Lecture 2

Review & intro to OMICS

RNA & proteins

Tools for studying DNA

RNA content of a cell

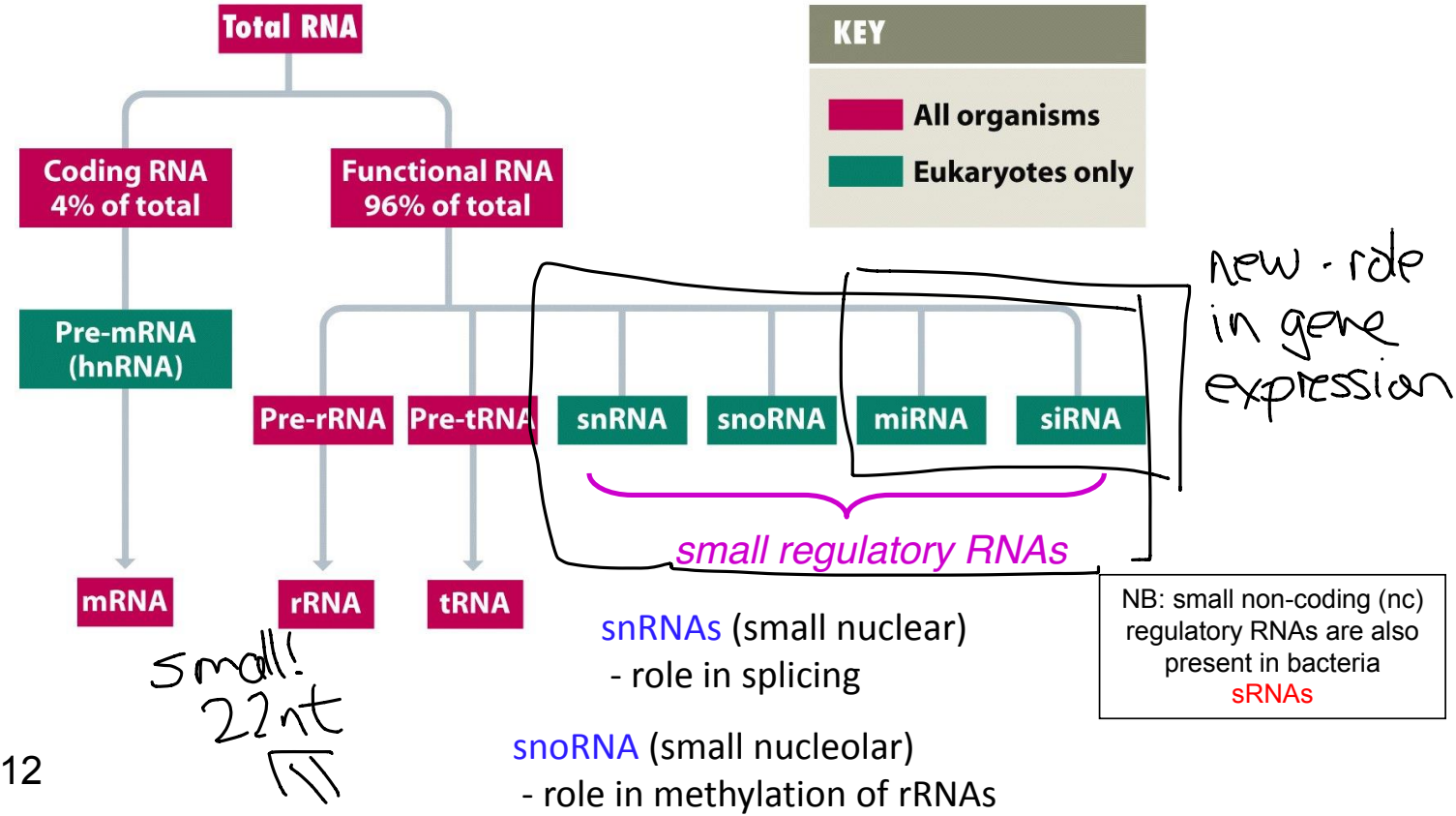


Fig.1.12

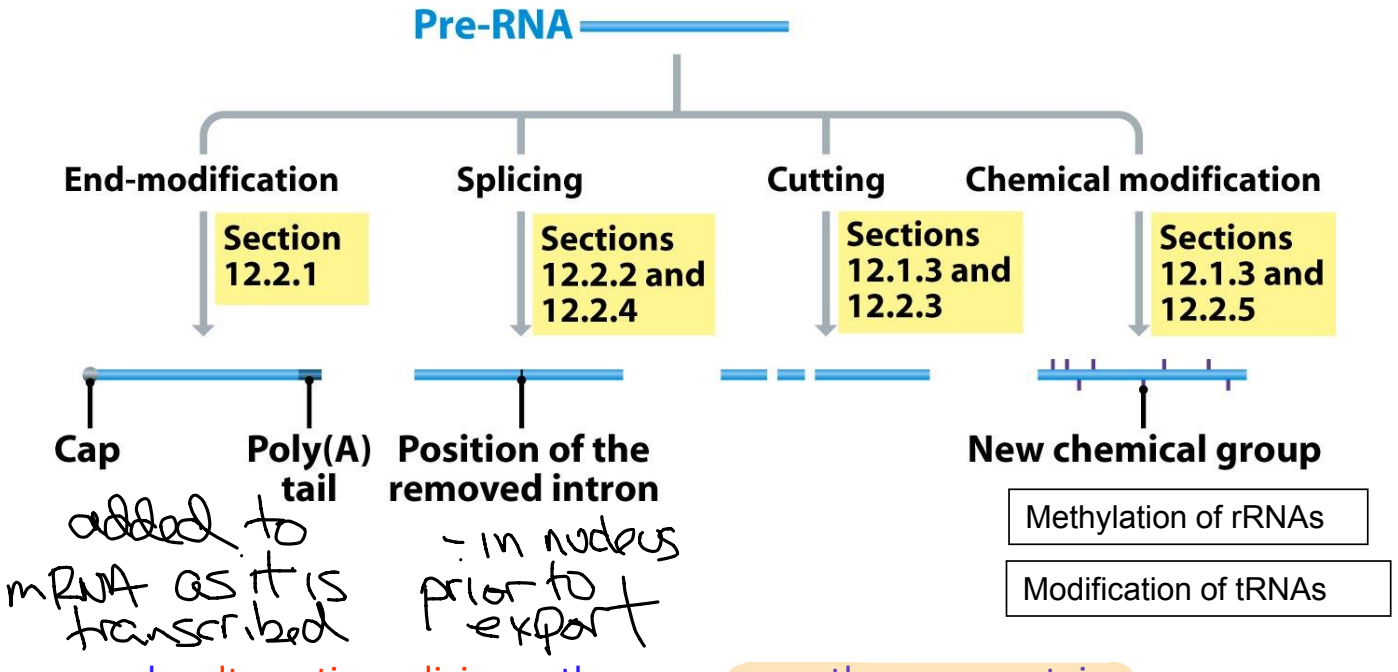
snRNAs (small nuclear)
- role in splicing

snoRNA (small nucleolar)
- role in methylation of rRNAs

miRNA (microRNAs) & siRNA (short interfering RNAs)
- role in regulation of expression of individual genes

lncRNA long non-coding RNAs (~200 nt or longer)
- role in regulation such as chromatin remodeling

RNA processing in eukaryotes

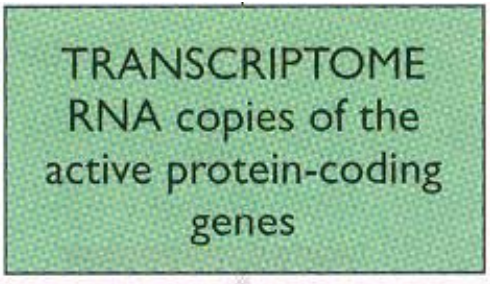


- may be alternative splicing pathways so more than one protein generated from one gene (Discussed later, Chapter 6)

- presence of long introns (& short exons) can make finding genes in eukaryotic DNA sequences difficult

Fig.1.13

Link between transcriptome & proteome



Second letter

First letter	U	C	A	G	Third letter
U	UUU Phe, UUC Phe	UCU Ser, UCC Ser, UCA Ser, UCG Ser	UAU Tyr, UAC Tyr	UGU Cys, UGC Cys	U
	UUA Leu, UUG Leu		UAA Stop, UAG Stop	UGA Stop, UGG Trp	C, A, G
C	CUU Leu, CUC Leu, CUA Leu, CUG Leu	CCU Pro, CCC Pro, CCA Pro, CCG Pro	CAU His, CAC His	CGU Arg, CGC Arg, CGA Arg, CGG Arg	U, C, A, G
			CAA Gln, CAG Gln		
A	AUU Ile, AUC Ile, AUA Ile	ACU Thr, ACC Thr, ACA Thr, ACG Thr	AAU Asn, AAC Asn	AGU Ser, AGC Ser	U, C, A, G
	AUG Met		AAA Lys, AAG Lys	AGA Arg, AGG Arg	
G	GUU Val, GUC Val, GUA Val, GUG Val	GCU Ala, GCC Ala, GCA Ala, GCG Ala	GAU Asp, GAC Asp	GGU Gly, GGC Gly, GGA Gly, GGG Gly	U, C, A, G
			GAA Glu, GAG Glu		

"standard code"

- can deduce amino acid sequence of protein from nt coding sequence

… using genetic code table

Fig.1.2

Fig.1.20

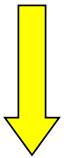
PROTEIN-CODING GENES

divided into codons

DNA

5’ …. ATG GGA TTG CCC GCC …. 3’ “coding strand”

3’ …. TAC CCT AAC GGG CGG …. 5’ “template strand”



mRNA 5’ …. AUG GGA UUG CCC GCC …. 3’

- *in research papers DNA usually shown as single-stranded with coding strand in 5’ to 3’ orientation (left to right)*

… so genetic code table can be used directly

Remember that although AUG is the standard initiation codon, there can also be AUG triplets within an ORF (open reading frame); — protein coding region.

… specifying internal Met residues in the protein

And when analyzing DNA data obtained in the lab, initiation codon might be located outside the sequenced region

Examples of deviation from the standard genetic code in mitochondria and microbes

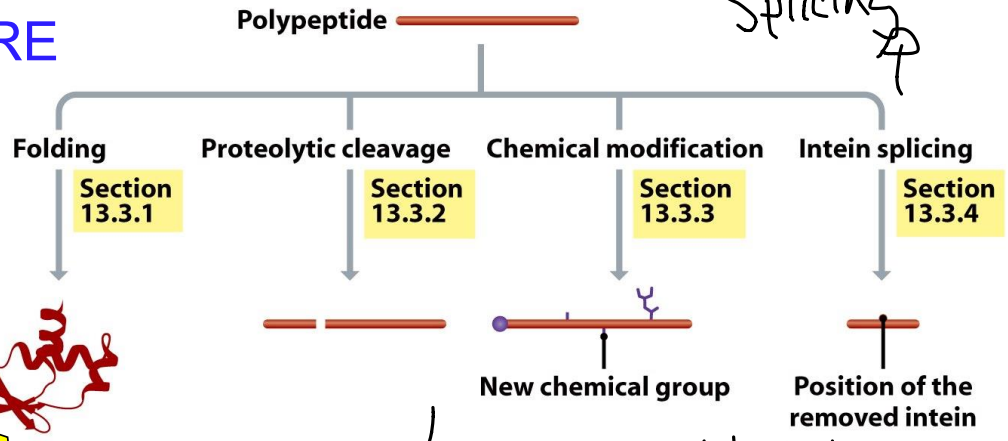
Organism	Codon	Should code for	Actually codes for
Mitochondrial genomes			
Mammals	UGA	Stop	Trp
	AGA, AGG	Arg	Stop
	AUA	Ile	Met
Drosophila	UGA	Stop	Trp
	AGA	Arg	Ser
	AUA	Ile	Met
Saccharomyces cerevisiae	UGA	Stop	Trp
	CUN	Leu	Thr
	AUA	Ile	Met
Fungi	UGA	Stop	Trp
Maize	CGG	Arg	Trp
Nuclear and prokaryotic genomes			
Several protozoa	UAA, UAG	Stop	Gln
Candida cylindracea	CUG	Leu	Ser
Micrococcus sp.	AGA	Arg	Stop
	AUA	Ile	Stop
Euplotes sp.	UGA	Stop	Cys
Mycoplasma sp.	UGA	Stop	Trp
	CGG	Arg	Stop

Actually RNA editing changes CGG to UGG in mRNA, so genetic code is NOT deviant in plant mitochondria

Table 1.3

PROTEIN SEQUENCE & STRUCTURE

akin to mRNA splicing



review

post-translational modifications
- phosphorylation
- glycosylation
- many others

Impact function

Fig.13.24

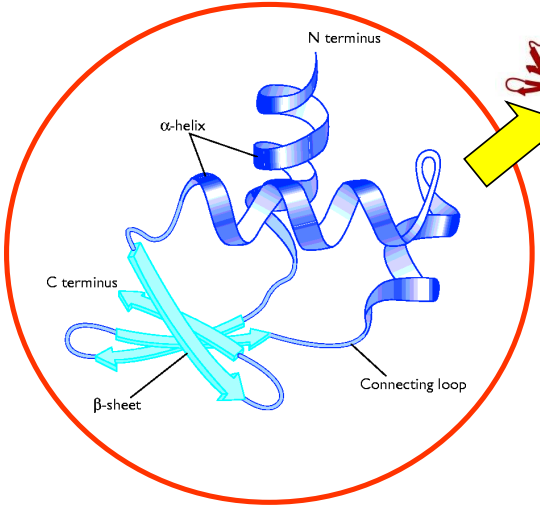


Fig.1.17

Different proteins can be generated from single precursor polypeptide through post-translational events

…so can have larger proteome (set of proteins) than predicted from number of genes in genome

& alternative splicing can also contribute to larger proteome

TECHNIQUES & TOOLS FOR STUDYING DNA

Genomes are very large…

- so need methods to obtain small (relatively speaking) sections of DNA in abundant & pure form for molecular analysis

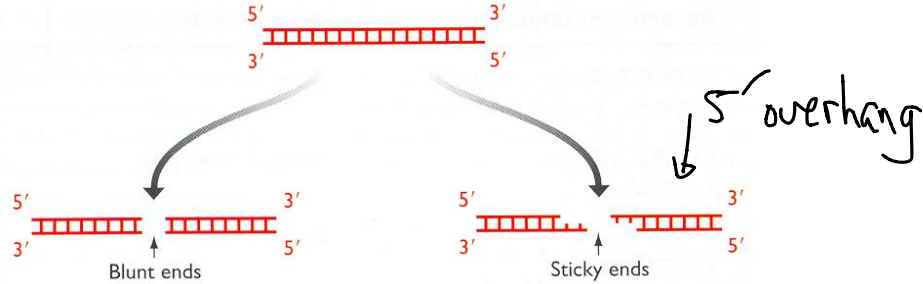
1. *Restriction enzyme cleavage* & agarose gel electrophoresis
2. *Southern hybridization* & *northern hybridization*
3. *PCR (polymerase chain reaction)* & *RT-PCR*
4. *Molecular cloning*

1. Restriction endonucleases - cleave DNA into specific, small fragments

- made by bacteria to cut viral DNA

DNA endonucleases cut double-stranded DNA at specific recognition sites

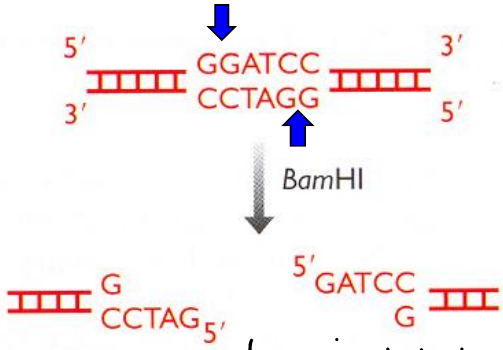
2 types of double strand cuts



"blunt" ends

"sticky" ends
Cohesive ends

Fig.2.10



*Bam*HI: staggered cut with 5' overhang

- some give 3' overhang

- most cut within recognition sequence. - Type IIS don't. Will see these later.

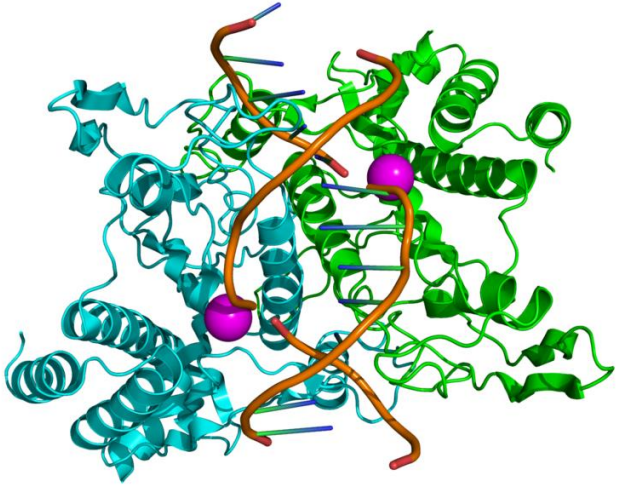
same on both strands

4 types (I-IV) restriction endonucleases occur naturally

5’-GAATTC-3’
3’-CTTAAG-5’

Most commonly used are Type II

Type II
- recognize a palindromic DNA sequence of 4-8 bp
- cut within the recognition site
- homodimeric
- require Mg2+ as cofactor



EcoRI bound to major groove of DNA

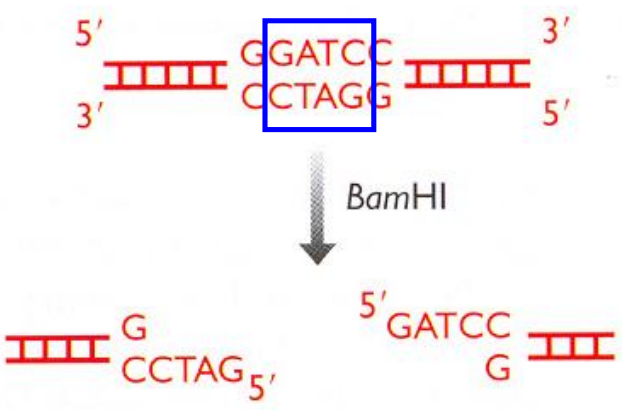
Nomenclature

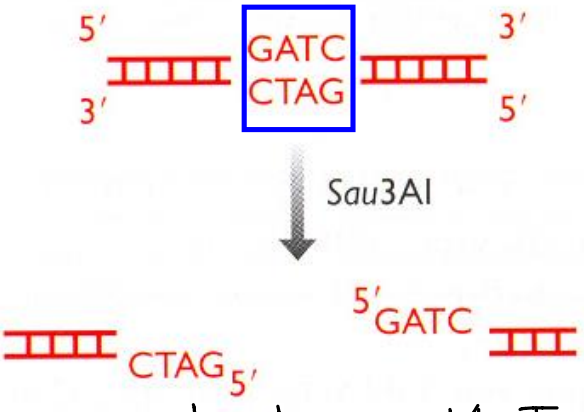
Example EcoRI

E genus of producing strain
co species of producing strain
R additional strain descriptor
I order of identification from strain

Recognition sequences often 4 or 6 bp, but also “rare cutters” (eg. NotI 5’ GCGGCCGC 3’) can be useful for generating very large fragments in genomic mapping

Two different restriction enzymes may generate same “sticky ends”





lookup XbaI SpeI AvrII NheI

-“compatible ends” useful for cloning
(eg. partial Sau3A genomic digest ligated into BamHI site in vector)

Isoschizomers – restriction enzymes with identical recognition sequences

… but may have different response to methylation state

MspI cleaves 5’ CCGG 3’ regardless of methylation state

HpaII does not cleave 5’ CCGG 3’ if 2d C is methylated

Useful for epigenetic studies (evaluating DNA methylation status near gene of interest)

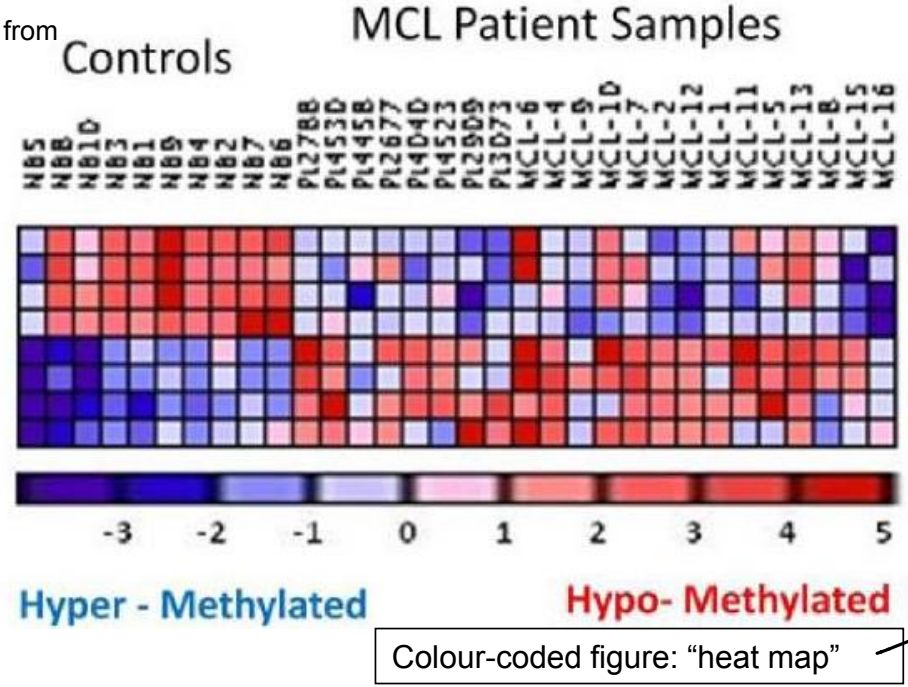
Example using isoschizomers to assess DNA methylation state of genes in cancer patients

"called HELP"

- used assay called "*Hpa*II tiny fragment Enrichment by Ligation–mediated PCR"
- found "significant aberrancy in promoter methylation patterns compared with normal NBCs"

Mantle Cell Lymphoma

NBC: naïve B cells (ie. from healthy people)



Colour-coded figure: "heat map"

→ very common in genomics

"Genome-wide DNA methylation analysis reveals novel targets for drug development in mantle cell lymphoma"

Leshchenko et al. *Blood* 116:1025, 2010

Agarose gel electrophoresis - to separate DNA fragments by size

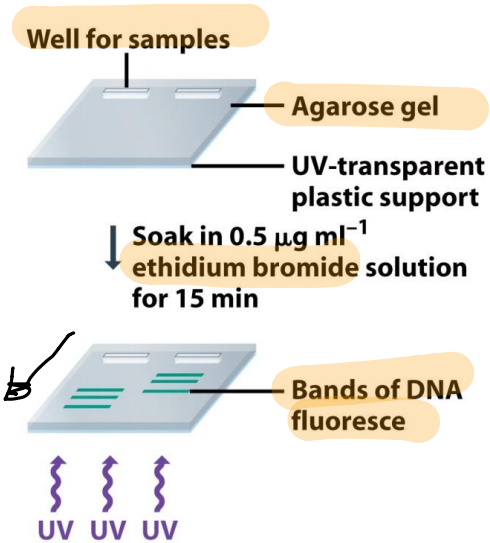


Fig.T2.1

Small fragments migrate more rapidly than large ones

- works great upto ≃ 20Kb

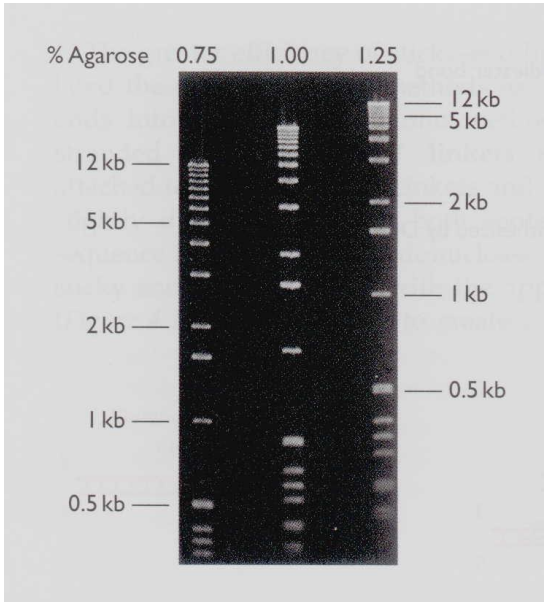


Fig.T2.2

Lanes with size markers, on gels of different % agarose

"Pulsed field" electrophoresis for separation of large DNA molecules (Fig.3.30)

For example : restriction fragments generated by "rare cutters"

- megaplasmids
- whole chromosomes (eg. yeast)

- changes orientation of field
- large DNA responds slowly

Because genomic DNA (bacterial or eukaryotic) is very complex, number of restriction fragments generated is too large to see discrete bands after electrophoresis...

so "continuum" of signals in lane

Why are different profiles expected for genomic DNA cleaved with *Bam*HI (6 bp cutter) vs. *Sau*3A (4 bp cutter)?

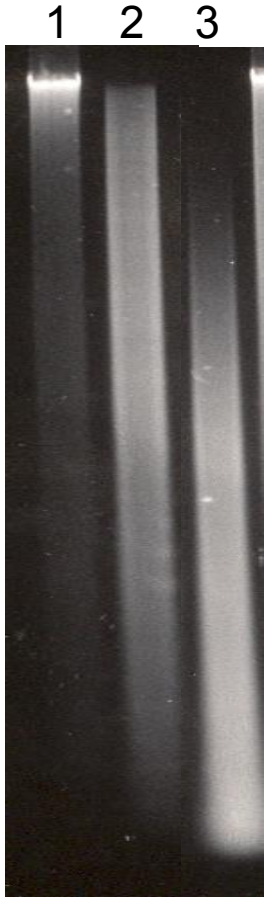
- 4bp site is more frequent

assume 50% GC content

need $4^4 = 256$ - site occurs ≈ every 250 bp!

need 4^6 for 6bp recognition site to occur by chance.

$4^6 = 4096$ ≈4kb

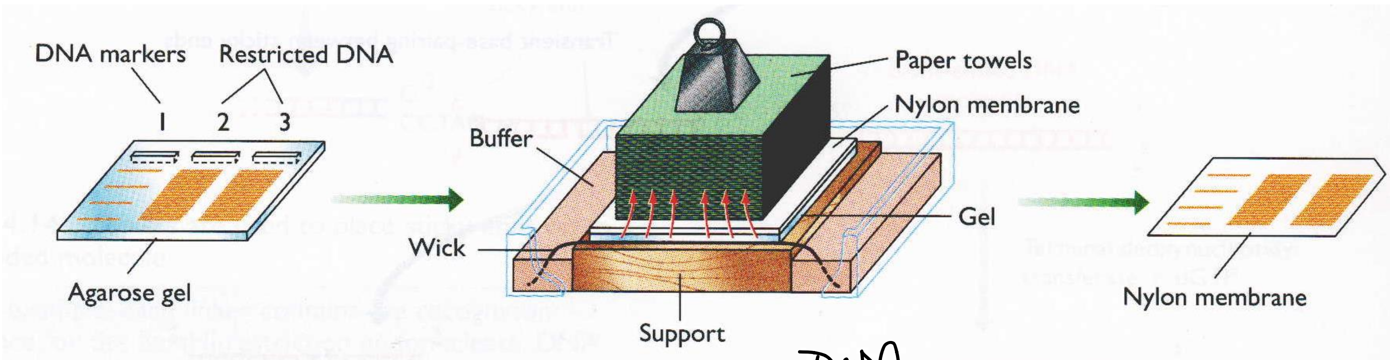


Lane 1 = uncut DNA
Lane 2 = 6 bp cutter
Lane 3 = 4 bp cutter

Developed by Edwin Southern in 1975

2. Southern hybridization

- to detect specific DNA fragment containing sequence (eg.gene) of interest (vs. all other fragments)

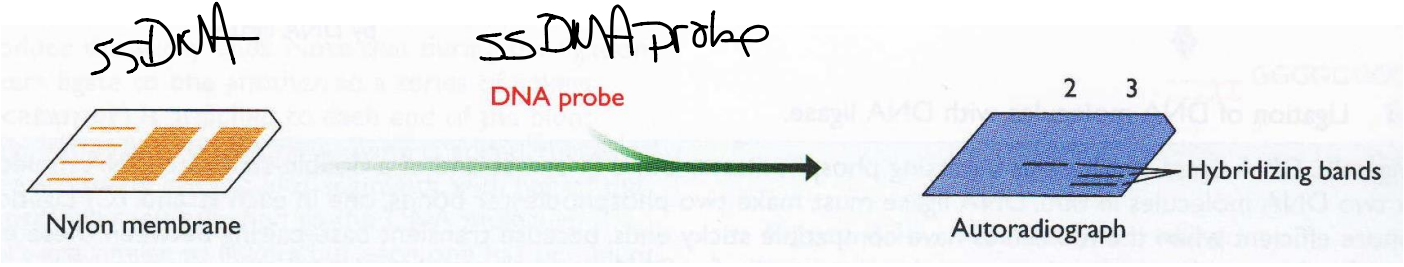


→ make ssDNA

Fig.2.11A

1. *After electrophoresis, <u>denature</u> DNA and transfer it from gel to membrane (eg. by capillary action or electroblotting…)*

…so that DNA fragments remain in same relative positions



- *probe will anneal with single-stranded DNA on blot, if sequences are complementary*

Fig.2.11B

2. Hybridize blot with "probe" (DNA, oligomer, cDNA, or RNA… which is tagged either radioactively or non-radiolabelled) and detect specific hybrid by autoradiography

Stability of hybrid depends on:

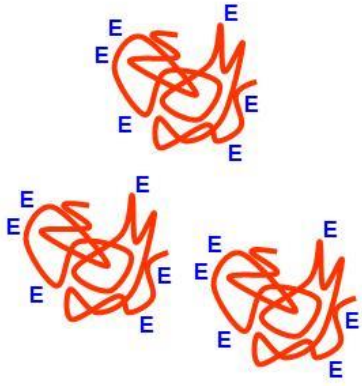
- length of hybrid, (eg for oligomer probes), GC content …
- hybridization conditions (such as temperature, ionic strength…)

15-20' mer probes can form stable hybrids under appropriate conditions

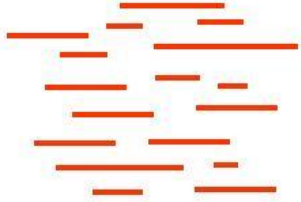
..and hybrids of ~ 50 bp or longer are stable under all standard conditions

Example

Isolate DNA

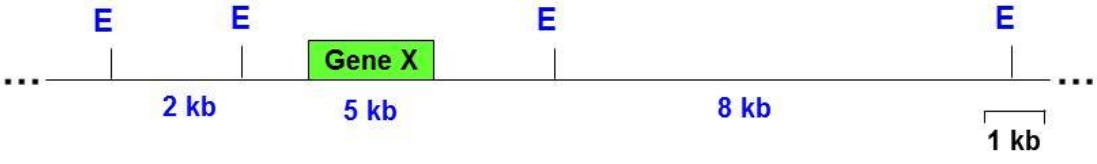


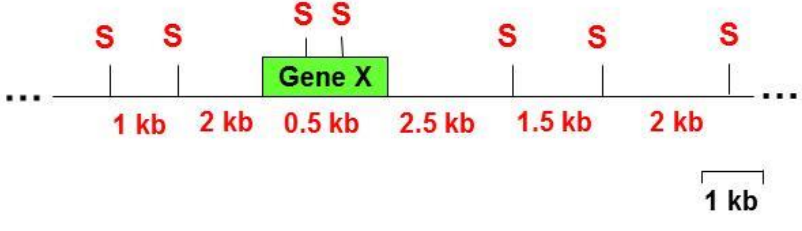
Restrict DNA with enzyme E

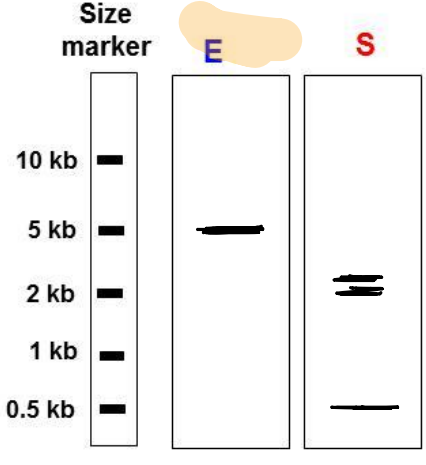


Electrophoresis of digested DNA, preparation of Southern blot

Probe blot with gene X (which is 2 kb long)







What do you expect to see?

Some applications of Southern blot analysis

- to identify restriction fragment carrying sequence (eg gene) of interest
- to identify gene copy number (eg. multi-gene families)
- to identify homologous genes in other organisms (zoo blot)

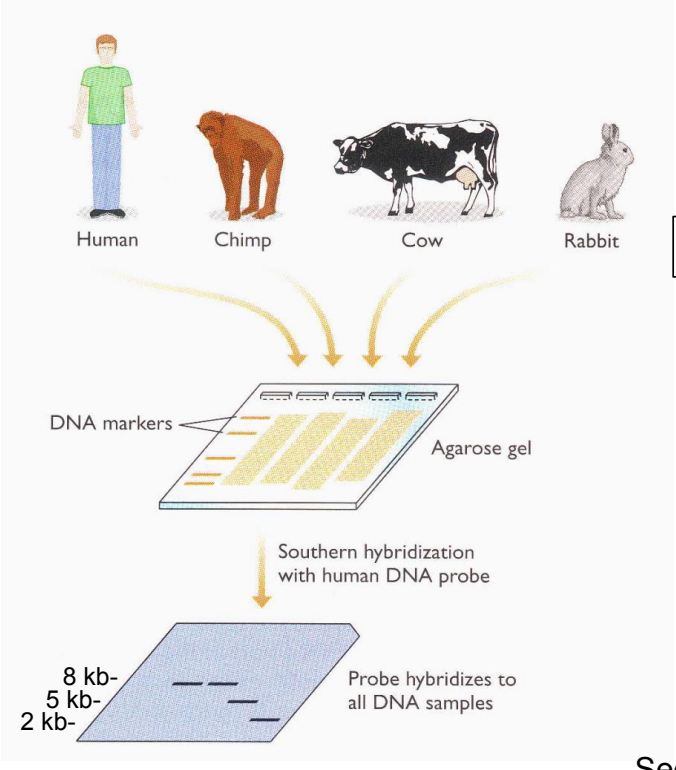


Fig. 5.12

Heterologous hybridization
- use conditions of "reduced stringency" (eg lower temp) so that duplex hybrids with some mismatches are stable

Interpretation of data shown in figure?

suppose probe is human gene X...

- ortholog of gene X in all 4 species
- restriction patterns are different in cow & rabbit

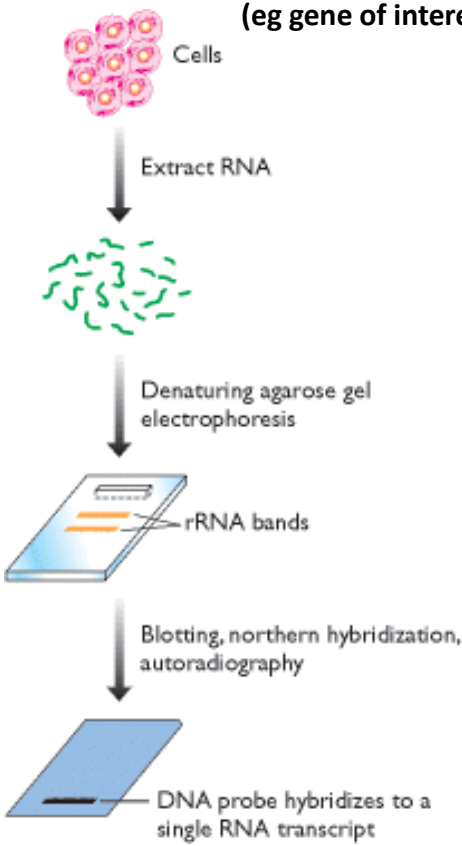
See also : Southern of pulsed-field gel , Boritsch *Nature Microbiol.* 1:15019, 2016

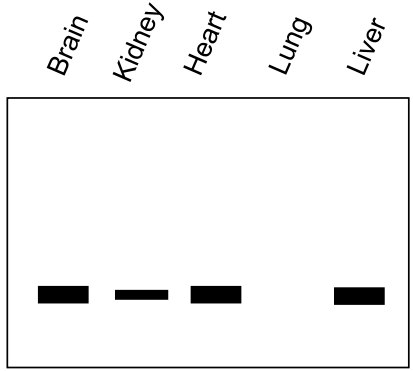
Northern hybridization (RNA blot) analysis

- to identify <u>expressed</u> regions of genomes

– RNA is electrophoresed, blotted to membrane and hybridized with probe (eg gene of interest)

- are transcripts are present or not?
- length of mRNA?
- abundance of mRNA?

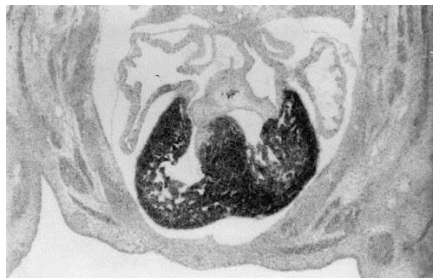




gene X probe

Probe: tagged *DNA (eg. PCR product, restriction fragment, cDNA clone…) in denatured form or oligomer or antisense (synthetic) RNA …*

In situ hybridization
- to determine cellular Location of mRNA



^{35}S-labeled β-myosin antisense probe hybridizing to heart ventricle in 13-day embryonic mouse

Strachan & Read Fig. 5.17

Fig. 5.11 **(NB: many identical copies of a particular mRNA are present on blot)**

See also: Northerns of new ncRNAs (van der Meulen *RNA Biol*. 13: 353, 2016)

Some protein encoding genes are constitutively expressed …

"housekeeping gene" products needed at all times

… whereas others are differentially expressed

in specific tissue type

during development

in response to environmental cues

Only a subset of genes are expressed at a given time
and mRNA levels can vary greatly among genes

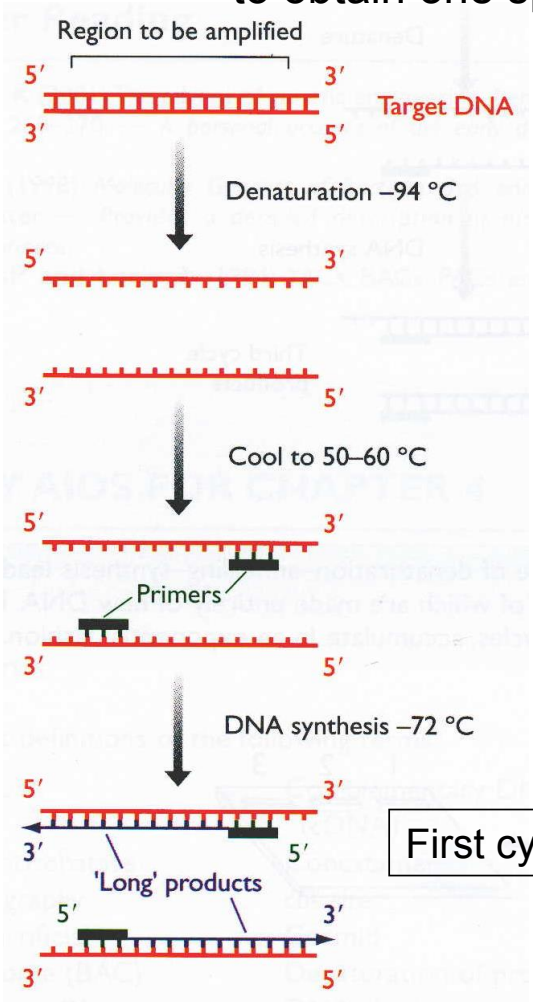
~10,000 – 15,000 different mRNAs present in "typical" mammalian cell type under given condition

(~ 20,000 different proteins present)

Aside: RNA-sequencing studies suggest ~ 8000 genes ubiquitously expressed in human tissues (Ramskold *PLoS* 2009)

3. PCR - polymerase chain reaction

- to obtain one specific DNA region in large copy number



First cycle

Fig.2.28

- rapid amplification of DNA region of interest by enzymatic reaction in test tube

1. Denaturation of duplex DNA

2. Annealing of 2 different primers
(synthetic oligomers, usually 15-25 nt)
- flank region of interest,
- in opposite orientation
… so anneal to opposite strands of DNA

3. Extension of complementary strands

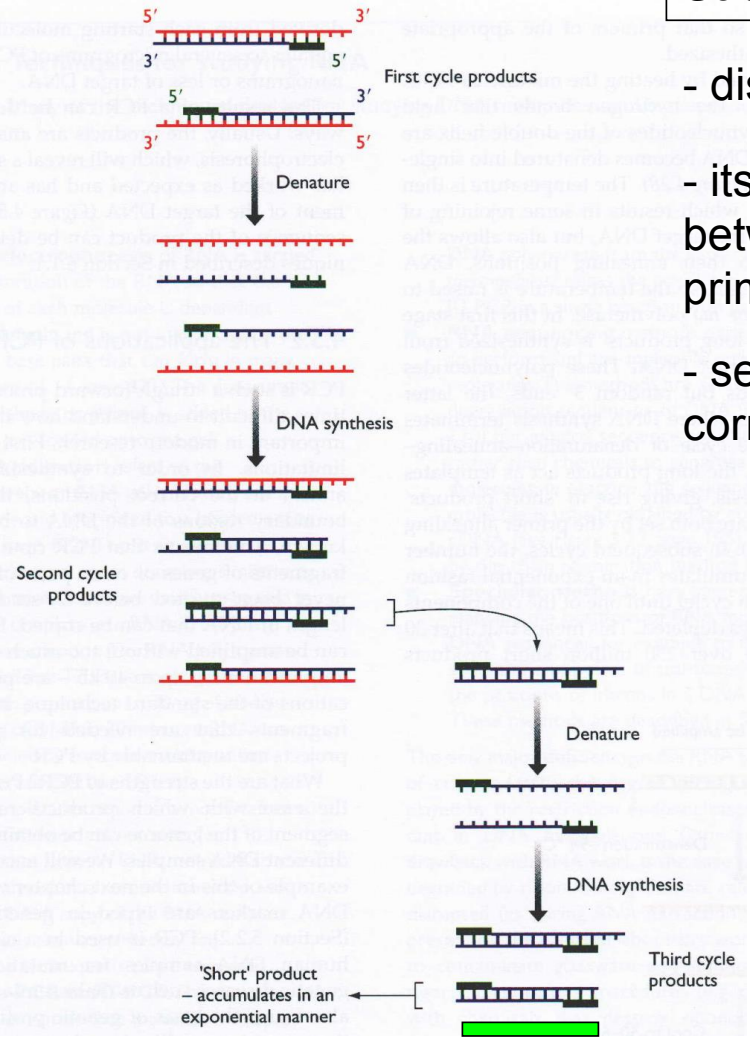
- cycle repeated 25-30 times

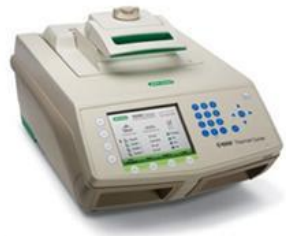
Subsequent PCR cycles

- discrete PCR product generated

- its length corresponds to distance between primers (including the primers)

- sequences at ends of amplicon correspond to the 2 primers used





Bio-Rad C1000 Thermal Cycler

NB: even though text shows both primers as black bars, they have different nt sequences

Fig.2.29

Powers & pitfalls of PCR

- rapid method to generate large amounts of specific segment of DNA (product usually < 10 kb in length)

"Long-range PCR"
~ 20-30 kb products

- need very small amount of template DNA

…but can lead to contamination problems

- need prior sequence info to design primers

Typically use 20-25 nt oligomers,
but for simplicity (as on a test) 6'mers are shown here

Why choose ~ 20'mers?

Tip: see Question 2.5 in text (p.61)

How to double-check that PCR product (amplicon) is correct one?

1. Is it the right size?

- agarose gel electrophoresis (with size markers)

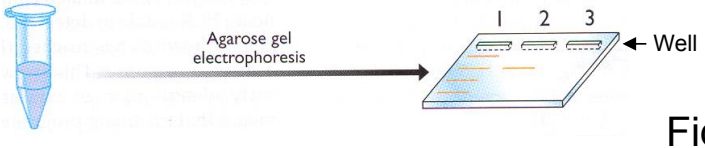
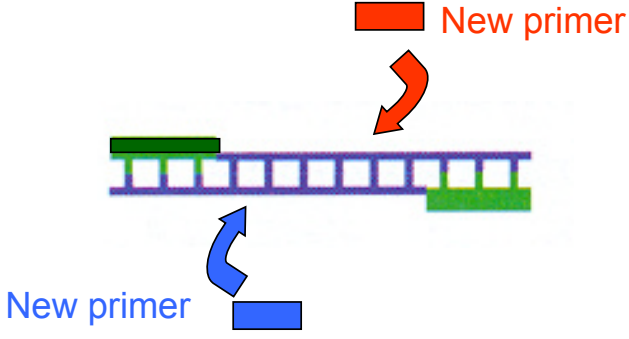


Fig. 2.30

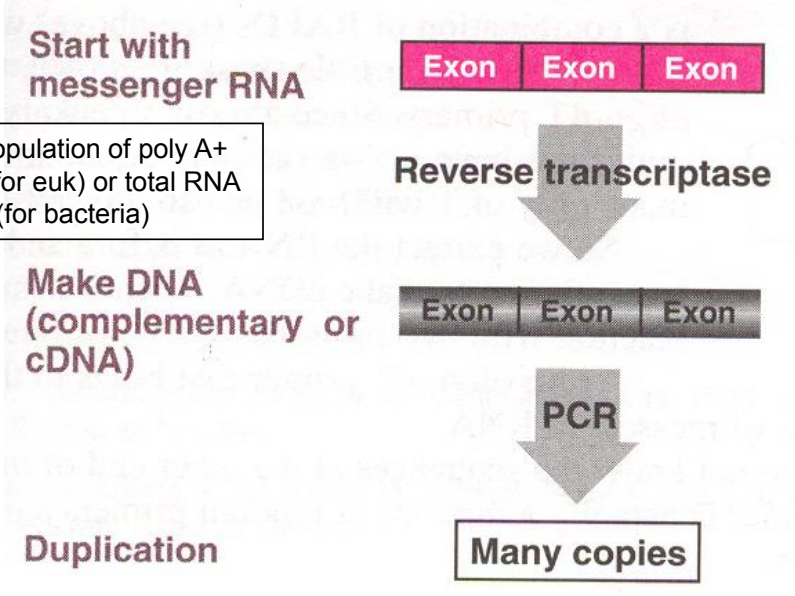
2. Does it contain the right sequence (eg gene X)?

- Southern hybridization - using gene X (eg. clone) as probe
- restriction analysis - are expected restriction sites present?
- nested PCR

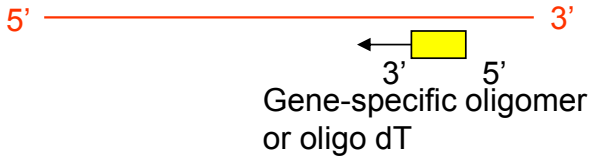
- design "internal" primers
to use in 2d PCR experiment with
1st PCR product as template DNA



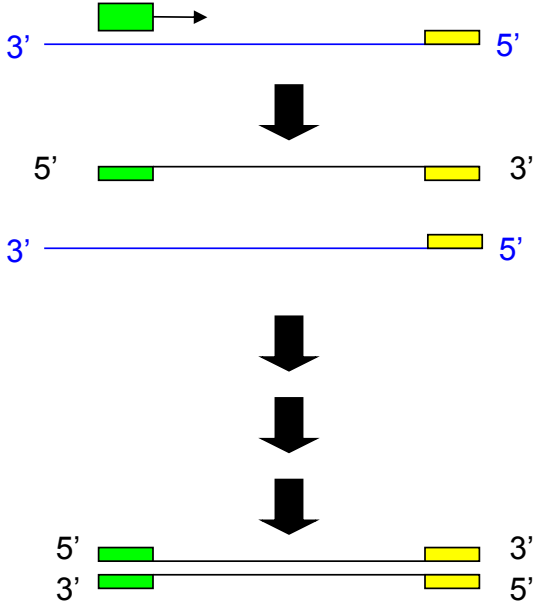
RT-PCR
Reverse Transcriptase



total population of poly A+ RNAs (for euk) or total RNA (for bacteria)



Gene-specific oligomer or oligo dT



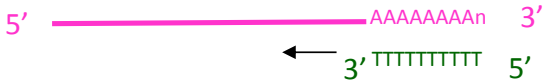
- then sequence RT- PCR product directly (or after cloning)

- (need sequence data to design primers for RT-PCR)

(see p.142, Chapter 5)

Choice of primer for cDNA synthesis

1. oligo dT - anneals to polyA tail of mRNA (if eukaryotic)

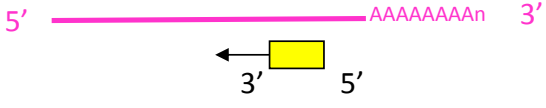


Generates copies of all mRNAs present

2. Random short primers (eg. 6’mers)

Generates partial copies of mRNAs

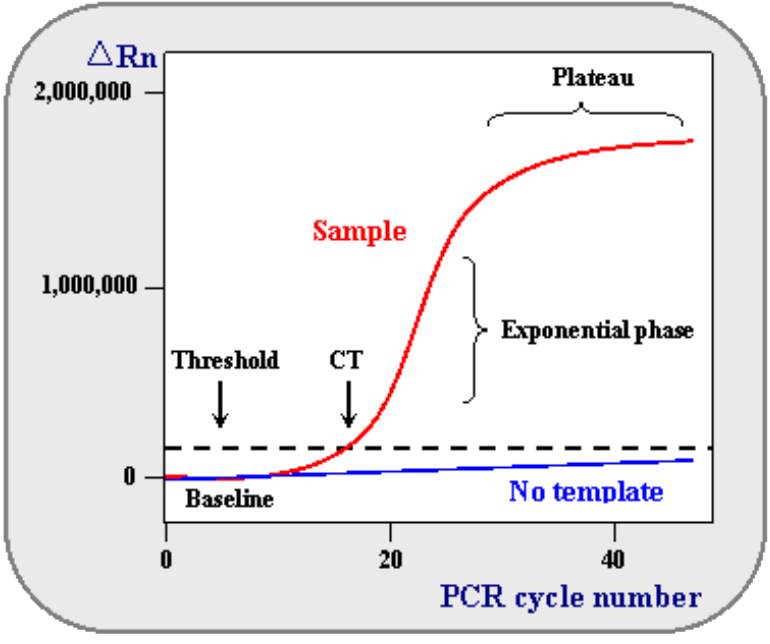
3. Sequence-specific primer



Generates cDNA just for specific gene of interest

Real-time quantitative qPCR

- to measure relative or absolute amount of mRNA present in different tissue types/developmental stages/environmental conditions…

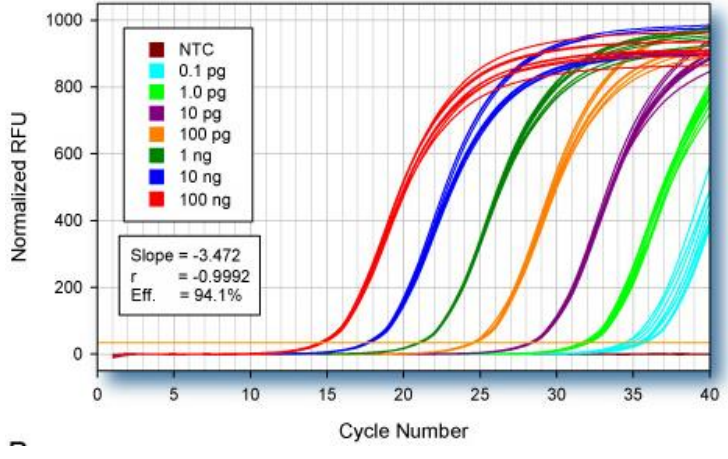


ΔRn : increment of fluorescent signal at each time point

CT : PCR cycle number where reporter fluorescence is greater than threshold

- detection and measurement of products generated during each cycle of PCR by using a reporter fluorescent probe

eg. SYBR green, TaqMan



RFU = relative fluorescence units
NTC = no template control

http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechQPCR.shtml
NCBI Technologies website