# The power of RNA-seq

## Introduction

Harm Nijveen & Dick de Ridder
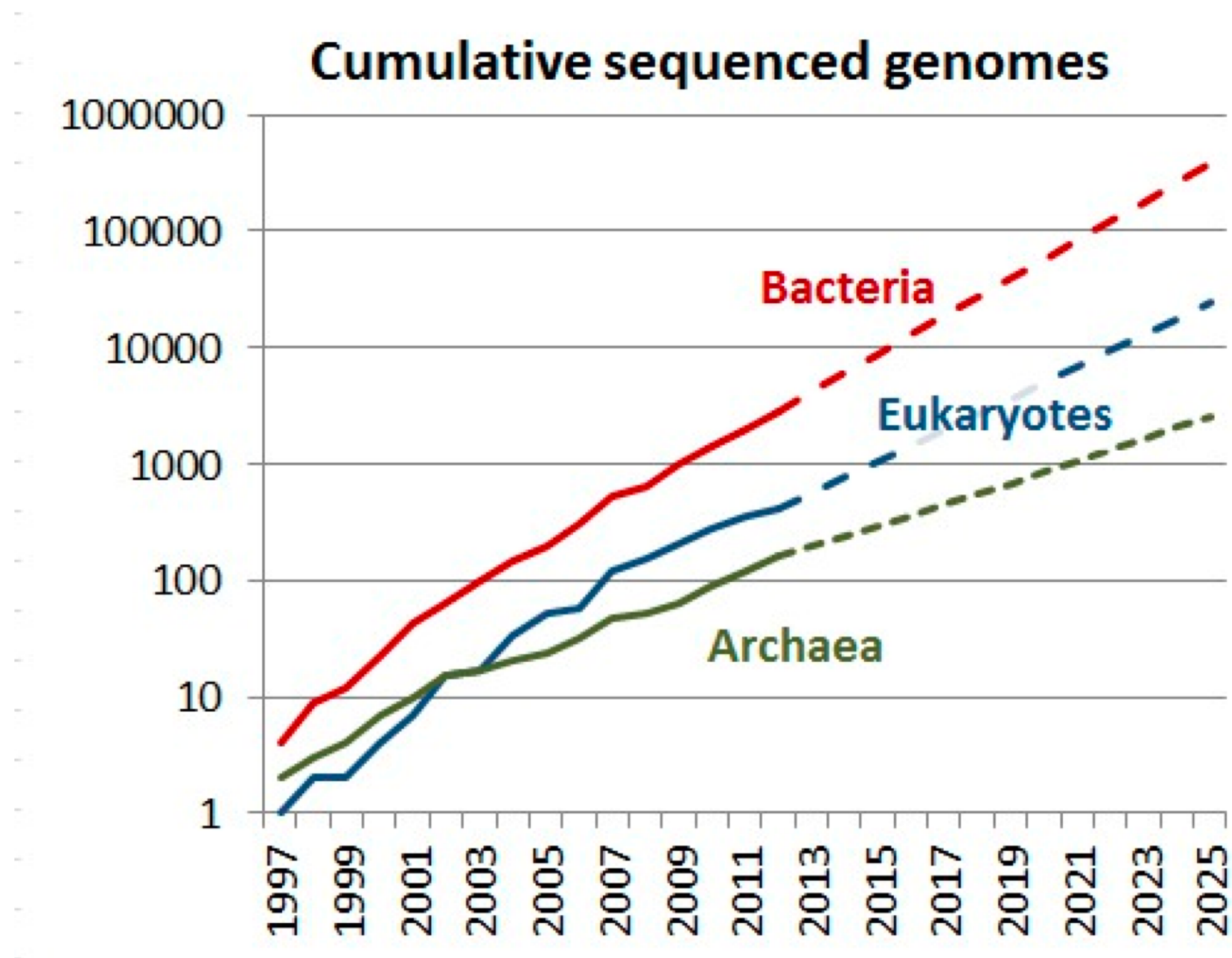
# Welcome!

| | | | | |
|---|---|---|---|---|
| Jun-11 | | 09:00 | Registration & coffee | |
| | | 09:30 | Introduction | |
| | 1a | 09:45 | **NGS** | Dick de Ridder |
| | | 10:30 | Coffee | |
| | 1b | 11:00 | **RNA-seq workflow** | Harm Nijveen |
| | | 11:45 | Practical: read mapping | |
| | | 12:30 | Lunch | |
| | 1c | 13:30 | **Assembly, annotation, quantification** | Harm Nijveen |
| | | 14:15 | Practical: quantification (using Excel) | |
| | | 15:00 | Coffee | |
| | | 15:30 | Practical: quantification (using R) | |
| | | 16:15 | Invited speaker | Richard Immink |
| | | 17:00 | End | |
| | | | | |
| Jun-12 | 2a | 09:30 | **Differential expression** | Harm Nijveen |
| | | 10:15 | Practical: differential expression | |
| | | 10:30 | Coffee | |
| | | 11:00 | Practical: differential expression | |
| | | 12:30 | Lunch | |
| | 2b | 13:30 | **Clustering & Visualisation** | Dick de Ridder |
| | | 14:15 | Practical: clustering & visualisation | |
| | | 15:00 | Coffee | |
| | | 15:30 | Practical: clustering & visualisation | |
| | | 16:15 | Invited speaker | Basten Snoek |
| | | 17:00 | End | |
| | | | | |
| Jun-13 | 3 | 09:30 | **Interpretation: enrichment, networks etc.** | Dick de Ridder |
| | | 10:15 | Practical: interpretation | |
| | | 10:30 | Coffee | |
| | | 11:00 | Practical: interpretation | |
| | | 12:30 | Lunch | |
| | | 13:30 | Topical talk | Gonda Buijs |
| | | 14:15 | Topical talk | Harm Nijveen |
| | | 15:00 | Coffee | |
| | | 15:30 | Topical talk | Rens Holmer |
| | | 16:15 | Drinks | |

**WAGENINGEN UNIVERSITY**
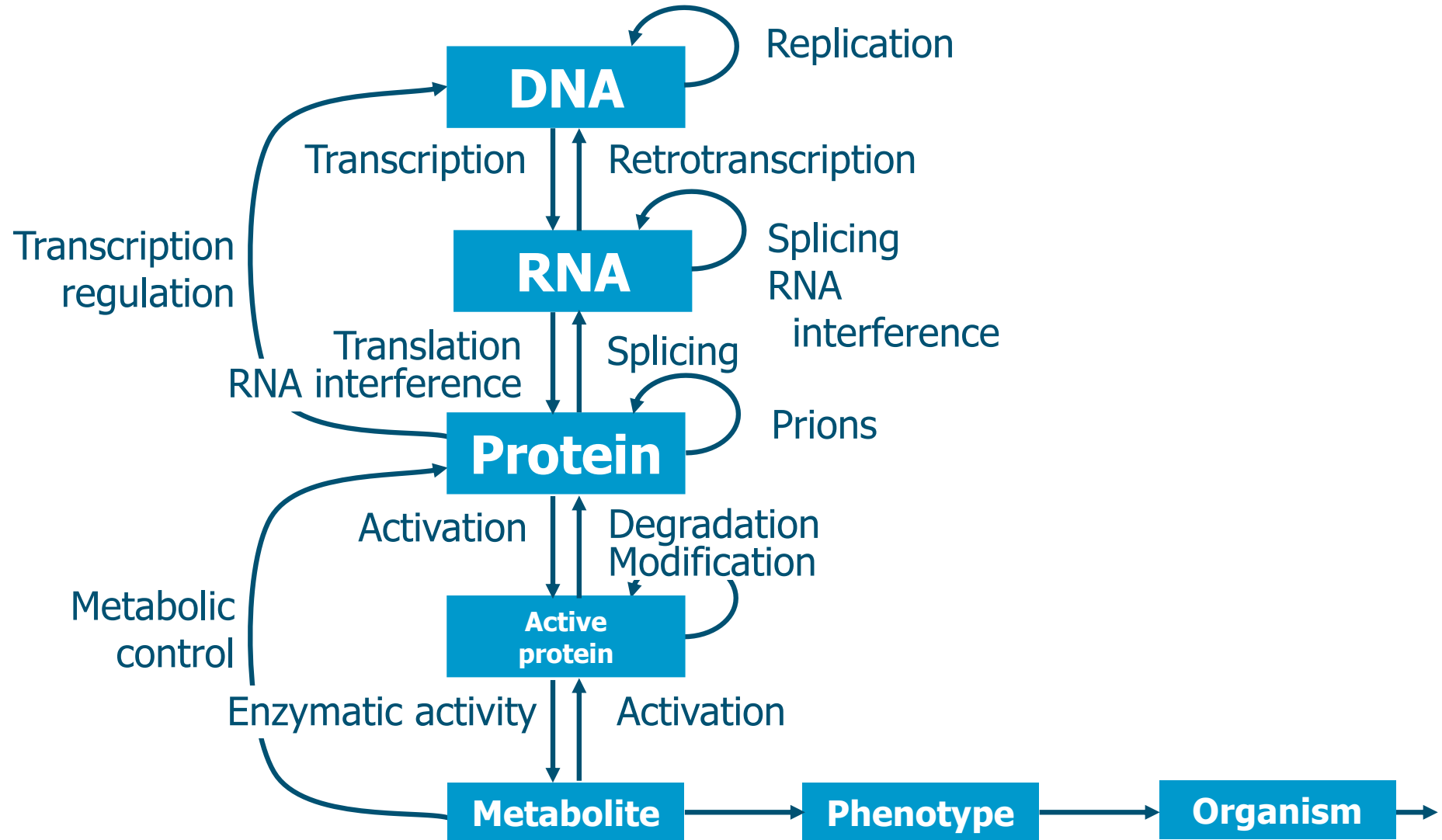WAGENINGEN UR

# Genomics



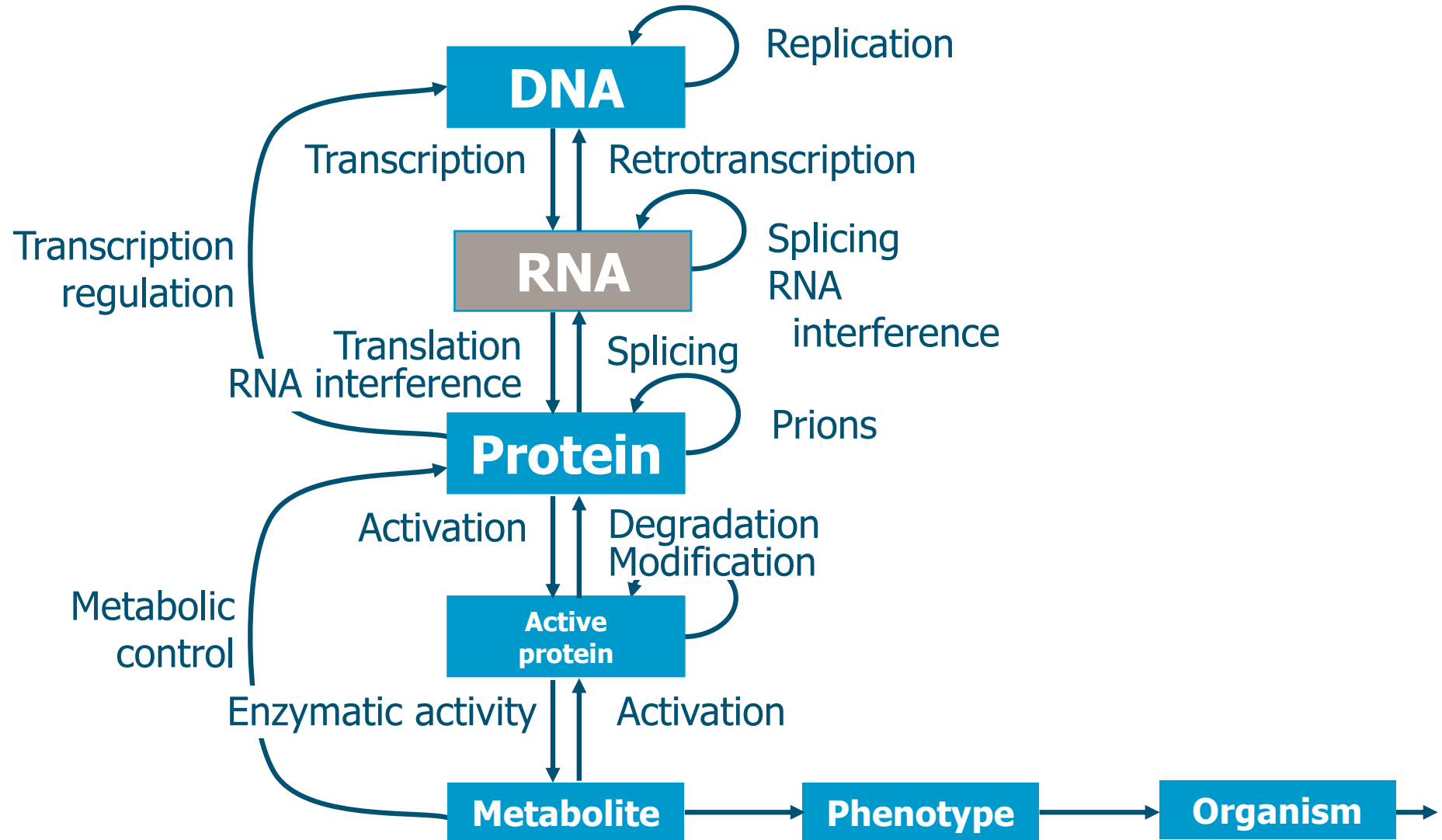**Cumulative sequenced genomes**

# Genomics (2)

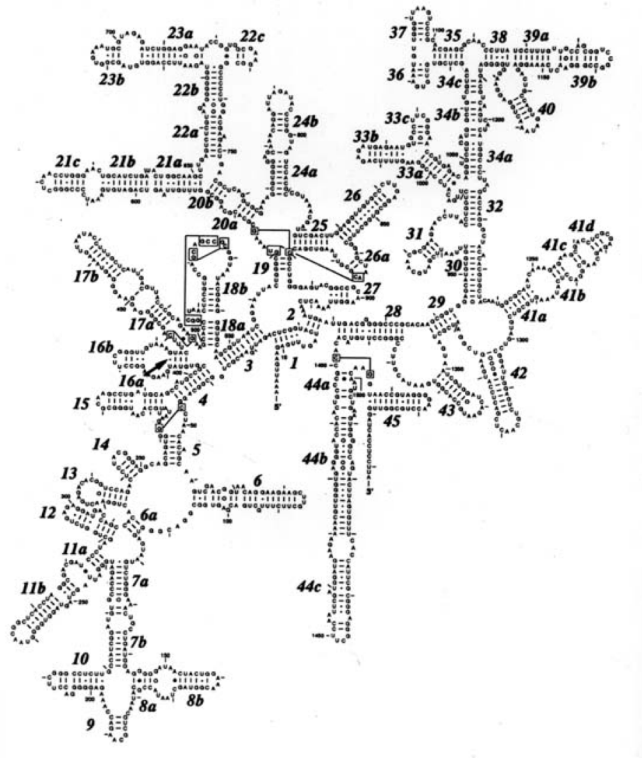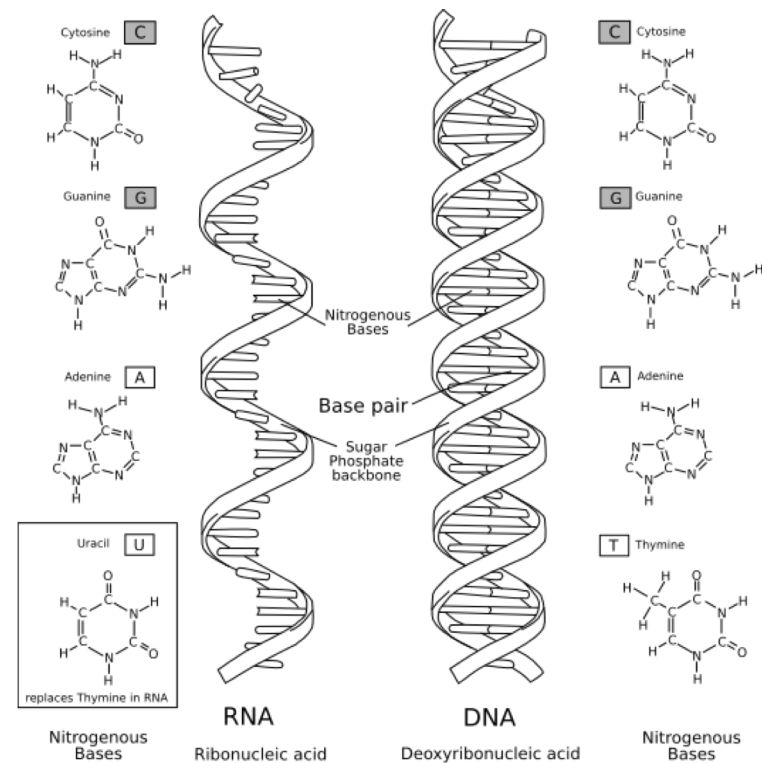# Between genes and proteins (functions)…

# Transcripts

# RNA vs. DNA

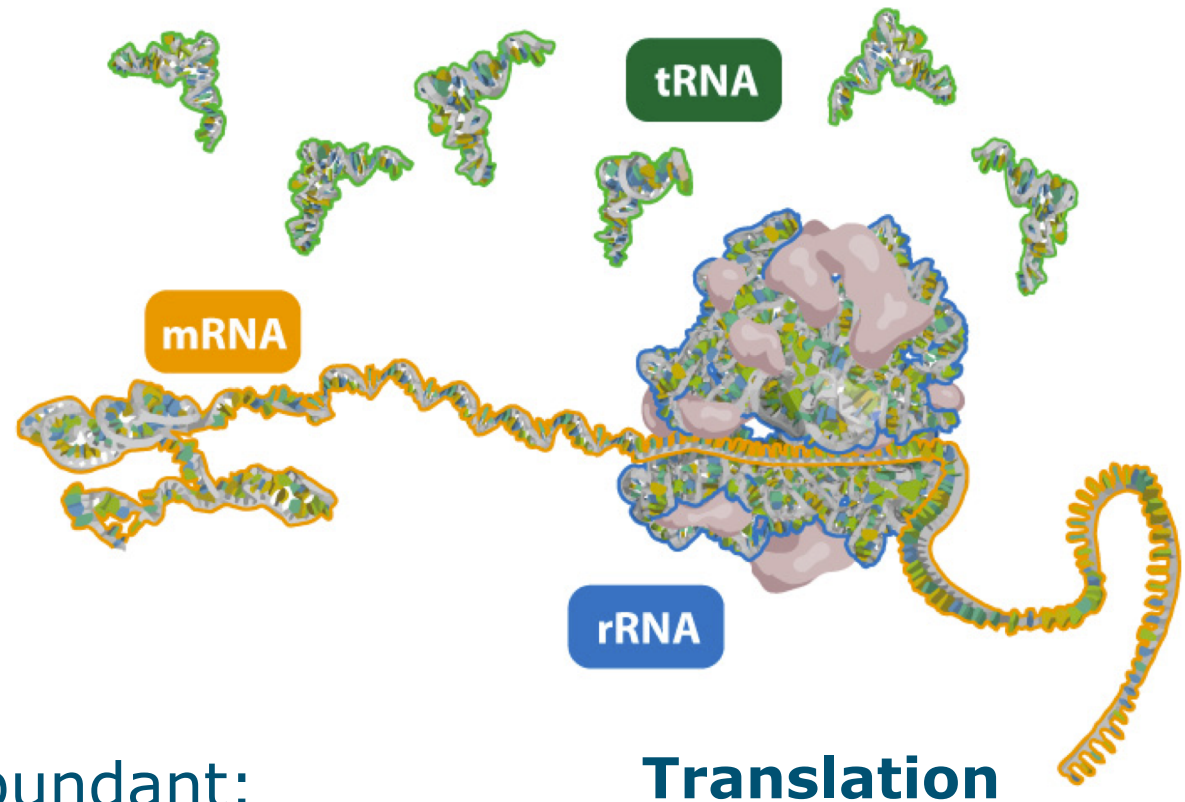- RNA is single stranded, folds into different structures
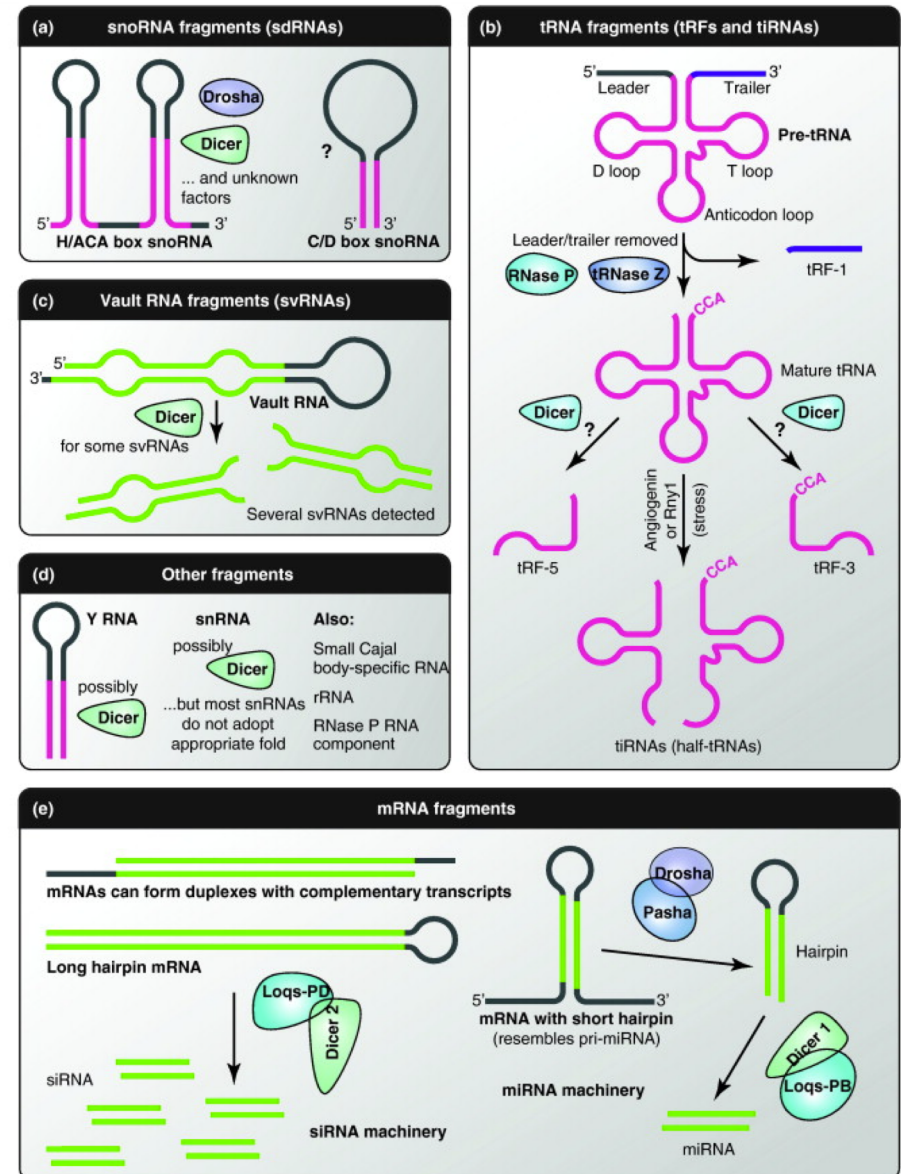
- Uracil vs. thymine

# Different classes of RNA

- mRNA:
  messenger RNA

- tRNA:
  transfer RNA

- rRNA:
  ribosomal RNA

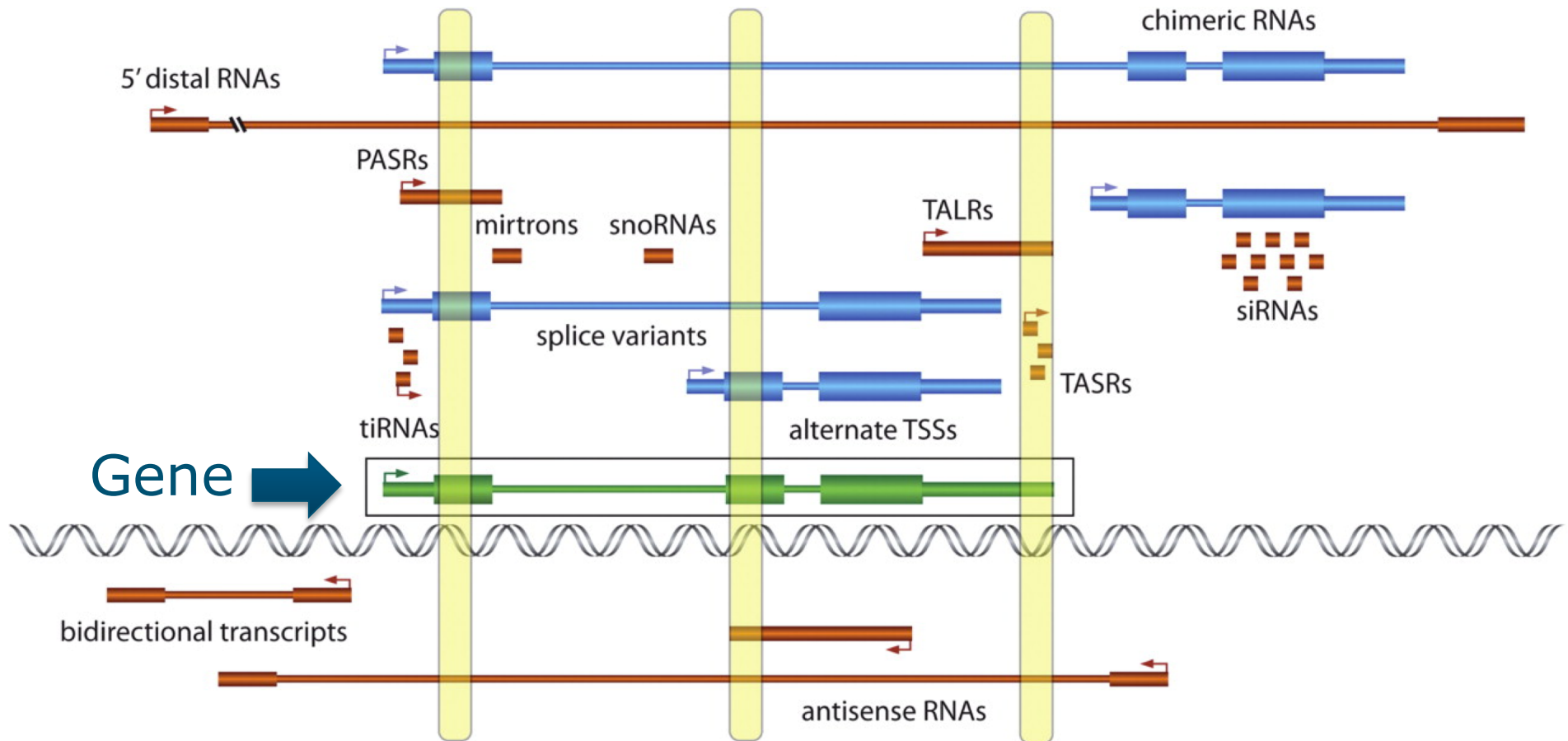- rRNA is the most abundant:
  >80% of total RNA

tRNA

mRNA

rRNA

**Translation**

# Traces of RNA world

- **miRNA:** microRNA
- **siRNA:** short interfering RNA
- **sn(o)RNA:** small nucle(ol)ar RNA
- **lncRNA:** long non-coding RNA
- **piRNA:** piwi-interacting RNA
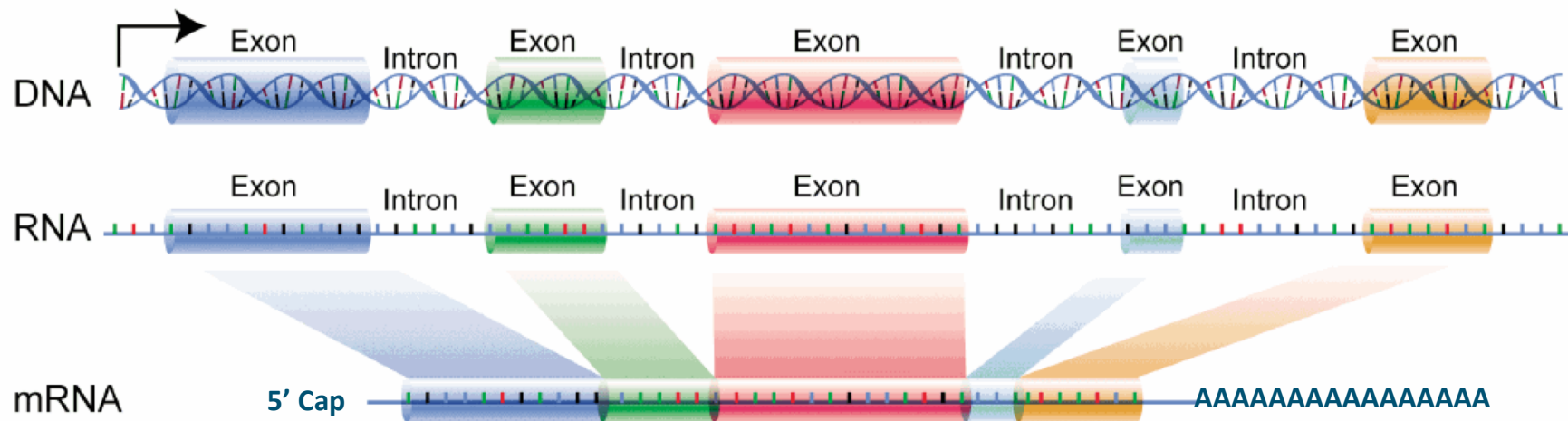- ...

# Pervasive transcription



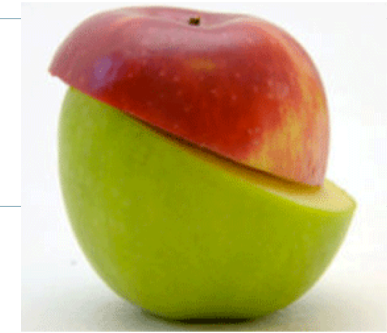- ENCODE project: >80% of human genome is transcribed

# mRNA processing

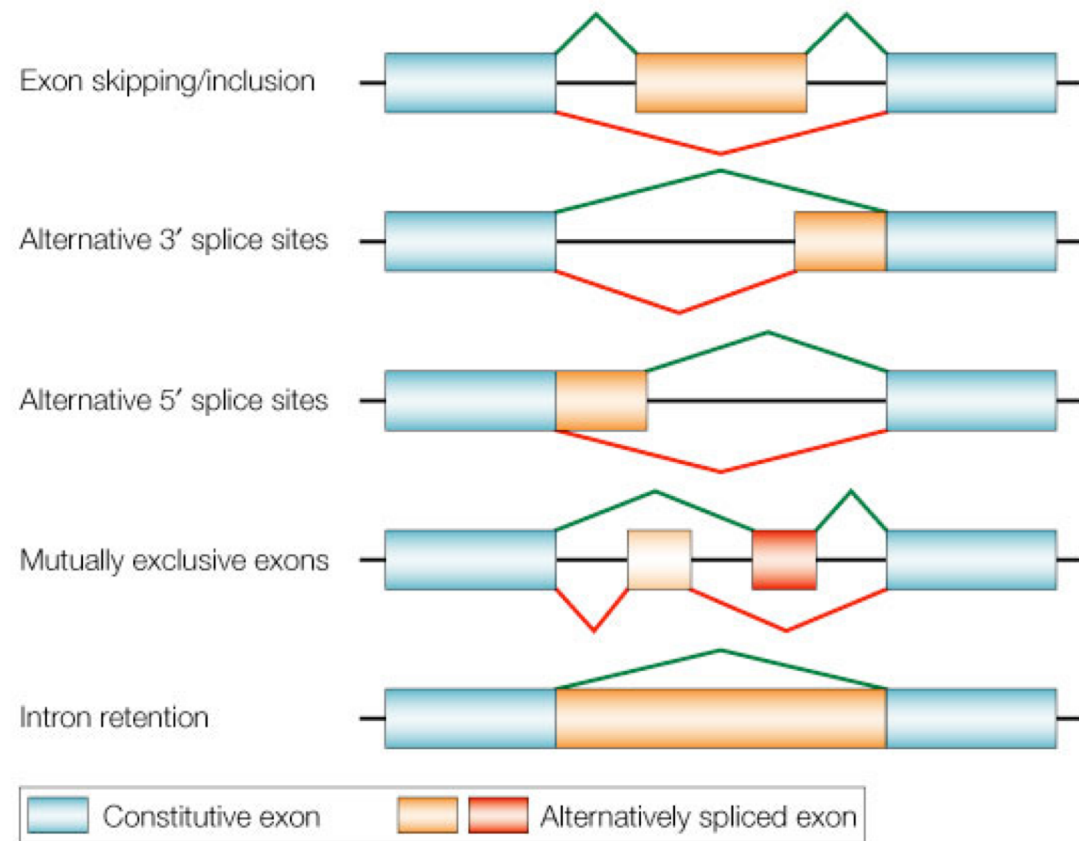- Pre-mRNA is spliced to form mature mRNA, capped (5') and polyadenylated (3')

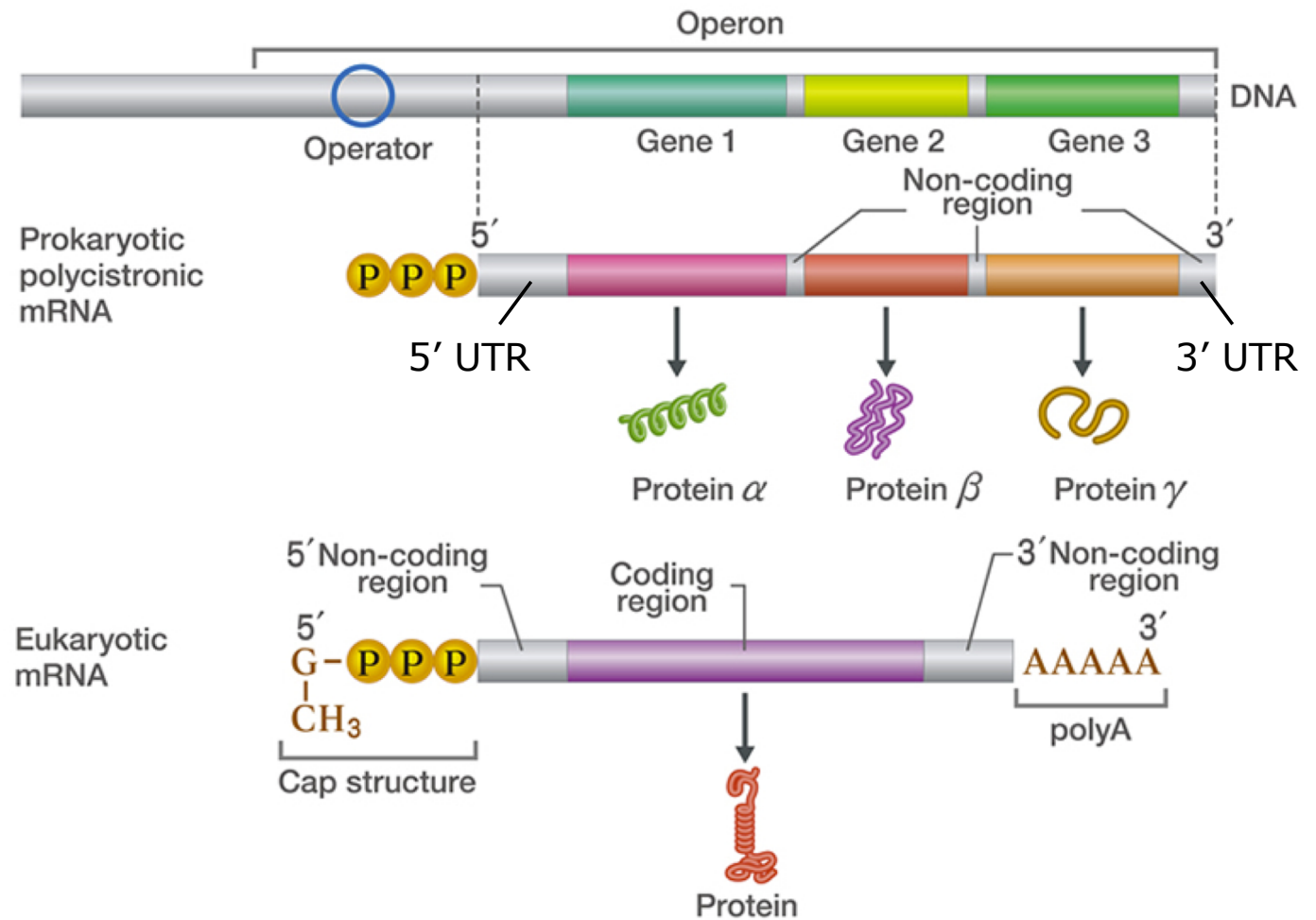  (In prokaryotes, only the latter occurs)

# Alternative splicing

- In eukaryotes, one gene can result in different mRNA transcripts: *isoforms*



Exon skipping/inclusion

Alternative 3' splice sites

Alternative 5' splice sites

Mutually exclusive exons

Intron retention

Constitutive exon    Alternatively spliced exon

# mRNA structure

- In bacteria, one mRNA can translate into multiple proteins: *polycistronic expression*



Operon

Operator — Gene 1 — Gene 2 — Gene 3 — DNA

Non-coding region

Prokaryotic polycistronic mRNA

P P P — 5′ — 3′

5′ UTR — 3′ UTR

Protein α — Protein β — Protein γ

Eukaryotic mRNA

5′ Non-coding region — Coding region — 3′ Non-coding region

G–P P P — 5′ — AAAAA — 3′

CH₃

Cap structure — polyA

Protein

© CSLS/The University of Tokyo
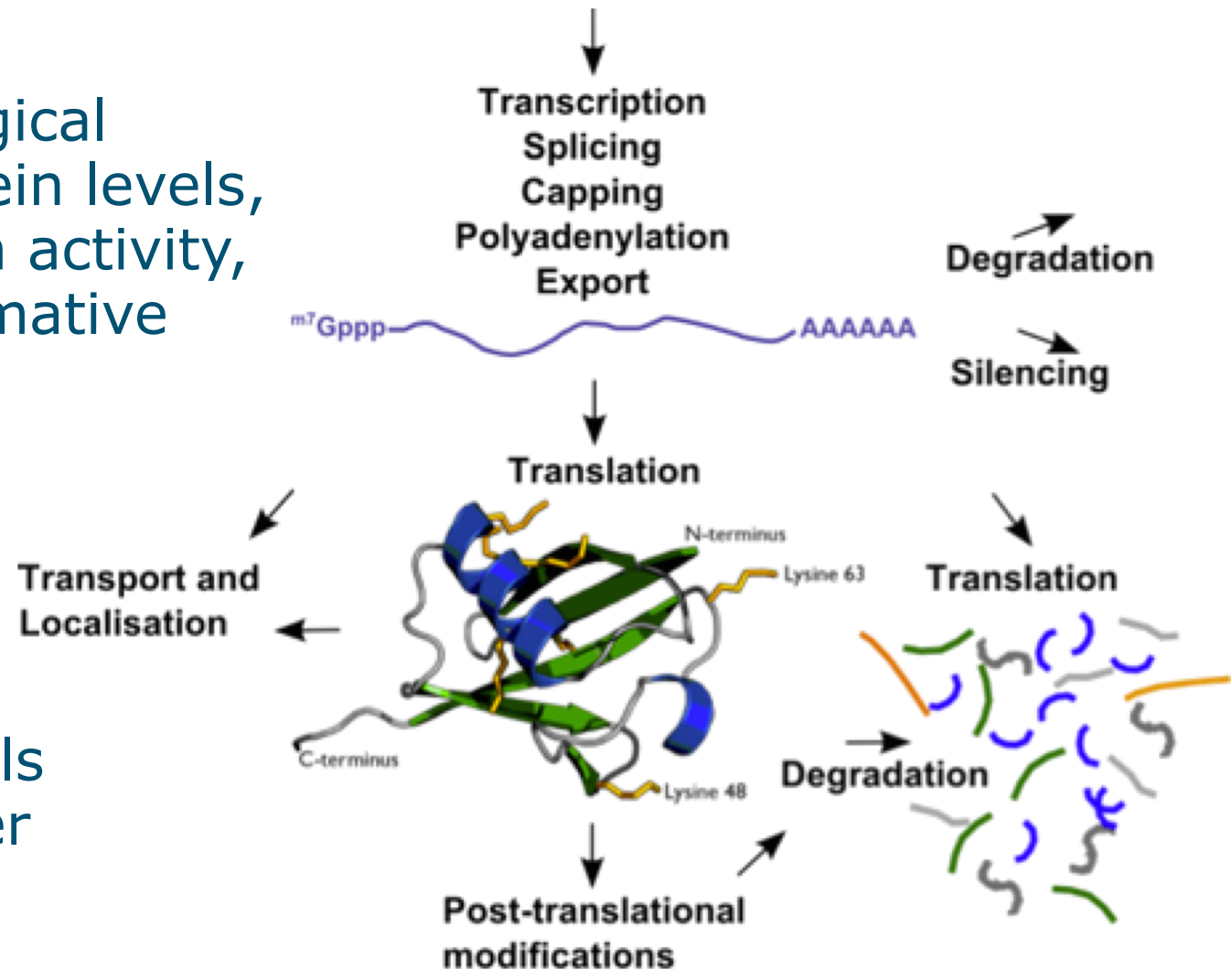
# Translation

- Ribosome (rRNA) uses transfer RNA (tRNA) to translate mRNA into a protein
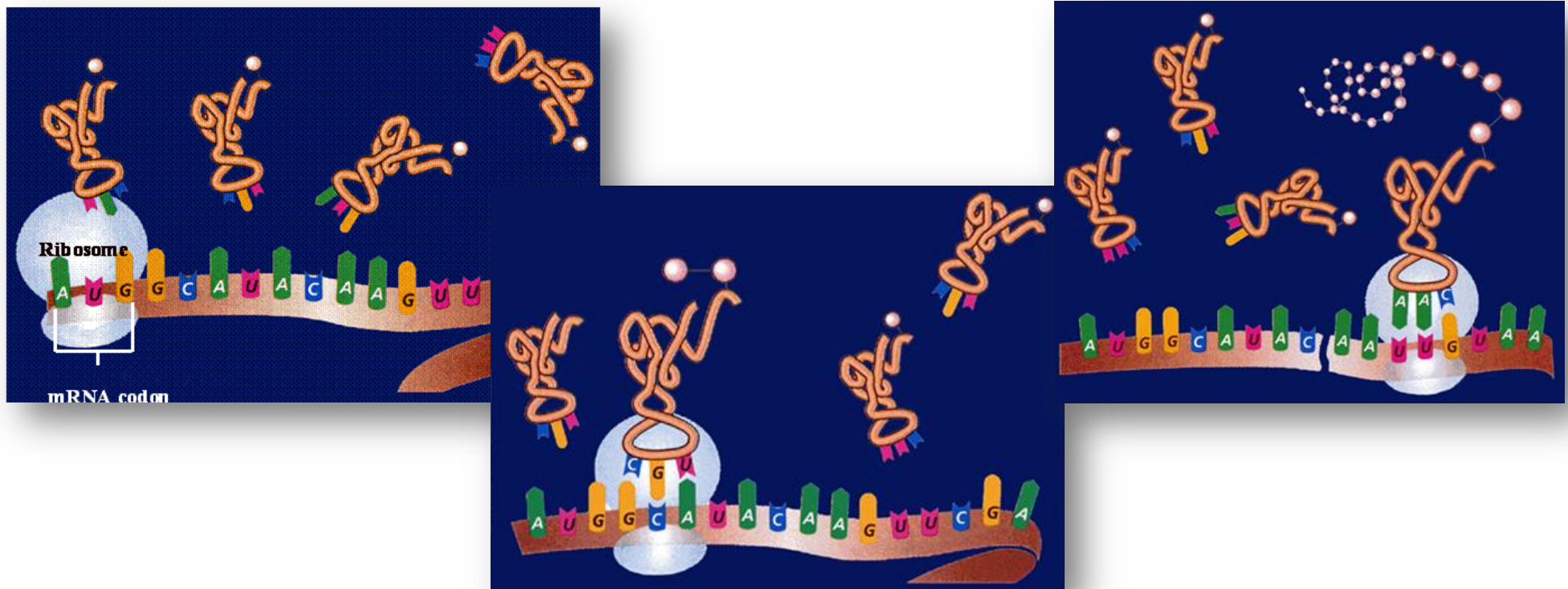
# Why measure mRNAs?

- For most biological questions protein levels, or even protein activity, are most informative
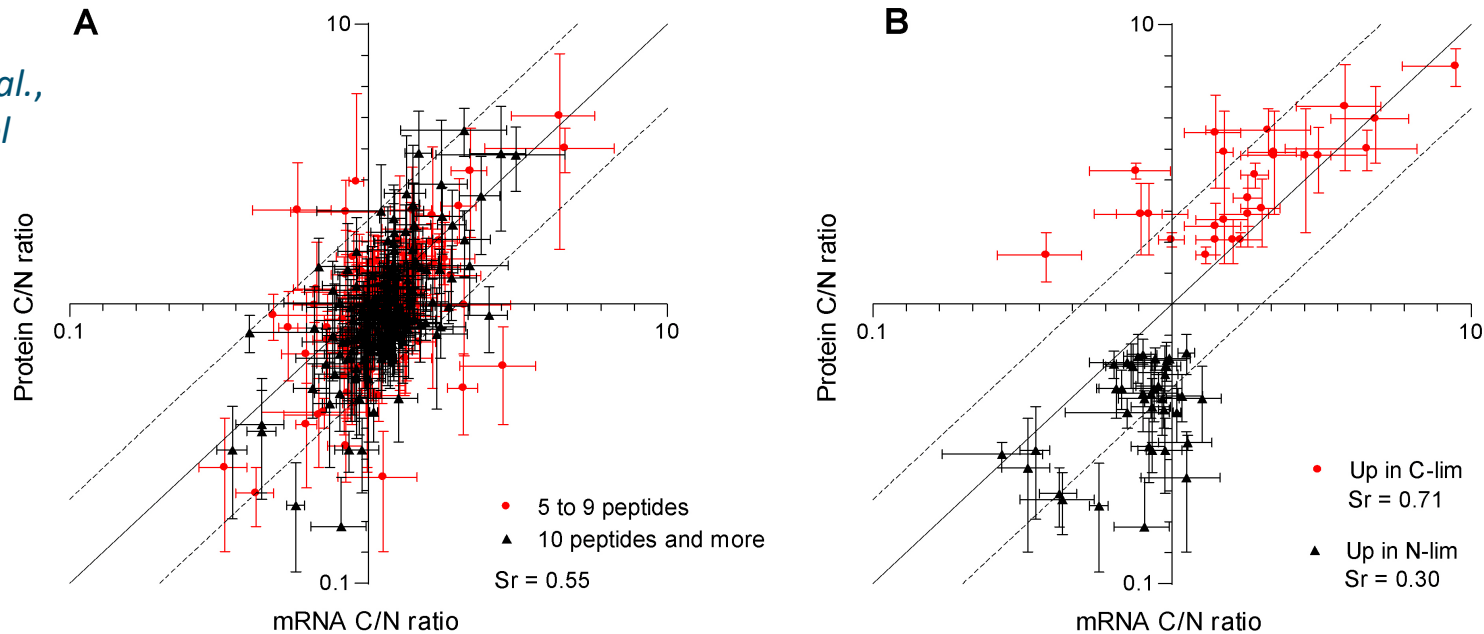
- But mRNA levels are much easier to measure



Transcription
Splicing
Capping
Polyadenylation
Export

$^{m7}$Gppp— — AAAAAA

Degradation

Silencing

Translation

Transport and Localisation

N-terminus
Lysine 63

C-terminus

Lysine 48

Translation

Degradation

Post-translational modifications

# Why measure mRNAs? (2)

- **But mRNA levels often do not correlate with protein levels**

Kolkman *et al.*,
*Mol Syst Biol*
2006



**A**

Protein C/N ratio — mRNA C/N ratio

- 5 to 9 peptides
- 10 peptides and more

Sr = 0.55

**B**

Protein C/N ratio — mRNA C/N ratio

- Up in C-lim
  Sr = 0.71
- Up in N-lim
  Sr = 0.30

- **… and protein levels often do not correlate with activity**

# mRNA measurement

- Transcript measurements:
    - is a specific transcript present?
    - and at what concentration?
- Transcriptomics:
    - what transcripts are present?
    - and at what concentrations?

- Note: mostly averaging over large numbers of cells, but single-cell analysis gaining traction

# How to measure mRNAs?

- Before 1995:
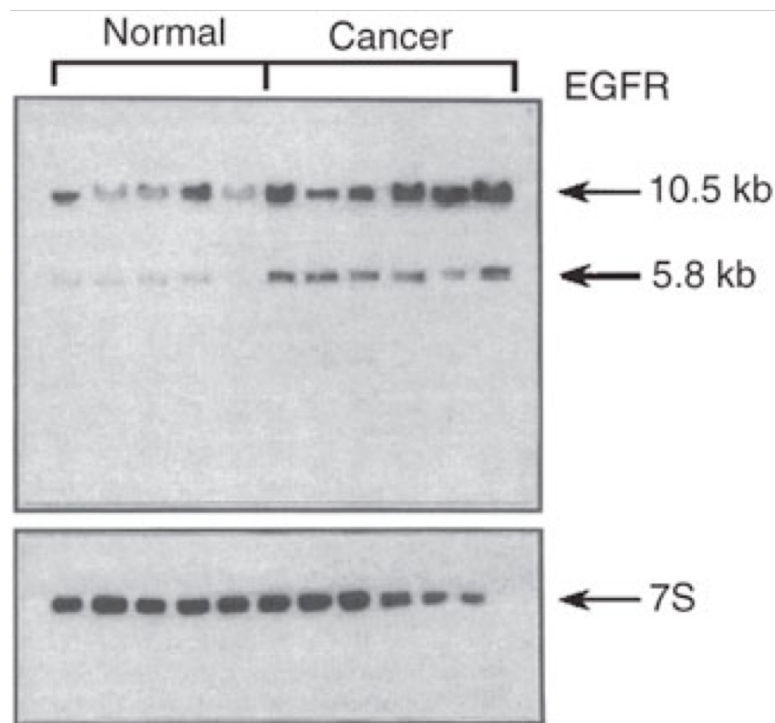


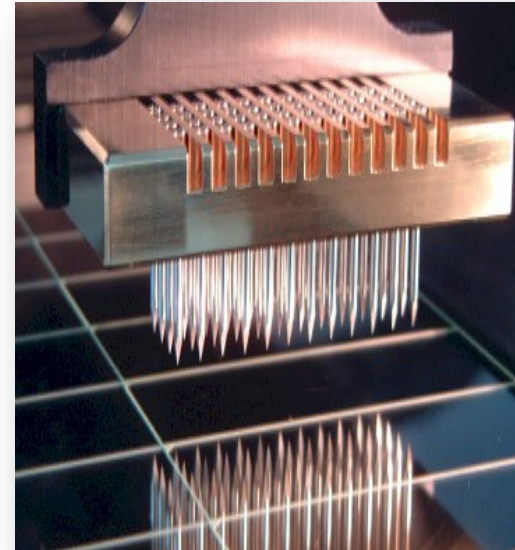Northern blot
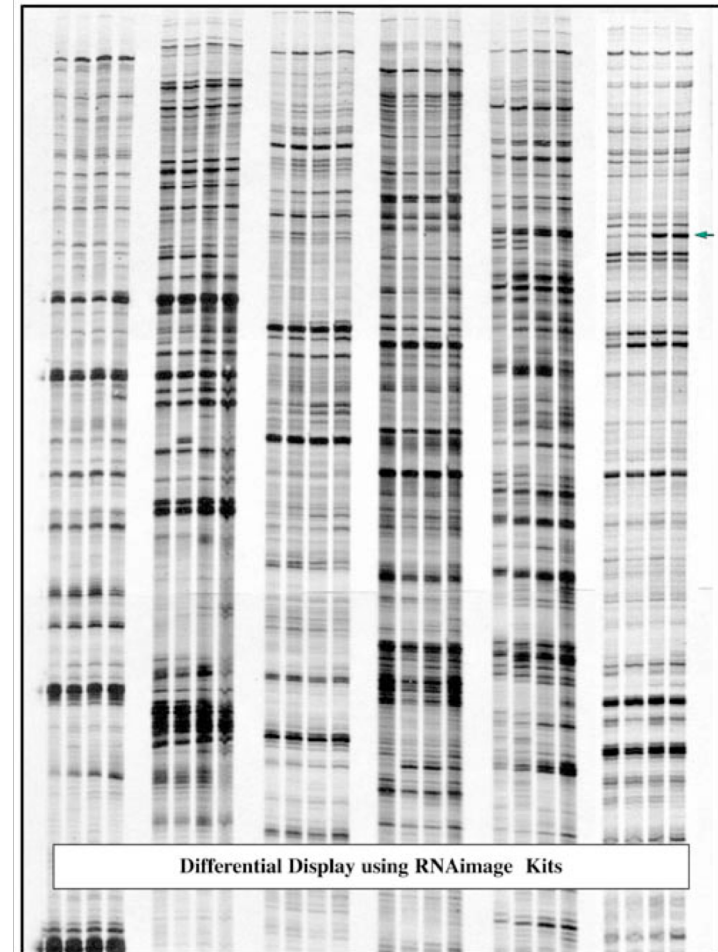


Differential display

# How to measure mRNAs? (2)

- qPCR
  - quantitative polymerase chain reaction
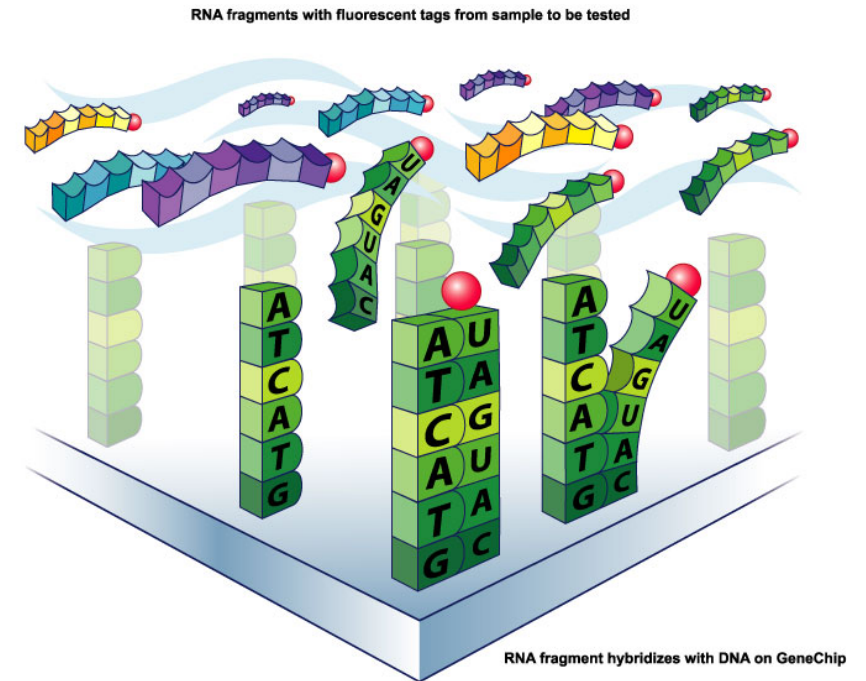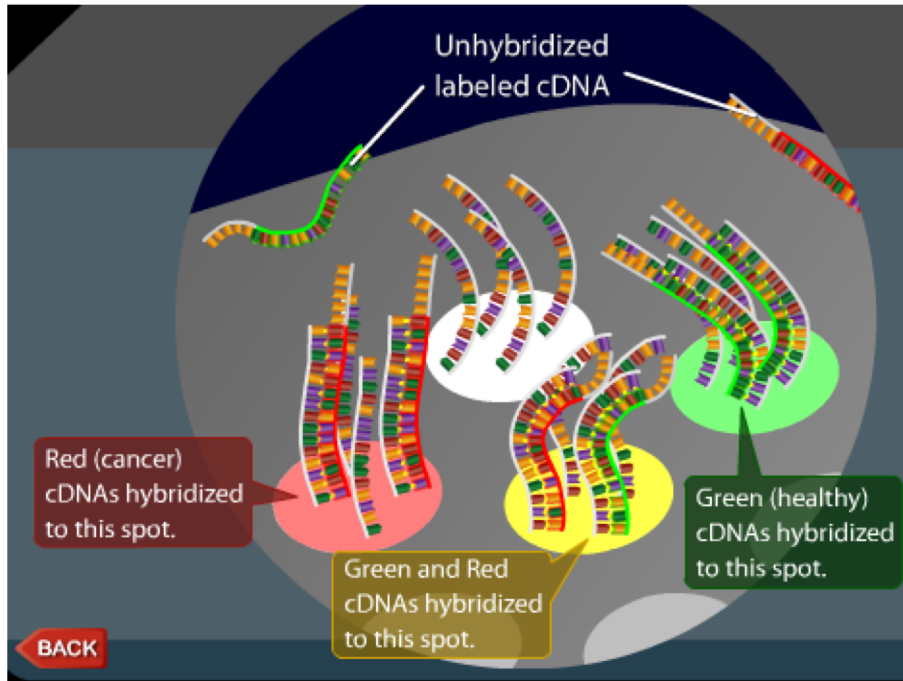  - targeted, only a few genes at a time

- Microarrays
  - hybridization using DNA probes
  - targeted, genome-wide

# Microarrays

- Widely used for gene expression profiling since 1995

- DNA probes fixed to a surface 'catch' specific DNA/RNA molecules that are labeled
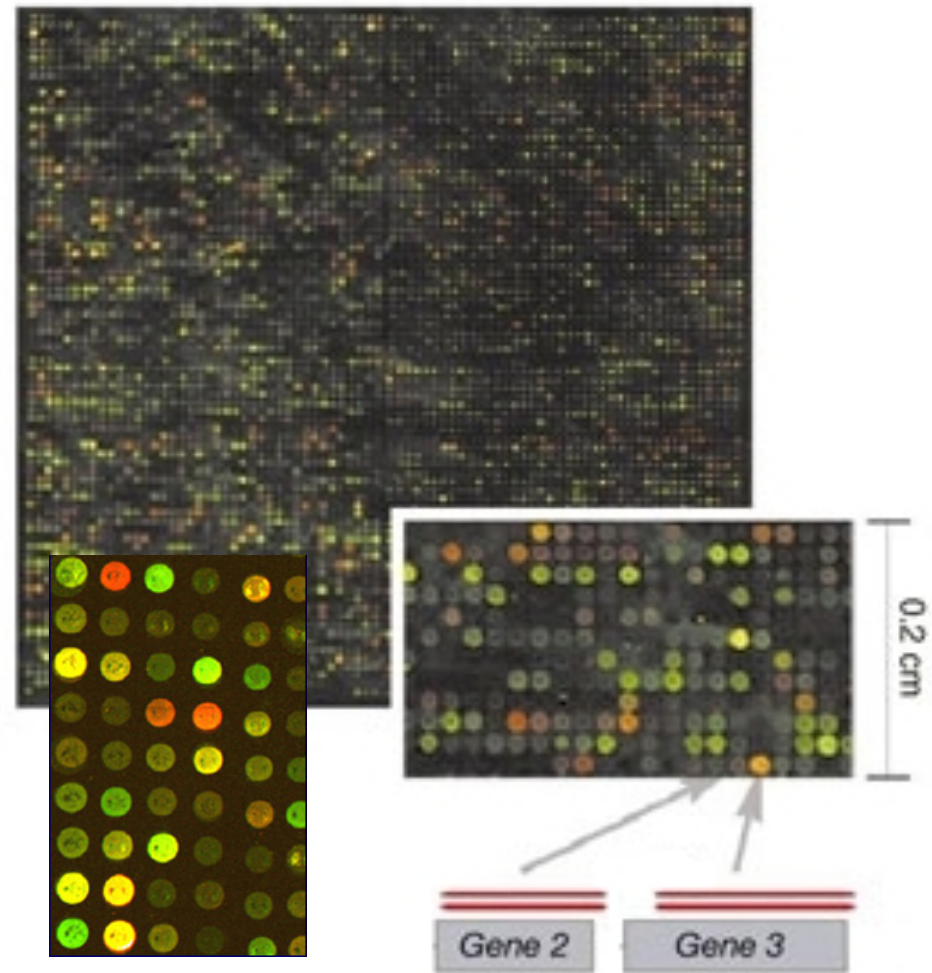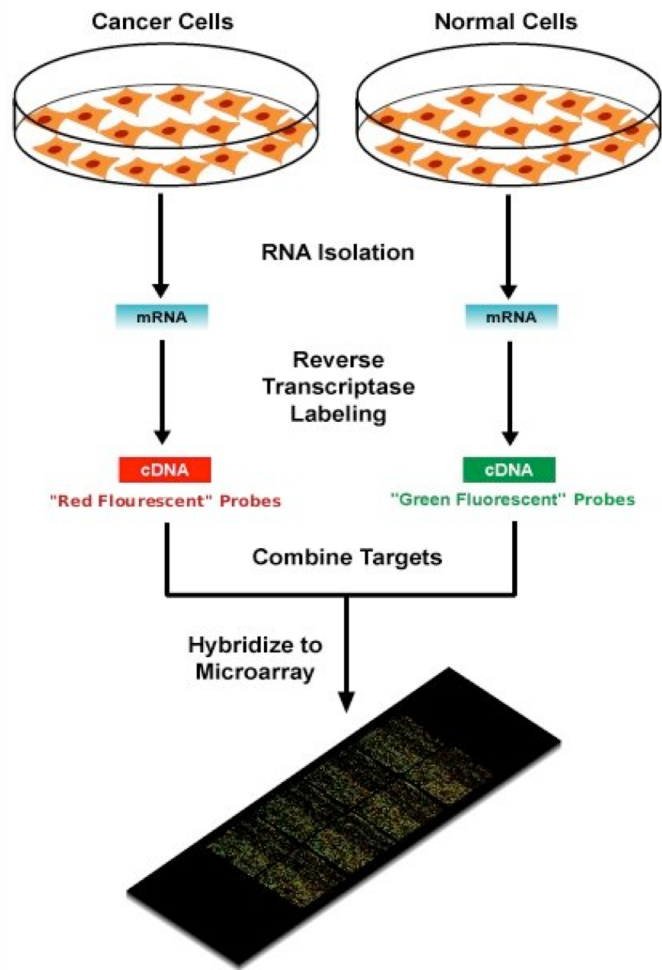
- The amount of label is measured by fluorescence



**AFFYMETRIX**

**GeneChip®**
Soybean Genome Array

P/N: 520119
Lot #: 4035168
Exp. Date: 10/05/08
For Research Use Only

WAGENINGEN UNIVERSITY
WAGENINGEN UR

# Microarrays (2)



Unhybridized labeled cDNA

Red (cancer) cDNAs hybridized to this spot.

Green and Red cDNAs hybridized to this spot.

Green (healthy) cDNAs hybridized to this spot.

BACK

RNA fragments with fluorescent tags from sample to be tested

RNA fragment hybridizes with DNA on GeneChip

- Two-color cDNA arrays:
  - 500-1000nt probes
  - competitive binding
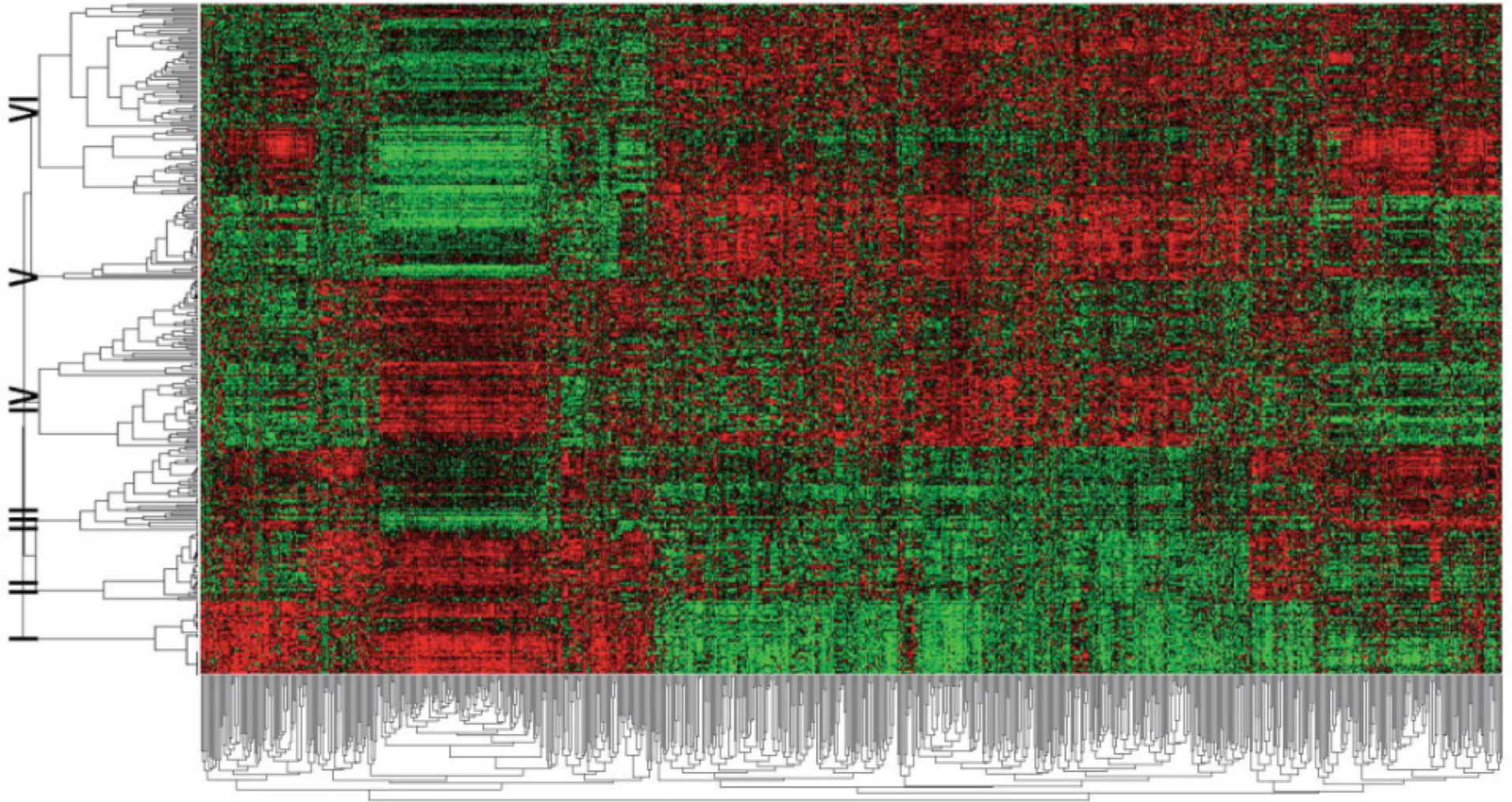  - color: relative expression, intensity: absolute expression

- Single-color oligo arrays:
  - multiple 25-60nt probes
  - intensity: absolute expression

WAGENINGEN UNIVERSITY
WAGENINGEN UR

# Microarrays (3)

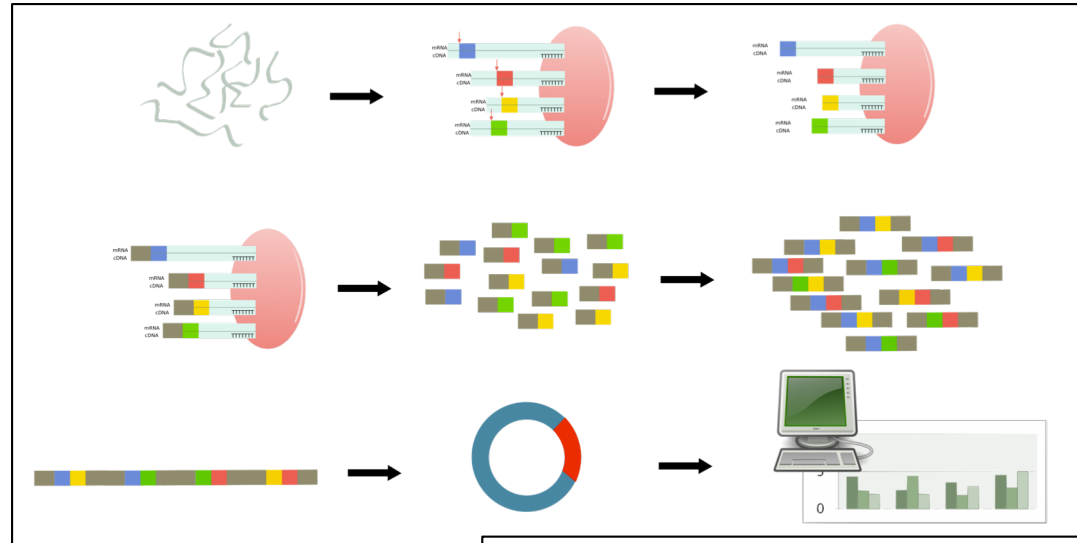# Example: breast cancer subtypes

# Microarray compendia

# Microarrays (4)

+ Highly standardized

+ Relatively cheap


– Gene sequence/structure should be known
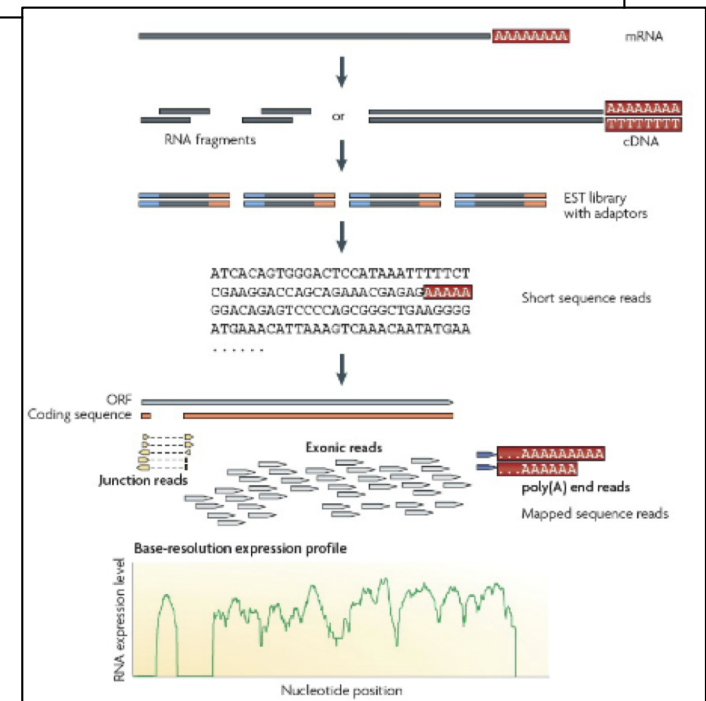
– Cannot detect new genes, isoforms

# mRNA sequencing

- SAGE
  - serial analysis of gene expression
  - genome-wide, reference-based

- **RNA-seq**
  - NGS-based
  - untargeted, genome-wide
  - affordable

# Take-home

- There are many types of RNA molecules

- The genome is pervasively transcribed,
  but mostly we are interested in mRNAs

- Transcript measurements are used as proxies
  for protein levels and subsequent functions

- Microarrays are still widely used as they are cheap
  and standardized, but require a genome

- RNA-seq is the current method of choice, particularly
  for unsequenced genomes and for isoforms