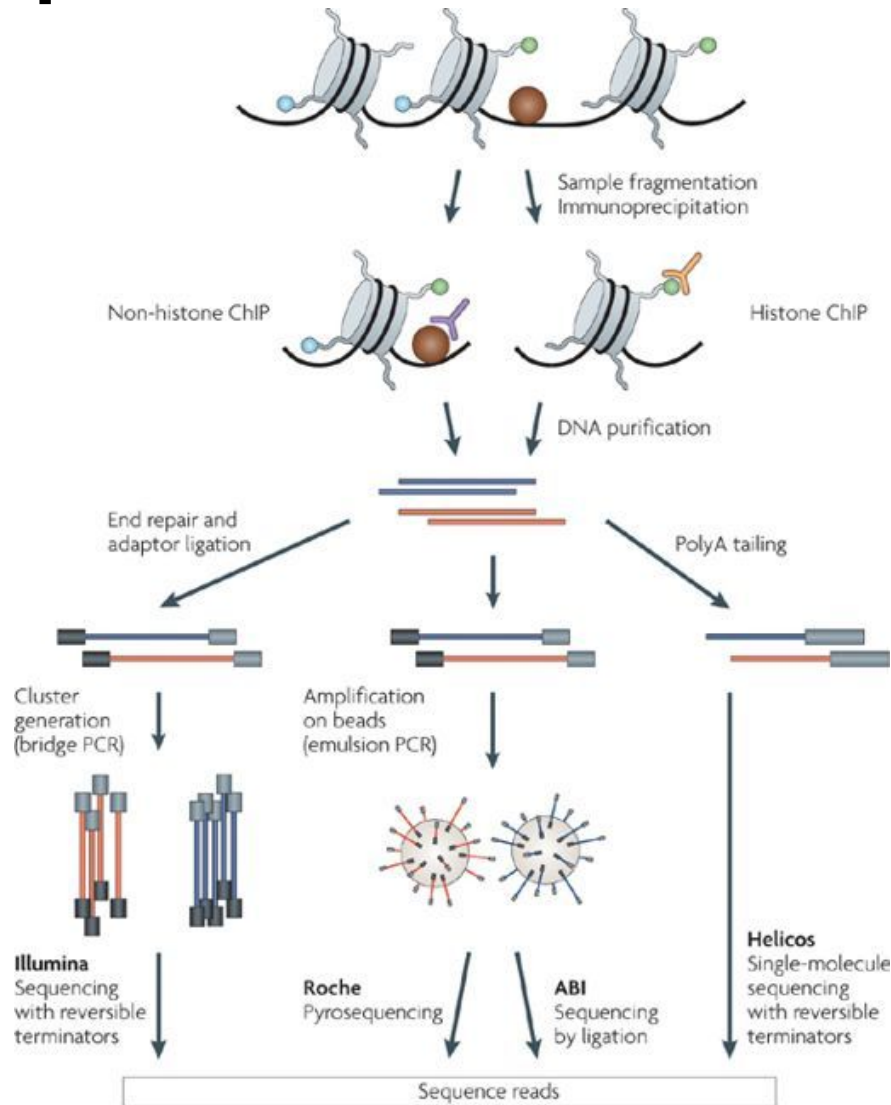
Barth & Imhof (2010) Trends in Biochemical Sciences

Histone modifications



Barth & Imhof (2010) Trends in Biochemical Sciences

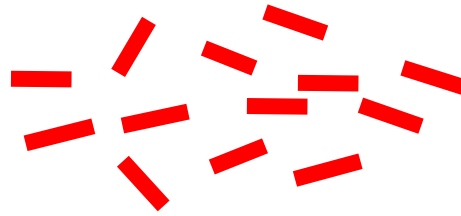
ChIP-seq: Chromatin ImmunoPrecipitation



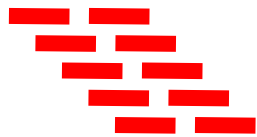
Park (2009) doi: 10.1038/nrg2641

Processing

Mapping



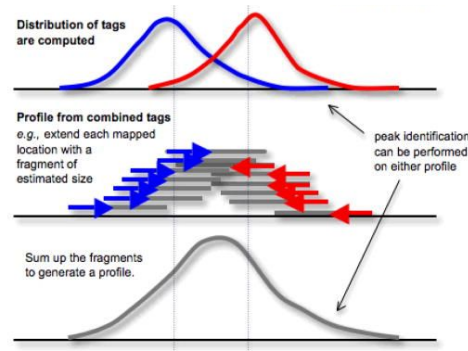
Mapping



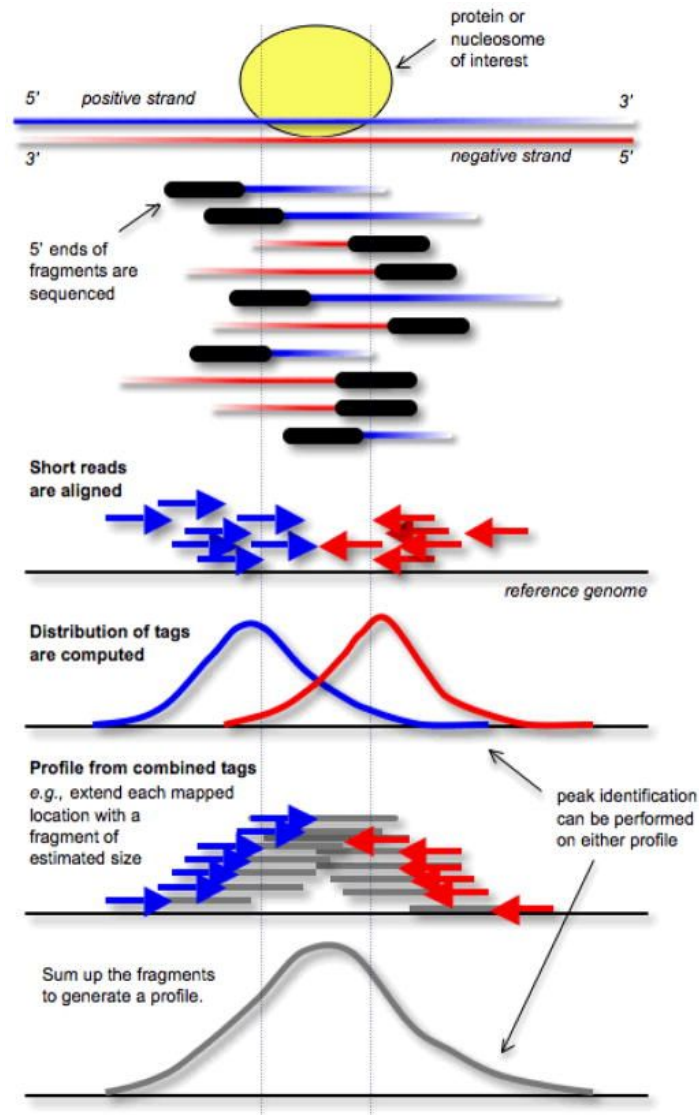
reference

Find a correspondence between the query sequences (ChIP-seq reads) and our prior knowledge (reference genome sequence).

Peak calling



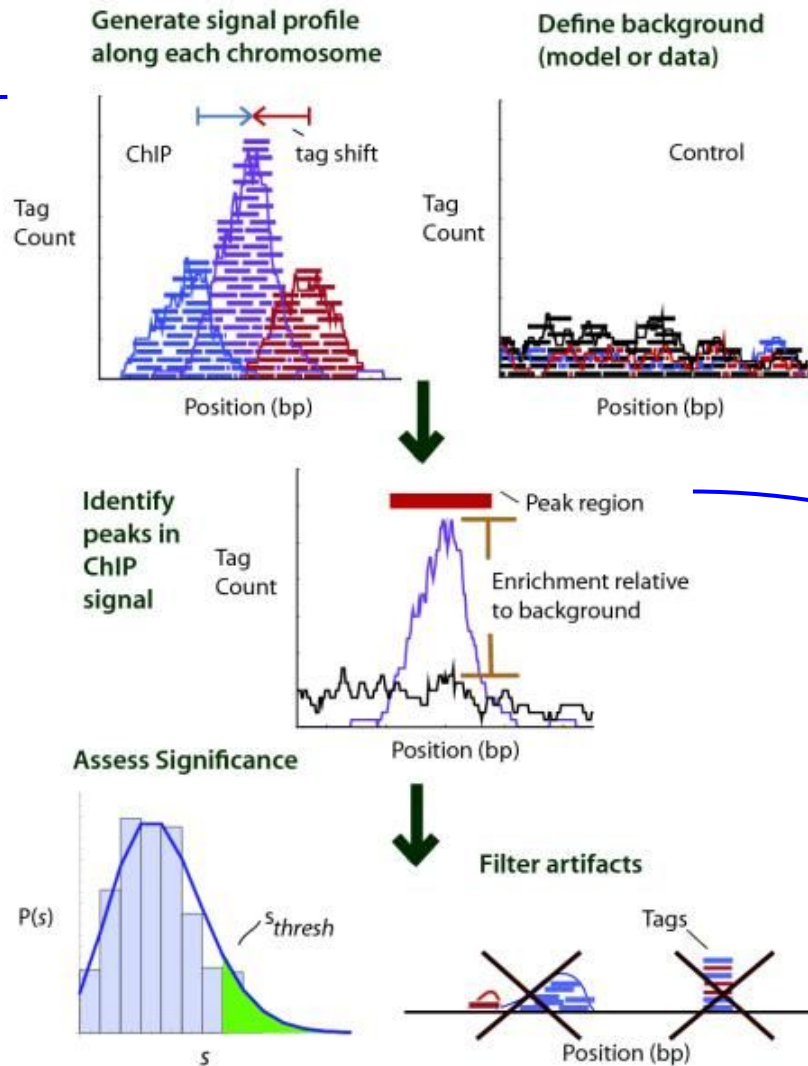
Peak calling



Park (2009) doi: 10.1038/nrg2641

Peak calling: MACS2

Signal profile
(bigWig)

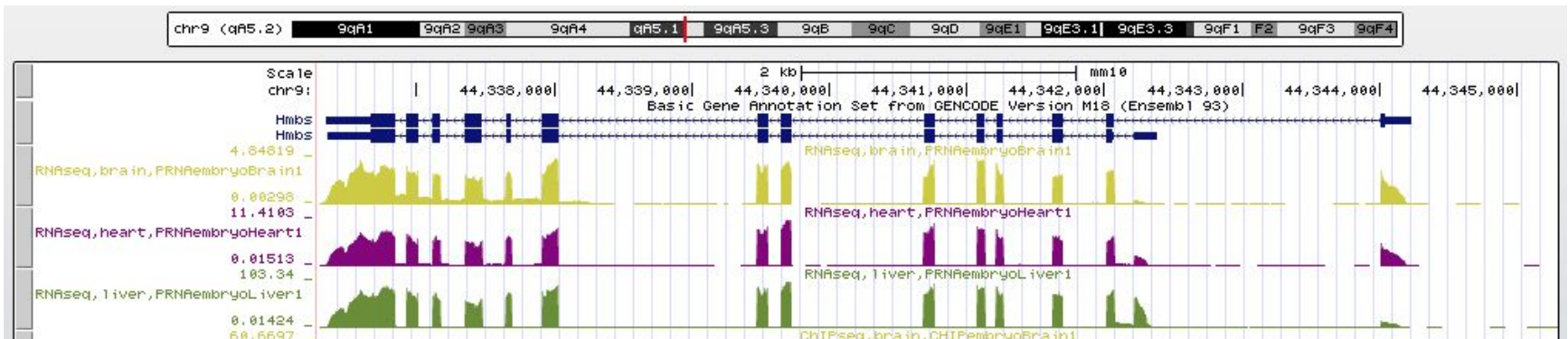


Peak regions
(BED)

Visualization

RNA-seq signal

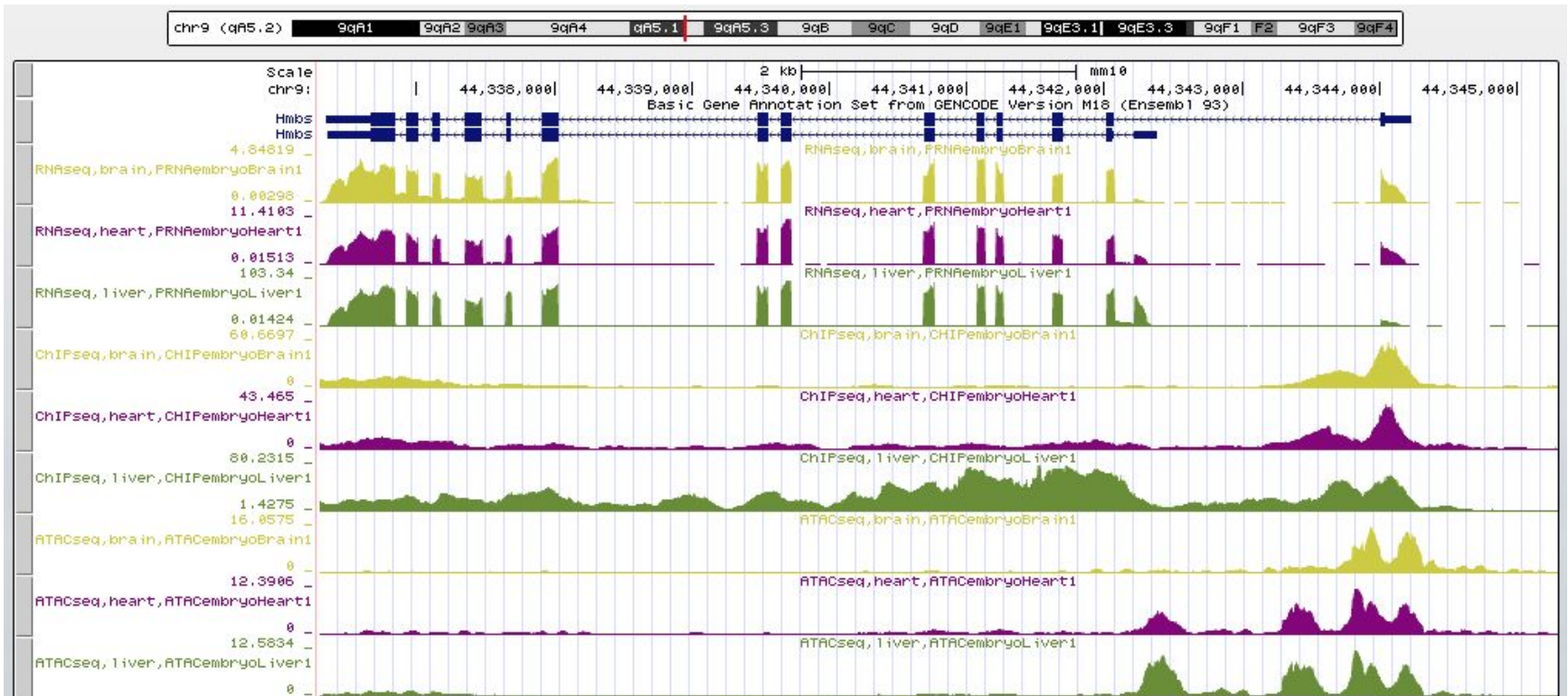
genome-euro.ucsc.edu



- expected read depth at each position in the genome
- can be normalized (e.g. RPM, reads per million reads)

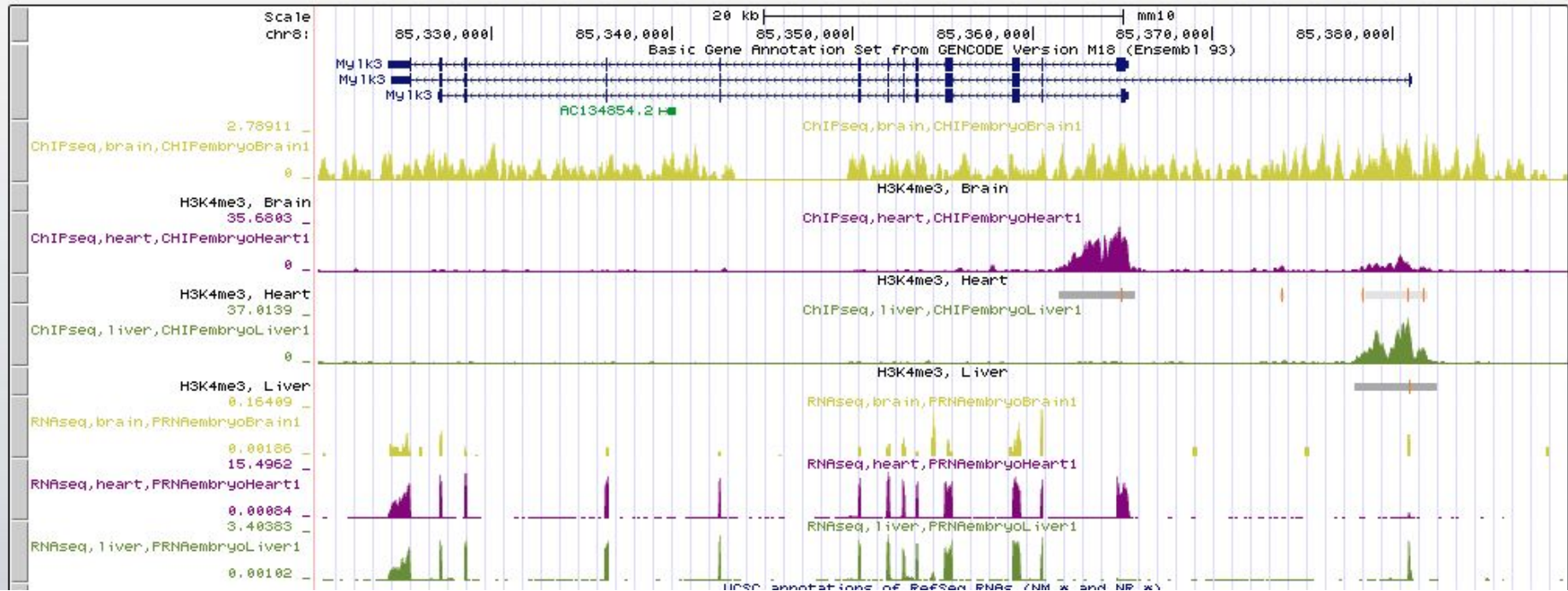
UCSC: signal files

genome-euro.ucsc.edu



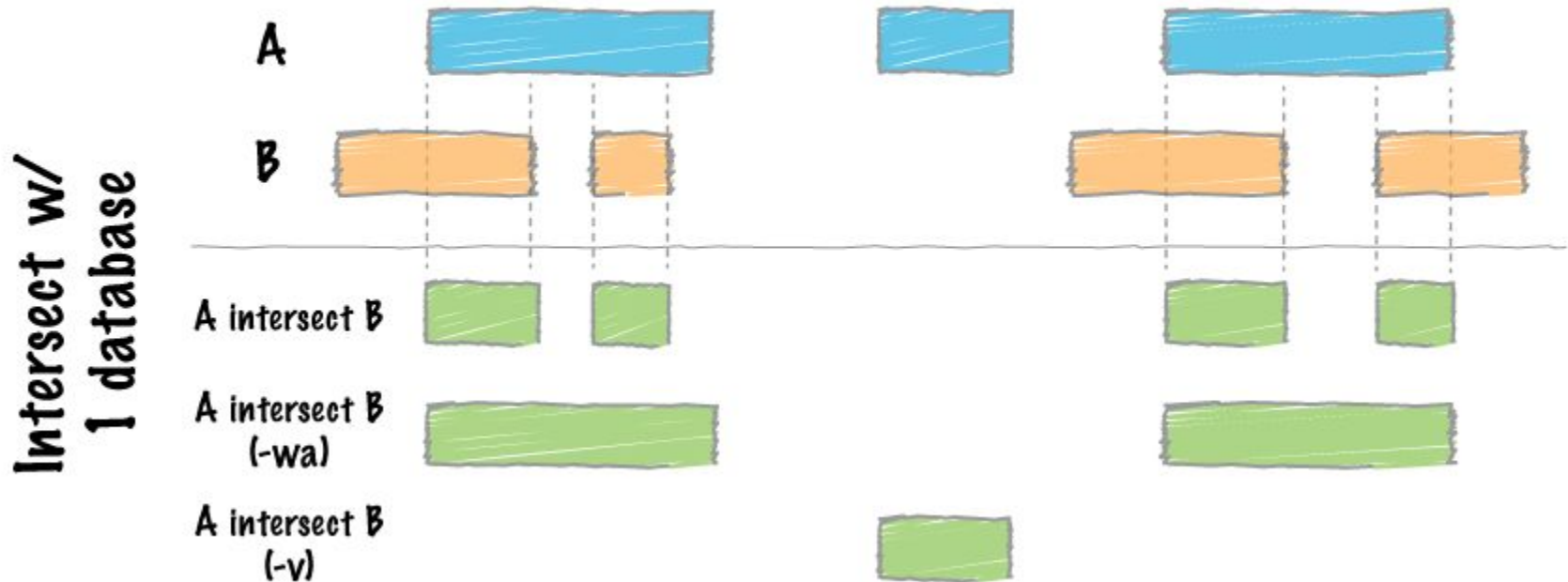
UCSC

chr8 (qC3) 8qA1.1 8qA2 8qA3 8qA4 B1.1 qB1.3qB2 8qB3.1 8qB3.3 8qC1 8qC2 8qC3 8qC5 8qD1 8qD3 8qE1 8qE2



Analysis

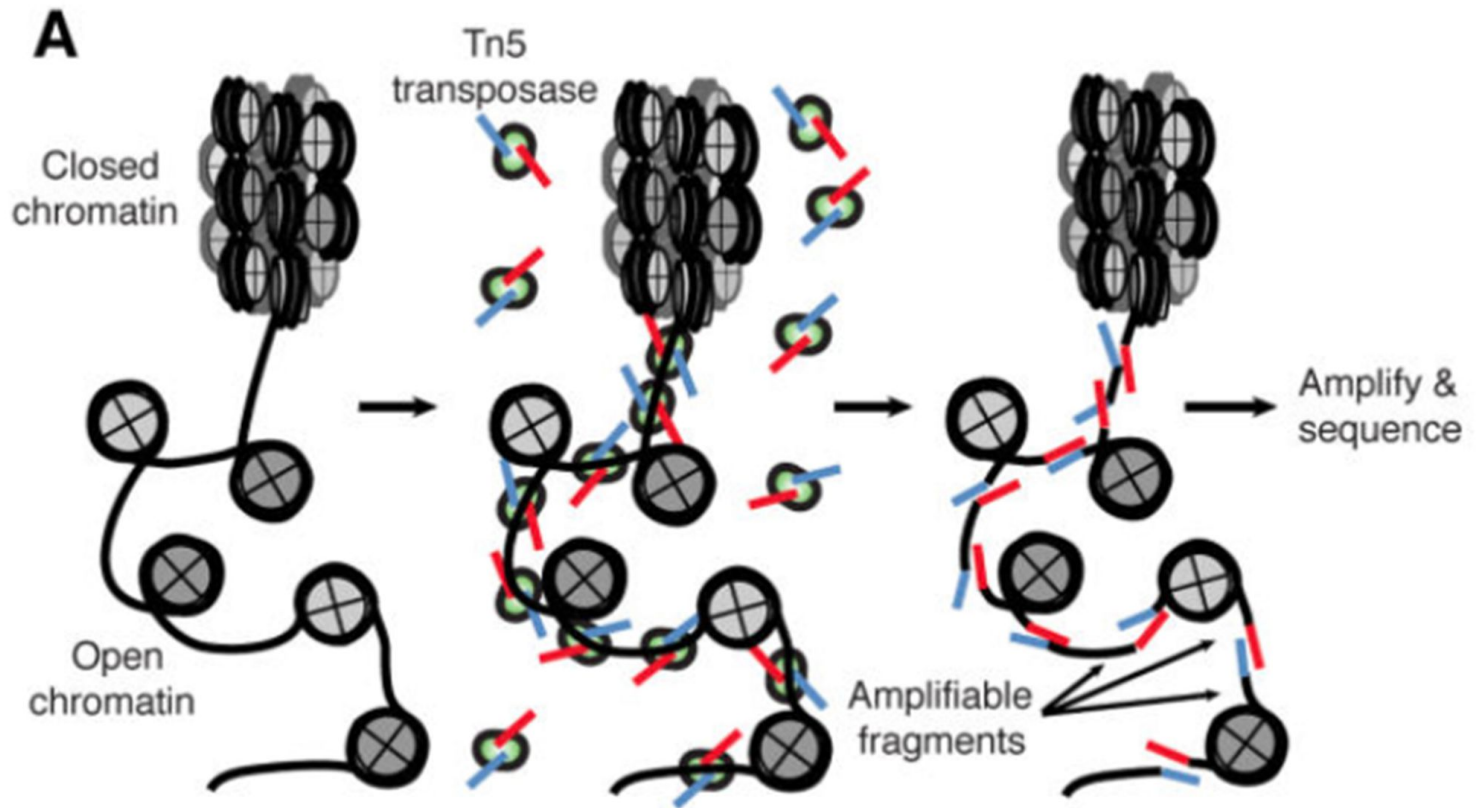
BEDTools intersect



<https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html>

ATAC-seq

Open chromatin



Buenrostro *et al.* (2015) doi:10.1002/0471142727.mb2129s109

Hands-on

Regulation of gene and isoform expression 6

https://public-docs.crg.es/rguigo/Data/cklein/courses/UVIC/handsOn/#_regulation_of_gene_and_isoform_expression