Bring Flow Cell

Housekeeping

- Me
- Names
- Apology
- Test Wifi and IBIEM Computing Environment

Morning Bootcamp Topics

- Microbiome Analysis Overview
- Computing Environment: RStudio
- Reproducible Research
  - Git
- Introduction to R

Microbiome Analysis Overview

Josh Granek
August 8, 2018

Metagenomics

	What	Information	Analogy	Target Size	Cost
Amplicon	Marker Gene	Who is Present	Name	100bp - 1kb	Low
Shotgun Metagenome	Genomes	What Genes are Present	CV	100kb - 100Mb	High
Shotgun Metatranscriptome	All RNA	What Genes are Expressed	Twitter Feed	100kb - 100Mb	High

granek JA - PubMed - NCBI

www.ncbi.nlm.nih.gov/pubmed/?term=granek+JA

NCBI Resources How To Sign in to NCBI

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed granek JA Search

Create RSS Create alert Advanced Help

Article types
Clinical Trial
Review
Customize ...

Text availability
Abstract
Free full text
Full text

PubMed Commons
Reader comments
Trending articles

Publication dates
5 years
10 years
Custom range...

Species
Humans
Other Animals

Clear all

Show additional filters

Summary 20 per page Sort by Most Recent Send to: Filters: Manage Filters

Search results

Items: 17

1. Evidence for distinct brain networks in the control of rule-based motor behavior.
Granek JA, Sergio LE.
J Neurophysiol. 2015 Aug;114(2):1296-309. doi: 10.1152/jn.00233.2014. Epub 2015 Jul 1.
PMID: 26133796
Similar articles

2. Rapid mapping of insertional mutations to probe cell wall regulation in Cryptococcus neoformans.
Esher SK, **Granek JA**, Alspaugh JA.
Fungal Genet Biol. 2015 Sep;82:8-21. doi: 10.1016/j.fgb.2015.06.003. Epub 2015 Jun 23.
PMID: 26112592
Similar articles

3. Integrating chemical mutagenesis and whole-genome sequencing as a platform for forward and reverse genetic analysis of Chlamydia.
Kokes M, Dunn JD, **Granek JA**, Nguyen BD, Barker JR, Valdivia RH, Bastidas RJ.
Cell Host Microbe. 2015 May 13;17(5):716-25. doi: 10.1016/j.chom.2015.03.014. Epub 2015 Apr 23.
PMID: 25920970 Free PMC Article
Similar articles

4. Antifungal drug resistance evoked via RNAi-dependent epimutations.
Calo S, Shertz-Wall C, Lee SC, Bastidas RJ, Nicolás FE, **Granek JA**, Mieczkowski P, Torres-Martínez S, Ruiz-Vázquez RM, Cardenas ME, Heitman J.
Nature. 2014 Sep 25;513(7518):555-8. doi: 10.1038/nature13575. Epub 2014 Jul 27.
PMID: 25079329 Free PMC Article
Similar articles 1 comment

5. Decoupled visually-guided reaching in optic ataxia: differences in motor control between canonical and non-canonical orientations in space.
Granek JA, Pisella L, Sternberger J, Vighetto A, Rossetti Y, Sergio LE.
PLoS One. 2013 Dec 31;8(12):e86136. doi: 10.1371/journal.pone.0086136. eCollection 2013.
PMID: 24392036 Free PMC Article
Similar articles

8. The genetic architecture of biofilm formation in a clinical isolate of Saccharomyces cerevisiae.
Granek JA, Murray D, Kayikçi Ö, Magwene PM.
Genetics. 2013 Feb;193(2):587-603. doi: 10.1534/genetics.112.142067. Epub 2012 Nov 19.
PMID: 23172852 Free PMC Article
Similar articles

7. The role of the caudal superior parietal lobule in updating hand location in peripheral vision: further evidence from optic ataxia.
Granek JA, Pisella L, Blangero A, Rossetti Y, Sergio LE.
PLoS One. 2012;7(10):e48619. doi: 10.1371/journal.pone.0048619. Epub 2012 Oct 5.
PMID: 23071689 Free PMC Article
Similar articles

9. Pleiotropic signaling pathways orchestrate yeast development.
Granek JA, Kayikçi Ö, Magwene PM.
Curr Opin Microbiol. 2011 Dec;14(6):676-81. doi: 10.1016/j.mib.2011.08.004. Epub 2011 Sep 23. Review.
PMID: 21962291 Free PMC Article
Similar articles

Find related data
Database: Select
Find items

Search details
granek JA [Author]
Search See more...

Recent Activity
Turn Off Clear
granek JA (17) PubMed
granek J (20) PubMed
Global patterns of 16S rRNA diversity at a depth of millions of sequences per sa... PubMed
Scott Harrison (71) PubMed
See more...

Amplicon Sequencing

PCR amplify and sequence a marker gene

	Marker Gene
Bacteria	16s rRNA
Fungi	18s or ITS rRNA
Archaea	16s rRNA
Protozoa	18s rRNA
Viruses	?????

Metagenomics

	What	Information	Analogy	Target Size	Cost
Amplicon	Marker Gene	Who is Present	Name	100bp - 1kb	Low
Shotgun Metagenome	Genomes	What Genes are Present	CV	100kb - 100Mb	High
Shotgun Metatranscriptome	All RNA	What Genes are Expressed	Twitter Feed	100kb - 100Mb	High

Metagenomics

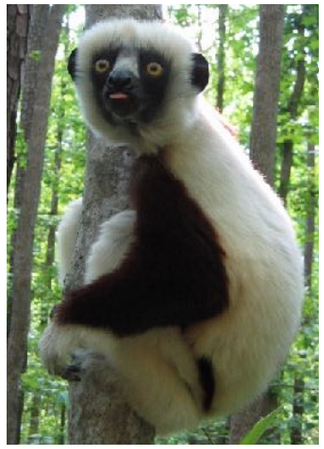
	What	Information	Analogy	Target Size	Cost	Discovery?
Amplicon	Marker Gene	Who is Present	Name	100bp - 1kb	Low	+/-
Shotgun Metagenome	Genomes	What Genes are Present	CV	100kb - 100Mb	High	++
Shotgun Metatranscriptome	All RNA	What Genes are Expressed	Twitter Feed	100kb - 100Mb	High	++

16s Amplicon Analysis Overview

Big Picture

1. What is present?
2. How much?
3. Are there differences between treatments, host species, …?
4. What are the differences?

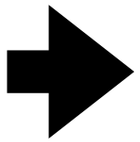
Molecular Biology





http://www.geograph.org.uk/photo/2847164
https://commons.wikimedia.org/wiki/File:Pcr.jpg
https://commons.wikimedia.org/wiki/File:Illumina_MiSeq_sequencer.jpg

Bioinformatic Analysis



```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0
TTACGCTAACAGGCGGTAGCCTGGCAGGGTCAGGAAATCAATTAACTCATCGGAAGTGGTGATCTGTTCCATCAAGCGTGCGGCATCGTCAAAACGCCC
+
ABBBABBBAFFFGGGGGGGGGGHGGHGGGCG2GF3FFGHHHHHHGGFGHEHHGGGEHHHHAGGHHGHHHFFDHFHHHGEGGGG@F@H?GHH/GBEFGGG
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CTGCCAGTTGAACGACGGCGAGCAGTTATAAGCCAGCAGTTTGCCCGGATATTTCGCGTGGATAGCTTGTGCAAAGCGACGCGCCAGTTCCAGATCCGGCG
+
AAABBFFFFFFFGGGGGGGGGGGGHHHHHHHHHGHGHGHHHHHGHHHGGGGGHHHHGGGGGGGHHHGHHFFHHHHHGHGGGGGGGGGGHHHHHHHHHHHGGG
@M00698:36:000000000-AFBEL:1:1101:15928:1413 1:N:0:0
GTAAAGTCCTGAGTGATACCGGCAACTTTTACCCCCAGTCCCACTTTCGAACCGGCAAACATATCGGCAAAAGAGGCCGTGCCTGATTTAAAGCCGTAGGT
+
```

	Sample 1	Sample 2	…	Sample N
Bacteria 1				
Bacteria 2				
…				
Bacteria N				

https://commons.wikimedia.org/wiki/File:Illumina_MiSeq_sequencer.jpg

Statistical Analysis

	Sample 1	Sample 2	…	Sample N
Bacteria 1				
Bacteria 2				
…				
Bacteria N				

1. What is present?
2. How much?
3. Are there differences between treatments, host species, …?
4. What are the differences?

Caveat

http://www.lebleu.com

Sequencing Details

DNA Sequencing Technologies (Abridged)

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

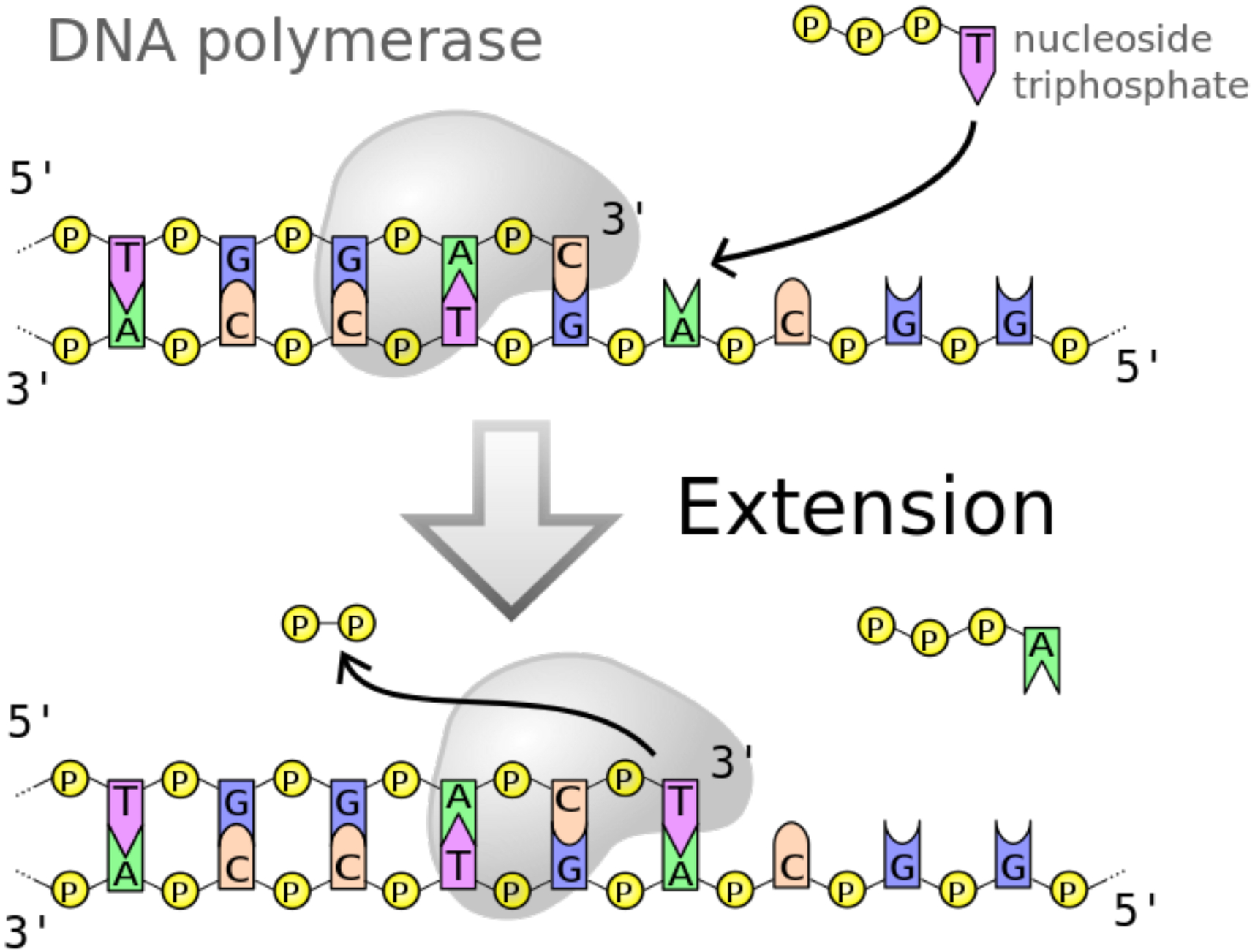
Sequencing by Synthesis

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

DNA Synthesis

- What are the minimum components for DNA Replication?

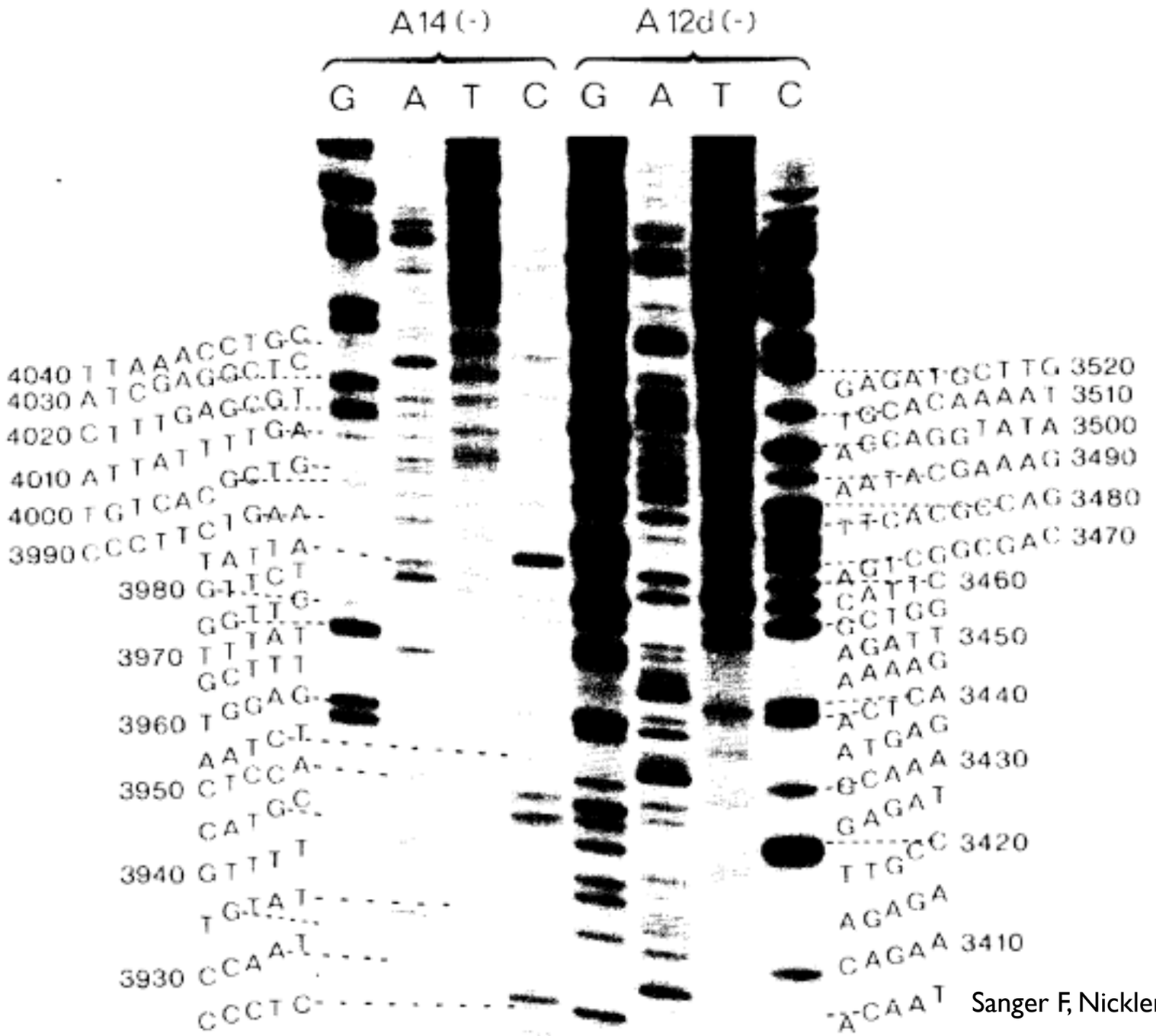
DNA Synthesis



https://upload.wikimedia.org/wikipedia/commons/thumb/6/6f/DNA_polymerase.svg/700px-DNA_polymerase.svg.png

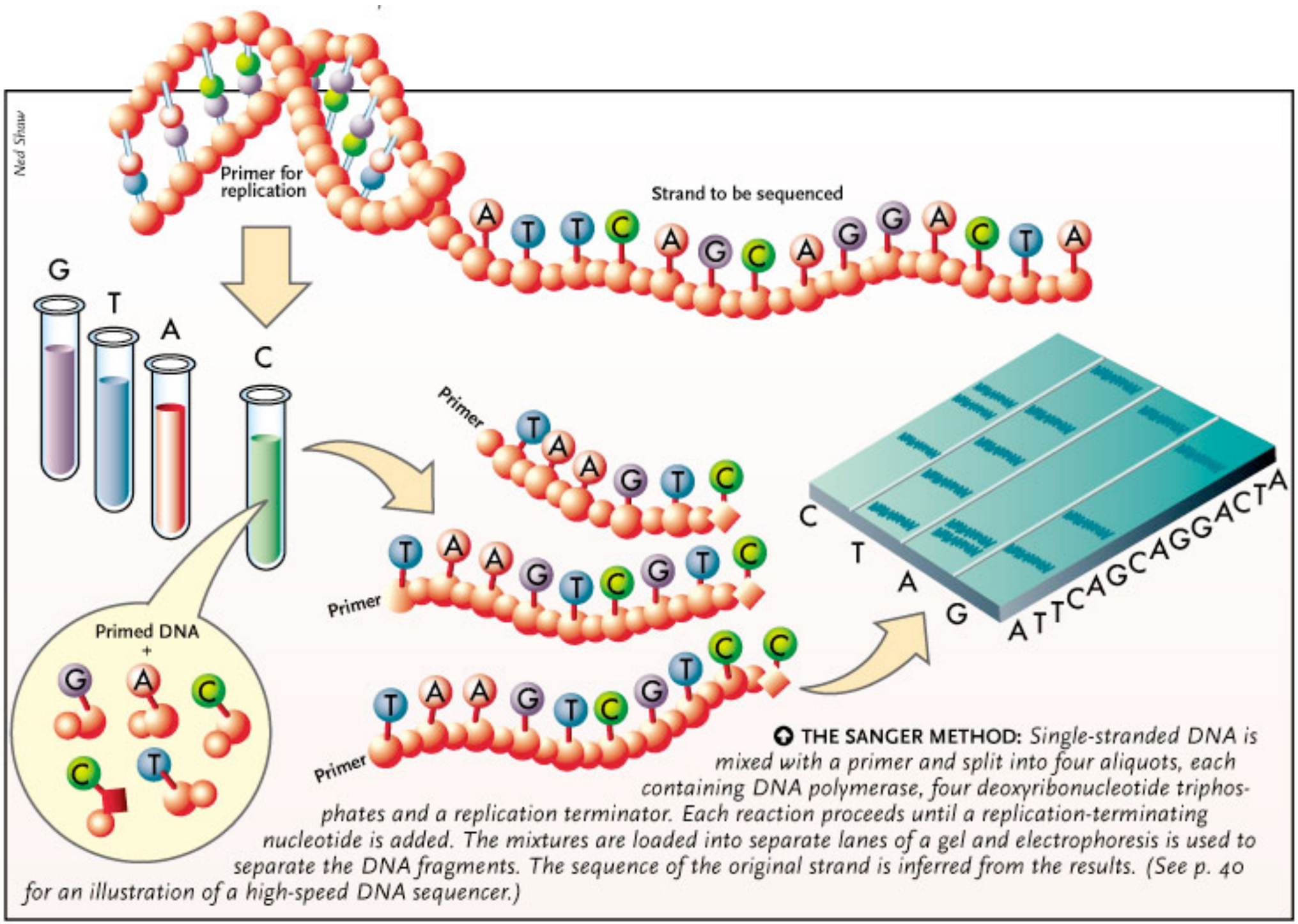
1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

Sanger Sequencing



Sanger F, Nicklen S, Coulson AR., *PNAS*, 1977

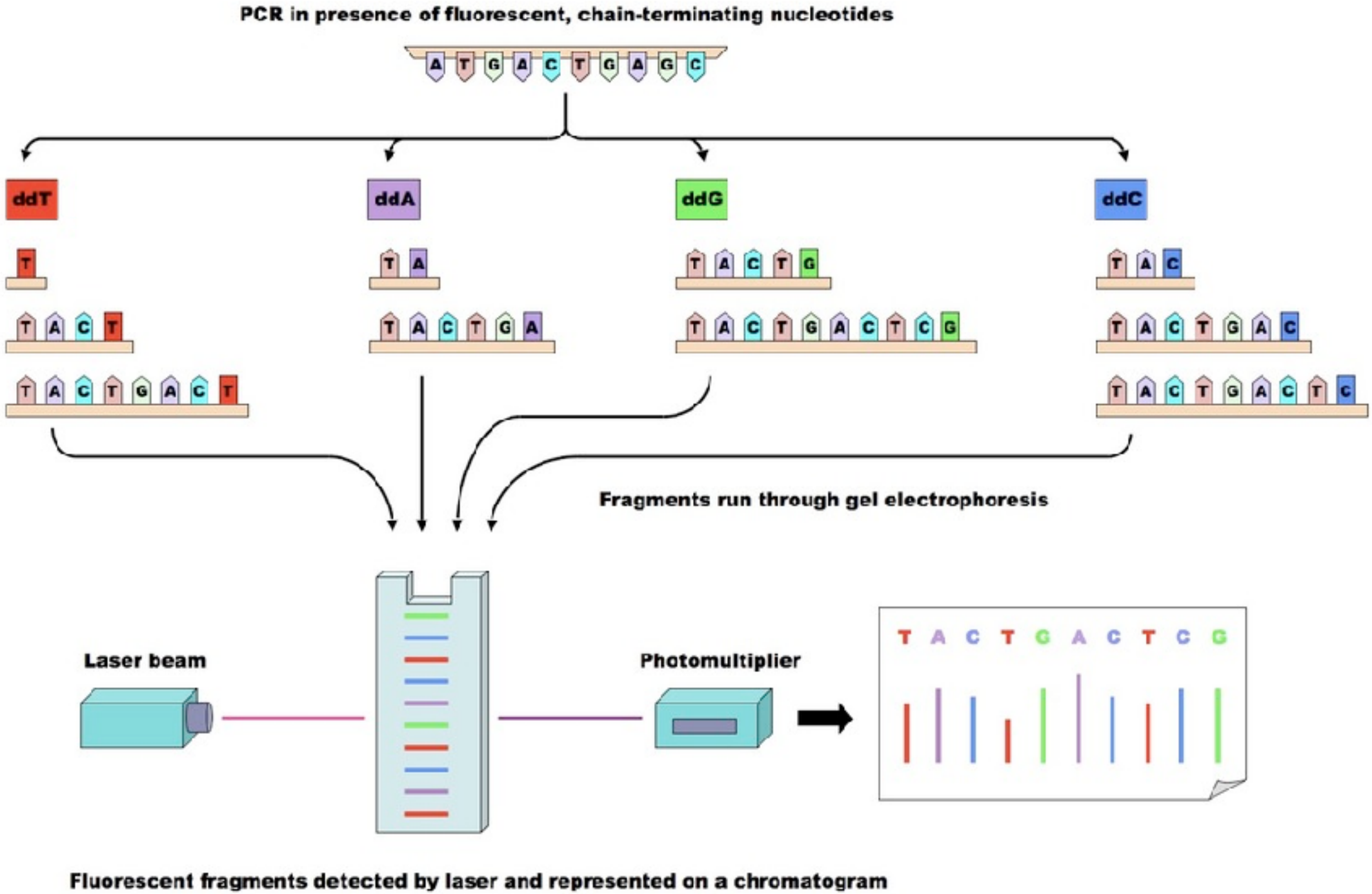
Sanger Sequencing



THE SANGER METHOD: *Single-stranded DNA is mixed with a primer and split into four aliquots, each containing DNA polymerase, four deoxyribonucleotide triphosphates and a replication terminator. Each reaction proceeds until a replication-terminating nucleotide is added. The mixtures are loaded into separate lanes of a gel and electrophoresis is used to separate the DNA fragments. The sequence of the original strand is inferred from the results. (See p. 40 for an illustration of a high-speed DNA sequencer.)*

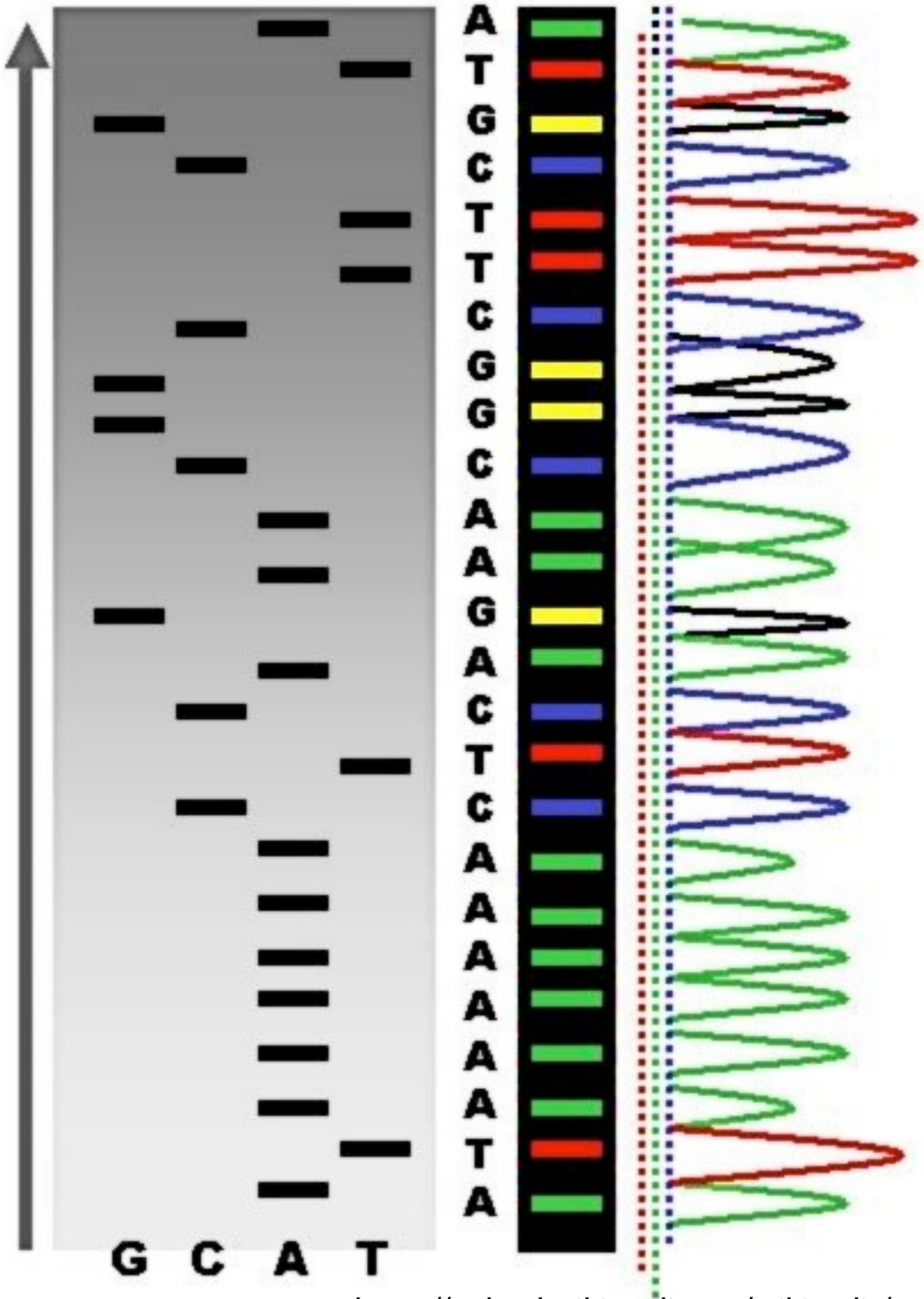
http://www.the-scientist.com/?articles.view/articleNo/15939/title/DNA-Sequencing-Industry-Sets-its-Sights-on-the-Future/

Dye-terminator



http://www.vce.bioninja.com.au/_Media/sequencing_med.jpeg

Radiolabel vs. Dye



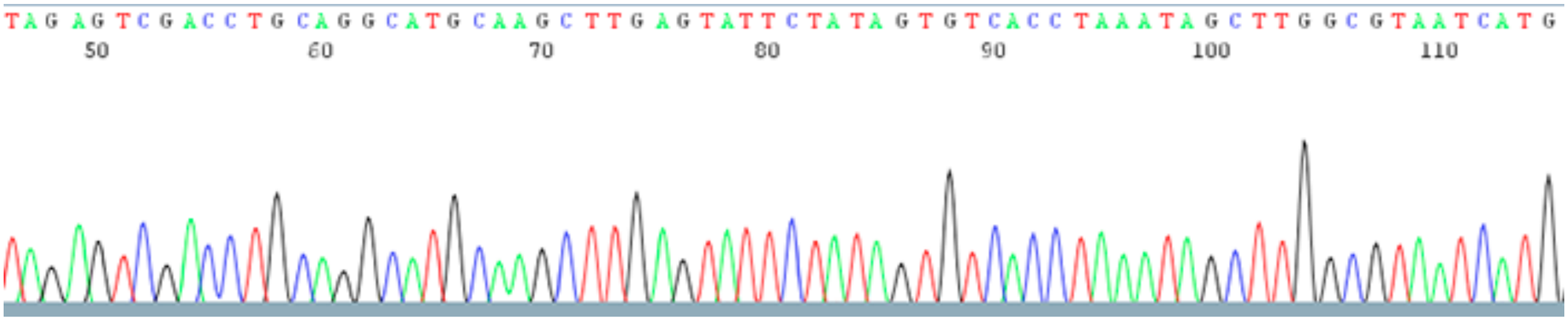
https://upload.wikimedia.org/wikipedia/commons/3/3d/Radioactive_Fluorescent_Seq.jpg

High-Throughput

How?

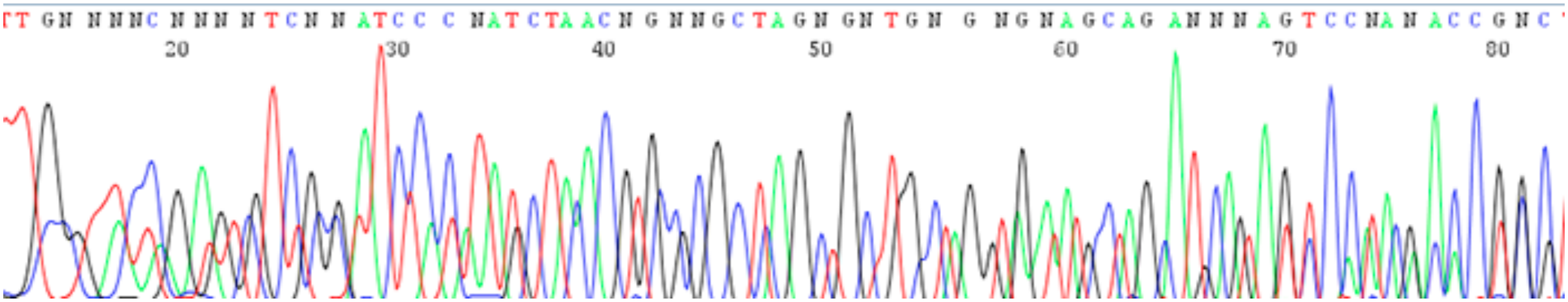
- Separate
- Detect

Dye-terminator Sanger Sequencing



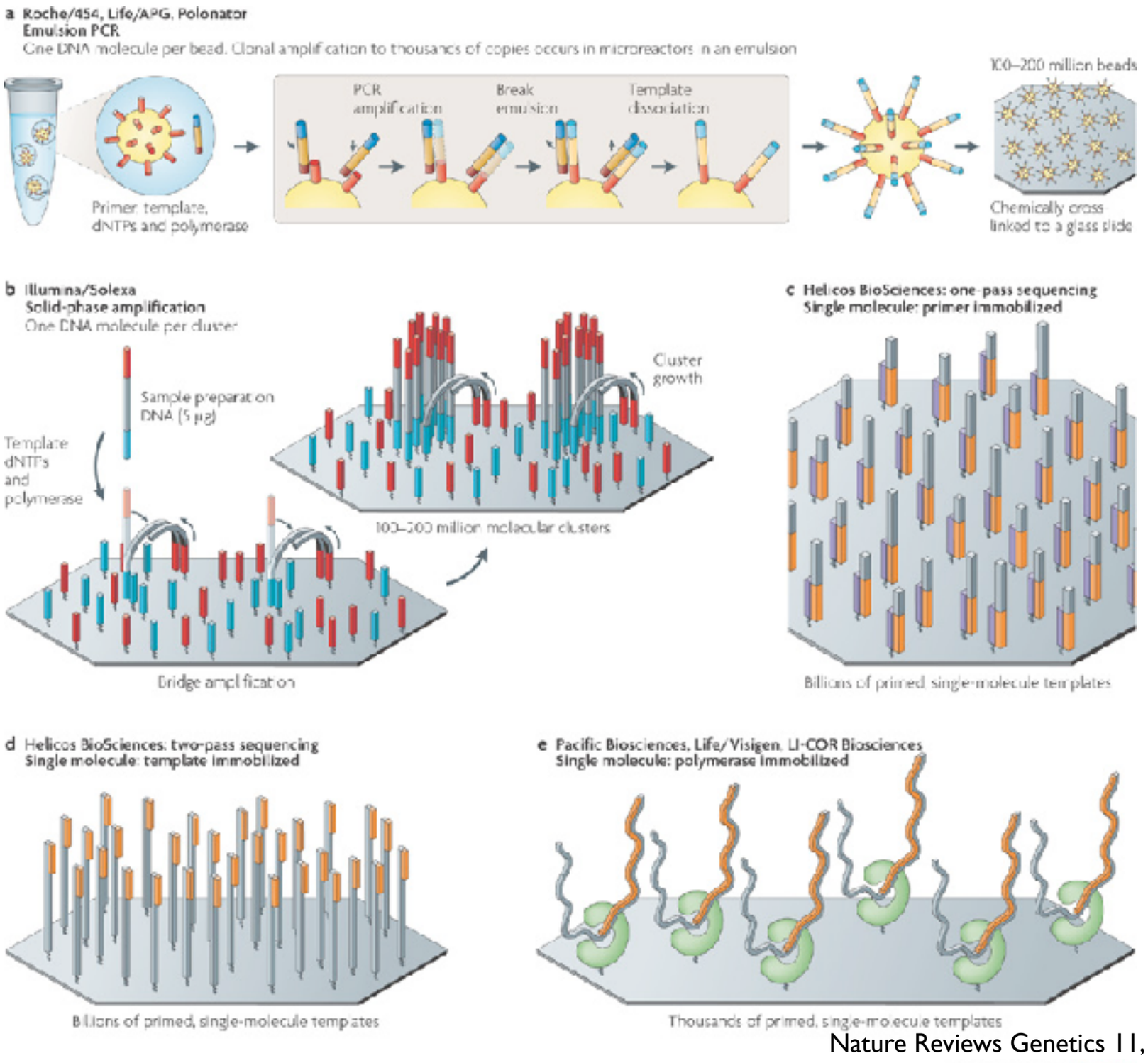
https://dnacore.mgh.harvard.edu/new-cgi-bin/site/pages/sequencing_pages/seq_troubleshooting.jsp

Bad Sanger Sequence



https://dnacore.mgh.harvard.edu/new-cgi-bin/site/pages/sequencing_pages/seq_troubleshooting.jsp

Template immobilization



Nature Reviews Genetics 11, 31-46, doi:10.1038/nrg2626

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

Library Preparation

DNA Fragment

200-1000 bp

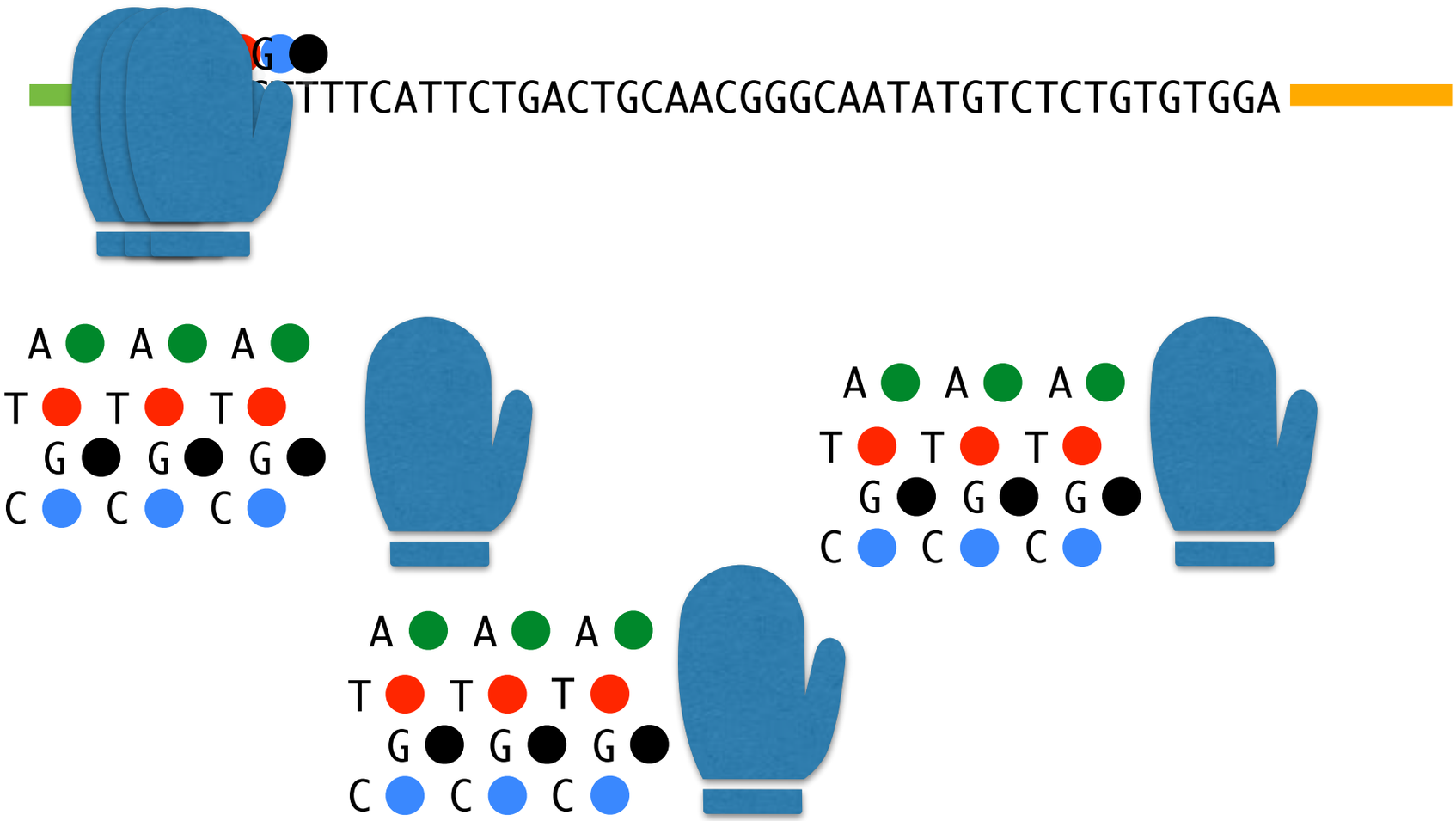
+

Adapters

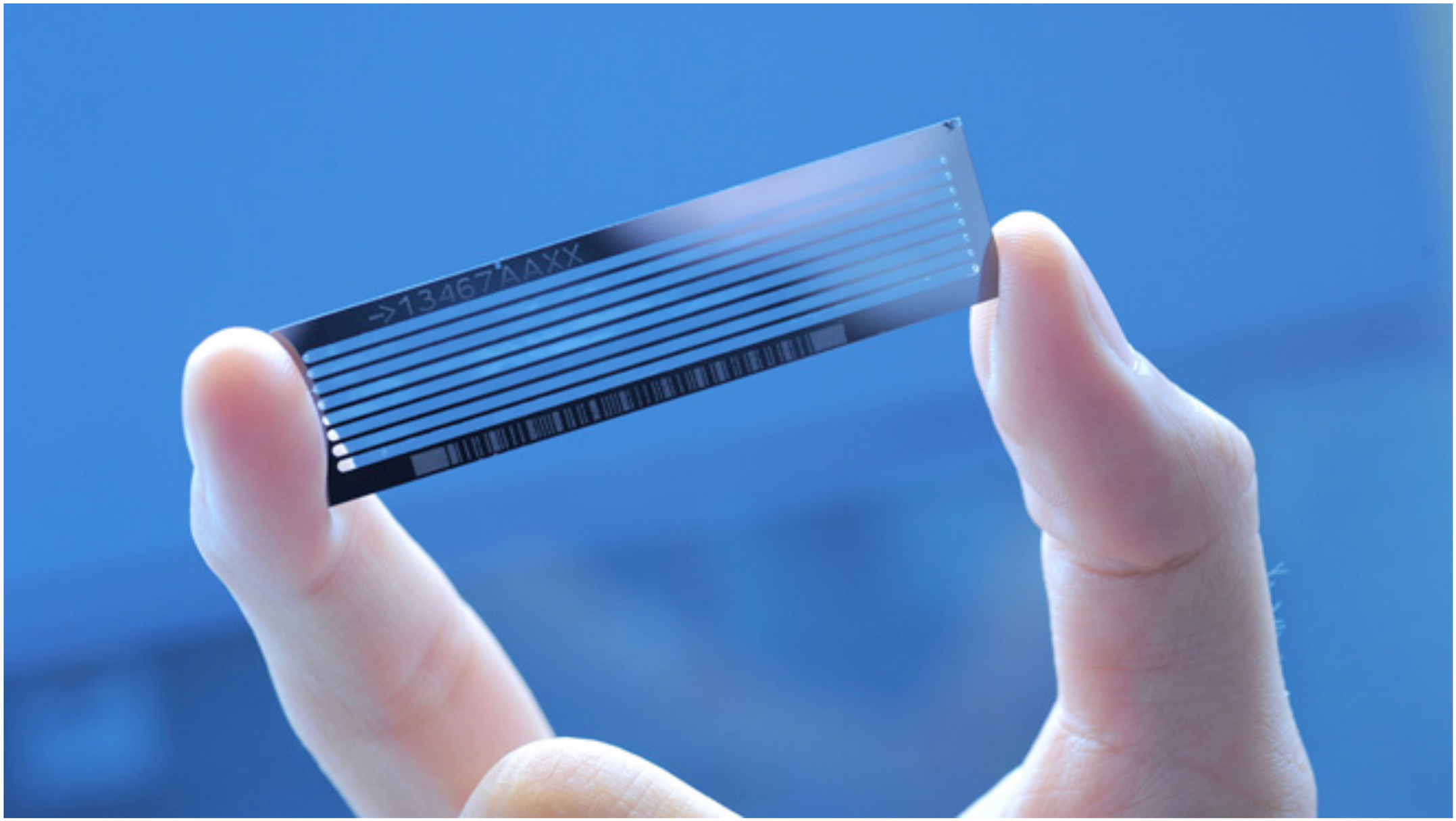
=

Sequencing "Library"

Sequencing


TTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGA
A A A
T T T
G G G
C C C
A A A
T T T
G G G
C C C
A A A
T T T
G G G
C C C

A Flow Cell



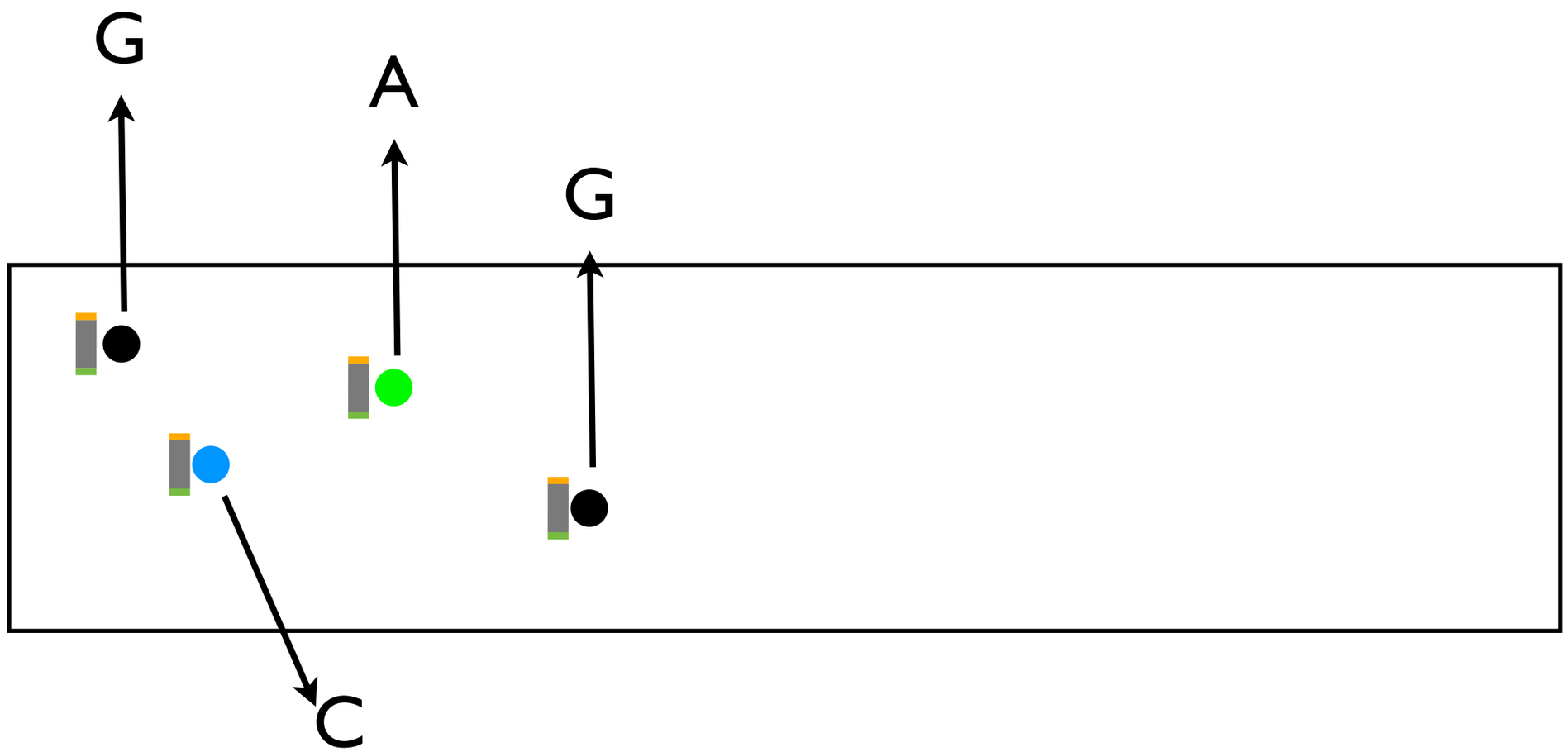
http://global.fncstatic.com/static/managed/img/Scitech/neanderthal%20genome%205.jpg

Show Flow Cell

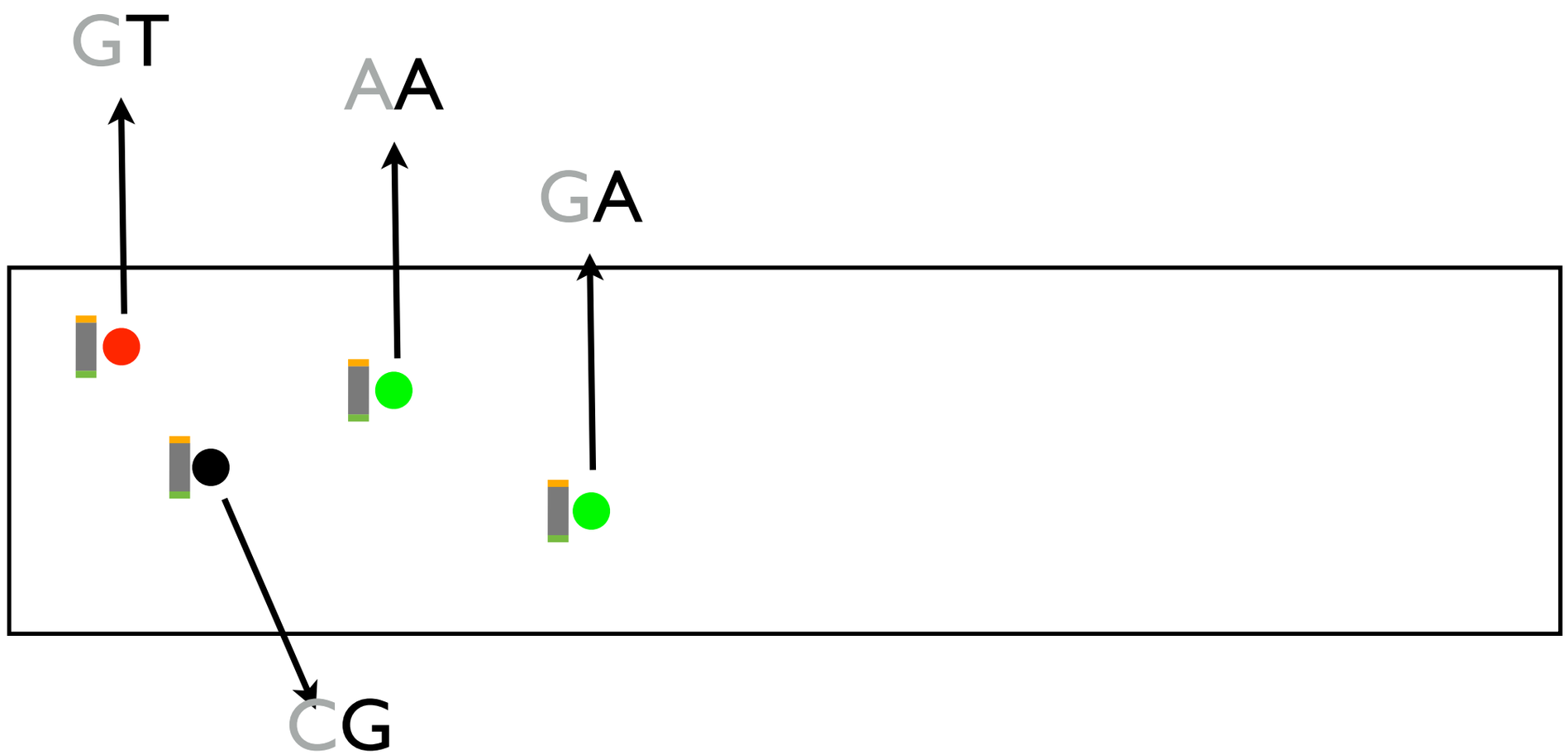
A Flow Cell

Bind Library

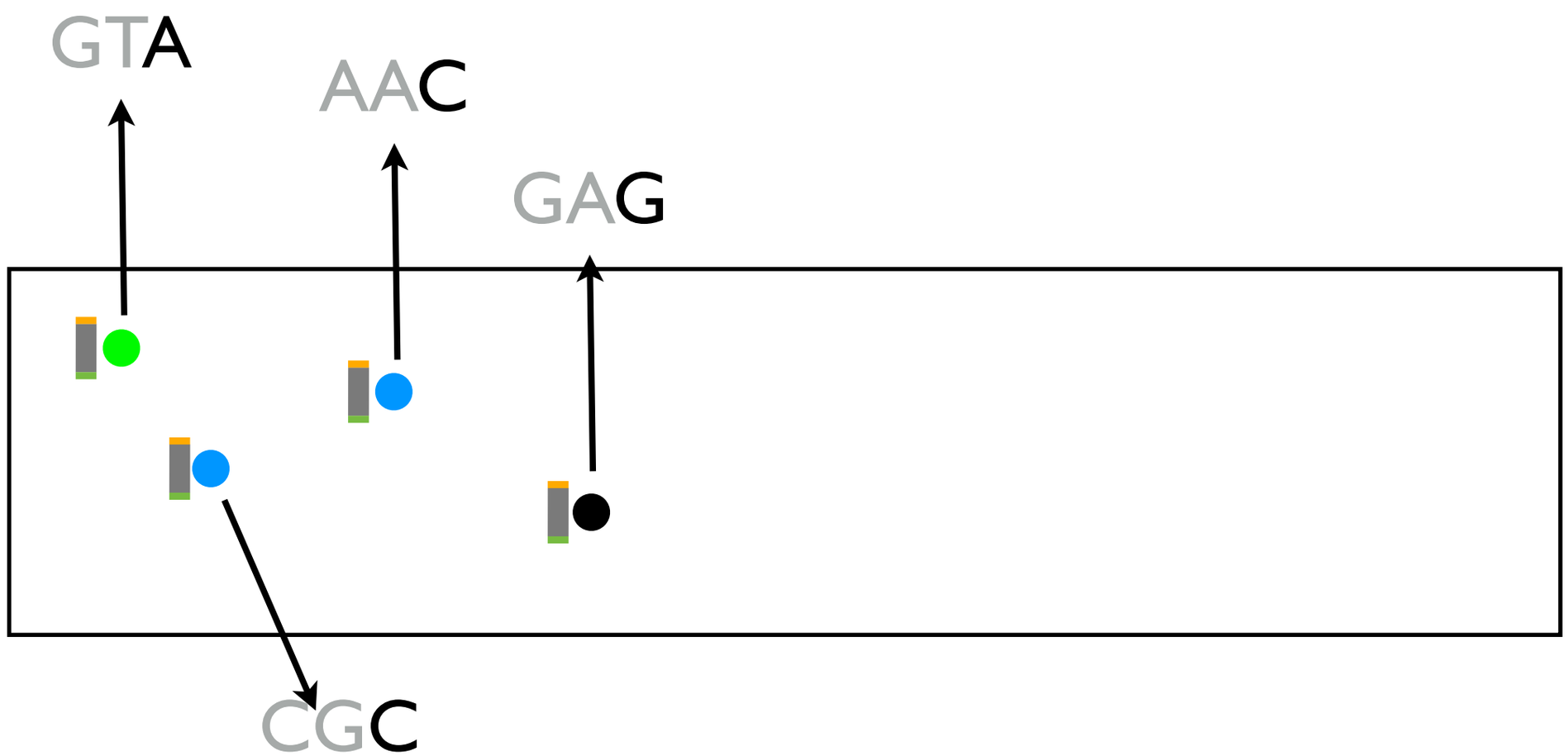
1st Cycle



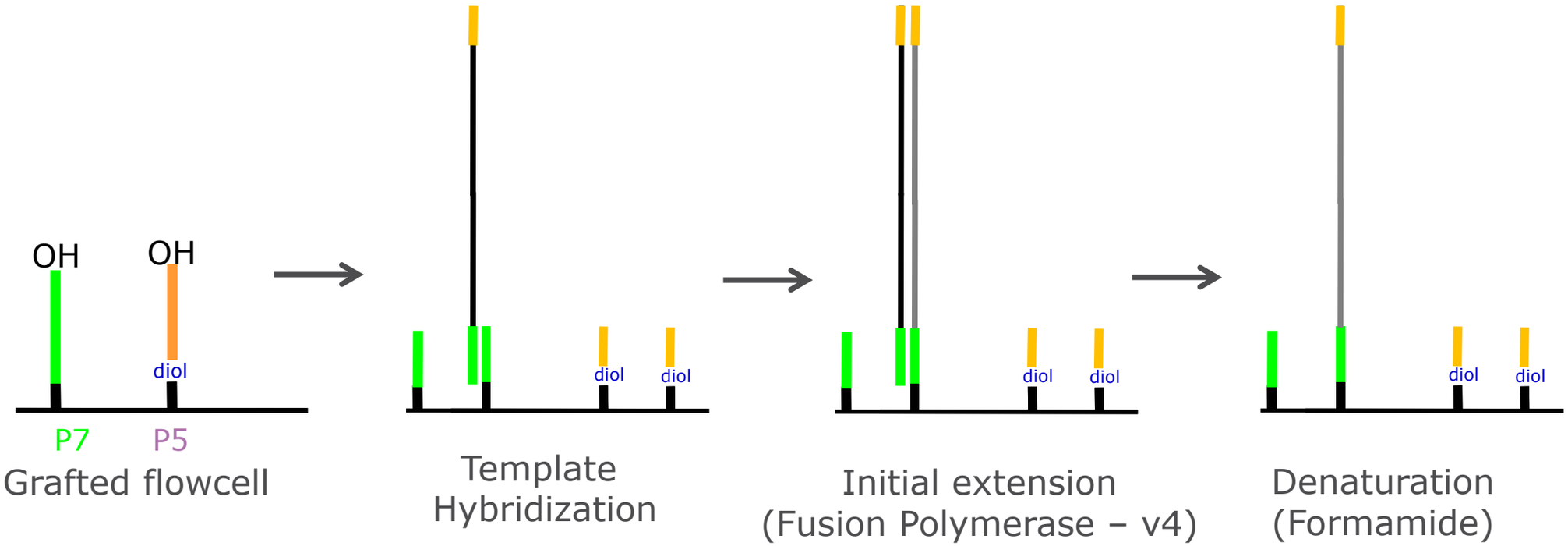
2nd Cycle



3rd Cycle



Cluster generation – hybridization and amplification



Grafted flowcell

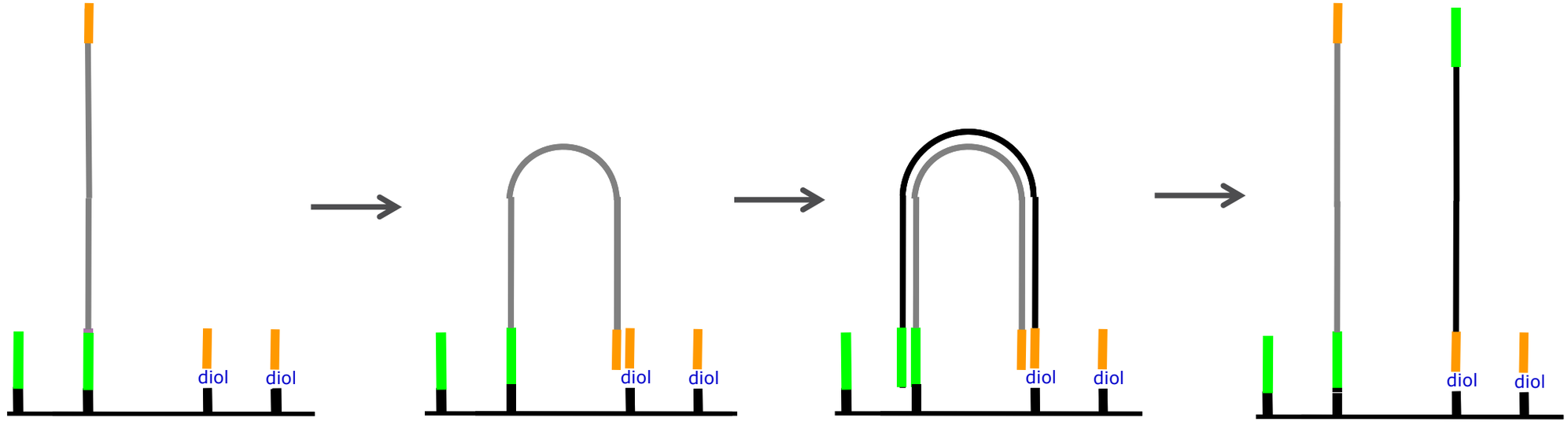
Template Hybridization

Initial extension (Fusion Polymerase – v4)

Denaturation (Formamide)

illumina

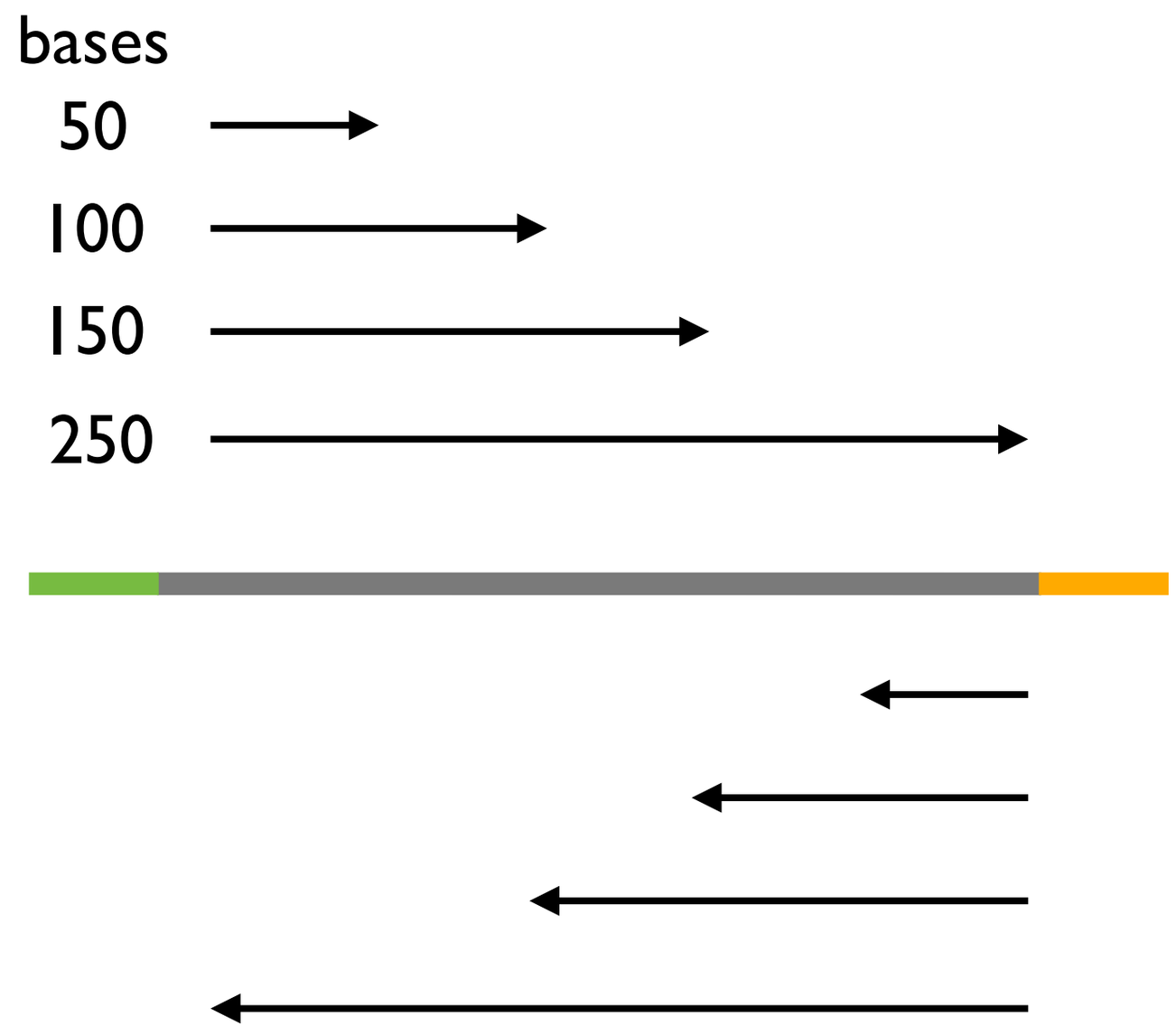
Cluster generation – hybridization and amplification



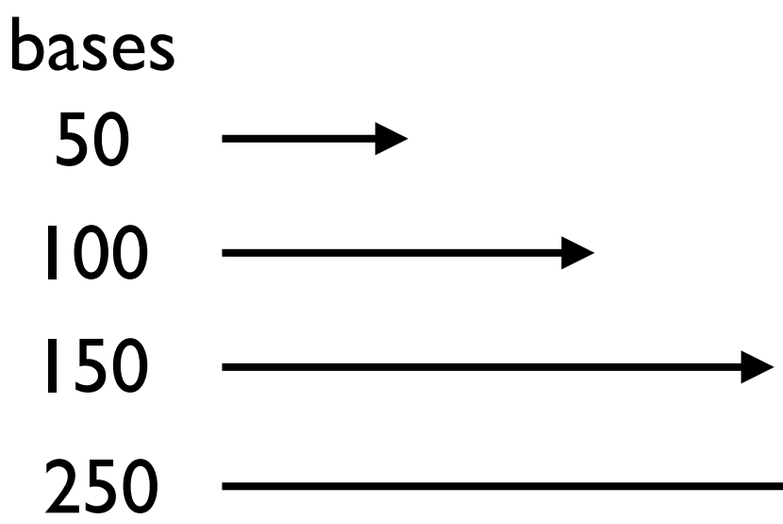
illumina

Paired End

Paired End



Paired End



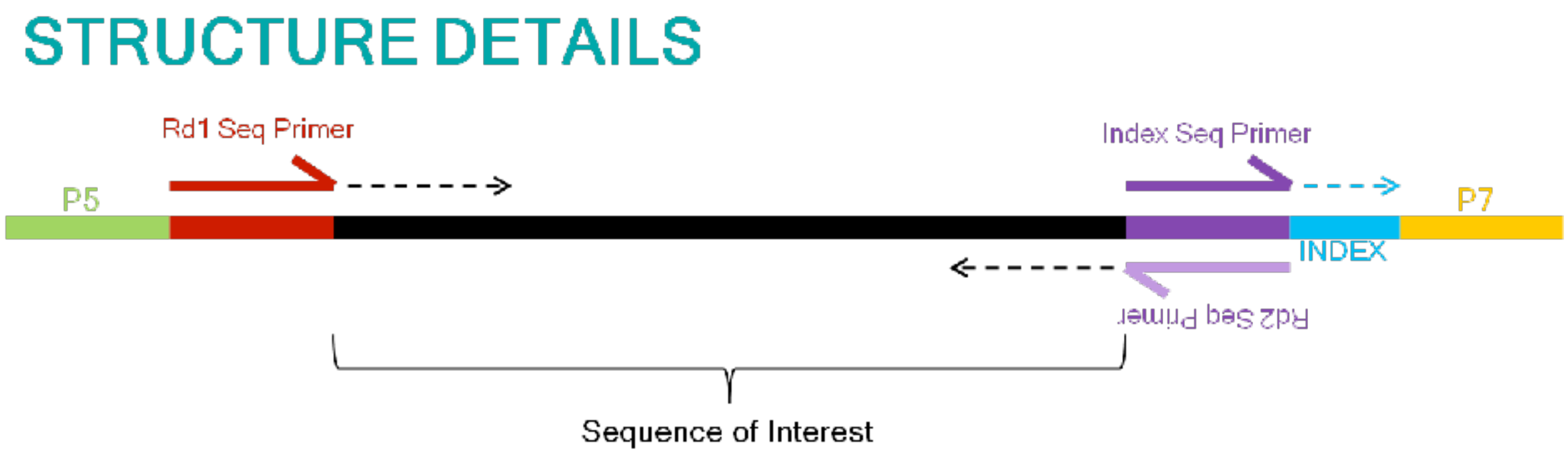
Sequencing Library

Amplicon Library

Shotgun Library

Stopped Here 8/9/2018

Multiplexing (Barcodes)



http://nextgen.mgh.harvard.edu/IlluminaChemistry.html

MiSeq, NextSeq, and More Seqs

	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 x 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

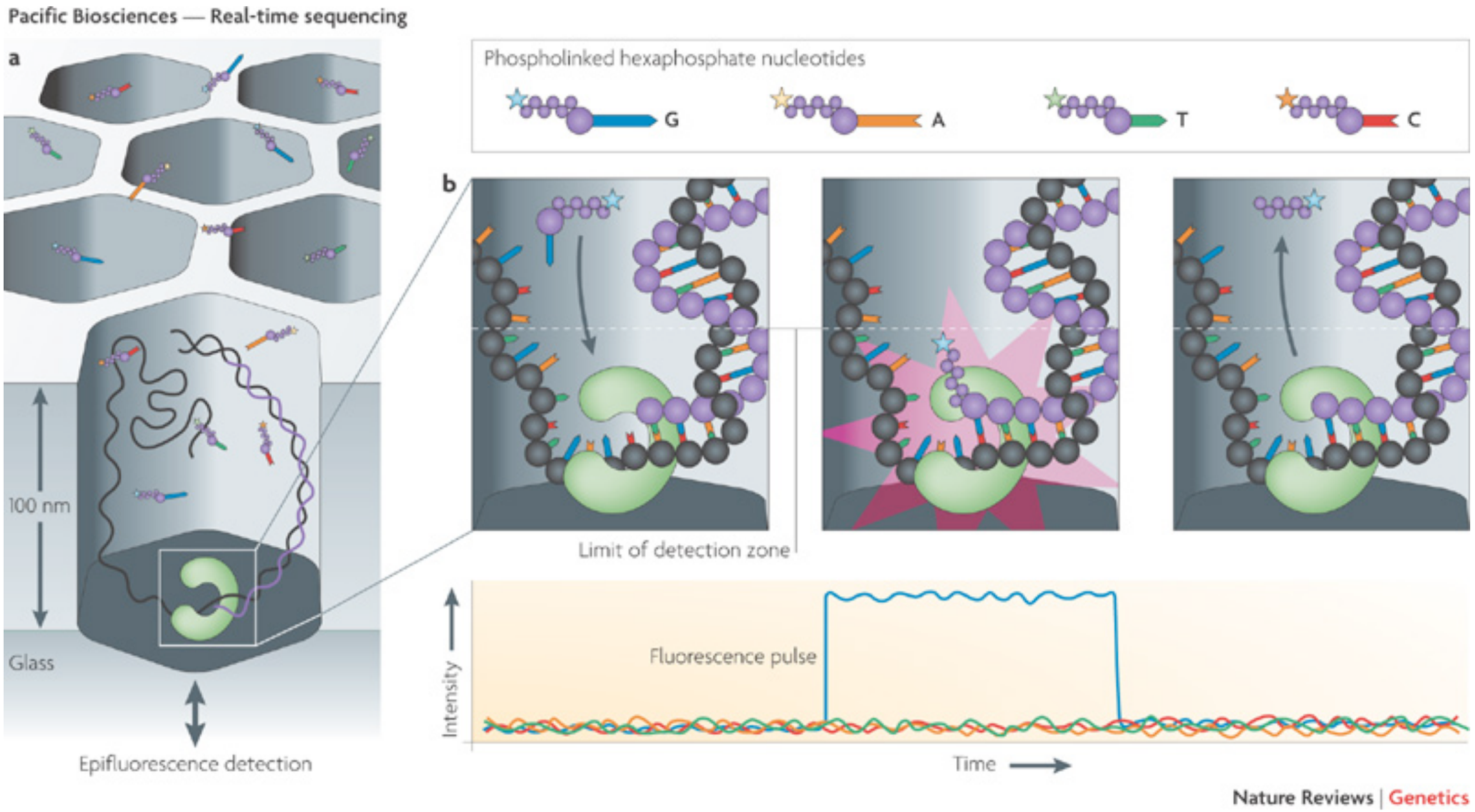
Illumina Video

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

Single Molecule Technologies

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

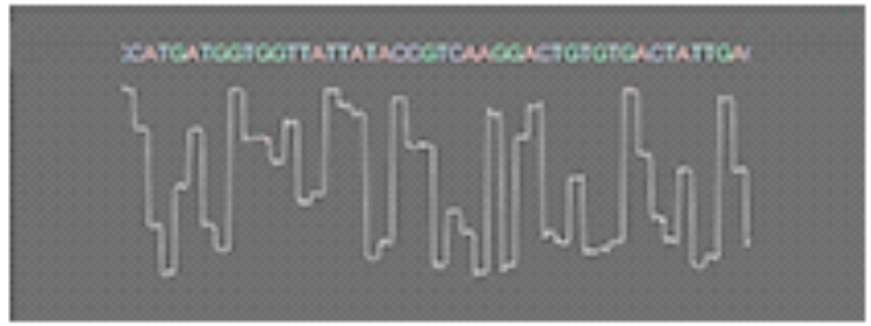
Pacific Biosciences



Nature Reviews Genetics 11, 31-46, doi:10.1038/nrg2626

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

Oxford Nanopore



https://www2.nanoporetech.com/science-technology/how-it-works

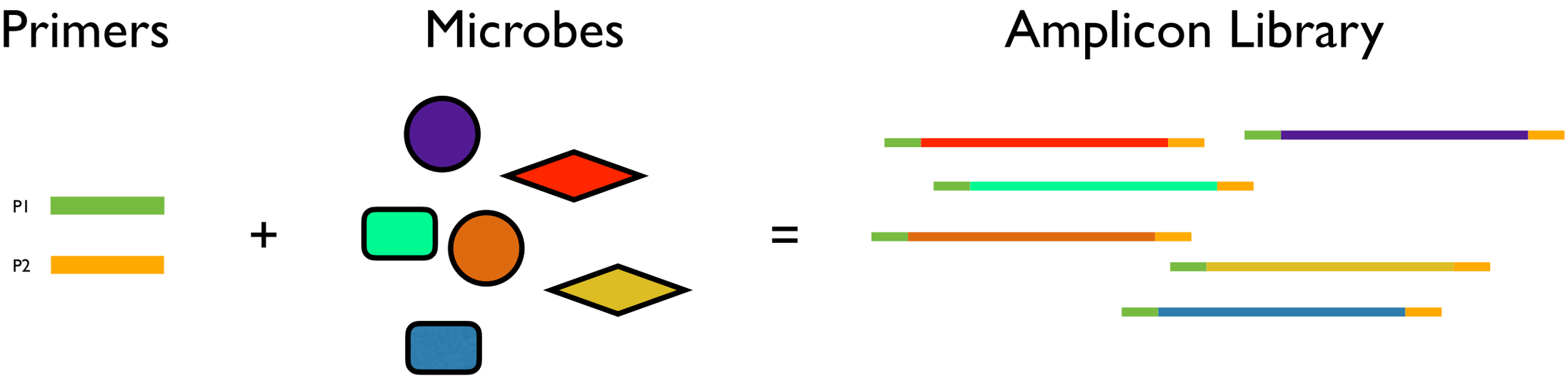
Sequencers

https://www2.nanoporetech.com/images/product-page/MinIOn-Banner.jpg
http://www.gatc-biotech.com/en/gatc/sequencing-technologies/pacbio-rs-ii.html
http://www.dnavision.com/illumina.php

Method	Read length	Accuracy	Reads per run	Max Output	Run Time	Cost ($/ Mb)	Pros	Cons
Sanger	400-900 bp	99.9%	1	900 bp	0.3-3 hours	$2400	Long reads.	Expensive. Low Output
454	700 bp	99.9%	1 million	700 Mb	24 hours	$10	Long reads Fast	Expensive. Homopolymer errors.
Ion Torrent	-400 bp	98%	up to 80 million	32 Gb	2 hours	$1	Cheap equipment Fast.	Homopolymer errors. Short reads
Illumina	50-600 bp	99.9%	1 million to 3 billion	1500 Gb	1-11 days	$0.05 to $0.15	High yield Cheap	Equipment expense. Short reads
PacBio	>10kb ave. >40kb max	87%*	50,000	500 Mb	0.5-4 hour	$0.13 to $0.60	Very Long reads Fast Detects base modification	Homopolymer errors. Moderate Output. Equipment expense.

https://en.wikipedia.org/wiki/DNA_sequencing

Amplicon Sequencing



FASTQ

- FASTA with Quality

- https://en.wikipedia.org/wiki/FASTQ_format

FASTQ: Filenames

“What’s in a name? that which we call a **FASTQ**
By any other name would smell as sweet”

— *Romeo and Juliet*, William Shakespeare

FASTQ: Filenames

Sample Name	Barcode	Lane #	Read #	File Part
1C	TAAGGCGA	L006	R1	001.fastq.gz
1C	TAAGGCGA	L006	R1	002.fastq.gz
1C	TAAGGCGA	L006	R1	003.fastq.gz
1C	TAAGGCGA	L006	R1	004.fastq.gz
1C	TAGGCATG	L003	R1	001.fastq.gz
1C	TAGGCATG	L005	R1	001.fastq.gz
2C	CGTACTAG	L003	R1	001.fastq.gz

FASTQ: Format

- **FASTA Format**

```
>NC_007779.1 Escherichia coli str. K-12 substr. W3110, complete genome
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGCTTCTGAACTG
GTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAATATAGGCATAGCGCACAGAC
AGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACCATTACCACCACCATCACCATTACCACAGGT
AACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGG
TAACGAGGTAACAACCATGCGAGTGTTGAAGTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTTGCCG
ATATTCTGGAAAGCAATGCCAGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCACCAACCACCTGGTG
GCGATGATTGAAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCCGAACTTTT
GACGGGACTCGCCGCCGCCCAGCCGGGGTTCCCGCTGGCGCAATTGAAAACTTTCGTCGATCAGGAATTTGCCCAAATAA
AACATGTCCTGCATGGCATTAGTTTGTTGGGGCAGTGCCCGGATAGCATCAACGCTGCGCTGATTTGCCGTGGCGAGAAA
```

Header Sequence

- **FASTQ Format**

```
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CTGCCAGTTGAACGACGGCGAGCAGTTATAAGCCAGCAGTTTGCCCGGATATTTCGCGTGGATAGCTTGTGCAAAGCGACGCGCCAGTTCCAGATCCGGCG
+
AAABBFFFFFFFGGGGGGGGGGGGHHHHHHHHHGHGHGHHHHHGHHHGGGGGHHHHGGGGGGGHHHGHHFFHHHHHGHGGGGGGGGGGHHHHHHHHHHHGGG
```

Quality Score

FASTQ: Format

```
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CTGCCAGTTGAACGACGGCGAGCAGTTATAAGCCAGCAGTTTGCCCGGATATTTCGCGTGGATAGCTTGTGCAAAGCGACGCGCCAGTTCCAGATCCGGCG
+
AAABBFFFFFFFGGGGGGGGGGGGHHHHHHHHHGHGHGHHHHHGHHHGGGGGHHHHGGGGGGGHHHGHHFFHHHHHHGHGGGGGGGGGGHHHHHHHHHHHGGG
@M00698:36:000000000-AFBEL:1:1101:15928:1413 1:N:0:0
GTAAAGTCCTGAGTGATACCGGCAACTTTTACCCCCAGTCCCACTTTCGAACCGGCAAACATATCGGCAAAAGAGGCCGTGCCTGATTTAAAGCCGTAGGT
+
1>AAAFFFFFFFGGGGGGGGCEGGGHHHHHFHHHHGGGHHHHHHHHHGHFA0FEGGGGGHGGHHHHEECEEHFEFH/E/E/>EGGHBHHHDGHHHHC<?E/1
@M00698:36:000000000-AFBEL:1:1101:15876:1414 1:N:0:0
CCGGAACAATCGCTGGCAGGCTTTTTACGTCCGGGTTATAGCGATAGCGGTCTCAATATTCATCAGCCCGCTTTCGCTGGCGGGTGTCGATTGTCGCCT
+
ABCCCCDFFFFEGGGGGGGGGGHHHHHHHGHHCGGGFFHHHHHHGGHGGHGGGGGHHHHHHHHHHHHHHGGGGGGHHHHGGGGGG<CHGHGHEHDGGF
@M00698:36:000000000-AFBEL:1:1101:15340:1413 1:N:0:0
CCACTAACAAACTAGCCTGATTAAGTTTTAACGCTTCAACCCCAGGCAGGGCTTCCACGCGATCTCTTTTGGGTTTGACCTCTCTTGATCCCCGTCCTAAG
+
AABABFFFFFFCGGGGGGGGGGFHHGFHHGFHGGGGGHHHHGGGGFHGGFGGHHHCFGFFEFCGHHFHHHHGGAFGGFFFFHHHHHHHHHHHHGEEGGGHHB
@M00698:36:000000000-AFBEL:1:1101:16045:1413 1:N:0:0
GTAGCATTATCAGAGAGTTGCCATTCACGCATTGGCTTAACCGCGCGCAGACCATCAACAGTCACTTTGGCGTCAAAGACATTAGGCGTGCAGTATTTTTT
+
?ABCCFFFFFFFGGGGGGGGGGHHHHHHHHHGGGHHHHHHHHHHGGGGGGGGGHHHHHHHHHHHHHHHHHHHHHGGGGGHHHHHHHHHHHGGGGHHGHHHHHHG
@M00698:36:000000000-AFBEL:1:1101:17191:1414 1:N:0:0
TCGGCACCAATATAGGTAACGCATGGTTCACCGTCTTCAGCTACGGCGGCGATTTTGGTCAGGATCGGTGCTACCAATTCATAGGCTTCTTTCTGGCCAC
+
ABAABBBBFBFFGGGGGGGGGGGGGHHHHHHHHGHGHGHHHHHHHHGGGGGGGGGHGHGGHHHHHHHHHGGGGGHHHHHHHHHHHHFHGHHHHHHHHHHBFHHG
@M00698:36:000000000-AFBEL:1:1101:16186:1414 1:N:0:0
TCAGGGTCGTCCGTGGAGCAAACATAGCCCTGAGGCTTATTGAGCATGAAGTAACGTGGACCGTGTTGCTGCGCCAGCGGGTTGCCATCGTAAGCGACAT
+
BCCCCFFFCCCCGGGGGGGGGGHHHHGHHHHHHHHGGHHHHHHHHHHHHHHHHHHHHHGHHGGHHGGHHHHHHHHHGGGGGGGGGGGGGHHHHHGHHHFHGGGGG
@M00698:36:000000000-AFBEL:1:1101:15394:1415 1:N:0:0
ACTTCTTGAAGGTACGTTGCAGGCTGGCACTTCTGCCGTTTCTGATAAGTTGCTTGATTTGGTTGGACTTGGTGGCAAGTCTGCCGCTGATAAAGGAAAGG
+
@BBCCFFFFFFFGGGGGGGGGGGGGHHDGHHHHHHHHHHGEGHGHHHGHHHHHHHHHHHEHHHHHHHGGHHHHHHHEGGHHGHHHHHHGGGGFHHHHHHHHHHH
@M00698:36:000000000-AFBEL:1:1101:14326:1417 1:N:0:0
TGGCGCAACTAACAGAACGTCTTGCGTTTTGTTGGCGAAGCCGCTGGTGTTTGTAAATTTATTAGTGATCGGCGTAGCGTTACGGGTTTCACCGTAGTTCG
+
ABCCCCCCCFFFGGGGGGGGGGGHHGGGGGHHHHHHGGGGGHGGGGGFGFFHHHHHHHHGHHHHHHHHHHGHGGGGGFFAE>EGGGGGEGGHHHHGGHGGHHG
@M00698:36:000000000-AFBEL:1:1101:15479:1417 1:N:0:0
GCCCCCCTCACTCTGACTTTAGTGTGCGCCTTTTGCCTGCGCCACGATCTCCTCGACATTTTCCGGCATTGGTTCATACGGCGTTTCTTTCGCCGTTGG
+
@BCCCCCCCFCFGGGGGGGGGGGHHHHGGGGGHHHHHHHHGGGGGGGHHHHHHHHGGGGGHHHHHGGGGGHHHHHHHHHHHGGGGGGHHHHHGGGGGGGH
@M00698:36:000000000-AFBEL:1:1101:16850:1418 1:N:0:0
CGTTTTCTTCATCGCGCTCTTTGCTGCCTAACAGCGTGCGCCAGCCTGCTTCAGCAAGAAAACGCGCTTTAGCGACAAATTTGCCTTTGGCAATGTCCAGT
+
AAACCCFFFFFFGGGGGGGGGGGHHHHHHHHHHHHHGGGGGGGGGGGHHHHHHHHHHHHHHHHHHHGGGGGGGHGHGGGGGFHHHHHHHHHGHHHHHHHHHHHHG
@M00698:36:000000000-AFBEL:1:1101:16255:1418 1:N:0:0
CCGGCTTGCTGGTTGCAGCCGTTGCTGTACTTGATGTCAGGCGTGCCGGTGCCGTATTTTCAAACGGTGATGCCGGACGCGACTCTACGGGTTTGACATCG
+
AAABCCCCCFFFGGGGGGGGGGGGGHHHHHHHHHHHHHHHHHHGGGGHGGGGGHGFGHHHHHHHHHHGGGHGHHHHGGGGGGGGGGGGHHHGGGCGGGHHHHH?
@M00698:36:000000000-AFBEL:1:1101:17071:1419 1:N:0:0
CACGCCACCTTTATCCAGTTTGCGGCTTTGCGAAGTGGCGAGCCACAGCGCGTTTTCTTGCTGGCTATAAGCCATTTCGTAGGCACCTTTACCTACCGCTT
+
AABBBBBBFFFFGGGGGGGGGGGHGGGGGHHHGGGGFHHGGGGGGGHHHHHGGGFGGGHHHHHHHHHHGHHHHHHHHHHHHHHHHHGGHGHGHHHFHHHHHHGGGG
@M00698:36:000000000-AFBEL:1:1101:15606:1420 1:N:0:0
CGGGTTTTTAACTTTCAGCCACGGGCCACCGTCGATCAGTTCACCGCCAAACTCTTCACGCGCCAGCTGGTAGCCCCAGTCTTTAAACGCTCCTTCGGTGA
+
BBCDBBCCDCFFGGGGGGGGGGGGGGGGGHGGGGGGHGHHHHHHHHGGGGGHHHHHHHHHHHHGGGGGGHHHHHHHHHHGGGHHHHHHHHHHGGGGGHHHHGFGG
@M00698:36:000000000-AFBEL:1:1101:15945:1421 1:N:0:0
TCATTGTTGCGCGGTATTCGCGCCCGTTGGTCGAGTAGCAGAAGGGGATTTTAAACCGTTGTTTGCCGCTGGTGTCCTGCCAGCTGGTTTCATACTCTGG
+
?ABABFFFFBBBFGEEGGGGGGGGGGGHGHCHGCGGGHGHHHHHHHGGGGHGHHHHGHGEGDGHHHHHGGG?EB2GGHGHHHGHGHHHHGHHFHHHHHGGH
@M00698:36:000000000-AFBEL:1:1101:16329:1421 1:N:0:0
GCCCCGACAGCTGTATGCATAGCGATAAATTCCAGCAGGCCGGGACGCCGGTCTATTTTGCCCCAGAGTGTCAATGTTAGACTTGACGGACATTGTGCAG
+
3ABCCCCCCCCCFGGGGGGGGGGHGGHGGHHHHHHHHHHHHGGGGGGGGGGGGGGGHHHHHHHHHHGGGHHHHHHHGHHHHHHHHHHHHHGGGGGHHHHHHHHH
```

Quality Scores

- Go to Notebook:

  Wk4_Day4_AM/quality_scores.ipynb

FASTQ: Read Files

Combined_R1.fastq.gz

```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0
TTACGCTAACAGGCGGTAGCCTGGCAGGGTCAGGAAATCAATTAACTCATCGGAAGTGGTGATCTGTTCCATCAAGCGTGCGGCATCGTCAAAACGCCC
+
ABBBABBBAFFFGGGGGGGGGHGGHGGGCG2GF3FFGHHHHHHGGFGHEHHGGGEHHHHAGGHHGHHHFFDHFHHHGEGGGG@F@H?GHH/GBEFGGG
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CTGCCAGTTGAACGACGGCGAGCAGTTATAAGCCAGCAGTTTGCCCGGATATTTCGCGTGGATAGCTTGTGCAAAGCGACGCGCCAGTTCCAGATCCGGCG
+
AAABBFFFFFFFGGGGGGGGGGGGHHHHHHHHHGHGHGHHHHHGHHHGGGGGHHHHGGGGGGGHHHGHHFFHHHHHGHGGGGGGGGGGHHHHHHHHHHGGG
```

Combined_R2.fastq.gz

```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 2:N:0:0
GGAAGATGCGGCGACGGCTGAAATTTCCCGTACCTCGATCTGGCAGTGGATCCATCATCAAAAAACGTTGAGCAATGGCAAACCGGTGACCAAAGCCTTGT
+
ABBABFFFFDBDGCG??FFGGGHGHFEG3EAEGGFHAE3GFBGGHGGGHHCFGHFGBGHFHHDFEGGHFHEFHHHH3BFGF0GFEGGGGGHHA/FGHFHHH
@M00698:36:000000000-AFBEL:1:1101:16483:1412 2:N:0:0
GCTTCTTCCGTACTCATGCGGGCATTGAGCAAGCGATCAGCCGTGGCCTGGCGTATGCGCCATATGCTGACCTGGTCTGGTGTGAAACCTCCACGCCGGAT
+
CCCCCFFFFBFFGGGGGGGGCECGHHHHHHHHHHGGHGGGGHGGCGCHHGFHGGGGHHGGGGGHHHHHHGHHHHHHHHHHHHHHEHGHHHGHHHHGGGGGGG
```

Combined_I1.fastq.gz

```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0
AGTTCC
+
CCCCDF
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CCTGTC
+
A11>>1
```