INTRODUCTION TO NEXT-GENERATION SEQUENCING

Presented by: Farid MUSA (252021030)

Izmir Institute of Technology, Bioengineering Department, Urla/IZMIR

OUTLINES

- Introduction DNA Sequencing
  - Principles of DNA Sequencing
  - Maxam–Gilbert Sequencing
  - Sanger Sequencing
  - Shotgun Sequencing
  - Primer Walking
- Introduction to Genome Sequencing
  - Hierarchical shotgun sequencing
  - Whole-genome shotgun sequencing
- Next-Generation Sequencing (NGS)
  - NGS Sequencing Technology
  - NGS Principles
  - Library Preparation
  - Amplification Methods
  - Sequencing Methods
  - NGS Platform Comparison
- Introduction to DNA Sequence Alignment
  - Needleman-Wunsch Algorithm

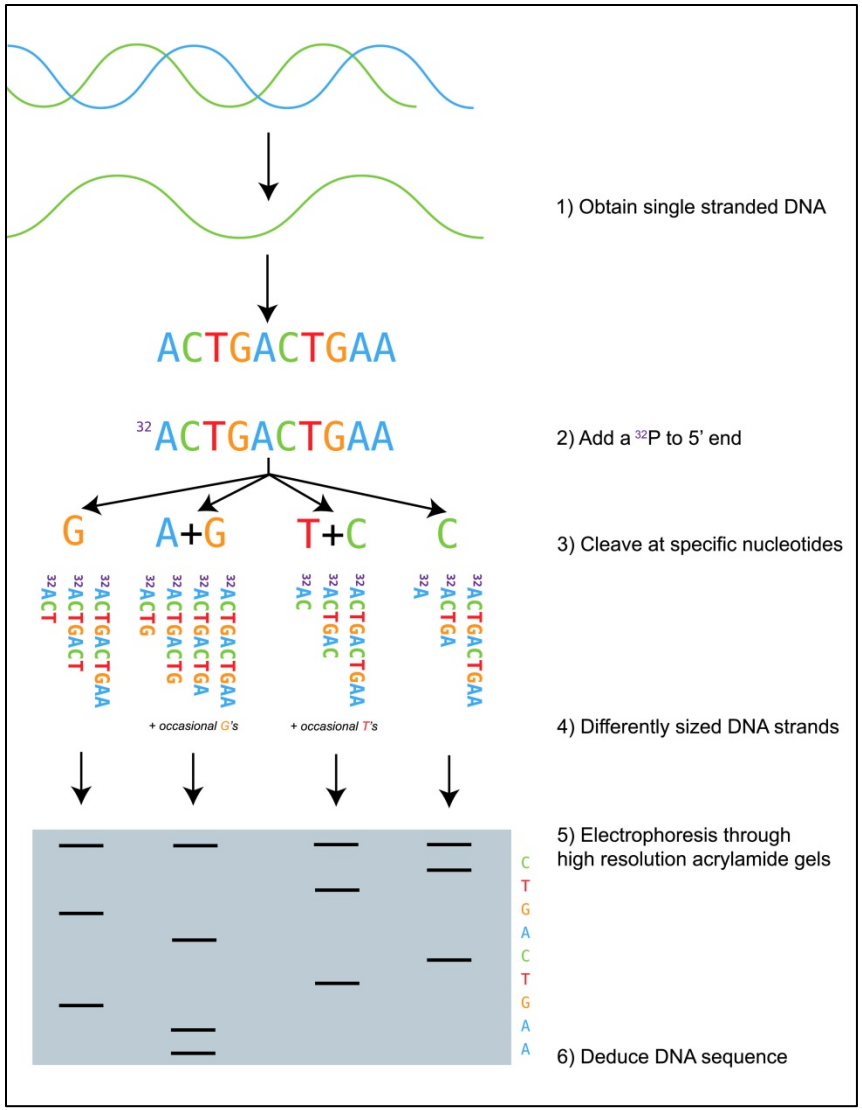
PRINCIPLE OF DNA SEQUENCING

- **DNA sequencing** is the process of finding the order or sequence of nucleotides in DNA molecule
- In 1977 two DNA sequencing methods were developed and published. Now known as **First Generation Sequencing:**
  - **Maxam–Gilbert Sequencing by chemical degradation method**
  - **Sanger Sequencing by chain termination method**
- Although initially Maxam–Gilbert was more popular eventually Sanger sequencing method become more preferred due to several technical and safety reasons
- *It is important to note that at that time there was no PCR, which was discovered in 1983*

(Pareek, Smoczynski, & Tretyn, 2011)

MAXAM–GILBERT SEQUENCING

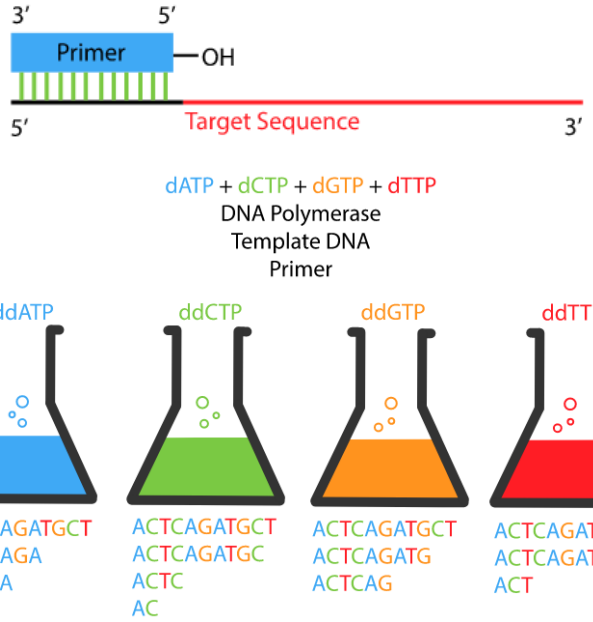
1. Double-stranded DNA is **denatured** to single-stranded
2. **5' end** of the DNA fragment that will be used for sequencing is **radioactively labeled** by a kinase reaction using gamma-^{32}P.
3. DNA strands are then **cleaved at specific positions using chemical reactions** in 4 different reaction tubes. For example
   - Dimethyl sulphate selectively attacks purine (A and G)
   - Hydrazine selectively attacks pyrimidines (C and T)
   - A+G means that attack will be on A but may also cleave at G
4. Cleaved fragments in each tube are then **passed to gel electrophoresis for size separation**
5. Under **X-ray film** gel reveal bands with radiolabeled DNA molecules.
6. Fragments are then **ordered by size** to help **derive the original sequence** of the DNA molecule

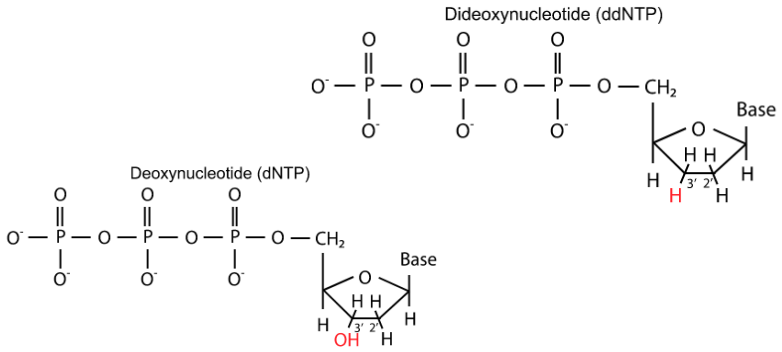


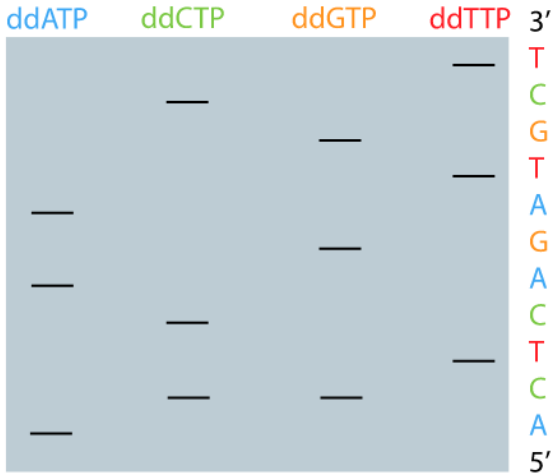
(Pareek, Smoczynski, & Tretyn, 2011) (SnipAcademy, 2017)

SANGER SEQUENCING

1. **Clone one DNA fragment** into vectors and **amplify**
2. **Attach primers** to the fragments
3. To each tube add:
   - **4 standard dNTPs**
   - Only **one type of ddNTPs**
4. Add **DNA polymerase to initiate** synthesis
   - DNA polymerase will add dNTPs and will **stop when encounters any ddNTP**
5. Separate single stranded DNAs in **gel electrophoresis**
6. **Read** the sequence from bands



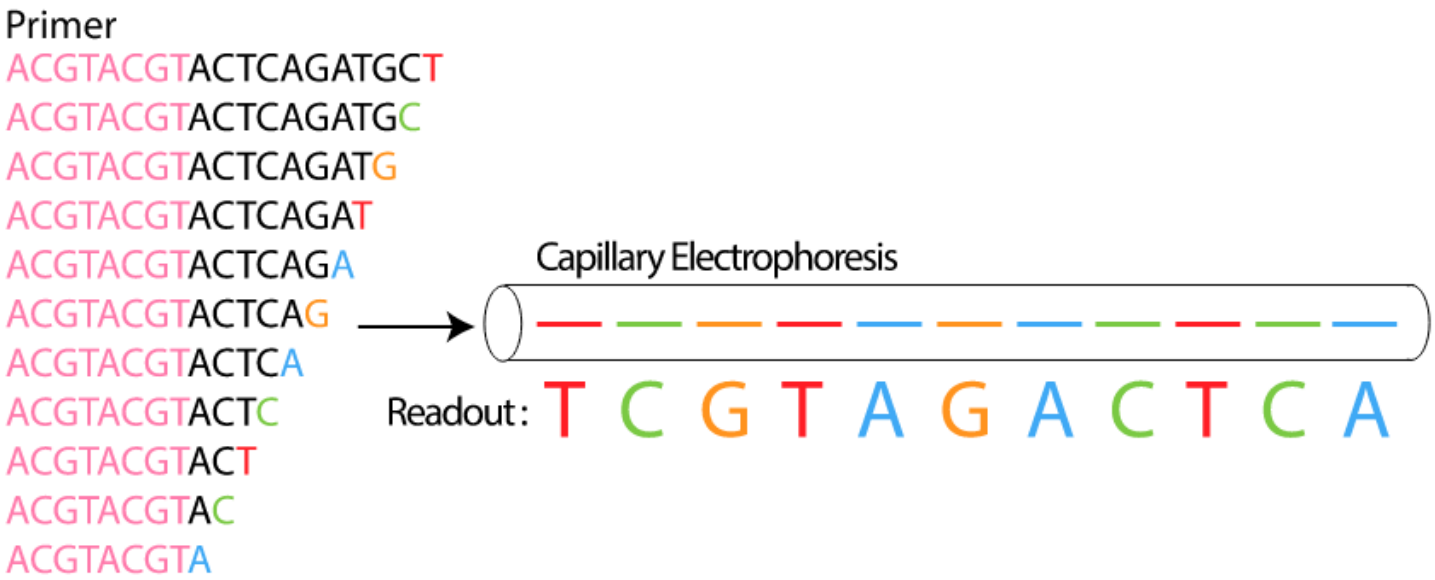




(Shendure & Ji, 2008) (Snipcademy, 2017)

AUTOMATED SANGER SEQUENCING

- Same process but **ddNTPs are fluorescently labeled with 4 different colors**
- Instead of gel electrophoresis **high-resolution capillary electrophoresis** is **coupled to four color detection of emission spectra** is used to **automatically read sequence** as it exits capillary



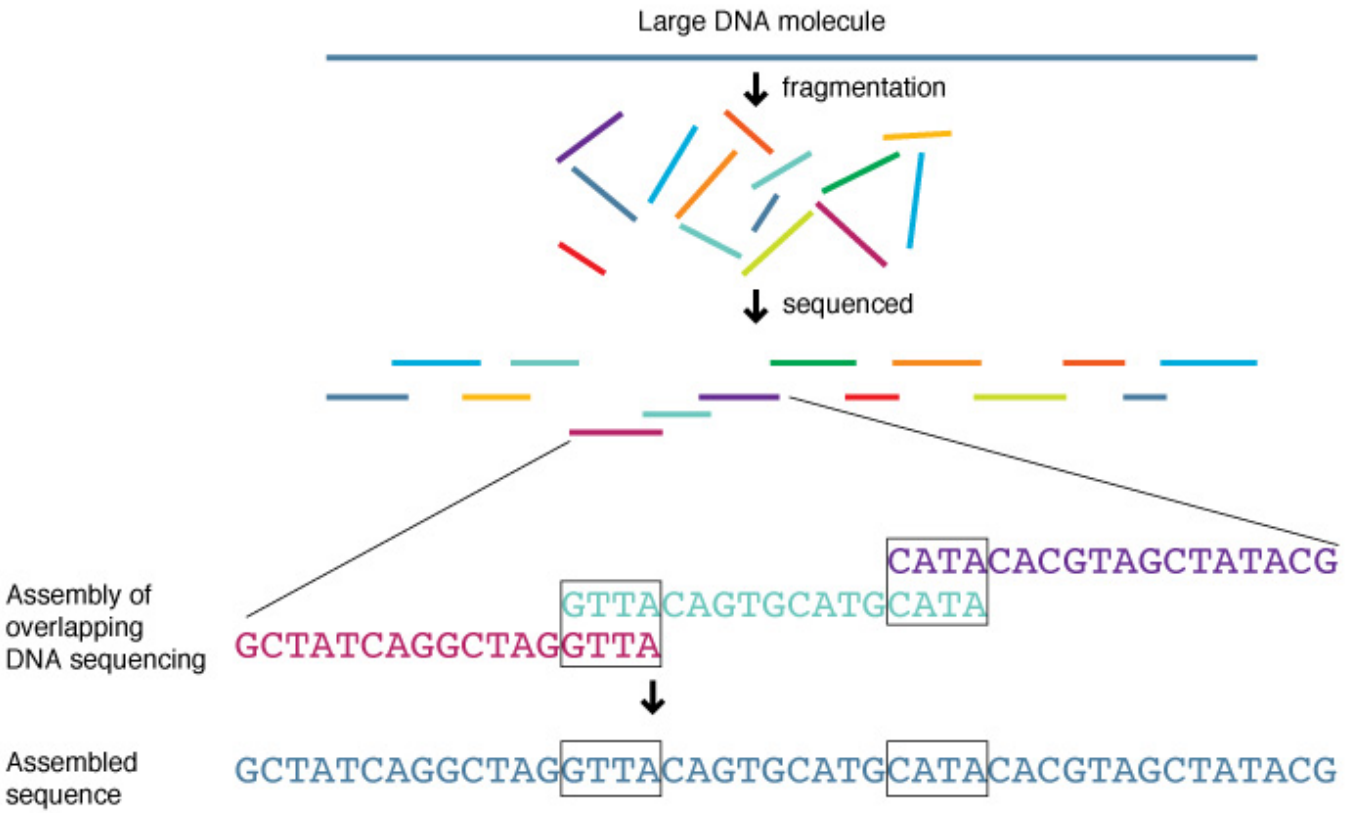
(Shendure & Ji, 2008) (Snipcademy, 2017)

SHOTGUN SEQUENCING

- **Long DNA** molecules **can not be directly sequenced** by Sanger sequencing method
- In such case **Shotgun sequencing** method must be used
- Method involve breaking down long DNA molecule mechanically or enzymatically into smaller random fragments of approximately 1000 bases long
- Followed by cloning individual fragments into **universal vector** and **amplification**
- Random fragments (**aka DNA inserts**) are then individually sequenced with universal primer of cloning vector
- Fragments that have overlapping regions are then called **contigs**(contiguous sequence) and are used to assemble original sequence

(Shendure et al., 2017) (Amy Rogers, 2008)

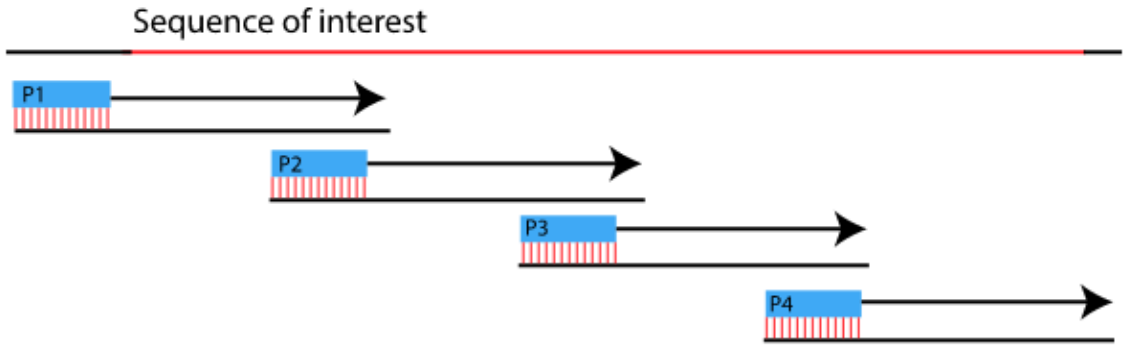
SHOTGUN SEQUENCING



https://upload.wikimedia.org/wikipedia/commons/5/54/Shotgun_sequencing_lg.jpg

PRIMER WALKING

- **Primer Walking** is required when:
  - Fragment is too long to be sequenced by Sanger method in a **single run**
  - Final assembly of shotgun sequencing have **gaps** that were not covered by **contigs**
- Simply primer walking is addition of new primer upstream(right-to-left) of first sequence
- New primers are synthesized once the first sequence is complete, annealed to the original fragment and sequencing is restarted
- If many gaps are present in final assembly method become time consuming but efficiently reveals missed fragments



(SnipAcademy, 2017)

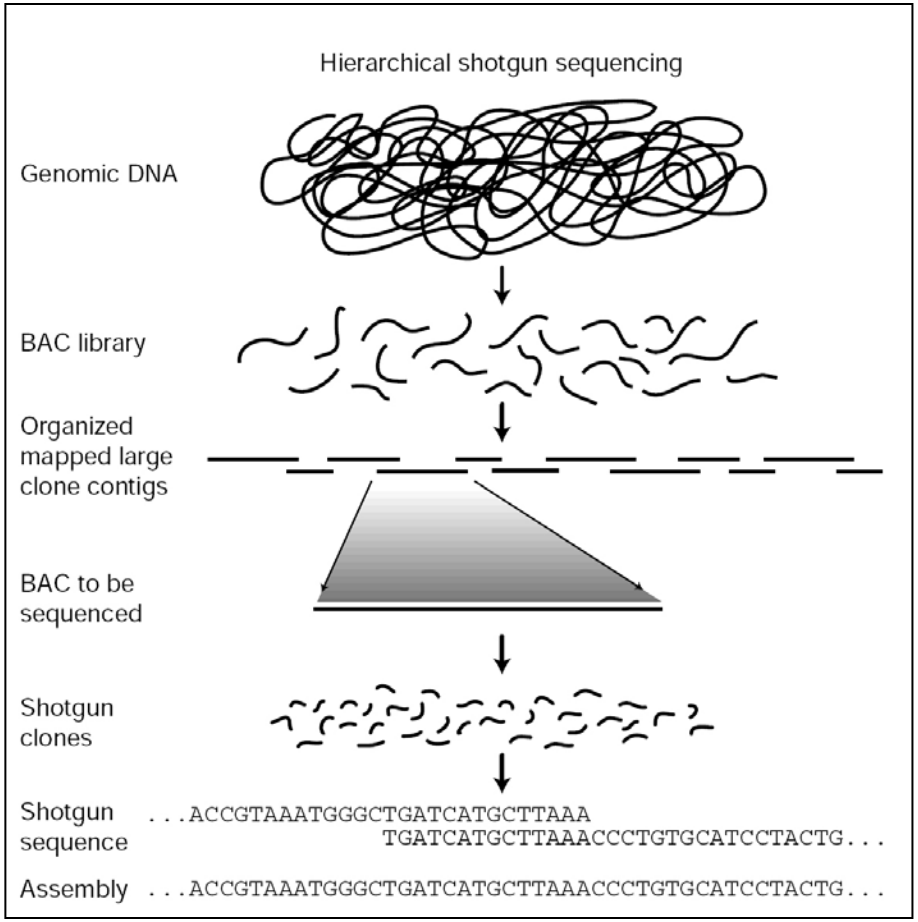
INTRODUCTION TO GENOME SEQUENCING

- Very large DNA molecules of such as chromosomal DNA or even whole genomes require slightly modified shotgun sequencing approaches also called **genome sequencing**
- There are three classical genome sequencing methods:
  - Hierarchical shotgun sequencing
  - Whole-genome shotgun sequencing
  - Double-barrel shotgun sequencing
- *Initially **Human Genome Project** used hierarchical shotgun sequencing later whole-genome shotgun sequencing method were developed. Both contributed to the sequencing of the first human genome.*

(McClean, 2005) (Chial, 2008)

HIERARCHICAL SHOTGUN SEQUENCING

- First **genomic DNA** is fragmented into relatively large pieces (~150Mb)
- Pieces are then inserted into bacterial clones BACs, grown in E. coli and recovered
- BAC libraries are then organized and mapped to find physical location of pieces, and form **physical map.** This step is also referred as **Golden Tiling Path.**
- BAC libraries are then **individually sequenced** using **standard shotgun sequencing**
- BAC sequences are **then assembled and aligned to physical map** to find original genomic DNA
- *BAC libraries are useful because each piece can be sequenced in different labs by different researchers*



(McClean, 2005) (Chial, 2008)

WHOLE-GENOME SHOTGUN SEQUENCING

- **Whole-genome shotgun sequencing** is simply shotgun sequencing but **without preparation of physical map.**
- In this method whole genomic DNA is fragmented into fragments of **relatively smaller size 100Kb**
- Fragments are then **cloned into plasmids**, **grown** in bacteria and **recovered**
- **Each fragment** in plasmids are fragmented again and then **sequenced to form contigs**
- **Long contiguous sequences** are then assembled to form **final DNA sequence**. This is also known as ***de novo sequencing***
- This method and **double-barrel shotgun sequencing** are very similar with only main difference that in the latter sequencing is performed **from both ends of DNA inserts**. This is also known as **pairwise-end sequencing**
- *Sequencing of large genomic DNAs using this method was only possible with technological advances both in DNA sequencing efficiency and computational tools that are used for assembly of contigs*

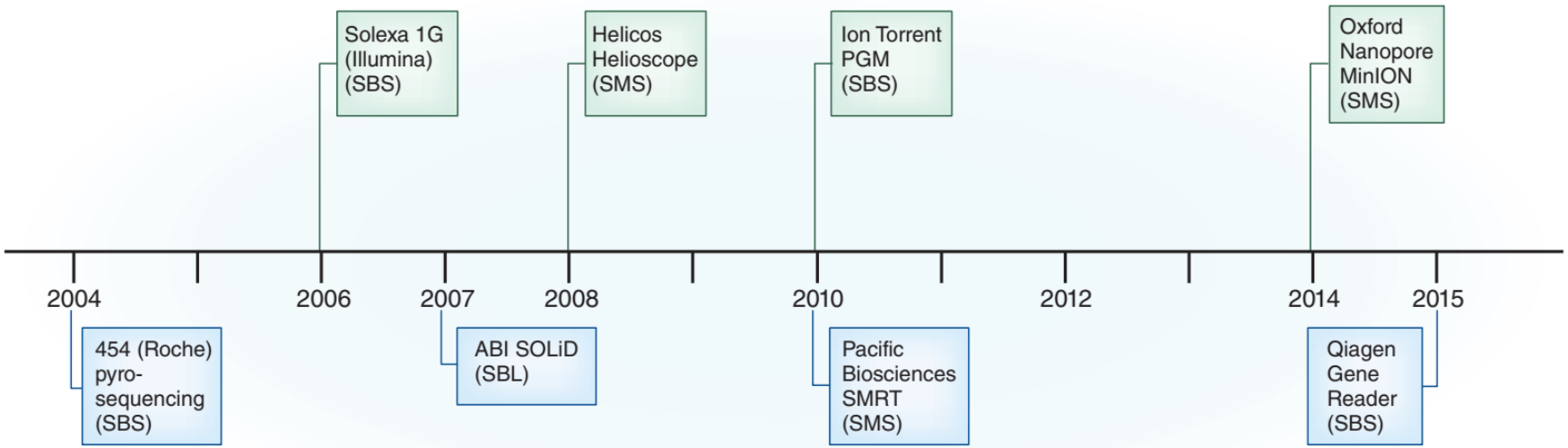
(McClean, 2005) (Chial, 2008)

NGS TECHNOLOGY

- Next Generation Sequencing (Second Generation)
  - Rosche/454 (Pyrosequencing)
  - Illumina/Solexa (Illumina Dye Sequencing)
  - ABI/SOLiD formerly Life/APG's (Support Oligonucleotide Ligation Detection)
  - Ion Torrent (Post-Light Sequencing)
  - Heliscope by Helicos BioSciences (Single Molecule Fluorescent Sequencing)
- Third Generation Sequencing
  - PacBio (Single Molecule Real Time Sequencing)
  - Oxford Nanopore (Nanopore Sequencing)

(Frese, Katus, & Meder, 2013; Heather & Chain, 2016; Kchouk, Gibrat, & Elloumi, 2017; van Dijk, Auger, Jaszczyszyn, & Thermes, 2014; Yohe & Thyagarajan, 2017)

NGS TIMELINE



(Mardis, 2017)

NGS PRINCIPLES

- Conceptually second generation NGS platforms share same work flow
  1. **LIBRARY PREPARATION**
  2. **AMPLIFICATION/ENRICHMENT**
  3. **SEQUENCING**
  4. **DATA PROCESSING**
- Methods involved in each step may differ by platform
- **Library Preparation** refers to DNA pretreatment process before sequencing and is typically combined with **amplification** step
- Actual **sequencing** is performed in the last step and the technology involved in this step is highly dependent on the preferred platform
- Generated data during sequencing step is then handled by bioinformatic tools and further analyzed as needed

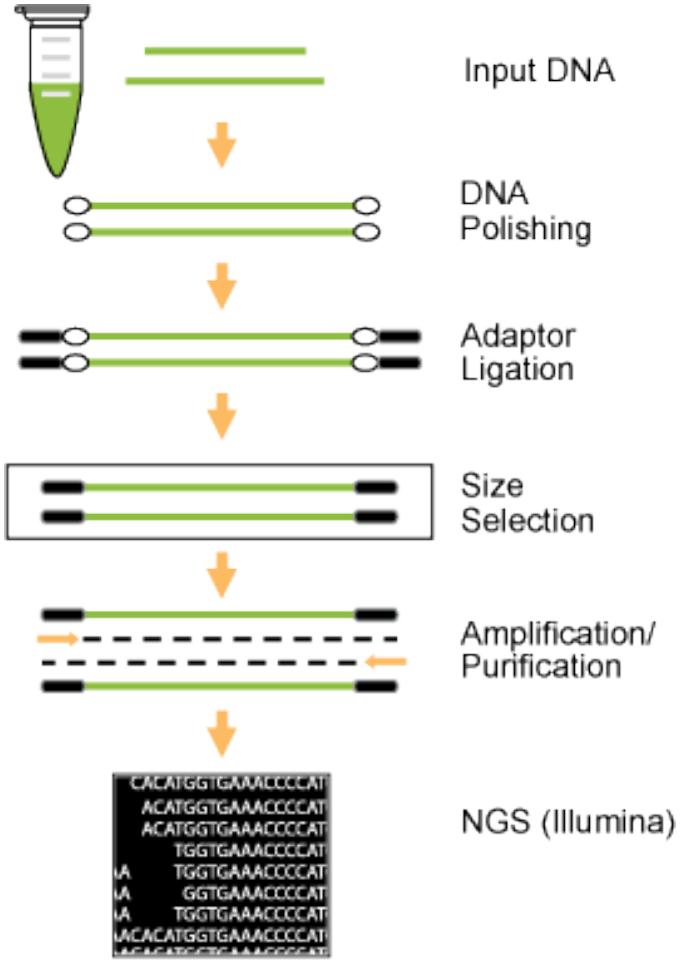
(Shendure & Ji, 2008) (Yohe & Thyagarajan, 2017)

LIBRARY PREPARATION

- *None of the second generation sequencing methods are capable of sequencing large whole DNA molecule at once*
- Prior to amplification DNA must be **randomly fragmented** and resulting fragments treated based on requirements of the platform
- Most NGS platforms have similar library preparation steps
- Libraries can be **single-end/pair-end(SE/PE)** or **mate-pair(ME)**
- Typically final DNA fragments, **so called DNA library**, must satisfy most of the following requirements depending on the platform:
  - Short DNA fragments of similar size
  - Blunt ended with no 3’ or 5’ overhangs
  - 5’ phosphate and 3’ hydroxyl groups at blunt ends
  - dA tailing (dAMP incorporation onto 3’ ends)
  - Ligated **adapters** (synthetic oligonucleotides)

(Shendure & Ji, 2008) (van Dijk, Jaszczyszyn, & Thermes, 2014)

LIBRARY PREPARATION



(EpiNext, 2017)

LIBRARY PREPARATION (SE/PE)

STEP 1 - FRAGMENTATION

- Initial DNA template must be randomly fragmented via one of the following approaches
  - Mechanically (shearing, sonication, etc.)
  - Enzymatically (restriction enzymes or transposases)
  - Chemically

STEP 2 - SIZE SELECTION

- DNA fragments (**Inserts Size**) of desired length are selected using either Solid-Phase Reversible Immobilization (SPRI) beads or Agarose gel electrophoresis

STEP 3 – END REPAIR (BLUNTING)

- Many fragments may end up with overhanging sticky 5’ or 3’ ends
- These sticky ends must be repaired by transforming sticky ends into blunt ends using **T4 DNA polymerase** and **DNA Polymerase I, Large (Klenow) Fragment** by
  - Filling of 5’ overhangs by 5’ to 3’ polymerase activity
  - Degrading 3’ overhangs by 3’ to 5’ exonuclease activity
- After blunting is complete, **5’ ends are phosphorylated** for ligation and **3’ dA tailing** is performed for Illumina or 454 platforms

STEP 4 - ADAPTER LIGATION

- Following **adapters are ligated** to **both ends** of the DNA fragments via **DNA ligase**
  - **Universal PCR primers** for PCR amplification
  - **Hybridization sequences** for surface immobilization
  - **Molecular barcodes** for multiplex pooled sequencing
  - **Recognition sites** to initiate sequencing

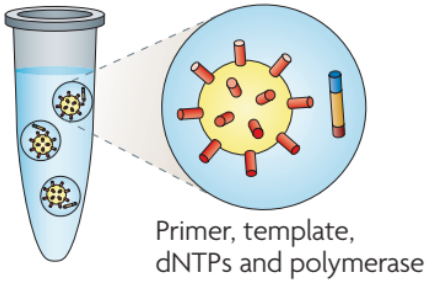
(Shendure & Ji, 2008) (van Dijk, Jaszczyszyn, & Thermes, 2014) (Head et al., 2014)

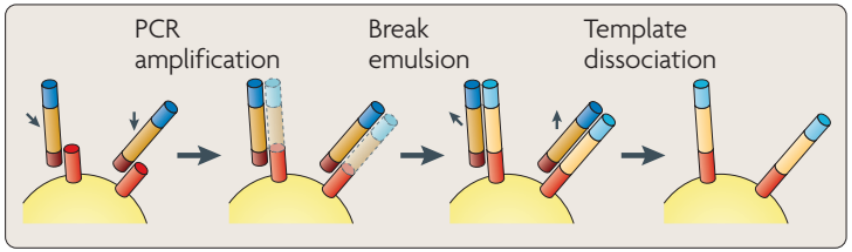
AMPLIFICATION/ENRICHMENT

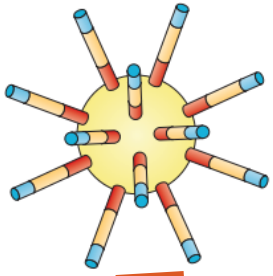
- During **amplification/enrichment** step **DNA libraries** are **clonally amplified**
- Amplification is necessary because most sequencing detection systems **are not capable of detecting one molecule.** *Except for Heliscope platform by Helicos BioSciences that filed for bankruptcy in 2012*
- Amplification methods depend on used platform but most common are:
  - **Emulsion PCR (emPCR) – 454, SOLiD and Ion Torrent**
  - **Solid-phase amplification – Illumina**

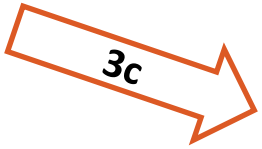
(Kchouk, Gibrat, & Elloumi, 2017) (Metzker, 2010)

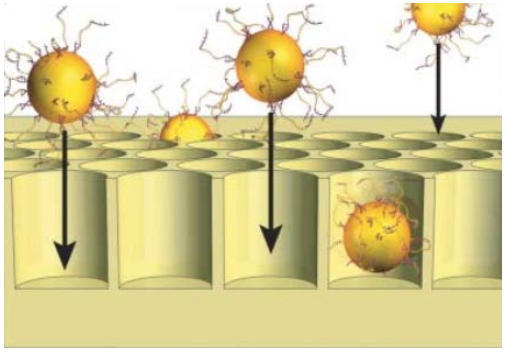
emPCR AMPLIFICATION



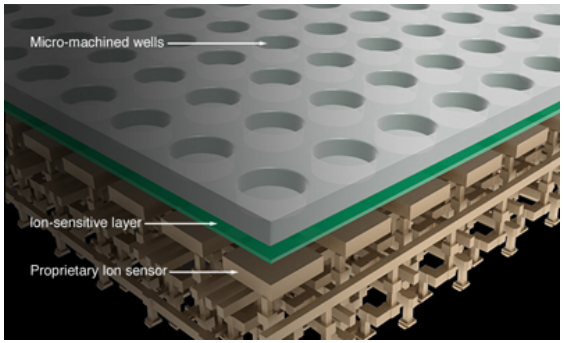




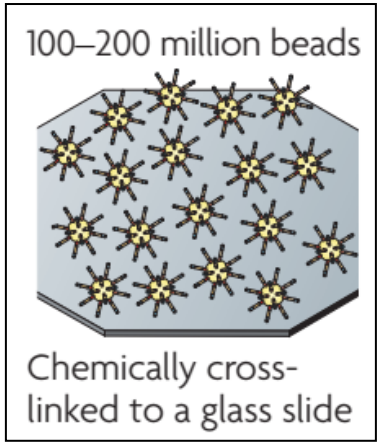




PicoTiterPlate or PTP
(Roche/454)



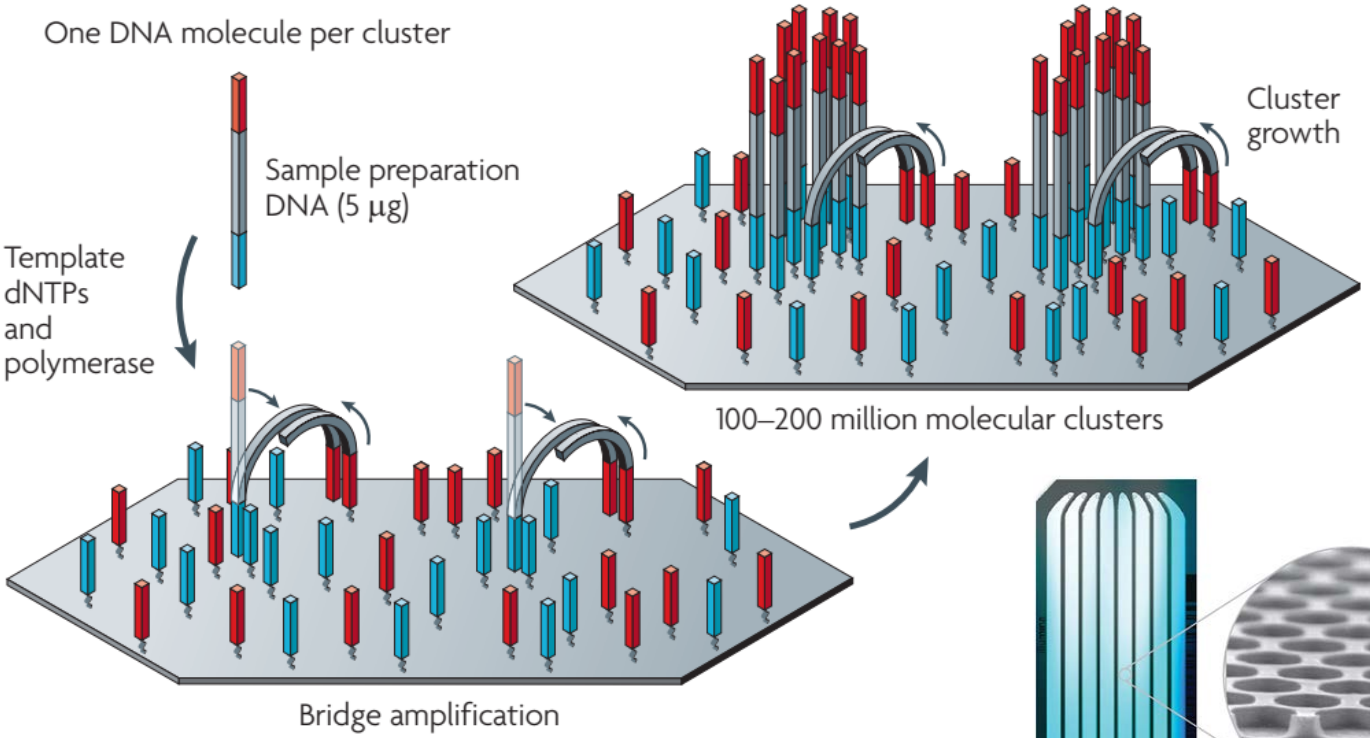
Ion Semiconductor Chip
(Ion Torrent)

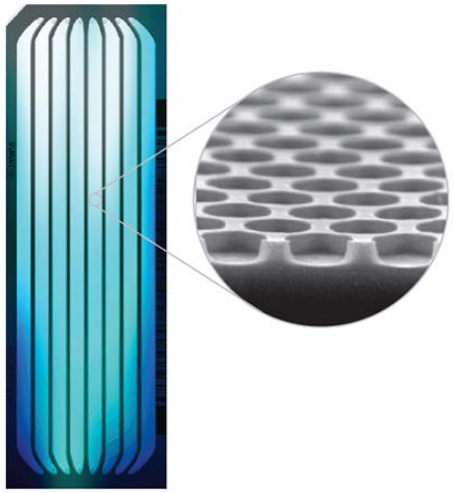


Aminocoated glass surface
(ABI/SOLiD)

(Metzker, 2010) (Kelly, Baret, Taly, & Griffiths, 2007) (Mardis, 2013) (Anthony, 2011)

SOLID-PHASE AMPLIFICATION





Illumina Flow Cell

Millions to billions of DNA clusters!

(Metzker, 2010) (Illumina, 2018)

SEQUENCING PRINCIPLES

- There are three main principles for next generation DNA sequencing
  - **Sequencing-by-Synthesis(SBS)**
  - **Sequencing-by-Ligation(SBL)**
  - **Sequencing-by-Hybridization(SBH)**
- Each has advantages and challenges but most common sequencing method is by synthesis followed by SBL
- Each sequencing platform have its own unique differences that are based on one or more than one of the principles above
- Although SBH is cost effective it is mostly used for genome-wide association studies and variant detection rather than *de novo* sequencing

(Kchouk, Gibrat, & Elloumi, 2017) (Shendure & Ji, 2008)

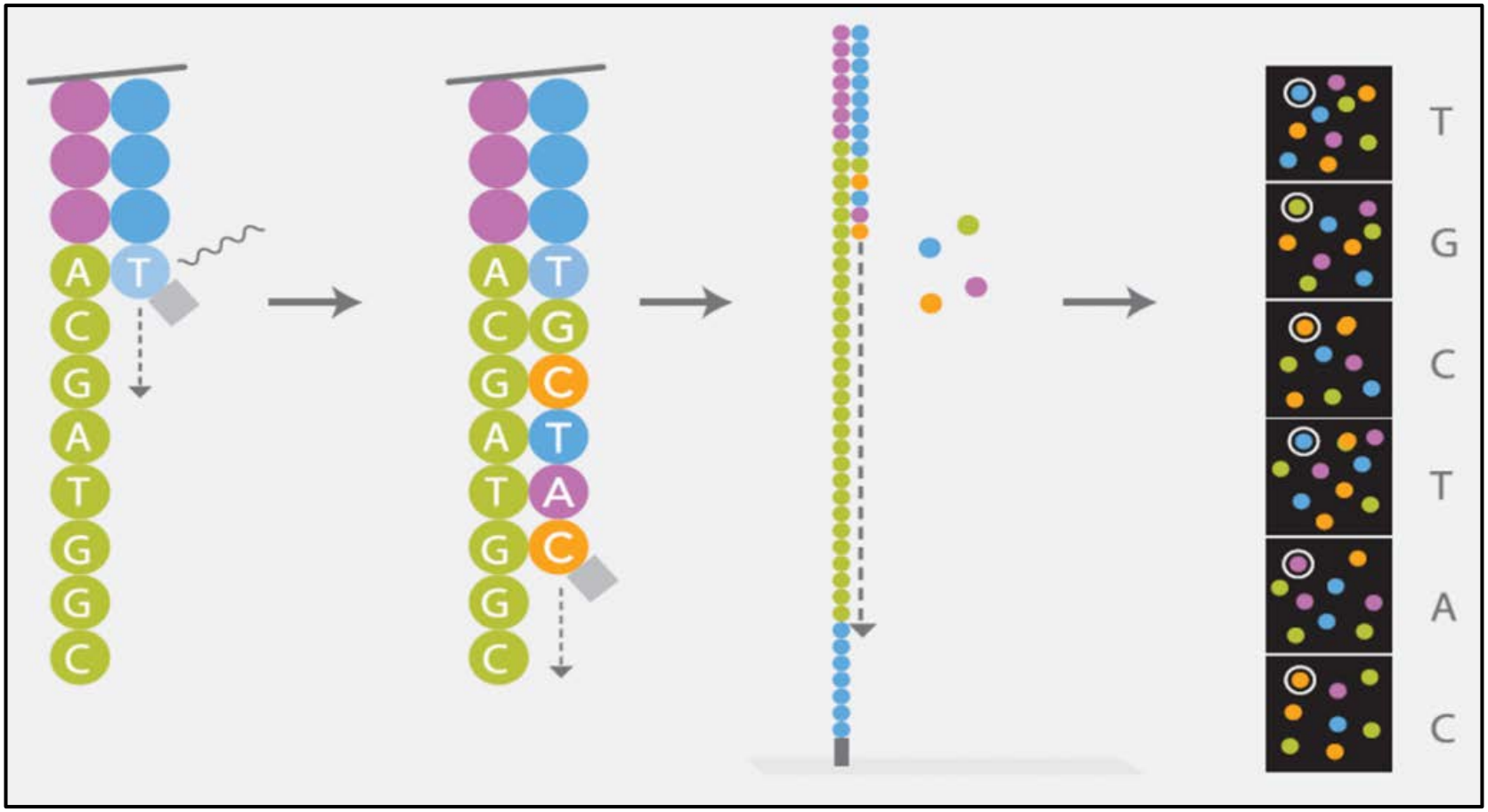
NGS TECHNICAL DIFFERENCES

Technology	Sequencing Method	Detection
Rosche/454	Pyrosequencing (SNA)	Luminescence (Pyrophosphate)
Illumina/Solexa	Illumina Dye Sequencing (CRT)	Fluorescence (4 color Fl-dNTP)
Ion Torrent	Post-Light Sequencing (SNA)	ΔpH/Proton Detection
Heliscope	Single Molecule Fluorescent Sequencing (CRT)	Fluorescence (1 color Fl-dNTP)
ABI/SOLiD	Support Oligonucleotide Ligation Detection	Fluorescence (4 color 1,2 dNTP probes)

CRT (Cyclic Reversible Termination), **SNA** (Single-nucleotide addition)

(Kchouk, Gibrat, & Elloumi, 2017) (Metzker, 2010) (Fuller et al., 2009)

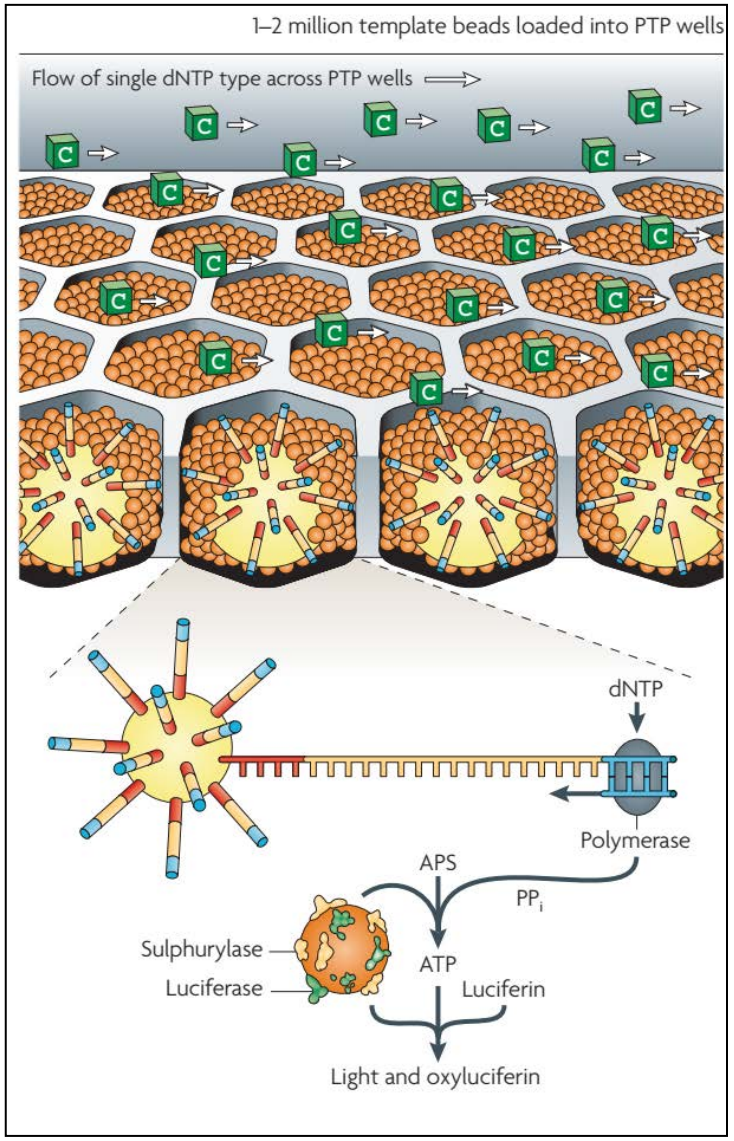
SEQUENCING-BY-SYNTHESIS



(CeGaT, 2018)

PYROSEQUENCING

- Components:
  - *single-strand DNA*
  - *one of dNTPs for each cycle*
  - *DNA polymerase*
  - *ATP sulfurylase, luciferase and apyrase*
  - *adenosine 5´ phosphosulfate (APS)*
  - *luciferin*
- Generated light is proportional to ATP and can be detected by high-resolution **charge-couple device(CCD)**
- Each cycle is completed when DNA polymerase runs out of dNTPs



DNA polymerase

$(DNA)_n + dNTP \text{ ----------} > (DNA)_{n+1} + PPi$

Sulfurylase

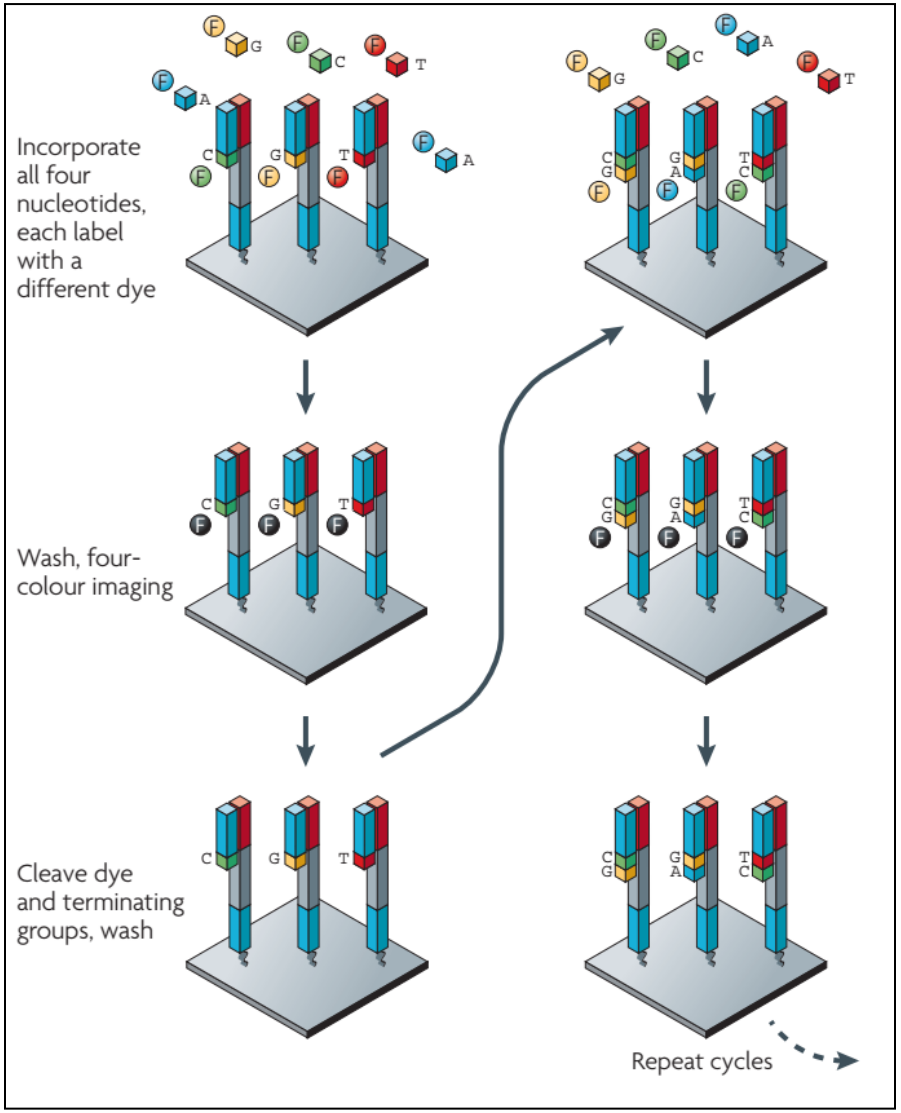
$PPi + \text{adenosine 5' phosphosulfate} \text{ ----------} > ATP + SO_4^{-2}$

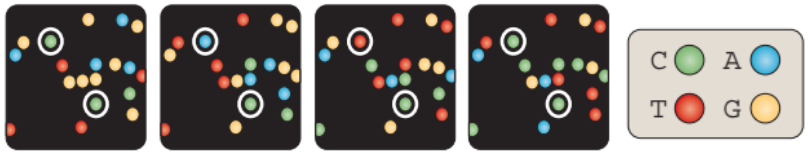
$ATP + \textit{luciferin} + O_2 \text{ ---------} > AMP + PPi + CO_2 + \text{one photon}$

(Metzker, 2010)

ILLUMINA DYE SEQUENCING

1. **DNA polymerase** used to incorporate one of four colored **Illumina's reversible terminator dNTPs with inactive 3'-hydroxyl group**
2. Wash four colored dNTPs and perform imaging
3. **Cleavage** step **removes fluorescent dyes** and **reducing agent tris(2-carboxyethyl)phosphine (TCEP)** is used to regenerate 3'-OH group to allow incorporation of next dNTPs

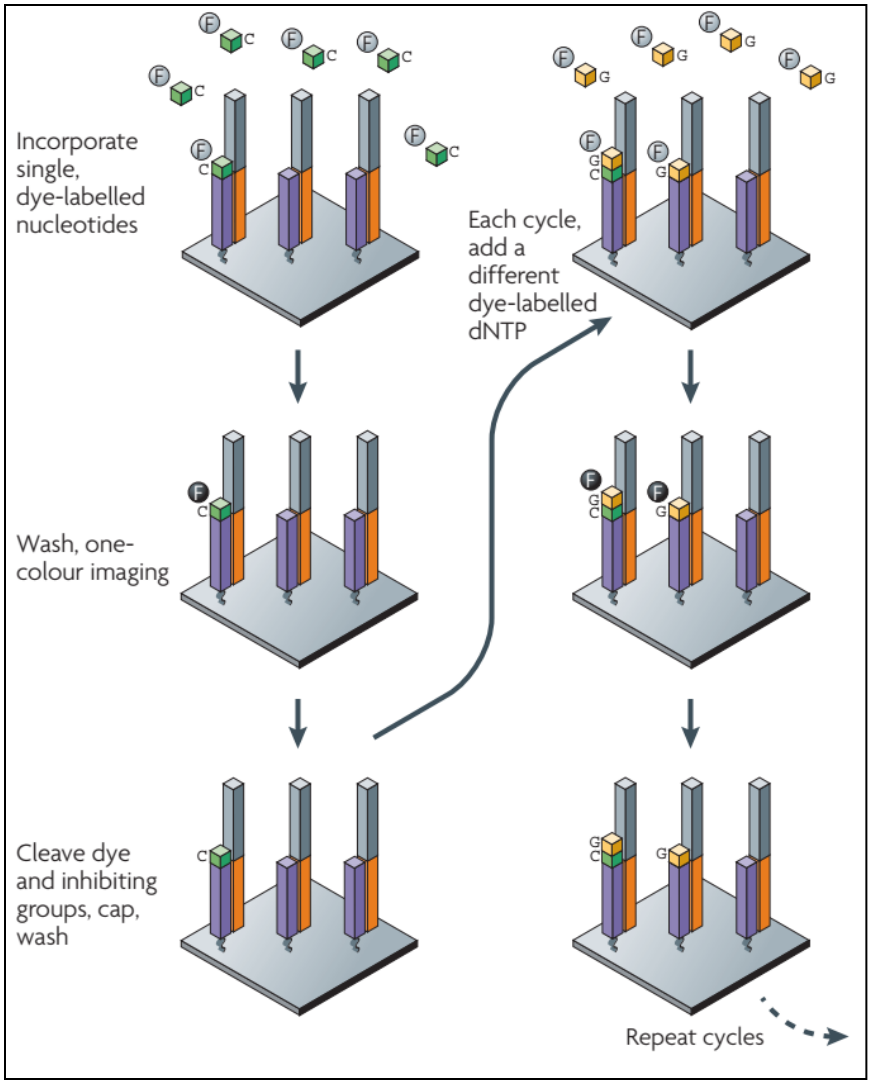




(Metzker, 2010)

SINGLE MOLECULE FLUORESCENT SEQUENCING

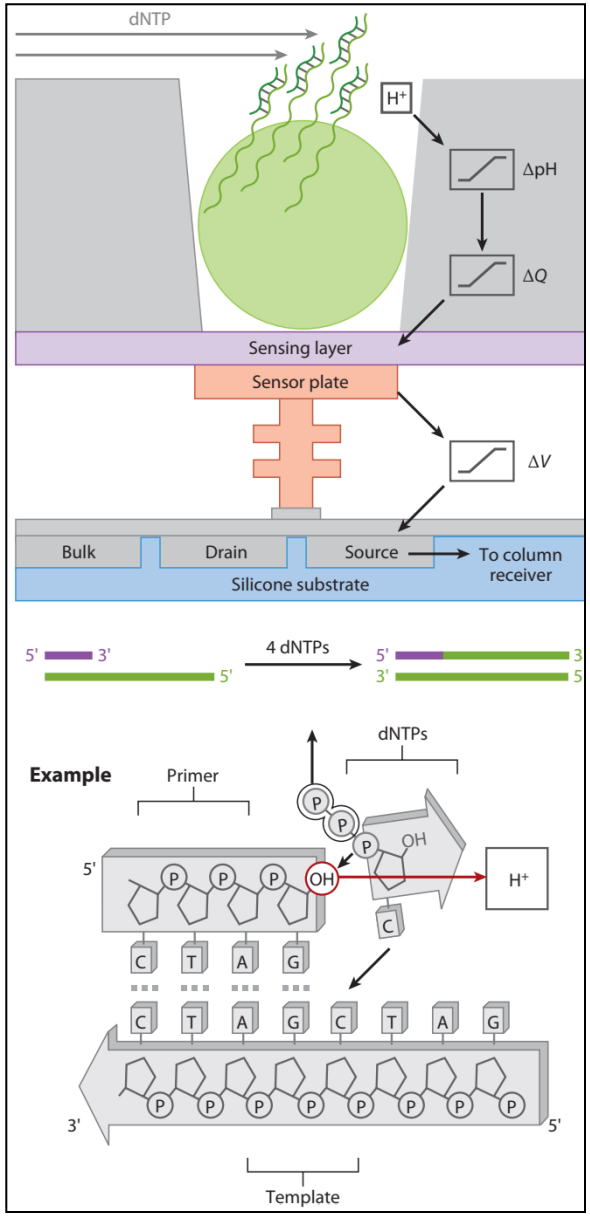
1. **DNA polymerase** used to incorporate one colored **Helicos Virtual Terminators Cy5-2′-deoxyribonucleoside triphosphate (dNTP)**
2. Wash one colored dNTPs and perform imaging
3. **Cleavage** step **removes fluorescent dyes** and **reducing agent tris(2-carboxyethyl)phosphine (TCEP)** is used to remove inhibitory group to allow incorporation of next dNTPs



(Metzker, 2010)

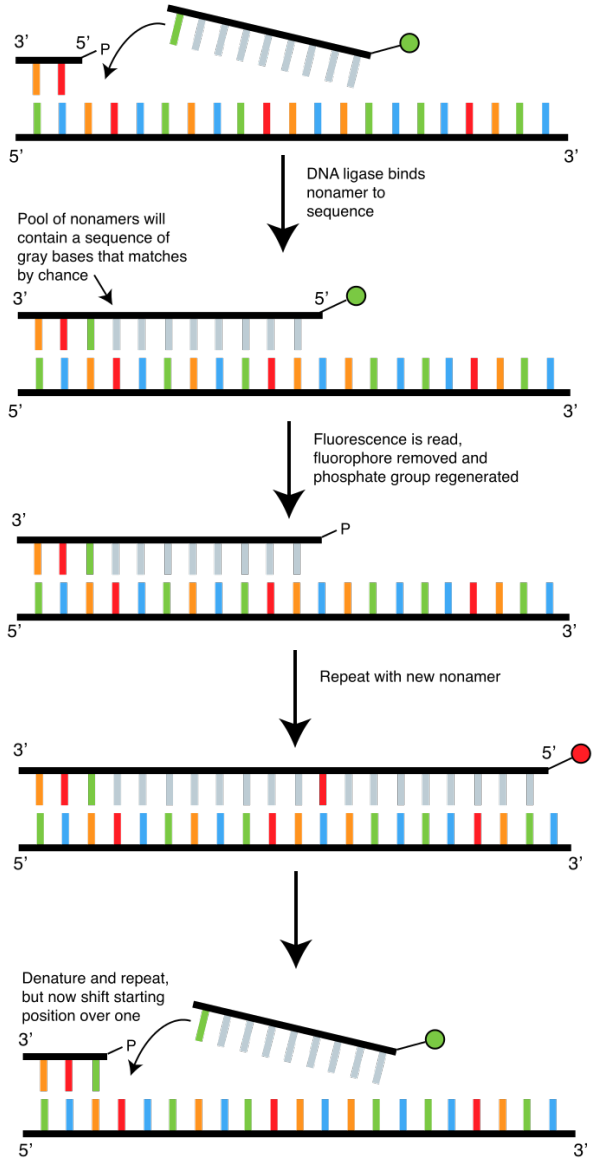
POST-LIGHT SEQUENCING

- Sequencing is performed on semiconductor **Ion Chip** with pH sensor
- DNA Polymerase incorporates one **non-modified dNTP at time**
- DNA Polymerase activity releases one pyrophosphate and **one hydrogen ion**
- Released H^+ ion **changes pH in the single well**
- Change in pH is instantly detected by **sensor plate(pH detector)**
- After detection chip is washed and new dNTP is added



(Mardis, 2013)

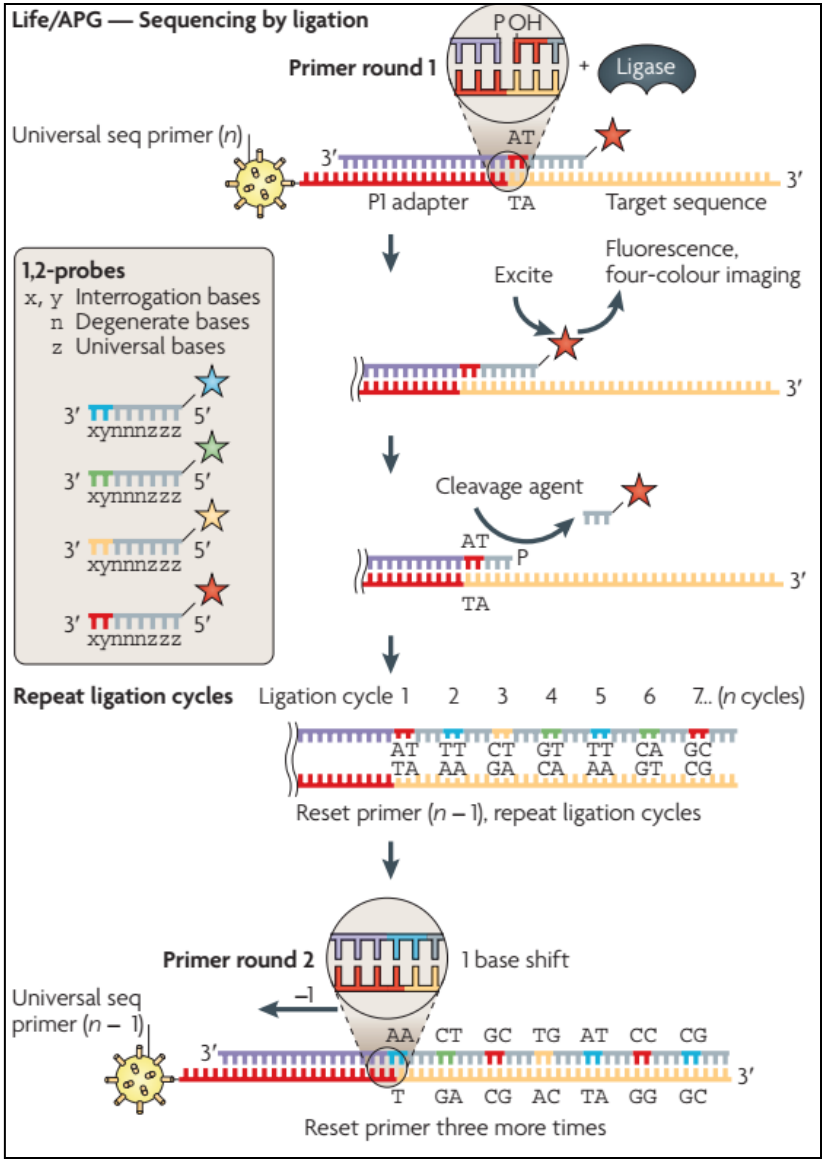
SEQUENCING-BY-LIGATION

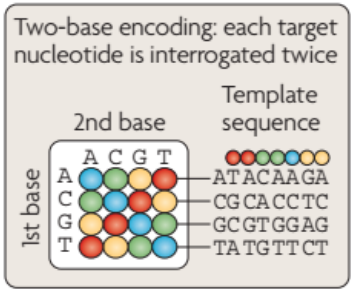


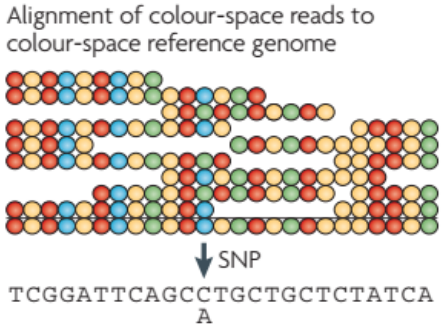
(SnipAcademy, 2017)

SUPPORT OLIGONUCLEOTIDE LIGATION DETECTION

- **For 50 bp read length** sequencing is performed in **5 primer rounds** and **10 ligation rounds**
- **4 octamer probes** are with **2 dNTPs (16 possible combinations)** and **6 degenerated dNTPs with one of 4 fluorescent labels**
- During ligation rounds one of four possible octamer probes is **hybridized with complementary template region** followed with **fluorescence detection**
- Next, **last 3 dNTPs** with fluorescent label are **removed** and **probe is ligated** with previous dNTP
- 9 more ligation rounds are performed **until template is completely covered**
- Extended **primer product is denatured and removed** to complete first primer round
- **Second primer round** is performed with **new (n-1) offset primer** with 10 ligation rounds
- This is performed until (n-5) primer round is complete
- Finally generated data is deconvoluted and final sequence is obtained for each DNA fragment bead







(Metzker, 2010) (Voelkerding, Dames, & Durtschi, 2009)

NGS PLATFORM COMPARISON

	Rosche/454	Illumina/Solexa	Ion Torrent	Heliscope	ABI/SOLiD
Read Length (bp)	200-1000	75-300	200-400	30-35	35/50/75
Reads per Run	100-1M	25M-6B	0.4M-80M	600M-1B	1B-6B
Appx. Data Generated	0.45Gb	1.8Tb	10Gb	37 Gb	160Gb
Error Type (Avg. % Rate)	Indel (1%)	Mismatch (0.1%)	Indel (1%)	Indel (1%)	Mismatch (0.06%)
Run Time	~10 h	~10 days	~7 days	~ 8 days	~12 days
Appx. Device Cost (~2012 Prices)	100,000 $ – 500,000 $	100,000 $ – 600,000 $	80,000 $ – 150,000 $	1 Million $	500,000 $
Avg. Cost per Gb (~2012 Prices)	10,000$	40-500	1000$	1000$	130$

(Kchouk, Gibrat, & Elloumi, 2017; Liu et al., 2012; Metzker, 2010; Quail et al., 2012; Reinert, Langmead, Weese, & Evers, 2015; van Dijk, Auger, Jaszczyszyn, & Thermes, 2014)(nextgenseek, 2012)

INTRODUCTION TO DNA SEQUENCE ALIGNMENT

- **Sequence Alignment** is a process of lining up two or more sequences of nucleotides or amino acids in order to find any global or regional similarity that may denote some structural, functional or evolutionary relationship
- There are **two type** of sequence alignments:
  - **Pair-wise alignment –** Only two sequences
  - **Multiple sequence alignment –** More than two sequences
- There are **two computational approaches** for alignment:
  - **Global alignment –** Complete end-to-end alignment
  - **Local alignment –** Regional alignment
- Alignment of NGS reads are called **Short Read Alignment** and is a subject of significant research due to alignment of massive amounts of short sequences to a references genome

(Diniz & Canduri, 2017)(Reinert, Langmead, Weese, & Evers, 2015)

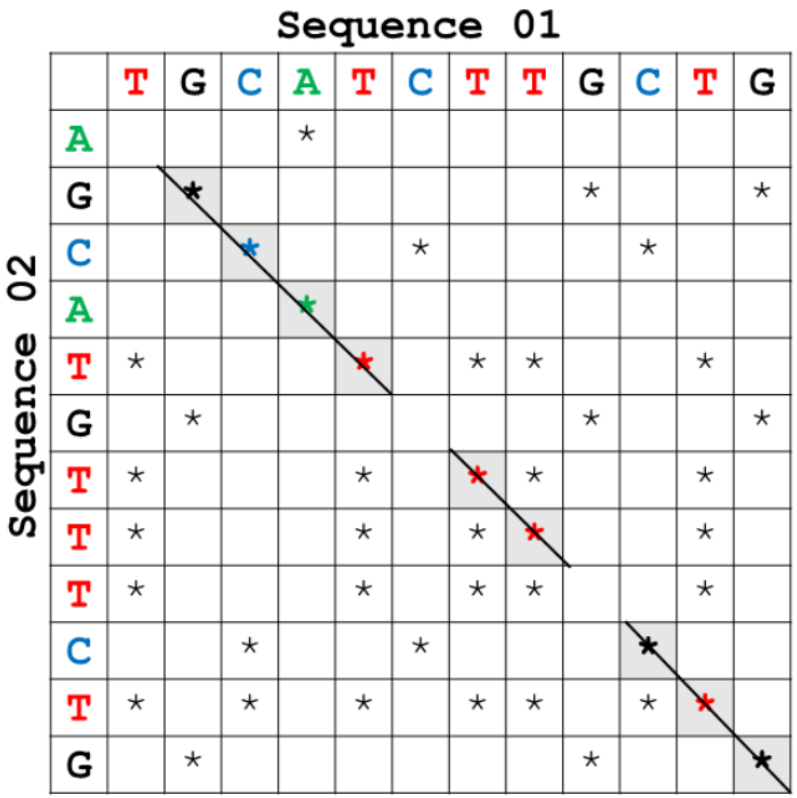
ALIGNMENT METHODS

There are four **main sequence alignment algorithms**:

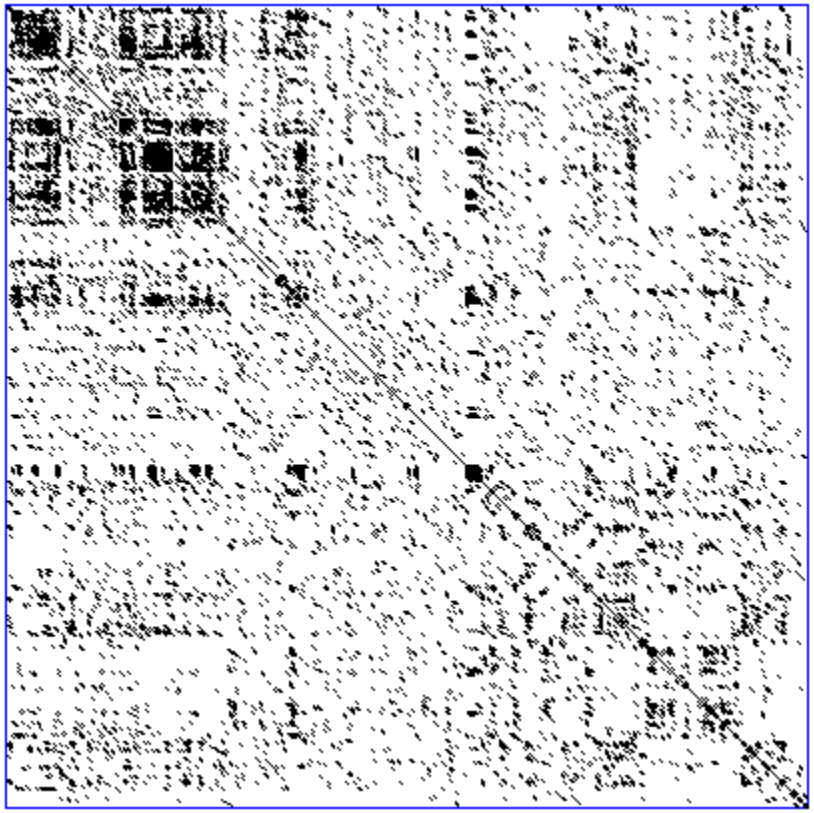
- **Brute Force**
  - Simplest and most inefficient
- **Dot-Matrix**
  - Can provide quick visualization of two sequences
- **Dynamic Programming**
  - Guarantee optimal alignment and provide scores but not efficient method for aligning long sequences
- **Word Method** (aka **k-tuple method**)
  - Use heuristic algorithms that do not guarantee optimal alignments but significantly more efficient than any other method. Uses seed-and-extent algorithm and is implemented in most well known tools such as BLAST and FASTA

(Diniz & Canduri, 2017)

DOT-MATRIX



Dot Plot Demo



Real Dot Plot

(Diniz & Canduri, 2017)

DYNAMIC PROGRAMMING

- There are two main dynamic programming algorithms for sequence alignment
  - **Needleman-Wunsch** algorithm for global alignments
  - **Smith-Waterman** algorithm for local alignment
- Both use dot matrix approach with distance measurements such as Manhattan distance to generate score matrix that is used to find optimal alignment
- Score are calculated using various scoring matrices such as
  - Simple identity matrix for nucleotide scoring
  - PAM and BLOSUM matrices for amino acid scoring

(Marketa & Jeremy, 2008)

NEEDLEMAN-WUNSCH ALGORITHM EXAMPLE

- Say that you want to globally align following sequences
  - Top Sequence (S1): GTCACATGCC – 10 base pairs
  - Side Sequence (S2): GCCGACAGT – 9 base pairs
- Alignment can be performed in following steps:
  1. **Matrix Initialization** – Select matrix scheme and initiate matrix
  2. **Matrix Fill** – Apply recurrence relations to every cell
  3. **Matrix Traceback** – Find path back to the most top-left (0,0) position starting from most lower-right position
- First let's set matrix scheme with following alignment parameters:
  - **Match**: 5 points ($\delta_{i,j}$)
  - **Mismatch**: -3 points ($\delta_{i,j}$)
  - **Gap Penalty**: -5 points (w)

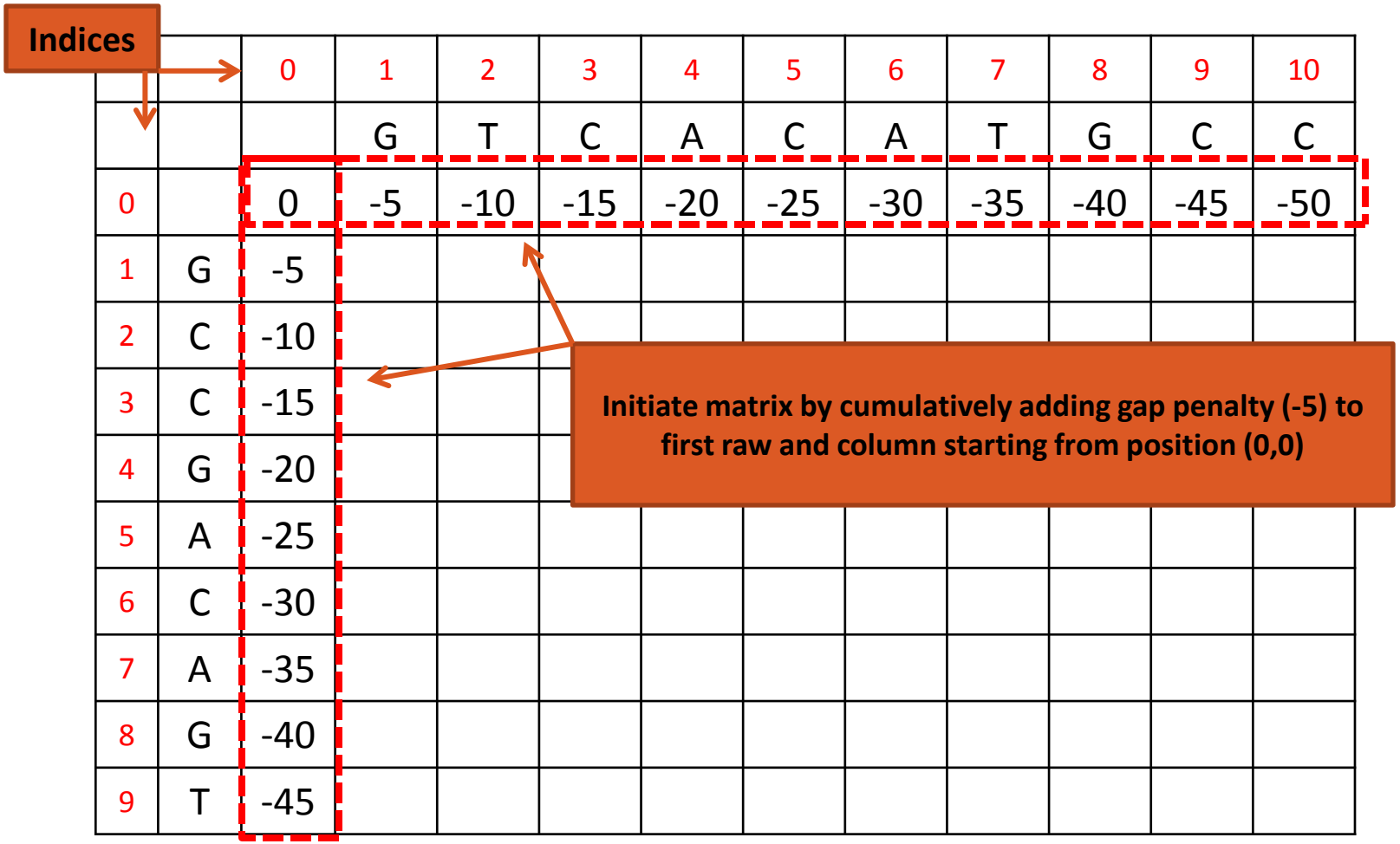
(Shu & Ouw, 2004)

MATRIX

Sequence S1 length was 9 and S2 length was 10. Resulting matrix size is (9+1)x(10+1) = **10x11**

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

MATRIX INITIALIZATION



MATRIX FILL

Recurrence relations are:

$$s_{i,j} = max \begin{cases} s_{i-1,j-1} + \delta_{i,j} \\ s_{i,j-1} + w \\ s_{i-1,j} + w \end{cases}$$

		0	1	2	3	4	5	6	7	8	9	10
			G	T	C	A	C	A	T	G	C	C
0		0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50
1	G	-5	5									
2	C	-10	0									
3	C	-15										
4	G	-20										
5	A	-25										
6	C	-30										
7	A	-35										
8	G	-40										
9	T	-45										

$$s_{1,1} = max \begin{cases} s_{1-1,1-1} + \delta_{1,1} = 0 + 5 = 5 \\ s_{1,1-1} + w = -5 - 5 = -10 = 5 \\ s_{1-1,1} + w = -5 - 5 = -10 \end{cases}$$

$$s_{2,1} = max \begin{cases} s_{2-1,1-1} + \delta_{2,1} = -5 - 3 = -8 \\ s_{2,1-1} + w = -10 - 5 = -15 \\ s_{2-1,1} + w = 5 - 5 = 0 \end{cases} = 0$$

Where

- i stands for rows of side sequence
- j stands for columns of top sequence
- $s_{i,j}$ is score of the at position i, j
- $\delta_{i,j}$ is +5 reward if i and j are match and -3 penalty if mismatch
- w is -5 points gap penalty

Above function is applied for every cell and maximal value out of three is chosen

MATRIX FILL

		0	1	2	3	4	5	6	7	8	9	10
			G	**T**	**C**	**A**	**C**	**A**	**T**	**G**	**C**	**C**
0		*0*	*-5*	*-10*	*-15*							*0*
1	**G**	*-5*	5	0	-5							0
2	**C**	*-10*	0	2	5							0
3	**C**	*-15*	-5	-3	7	2	5	0	-5	-10	-15	-20
4	**G**	*-20*	-10	-8	2	4	0	2	-3	0	-5	-10
5	**A**	*-25*	-15	-13	-3	7	2	5	0	-5	-3	-8
6	**C**	*-30*	-20	-18	-8	2	12	7	2	-3	0	2
7	**A**	*-35*								7	2	-3
8	**G**	*-40*	-30	-28	-18	-8	2	12	14	17	12	7
9	**T**	*-45*	-35	-25	-23	-13	-3	7	17	12	14	**9**

$$s_{4,2} = max \begin{cases} s_{4-1,2-1} + \delta_{4,2} = -5 - 3 = -8 \\ s_{4,2-1} + w = -10 - 5 = -15 \\ s_{4-1,2} + w = -3 - 5 = -8 \end{cases} = -8$$

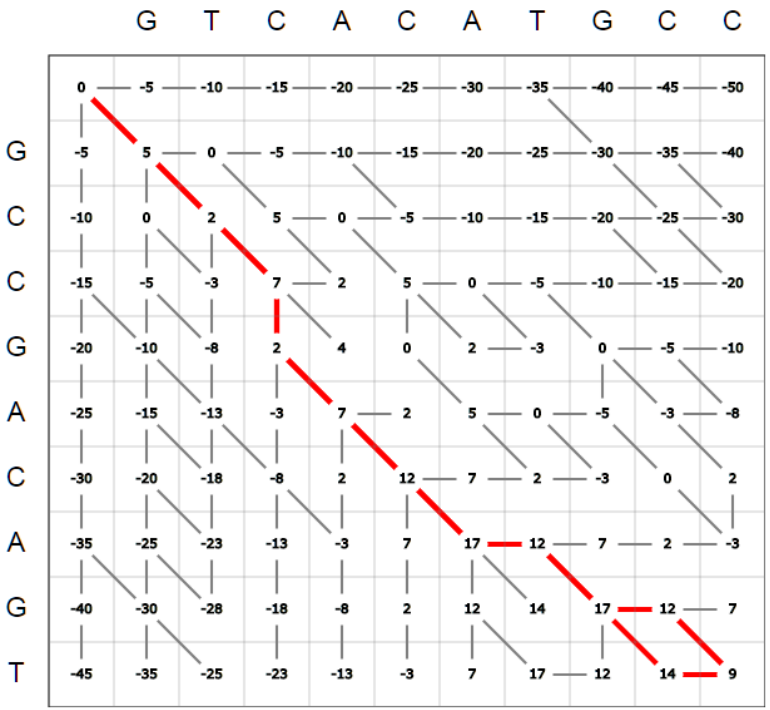
FINAL POSITION SCORE = ALIGNMENT SCORE

MATRIX TRACEBACK

For given alignment parameters there are two possible optimal alignments with same alignment score of 9 points

GTC-ACATGCC	GTC-ACATGCC
GCCGACA-G-T	GCCGACA-GT-

		G	T	C	A	C	A	T	G	C	C
	0	-5	-10	-15	-20	-25	-30	-35	-40	-45	-50
G	-5	5	0	-5	-10	-15	-20	-25	-30	-35	-40
C	-10	0	2	5	0	-5	-10	-15	-20	-25	-30
C	-15	-5	-3	7	2	5	0	-5	-10	-15	-20
G	-20	-10	-8	2	4	0	2	-3	0	-5	-10
A	-25	-15	-13	-3	7	2	5	0	-5	-3	-8
C	-30	-20	-18	-8	2	12	7	2	-3	0	2
A	-35	-25	-23	-13	-3	7	17	12	7	2	-3
G	-40	-30	-28	-18	-8	2	12	14	17	12	7
T	-45	-35	-25	-23	-13	-3	7	17	12	14	9



http://experiments.mostafa.io/public/needleman-wunsch/

https://gtuckerkellogg.github.io/pairwise/demo/

REFERENCES

- Amy Rogers, M. D., Ph.D. (2008). Shotgun sequencing. Retrieved from https://www.csus.edu/indiv/r/rogersa/Bio181/SeqShotgun.pdf
- Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *N Biotechnol, 25*(4), 195-203. doi:10.1016/j.nbt.2008.12.009
- Anthony, S. (2011). New Silicon Chip Sequences Complete Genome in Three Hours. Retrieved from https://www.extremetech.com/computing/84340-new-silicon-chip-sequences-complete-genome-in-three-hours
- ATDBio. Next generation sequencing. Retrieved from https://www.atdbio.com/content/58/Next-generation-sequencing
- Brown, T. A. (2016). *Gene cloning and DNA analysis: an introduction*: John Wiley & Sons.
- CeGaT. (2018). Next-Generation Sequencing Services. Retrieved from https://www.cegat.de/en/services/next-generation-sequencing/
- Chial, H. (2008). DNA sequencing technologies key to the Human Genome Project. *Nature Education, 1*(1), 219.
- Diniz, W. J., & Canduri, F. (2017). REVIEW-ARTICLE Bioinformatics: an overview and its applications. *Genet Mol Res, 16*(1). doi:10.4238/gmr16019645
- EpiNext. (2017). EpiNext High-Sensitivity DNA Library Preparation Kit (Illumina) Retrieved from https://www.epigentek.com/catalog/epinext-high-sensitivity-dna-library-preparation-kit-illumina-p-3632.html
- Frese, K. S., Katus, H. A., & Meder, B. (2013). Next-generation sequencing: from understanding biology to personalized medicine. *Biology (Basel), 2*(1), 378-398. doi:10.3390/biology2010378
- Fuller, C. W., Middendorf, L. R., Benner, S. A., Church, G. M., Harris, T., Huang, X., . . . Vezenov, D. V. (2009). The challenges of sequencing by synthesis. *Nat Biotechnol, 27*(11), 1013-1023. doi:10.1038/nbt.1585
- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: overviews and challenges. *Biotechniques, 56*(2), 61-64, 66, 68, passim. doi:10.2144/000114133
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics, 107*(1), 1-8. doi:10.1016/j.ygeno.2015.11.003
- Hui, P. (2014). Next generation sequencing: chemistry, technology and applications. *Top Curr Chem, 336*, 1-18. doi:10.1007/128_2012_329
- Illumina. (2018). Patterned Flow Cell Technology. Retrieved from https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/patterned-flow-cells.html
- Illumina, I. (2015). An introduction to next-generation sequencing technology.
- Kchouk, M., Gibrat, J.-F., & Elloumi, M. (2017). Generations of sequencing technologies: From first to next generation. *Biology and Medicine, 9*(3).
- Kelly, B. T., Baret, J. C., Taly, V., & Griffiths, A. D. (2007). Miniaturizing chemistry and biology in microdroplets. *Chem Commun (Camb)*(18), 1773-1788. doi:10.1039/b616252e
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., . . . Law, M. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol, 2012*, 251364. doi:10.1155/2012/251364
- Mardis, E. R. (2013). Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif), 6*, 287-303. doi:10.1146/annurev-anchem-062012-092628
- Mardis, E. R. (2017). DNA sequencing technologies: 2006-2016. *Nat Protoc, 12*(2), 213-218. doi:10.1038/nprot.2016.182
- Marketa, Z., & Jeremy, O. (2008). Understanding bioinformatics. *Garland Science*.
- McClean, P. (2005). Genome Sequencing Retrieved from https://www.ndsu.edu/pubweb/~mcclean/plsc411/Genome%20Sequencing%20Complete.pdf
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet, 11*(1), 31-46. doi:10.1038/nrg2626
- nextgenseek. (2012). Comparing Price and Tech. Specs. of Illumina MiSeq, Ion Torrent PGM, 454 GS Junior, and PacBio RS. Retrieved from http://nextgenseek.com/2012/08/comparing-price-and-tech-specs-of-illumina-miseq-ion-torrent-pgm-454-gs-junior-and-pacbio-rs/
- Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. *J Appl Genet, 52*(4), 413-435. doi:10.1007/s13353-011-0057-x
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., . . . Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics, 13*(1), 341.
- Reinert, K., Langmead, B., Weese, D., & Evers, D. J. (2015). Alignment of Next-Generation Sequencing Reads. *Annu Rev Genomics Hum Genet, 16*, 133-151. doi:10.1146/annurev-genom-090413-025358
- Rodney E. Shackelford, D. O., Ph.D. (30 May 2018). Pathology Outlines - Pyrosequencing. Retrieved from http://www.pathologyoutlines.com/topic/molecularpathdnaseqpyro.html
- Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature, 550*(7676), 345-353. doi:10.1038/nature24286
- Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol, 26*(10), 1135-1145. doi:10.1038/nbt1486
- Shu, J. J., & Ouw, L. S. (2004). Pairwise alignment of the DNA sequence using hypercomplex number representation. *Bull Math Biol, 66*(5), 1423-1438. doi:10.1016/j.bulm.2004.01.005
- SnipAcademy. (2017). Bioinformatics Lessons. Retrieved from https://binf.snipcademy.com/lessons/
- Sterky, F., & Lundeberg, J. (2000). Sequence analysis of genes and genomes. *Journal of Biotechnology, 76*(1), 1-31.
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet, 30*(9), 418-426. doi:10.1016/j.tig.2014.07.001
- van Dijk, E. L., Jaszczyszyn, Y., & Thermes, C. (2014). Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res, 322*(1), 12-20. doi:10.1016/j.yexcr.2014.01.008
- Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clin Chem, 55*(4), 641-658. doi:10.1373/clinchem.2008.112789
- Yohe, S., & Thyagarajan, B. (2017). Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med, 141*(11), 1544-1557. doi:10.5858/arpa.2016-0501-RA