



Sequencing of Nucleic Acids

JONATHAN POHL, MLS (ASCP)^{CM} MB^{CM}

CLEVELAND CLINIC CENTER FOR PATHOLOGY EDUCATION

Disclaimer

This presentation was developed by Cleveland Clinic's Center for Pathology Education (CPE) for use in Cleveland Clinic's Molecular Biology Tech Categorical Exam Preparation Program (TCEPP).

This presentation is intended for educational purposes only.

Information in this presentation is not meant to substitute or replace the policies and standard operating procedures of Cleveland Clinic or RT-PLMI.

Images of and references to specific brands of instrumentation, consumables, reagents, or other purchasable items are used as examples and are not endorsements of those specific brands or items.

Objectives

- Extensively detail sequencing methods used in Cleveland Clinic's molecular labs.
 - Dideoxy Chain Termination (Sanger, Automated Fluorescent)
 - Pyrosequencing
 - Methylation-Specific Sequencing (Pyro and Sanger)
 - Next Generation Sequencing (Reversible Dye Terminator)
- Go over the basic principle of other clinical sequencing methods not used in Cleveland Clinic's molecular labs.
 - Ion-Conductance Sequencing

References and Additional Resources

- Buckingham, Lela. *Molecular Diagnostics: Fundamentals, Methods, and Clinical Applications*. 3rd ed., F.A. Davis, 2019.
 - Chapter 9: DNA Sequencing
 - Chapter 6: Nucleic Acid Amplification (sections on Bridge PCR and Emulsion PCR)
- Rifai, Nader, et al. *Principles and Applications of Molecular Diagnostics*. 3rd ed., F.A. Davis, 2019.
 - Chapter 4: Nucleic Acid Techniques
- Additional publicly available educational material from Illumina™ and ThermoFisher™ websites regarding their sequencing methodologies.

Sequencing

Sequencing is the determination of the exact order of nucleotides in a sample nucleic acid.

- **Targeted sequencing:** sequencing of specific genes or other sequences of interest
- **Whole exome sequencing (WES):** looks at an organism's exons, the regions of the genome that code for RNA and protein products
- **Whole genome sequencing (WGS):** looks at every base pair of an organism's genome, including both exons and introns
- **Whole transcriptome sequencing (RNAseq):** sequences mRNA transcripts in addition to DNA

Sequencing by Synthesis

Sequencing by synthesis refers to methods that determine a sequence by detecting each dNTP polymerase incorporates into a growing strand of DNA.

Three sequencing by synthesis methods are used in Cleveland Clinic's molecular labs:

- Dideoxy Chain Termination Sequencing (Sanger / Automated Fluorescent)
- Pyrosequencing (Methylation-Specific Sequencing)
- Reversible dye terminator sequencing (Illumina™)

Additional sequencing by synthesis methods not used in Cleveland Clinic's molecular labs:

- Ion-Conductance, aka Semiconductor Sequencing (Ion-Torrent™)

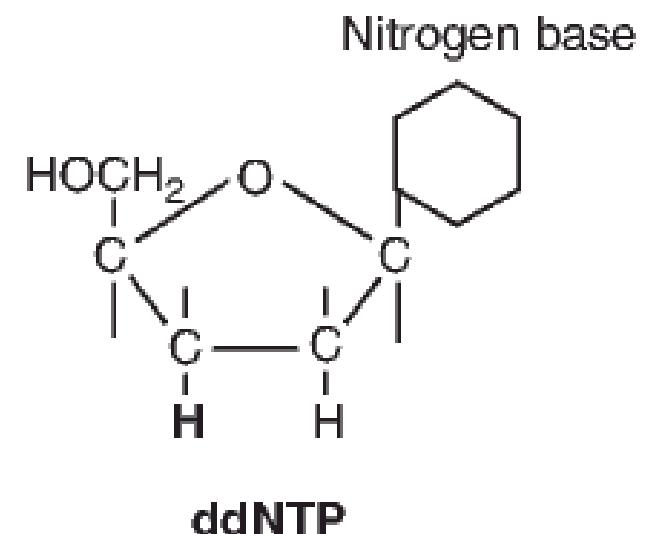
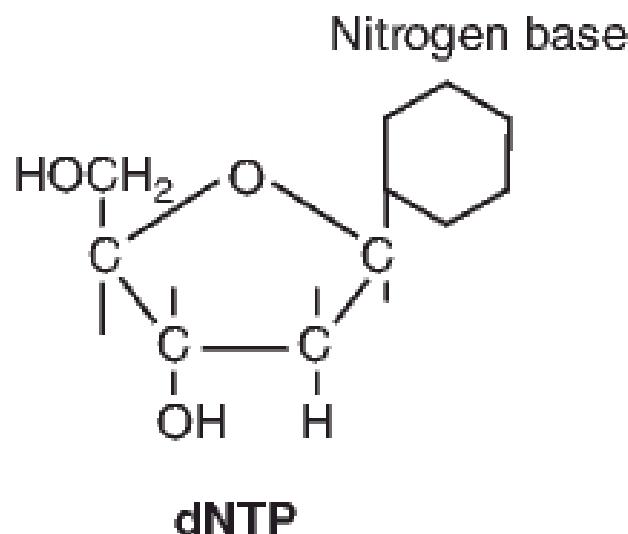
Dideoxy Chain Termination Sequencing

- Sanger Sequencing
- Automated fluorescent sequencing

Named for its inventor, Frederick Sanger, **Sanger sequencing** can be more descriptively referred to as **dideoxy chain termination sequencing**.

Utilizes dideoxynucleotides (**ddNTPs**) to stop DNA synthesis.

- ddNTPs lack the 3' hydroxyl group normally found in dNTPs

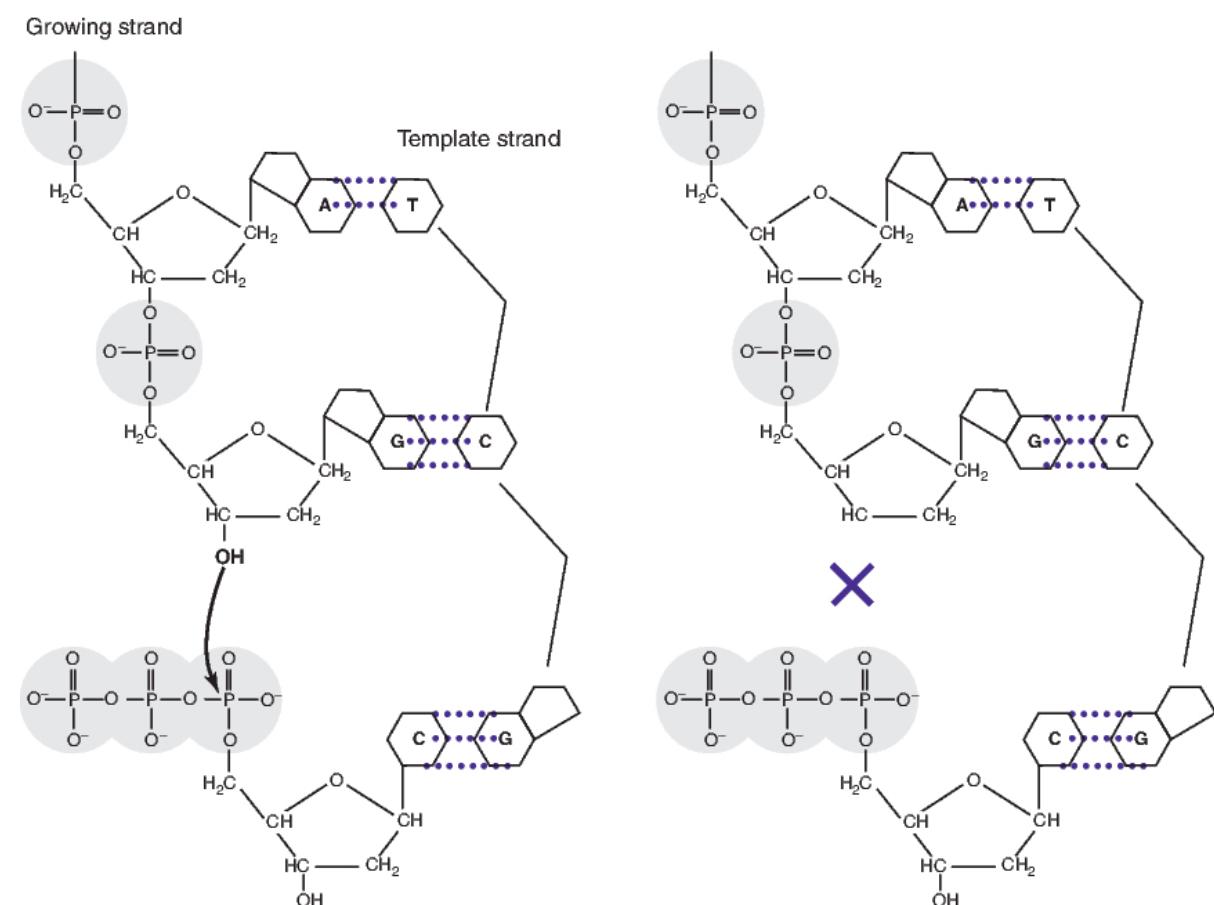


Sanger Sequencing

Absence of the 3' hydroxyl group prevents the 5' phosphate group of an incoming nucleotide from forming a phosphodiester bond.

This means that incorporation of ddNTP into growing strand will functionally terminate synthesis.

By combining dNTPs and ddNTPs into a common PCR reaction, random synthesis termination will produce amplicons of every possible bp size for a sequence.



Sanger Sequencing

The original Sanger process:

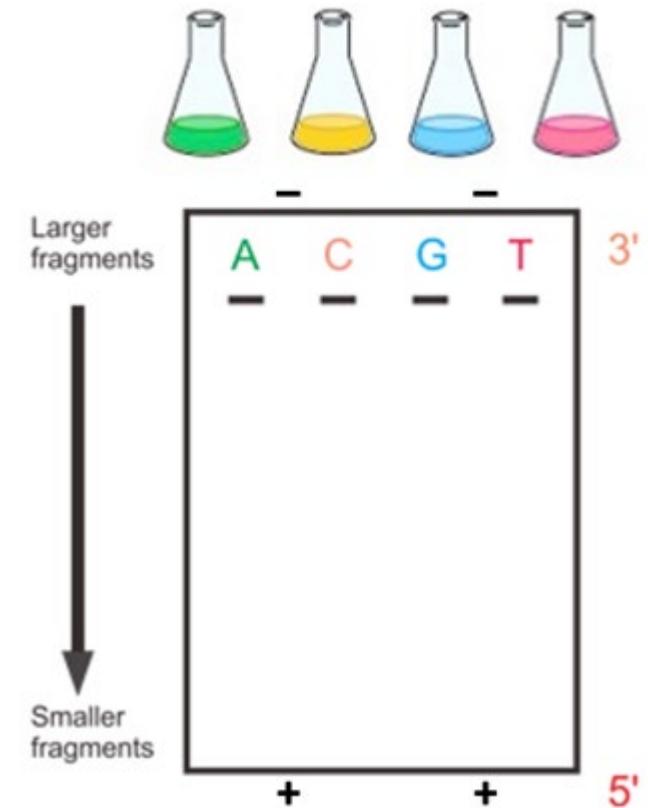
- 4 Reaction tubes: A, C, G, and T
 - Master mix for each tube has one of the four ddNTPs included, in addition to normal dNTPs used in PCR
 - Master mix also includes primers that target sequence of interest, buffer, and polymerase
- Tubes undergo PCR cycling
 - Polymerase will randomly incorporate dNTPs or ddNTPs as it synthesizes new strands of DNA
 - Each time ddNTP is incorporated, synthesis stops
 - This creates fragments of every possible base-pair length with a single ddNTP at their 3' ends



Sanger Sequencing

The original Sanger process:

- PCR products are denatured and loaded onto a polyacrylamide gel for electrophoresis.
- Applying different reaction mixes in different lanes of the gel allows for differentiation of which ddNTP is incorporated at the end of each fragment.
- Fragments will separate based on base pair size, with smaller fragments migrating towards the anode faster than larger fragments.



Sanger Sequencing

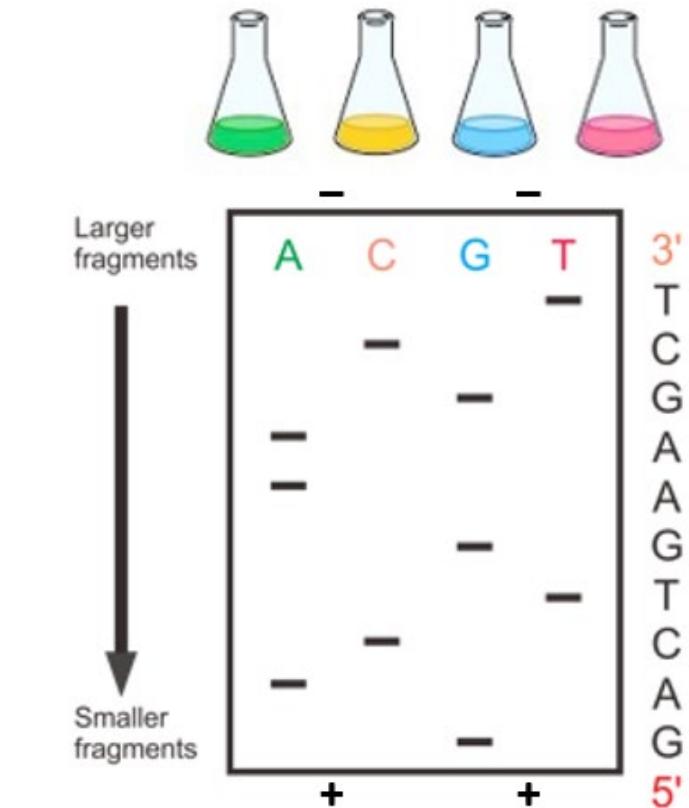
The original Sanger process:

- After migration, fragment migration can be visualized radiometrically or fluorescently.
 - Radiometric detection: radioactively-labelled (^{32}P) primers or ddNTPs and X-ray visualization
 - Fluorescent detection: fluorescent staining (EtBR, etc.) and UV visualization
- Final result is a **sequencing ladder**, which is read from anode-to-cathode (bottom-to-top).

The sequencing ladder to the right is read as:

- Sequenced Strand: 5'-GACTGAAGCT-3'
- Original Template: 5'-AGCTTCAGTC-3'

5' AGCTTCAGTC 3'
 G 5'
 AG 5'
 CAG 5'
 TCAG 5'
 GTCAG 5'
 AGTCAG 5'
 AAGTCAG 5'
 GAAGTCAG 5'
 CGAAGTCAG 5'
 TCGAAGTCAG 5'



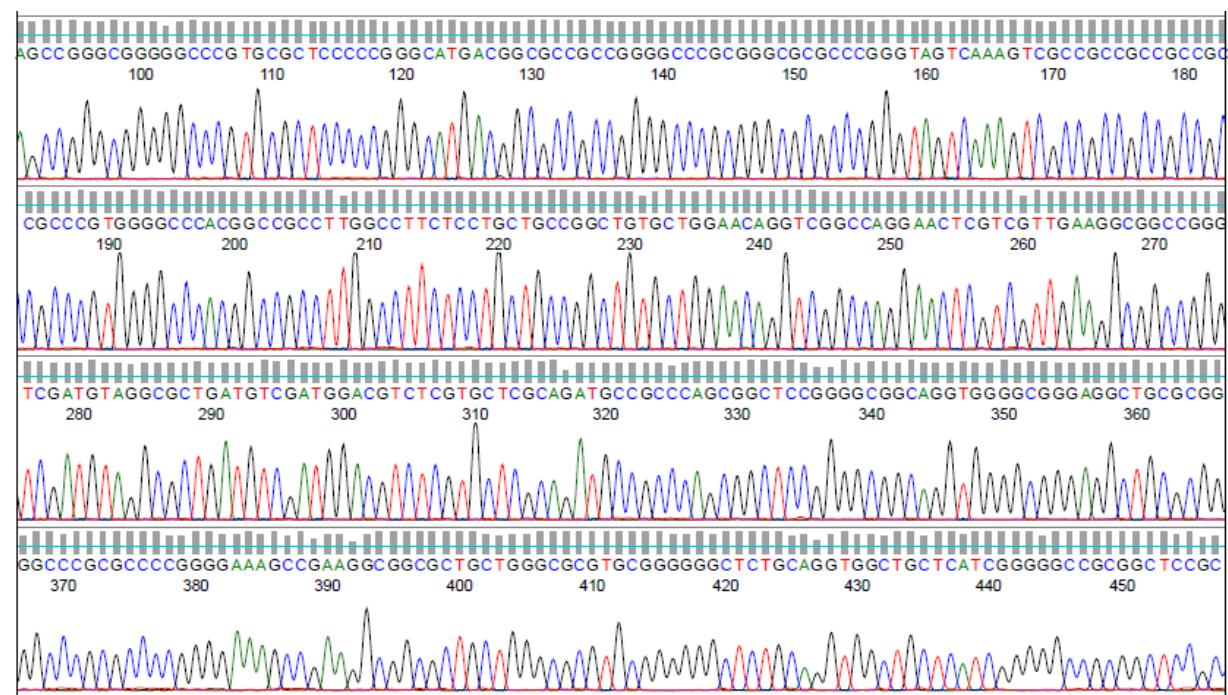
Automated Fluorescent Sequencing

In automated fluorescent sequencing, dideoxy-terminated fragments are resolved by capillary electrophoresis and differentiated with fluorescent labels.

Result electropherogram is called a **chromatogram**, with each fragment peak having one of four unique colors corresponding to the ddNTP incorporated at its 3' end.

Two formats:

- Dye primer
- Dye terminator



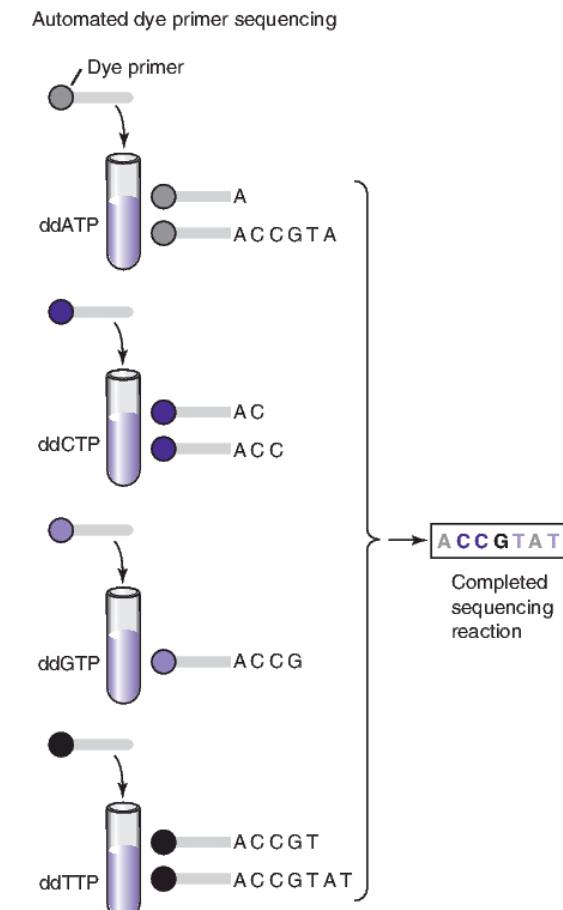
Automated Dye Primer Sequencing

In **automated dye primer sequencing**, the 5'-end of the sequencing primer for each of four ddNTP reaction wells is labelled with a unique fluorophore.

Following sequencing reaction, all four reaction mixtures are combined into a single well and resolved by capillary electrophoresis.

Fluorophore detected indicates which reaction well a fragment came from, indicating the ddNTP terminator incorporated.

This has been mostly replaced by dye terminator sequencing.



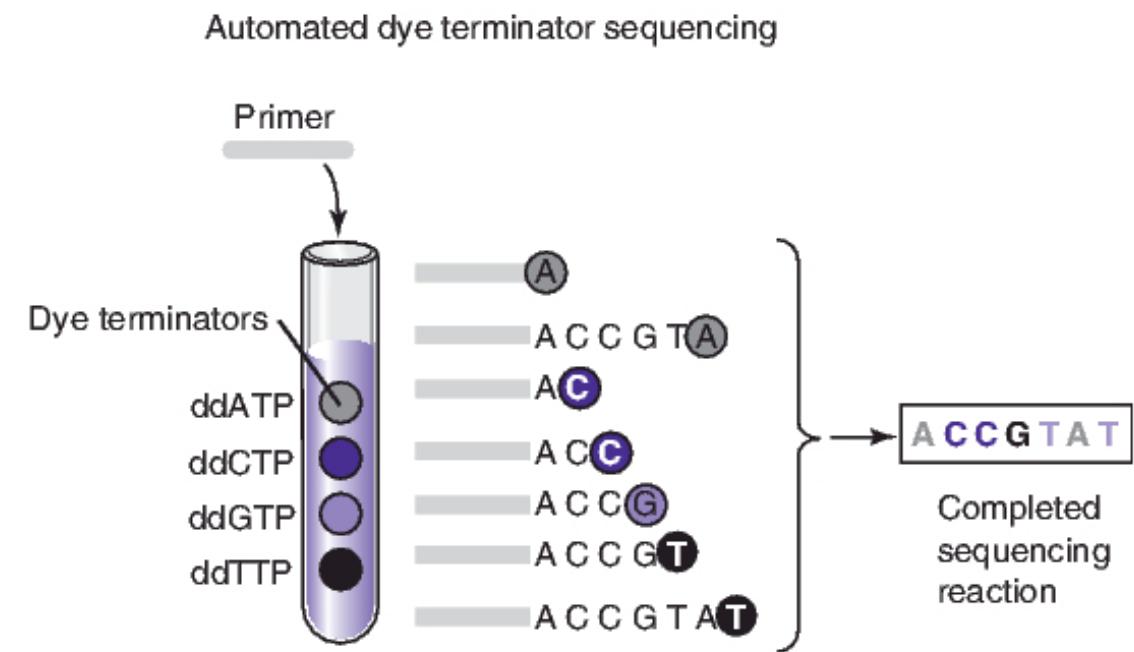
Automated Dye Terminator Sequencing

In automated dye terminator sequencing, a unique fluorophore is added to each of the four ddNTPs.

This means that the entire sequencing reaction can be performed in a single reaction well.

Fragments are then resolved by capillary electrophoresis and differentiated by the fluorescently-tagged ddNTP incorporated at their 3' end.

This is the “Sanger sequencing” method of choice.

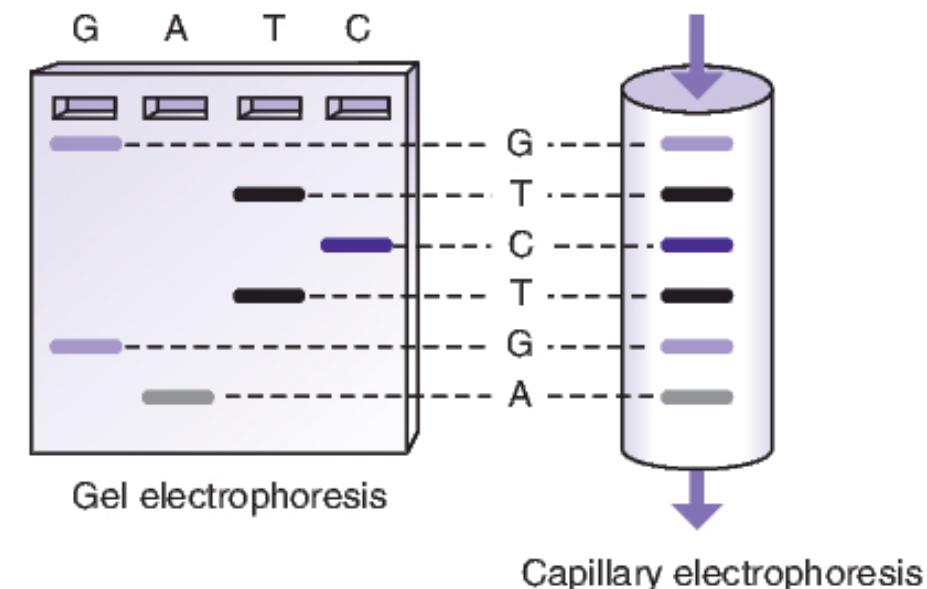


Automated Fluorescent Sequencing

Fragment resolution of the ddNTP-generated sequencing ladder via capillary electrophoresis is similar to gel-based ladders.

- Small (bp) fragments migrate through capillary faster than large fragments.
- Sequencing ladder is constructed 5'-3' as fragments pass the detector.
- ddNTP incorporated at 3' end differentiated by primer/ddNTP fluorophore emission spectra.

Sequence to right = 5'-AGTCTG-3'



Interpreting the Chromatogram

Sequencing ladder base calling is performed automatically by capillary electrophoresis software.

Software calculates a **Phred score** for each base call, a quality value (**Q-score**) that predicts the probability of base call error using the following formula:

$$Q = -10\log_{10}(p)$$

Where **Q** is the calculated quality score, and **p** is the estimated error.

The higher the Q-score (aka Phred score), the lower the probability a base is mis-called.

Q-Score	Incorrect Call Probability	Base Call Accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

In general, a Q-score ≥ 20 is considered acceptable.

Interpreting the Chromatogram

Phred score is graphically visualized in sequencing software with bars placed just above the base call.

- Larger bars indicate a higher Q-score and higher confidence in the base call.
- Smaller bars indicate a lower Q-score and lower confidence in the base call.

Bases with too low of a Phred score are uncalled by analysis software and labelled generically as “N” for nucleotide.

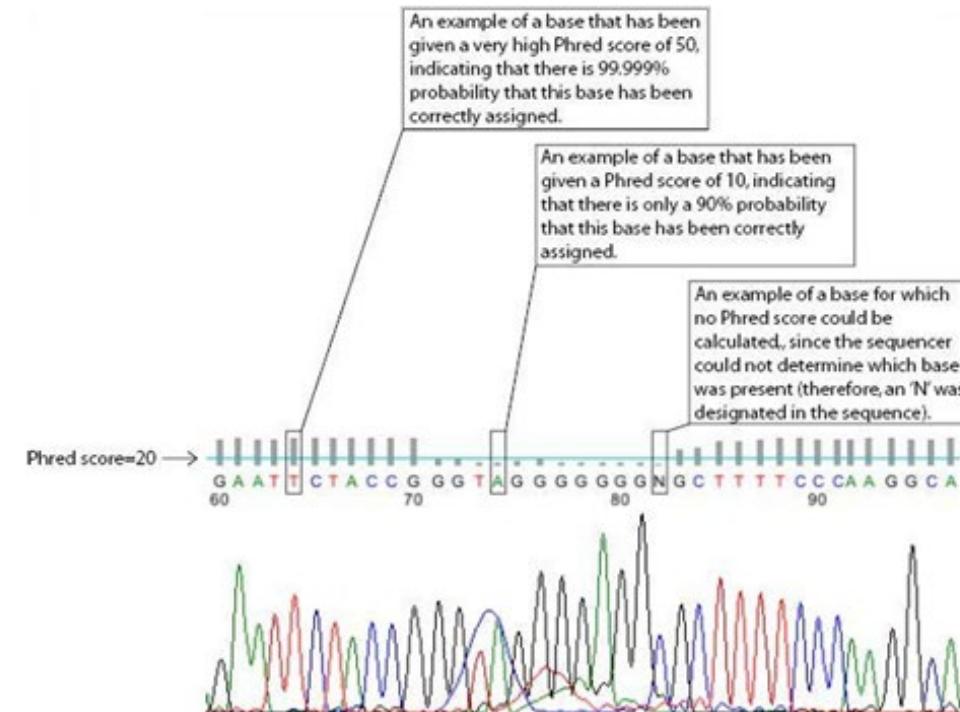


Figure 1. An example of a DNA sequence tracing and the Phred score (grey bars) corresponding to each colored peak. The colored peaks on the trace correspond to each DNA letter. For example 'T' bases are represented in red, and this sequence has four 'T' bases on a row, as viewed by the four red peaks in the sequence. The aqua horizontal line placed across the grey bars represents a Phred score of 20 which is considered an acceptable level of accuracy. As indicated in Table 1, a Phred score of 20 corresponds to a 99% accuracy in the base call. Therefore, bars above this line indicate base calls that have a higher than 99% probability of being correct. Those below have less than a 99% probability of being correct. Sequence tracing program is courtesy of FinchTV (www.geospiza.com).

Interpreting the Chromatogram

Dye blobs are common artifacts caused by unincorporated dye-labelled terminators.

Comigrate with amplicons and create large flashes of fluorescence that obscure the sequencing ladder and make it impossible to accurately call bases.

Can be prevented by cleaning-up sequencing product prior to CE.

- “Salting out” DNA precipitation
- Spin column/magnetic bead DNA purification
- Bead-driven sequestration and removal of dye terminators

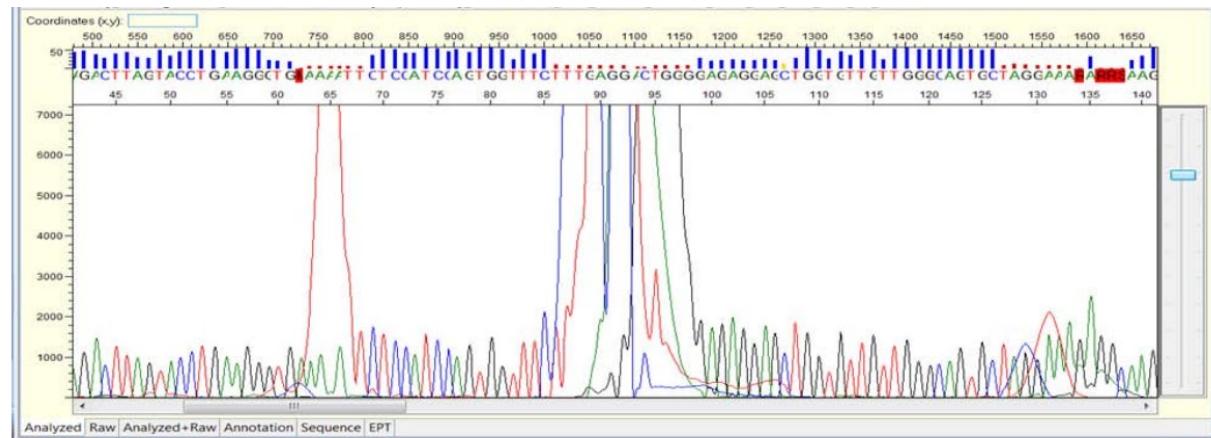


Figure 10 Severe dye blobs in the 60–65bp and 125–140bp regions

Interpreting the Sequence

For most bases in a sequence, an A/C/G/T designation can be given to indicate the expected or observed base at any given position.

The International Union of Biochemistry and Molecular Biology (**IUB**) have also established standardized symbols to capture mixed bases, normal “wobble” and other polymorphisms.

For example:

- Reference sequence:
5'-CACATCCYACTGCAA-3'
- Possible observed variants that still “match” the reference:
5'-CACATCCC~~A~~CTGCAA-3'
5'-CACATCCTACTGCAA-3'

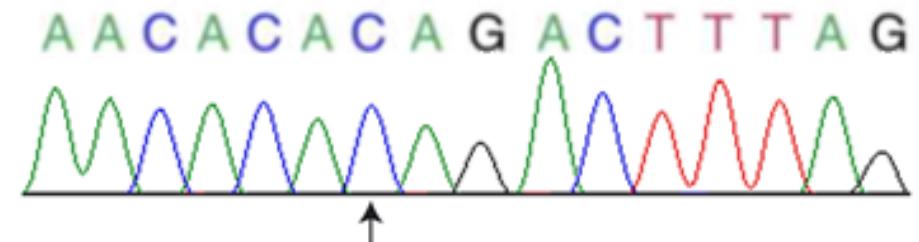
TABLE 9.5 IUB Universal Nomenclature for Mixed Bases

Symbol	Bases	Mnemonic
A	Adenine	Adenine
C	Cytosine	Cytosine
G	Guanine	Guanine
T	Thymine	Thymine
U	Uracil	Uracil
R	A, G	puRine
Y	C, T	pYrimidine
M	A, C	aMino
K	G, T	Keto
S	C, G	Strong (3 H bonds)
W	A, T	Weak (2 H bonds)
H	A, C, T	Not G
B	C, G, T	Not A
V	A, C, G	Not T
D	A, G, T	Not C
N	A, C, G, T	aNy
X, ?	Unknown	A or C or G or T
O, -	Deletion	

Interpreting the Sequence

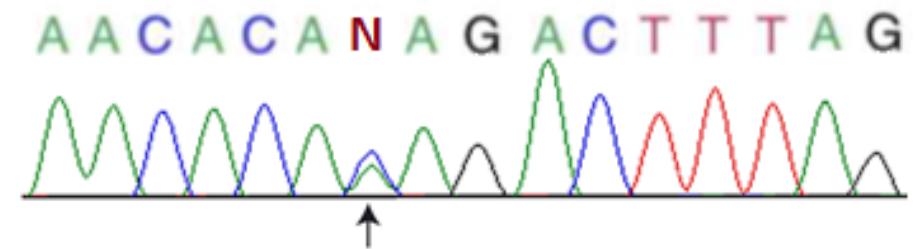
WT sequence (top):

- Both alleles: 5'-AACACACAGACTTTAG-3'



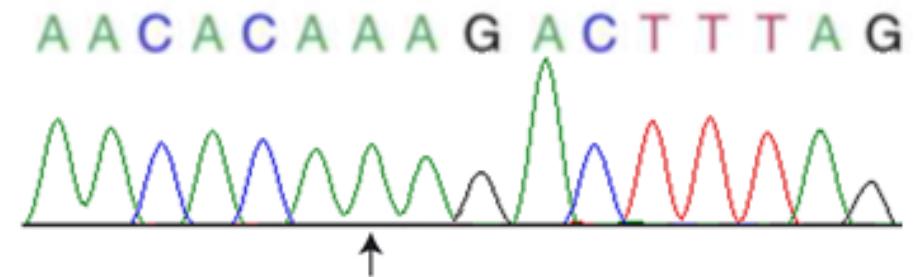
Heterozygous C>A variant (middle)

- Allele 1: 5'-AACACACAGACTTTAG-3'
- Allele 2: 5'-AACACAAAGACTTTAG-3'
- Combo: 5'-AACACACMAGACTTTAG-3'



Homozygous C>A variant (bottom)

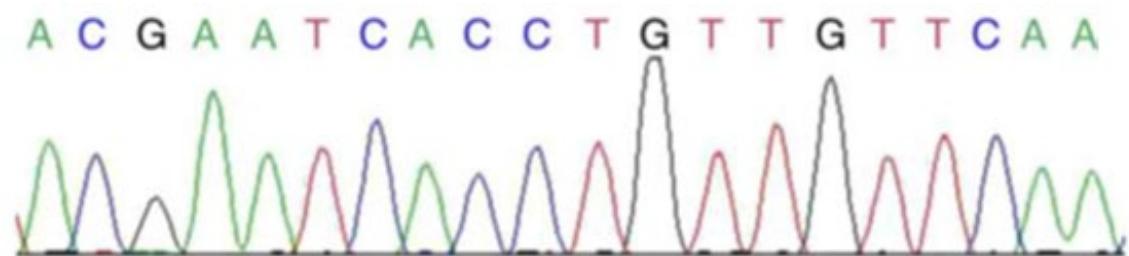
- Both alleles: 5'-AACACAAAGACTTTAG-3'



Interpreting the Sequence

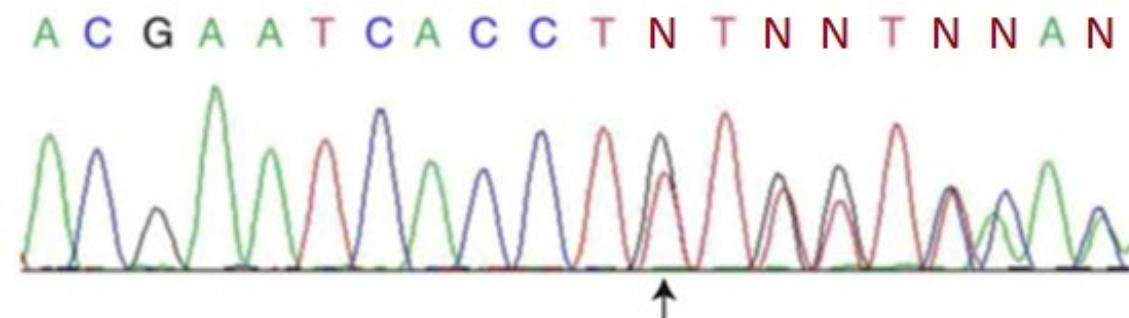
WT Sequence (top)

- 5'-ACGAATCACCTGTTGTTCAA-3'



Heterozygous frameshift mutation caused by single guanine deletion (bottom)

- Allele 1: 5'-ACGAATCACCT**G**TGTTCAA-3'
- Allele 2: 5'-ACGAATCACCT**O**TTGTTCAA-3'



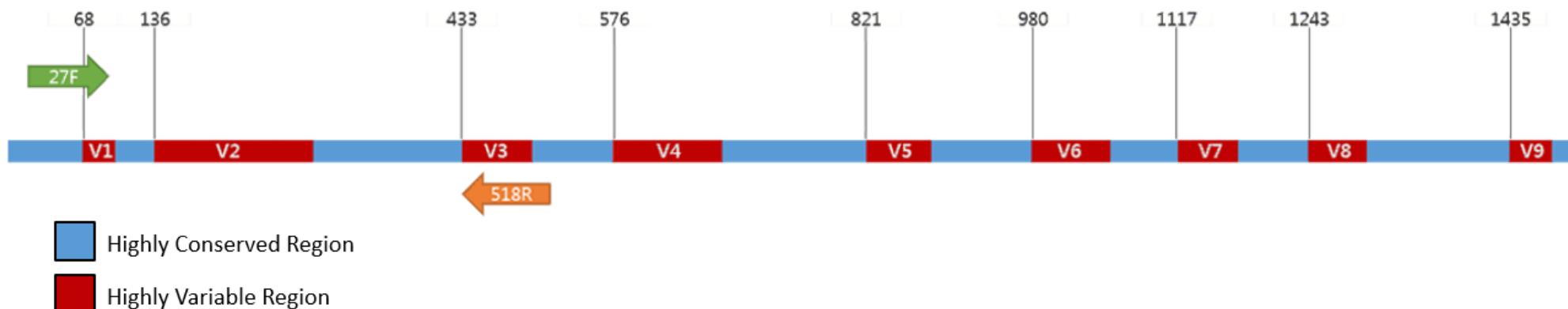
CCF's 16S rDNA Sequencing

CCF sequences the 16S ribosomal RNA gene of unknown bacterial isolates as a genetic means of identification.

- Highly conserved regions = ideal for primer design
- Highly variable regions = ideal for species differentiation

A 500bp region of the 16s RNA gene (variable regions V1-V3) is PCR-amplified using universal forward and reverse primers.

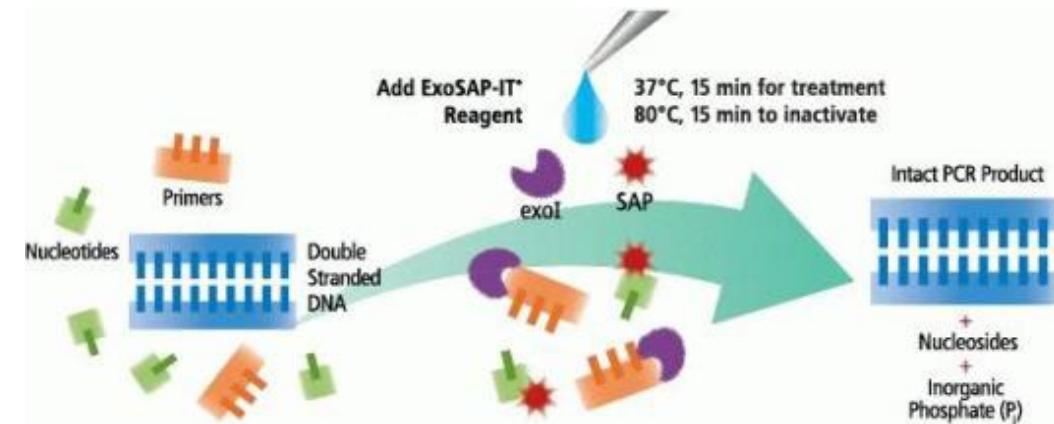
- dNTPs only for this PCR step



CCF's 16S rDNA Sequencing

Following universal amplification, PCR product is then “cleaned up” using an alkaline phosphatase enzyme to digest residual dNTPs and unextended primers.

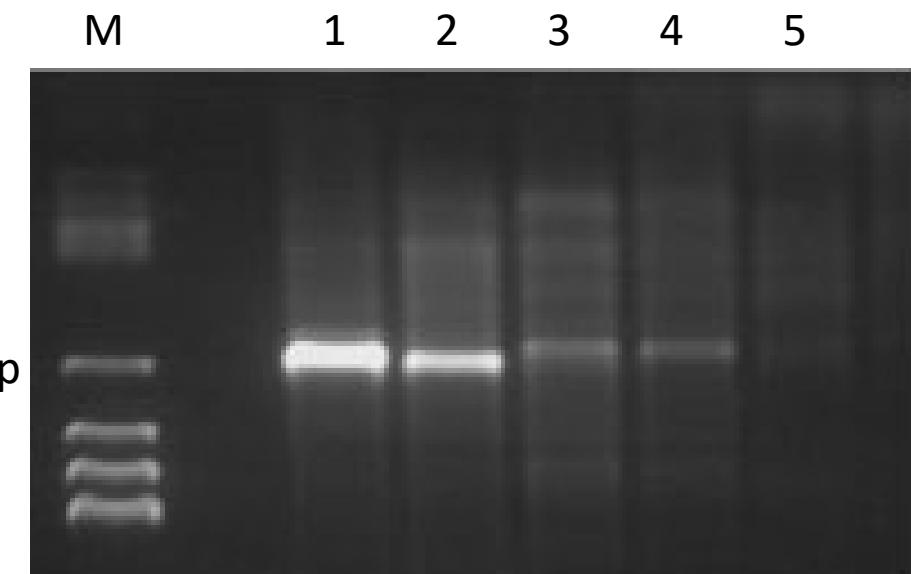
- Shrimp Alkaline Phosphatase (SAP) dephosphorylates free nucleotides, producing nucleosides and free pyrophosphate
- Without the triphosphate group, nucleosides can no longer be incorporated
- Heat denaturation then inactivates SAP enzyme prior to sequencing with ddNTPs



CCF's 16S rDNA Sequencing

Cleaned up PCR-product is analyzed for sequencing suitability using agarose gel electrophoresis.

- PCR-product is loaded into E-Gel alongside sizing ladder.
- PCR-product should ideally show an intense band between 420-540bp in length.
- E-Gel image to the right:
 - M = size ladder
 - Samples 1, 2 = good size and quantity
 - Samples 3, 4 = good size, poor quantity
 - Sample 5 = no amplification



CCF's 16S rDNA Sequencing

Cycle sequencing is then performed on the cleaned-up product, using:

- Forward and reverse sequencing primers to capture both complementary strands.
- Unlabeled dNTPs and fluorescently-labelled ddNTPs (dye terminator method) to generate amplicons of every possible bp length.

ddCTP = blue dye

ddATP = green dye

ddGTP = yellow dye (appears black in software)

ddTTP = red dye

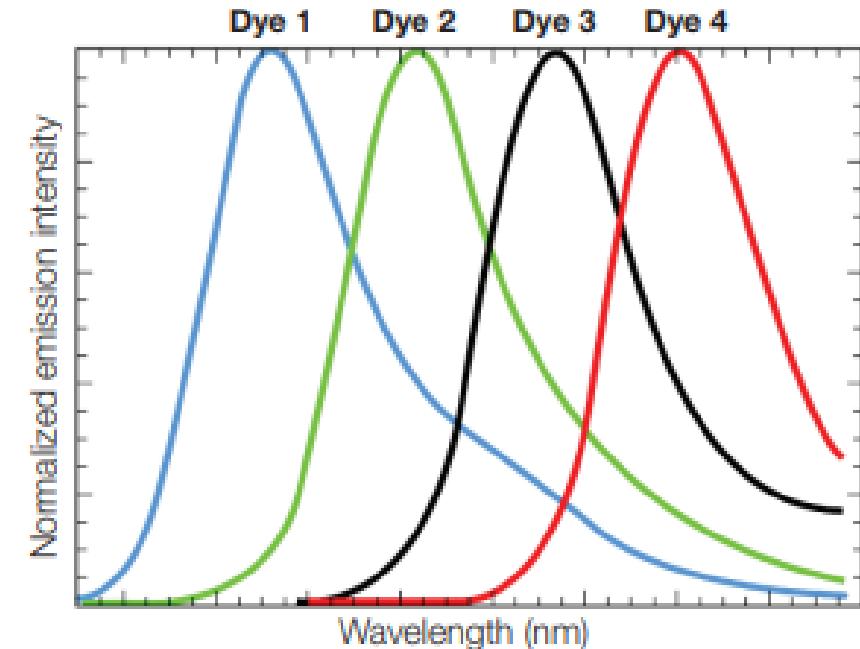


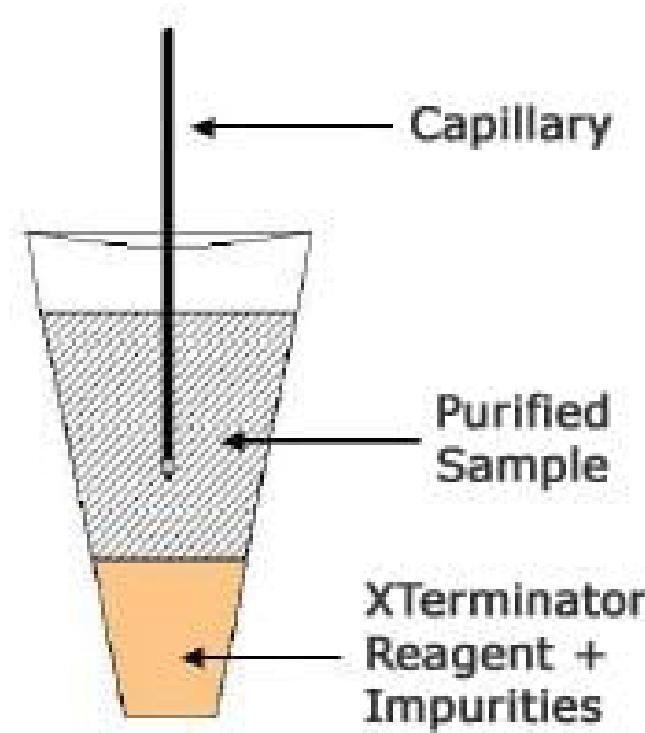
Figure 6. Emission spectra of the four BigDye dyes.
Dye 1 = Big-d110, Dye 2 = R6G, Dye 3 = Big-dTAMRA, and
Dye 4 = Big-dROX.

CCF's 16S rDNA Sequencing

Cycle sequencing product is cleaned up through bead-driven removal of unincorporated dye terminators to prevent dye blob artifact.

- BigDye Xterminator Purification Kit
- Sand-like beads capture impurities and sediment at the bottom of sample well.

Final cleaned-up sequencing product is loaded into CE plate and resolved using the ABI 3730 DNA Analyzer.



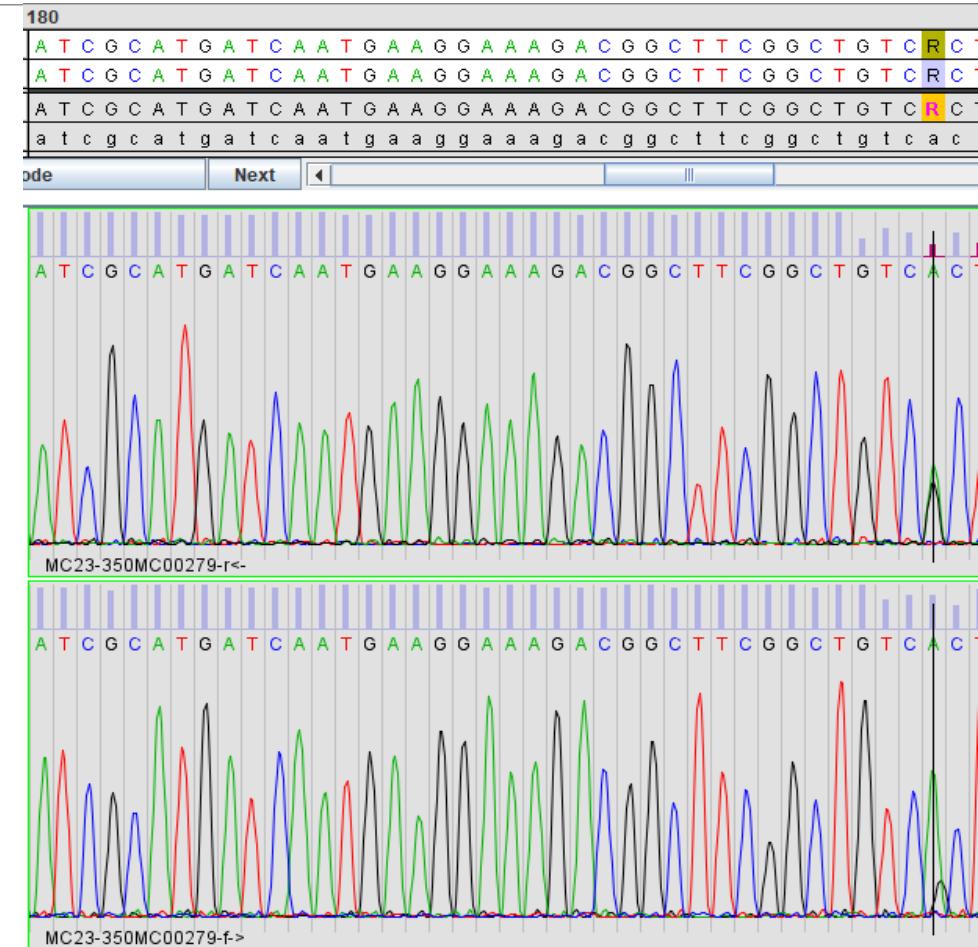
CCF's 16S rDNA Sequencing

Sequencing technologist manually reviews forward and reverse sequencing ladders for errors in software's auto base calling.

Image to the right:

- For both forward and reverse reads, software originally miscalled position 218 as "A" despite the presence of both adenine and guanine peaks.
- Position 218 was manually changed to "R" at the top to capture both peaks.

Note: Software makes read comparison easier by automatically flipping the reverse ladder to its complement.



CCF's 16S rDNA Sequencing

Once manual review of the sequencing ladder is complete, a secondary sequencing software is used to compare finalized sample sequence against reference sequences to identify the best possible organism match.

16s Sequence

```
GACGAACGCTGGCGCGTGCCTAATACATGCAAGTAGAACGCTGAAGGAGGAGCTTC  
TTCTGAATGAGTTGCGAACGGGTGAGTAACCGTAGGTAACCTGCCGTAGCGGGGGGAT  
AACTATTGAAACGATAGCTAATACCGCATAATAGTAGATGTCATGRCATTTGCTTGA  
AAGGTGCAATTGCACTACCACTACAGATGGACCTGGCTGTATTAGTAGTTGGTGGGTA  
CGGCTCACCAGGGCACGATACTAGCCGACCTGAGAGGGTGATGCCACACTGGACT  
GAGACAGCGCCAGACTCTACGGGAGGGAGCAGTAGGAACTTCGGCAATGGACGGAA  
GTC TGACCGAGCAACGCCGCGTGA GTGA AGAAGGTT CGGATCGTAAAGCTCTGTTGA  
AGAGAAGAACGAGTGTGAGAGTGGAAAGTTCAACTGTGACGGTATCTTACAGAAAGGG  
ACGGCTAACTACGTG
```

Paste nucleotide sequence here

16s Sequence Length Number of nucleotides in the final sequence generated from this specimen

Which database(s) did you use?

SmartGene IDNS 16s Centroids
 SmartGene IDNS 16s Eubacteria
 SmartGene IDNS 16s Mycobacteria
 SmartGene NIH 16s Nocardia
 NCBI (nr/nt)
 Other

Upload 16s rRNA gene bacterial sequencing PDF results here [smg_ALT102923-1_Strep.pdf \(0.08 MB\)](#)
* must provide value [Upload new version](#) or [Remove file](#) or [Send-it](#)

Name of top hit organism for 16s

Percent identity (%)

Number of mismatches

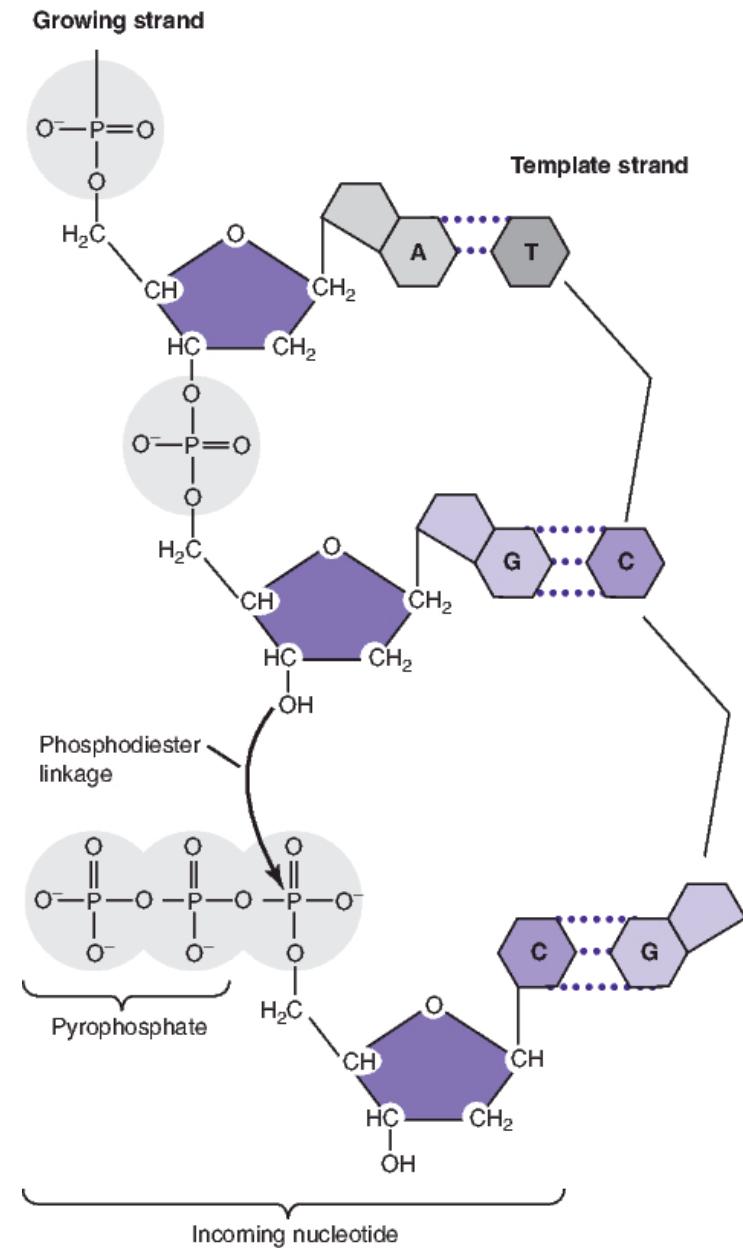
Match length

Score

Pyrosequencing

Recall the chemistry of DNA synthesis:

- DNA polymerase reads the template strand and incorporates complementary deoxynucleoside triphosphates (**dNTPs**) into the growing strand.
- Whenever a phosphodiester linkage is formed between the 3' hydroxyl of the growing strand and the 5' phosphate group of an incoming dNTP, a **pyrophosphate (PPi)** is released.



Pyrosequencing

Pyrosequencing is a sequencing-by-synthesis methodology that works by detecting the release of pyrophosphate molecules during nucleotide incorporation.

In pyrosequencing, PPi release is detected through a chemiluminescent reaction.

By adding nucleotides sequentially and monitoring for luminescence with each addition, sequence can be determined.



Fun fact: fireflies glow using the same chemiluminescent reaction as pyrosequencing.

What do I need for Pyrosequencing?

Generic components:

- Sequencing primer
- Polymerase
- Deoxynucleoside triphosphates (dNTPs)
- Buffer

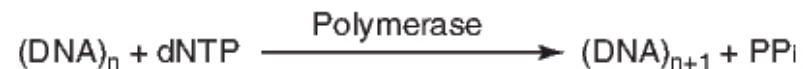
Additional enzymes:

- Sulfurylase
- Luciferase
- Apyrase

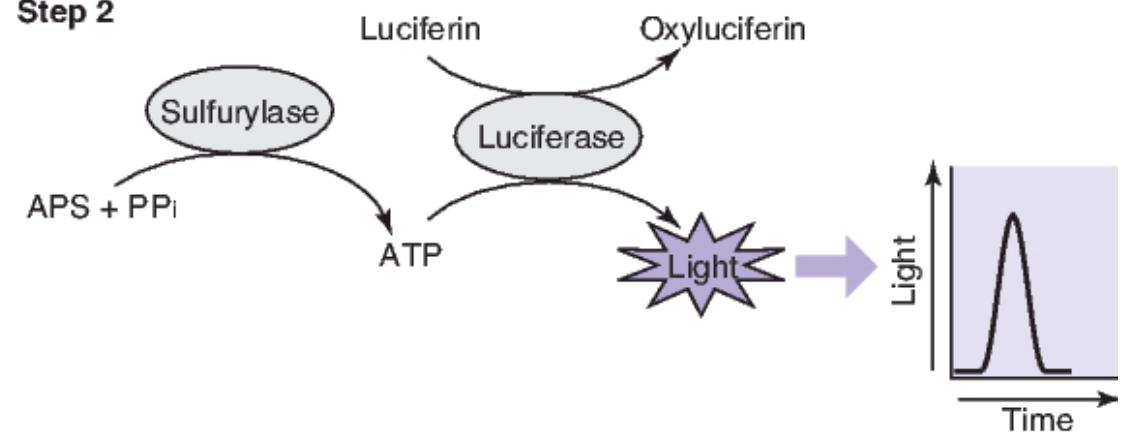
Additional substrates:

- Adenosine 5' phosphosulfate (APS)
- Luciferin

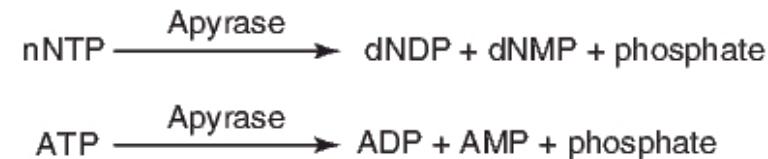
Step 1



Step 2



Step 3

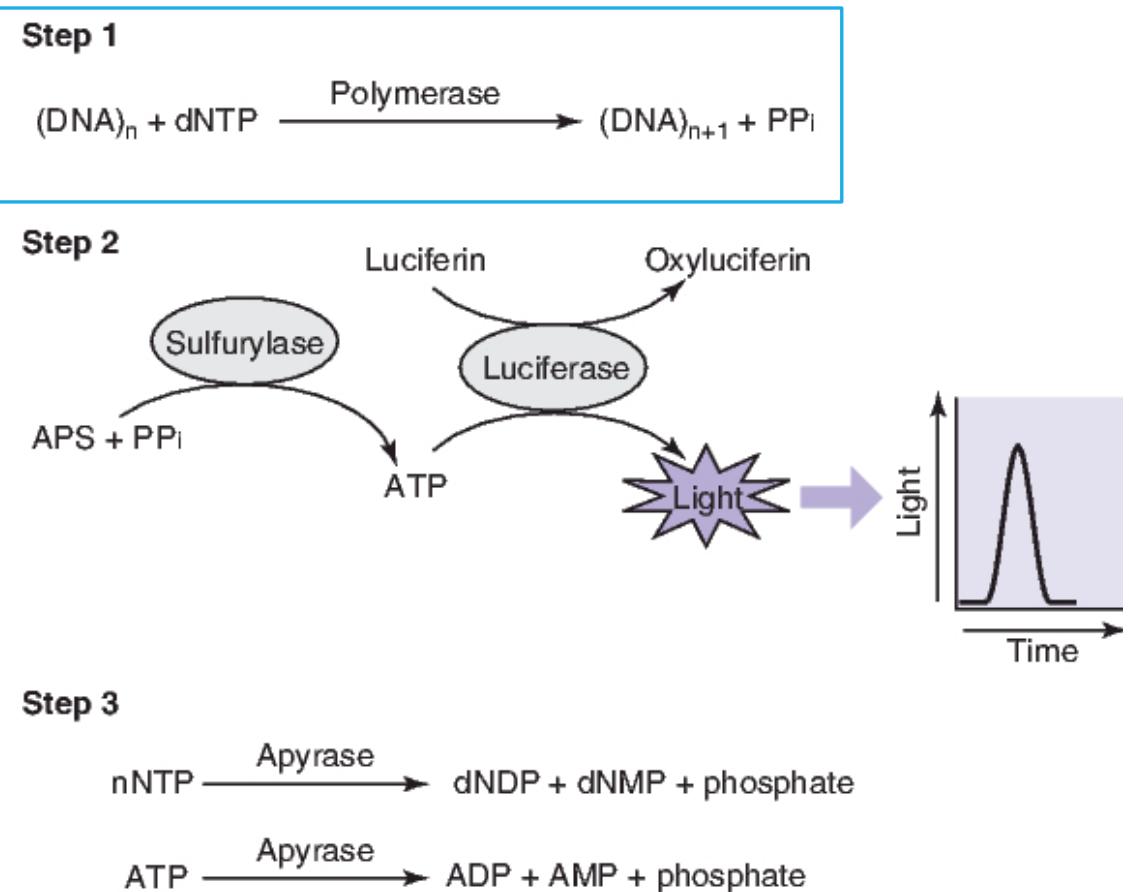


Step 1: Nucleotide Incorporation

One of four **dNTPs** is added to reaction mixture (dATP, dCTP, dGTP, dTTP).

If complementary to the next base in the template strand, **polymerase** will incorporate the dNTP into the growing strand of DNA.

dNTP incorporation releases a **pyrophosphate** molecule.

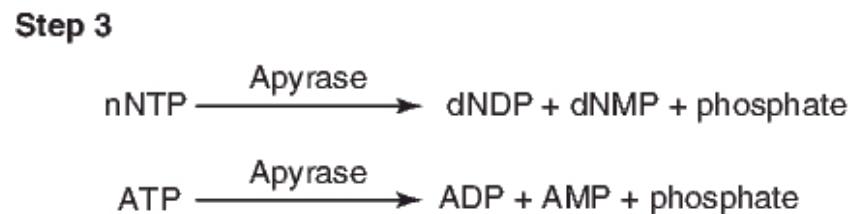
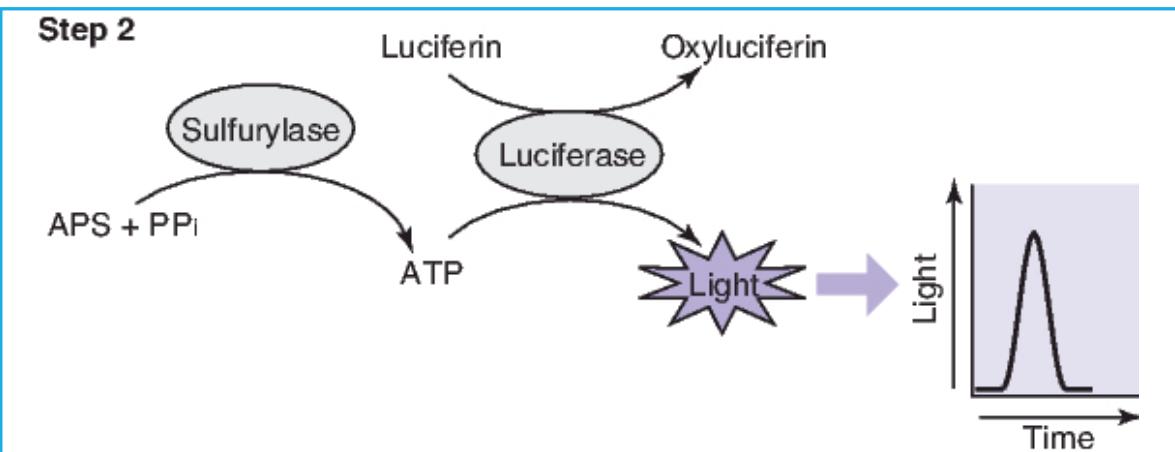
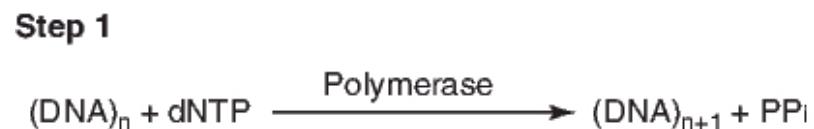


Step 2: Luminescence and Detection

Sulfurylase uses the free pyrophosphate to convert adenosine 5' phosphosulfate (**APS**) to adenosine triphosphate (**ATP**).

Luciferase uses the energy from ATP to catalyze the oxidation of **luciferin** to **oxyluciferin**.

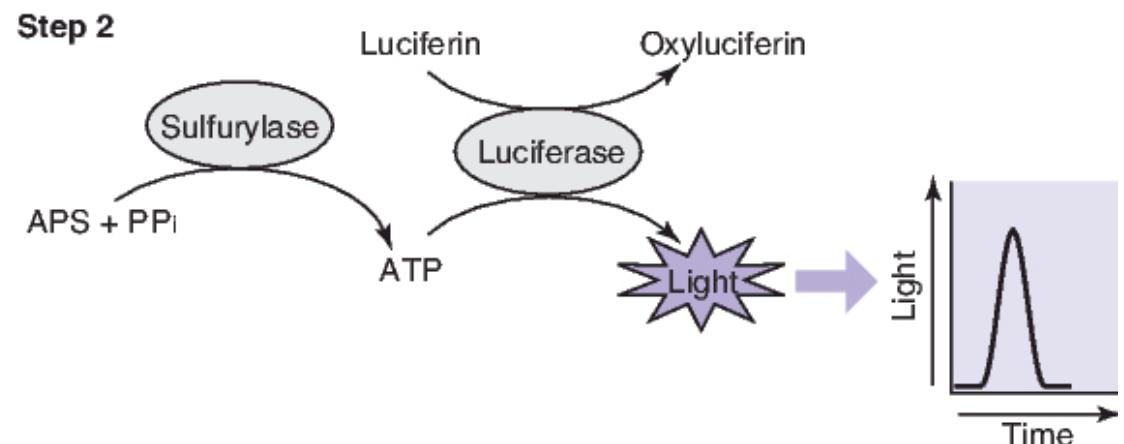
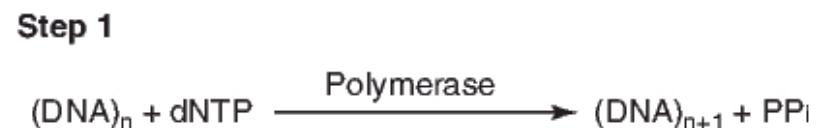
This oxidative reaction produces a flash of light, which is detected by a camera and visualized on a **pyrogram**.



Step 3: System Regeneration

Apyrase degrades unincorporated dNTPs and ATP molecules.

This resets the reaction mixture for subsequent nucleotide addition.

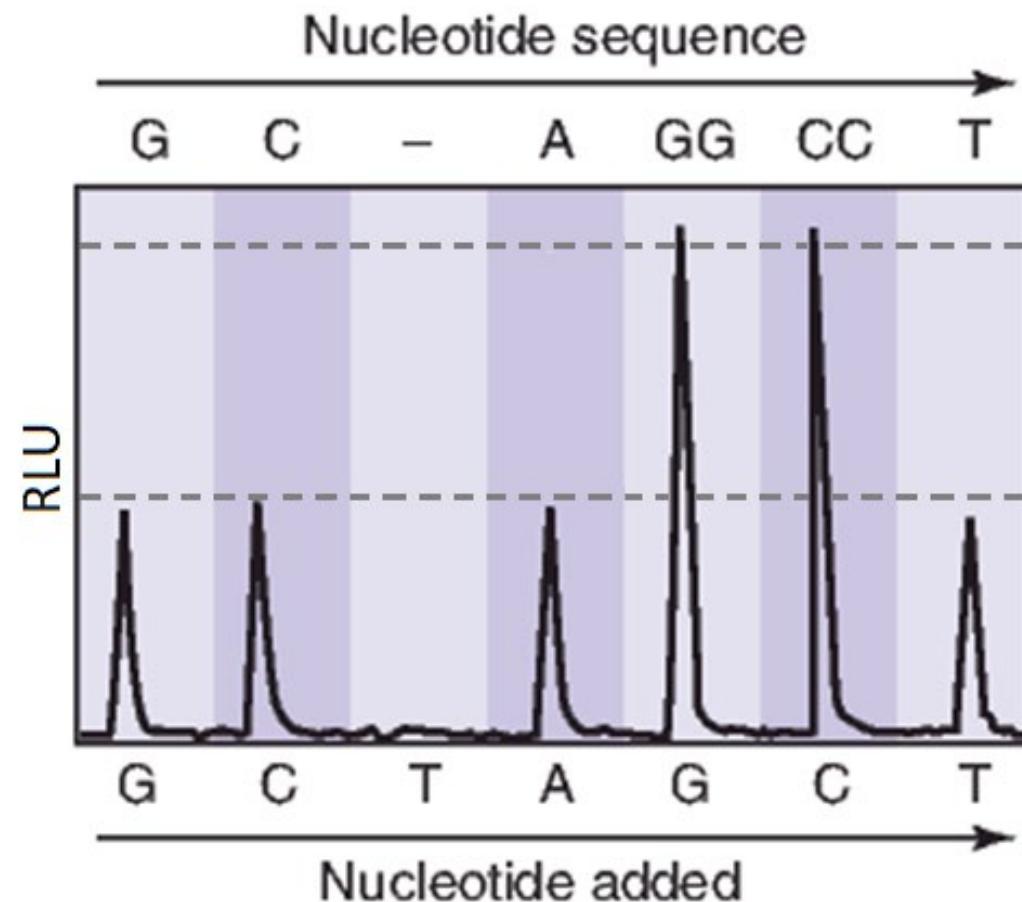


Reading the Pyrogram

A **pyrogram** plots relative light units (**RLU**) on the y-axis, and the specific nucleotide added at each cycle of the reaction on the X-axis.

The amount of light generated by nucleotide addition is proportional to the number of nucleotides that are incorporated into the growing strand:

- No dNTP incorporated = 0x RLU
- One dNTP incorporated = 1x RLU
- Two dNTPs incorporated = 2x RLU
- Etc...



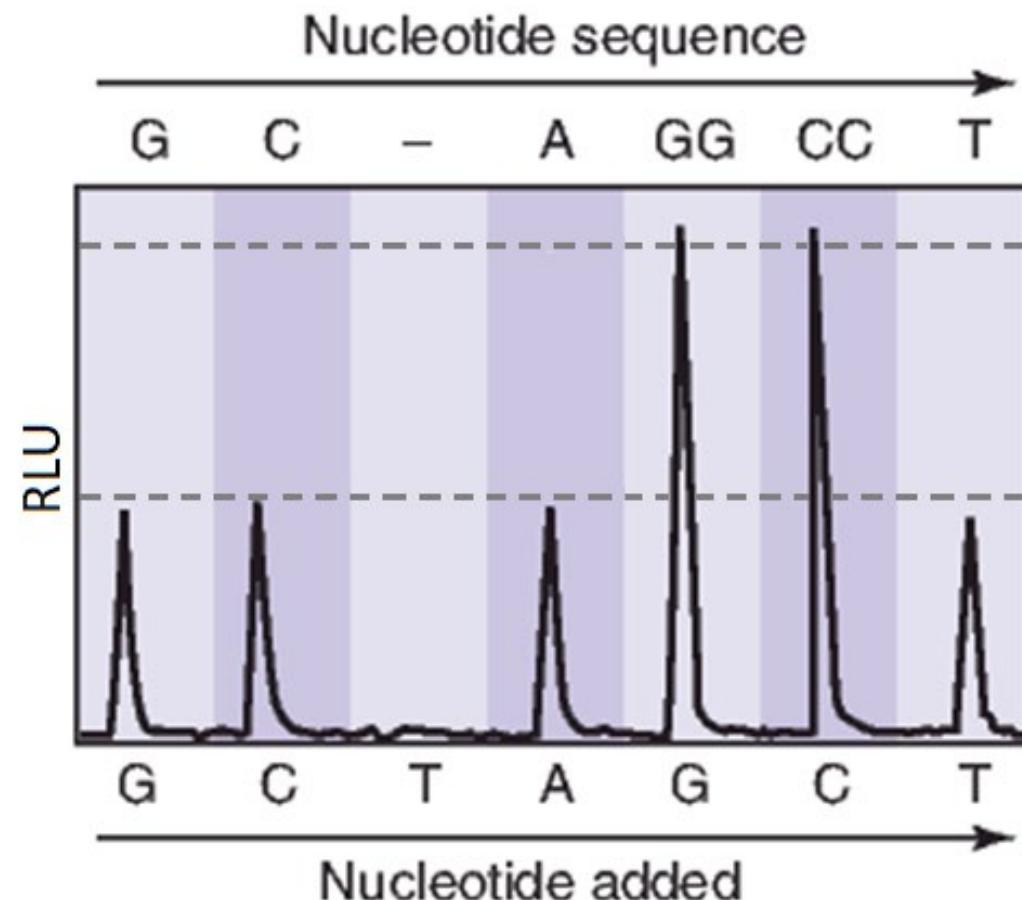
Reading the Pyrogram

Sequence is synthesized and graphed 5' to 3'.

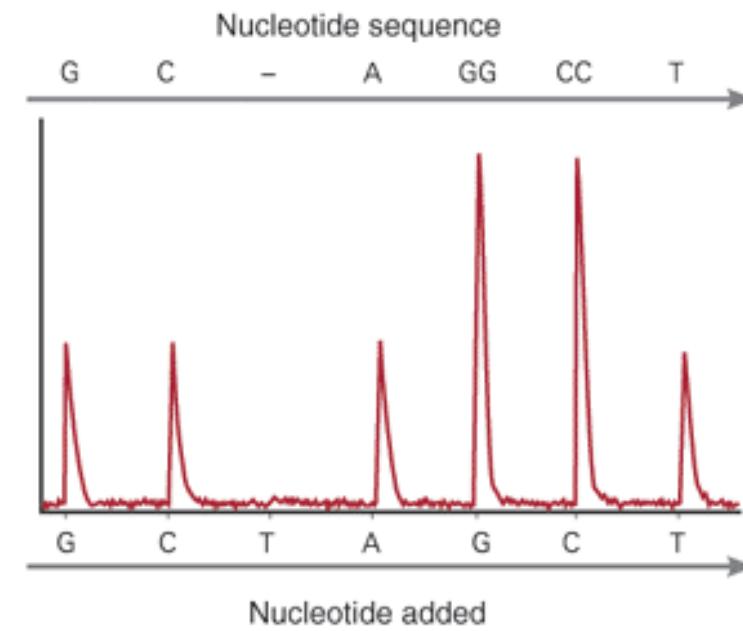
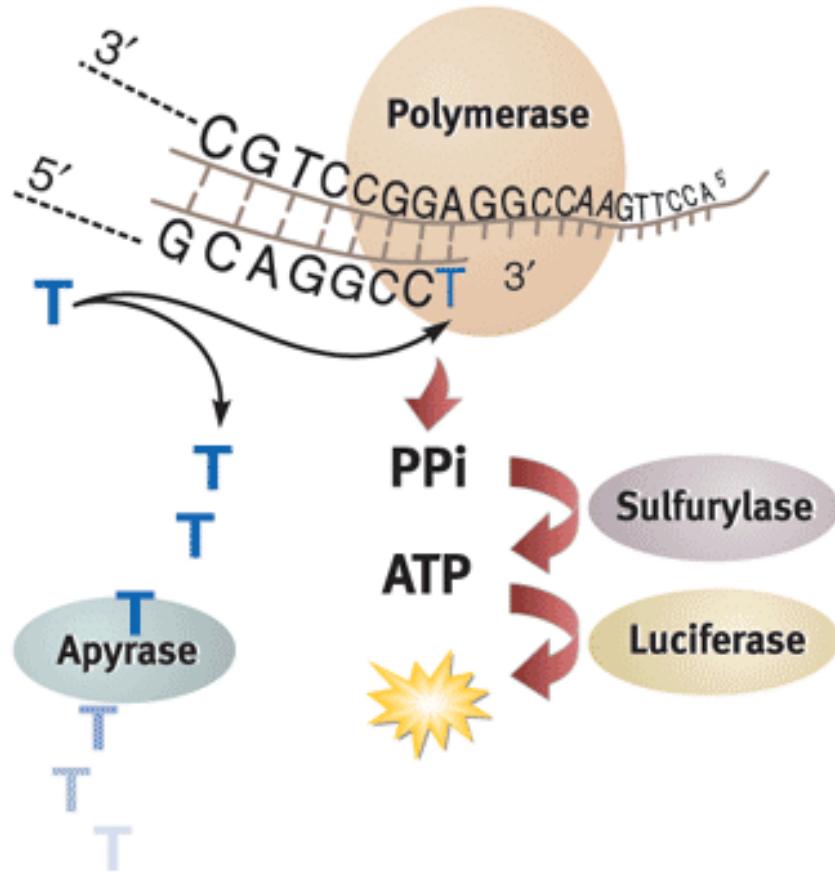
- The sequence in the pyrogram to the right can be read as 5'-GCAGGCCT-3'.

The sequence generated is complementary to the sequence being read by polymerase.

- The original template strand can be inferred as 5'-AGGCCTGC-3'
- Sequencing primer design will determine if sense or antisense strand is synthesized.



Pyrosequencing in one picture

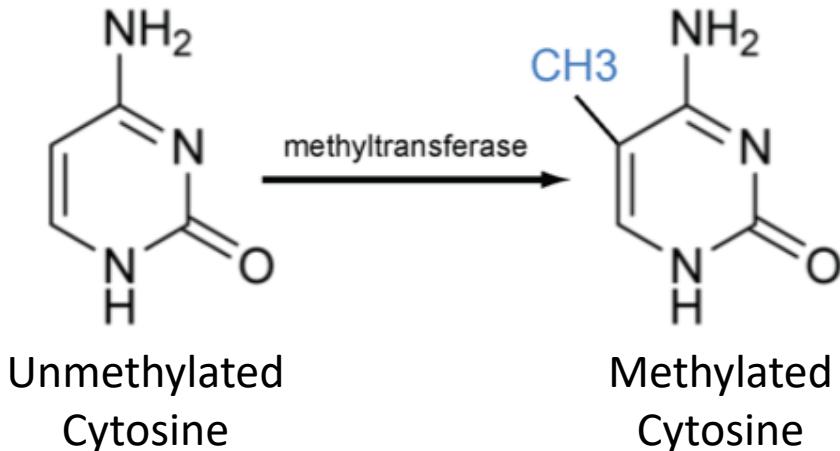


Methylation Specific Sequencing

- Bisulfite-Pyrosequencing
- Bisulfite-Sanger Sequencing

Recall the epigenetic effects of methylation:

- **CpG islands:** cytosine-guanine-rich sequences found in first exons, promoter regions and toward the 3' ends of genes.
 - CpG = cytosine followed immediately by guanine
- **Methylation** of cytosine residues prevents transcription of a gene, thereby down-regulating expression of its RNA/protein product.
- Aberrant methylation/demethylation of cytosines can lead to silencing/activation of a gene, and subsequent under/over-expression of RNA and proteins products.



Methylation-Specific Sequencing

Methylation-Specific Sequencing uses sequencing methodologies to analyze the methylation status of a specific gene target to predict the likely expression of that gene.

Also called **bisulfite sequencing** due to its first analytic step: **bisulfite conversion**.

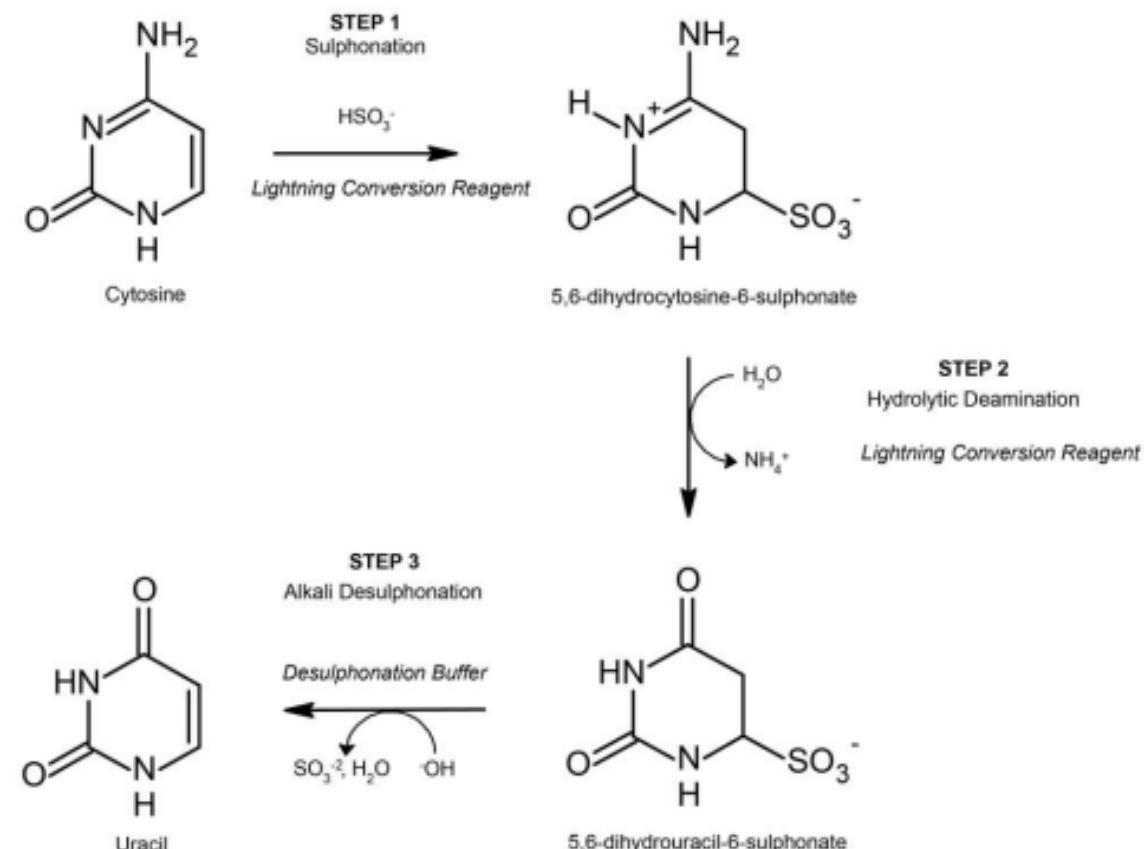
Bisulfite-converted product is then sequenced using one of the sequencing methods previously described:

- Pyrosequencing
- Sanger Sequencing

Bisulfite Conversion

Bisulfite Conversion of Cytosine to Uracil

- Step 1: **nonmethylated** cytosine residues, are sulphonated
- Step 2: Amine group is removed by hydrolysis
- Step 3: Base introduction (NaOH) desulphonates the molecule, leaving behind uracil



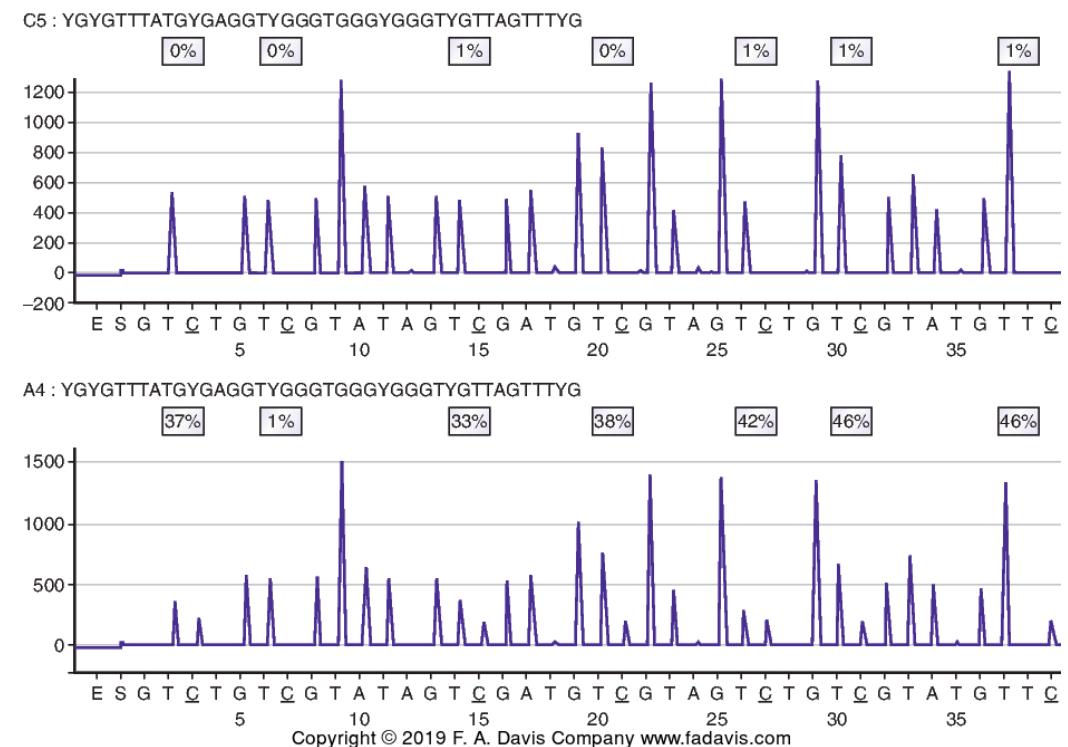
Methylated cytosine is unable to undergo bisulfite conversion and will remain as cytosine when sequenced.

Bisulfite-Pyrosequencing

Bisulfite-converted DNA is then sequenced to quantify the amount of converted/unconverted CpG islands.

With pyrosequencing, dTTP and dCTP are added sequentially at each CpG site.

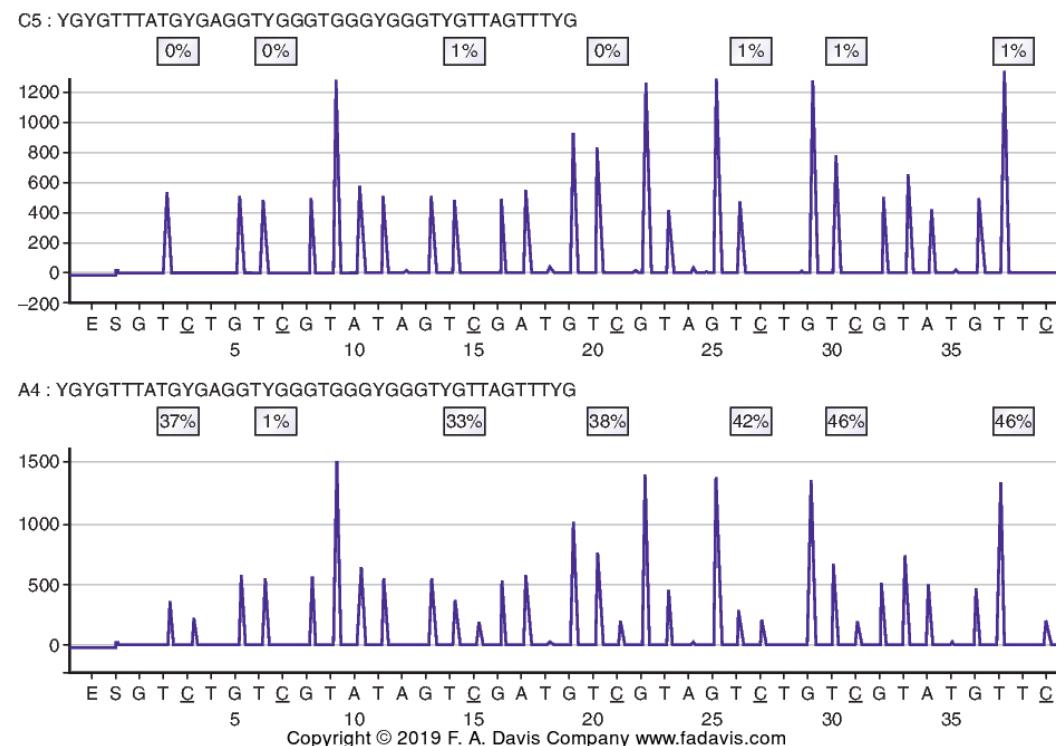
- Note: Thymine is analogous to uracil and substitutes it in the pyrosequencing reaction.



Bisulfite-Pyrosequencing

By comparing the amount of light (RLU) produced when dTTP and dCTP is added, a percent methylation can be calculated for each CpG site.

- Lots of light during dTTP addition = lots of converted cytosines = low methylation
- Lots of light during dCTP addition = lots of unconverted cytosines = high methylation



Bisulfite-Sanger Sequencing

Pyrosequencing, the method of choice for CCF bisulfite sequencing, isn't the only available method.

Dideoxy chain termination sequencing can also be used to resolve cytosine-uracil conversion.

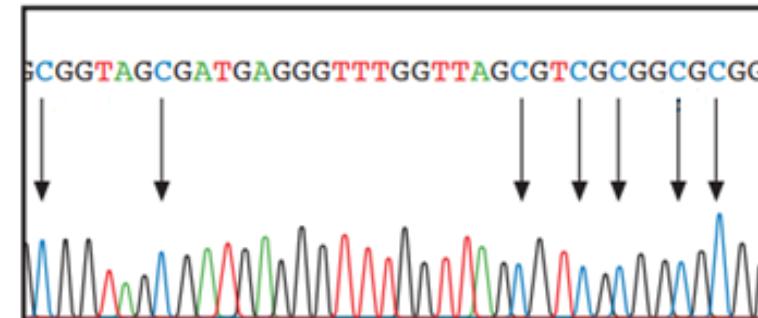


Figure 18. DNA sequence from untreated DNA. Arrows show locations of nonmethylated cytosine positioned before guanine. After bisulfite treatment, nonmethylated cytosine is converted to T.

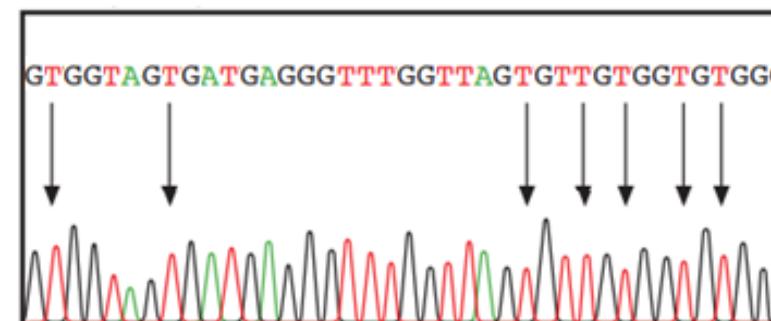


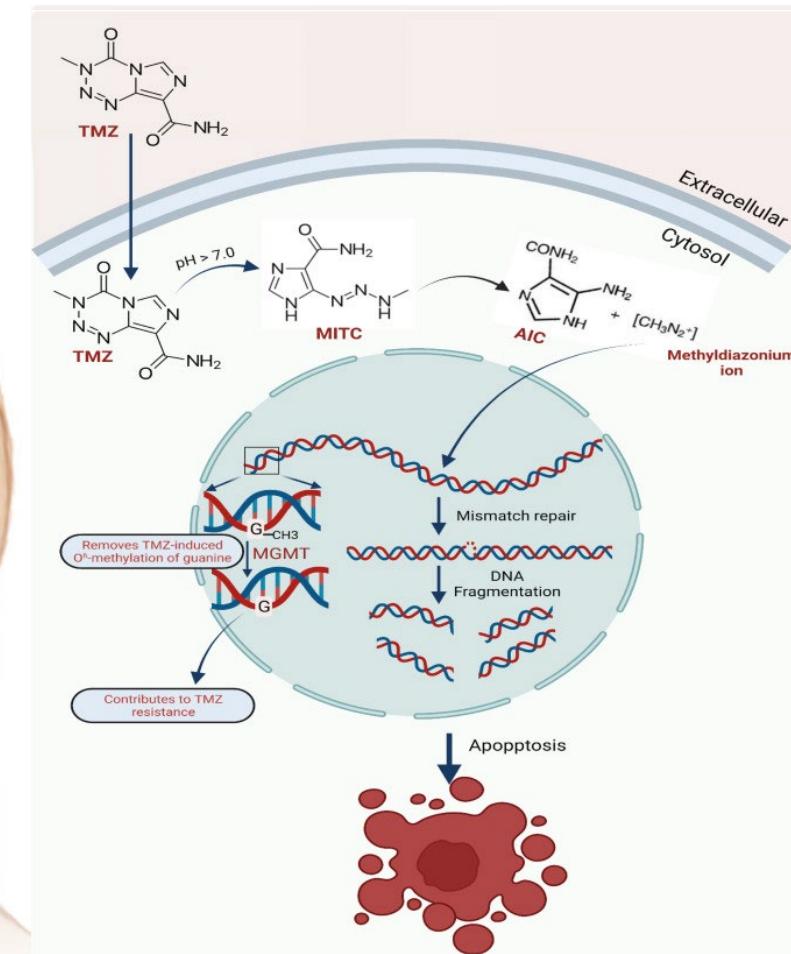
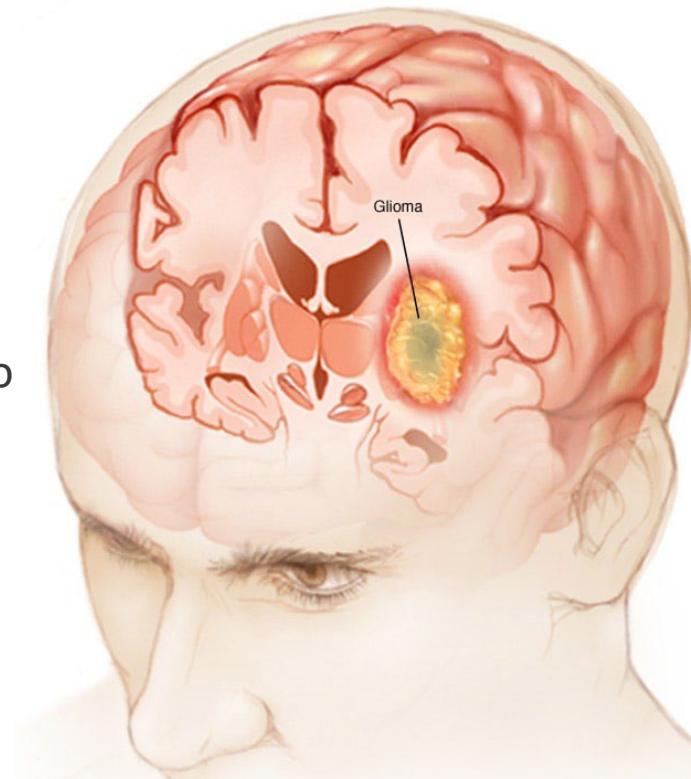
Figure 19. DNA sequence from bisulfite-treated DNA. Arrows show locations of nonmethylated cytosine converted to thymine after bisulfite sequencing.

MGMT Methylation Analysis

Let's use MDx's MGMT Methylation Analysis test as an example.

Background:

- O⁶-Methylguanine-DNA Methyltransferase (MGMT) is a DNA repair enzyme that reverses alkylation.
- Tumors that produce this protein have resistance to alkylating agents (temozolomide) by repairing damage to the tumor.
- Tumors with MGMT promoter methylation are inhibited from producing as much of this protein, making them more sensitive to treatment with alkylating agents.



MGMT Methylation Analysis

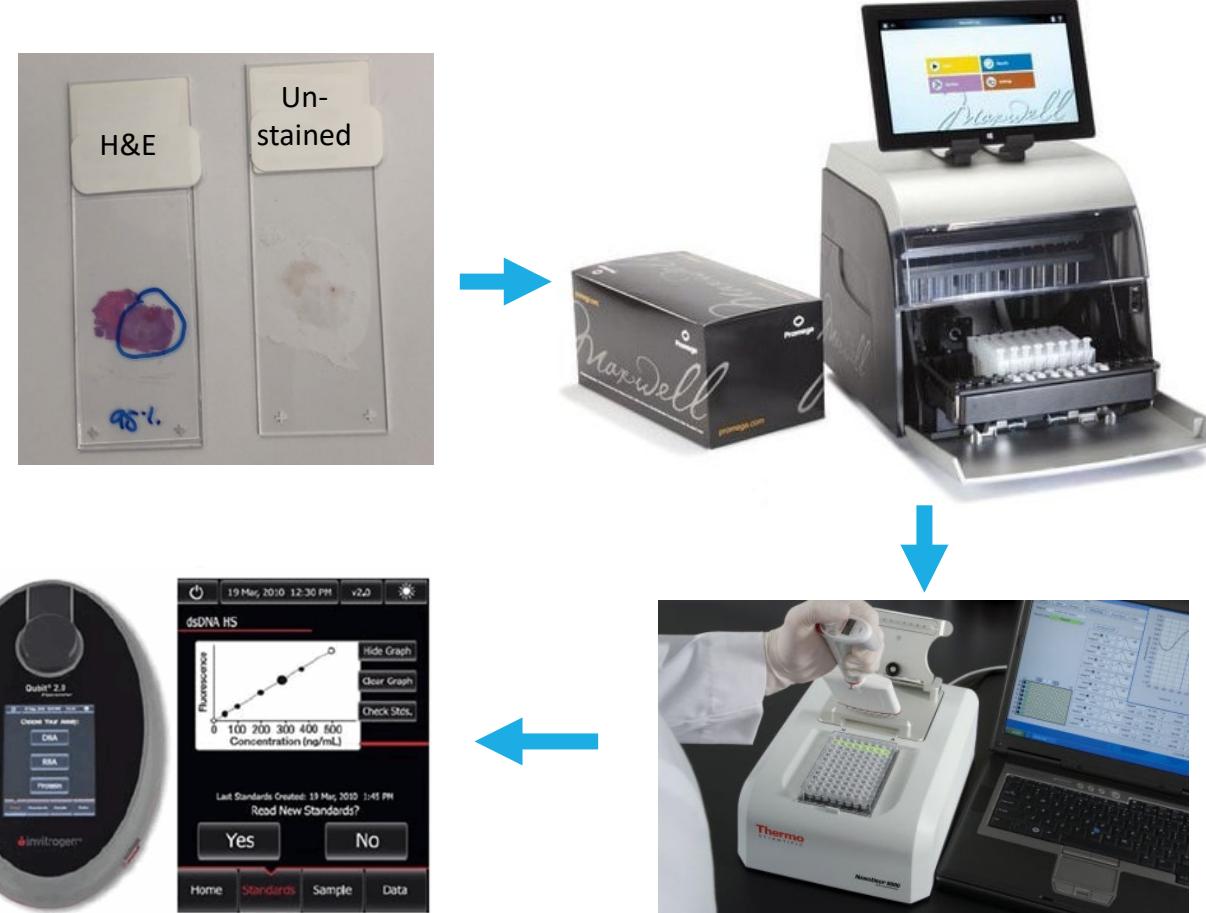
DNA is extracted from FFPE of patient tumor using a solid phase extraction methodology.

- H&E-stained slide, circled by AP, is used as template to isolate tumor from unstained slide

Extracted DNA is quantified/qualified using both spectrophotometry and fluorometry.

- Spec = quality indicators (230 and 280 ratios)
- Fluor = more specific quant

Sample DNA is diluted (as needed) with nuclease-free water.



MGMT Methylation Analysis

Diluted DNA is then bisulfite-converted using EZ DNA Methylation-Lightning Kit, containing:

- CT-Lightning Conversion Reagent
- Binding Buffer
- Spin Column with silica filter
- Wash Buffer (EtOH-based)
- Desulphonation Buffer
- Elution Buffer



MGMT Methylation Analysis

Bisulfite-converted DNA is amplified using forward and reverse primers targeting Exon 1 of the MGMT gene.

- Region contains 5 CpG sites (Y) and one CpA (Q) used as an internal quality control.

PCR-product is loaded onto the PyroMark Q48 instrument.

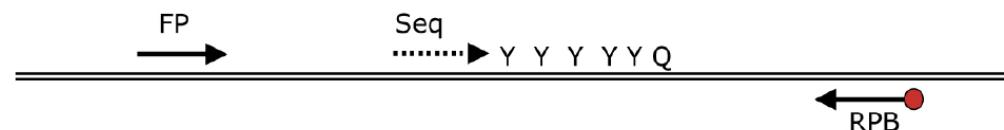
- Instrument uses silica beads to bind and retain amplicons, while washing away unincorporated PCR reagents (dNTPs, primers, etc.)



MGMT Methylation Analysis

Washed PCR-product is then sequenced by the same Pyromark Q48 instrument.

- Sequencing reaction is primed with a single nested sequencing primer which hybridizes closer to 5 CpG sites and internal control.
- Sequence to analyze:
5'-**Y**GTTTG**Y**GTTYG**A**-3'
 - Enzyme mixture (containing polymerase, sulfurylase, luciferase, and apyrase) is dispensed first.
 - Substrate mixture (containing APS and luciferin) is dispensed second.
 - dNTPs are then dispensed sequentially according to expected sequence above, with every "Y" including both a thymine dispensation and a cytosine dispensation.

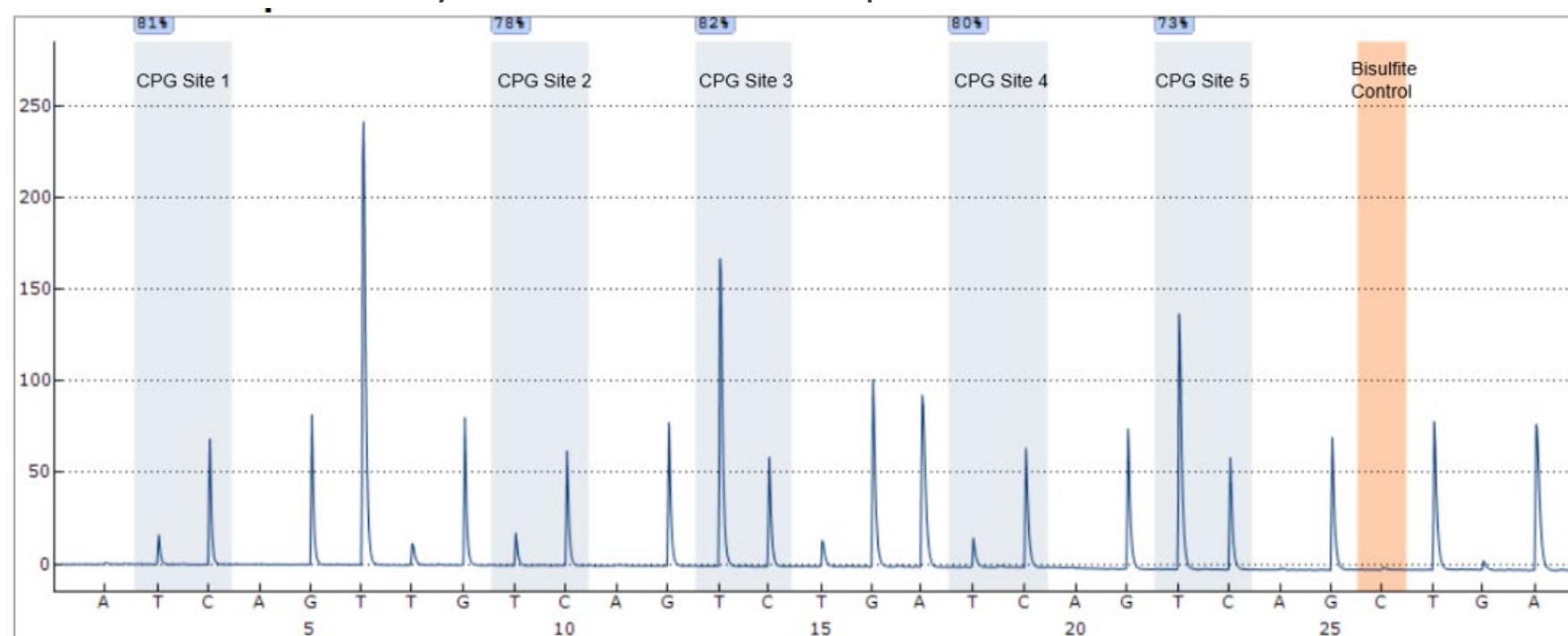


MGMT Methylation Analysis

As sequencing reaction occurs, instrument luminometer tracks and graphs luminescence generated from each dNTP incorporation on a pyrogram.

Methylation% at each CpG site is determined by the ratio of C:T incorporated.

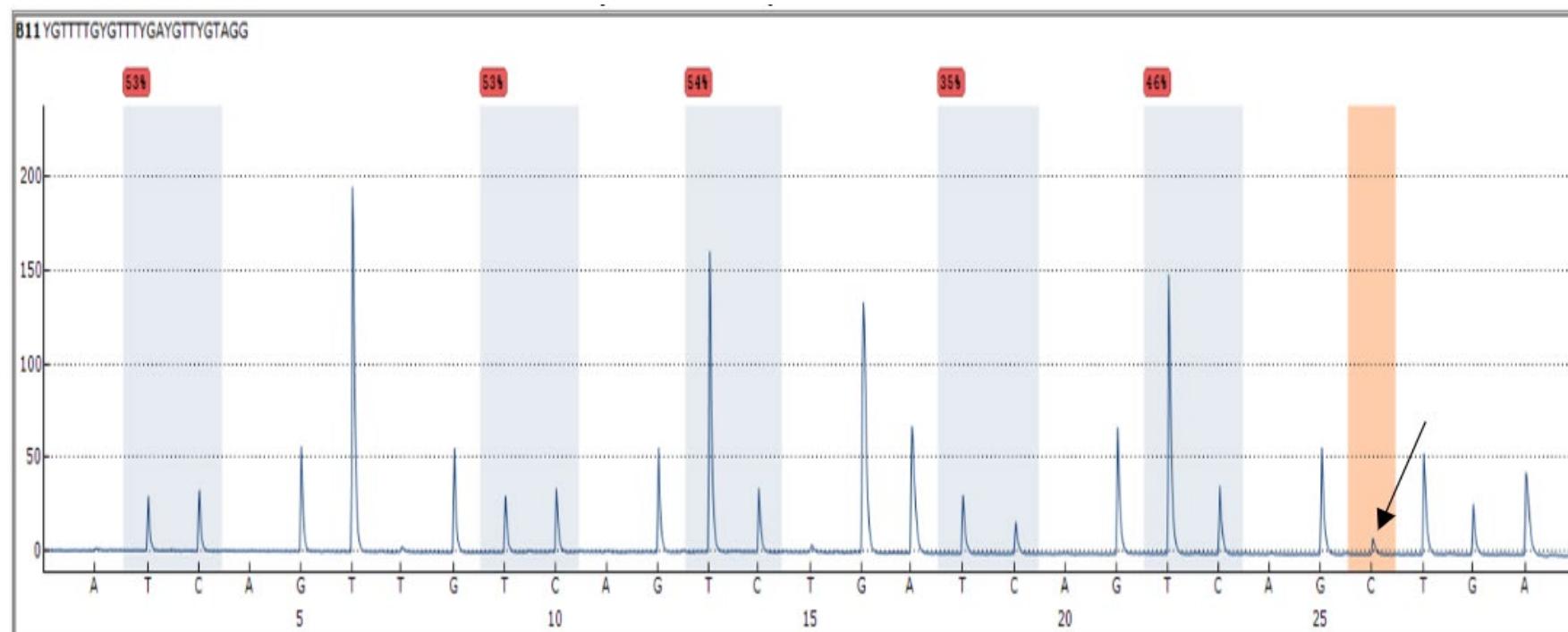
Mean methylation is calculated as an average of all 5 CpG sites.



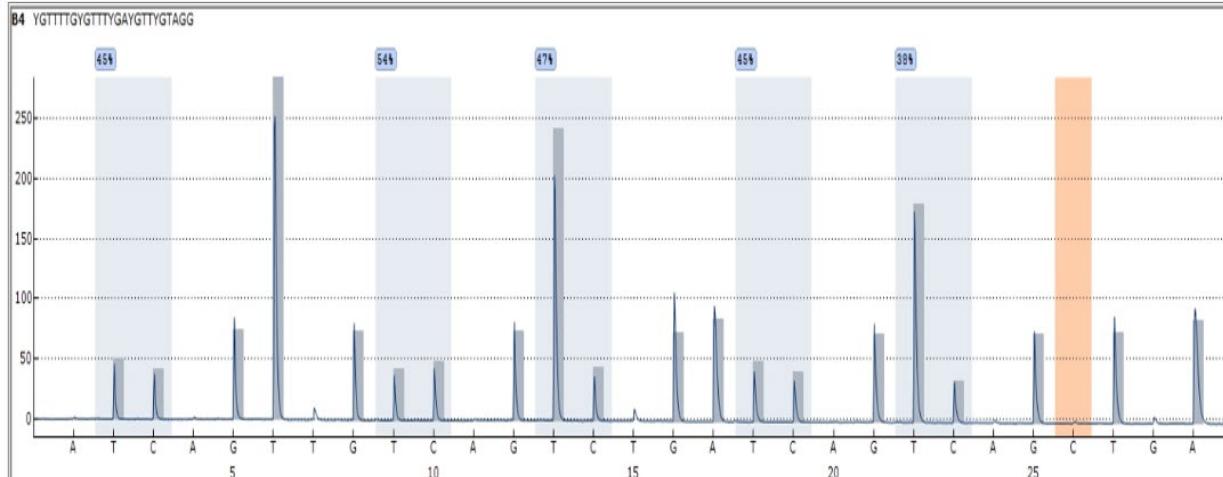
MGMT Methylation Analysis

CpA site towards the end of the sequence is used as an intrinsic internal control.

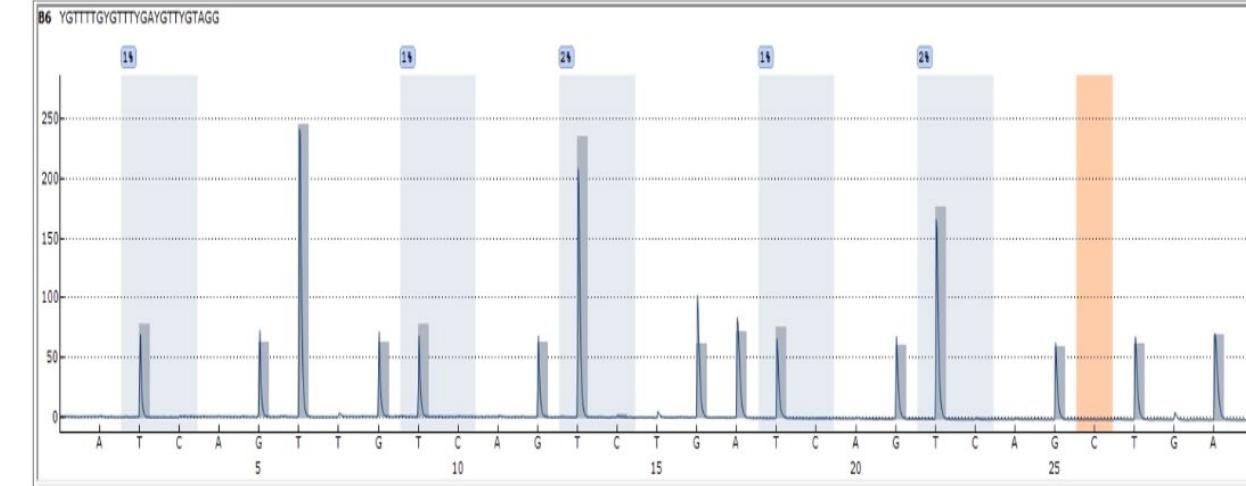
- CpAs are not sites for methylation and should be completely converted to uracil during bisulfite conversion and detected as thymine during sequencing.
- Unconverted cytosine at this site indicates failure of bisulfite conversion.



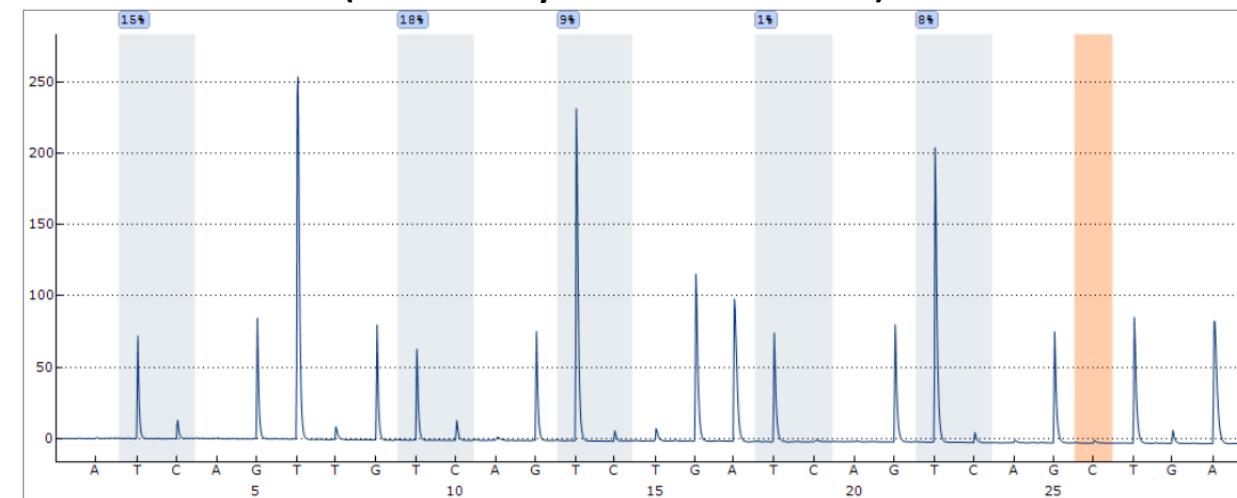
MGMT Positive Hypermethylated (Mean Methylation \geq 12%)



MGMT Negative Not-Hypermethylated (Mean Methylation ≤ 7%)



MGMT Indeterminate (Mean Methylation >7% and <12%)



Next Generation Sequencing

- Reversible Dye Terminator Sequencing

Next generation sequencing (NGS) is a catch-all term that refers to any of a number of methods designed to sequence large numbers of templates carrying millions of bases simultaneously.

- Also called **massive parallel sequencing (MPS)**.

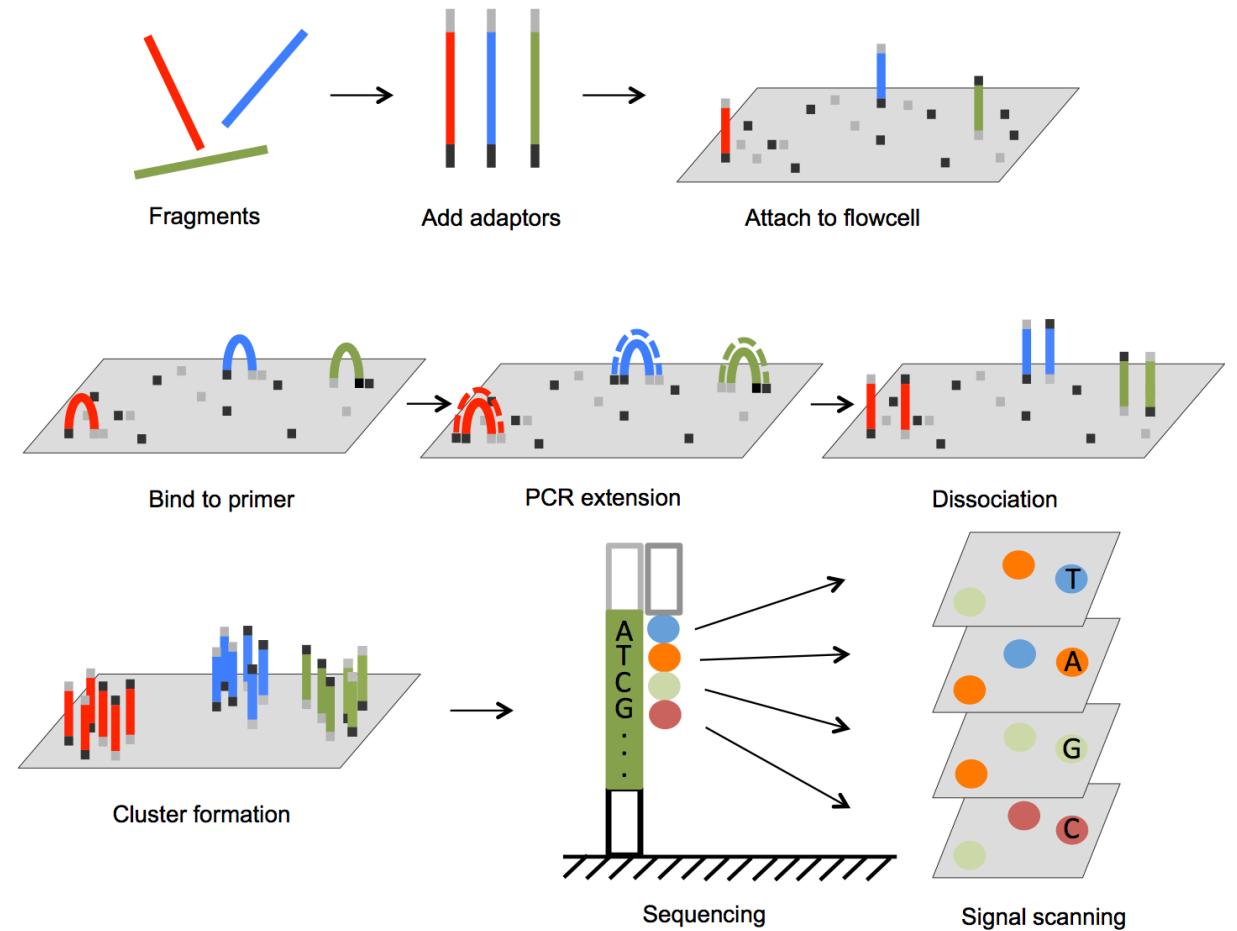
At CCF, we perform Illumina's™ **reversible dye terminator sequencing** method.



Next Generation Sequencing

The major steps of **reversible dye terminator sequencing** (Illumina™):

- Library preparation
- Cluster generation by bridge PCR
- Sequencing
- Data Analysis via bioinformatics pipeline

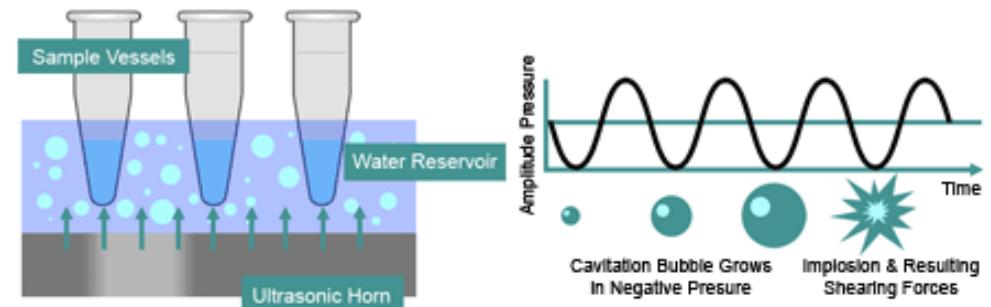


Library Prep: Fragmentation

Genomic DNA/cDNA is first fragmented to produce pieces 100-1,000bp in length.

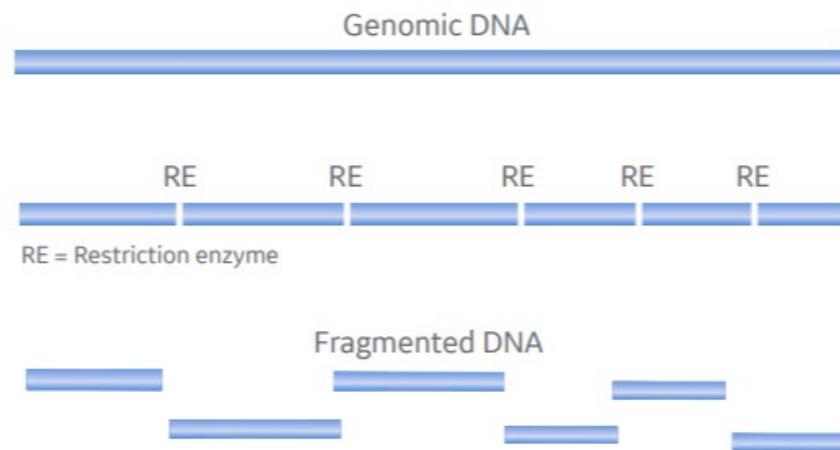
- **Mechanical shearing**

- **Sonication** using acoustic waves (top image)
- **Nebulization** using compressed gas to force NAs through a small hole



- **Enzymatic shearing**

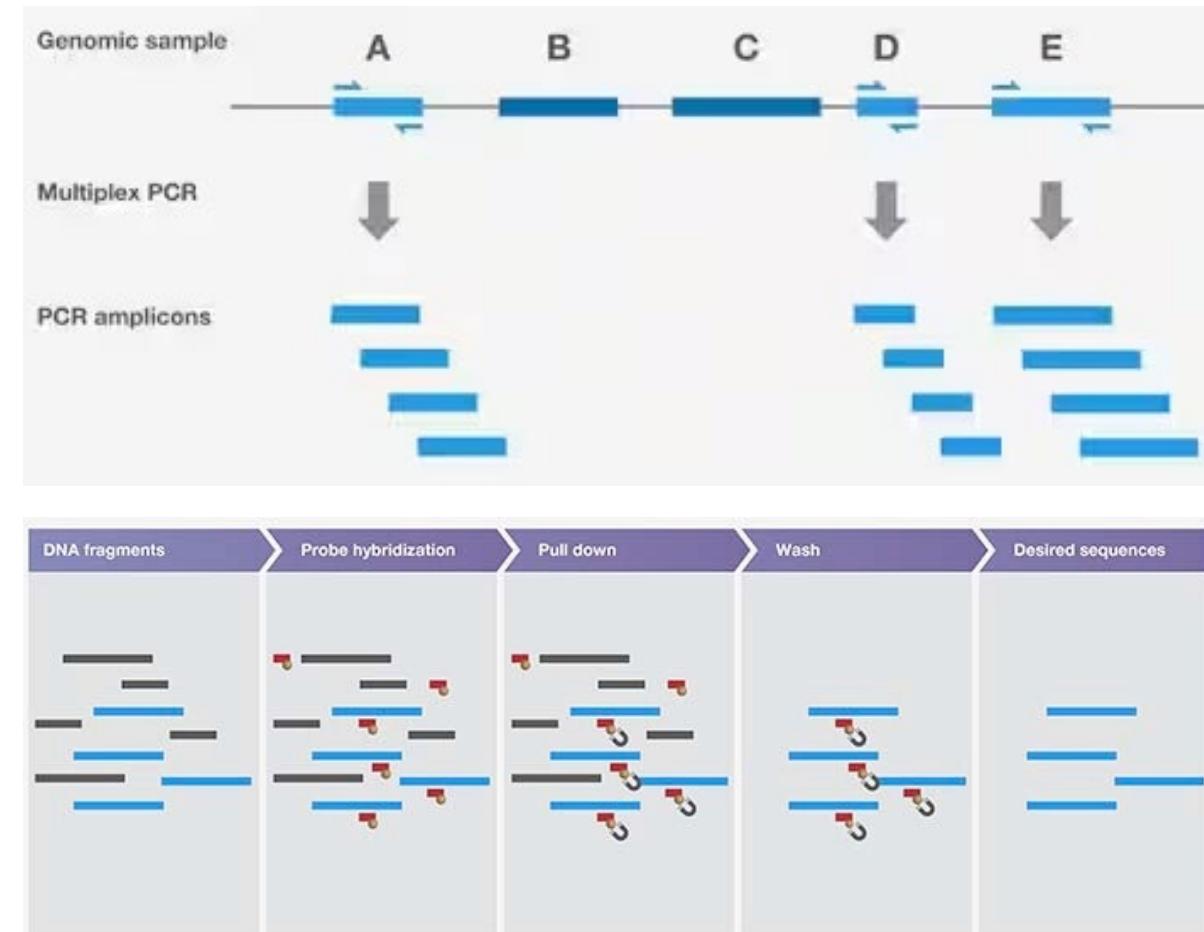
- Restriction endonucleases (bottom image) or nicking enzymes



Library Prep: Target Enrichment

For targeted panels, fragmented DNA can be further refined using several methods.

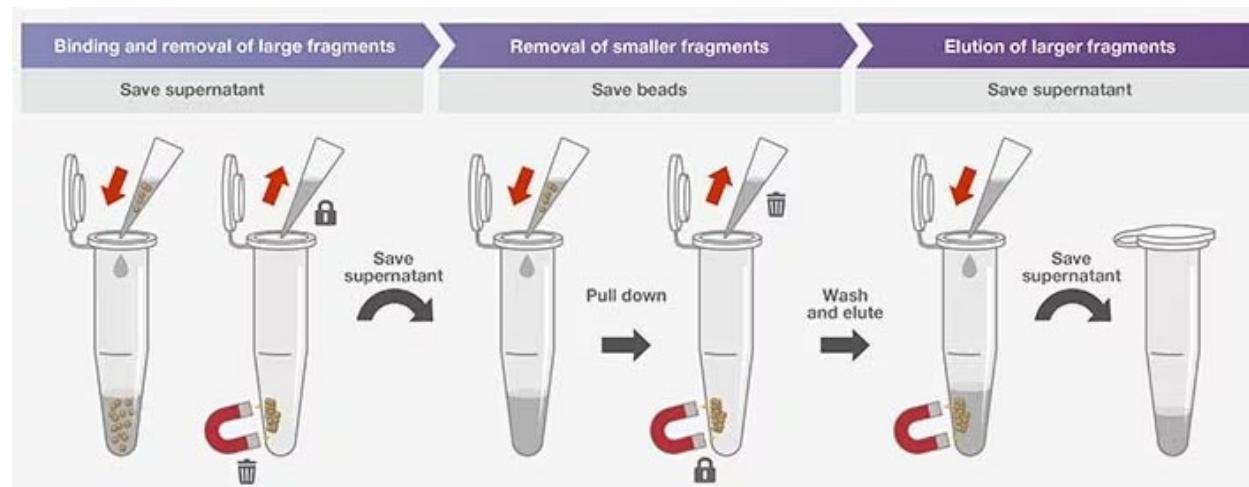
- **Primer-driven amplification** of sequence targets using multiplexed PCR (top image).
- **Hybrid capture** using oligonucleotide probes coupled with magnetic beads (bottom image), or biotinylated oligonucleotide probes captured with streptavidin beads.



Library Prep: Size Selection

Magnetic beads can also be used to preferentially select larger fragments of DNA for amplification/sequencing.

- Buffer concentration (stringency!) encourages binding and retention of hmwDNA to beads.
- Magnet retains beads (with captured hmwDNA) while supernatant (containing lmwDNA) is removed.
- Elution buffer is then added to release hmwDNA from beads and transfer to a clean tube.



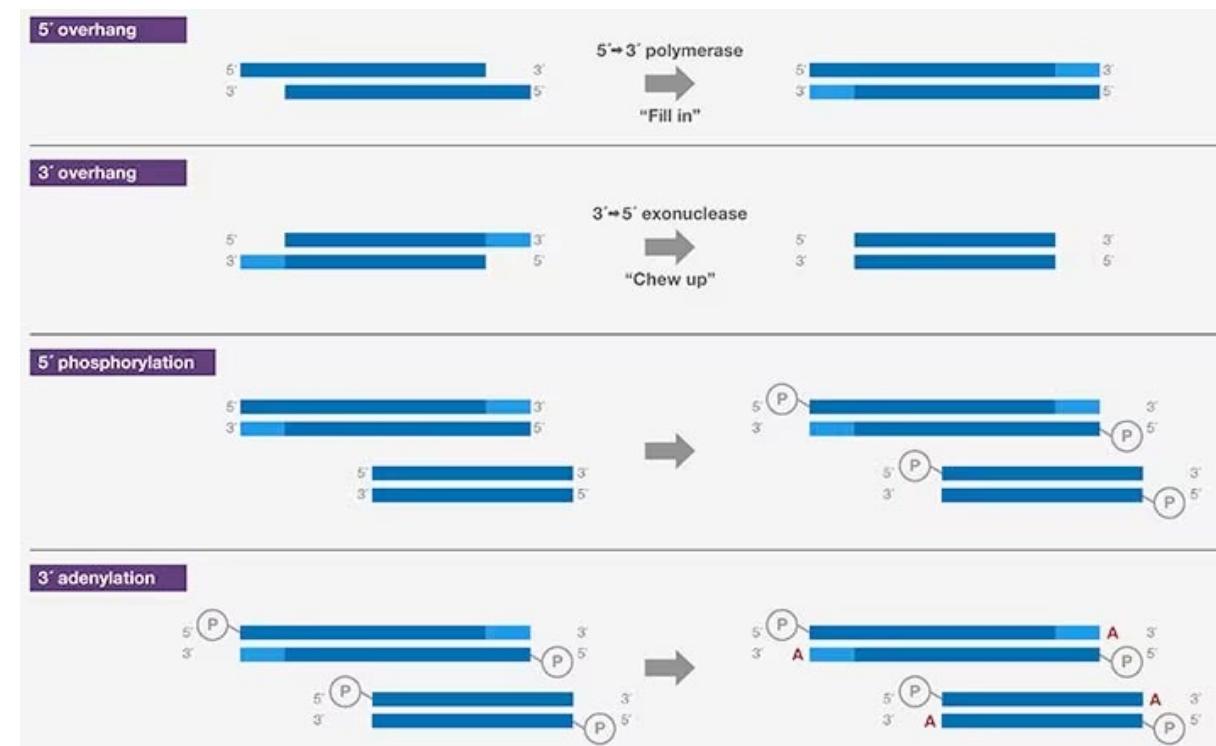
Library Prep: End Repair

Fragmentation leaves “sticky” ends on DNA.

In the **end repair** step of library prep, polymerase enzyme uses both its 5'-3' polymerase activity and 3'-5' exonuclease activity to blunt the ends of DNA.

Additional enzymes **phosphorylate the 5' ends** and **adenylate the 3' ends (A-tailing)**

Phosphorylation/adenylation aids ligation in the next step.



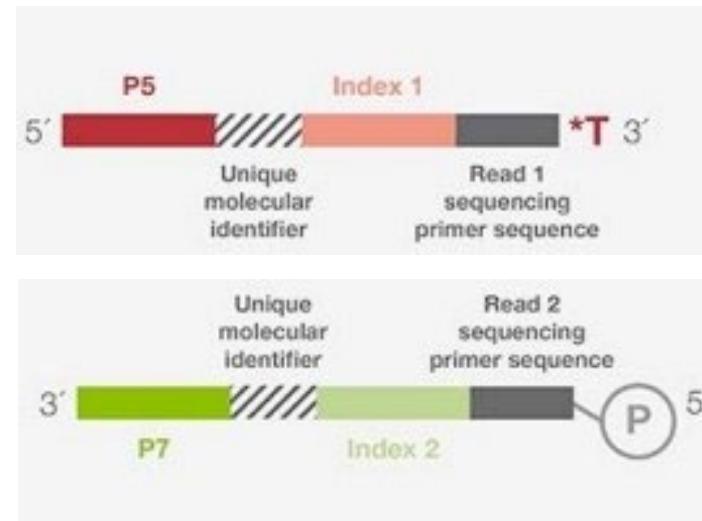
Library Prep: Adapter Ligation

Following end repair, DNA samples must be adapted and indexed.

- **Adapters** are pairs of short oligonucleotide sequences complementary to the oligos embedded on the flowcell as well as the sequencing primers.
- **Indexes** are unique molecular barcodes that are used to differentiate patients in the final pooled library.

Ligase is used to ligate adapters/indexes to both ends of patient DNA.

- Adapters and indexes may be separate molecules that require their own separate ligation step, or combined into a single molecule (images to the right) for a single ligation step.

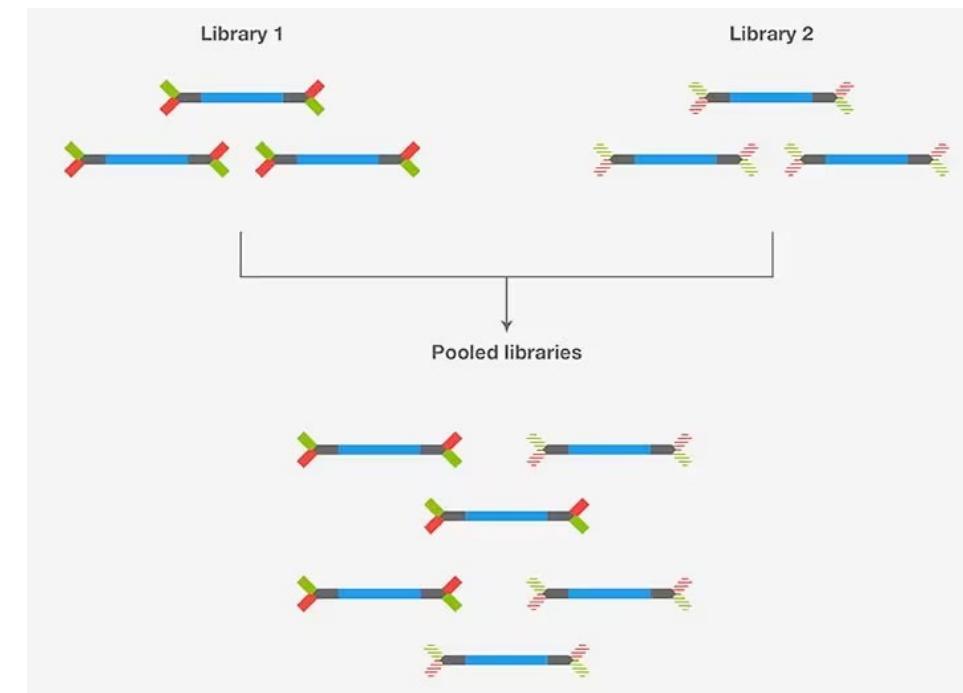


Library Prep: Library Pooling

By providing each patient with a unique index during adapter ligation, patient samples may be combined into a final **pooled library**.

Prior to pooling, patient samples are quantified and diluted as necessary to prevent over-/underclustering in the next step.

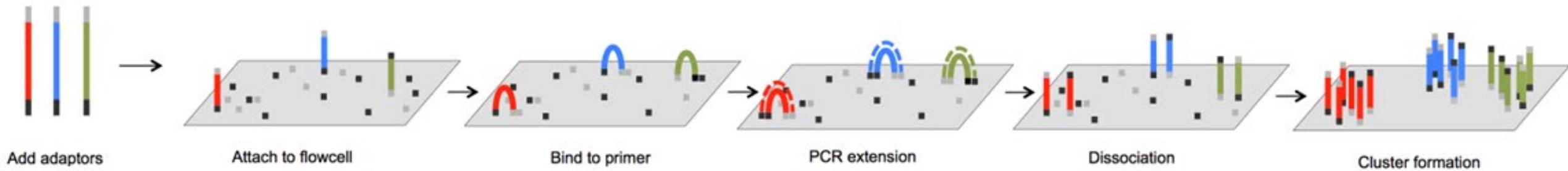
- Fluorometry (Qubit)
- Microfluidics (Bioanalyzer)
- Electrophoresis (TapeStation)
- qPCR



Cluster Generation

Pooled library is then loaded into a **flowcell** for clonal amplification by **bridge PCR**.

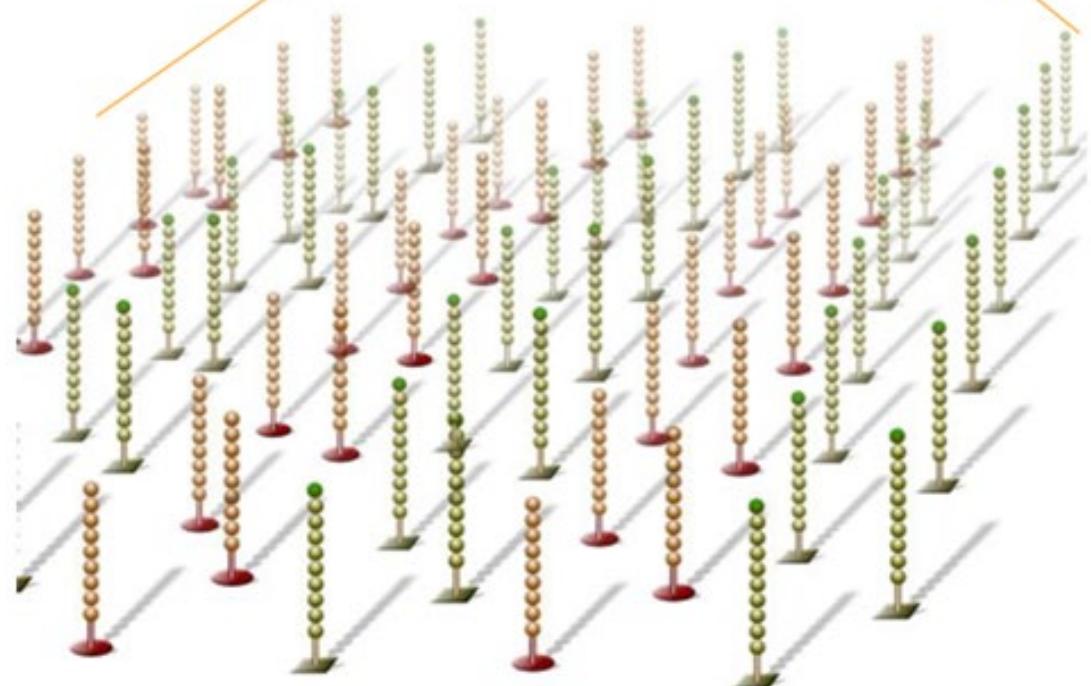
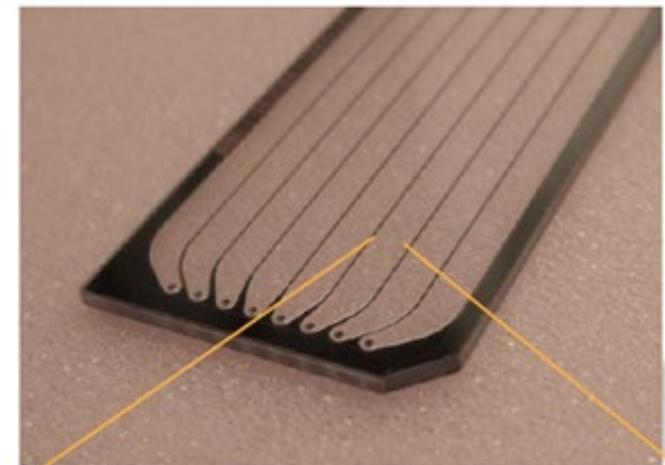
Bridge PCR forms millions of localized clusters, or **polonies**, containing thousands of copies of patient DNA fragments.



Cluster Generation

The flowcell is a thick glass slide with multiple lanes.

Each lane is coated with a lawn of embedded oligonucleotides that are complementary to the adapters added in library preparation.



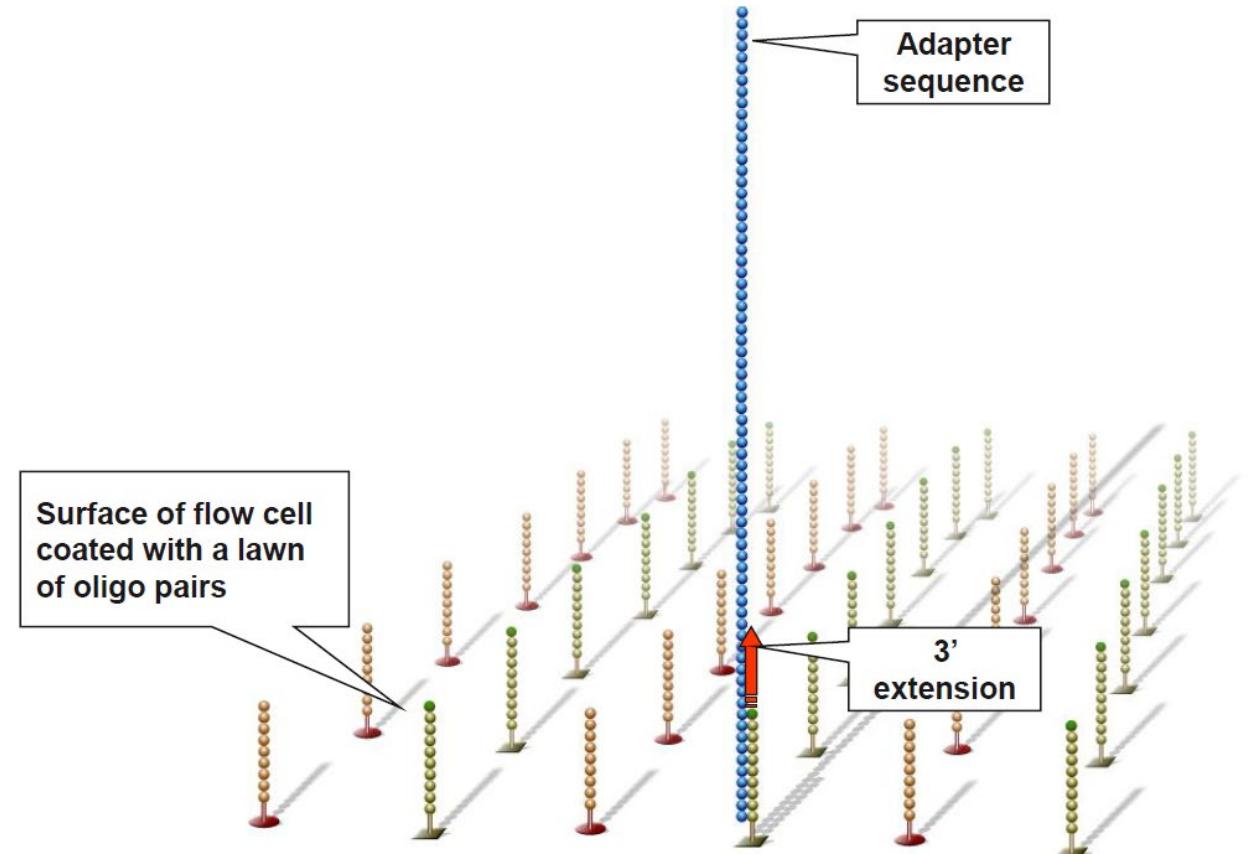
illumina®

Cluster Generation: Step 1

In the first step of cluster generation, the adapter sequence (ligated to patient DNA during library prep) hybridizes to its complementary embedded oligonucleotide.

Polymerase enzyme then extends the embedded oligo by reading the hybridized patient strand.

Newly synthesized strand will be complementary to the patient template DNA, including the adapter and barcode sequences added during library prep.

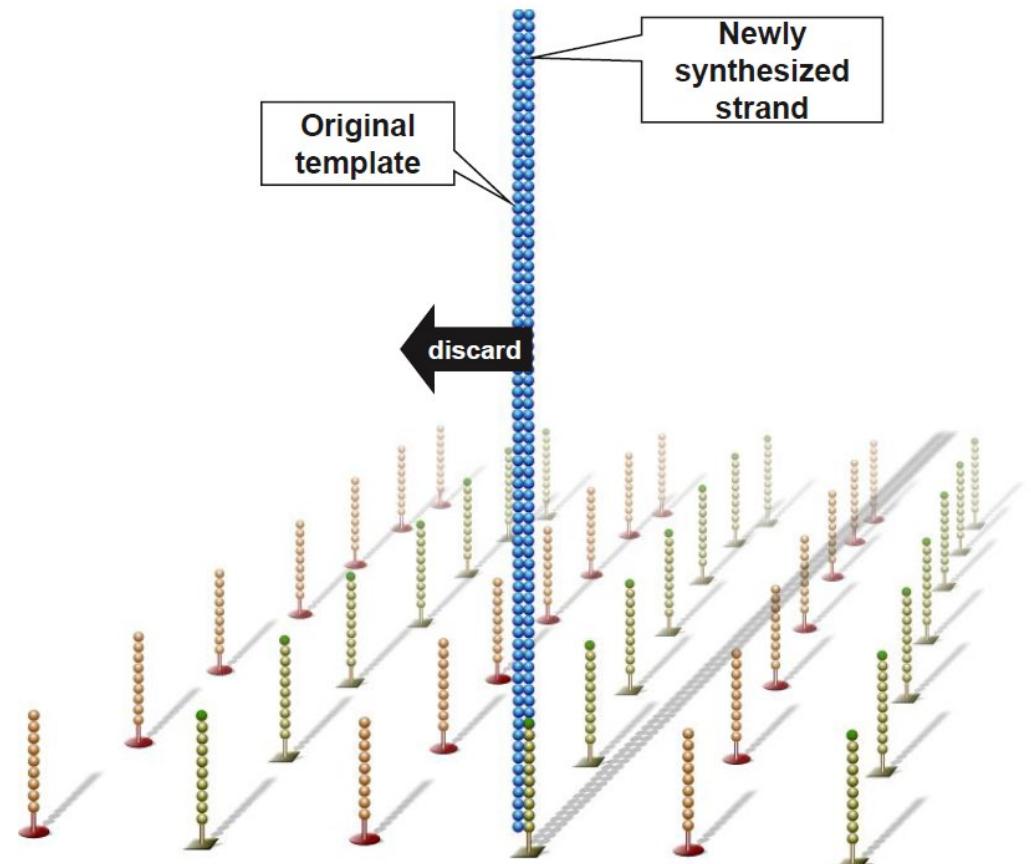


illumina®

Cluster Generation: Step 2

In second step of cluster generation, the flowcell is flooded with a denaturing wash solution.

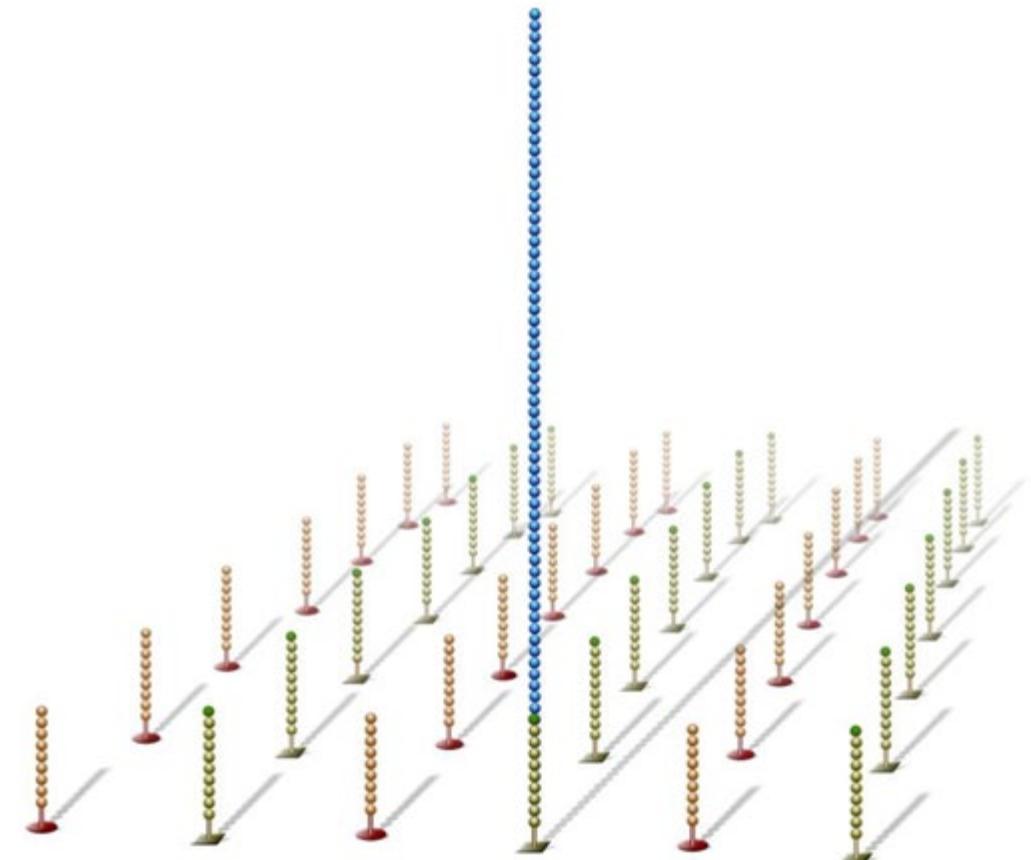
This causes the DNA duplex to denature and the original patient template DNA to be washed away.



illumina®

Cluster Generation: Step 2

The newly synthesized strand complementary to the patient template remains, since it is still embedded on the surface of the flowcell.

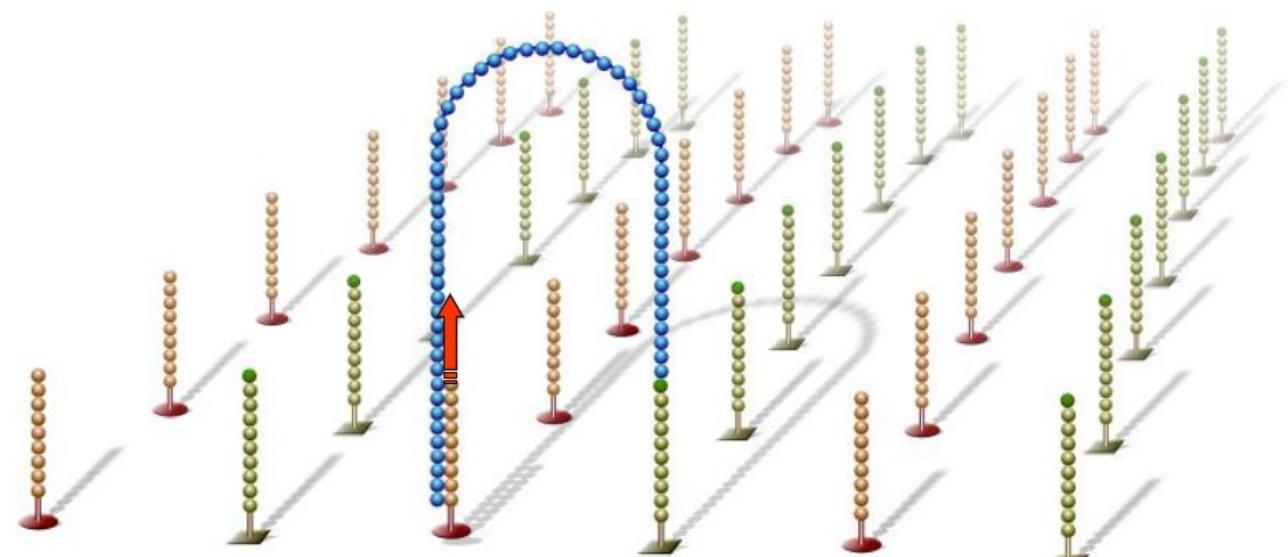


illumina®

Cluster Generation: Step 3

In the third step of cluster generation, annealing conditions are reestablished and the new strand bends over to hybridize to a second embedded oligonucleotide.

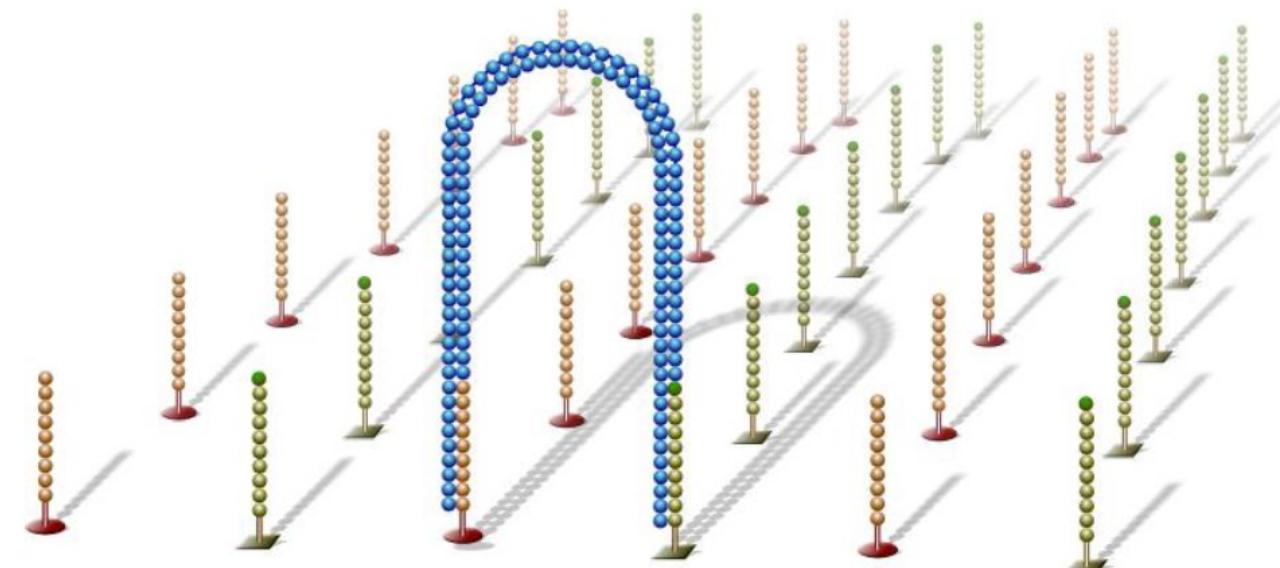
Polymerase enzyme extends the second embedded oligo by reading the hybridized strand as a template.



illumina®

Cluster Generation: Step 3

This creates a characteristic double-stranded DNA bridge, from which bridge PCR gets its name.

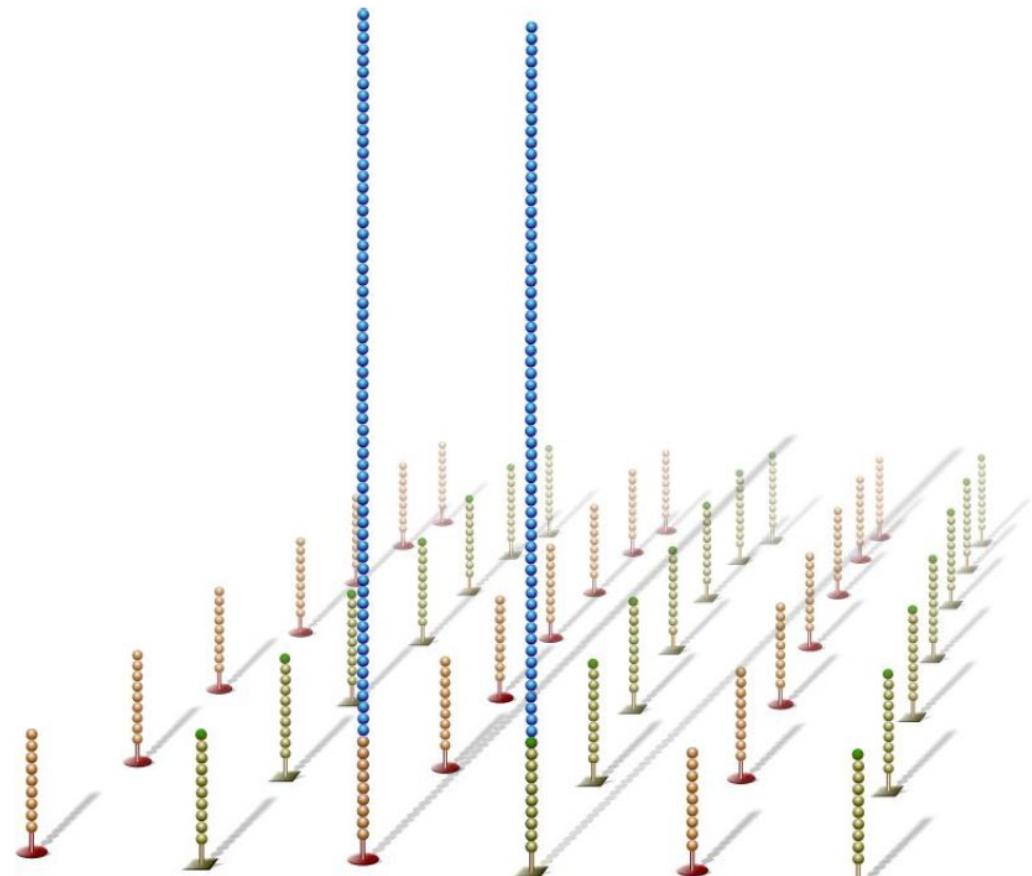


illumina®

Cluster Generation: Step 4

In the fourth step, the flowcell is again flooded with a denaturing wash solution.

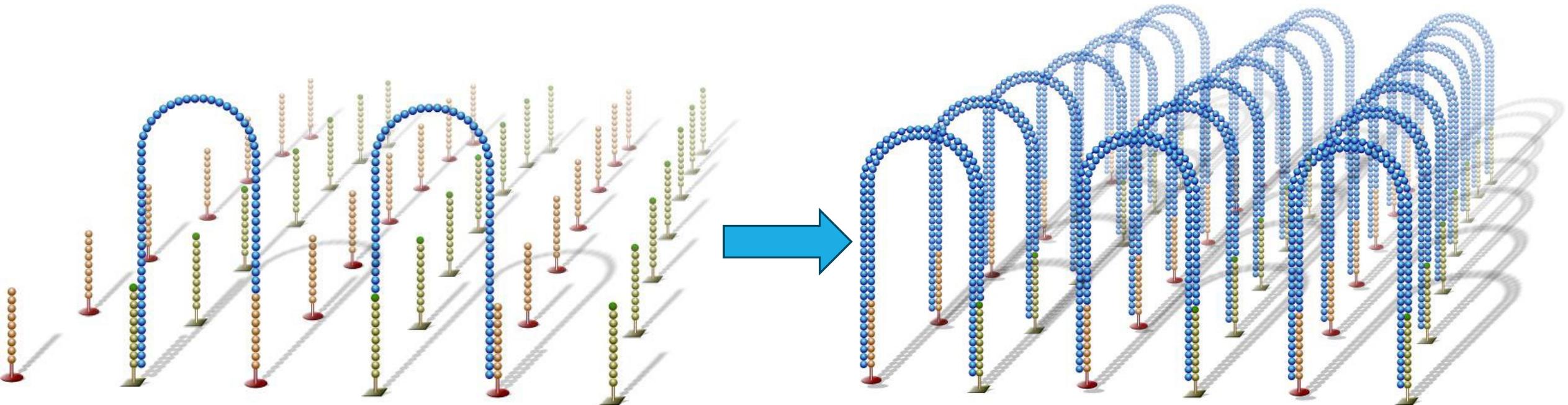
This causes the double-stranded bridge to denature, leaving behind copies of both forward and reverse strands of patient DNA embedded on the surface of the flowcell.



illumina®

Cluster Generation: Step 5

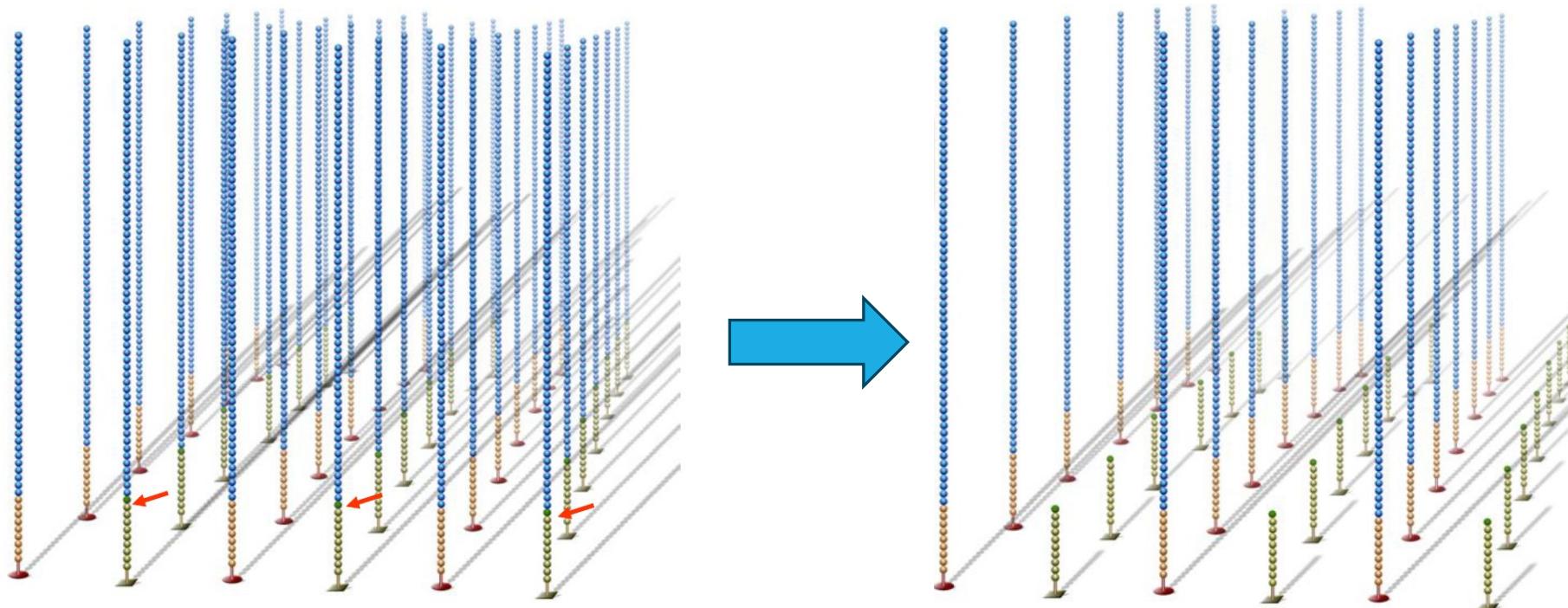
Denaturing/hybridization/extension steps are then repeated to form a localized clusters of forward and reverse strands of patient DNA.



Sequencing: Read 1

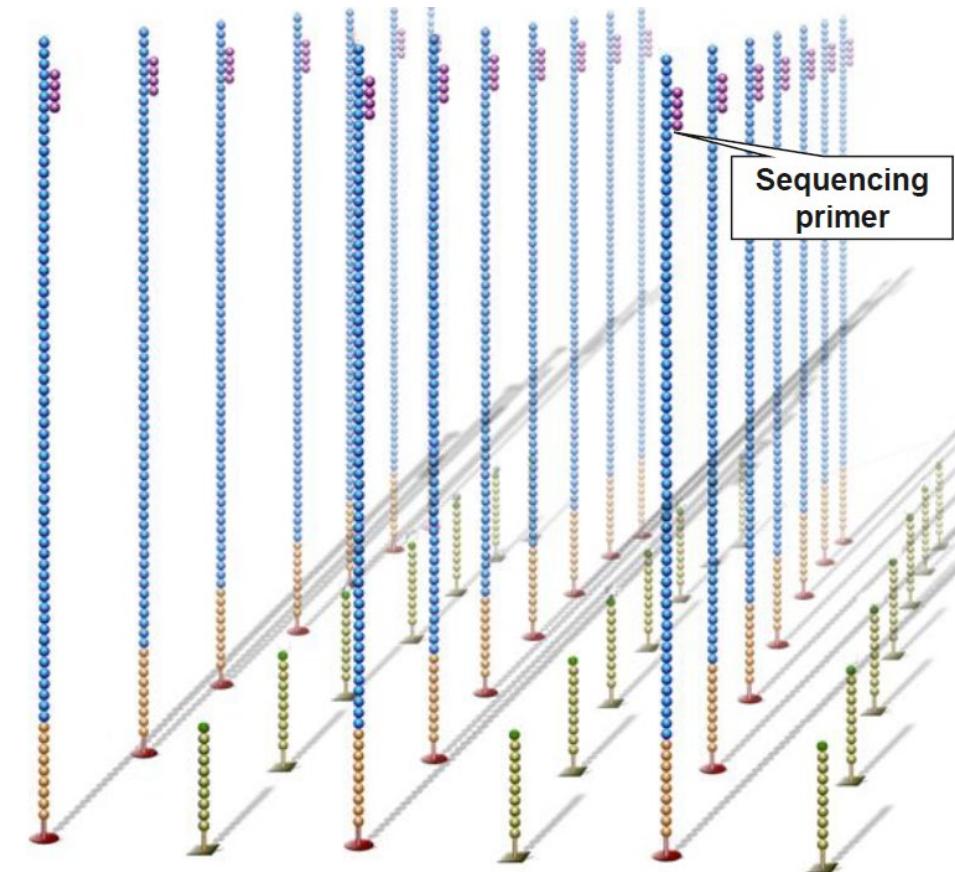
Following cluster generation, double-stranded DNA bridges are denatured and linearized.

The reverse strands are then cleaved from their embedded oligonucleotides, leaving behind a cluster of only forward strands.



Sequencing: Read 1

The Read 1 sequencing primer is then hybridized to the forward strands.



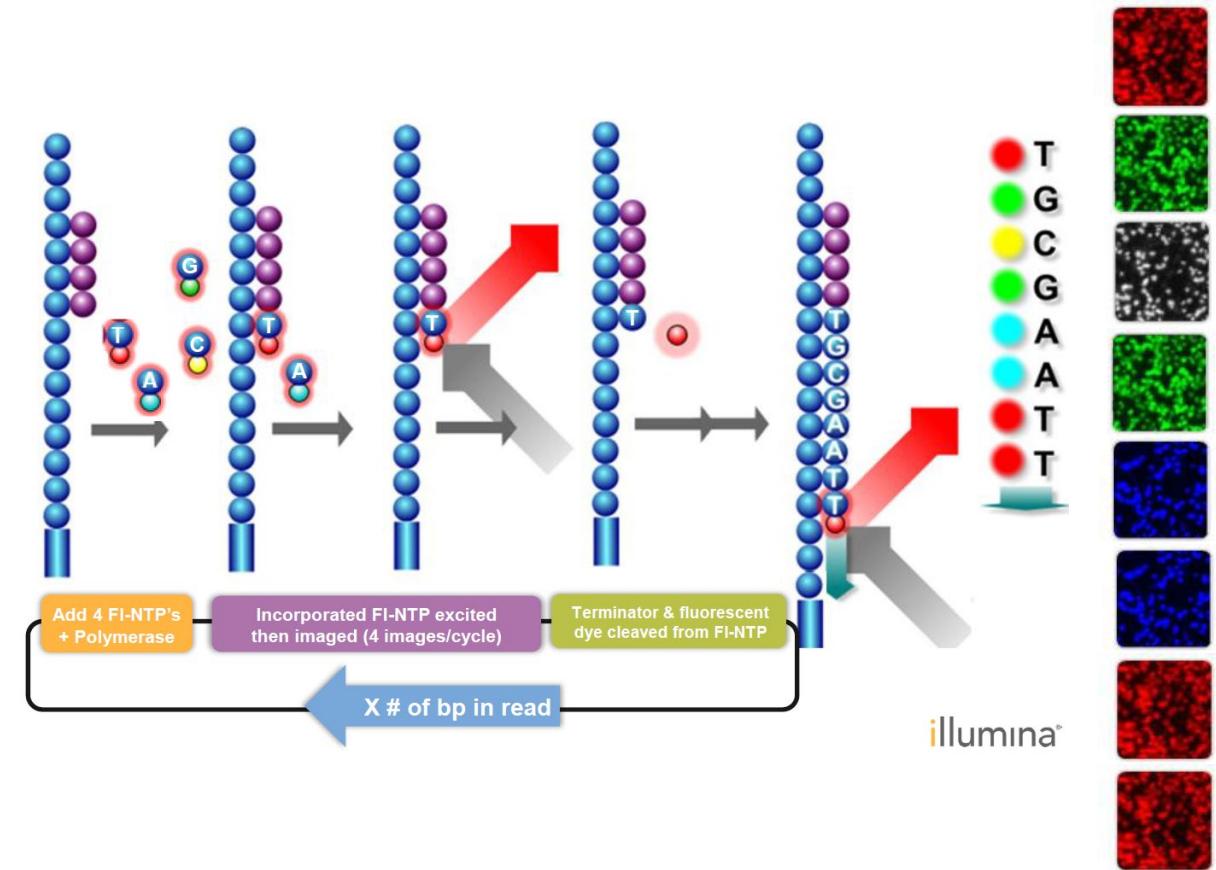
illumina®

Sequencing: Read 1

Reversible dye terminator sequencing:

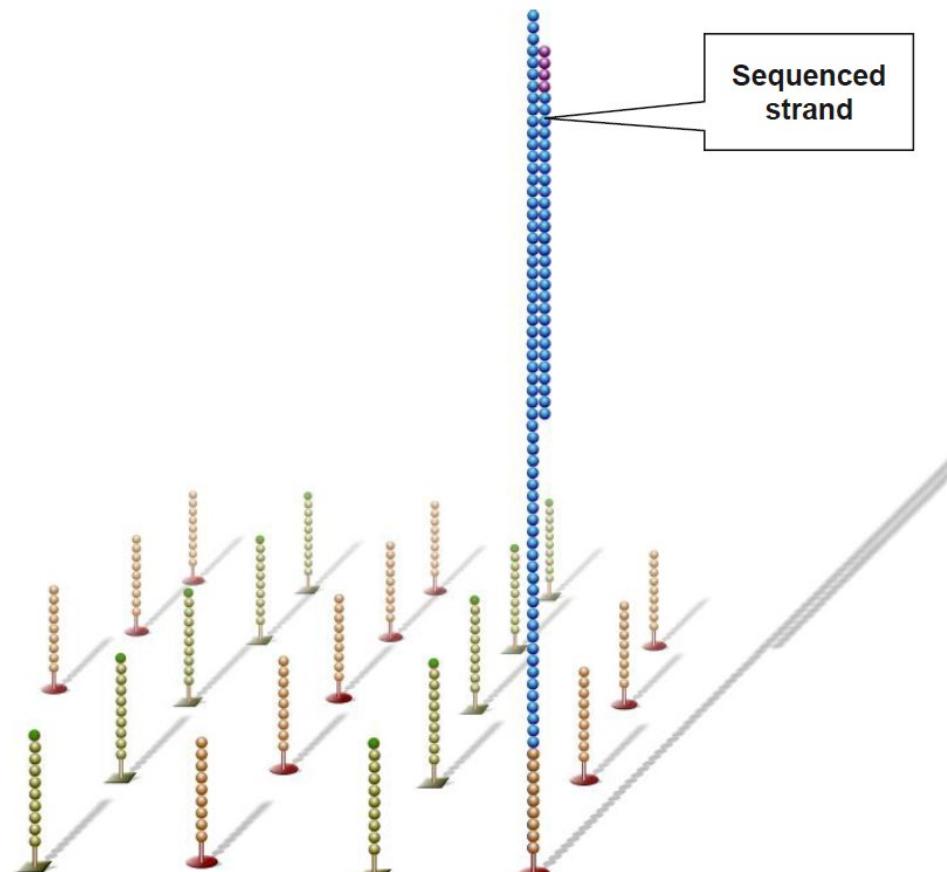
- The flowcell is flooded with nucleotides that are labelled with unique dye terminators.
- Polymerase incorporates the complementary nucleotide-dye terminator to the primer, stopping synthesis after a single base addition.
- UV light excites fluorescent dye, which is detected by a camera and recorded.
 - Wavelength of emission differentiates nucleotide added.
- Dye terminator is cleaved from the incorporated nucleotide, preparing the strand for the next addition.

Process is then repeated for the entire length of the forward strand.



Sequencing: Read 2

Once the sequence from Read 1 has been gathered, the sequenced strand is denatured and washed away.

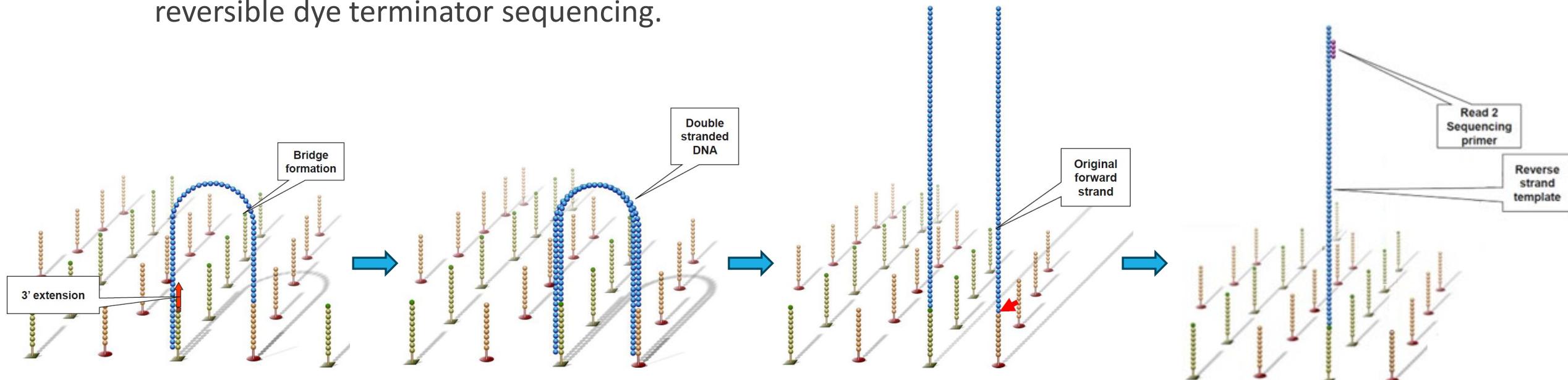


Sequencing: Read 2

Bridge PCR is cycled one more time to re-generate the reverse strand.

Double-stranded DNA bridge is linearized, and forward strands are cleaved and washed away.

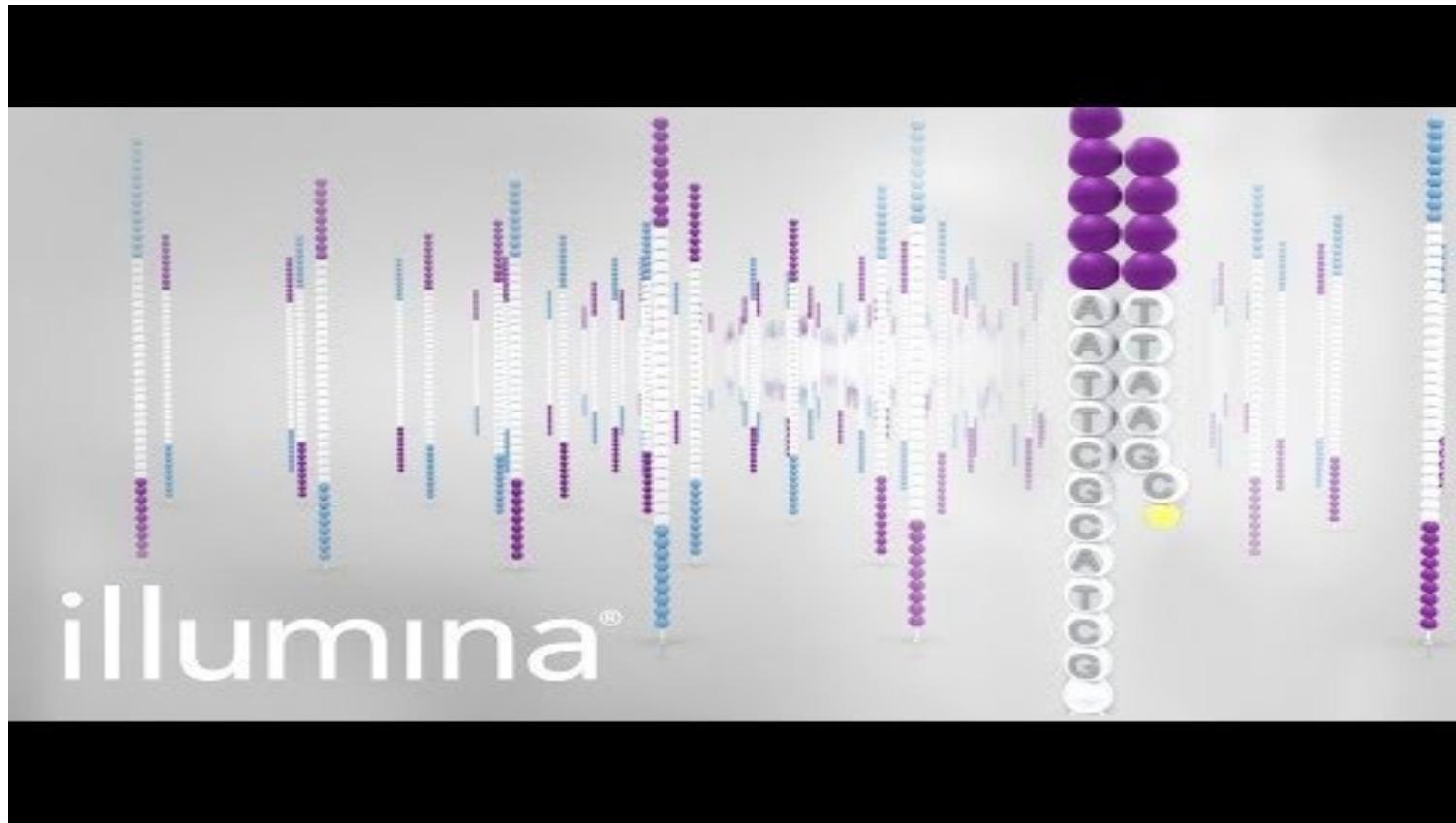
Read 2 sequencing primer binds to the remaining reverse strands for the second round of reversible dye terminator sequencing.



Illumina™ NGS Video

YouTube Link:

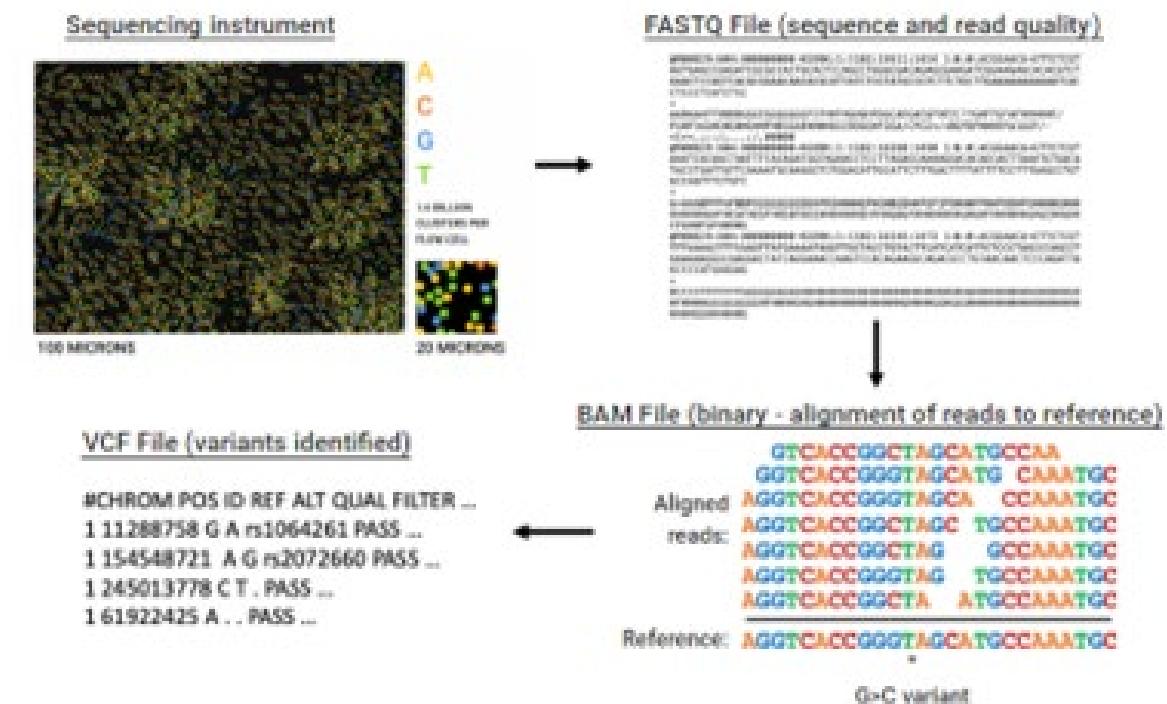
<https://www.youtube.com/watch?v=fCd6B5HRAZ8>



Bioinformatics Pipeline

There are three main file types generated and stored for NGS bioinformatics:

- Raw sequencing reads and quality data gathered from the instrument is exported as a **FASTQ File**.
- Sequencing reads from the FASTQ file are aligned to the reference genome using secondary software and stored as a **BAM File**.
- Tertiary software analysis of the BAM file generates a final **VCF File**, containing variants identified and any additional annotations regarding the sequence data (clinical significance, read quality, etc.).

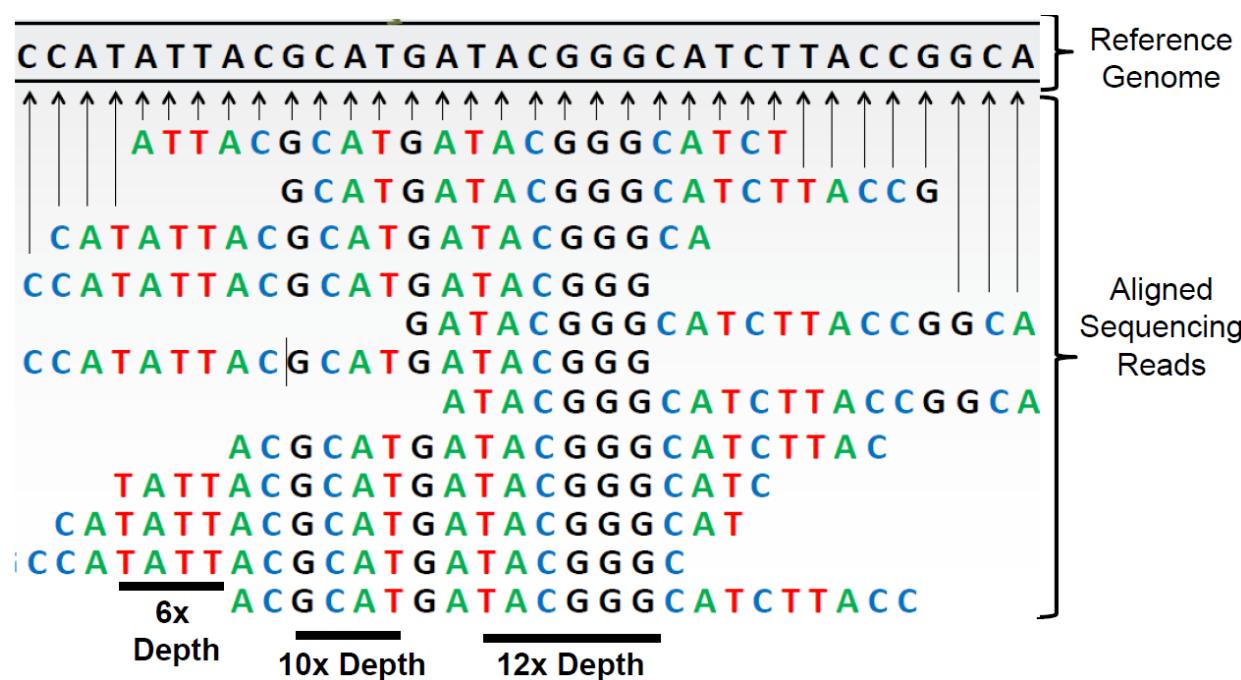


Sequence Alignment

Individual reads gathered during sequencing are sorted to the appropriate patient based on their unique index sequence.

Secondary analysis software compares each read to a reference and stacks them on top of each other in a process called **alignment**.

Alignment enables full gene/genome sequences to be assembled from overlapping, shorter individual reads.



Quality Metrics

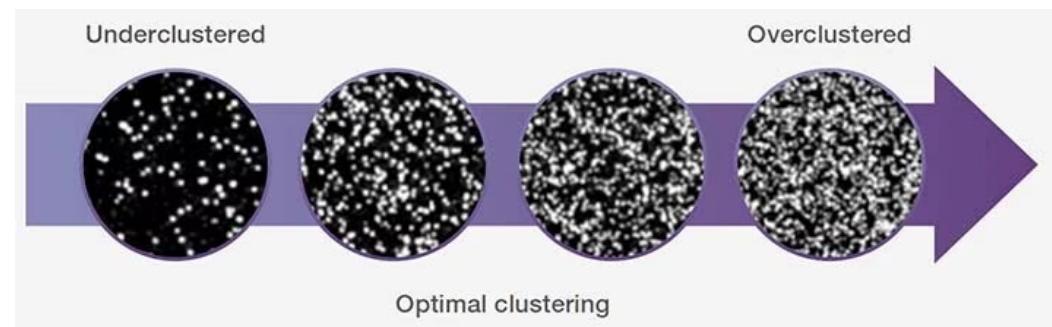
There are several important quality metrics associated with NGS sequencing data

Q-score is quality score similar to Phred scores used in Sanger

- Reported as %>Q20 and %>Q30, where Q20 is a 1-in-100 error rate and Q30 is a 1-in-1000 error rate.

Cluster density

- Underclustering means fewer reads
- Overclustering lowers quality score



NGS Quality Metrics

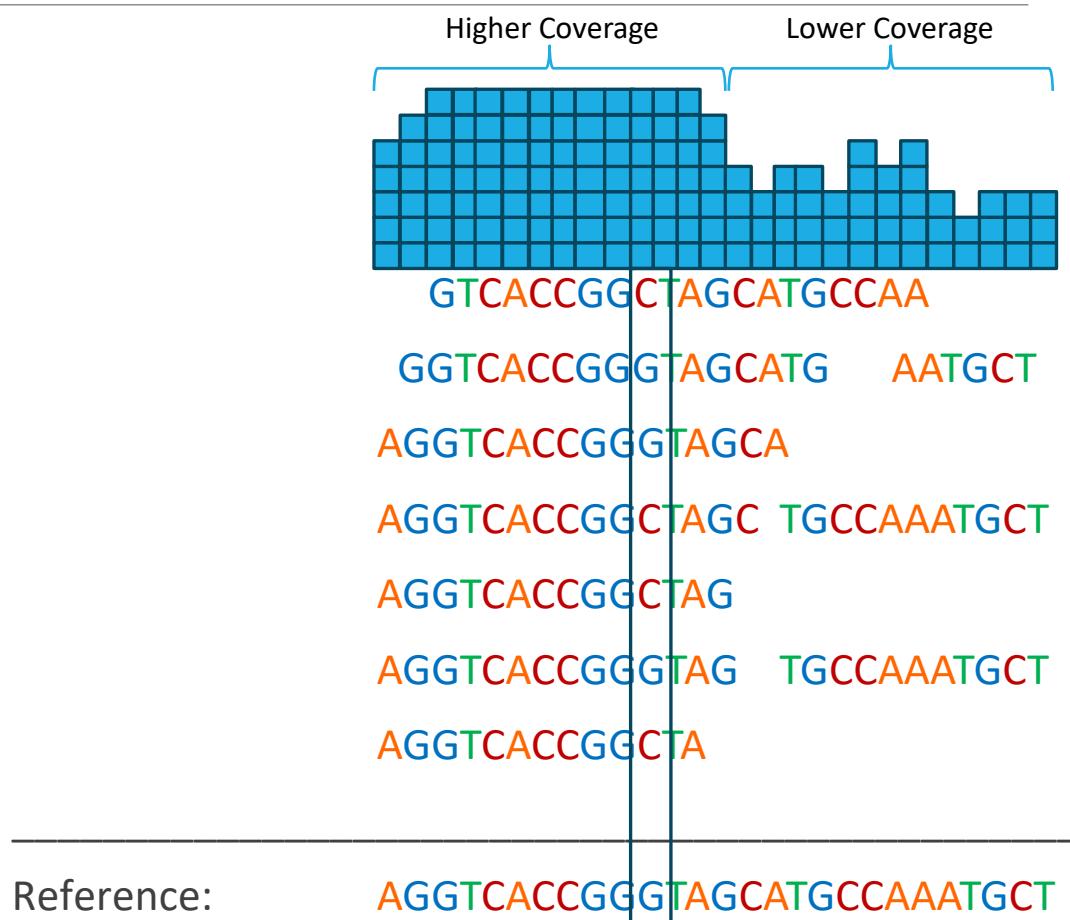
Read length refers to the number of base pairs sequenced from a DNA fragment.

Read depth is the total number of reads which included a specific base from the reference sequence.

Coverage is the number of times a region containing a particular variant has been sequenced from independent fragments.

- Note: in practice “depth” and “coverage” are often used interchangeably or in combination as “depth of coverage” to describe the same idea.

Example to the right shows heterozygous G>A substitution with 7X coverage.



Reference:

AGGT CACC GGG TAG C ATGCCAAATGCT

Next Generation Sequencing

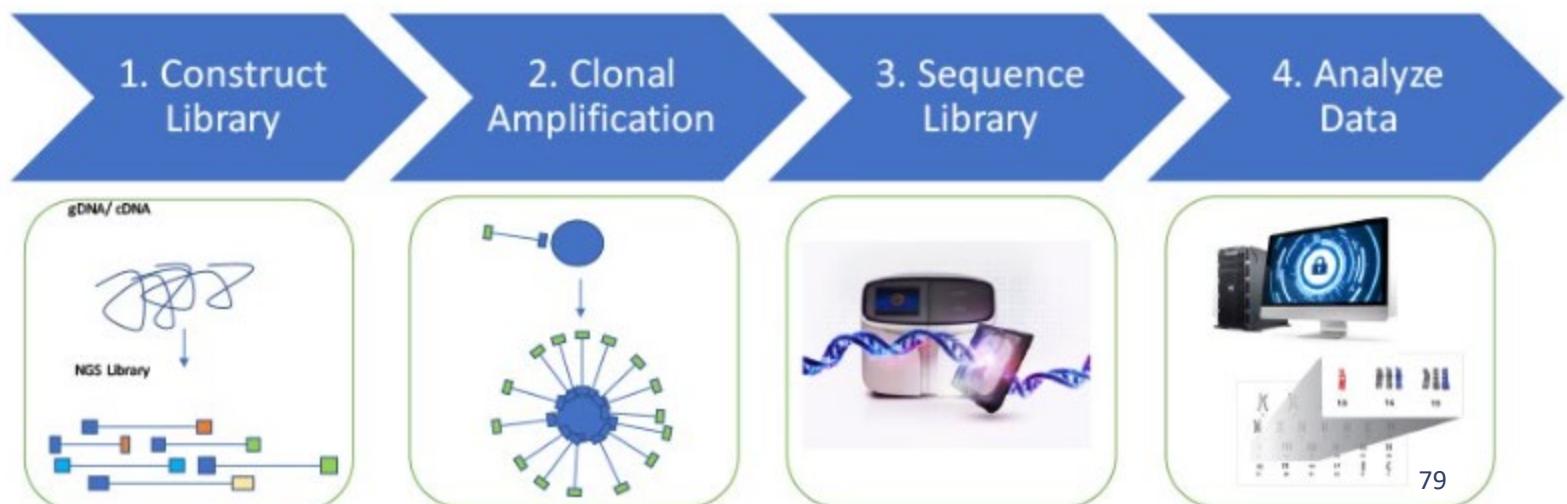
- Ion-Conductance Sequencing

Illumina™ and its bridge amplification/reversible dye terminator NGS method isn't the only one used in medical laboratories.

Ion-Torrent™ instruments use two other techniques to accomplish the same goal of massively parallel sequencing by synthesis.

- Emulsion PCR
- Ion-Conductance Sequencing

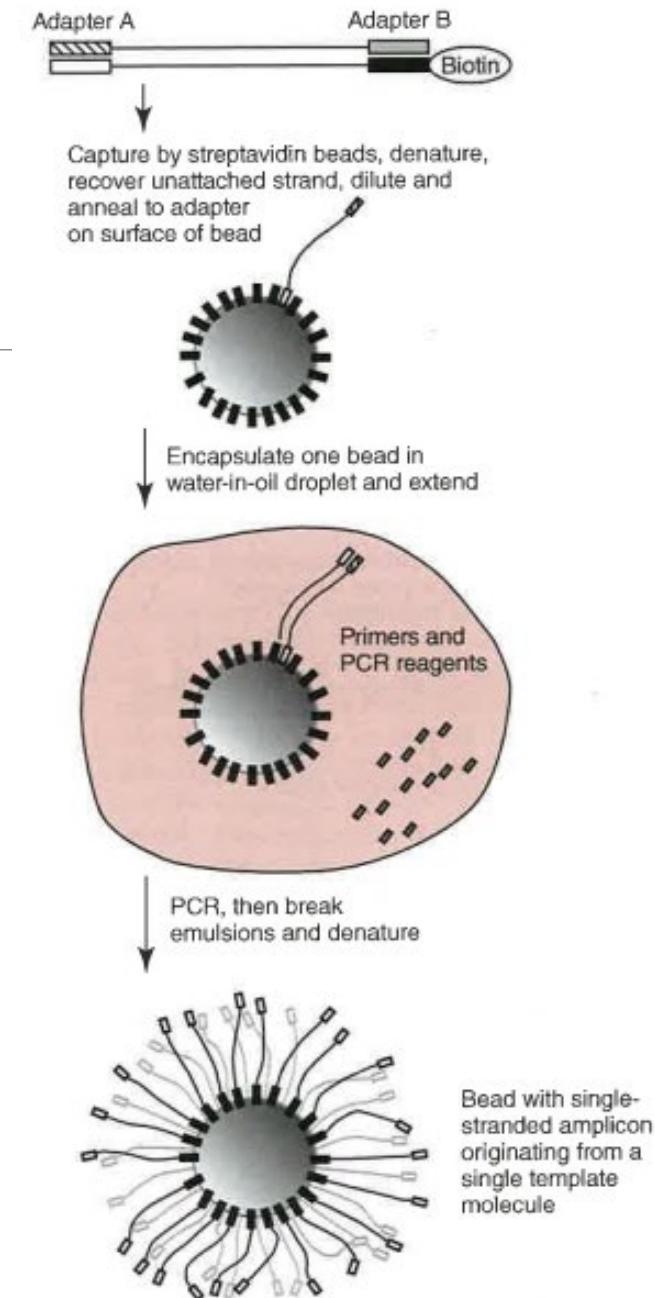
The workflow has very similar steps to the method previously discussed:



Emulsion PCR

Ion-Torrent™ systems clonally amplify their indexed libraries by solid-phase **emulsion PCR**.

- Biotinylated adapters are ligated to DNA fragments.
- Fragments are captured on streptavidin beads by the biotinylated adapter.
- Beads and PCR reaction components (dNTPs, polymerase, etc) are emulsified with oil into thousands of individual droplets.
- Oil droplets undergo PCR to amplify captured DNA fragments.
- Emulsion is broken, beads are denatured and applied to a solid surface (chip) for sequencing.



Ion-Conductance Sequencing

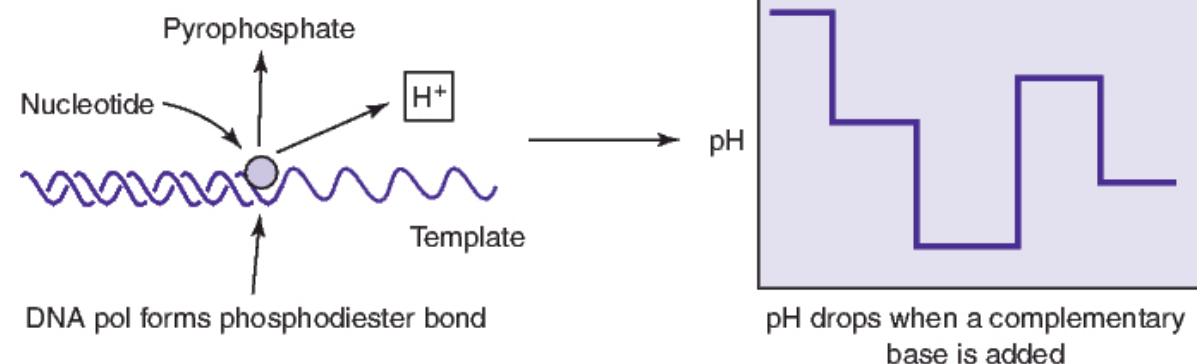
Ion-conductance sequencing, also called **semiconductor sequencing**, is similar in principle to pyrosequencing.

When a dNTP is incorporated into a growing strand, two molecules are released.

- PPi, which is detected in pyrosequencing.
- And H⁺, which is detected in ion-conductance sequencing.

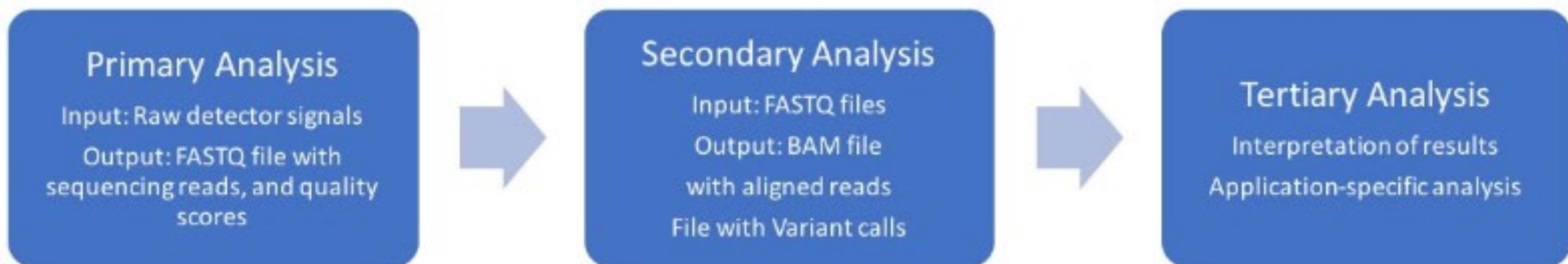
The release of the H⁺ ion causes a drop in pH, which is detected by a pH meter.

Sequential addition/degradation of dNTPs allows for 5'-3' sequence determination.



Data Analysis

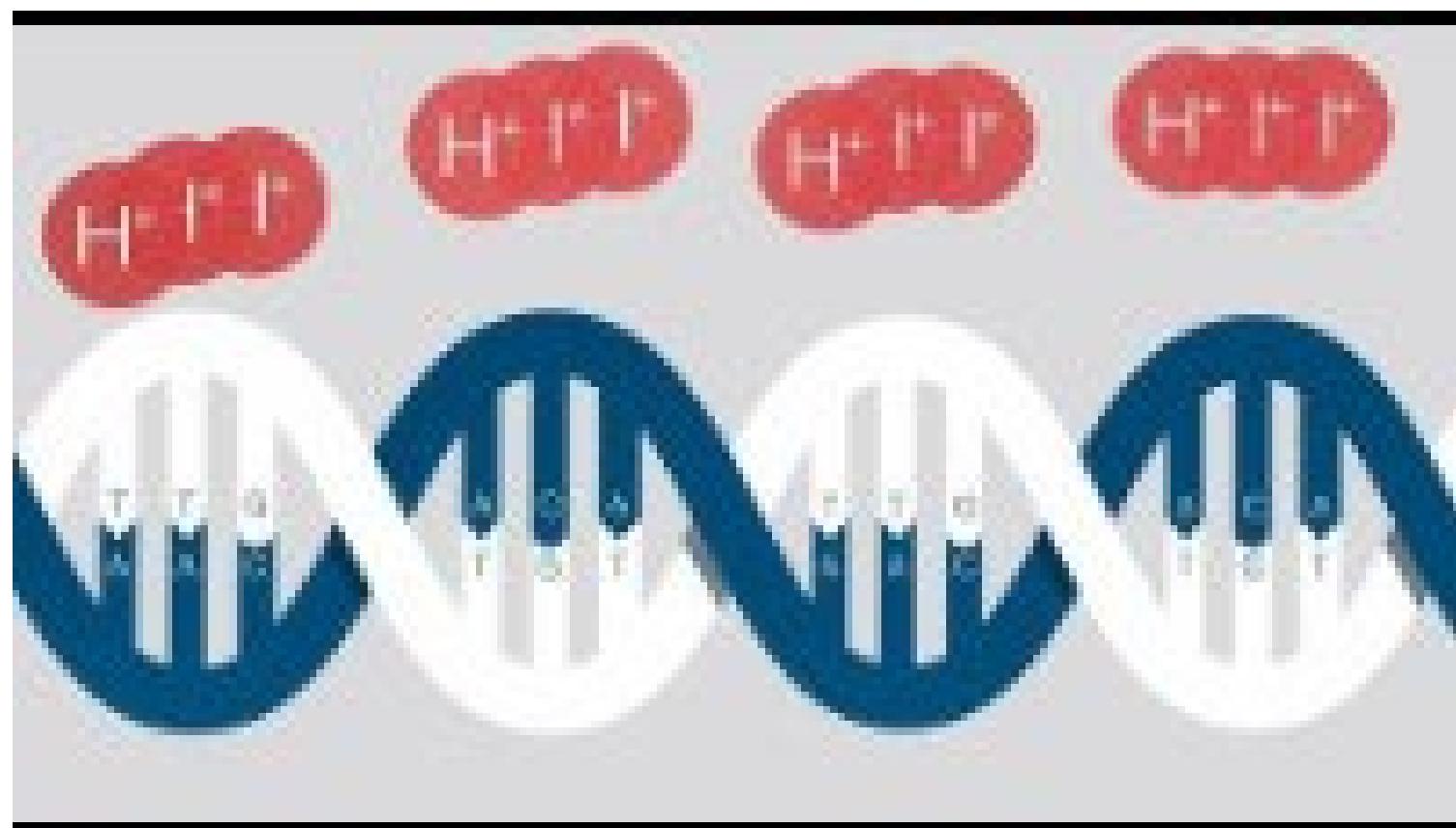
As with Illumina™, Ion-Torrent sequencing data then goes through a similar bioinformatics pipeline to align and interpret sequence reads.



Ion-Torrent™ NGS Video

YouTube Link:

<https://www.youtube.com/watch?v=zBPKj0mMcDg>



Other Sequencing Methods

There are a couple other sequencing methods mentioned in the textbook that aren't defined as "sequencing by synthesis."

- Sequencing by ligation
- Single-molecule sequencing

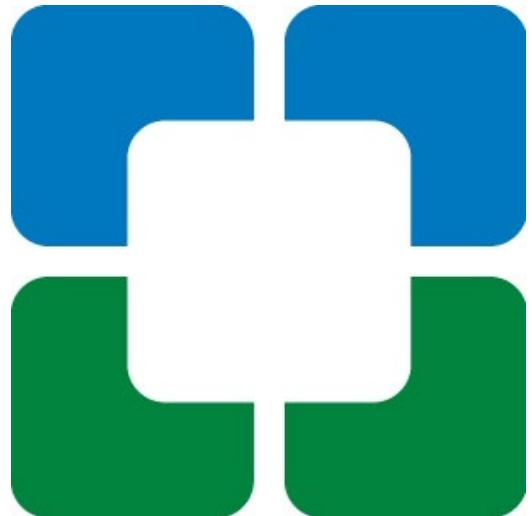
Not commonly used in medical labs, more common in research applications.

- Read about them in the book but focus more on the methods covered in this presentation.

Questions?

EMAIL: POHLJ@CCF.ORG

WORK PHONE: 216 308-0801



This concludes the presentation.

CLEVELAND CLINIC CENTER FOR PATHOLOGY EDUCATION