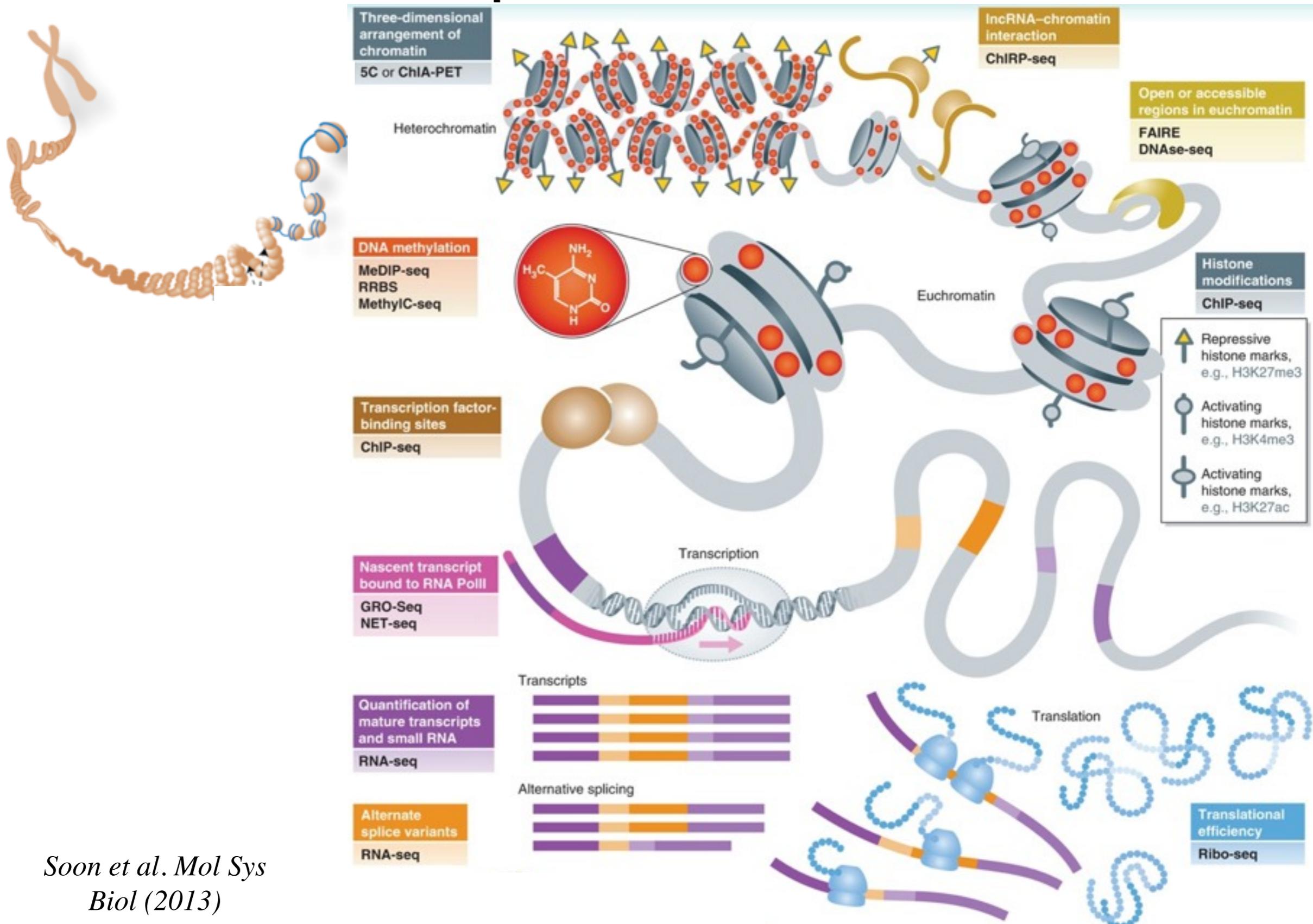
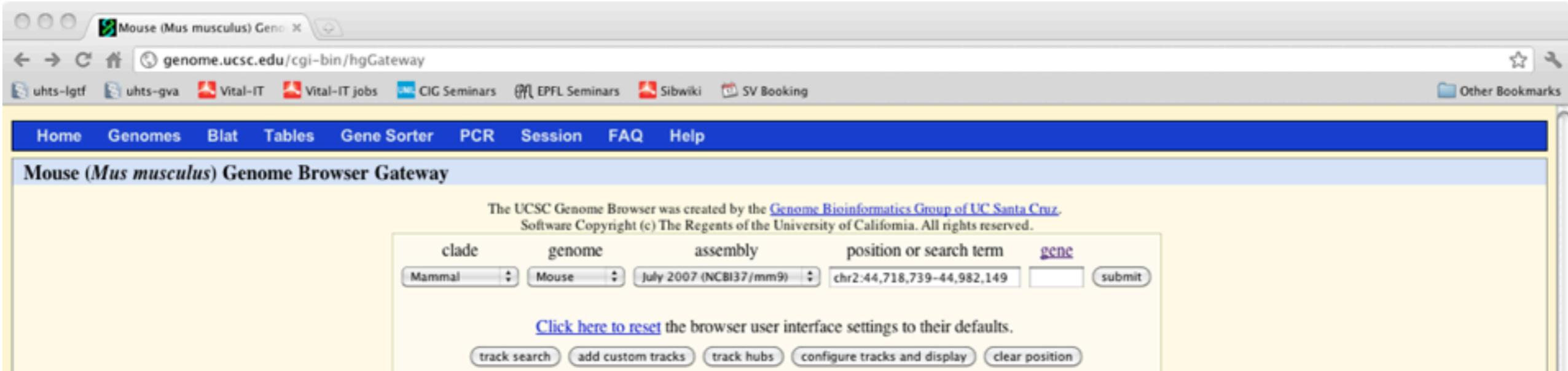


A dynamic picture of the genome from multiple functional data



The UCSC genome browser and database



Mouse (*Mus musculus*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade genome assembly position or search term gene
Mammal Mouse July 2007 (NCBI37/mm9) chr2:44,718,739-44,982,149 gene
[submit](#)

[Click here to reset](#) the browser user interface settings to their defaults.
[track search](#) [add custom tracks](#) [track hubs](#) [configure tracks and display](#) [clear position](#)

About the Mouse July 2007 (NCBI37/mm9) assembly ([sequences](#))

The July 2007 mouse (*Mus musculus*) genome data were obtained from the Build 37 assembly by [NCBI](#) and the [Mouse Genome Sequencing Consortium](#).

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic region, an mRNA or EST, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the Mouse genome. See the [User's Guide](#) for more information.

Request: Genome Browser Response:

chr16	Displays all of chromosome 16
chr16:1-5000000	Displays first 5 million bases of chr 16
chr16:1000000+2000	Displays a region of chr 16 that spans 2000 bases, starting with position 1000000
AC072048.4	Displays region of contig AC072048.4
D16Mit120	Displays region around STS marker DMit16120 from the MGI consensus genetic map, including 100,000 bases on each side as well
D16Mit120:D16Mit149	Displays region between STS markers D16Mit120 and D16Mit149
BC012683	Displays alignment location of mRNA with GenBank accession BC012683
AW045217	Displays alignment location of EST with GenBank accession AW045217
Ncam2	Displays region of genome with official MGI mouse genetic nomenclature Ncam2
pseudogene mRNA	Lists transcribed pseudogenes but not cDNAs
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
Smith,D.	Lists mRNAs deposited by co-auth

Use this last format for author queries. Although GenBank re



Mus musculus

(Photo courtesy of [The Jackson Laboratory](#))

<http://genome.ucsc.edu>

The UCSC genome browser and database

Screenshot of the UCSC Genome Browser Gateway for Mouse (*Mus musculus*). The interface includes a header with links to Home, Genomes, Blat, Tables, Gene Sorter, PCR, Session, FAQ, and Help. Below the header is a search bar with dropdown menus for species (Mammal, Mouse), assembly (July 2007 (NCBI37/mm9)), and a position or search term input field. A circled area highlights the assembly selection dropdown and the position/search term input field.

Genome collection

About the Mouse July 2007 (NCBI37/mm9) assembly ([sequences](#))

The July 2007 mouse (*Mus musculus*) genome data were obtained from the Build 37 assembly by [NCBI](#) and the [Mouse Genome Sequencing Consortium](#).

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic region, an mRNA or EST, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the Mouse genome. See the [User's Guide](#) for more information.

Request: **Genome Browser Response:**

chr16	Displays all of chromosome 16
chr16:1-5000000	Displays first 5 million bases of chr 16
chr16:1000000+2000	Displays a region of chr 16 that spans 2000 bases, starting with position 1000000
AC072048.4	Displays region of contig AC072048.4
D16Mit120	Displays region around STS marker DMit16120 from the MGI consensus genetic map, including 100,000 bases on each side as well
D16Mit120:D16Mit149	Displays region between STS markers D16Mit120 and D16Mit149
BC012683	Displays alignment location of mRNA with GenBank accession BC012683
AW045217	Displays alignment location of EST with GenBank accession AW045217
Ncam2	Displays region of genome with official MGI mouse genetic nomenclature Ncam2
pseudogene mRNA	Lists transcribed pseudogenes but not cDNAs
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
Smith,D.	Lists mRNAs deposited by co-auth

Use this last format for author queries. Although GenBank re



Mus musculus

(Photo courtesy of [The Jackson Laboratory](#))

<http://genome.ucsc.edu>

The UCSC genome browser and database

Mouse (*Mus musculus*) Geno X genome.ucsc.edu/cgi-bin/hgGateway

Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help

Mouse (*Mus musculus*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#). Software Copyright (c) The Regents of the University of California. All rights reserved.

genome position or search term gene

Mammal Mouse July 2007 (NCBI37/mm9) chr2:44,718,739-44,982,149 submit

Click here to reset the browser user interface settings to their defaults.

track search add custom tracks track hubs configure tracks and display clear position

Genome collection

position

About the Mouse July 2007 (NCBI37/mm9) assembly ([sequences](#))

The July 2007 mouse (*Mus musculus*) genome data were obtained from the Build 37 assembly by [NCBI](#) and the [Mouse Genome Sequencing Consortium](#).

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic region, an mRNA or EST, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the Mouse genome. See the [User's Guide](#) for more information.

Request: Genome Browser Response:

chr16	Displays all of chromosome 16
chr16:1-5000000	Displays first 5 million bases of chr 16
chr16:1000000+2000	Displays a region of chr 16 that spans 2000 bases, starting with position 1000000
AC072048.4	Displays region of contig AC072048.4
D16Mit120	Displays region around STS marker DMit16120 from the MGI consensus genetic map, including 100,000 bases on each side as well
D16Mit120:D16Mit149	Displays region between STS markers D16Mit120 and D16Mit149
BC012683	Displays alignment location of mRNA with GenBank accession BC012683
AW045217	Displays alignment location of EST with GenBank accession AW045217
Ncam2	Displays region of genome with official MGI mouse genetic nomenclature Ncam2
pseudogene mRNA	Lists transcribed pseudogenes but not cDNAs
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
Smith,D.	Lists mRNAs deposited by co-auth

Use this last format for author queries. Although GenBank re

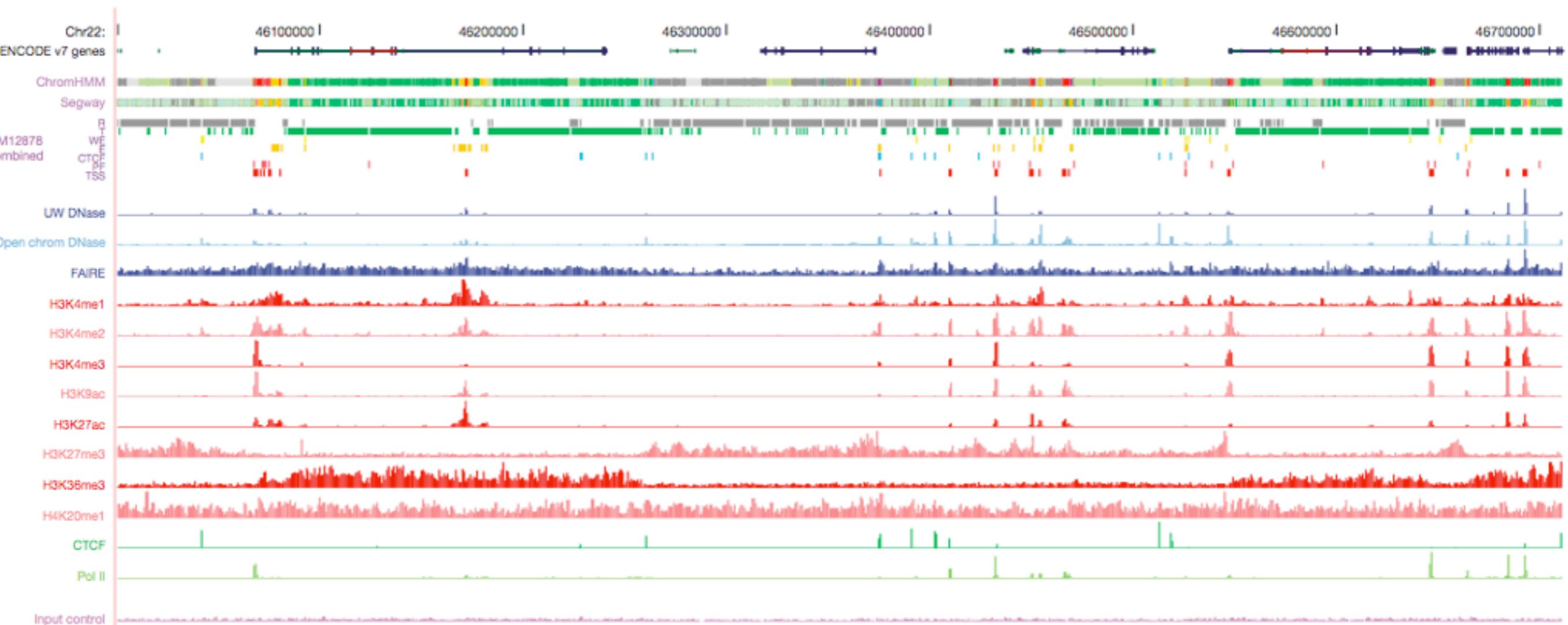


Mus musculus

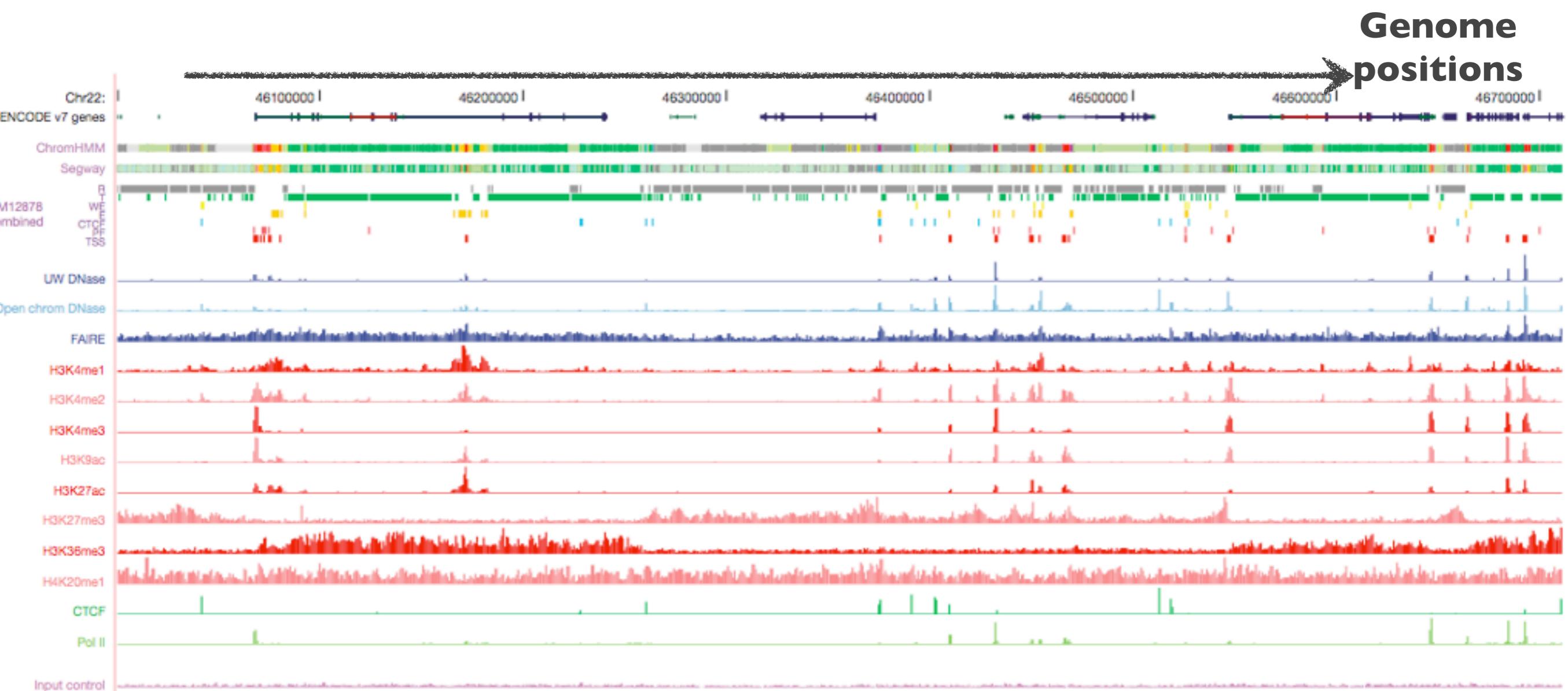
(Photo courtesy of [The Jackson Laboratory](#))

<http://genome.ucsc.edu>

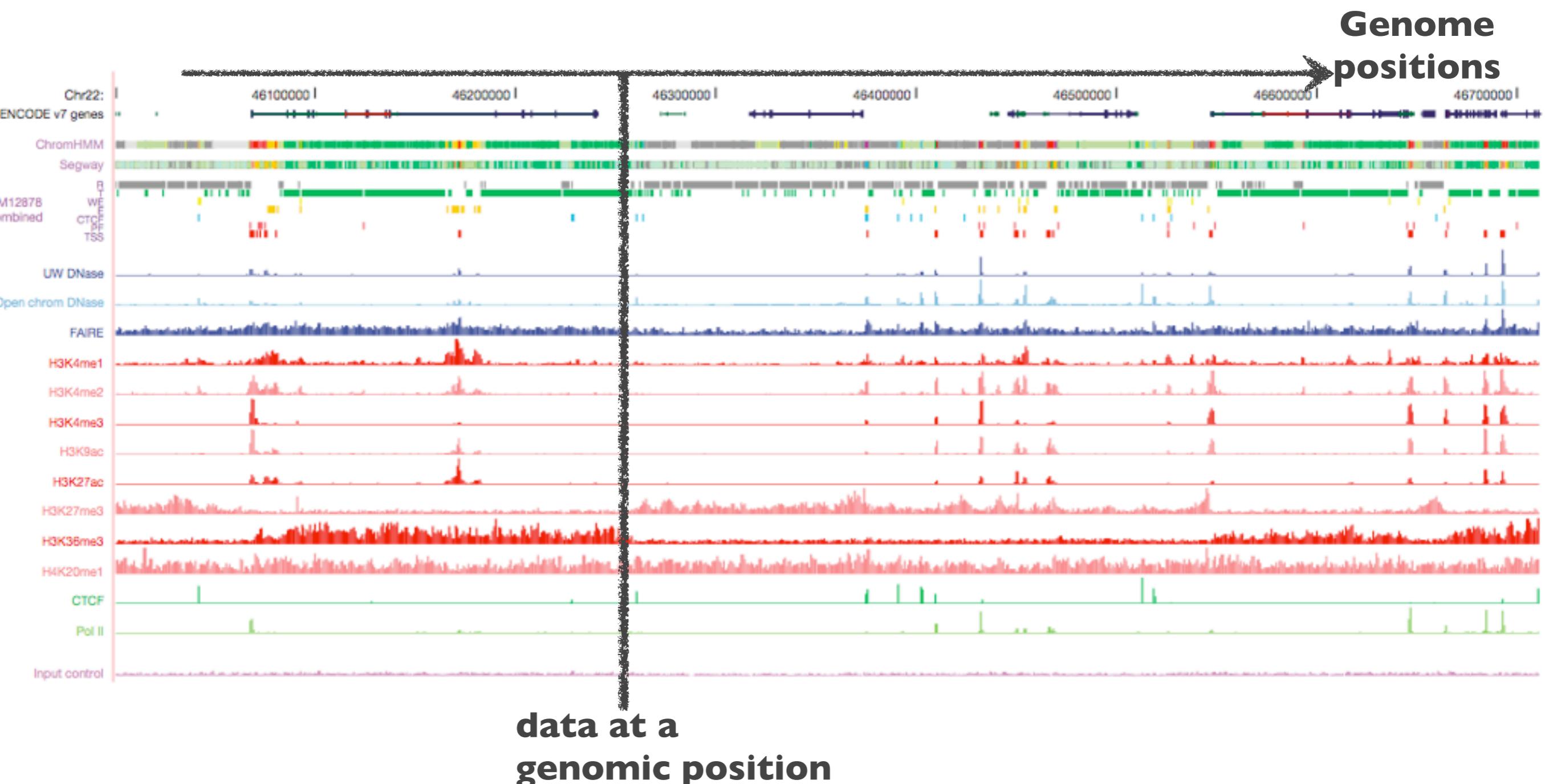
Transcriptional landscape



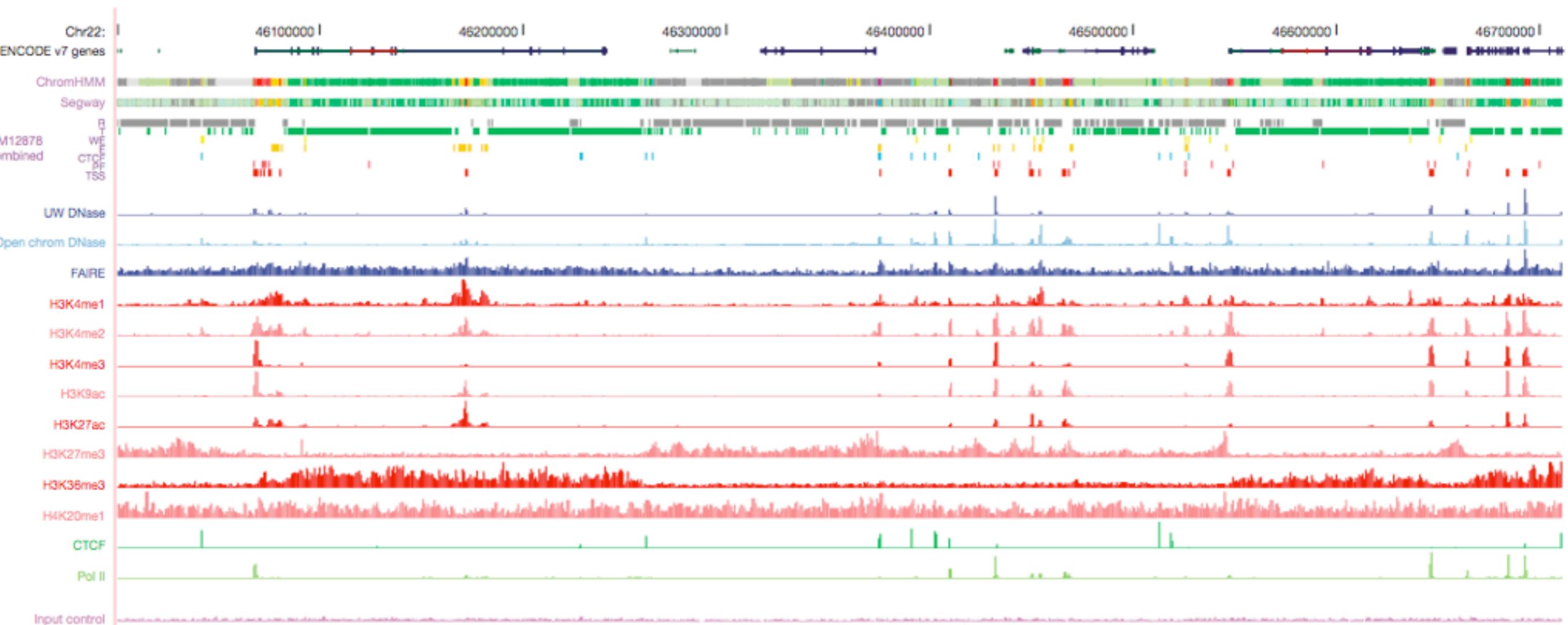
Transcriptional landscape



Transcriptional landscape



Transcriptional landscape



Objects and nomenclature

- Genome feature: segment of the genome

```
chr11 101,350,078 101,413,269
```

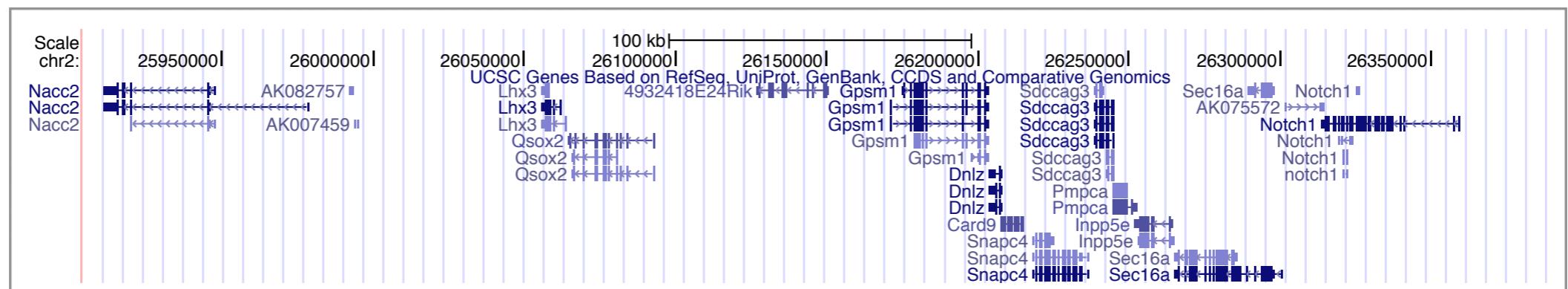
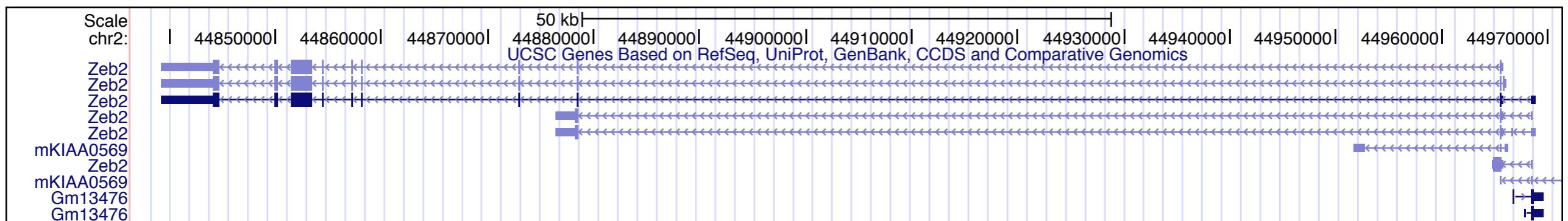
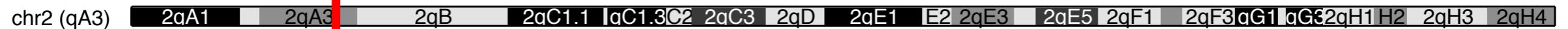
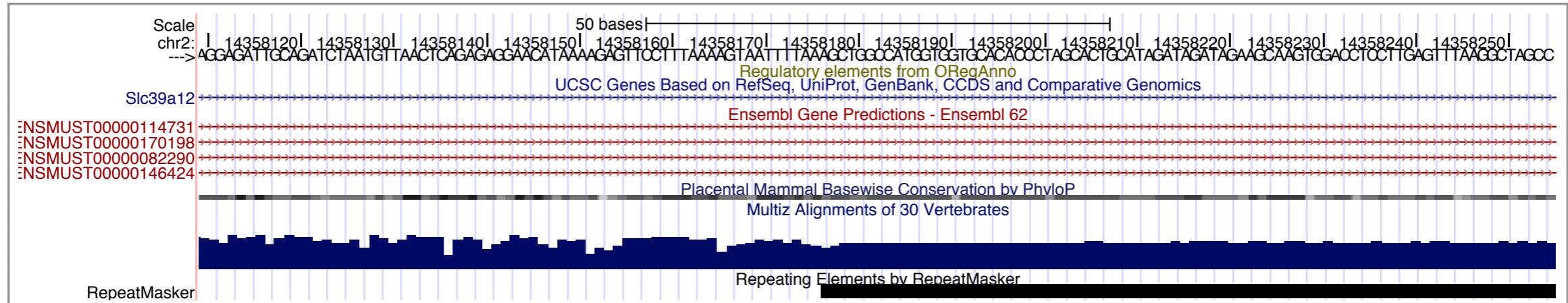
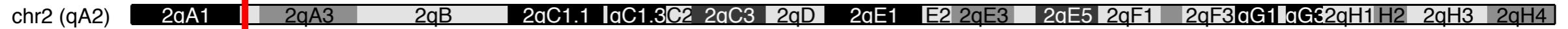
- Annotated genome feature: adding properties (name, class, value, substructure)

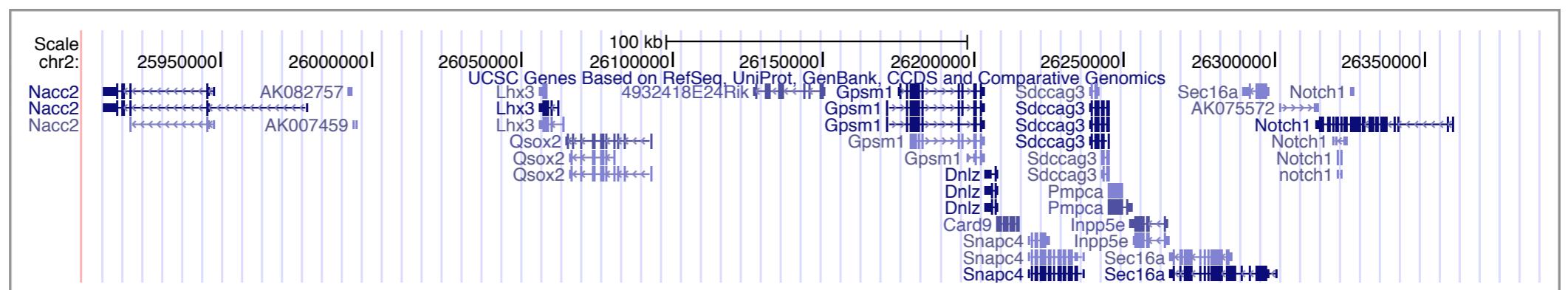
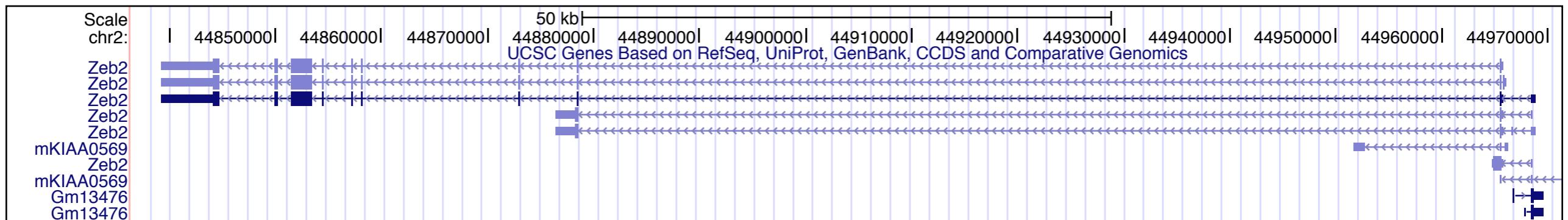
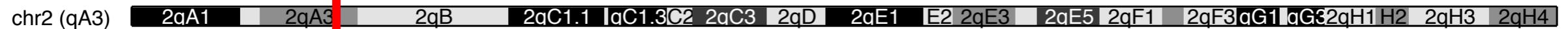
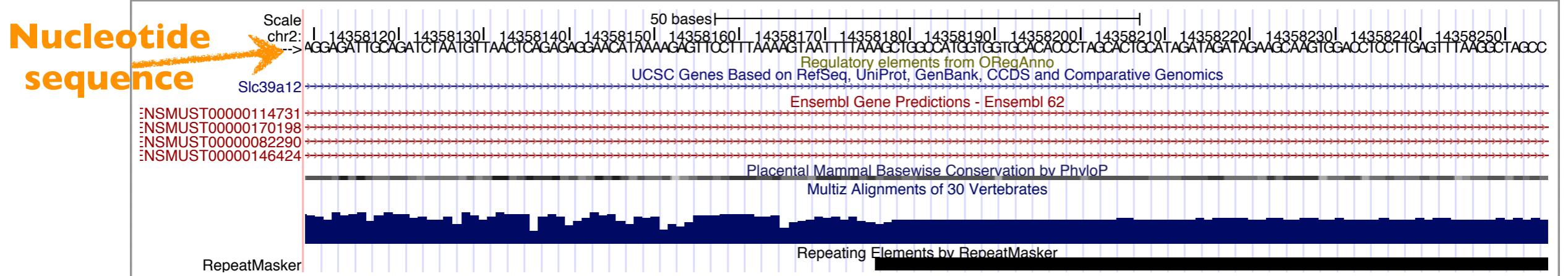
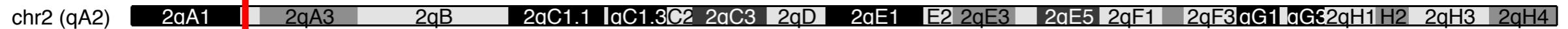
```
chr11 101,350,078 101,413,269 Brca1 protein_coding_ge3ne -
```

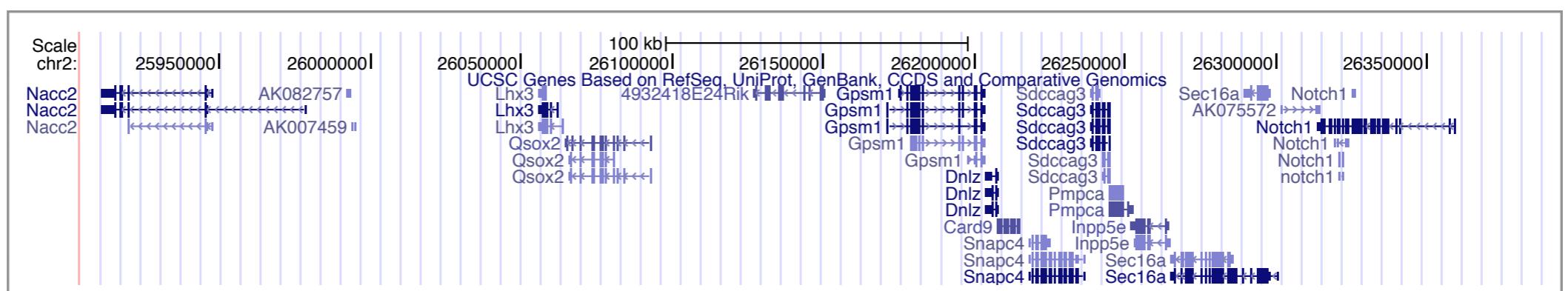
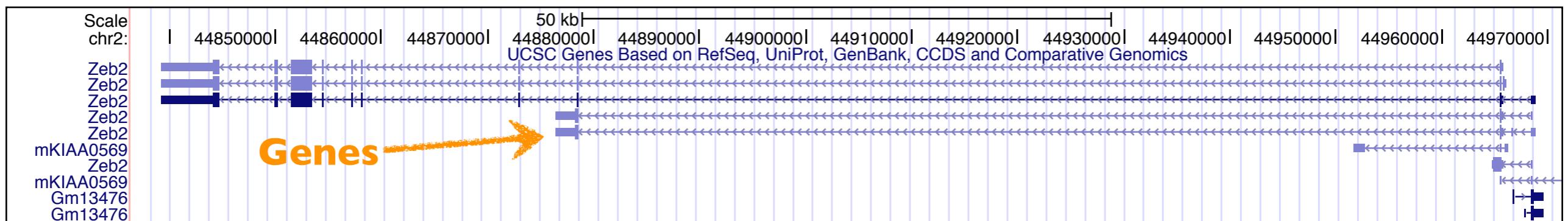
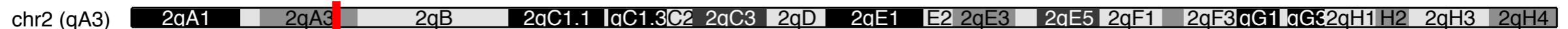
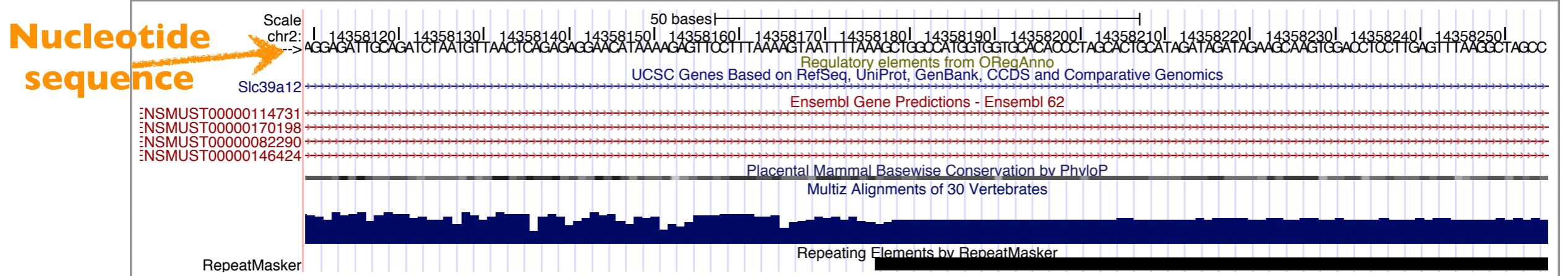
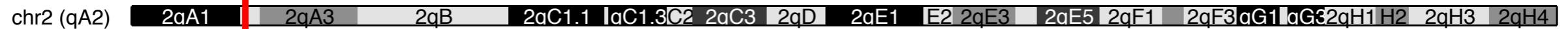
- Genome track: a set of features on a particular (version of a) genome

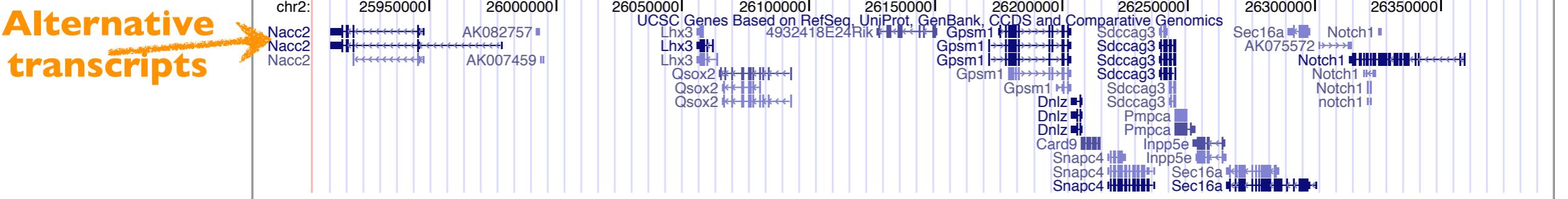
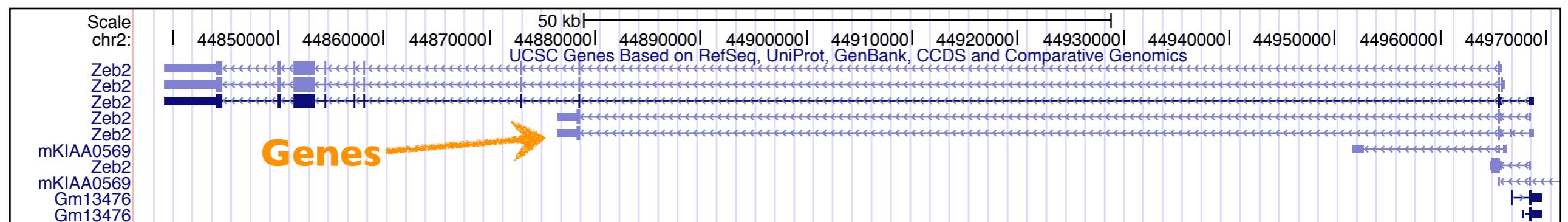
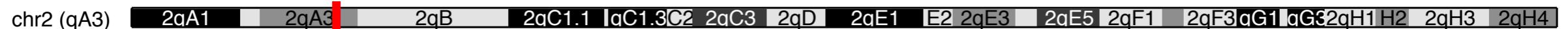
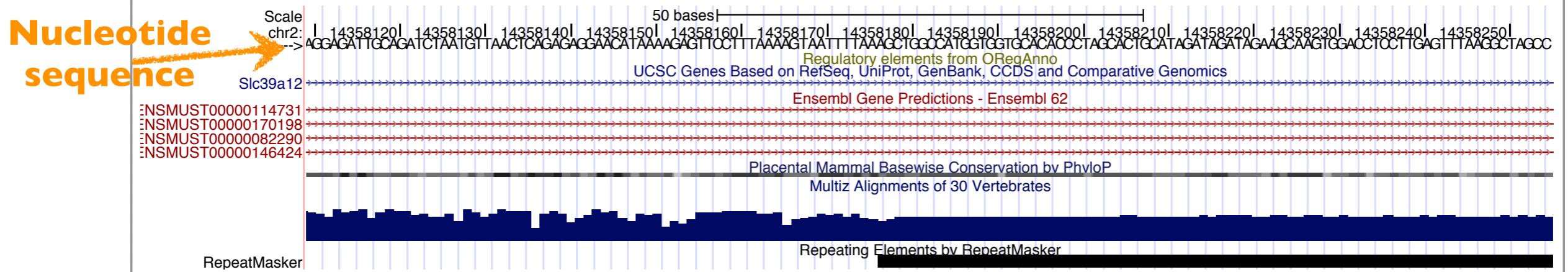
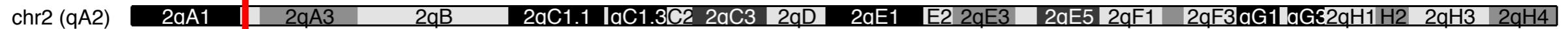
```
chr11 101350077 101413269 uc007lpd.2 10 -
chr11 101357633 101369541 uc007lpf.1 14 -
chr11 101384638 101413269 uc011yfm.1 0 -
chr11 101413468 101443259 uc007lpg.1 0 +
chr11 101414171 101443259 uc007lph.1 22 +
```

- Quantitative signal: a value associated with every nucleotide in the genome (or within a set of regions)

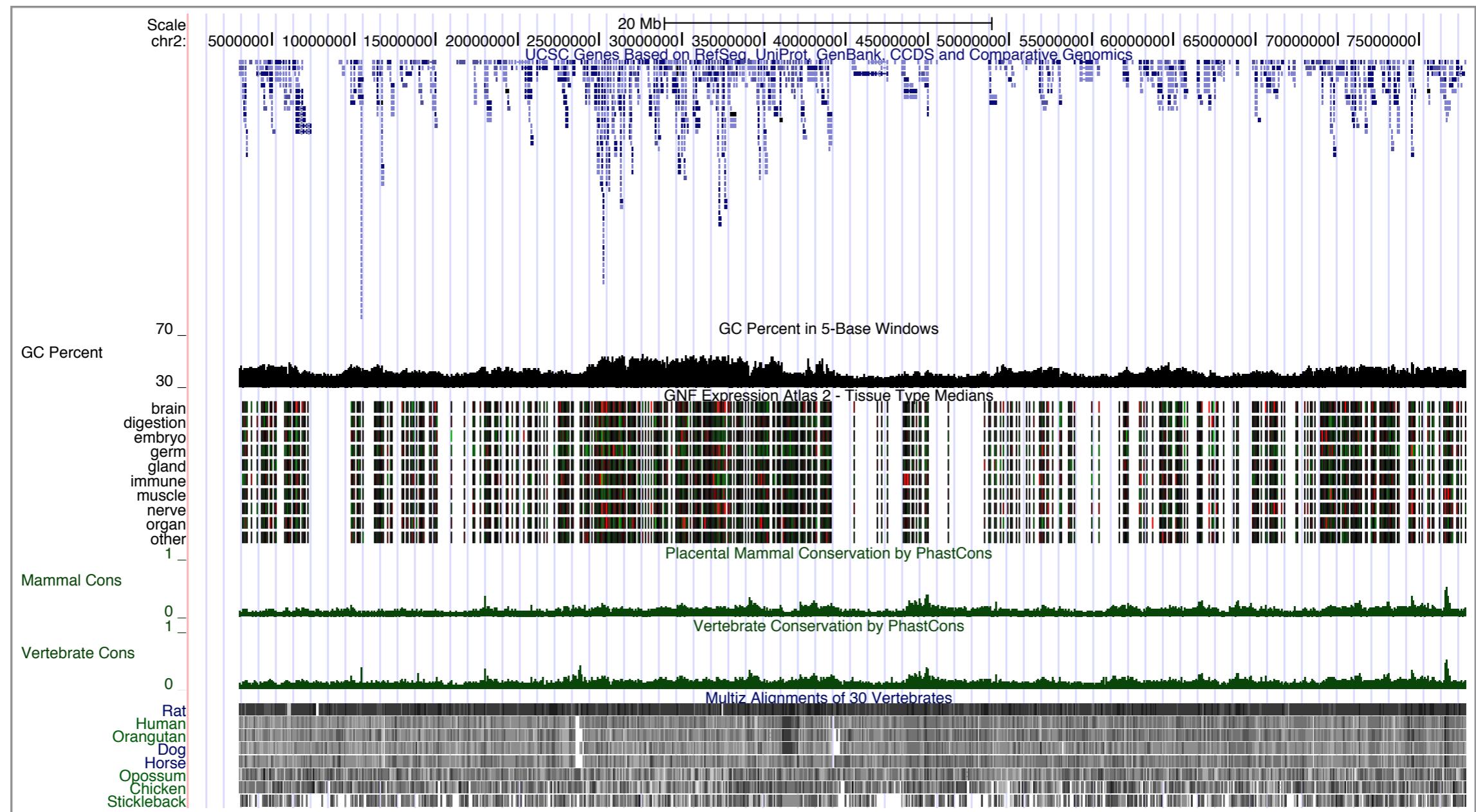




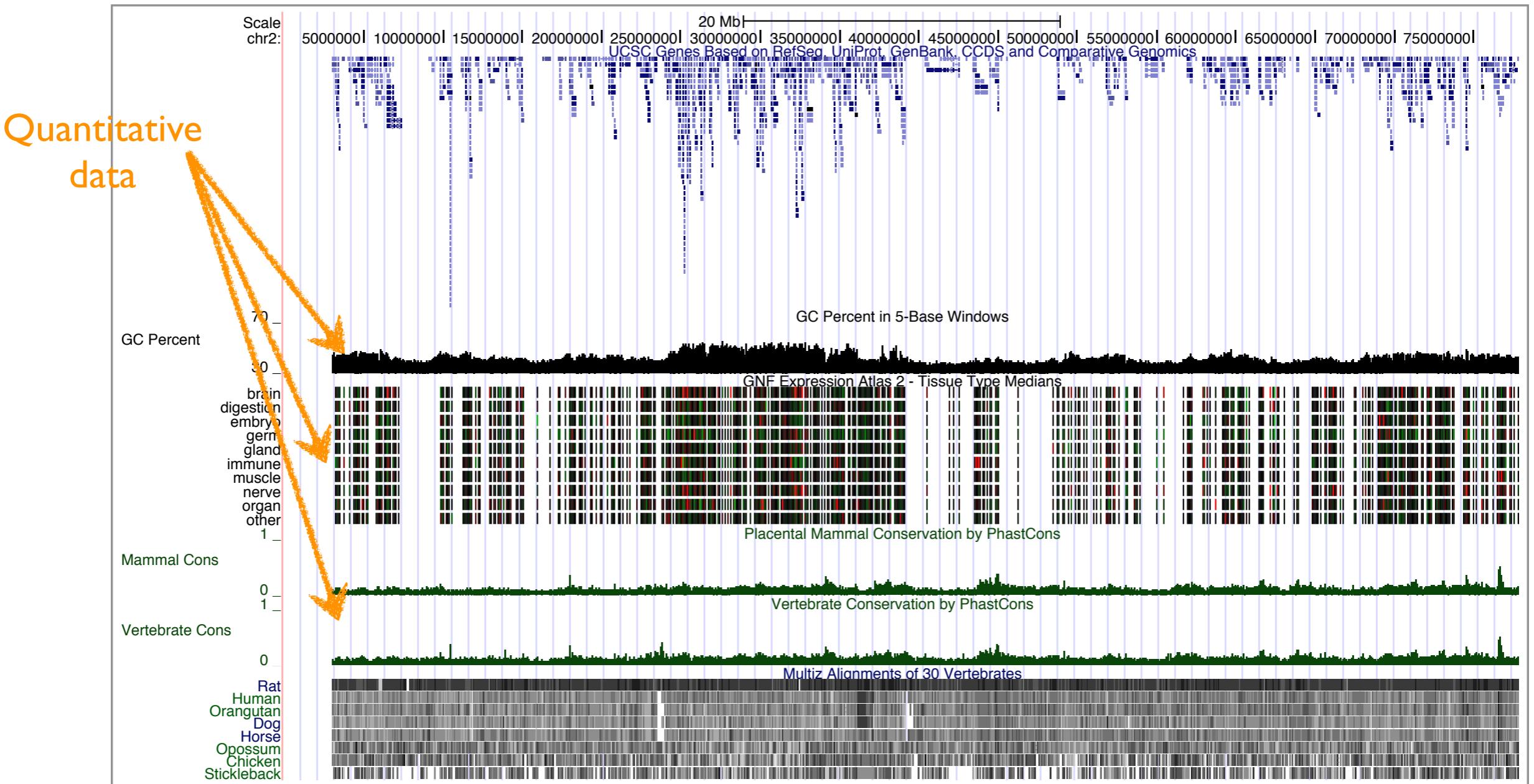


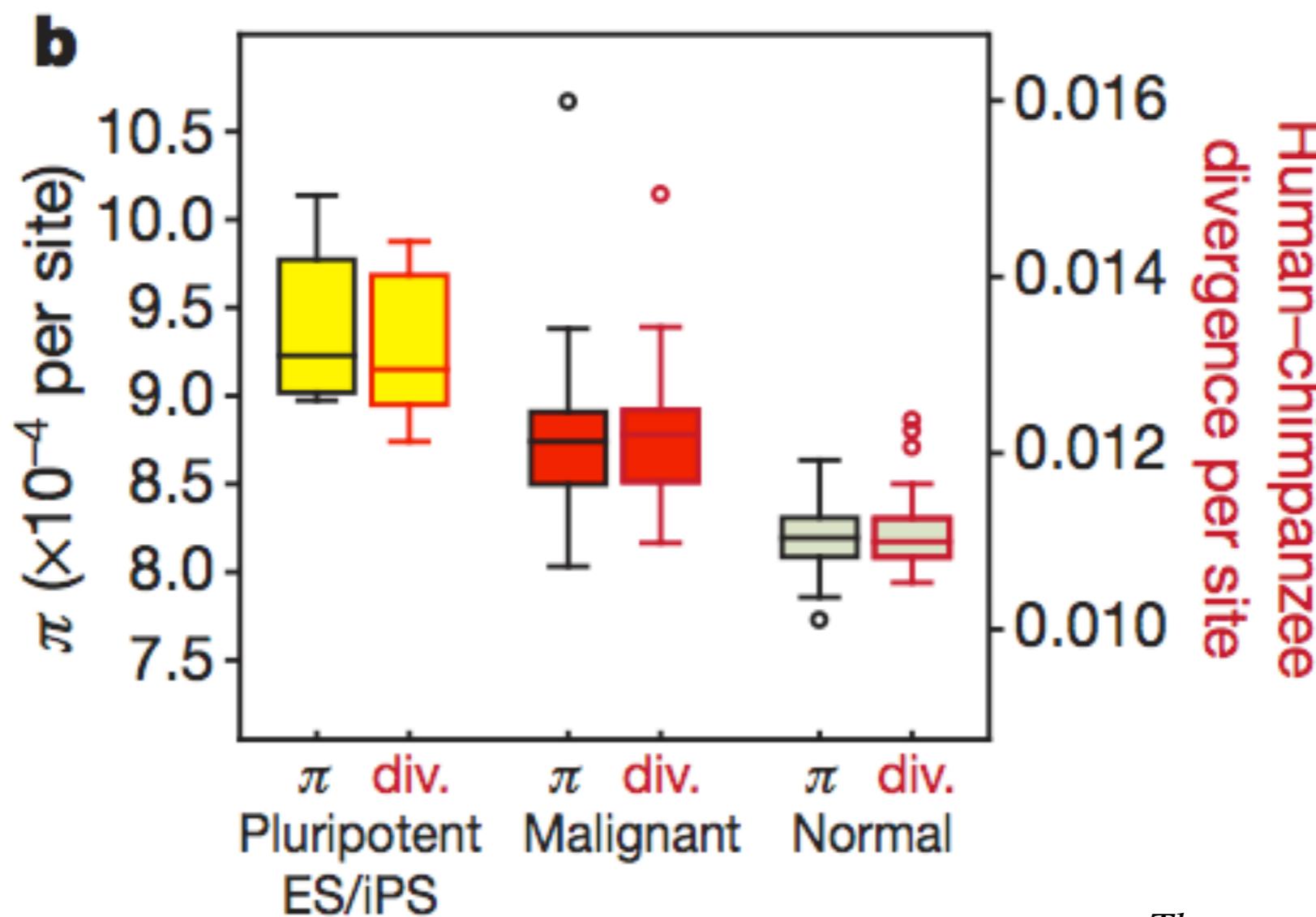


chr2 (qA1-qC3) 2qA1 2qA3 2qB 2qC1.1 C1.3 C2 2qC3 2qD 2qE1 E2 2qE3 2qE5 2qF1 2qF3 qG1 qG3 2qH1 H2 2qH3 2qH4



chr2 (qA1-qC3) 2qA1 2qA3 2qB 2qC1.1 C1.3 C2 2qC3 2qD 2qE1 E2 2qE3 2qE5 2qF1 2qF3 qG1 qG3 2qH1 H2 2qH3 2qH4



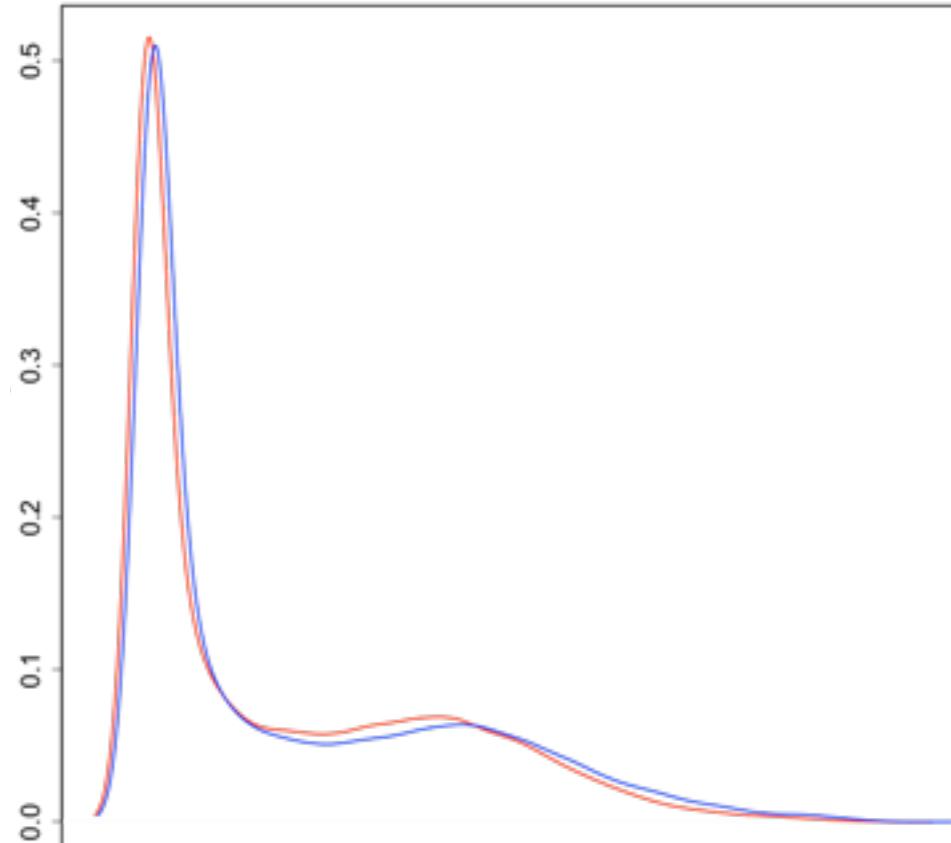
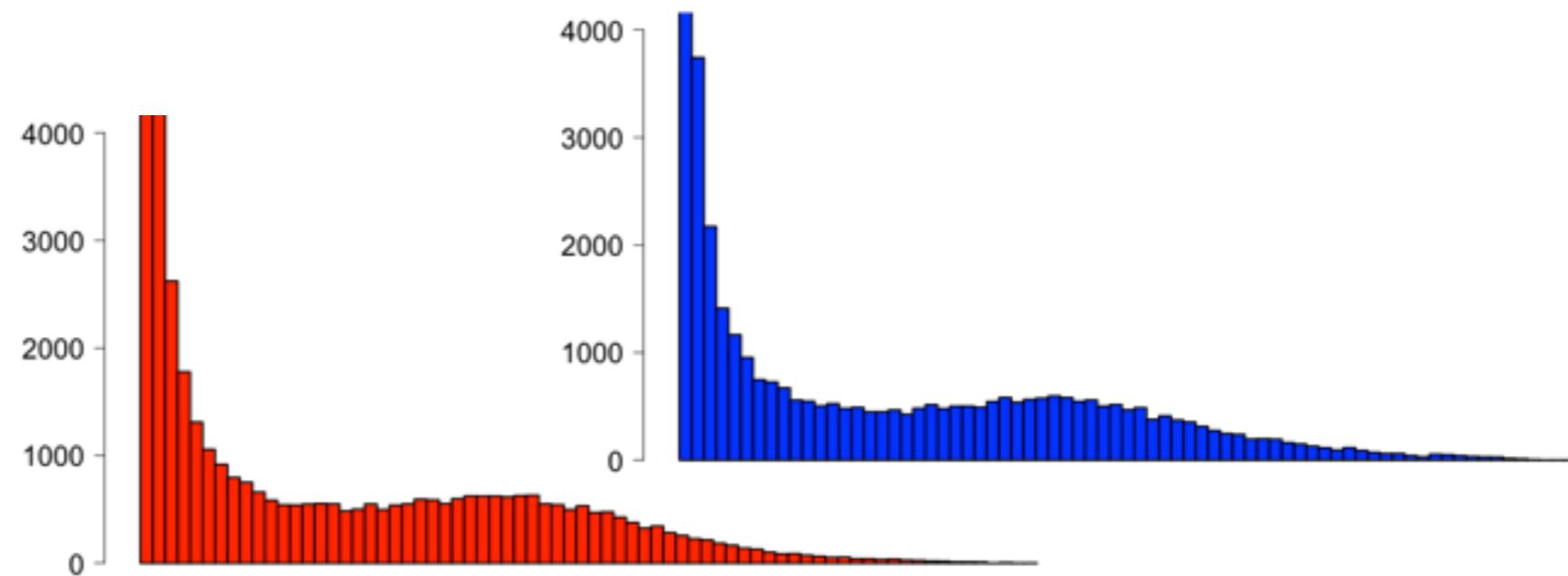


Thurman, RE et al. *Nature* (2012).

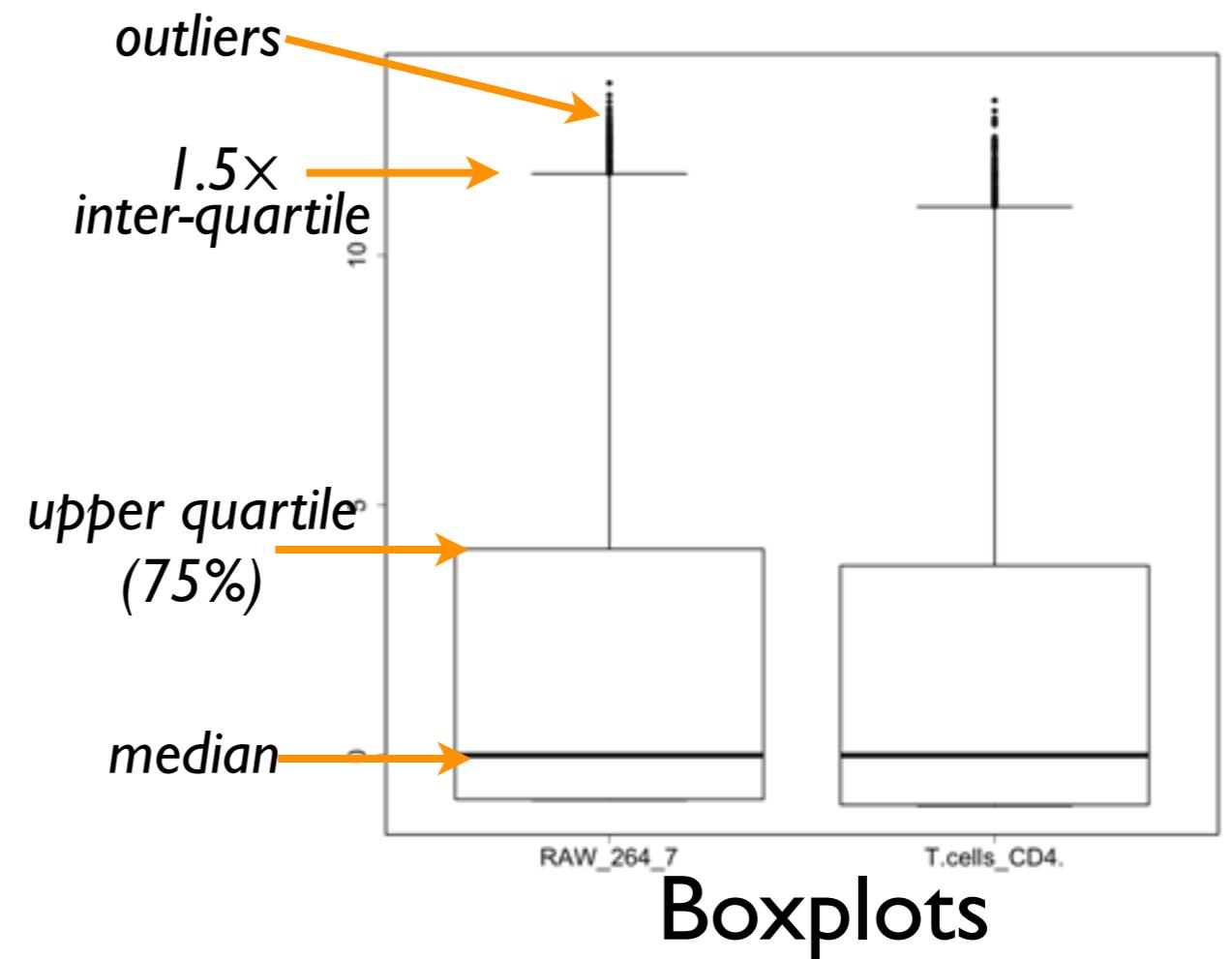
Genetic/evolutionary diversity
within active regulatory elements

Graphical representations: distributions

Histograms

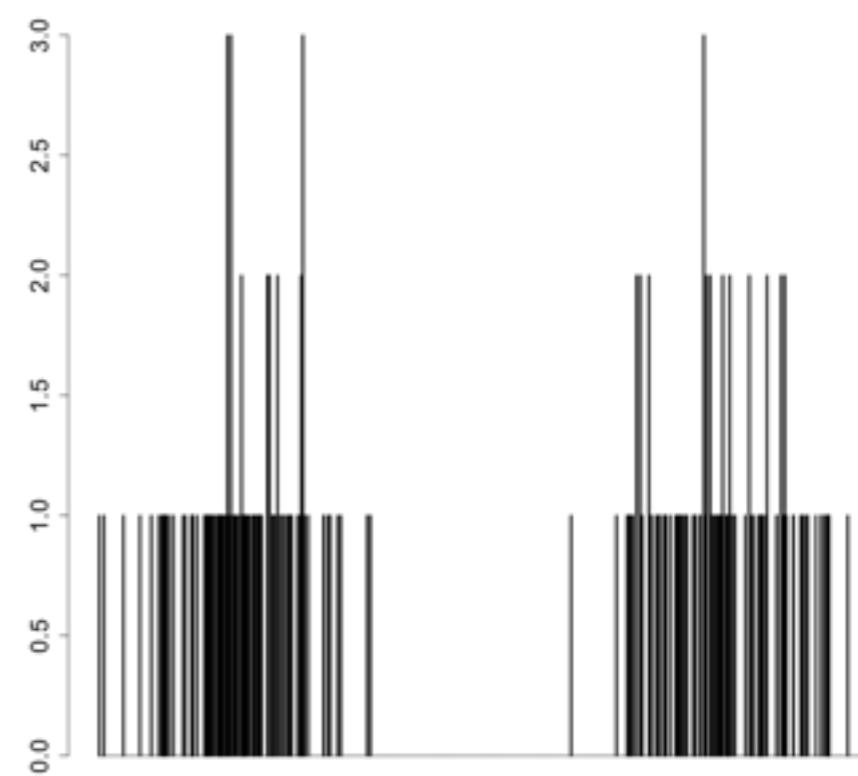
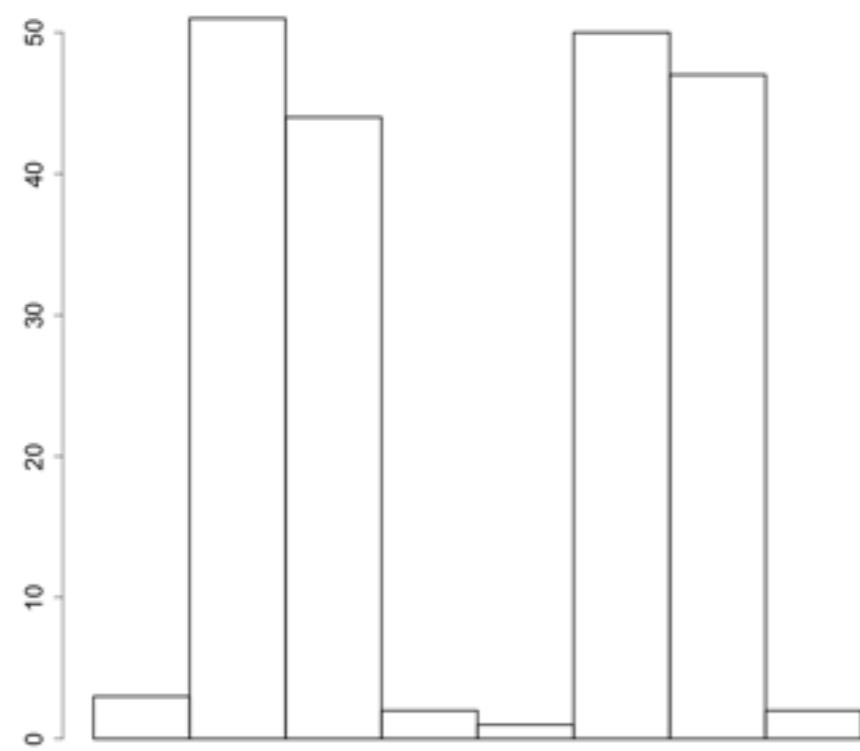
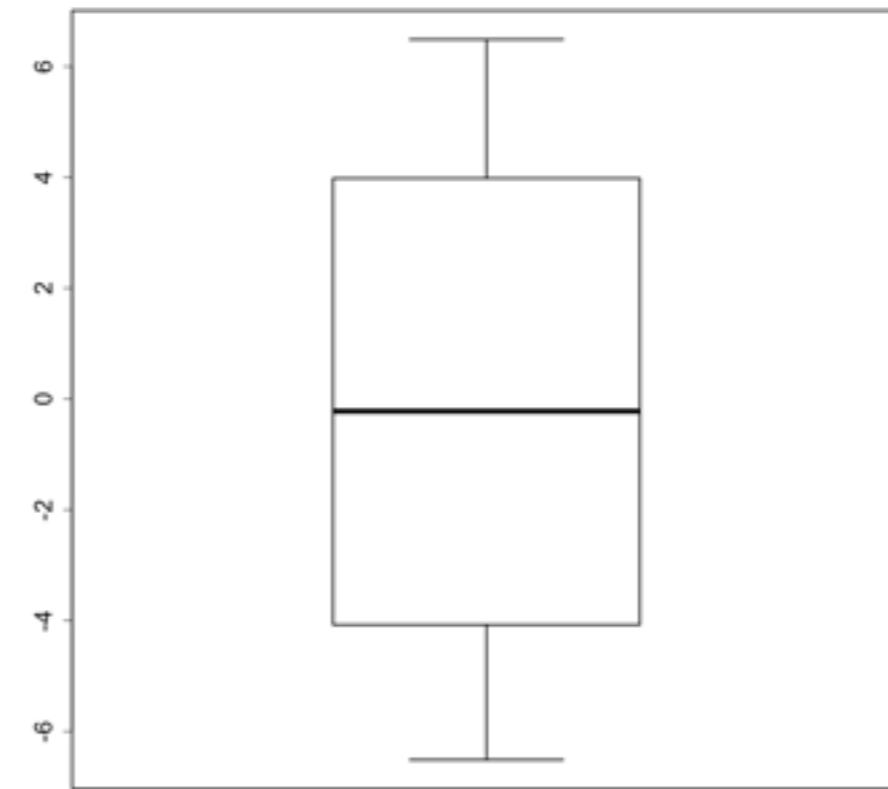
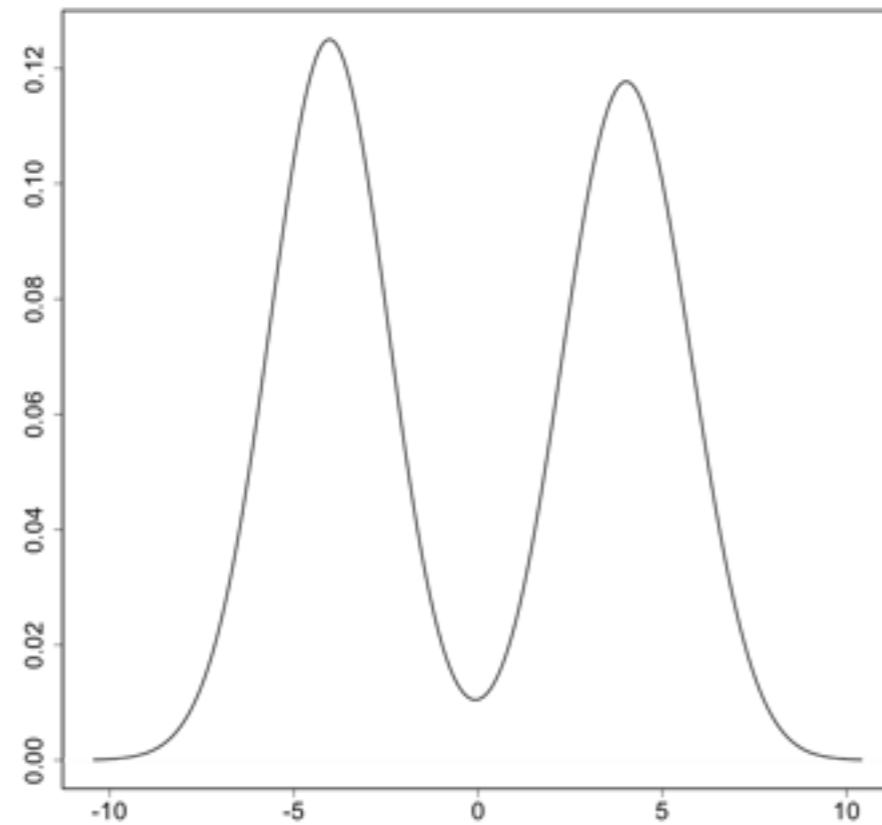


Smoothed density

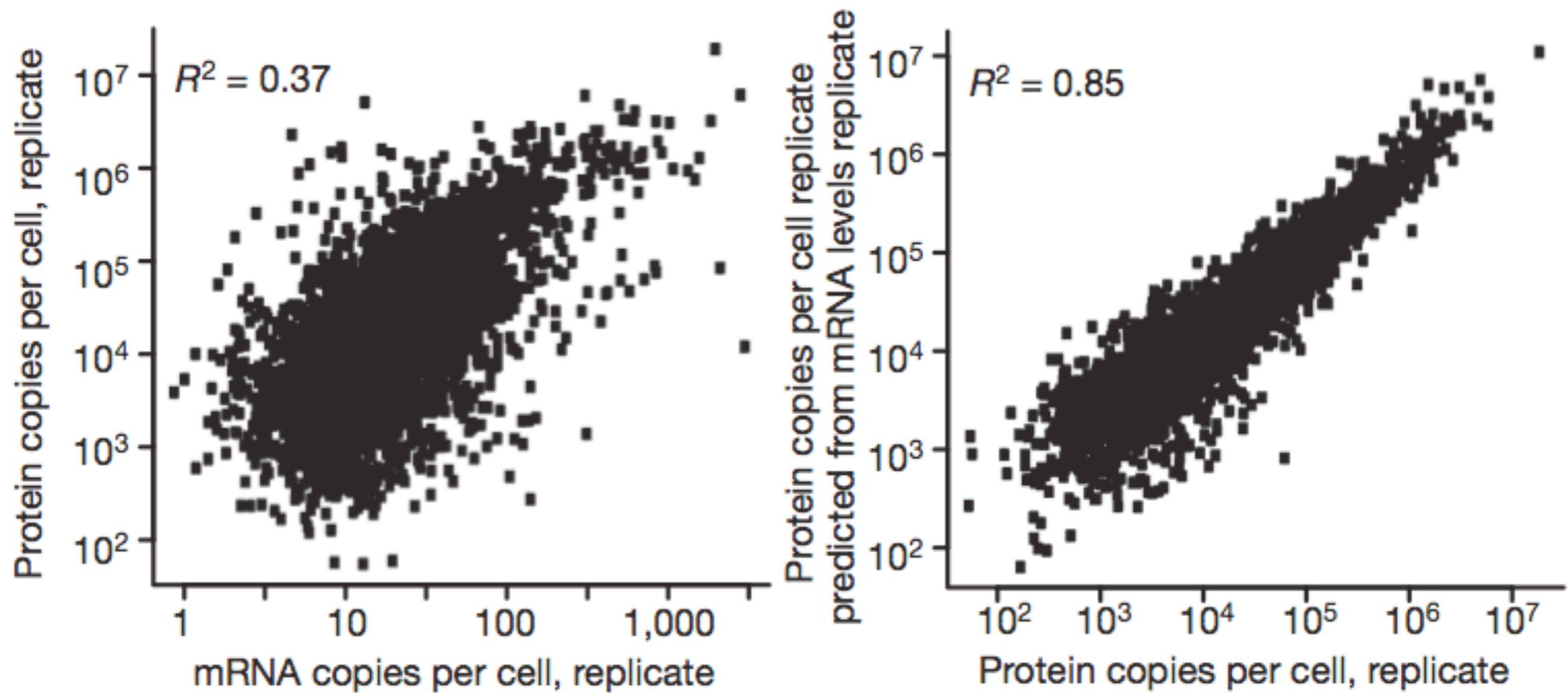


Boxplots

Graphical representations: distributions



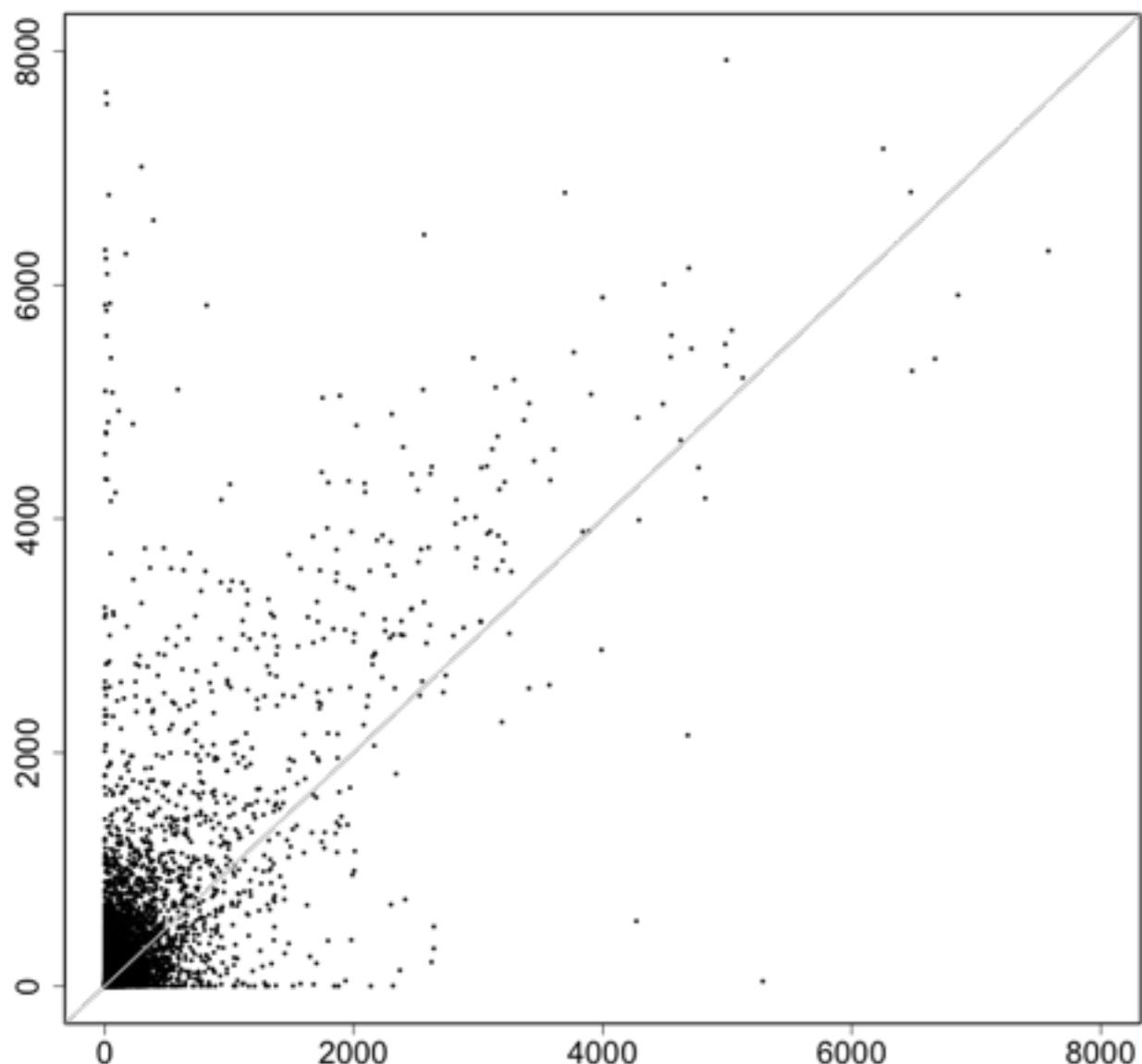
Predicting protein levels from gene expression



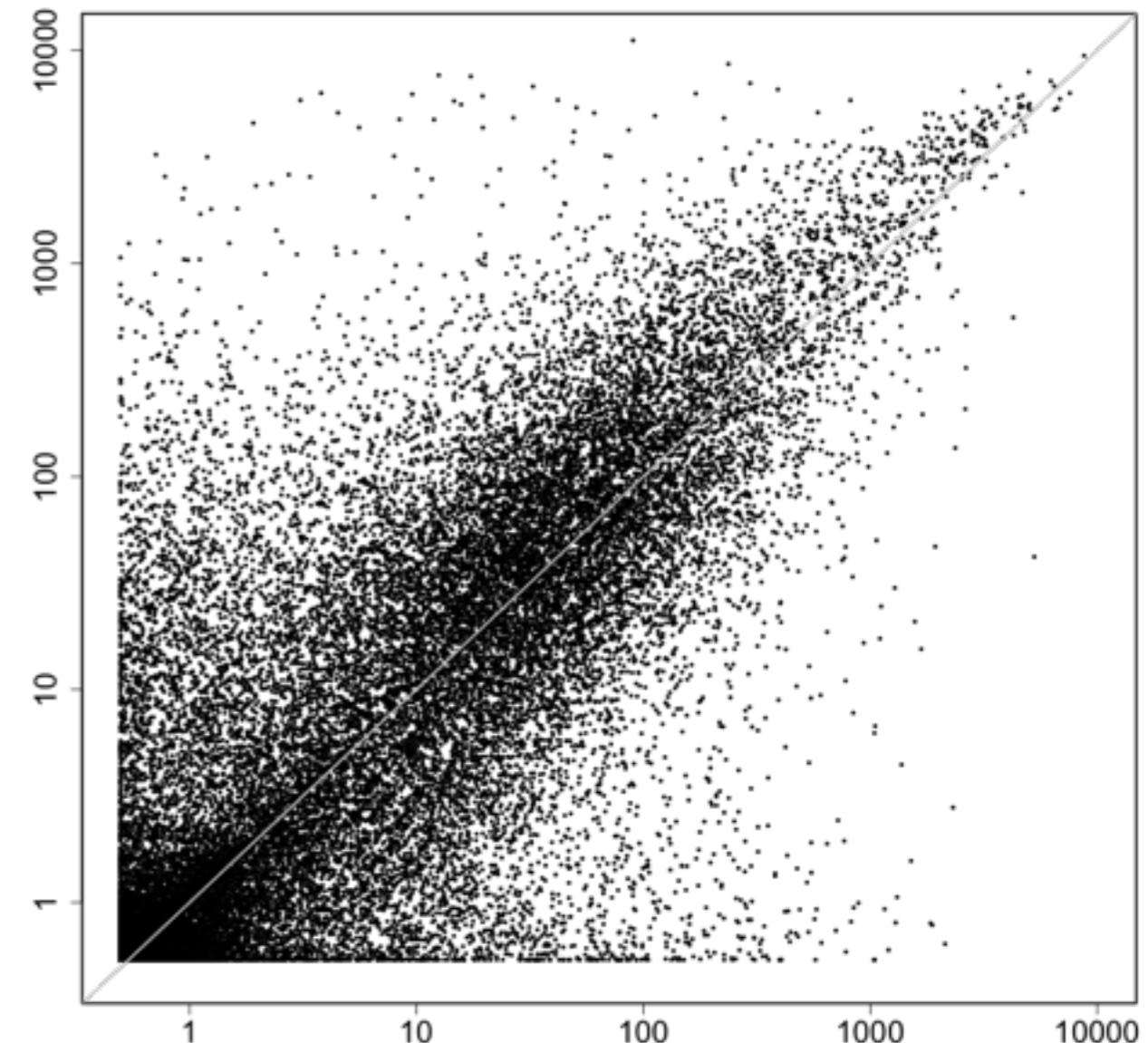
Schwanhäusser et al. *Nature* (2011).

Graphical representations: scatterplot

N=45,101

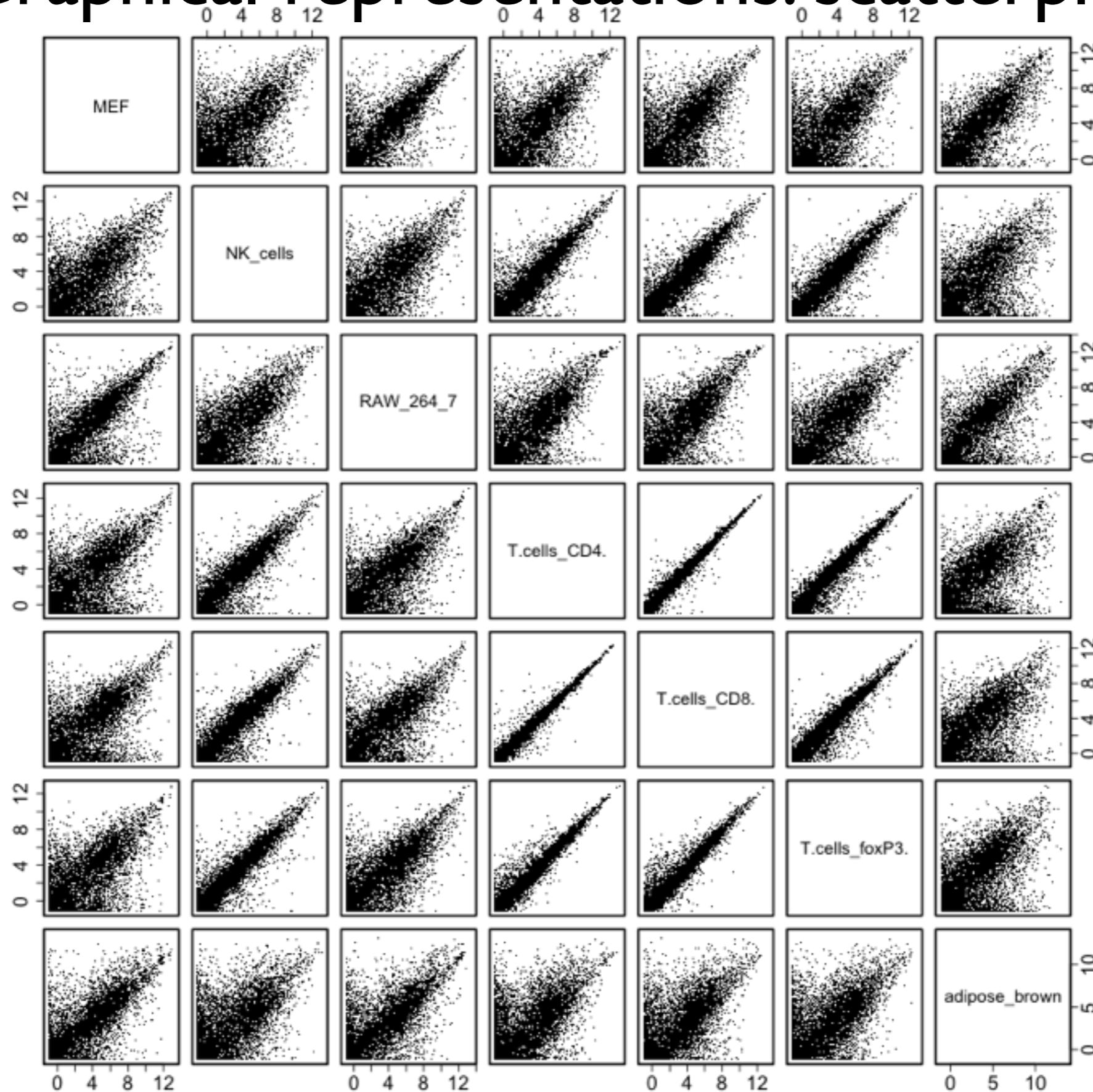


Linear scale:
correlation=0.73

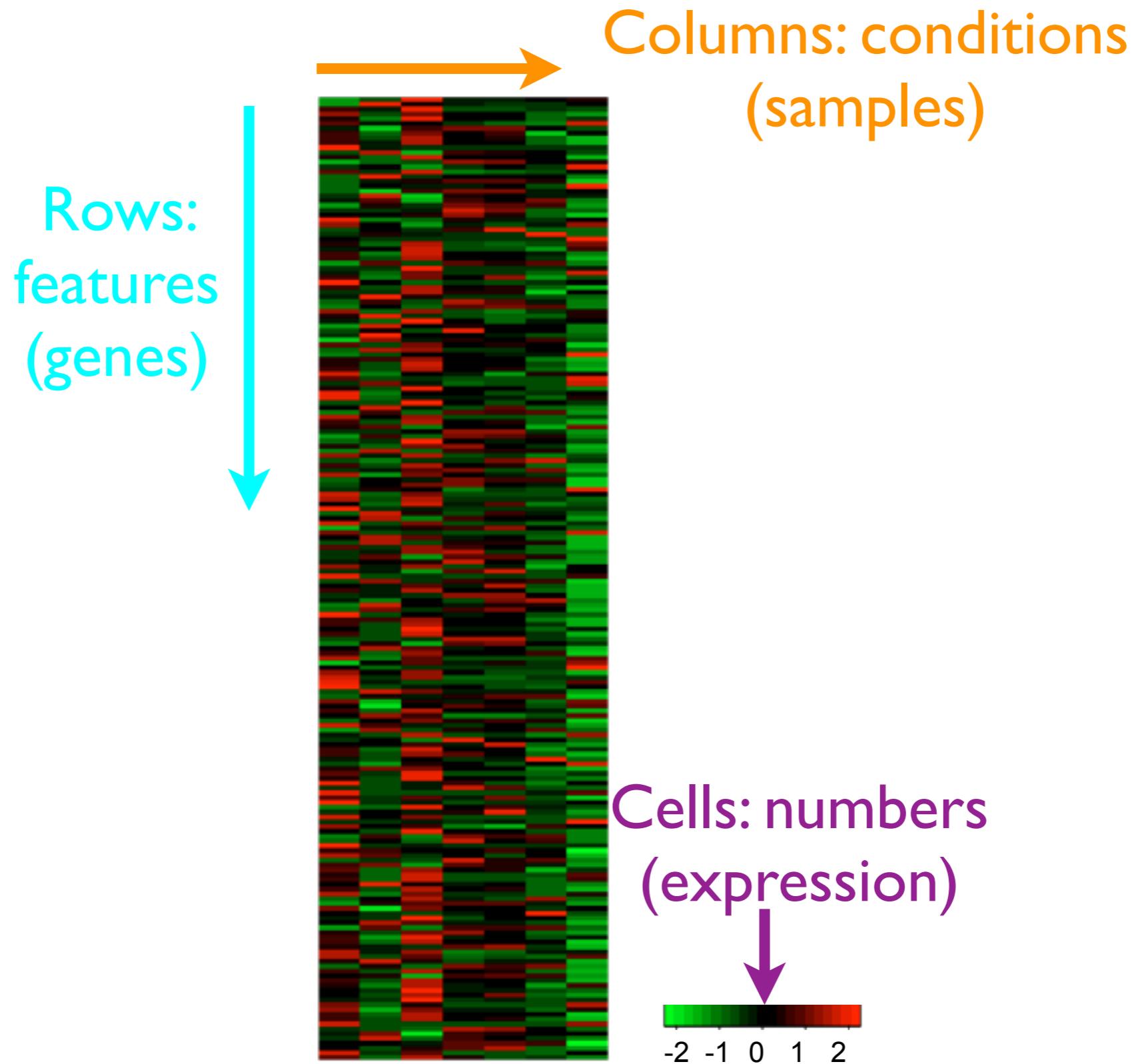


Log scale:
correlation=0.83

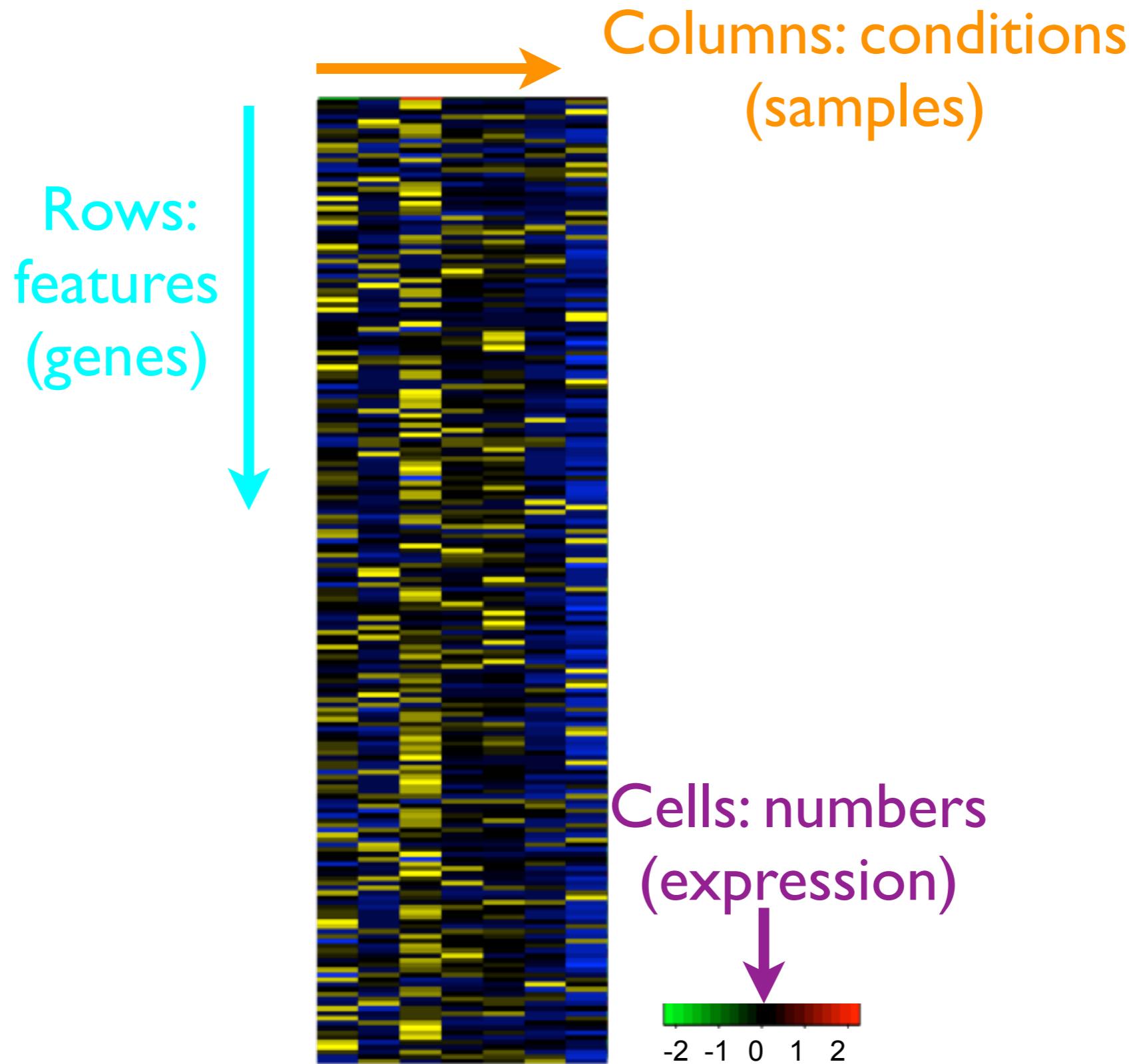
Graphical representations: scatterplot



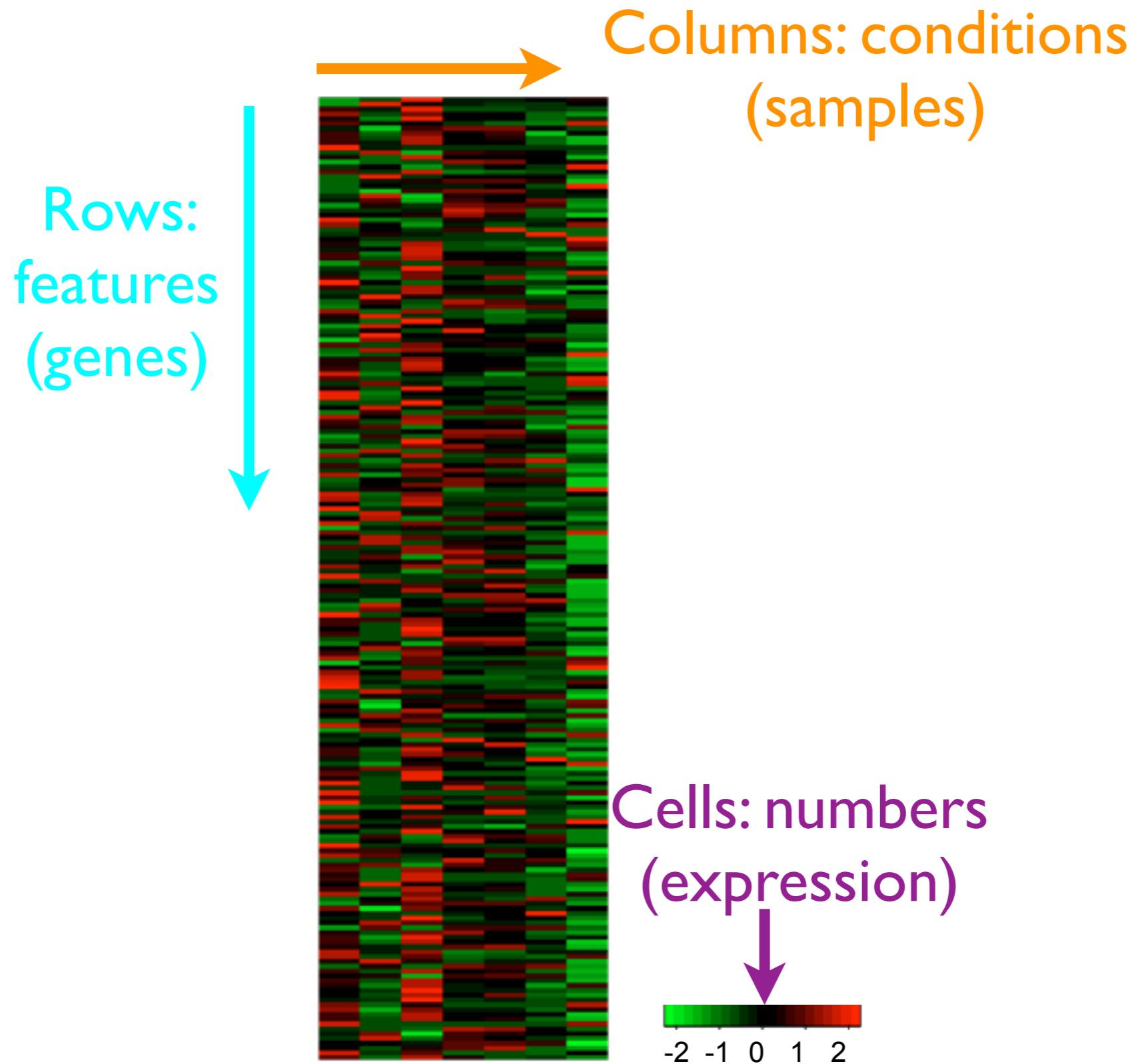
Graphical representations: heatmaps



Graphical representations: heatmaps



Graphical representations: heatmaps

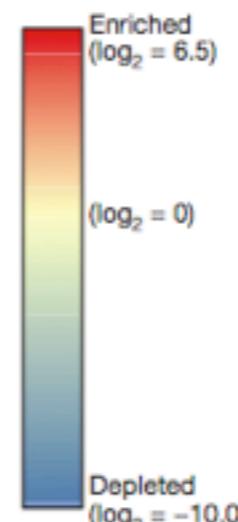
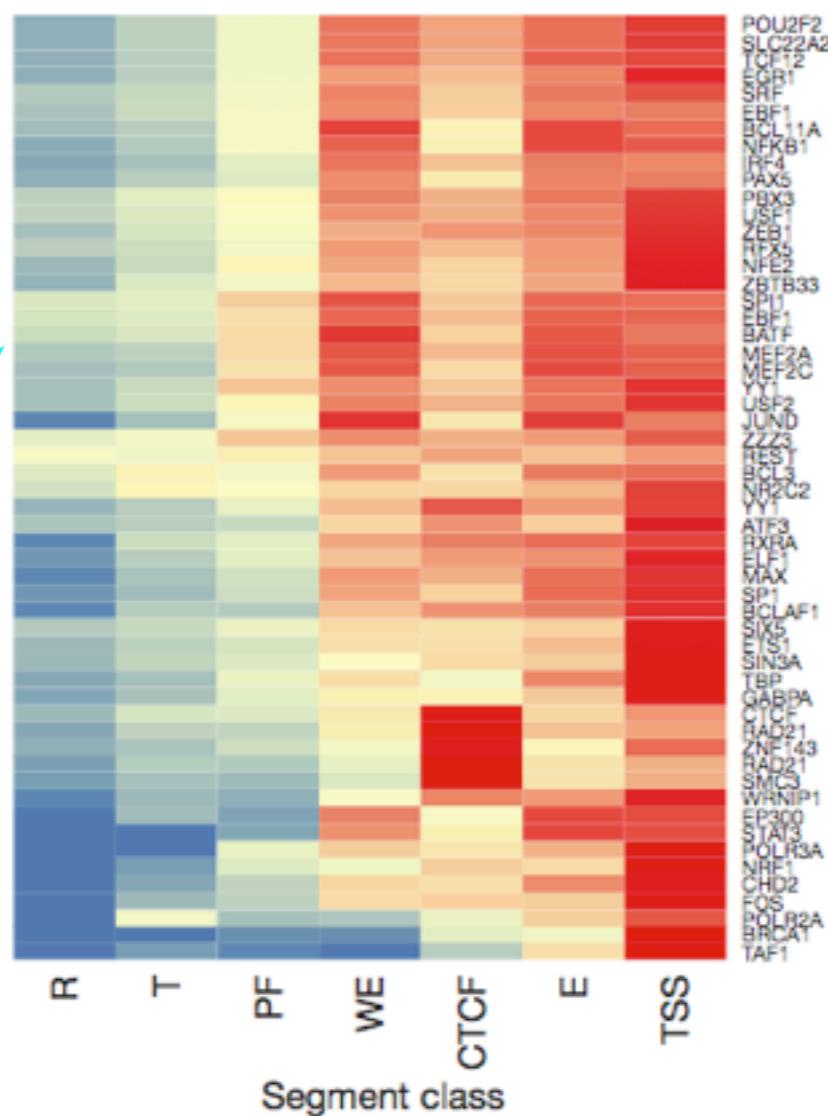


Graphical representations: heatmaps

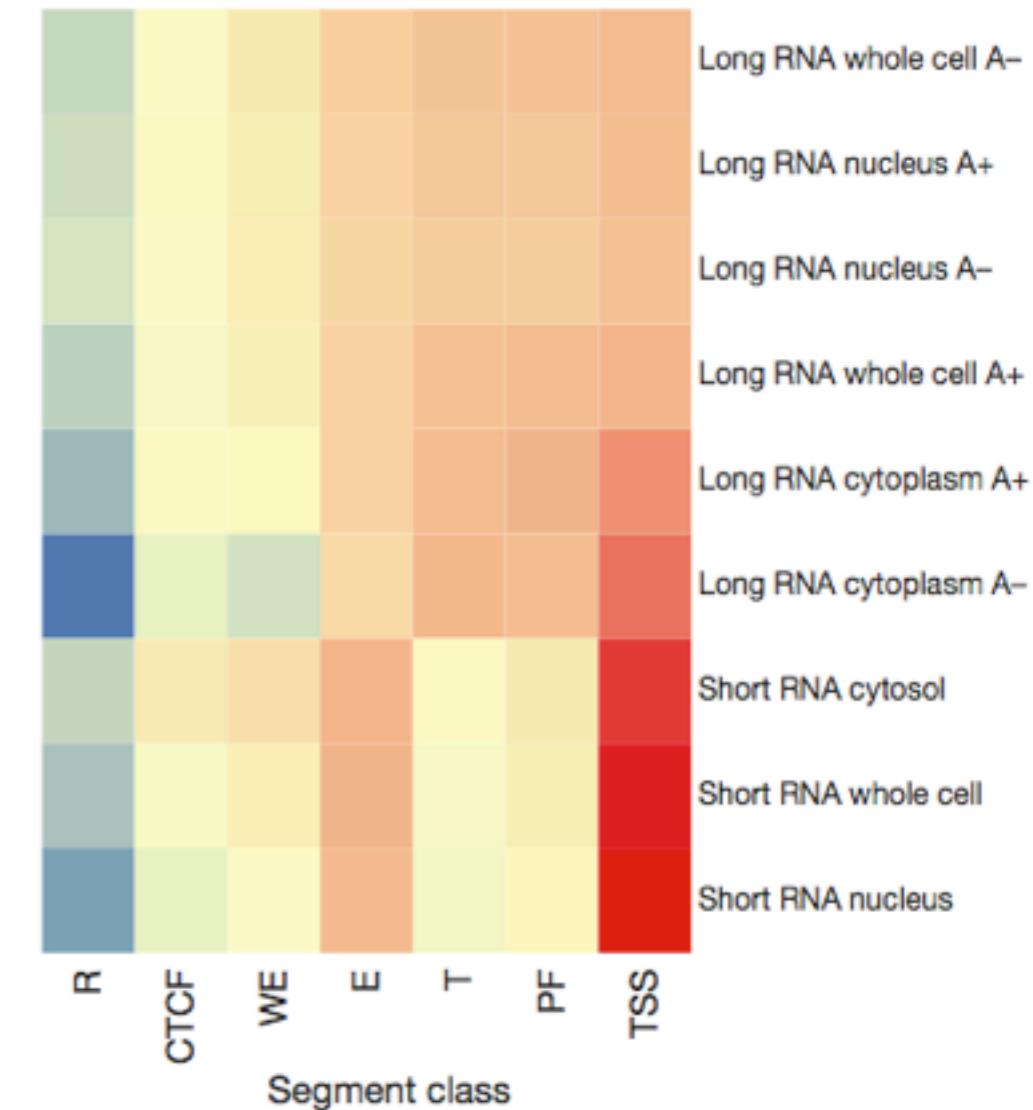
Columns: conditions
(class)

Rows:
features
(TF)

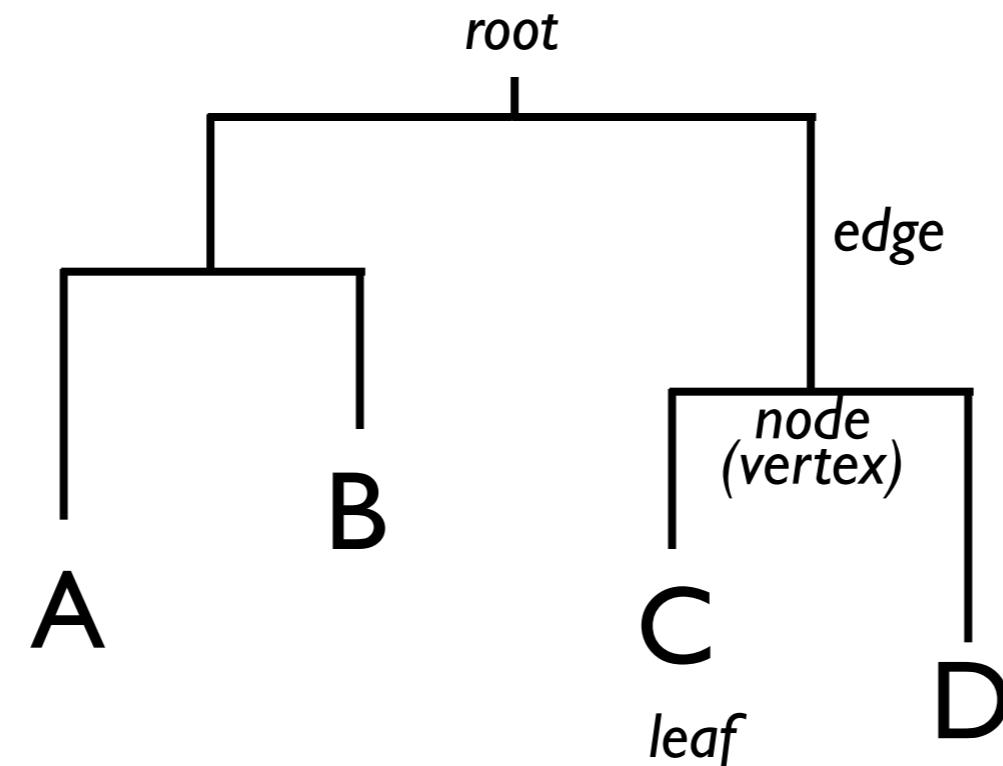
GM12878 segment overlaps with transcription factors (obs./exp. coverage)



GM12878 overlaps with RNAs (obs./exp.)



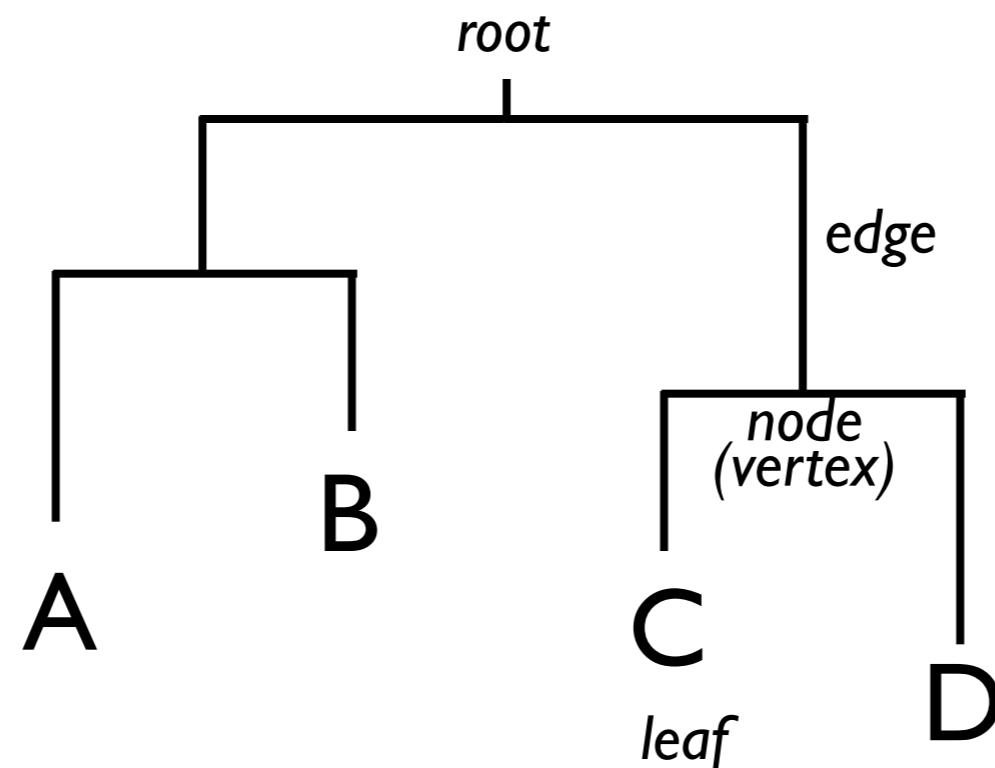
Graphical representations: (binary) trees



A tree represents a set of distances/similarities

Graphical representations: (binary) trees

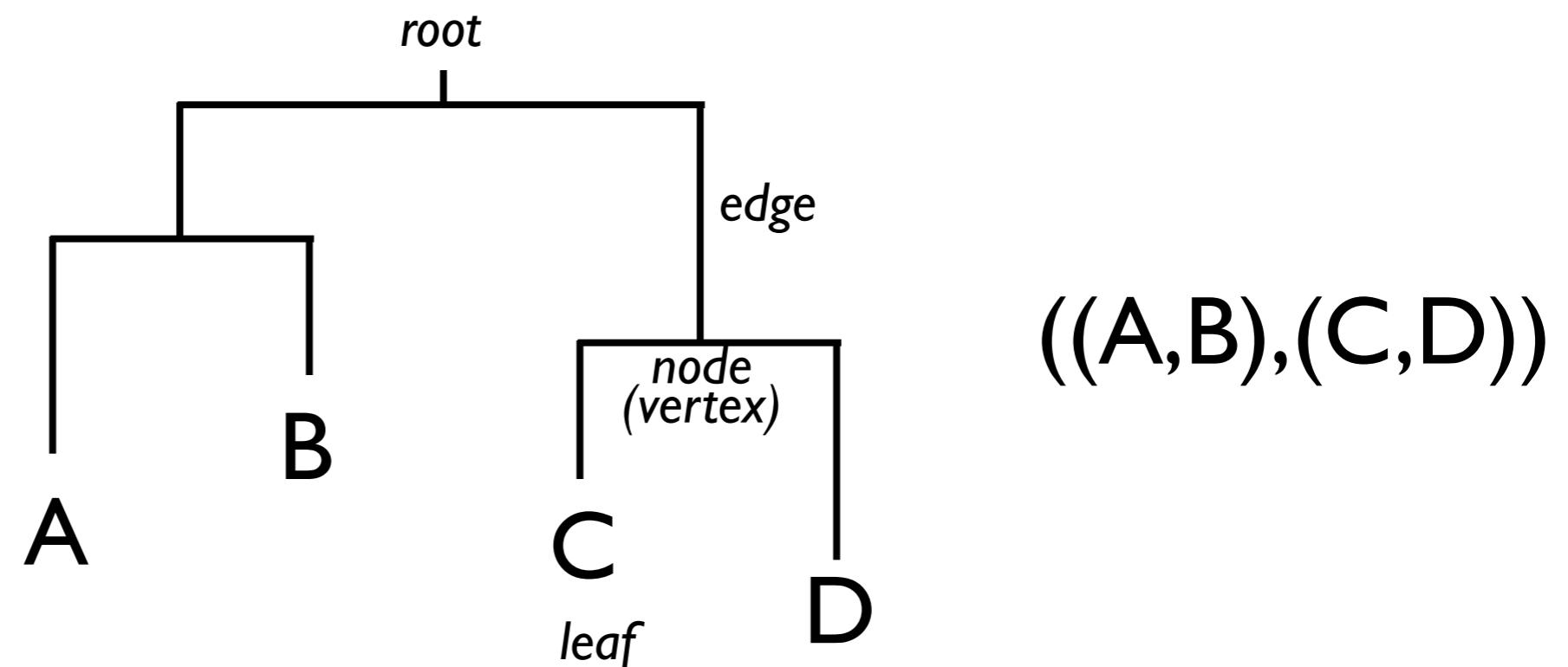
	A	B	C	D
A	0	3	6	7
B	3	0	5	6
C	6	5	0	3
D	7	6	3	0



A tree represents a set of distances/similarities

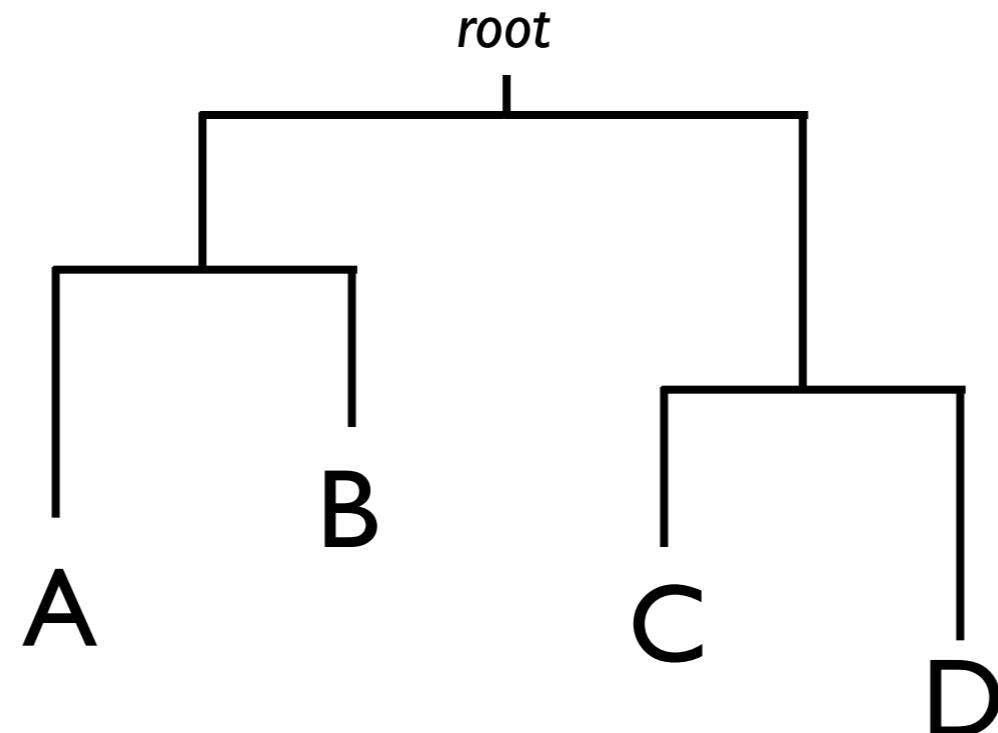
Graphical representations: (binary) trees

	A	B	C	D
A	0	3	6	7
B	3	0	5	6
C	6	5	0	3
D	7	6	3	0



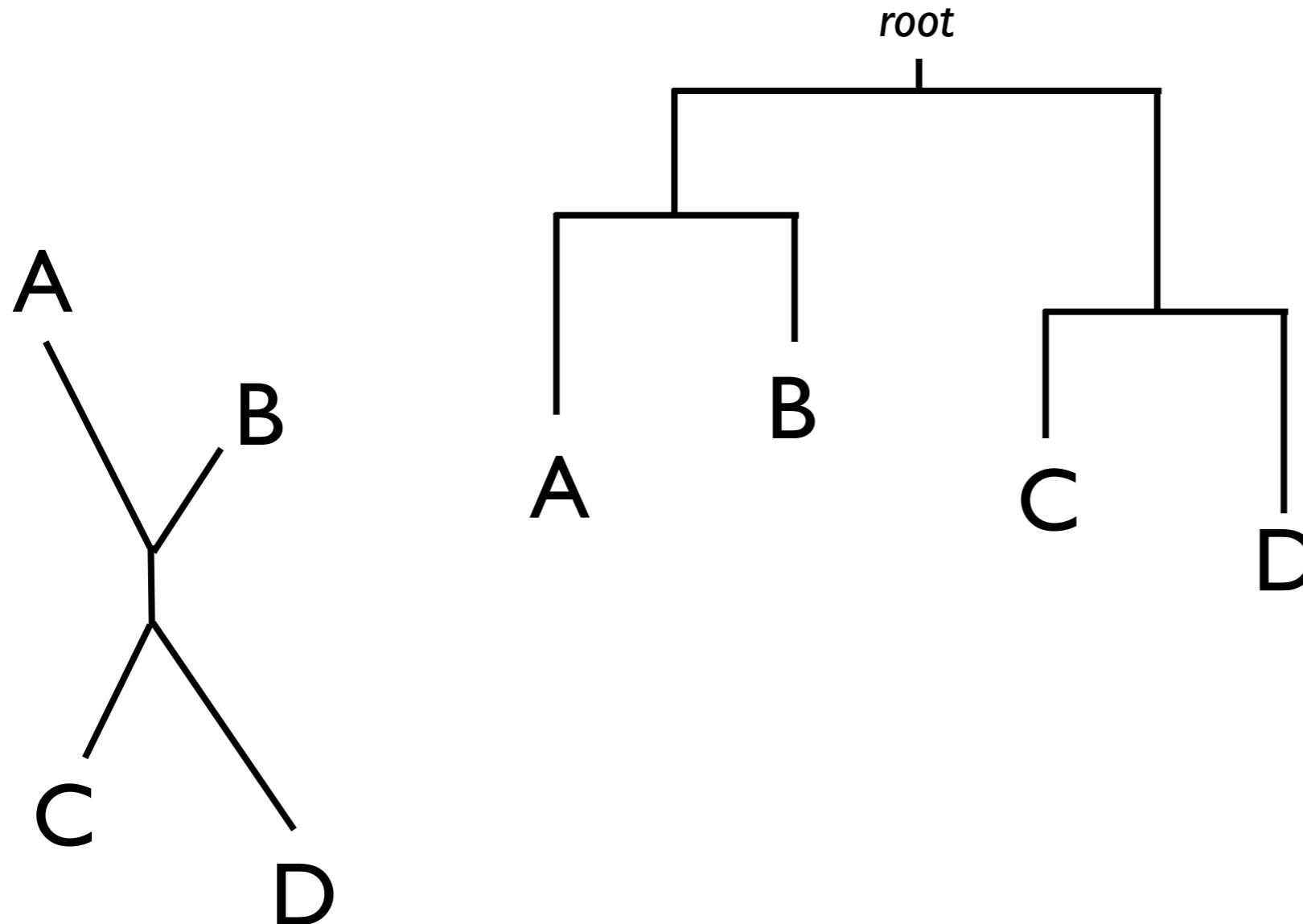
A tree represents a set of distances/similarities

Graphical representations: (binary) trees



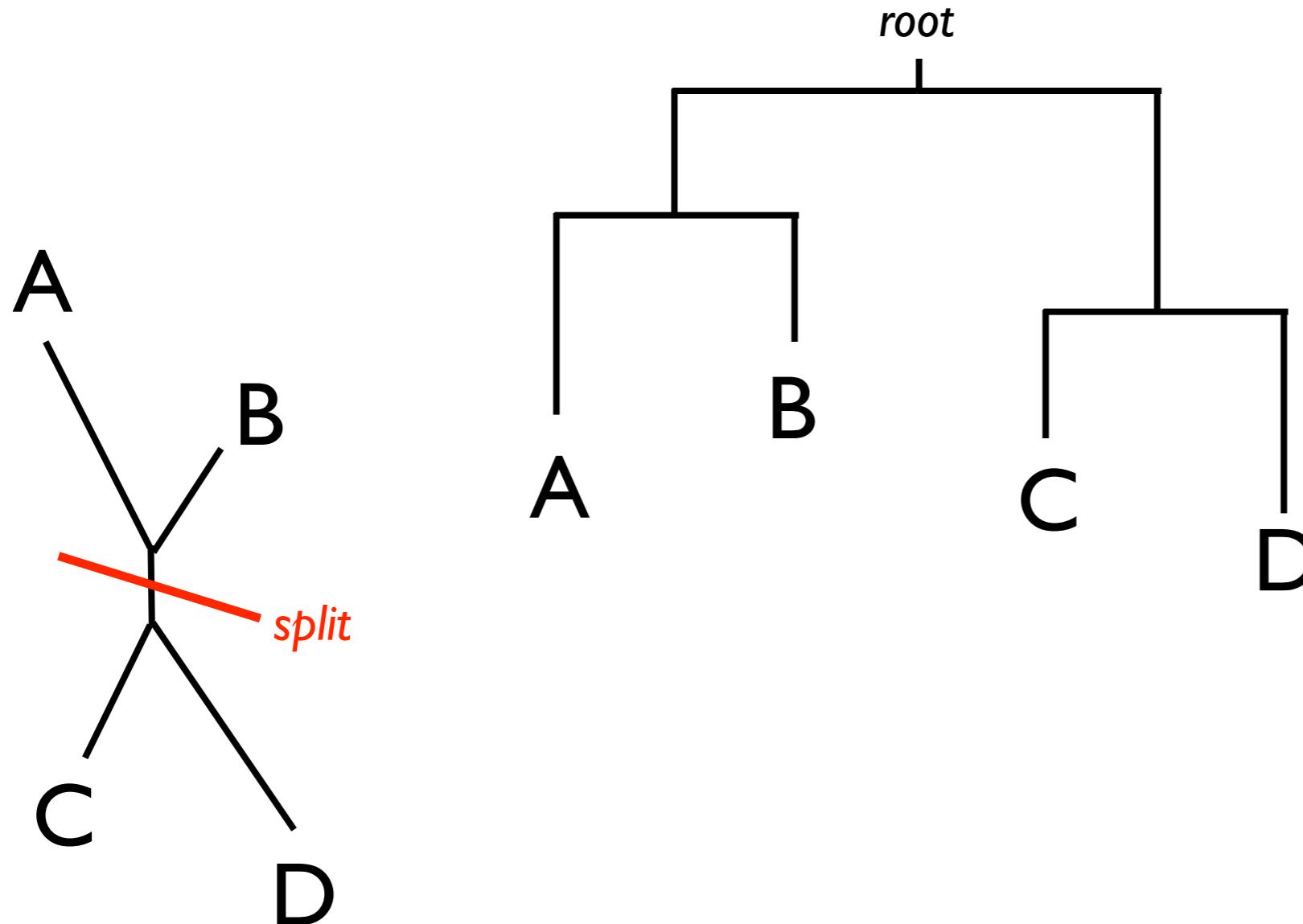
A tree represents a set of distances/similarities

Graphical representations: (binary) trees



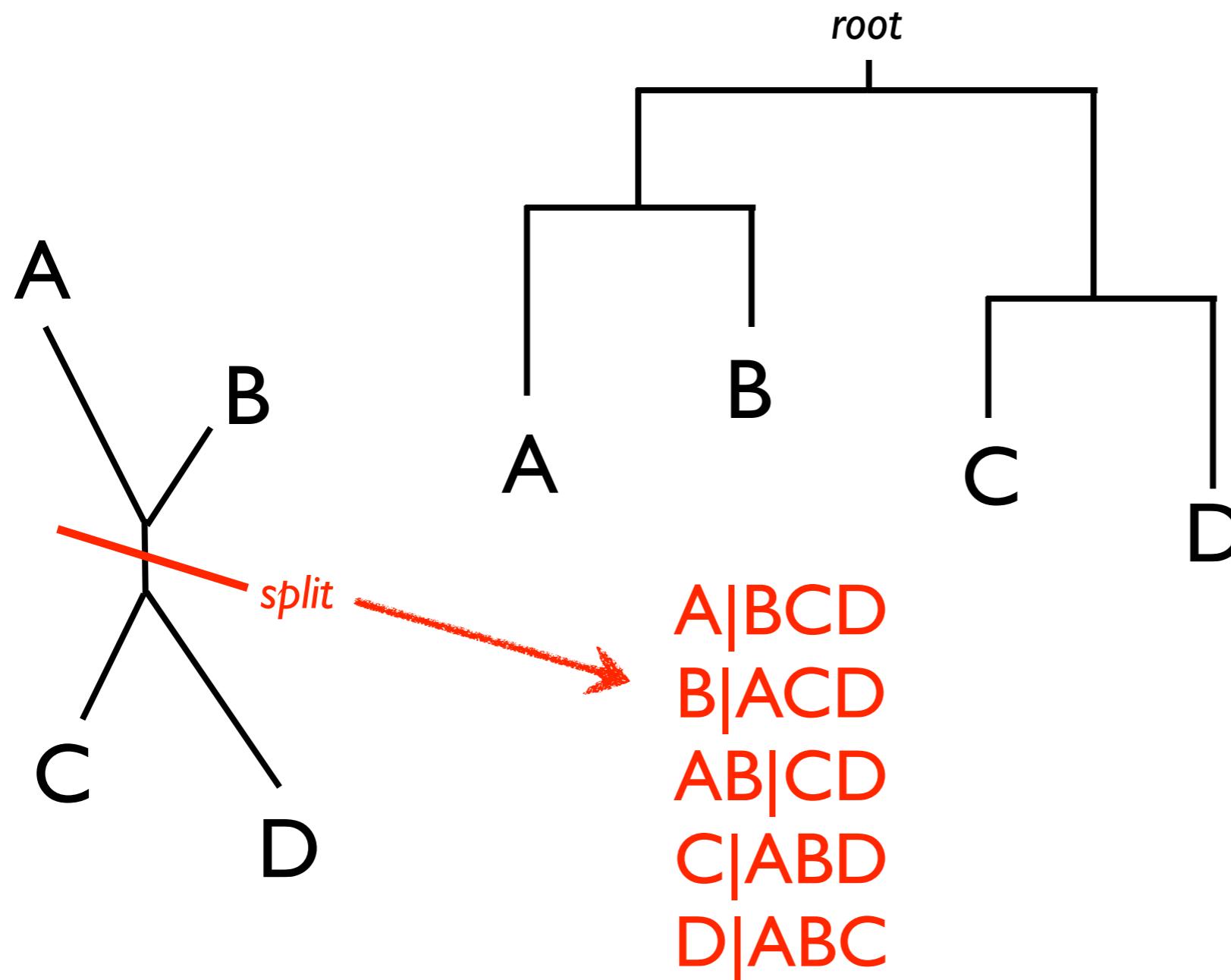
A tree represents a set of distances/similarities

Graphical representations: (binary) trees



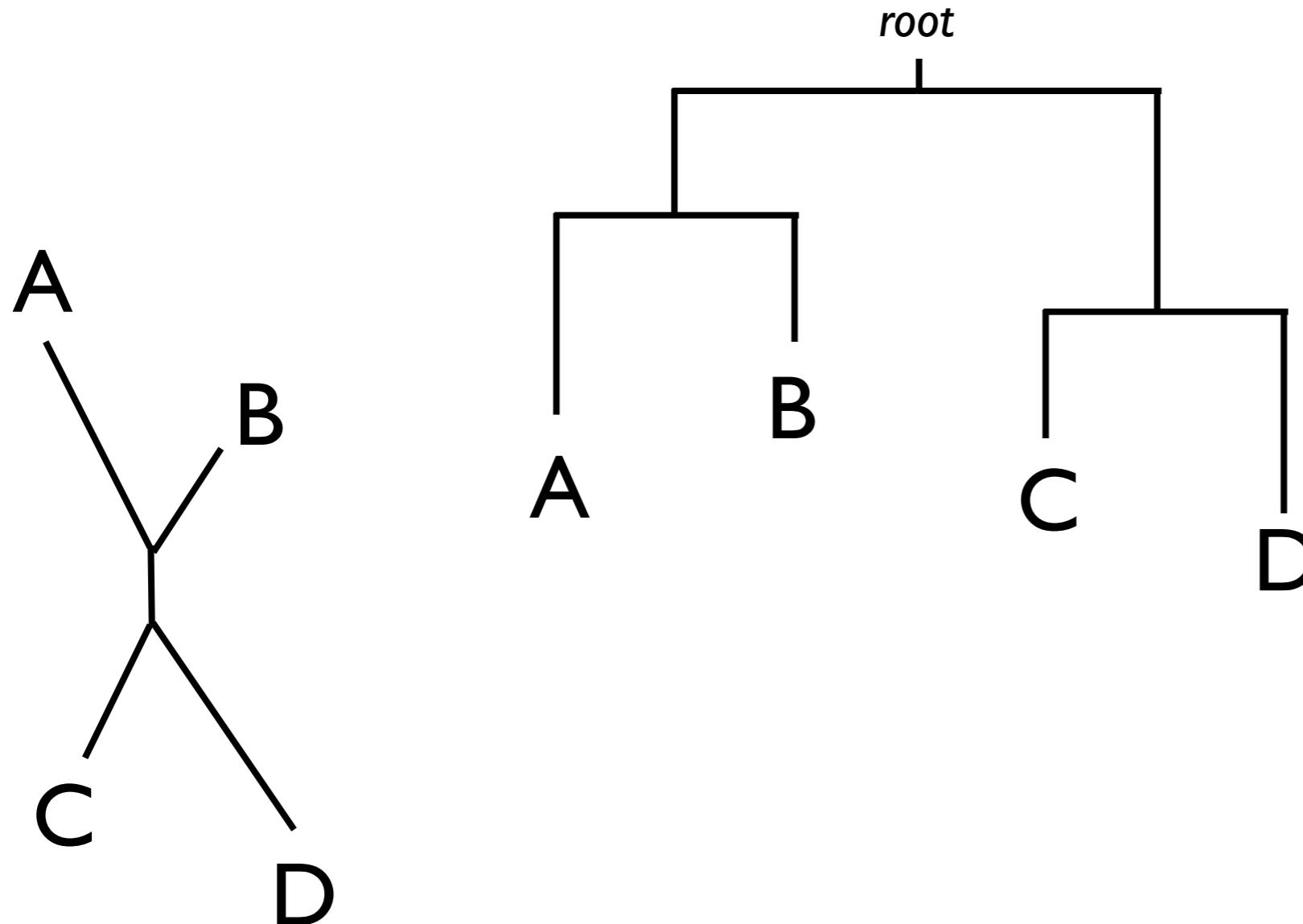
A tree represents a set of distances/similarities

Graphical representations: (binary) trees



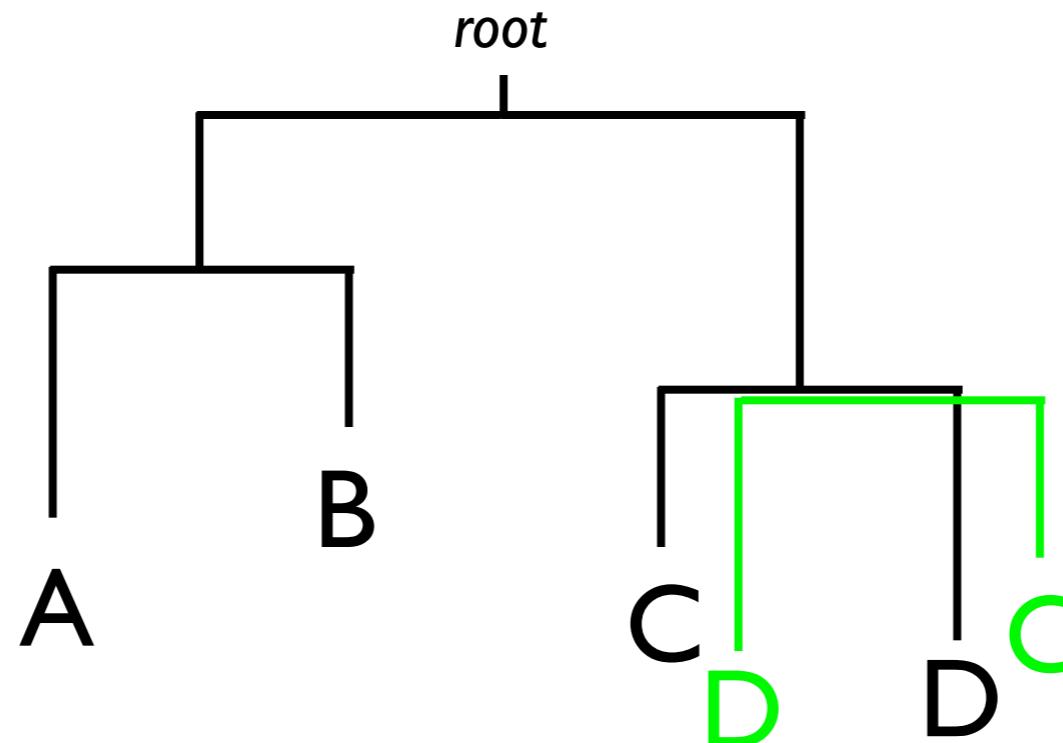
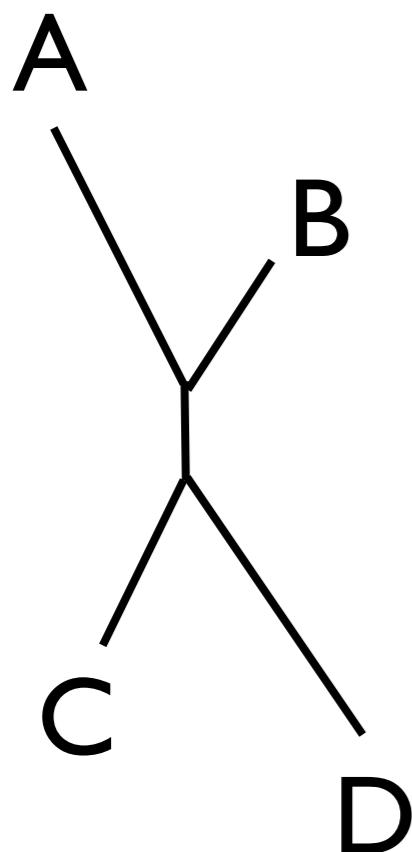
A tree represents a set of distances/similarities

Graphical representations: (binary) trees



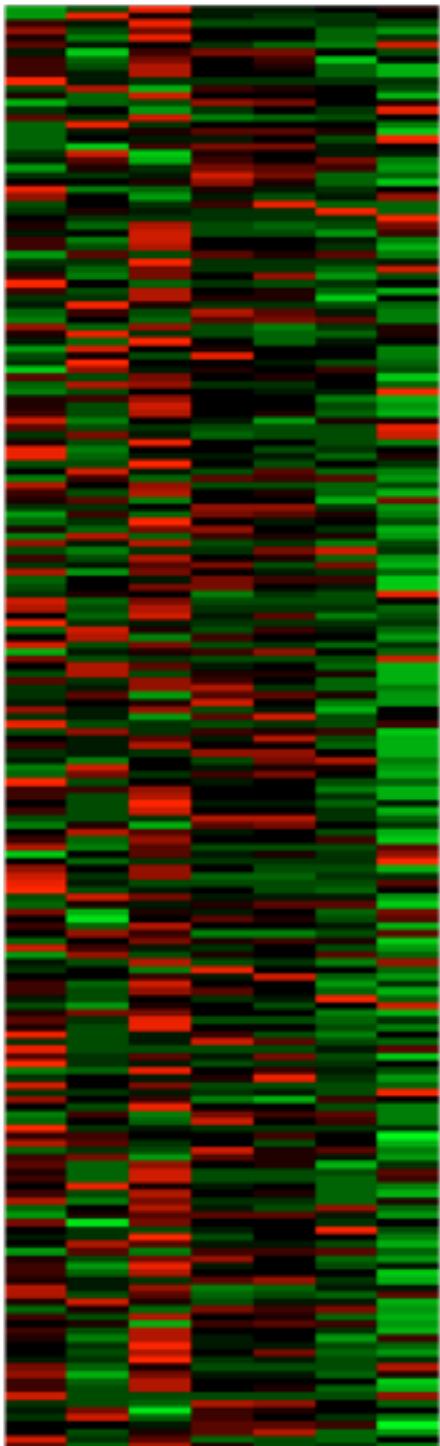
A tree represents a set of distances/similarities

Graphical representations: (binary) trees

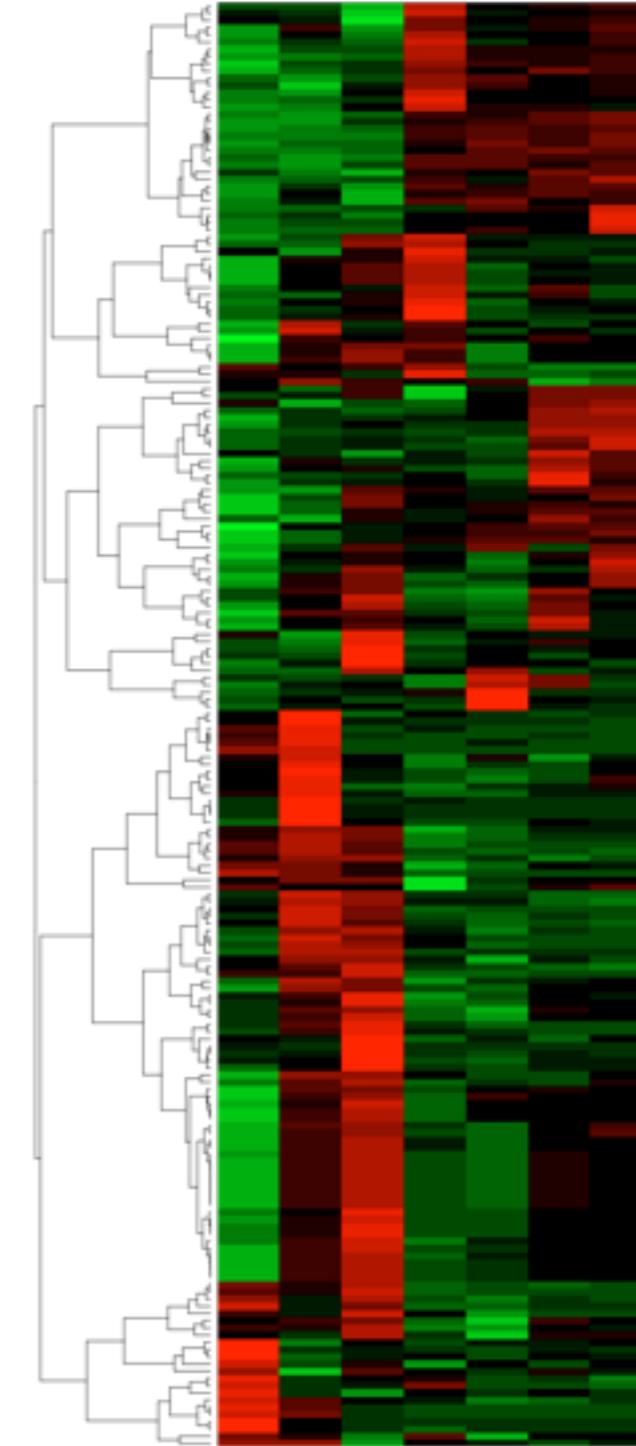


A tree represents a set of distances/similarities

Classification

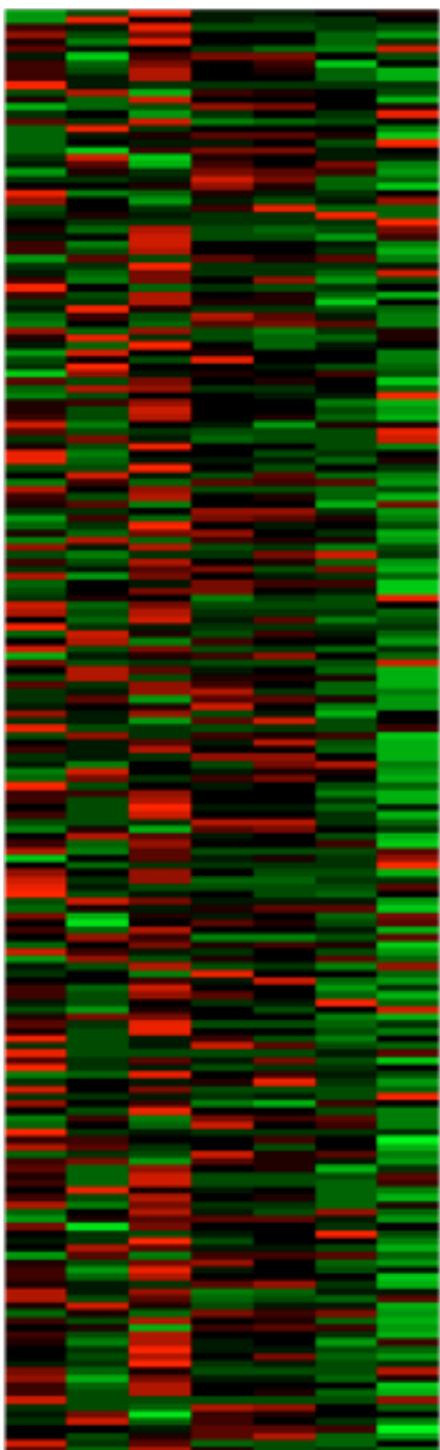


Re-order rows/columns

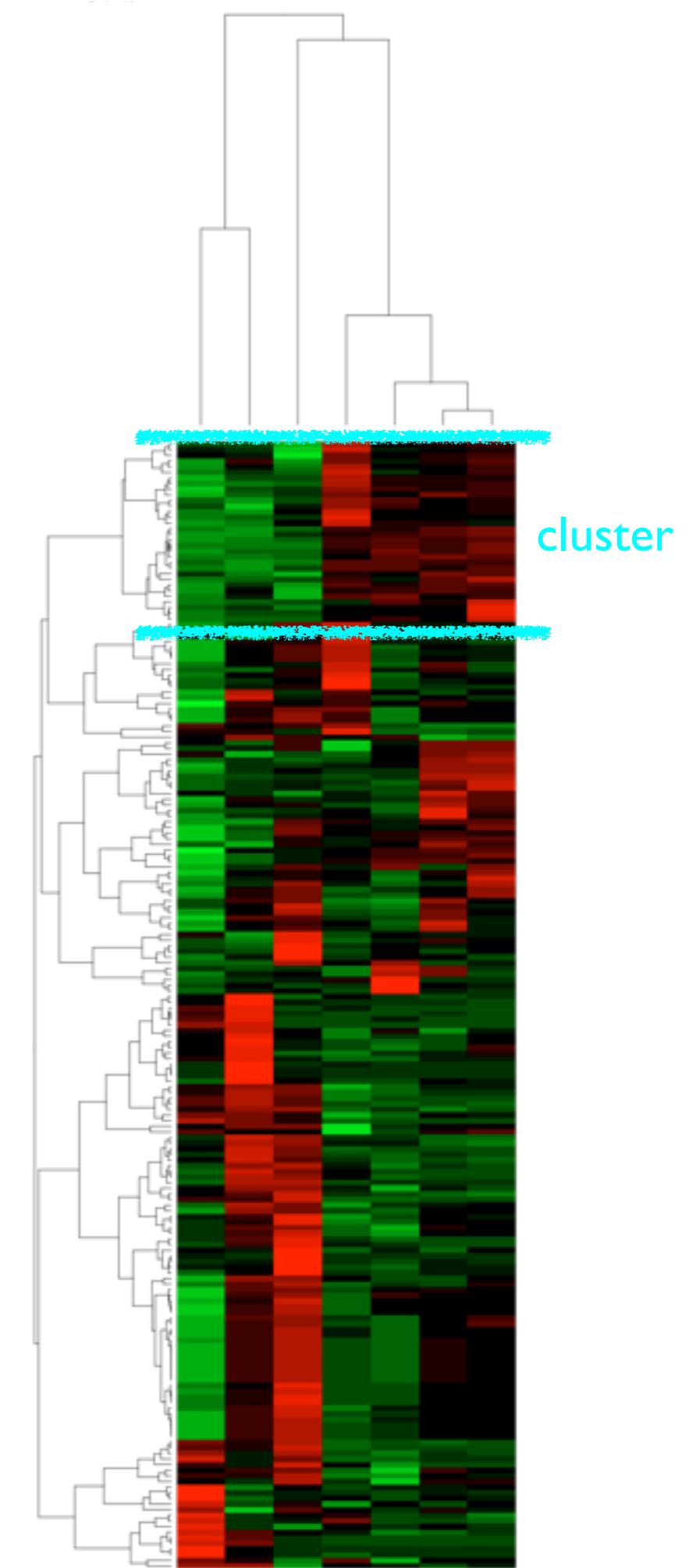


data: <http://biogps.org/downloads/>

Classification

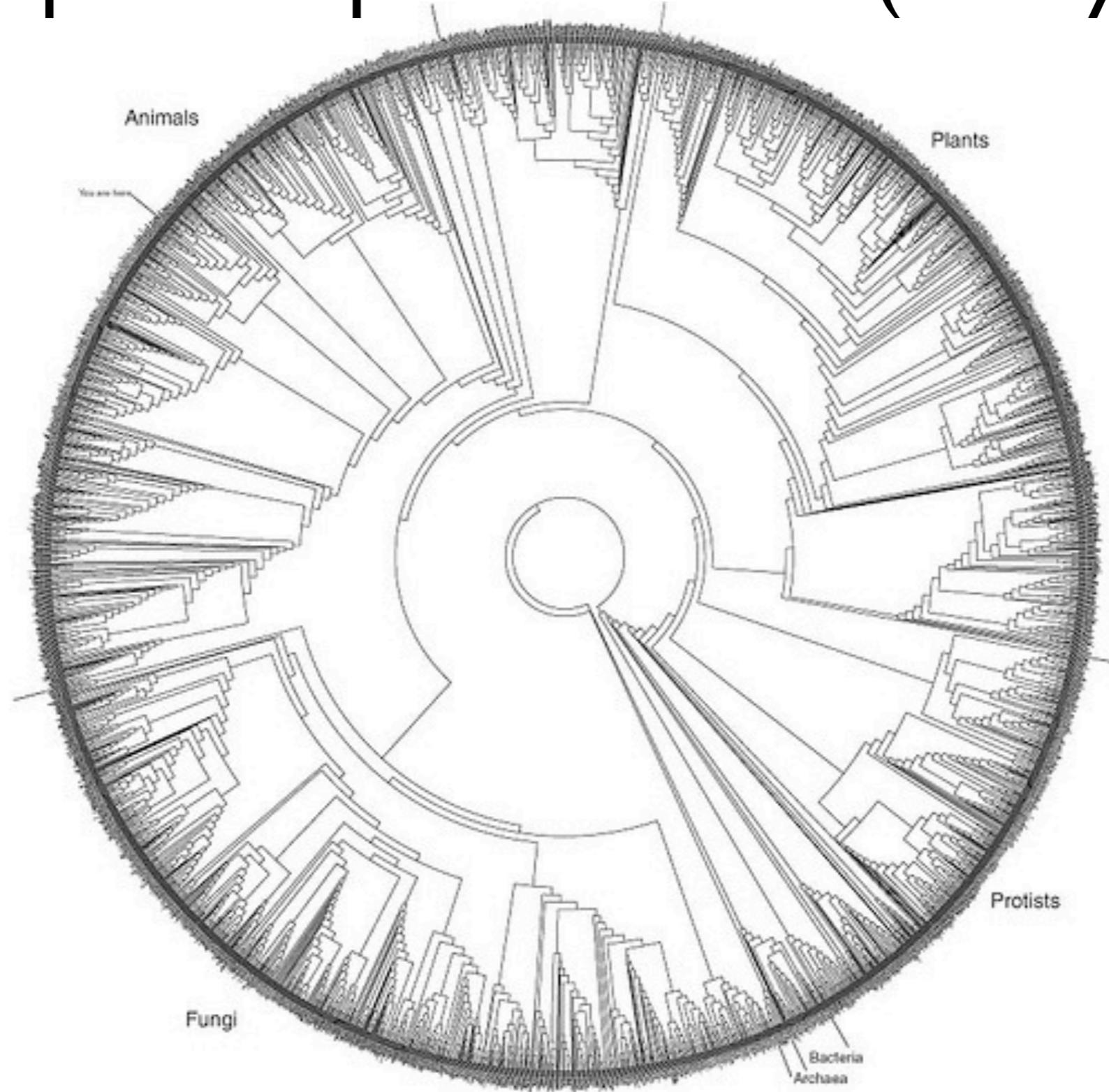


Re-order rows/columns
→



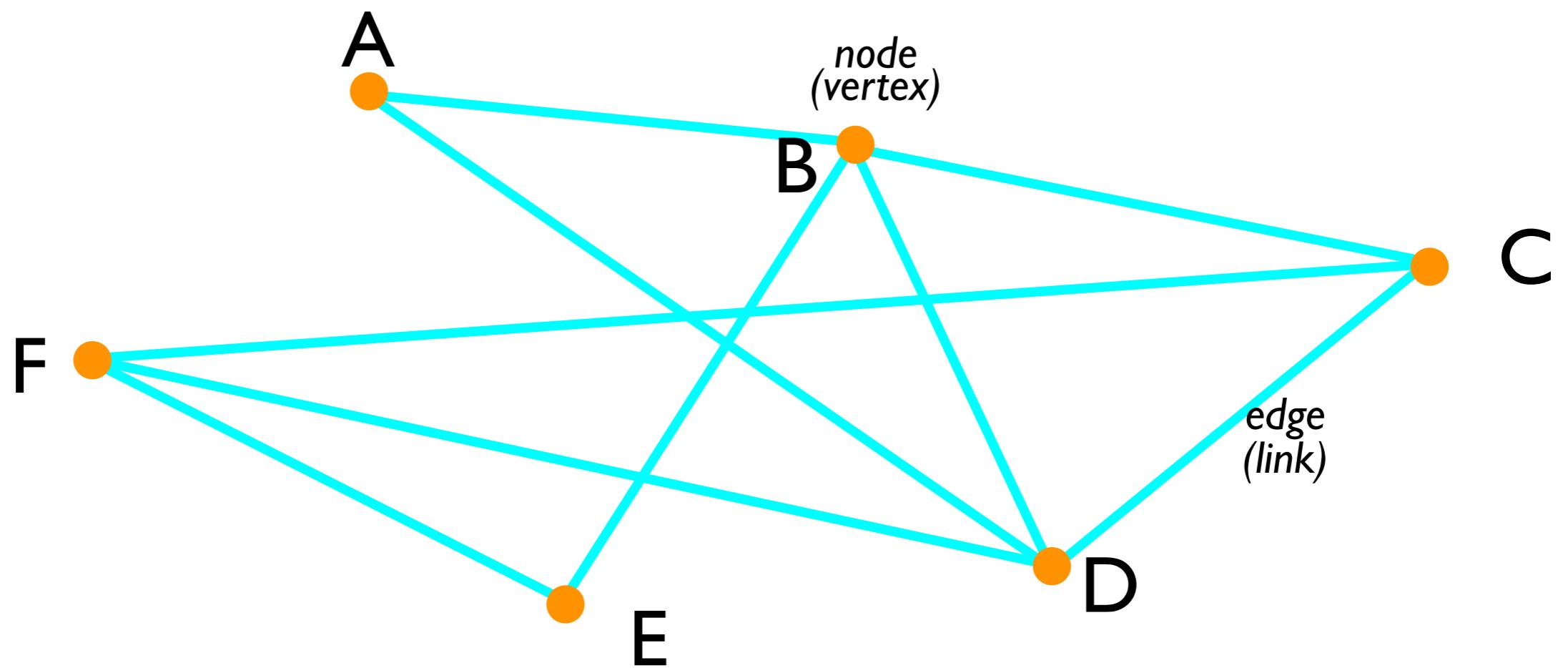
data: <http://biogps.org/downloads/>

Graphical representations: (binary) trees



3000 species, rRNA sequences

Graphs (Networks)

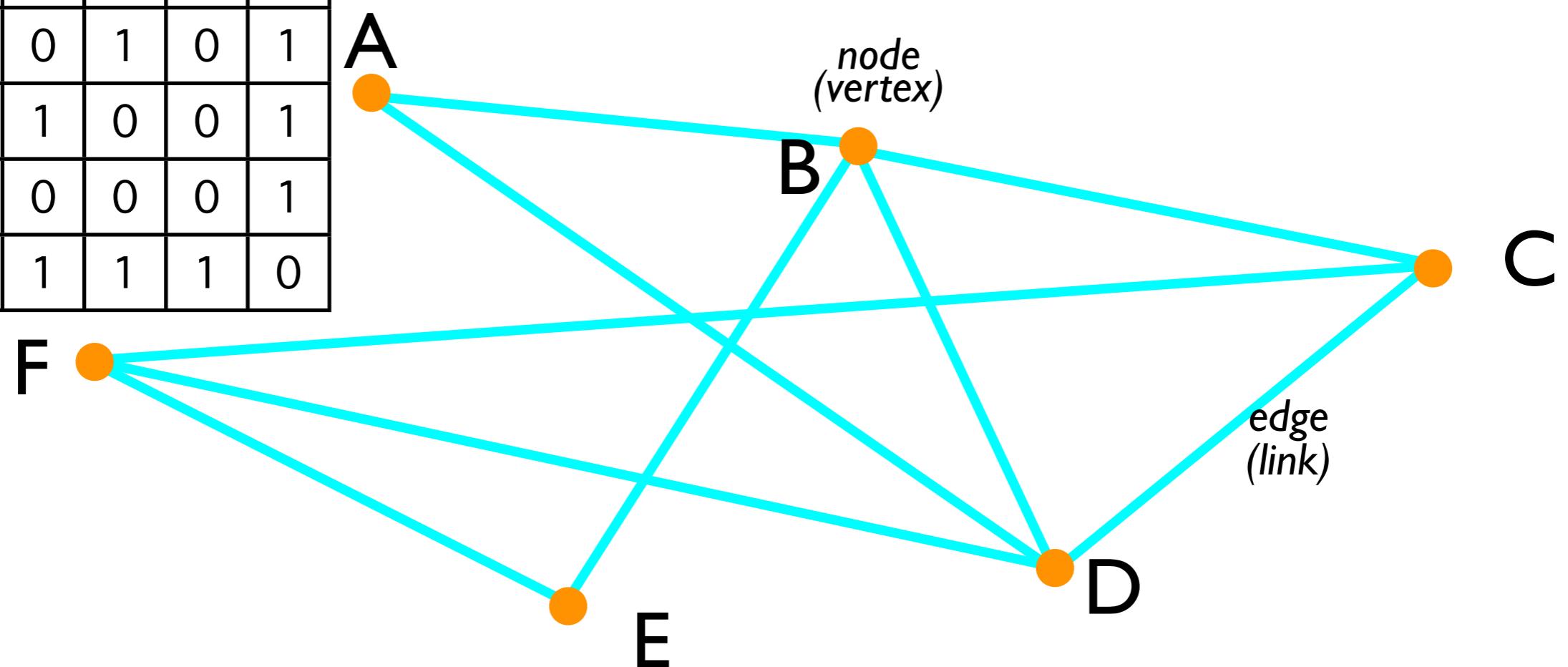


A graph represents a set of relations

Graphs (Networks)

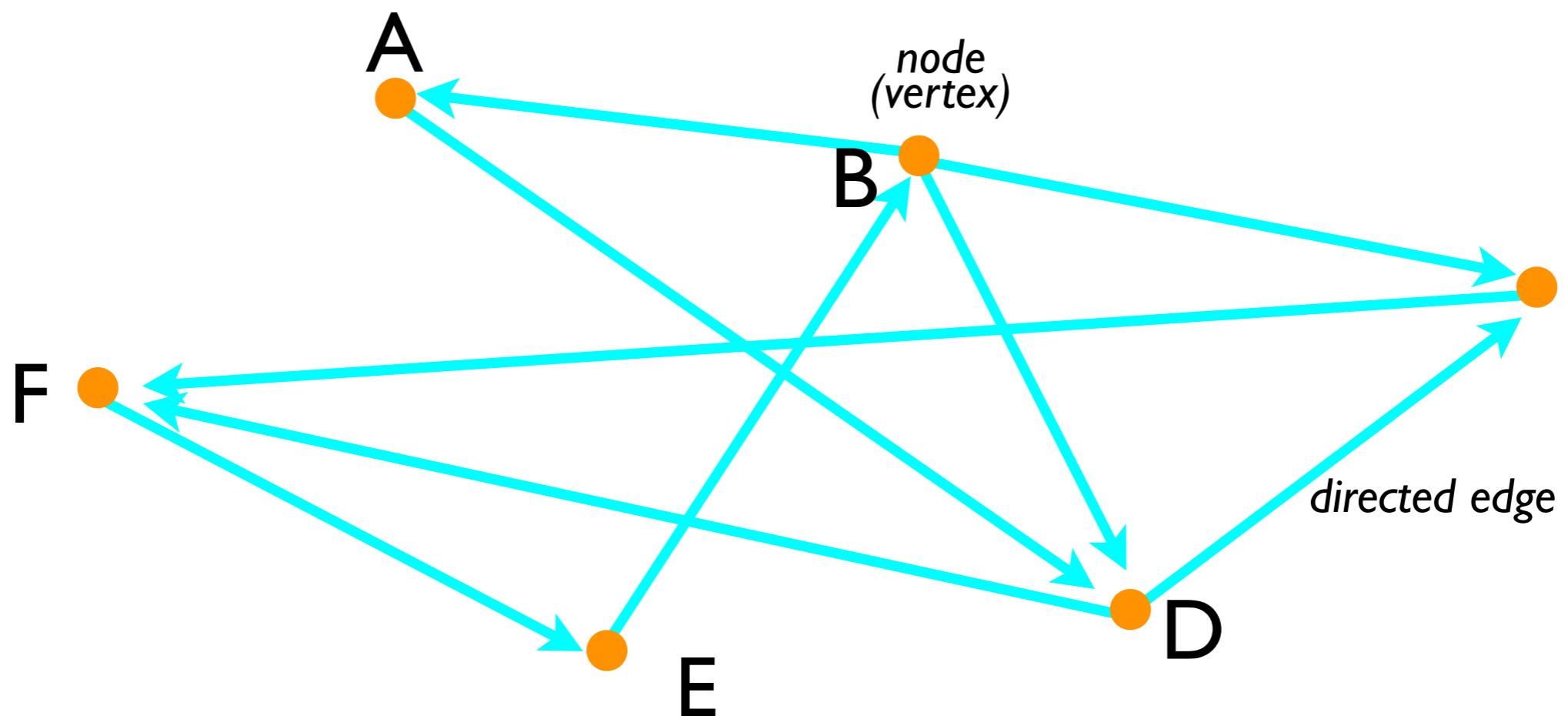
incidence matrix

	A	B	C	D	E	F
A	0	1	0	1	0	0
B	1	0	1	1	1	0
C	0	1	0	1	0	1
D	1	1	1	0	0	1
E	0	1	0	0	0	1
F	0	0	1	1	1	0



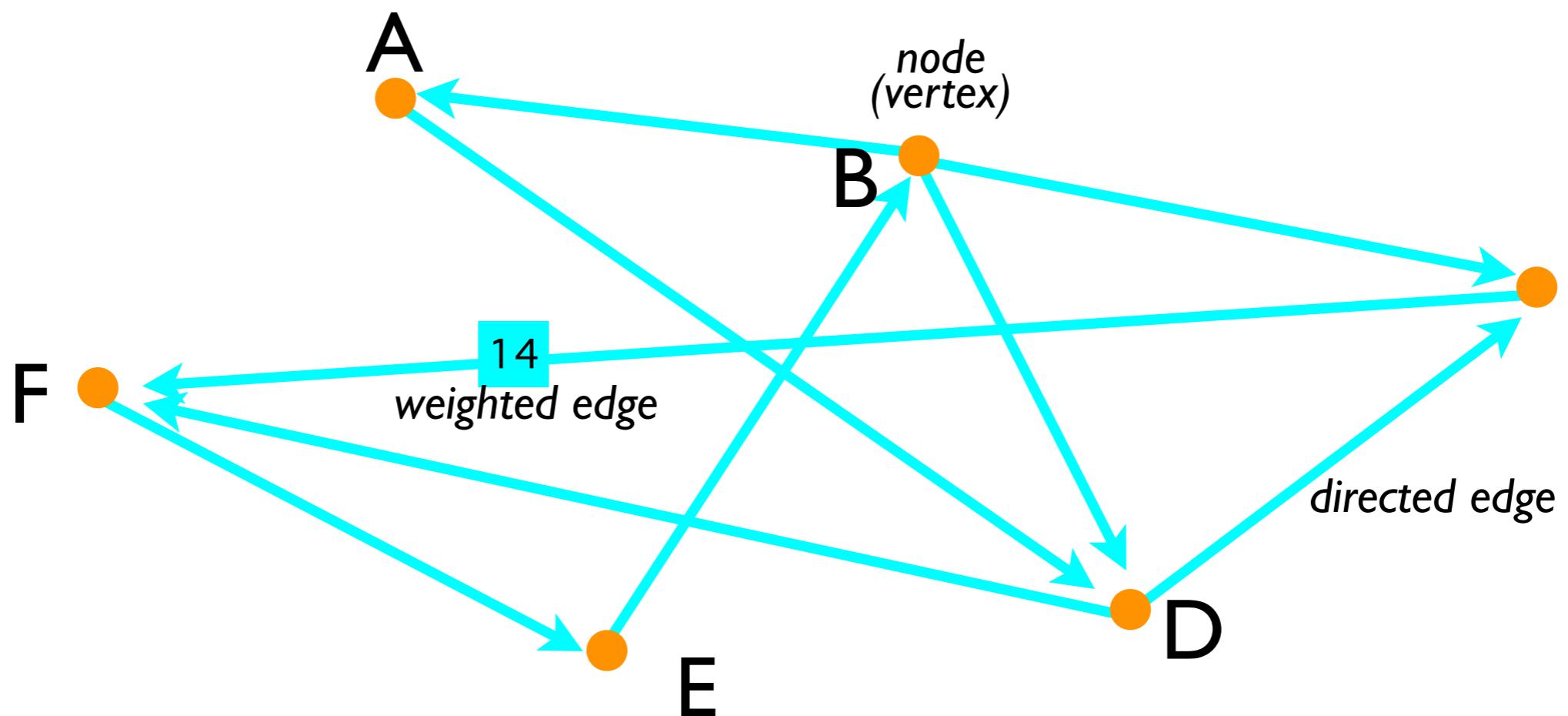
A graph represents a set of relations

Graphs (Networks)



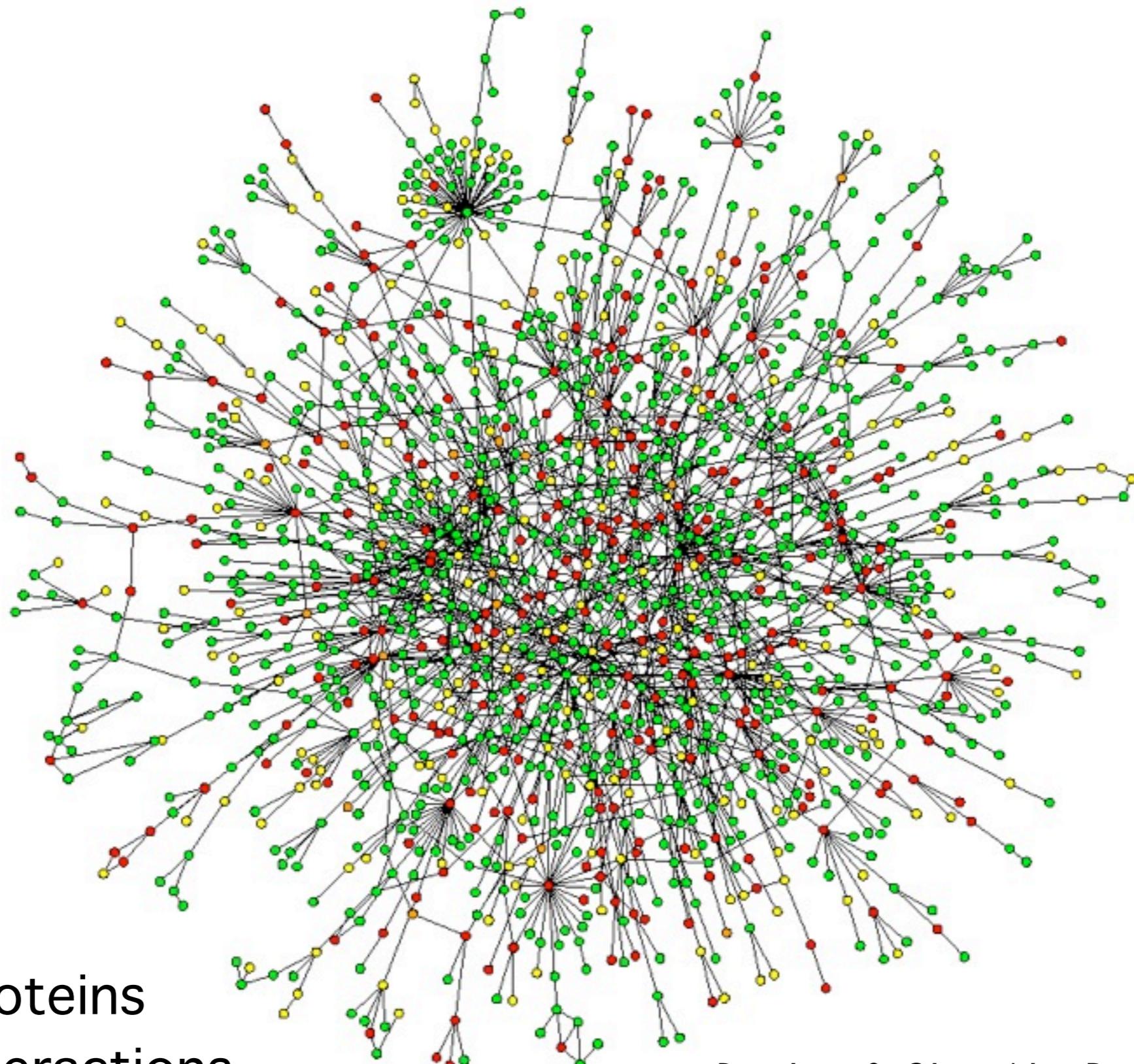
Directed graph

Graphs (Networks)



Directed graph

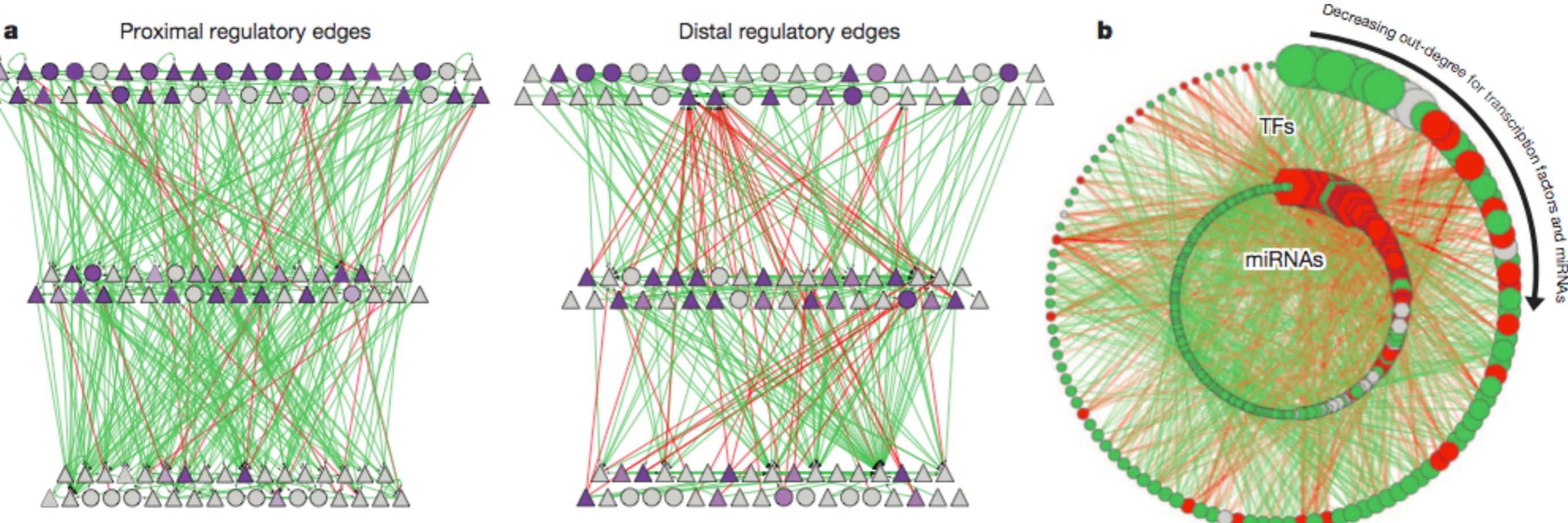
Graphs (Networks)



1870 yeast proteins
make 2240 interactions

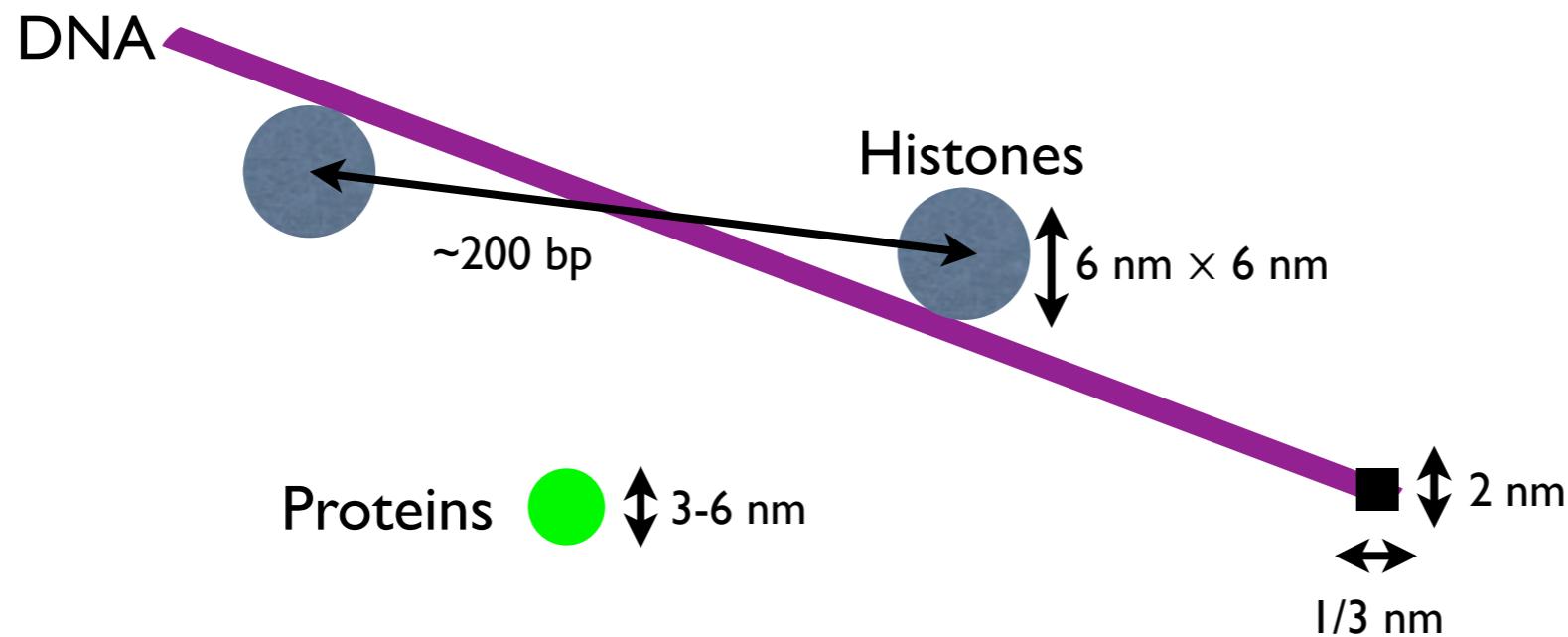
Barabasi & Oltvai, Nat. Rev. Genet. (2004)

Graphs (Networks)



Important numbers

- Human: 10^{14} cells
- Size of a mammalian cell: $100\text{-}10,000 \mu\text{m}^3$ (yeast: $20\text{-}100 \mu\text{m}^3$)
- Size of a nucleus: $300 \mu\text{m}^3$ (10% of cell)



Important numbers

- Cell cycle time: ~1 day (yeast: 200 min.)
- Transcription: max 70 nt/sec., average 1-2kb/min
- Translation: 10-20 aa/sec
- Number of mRNA per cell: 300,000 (human), 15,000 (yeast), 1380 (e. coli)
- Mutation rate: 10^{-8} per basepair, per duplication
- Misincorporation rate: 10^{-4} - 10^{-5} per nucleotide (transcription)/
 10^{-3} - 10^{-4} per amino acid (translation)

<http://bionumbers.hms.harvard.edu>

References

- History and philosophy:
 - Michel Morange, *Histoire de la biologie moléculaire*, La Découverte (2003)
 - Evelyn Fox Keller, *Making sense of life*, Harvard UP (2002)
- Numbers and orders of magnitude:
 - Rob Phillips et al, *Physical Biology of the Cell*, Garland (2008)
 - “SnapShot”, *Cell* 141 (2010)