

# Introduction to Functional Genomics

RORY STARK  
4 OCTOBER 2015

# Agenda

- Where does Functional Genomics fit in the MPhil?
- What is Functional Genomics?
- RNA Transcription/Gene Expression
- Measuring Gene Expression
  - Microarrays
  - High-throughput Sequencing
- Beyond Gene Expression: Transcriptional Regulation
  - Transcription factors
  - Epigenetics
  - Post-transcriptional regulation
- Course Outline



UNIVERSITY OF  
CAMBRIDGE

2



CANCER  
RESEARCH  
UK

CAMBRIDGE  
INSTITUTE

# Where does Functional Genomics fit?



UNIVERSITY OF  
CAMBRIDGE

3



CANCER  
RESEARCH  
UK

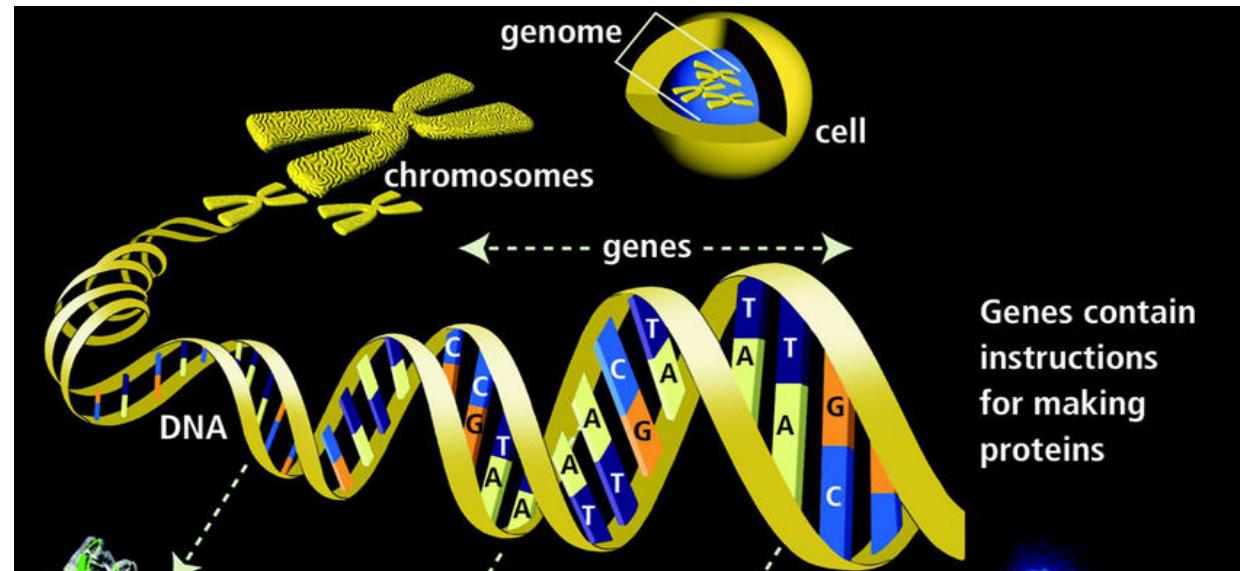
CAMBRIDGE  
INSTITUTE

# Computational Biology MPhil classes

## GENOMICS

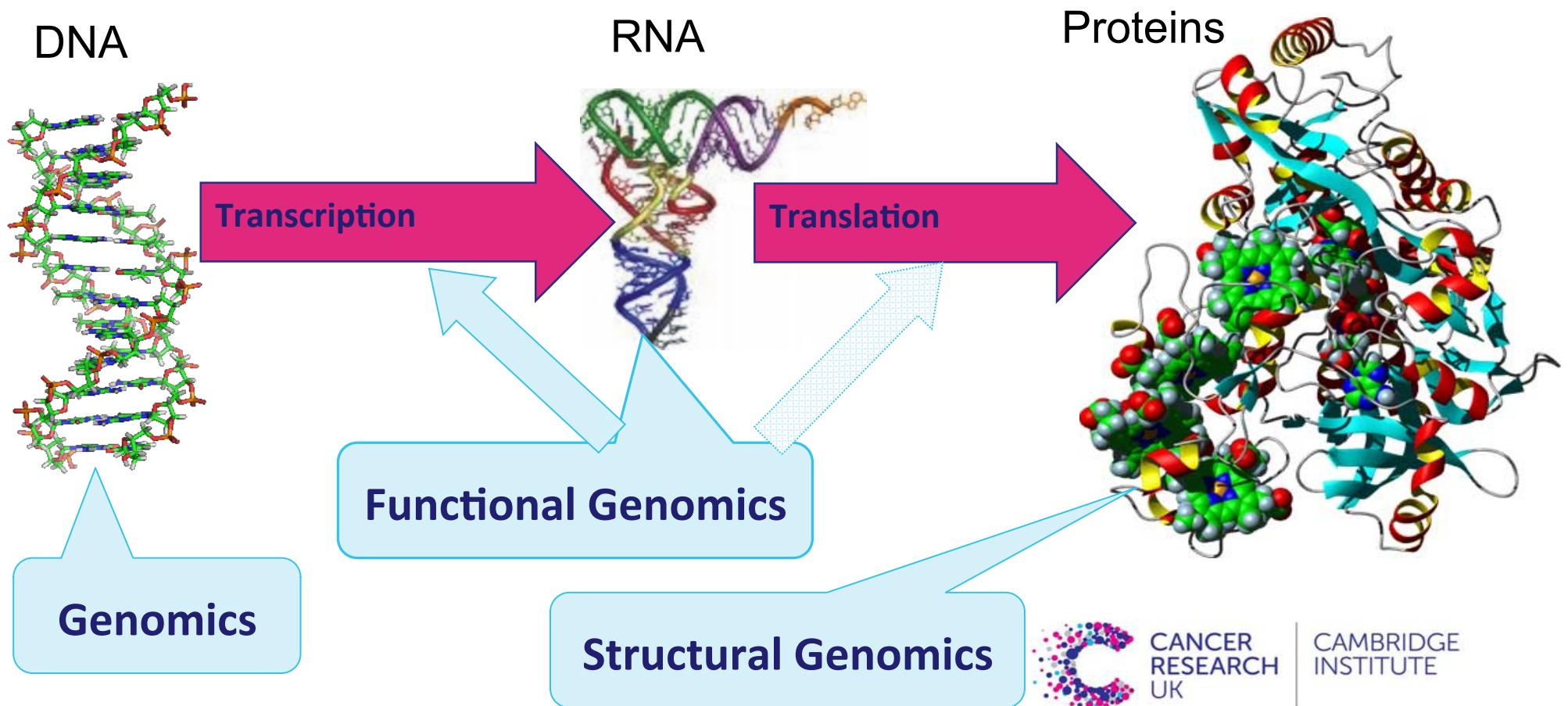
- Sequence-based
  - Genome Informatics
  - Genome Sequence Analysis
  - Population Genetic Analysis of Genomic Data
- Structure-based
  - Structural Biology
  - Biological Imaging and Analysis
- Function-based
  - Functional Genomics

# The Genome



- Each cell contains a complete copy of the genome, distributed along chromosomes (compressed and entwined DNA)
- $3 \times 10^9$  (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...)

# The Central Dogma of Molecular Biology



# Functional Genomics: Sequence vs. Function



UNIVERSITY OF  
CAMBRIDGE

7



CANCER  
RESEARCH  
UK

CAMBRIDGE  
INSTITUTE

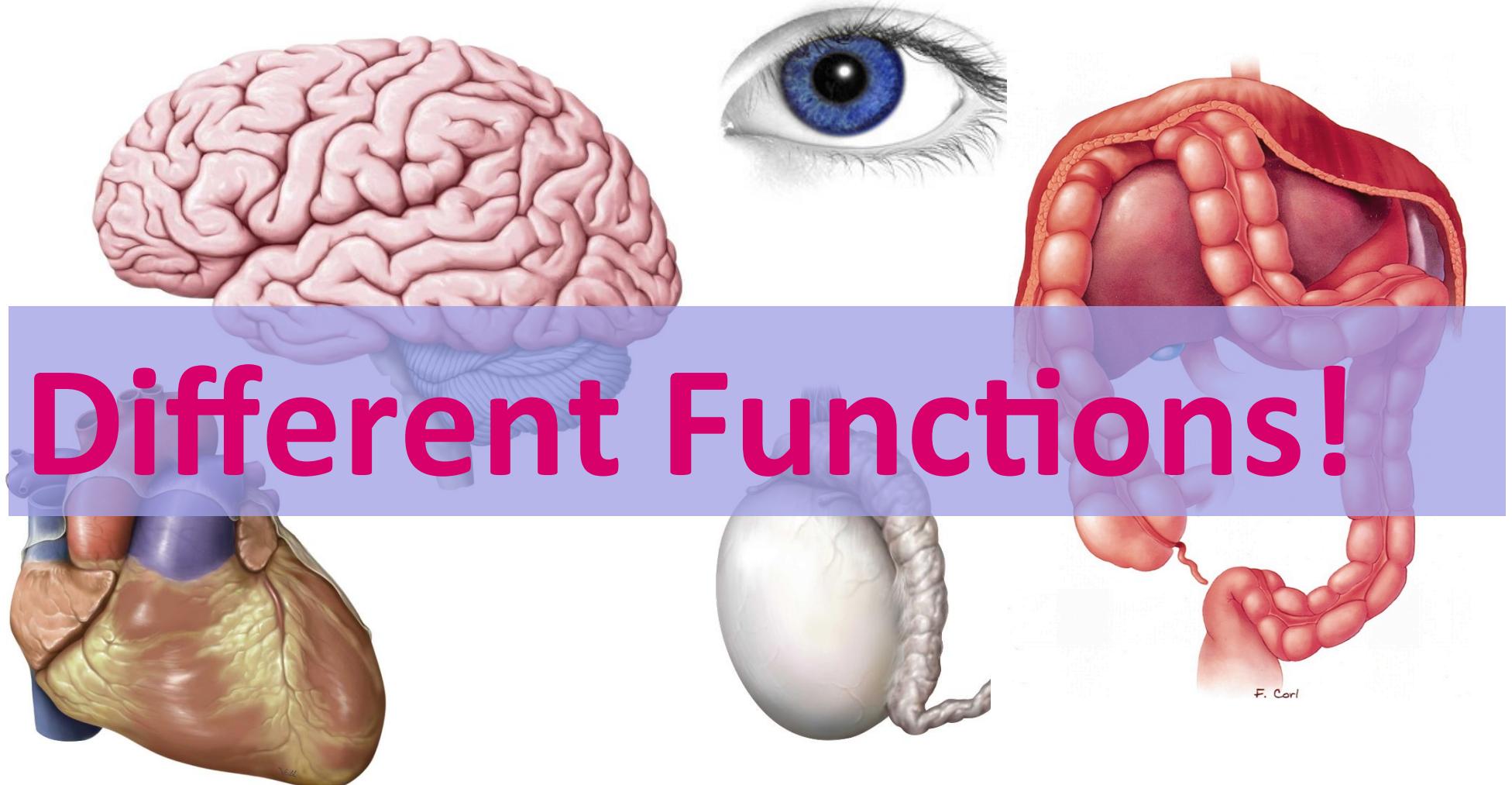
# What accounts for the difference in phenotype?



## Different Genomes!



# What accounts for the difference in phenotype?



Different Functions!



UNIVERSITY OF  
CAMBRIDGE

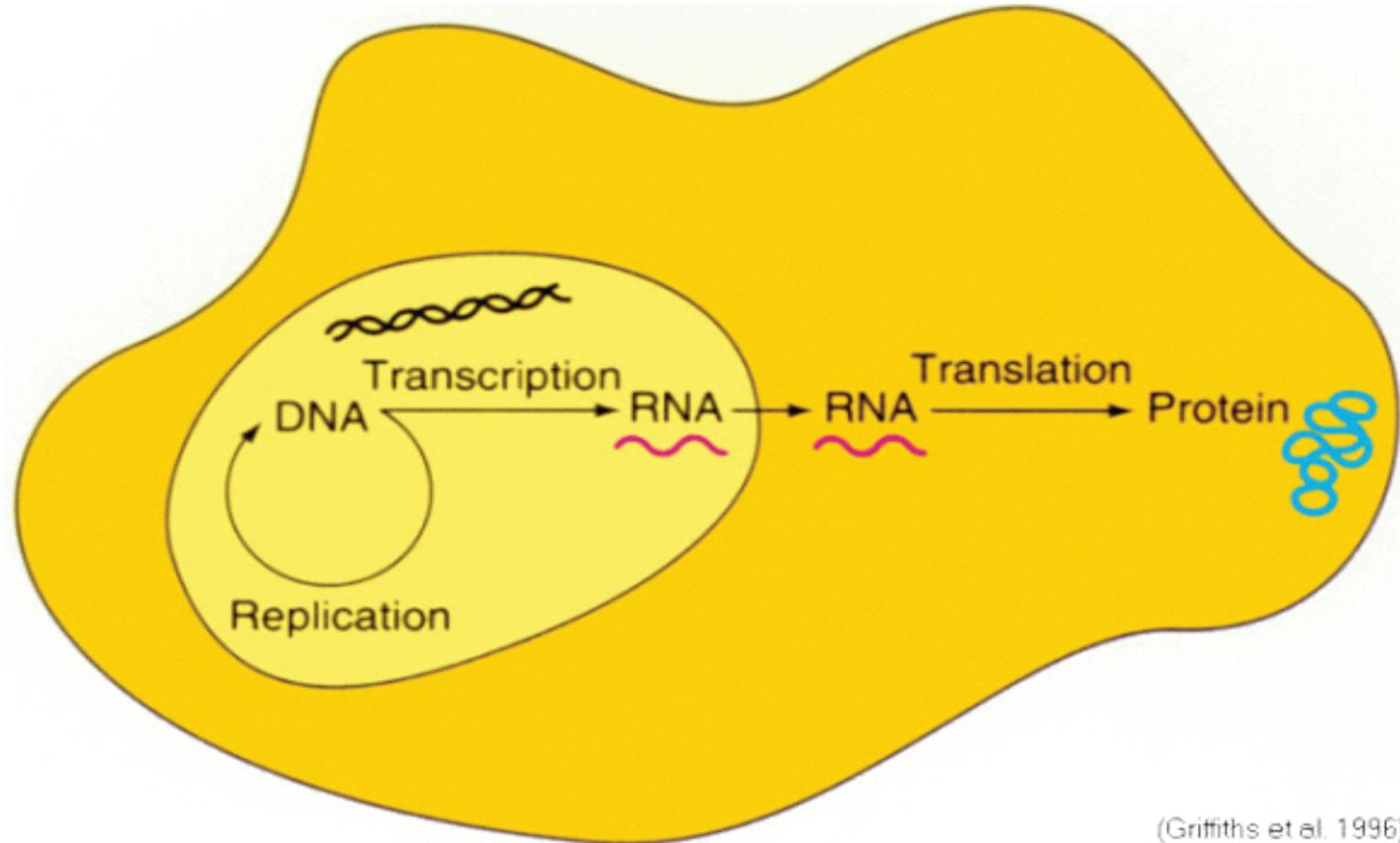
9



CANCER  
RESEARCH  
UK

CAMBRIDGE  
INSTITUTE

# Gene expression

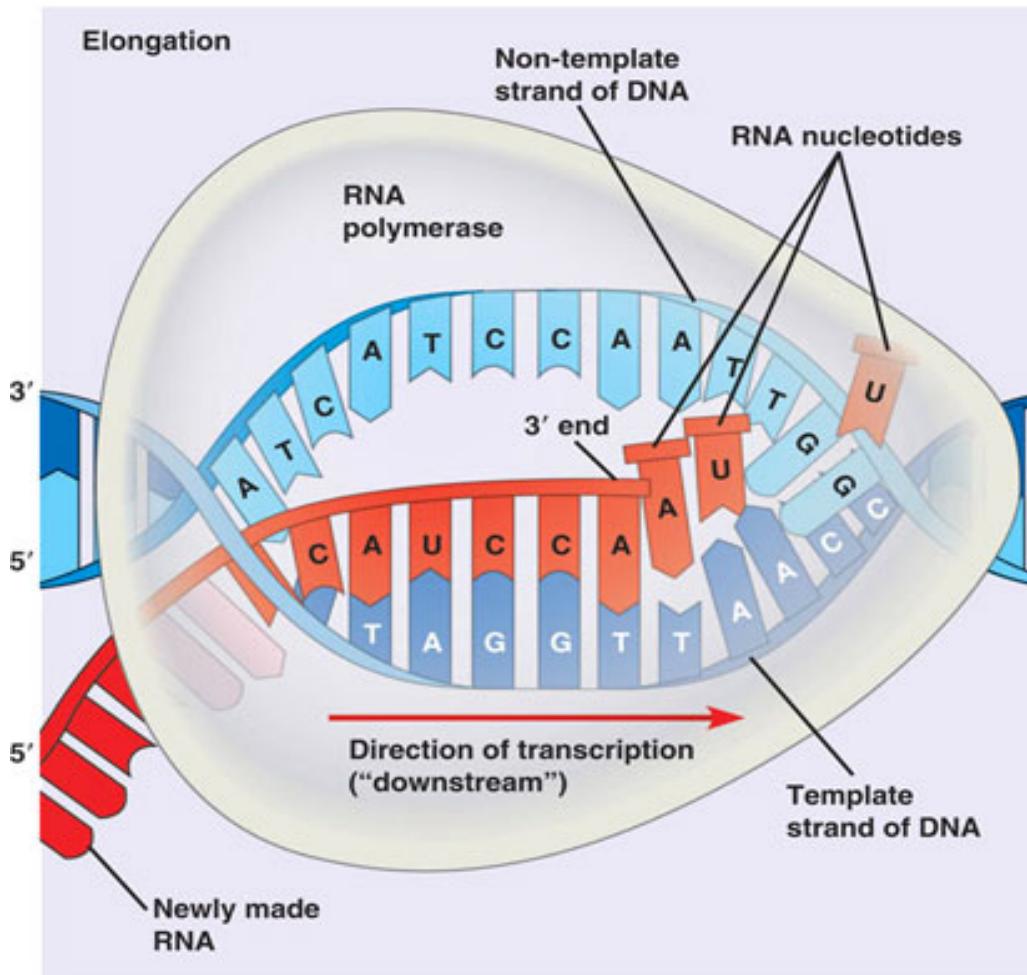


(Griffiths et al. 1996)

A gene is a segment of chromosomal DNA that directs the synthesis of a protein

An intermediate step is the gene being transcribed or expressed

# Gene transcription (DNA → RNA)



<http://fig.cox.miami.edu/~cmallery/150/gene/c7.17.7b.transcription.jpg>

11

CAMBRIDGE  
INSTITUTE

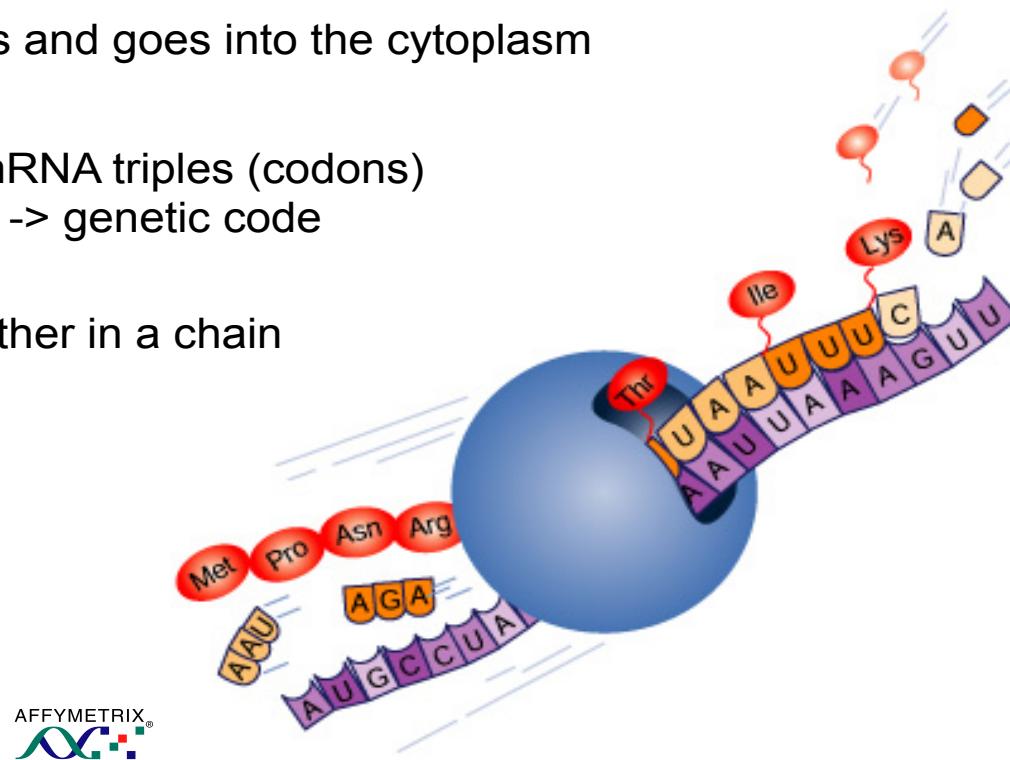
Transcript may undergo *post-transcriptional modification* (e.g. miRNA)

# Gene translation (RNA -> protein)

mRNA leaves nucleus and goes into the cytoplasm

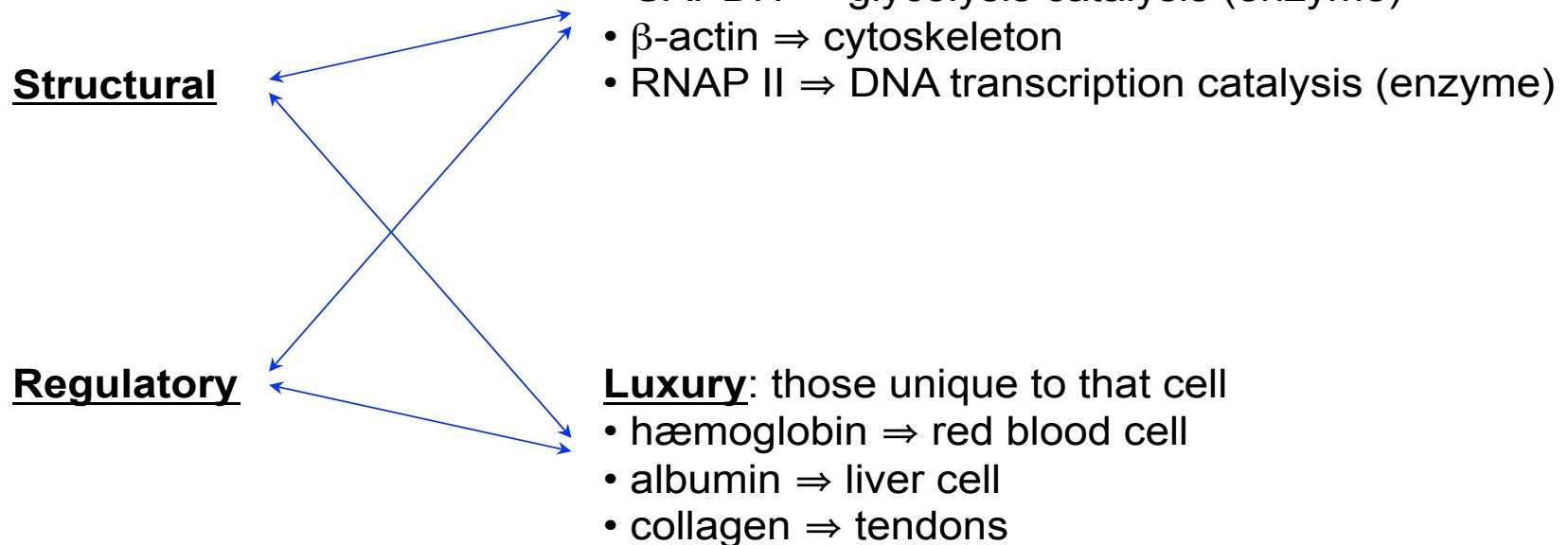
Ribosomes convert mRNA triples (codons) into 1/20 amino acids -> genetic code

Amino acids link together in a chain to make a protein



Protein undergoes “post-translational modification”.

# Proteins: the building blocks of cells



# 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)



UNIVERSITY OF  
CAMBRIDGE

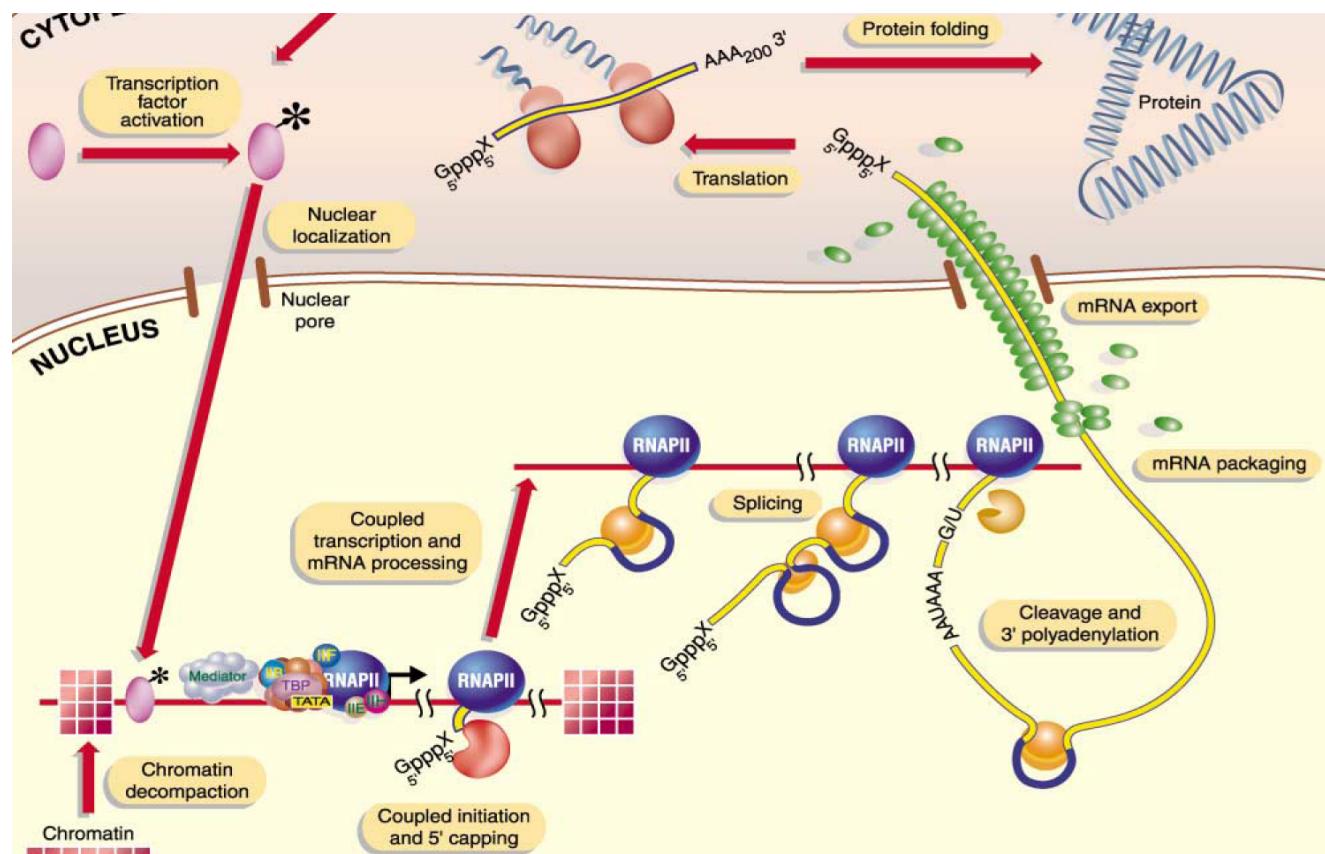
15



CANCER  
RESEARCH  
UK

CAMBRIDGE  
INSTITUTE

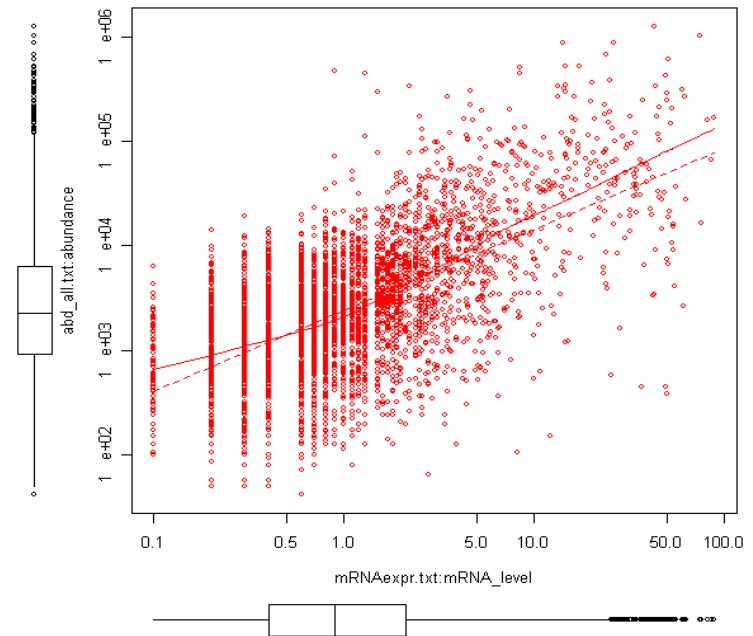
# Cell-specific function



# 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



UNIVERSITY OF  
CAMBRIDGE

19



CANCER  
RESEARCH  
UK

CAMBRIDGE  
INSTITUTE

# Reverse transcription (mRNA -> cDNA)

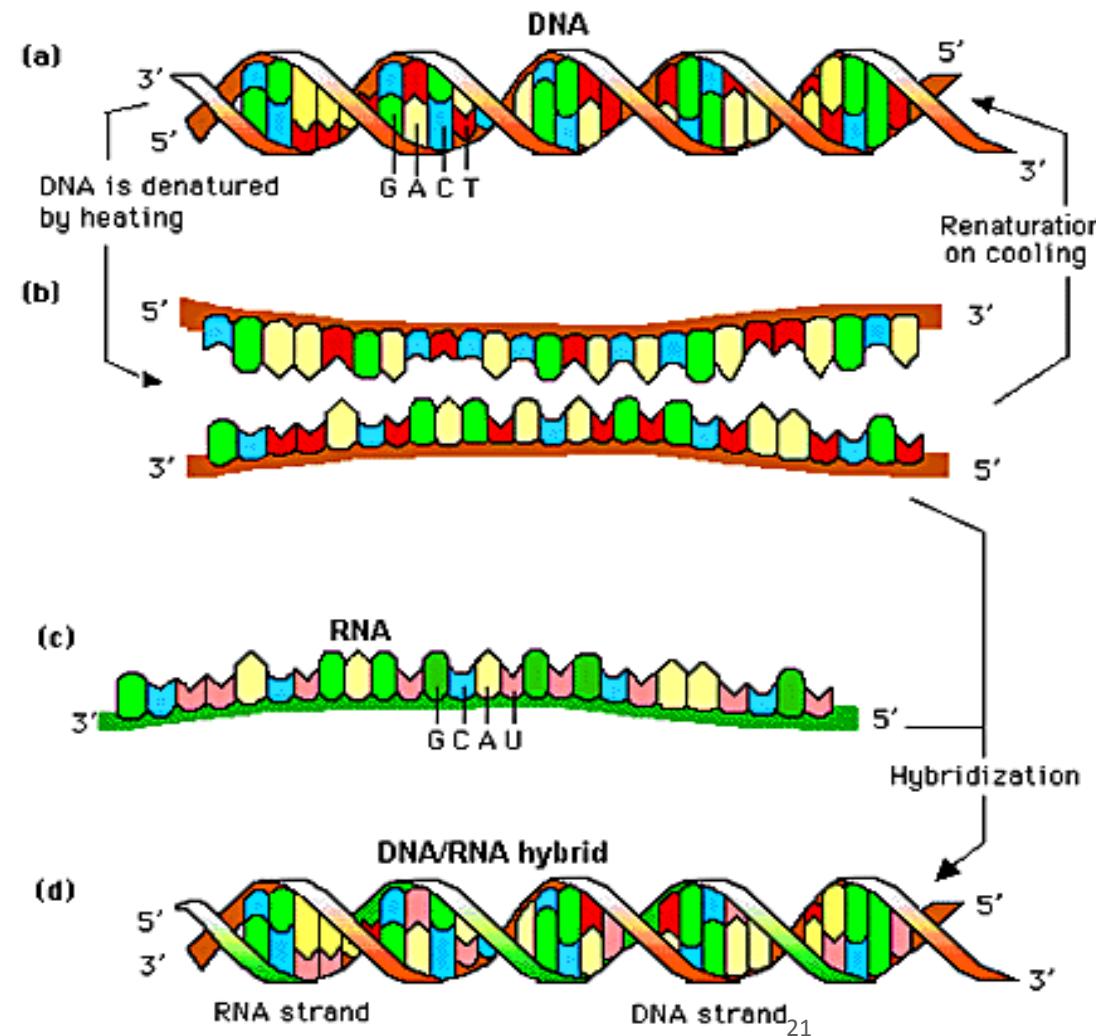


Clone cDNA strands,  
complementary to the mRNA

Rafa Irizarry



# Measuring RNA: cDNA hybridization



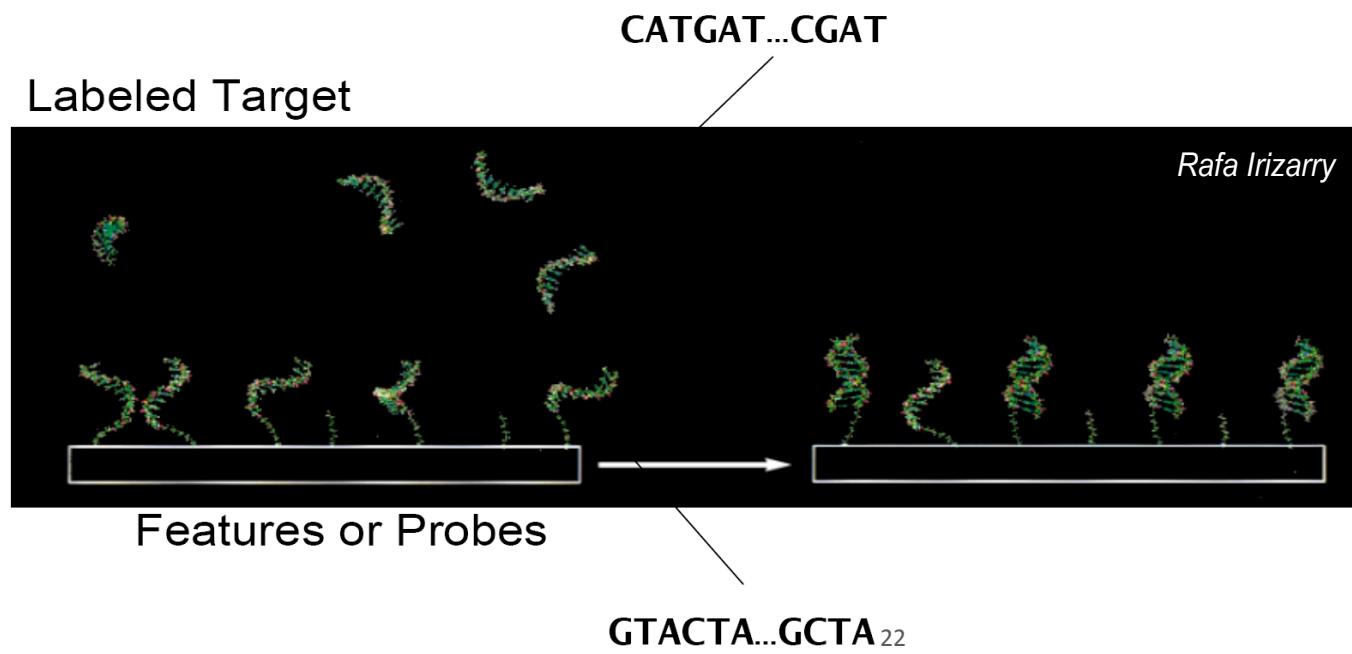
# Measuring cDNA: Microarrays

Use hybridization to measure abundance of mRNA transcripts

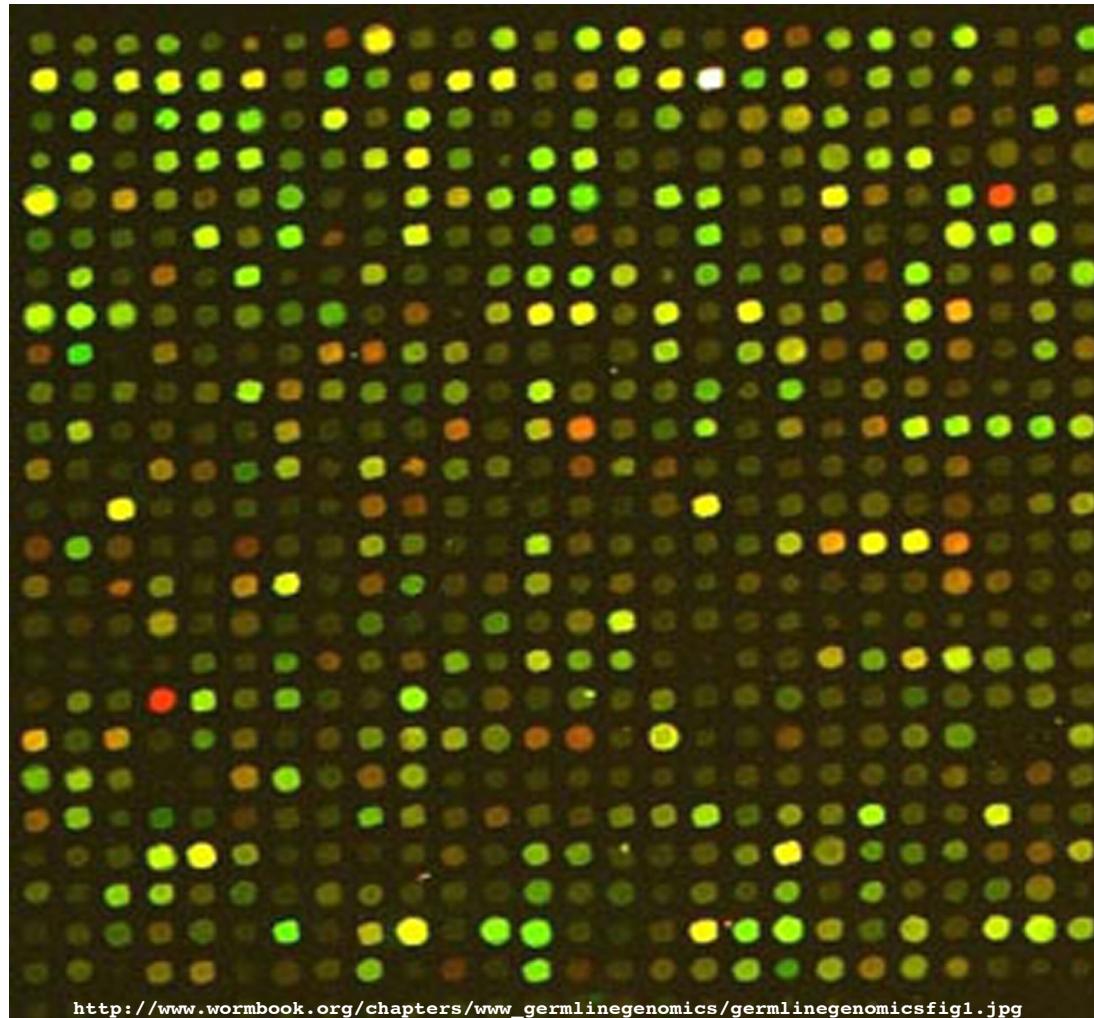
Fix “probes” to a solid support

Hybridize labeled samples of mRNA to probes

Use labels to measure hybridization intensity



# Microarrays: Scanning



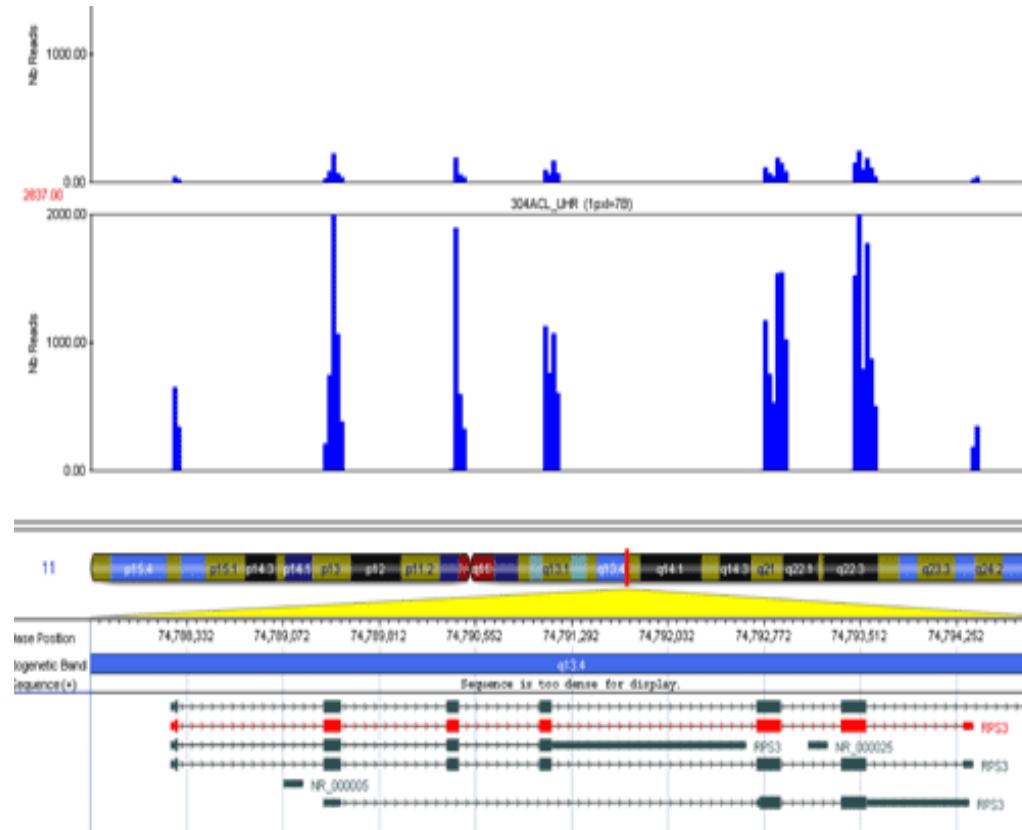
[http://www.wormbook.org/chapters/www\\_germlinegenomics/germlinegenomicsfig1.jpg](http://www.wormbook.org/chapters/www_germlinegenomics/germlinegenomicsfig1.jpg)

23

Typically less than 1 inch width, spot diameter  $\approx$  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 “mapped” back to a reference genome/transcriptome to see where it was transcribed from
- We can count how many transcripts came from each gene



# Beyond Gene Expression: Transcriptional Regulation



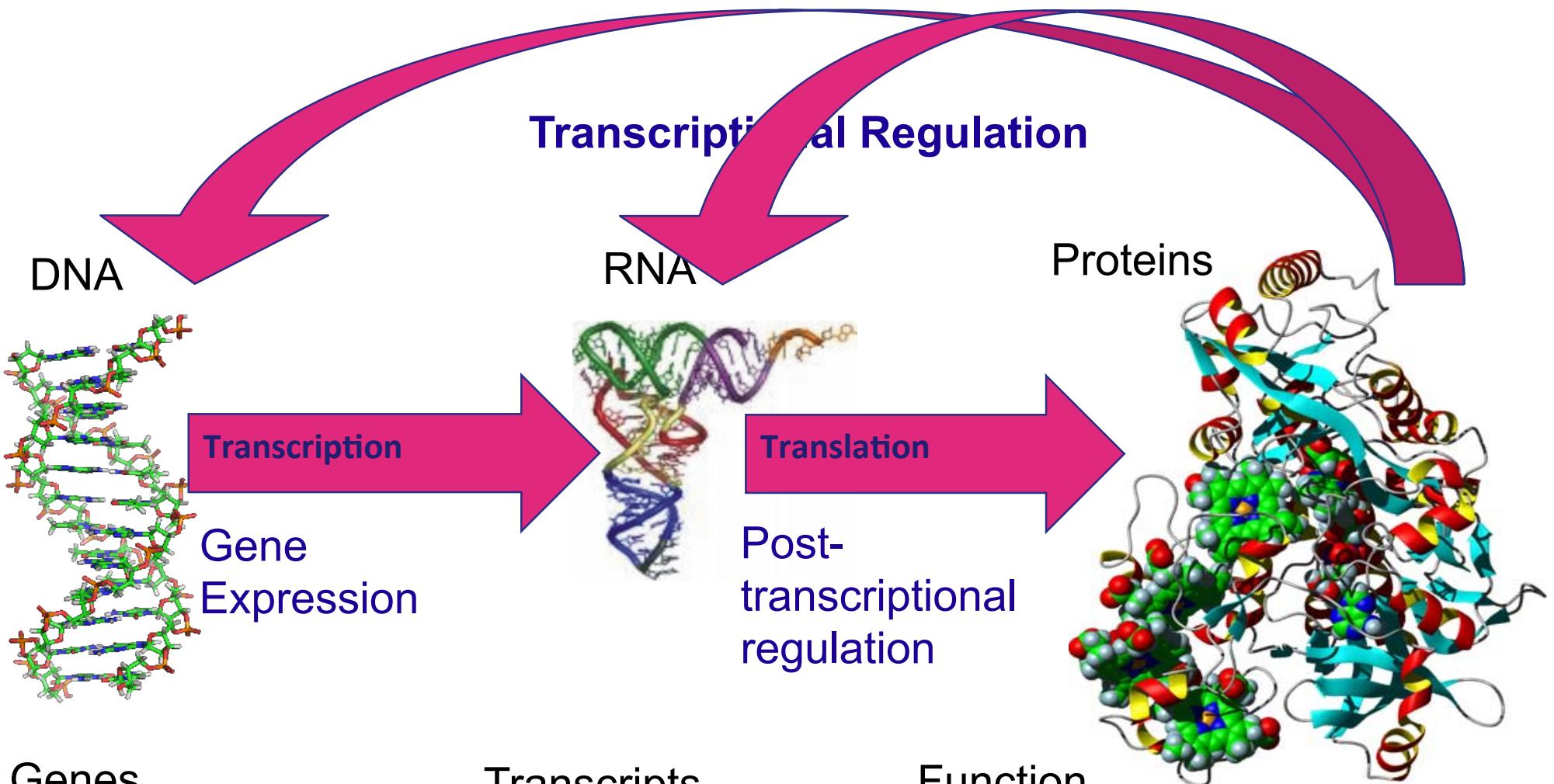
UNIVERSITY OF  
CAMBRIDGE

25



CANCER  
RESEARCH  
UK

CAMBRIDGE  
INSTITUTE



Genes  
Genomes  
ACGT

Transcripts  
Transcriptomes  
ACGU

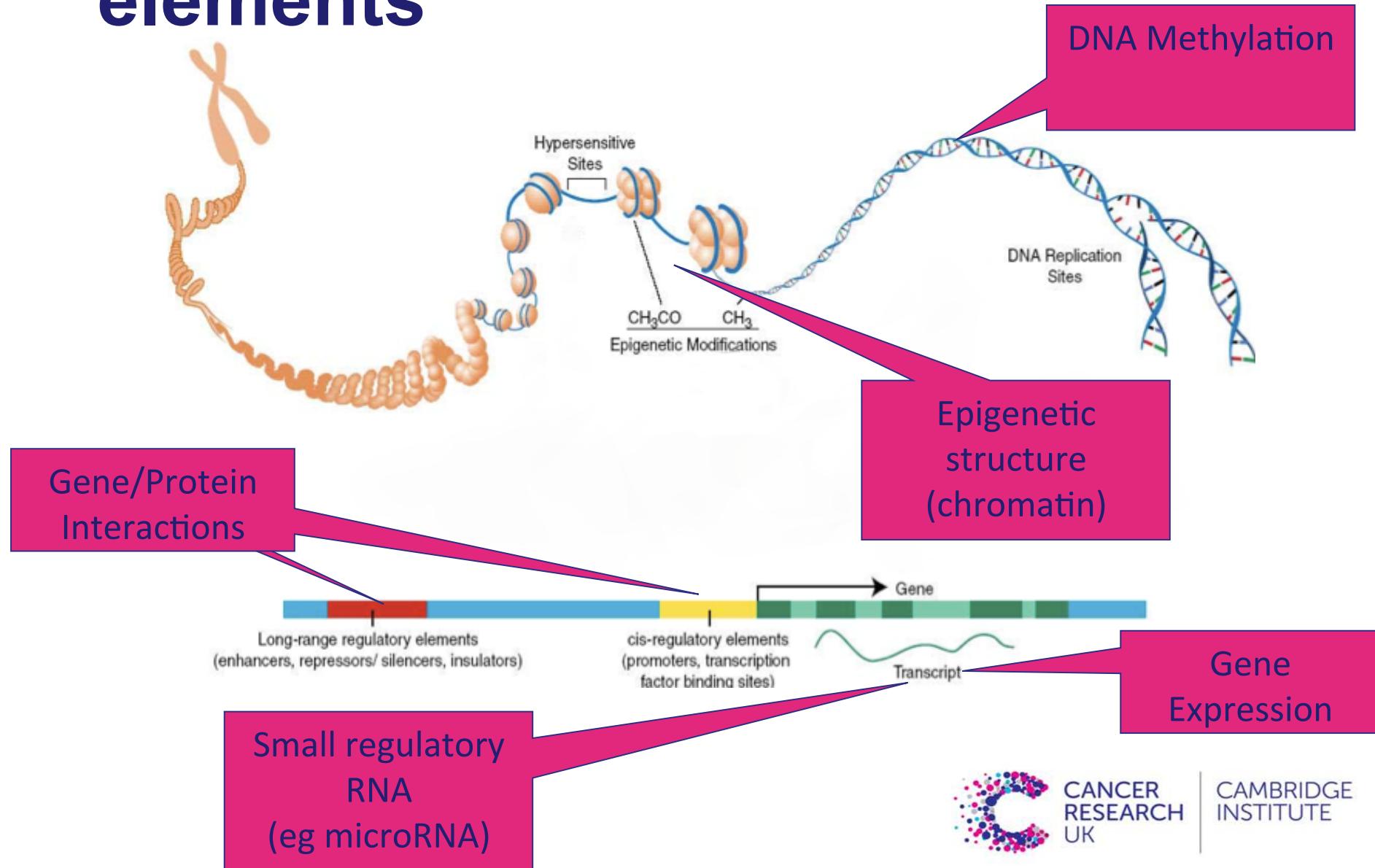
Function  
Proteome  
Amino Acids  
HKDESTNQCUGPAVILMFYW



CANCER  
RESEARCH  
UK

CAMBRIDGE  
INSTITUTE

# Transcriptional regulatory elements



# Regulatory elements of interest include...

## TRANSCRIPTION FACTORS

- ChIP

## HISTONE MARKS

- ChIP

## DNA METHYLATION

- MeDIP etc.

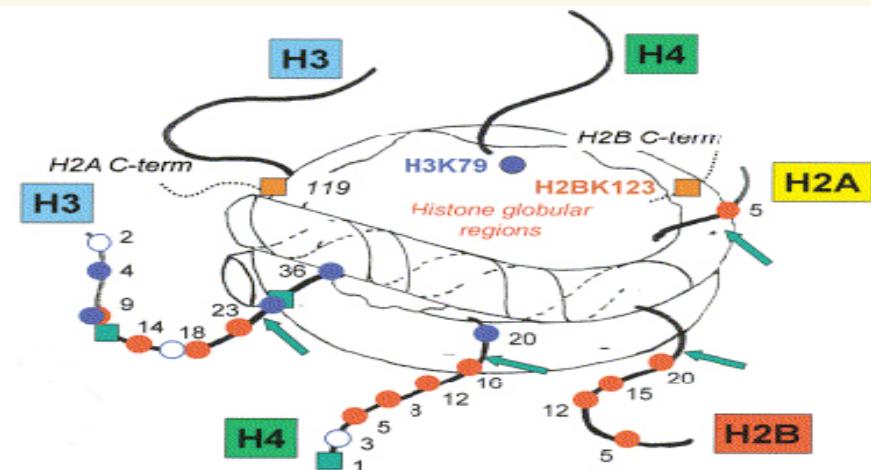
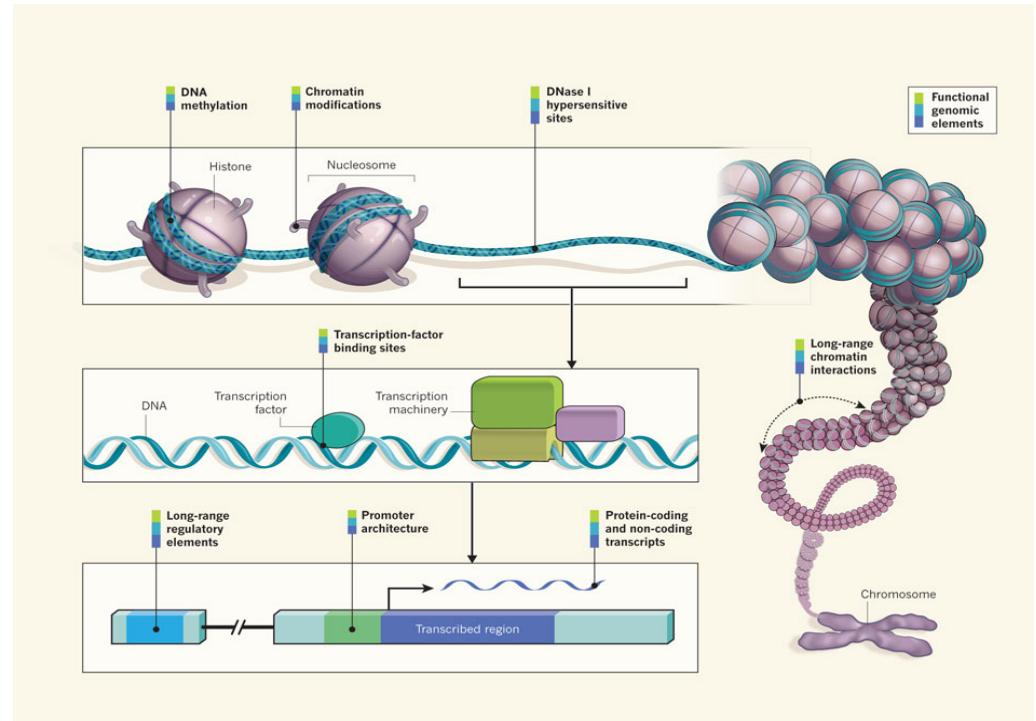
## NUCLEOSOMES

## RNA POLYMERASE

- Pol II ChIP

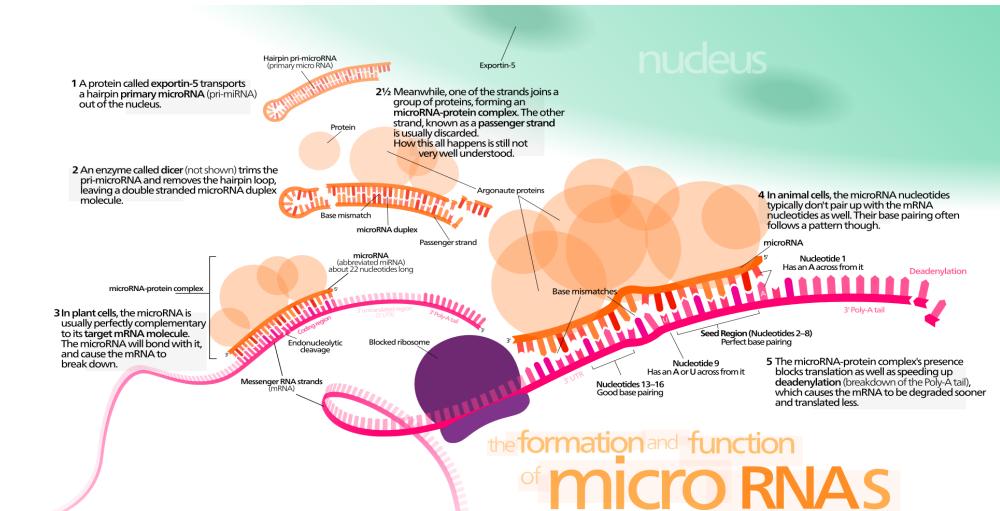
## OPEN CHROMATIN

- DNase Hypersensitivity

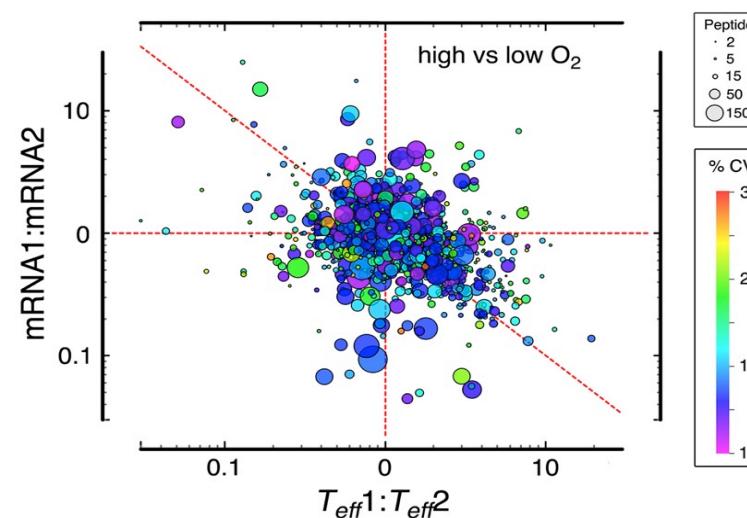


# And Beyond...

## – Post-transcriptional Regulation



## – Translational Efficiency



29

# Functional Genomics: Course Outline



UNIVERSITY OF  
CAMBRIDGE

30



CANCER  
RESEARCH  
UK

CAMBRIDGE  
INSTITUTE

# Class Schedule

- 1: Introduction
  - Rory Stark
- 2-4: Differential gene expression and Microarrays
  - Oscar Rueda/ Maurizio Callari
- 5-10: High throughput sequencing
  - Next-gen sequencing: Rory Stark
  - RNA-Seq, three lectures: Ernest Turro
  - ChIP-seq, two lectures: Rory Stark
- 11-14: Integrative and Downstream Analysis
  - Shamith Samarajiwa
- 14: Copy number analysis
  - Geoff Macintyre
- 15-16: Clustering, classification, and survival analysis
- 17: Open Q/A and Feedback Session

# Class structure

## LECTURERS:

- Rory Stark
- Oscar Rueda
- Maurizio Callari

## LECTURERS:

- Ernest Turro
- Shamith Samarajiwa
- Geoff Macintyre

## LECTURES AND PRACTICALS:

- Wednesdays and Fridays, 10:00 – 11:00, through 2 December
- Practical demonstrations Wednesdays 11-12 *in MR15*
- Lecture notes and Practicals available on Moogle

## ASSESSED WORK:

- Two problem sets due during term
- Term project due after New Year
- Presentations beginning of Lent Term

# 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