Measuring gene expression

Ribosome profiling

What is ‘gene expression’?

Wikipedia: Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. These products are often proteins, but many genes have been discovered to encode RNAs that are not translated, but exert their functions as RNAs.

These RNAs are called non-coding RNAs.

Examples: transfer RNA (tRNA – function in mRNA translation), small nuclear RNA (snRNA - splicing), small nucleolar RNAs (snoRNAs – rRNA modification), microRNAs (miRNAs – regulation of mRNA expression).

‘Measuring’ gene expression would therefore mean determining the mass/concentration of functional gene products within different types of cells and in different conditions.

Methods for measuring RNA level gene expression gene expression

Low throughput
- Northern blot
- Differential display
- Quantitative PCR

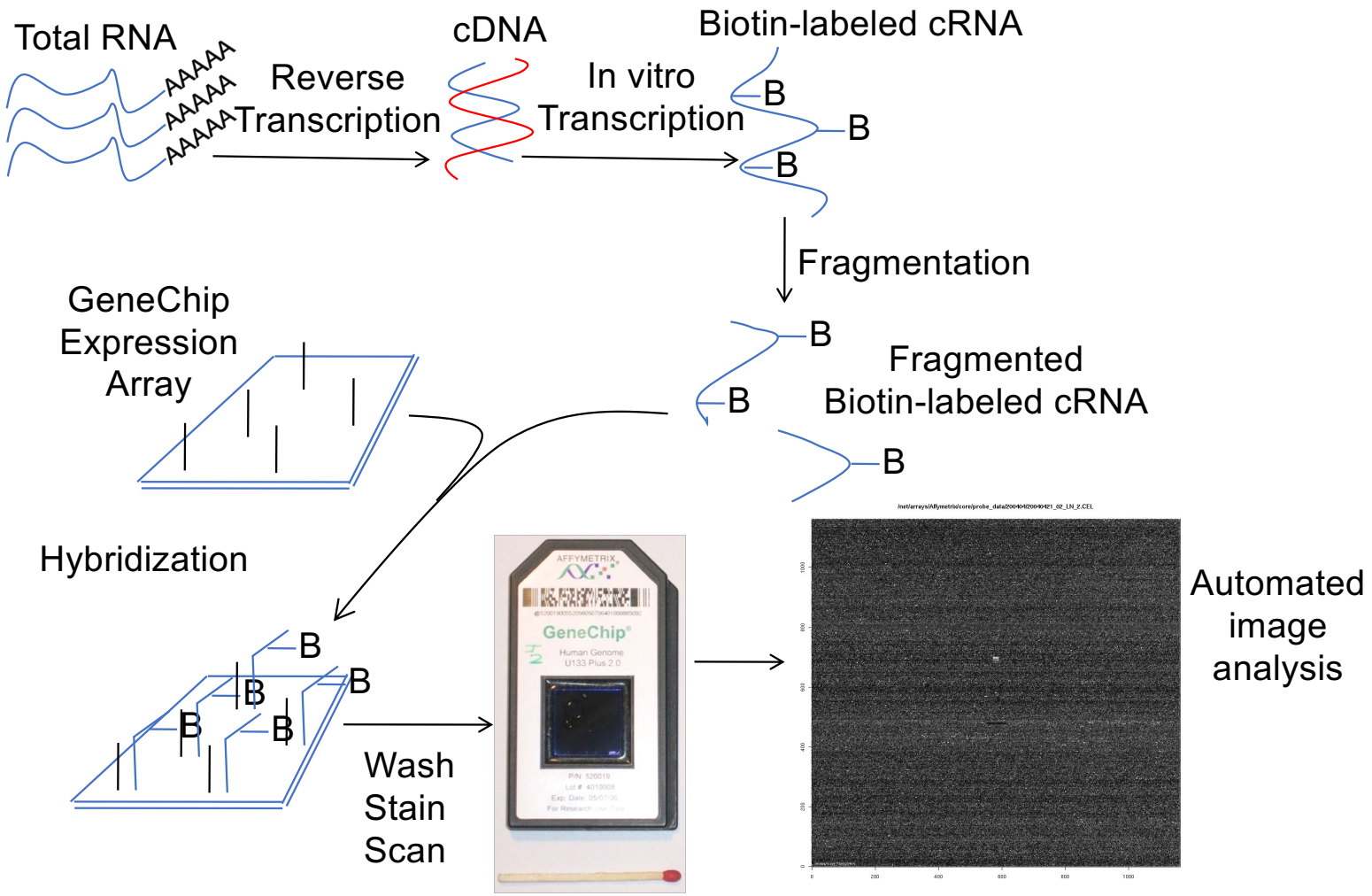
High throughput
- Microarrays
- Sequencing

Methods for measuring RNA level gene expression gene expression

Affymetrix microarrays

- Probes: short (25 nucleotides-long), high-density sequences
- Probe sets: 11-20 probes corresponding to one gene
- 54’000 probe sets per chip

Measuring gene expression with Affymetrix microarrays



Sequencing-based methods

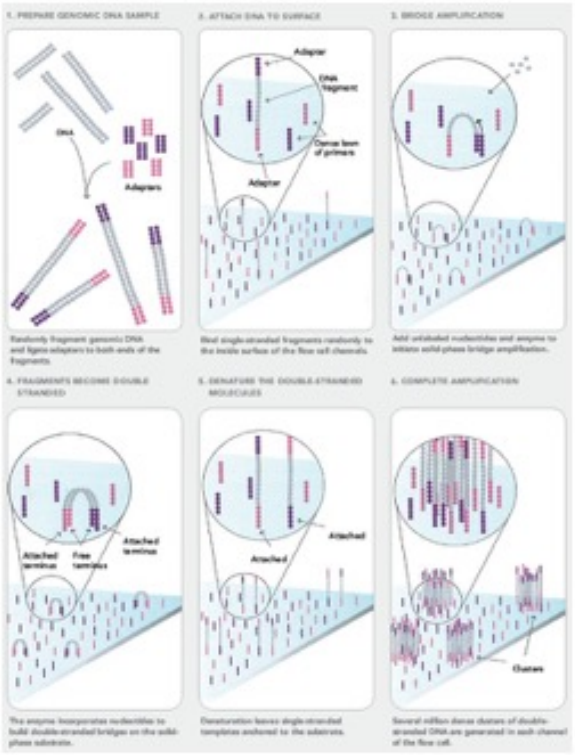
Some limitations of microarrays:

1. Need a fair amount of material (RNA)
2. Hybridization-based method -> cannot detect transcripts that are in low abundance or are repetitive
3. Need to design the probes -> can only detect the genes that are known, for which probes can be designed.

Sequencing-based methods:

1. Can detect expression of yet unknown transcripts
2. Currently, expression profiling can be done from single cells.

Short read sequencing on the Illumina platform



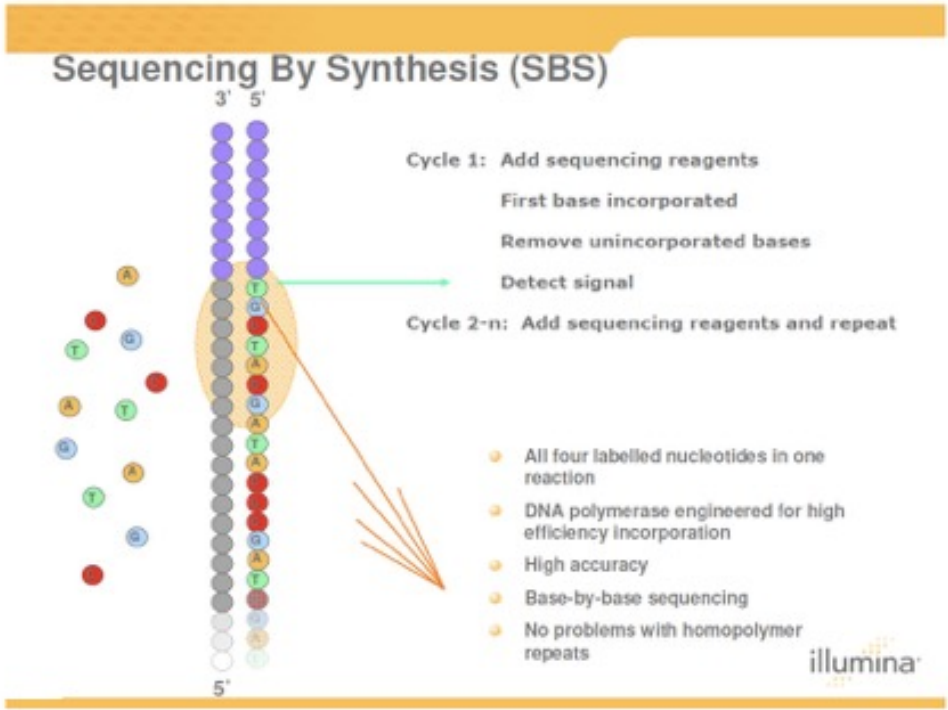
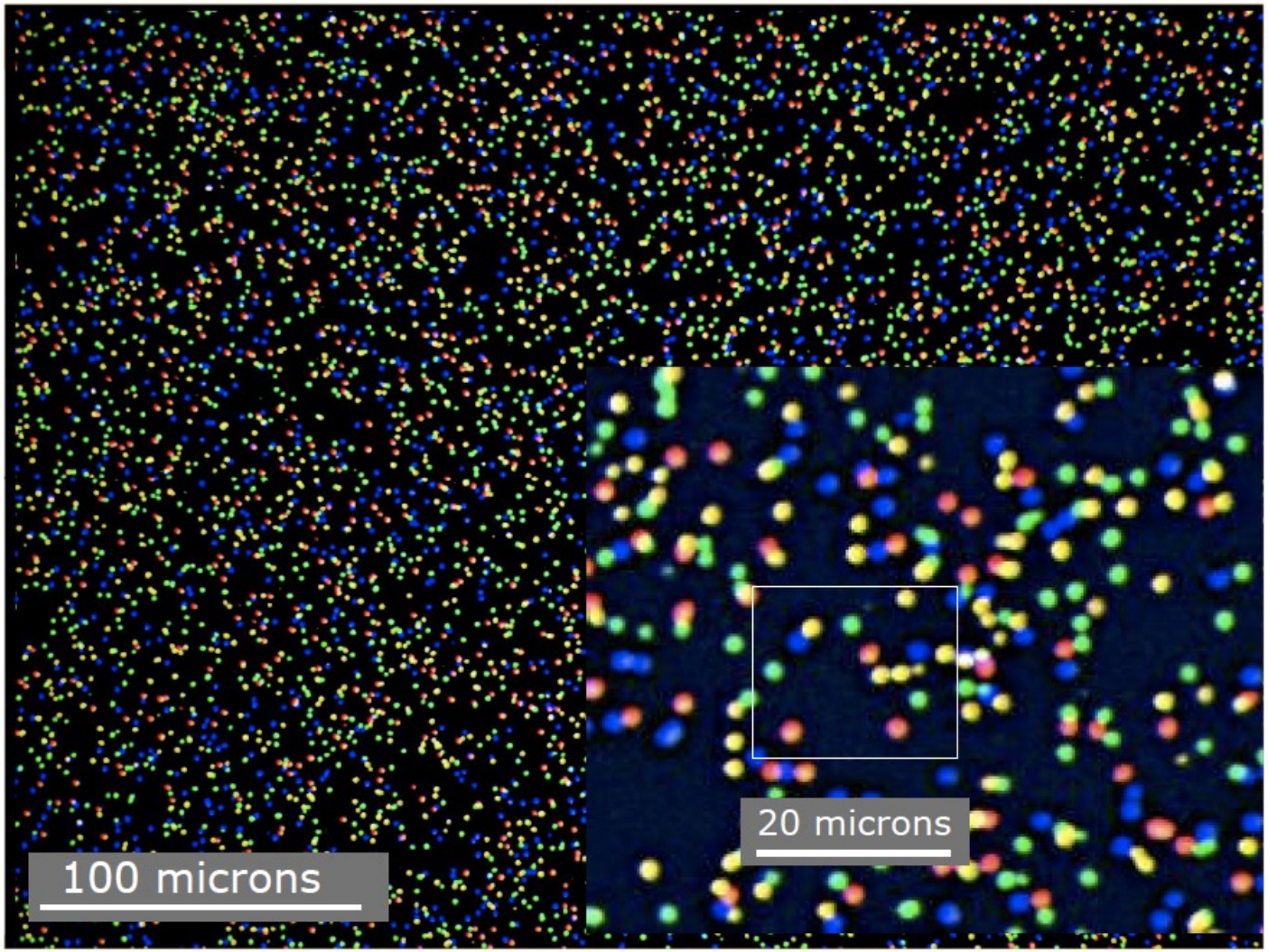
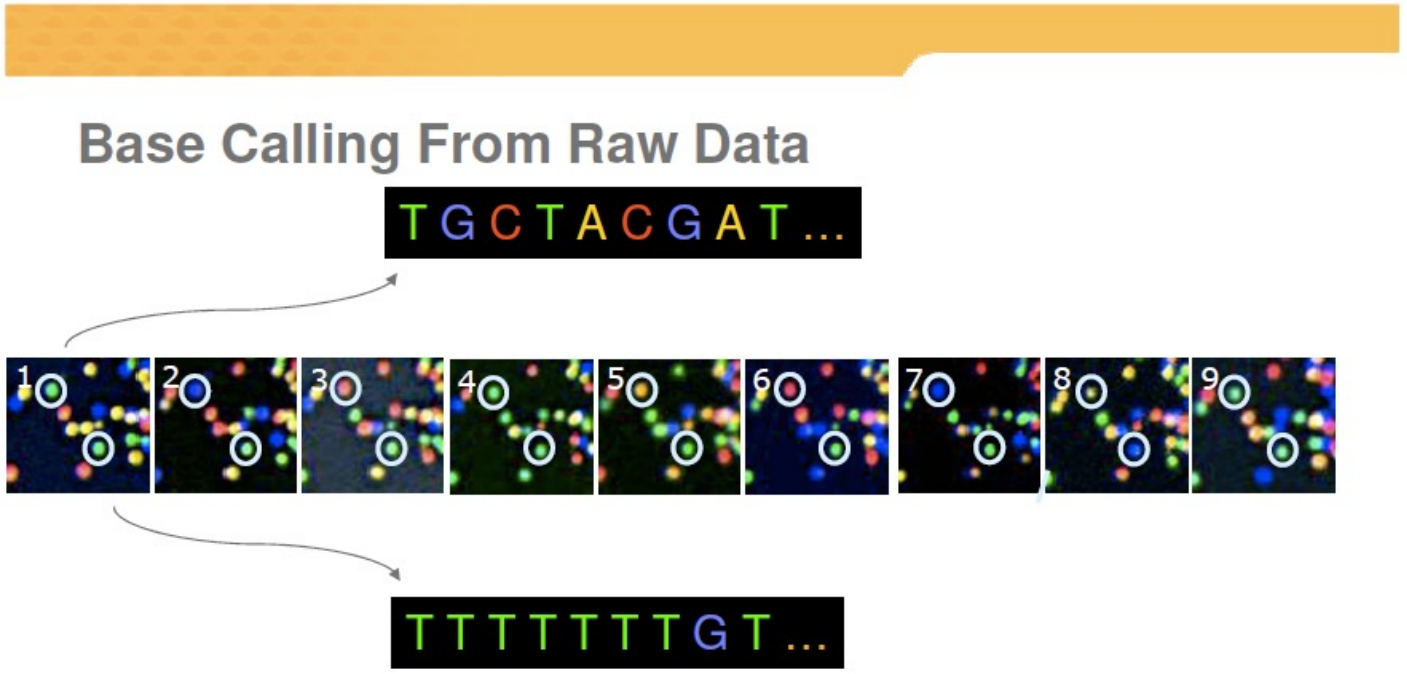


Image analysis



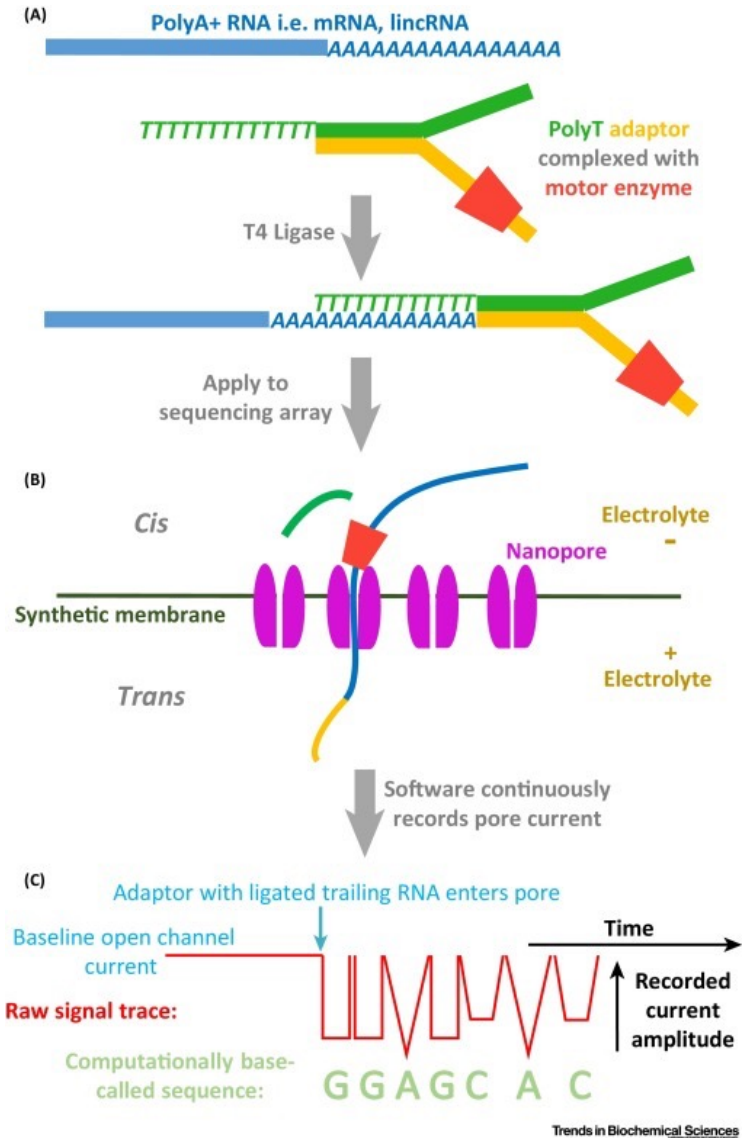
Software associated with the sequencer

From images to sequences



The identity of each base of a cluster is read off from sequential images

illumina

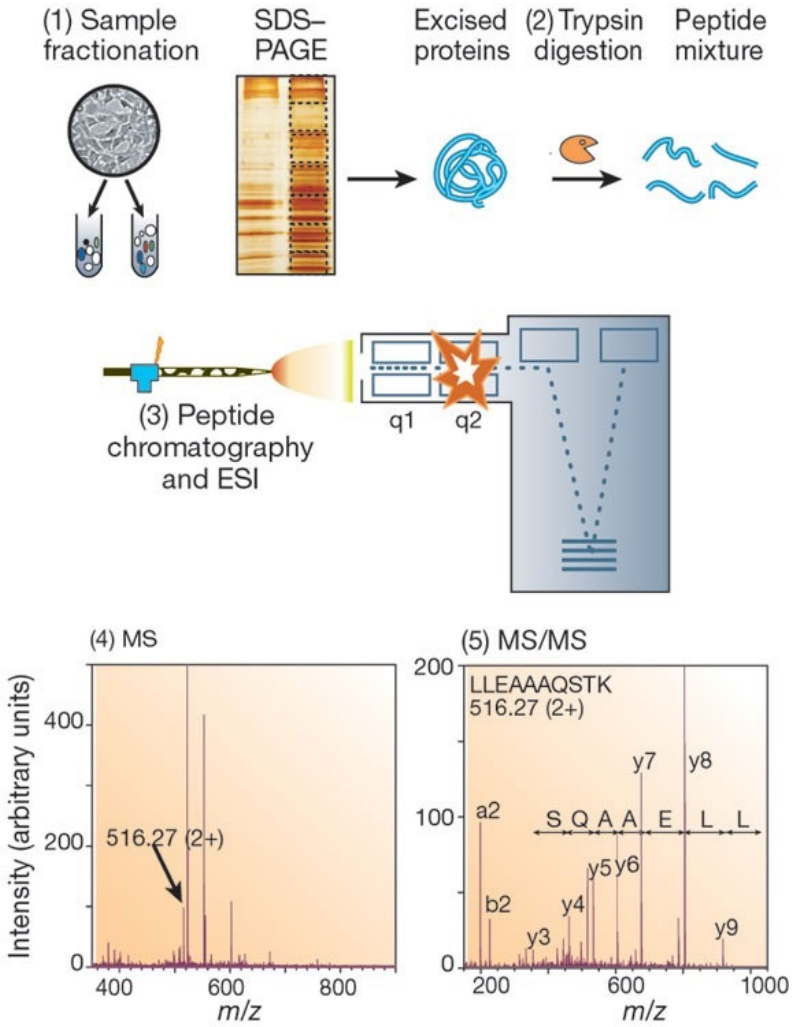


43(4):225-227 (2018)

Direct ‘reading’ of RNAs with long read sequencing

- Can detect also RNA modifications
- So far throughput is much lower compared to DNA sequencing

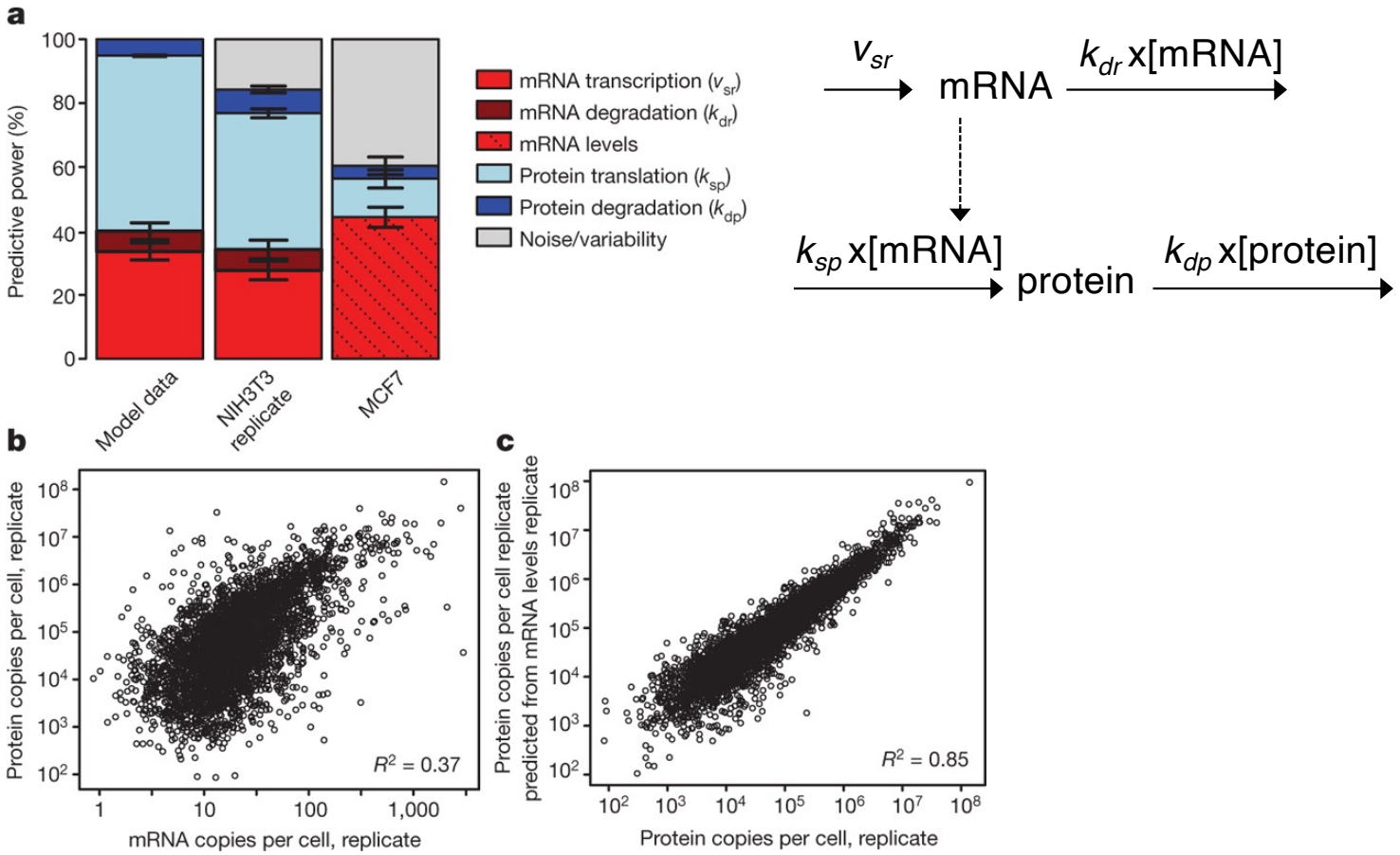
High-throughput measurement of protein-level expression



Nature 422: 198-207 (2003)

- Many transformations of the sample, hard to get reproducible estimates across experiments
- More limited coverage of the expressed genes compared to RNA-seq

Is the mRNA abundance a good ‘proxy’ for protein abundance?



Schwanhäusser et al. Nature 2011

Is the mRNA abundance a good ‘proxy’ for protein abundance?

Not really…

What to do about this?

- Measure directly protein levels by tandem mass spectrometry -- drawbacks
  - Limited coverage of the expressed proteome
  - Dependence on database of ions
  - Complex issues with sample preparation for different tissue types

- Ribosome profiling

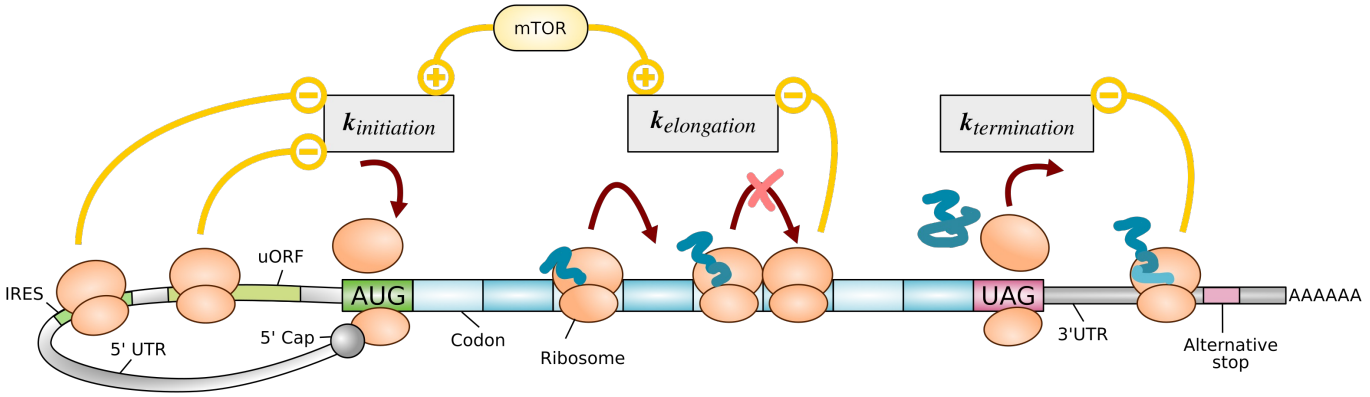
– potential advantages

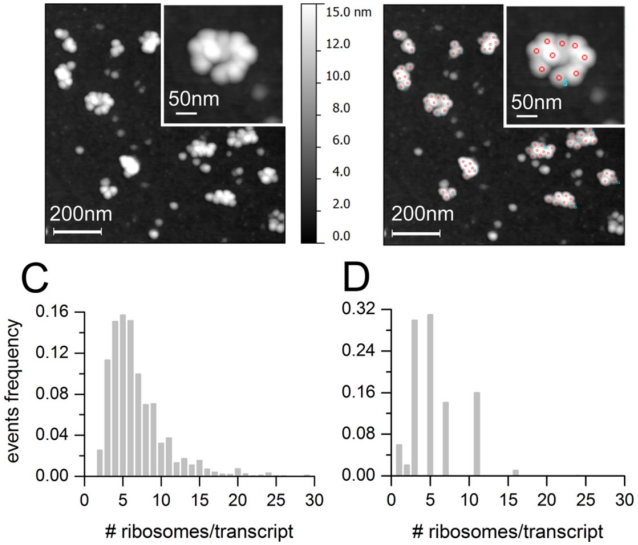
  - Better coverage of the set of expressed genes
  - Any genomic region undergoing translation can be detected (no need to rely on a database)

– potential drawbacks

  - Assumes minimal ribosome ‘queueing’
  - Assumes minimal variation in elongation rate
  - Provides estimates of ribosome densities along transcripts, not of protein levels or synthesis rates

mRNA translation



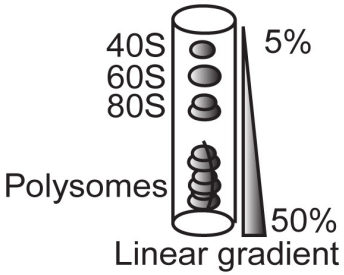


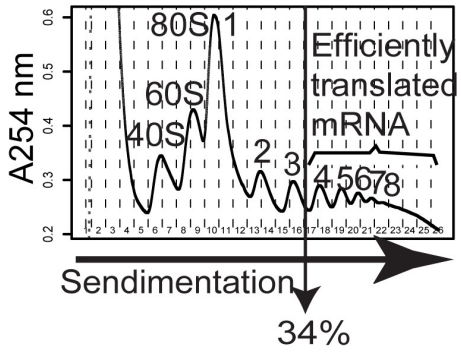
Lauria et al. Nucl. Acids Res. 2015

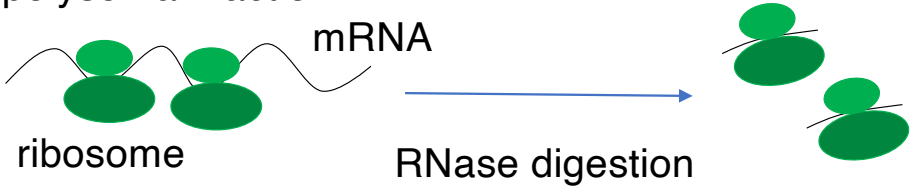
Polysome and ribosome profiling

A

Liang *et al. Nucleic Acids Res* 46:9 January 2018

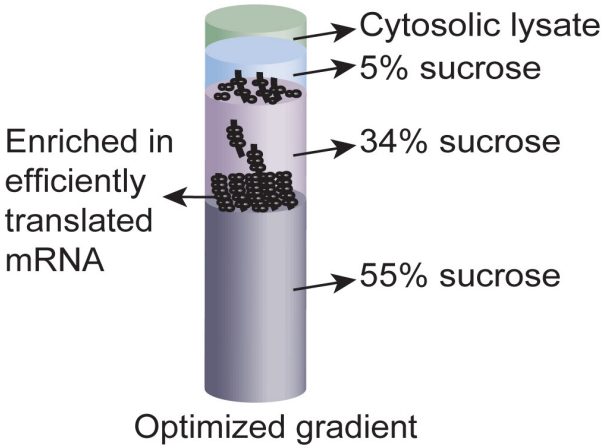






Sequencing,
Analysis

B



Genome browser view of mapped reads