

An aerial, top-down view of a large, diverse crowd of people of various ages and ethnicities, scattered across a white surface. The people are wearing a wide range of colorful clothing, and their shadows are cast on the ground, suggesting a bright, overhead light source. The crowd is more densely packed in the upper right and lower right areas, with more space in the upper left and lower left.

Functional Genomics Overview

RORY STARK

PRINCIPAL BIOINFORMATICS ANALYST

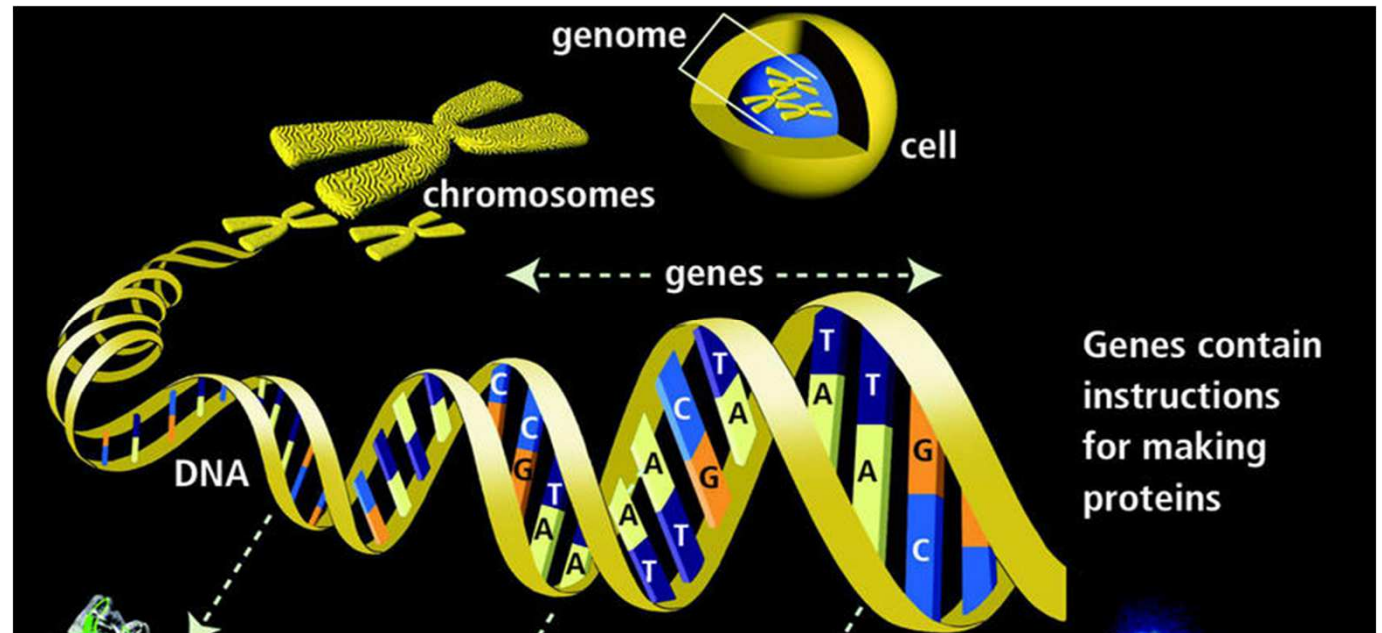
CRUK – CAMBRIDGE INSTITUTE

18 SEPTEMBER 2017

Agenda

- What is Functional Genomics?
- RNA Transcription/Gene Expression
- Measuring Gene Expression
 - Microarrays
 - High-throughput Sequencing
- Transcriptional Regulation
 - Transcription factors
 - Epigenetics
 - Post-transcriptional regulation

The Genome



- Each cell contains a complete copy of the genome, distributed along chromosomes (compressed and entwined DNA)
- 3×10^9 (3Gb) base pairs in human DNA: 6 meters in each cell!
- Encodes blueprint for all cellular structures and activities and which cells go where (somehow...)

Functional Genomics: Sequence vs. Function

What accounts for the difference in phenotype?



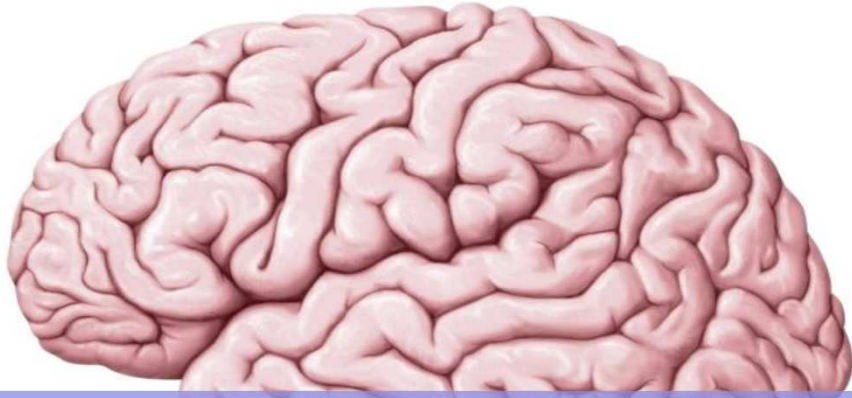
Different Genomes!



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE

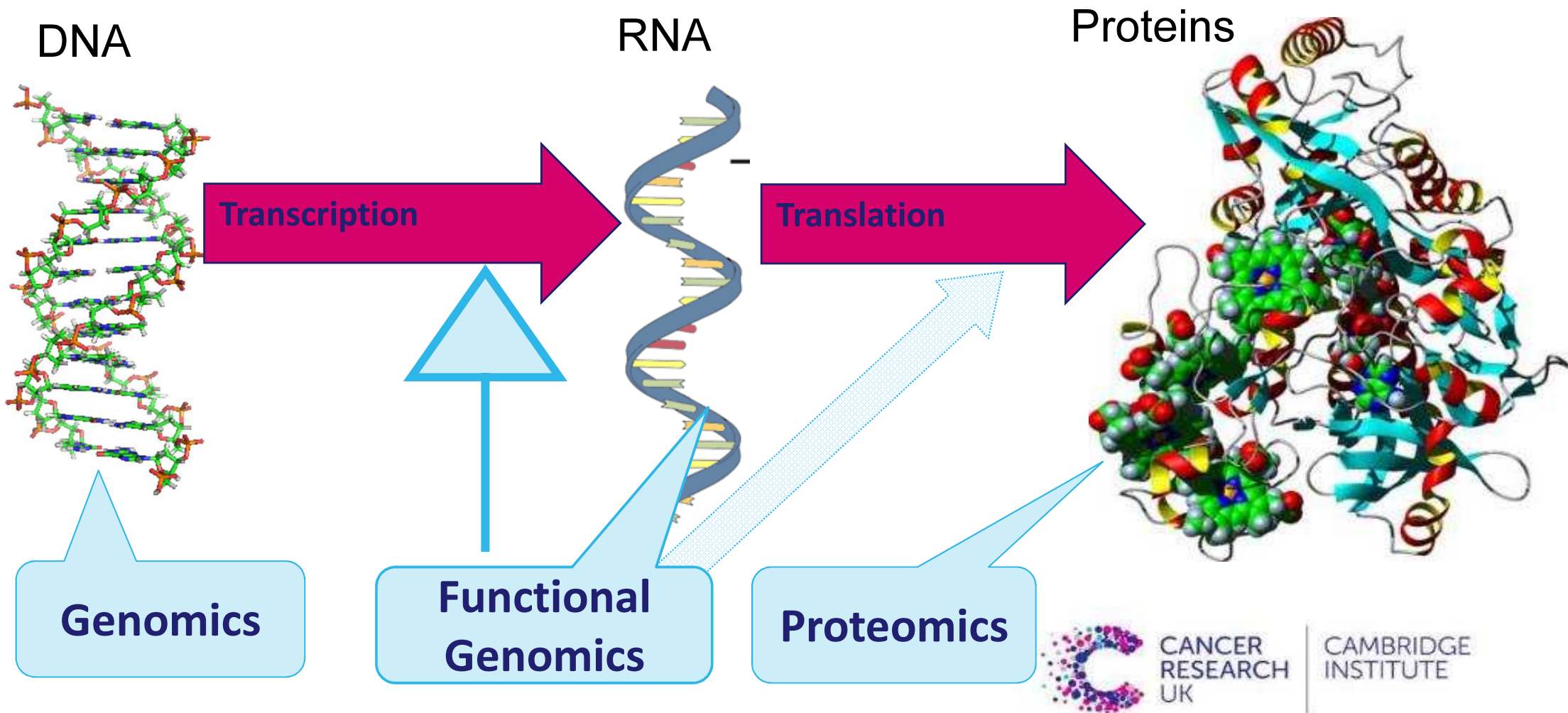
What accounts for the difference in phenotype?



Different Functions!



The Central Dogma of Molecular Biology



So, what is functional genomics?

- Where *sequence-based genomics* looks at the structure and components of genomes, and analyses the similarities and differences between genomes...
- *Functional genomics* looks at **how genomes result in cellular phenotypes**, and analyses differences in how the same genome functions differently in different cells, and how changes in genomes alter function

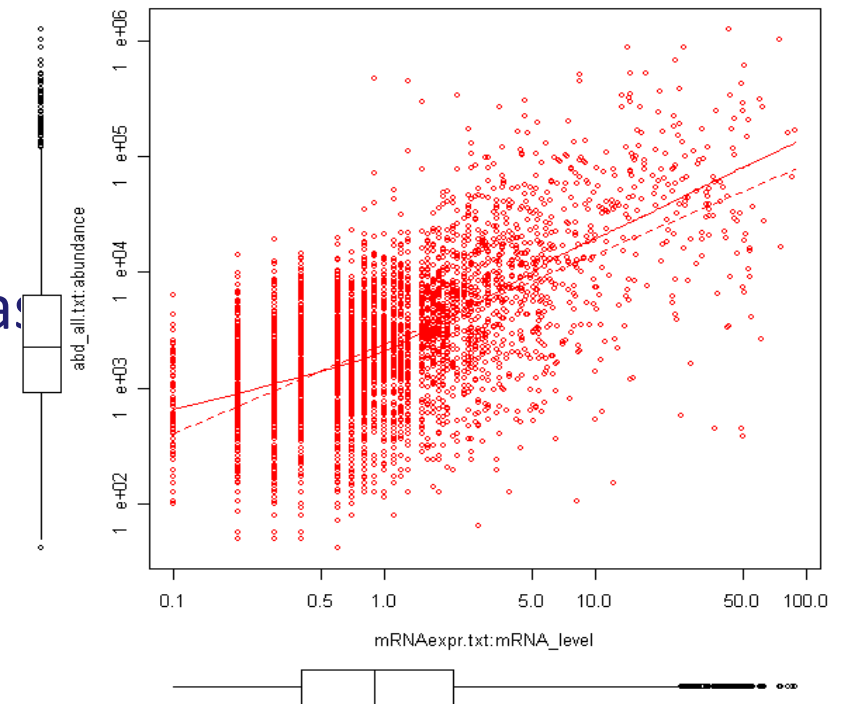


Gene Expression (RNA Transcripts)

Gene expression experiments

- Measure the **expression** levels of many genes in parallel
- Ideally, we'd measure all protein levels
- However, proteomics is difficult!
- Instead, use mRNA (“transcript”) levels as a **proxy** for protein levels
- (How good a proxy is RNA?)
- Several good ways to measure RNA
- Analyses:
 - Expression levels
 - **Differences in expression levels (DE)**
 - Patterns of expression
 - Splicing and isoforms

mRNA vs Protein levels in Yeast
R=0.44



Ghaemmaghami et al Nature 2003



What kinds of samples are we interested in?

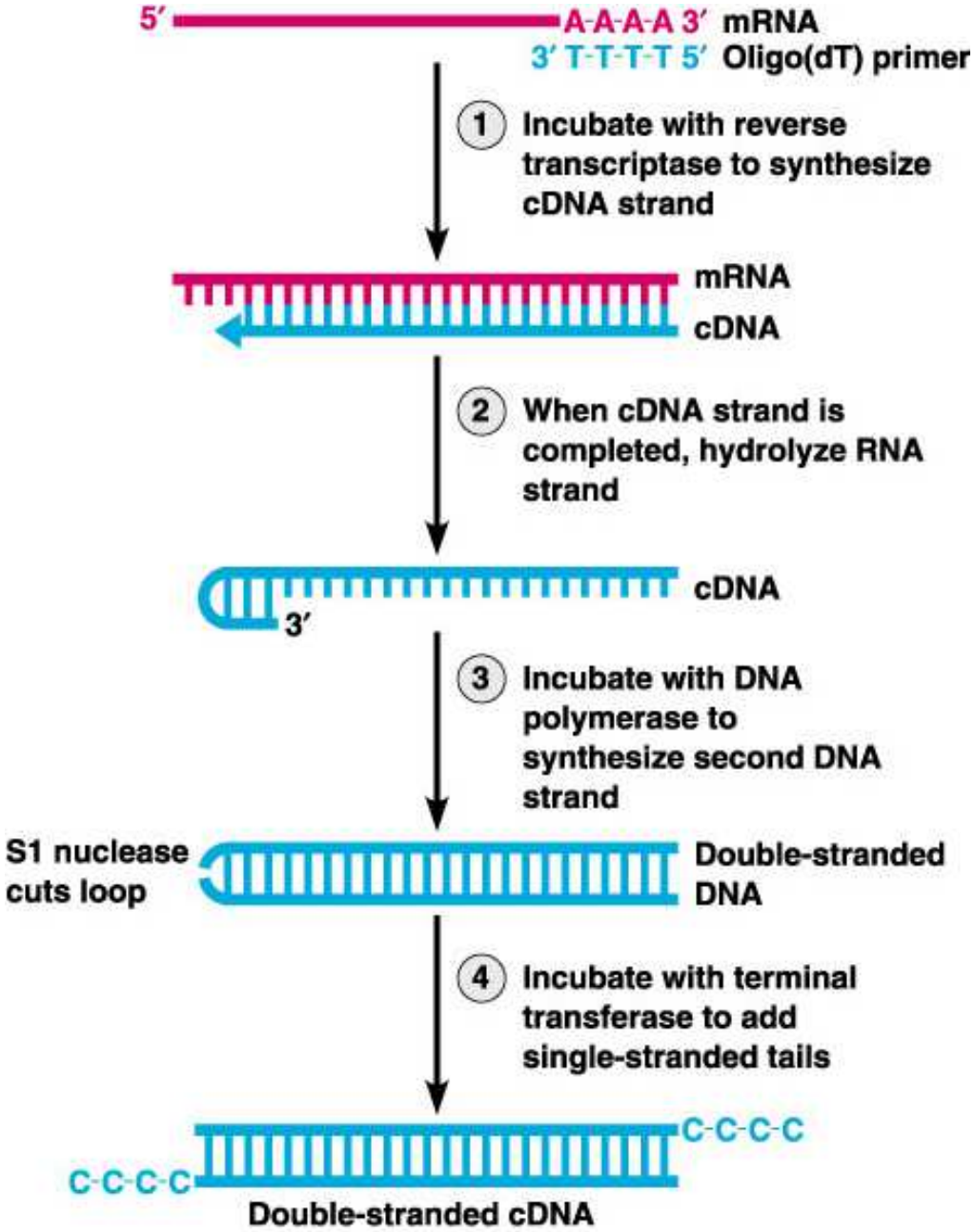
- Different **tissues**, same organism
 - human brain/human liver
- Same tissue, different **organism**
 - human liver/mouse liver
 - wt/ko
- Same tissue, same organism, different **condition**
 - benign/tumour
 - treated/untreated
- **Time** course (effect of treatment over time)
- *In vivo* vs *In vitro*
- ...



Measuring Gene Expression

Reverse transcription (mRNA -> cDNA)

- Most RNA-seq involves large **populations** of cells (10^{6-7})
- Most RNA-seq involves sequencing **cDNA** synthesized using reverse transcription
- Most RNA-seq involves significant **amplification** of cDNA molecules via PCR



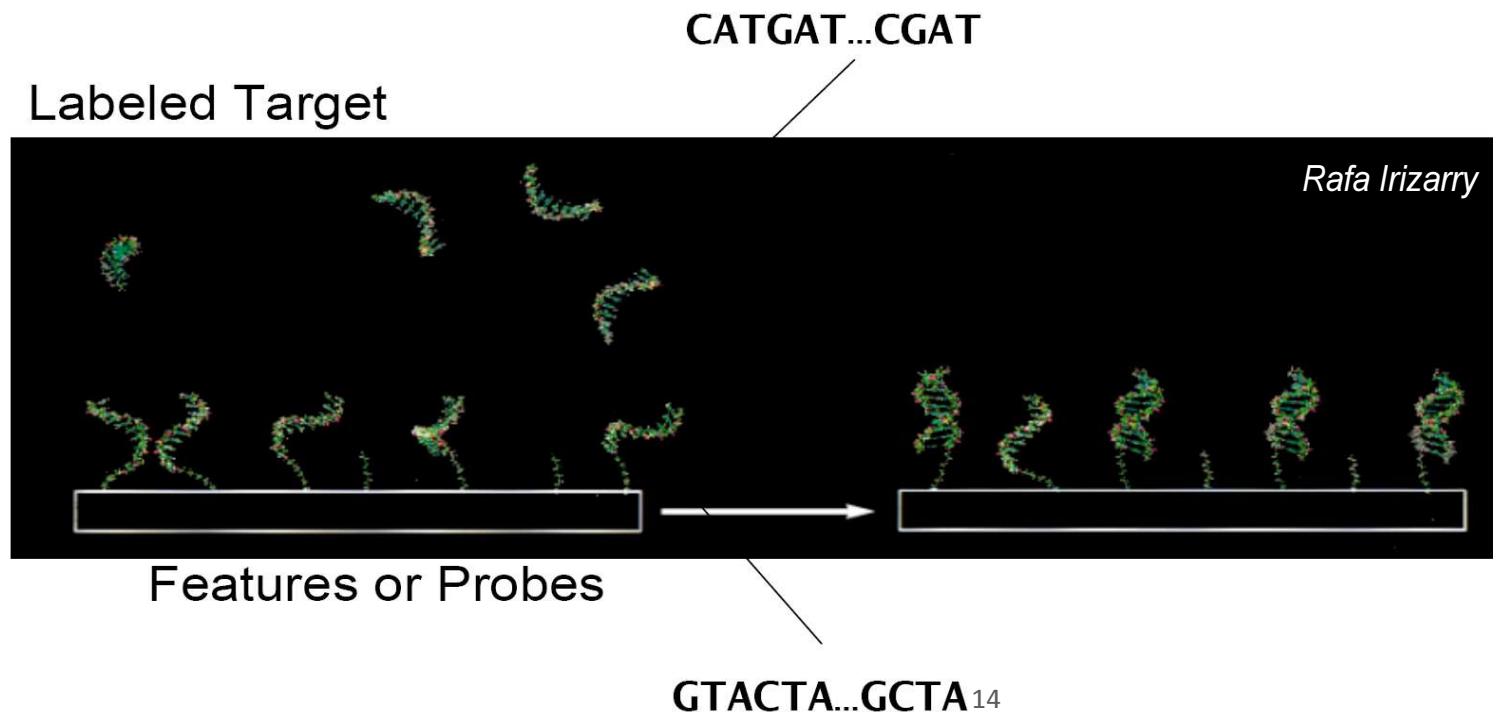
Measuring cDNA: Microarrays

Use hybridization to measure abundance of mRNA transcripts

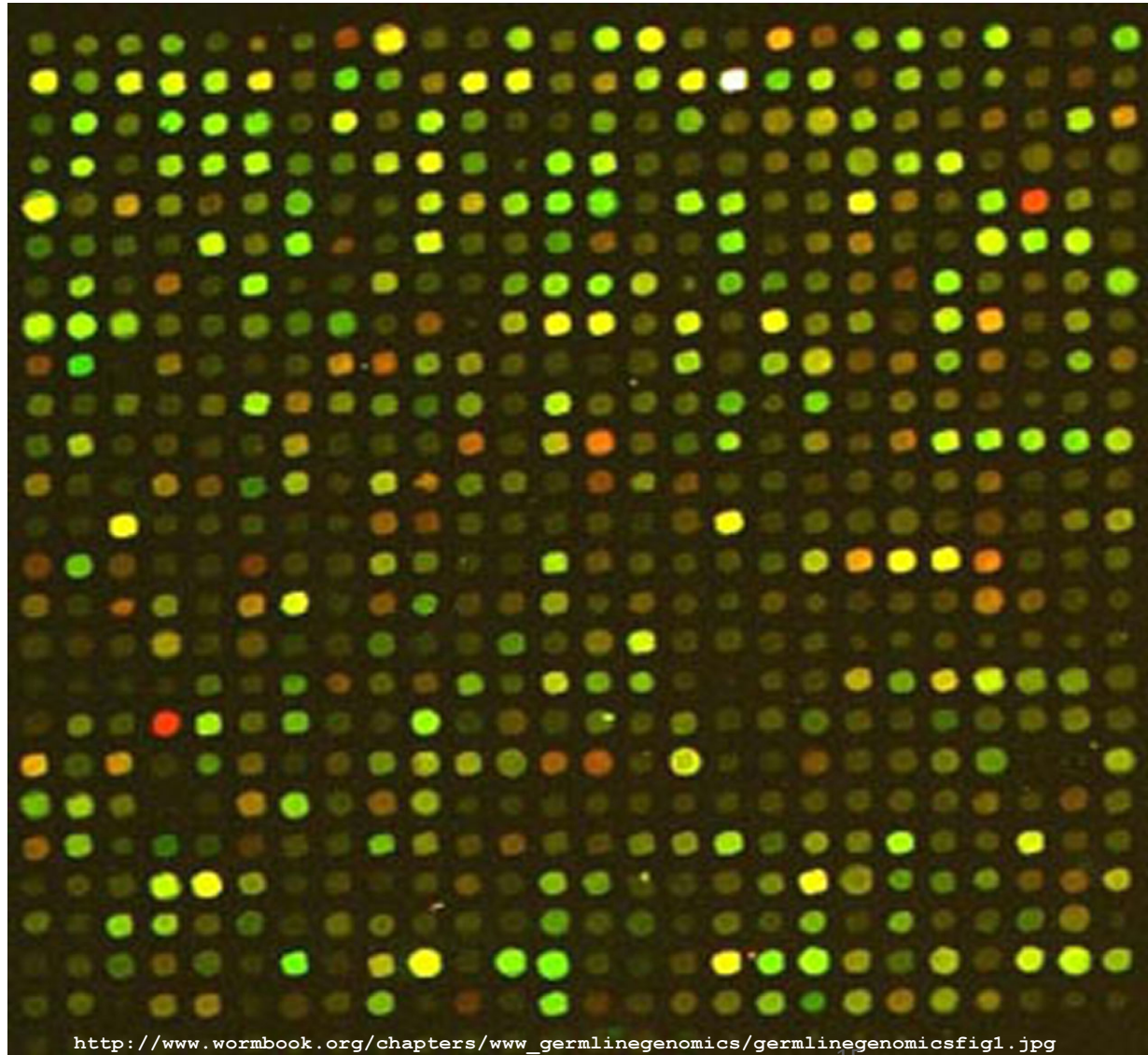
Fix “probes” to a solid support

Hybridize labeled samples of mRNA to probes

Use labels to measure hybridization intensity



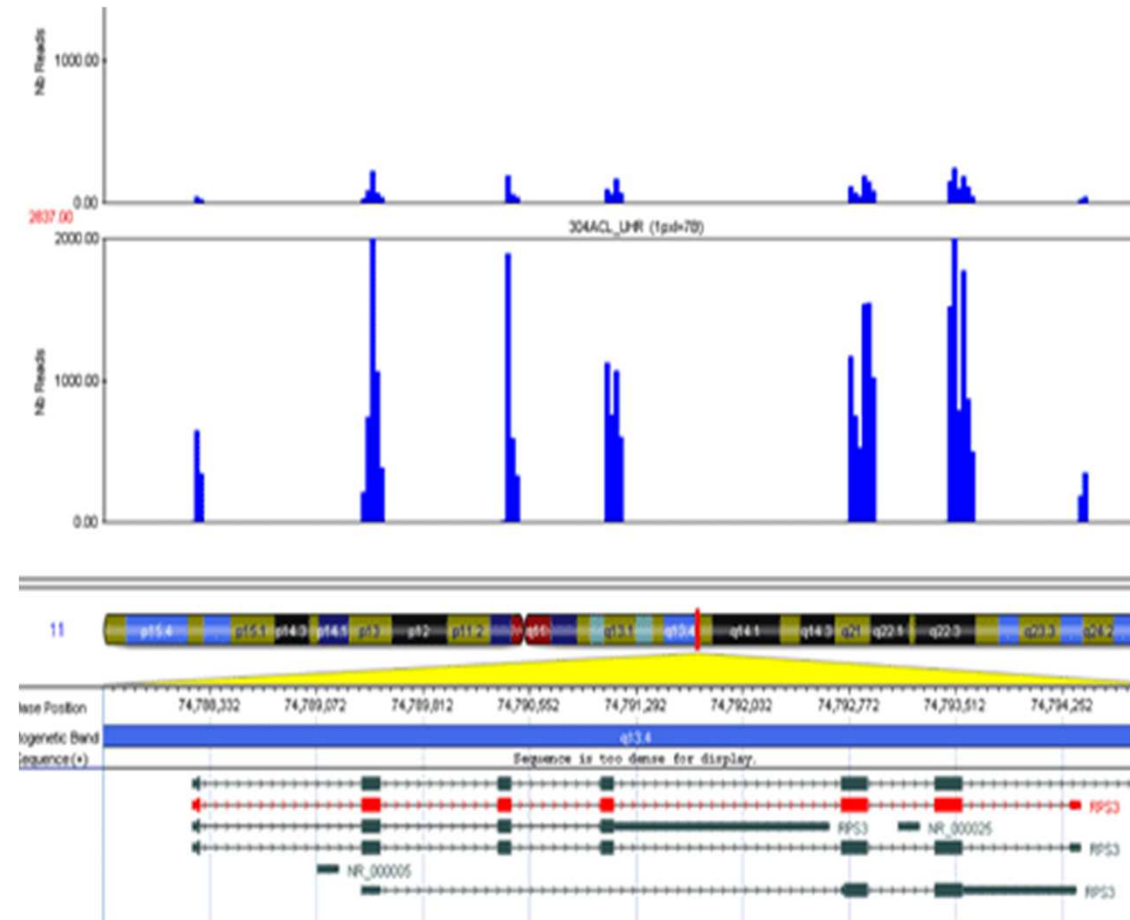
Microarrays: Scanning



Typically less than 1 inch width, spot diameter ! 0.1 mm

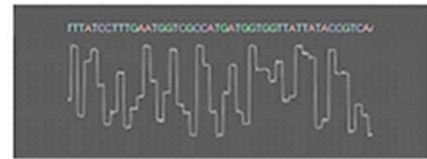
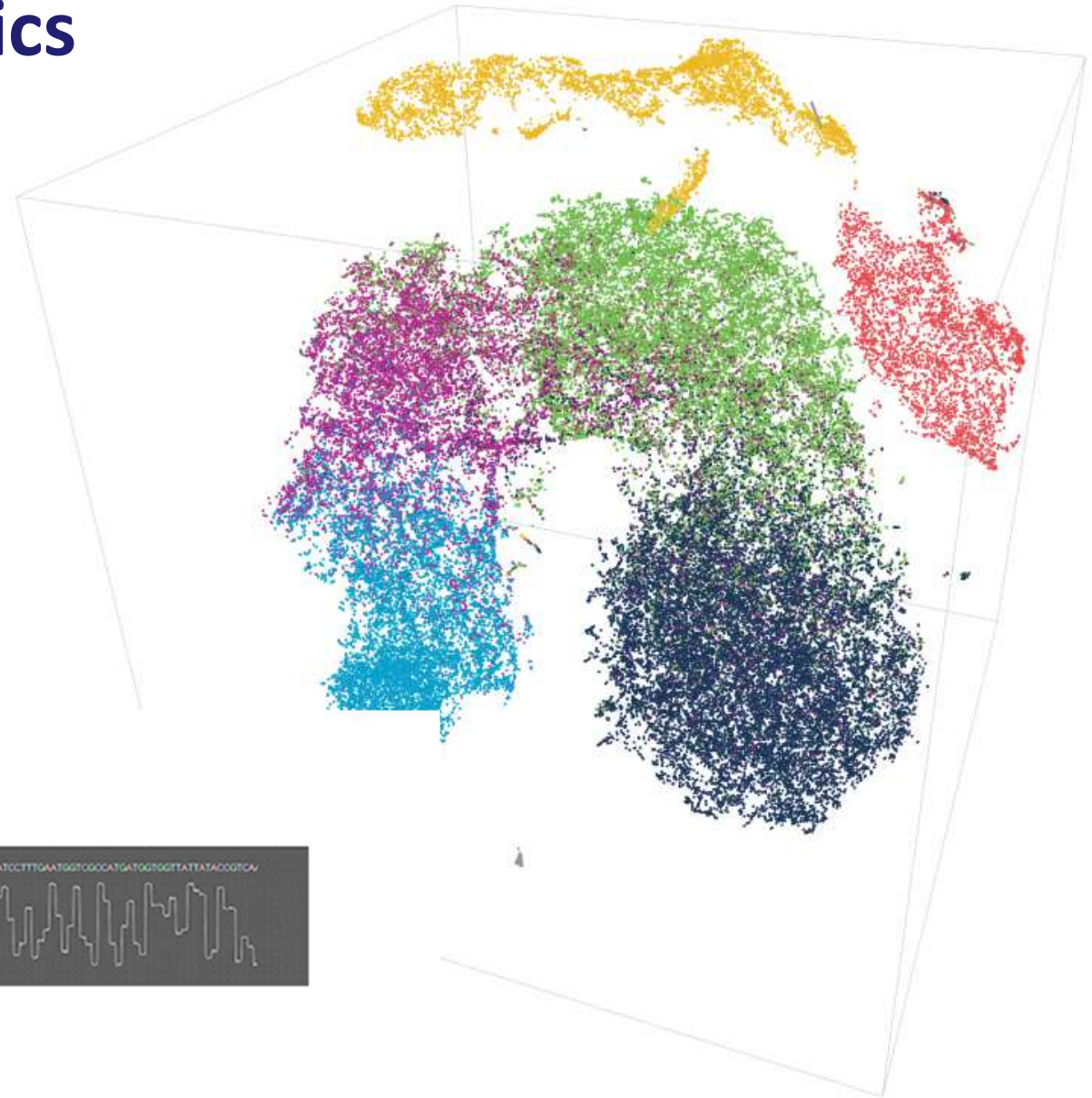
Measuring cDNA: RNA-seq

- High-throughput sequencing allows us to sequence a representative **sample** of the cDNA population “directly”
- Each sequence “read” is **aligned** back to a reference genome/transcriptome to see where it was transcribed from
- We can **count** how many transcripts came from each gene



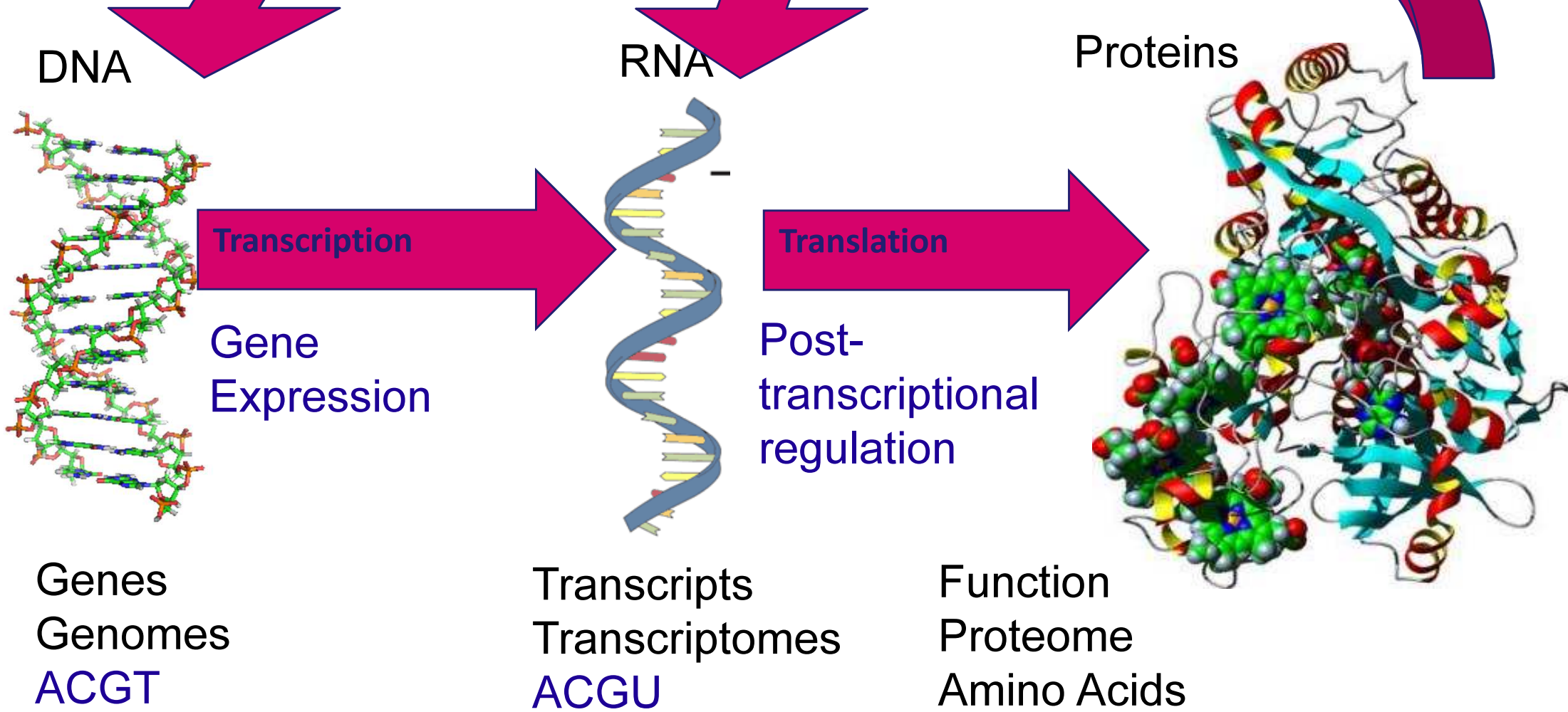
Trends in Transcriptomics

- Single-cell sequencing
- Nanopore Sequencing
 - Full-length transcript sequencing
 - Direct RNA sequencing

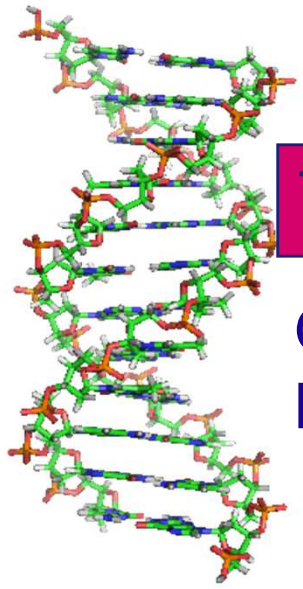


Beyond Gene Expression: Transcriptional Regulation

Transcriptional Regulation



DNA



Transcription

Gene Expression

Genes
Genomes
ACGT

RNA

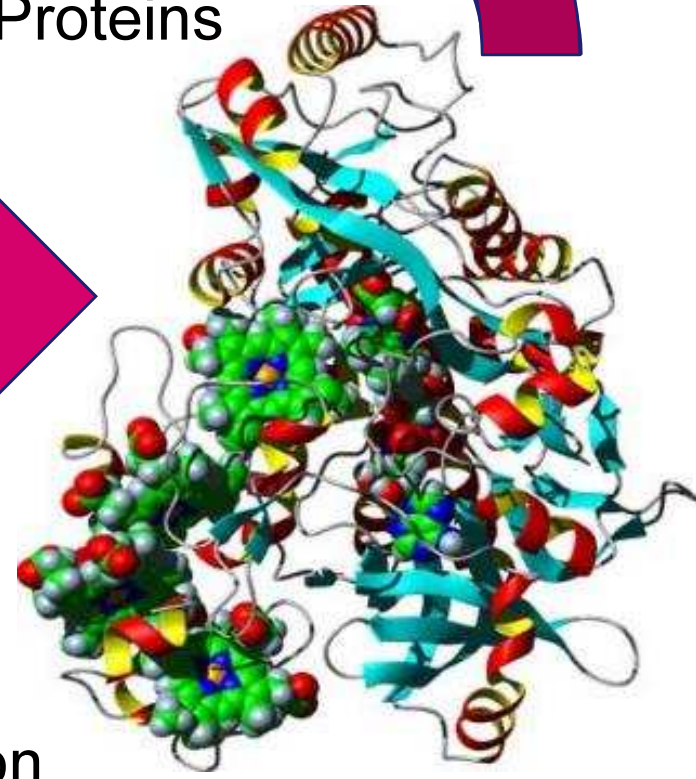


Translation

Post-transcriptional regulation

Transcripts
Transcriptomes
ACGU

Proteins



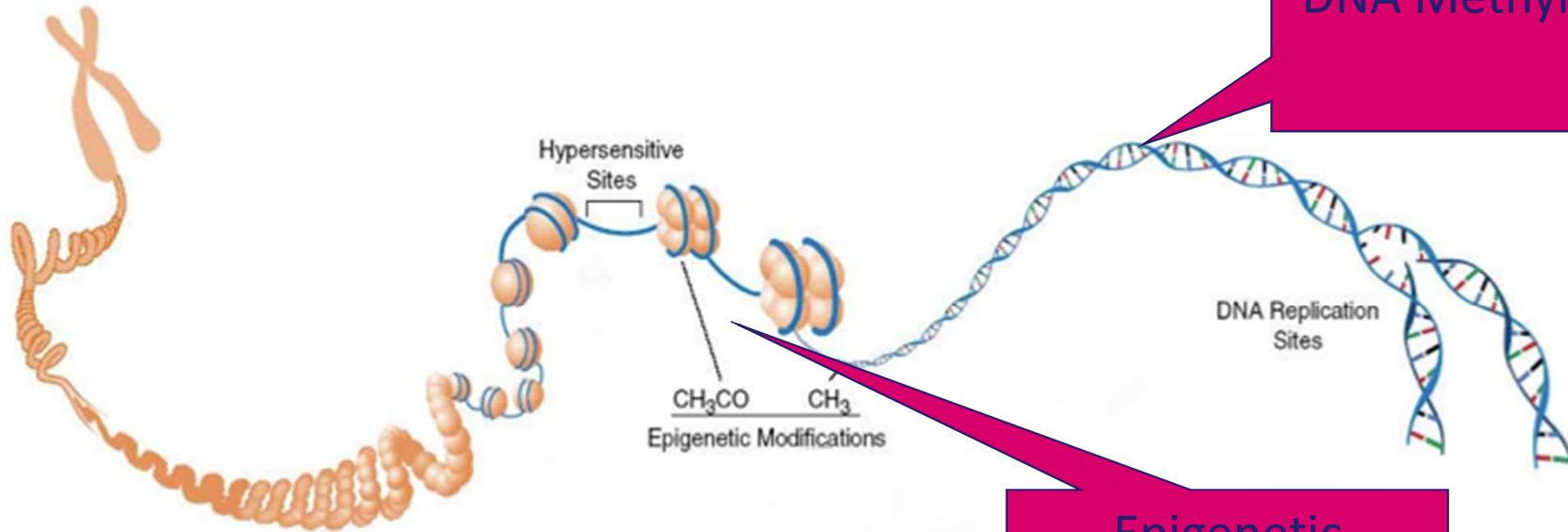
Function
Proteome
Amino Acids
HKDESTNQ CUGPAVILMFYW



CANCER RESEARCH UK

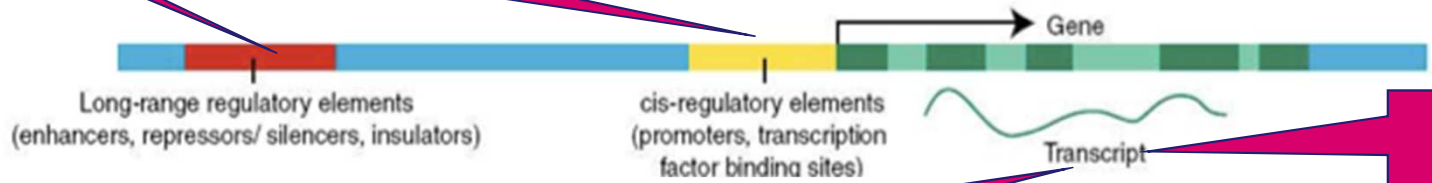
CAMBRIDGE INSTITUTE

Transcriptional regulatory elements



Gene/Protein Interactions

Epigenetic structure (chromatin)



Gene Expression

Small regulatory RNA (eg microRNA)

Regulatory elements of interest include...

TRANSCRIPTION FACTORS

- ChIP

HISTONE MARKS

- ChIP

DNA METHYLATION

- RRBS

- MeDIP

OPEN CHROMATIN

- DNase Hypersensitivity

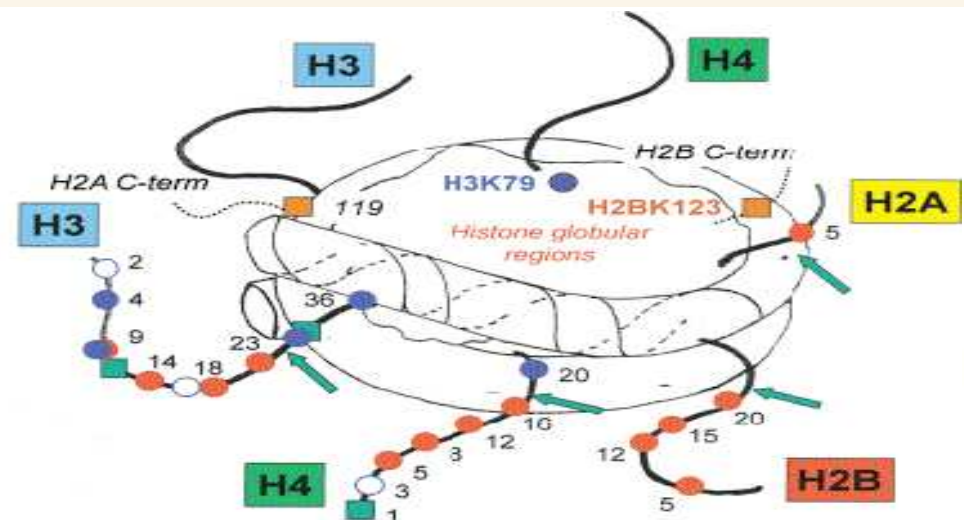
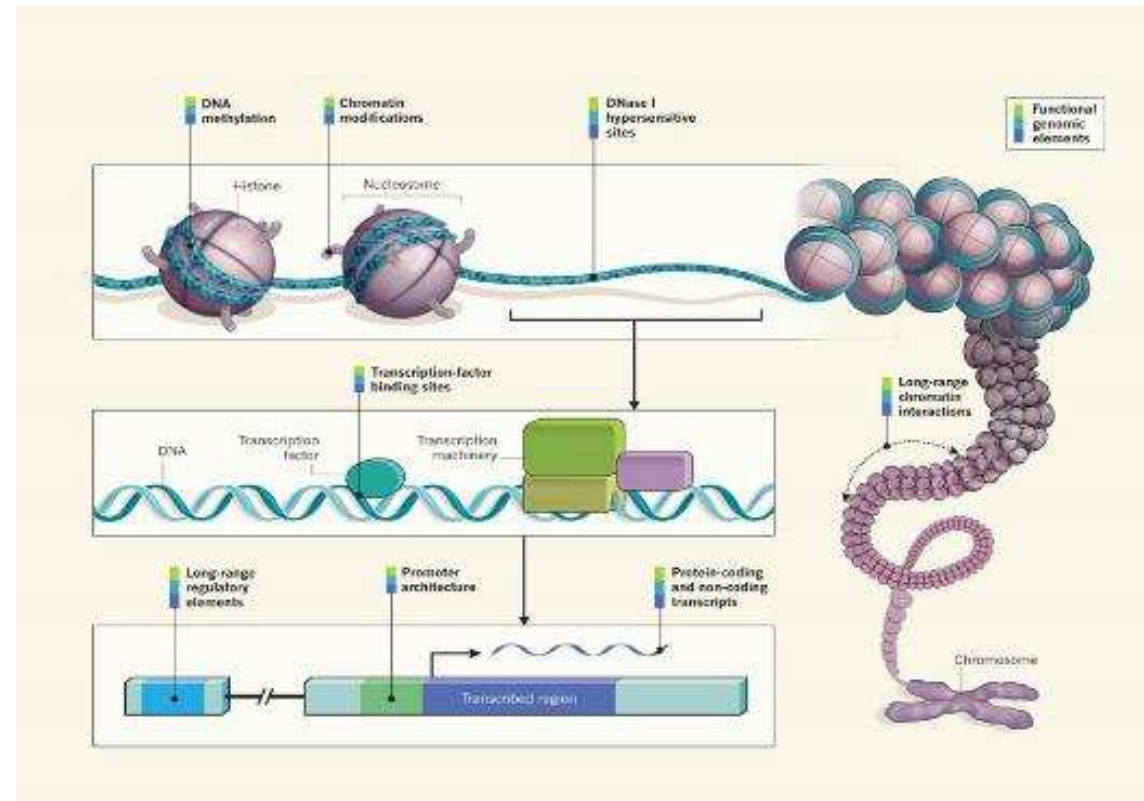
- ATAC

CHROMATIN STRUCTURE

- HiC

RNA POLYMERASE

- Pol II ChIP



Cell differentiation

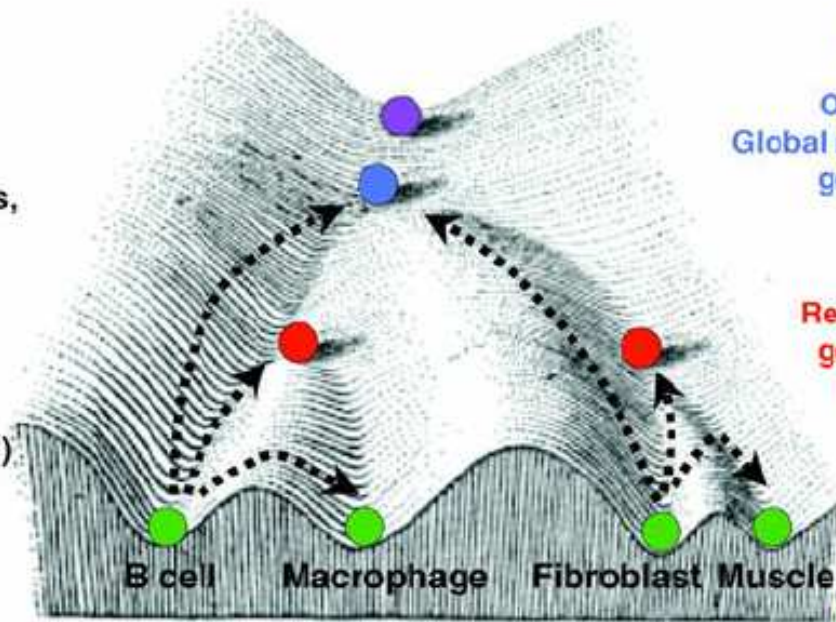
Developmental potential

Totipotent
Zygote

Pluripotent
ICM/ES cells, EG cells,
EC cells, mGS cells
iPS cells

Multipotent
Adult stem cells
(partially
reprogrammed cells?)

Unipotent
Differentiated cell
types



Epigenetic status

Global DNA demethylation

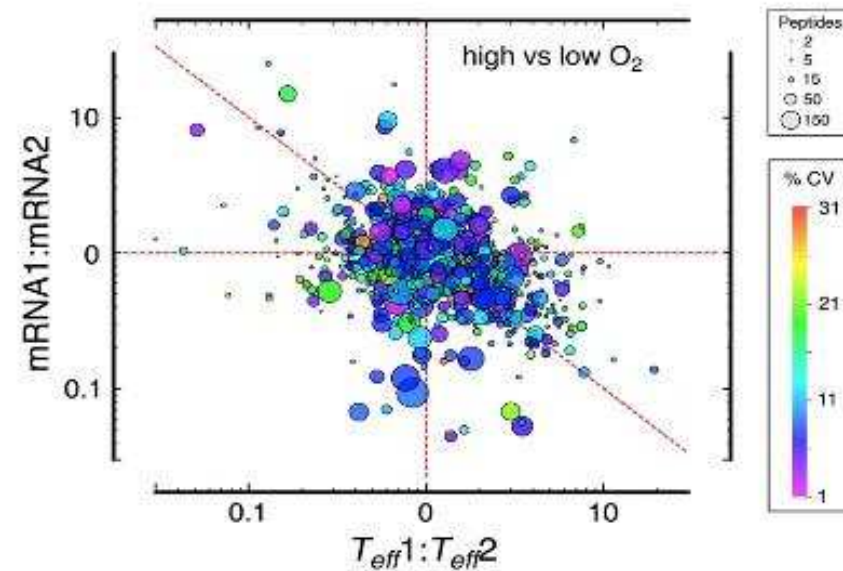
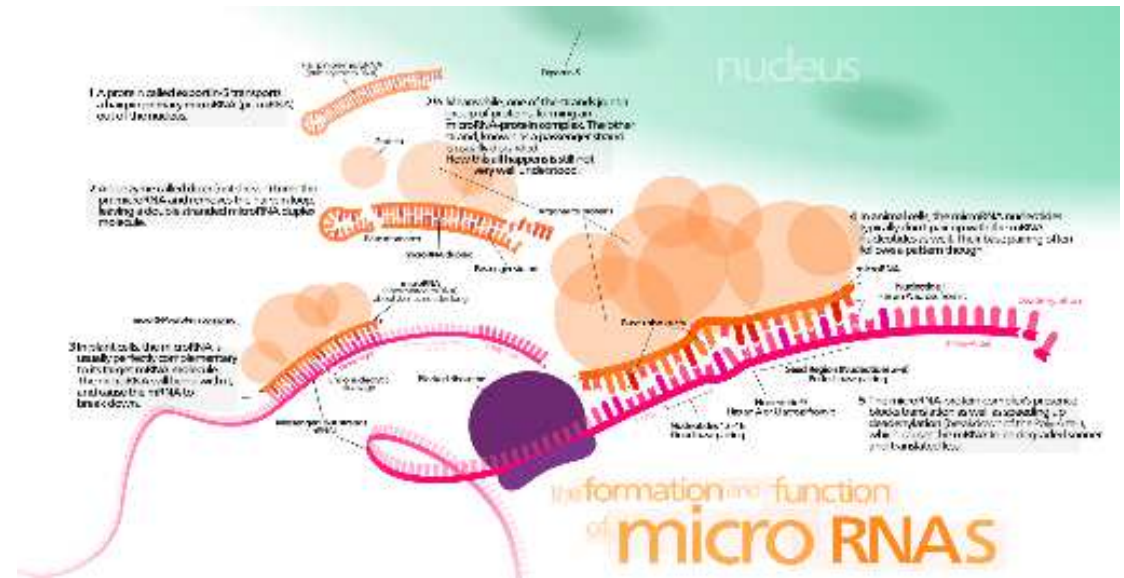
Only active X chromosomes;
Global repression of differentiation
genes by Polycomb proteins;
Promoter hypomethylation

X inactivation;
Repression of lineage-specific
genes by Polycomb proteins;
Promoter hypermethylation

X inactivation;
Derepression of
Polycomb silenced
lineage genes;
Promoter hypermethylation

And Beyond...

- Post-transcriptional Regulation
- Translational Efficiency
- DATA INTEGRATION



Acknowledgements

Former lecturers who
contributed to these
lecture notes:

Benilton S Carvalho

Nuno Barbosa-Morais

Stephen J Eglan

Natalie Thorne



UNIVERSITY OF
CAMBRIDGE



CANCER
RESEARCH
UK

CAMBRIDGE
INSTITUTE