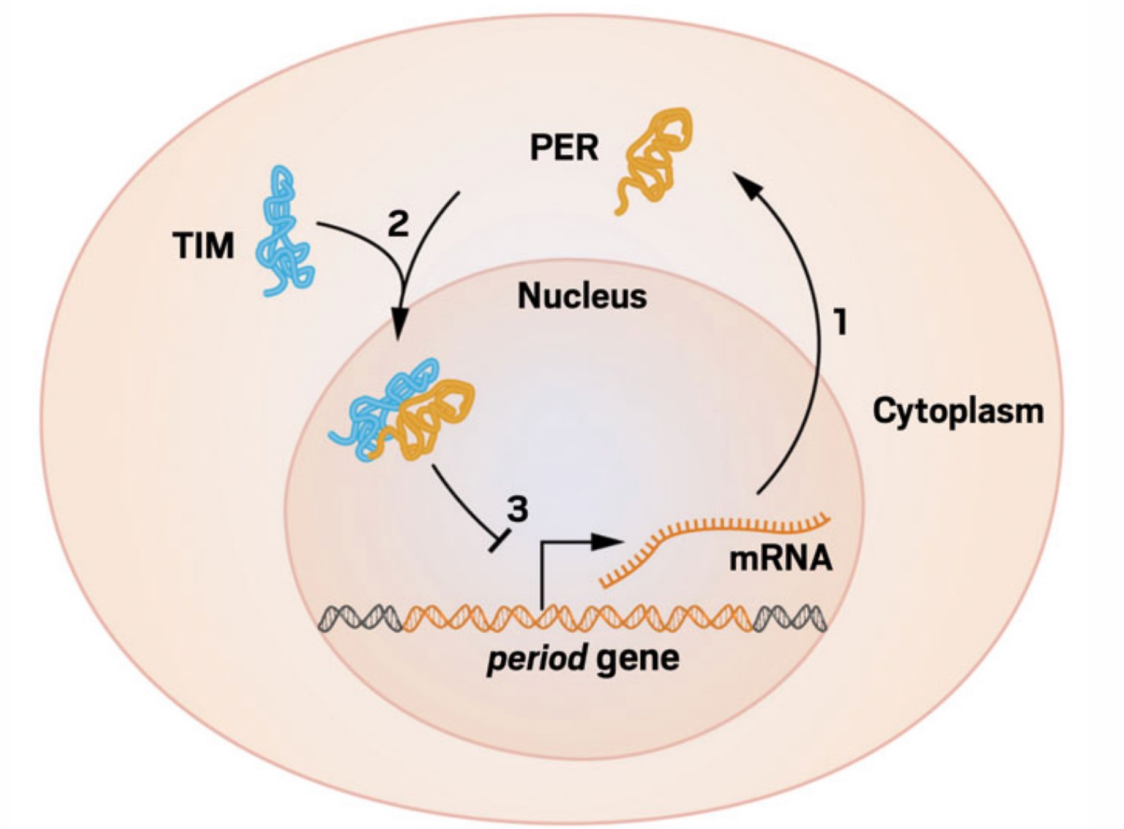
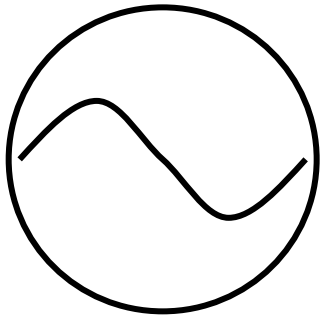
Time-course experiments and Fourier analysis

Neuro-genomics
Class 4

Example of the need for time-course experiments: circadian biology

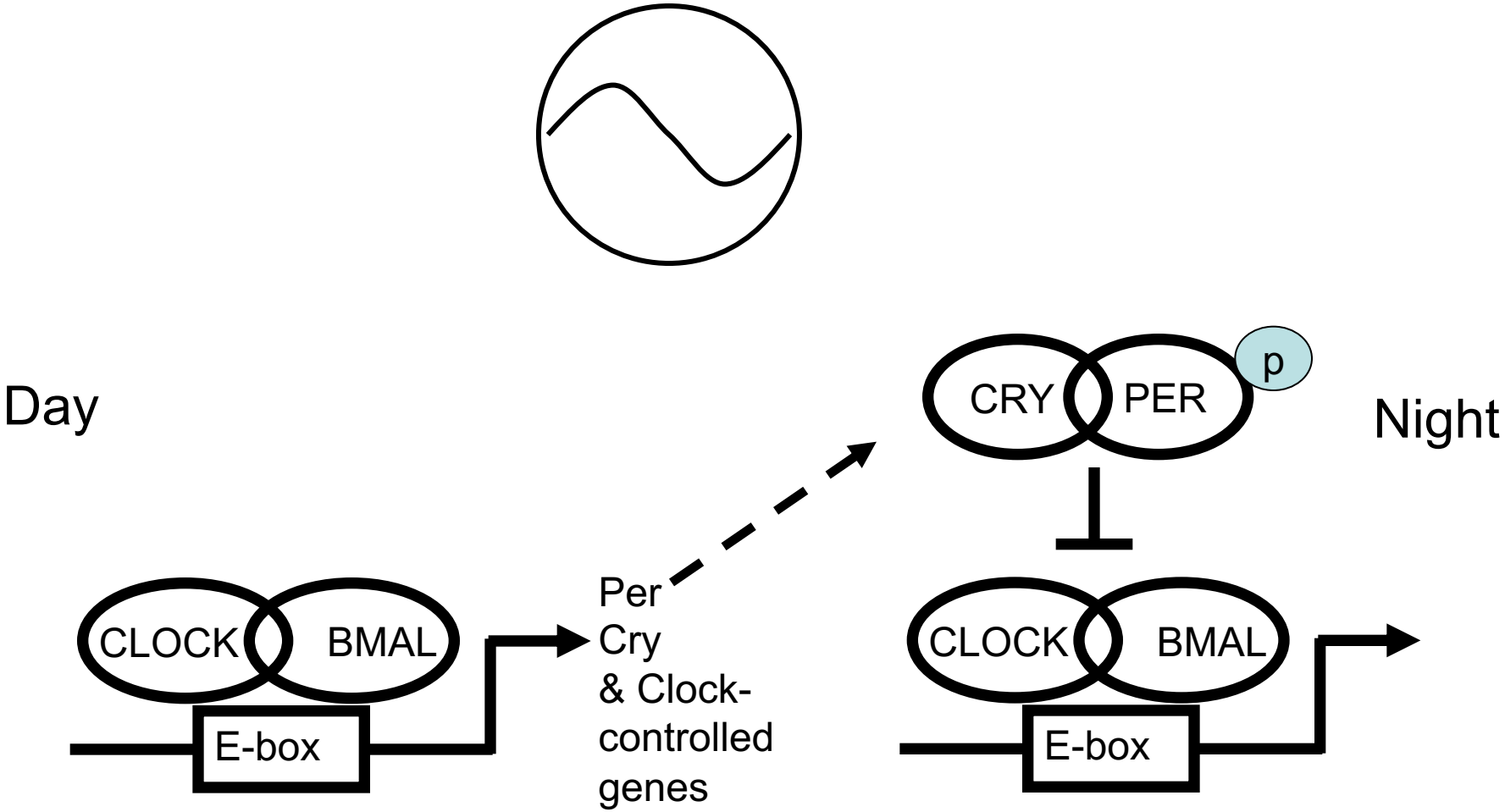




Credit: Nobel Assembly at Karolinska Institutet

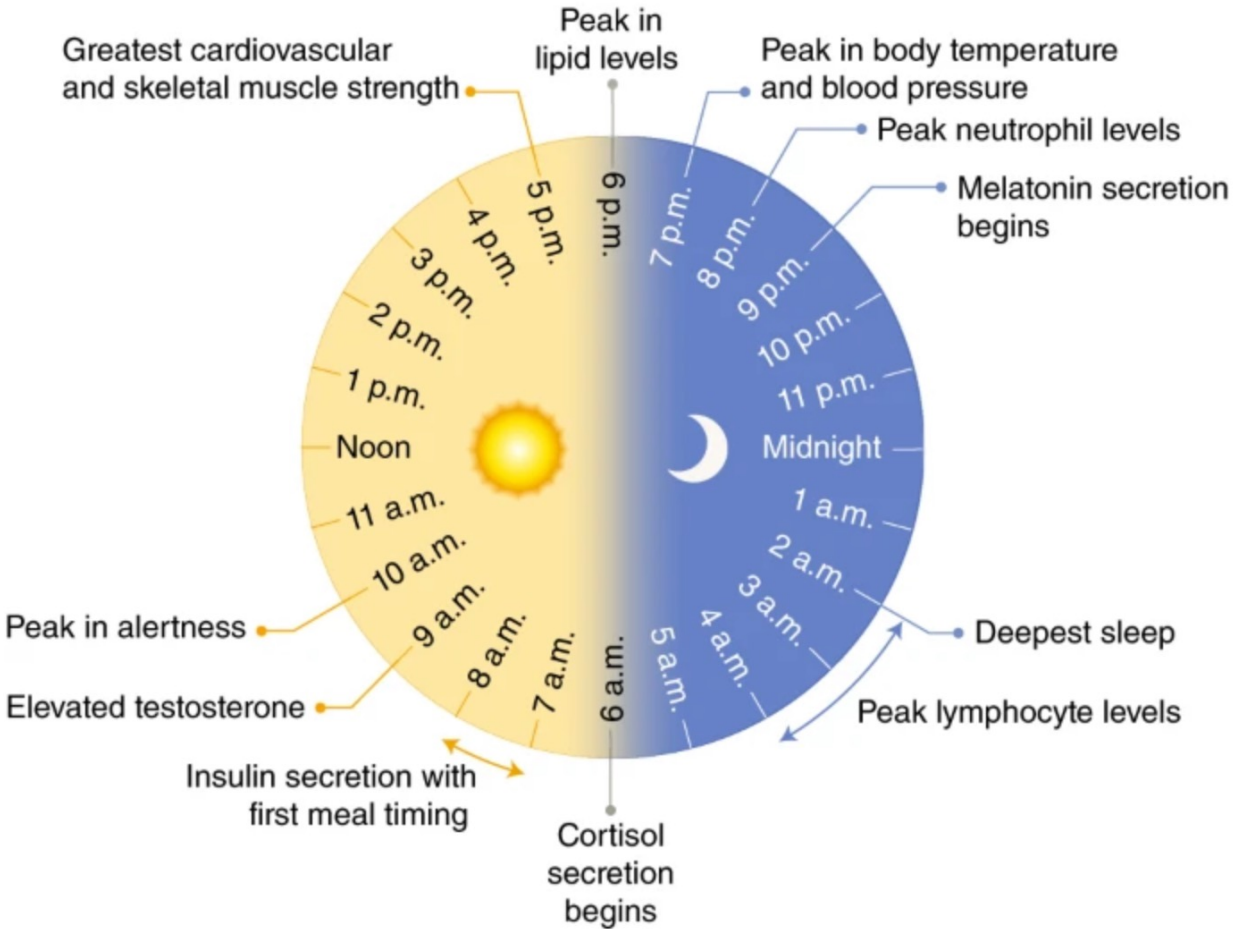
Organisms' biological clocks are controlled by oscillations in the level of PER, a protein produced from the *period* gene (1). When complexed with the TIMELESS protein, or TIM (2), PER accumulates in the nucleus, where it inhibits *period* gene activity (3).

The circadian clock in vertebrates



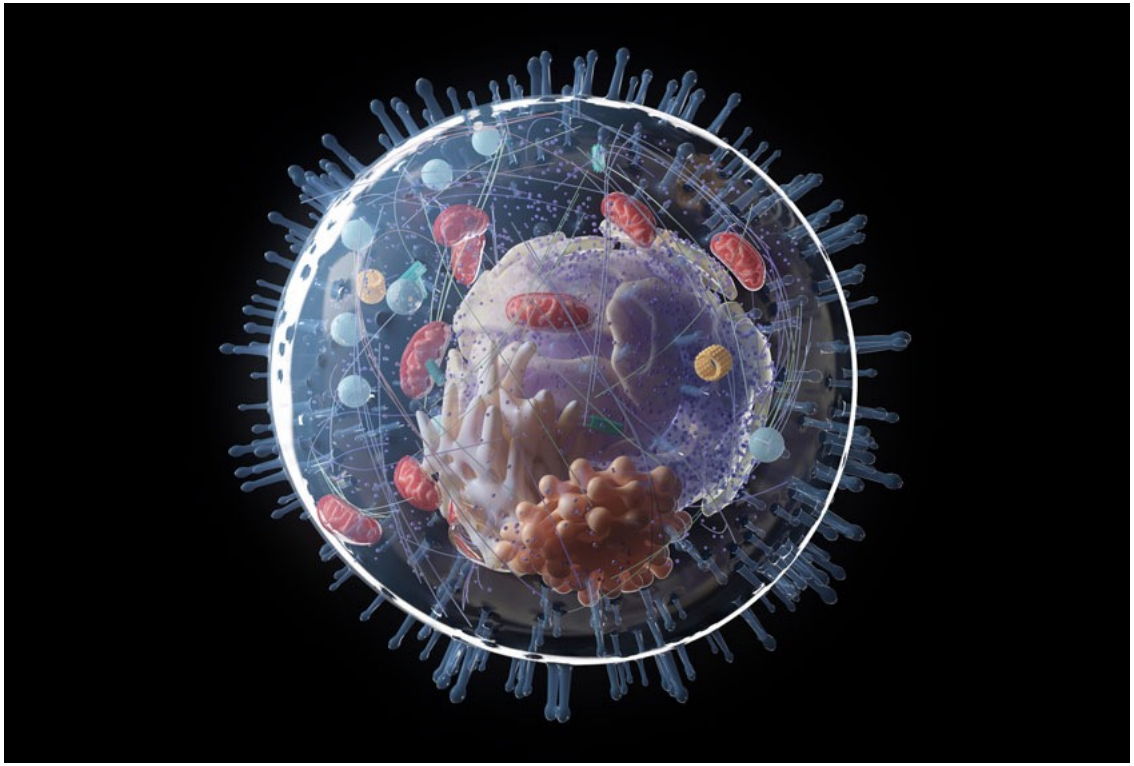
What makes circadian biology so important?

Fig. 1: The mammalian circadian clock.

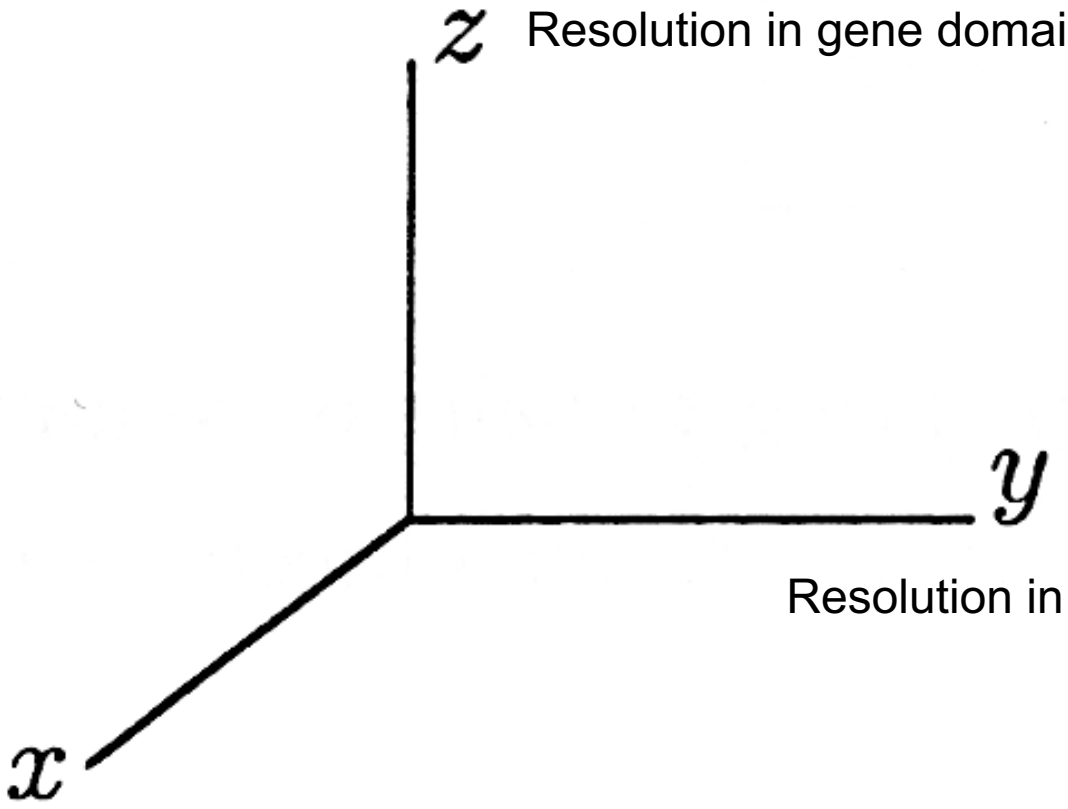


Masri & Sassone-Corsi, Nat Med, 2018

What makes circadian biology so important?



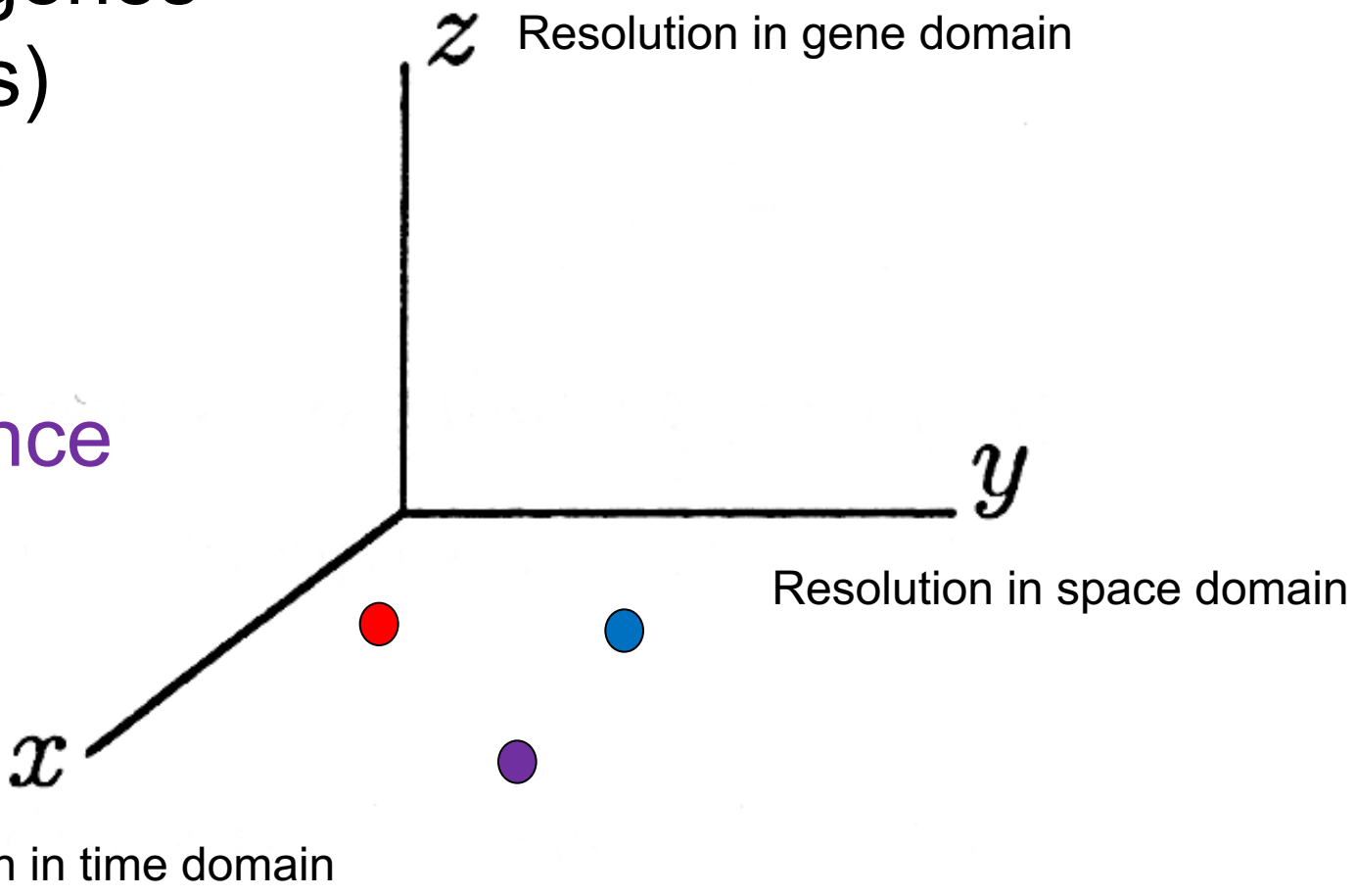
What is the ideal circadian measurement?



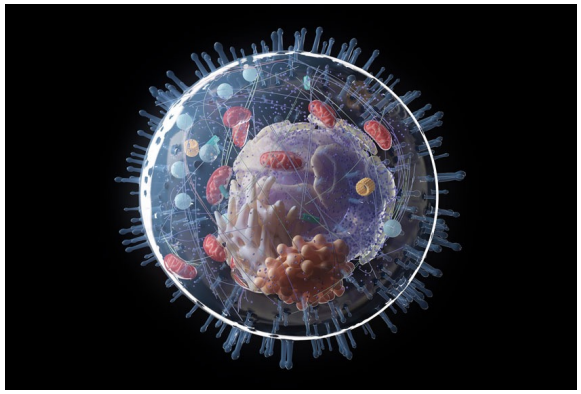
Molecular methods for measuring the circadian clock

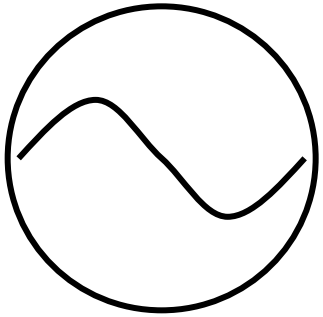
One or few genes (until ~2000s)

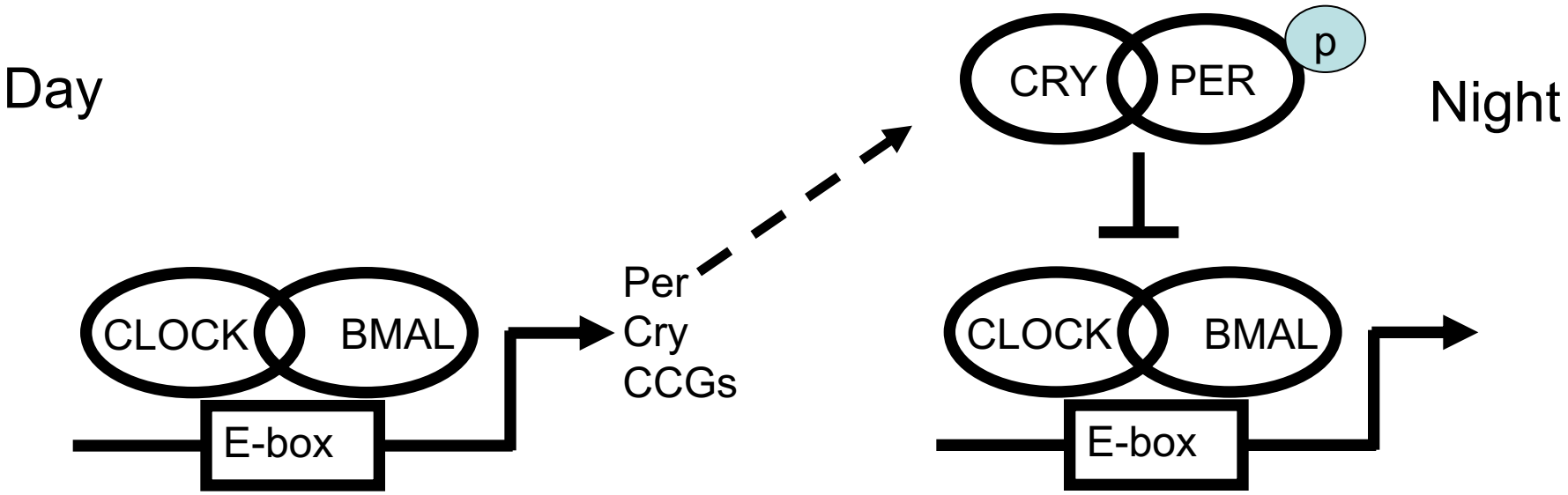
- qRT-PCR
- ISH
- Fluorescence rhythms



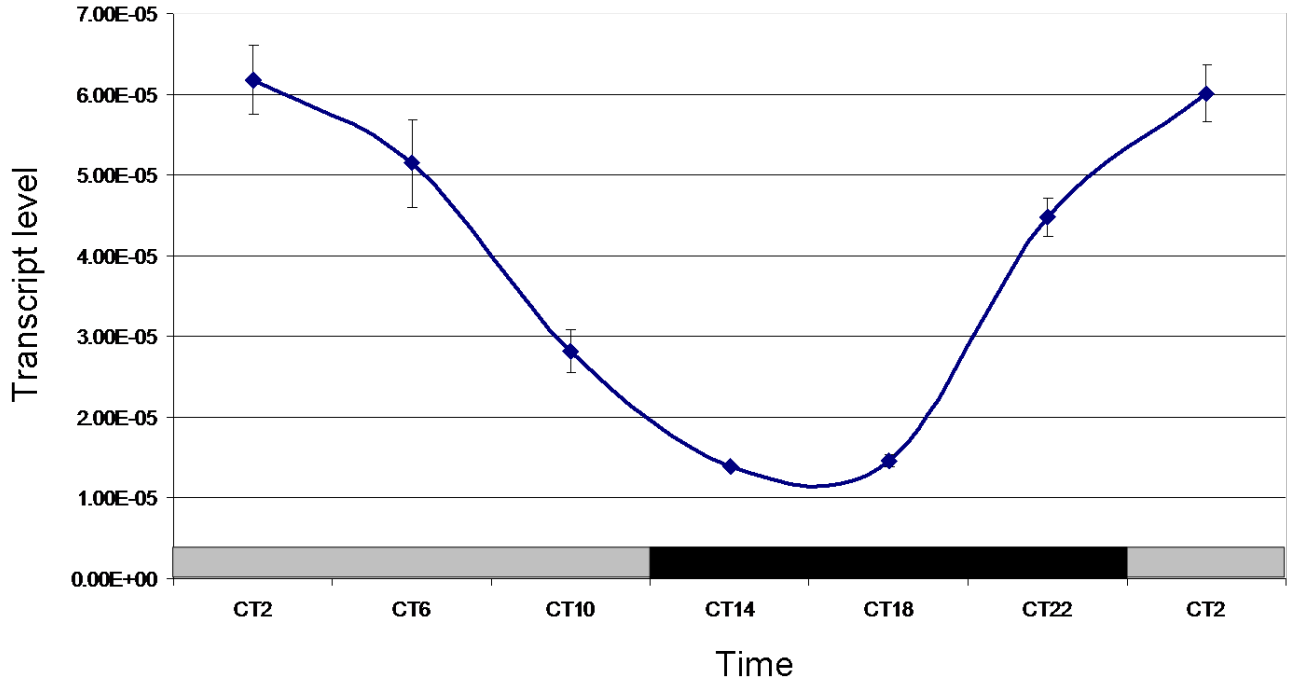
Which tissues are circadian?





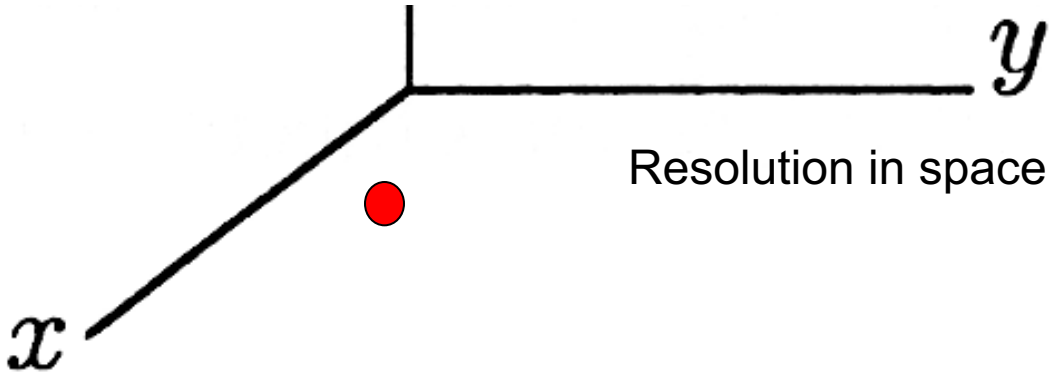


qRT-PCR



Clock gene in the zebrafish pineal

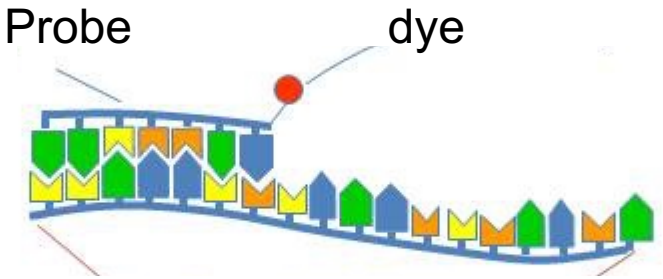
- Advantages
- Disadvantages

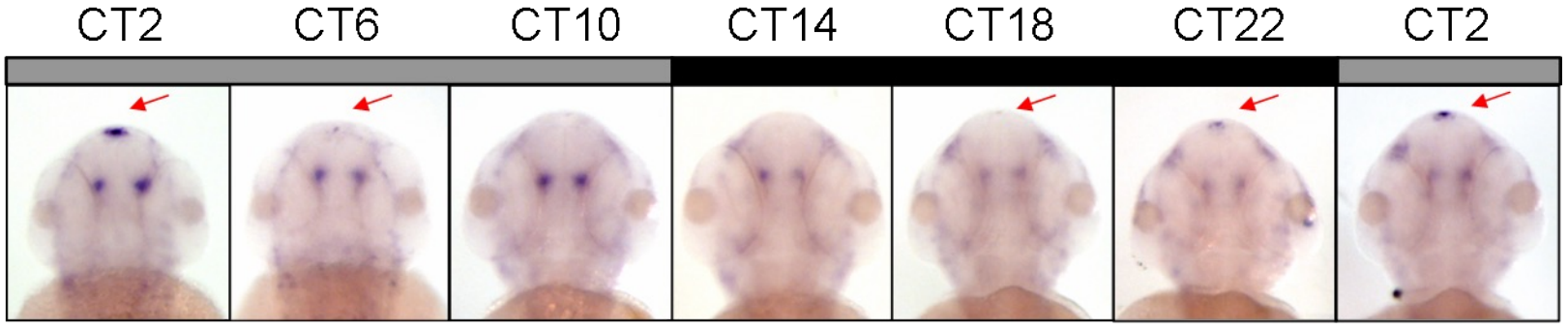


Resolution in space domain

Resolution in time domain

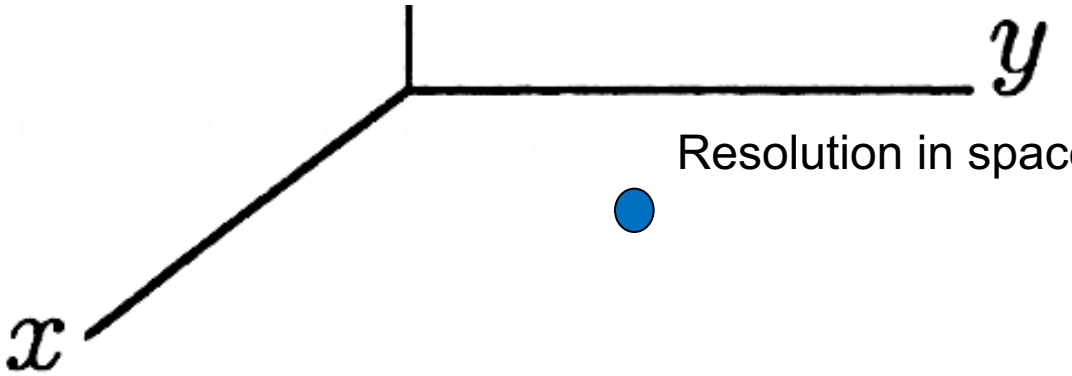
In Situ Hybridization (ISH)





Clock gene in the zebrafish pineal

- Advantages
- Disadvantages

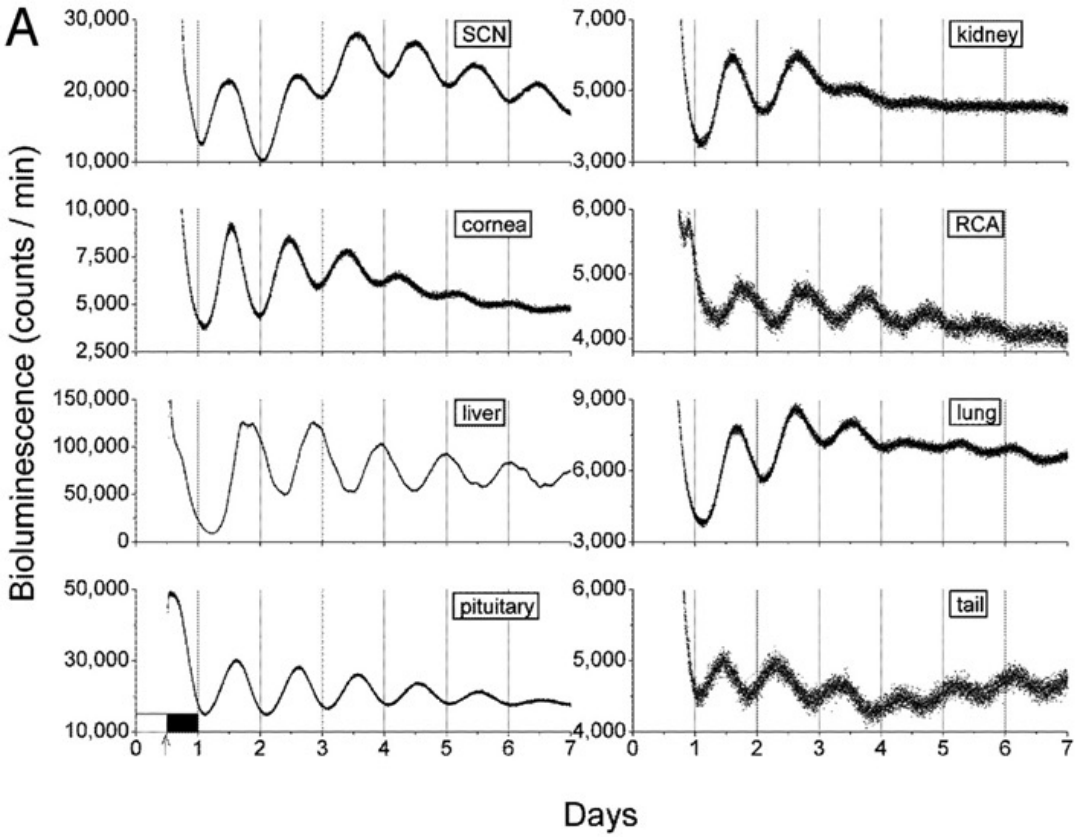


Resolution in space domain

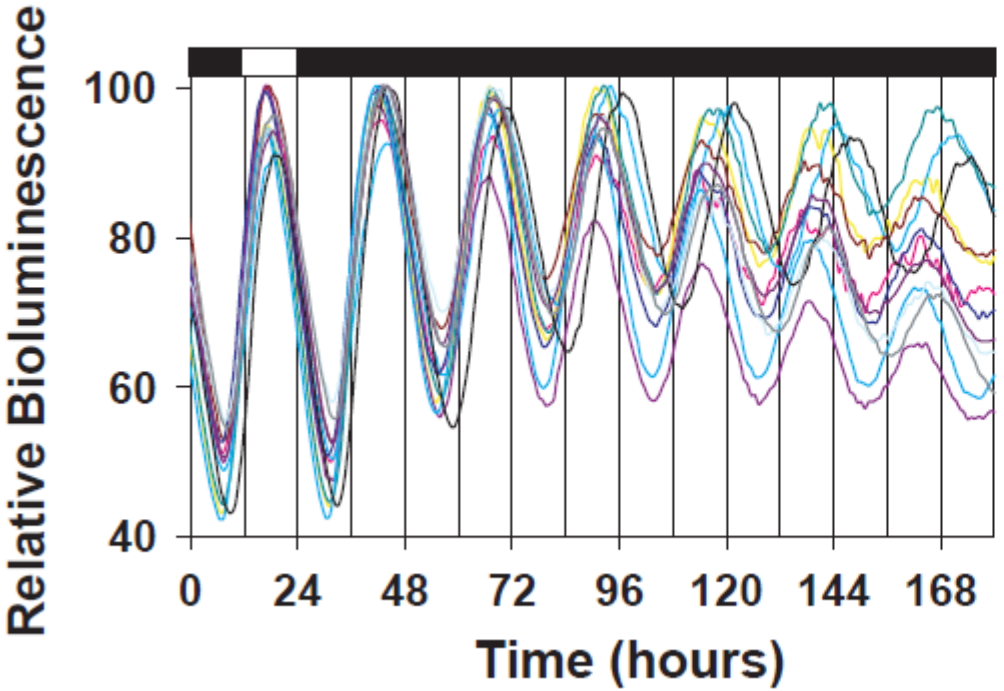
Resolution in time domain

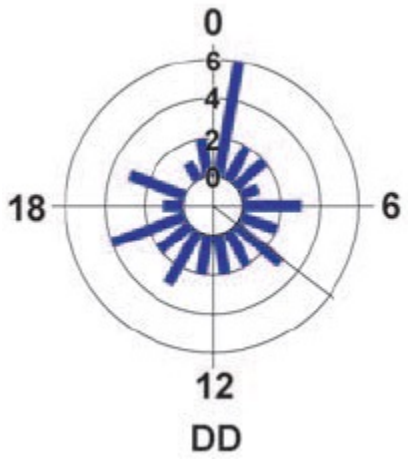
Bioluminescence and fluorescence rhythms

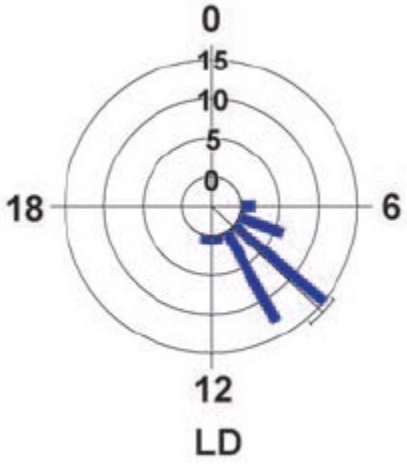
- Rhythms in all examined tissues (Yoo *et al*., PNAS, 2004)
- And if we don't see rhythm?



Single-cells measurements





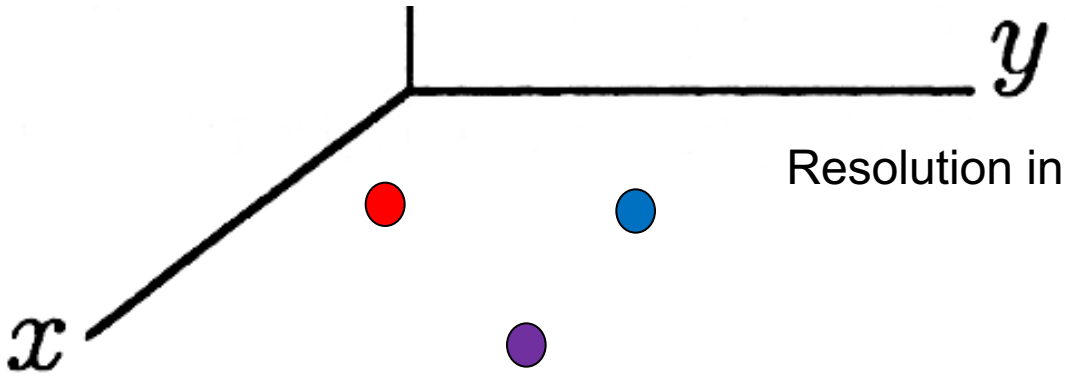


Carr and Whitmore,
Nature Cell Biology, 2005

Molecular methods for measuring the circadian clock

One or few genes (until ~2000)

- qRT-PCR
- ISH
- Bioluminescence and fluorescence rhythms

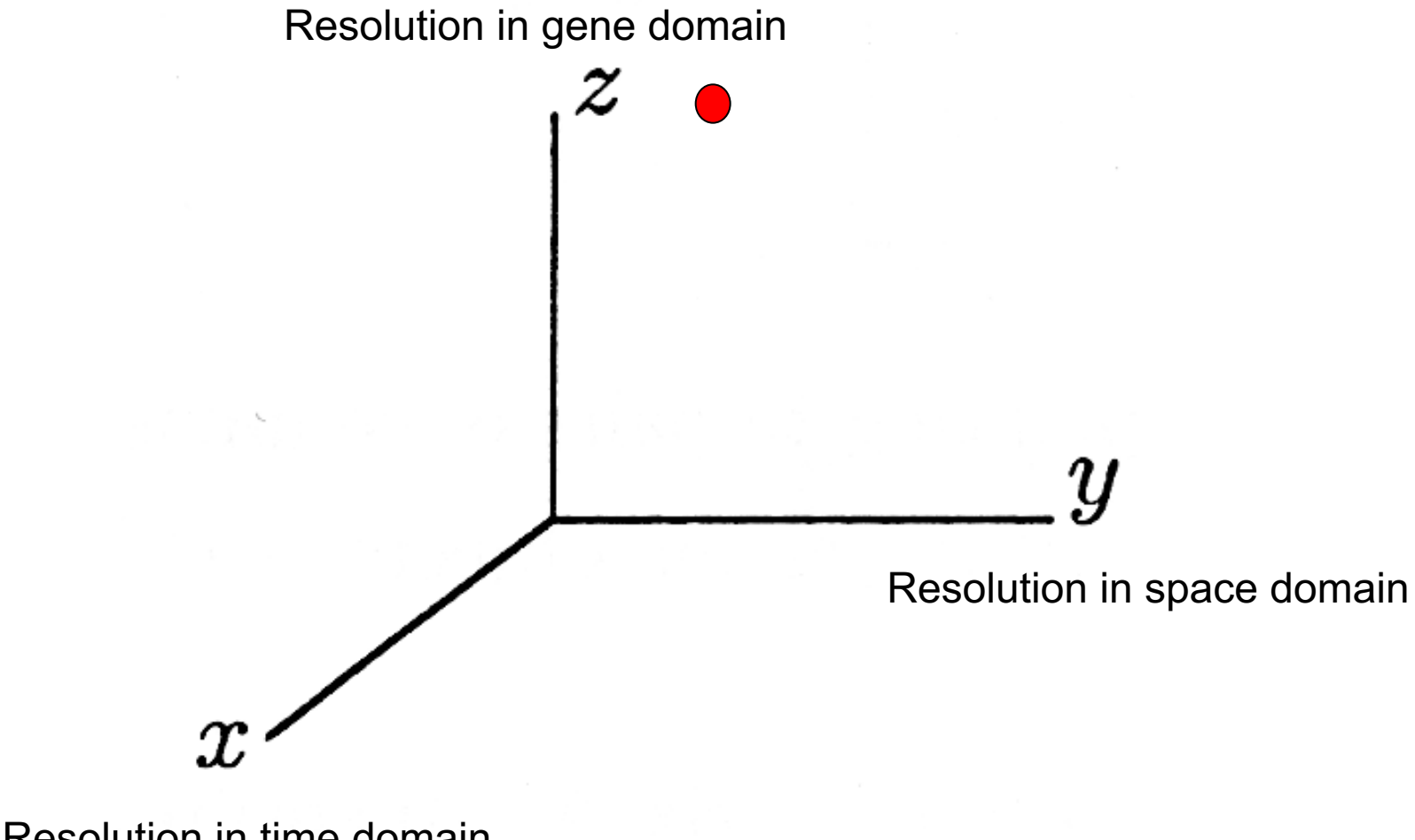


Resolution in space domain

Resolution in time domain

Molecular methods for measuring the circadian clock

- Next-generation sequencing



Next-generation sequencing of mRNA (mRNA-seq)

Tissue / whole animal

mRNA isolation

mRNA sequencing

Sequencing Reads

1) ACTAGTCAACCTCGGGGAATCA
2) TGTGACGTACACGTCACA
3) TTGGCTCCACACTGCC
4) AAACACACGTGCGTGCACG
5) GTGTCACGTGCACCACGTGTG
6) GTAAACACGTGTCTGCGTCA
7) TGTTGACGTAACACACTGT
8) …
9) …
…
200,000,000) …
Each read is ~100 bases

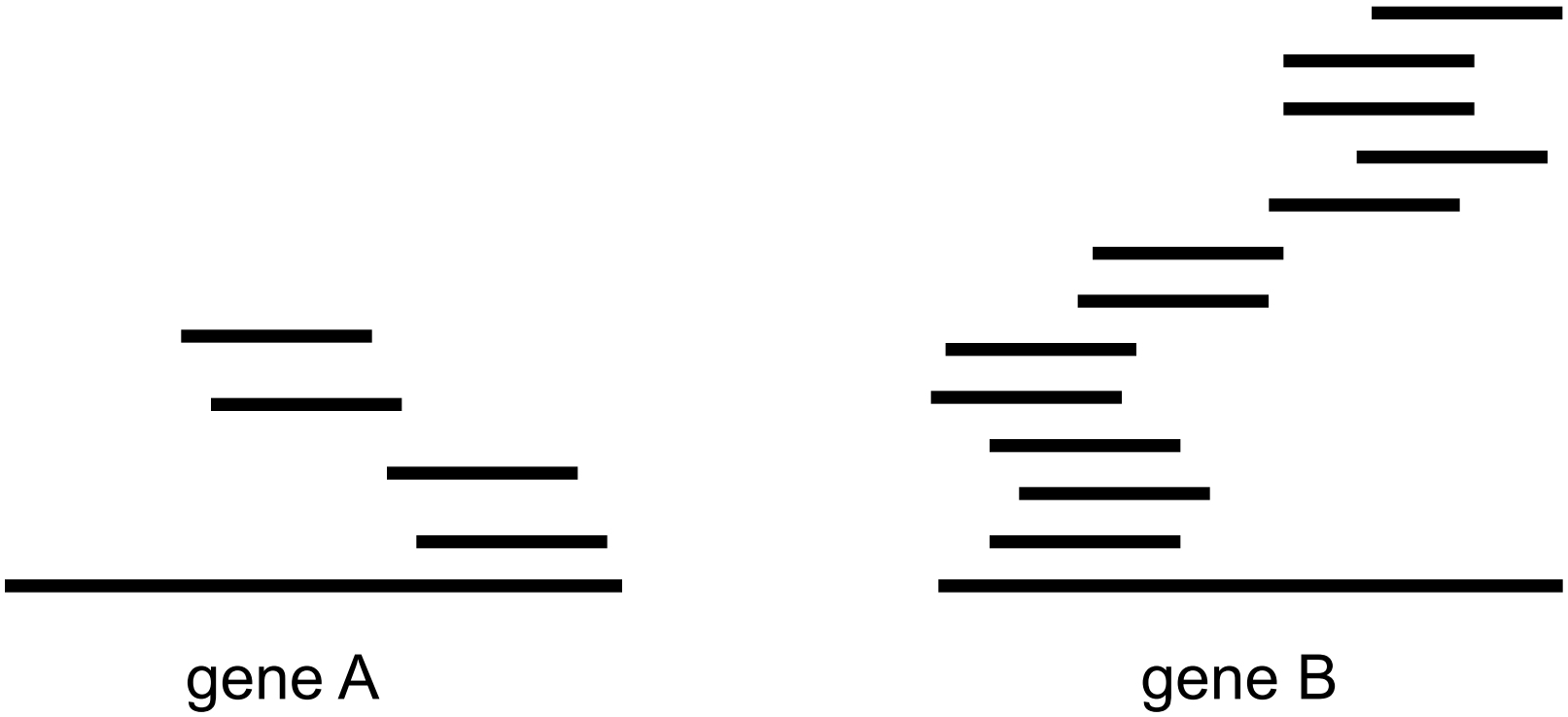
mRNA-seq – measuring gene expression

Sequencing reads

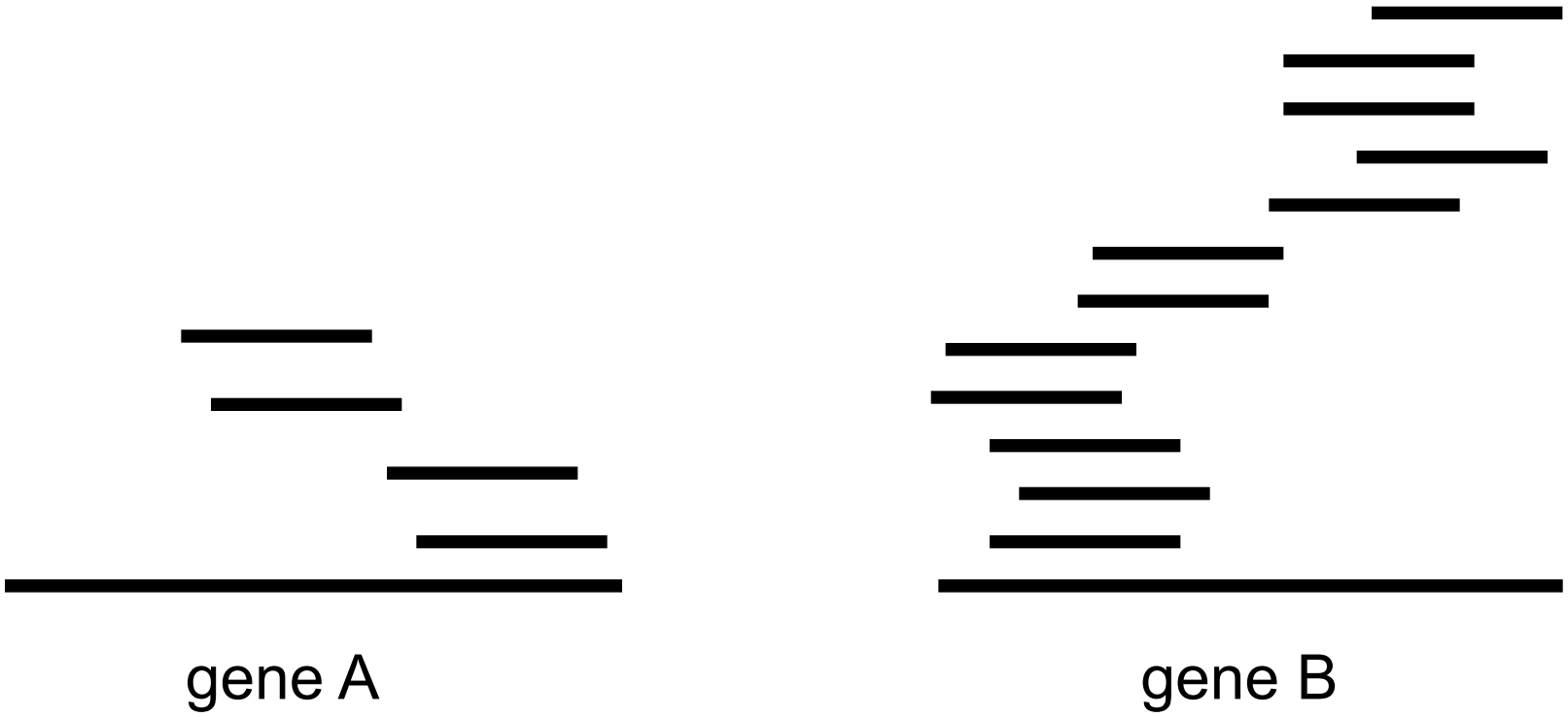
```
AGTCTTCCTCGA
   CTTCCTCGAGATA
          GAGATACGATA
                 ATATACCGCC
                        CCATTTAGT
                             TAGTTTTTTGAG
                                    TGAGAGACG
                                         ACGCGCAGAGA
                                               GAGAGGA

AGTCTTCCTCGAGATACGATATACCGCCCATTTAGTTTTTGAGAGACGCGCAGAGAGGA
```

mRNA sequence

mRNA-seq – measuring gene expression



mRNA-seq – measuring gene expression

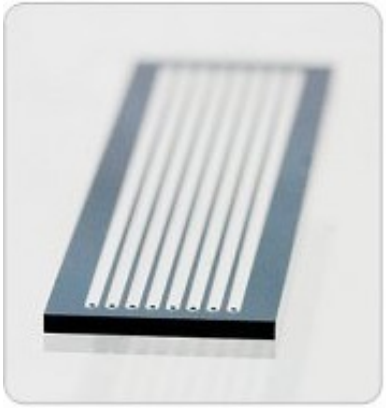


- The number of reads for each gene is proportional to the gene expression level and ?

Measuring genome-wide gene expression

Next-generation sequencing

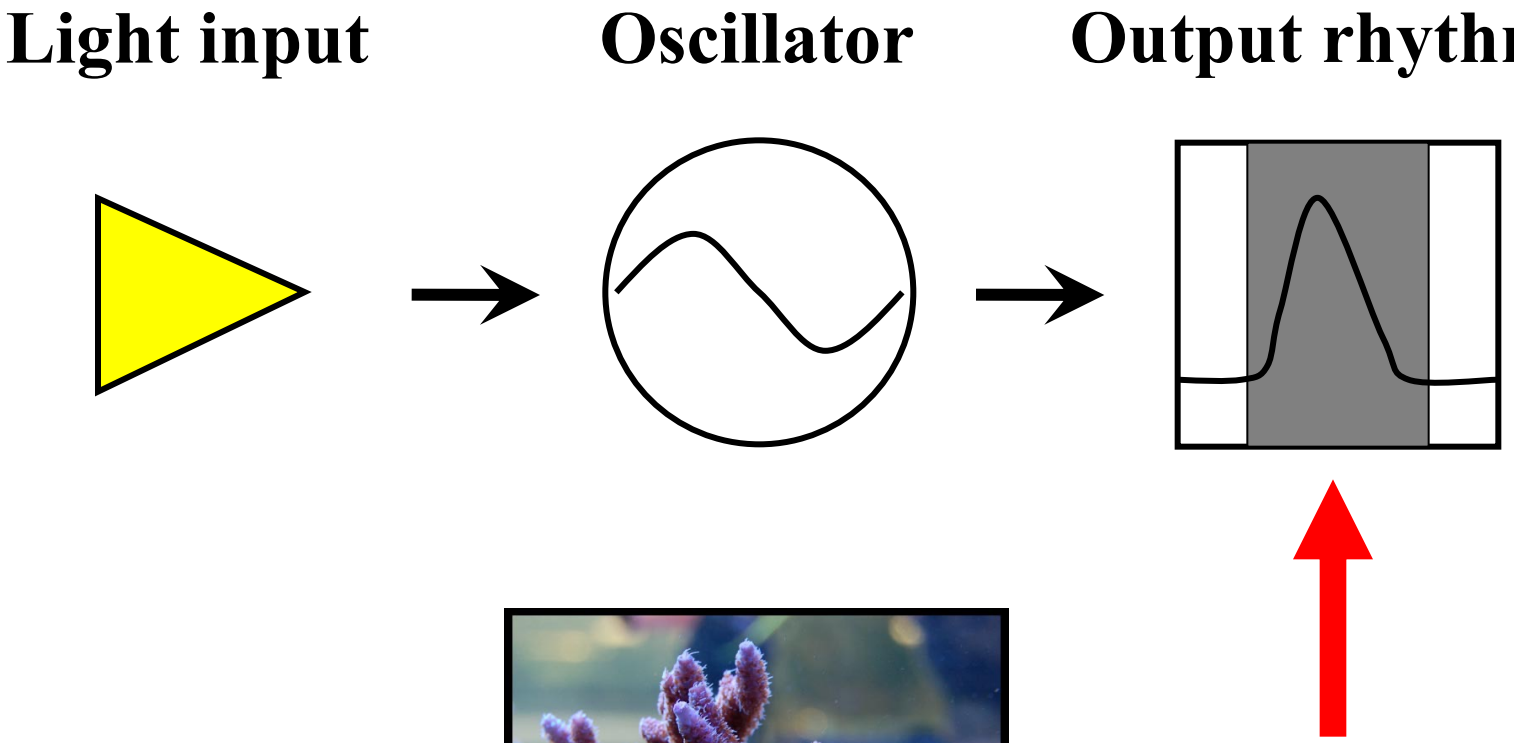
FIGURE 1: ILLUMINA GENOME ANALYZER FLOW CELL



Up to eight samples can be loaded onto the flow cell for simultaneous analysis on the Illumina Genome Analyzer.

- Simultaneous measurements of the RNA levels of all the genes!

Which genes are clock-controlled?



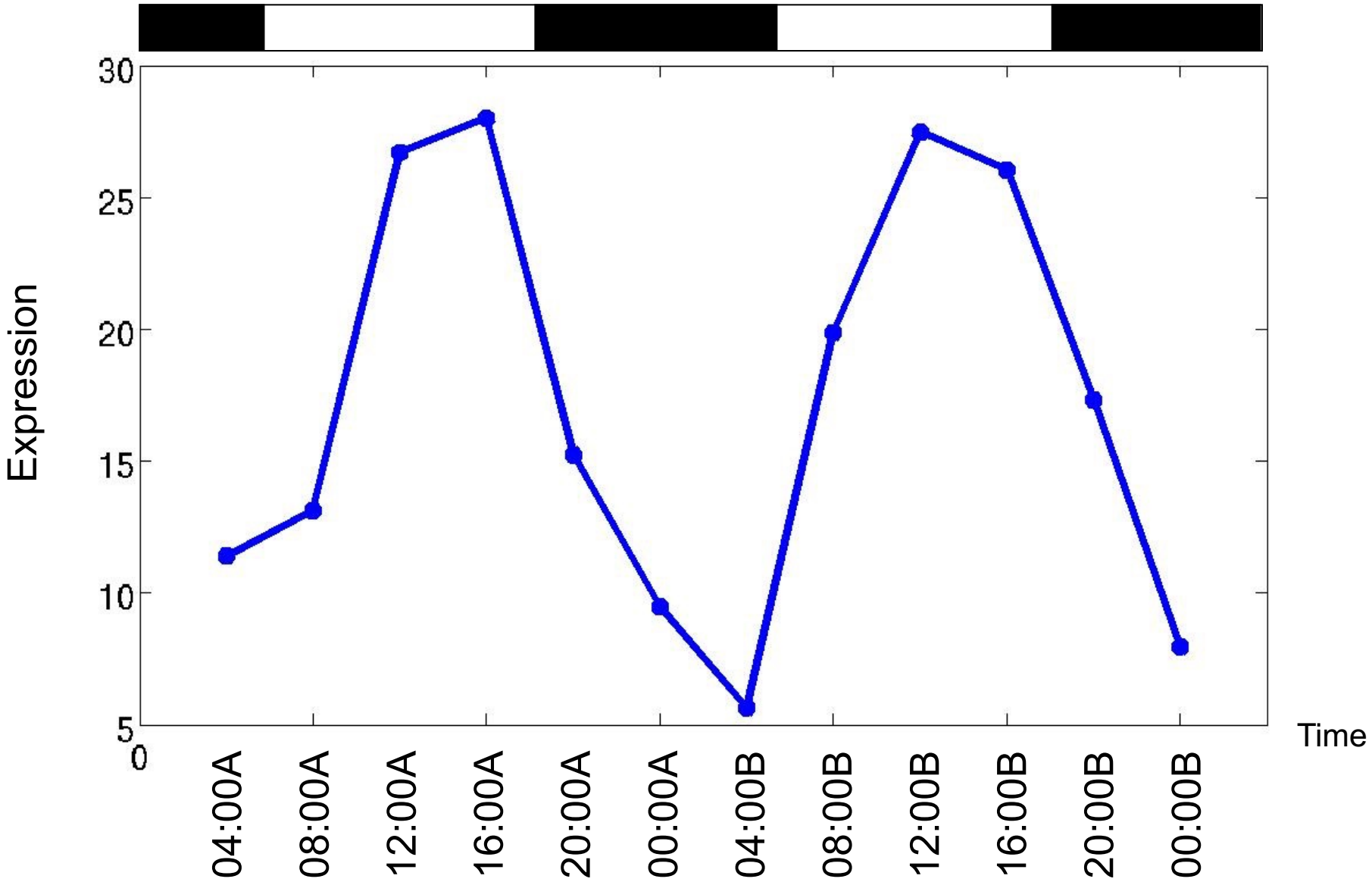
Experimental design

- One sample at night (24:00),
  one sample at day (12:00)
- We will measure the expression levels of all the genes and look for changes
- What can we lose ?

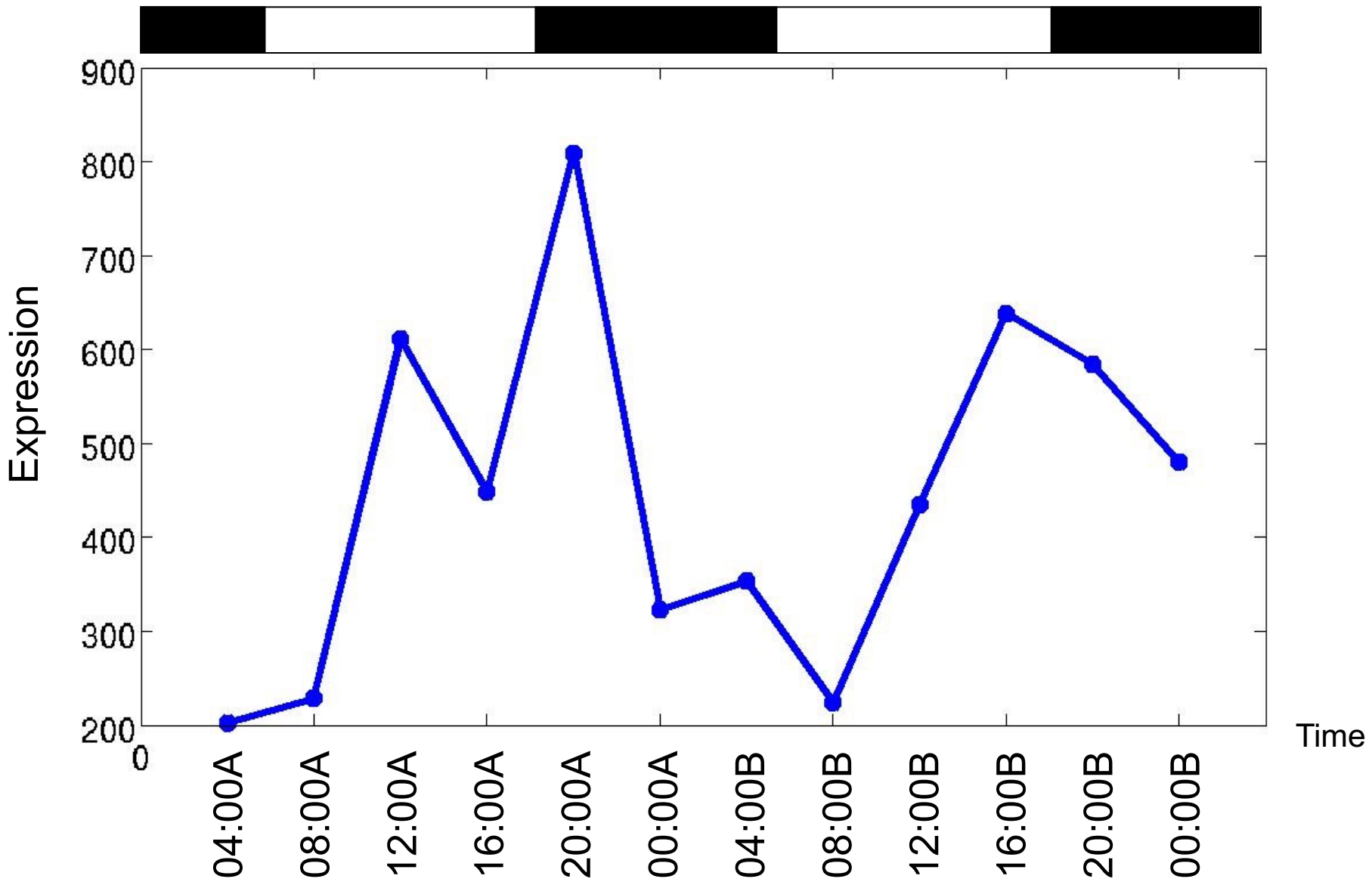
Experimental design

- One sample at night (24:00),
  one sample at day (12:00)
- We will measure the expression levels of all the genes and look for changes
- What can we lose ?
- One sample every 4-hr through the daily cycle
- Measure two consecutive days

Gene A

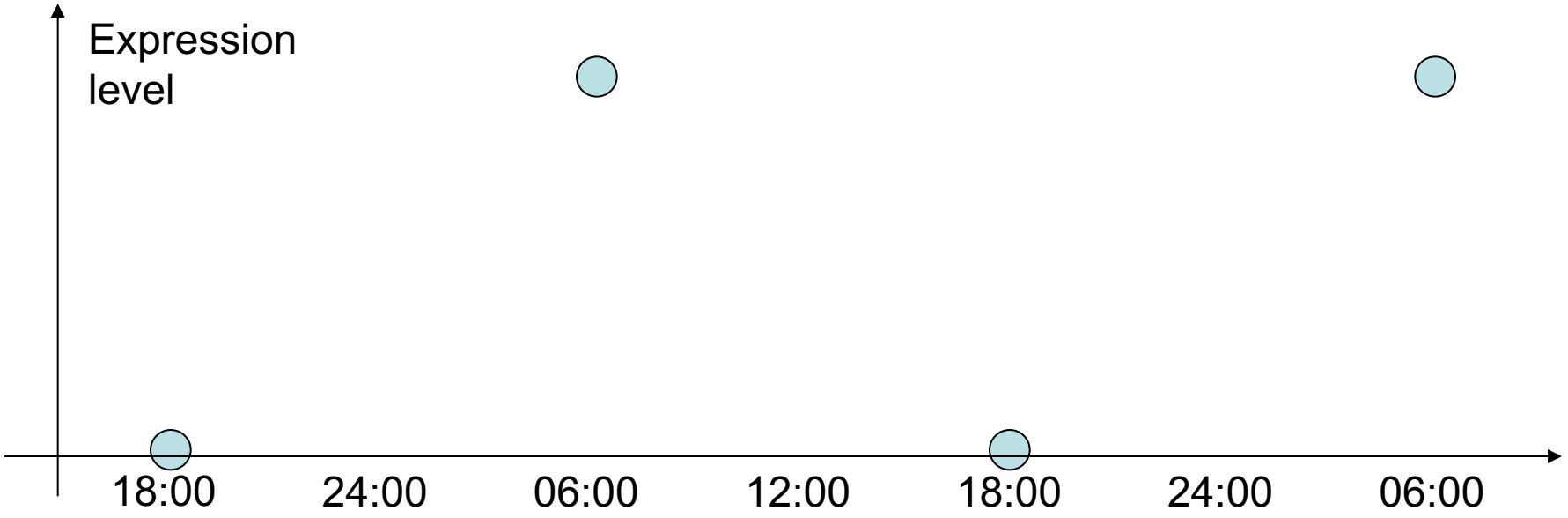

30
25
20
15
10
5
0
Expression
04:00A
08:00A
12:00A
16:00A
20:00A
00:00A
04:00B
08:00B
12:00B
16:00B
20:00B
00:00B
Time

Gene B


900
800
700
600
500
400
300
200
0
Expression
Time
04:00A
08:00A
12:00A
16:00A
20:00A
00:00A
04:00B
08:00B
12:00B
16:00B
20:00B
00:00B

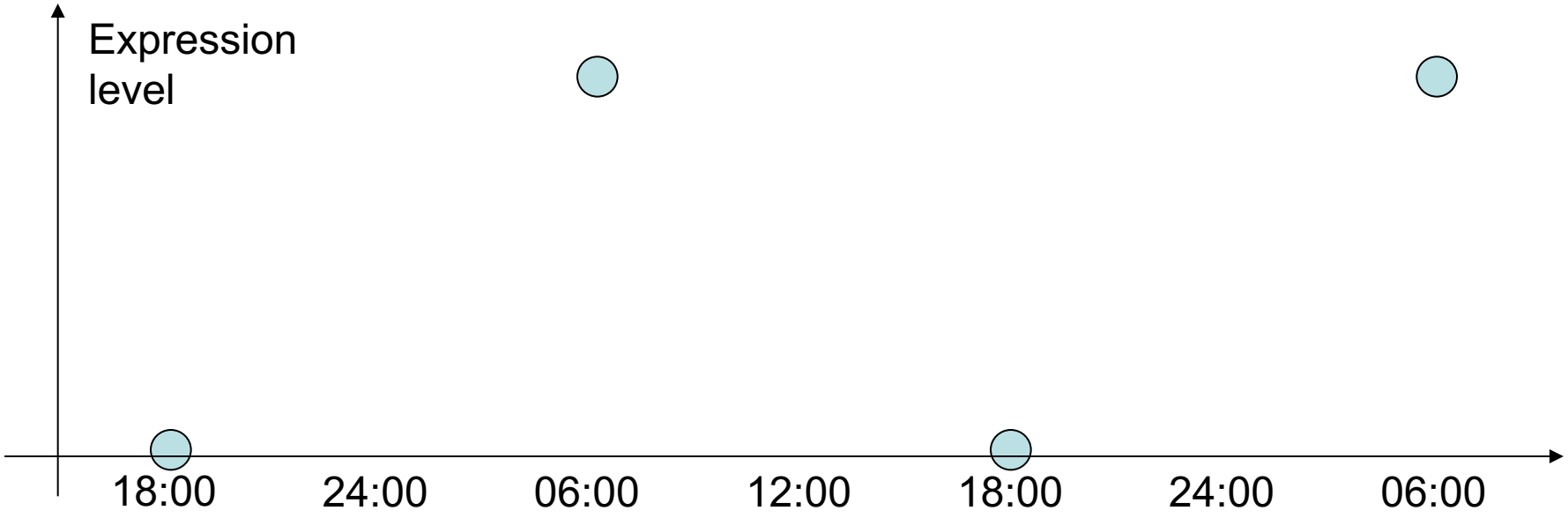
Identifying circadian genes

- How ‘circadian’ looks like?



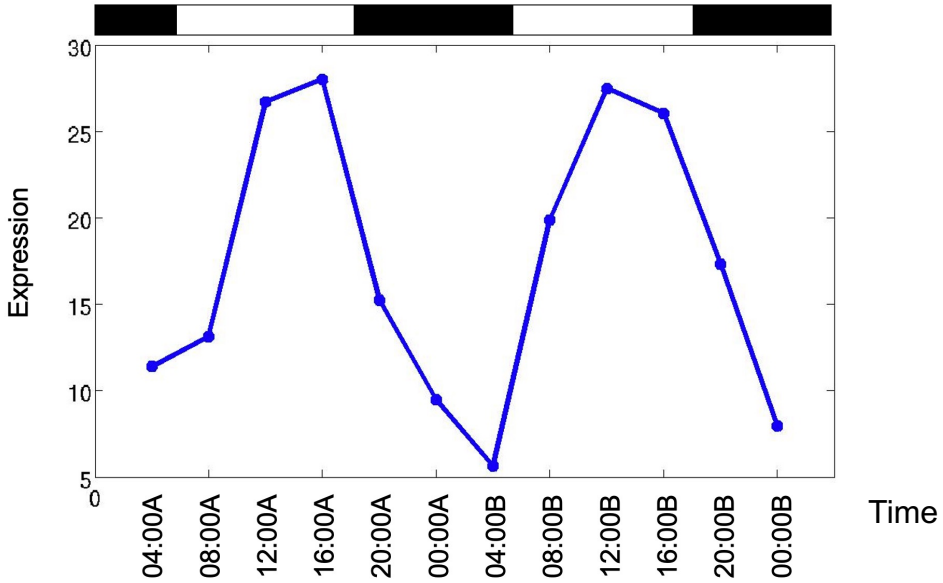
Identifying circadian genes

- How ‘circadian’ looks like?



- The ‘visual inspection’ method
- Time domain (fit to sine) and frequency domain (Fourier analysis)

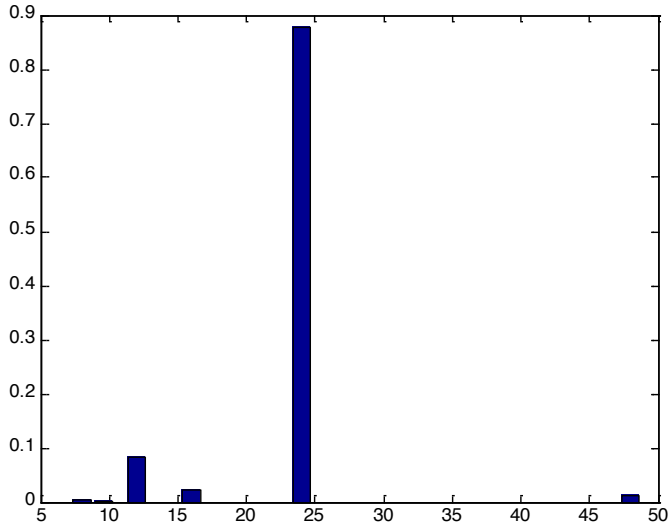
Identifying circadian genes



G-factor

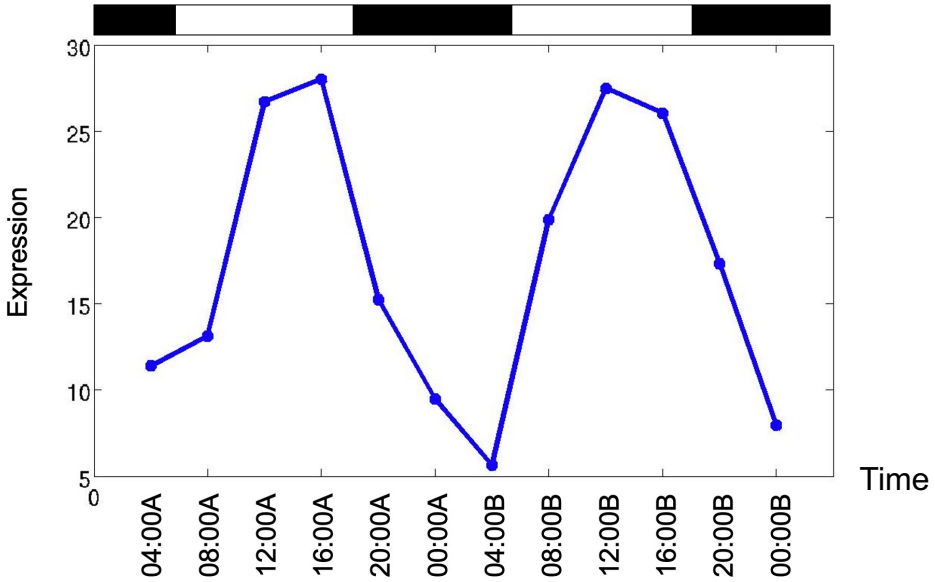
Fourier analysis

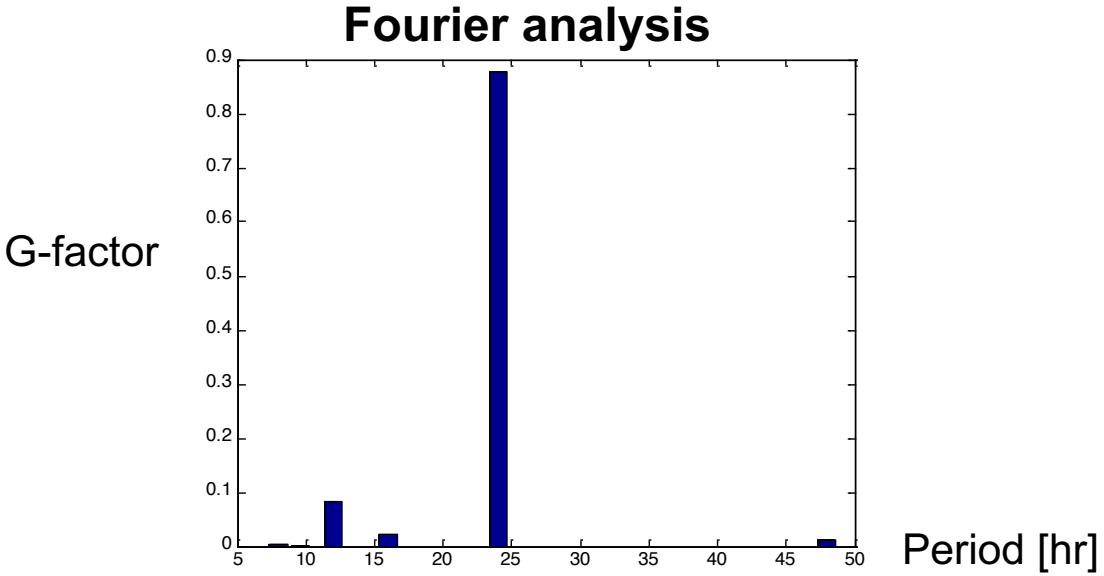
$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi}{N}kn} \qquad \text{(Eq.1)}$$
$$= \sum_{n=0}^{N-1} x_n \cdot \left[\cos\left(\frac{2\pi}{N}kn\right) - i \cdot \sin\left(\frac{2\pi}{N}kn\right)\right],$$



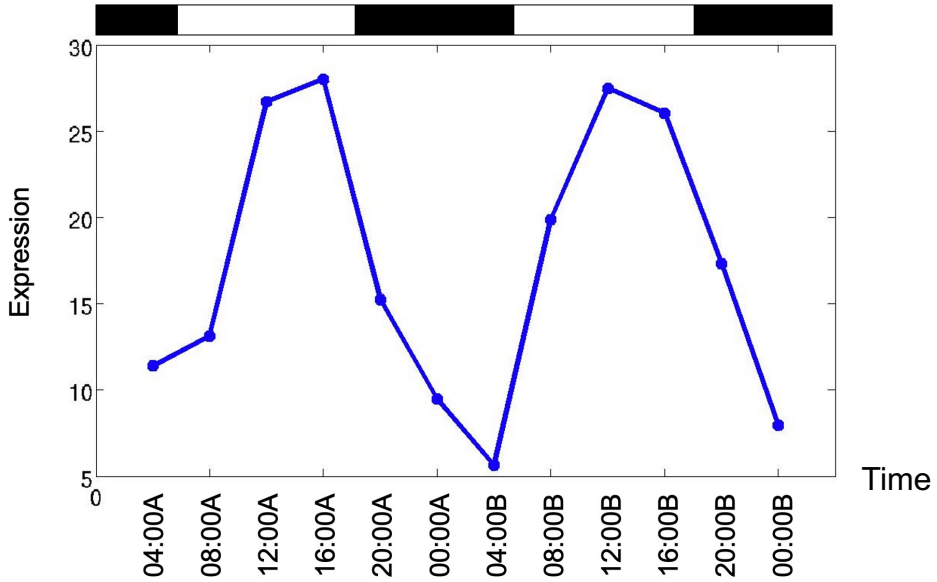
G-factor = Fourier intensity in one frequency divided by the sum of all intensities

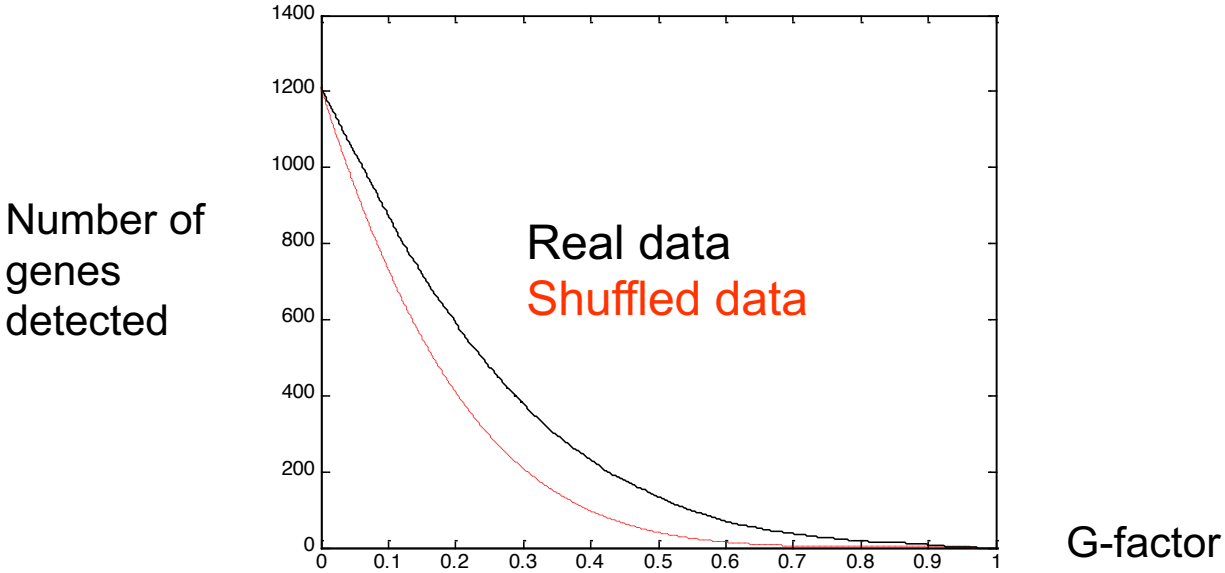
How can we know the statistical significance of a given G-factor?



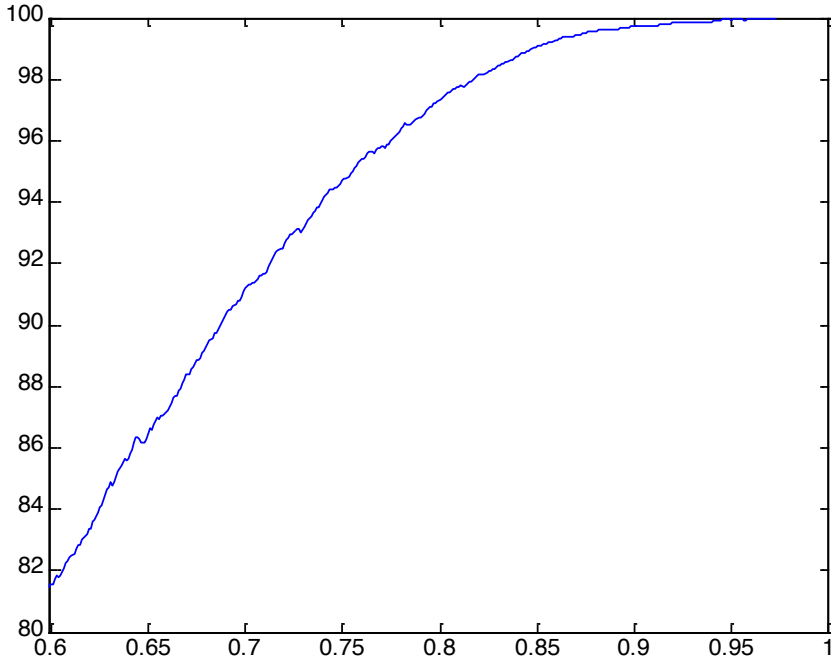


How can we know the statistical significance of a given G-factor?

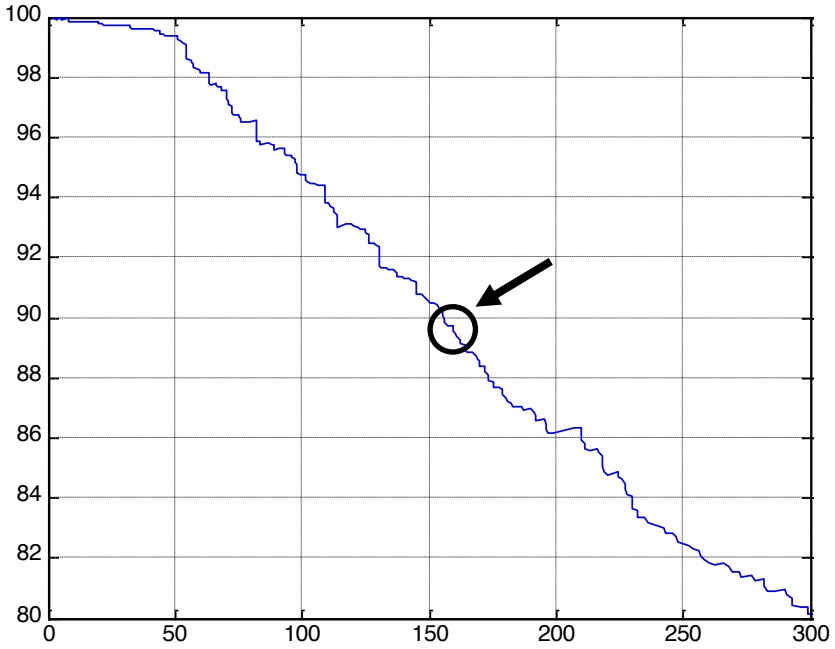




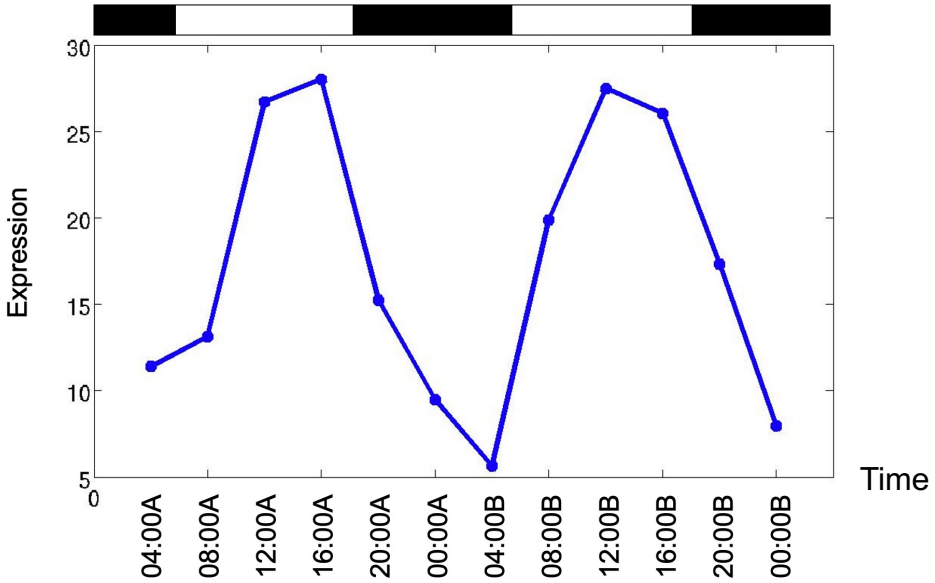
True-Positive [%] versus G-factor

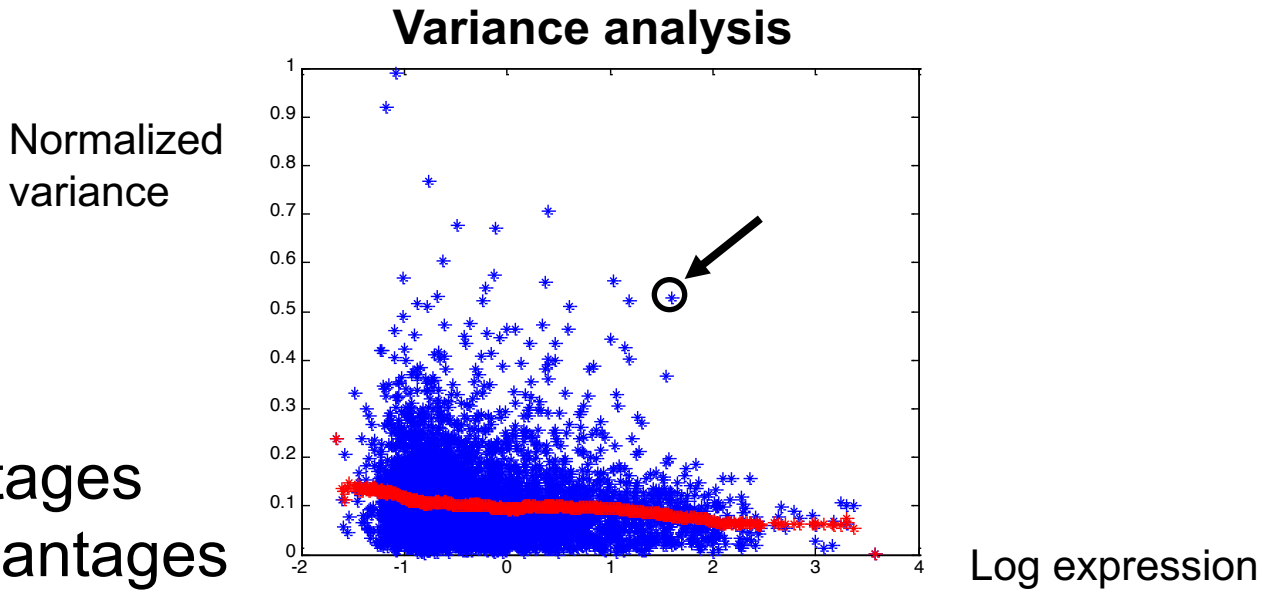


True-Positives [%] versus list length



Variance analysis





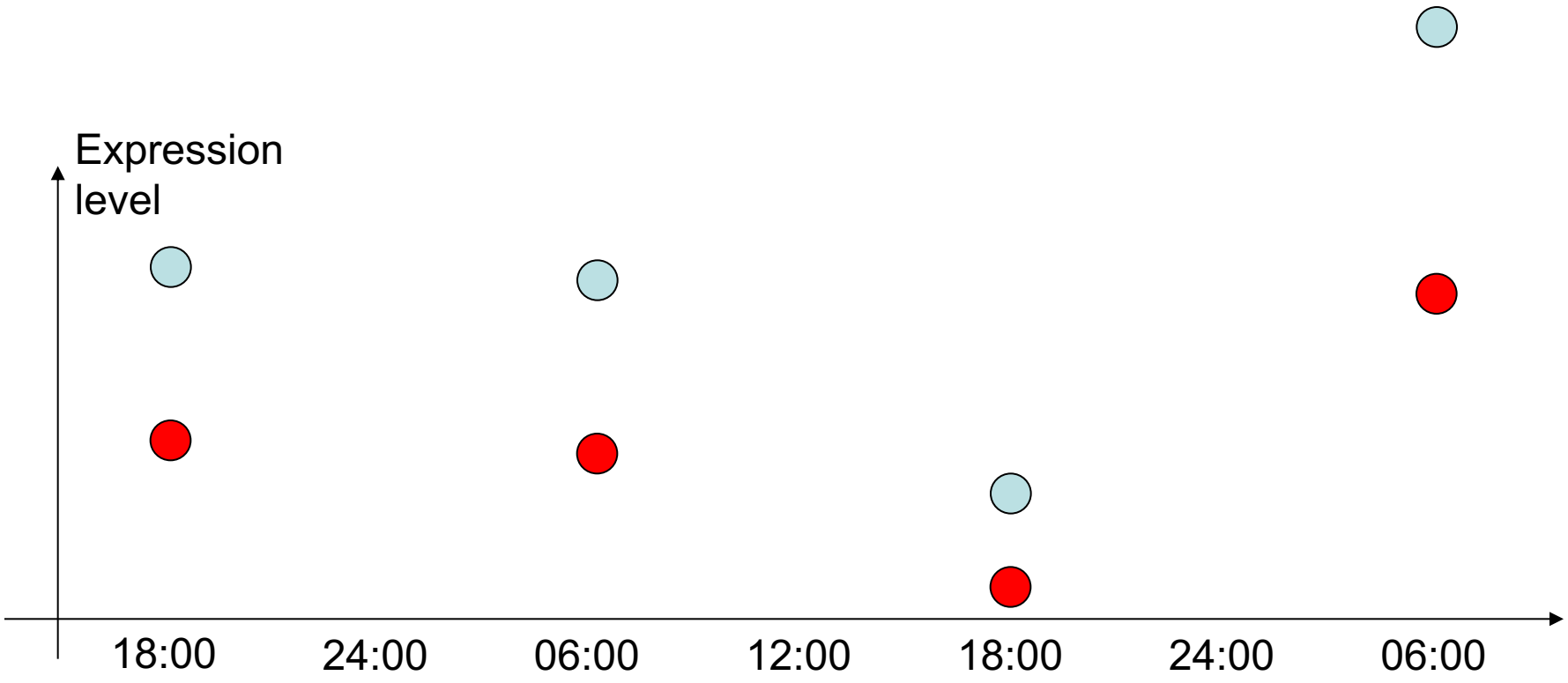
- Advantages
- Disadvantages

We got tens of circadian genes, now what?

G-factor	Gene symbol	Gene name	Peak time	Annotation
0.7833	GS01YF04	novel	4:00	little similarity found
0.7756	B033-H1	novel	4:00	splicing factor, arginine/serine-rich, marked as splicing factor, arginine/serine-rich B (sfrs-b)
0.7753	B043-D8	novel	4:00	THO complex 4 (chaperone, binds mRNA and process it by splicing and export)
0.7706	B027-B1	heat shock protein 90	16:00	
0.7696	D024-F7	GRP94 (hsp90)(gp96)	16:00	
0.762	A008-E3	calreticulin	16:00	(binds melatonin in the cell nuclear, has chaperone activity)
0.7616	B026-B3	novel	4:00	splicing factor, arginine/serine-rich, marked as splicing factor, arginine/serine-rich B (sfrs-b)
0.7567	D015-F1	heat shock protein 90	16:00	
0.7559	B035-F2	GRP94 (hsp90)(gp96)	16:00	
0.7554	A004-E3	novel	4:00	no similarity found
0.7447	A031-H7	heat shock protein 90	16:00	
0.7443	C003-G9	NA	4:00	NA
0.7424	A049-A7	novel	16:00	little similarity found
0.7344	A032-E10	NA	20:00	

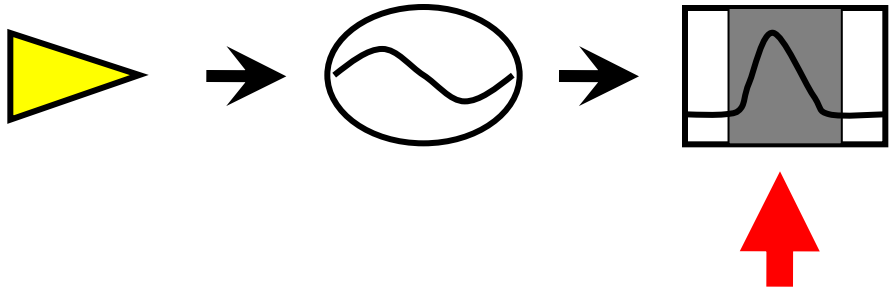
The clustering concept

- Eisen et al., PNAS, 1998
- ~20,000 citations
- What is ‘similar’?
- Pearson definition

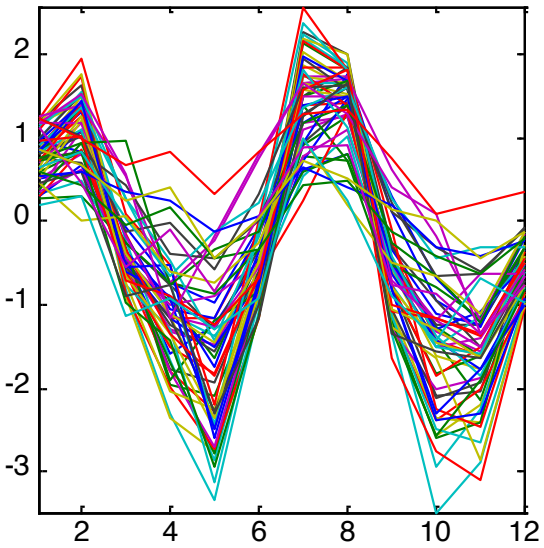


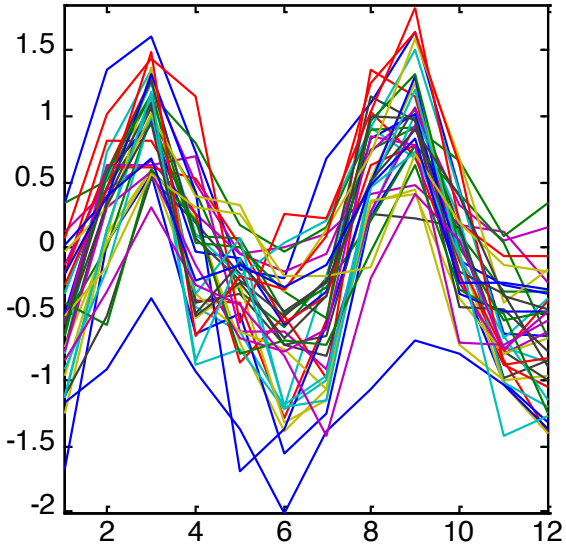
Research question

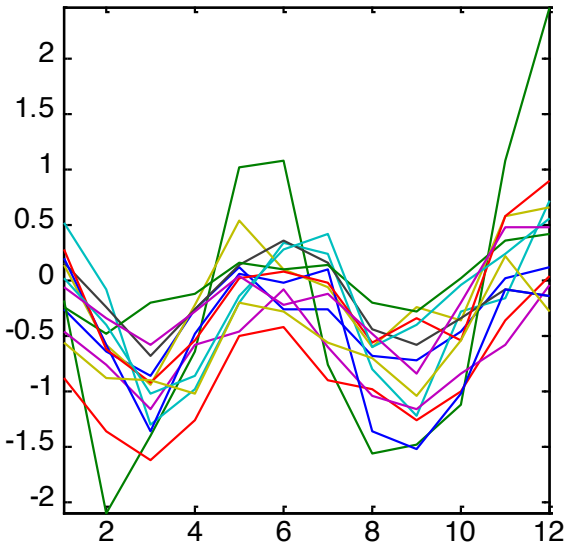
- Clock-controlled genes?

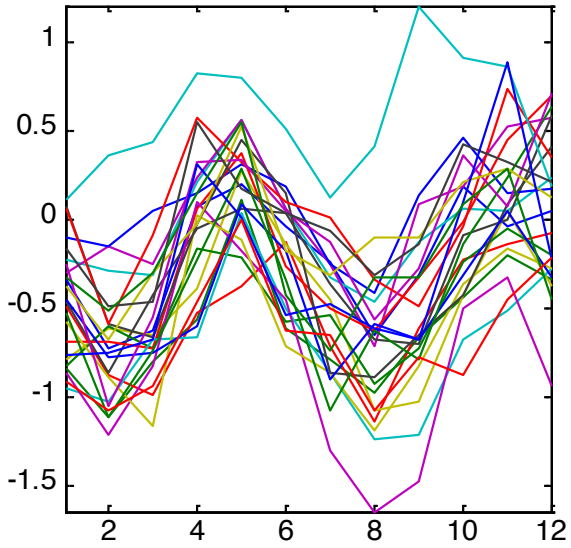


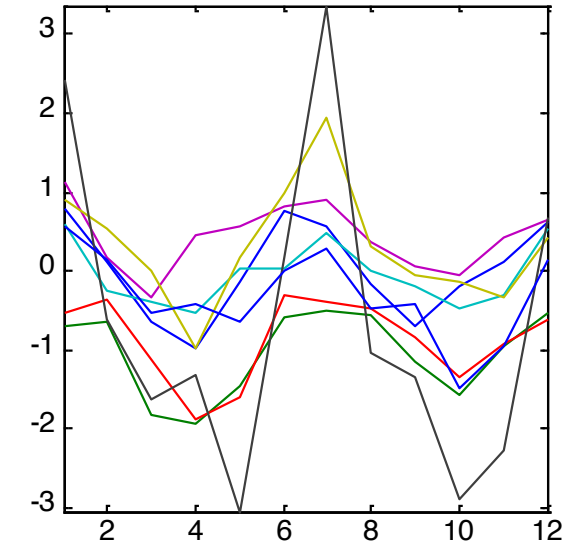
K-means clustering

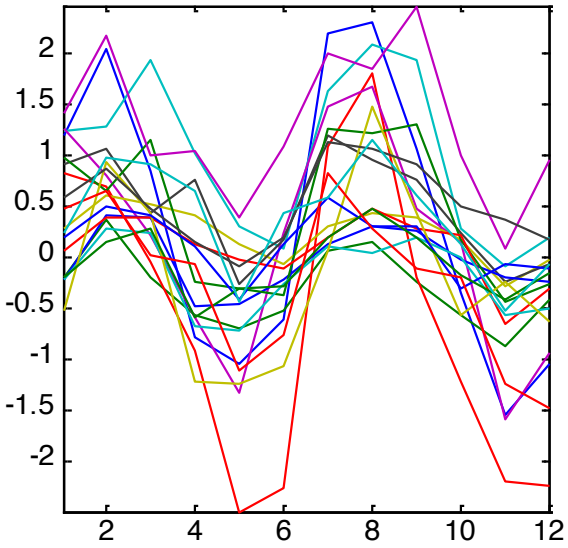




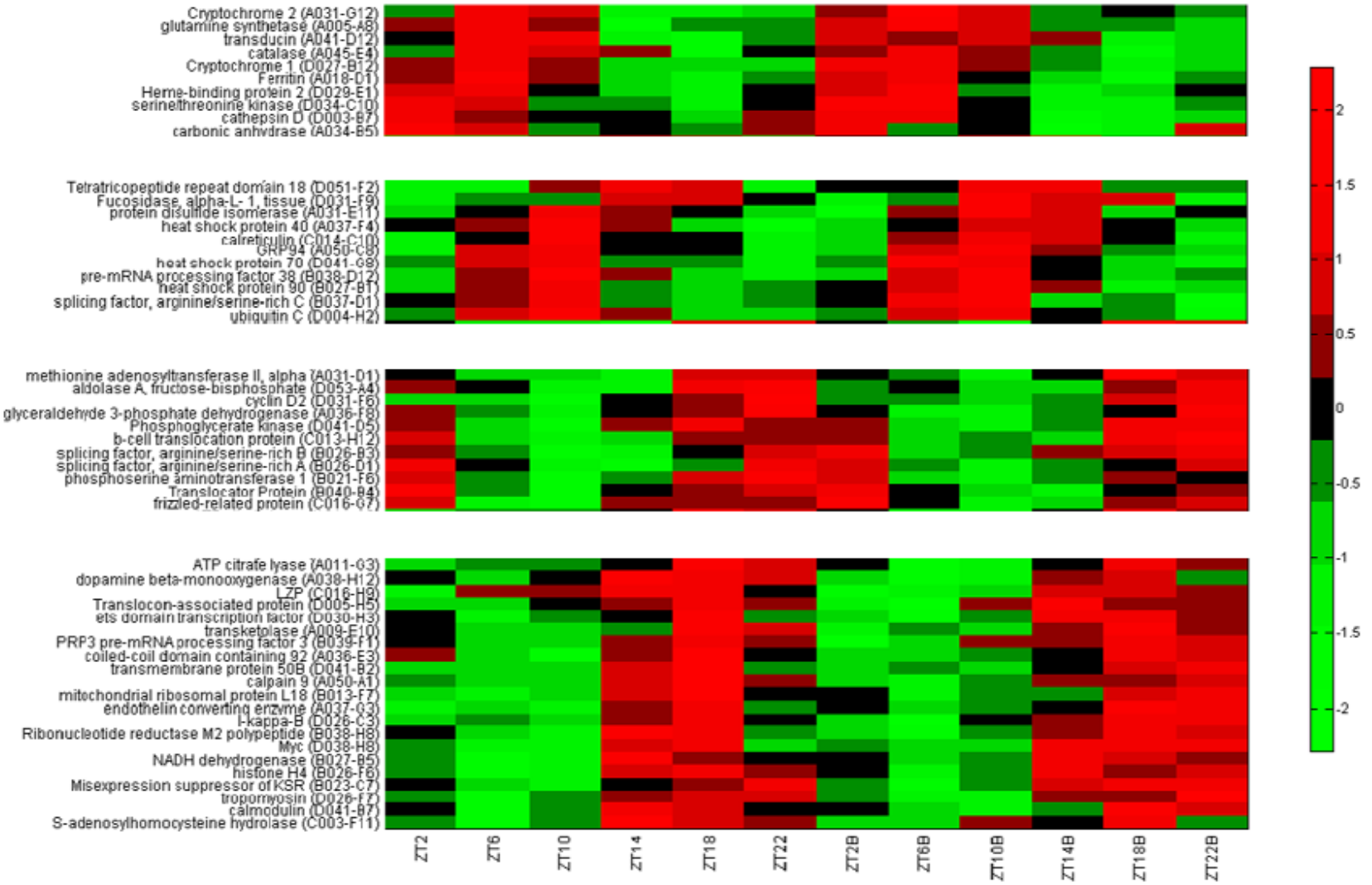






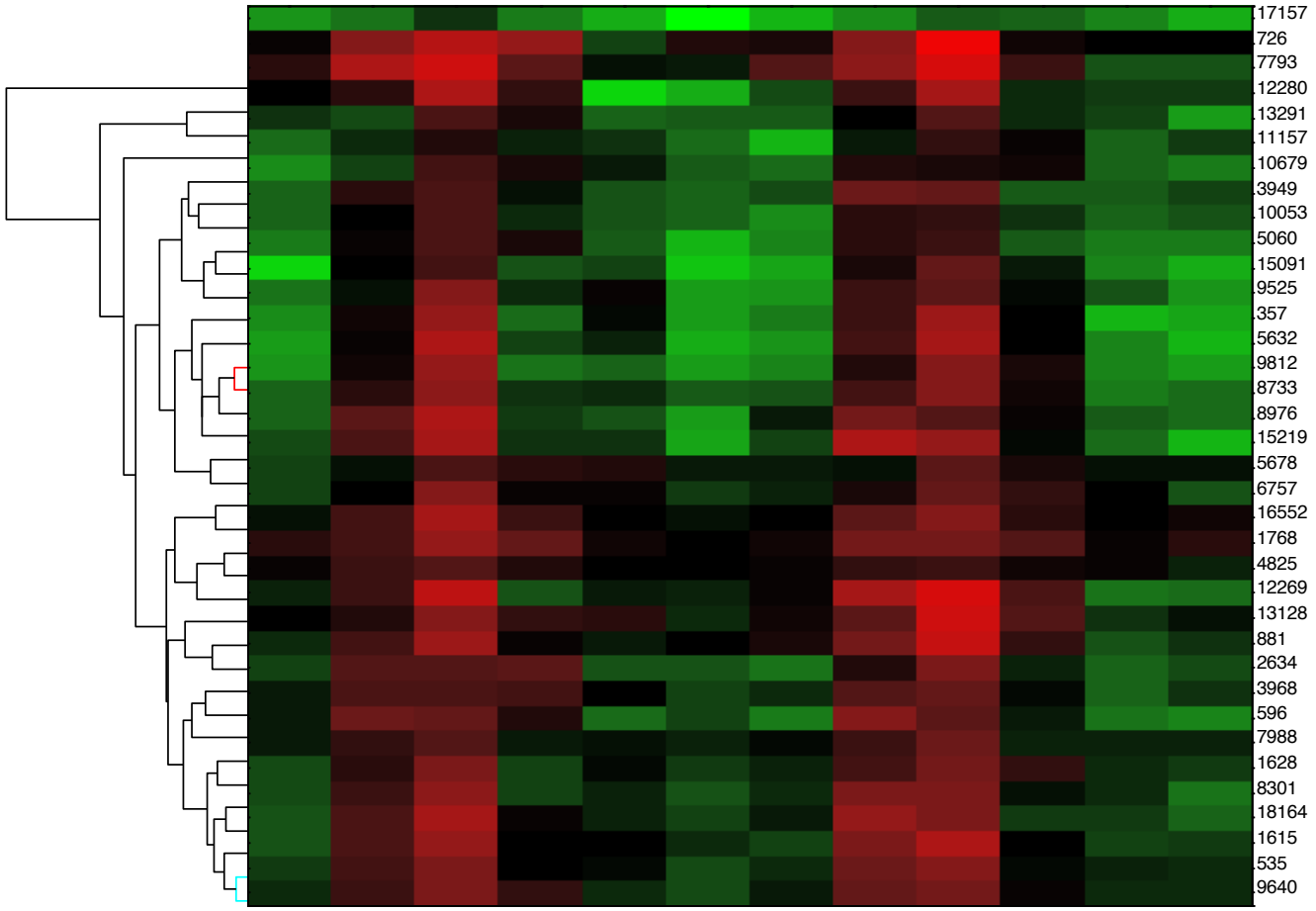


Genes in the same cluster share the same function



Levy et al., Science, 2011

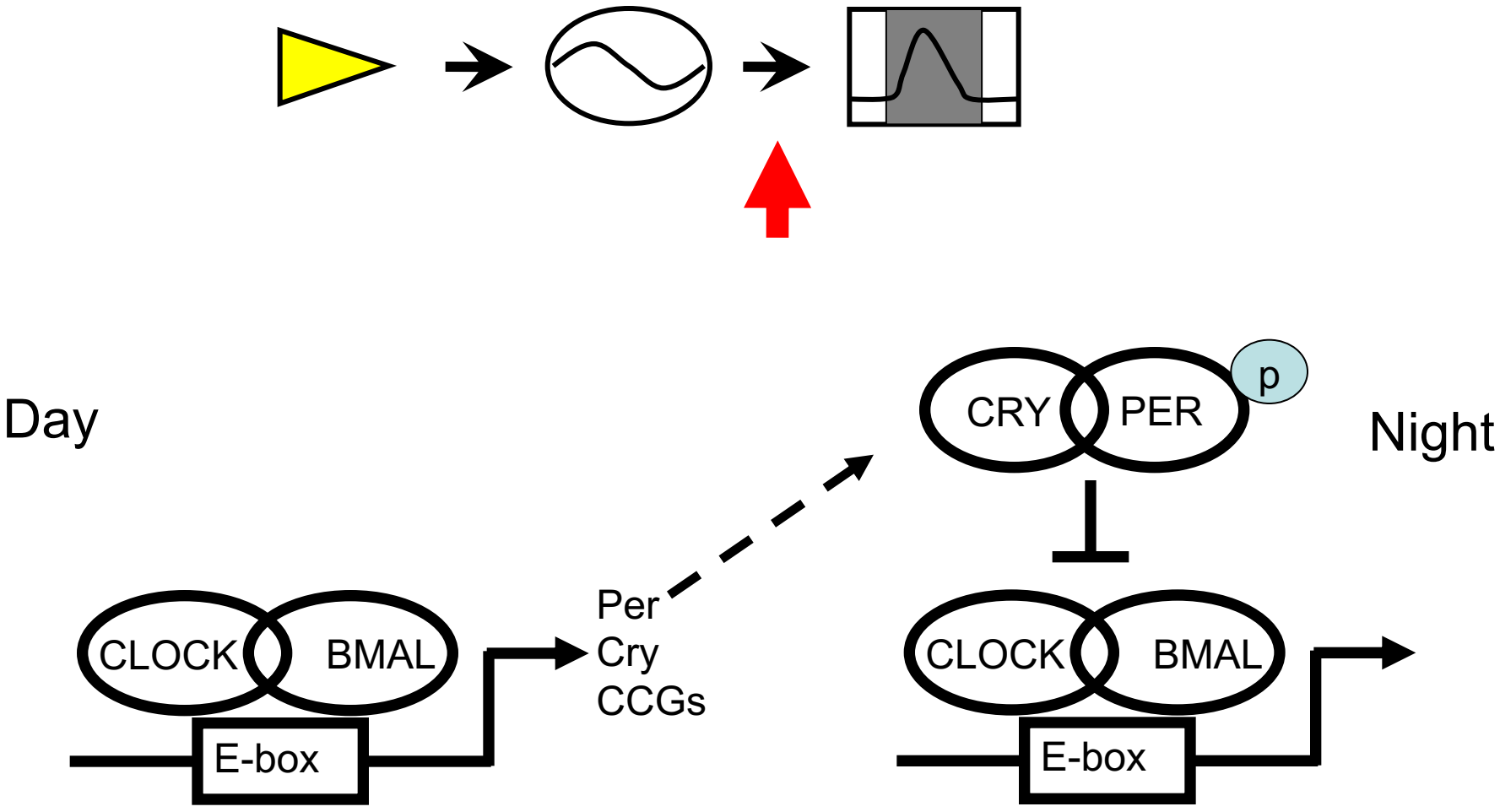
Group with peak at 16:00



- Almost all the genes in this group have the same function
- Unknown genes? Solution to a very hard problem

Levy et al., Science, 2011

Regulatory mechanisms



Thinking outside the (E-)box

- Key assumption:
  similar expression = similar regulation
- The third concept that resulted from Eisen paper

Promoter 1: ATCTGCCCTAGGCACGAAAATCGCG
Promoter 2: CCCACGAAATAGGCTCAAGGTAGTA
Promoter 3: GGTTCCACAACCATGCACACGAAGA
Promoter 4: GCGCATCGCACGAAATATTGATGCG
Promoter 5: CCCACCATTACCAGACACACGAAGA
Promoter 6: CGCACGAAATTGCCTCATGGTAGTA
Promoter 7: CCAGGAAGAGGCACGAAAGAGGGC

- *ab initio* motif detection programs
- need DNA sequences …