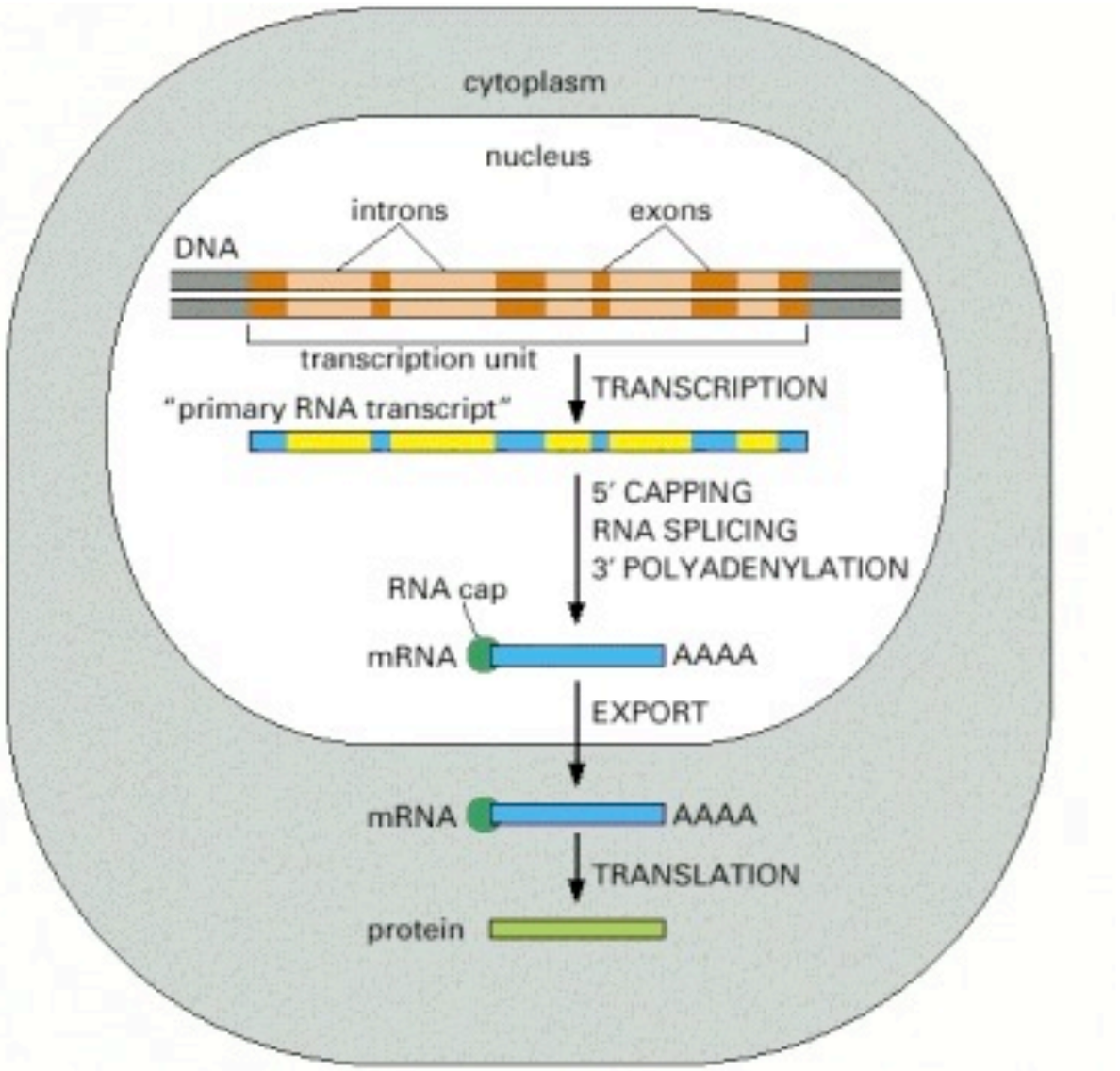
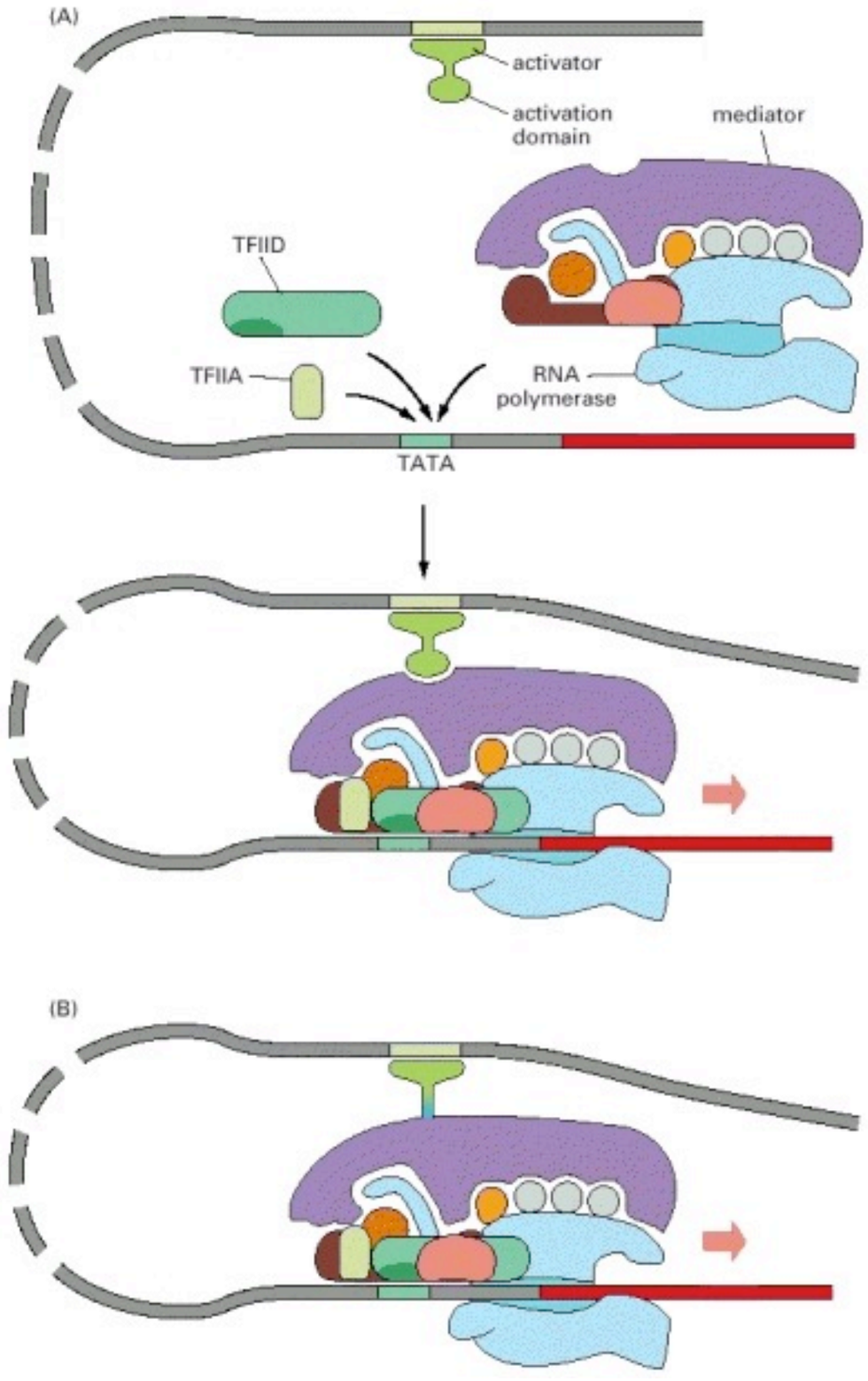
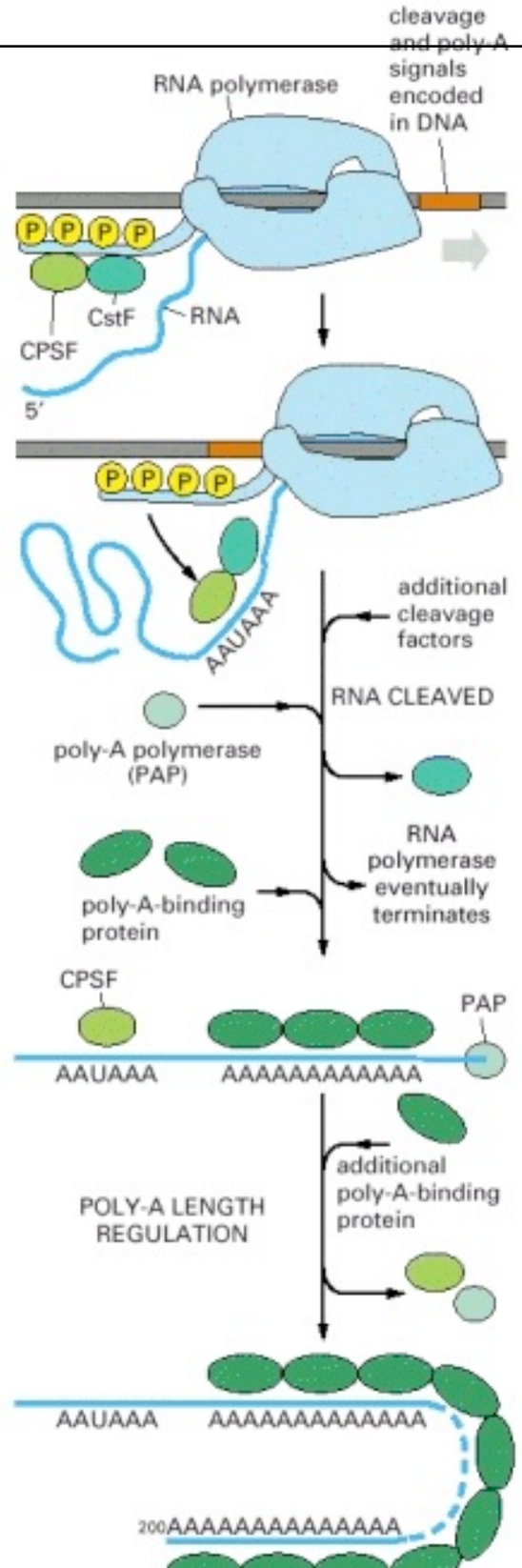
Transcription

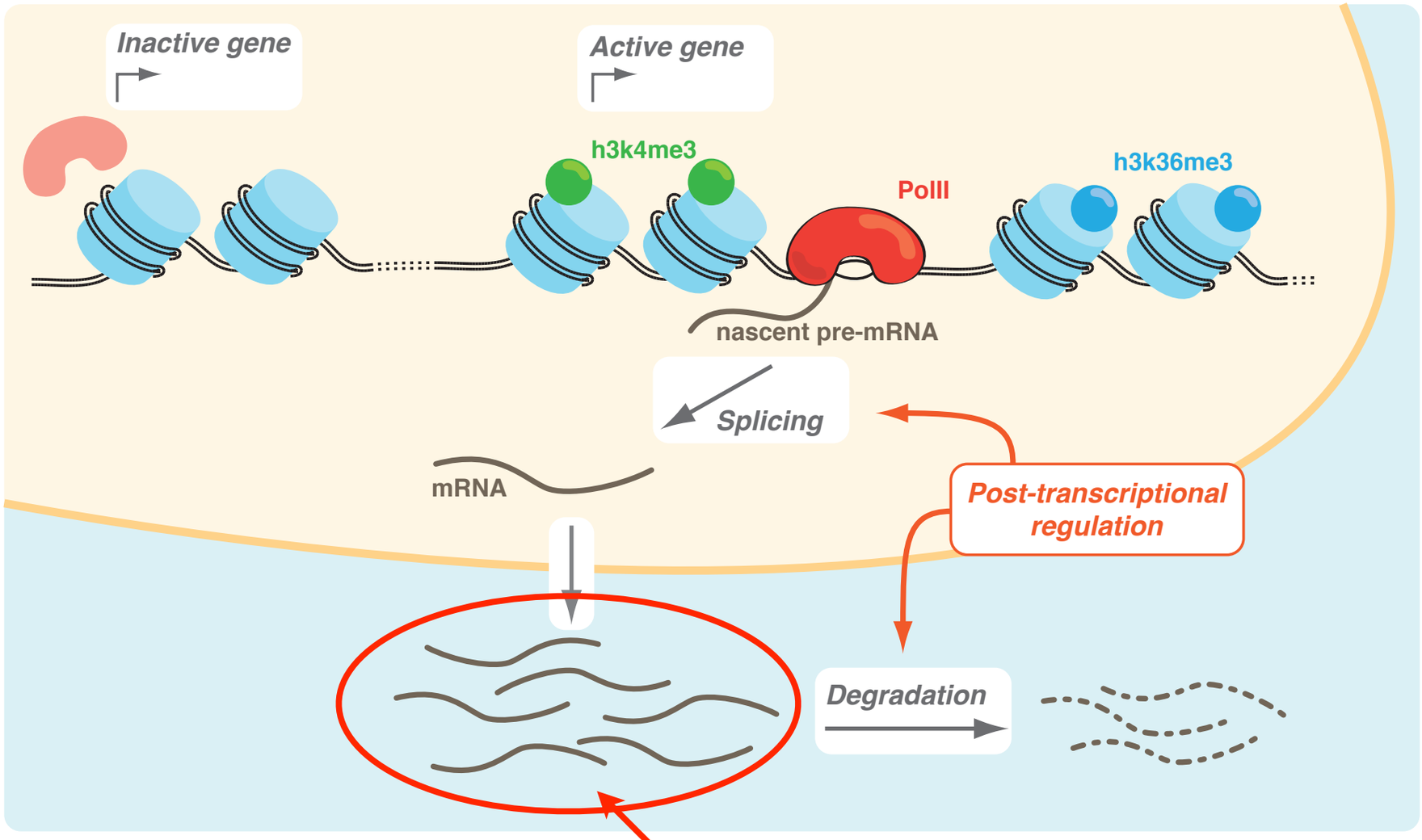


Transcription





Transcription



Equilibrium pool of mRNA for each gene

Simplest model

Transcription rate: $P(t)$ [#mRNA/time]
mRNA pool: $m(t)$ [#mRNA]
Degradation rate: γ [1/time]

$$\dot{m}(t) \;=\; P(t) - \gamma m(t) \;.$$

Case 1: $P(t) = P_0$

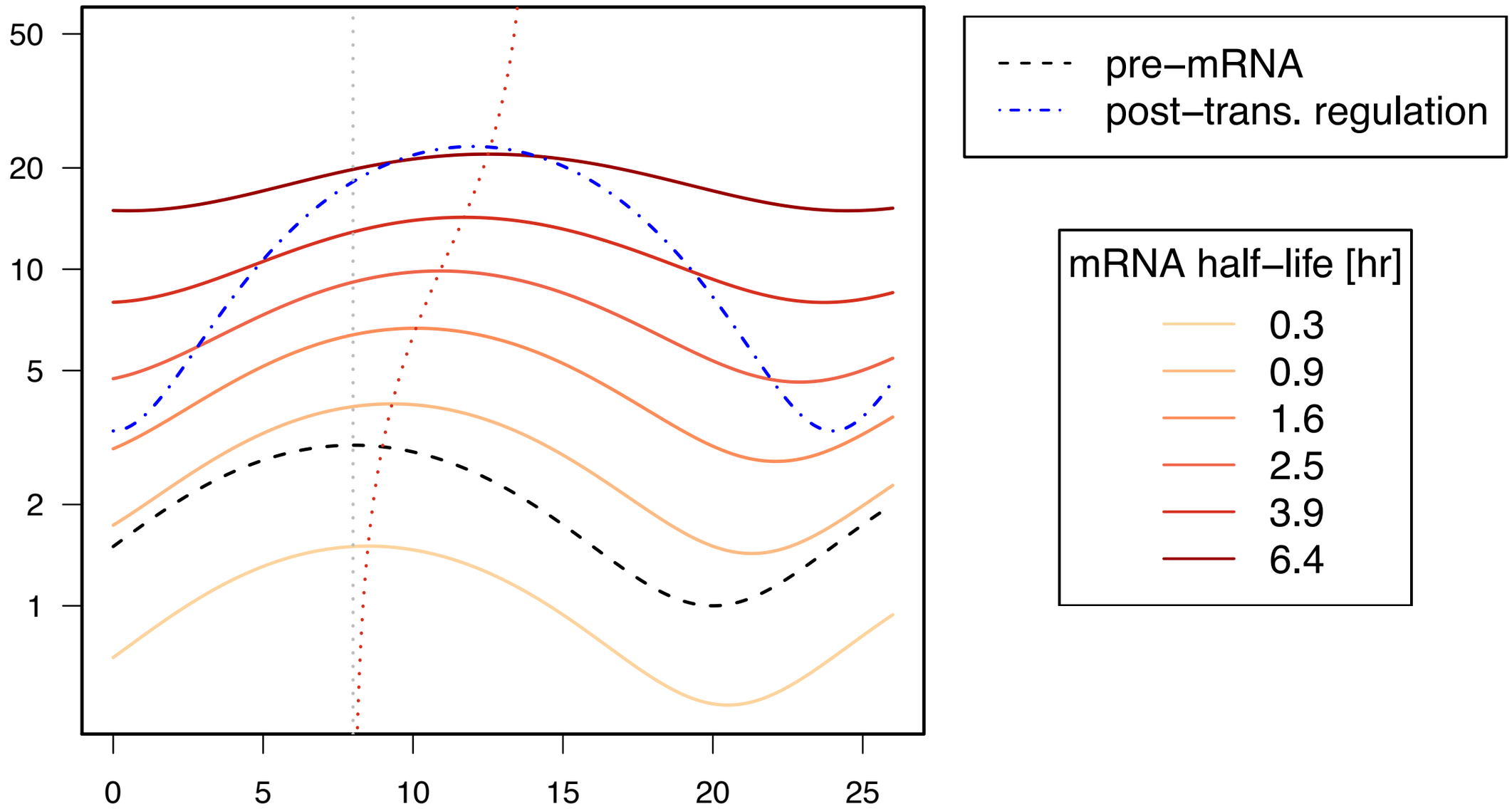
$$m(t) \;=\; \frac{P_0}{\gamma}(1 - e^{-\gamma t}) + e^{-\gamma t} m(0) \;\to\; \frac{P_0}{\gamma} \;.$$

Case 2: $P(t) = P_0 + \cos(\omega t)$

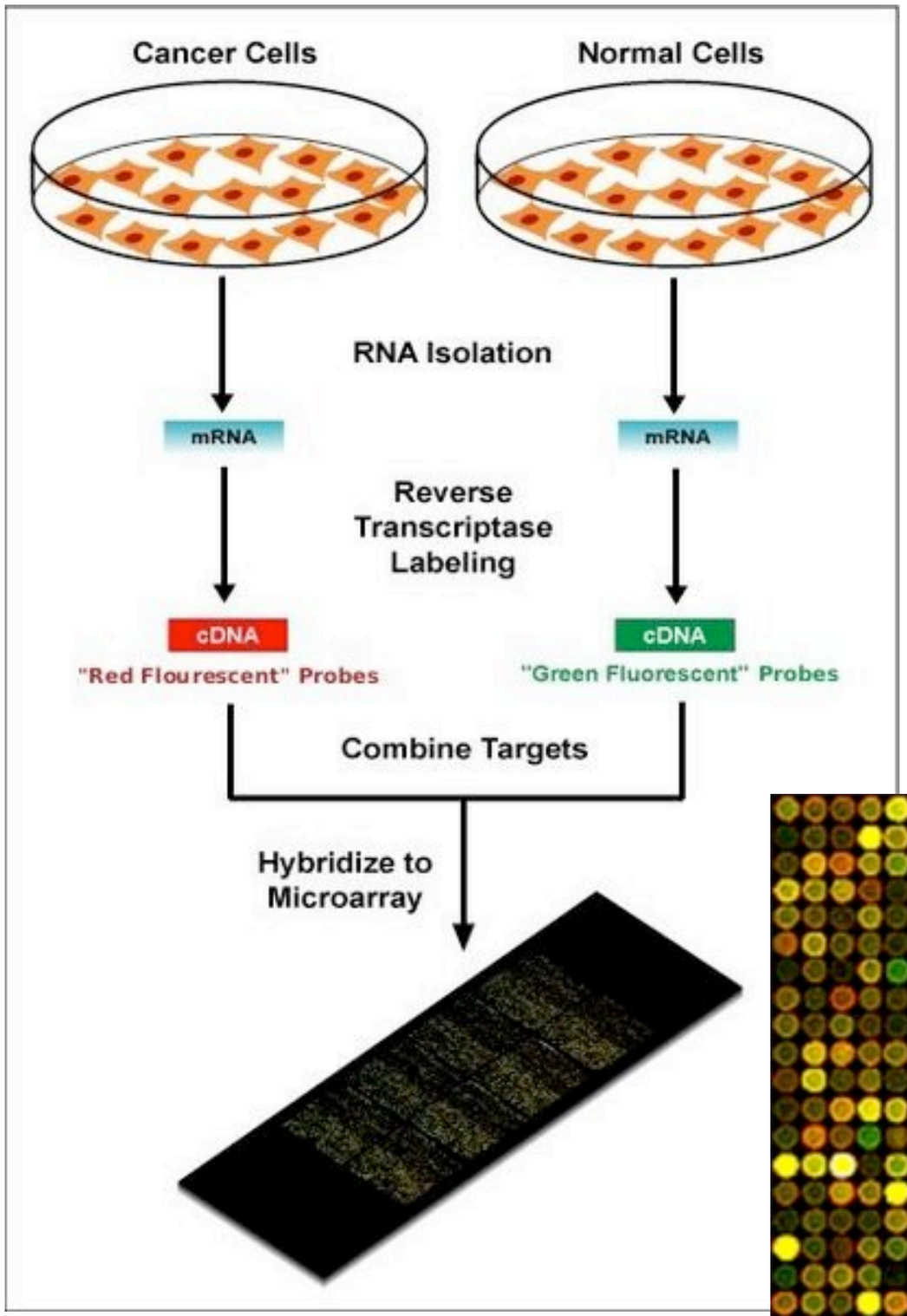
$$m(t) \;\to\; \frac{P_0}{\gamma} + \frac{1}{\sqrt{\gamma^2 + \omega^2}} \cos(\omega(t - \tau)) \;.$$

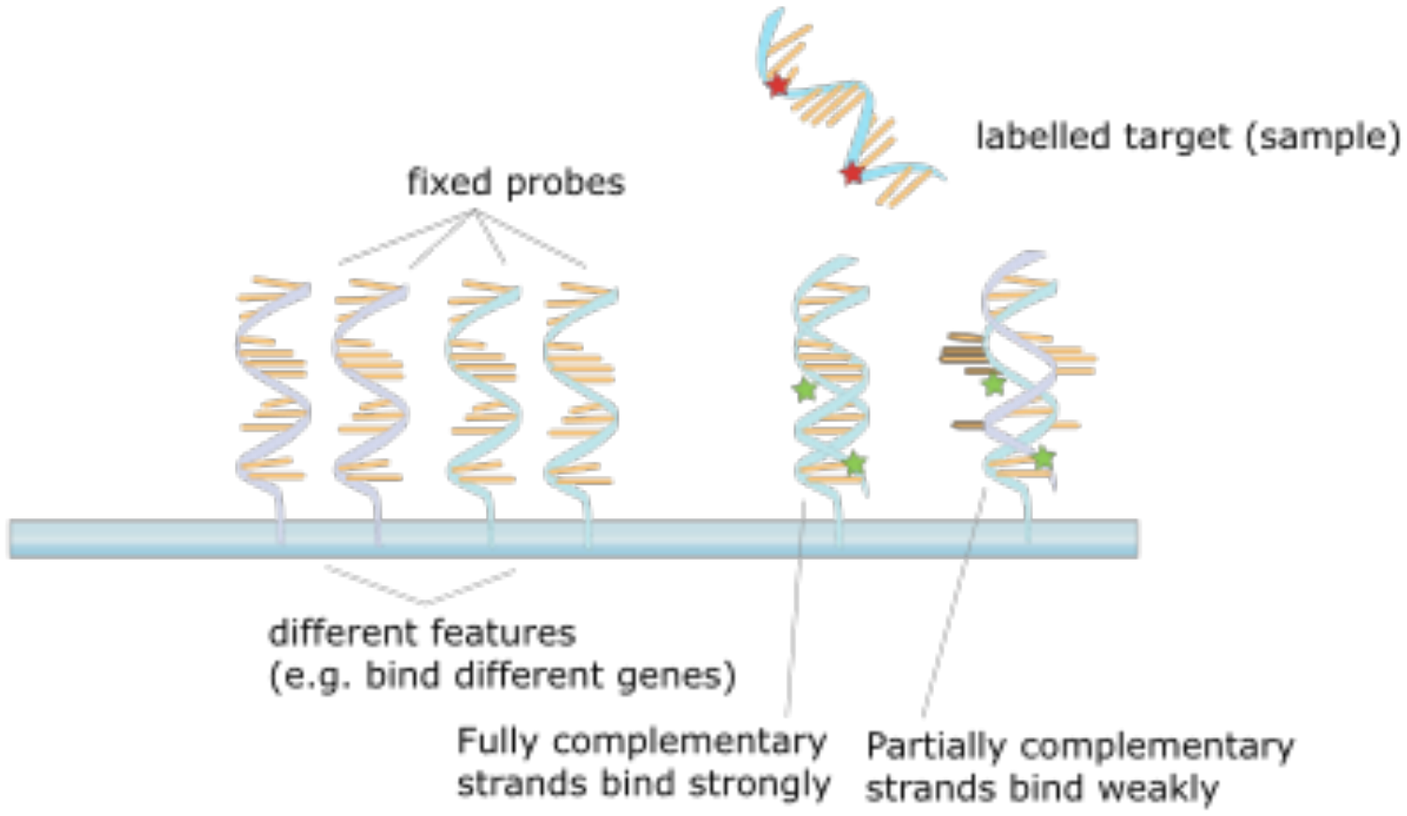
with $\sin \omega\tau = \frac{\omega}{\sqrt{\gamma^2+\omega^2}}$

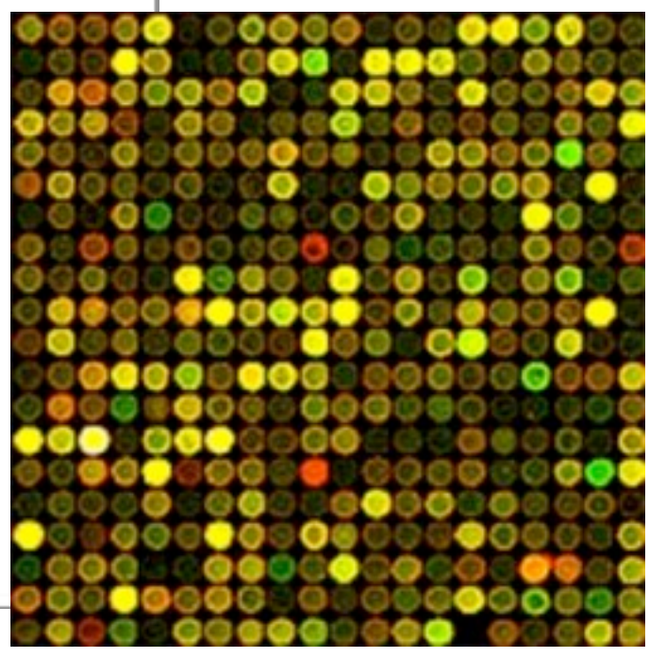
Simplest model



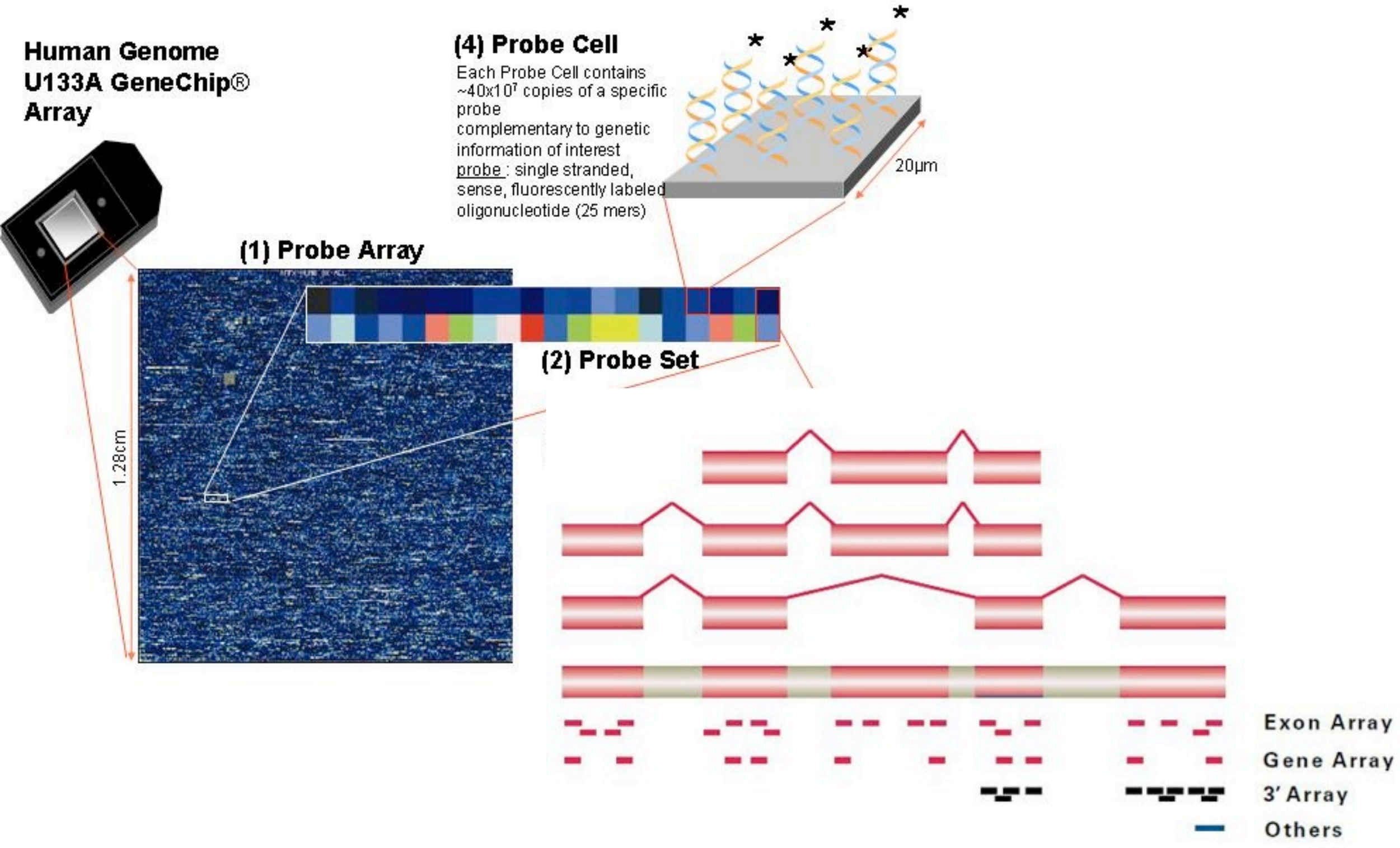
Microarrays







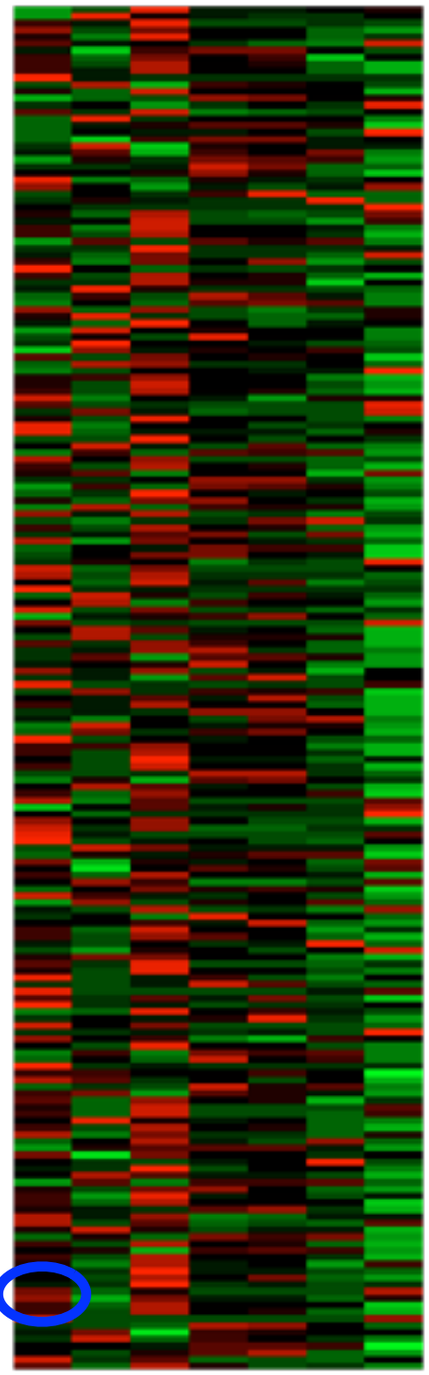
Microarrays



Output

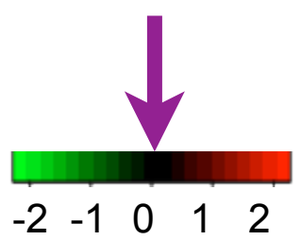
Columns: conditions (experiments)

Rows: features (genes)



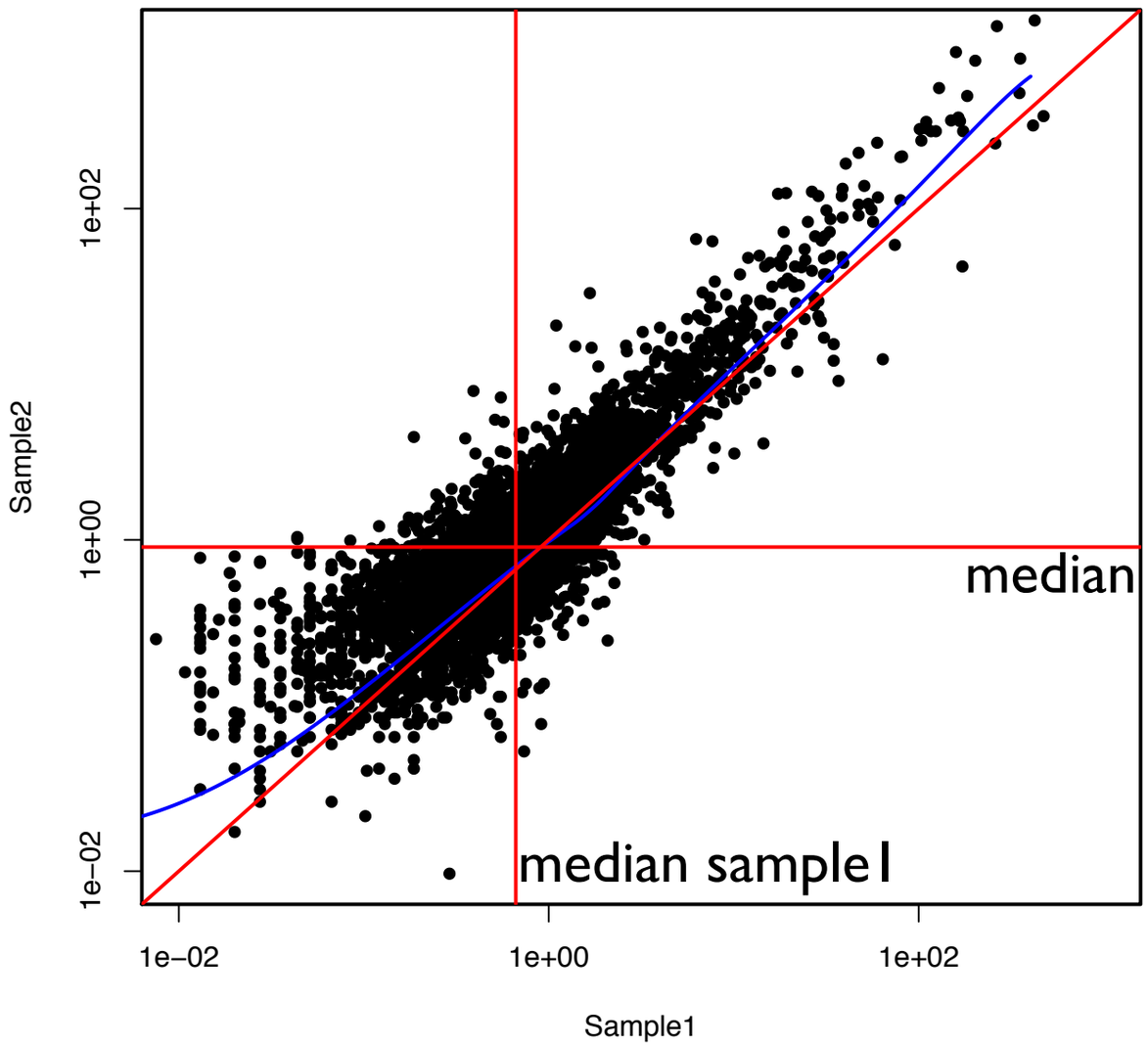
Signal is proportional to concentration of probe times concentration of mRNA

Cells: numbers (measured signal)



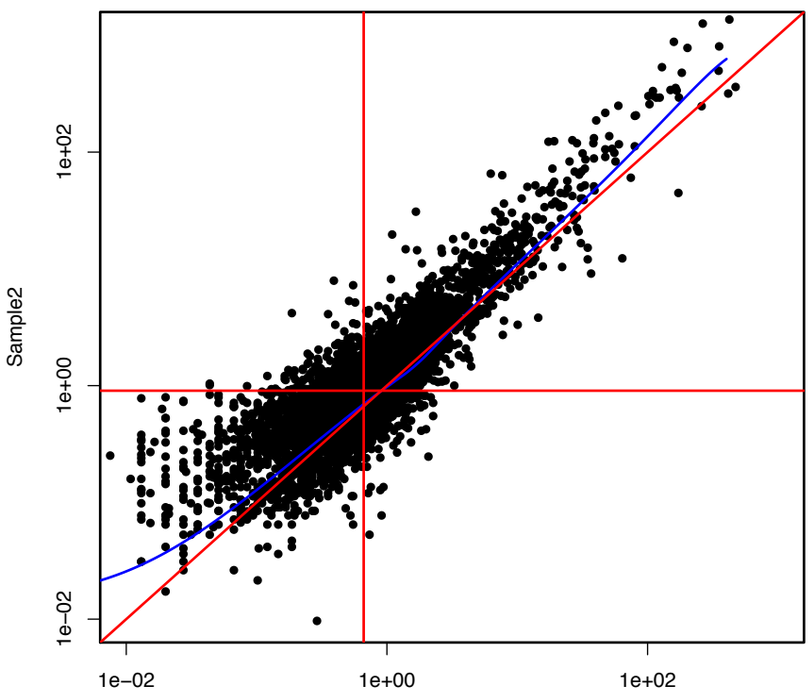
Normalization

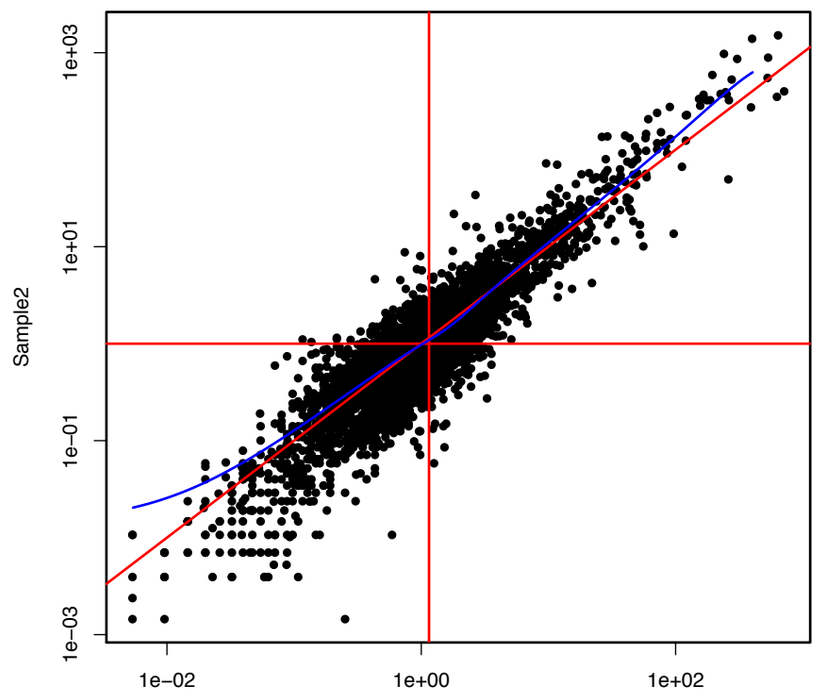
Average variation between any 2 conditions is 0:
Systematic variation MUST BE technical artifact

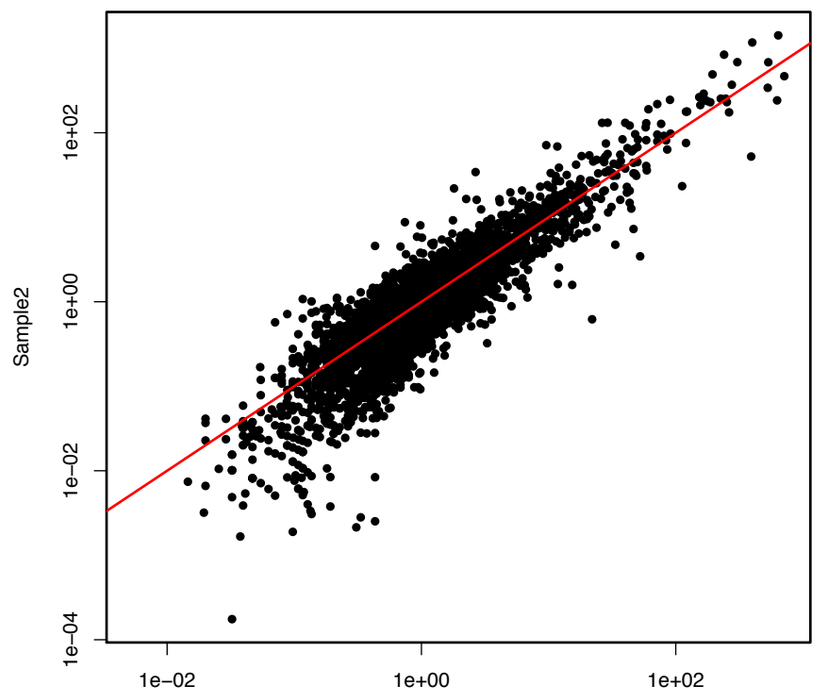


- Normalization consists in removing systematic variations
- Work in log-log coordinates
- Shift in medians
- Curved shaped

Normalization

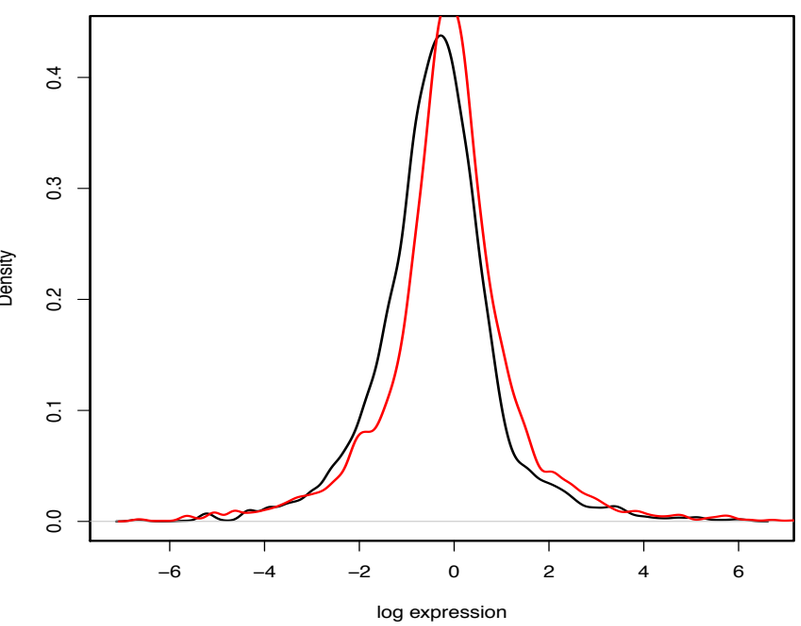


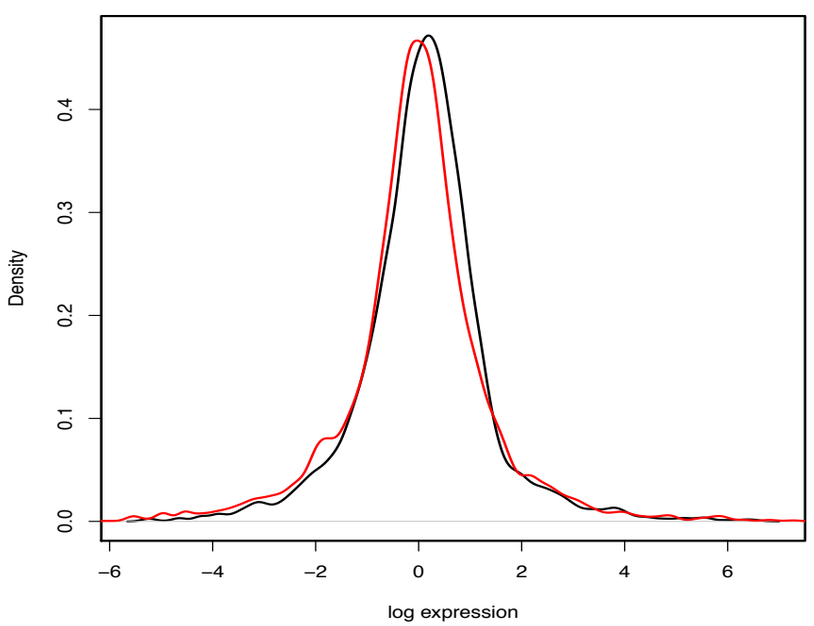


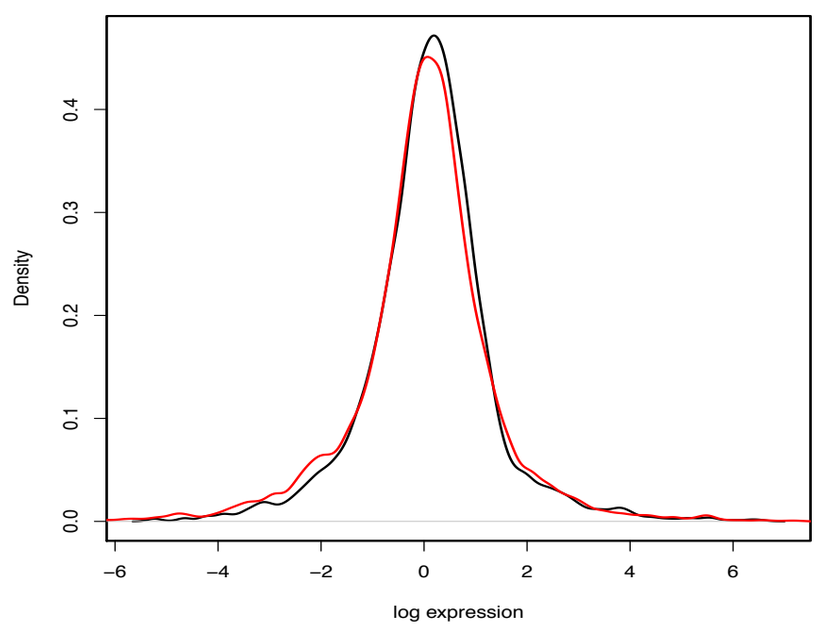


sample=sample/median(sample)

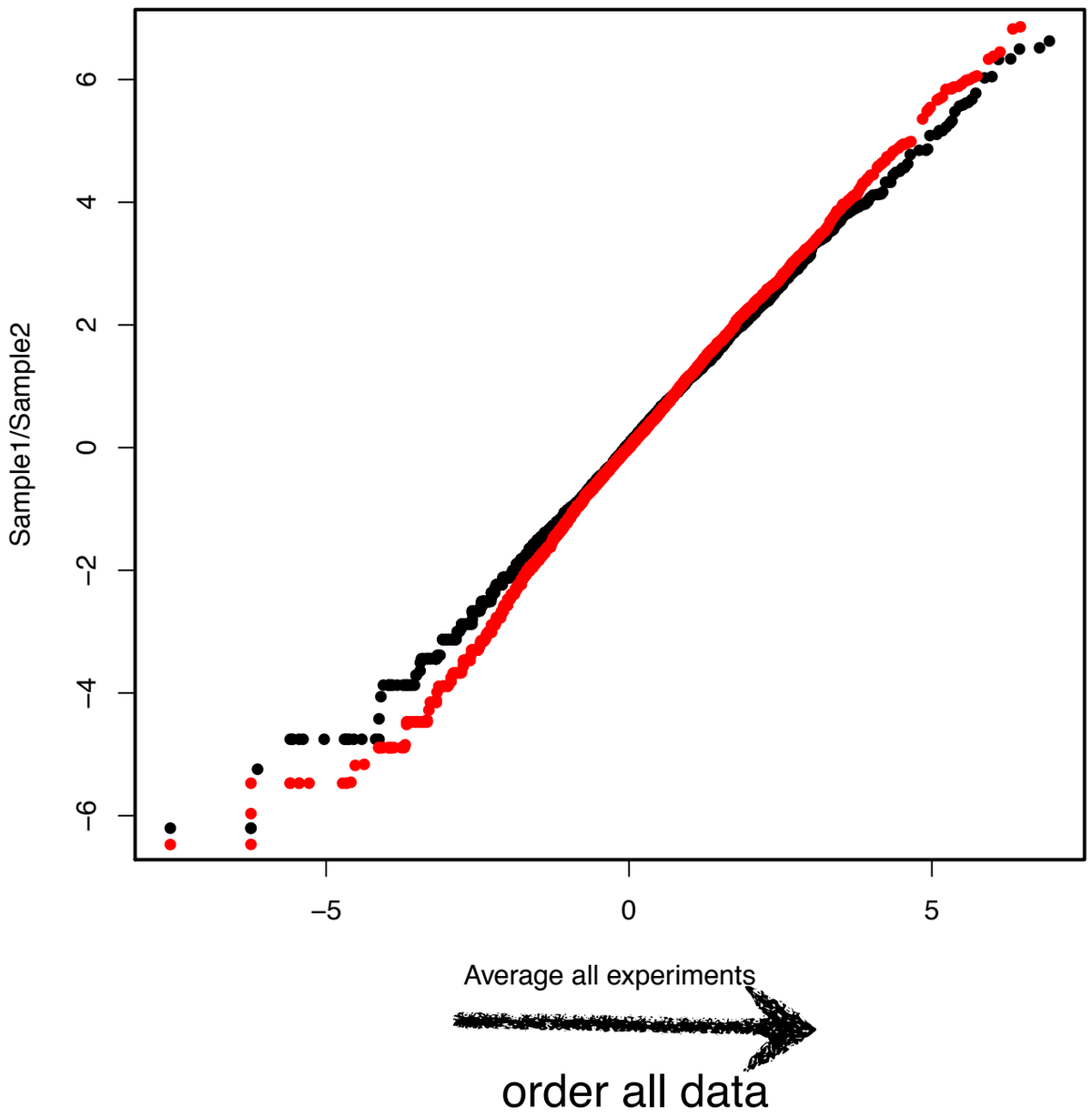
sample2=sample2/fit*sample1





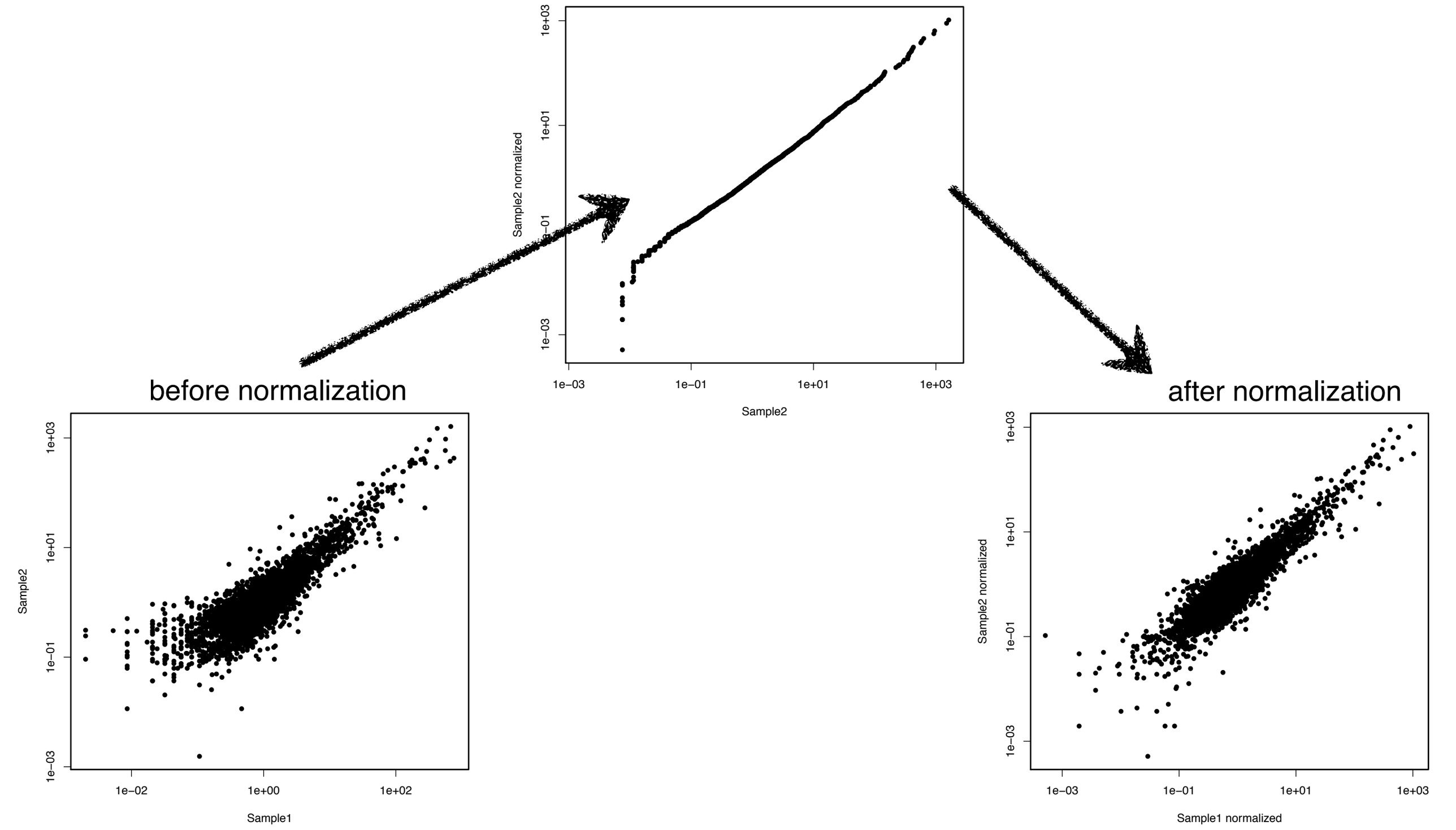


Quantile normalization



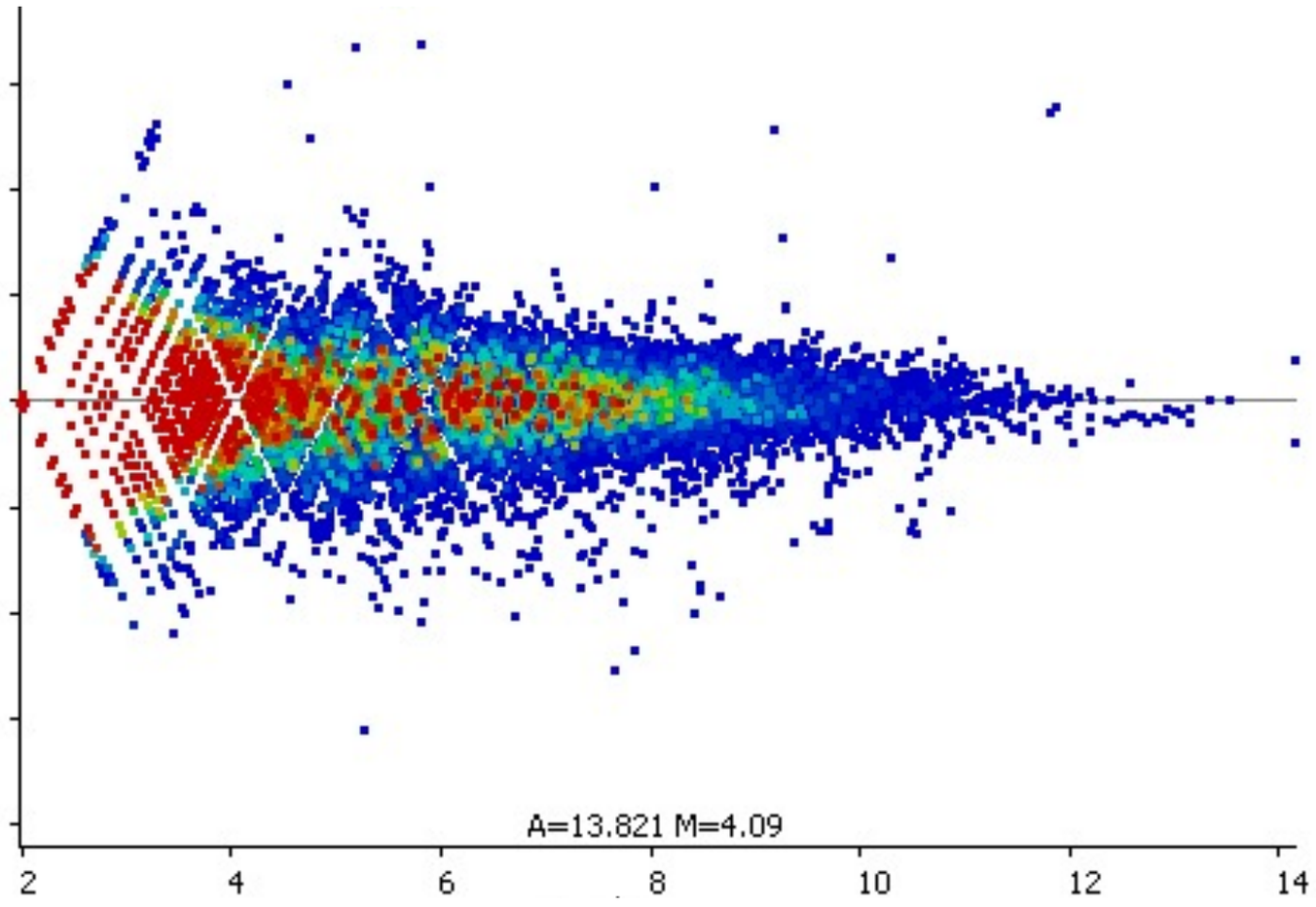
- Substitute ordered values from every sample with ordered values from average (or from specific distribution, e.g. gaussian)

Quantile normalization



MA plot: differential expression

Log2(Sample1/Sample2)

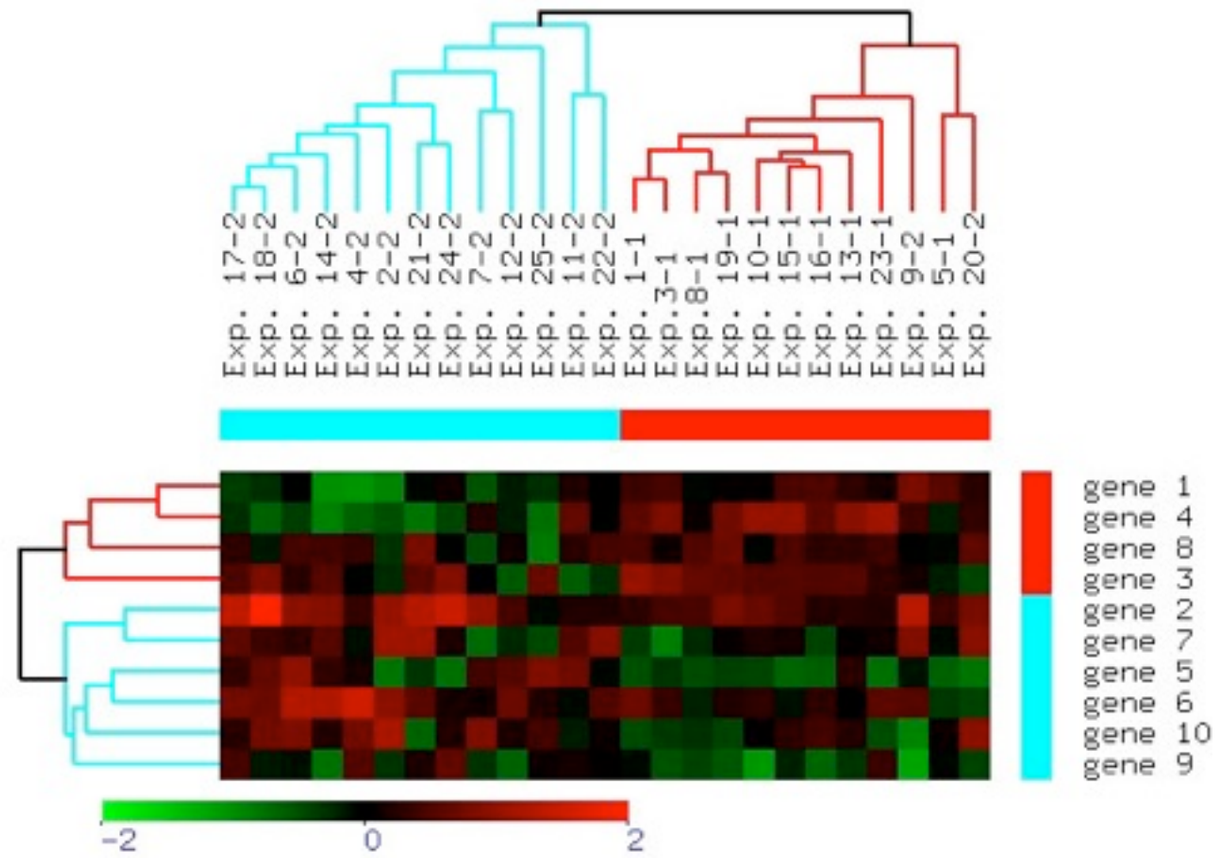


differential expression

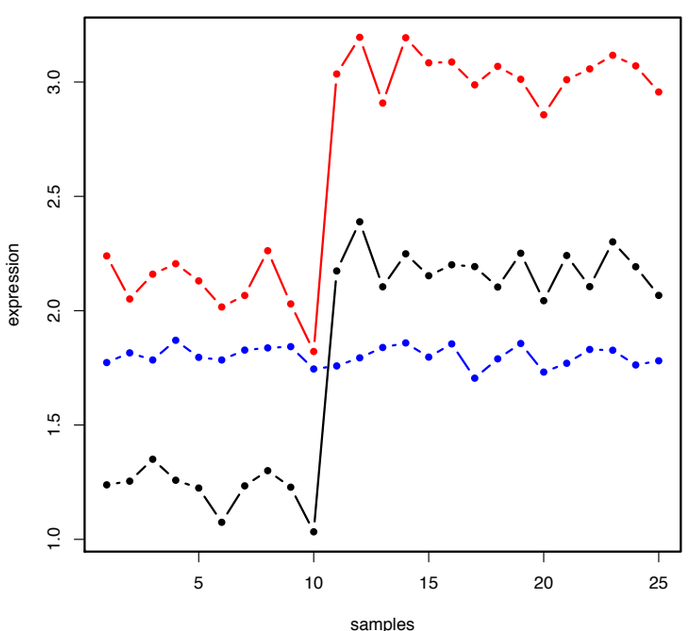
Log10(Sample1*Sample2)/2

absolute expression

Clustering



- Same algorithm as UPGMA
- Distance matrix is 1-cor(row1,row2)
- Update matrix with distance to average of two groups weighted by size
- Do the same for columns (rotate matrix)



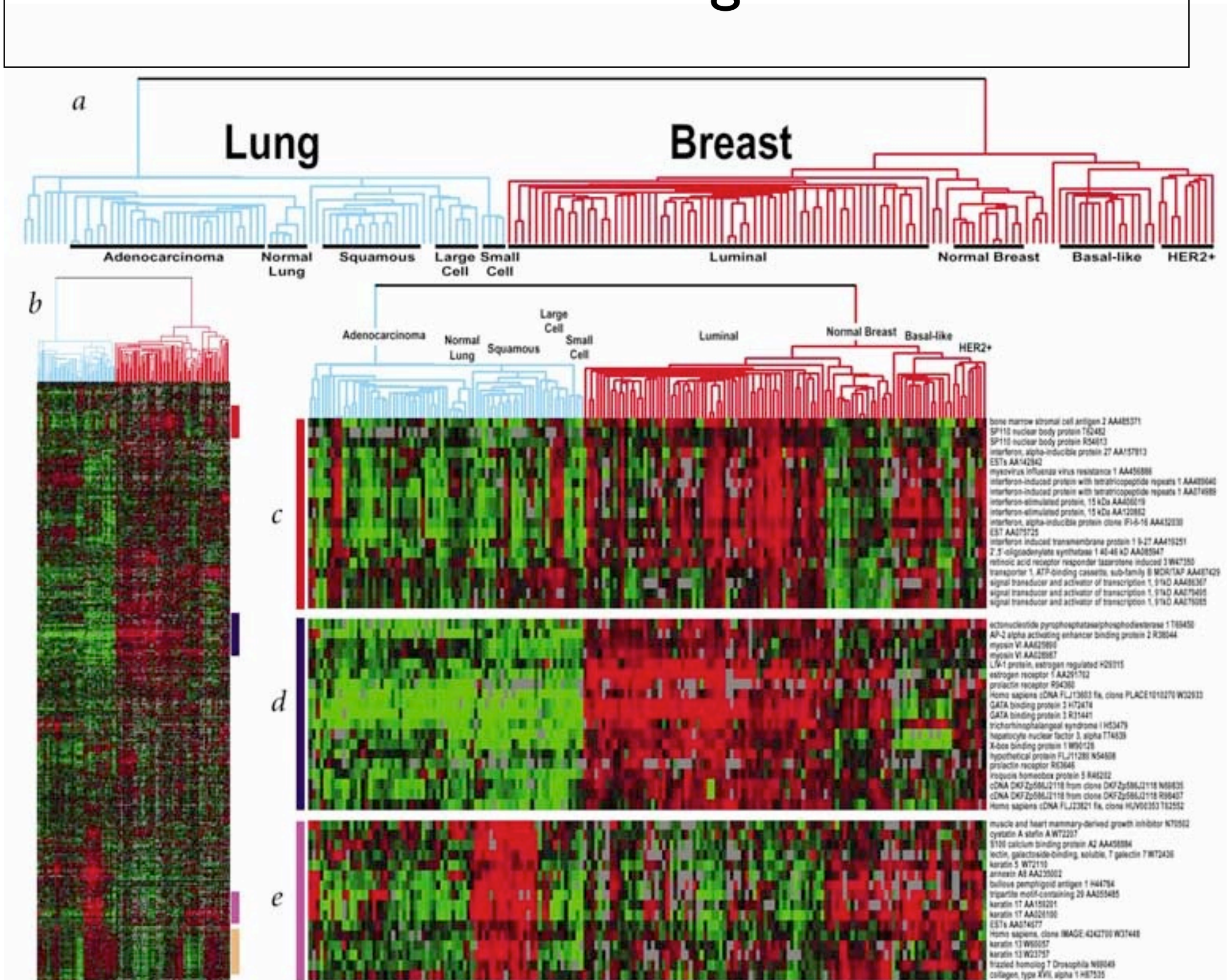
distance

	1	2
2	4.3	
3	2.4	4.9

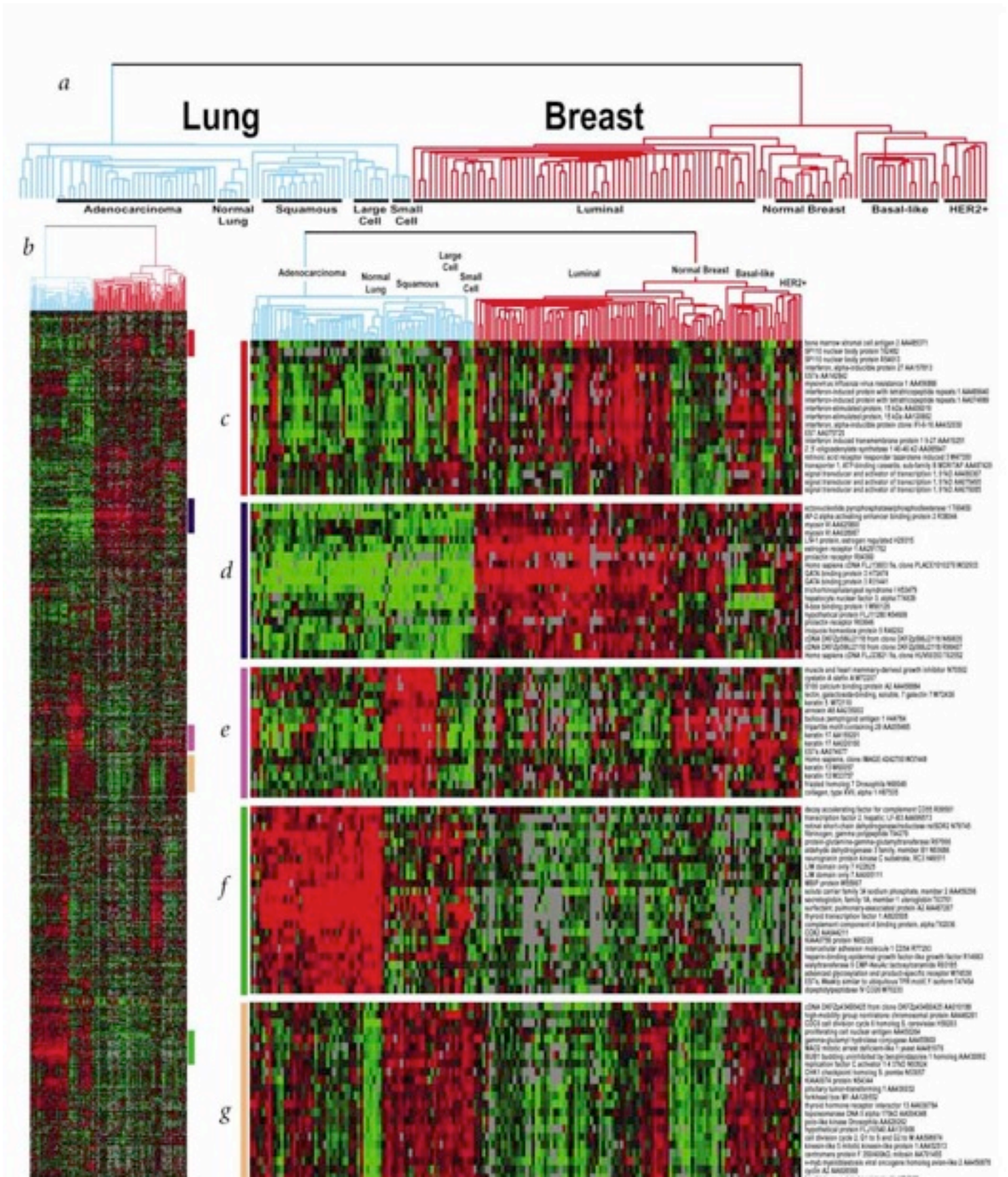
correlation

	1	2
2	0.99	
3	-0.06	-0.03

Clustering


a
Lung
Breast
Adenocarcinoma
Normal Lung
Squamous
Large Cell
Small Cell
Luminal
Normal Breast
Basal-like
HER2+
b
c
d
e

Clustering



- Similar expression patterns across many conditions probably imply a common set of regulators
- Looking for a shared set of functional annotations can help find the regulators

Linear models

	treat+WT	treat+KO	no treat+WT	no treat+KO
g	M11	M12	M13	M14

For each gene, make a linear relation between effect (expression) and factors (conditions)

$$\log(M_{cg}) = \alpha_g + \sum_n I_{cn}\beta_{ng} + \epsilon_g \ ,$$

$$\langle \log(M_{cg}) \rangle = \alpha_g + (I \cdot \beta)_{cg} \ ,$$

Design matrix:

	1	2	3	4
treat	1	1	0	0
KO	0	1	0	1