

Sequencing Facility Tour

- Jul 27, 2018 10:30 AM to 11:30 AM
- Bus from hospital around 10 AM
- How Many?: 14

RNA-Seq Sample Preparation Theory

From Sample to Data

Rna-Seq Applications

- Differential Expression
- Transcriptome
- Genome Annotation
- “Bargain Exome”
- SNPs
- Gene Fusions

Overview

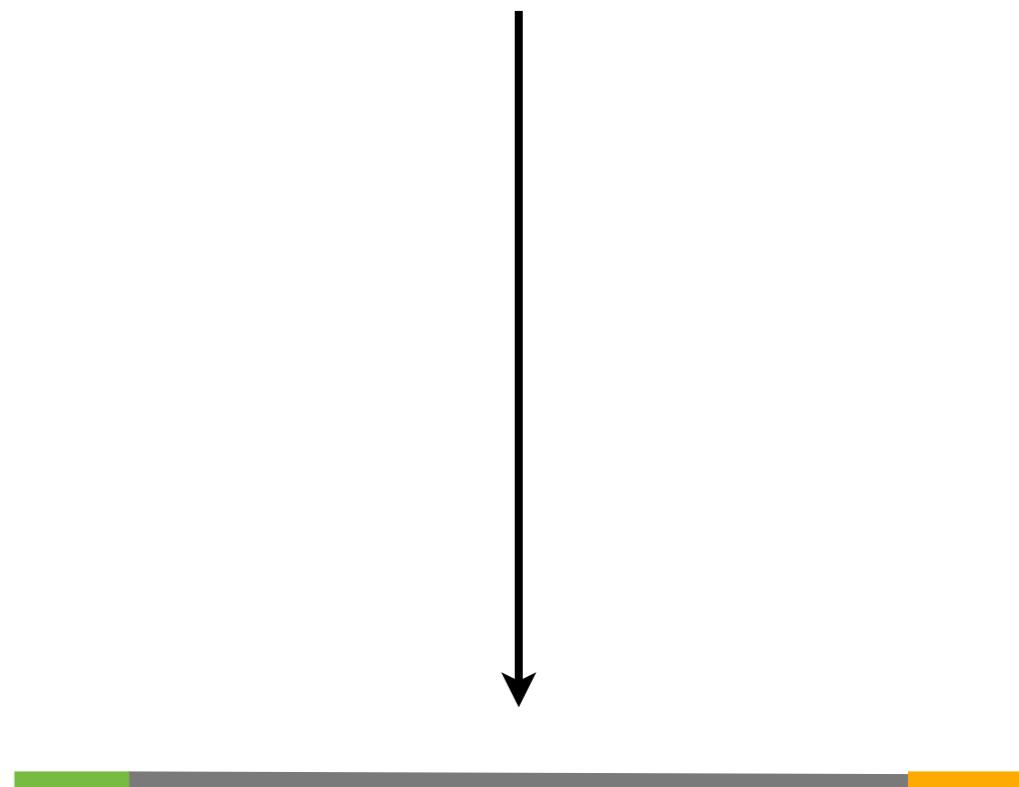
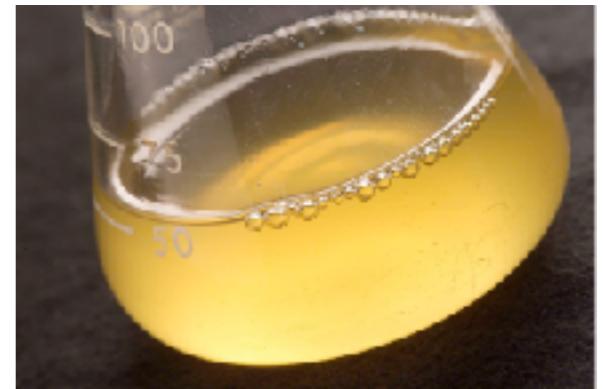
- From Library to Data
- Illumina Sequencing

Not Appearing Today

- Data Analysis
- Software

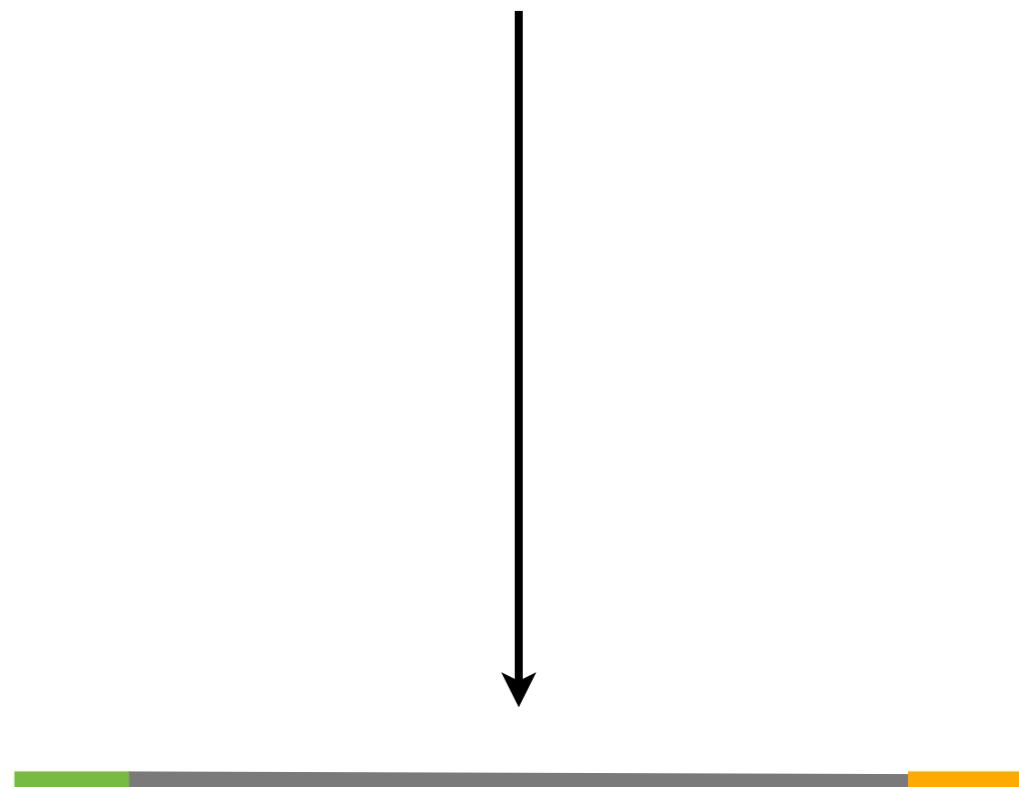
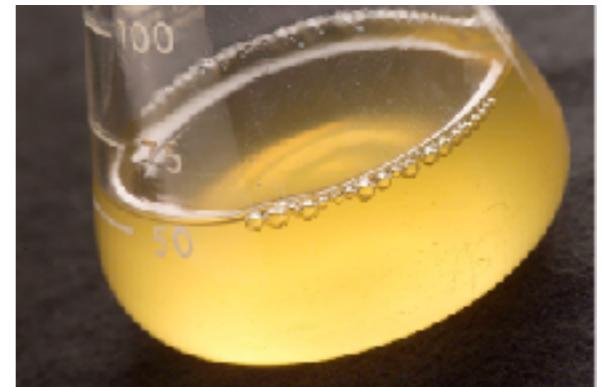
Major Experiment Components

- 1.Growth
- 2.Sample Collection
- 3.RNA Extraction
- 4.mRNA Enrichment
- 5.Library Preparation



Major Experiment Components

- 1.Growth**
- 2.Sample Collection**
- 3.RNA Extraction**
- 4.mRNA Enrichment**
- 5.Library Preparation**



Growth and Sample Collection

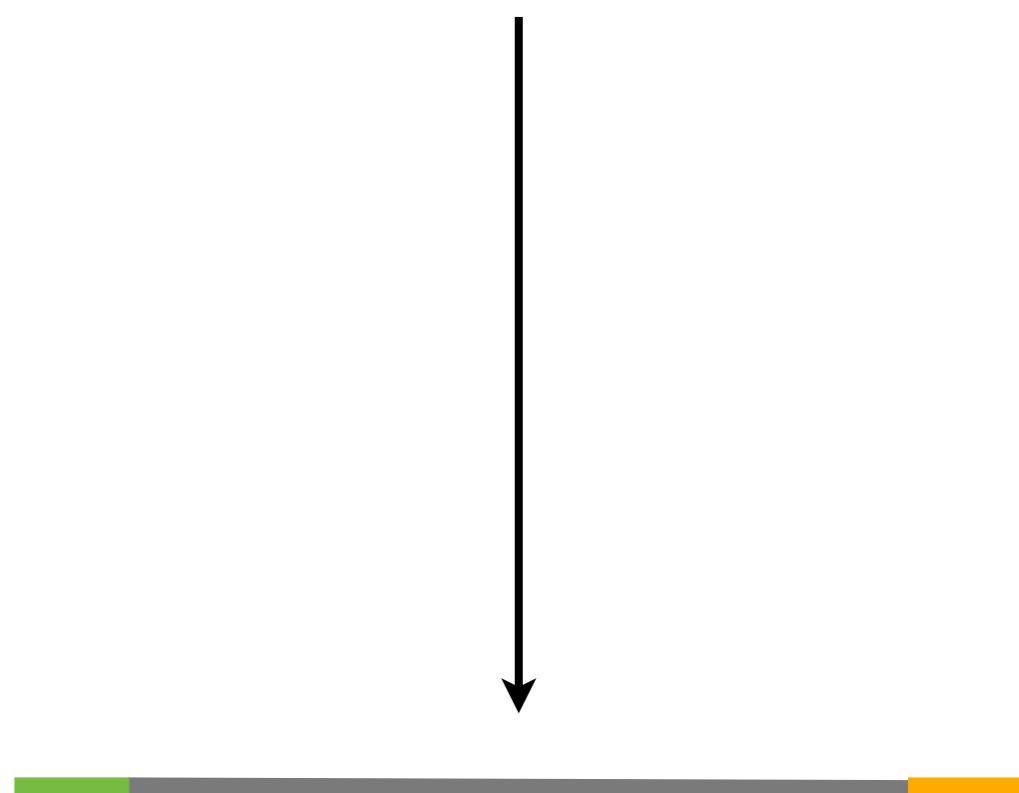
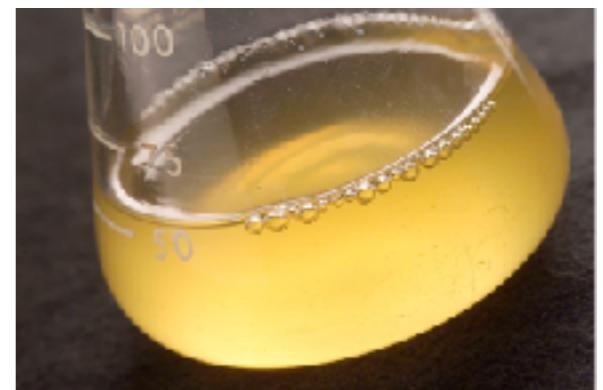
- Avoid Confounding Factors!
- System Specific
- Experiment Specific
- **Avoid RNA response to sample collection!**

Sample Collection Options

- Flash freeze
- RNA stabilizers
 - RNA protect
 - RNAlater
- Phenol (hot acid phenol, trizol, etc)

Major Experiment Components

- 1.Growth
- 2.Sample Collection
- 3.RNA Extraction
- 4.mRNA Enrichment
- 5.Library Preparation



RNA Extraction: Why?

- Have cells, need RNA

RNA Extraction Options

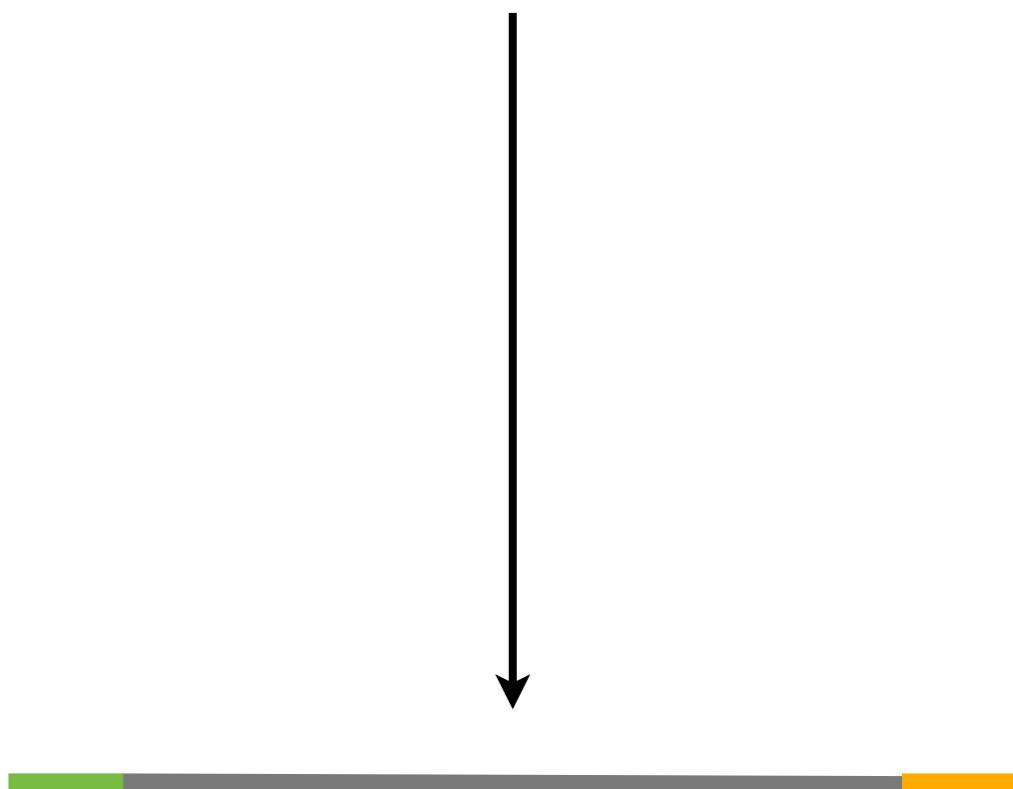
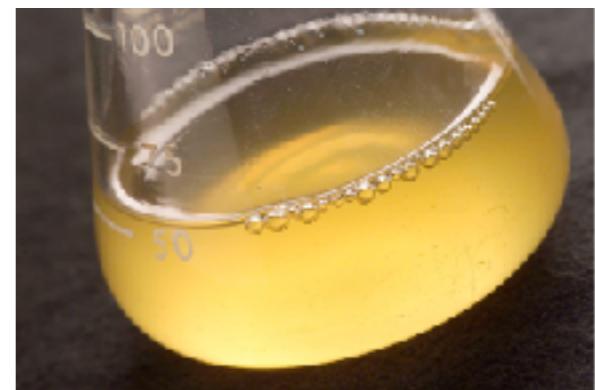
- Kits
 - Qiagen RNeasy Mini Kit
 - Etc
- Phenol (hot acid phenol, trizol, etc)

Our Samples

1. Collect cells (centrifuge liquid culture)
2. Flash freeze
3. Lyophilize overnight
4. Bead beating to break open cells
5. Qiagen RNEasy: “Purification of Total RNA from Plant Cells and Tissues and Filamentous Fungi”

Major Experiment Components

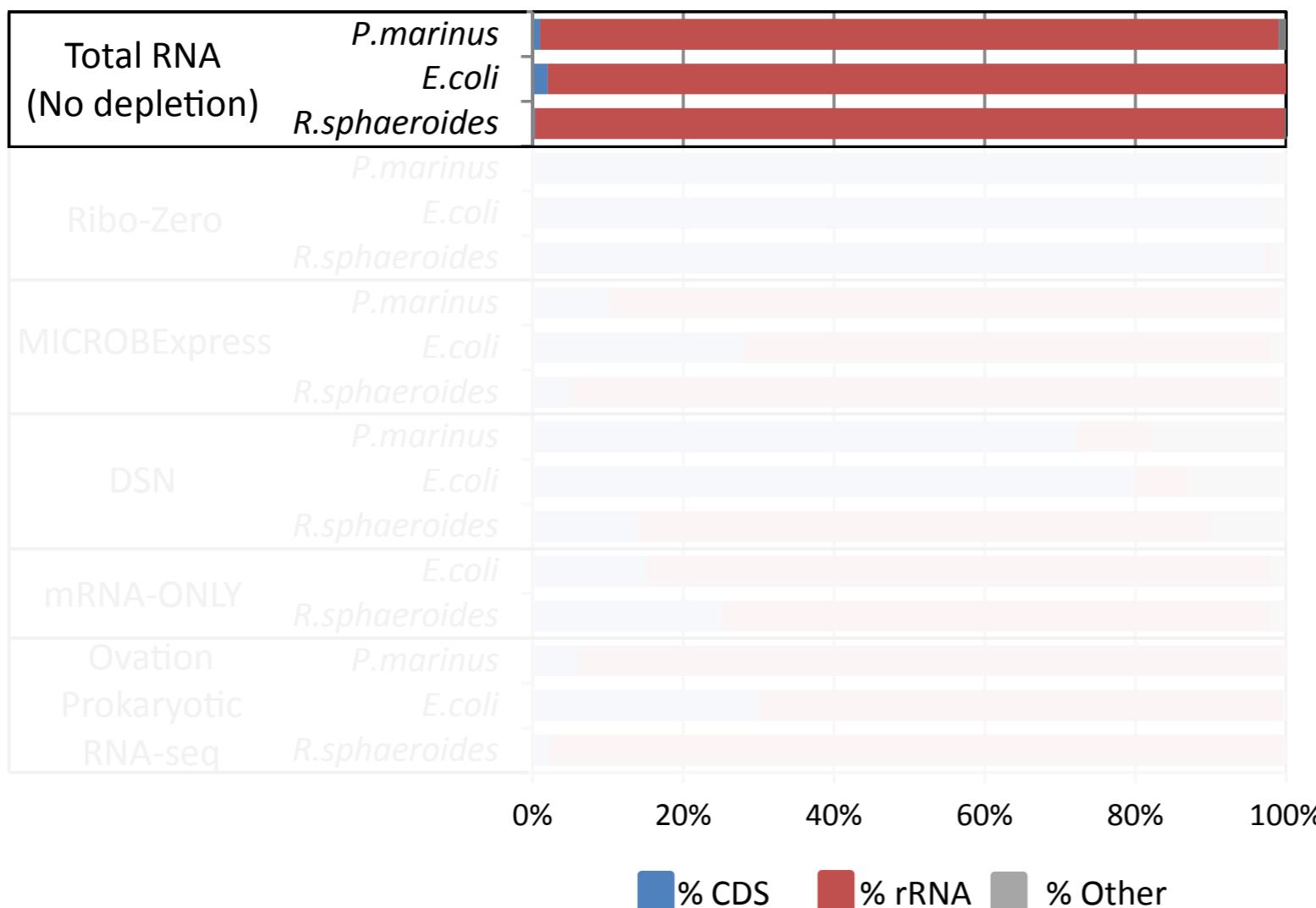
- 1.Growth
- 2.Sample Collection
- 3.RNA Extraction
- 4.mRNA Enrichment
- 5.Library Preparation



mRNA Enrichment: Why?

mRNA Enrichment: Why?

(a)

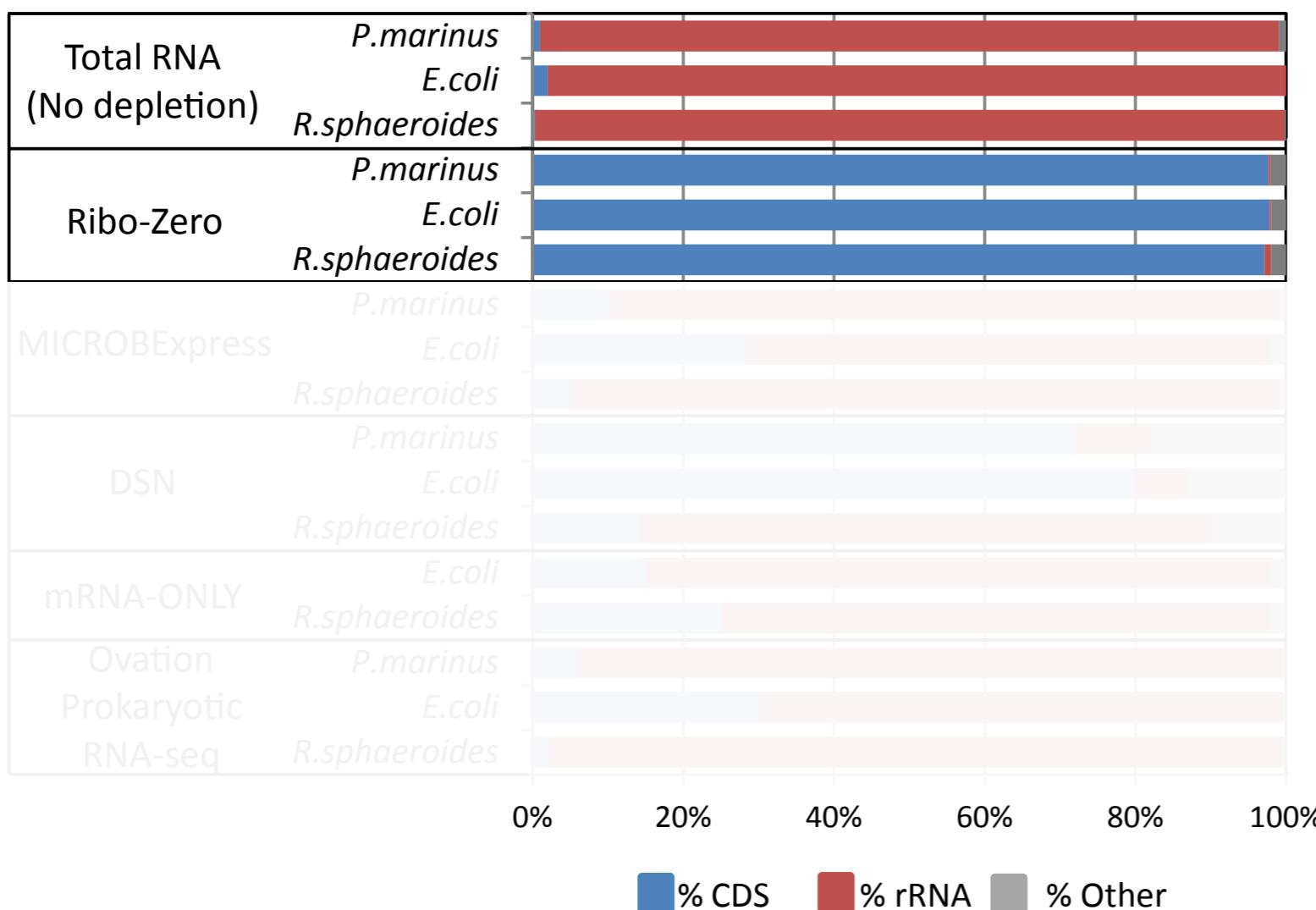


(b)



mRNA Enrichment: rRNA Depletion

(a)



(b)



rRNA Depletion: How?

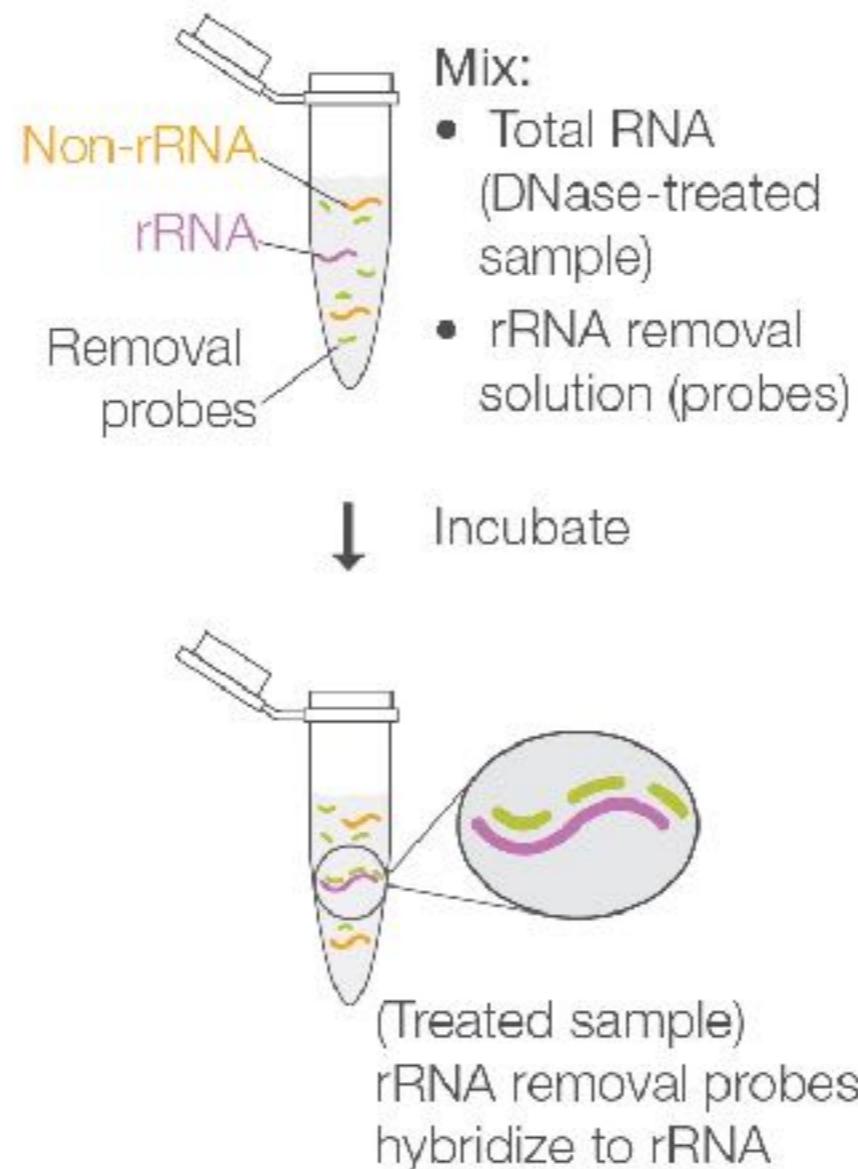
Part I:

Pellet and wash
magnetic beads (x2)

Magnet

Resuspend magnetic
beads (+ optional
RiboGuard
RNase Inhibitor)

Treat Sample With rRNA Removal Solution

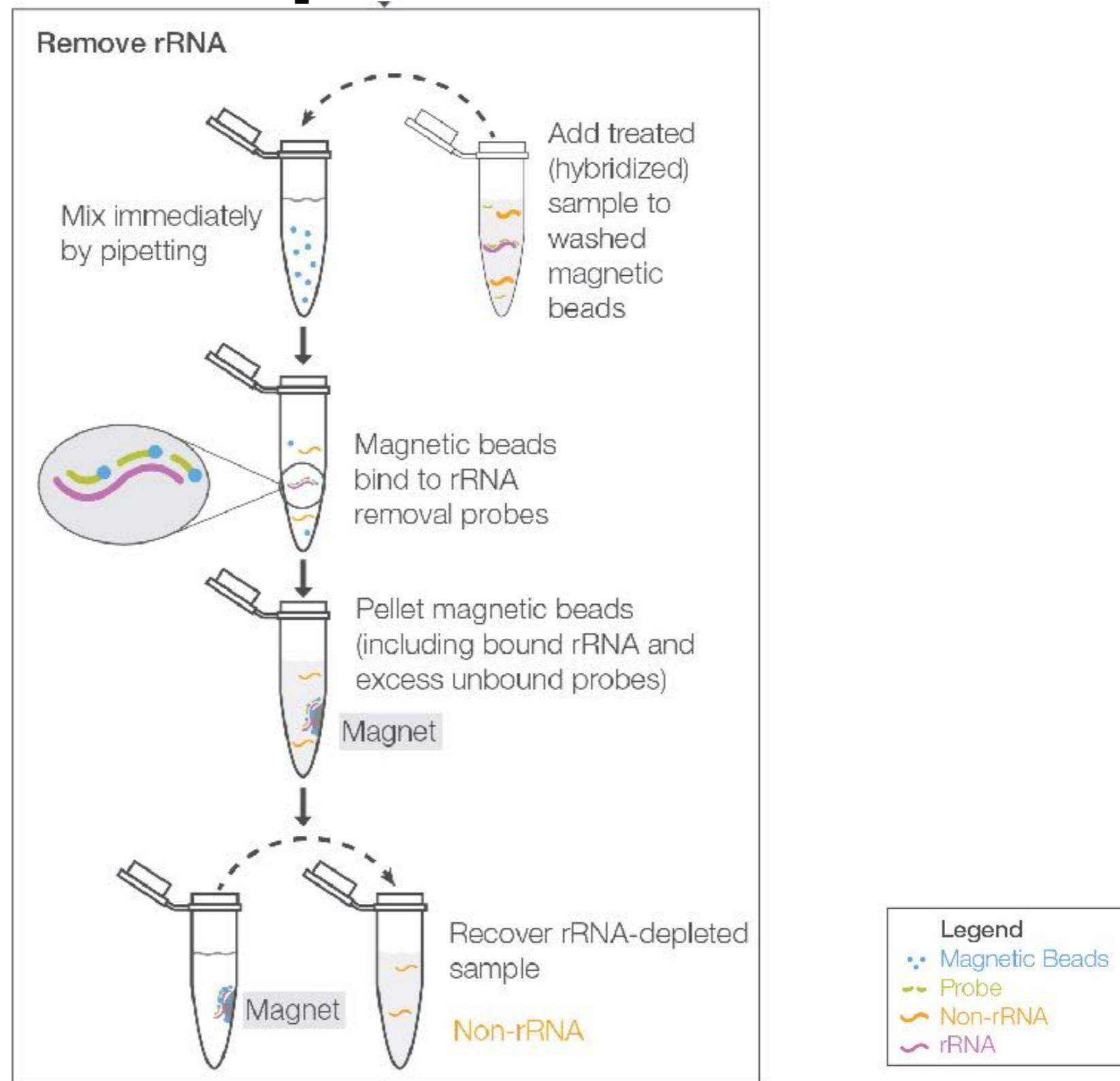


Legend

- Magnetic Beads
- Probe
- Non-rRNA
- rRNA

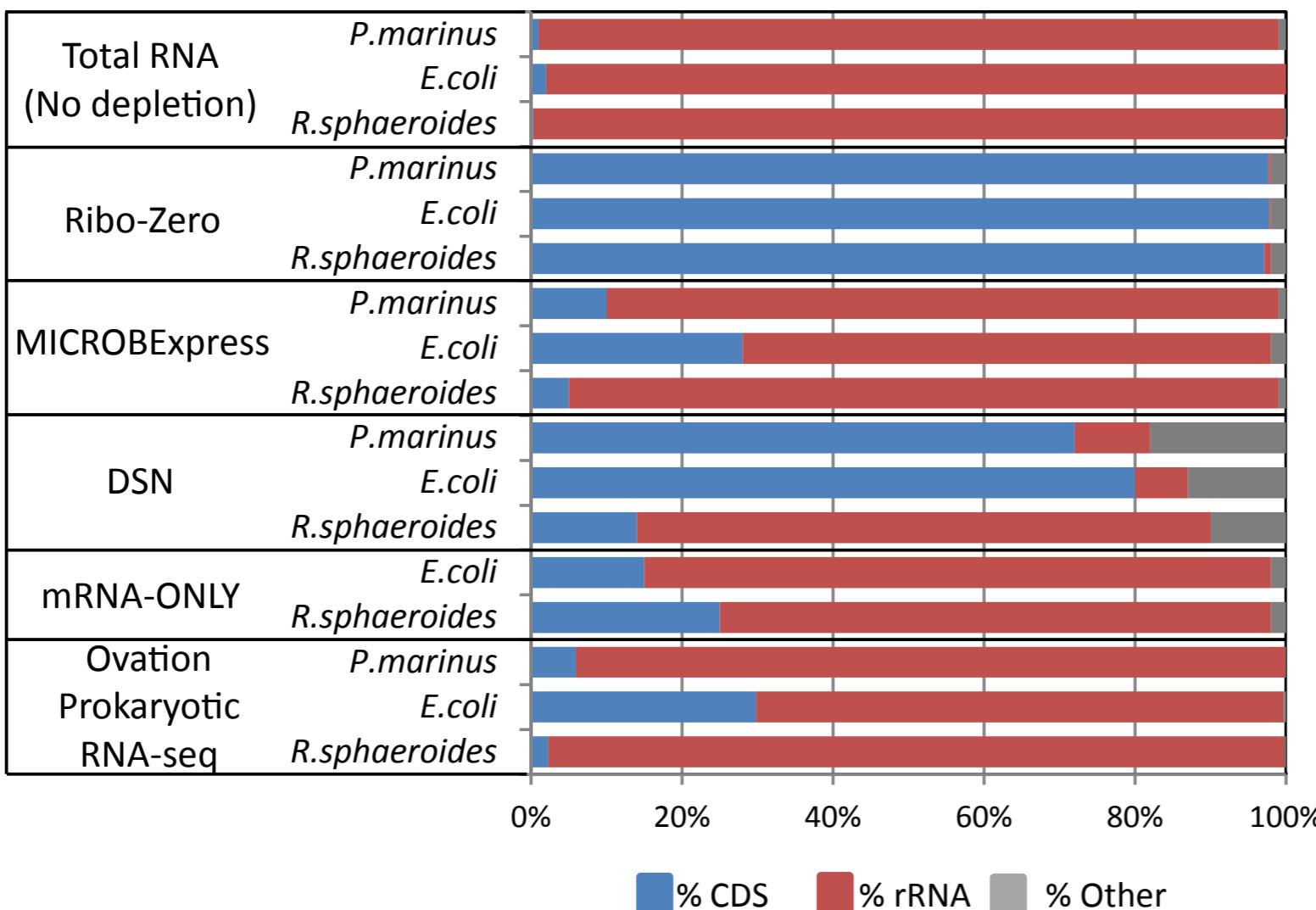
rRNA Depletion: How?

Part 2:



mRNA Enrichment: Alternatives

(a)

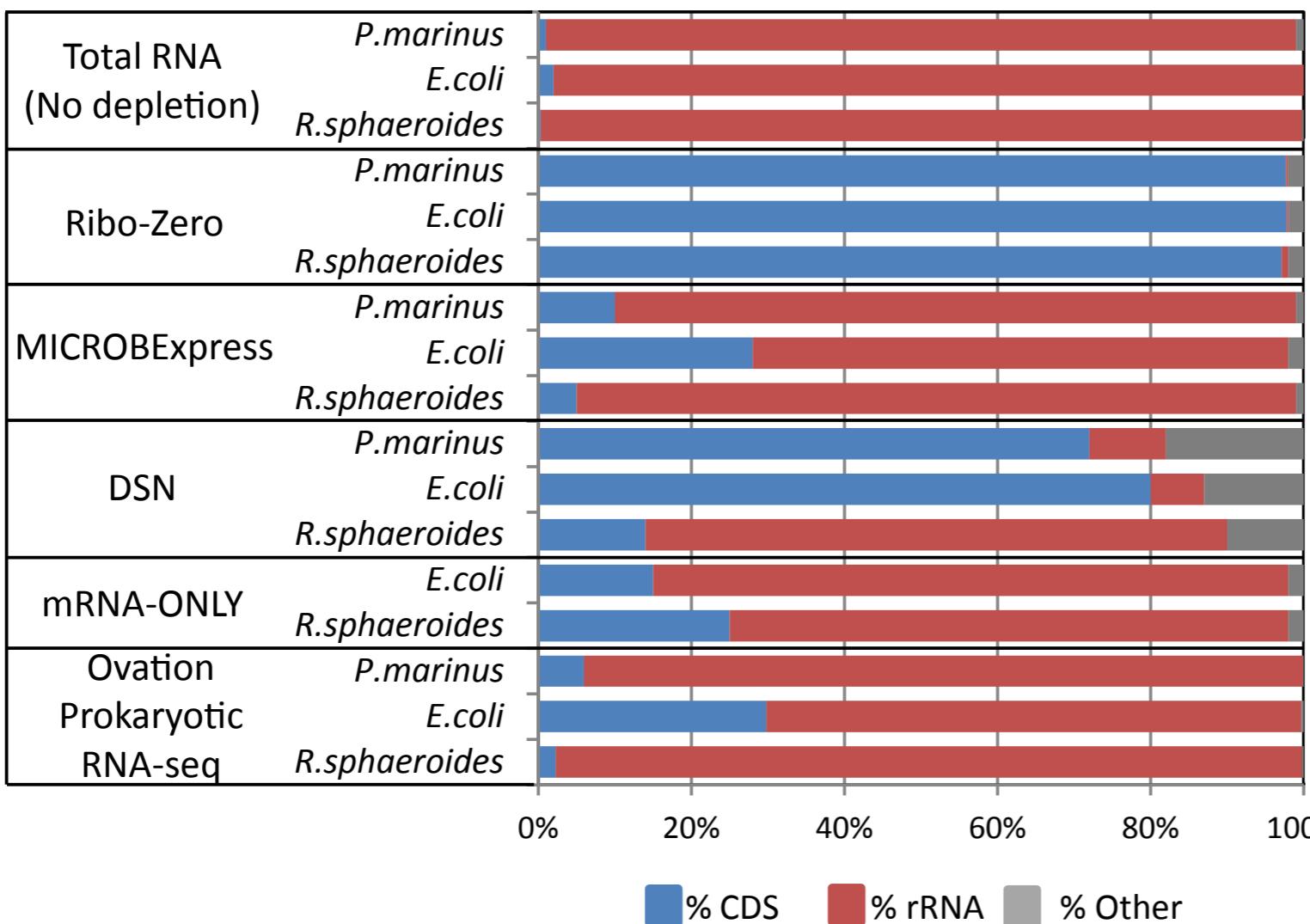


(b)



mRNA Enrichment: Alternatives

(a)



(b)



mRNA Enrichment: Alternatives

- Selection for desired RNA
 - poly(A) mRNA enrichment
 - Selective polyadenylation of mRNAs
 - Antibody capture of RNAs that interact with a specific protein
 - Non-random priming
- Selection against non-desired RNA
 - Ribosomal RNA capture
 - Duplex-specific nuclease (DSN) normalization
 - Degradation of processed RNA

mRNA Enrichment: Alternatives

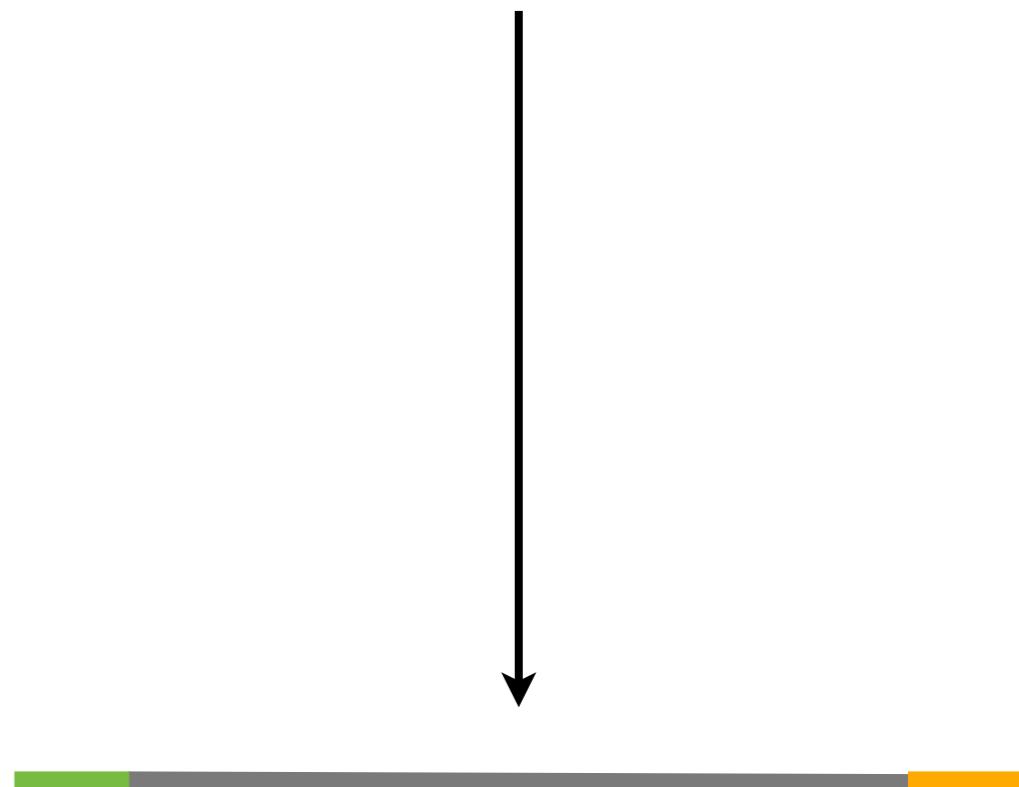
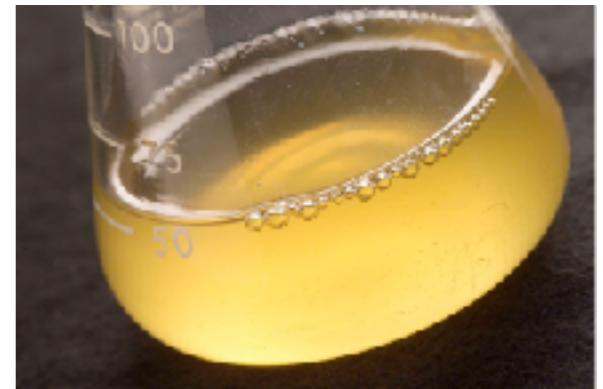
- Selection for desired RNA
 - poly(A) mRNA enrichment
 - Selective polyadenylation of mRNAs
 - Antibody capture of RNAs that interact with a specific protein
 - Non-random priming
- Selection against non-desired RNA
 - Ribosomal RNA capture
 - Duplex-specific nuclease (DSN) normalization
 - Degradation of processed RNA

mRNA Enrichment: Alternatives

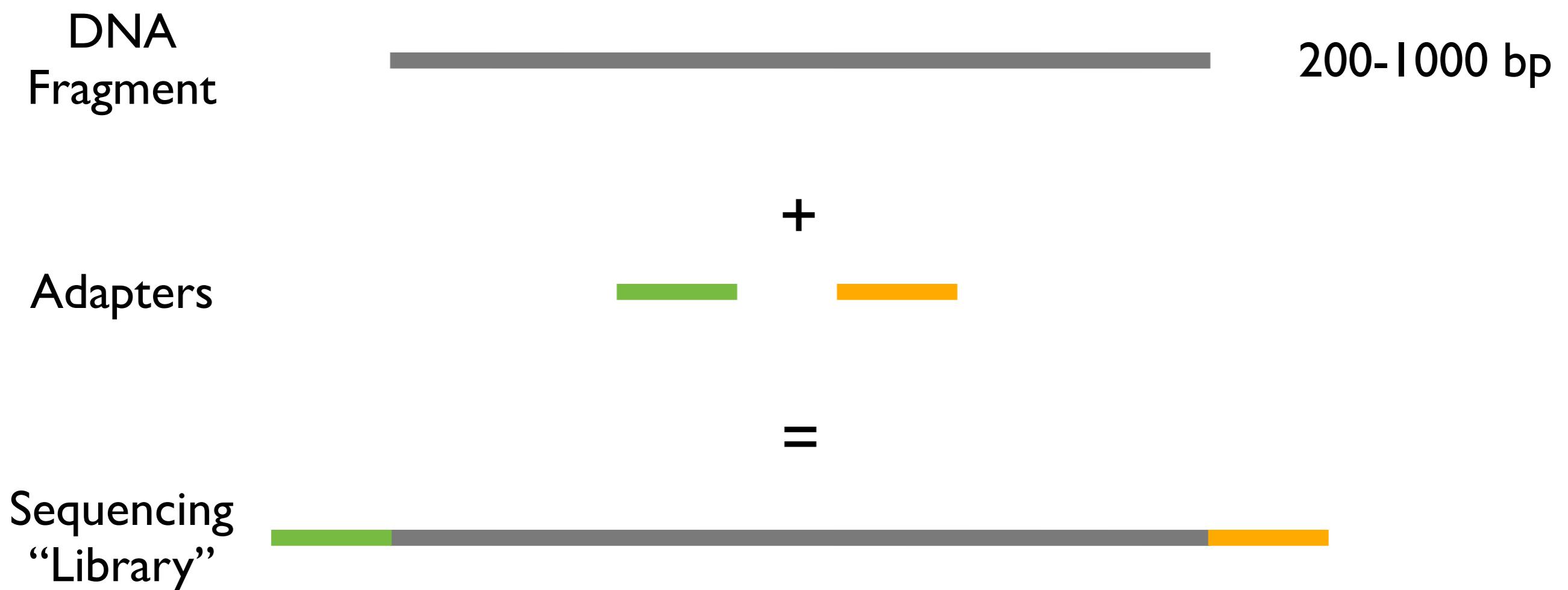
- Selection for desired RNA
 - poly(A) mRNA enrichment
 - Selective polyadenylation of mRNAs
 - Antibody capture of RNAs that interact with a specific protein
 - Non-random priming
- Selection against non-desired RNA
 - Ribosomal RNA capture
 - Duplex-specific nuclease (DSN) normalization
 - Degradation of processed RNA

Major Experiment Components

- 1.Growth
- 2.Sample Collection
- 3.RNA Extraction
- 4.mRNA Enrichment
- 5.Library Preparation



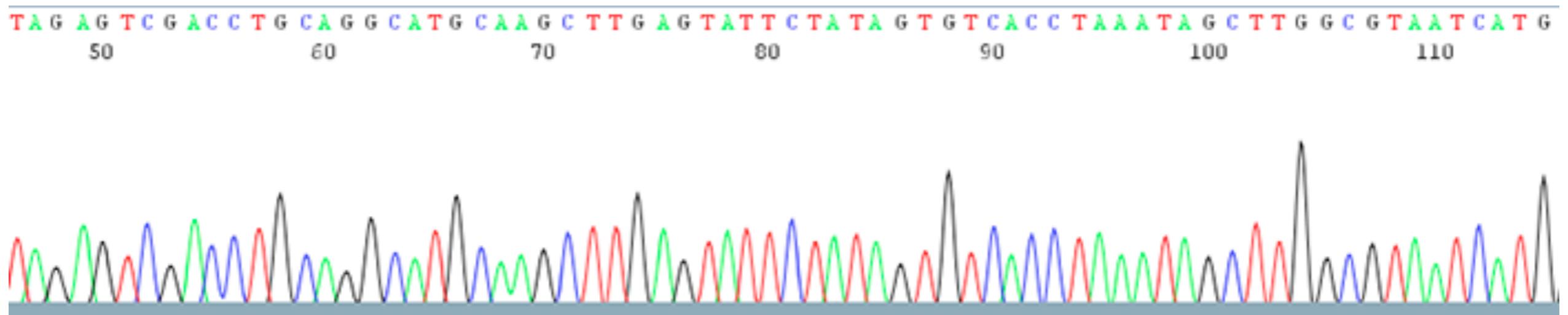
Library Preparation



The End!

Just Kidding!

Dye-terminator Sanger Sequencing



Sequencing

AGCTTTCAATTGACTGCAACGGGCAATATGTCTGTGTGGA

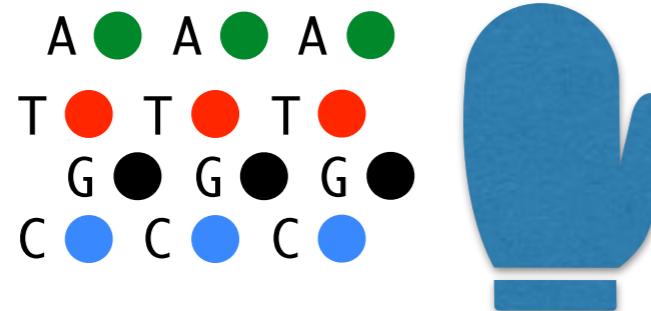
Sequencing



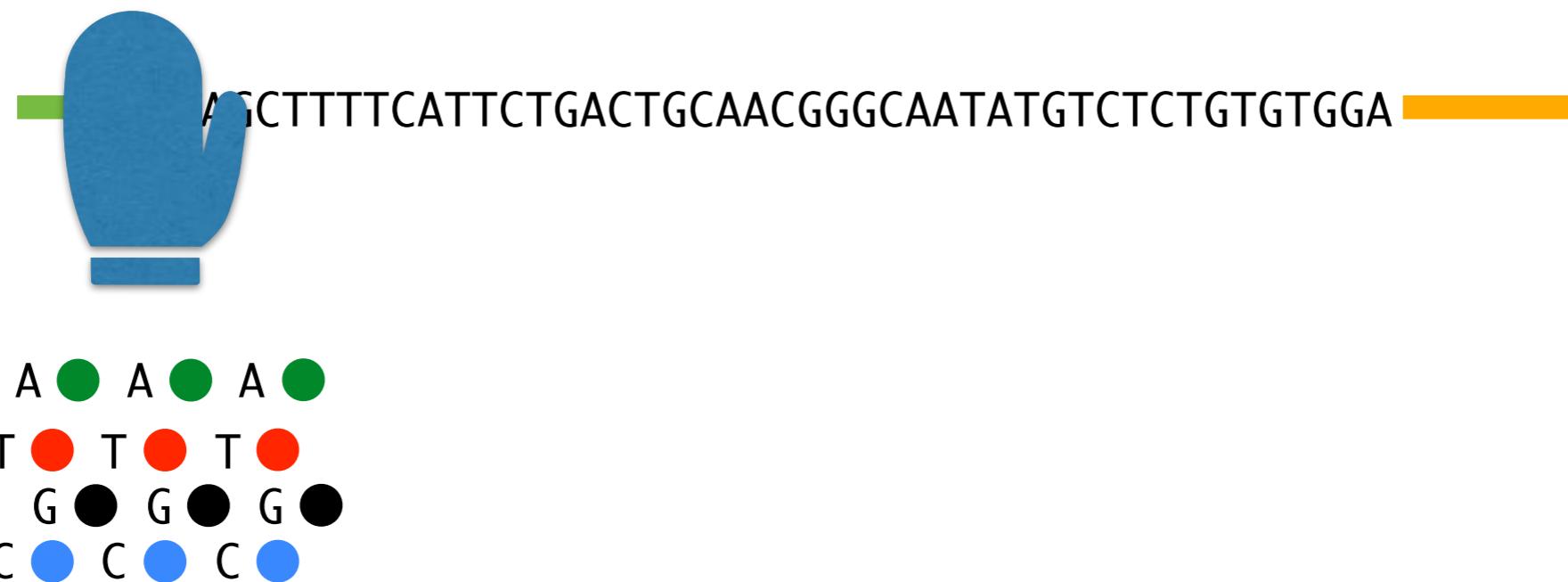
AGCTTTCAATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

Sequencing

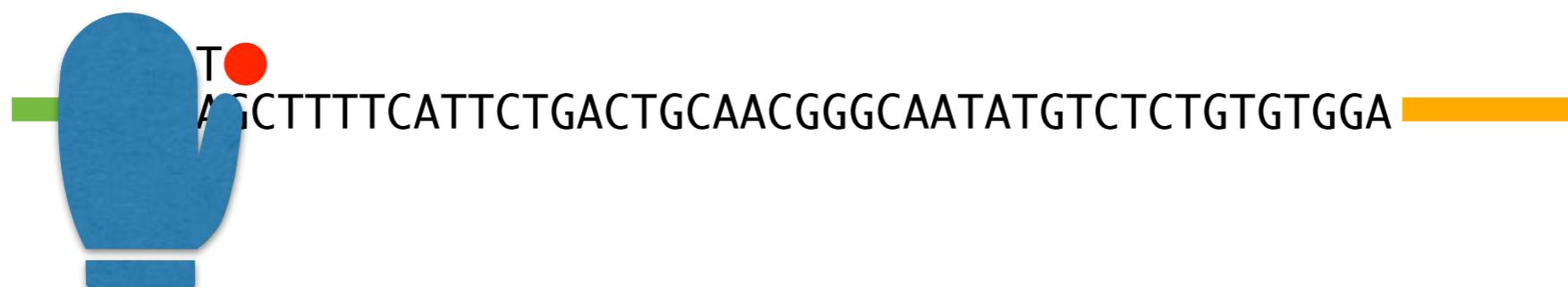
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA



Sequencing



Sequencing



A ● A ● A ●
T ● T ● T ●
G ● G ● G ●
C ● C ● C ●

Sequencing

T
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTGTGTGGA

Sequencing

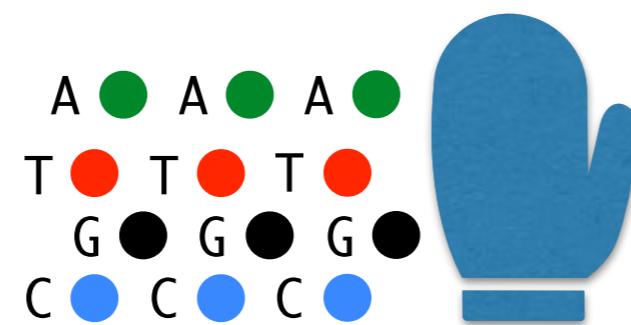
T
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTGTGTGGA

Sequencing

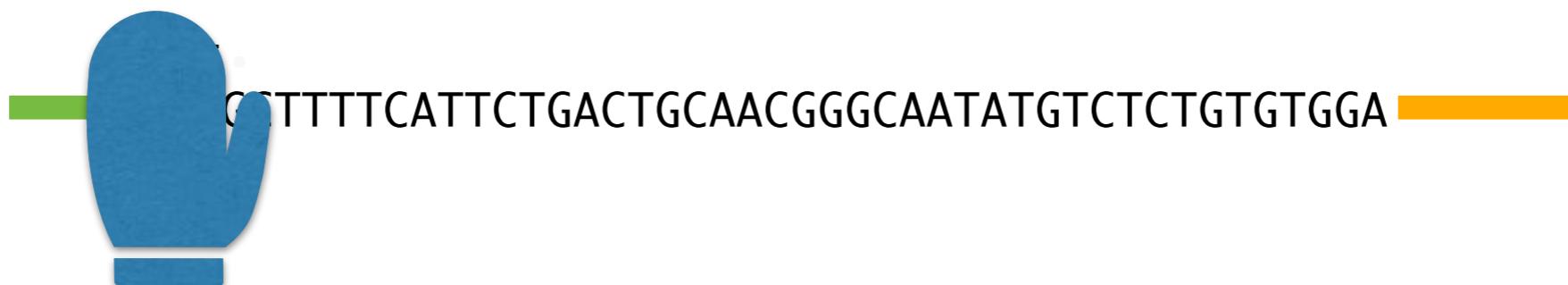
T
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

Sequencing

T
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

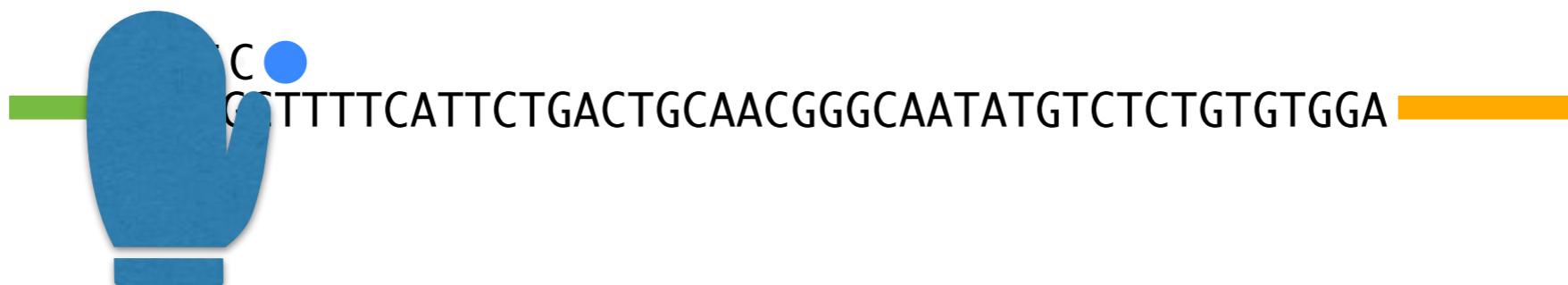


Sequencing



A ● A ● A ●
T ● T ● T ●
G ● G ● G ●
C ● C ● C ●

Sequencing



A ● A ● A ●
T ● T ● T ●
G ● G ● G ●
C ● C ● C ●

Sequencing



A sequence logo visualization for a DNA sequence. The logo consists of four vertical bars of different heights: a green bar (T), a blue bar (C), a red bar (G), and a black bar (A). Below the logo, the corresponding DNA sequence is written: AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA.

AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

Sequencing



A sequence logo visualization for a DNA sequence. The logo consists of four vertical bars of different heights, each representing a nucleotide: Thymine (T, green), Cytosine (C, blue), Guanine (G, yellow), and Adenine (A, red). Below the logo, the corresponding DNA sequence is written: AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA.

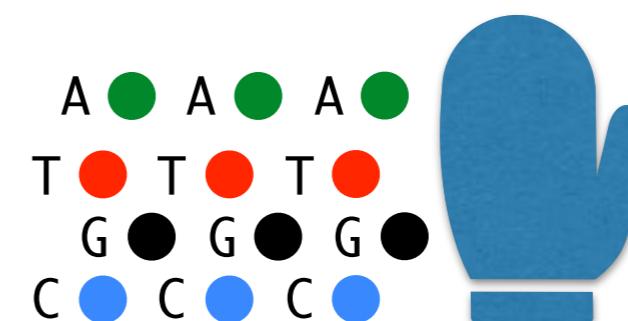
TCGAGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

Sequencing

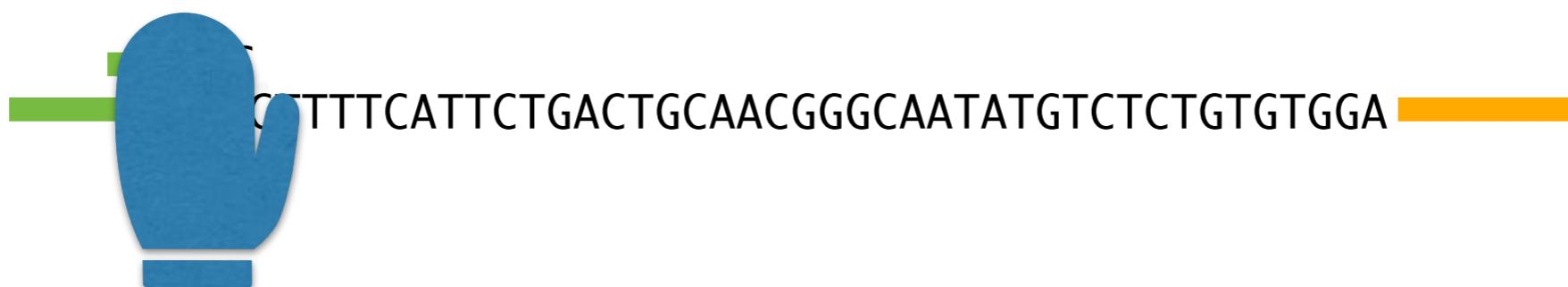
TC
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

Sequencing

TC
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA

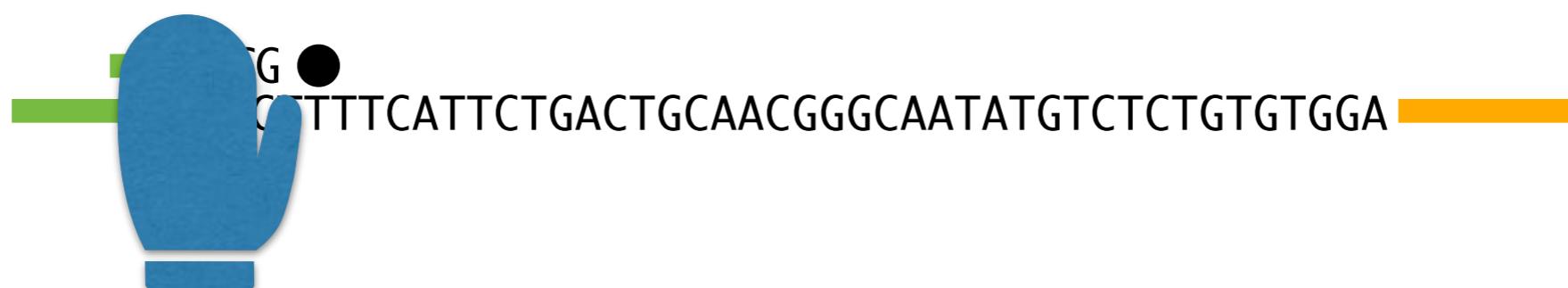


Sequencing



A ● A ● A ●
T ● T ● T ●
G ● G ● G ●
C ● C ● C ●

Sequencing



A ● A ● A ●
T ● T ● T ●
G ● G ● G ●
C ● C ● C ●

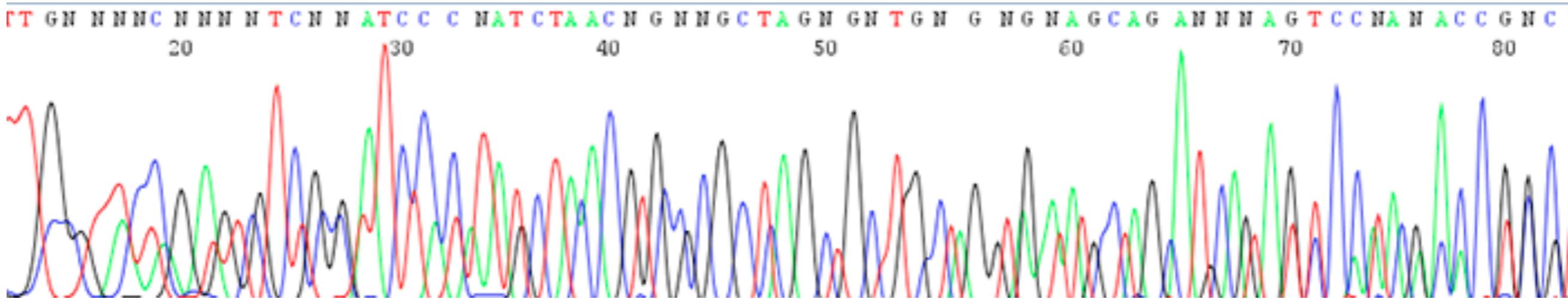
Sequencing

T C G ●
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTGTGTGGA

Sequencing

T C G ●
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTGTGTGGA

Dye-terminator Sanger Sequencing



How?

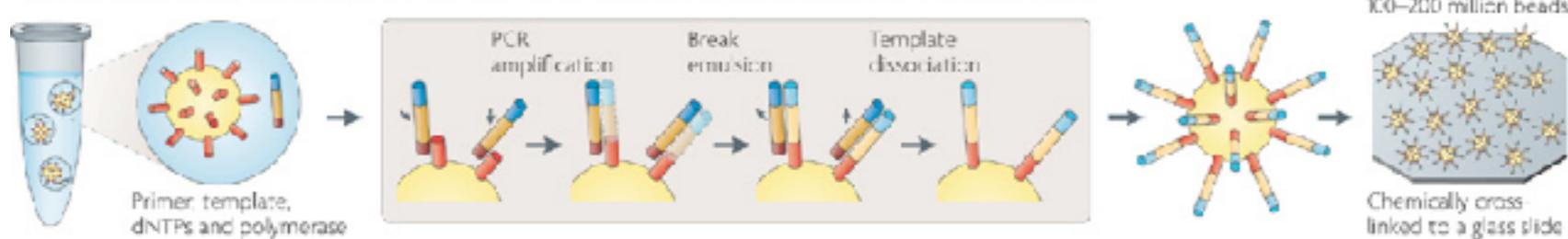
How?

- Separate
- Detect

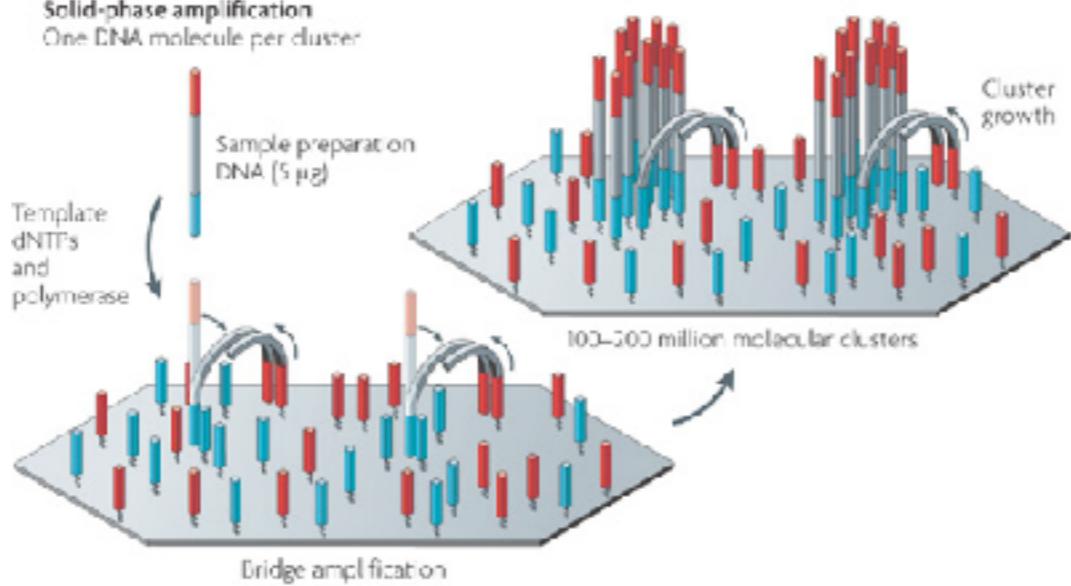
Template immobilization

a Roche/454, Life/APG, Polonator
Emulsion PCR

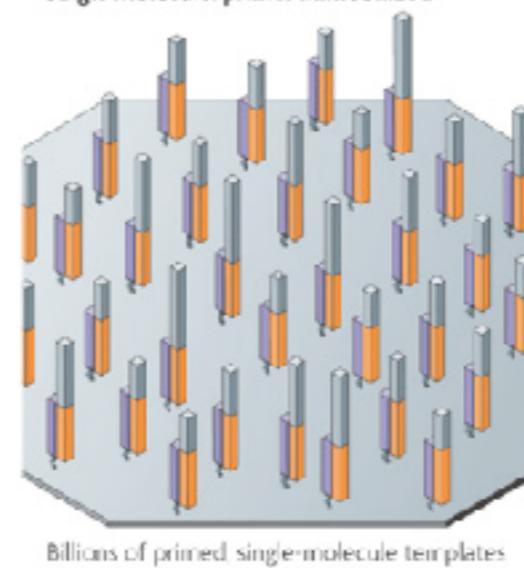
One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion



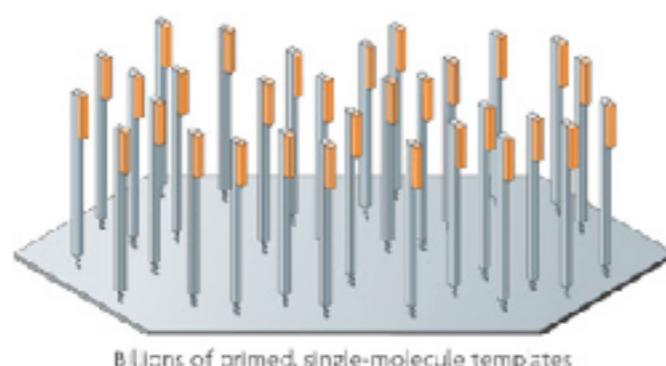
b Illumina/Solexa
Solid-phase amplification
One DNA molecule per cluster



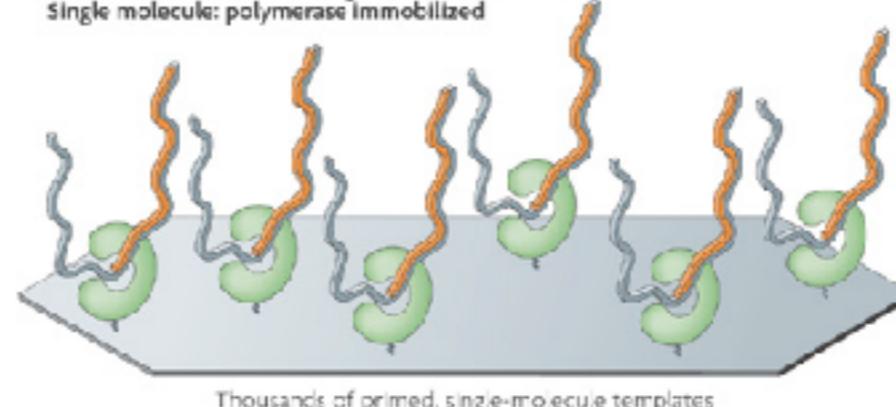
c Helicos BioSciences: one-pass sequencing
Single molecule: primer immobilized



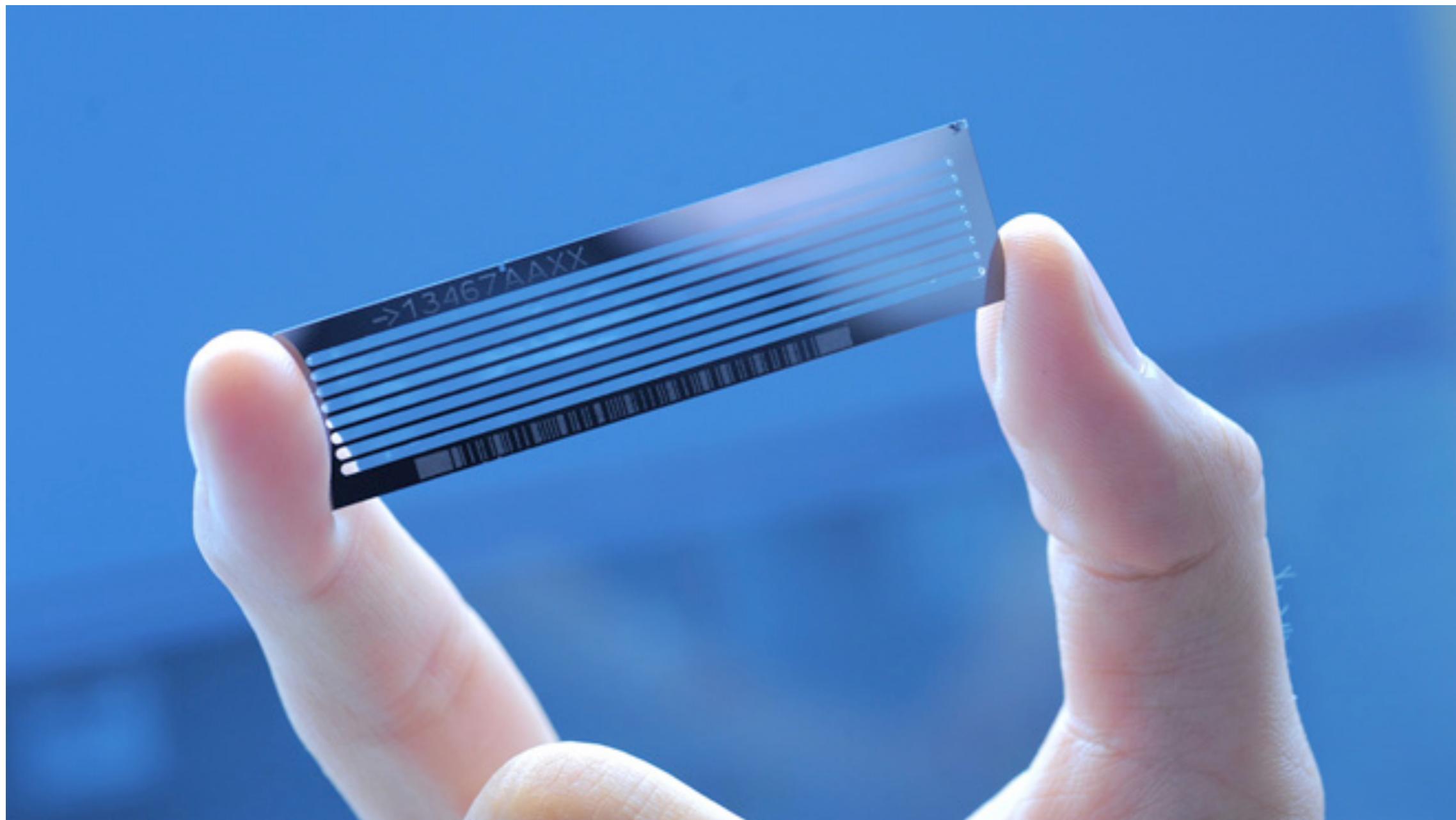
d Helicos BioSciences: two-pass sequencing
Single molecule: template immobilized



e Pacific Biosciences, Life/Visigen, LI-COR Biosciences
Single molecule: polymerase immobilized

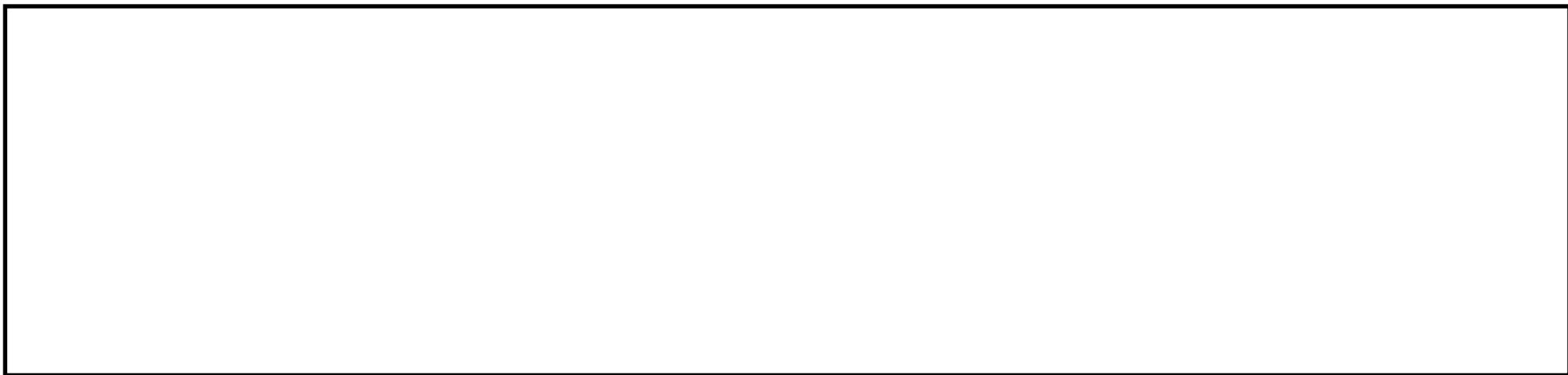


A Flow Cell



**Pass Around Flow
Cells!!!**

A Flow Cell



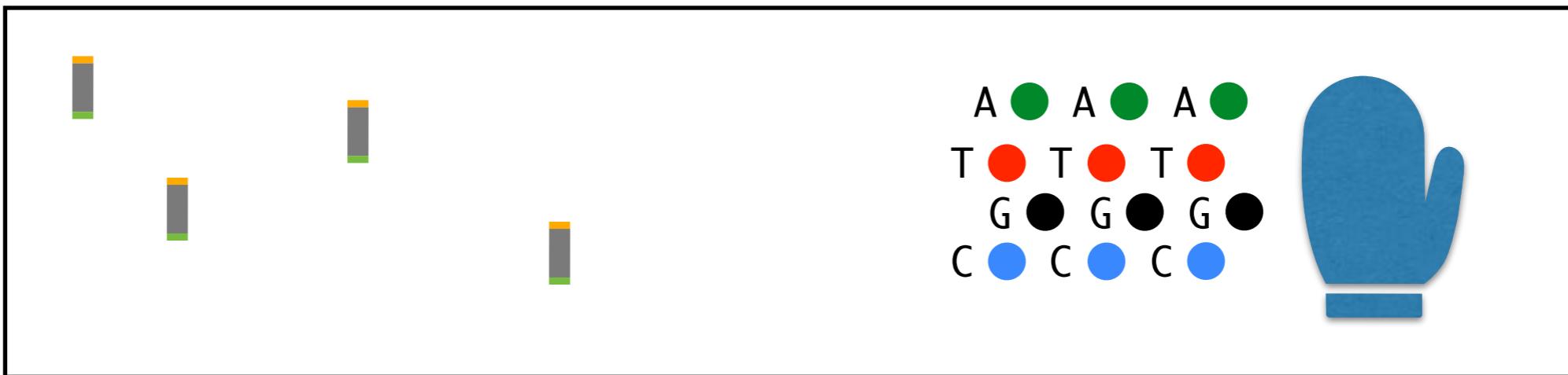
SBS: Sequencing by Synthesis

An Illumina Story

Bind Library



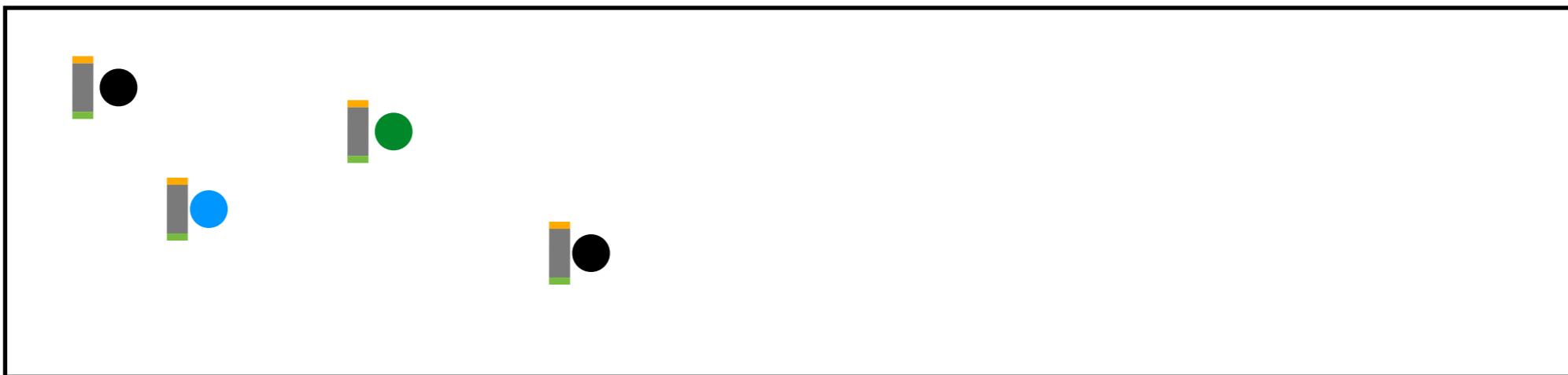
1st Cycle



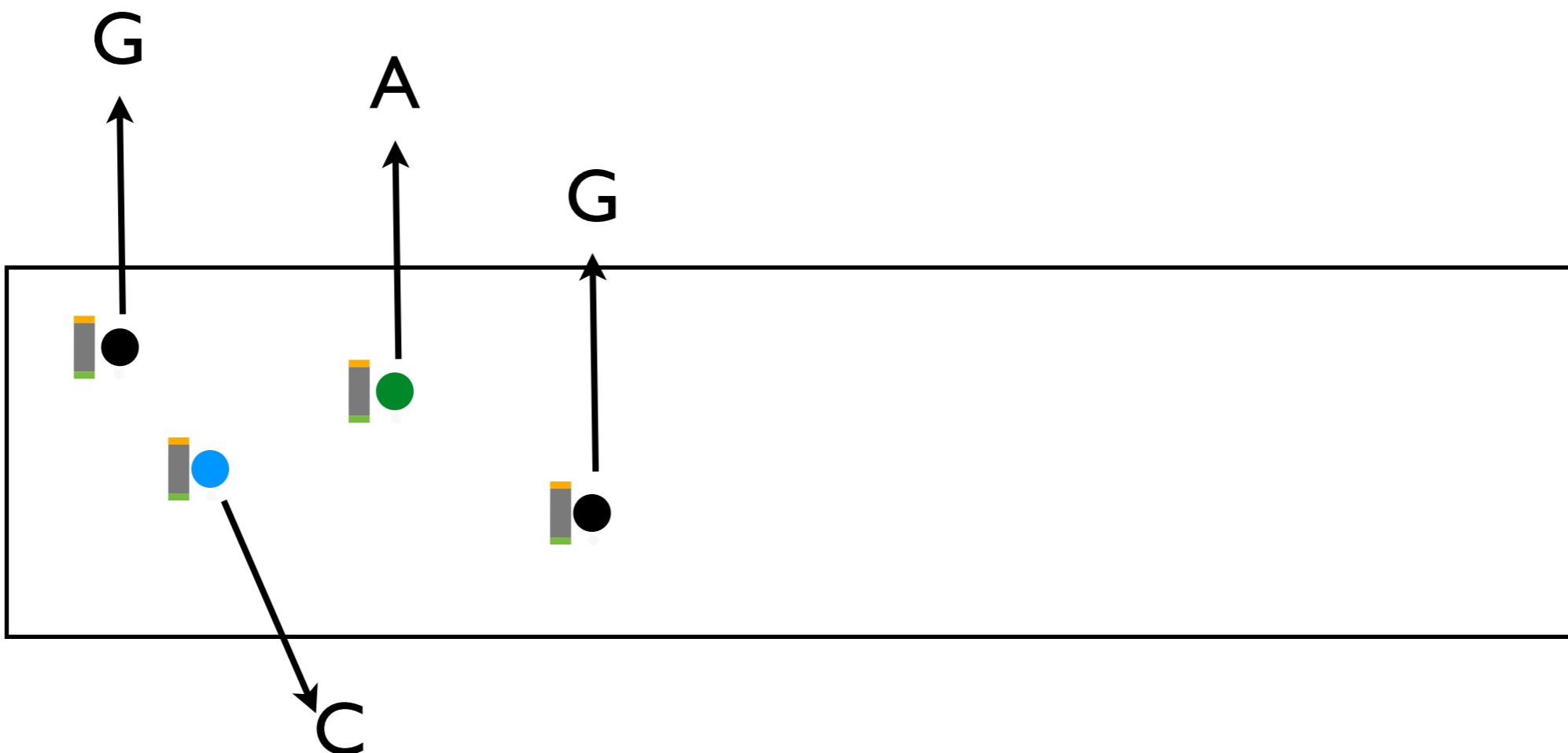
1st Cycle



1st Cycle



1st Cycle



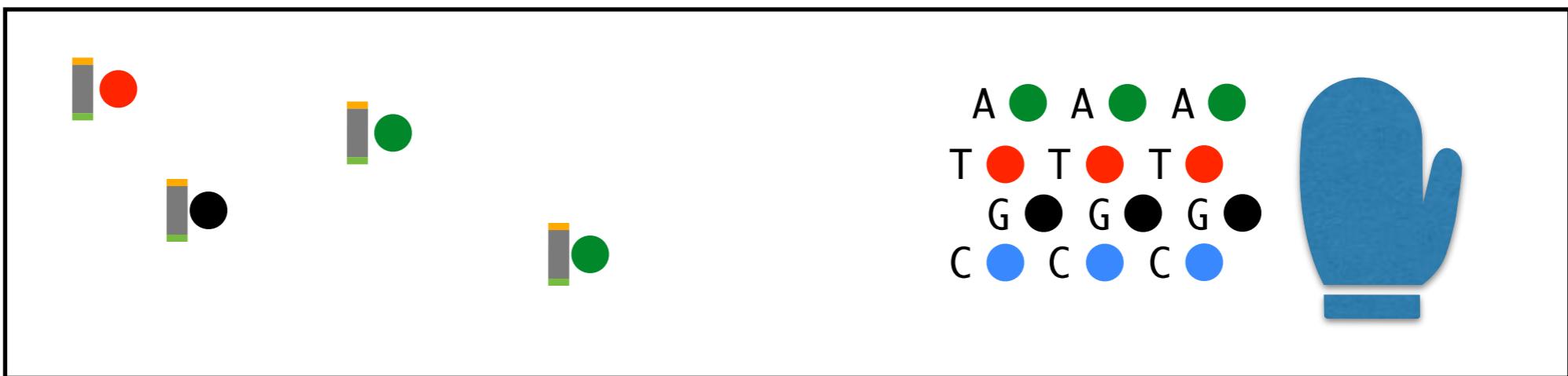
2nd Cycle



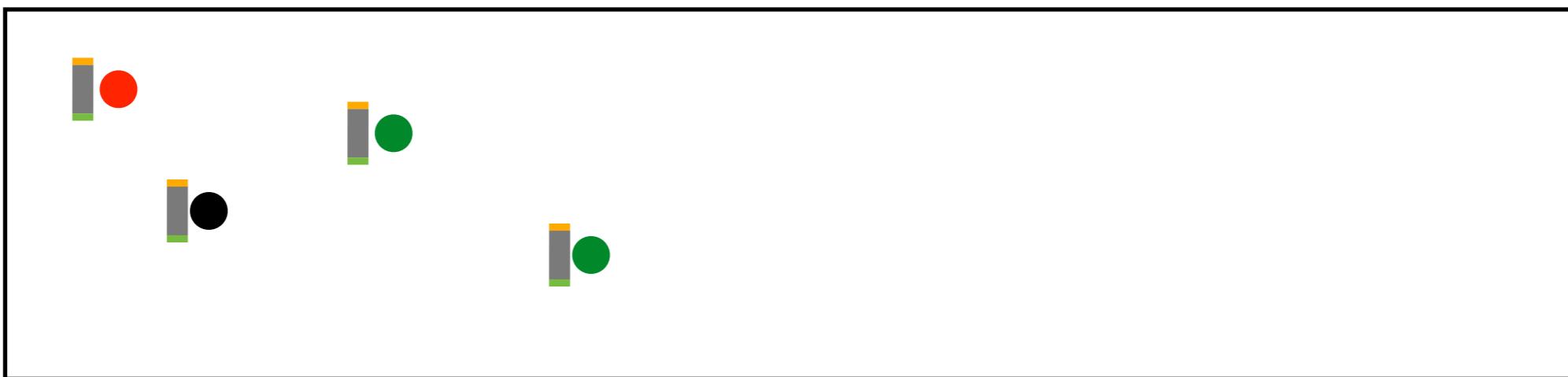
2nd Cycle



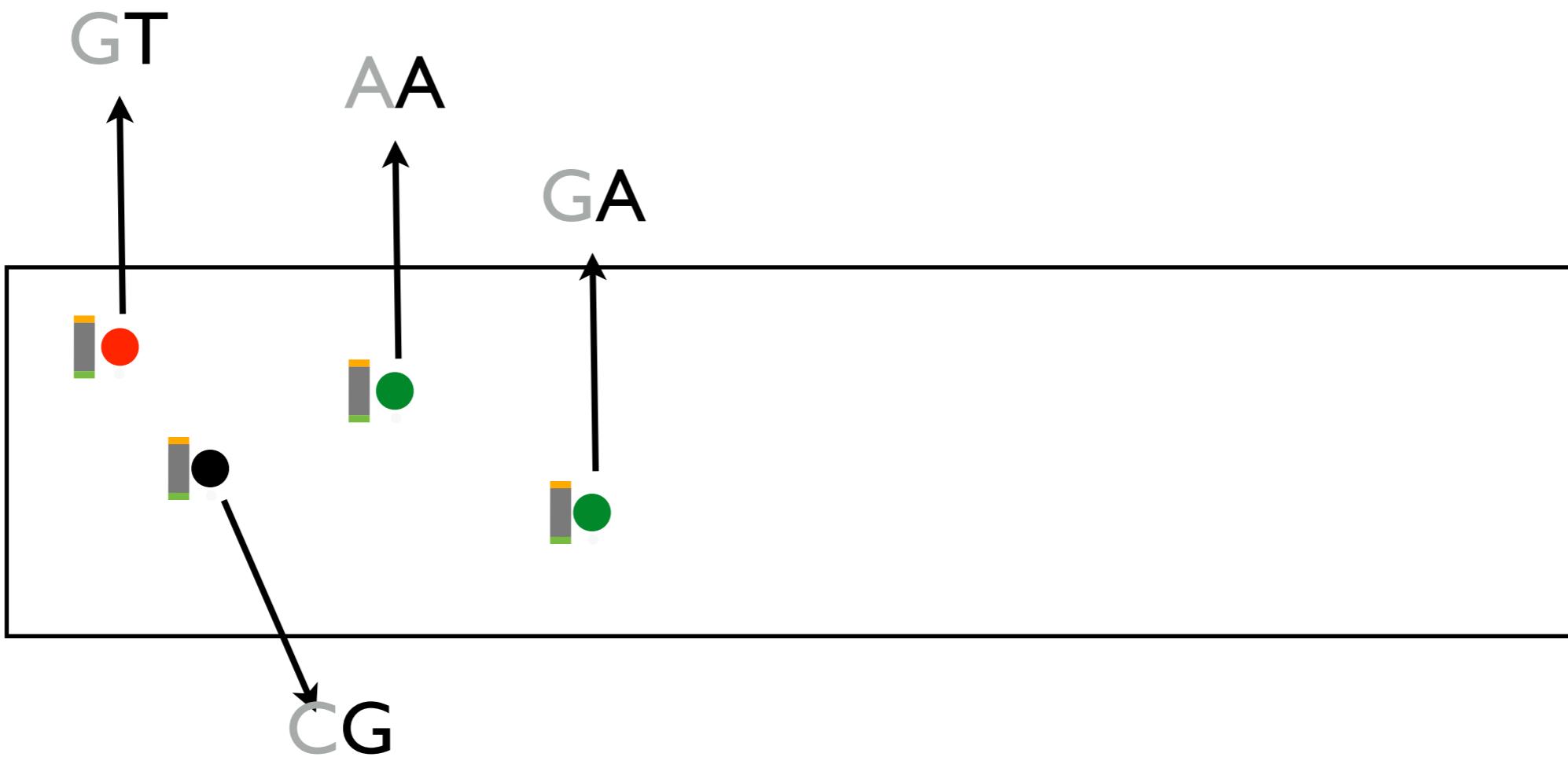
2nd Cycle



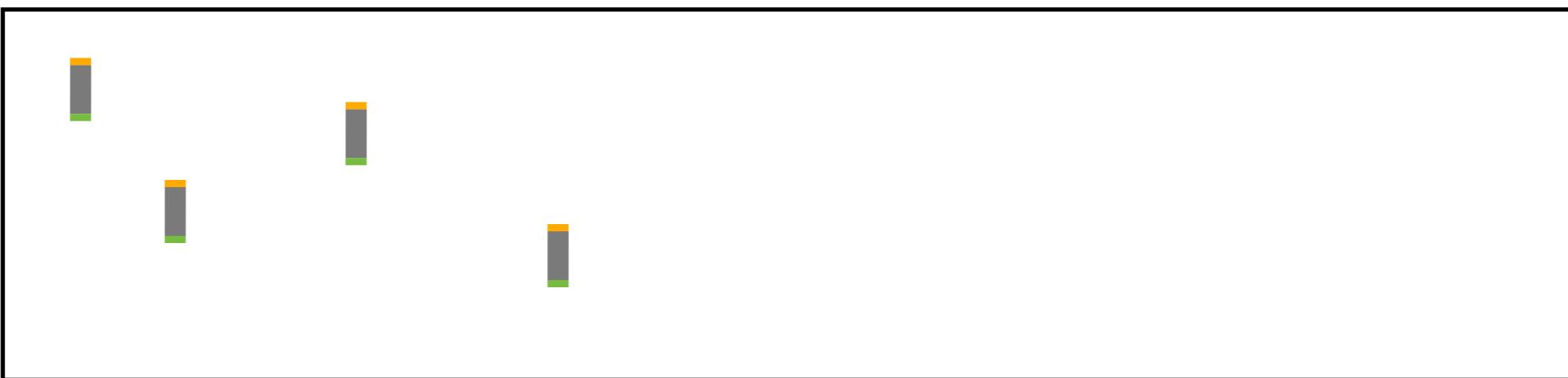
2nd Cycle



2nd Cycle



3rd Cycle



3rd Cycle



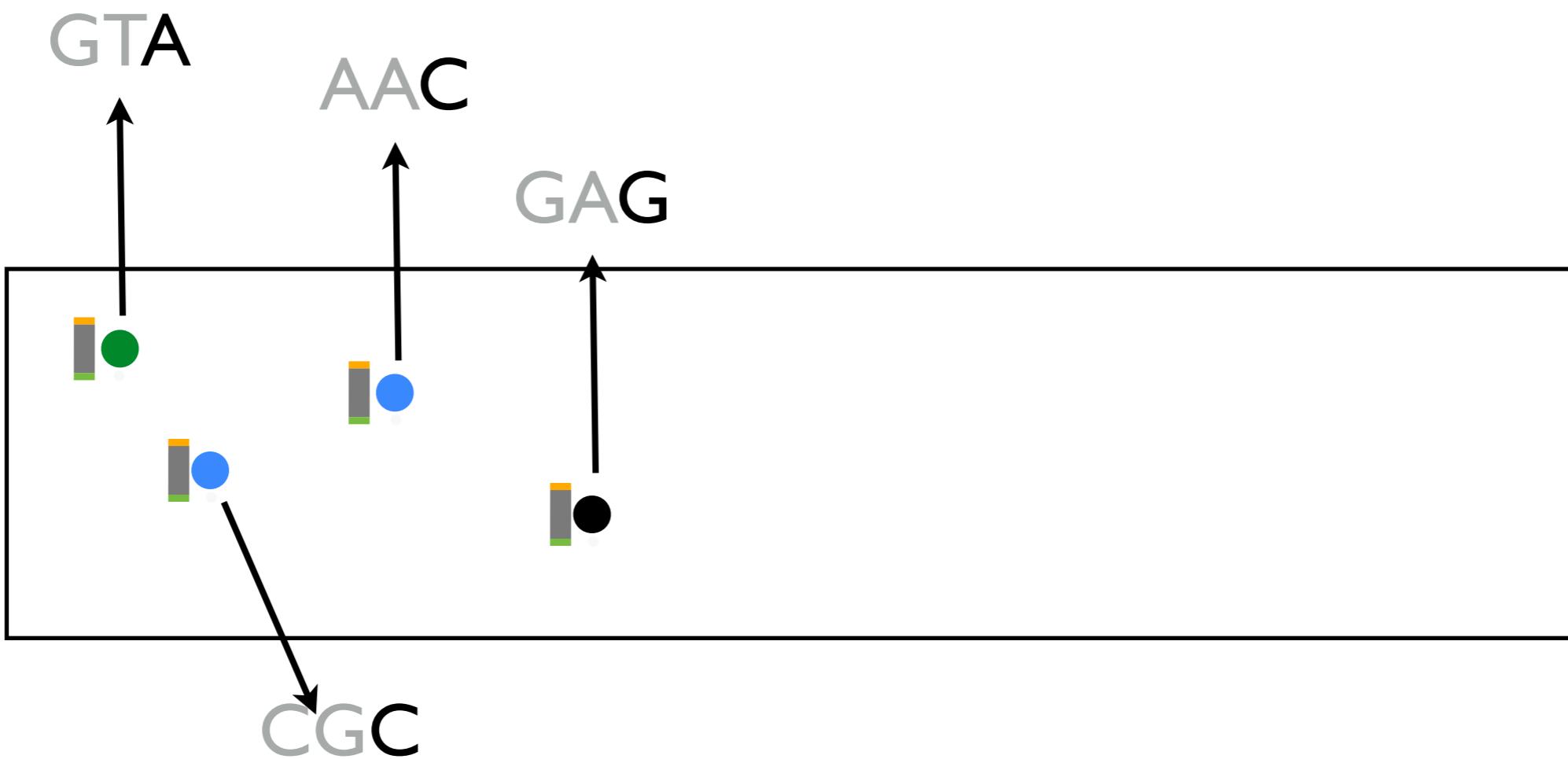
3rd Cycle



3rd Cycle



3rd Cycle

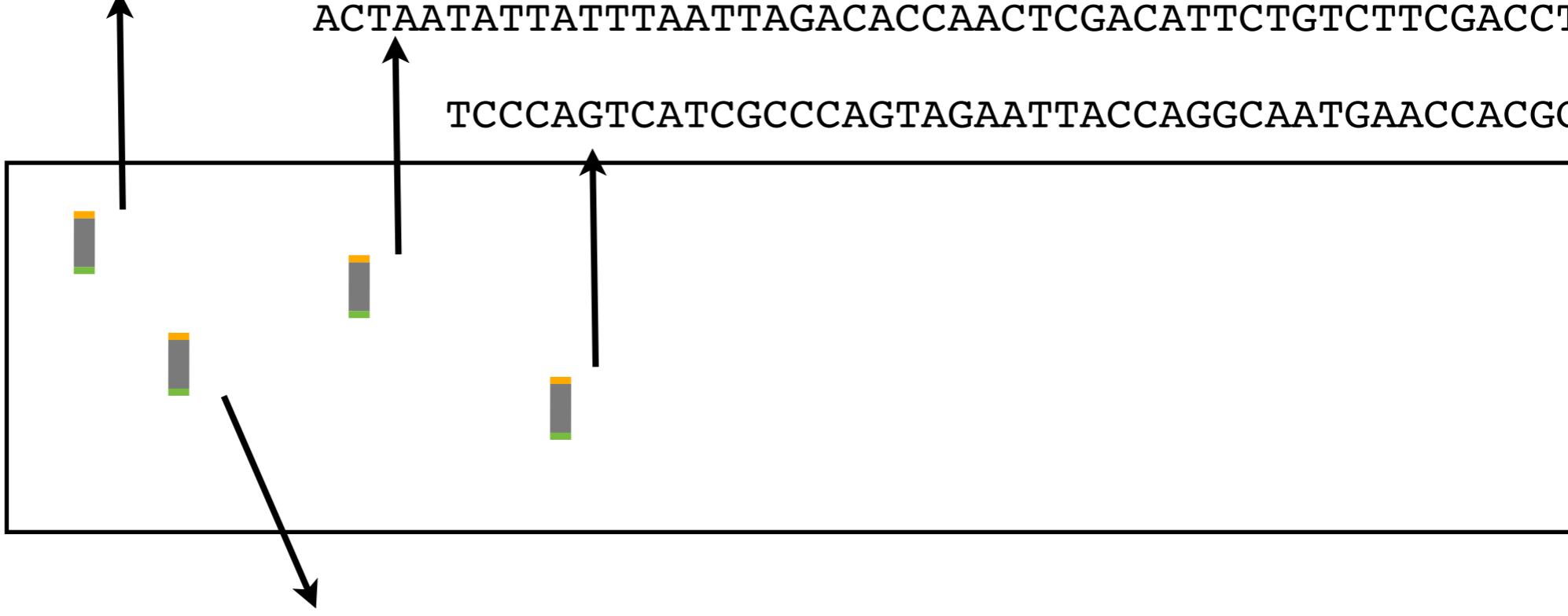


50th Cycle

GAATTCTAAAACAGTTGCATTCTATAATTACAAAATAATTGAAACACTTC

ACTAATATTATTAAATTAGACACCAACTCGACATTCTGTCTTCGACCTAT

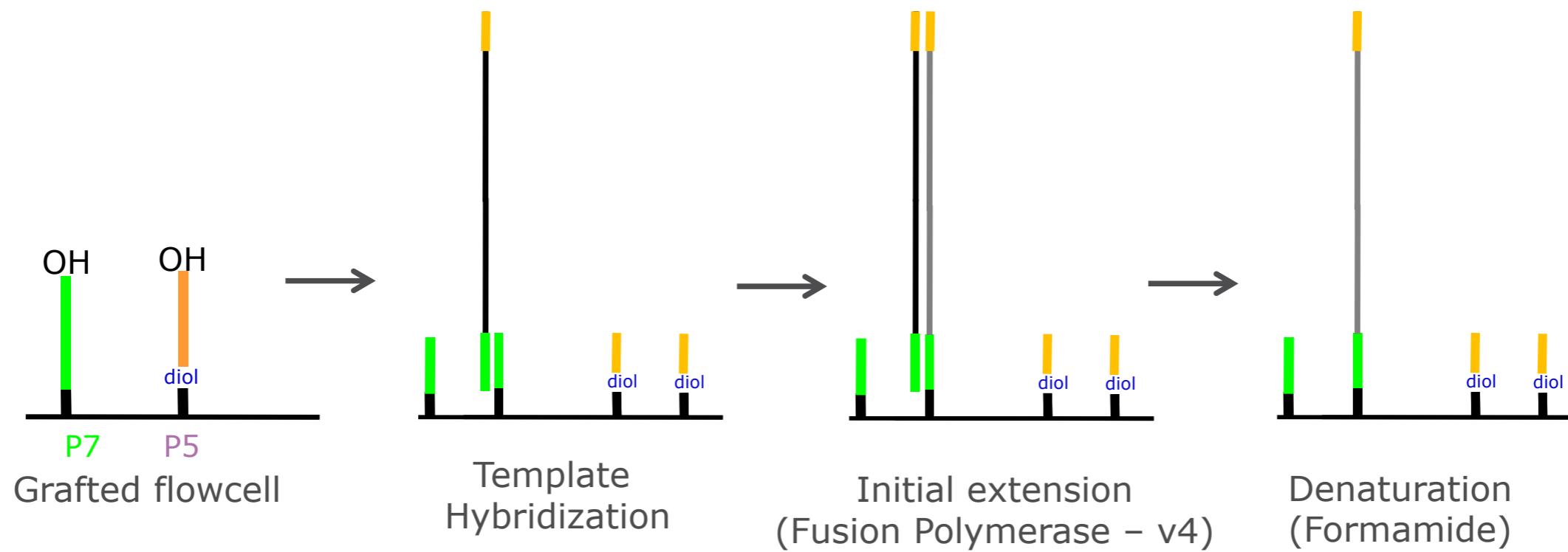
TCCCAGTCATCGCCCCAGTAGAATTACCAGGCAATGAACCACGGCCTTCA



Illumina Short Reads

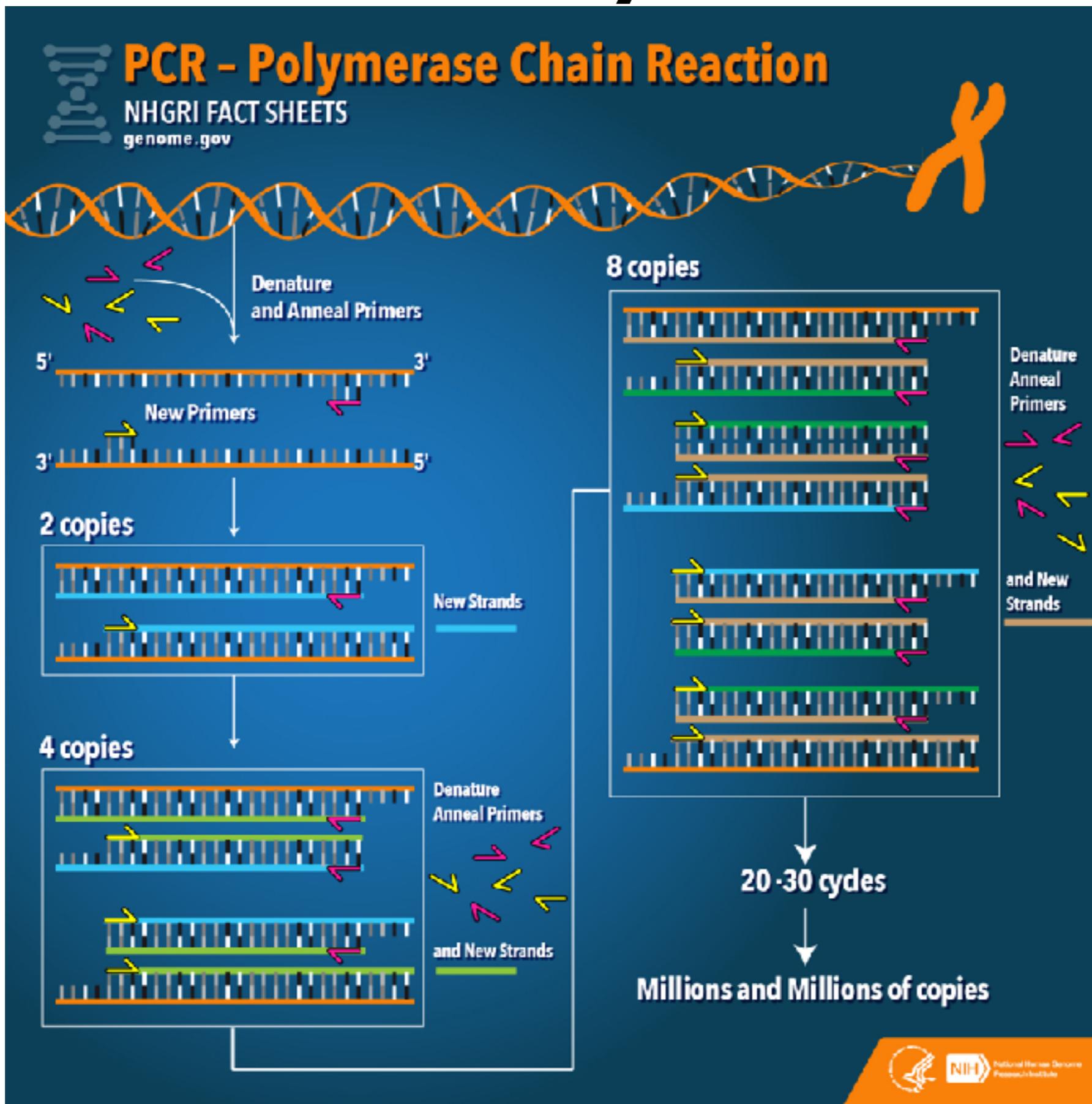
- 50 - 300bp

Cluster generation – hybridization and amplification

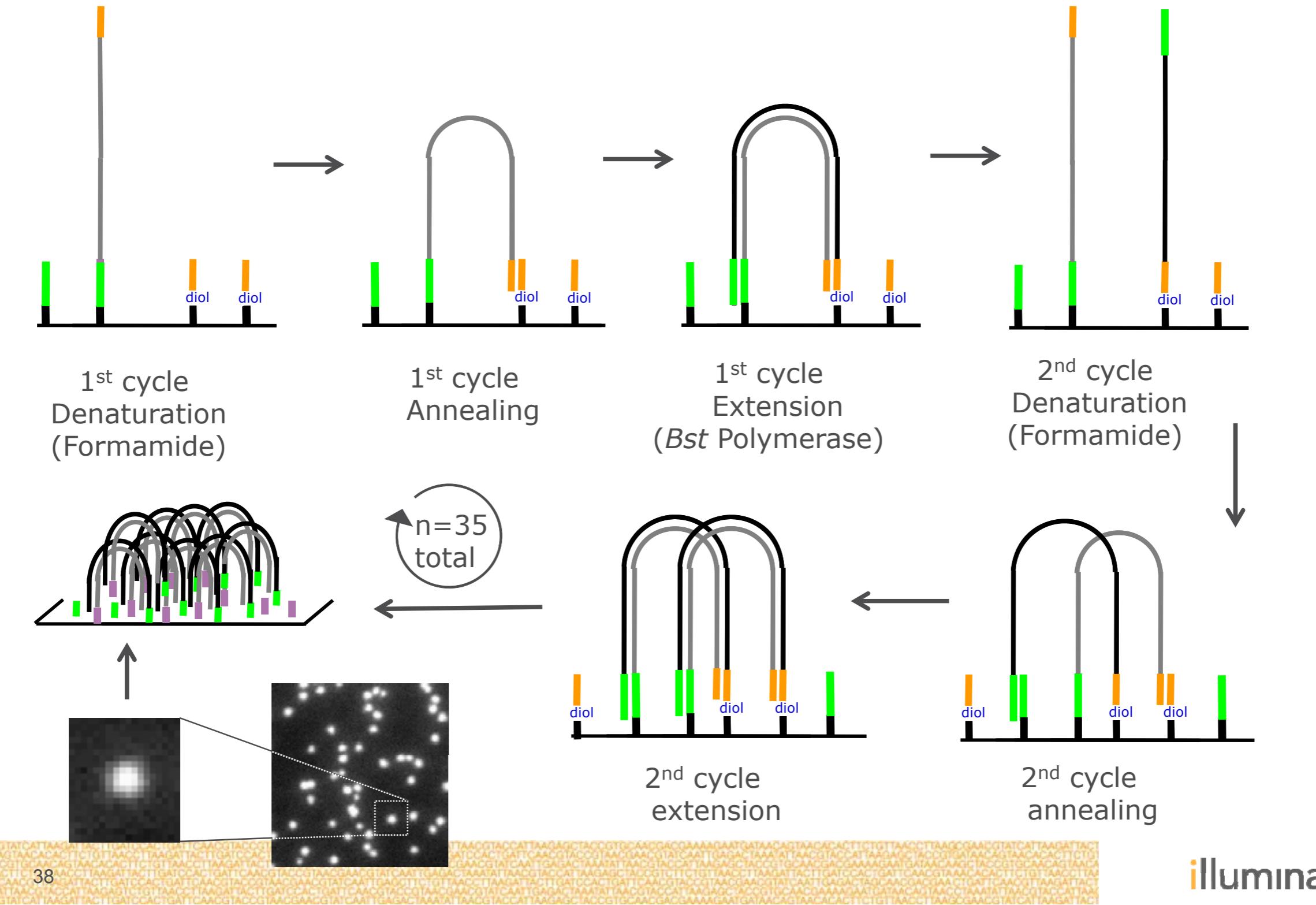


SBS: More Detail

PCR Anyone?



Cluster generation – hybridization and amplification



Library Preparation: Why?

We have RNA. We need a DNA library.



Library Preparation: How?

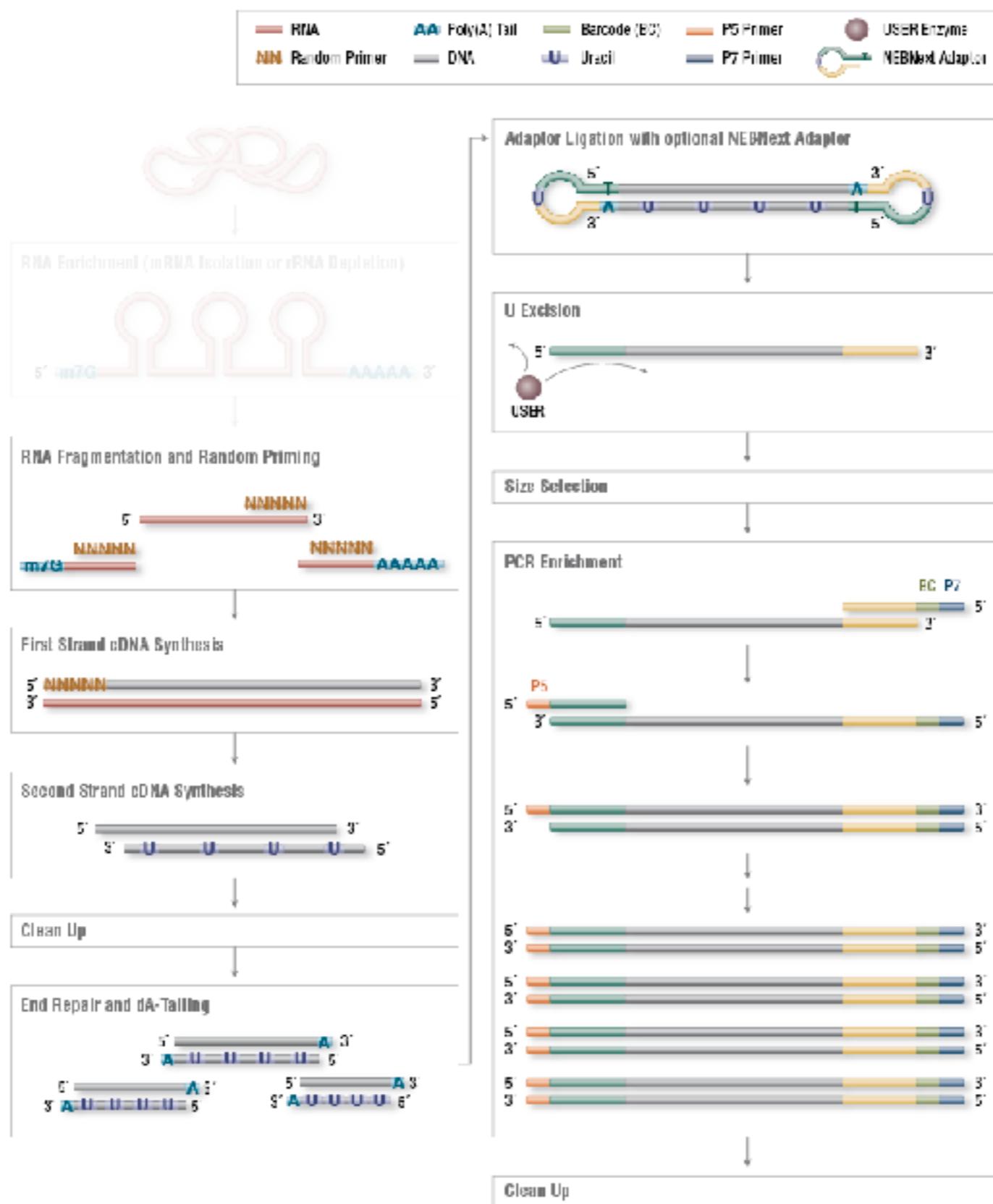
NEBNext® Ultra™ II Directional RNA
Library Prep with Sample Purification Beads

Catalog #: E7765

Library Preparation: How?

- 1.RNA Fragmentation
- 2.cDNA Synthesis
- 3.Adapter Ligation
- 4.PCR Enrichment

Library Prep Workflow



Fragmentation



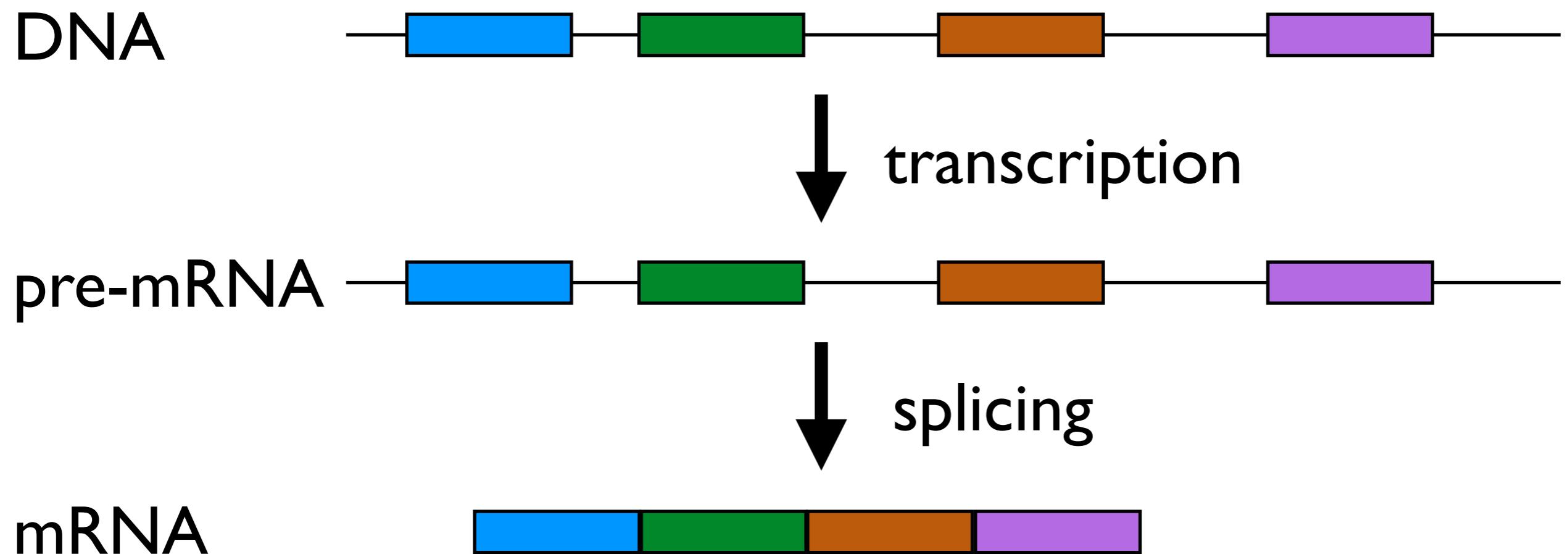
Library Prep

Fragmentation: Why?

- Efficient cluster generation and sequencing
- Distribution of reads across mRNA

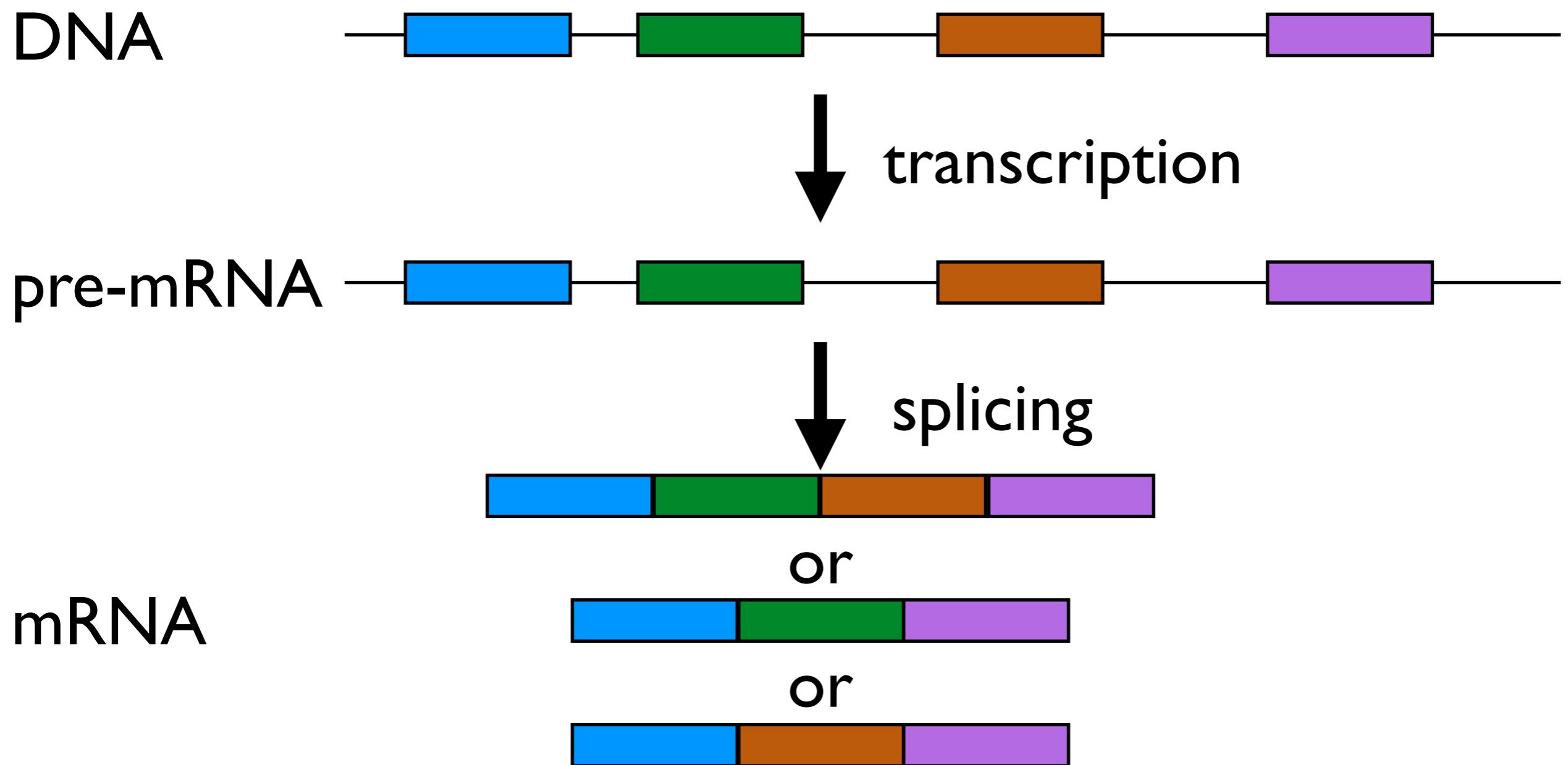
Differential Splicing

Splicing*



* Does not occur in bacteria

Differential Splicing*

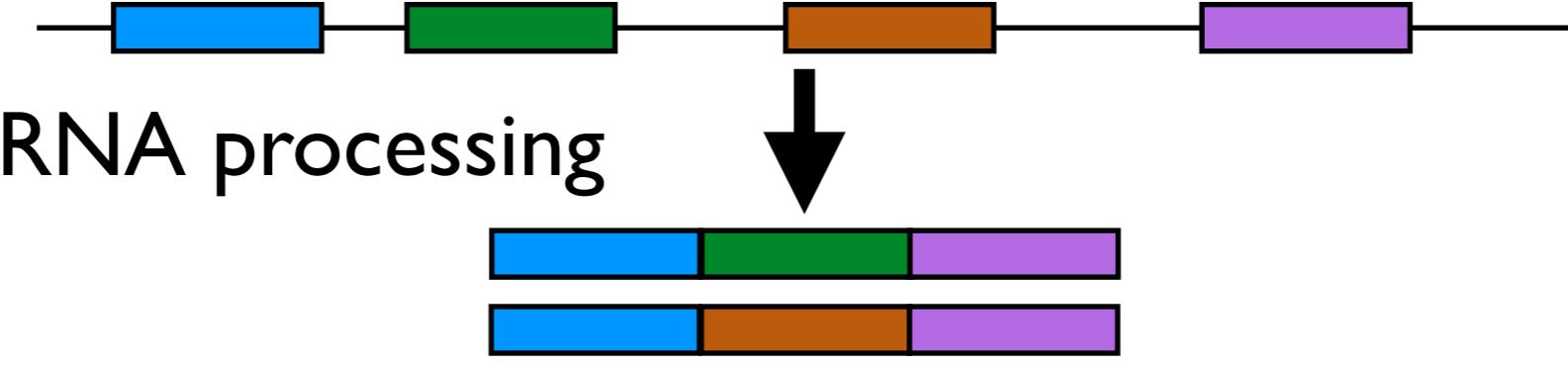


* Does not occur in bacteria

Library Prep

Fragmentation: Why?

pre-mRNA ——————



↓

RNA processing

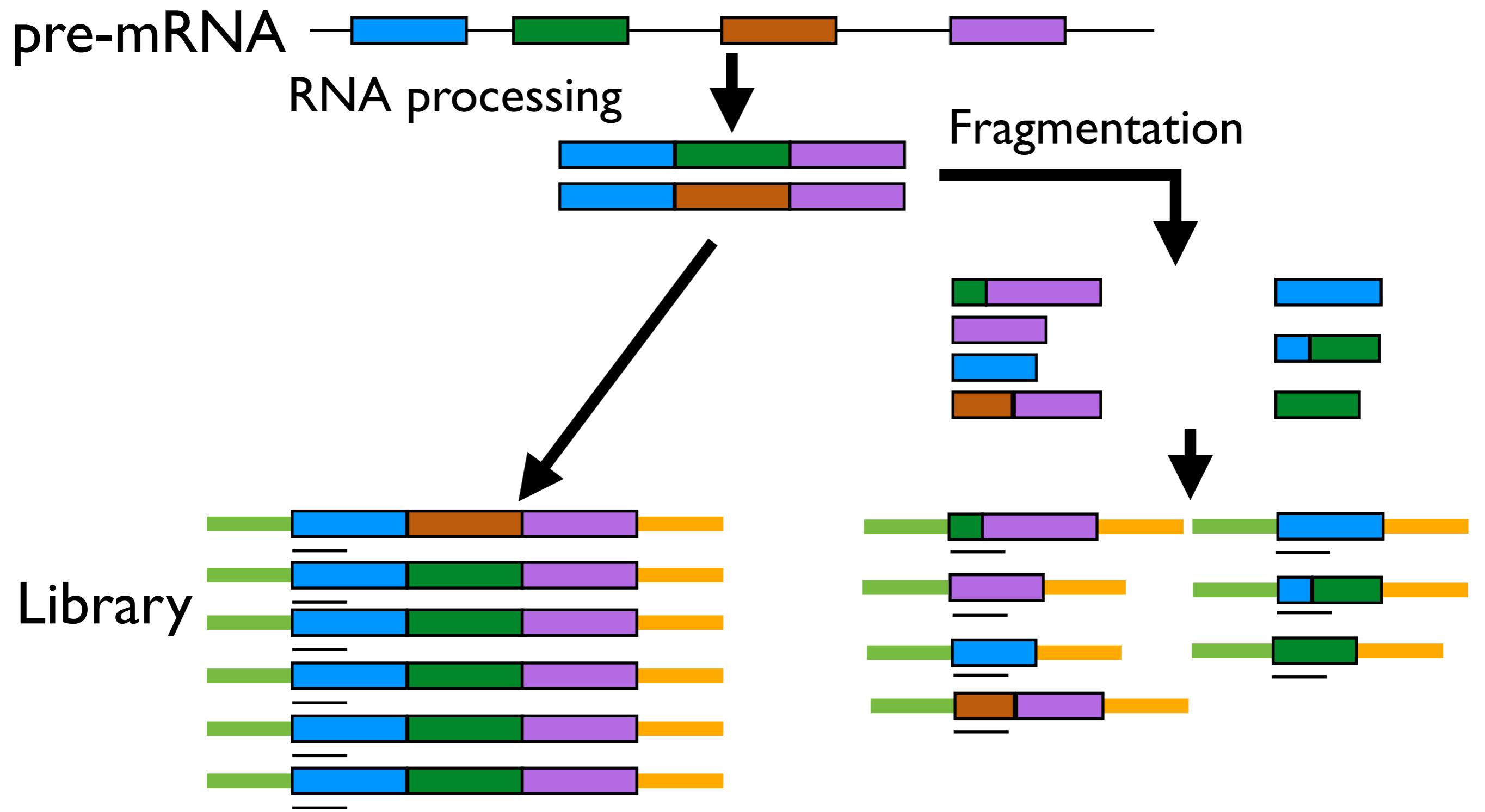


Library



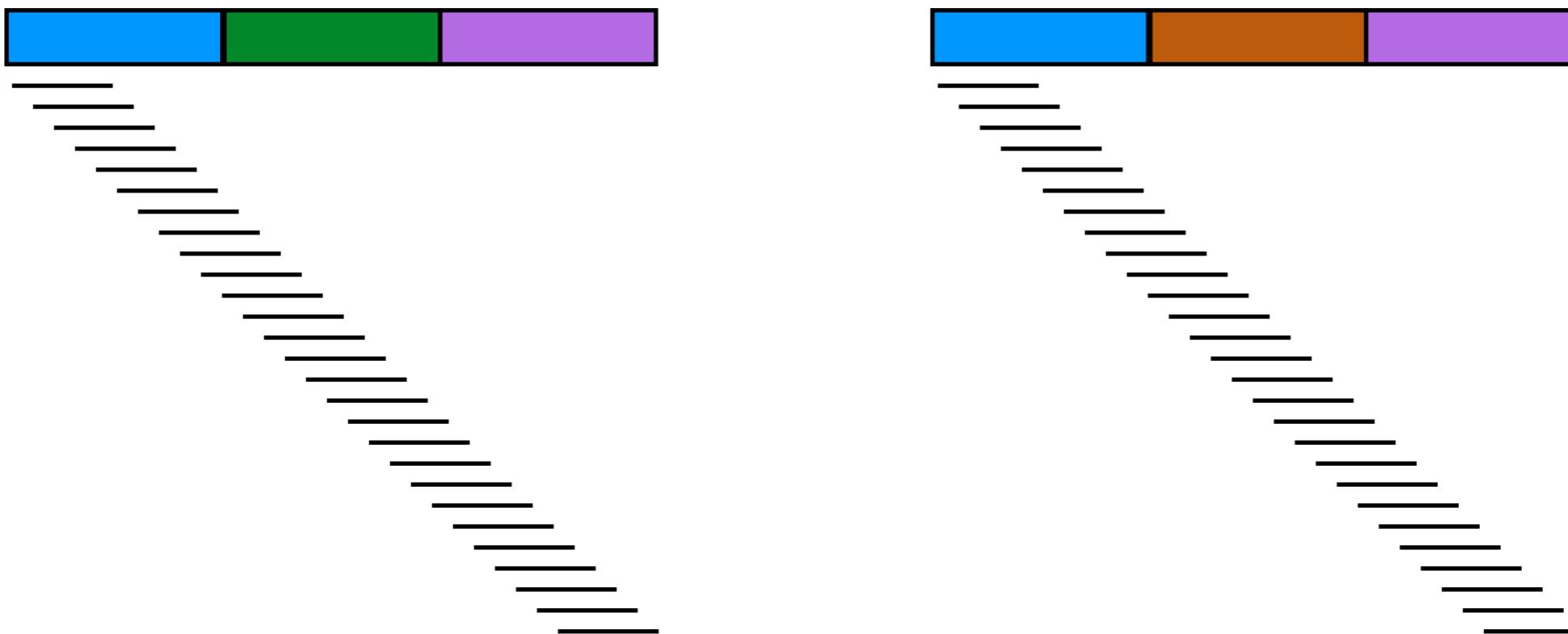
Library Prep

Fragmentation: Why?



Library Prep

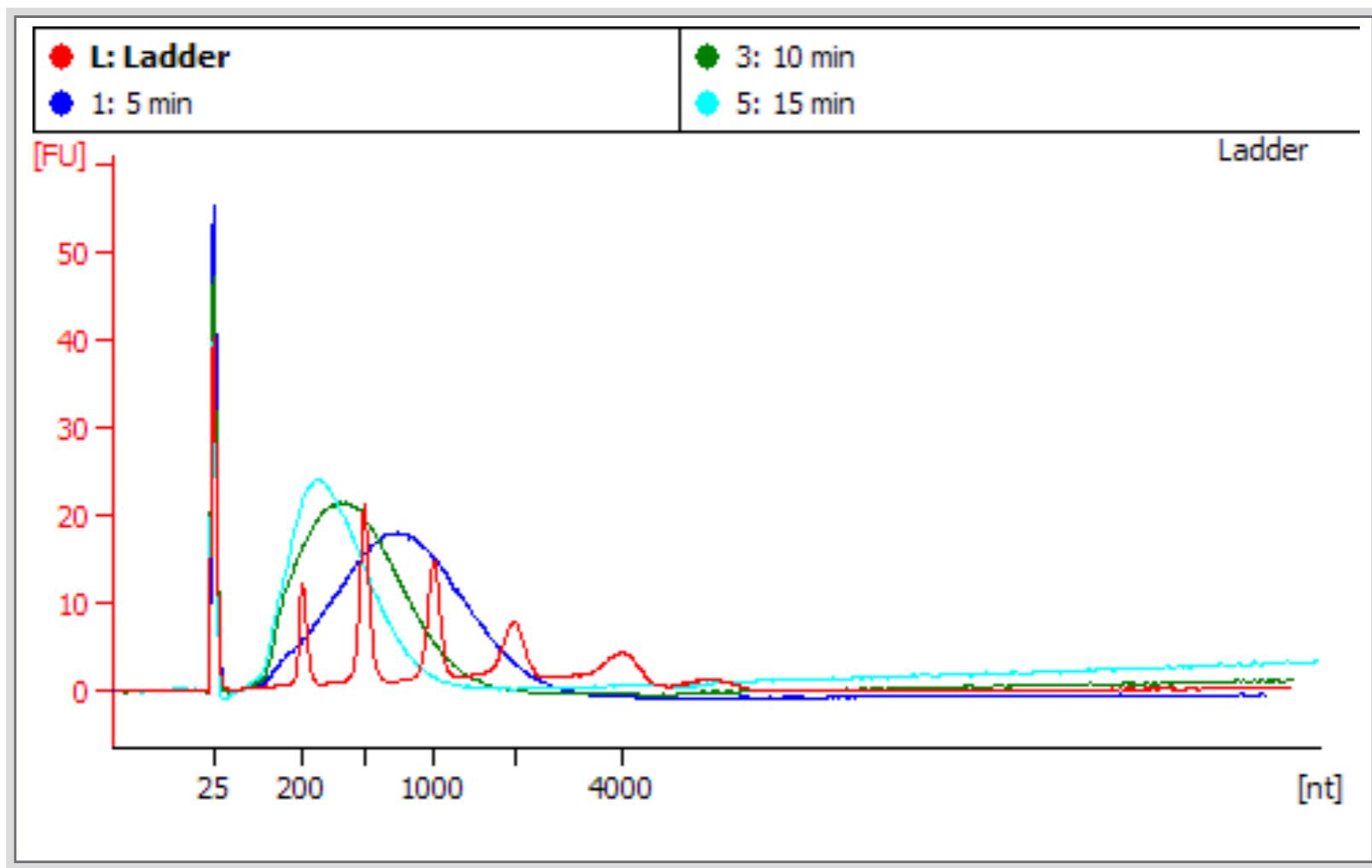
Fragmentation: Why?



Library Prep

Fragmentation: How?

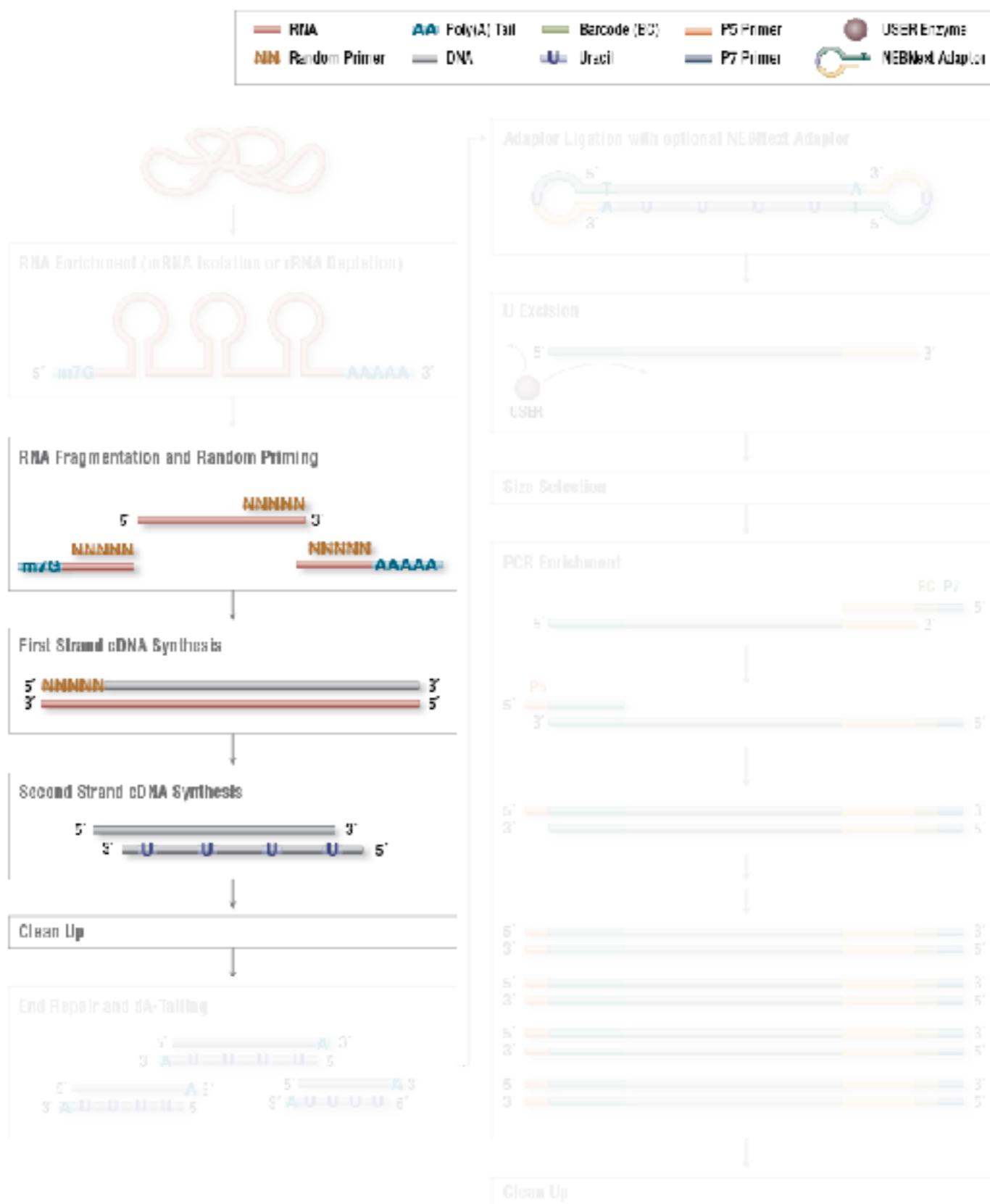
- Heat with divalent metal cation
(Chemical)



Library Prep Fragmentation: Alternatives?

- Degraded RNA
- Small RNAs
- DNA Fragmentation uses Physical or Enzymatic methods
- Needs to be Random!!!

cDNA Synthesis



cDNA Synthesis: Why?

- Have RNA, need DNA

cDNA Synthesis

RNA Fragmentation and Random Priming



First Strand cDNA Synthesis



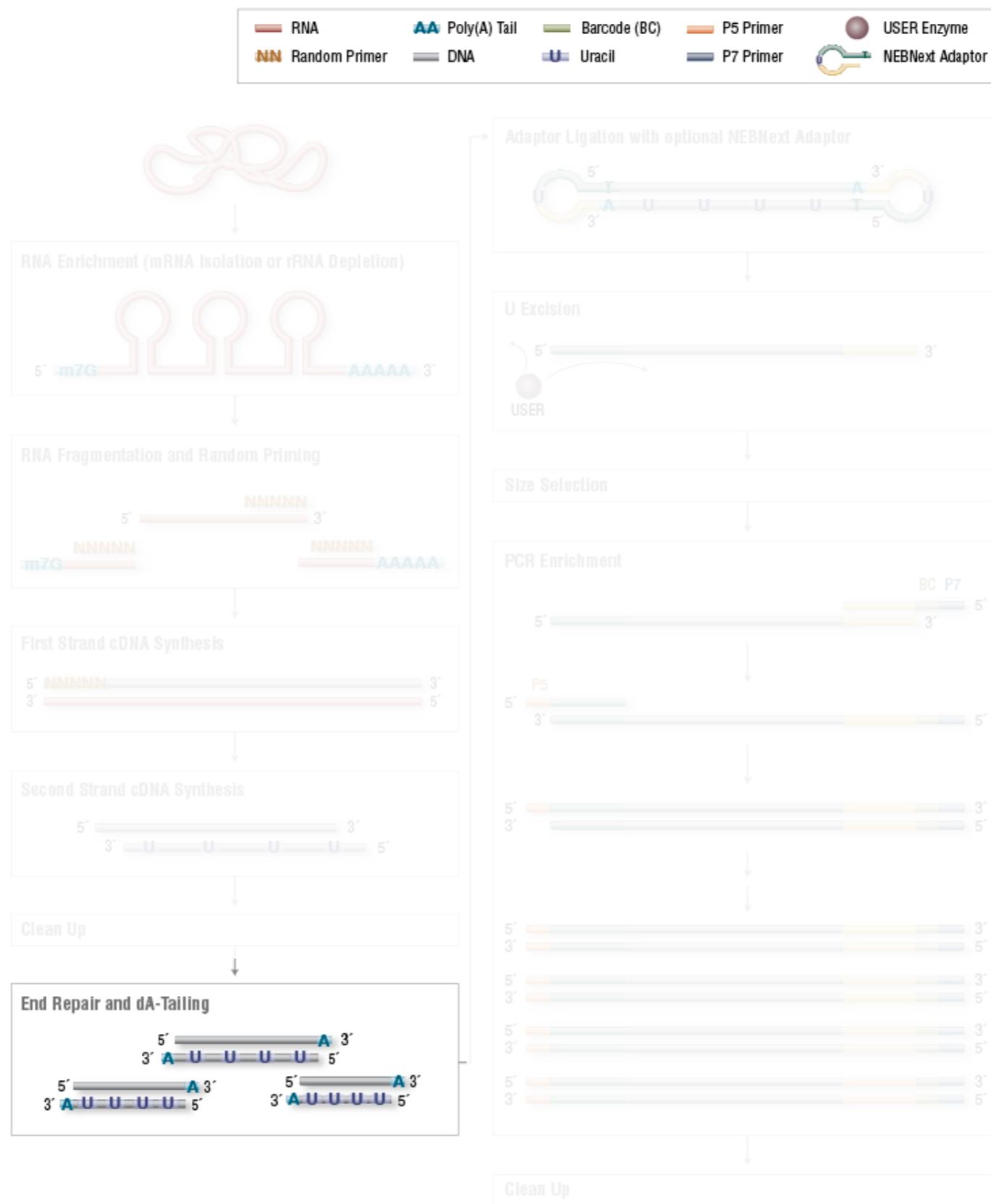
Second Strand cDNA Synthesis



cDNA Synthesis: How?

- First Strand:
 - Reverse Transcriptase
 - Random Primers
 - dNTPs
- Second Strand:
 - RNase H: generate RNA primers
 - DNA polymerase I: DNA synthesis
 - DNA ligase: ligate fragments
 - dNTPs

End-Repair and dA-Tailing



End-Repair and dA-Tailing

Prepare fragments for adapter ligation:

- Generate blunt ends
- Then generate 3' A overhang

DNA Ligation

DNA Ligation: What

Join two or more fragments of DNA into a single

*Do not confuse with hybridization

DNA Ligation: How

- Requirements:
 - two or more DNA fragments
 - DNA Ligase
 - Phosphate/Energy

DNA Ligation: Basics

5' -**CTGATCTGACTGA**-3'

3' -**GACTAGACTGACT**-5'

+

5' -**TGCGTATGCTAGT**-3'

3' -**ACGCATACGATCA**-5'

+

Ligase + ATP

=

5' -**CTGATCTGACTGATGCGTATGCTAGT**-3'

3' -**GACTAGACTGACTACGCATACGATCA**-5'

Derived from:

https://en.wikipedia.org/wiki/Sticky_and_b blunt_ends

Blunt End Ligation

5' -**CTGATCTGACTGA**-3'

3' -**GACTAGACTGACT**-5'

+

5' -**TGCGTATGCTAGT**-3'

3' -**ACGCATACGATCA**-5'

+

Ligase + ATP

=

5' -**CTGATCTGACTGATGCGTATGCTAGT**-3'

3' -**GACTAGACTGACTACGCATACGATCA**-5'

Derived from:

https://en.wikipedia.org/wiki/Sticky_and_b blunt_ends

Sticky End Ligation

5' - **CTGATCTGACT** - 3'
3' - **GACTAGACTGACTAC** - 5'

+

5' - **GATGCGTATGCTAGT** - 3'
3' - **GCATACGATCA** - 5'

+

Ligase + ATP

=

5' - **CTGATCTGACTGATGCGTATGCTAGT** - 3'
3' - **GACTAGACTGACTACGCATACGATCA** - 5'

Derived from:

https://en.wikipedia.org/wiki/Sticky_and_b blunt_ends

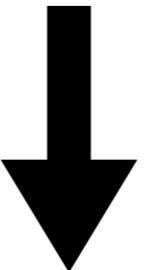
End Repair

End Repair: What

Fix overhanging ends so they are double-stranded

End Repair: What

5' -CTGATCTGACT -3'
3' -GACTAGACTGACTAC-5'



5' -CTGATCTGACT**GATG**-3'
3' -GACTAGACTGACTAC-5'

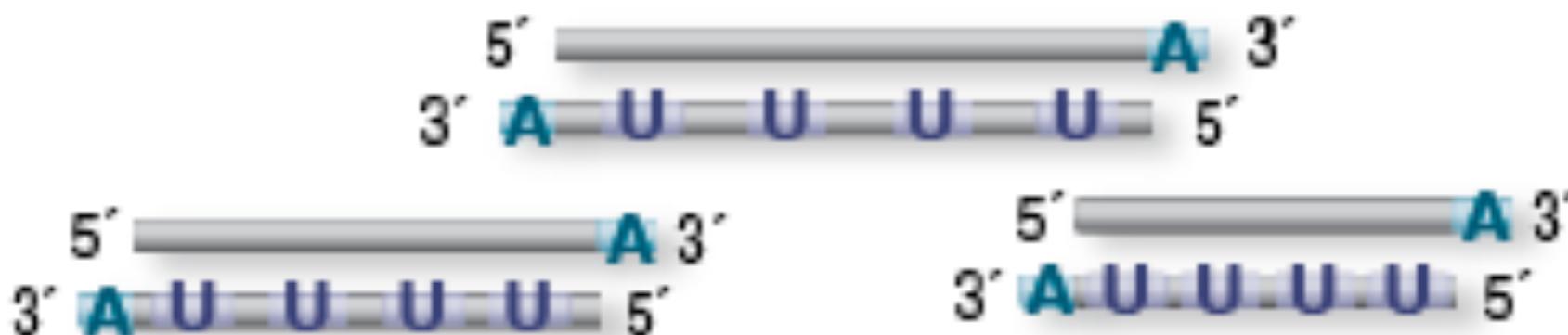
Why are ends NOT blunt?

Second Strand cDNA Synthesis



Clean Up

End Repair and dA-Tailing



Why are ends NOT blunt?

First Strand cDNA Synthesis

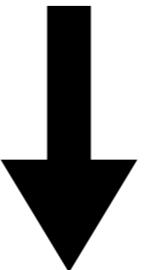


Second Strand cDNA Synthesis



End Repair: What

5' -CTGATCTGACT -3'
3' -GACTAGACTGACTAC-5'



5' -CTGATCTGACT**GATG**-3'
3' -GACTAGACTGACTAC-5'

End Repair: Why

- Allow blunt end ligation

End Repair: How

5' -CTGATCTGACT -3'
3' -GACTAGACTGACTAC-5'

+

?

=

5' -CTGATCTGACT**GATG**-3'
3' -GACTAGACTGACTAC-5'

End Repair: How

5' -CTGATCTGACT -3'
3' -GACTAGACTGACTAC-5'

+

DNA Polymerase

=

5' -CTGATCTGACT**GATG**-3'
3' -GACTAGACTGACTAC-5'

End Repair: How

- Requirements:
 - DNA with overhanging end
 - DNA Polymerase
 - dNTPs

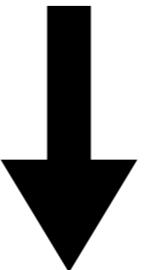
dA-Tailing

dA-Tailing: What

- Add a 3' “A” to blunt end fragments

dA-Tailing: What

5' -CTGATCTGACTGATG- 3'
3' -GACTAGACTGACTAC- 5'



5' - CTGATCTGACTGATGA- 3'
3' -**A**GACTAGACTGACTAC - 5'

dA-Tailing: Why

Allow sticky-end ligation to a “universal fragment”

dA-Tailing: Why

5' - GATGATTGCTGAAGA**A**-3'
3' -**A**CTACTAACGACTTC -5'

5' - AGTACTGTTCTTATA**A**-3'
3' -**A**TCA TGACAAGAAATA -5'

+

5' - CCATG-3'
3' -**T**GGTAC-5'

=

5' - GATGATTGCTGAAG**ACCATG**-3'
3' -**A**CTACTAACGACTTC**TGGTAC**-5'

5' - AGTACTGTTCTTAT**ACCATG**-3'
3' -**A**TCATGACAAGAAATA**TGGTAC**-5'

dA-Tailing: Why?



NEBNext Adaptor

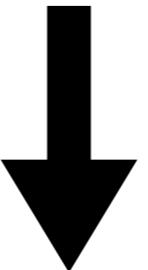


End Repair: How

- Requirements:
 - Blunt-end DNA Fragment
 - Klenow Fragment (3'→5' exo-)
 - Taq DNA Polymerase with ThermoPol® Buffer
 - ATP

dA-Tailing: What

5' -CTGATCTGACTGATG- 3'
3' -GACTAGACTGACTAC- 5'



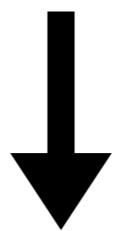
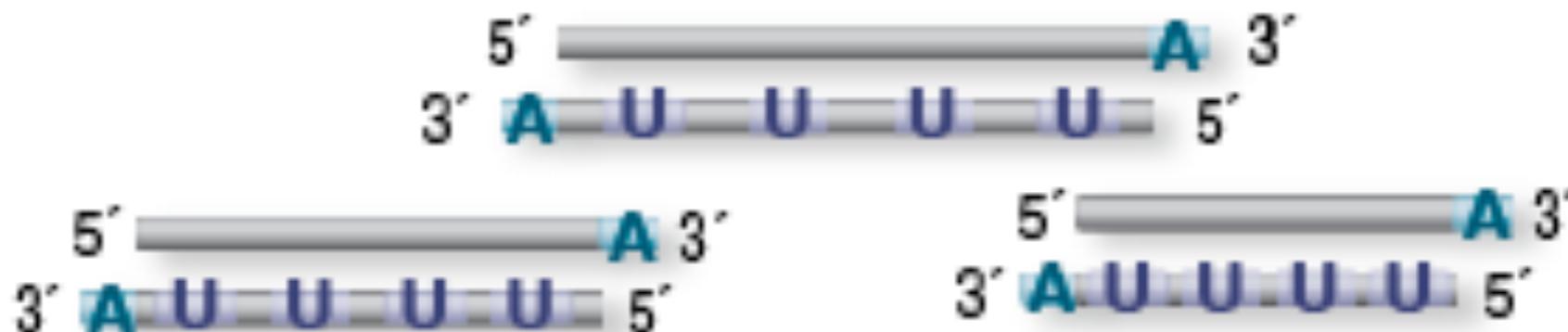
5' - CTGATCTGACTGATGA- 3'
3' -**A**GACTAGACTGACTAC - 5'

Adapter Ligation



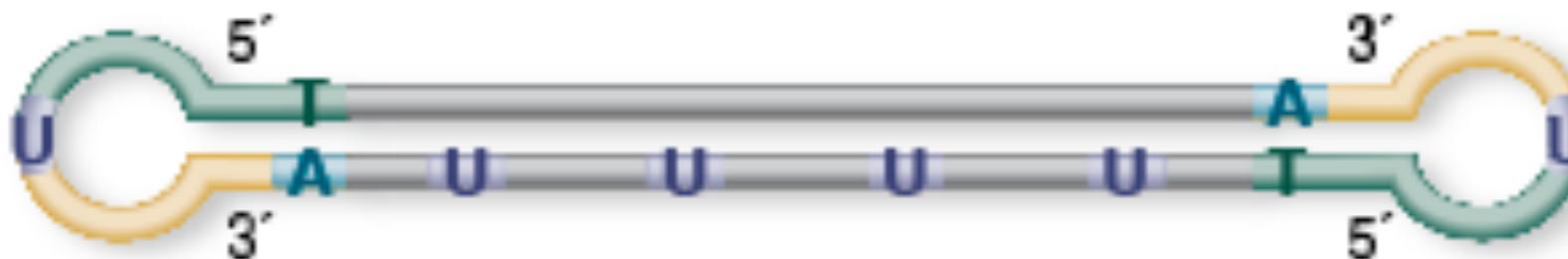
Adapter Ligation

End Repair and dA-Tailing



NEBNext Adaptor

Adaptor Ligation with optional NEBNext Adaptor



U Excision

First Strand cDNA Synthesis



Second Strand cDNA Synthesis

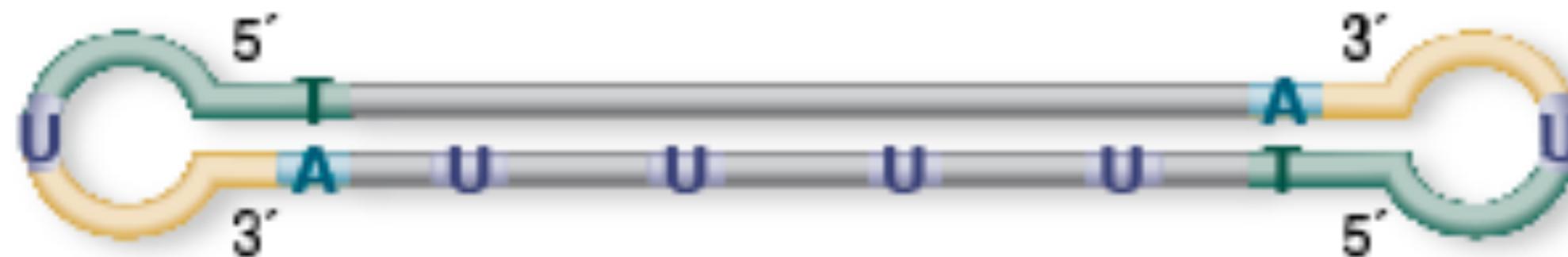


U Excision

Why?

U Excision

Adaptor Ligation with optional NEBNext Adaptor



U Excision



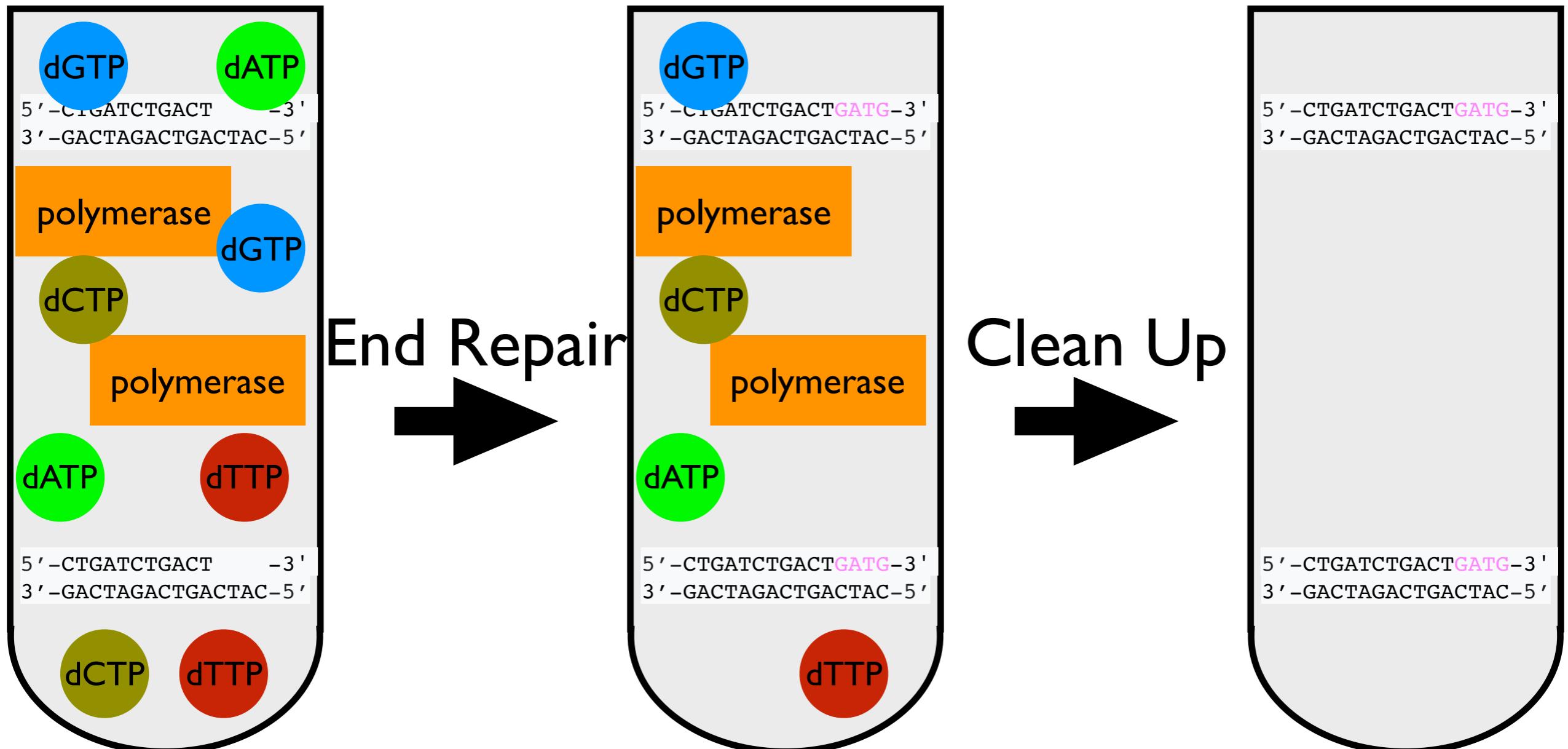
Size Selection (& Cleanup)



Clean Up and Size Selection: Why?

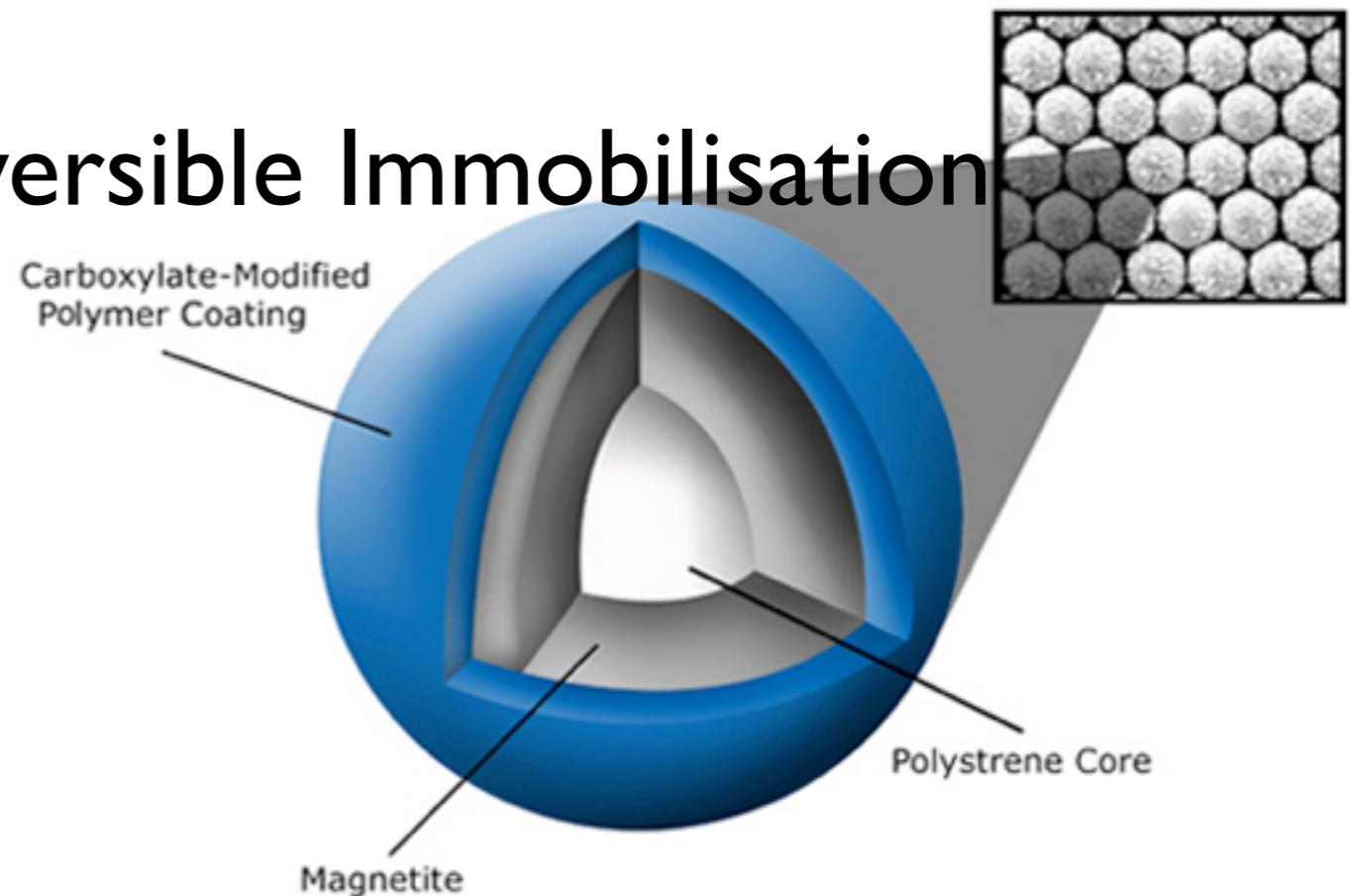
- Remove regents from previous step
- Eliminate unwanted fragments
 - Unligated adapter
 - adapter dimers
 - fragments without adapter
- Efficient cluster generation and sequencing

Sample Clean Up

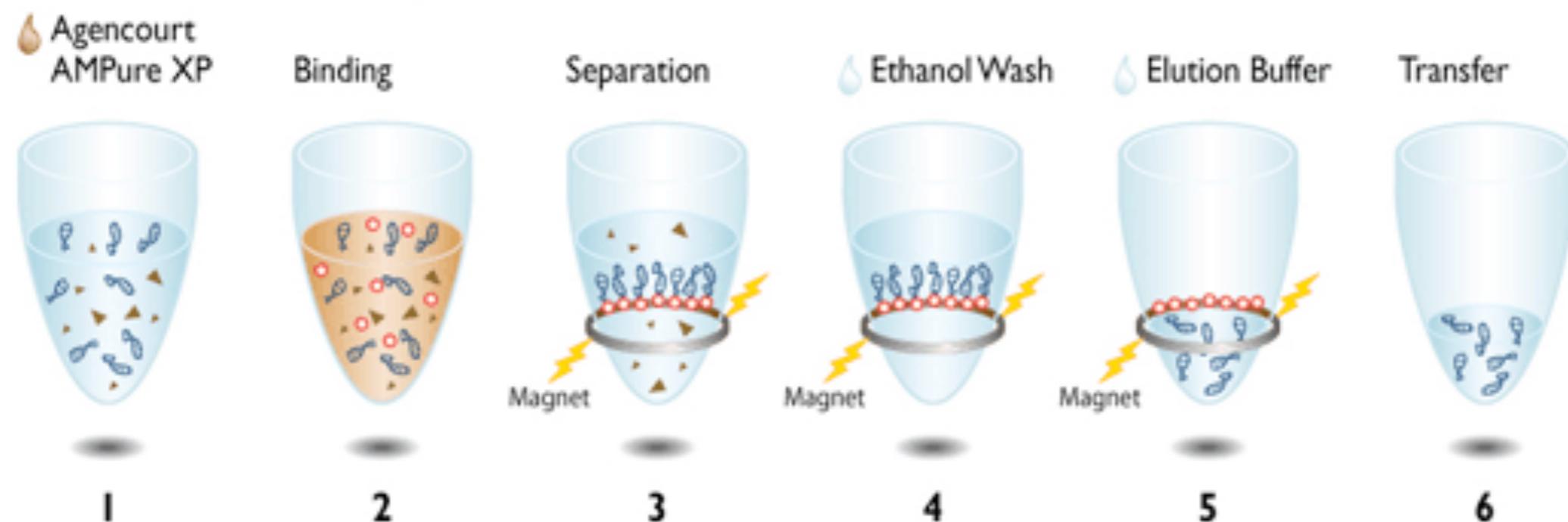


Clean Up and Size Selection: How?

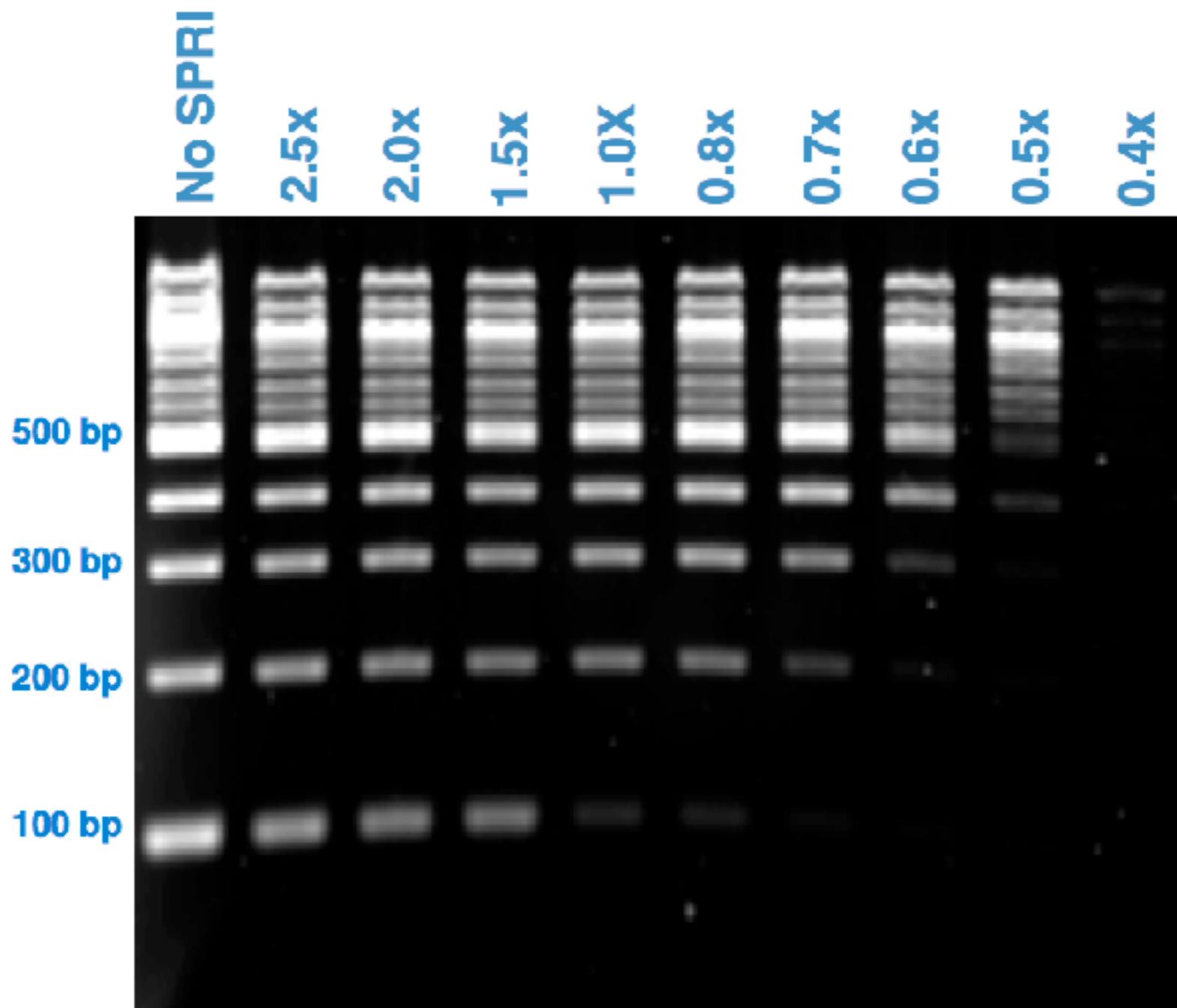
- Solid Phase Reversible Immobilisation (SPRI) beads



Clean Up and Size Selection: How?



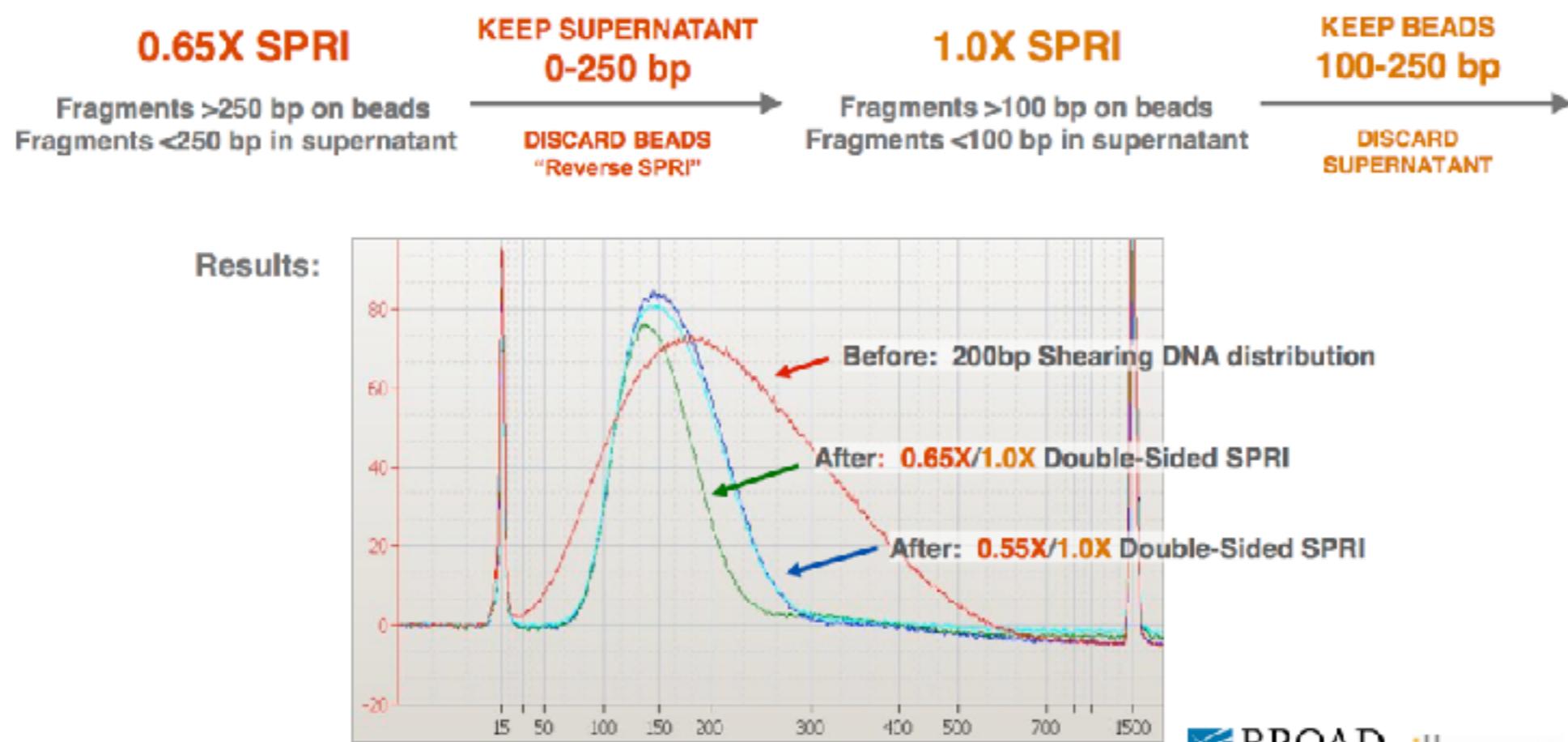
Size Selection: How?



Size Selection: How?

Option 2: Double-Sided SPRI

- By implementing a combination of good shearing with SPRI and “reverse” SPRI, one can select a fairly tight size range *with no gel*:



Clean Up and Size Selection: Alternatives

- Spin Columns
- Gel Purification
- DIY SPRI

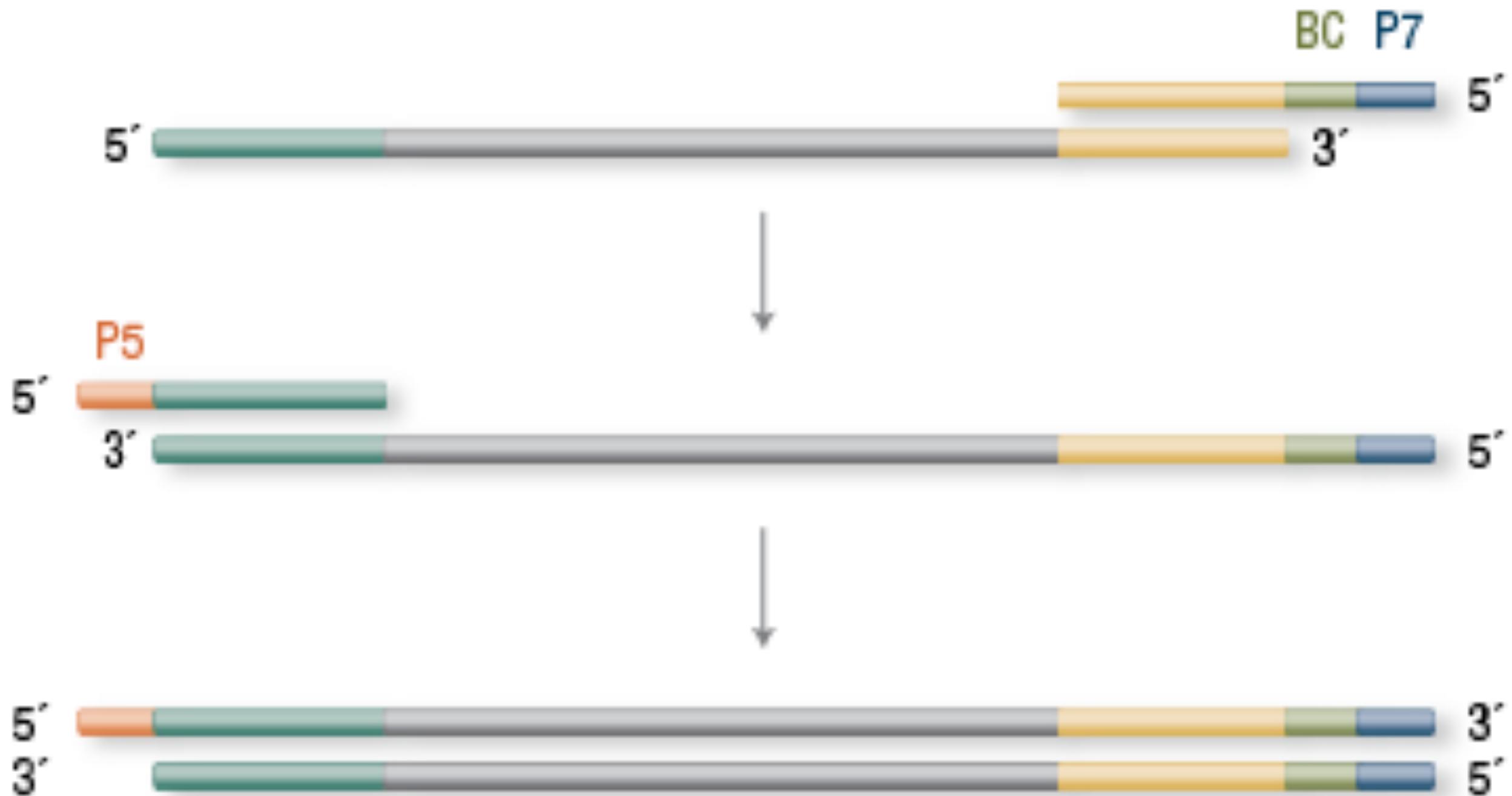
PCR Enrichment



PCR Enrichment: Why?

- I. Extend adapter to full length
 - A. add barcodes
 - B. add priming sites
2. Amplify library
 - A. Make more of the good fragments
 - B. Leave the garbage in the dust

PCR Enrichment

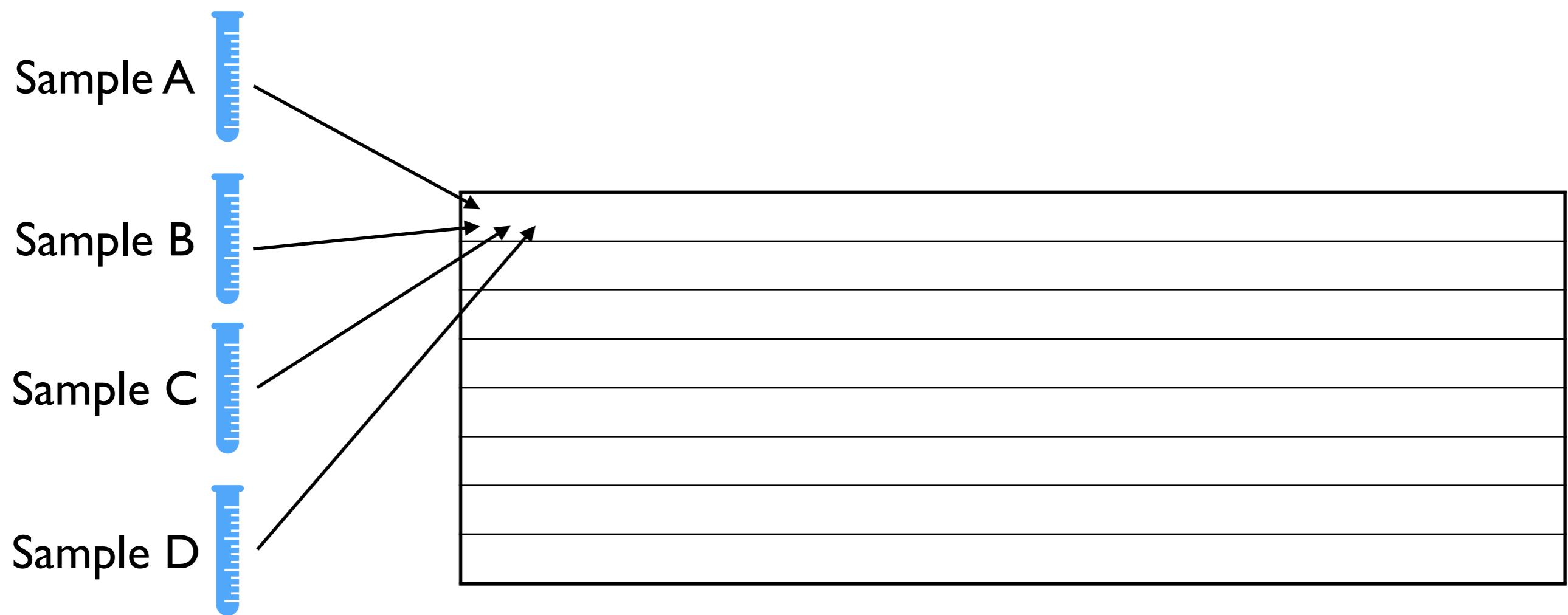


PCR Enrichment

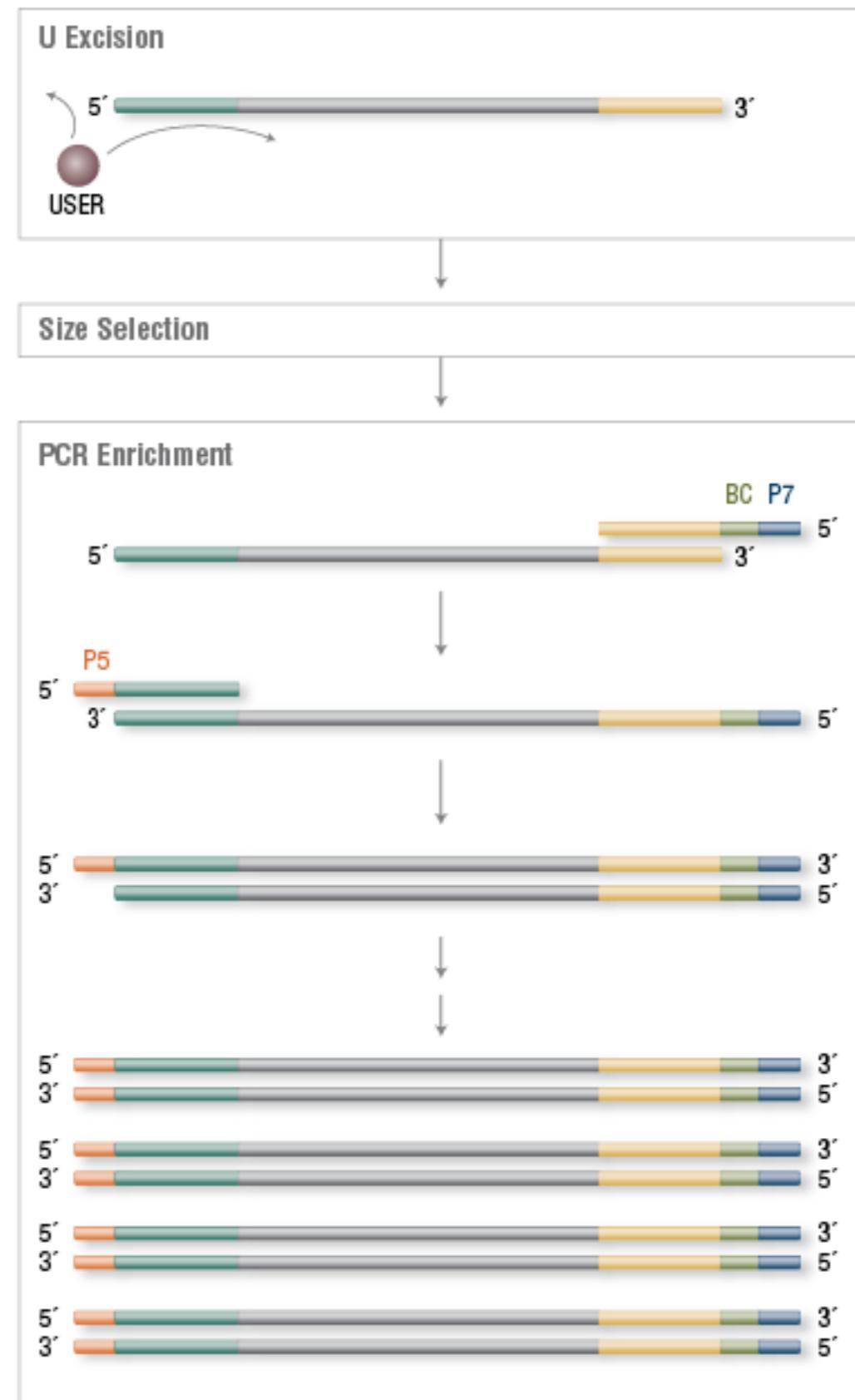


Barcodes: Why?

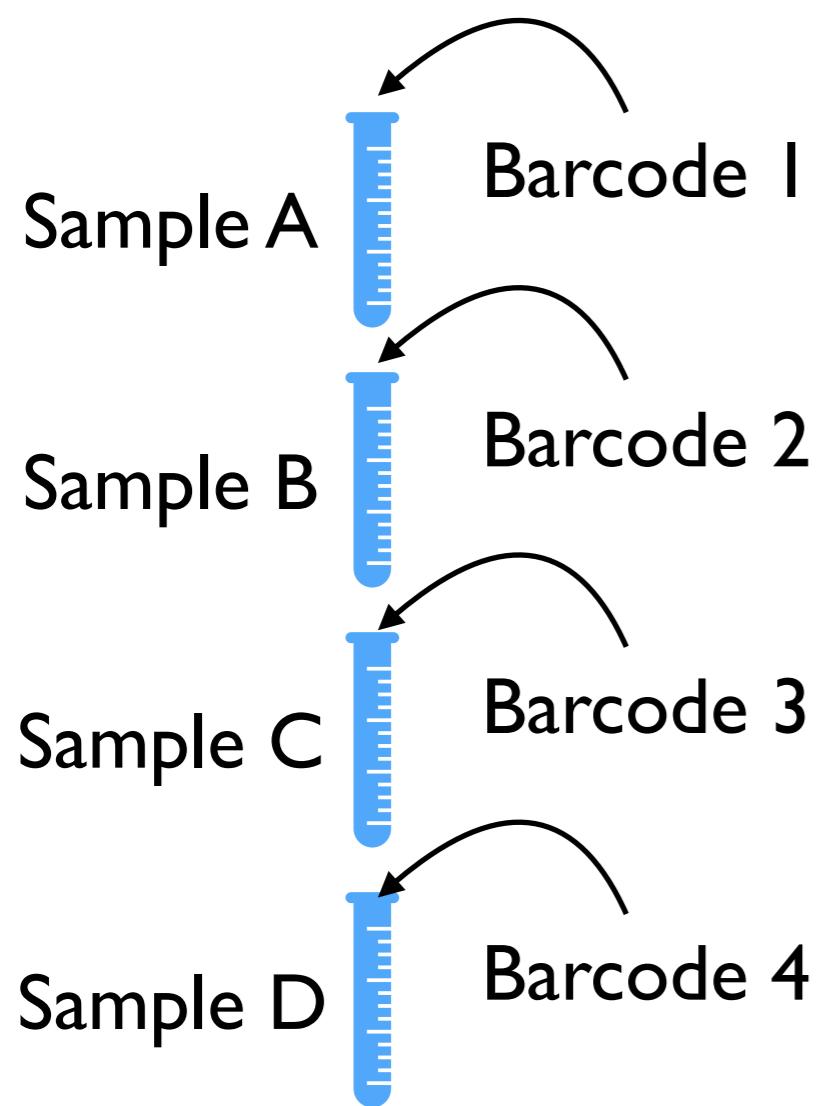
Multiplexing: Combine multiple samples in a lane



Barcodes



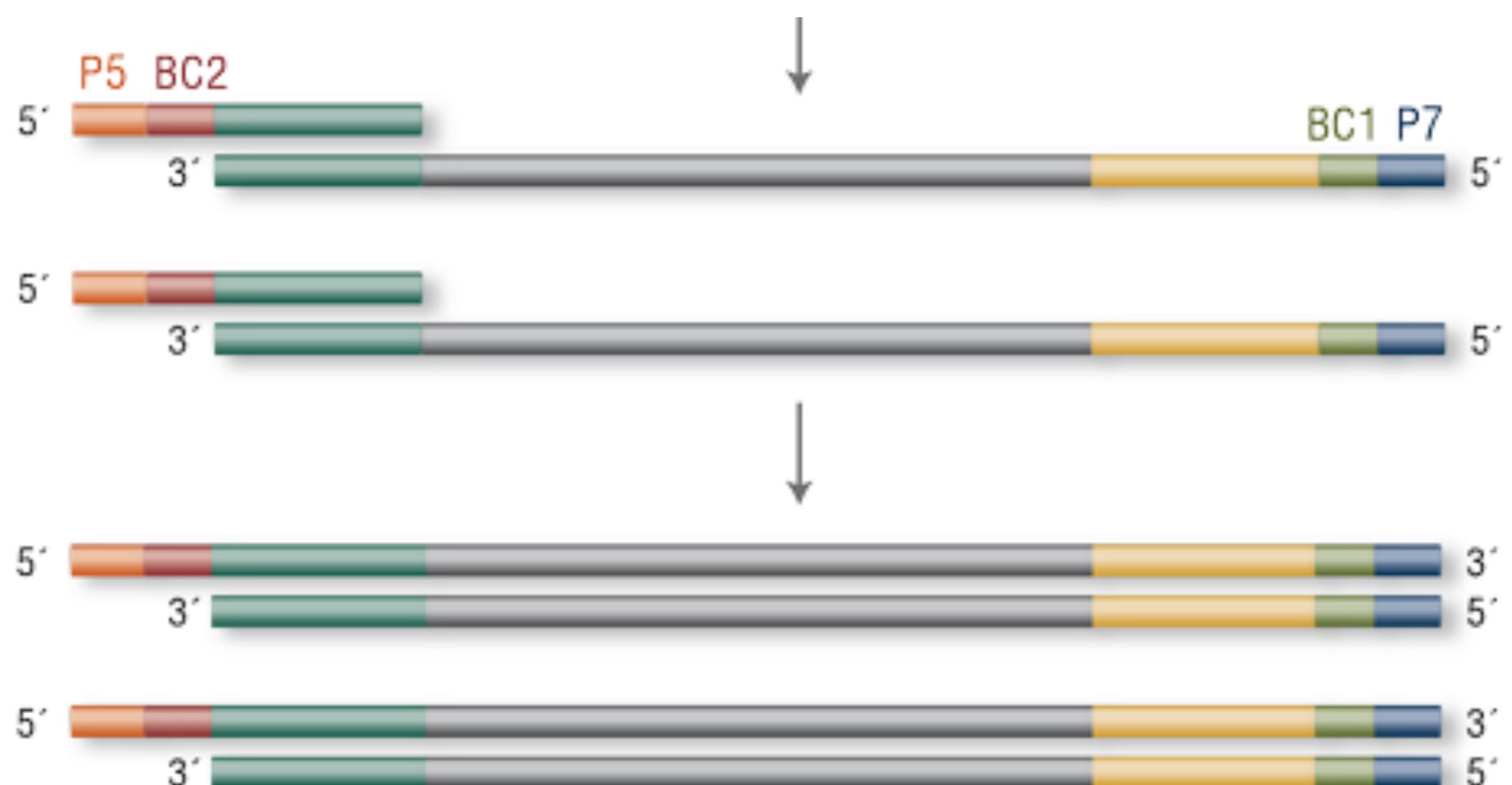
Barcodes



Barcodes

Sample_Name	I7_Index_ID	index
1_A	P49-E1	AAGACCGT
2_A	P50-E2	TTGCGAGA
3_A	P51-E3	GCAATTCC
4_A	P52-E4	GAATCCGT
5_A	P53-E5	CCGCTTAA
6_A	P54-E6	TACCTGCA
7_B	P55-E7	GTCGATTG
8_B	P56-E8	TATGGCAC
9_B	P57-E9	CTCGAACCA
10_B	P58-E10	CAACTCCA

Barcodes: Dual Index



[https://www.neb.com/-/media/nebus/page-images/
products/library-preparation-for-next-generation-
sequencing/dna_illumina_loopadapte6440.png?
la=en&device=modal](https://www.neb.com/-/media/nebus/page-images/products/library-preparation-for-next-generation-sequencing/dna_illumina_loopadapte6440.png?la=en&device=modal)

Barcodes: Dual Index

	BC1: A	BC1: B	BC1: C
BC2: W	Sample 1 A,W	Sample 2 B,W	Sample 3 C,W
BC2: X	Sample 4 A,X	Sample 5 B,X	Sample 6 C,X
BC2: Y	Sample 7 A,Y	Sample 8 B,Y	Sample 9 C,Y
BC2: Z	Sample 10 A,Z	Sample 11 B,Z	Sample 12 C,Z

Nasty Stuff

- Sodium Azide
- Actinomycin D

Library Preparation: Alternatives

- 1.Illumina Kits
- 2.Other Kits
- 3.DIY

Additional Sequencing Details

Paired-End

TCGAAAAG
AGCTTTCTGACTGCAACGGGCAATATGTCTGTGTGGA

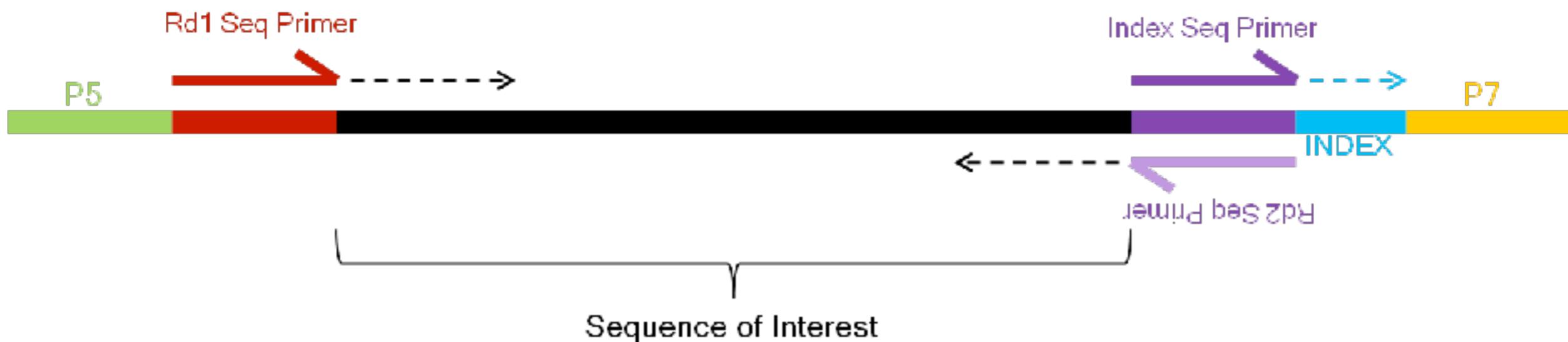
Paired-End

TCGAAAAG
AGCTTTCACTGCAACGGCAATATGTCTGTGGA

AGCTTTCACTGCAACGGCAATATGTCTGTGGA
GACACACCT

Multiplexing (Barcodes)

STRUCTURE DETAILS



HiSeq vs. MiSeq

	MiSeq	NextSeq	HiSeq 4000	NovaSeq 6000
Maximum Output	15 Gb	120 Gb	750 Gb	3000 Gb
Maximum Reads per Run	25 million	400 million	2.5 billion	10 billion
Maximum Read Length	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp
Run Time	4-56 hours	15-29 hours	< 1–3.5 days	13-45 hours
Cost*	\$1,787	\$4,695	\$19,206	\$35,538
Cost/Mbp*	\$0.119	\$0.039	\$0.026	\$0.012

* Duke Sequencing and Genomic Technologies Shared Resource, July 2018

Illumina Video

<https://www.youtube.com/watch?v=HMyCqWhwB8E>

Acknowledgements

- NEB
- Illumina

Evaluation!

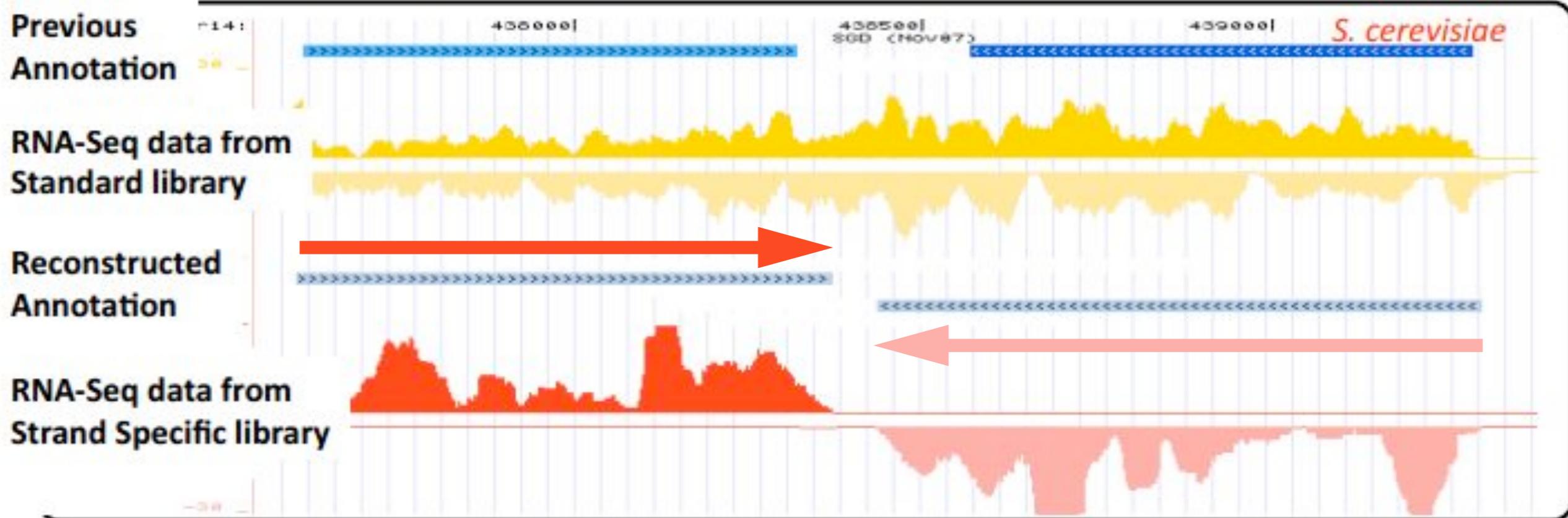
Extra Stuff

Strand-Specific Library

- Why Bother?

Strand-Specific Library

Strand-specific libraries



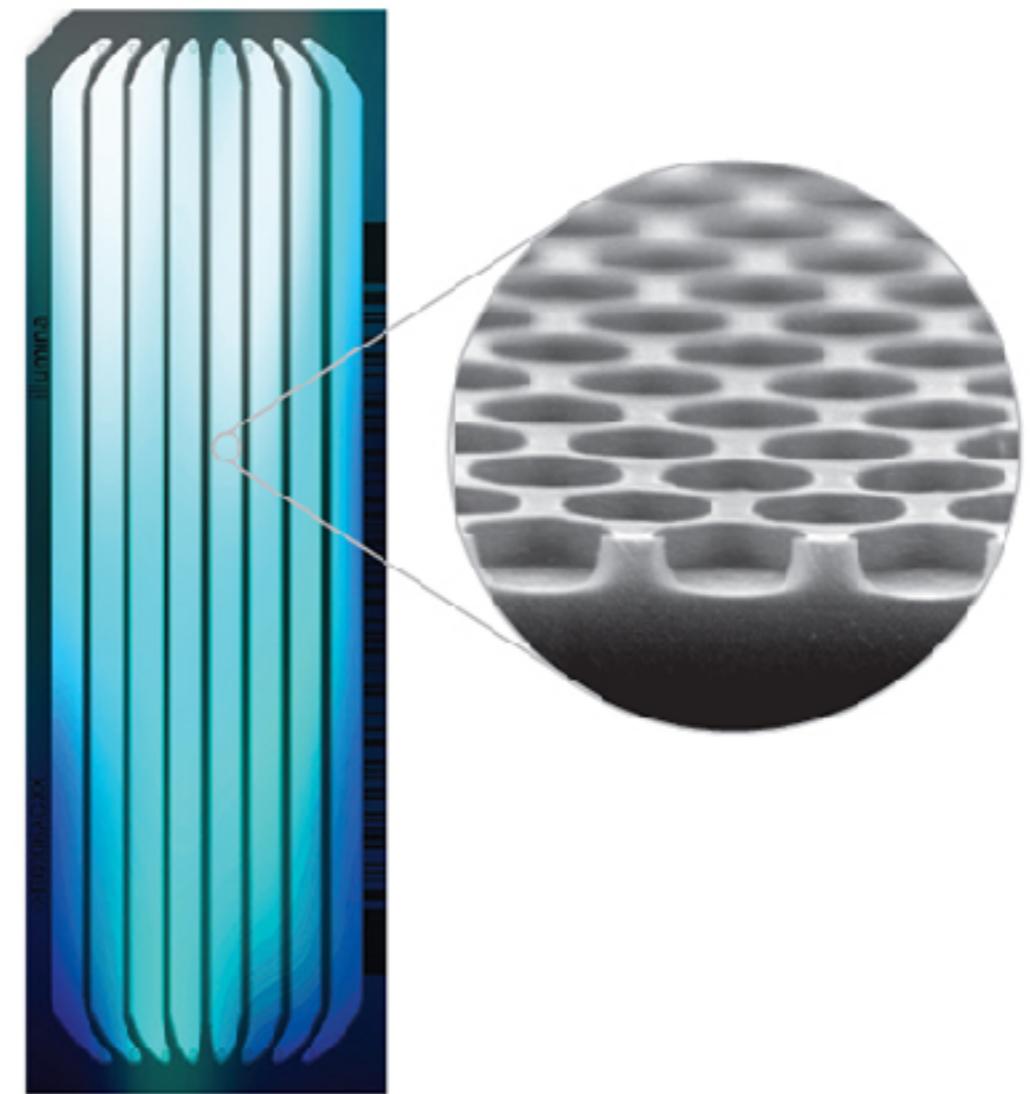
Joshua Levin and Moran Yassour

Strand Specific Prep



Patterned Flow Cells

- ExAmp
- Machines
 - HiSeq X
 - HiSeq 3000/4000
 - NovaSeq 6000



WhY Adapter?

DNA
Fragment



200-1000 bp

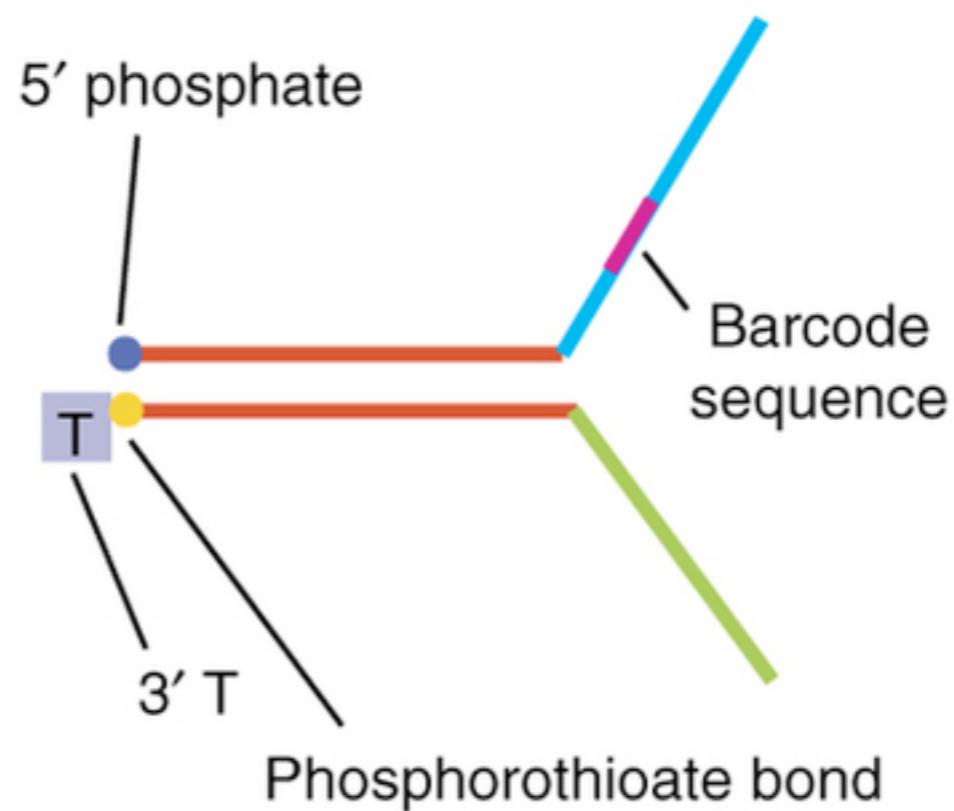
Adapters



Sequencing
“Library”



Y Adapter?



Uracil DNA glycosylase
and
DNA lyase

Uracil DNA glycosylase: What

- Remove Uracil base from DNA

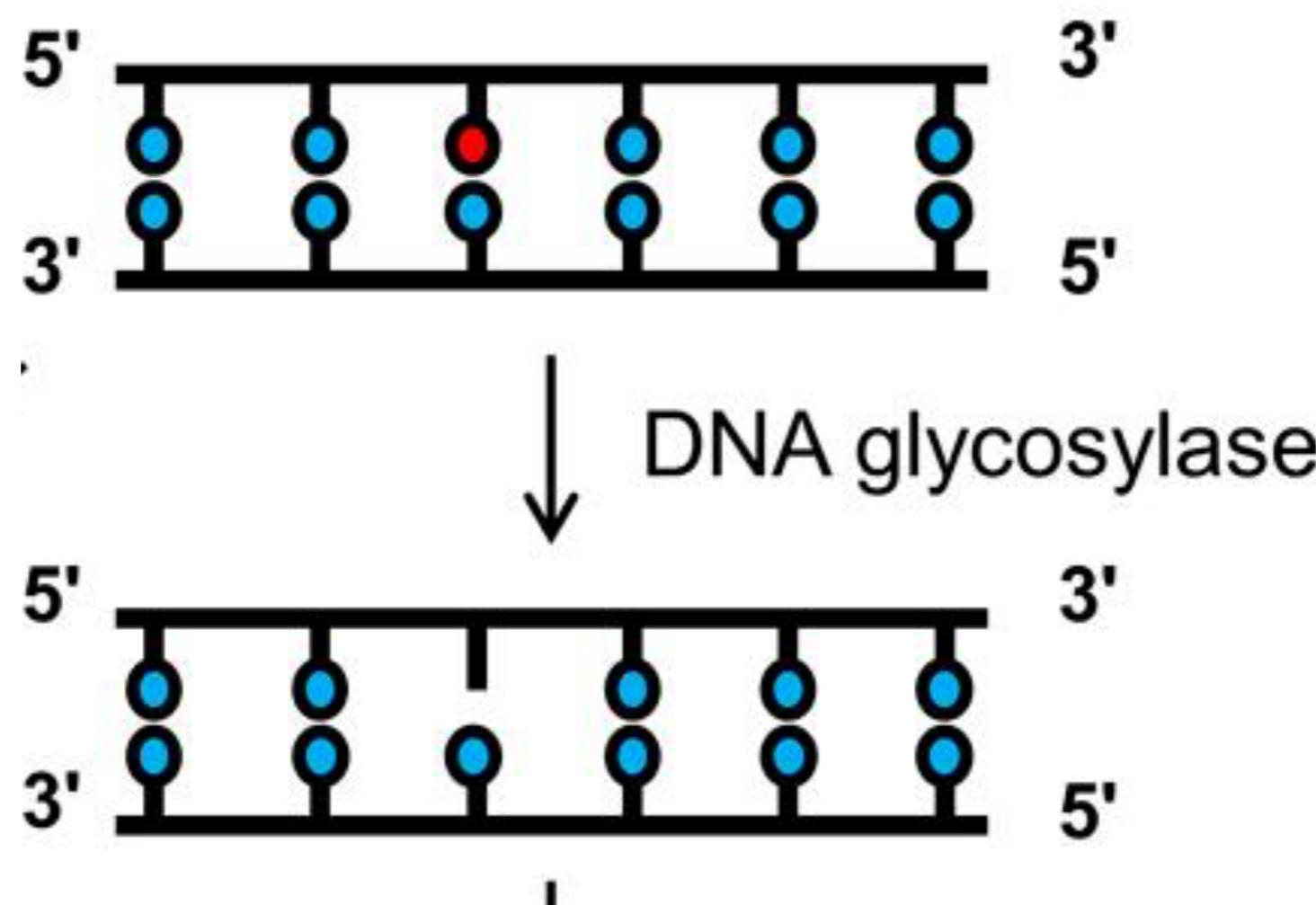
Uracil DNA glycosylase: What

5' -CTGATC**UGACTGATG**-3'
3' -**GACTAGACTGACTAC**-5'



5' -CTGATC-**GACTGATG**-3'
3' -**GACTAGACTGACTAC**-5'

Uracil DNA glycosylase: What



DNA Lyase: What

- Cleave DNA backbone at abasic site

DNA Lyase: What

5' -CTGATC-**GACTGATG**-3'
3' -**GACTAGACTGACTAC**-5'



5' -CTGATC GACTGATG-3'
3' -**GACTAGACTGACTAC**-5'