Lecture 4

DNA recombinant techniques (part 2):

- PCR
- Real-time PCR (aka qPCR)
- DNA Sequencing
- Next gen DNA Sequencing, RNA sequencing
- RT-PCR

Required reading for lectures 3-5: Chapters 7 and 8

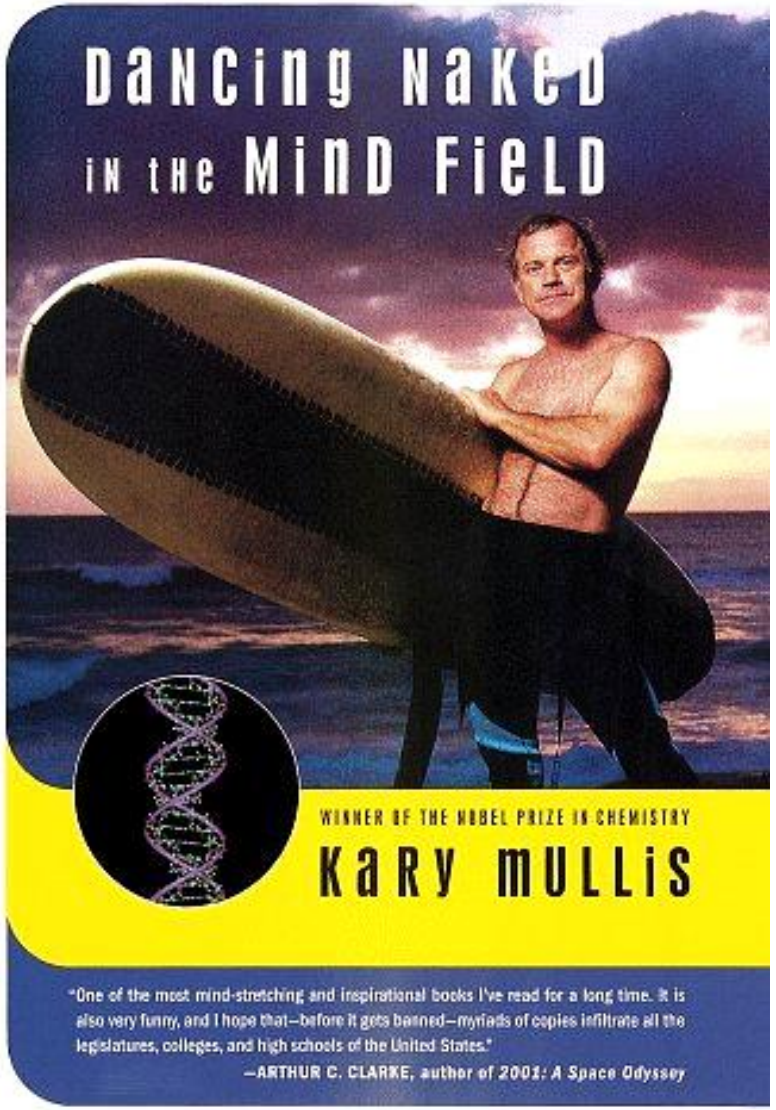
PCR: A BIOTECH REVOLUTION

- **Kary Mullis & Michael Smith, UBC (Nobel Prize 1993)**

  "Beginning with a single molecule of the genetic material DNA, the PCR can generate 100 billion similar molecules in an afternoon. The reaction is easy to execute. It requires no more than a test tube, a few simple reagents, and a source of heat"

- Small DNA amounts amplified by successive rounds of replication

Reaction		
Pre-denaturation		94°C for 10 min
Thermal cycling		
30-35	Denaturation	94°C for 30 sec
	Annealing	40-60°C for 30 sec
	Extension	72°C for 30 sec
Final extension		72°C for 5-10 min



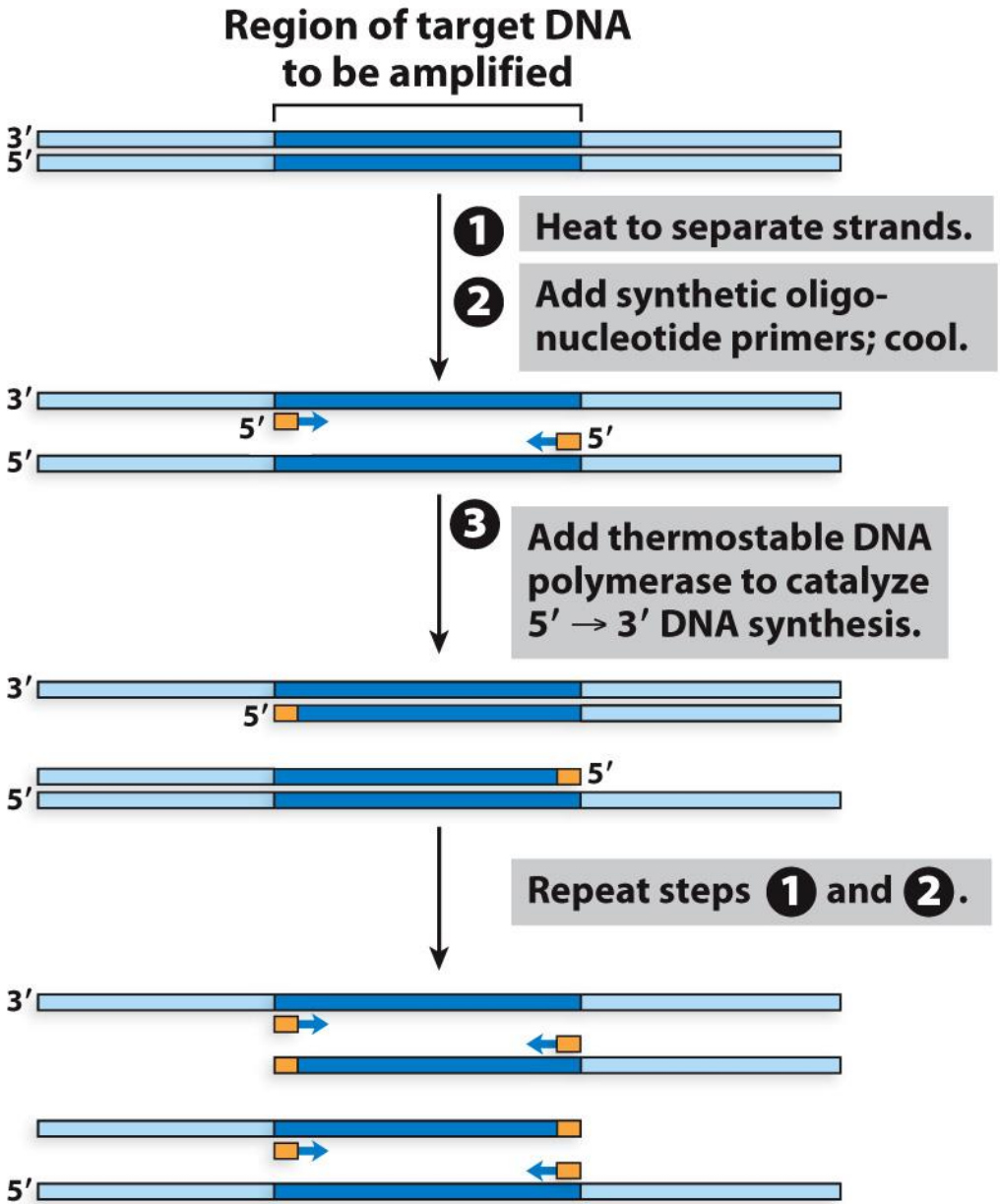


Figure 7-9a part 1

Molecular Biology: Principles and Practice, Second Edition

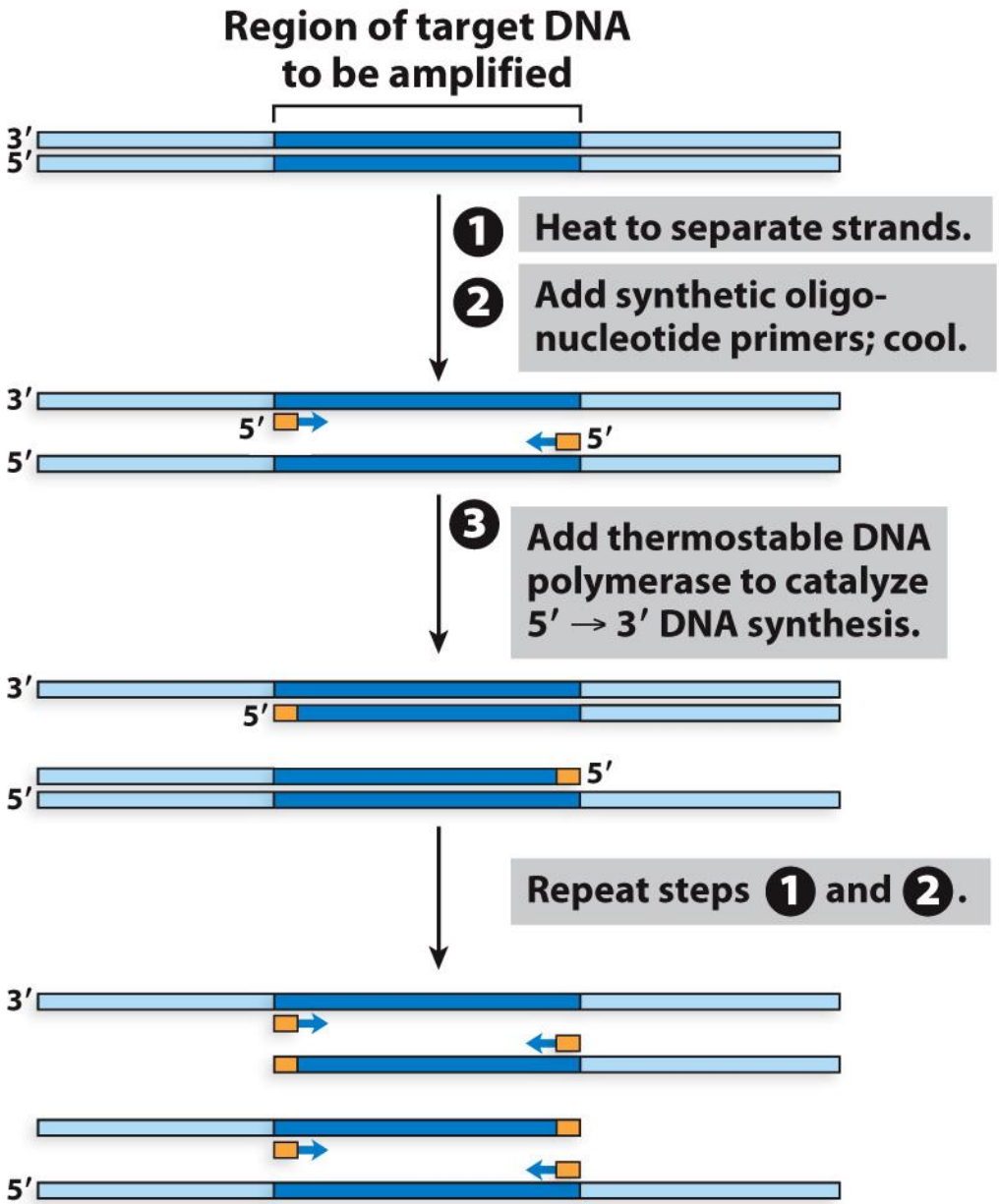


Figure 7-9a part 1
Molecular Biology: Principles and Practice, Second Edition

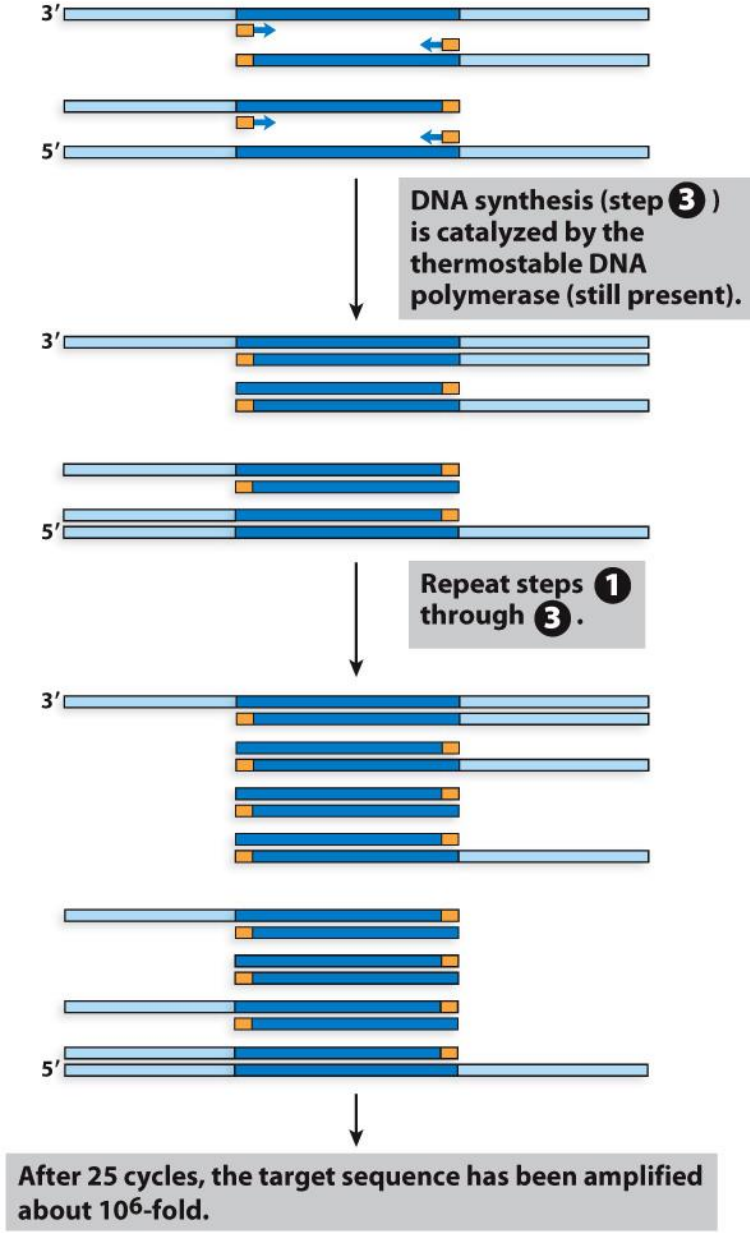
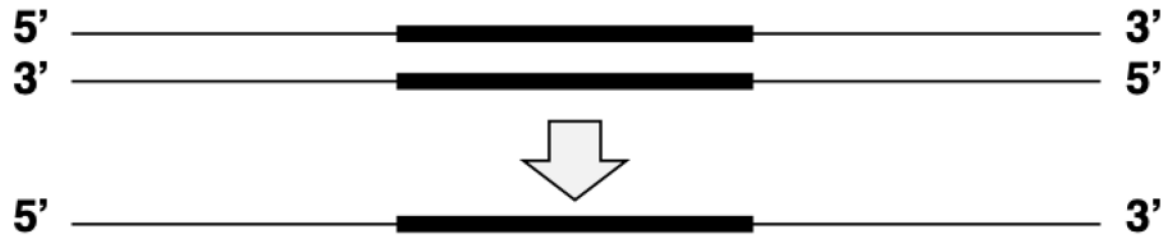


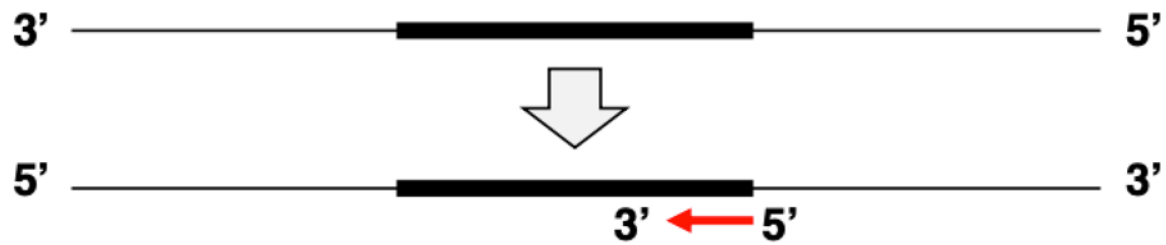
Figure 7-9a part 2
Molecular Biology: Principles and Practice, Second Edition

The first cycle

Template DNA (long)

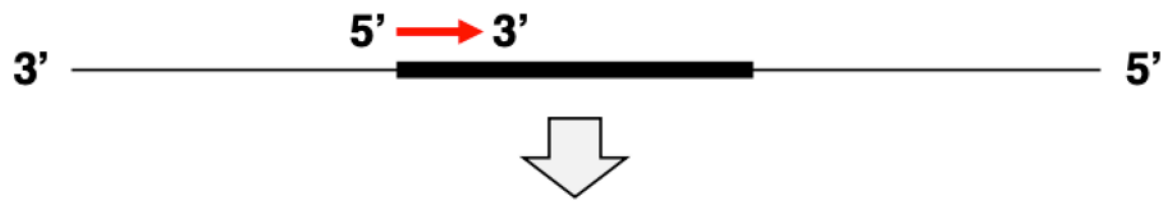


Denaturation in 94°C



Annealing in 40-60°C

Primers

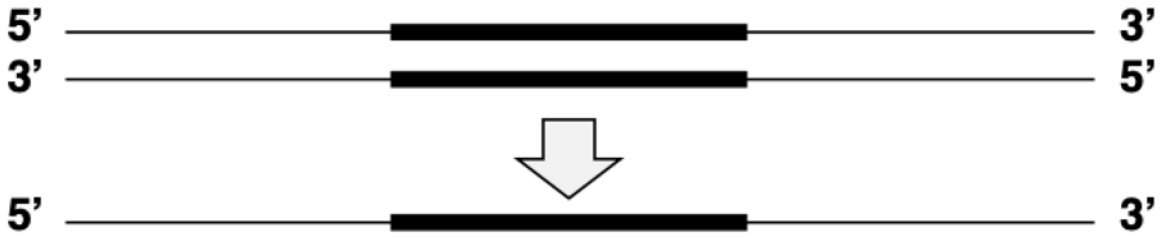


Extension in 72°C

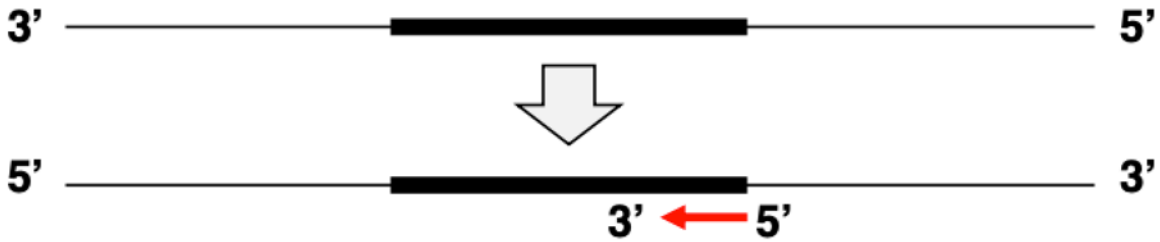
DNA polymerase + dNTP

The first cycle

Template DNA (long)

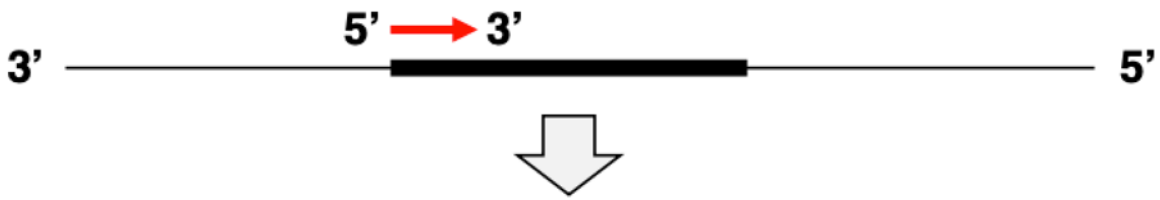


Denaturation in 94°C



Annealing in 40-60°C

Primers



Extension in 72°C

DNA polymerase + dNTP

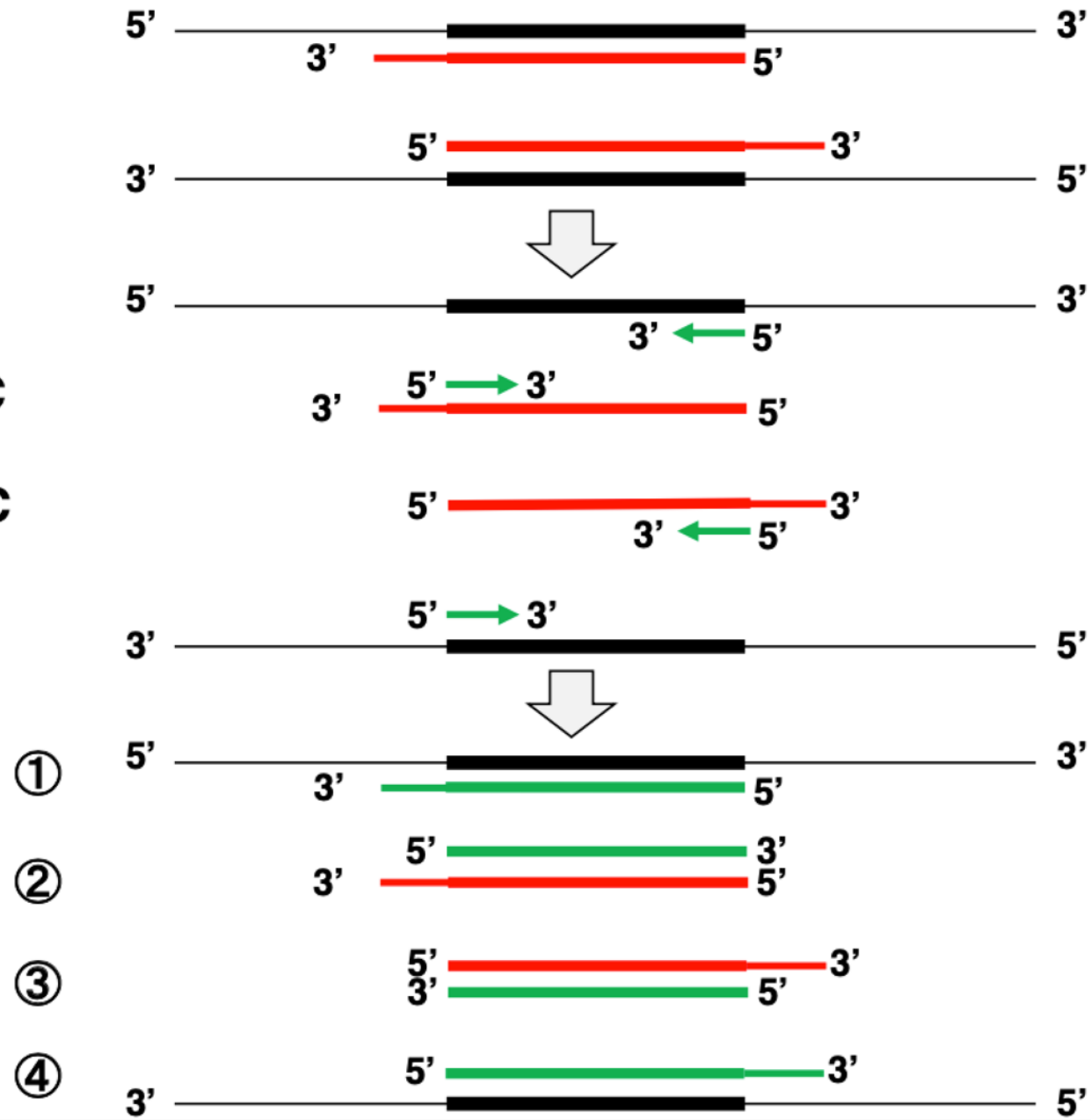
*Length of extension at the 3´end is determined by the time the thermocycler stays at 72ºC

The second cycle

Template DNA + product of the first cycle

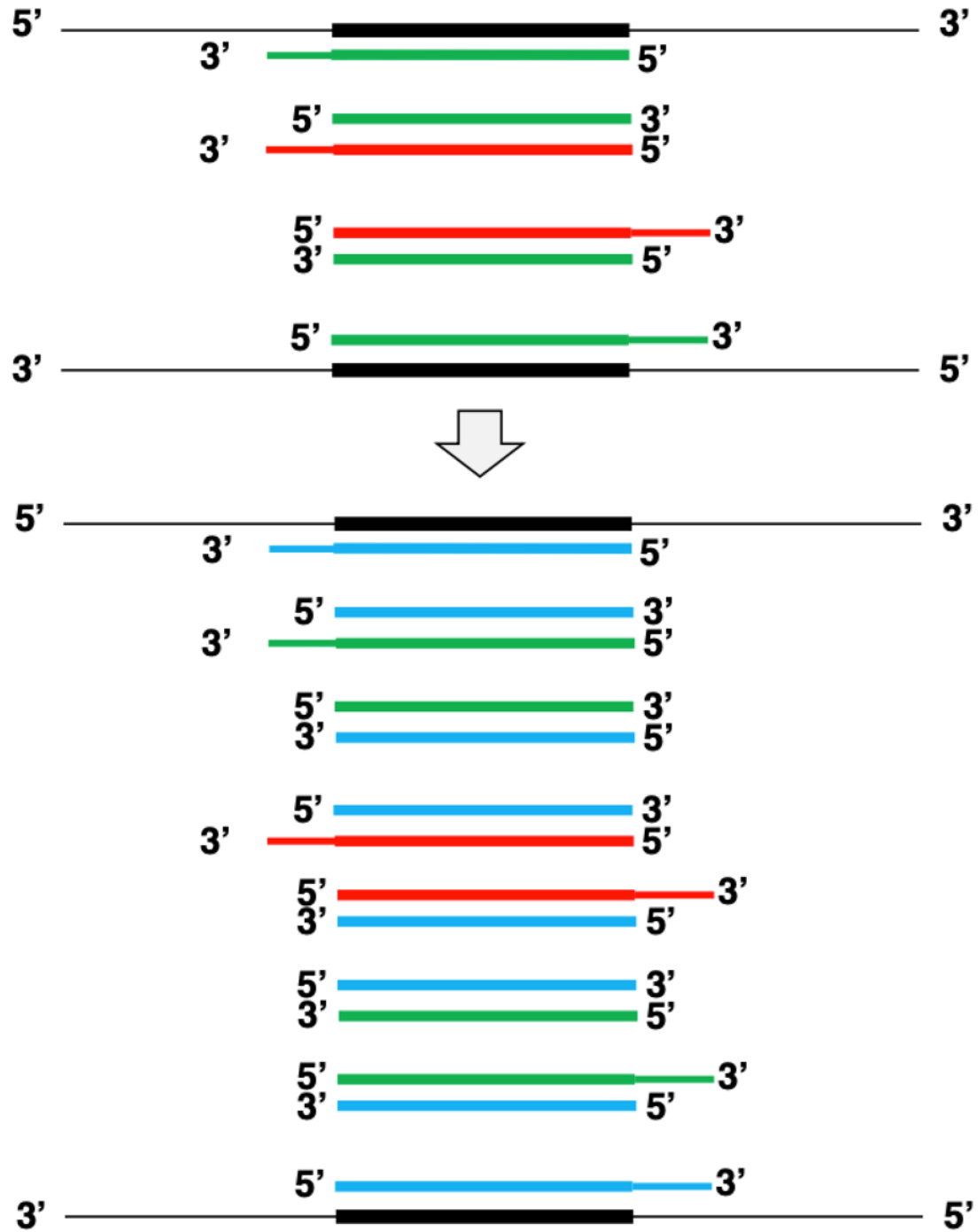
Denaturation in 94°C
&
Annealing in 40-60°C

Extension in 72°C

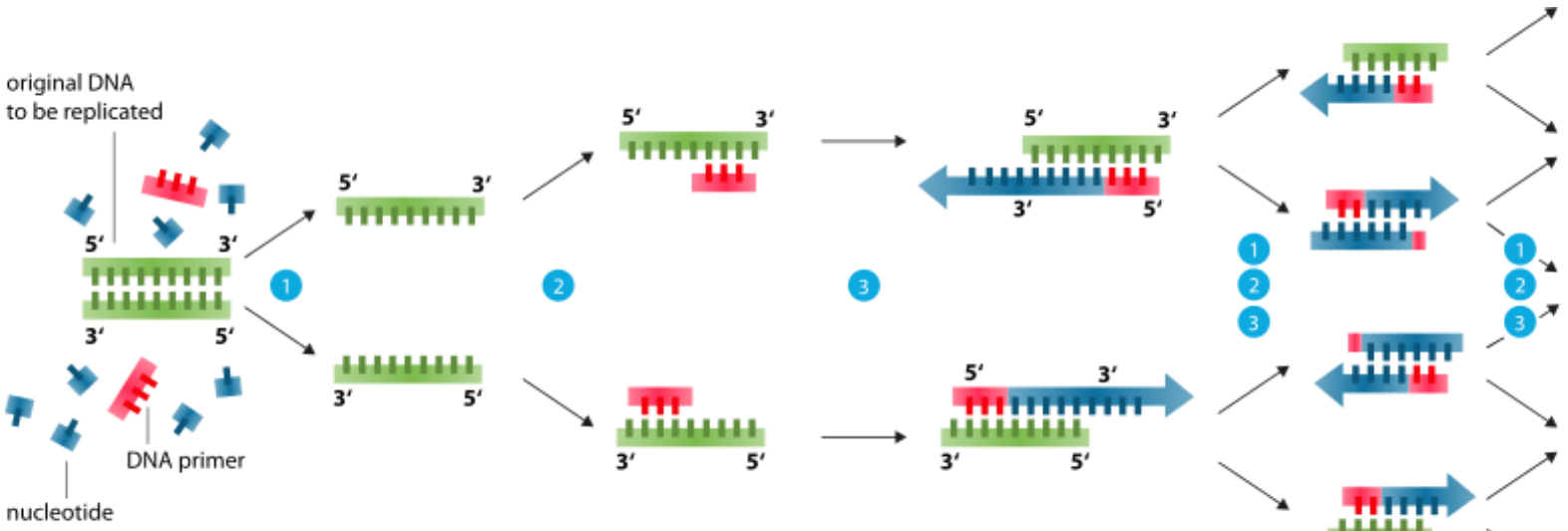


The third cycle

Template DNA + product of the second cycle



PCR set up and primer design



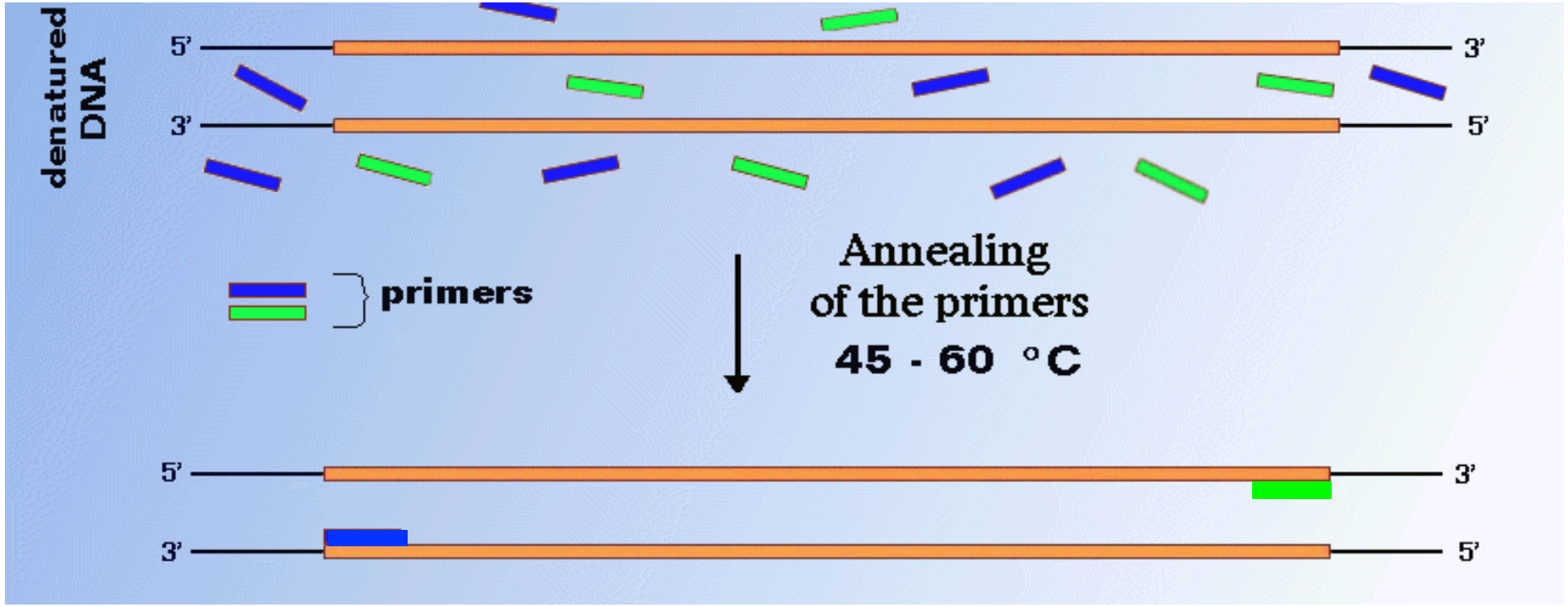
1. **Denaturation** at 94-96°C
2. **Annealing** at ~68°C
3. **Elongation** at ca. 72 °C

Amplification of specific DNAs

Requirements:

1. DNA template
2. Designed DNA primers: T_m - 5˚C

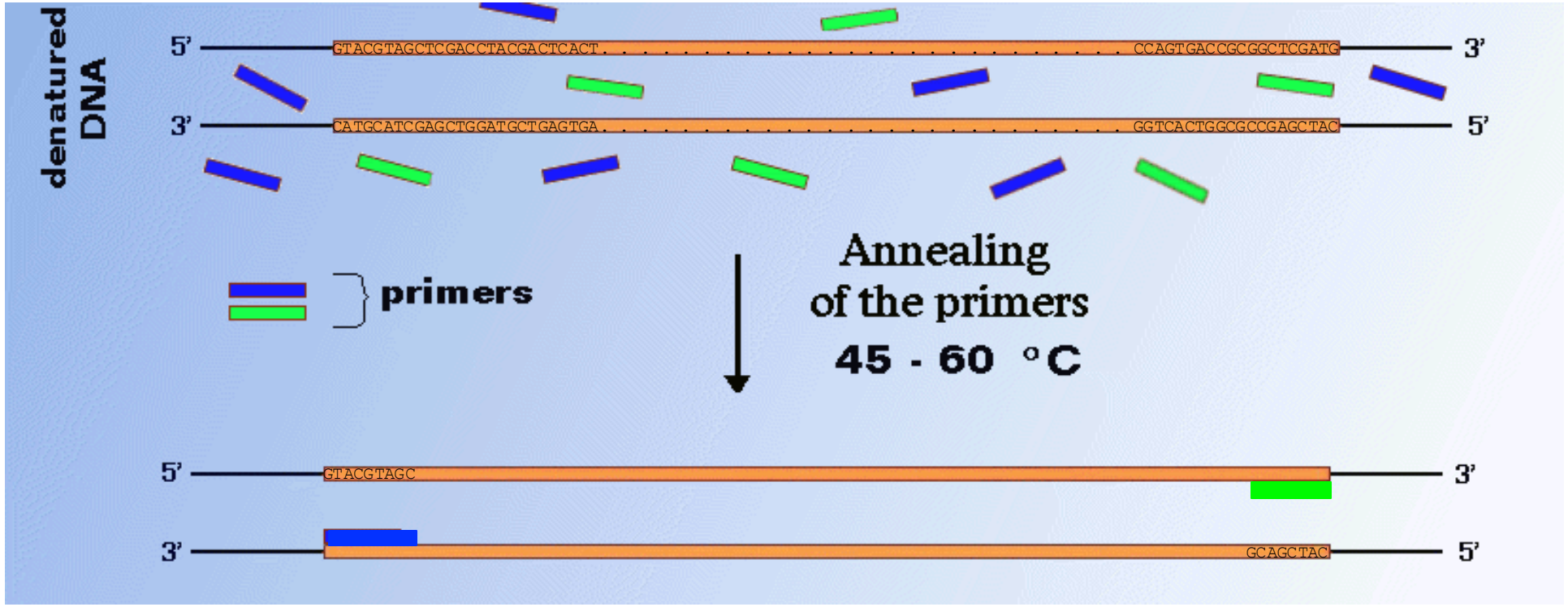
   $T_m = 2(\#A+T) + 4(\#G+C)$
3. dATP, dCTP, dTTP, dGTP (dNTPs)
4. Thermostable DNA polymerase (Taq DNA pol)



To design primers for PCR, you must have some sequence information.

To amplify the sequence in orange, the FORWARD and REVERSE primers (above) must contain the sequences at the 5´ and 3´ ends, respectively.

Remember, dsDNA is the starting template and primers anneal to complementary target sequences.



To design primers for PCR, you must have some sequence information.

To amplify the sequence in orange, the FORWARD and REVERSE primers (above) must contain the sequences at the 5´ and 3´ ends, respectively.

Remember, dsDNA is the starting template and primers anneal to complementary target sequences.

General considerations for primer design

1. **Primers must be long enough to bind to the template with high specificity and a sufficient high T_m**
   - Optimal length is 21 - 30 bp of homology
   - T_m Range: 45 - 70˚C (remembering that T_A is T_m – 5)

2. **Greater than 50% GC content** (to enhance T_m; also consider that shorter oligos cost less)
   - Designing primers that with several Gs or Cs at the 3´ end can be helpful for anchoring primers onto the template

3. **Primer pairs should have sufficiently similar T_ms**
   - Sometimes this is not possible (for example when using long and short primers): in these cases, make sure you use the lowest T_m primer to work out your annealing temperature.

4. **Check that your primers don't have significant homology with other regions of the template DNA**

5. **Avoid palindromic sequences**
   - You do not want primers or anneal with itself

Many online programs are available to assist in primer design – these help avoid issues mistakes with primer design

Additional DNA sequences can be added with carefully designed primers

- Inclusion of additional sequences at each end that are not present in the piece of DNA being targeted

- ***When would you want to do this?***
  - Most commonly done when including restriction endonuclease cleavage sites to facilitate the cloning of the amplified DNA

Bringing in material discussed in last lecture

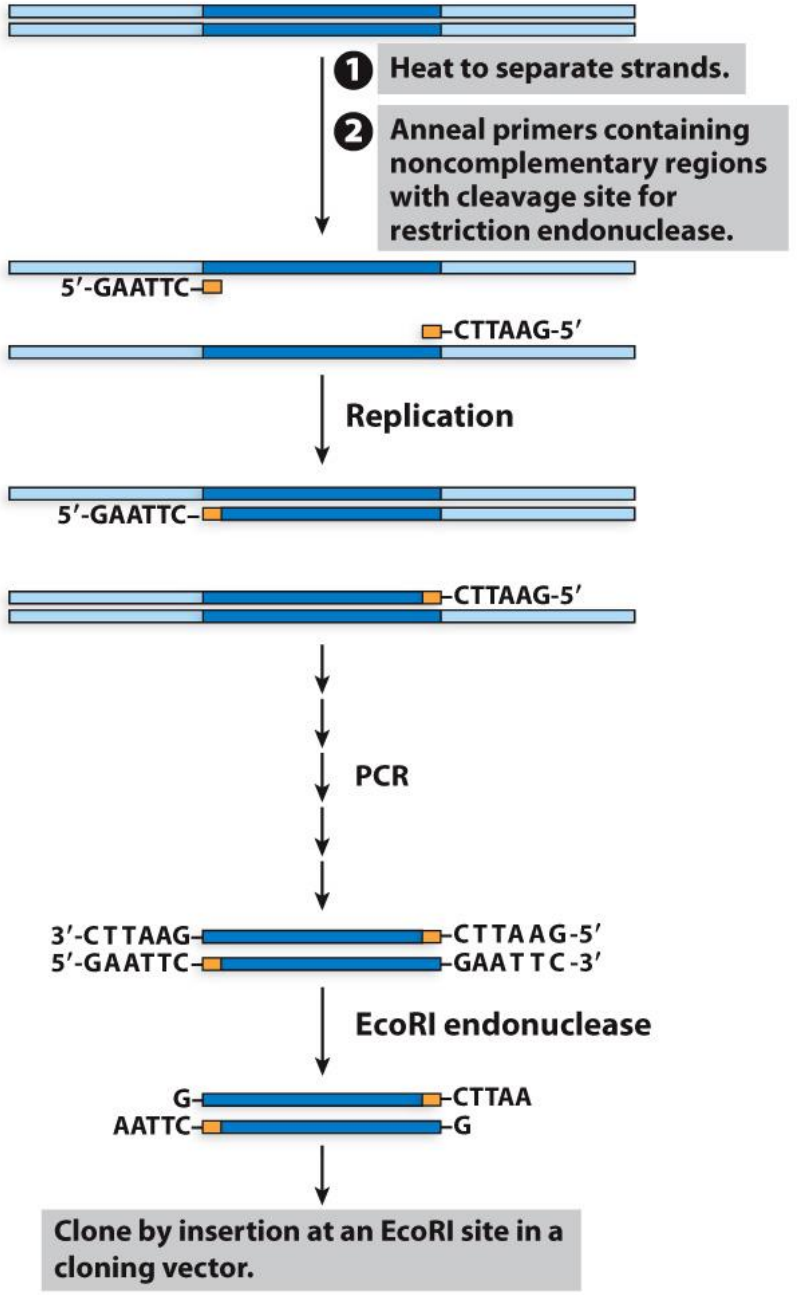
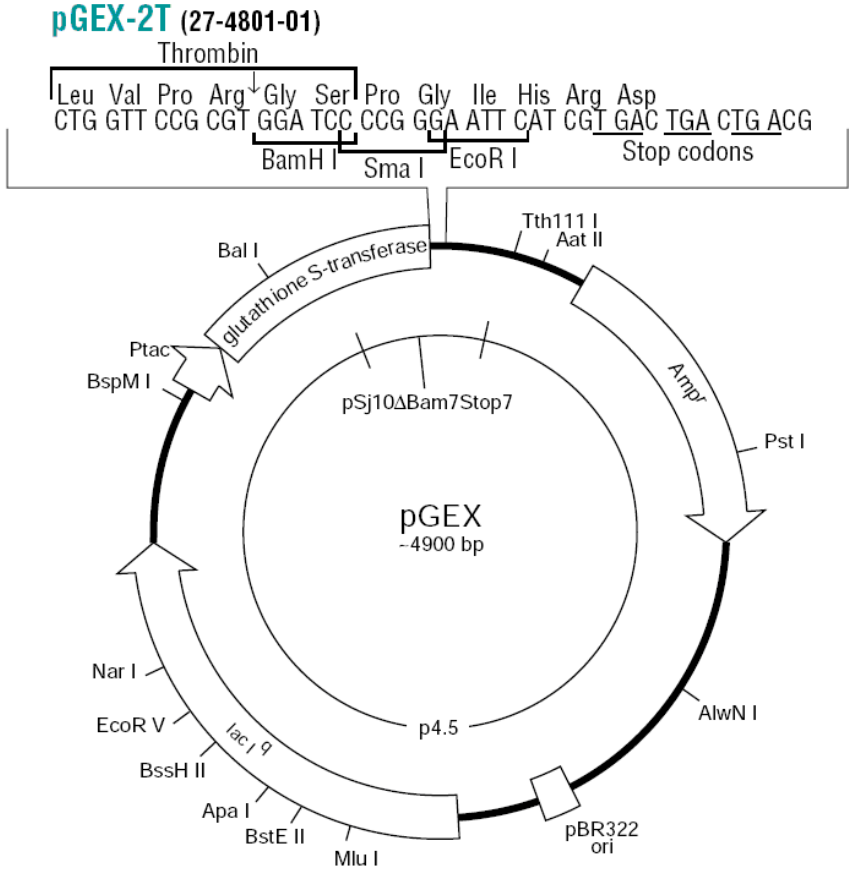


Figure 7-9b
Molecular Biology: Principles and Practice, Second Edition

Additional DNA sequences can be added with carefully designed primers

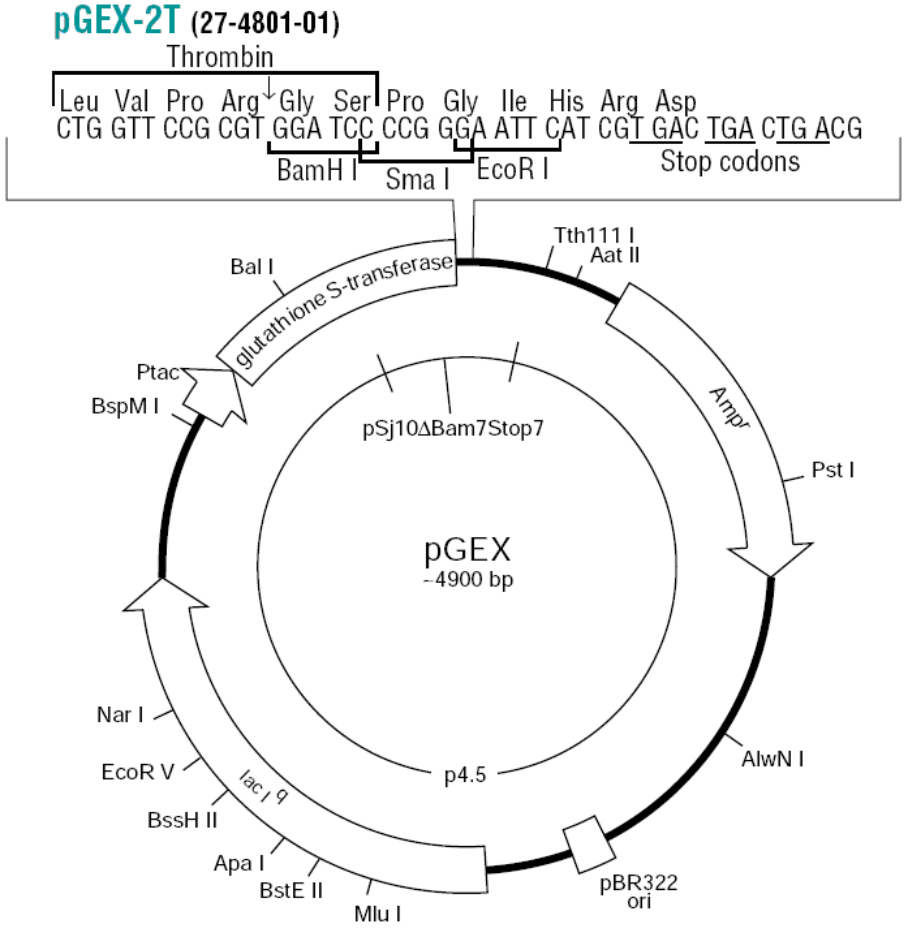


- Below are the 5´ and 3´ ends of the cDNA you want to clone into pGEX-2T
- You want to clone it using BamHI and EcoRI
- Both RE recognition sequences are not found anywhere in your cDNA.
- In order for this to work, you must **add the BamHI** recognition sequence to the 5´ end **and the EcoRI** recognition sequence to the 3´ end of your cDNA **via PCR**

cDNA of interest

5′ ATG GCA ATT TAC GCA GCC AGC …(500bp)… ATA GCA TAC CCT CAC AGC TAA 3′

Designing primers for cloning into vectors



Design PCR primers for cloning the following DNA sequence **in-frame** into the *Bam*HI and *Eco*RI sites in pGEX-2T vector:

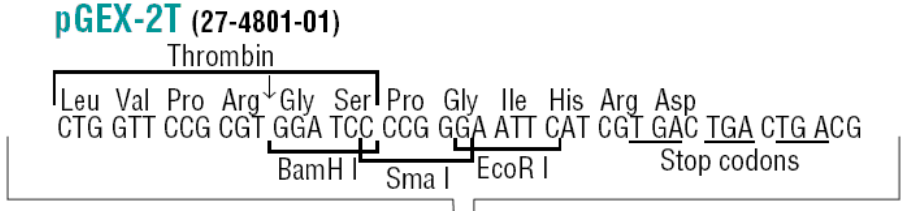
5′ ATG GCA ATT TAC GCA GCC AGC …(500bp)… ATA GCA TAC CCT CAC AGC TAA 3′

Designing primers for cloning into vectors

MAKING THE FORWARD PRIMER

1. If not selected already, choose restriction sites you are going to use to clone DNA (one 5', one 3')
   - DO NOT CHOOSE OVERLAPPING SITES
2. Add a couple of bases at the 5' end as a cap
3. Add the restriction site sequence
4. Add additional bases (1 or 2) **if required** to make an **in-frame** fusion
5. Add 21 base pairs of your sequence
   - The sequence you add is your exact template sequence!

Designing primers for cloning into vectors

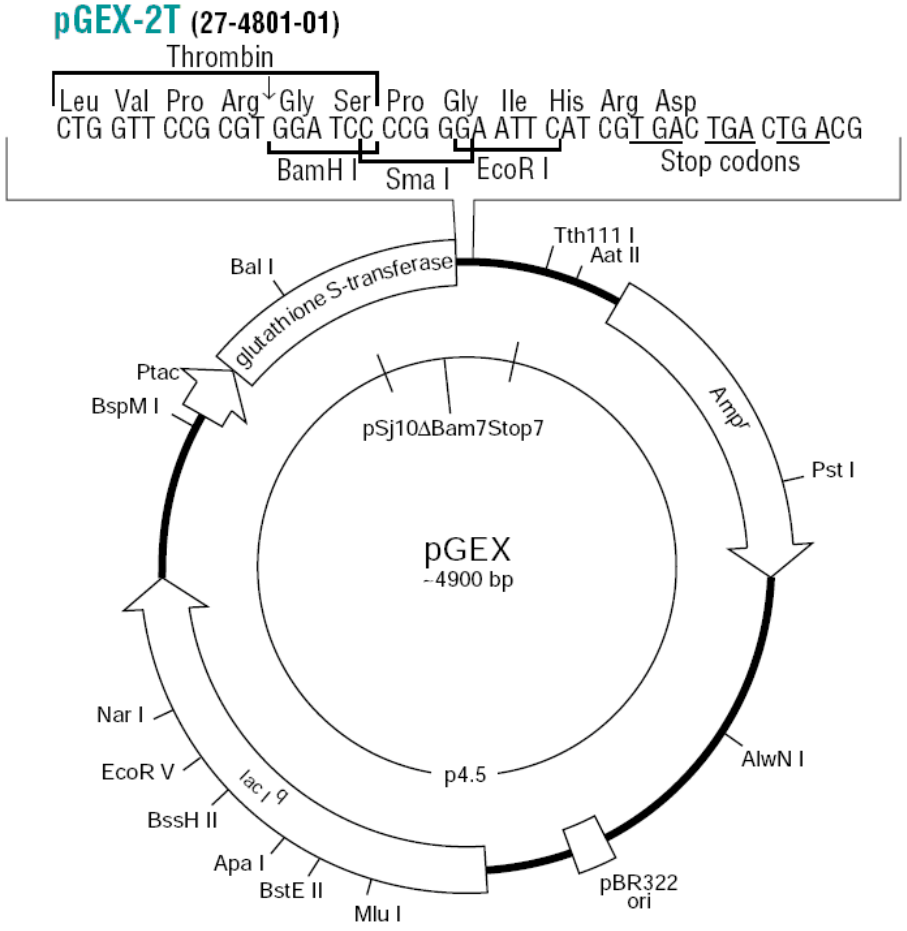


5’ ATG GCA ATT TAC GCA GCC AGC …(500bp)… ATA GCA TAC CCT CAC AGC TAA 3’

FORWARD PRIMER

1. Choose restriction site you are going to use to clone your DNA (one 5’, one 3’)
   - *Bam*HI (GGATCC)
2. Add a couple of bases at the 5’ end as a cap
   - 5’- GC-
3. Add the restriction site sequence
   - 5’-GC-GGA TCC-
4. Add additional bases (1 or 2) if required to make an **in-frame** fusion
   - Not required in this case, but could add 3 if so desire
5. Add 21 base pairs of your sequence
   - 5’-GC-GGA TCC ATG GCA ATT TAC GCA GCC AGC-3’

Designing primers for cloning into vectors



Design PCR primers for cloning the following DNA sequence **in-frame** into the *Bam*HI and *Eco*RI sites in pGEX-2T vector:

5’ ATG GCA ATT TAC GCA GCC AGC …(500bp)… ATA GCA TAC CCT CAC AGC TAA 3’

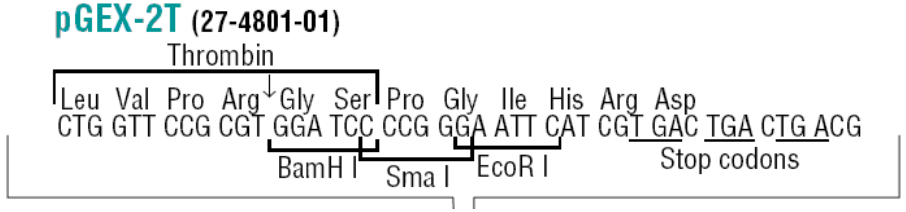
Designing primers for cloning into vectors

MAKING THE REVERSE PRIMER

You are working with the complementary strand, so if it helps write out the complementary sequence for your DNA sequence of interest, then write it in reverse

1. Add a couple of bases at the 5' end as a cap
2. Add the restriction site sequence
3. Add 21 base pairs of your sequence
   - This will be the reverse complement of your template sequence
   - No need to put reverse primer *in-frame* as transcription starts in the other direction

Designing primers for cloning into vectors



5’ ATG GCA ATT TAC GCA GCC AGC …(500bp)… ATA GCA TAC CCT CAC AGC TAA 3’

3’ …(500bp)… TAT CGT ATG GGA GTG TCG ATT 5’

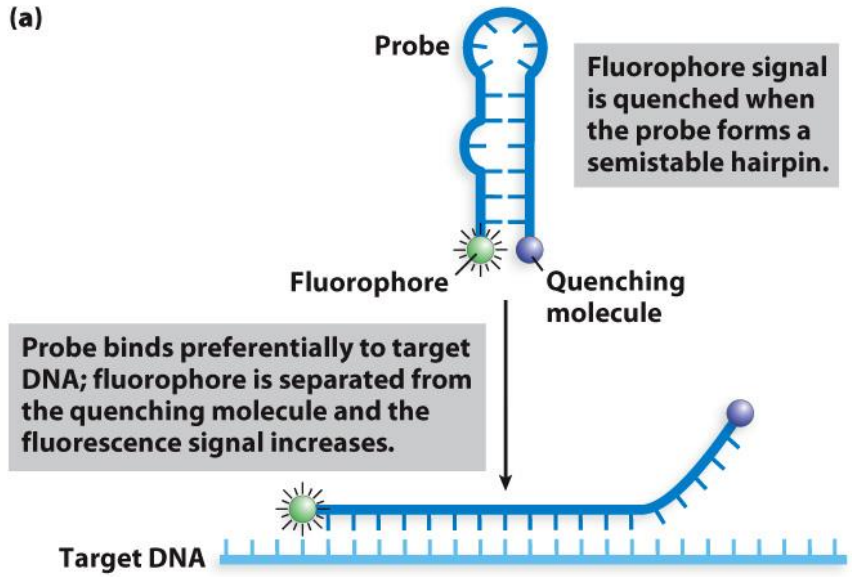
5’-TTA GCT GTG AGG GTA TGC TAT-3’

REVERSE PRIMER

1. Add a couple of bases at the 5’ end as a cap
   - 5’-GC-
2. Add the restriction site sequence
   - 5’-GC GAA TTC-
3. Add 21 base pairs of your sequence
   - This will be the reverse complement of your template sequence
   - 5’-GC GAA TTC TTA GCT GTG AGG GTA TGC TAT-3’

Quantitative PCR (qPCR)

- Also known as real-time PCR
- Used to quantify your gene in real time by monitoring the amplification of a target DNA molecule during PCR
- PCR is carried out in the presence of a probe that emits a fluorescent signal when the PCR product is present
- If the sequence of interest is present at higher levels than the other sequences in the sample, the PCR signal will reach a predetermined threshold faster
- qPCR can be combined to determine relative transcription levels of genes under specific conditions



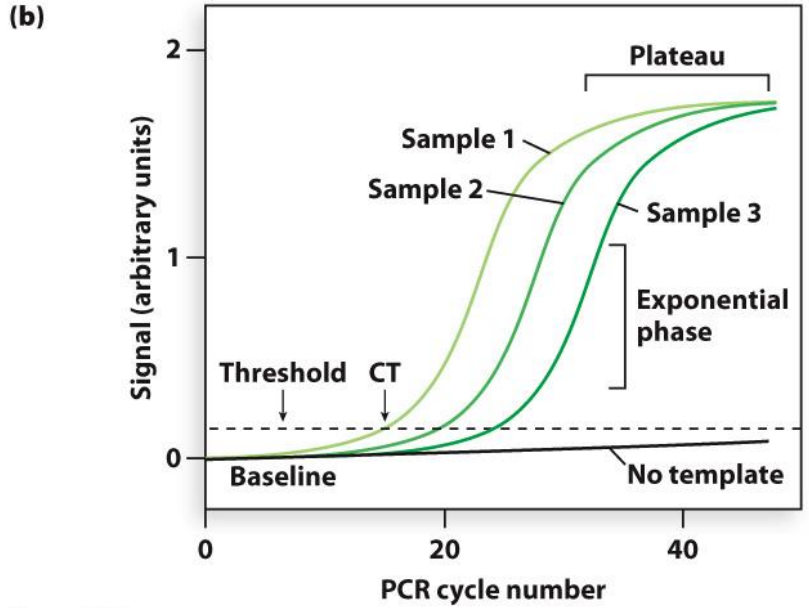
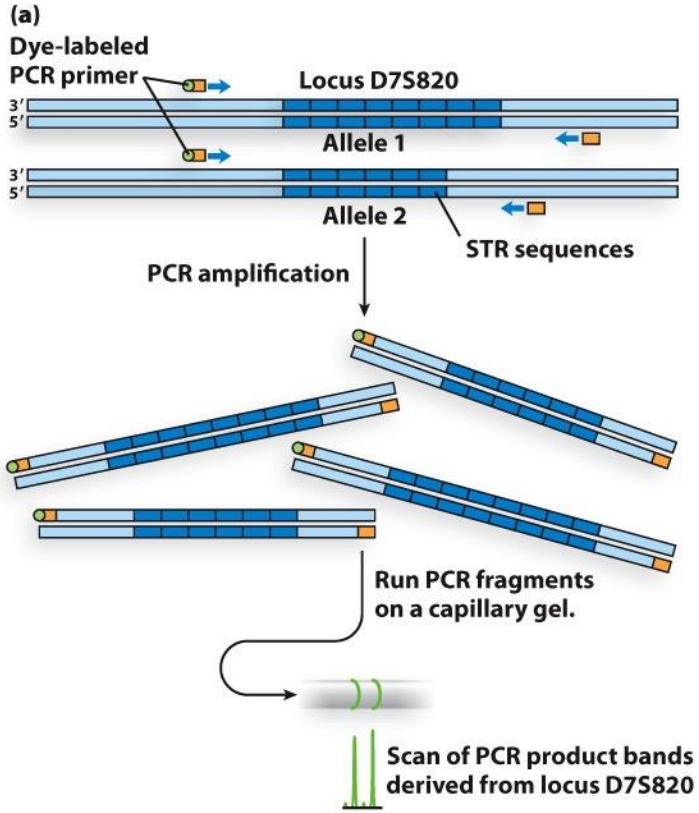


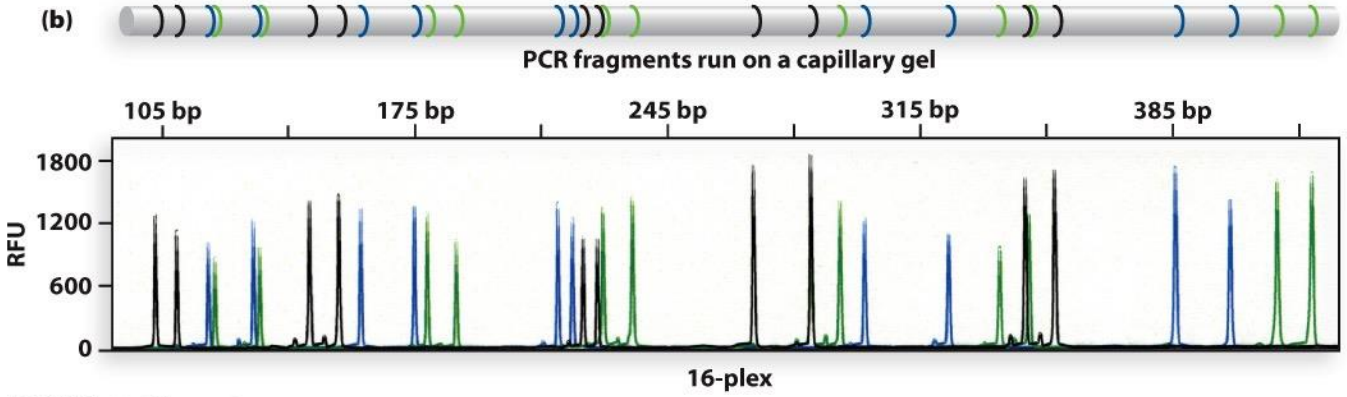
Figure 7-10
Molecular Biology: Principles and Practice, Second Edition

Some PCR uses (among MANY others)

- DNA fingerprinting
- Amplifying DNA and genes that correspond to purified proteins
- Molecular paleontology: our and other species genetics history (textbook pg. 280-290)
- RT-PCR (PCR Covid Tests)
- DNA sequencing
- Gene fusions

DNA fingerprinting gets an update with PCR

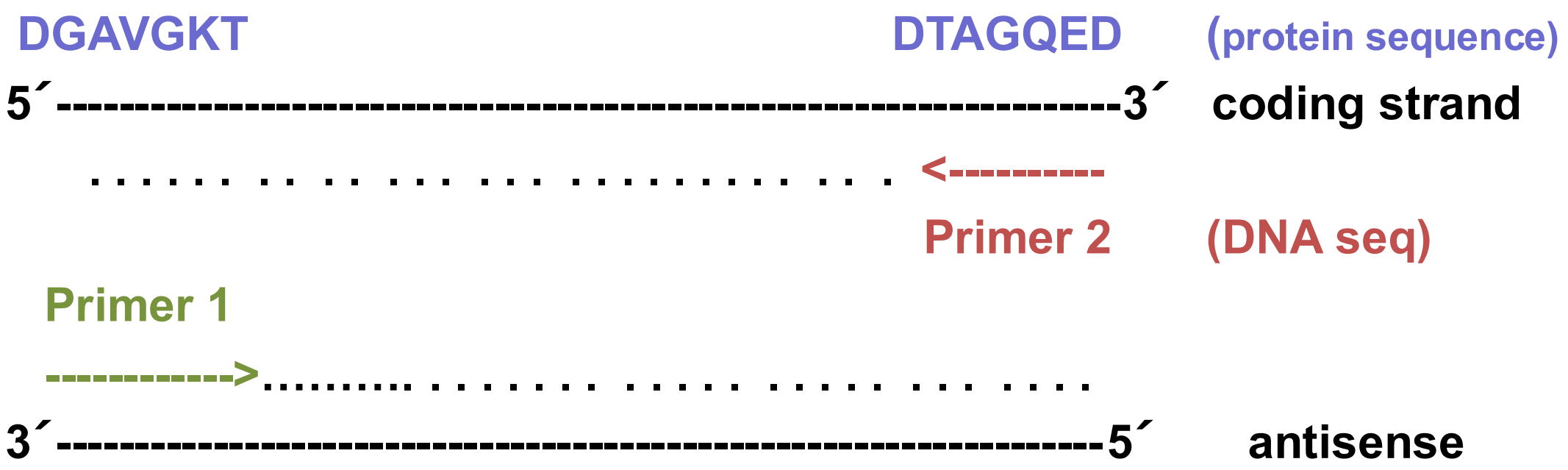




Highlight 7-1 Figure 1
Molecular Biology: Principles and Practice, Second Edition

- Forensic work now focuses on differences in the lengths of **short tandem repeat (STR)** sequences (as opposed to the differences in the restriction fragment length)
- STRs are usually 4bp in length
- STRs are repeated many times in tandem at a specific location (locus) in a chromosome
- More than 20 000 tetranucleotide STR loci have been characterized in the human genome
- The length of a particular STR in an individual can be determined with PCR
  - The sequence surrounding/flanking an STR are unique to each type of STR and are identical in all humans
  - PCR primers targeted to this flanking DNA are designed to amplify the DNA across the STR

Amplifying DNA and genes that correspond to purified proteins



- **primer 1** builds coding strand by annealing to the antisense strand and has a sequence corresponding to the codons for DGAVGKT amino acids (conserved at N-terminal end of consensus protein)
- **primer 2** builds antisense strand by annealing to the coding strand and has sequence corresponding to the reverse complement of the codons for DTAGQED amino acids (conserved at C-terminal end of consensus protein)

Amplifying DNA and genes that correspond to purified proteins

CONSERVED N-TERMINAL PROTEIN AMINO ACIDS

D G A V G K T---

5´ GA(T/C) GG(T/C/A/G) GC(N) GT(N) GG(N) AA(A/G) AC(N) 3´

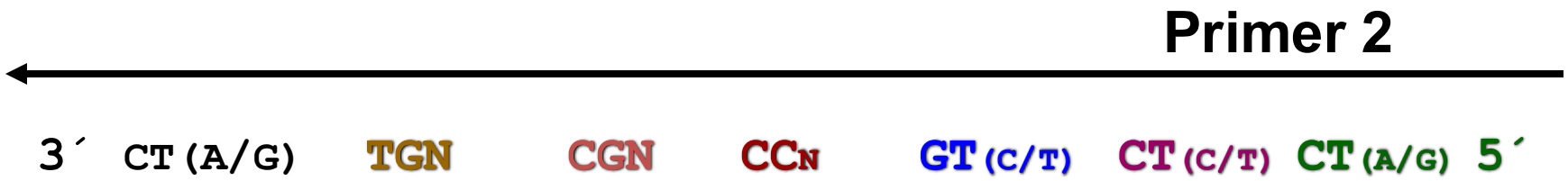
Primer 1

primer degeneracies

First position (5'-end)	Second position U	Second position C	Second position A	Second position G	Third position (3'-end)
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA STOP	UGA STOP	A
	UUG Leu	UCG Ser	UAG STOP	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

CONSERVED C-TERMINAL AMINO ACIDS FROM CONSENSUS PROTEIN

to get Primer 2 sequence take complement and reverse:

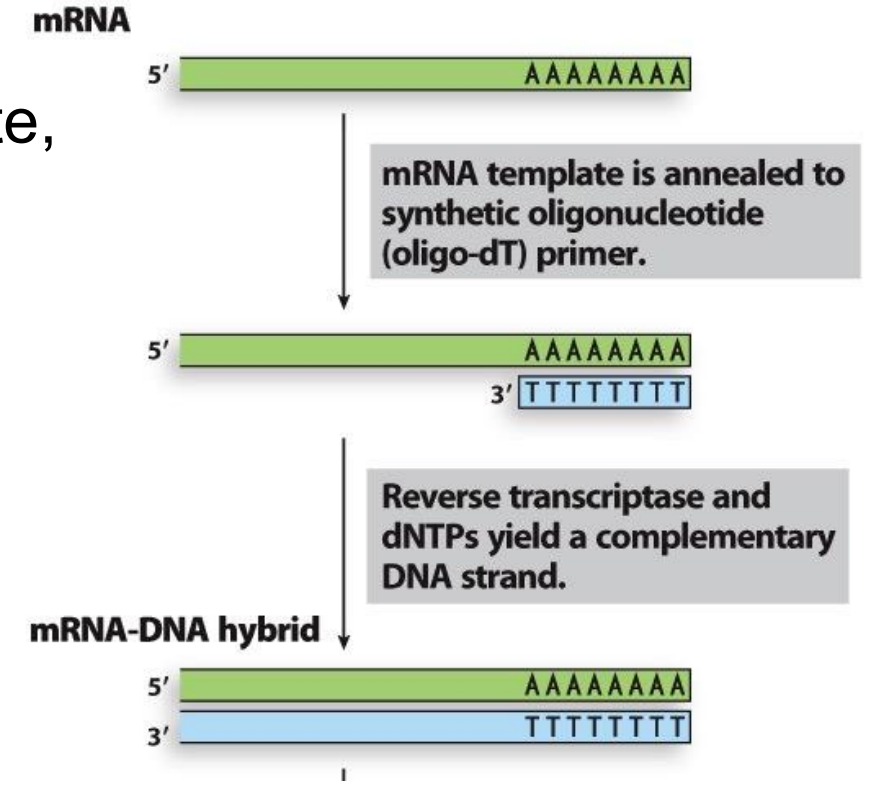


Primer sequence as written for ordering from supplier:

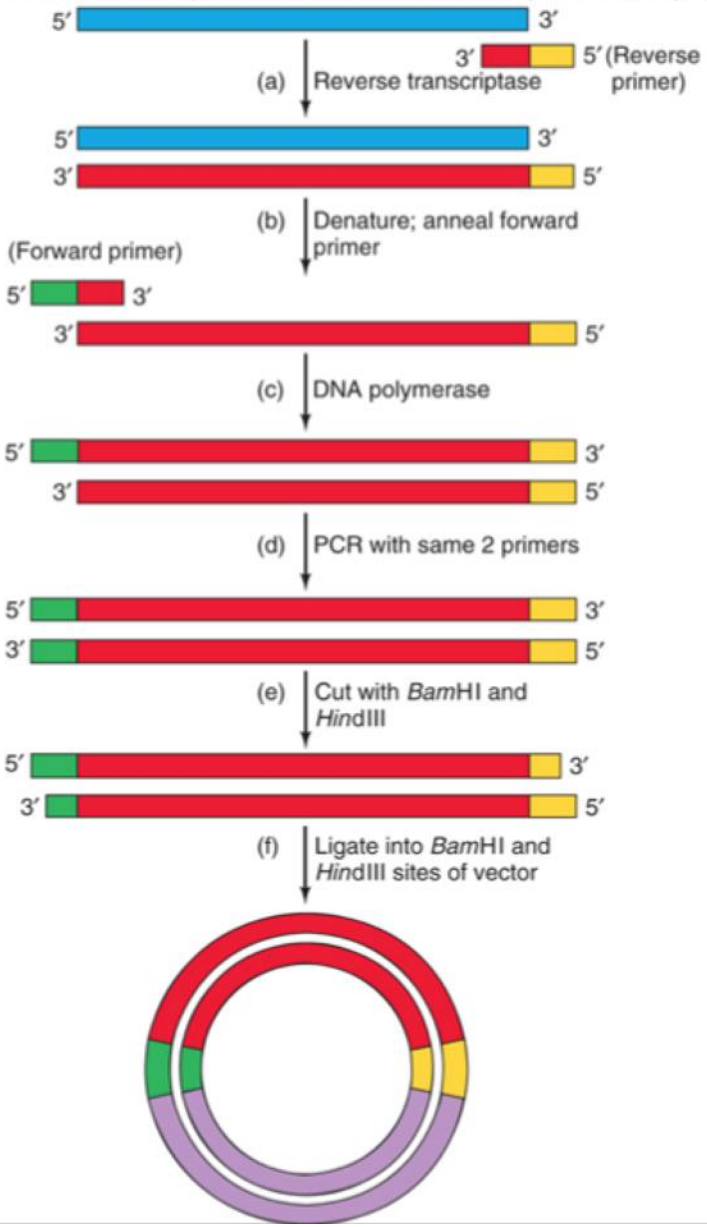
5´ (G/A)TC (T/C)TC (T/C)TG NCC NGC NGT (A/G)TC 3´

Reverse transcriptase PCR (RT-PCR) can be used to detect RNA sequences derived from living cells

- Sequences in RNA can be amplified by PCR if reverse transcriptase is used in the first PCR cycle
- After the DNA strand is made using RNA as a template, the remaining cycles can be carried out with a thermostable DNA pol
- allows the detection and quantification of mRNA
- very sensitive method that shows whether or not a specific gene is being expressed in a given sample
- Reverse transcriptase + PCR used to detect *small* amounts of mRNA
- Sequence of the mRNA must be known



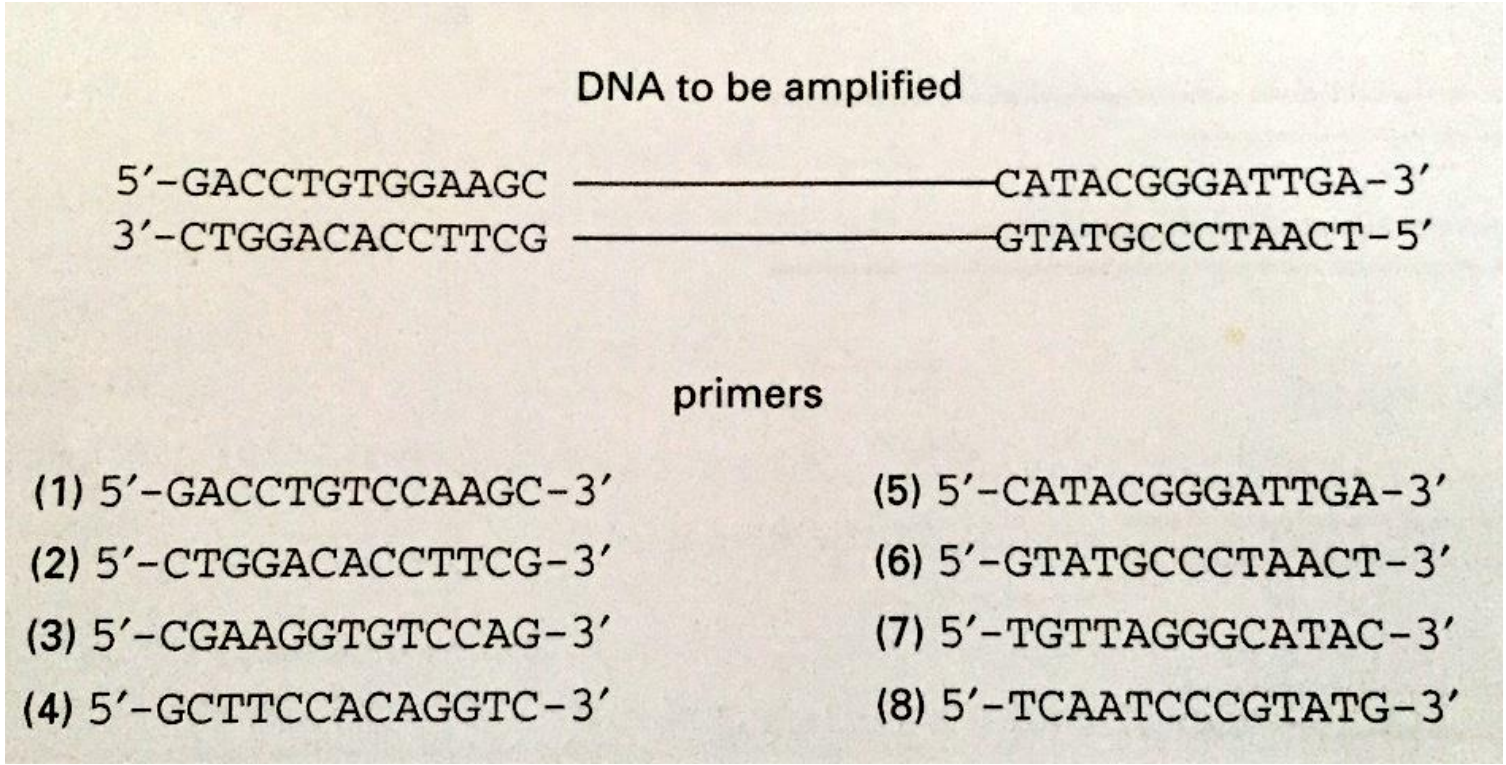
RT-PCR (in cDNA cloning)



- Reverse transcriptase + PCR used to detect ***small* amounts of mRNA**
- Use **a reverse primer** with *Hind*III site at its 5´ end and reverse transcriptase
- Denature the **mRNA-cDNA** hybrid and anneal to **forward primer** with a *Bam*HI site at its 5´ end
- The **forward primer** initiates second-strand **cDNA** synthesis with DNA polymerase
- Continue PCR with the same two primers to amplify the dsDNA
- Cut the cDNA with ***Bam*HI** and ***Hind*III** sites and ligate into **suitable vector** with complementary overhangs

Practice problem (solve on your own)

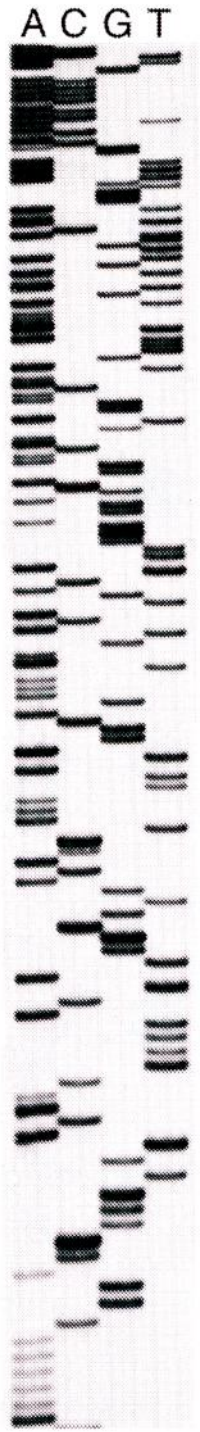
You want to amplify the DNA between the two stretches of sequence shown in Figure below. Of the listed primers choose the pair that will allow you to amplify the DNA by PCR.



DNA SEQUENCING

5’ ATGTAGCTAGTACGGCGCTATTTATGCTAGAGATCAGCGGC 3’

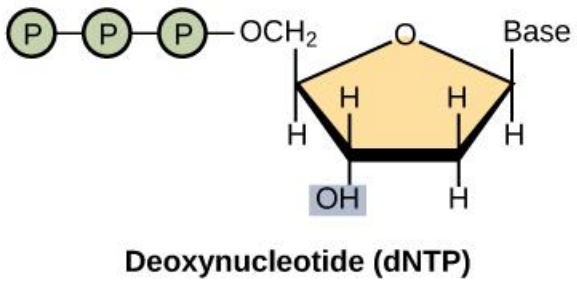
How would you sequence this DNA?



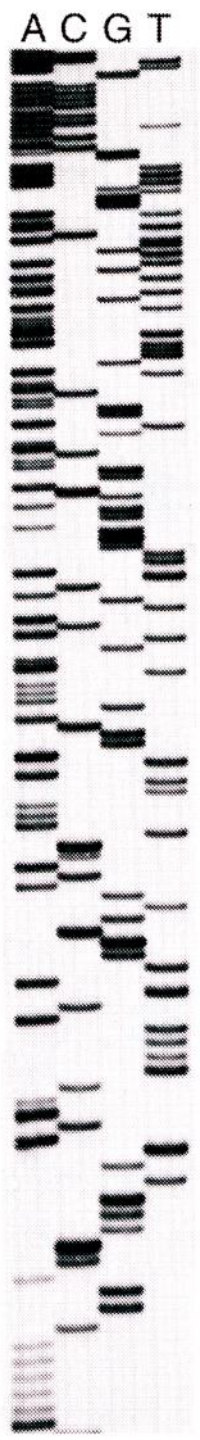
DNA SEQUENCING

5’ ATGTAGCTAGTACGGCGCTATTTATGCTAGAGATCAGCGGC 3’

We need both!



Deoxynucleotide (dNTP)



SANGER CHAIN TERMINATION (review)

Dideoxy chain termination method

What's in the reaction?

Template DNA

primer

Dideoxy-nucleotide **(chain terminators)**

dNTPs

DNA polymerase

all required co-factors (e.g., Mg^{2+}) in buffer

ddCTP stops replication at "C"

ddATP stops at "A"s

ddGTP stops at "G"s

ddTTP stops at "T"s

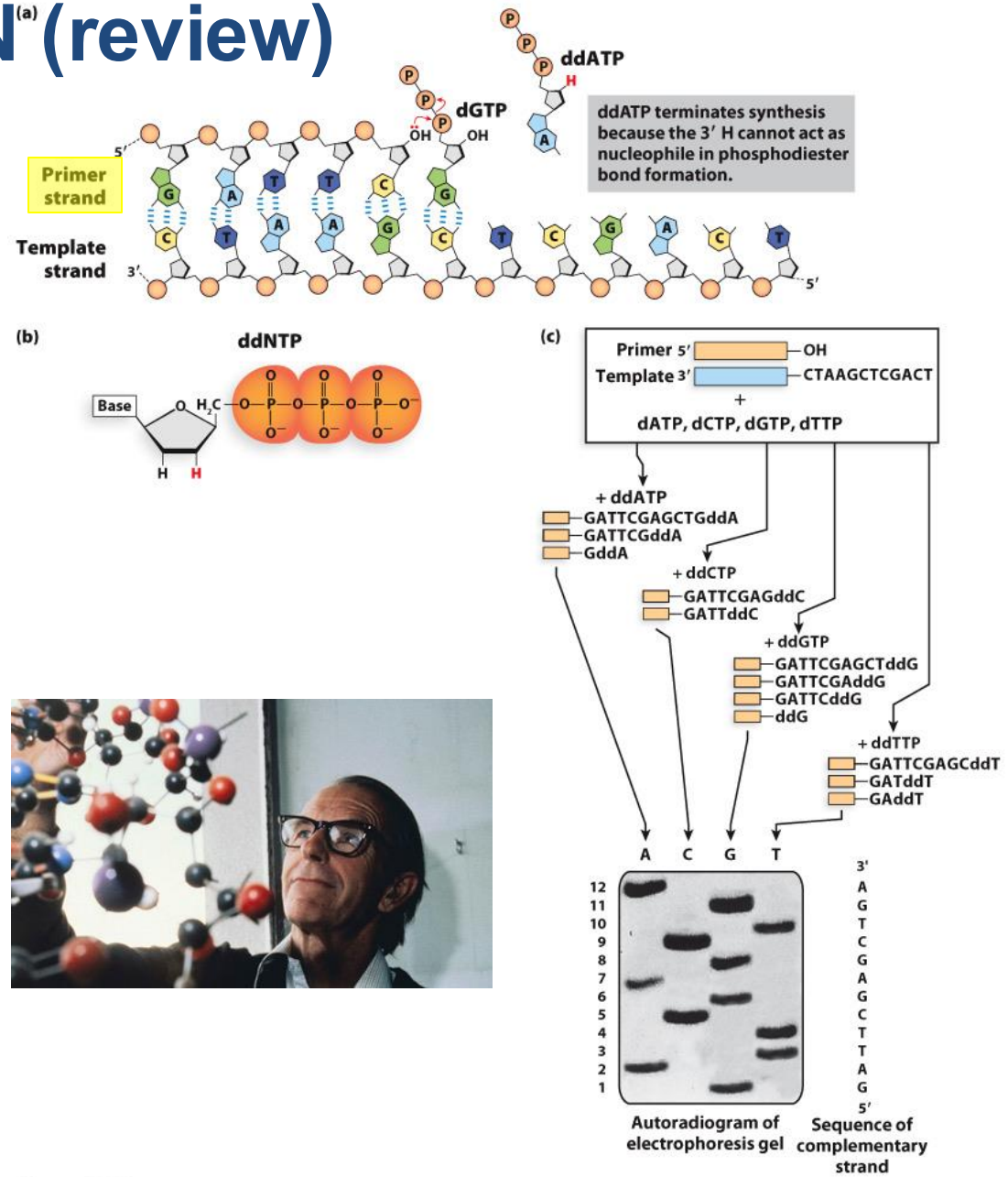


Figure 7-11

Molecular Biology: Principles and Practice, Second Edition

AUTOMATED DNA SEQUENCING

- DNA sequencing was first automated by a variation of the Sanger method
- Each of the four dideoxynucleotides used for the reaction was labeled with a differently coloured fluorescent tag
- Thousands of nucleotides sequences in a few hours
- Many genome projects were sequenced in this way (including the human genome project)
- Most research labs are still using this for sequencing genes of interest (0.5 – 1kb)

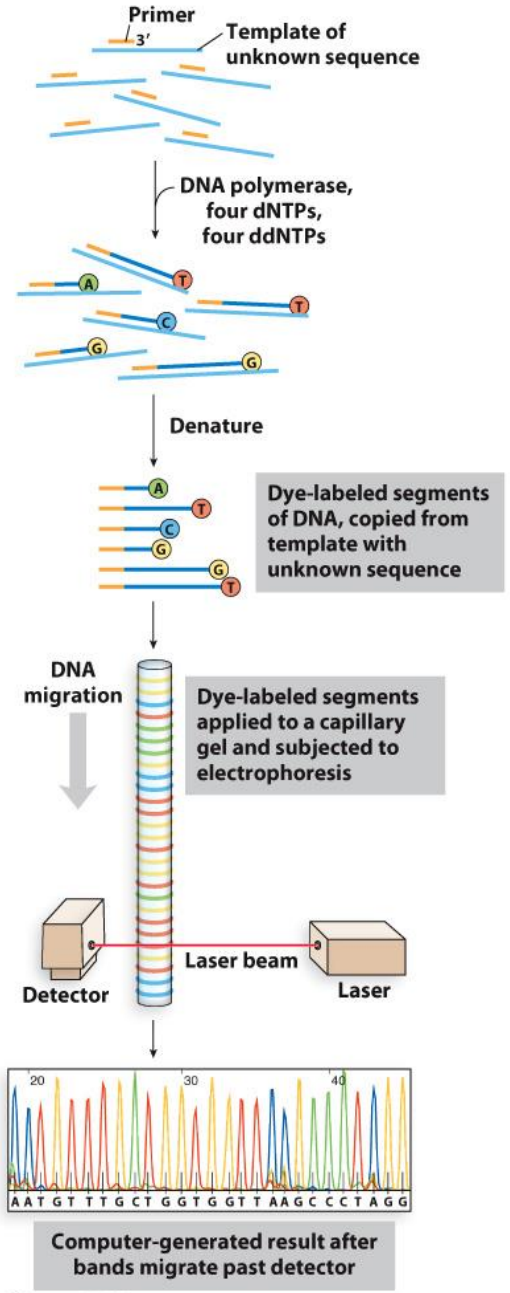
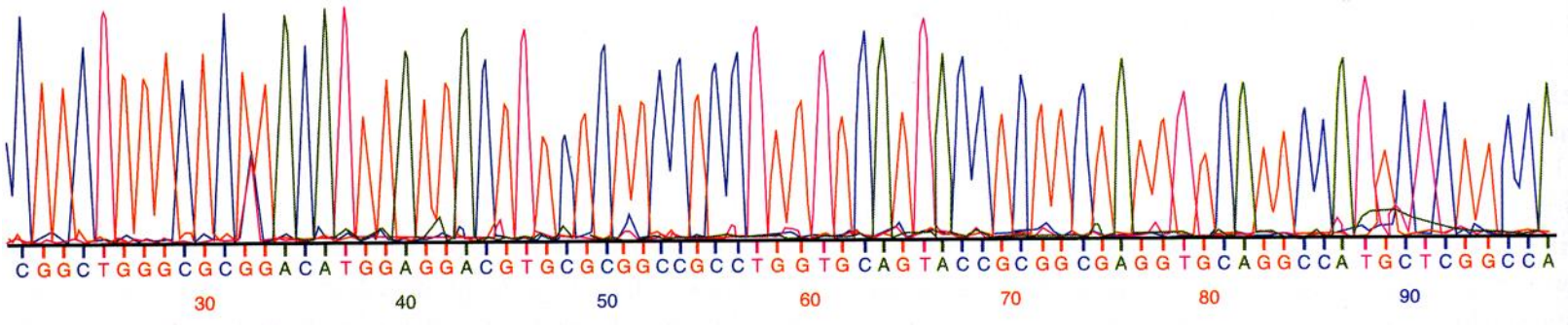
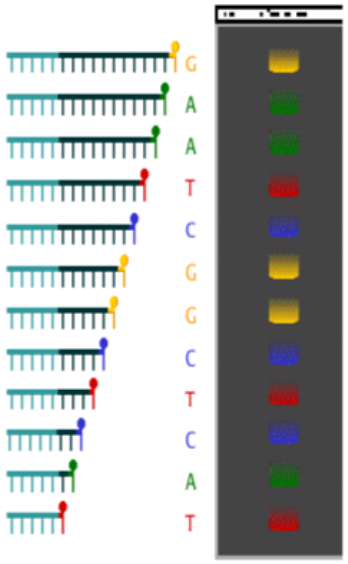
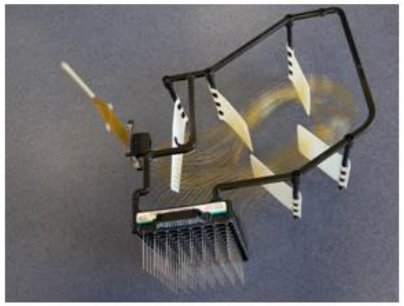


Figure 7-12
Molecular Biology: Principles and Practice, Second Edition



Sanger Sequencing

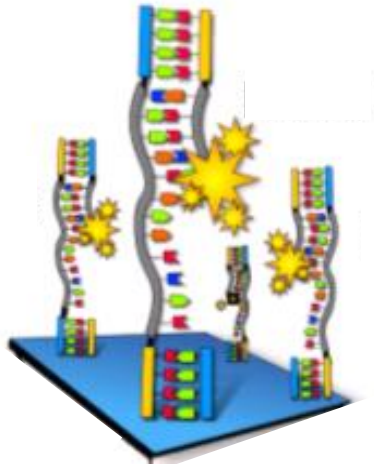


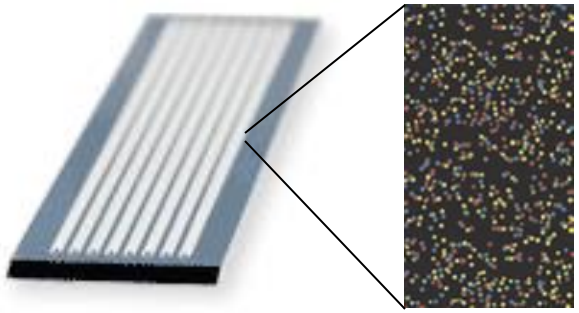


• Sequencing reaction products generated as a batch, then separated by capillary electrophoresis to read sequence

•One tube → one read

Flow Cell Sequencing



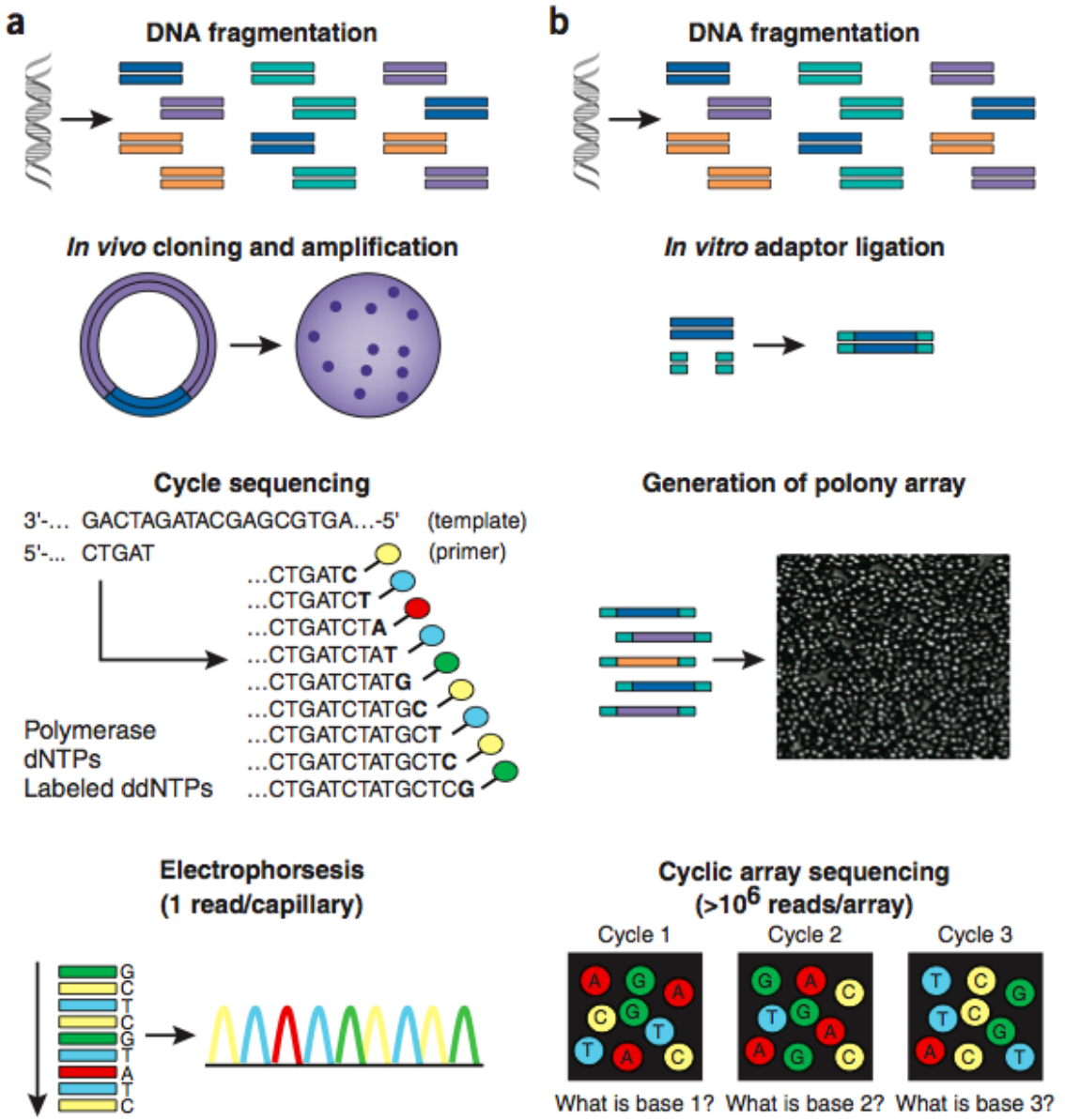


• A separate sequencing reaction for each nucleotide added. Sequence is read as it is "synthesized"

•One tube → millions of reads

PCR-based next generation DNA sequencing

Can you see a connection here with last lecture?



Next-generation DNA sequencing

PERSPECTIVE RESEARCH

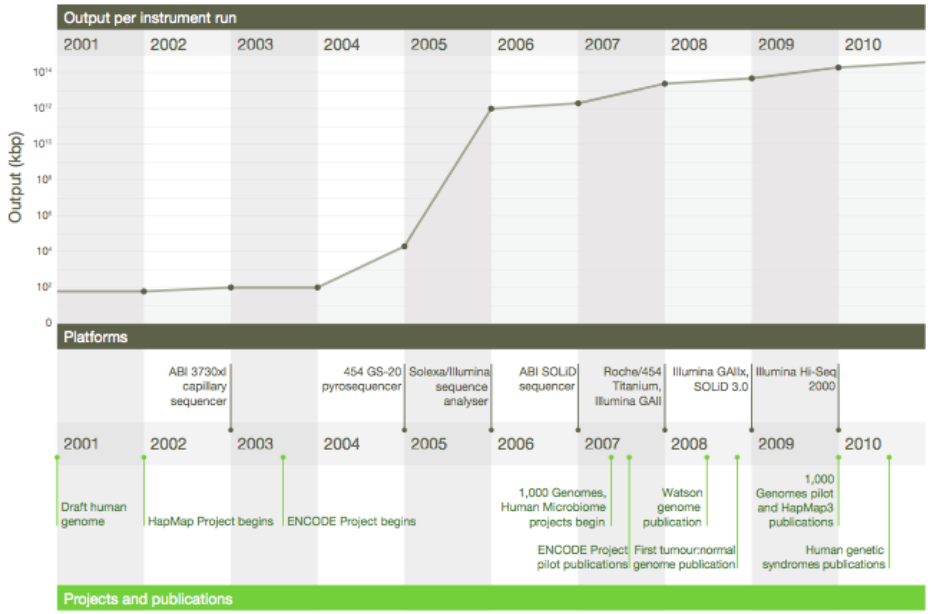
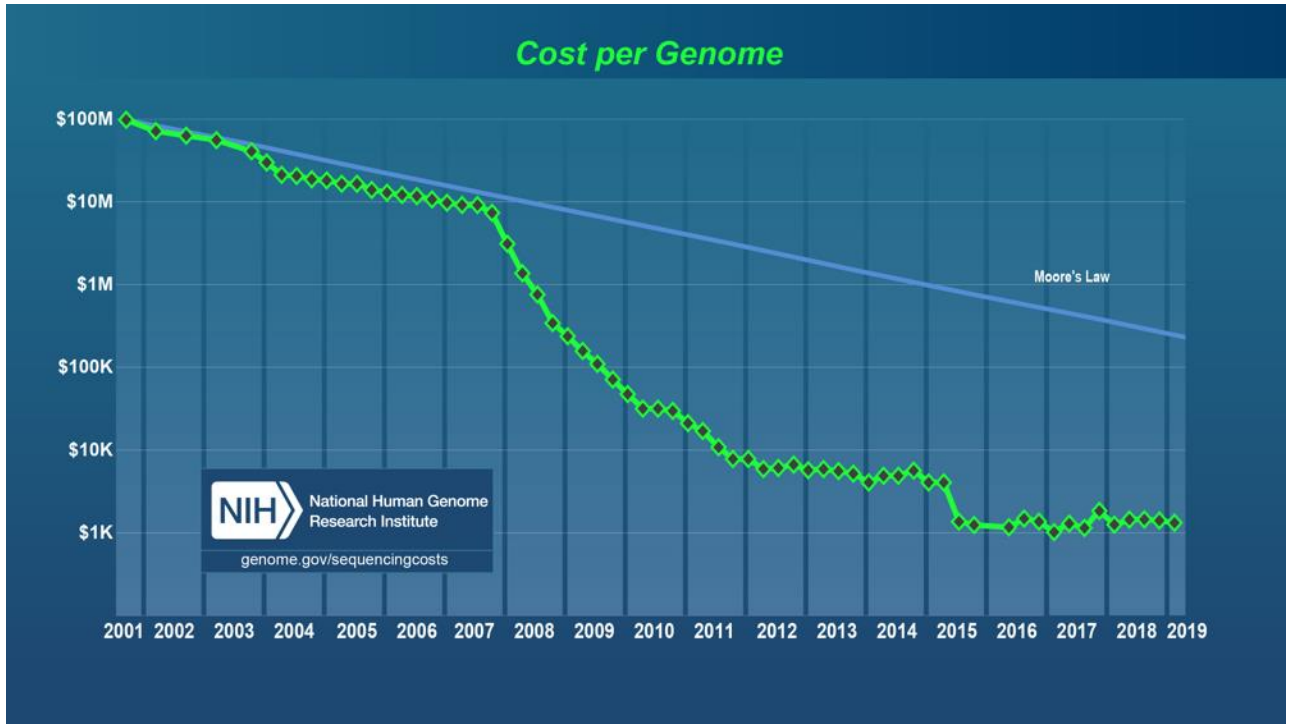
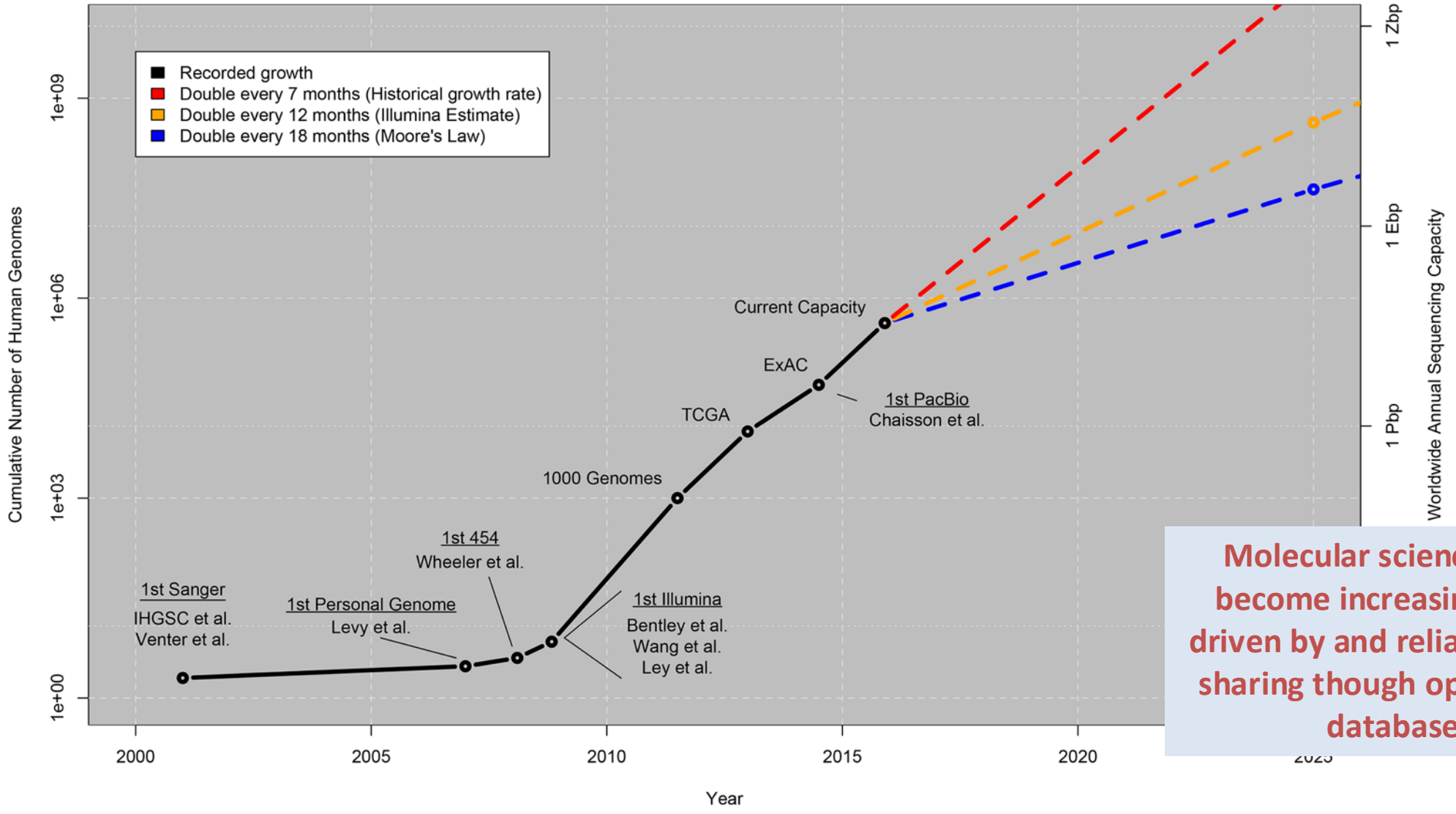


Figure 1 | Changes in instrument capacity over the past decade, and the timing of major sequencing projects. Top, increasing scale of data output per run plotted on a logarithmic scale. Middle, timeline representing major milestones in massively parallel sequencing platform introduction and instrument revisions. Bottom, the timing of several projects and milestones described in the text.



Hundreds of thousands to hundreds of millions of reactions imaged per instrument run = "massively parallel sequencing"

Growth of DNA Sequencing



Molecular sciences have become increasingly data driven by and reliant on data sharing though open-access databases

Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, et al. (2015) Big Data: Astronomical or Genomical?. PLOS Biology 13(7): e1002195. https://doi.org/10.1371/journal.pbio.1002195
http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002195

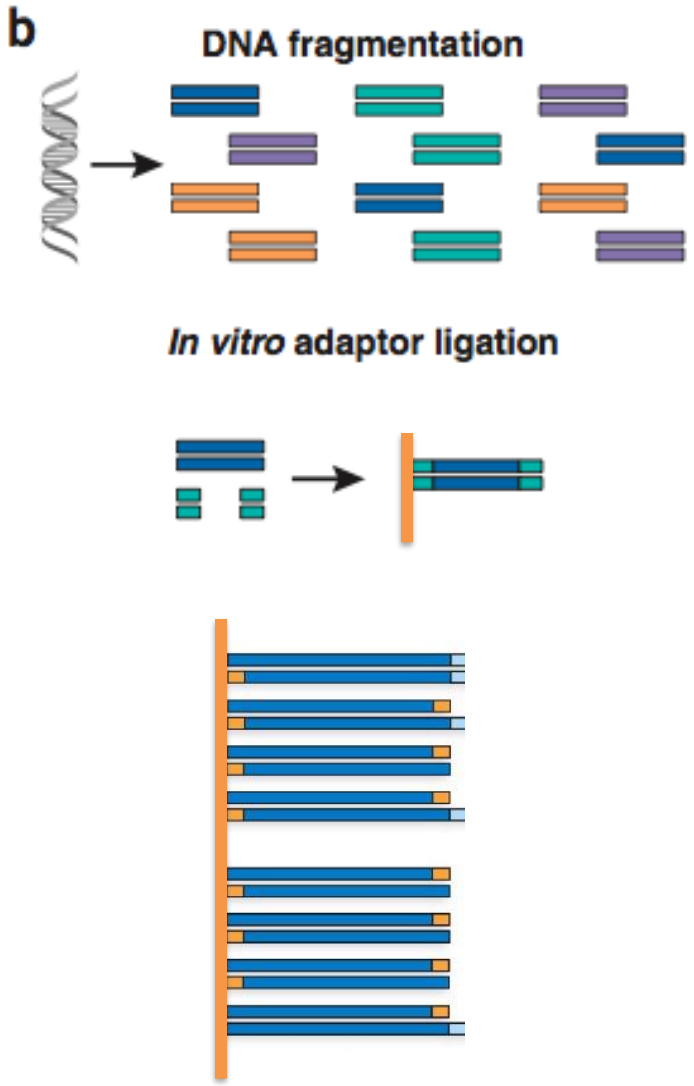
PLOS | BIOLOGY

Next-generation DNA sequencing instruments

Most commercially available sequencers have the following shared features:

- **Library preparation**
- **Cluster amplification**
- **Sequencing**
- **Alignment & Data analysis**

PCR-based DNA SEQUENCING

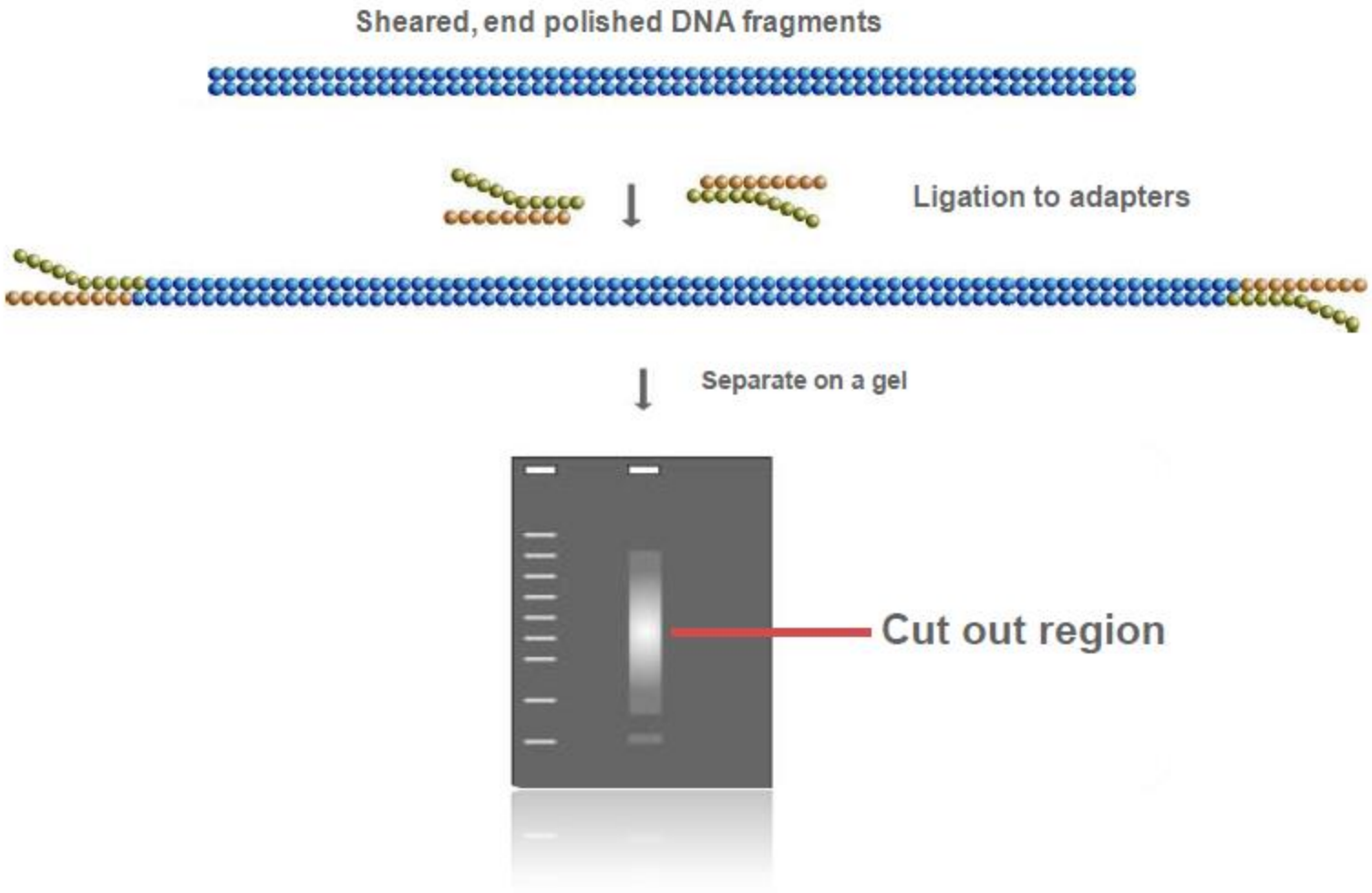


Advances made in DNA sequencing technologies allow us to obtain a person's genome in a day or two.

Set up:

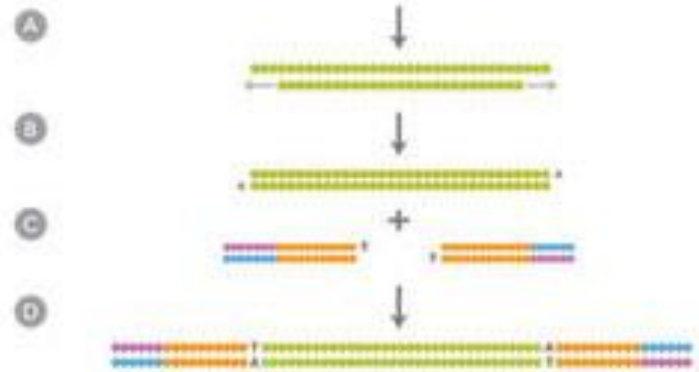
1. Genomic DNA is broken at random locations (a few hundred bp long)
2. Synthetic oligonucleotides are ligated to the ends of all fragments. This provides a reference point (and a starting point) for every DNA molecule
3. Fragments are immobilized on a solid surface and amplified by PCR

Genomic DNA Library Prep



1 LIBRARY PREPARATION

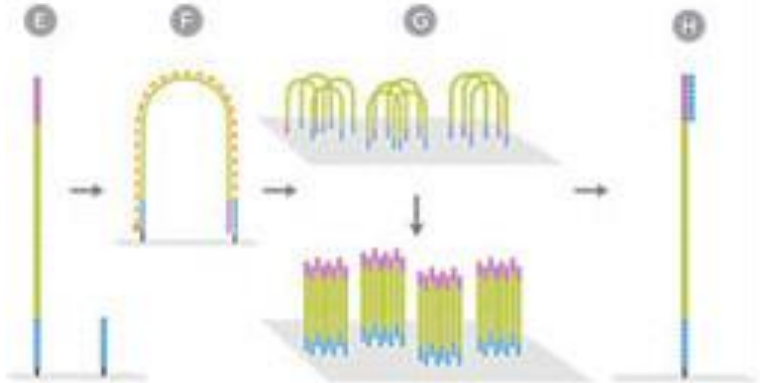
6 hours
3 hours hands-on time



- A Fragment DNA
- B Repair ends
  Add A overhang
- C Ligate adapters
- D Select ligated DNA

2 CLUSTER GENERATION

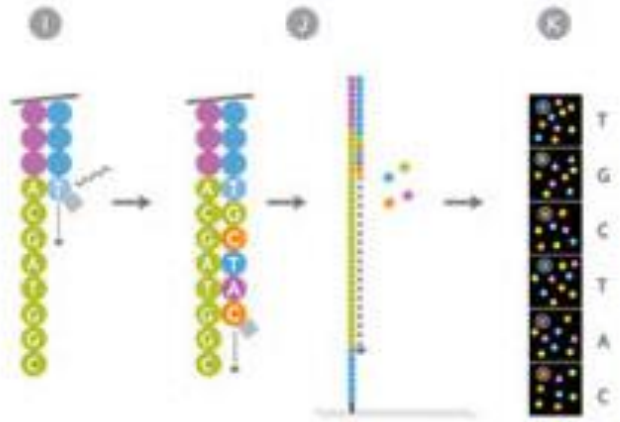
4 hours
30 minutes hands-on time
1–96 samples



- E Attach DNA to flow cell
- F Perform bridge amplification
- G Generate clusters
- H Anneal sequencing primer

3 SEQUENCING

1–3 days single-read run
3–7 days paired-end run
30 minutes hands-on time
1–96 samples



- I Extend first base, read, and deblock
- J Repeat step above to extend strand
- K Generate base calls

Illumina sequencing

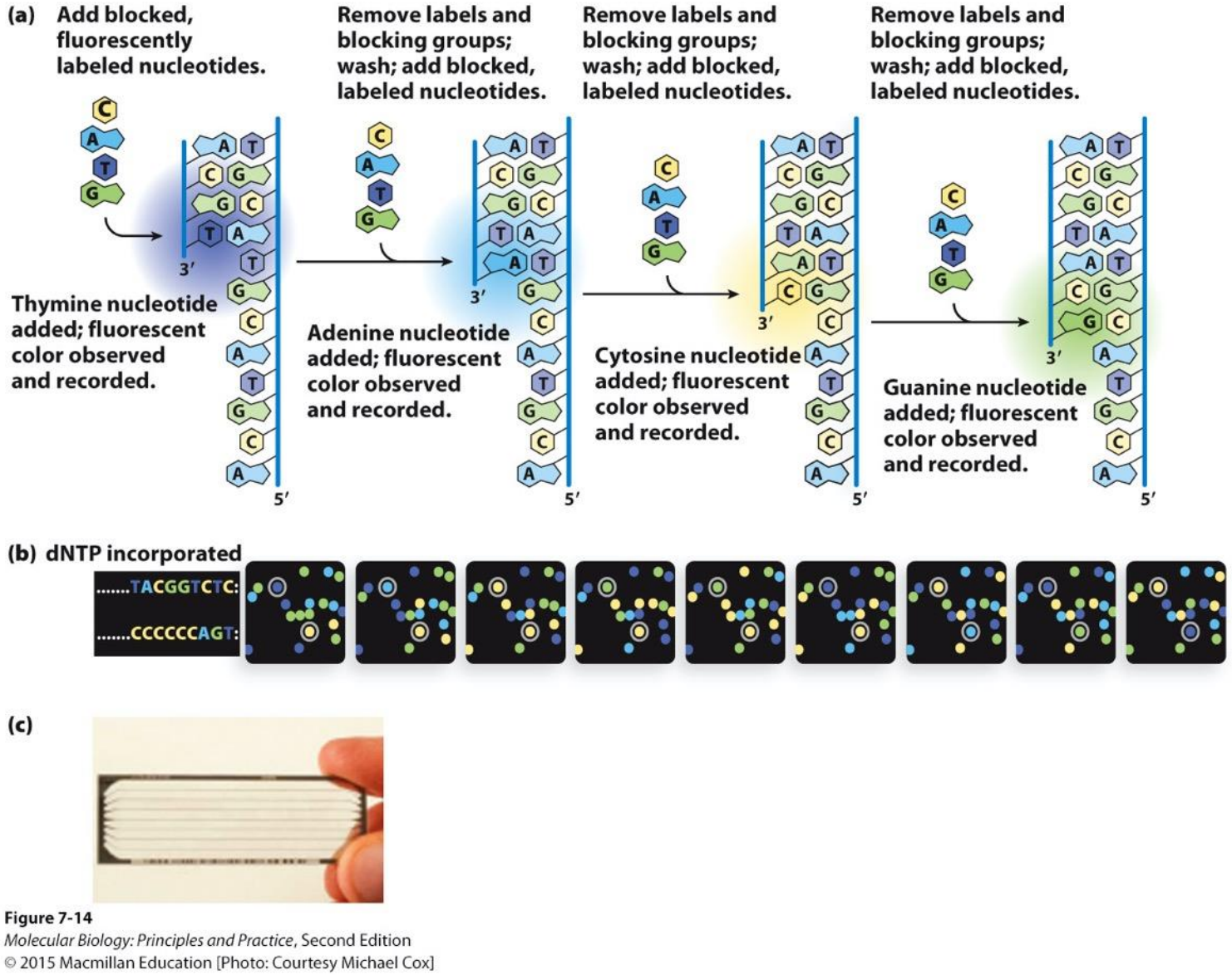


Figure 7-14
Molecular Biology: Principles and Practice, Second Edition

- Special sequencing primer is added that is complementary to the *in vitro* adapters added to each fragment previously
- Fluorescently labeled reversible terminator nucleotides and DNA polymerase are added
- DNA polymerase adds the appropriate nucleotide, each carrying a different fluorescent label
- The terminator nucleotides have blocking groups attached to the 3´ ends that act as temporary chain terminators
- Lasers excite the fluorescent labels, and an image of the entire surface reveals the colour (base) added to each cluster
- Fluorescent label and blocking group are removed in preparation for the next nucleotide to be added

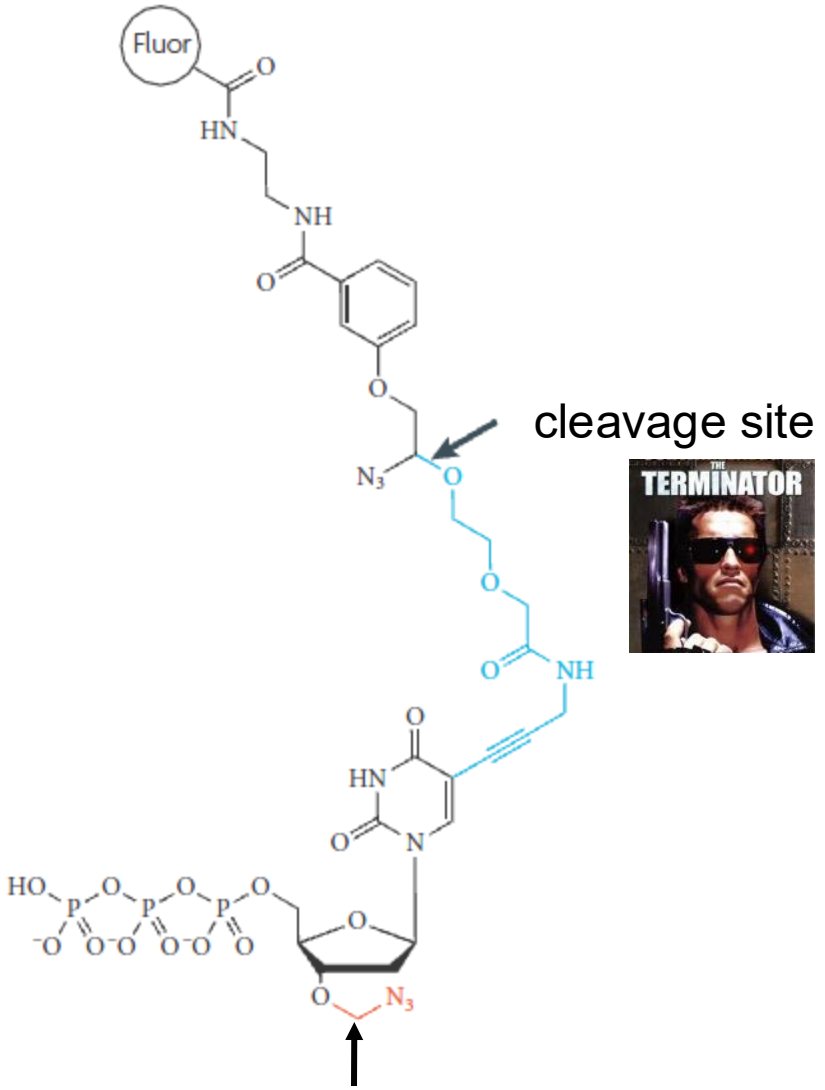
Reversible Terminators

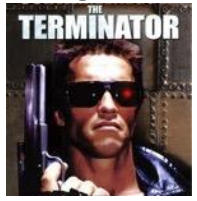
- ***Sequencing by synthesis*** **relies on reversible fluorescent terminators**

- **Unlike the dideoxy terminators used in Sanger sequencing, these terminators have a 3' azide group as the terminator**

- **They also have an azide group linking the nucleotide to the fluorophore**

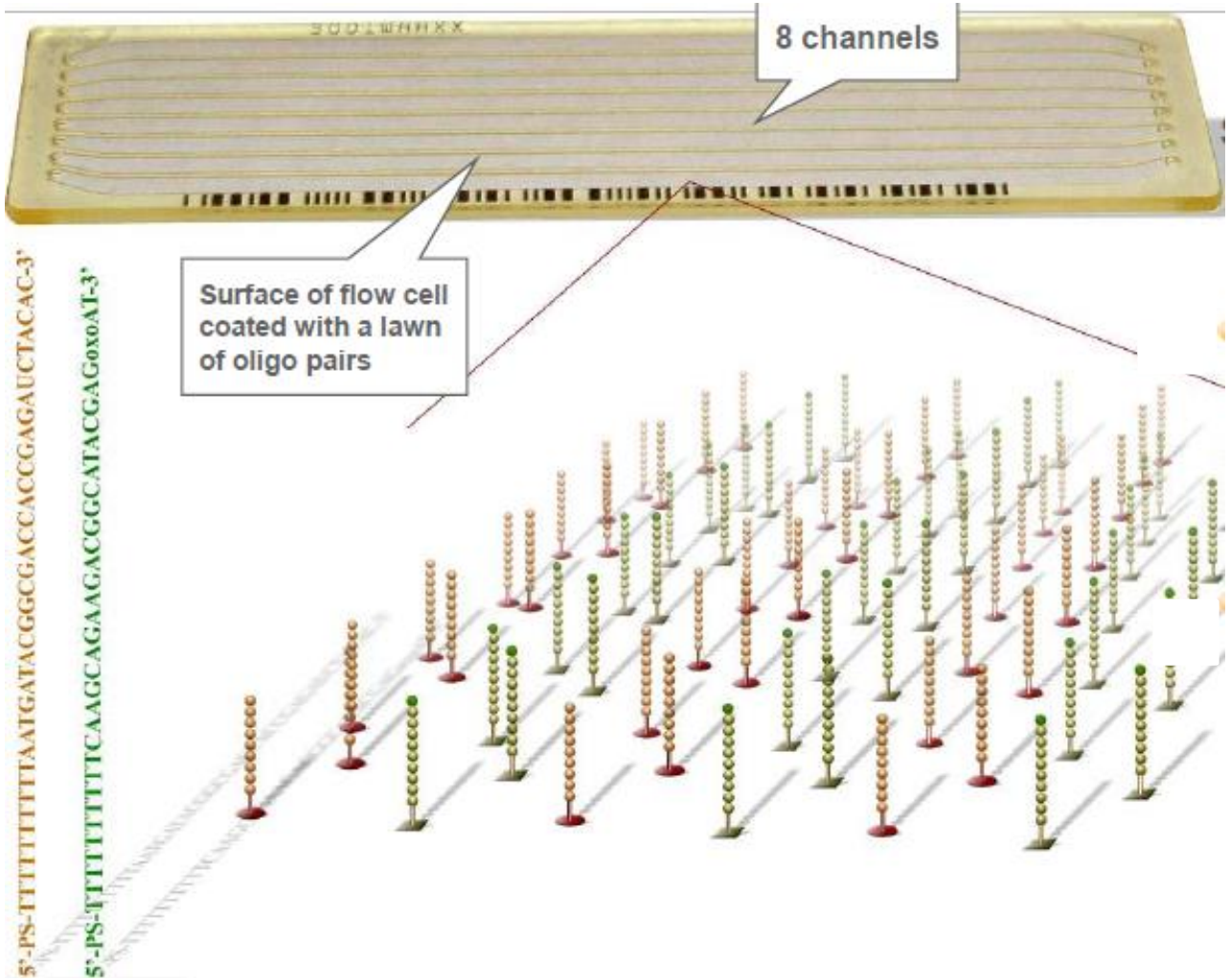
- **These azide groups can be chemically cleaved to allow incorporation of the next nucleotide**





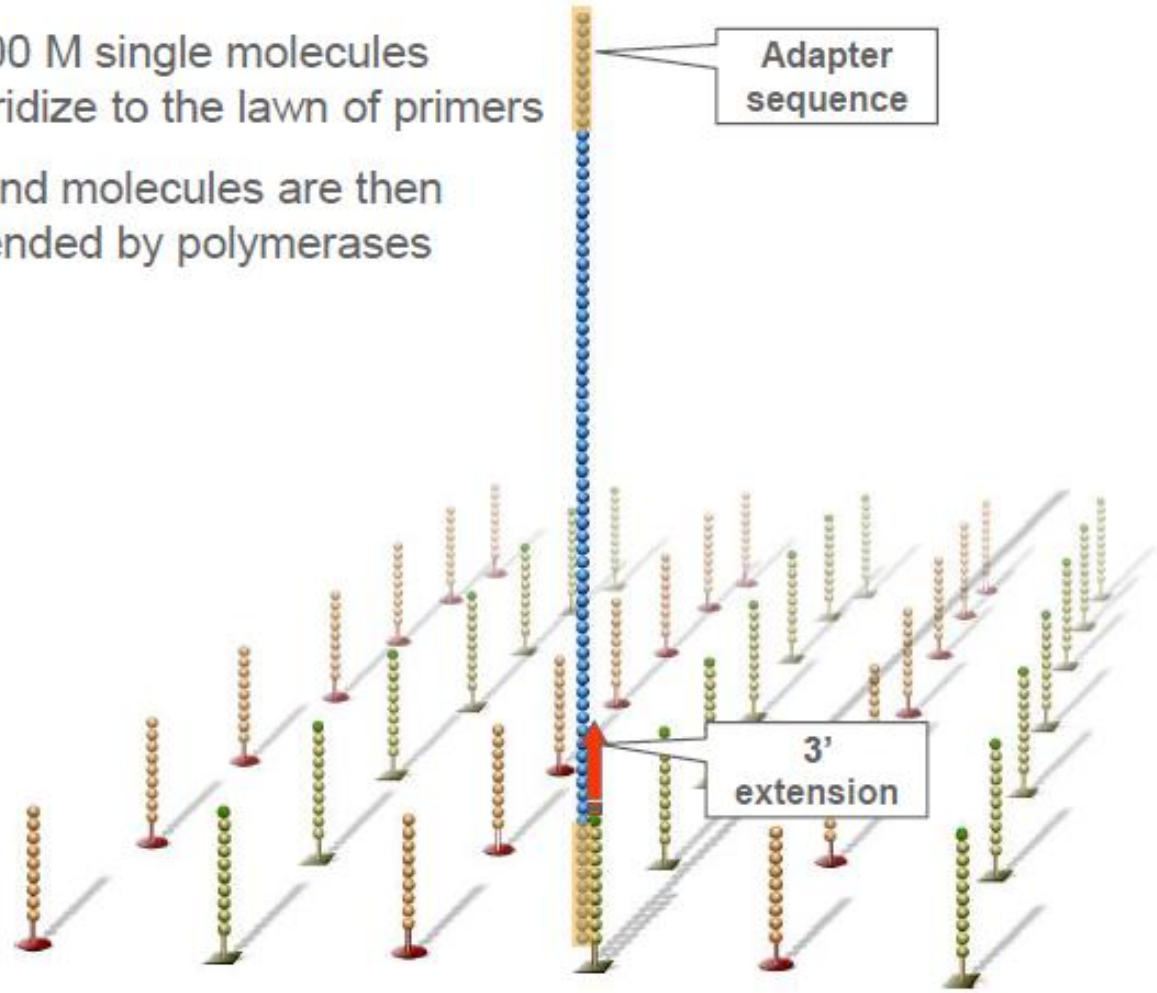
No, you do not need to know these structures!

Flow Cell


8 channels
Surface of flow cell coated with a lawn of oligo pairs
5'-PS-TTTTTTTTTTAATGATACGGCGACCACCGAGATCTACAC-3'
5'-PS-TTTTTTTTTTCAAGCAGAAGACGGCATACGAGoxoAT-3'

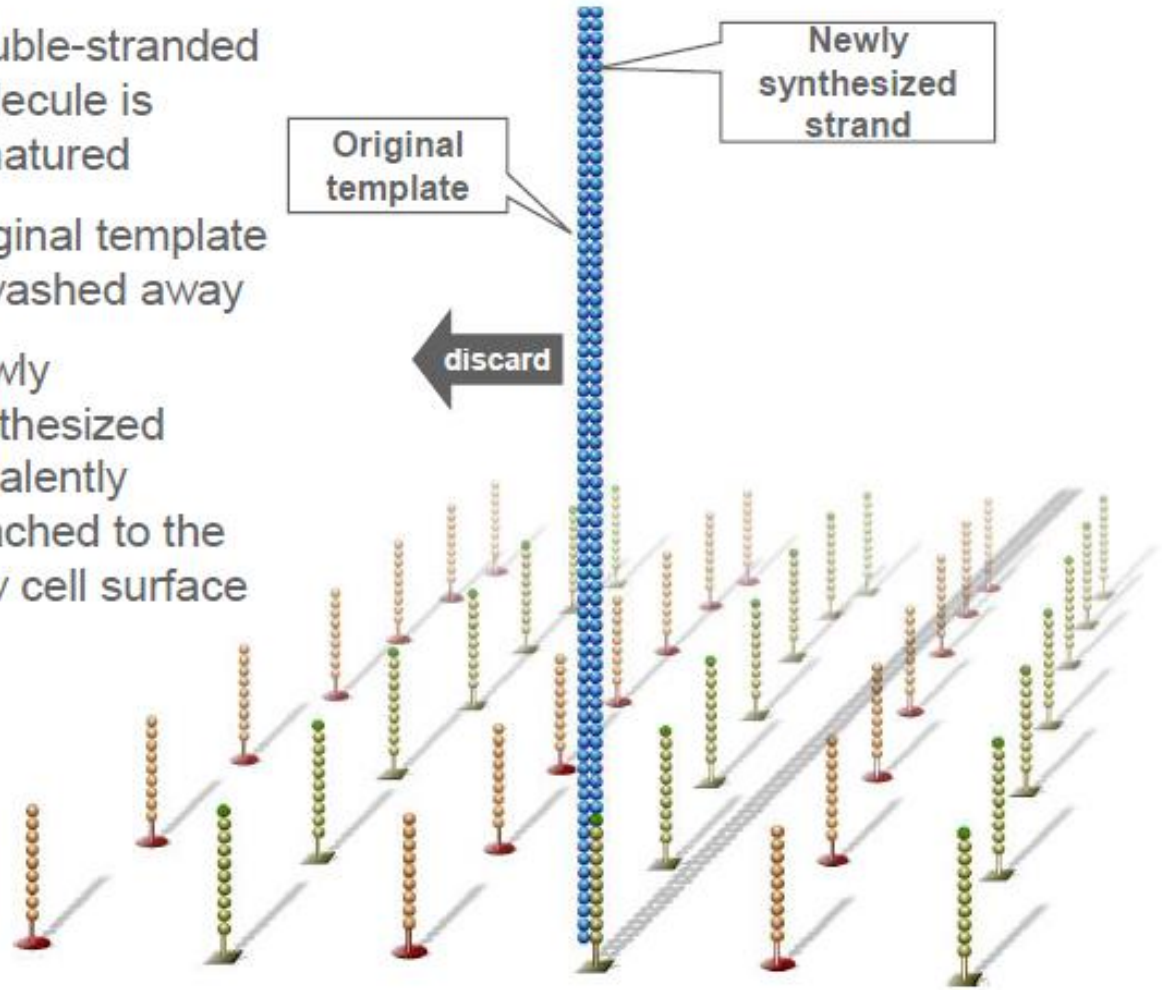
Cluster Generation:
Hybridize Fragment & Extend

- > 100 M single molecules hybridize to the lawn of primers
- Bound molecules are then extended by polymerases



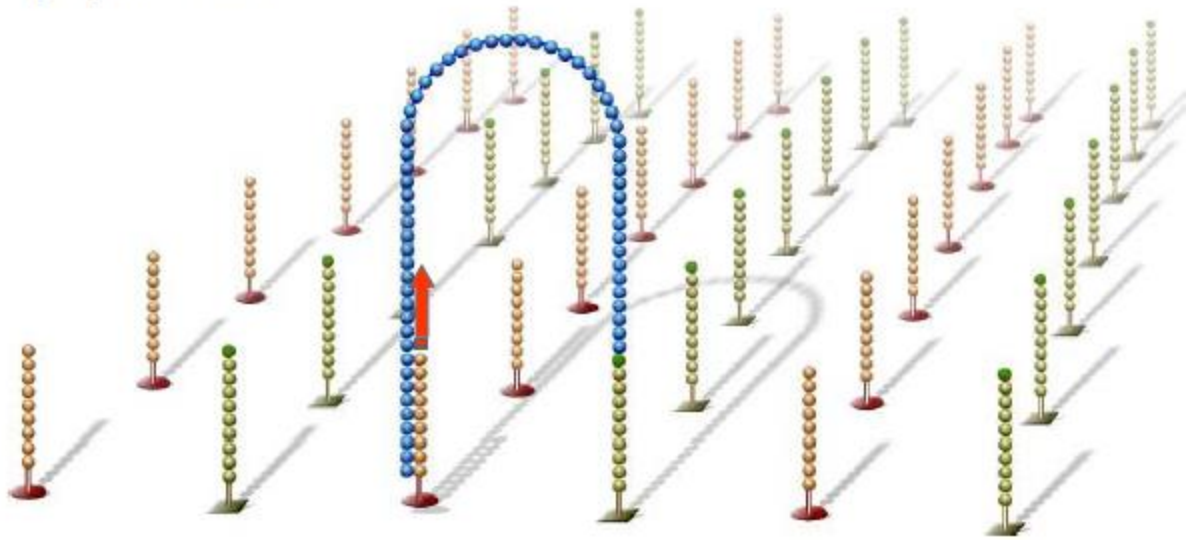
Cluster Generation: *Denature Double-stranded DNA*

- Double-stranded molecule is denatured
- Original template is washed away
- Newly synthesized covalently attached to the flow cell surface



Cluster Generation: *Bridge Amplification*

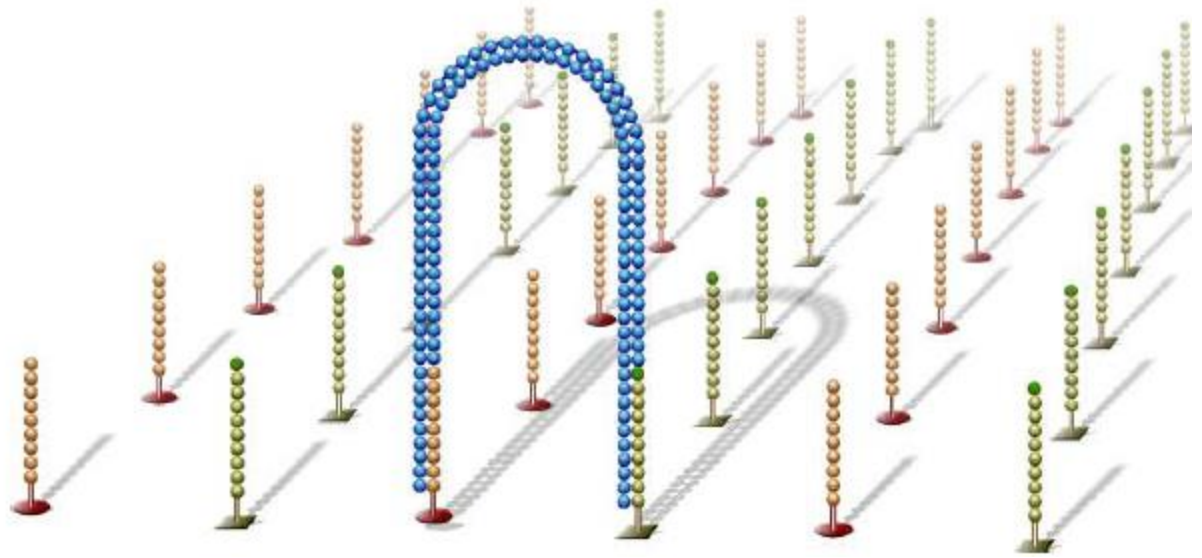
- Single-strand flips over to hybridize to adjacent primers to form a bridge
- Hybridized primer is extended by polymerases



Note that Bridge amplification is basically PCR on a solid support. The purpose is to generate thousands of copies of each initial template molecule, so that there is sufficient signal for detection during sequencing steps.

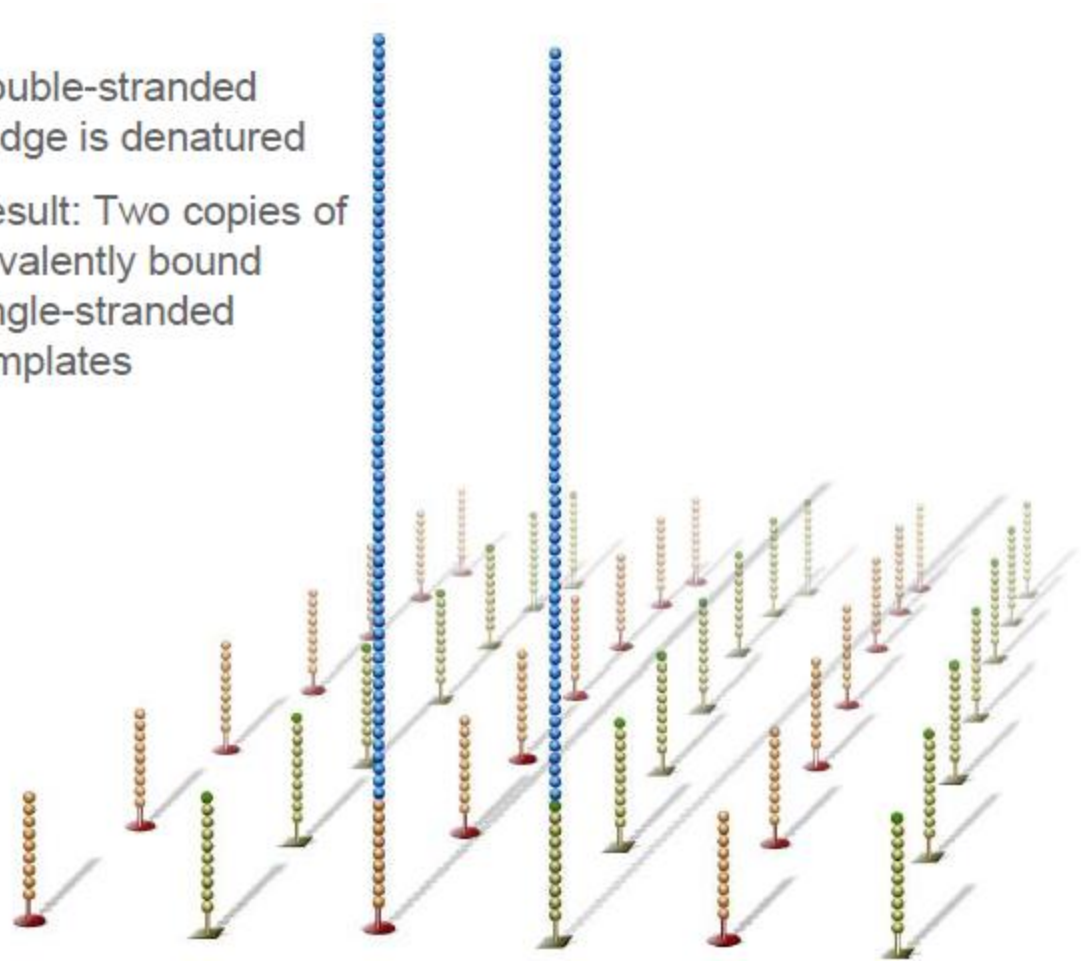
Cluster Generation:
Bridge Amplification

- Double-stranded bridge is formed



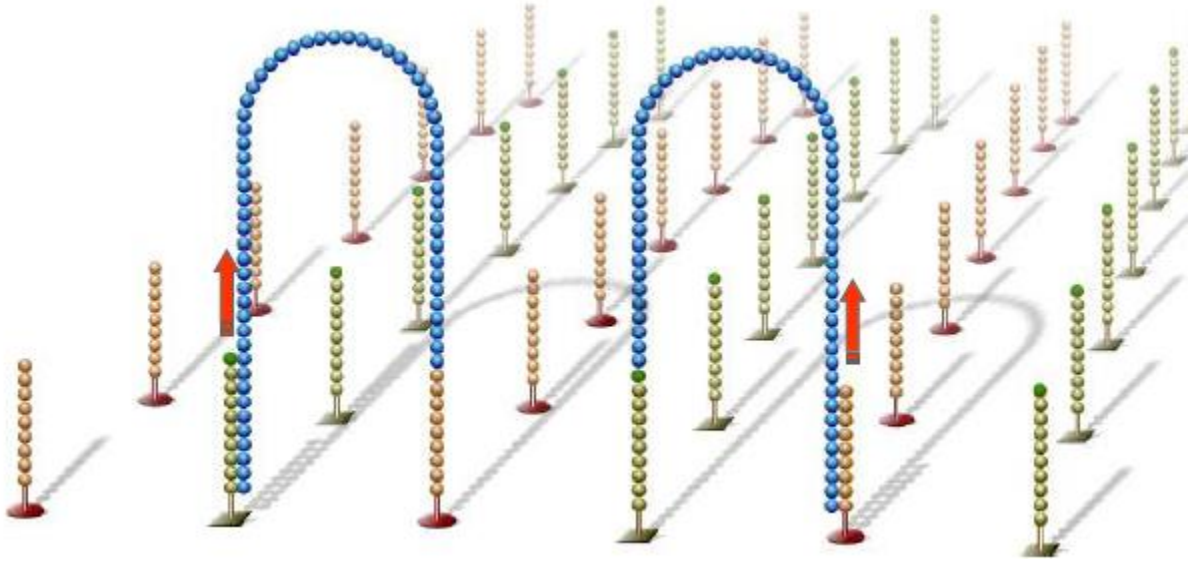
Cluster Generation: *Bridge Amplification*

- Double-stranded bridge is denatured
- Result: Two copies of covalently bound single-stranded templates



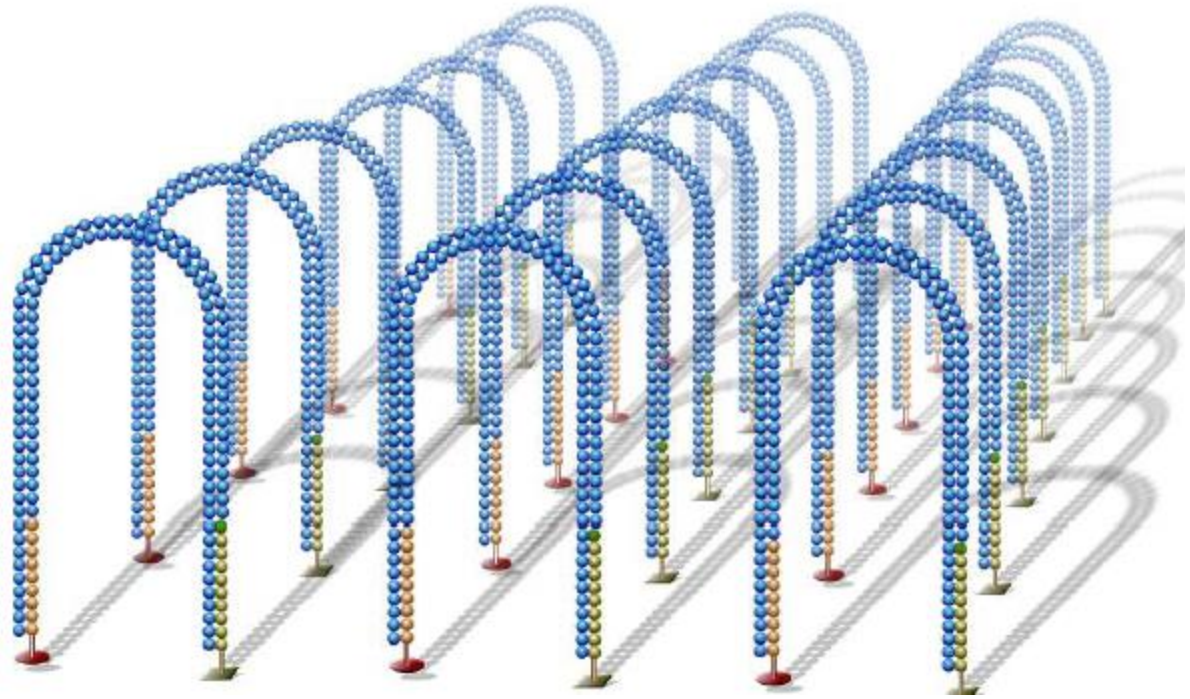
Cluster Generation: *Bridge Amplification*

- Single-strands flip over to hybridize to adjacent primers to form bridges
- Hybridized primer is extended by polymerase

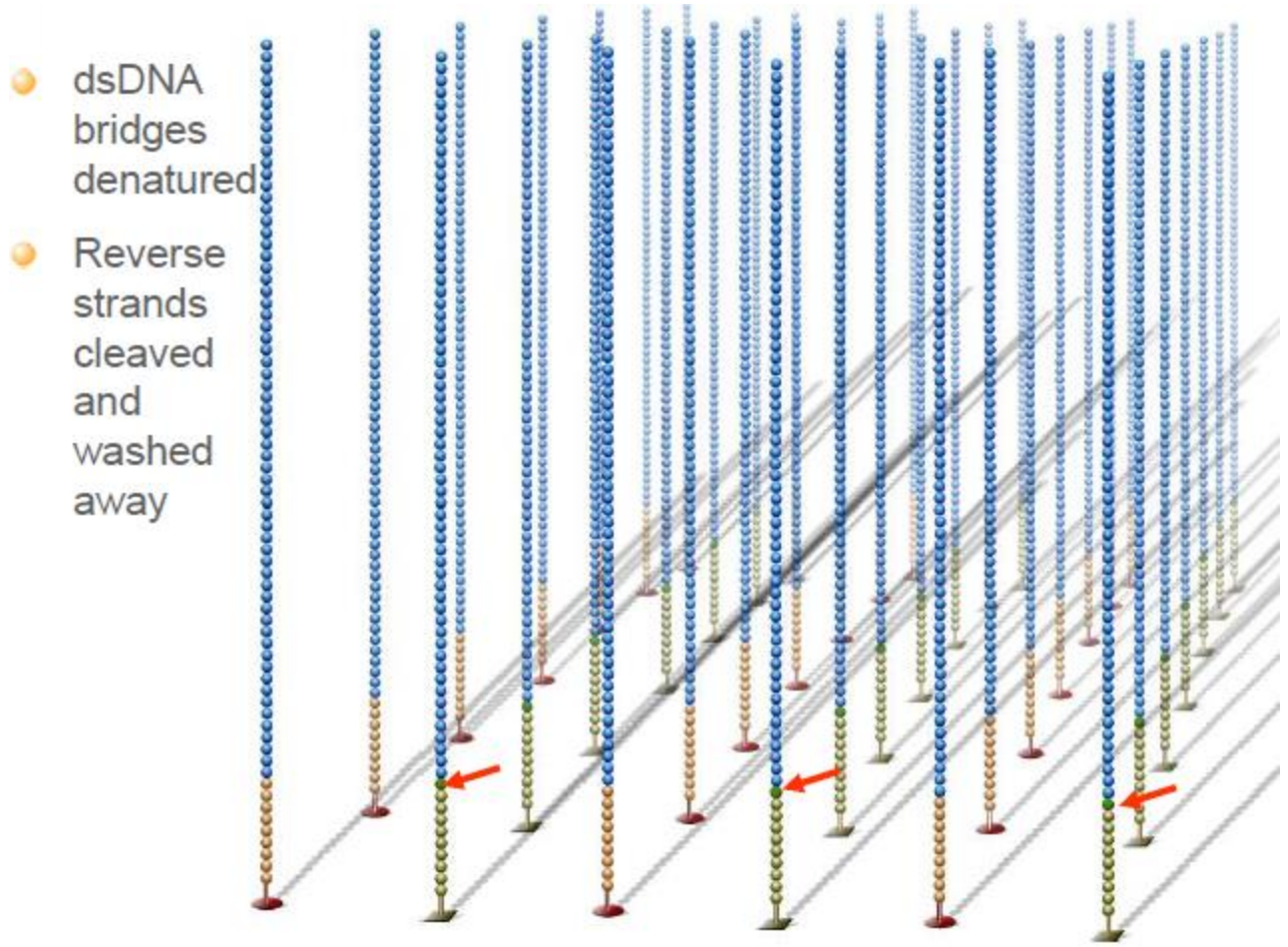


Cluster Generation: *Bridge Amplification*

- Bridge amplification cycle repeated until multiple bridges are formed



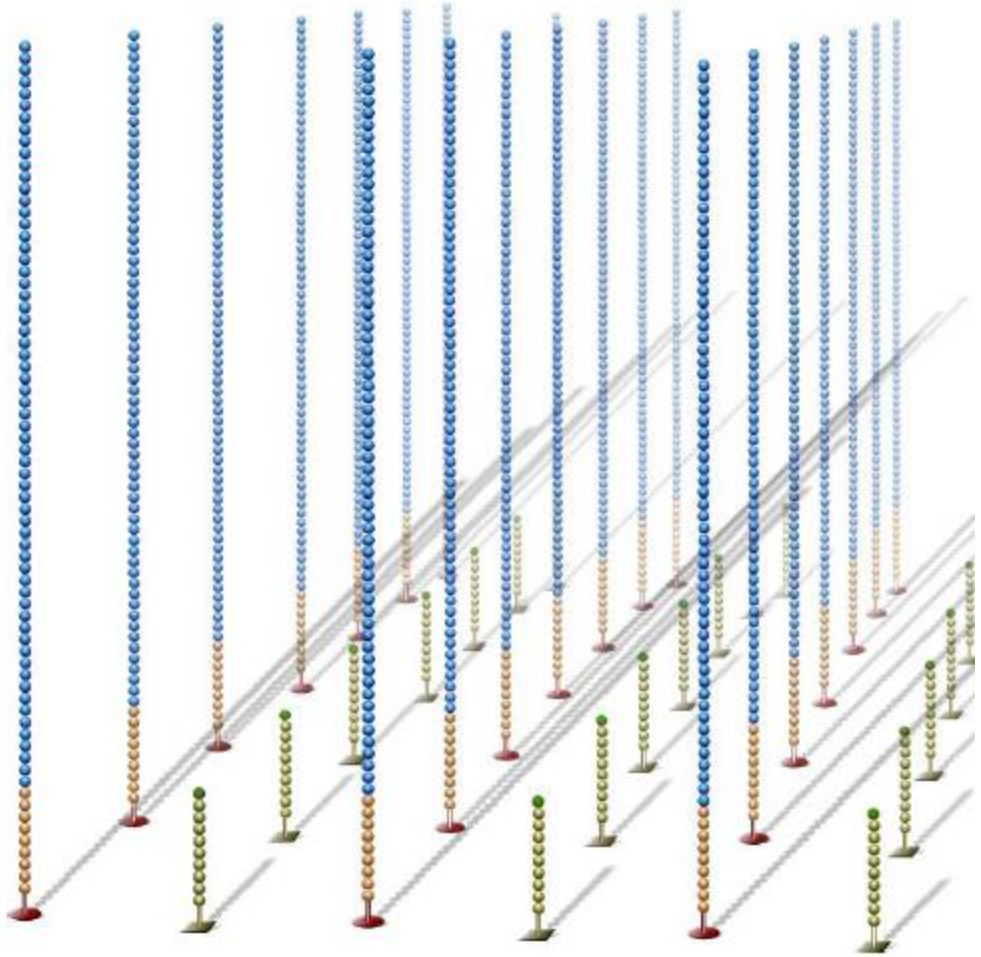
Cluster Generation



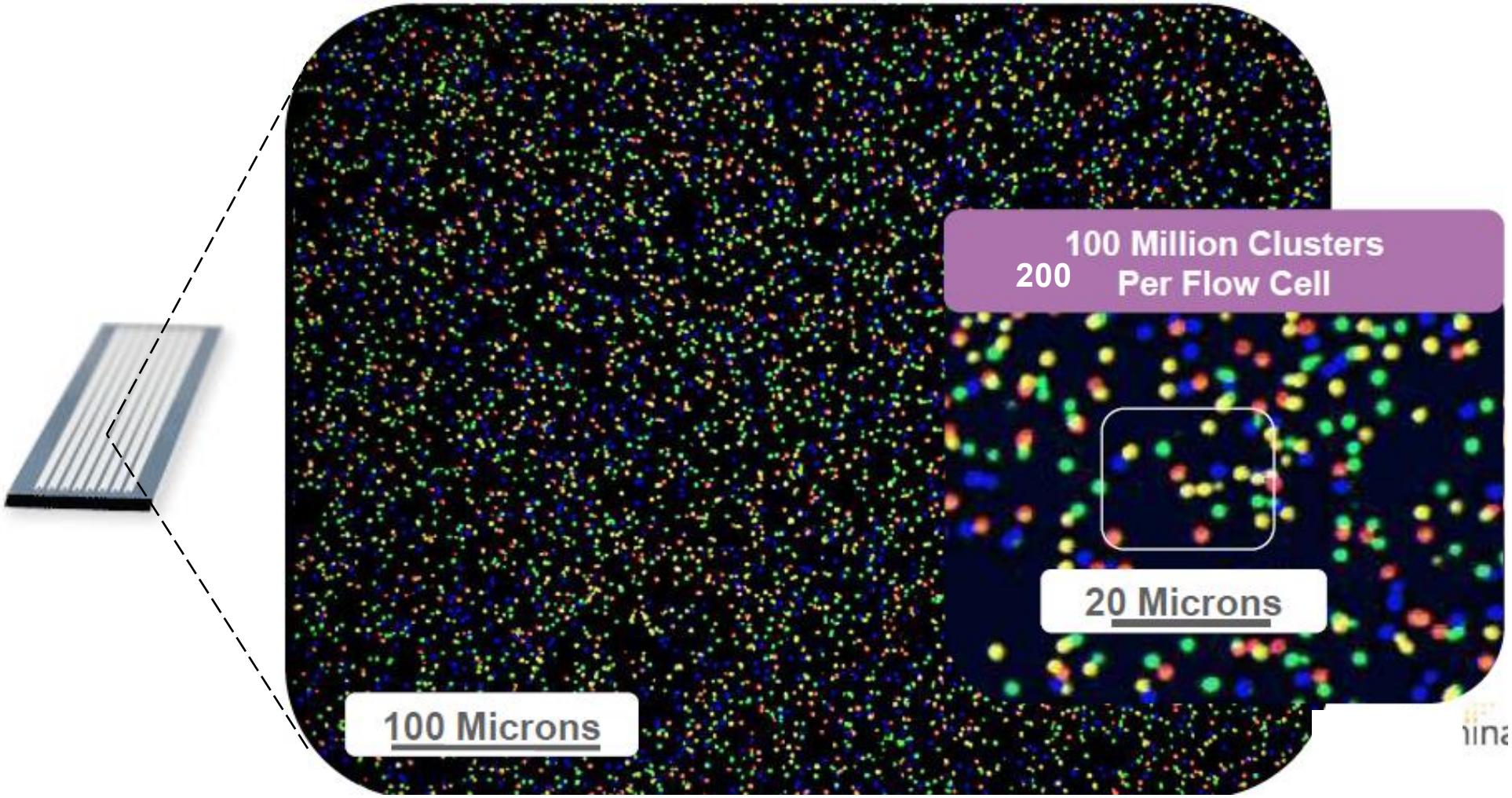
• there is a chemical modification to anchor oligos that allows selective cleaveage of the “forward” or “reverse” strands

Cluster Generation

- …leaving a cluster with forward strands only



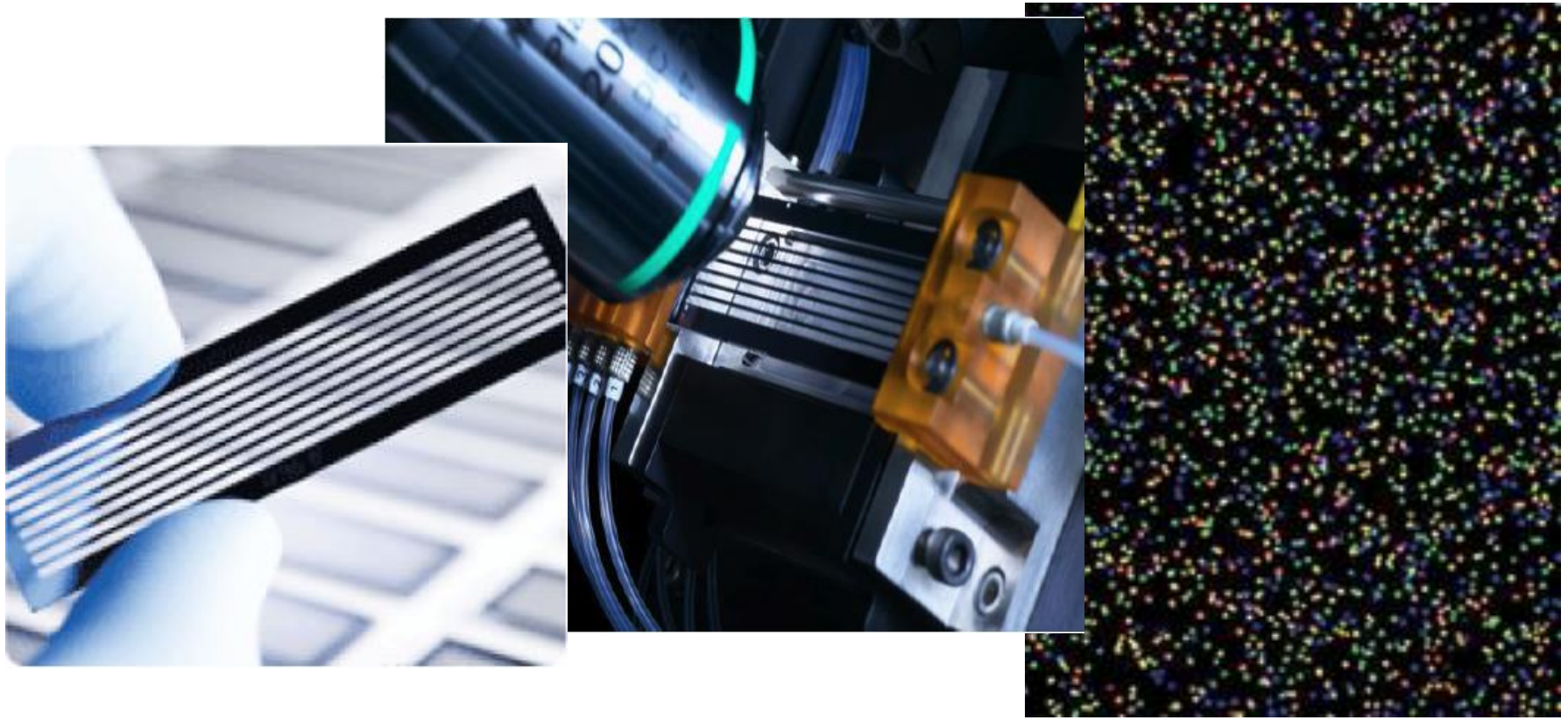
Now the DNA is ready to be sequenced!

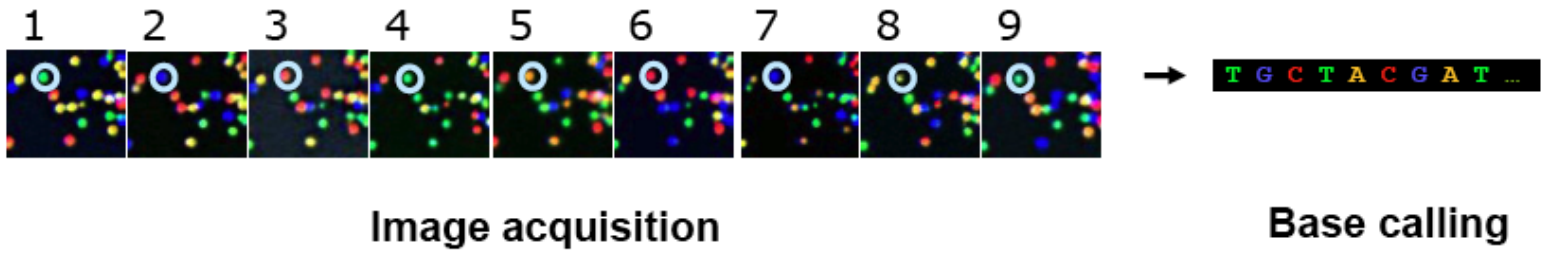


After Bridge PCR each lane has hundreds of millions of clusters

Each cluster contains thousands of copies of the template molecule that seeded it

BRIDGE PCR SEQUENCING: “HIGH-THROUGHPUT MASSIVELY PARALLEL WHOLE GENOME SEQUENCING”





Sequences assembly

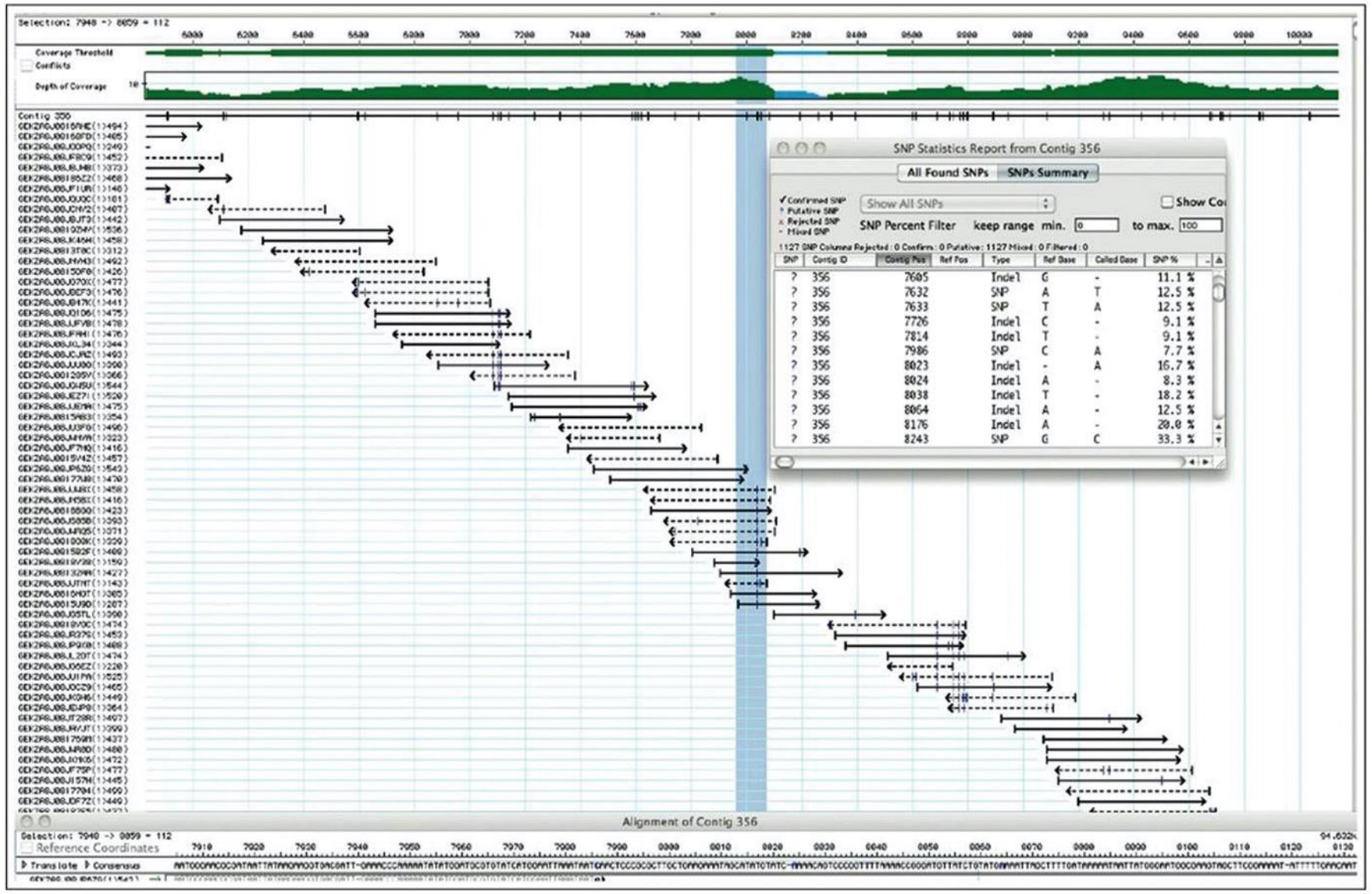
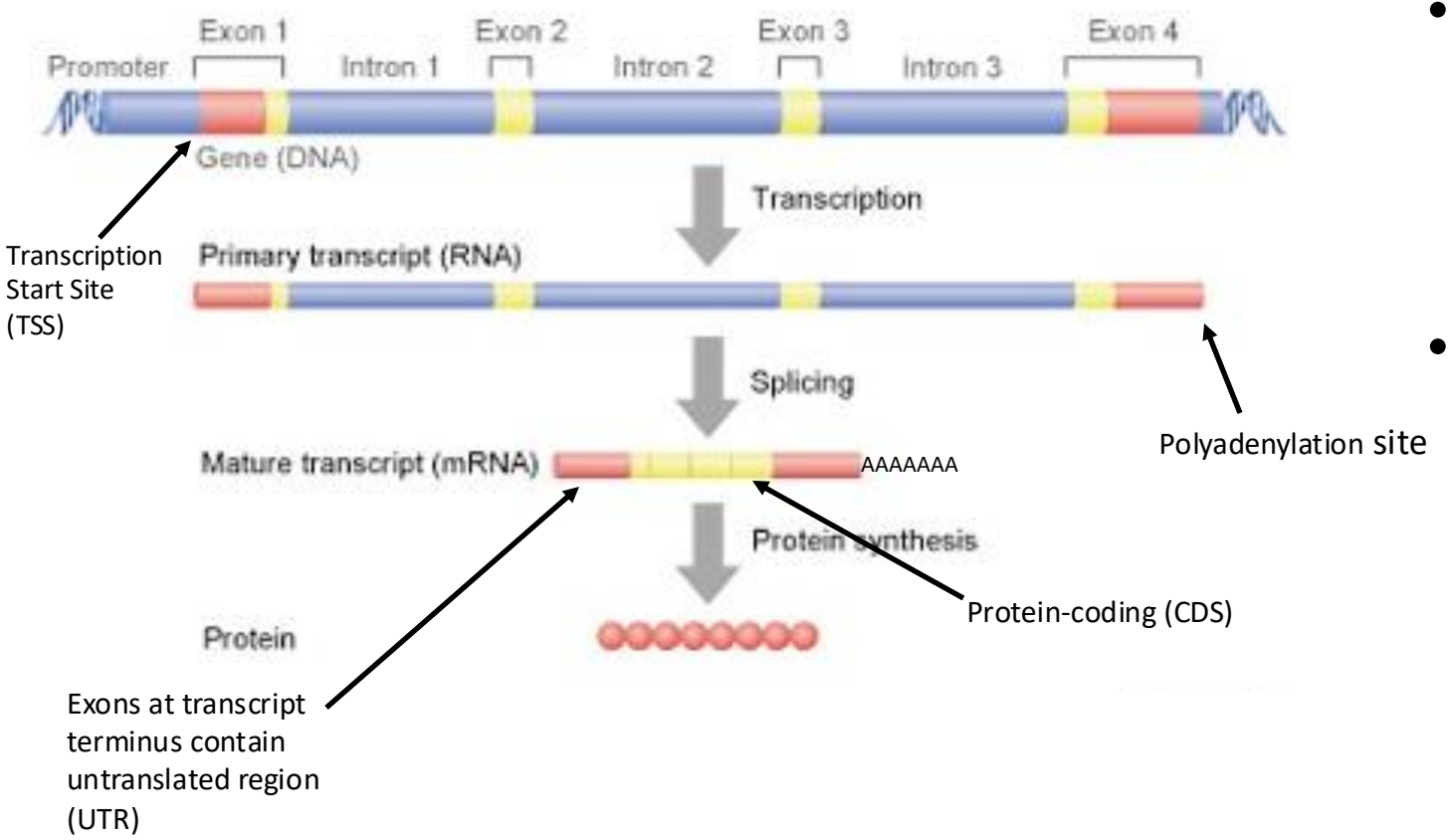


Figure 7-15

Molecular Biology: Principles and Practice, Second Edition

Genome vs transcriptome

The genes expressed in a cell under a given set of conditions constitutes its **transcriptome**



- The transcriptome may change in response to environmental changes or cellular signals.
- Microarrays provide one picture of cellular transcriptomes. The **RNA-Seq** approach is even more effective in generating a detailed transcriptome.

Microarrays can be used to show a selection of genes in an organism

- Can be used to look at genes at a specific developmental stage or those under a specific set of environmental conditions
- cDNA are synthesized with nucleotides that fluoresce in different colours for each sample and used to probe the microarray
- Limitation: genes you are tested for is limited to the number of spots available on the microarray
- Advances here… RNA-Seq

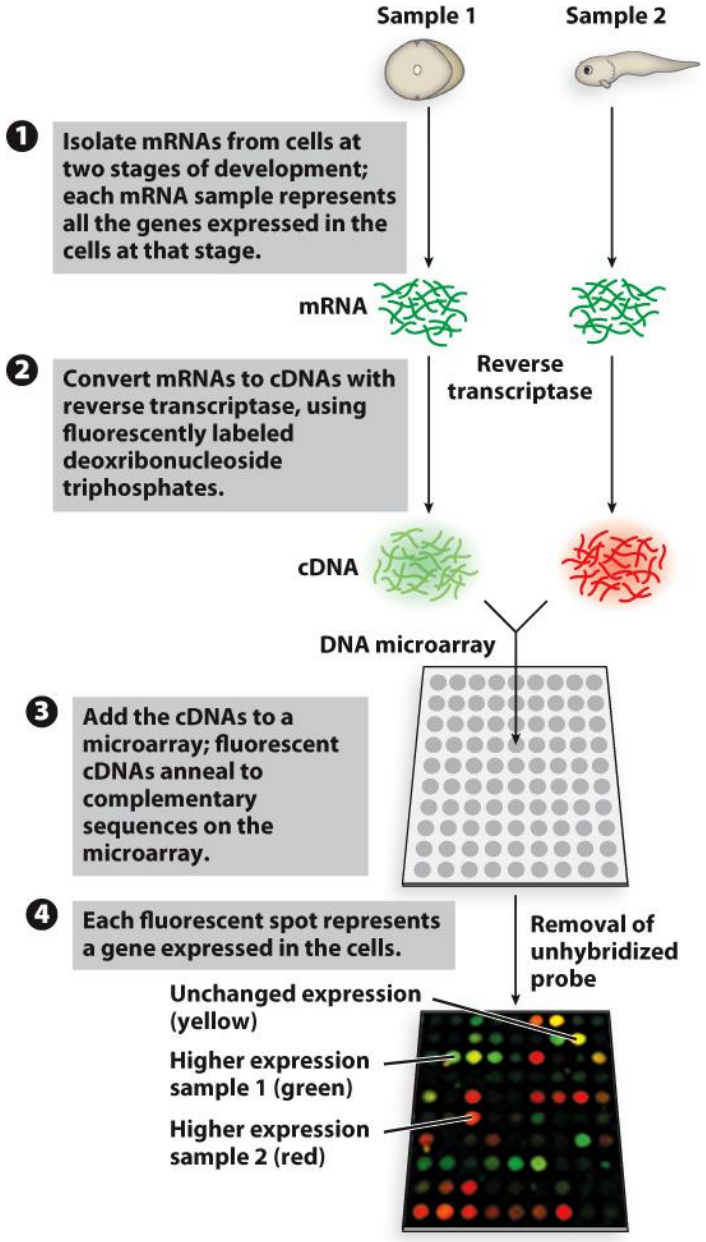


Figure 7-31
Molecular Biology: Principles and Practice, Second Edition

RNA-Seq: High Throughput RNA sequencing

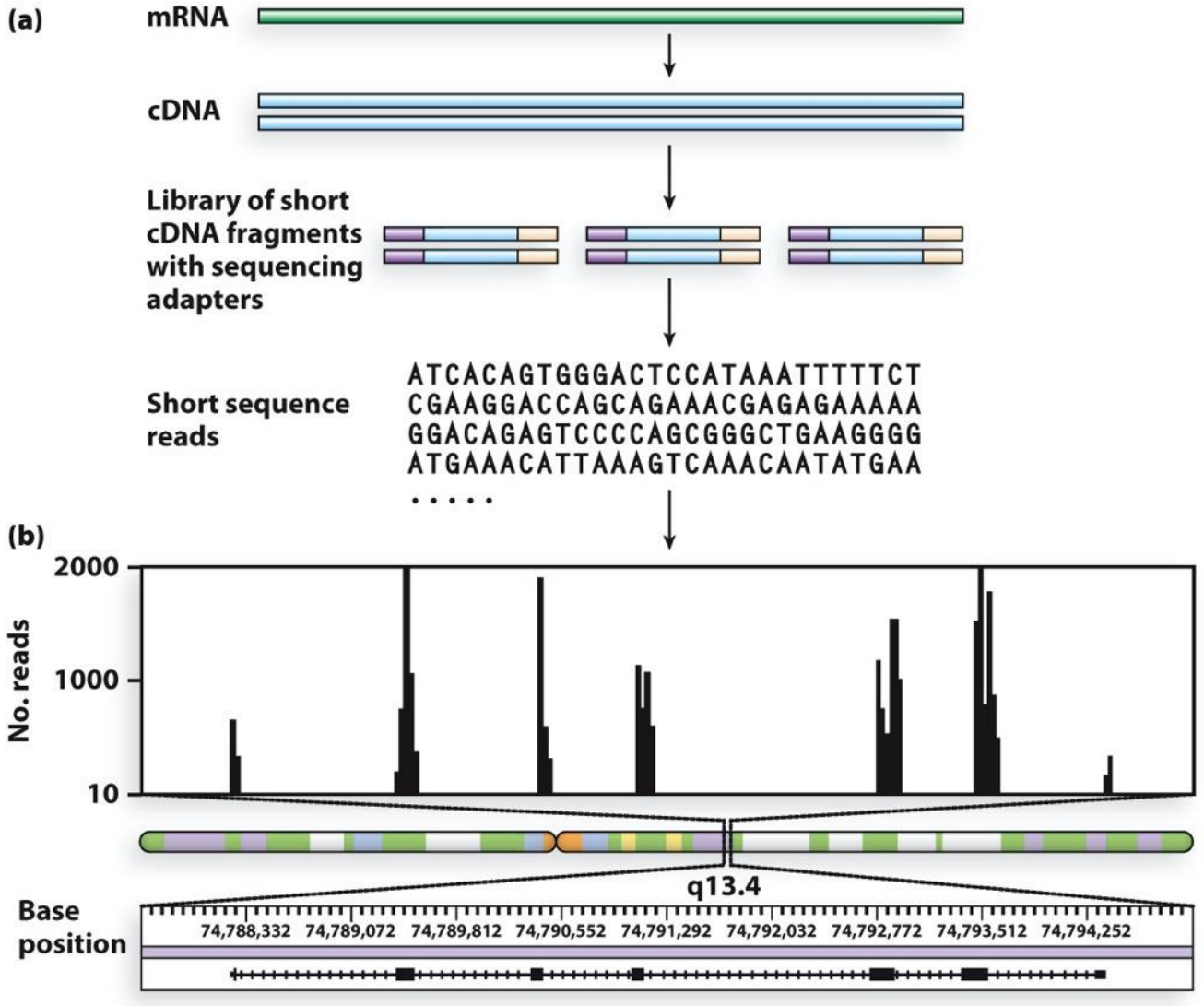


Figure 8-12
Molecular Biology: Principles and Practice, Second Edition

- provides information on gene expression levels with a much greater dynamic range and has proved highly accurate
- The direct sequencing also provides additional information, showing the exact transcriptional boundaries of genes and revealing how exons are linked together in transcripts.
- In some organisms, RNA transcripts are edited, producing new sequences not present in the DNA genes