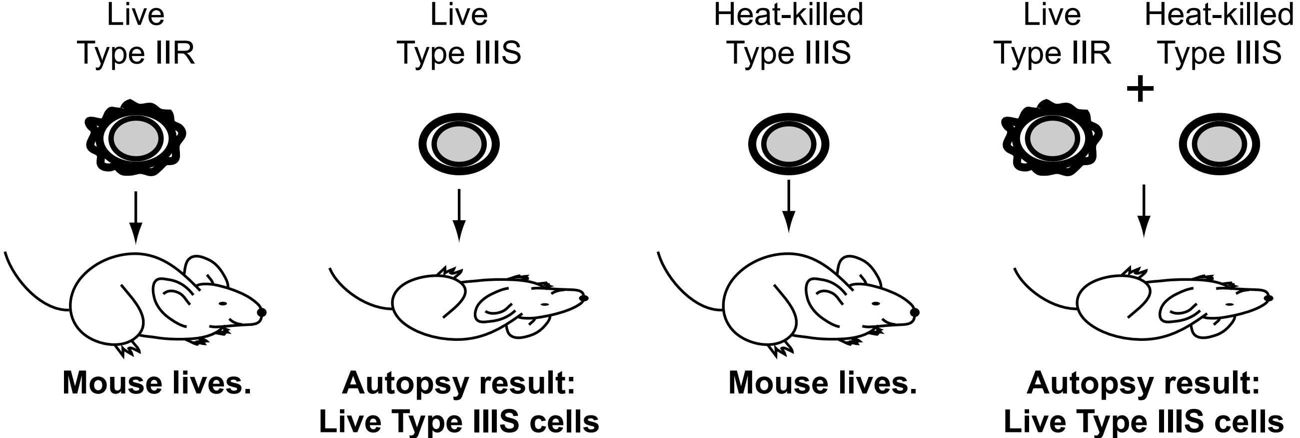
Griffith and the Transforming Principle

A. The Concept

The experiments of Griffith and Avery, MacLeod and McCarty are closely related. Griffith developed the concept of the ***transforming principle***. The prinicple was able to transform a non-pathogenic bacteria into a pathogenic strain. Changing phenotype is one of the characteristics of the hereditary material. Griffith called the factor that changed the phenotype the tranforming principle. Avery, McCarty, and MacLeod performed a series of experiments that demonstrated ***the hereditary materials was DNA***..



Fred Griffith's experiments provided the experimental platform for Avery, McCarty, and MacLeold to prove the DNA was the genetic material. He worked with the pathogenic bacteria Streptococcus pneumoniae that is lethal to mice. But not all types of the bacteria all lethal: type R is non-lethal, whereas type S is lethal. In addition, there are type II an III strains of the bacteria. Each of these can be either R or S. So a Type IIIS strain is lethal, whereas a type IIR is non-lethal.

Griffith was able to show that if you heat kill a Type IIIS strain and injected it into the mouse, the mouse lived. But if you mixed the heat-killed type IIIS material with live type IIR bacteria, the mouse would die. Furthermore, the autopsy showed that the mouse became infected with the Type IIIS strain. These meant that some material from the Type IIIS strain was taken up by the Type IIR strain to convert it into the Type IIIS strain. Griffith termed the material the ***transforming principle***.

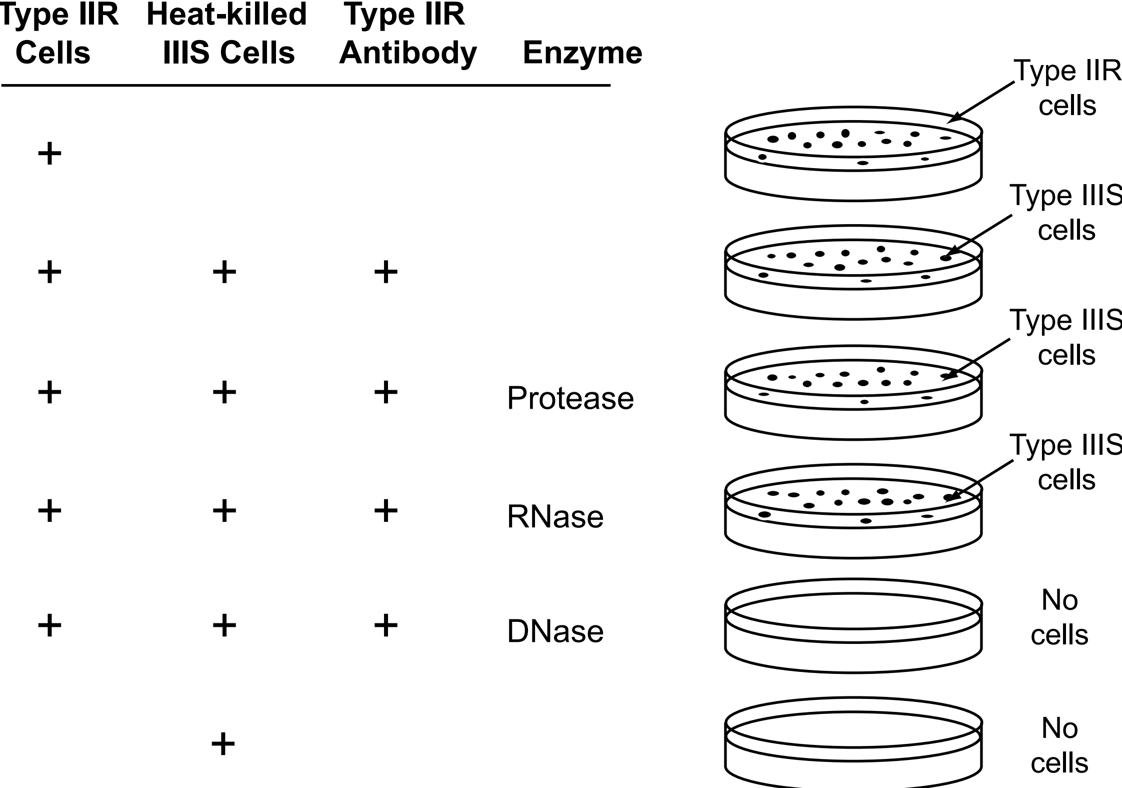
One feature of the genetic material is its ability to control phenotype. In Griffith's experiment, the bacterial strains have several phenotypes. The R types are not only non-lethal, and they have a rough (R) appearance on a blood agar plate. The S type are distinct from the R type: they are lethal and have a smooth morphology on the plates. The S types have a polysaccharide capsule that is lacking in the R types. Each capsule type is distinguished using antibodies; the type II capsule is antigenically distinct from the type III. The transformation from type II to type III and the conversion of type R to S are each distinct phenotypic changes. Therefore if the chemical nature of the transforming principle could be determined, then we would know the nature of the genetic material. Avery, MacLeod and McCarty found the answer.

Figure 1. The experiment of Griffith that demonstrated the concept of the transforming principle.

Avery, MacLeod and McCarty: DNA Is The Genetic Material

A. The Concept

Avery, MacLeod and McCarty extended the work of Griffith. They used his system, but rather than working with the mice they only studied the bacterial phenotypes relative to the material from the dead type IIIS. They performed careful analysis and proved that DNA, and not protein or RNA, was the genetic material.



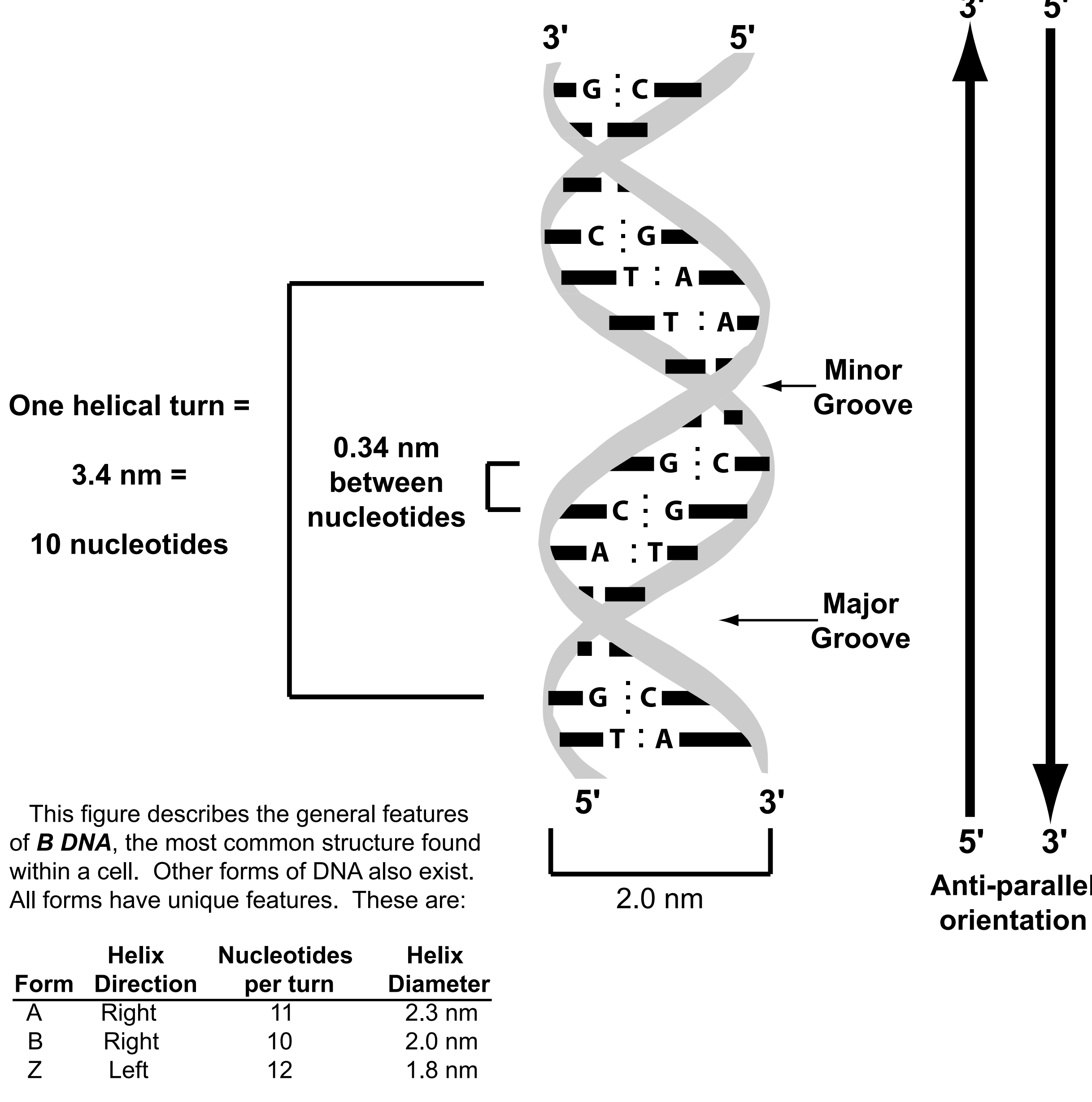
Rather than work with mice, Avery, MacLeod and McCarty used the phenotype of the *Streptococcus pneumoniae* cells expressed on blood agar. To ensure, a few potentially live cells did not escapte the heat treatment, they also precipitated those cells out of culture using an antibody to the type IIR cells. Finally, they included an enzyme treatment of the the material from the heat-killed cells. Each of these enzyme destroyed either proteins (protease), RNA (RNase), or DNA (DNase). These are the three main components of the heat-killed cells. As you can see above, the only treatement that prevented the conversion of the type IIR cells to type IIIS was DNase. This demonstrated conclusively that DNA was the transforming principle and the heredity chemical of life.

Figure 2. The experiment of Avery, MacLeod and McCarty that demonstrated that DNA was the genetic material.

DNA Structure

A. The Concept

DNA has a regular structure. It's orientation, width, width between nucleotides, length and number of nucleotides per helical turn is constant. All of these features were described by Watson and Crick. Adenine is always opposite thymine, and cytosine is always oppostie guanine. The two strands are held together by hydrogen bonds: two bonds between adeninine and thymine and three bonds between guanine and cytosine.



This figure describes the general features of ***B DNA***, the most common structure found within a cell. Other forms of DNA also exist. All forms have unique features. These are:

Form	Helix Direction	Nucleotides per turn	Helix Diameter
A	Right	11	2.3 nm
B	Right	10	2.0 nm
Z	Left	12	1.8 nm

Figure 3. The structure of common DNA molecules.

Deoxyribonucleotide Structure

A. The Concept

DNA is a string of deoxyribonucleotides. These consist of three different components. These are the ***dexoyribose sugar***, a ***phosphate group***, and a ***nitrogen base***. Variation in the nitrogen base composition distingushes each of the four deoxyribonucleotides.

Basic deoxyribonucleotide components

The basic building block is the ***deoxyribose sugar***. This sugar is distinguished because it contains a hydrogen (H) atom at the number 2' carbon. Normal ribose has a hydorxyl (-OH) group at this position.

Attached to the 5' carbon is a triphosphate group. This group is important because in a DNA chain it undergoes a reaction with the 3' OH group to produce polydeoxynucleotide.

The final feature of the molecule is a ***nitrogen base***. These are attached to the 1' carbon. Four bases are possilbe. Two pyrimidines (thymine and cytosine) and two purines (adenine and guanine). The double stranded DNA molecule is held together by hyrodgen bonds. Pairing involves specific atoms in each base. Adenine pairs with the thymine, and guanine pairs with cytosine. These pairings and the atoms involved are shown to the right.

You have probaly heard of ATP, the energy molecule. It is the deoxyribonucleotide to which adenine is attached. This molecule serves two very important functions in biological organisms.

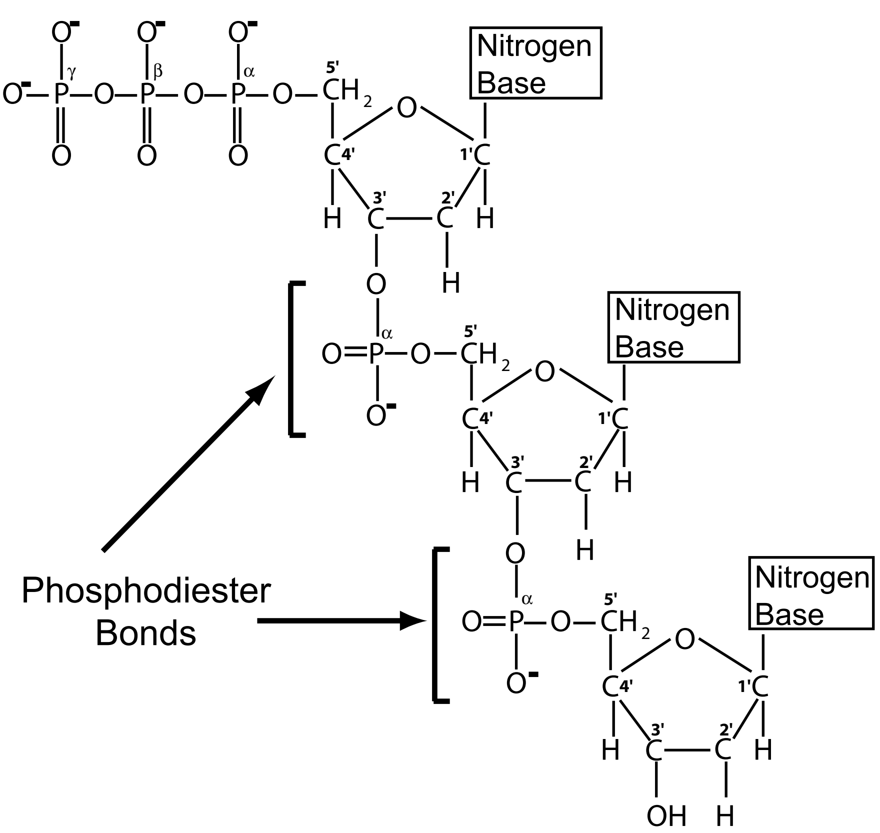
Figure 4. The structure of deoxyribonucleotides and base pairing among N bases.

A Single Strand Molecule of DNA

A. The Concept

Each strand of the double-stranded DNA molecule has the same basic structure. It is a series of series of deoxyribonucleotides linked together by phophodiester bonds.

5' end



3' end

DNA is a polynucleotide. It consists of a series of deoxyribonucleotides that are joined by phosphodiester bonds. This bond joins the a phosphate group to the 3' carbon of the deoxyribose sugar.

Each strand is complementary to the opposite strand. If one strand has an adenine at a position, its anti-parallele strand would have a thymine at the the corresponding position. Likewise, guanine and cytosine would be complementary.

Fig. 5. The single strand structure of DNA.

Making a Phosphodiester Bond/ Growing the DNA Chain

A. The Concept

The addition of a new nucleotide to a DNA molcule creates a phosphodiester bond. This requires the DNA chain that is being elongated and a deoxyribo-nucleotide.

5' end

3' end

5' end

3' end

(Pyrophosphate)

Phosphodiester bonds are formed when a new dideoxynucleotide is added to a growing DNA molecule. During the reaction, a condensation reaction occurs between the α phosphate of the nucleotide and the hyroxyl group attached to the 3' carbon. This reaction is performed by the enzyme DNA polymerase. This is also an energy requiring reaction. The energy is provided by the breaking of the high-energy phophate bond in the nucleotide. This results in the release of a pyrophosphate molecule.

Figure 6. The formation of the phosphodiester bond that grows the DNA chain.

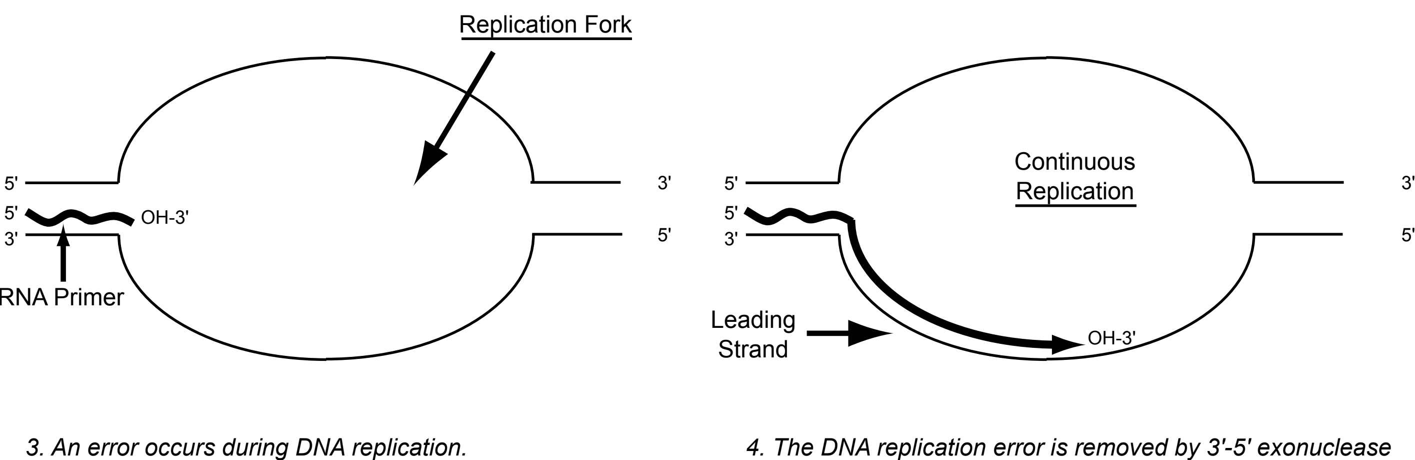
Steps of DNA Replication
(Part 1)

A. The Concept

DNA replication is an essential biological process. It's primary function is to produce new DNA for cell division. The process has several distinct steps that are important to understand. The factors that are absolute requirements for DNA replication to begin are a ***free 3'-OH group*** and a ***DNA template***. A RNA primer provides the free 3'-OH group. The DNA to be replicated serves as the template. It is important to remember that ***all*** DNA replication proceeds in the 5'-3' direction.

1. The replication fork is formed; RNA primer added.

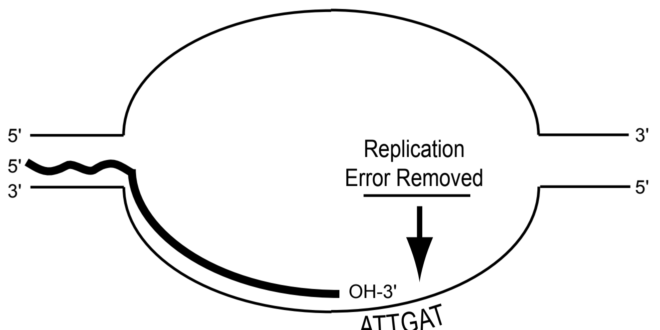
2. DNA is replicated by the 5'-3' synthesis function of DNA polymerase using the leading strand in a continuous manner.



3. An error occurs during DNA replication.

5'
5'
3'
Replication
Error
3'
5'
TAACTT
ATTGAT

4. The DNA replication error is removed by 3'-5' exonuclease function of DNA polymerase.



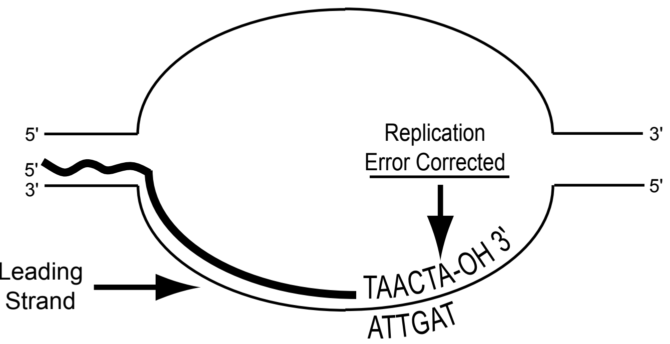
Notes on *E. coli* replication:

DNA Polymerase I and III. Pol III is the primary replicase enzyme that performs the elongation of the DNA strand. It adds nucleotides first to the RNA primer and then grows the chain by creating the phosphodiester bonds. It also has a 3'-5' proofreading (exonulcease) function that removes incorrecly incorporated nucleotides. DNA Pol I also has the 5'-3' replicase function, but it is primarily used to fill the gaps in the replicated DNA that occur when the RNA primer is removed. This enzyme also has a 5'-3' exonuclease function that is used to remove the RNA primer.

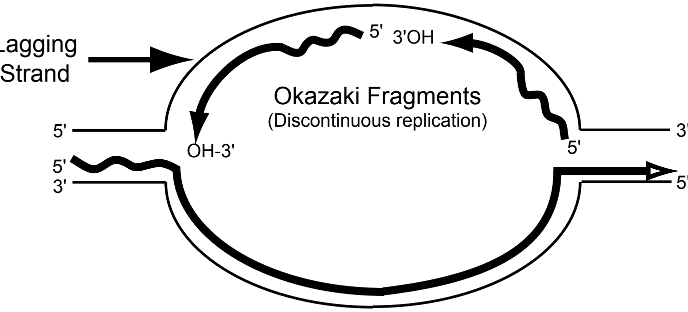
Figure 7. The steps of DNA replication.

Steps of DNA Replication
(Part 2)

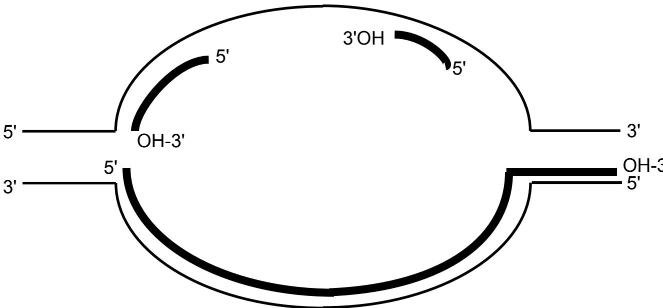
5. The DNA replication error is corrected.



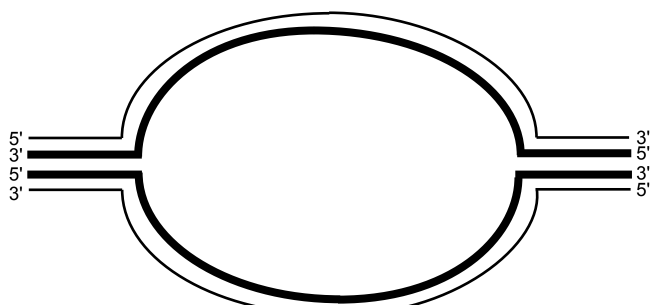
6. Meanwhile, Okazaki fragments are synthesized using the lagging strand in a discontinuous manner while the leadng strand is completed simultaneously.



7. The RNA primers are removed by 5'-3 exonuclease function of DNA polymerase.



8.. Replication is completed by the filling in the gaps by DNA polymerae and DNA ligase.



Notes on replication:

Okazaki fragments: Both prokaryotic and eukaryotic DNA replication proceed in the 5'-3' direction. This poses a problem because the replication fork only moves in that one direction. The problem relates to what is called the ***lagging strand***. It must be replicated in a direction that is opposite of the direction of the replication fork. This problem was solved by the discovery of Okazaki fragments (named after the person who discovered the process). In contrast to the ***leading strand***, in which DNA is replicated as a single molecule in a ***continuous*** manner, DNA is replicated in a ***disocontinuous*** manner on the lagging strand. Each fragment requires a RNA primer, and DNA Pol III in *E. coli* makes short stretches of DNA. These fragments are then stitched together when the primer is removed, and the strands are completed by the action of DNA Pol I and ligase.

Figure 7 (cont.). The steps of DNA replication.

Chain Termination Sequencing: the Sanger Technique

A. The Concept

DNA sequencing is the most important technique of genomics. By collecting the sequence of genes and genomes we begin to understand the raw material of phenotype development. The most common DNA sequencing technique is called ***chain termination sequencing*** or the ***Sanger technique*** (named after the person who created it). It is called chain termination because the incorporation of a ***dideoxynucleotide*** terminates the replication process because this nucleotide lacks the required 3'-OH group.

a. A dideoxynucleotide

O⁻ O⁻ O⁻
γ β α 5'
O⁻–P–O–P–O–P–O–CH_2
Nitrogen Base
4' 1' 3' 2'
Phosphate Group
Sugar Moiety
Note: neither the 2' or 3'carbon has an OH group

b. The reaction reagents

DNA template
sequencing primer
dNTPs
ddNTPs (low concentration)
DNA polymerase
salts

c. The sequencing reaction result: fragments that differ by one nucleotide in length

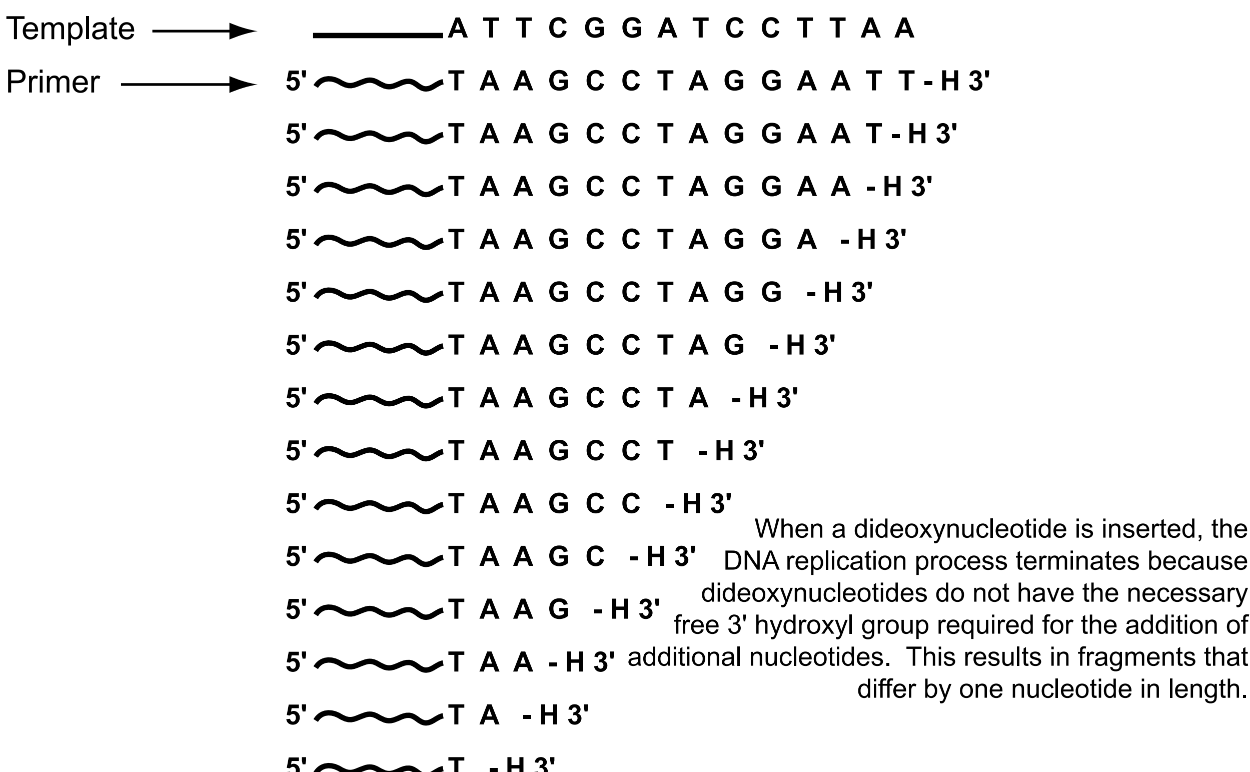


Figure 8. The chain termination (Sanger) DNA sequencing technique.

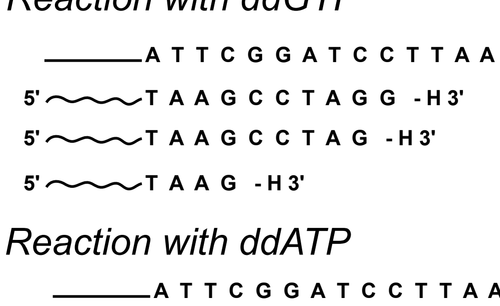
Gel-based Detection of DNA Sequences

A. The concept

Four DNA sequencing reactions are performed. Each contains only one of the four dideoxynucleotides. Each reaction is added to a single lane on the gel. Since one of the dNTPs is radioactive, the gel in which the fragments are separated, can be used to expose an x-ray film and read the sequence.

a. The sequencing products

Reaction with ddGTP



Reaction with ddATP

A T T C G G A T C C T T A A
5' T A A G C C T A G G A A -H 3'
5' T A A G C C T A G G A -H 3'
5' T A A G C C T A -H 3'
5' T A A -H 3'
5' T A -H 3'

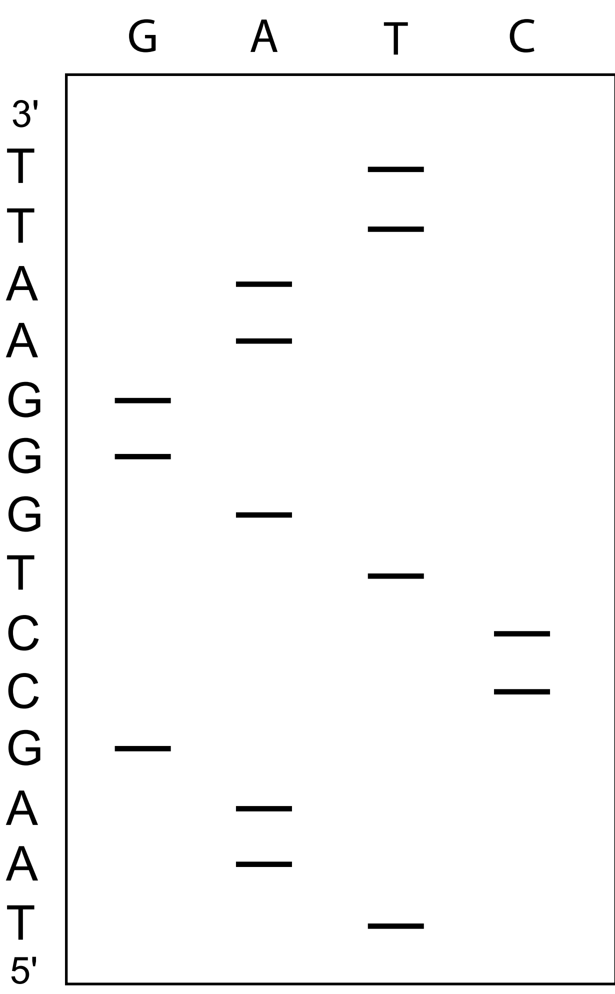
Reaction with ddTTP

A T T C G G A T C C T T A A
5' T A A G C C T A G G A A T T -H 3'
5' T A A G C C T A G G A A T -H 3'
5' T A A G C C T -H 3'
5' T -H 3'

Reaction with ddCTP

A T T C G G A T C C T T A A
5' T A A G C C -H 3'
5' T A A G C -H 3'

b. The sequencing gel



The sequencing reactions are separated on a polyacrylamide gel. This gel separates the fragments based on size. The shorter fragments run further, the longer fragments run a shorter distance. This allows the scientists to read the sequence in the 5'-3' direction going from the bottom to the top of the gel.

Figure 9. Gel-based detection of DNA sequencing products.

Fluorescent Sequencing and Laser Detection

A. The Concept

Rather than using four different reactions, each with a single dideoxynucleotide, the advent of fluorescently labeled dideoxynucleotide enabled 1) the sequencing reaction to be performed in a single tube, and 2) the fragment could be detected by laser technology. Originally, the products were separated in a polyacrylamide gel prior to laser detection. The introduction of capillary electrophoresis, coupled with laser detection enabled the detection of up to 96 products at a time.

B. The Reaction Products and Analysis

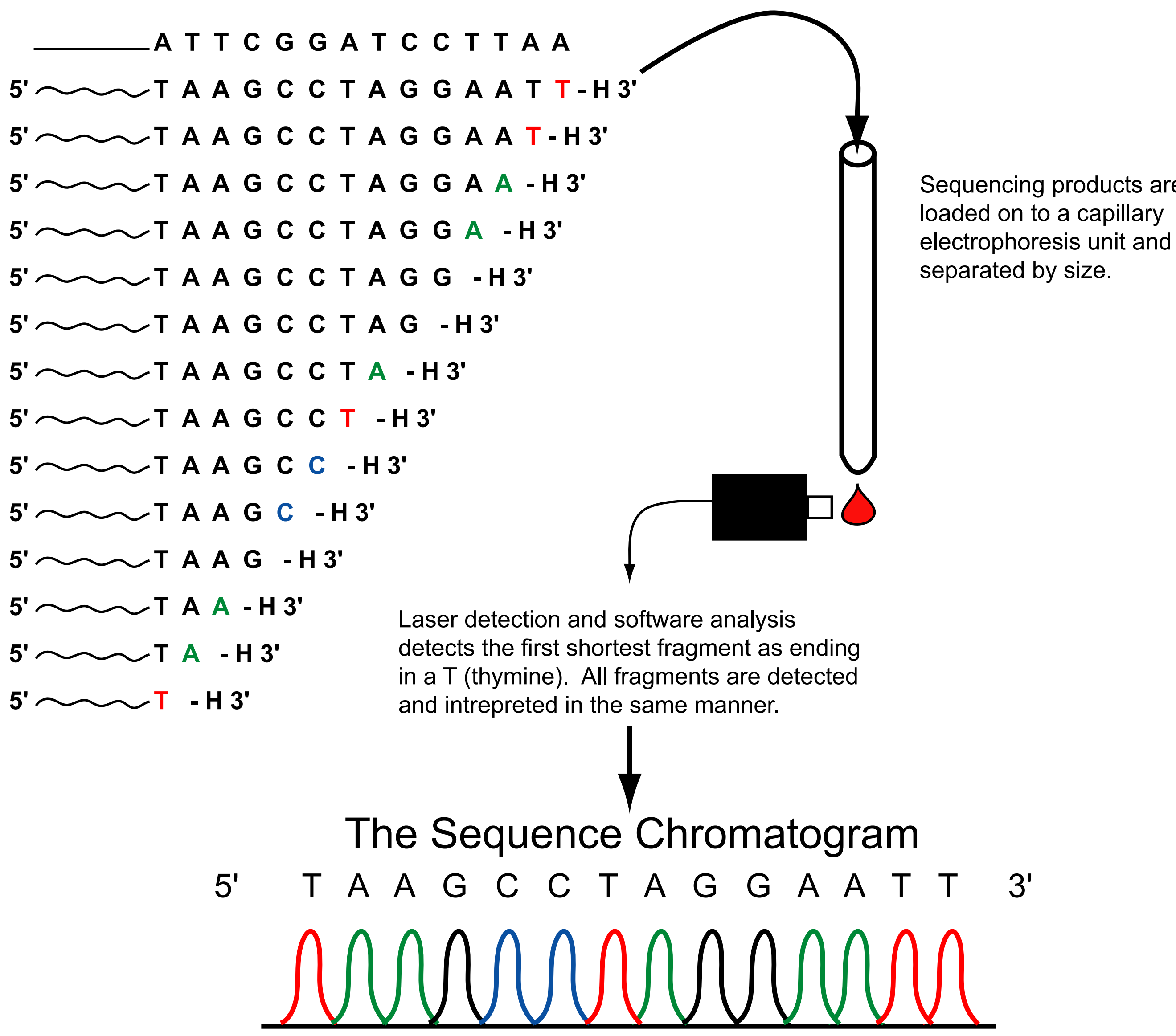
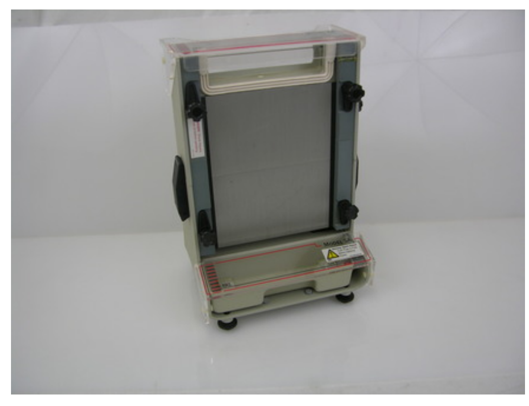
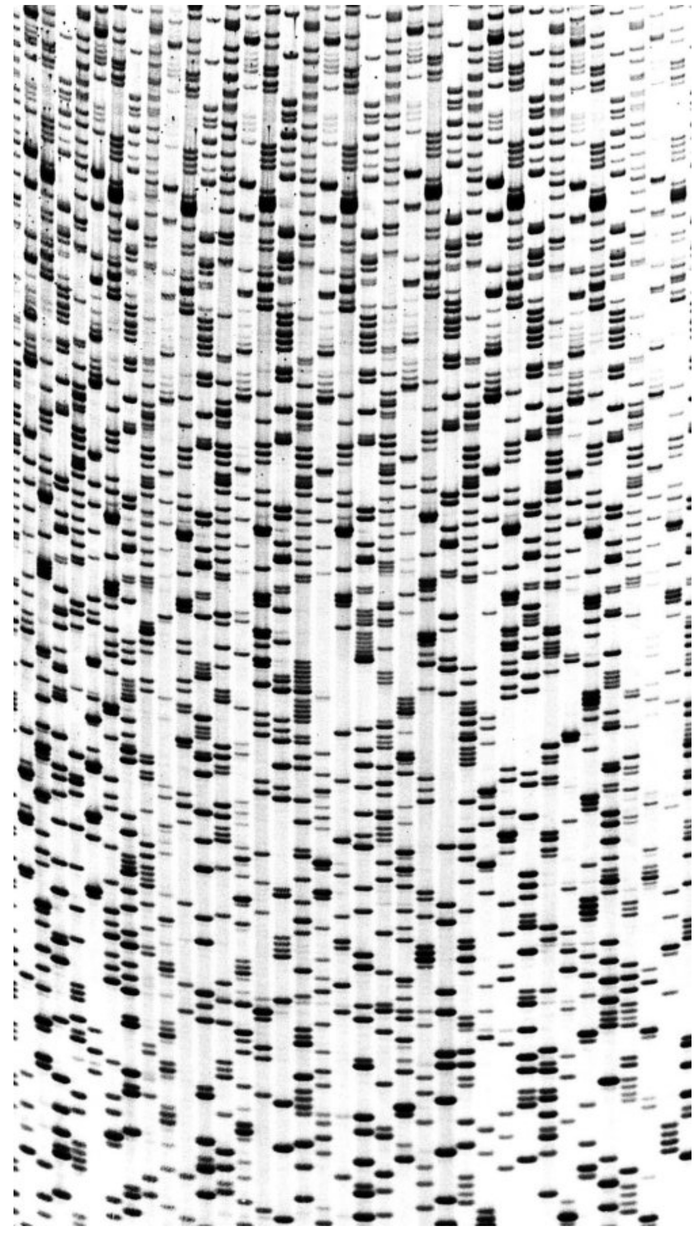


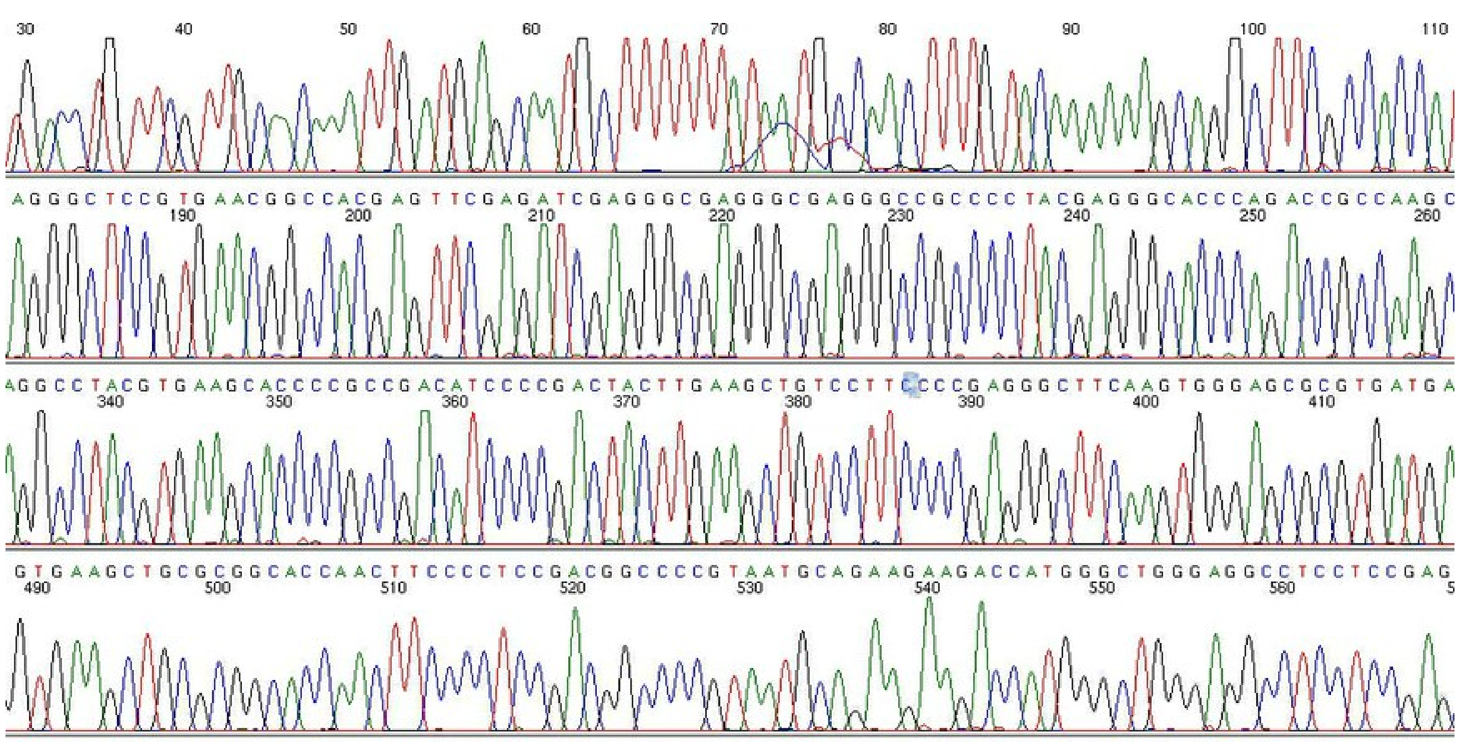
Figure 10. The fluorescent sequencing and laser detectiion process of DNA sequencing.

DNA Autoradiogram: Original Output Data from Sequencing





Output from Automated DNA Sequencer



Cheaper Sequencing: The Wave of the Future

Based on:

Bentley (2006) Current Opinion in Genetics and Development

What are the types of sequencing projects?

- Whole genome
    - Collecting the full sequence of a genome
        - Draft
            - 90-95% coverage
        - Finished
            - 98-99% coverage
- Resequencing
    - Targeted sequencing for localized regions
        - Discover localized changes
        - ***Requires reference sequence***

Human Genome Project

- $1 billion to complete
    - 99% of euchromatic region of the genome
    - 99.995% accuracy

2004 Costs

- Sanger technique
- $10 million for individual

What if it cost:

- $100,000 (100X cost reduction)
    - Sequence 24-48 humans
    - Discover genome-wide SNP
- $10,000 (1000X cost reduction)
    - Sequence about 500 humans
        - Discover evolution patterns
            - Genes undergoing natural selection can be discovered
            - Presumably key genes that “make humans humans”
    - Deep sequencing of specific disease gene
        - Discover key mutations associated with onset of disease
- $1000 (10,000X cost reduction)
    - Personal genome characterization

What Was Needed for All New Approaches

Reducing Cost

- How: Parallel sequencing
    - Large number of sequencing reactions occurring simultaneously
        - Requires high density reactions matrix
            - Many reactions in a small space
            - Miniaturization of reaction unit or space
    - Reduce reagent cost
        - Accomplished when above factors achieved

Throughput

- Many reactions occurring simultaneously
    - Current Sanger macrocapillary system
        - 96-384 reactions per machine
        - Sequencing centers have 30-60 machines
        - ***New approaches must have significantly greater throughput***

Sequence Accuracy Must Be Maintained

- Sanger procedure highly accurate
    - Well understood Phred scores reported
        - ***New systems will require quantifiable accuracy scores***

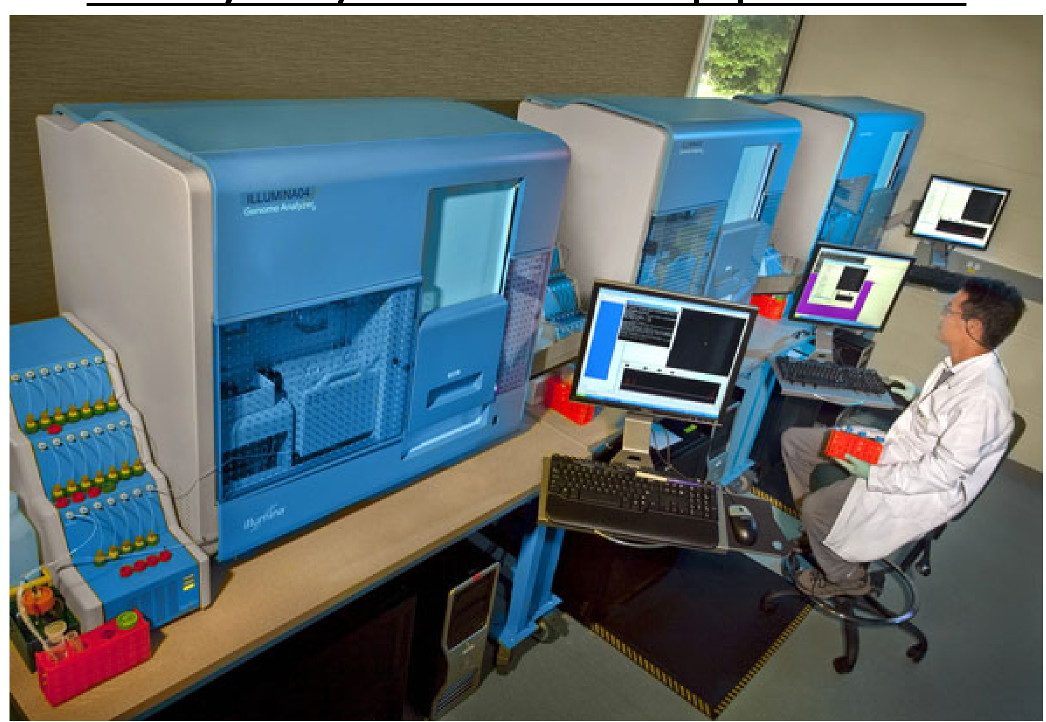
Completeness

- Read length issue
    - Sanger technology with capillary detection
    - 500-700 nt
        - Allows for assembly into
            - Contigs
            - Supercontigs
- Emerging technologies
    - Length requirement
    - Must be long enough to align accurately
        - 25-100 nt read length
            - Sufficient for resequencing with a reference genome
- Whole genome sequencing
    - 100 nt (or longer) needed for smaller genomes
    - Other advances needed for larger genomes

How Large Scale Sequencing Has Changed Over Time From a Centers Perspective

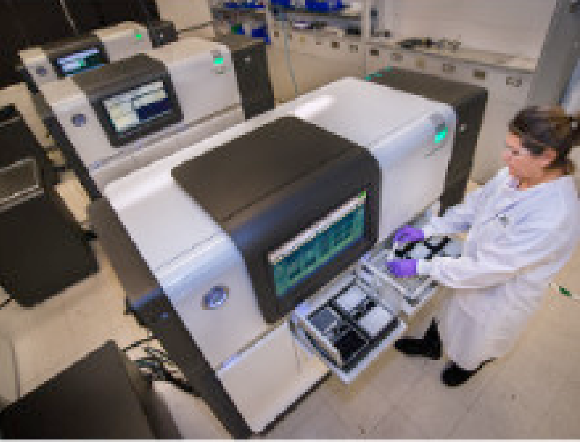
Then: DOE/JGI Sanger Sequencing Equipment Room

Recently: DOE/JGI Illumina GAII Equipment Room

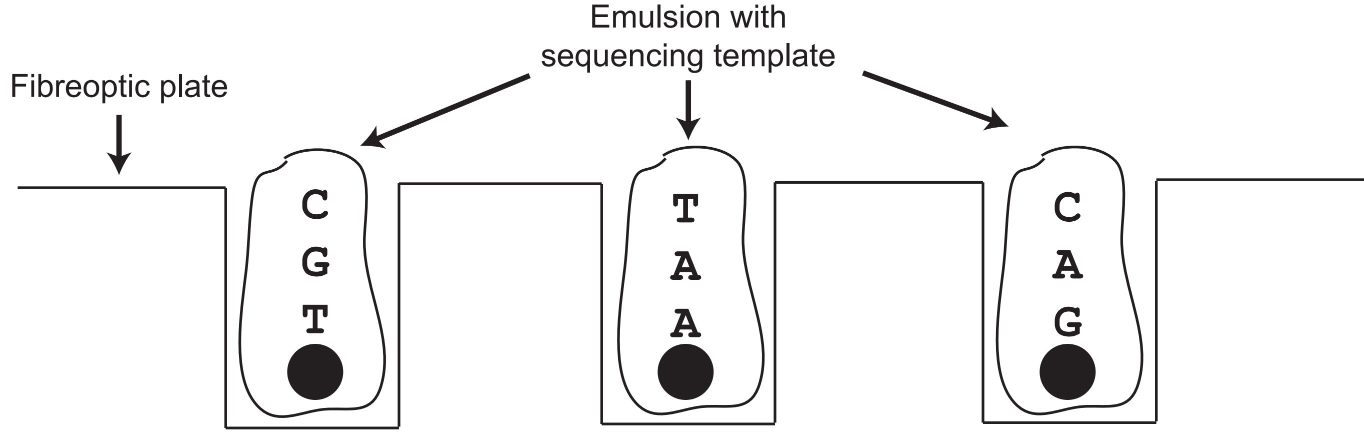


Now: DOE/JGI Illumina HiSEQ Equipment Room

Now: DOE/JGI PacBIO Equipment Room



454 Sequencing Reaction Principles



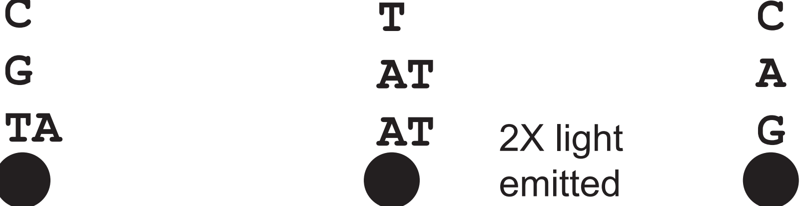
Step 1: add dATP
and other reagents

C G TA — 1X light emitted

T A A

C A G

Step 2: add dtTP
and other reagents



Step 3: add dCTP
and other reagents

C GC TA — 1X light emitted

T AT AT

C A GC — 1X light emitted

The Sequencing Technology

*Enhance your sequencing process —
from genome to sequence in record time*

Process Steps

Generate tens of millions of bases per run with the straightforward workflow of the Genome Sequencer 20 System (Figures 3-6).

DNA Library Preparation

Sample preparation is dependent on the type of starting material used. The preparation process comprises a series of enzymatic steps to produce single-stranded template DNA (sstDNA) incorporating primer and binding adaptors. For example, genomic DNA (gDNA) is fractionated into smaller fragments (300-800 base pairs) that are subsequently polished (blunted). Short Adaptors (A and B) are then ligated onto the ends of the fragments. These adaptors provide priming sequences for both amplification (emPCR) and sequencing of the sample-library fragments, and contain a streptavidin binding site for sample purification. Low molecular weight DNA is used without fragmentation and sample preparation begins with adaptor ligation. The A and B adaptors can also be added during PCR by using the appropriate primers (provided in GS emPCR Kit II (Amplicon A, Paired End) and GS emPCR Kit III (Amplicon B). The sstDNA library produced at the end of this preparation step is assessed for its quality, and the optimal amount (DNA copies per bead) needed for emPCR is determined by a titration run.

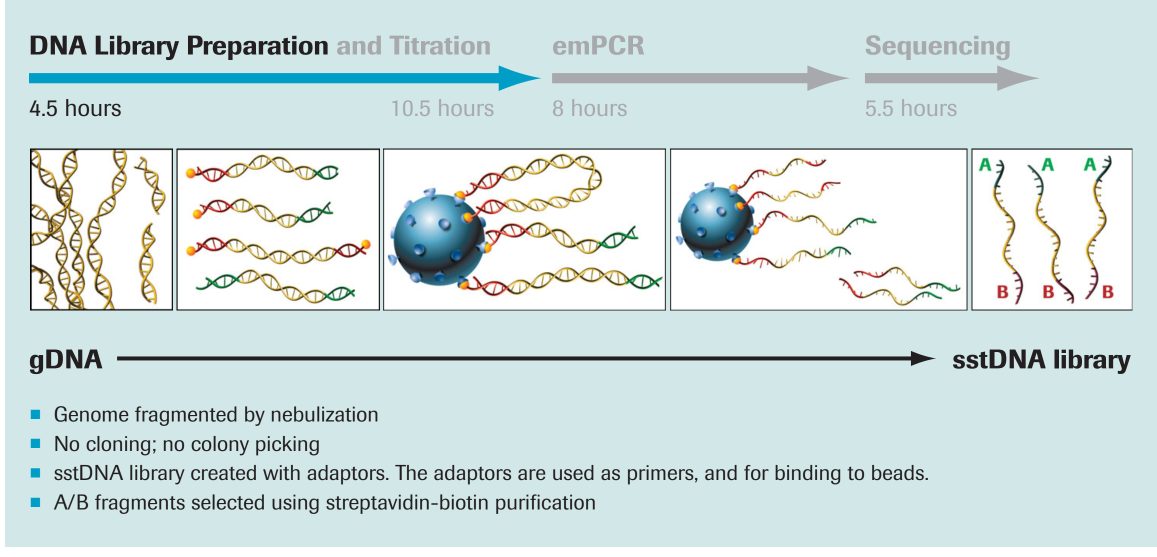


Figure 3: DNA library preparation with the Genome Sequencer 20 System.

emPCR Amplification

The sstDNA library is immobilized onto specially designed DNA Capture Beads. Each bead carries a single sstDNA library fragment. The bead-bound library is emulsified with amplification reagents in a water-in-oil mixture. Each bead is separately captured within its own microreactor for PCR amplification. Amplification is performed in bulk, resulting in bead-immobilized, clonally amplified DNA fragments that are specific to each bead.

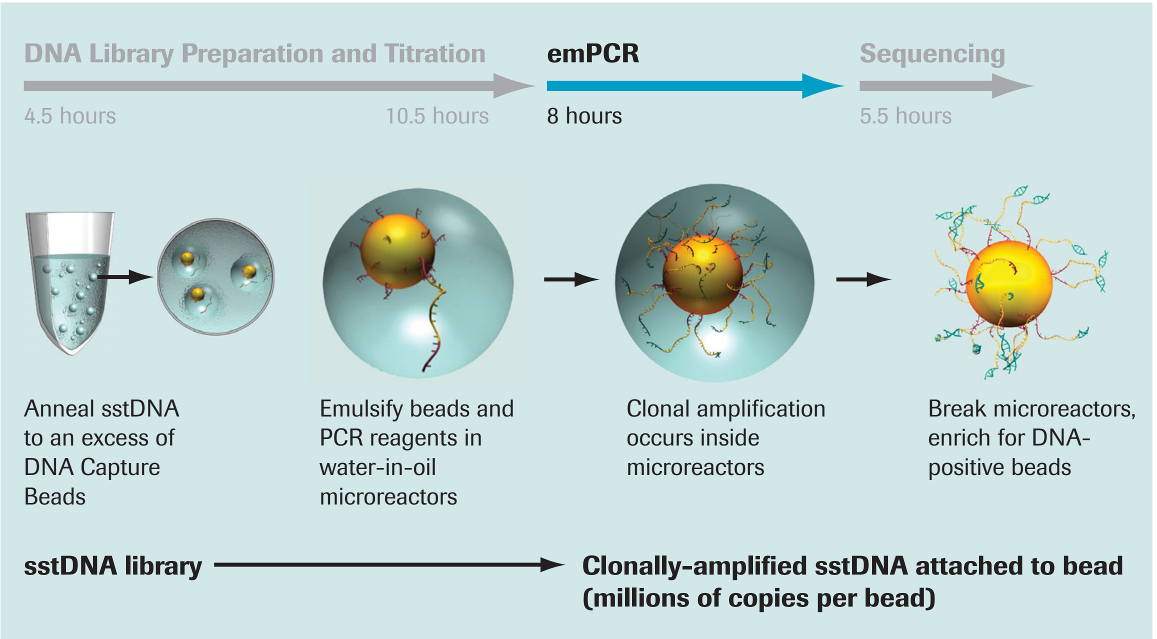


Figure 4: Overview of emulsion-based clonal amplification (emPCR) with the Genome Sequencer 20 System.

Process Steps

The Sequencing Technology

*Enhance your sequencing process —
from genome to sequence in record time*

Process Steps

Sequencing-by-Synthesis

Sequencing starts with the preparation of a PicoTiterPlate device; during this step, a combination of beads, sequencing enzymes, and an sstDNA library is deposited into the wells of the device. The bead-deposition process maximizes the number of wells that contain an individual sstDNA library bead.

The loaded PicoTiterPlate device is placed into the Genome Sequencer 20 Instrument. The fluidics subsystem flows sequencing reagents (containing buffers and nucleotides) across the wells of the plate. Each sequencing cycle consists of flowing individual nucleotides in a fixed order (TACG) across the PicoTiterPlate device. During the nucleotide flow, each of the hundreds of thousands of beads with millions of copies of DNA is sequenced in parallel.

If a nucleotide complementary to the template strand is flowed into a well, the polymerase extends the existing DNA strand by adding nucleotide(s). Addition of one (or more) nucleotide(s) results in a reaction that generates a chemiluminescent signal that is recorded by the CCD camera in the Genome Sequencer 20 Instrument. The signal strength is proportional to the number of nucleotides incorporated in a single nucleotide flow.

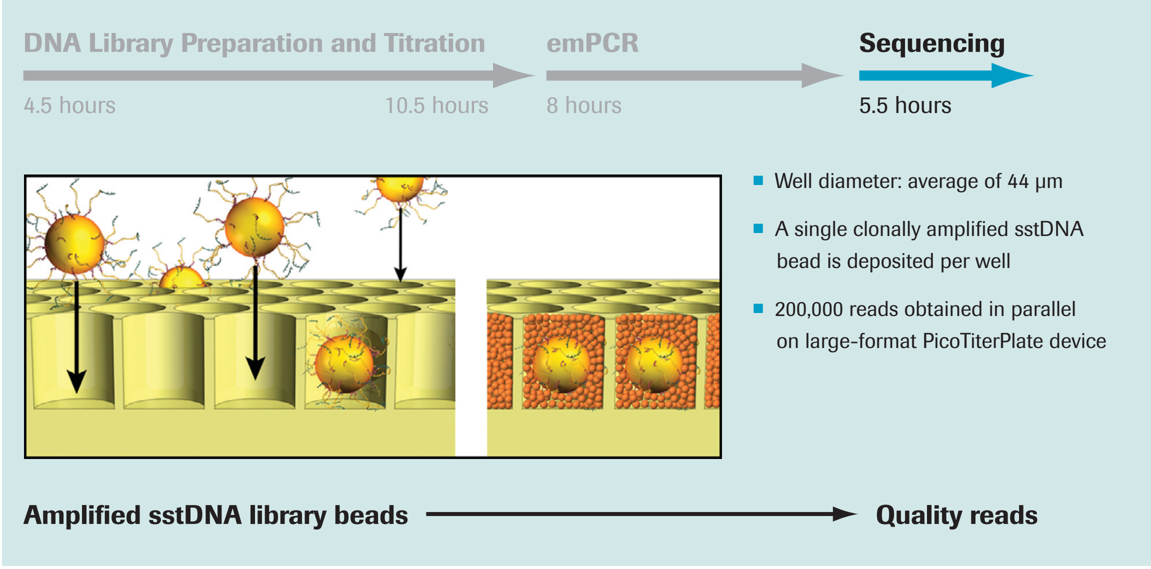


Figure 5: Deposition of DNA beads into the PicoTiterPlate device.

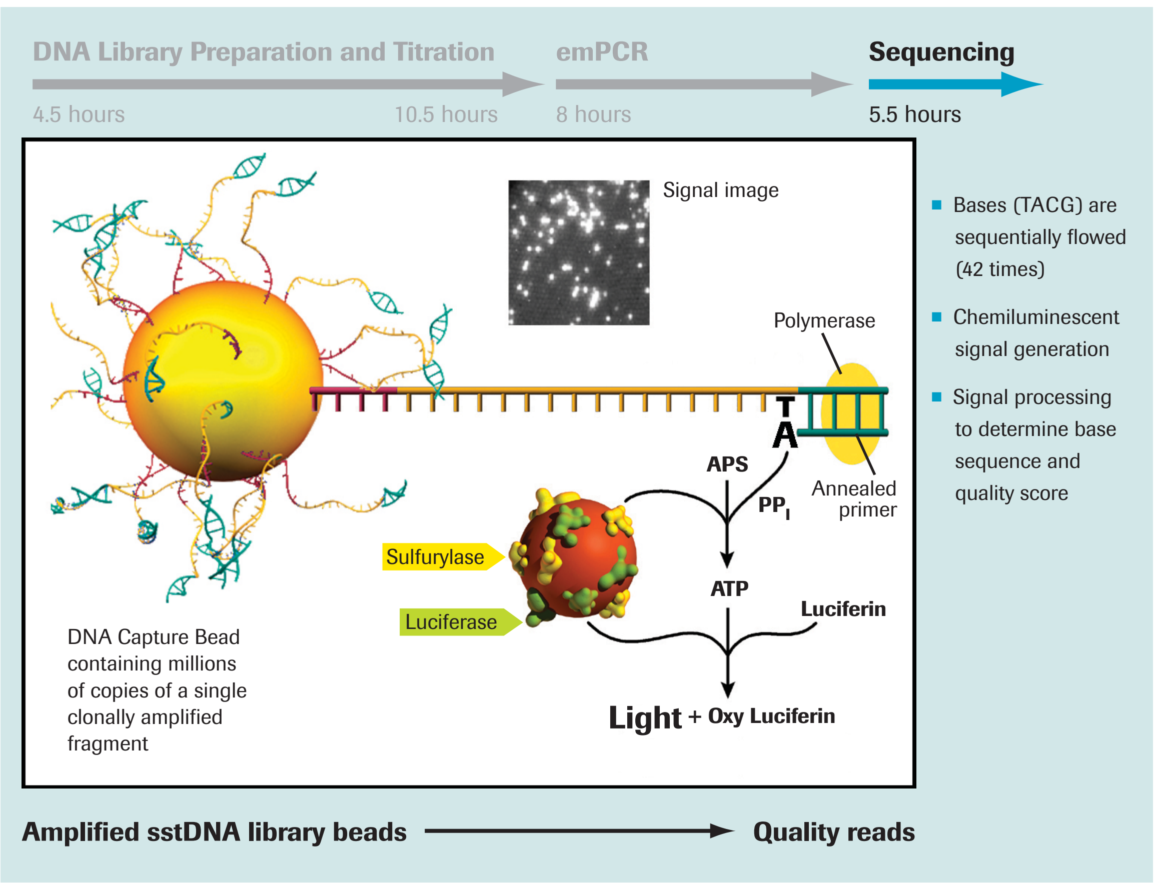


Figure 6: Sequencing reaction of the Genome Sequencer 20 System.

The Genome Sequencer 20 Instrument

Perform ultra-high-throughput DNA sequencing

Genome Sequencer 20 System

The Genome Sequencer 20 System revolutionizes DNA sequencing, delivering sequence data in a massively parallel fashion.

The Genome Sequencer 20 System includes:

- Instrument and accessories
- Reagents and consumables for library construction, amplification, and sequencing
- Analysis software for resequencing, *de novo* assembly, and amplicon sequencing.

The instrument (Figure 7) is the centerpiece of the Genome Sequencer 20 System. It comprises both optics and fluidics subsystems, which are controlled by a computer subsystem.

The fluidics subsystem consists of a reagents cassette, a sipper manifold, pumps, valves, and debubblers. It ensures accurate reagent dispensing and flows the sequencing reagents across the wells of the PicoTiterPlate device.

The optics subsystem includes a CCD camera, which captures the light signal resulting from the sequencing reaction (Figure 8).

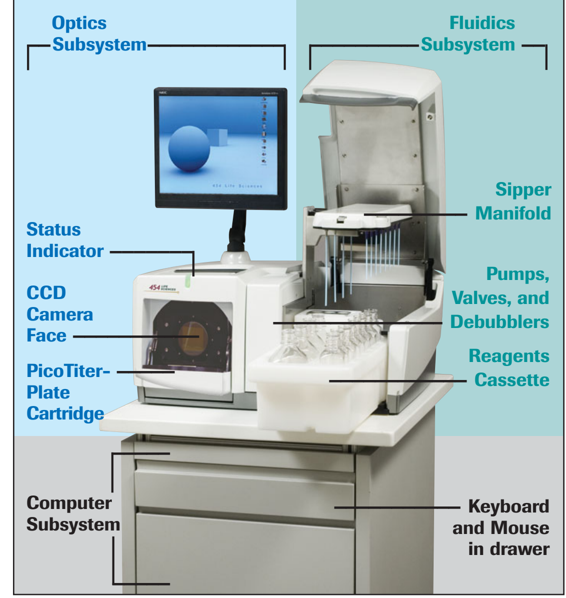


Figure 7: Open view of the Genome Sequencer 20 Instrument.

Instrument space requirement: 30 in (77 cm) Wide x 36 in (92 cm) Deep x 69 in (176 cm) High (cart and instrument).

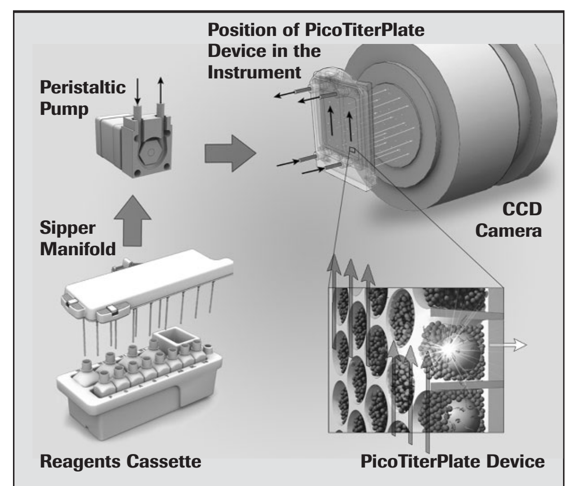
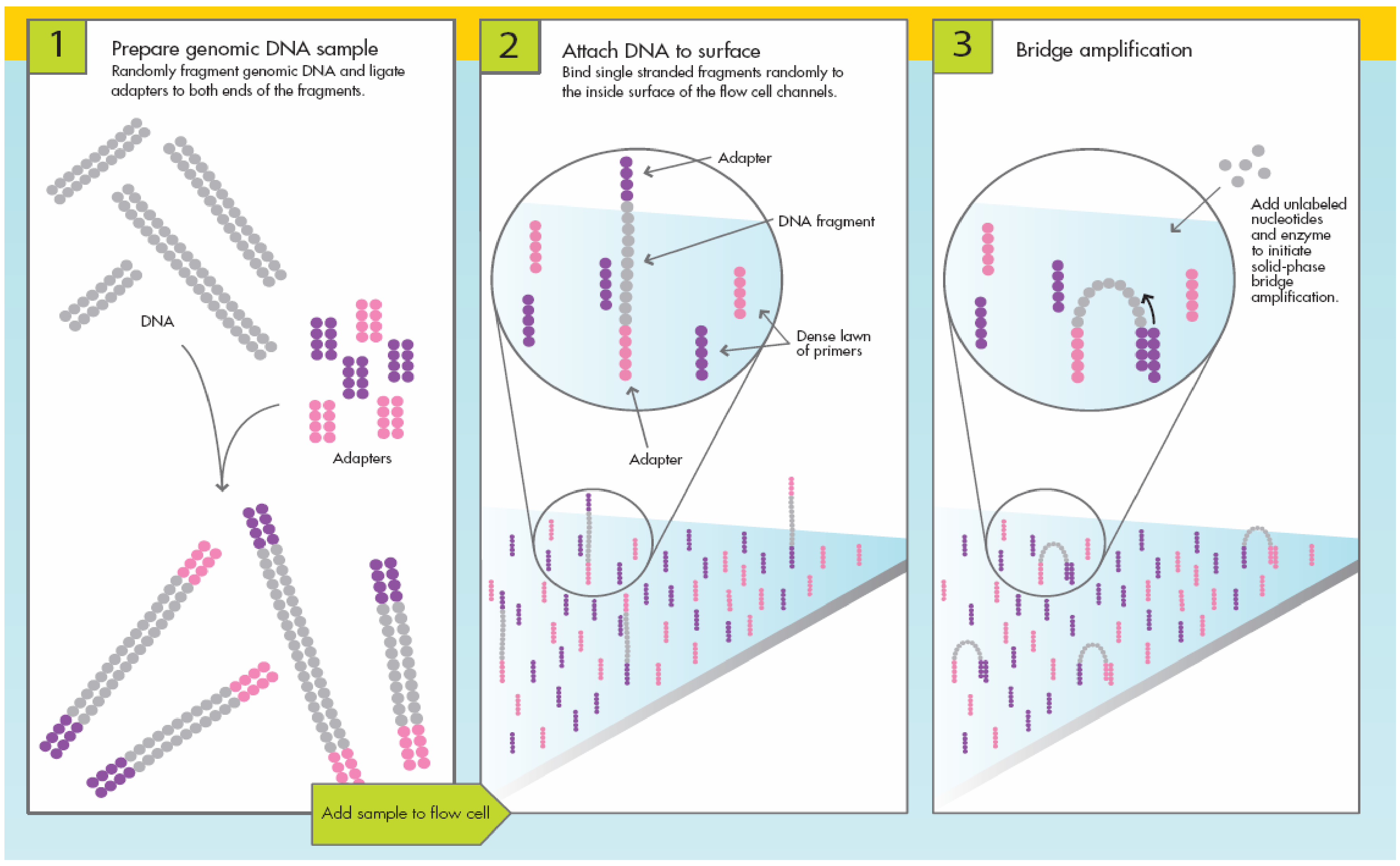
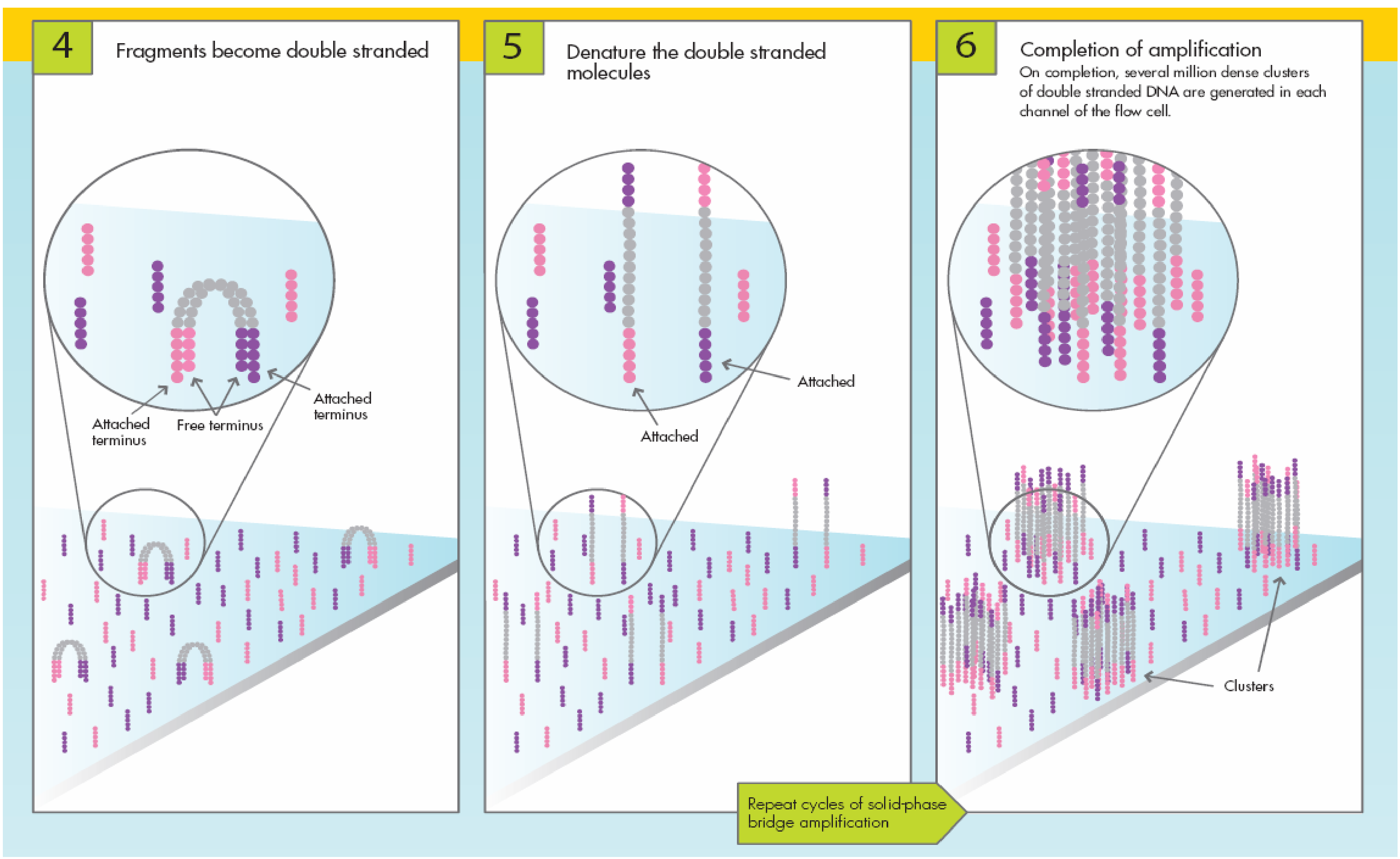


Figure 8: Expanded view of the Genome Sequencer 20 Instrument components. Arrows represent reagent flow.

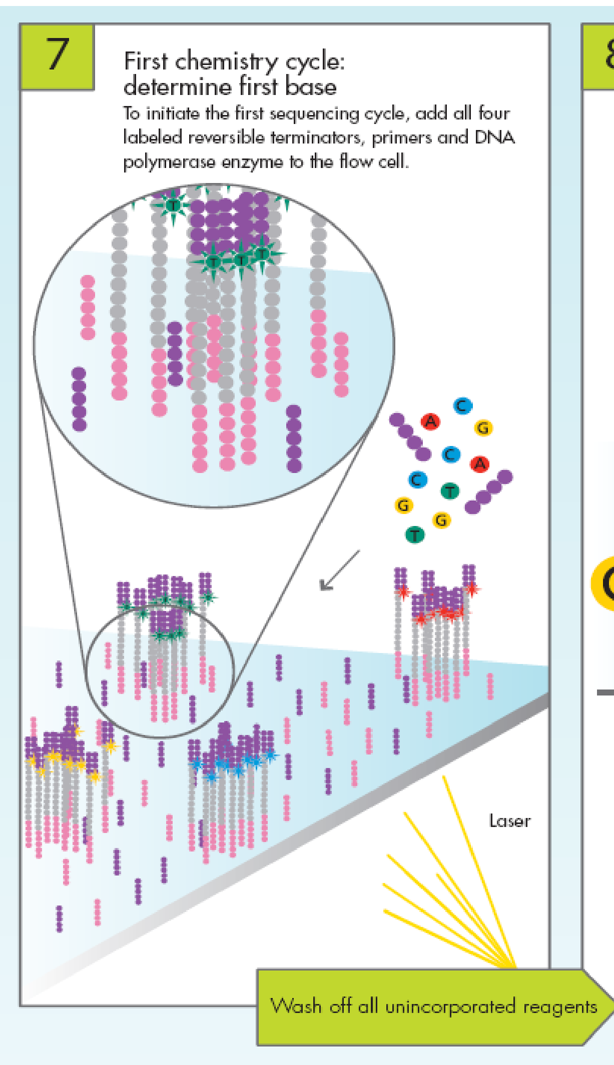
Illumina Sequencing by Synthesis Technology




4
Fragments become double stranded
Attached terminus
Free terminus
Attached terminus
5
Denature the double stranded molecules
Attached
Attached
6
Completion of amplification
On completion, several million dense clusters of double stranded DNA are generated in each channel of the flow cell.
Clusters
Repeat cycles of solid-phase bridge amplification

7 First chemistry cycle: determine first base

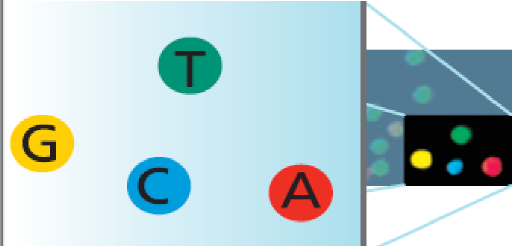
To initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.



Wash off all unincorporated reagents

8 Image of first chemistry cycle

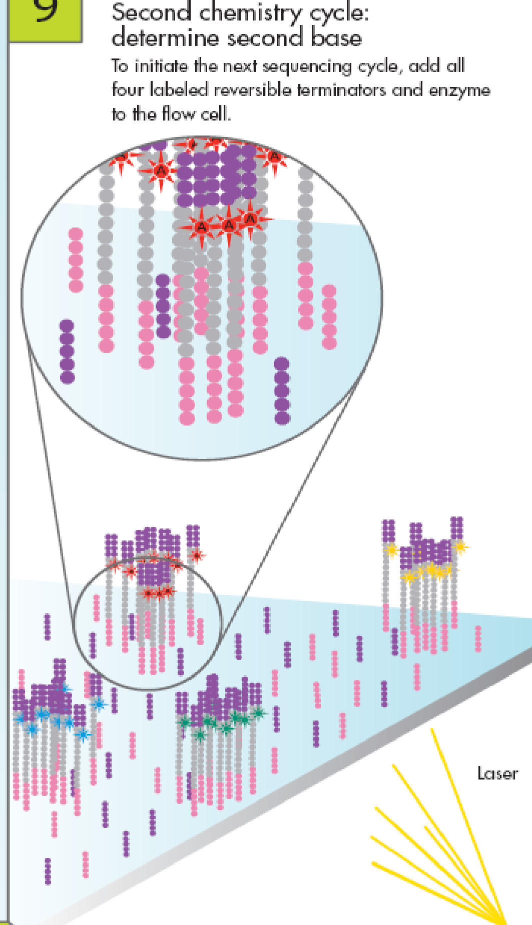
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

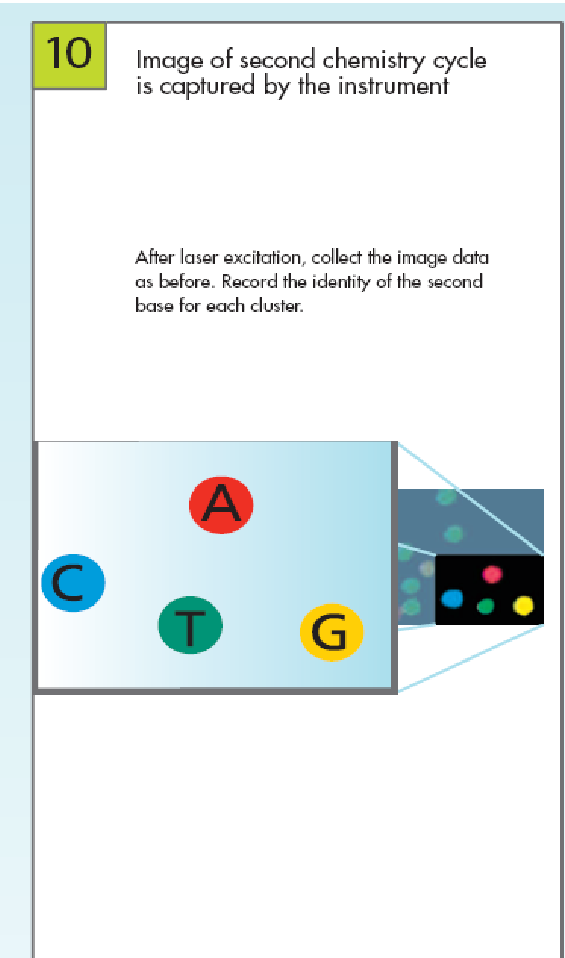


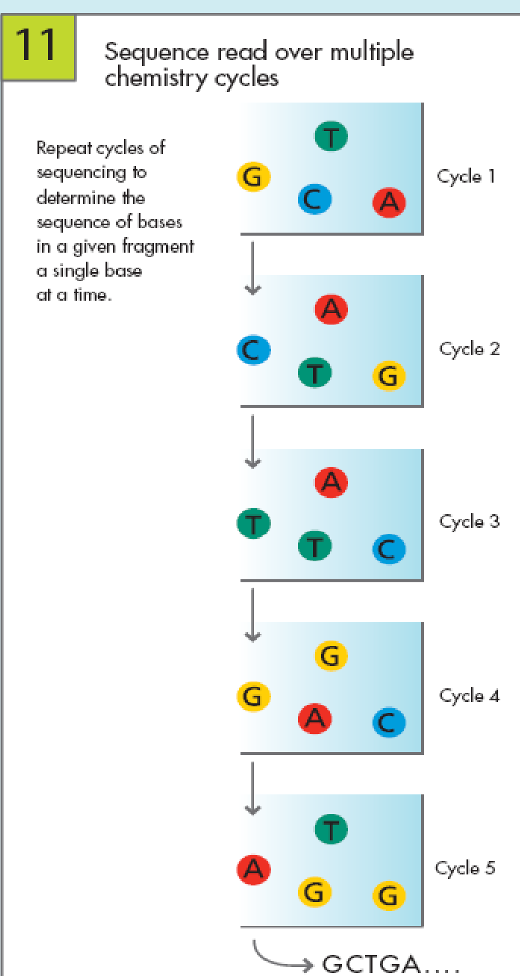
Remove the blocked 3′ terminus and the fluorophore from each incorporated base

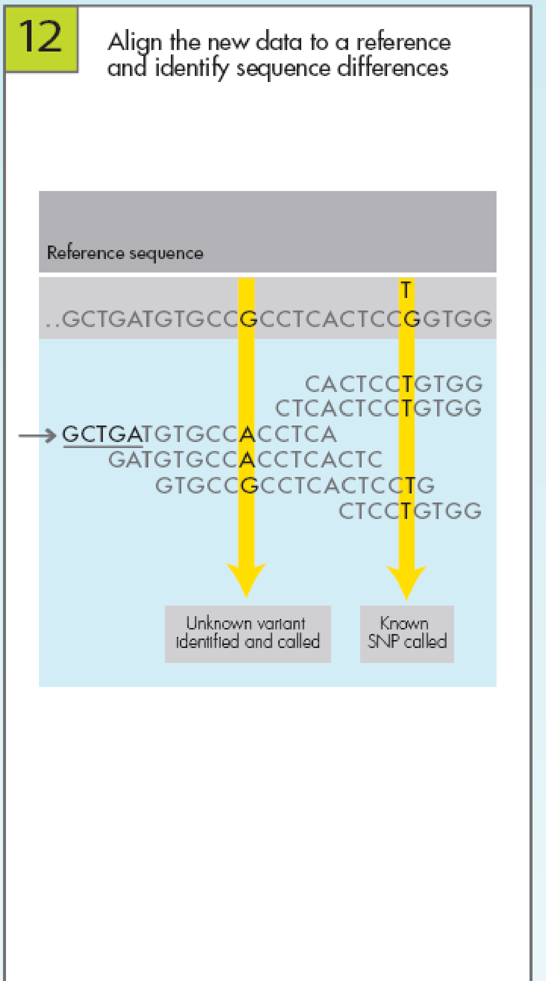
9 Second chemistry cycle: determine second base

To initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.




10
Image of second chemistry cycle is captured by the instrument
After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.
A
C
T
G


11
Sequence read over multiple chemistry cycles
Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.
Cycle 1
Cycle 2
Cycle 3
Cycle 4
Cycle 5
GCTGA....


12
Align the new data to a reference and identify sequence differences
Reference sequence
T
..GCTGATGTGCCGCCTCACTCCGGTGG
CACTCCTGTGG
CTCACTCCTGTGG
GCTGATGTGCCACCTCA
GATGTGCCACCTCACTC
GTGCCGCCTCACTCCTG
CTCCTGTGG
Unknown variant identified and called
Known SNP called

Illumina Sequencers Over Time: Today’s Workhorse

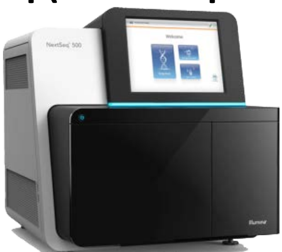
Illumina GAII [Maximum (Max) output: 25 gigabases (Gb)]

Illumina HiSeq 2500 (Max output; 500 Gb; Rapid Run Mode: 150 Gb)

Today's Illumina Models

(Mostly chemistry and reads per flow cell differences)

Illumina NextSeq (Max output: 120 gigabases)



Illumina HiSeq X10 (Max output: 1.8 Tb) GENOMES ONLY

Illumina HiSEQ 4000 (Max output: 1.5 Tb) Most other sequencing

Illumina NovaSeq (Next Generation; 2017 release; Max output: 1.5 Tb)

Single Polymerase Real Time DNA Sequencing

Developed by Pacific Biosciences
Sequences occurs at the rate of ***10 nt per second***

Principle

Reaction Cell

- A single DNA polymerase is immobilized on the bottom of a reaction cell
  - Reaction cell called a ZMW (Zero-mode waveguide)
- Φ29 DNA polymerase is used
  - Fast single subunit enzyme.
- Each sequencing plate contains ~3000 individual cells
  - Each holds only a single DNA molecule

Chemistry

- A phospholinked dNTP is used
  - Each dNTP contains a different fluorophore
- During sequence
  - A single labeled dNTP enters the polymerase
  - dNTP held in place shortly
  - Fluorescence signal is emitted in the ZMW for a short period of time
  - dNTP leaves and new dNTP enters

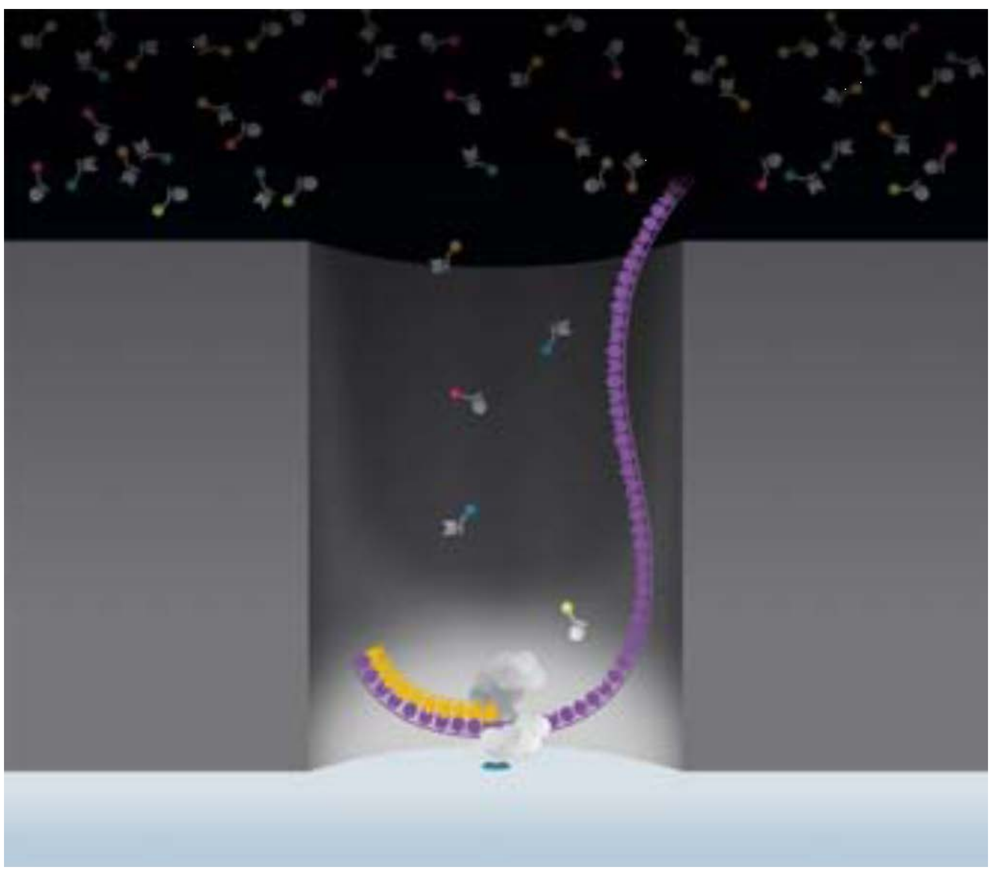
Detection and sequence determination

- Fluorescence signals for each ZMW collected
  - Data is collected as a movie of the sequential signals
    - Each individual signal is measured as a short pulse of light
  - Successive fluorescence signal data is collected
  - DNA sequence of single molecule is determined by sequence of light pulses

Images and Notes Below From:

Pacific Biosciences Technology Backgrounder (11/24/2008)

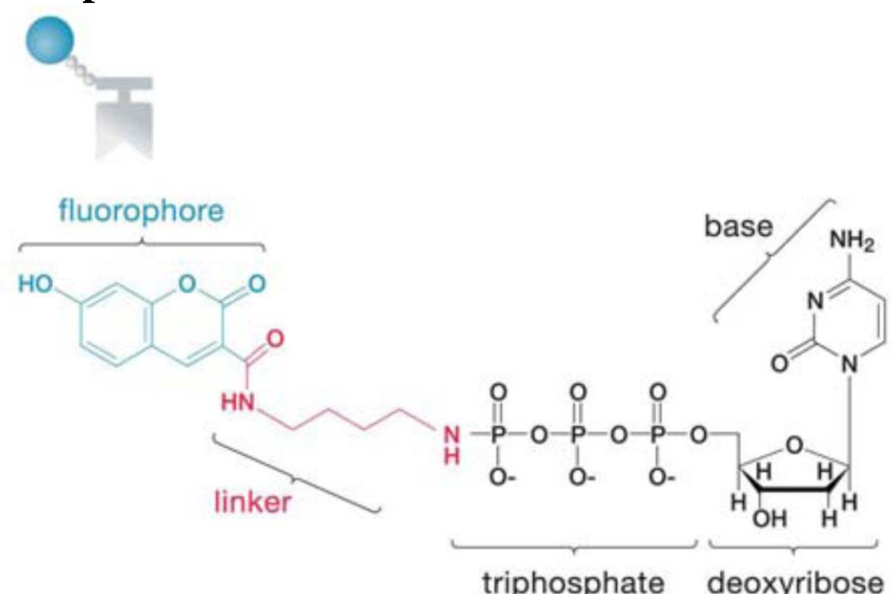
Title: Pacific Biosciences Develops Transformative DNA Sequencing Technology: Single Molecule Real Time (SMRT) DNA Sequencing



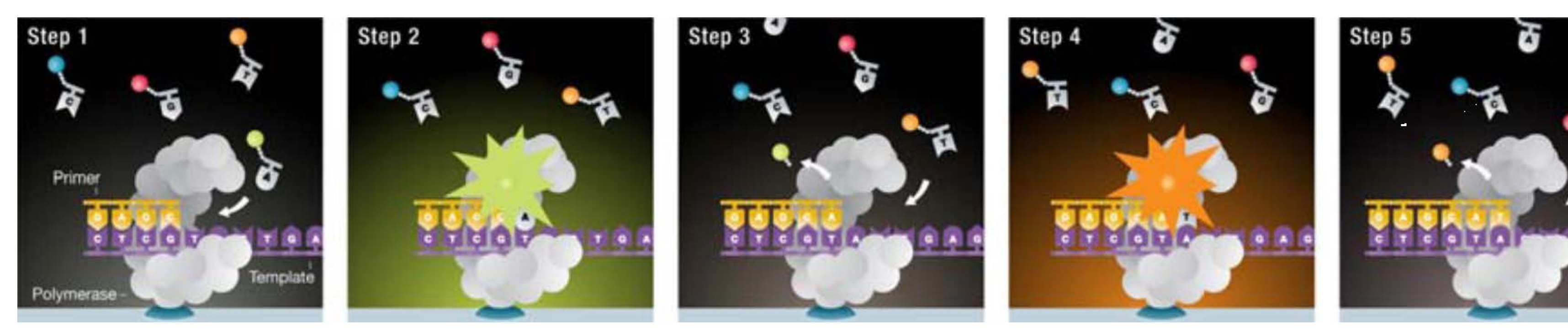
ZMW (Zero-mode waveguide) with Φ29 DNA polymerase and DNA template

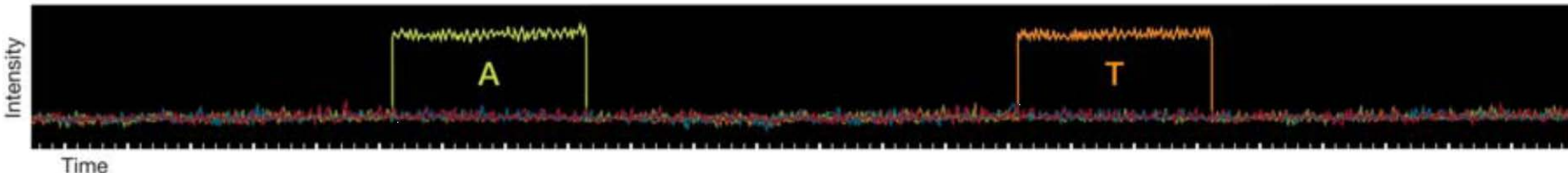
Base-labeled dNTP

Phospho-labeled dNTP

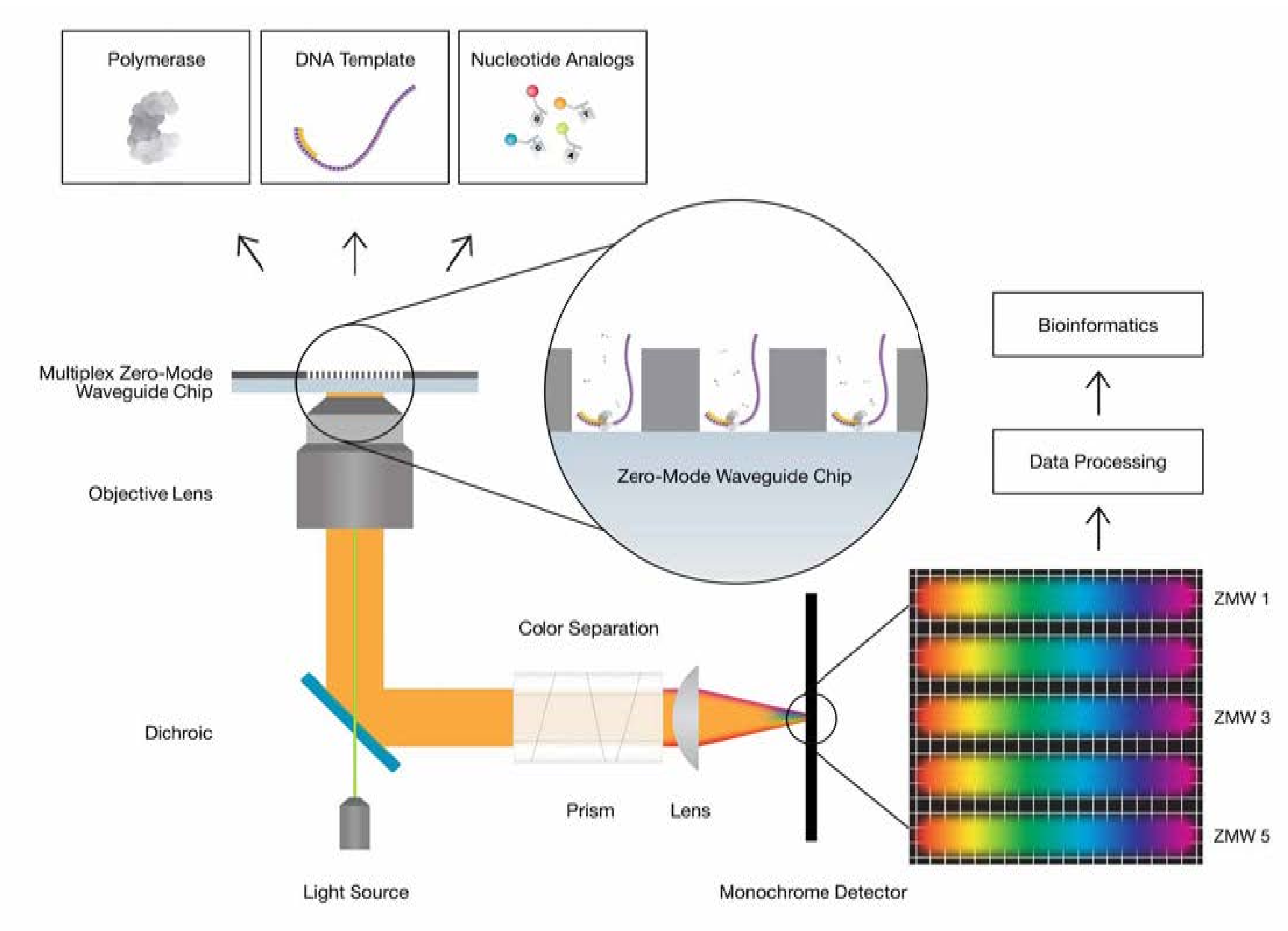


Single Polymerase DNA Sequencing

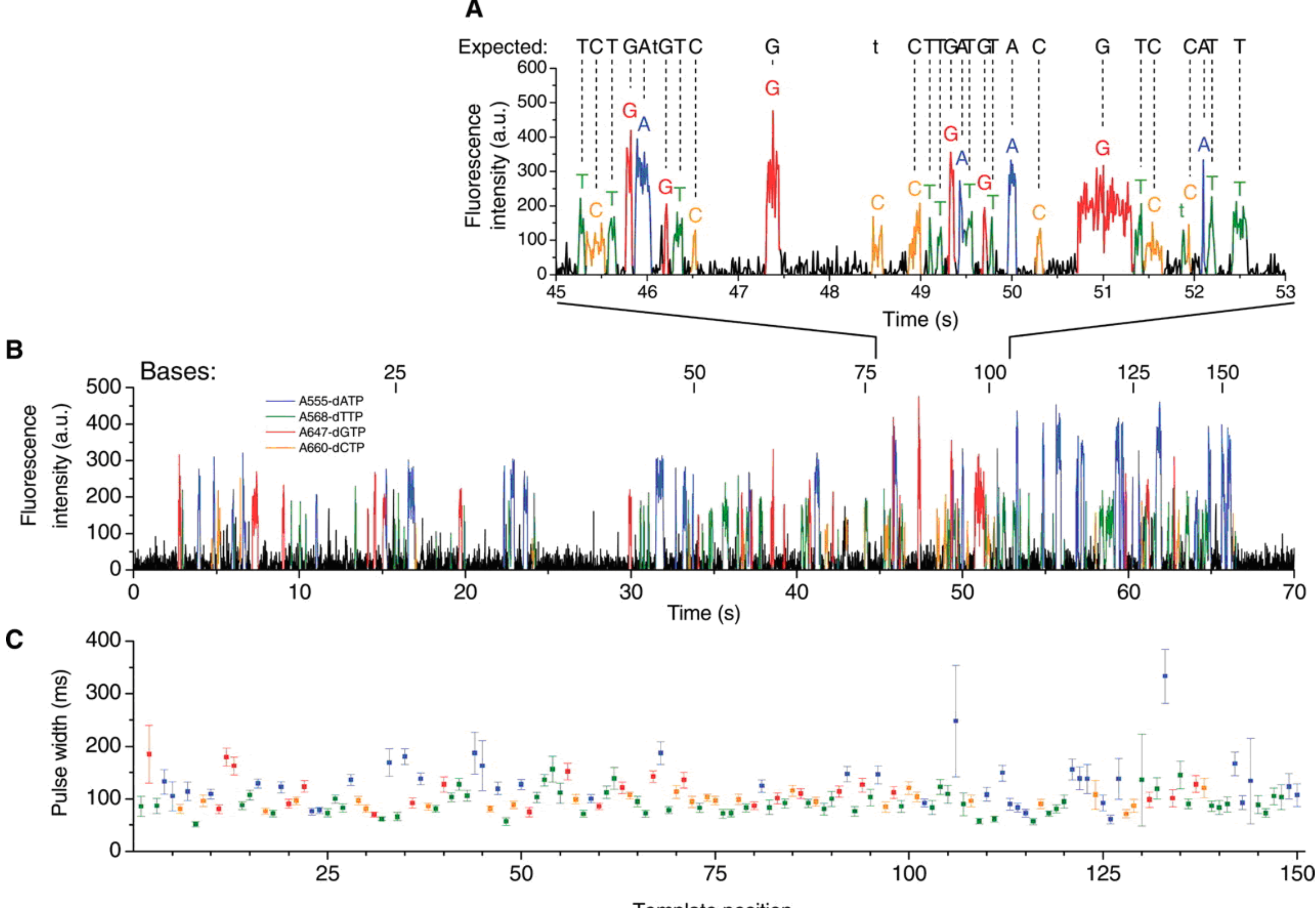




Step 1: Fluorescent phospholinked labeled nucleotides are introduced into the ZMW.
Step 2: The base being incorporated is held in the detection volume for tens of milliseconds, producing a bright flash of light.
Step 3: The phosphate chain is cleaved, releasing the attached dye molecule.
Step 4-5: The process repeats.



Polymerase
DNA Template
Nucleotide Analogs
Multiplex Zero-Mode
Waveguide Chip
Zero-Mode Waveguide Chip
Objective Lens
Bioinformatics
Data Processing
Color Separation
Dichroic
Prism
Lens
Light Source
Monochrome Detector
ZMW 1
ZMW 3
ZMW 5


A
Expected:
Fluorescence intensity (a.u.)
Time (s)
B
Bases:
A555-dATP
A568-dTTP
A647-dGTP
A660-dCTP
Fluorescence intensity (a.u.)
Time (s)
C
Pulse width (ms)
Template position

Original Potential Advantages

Speed

- 10 nt per second

Length

- 1000-2000 nt
- This is a claim
    - Not fully proven

Assembly

- Much easier to assembly longer fragments

Cost

- Company claim
    - $10/human genome

2010

- Only a few of these machines have been delivered
    - ***Advantages remain to be determined fully***

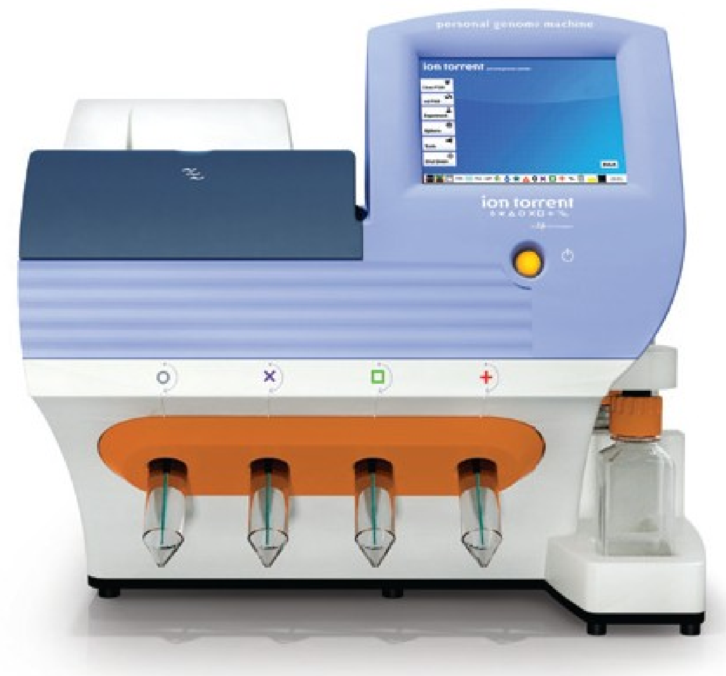
2017 Results

- Seeing 30 kb reads today
    - Company is increasing throughput with new machine
- Many genomes are being sequenced by this technology
    - This technology solves the problem of assembling repeat regions
    - Very cost effective
        - Today
            - $35,000 genomes primarily based on PacBio sequence data

Other Devices Available Today

Illumina MiSeq Personal Sequencer (Sequencing-by-Synthesis)

Ion Torrent (pH Sequencing)



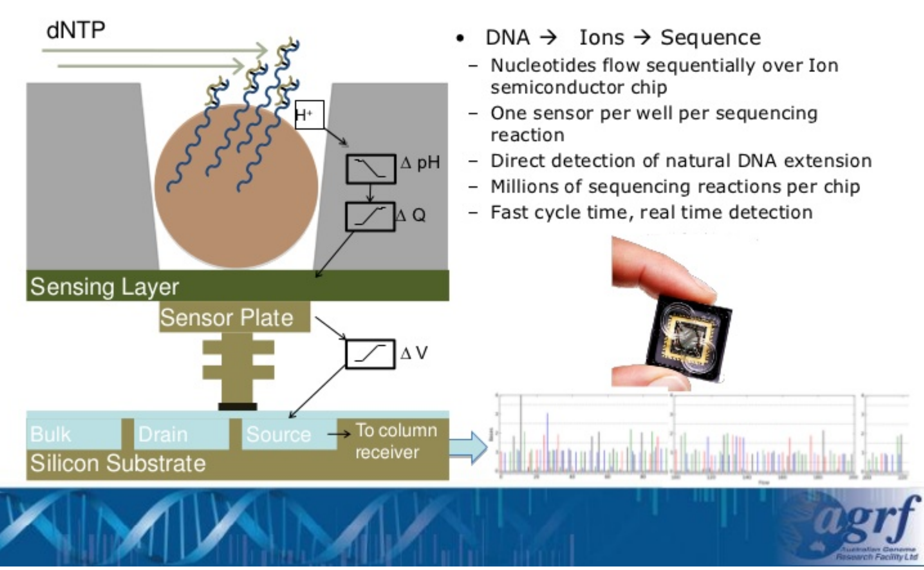
Semi-conductor Sequencing
Ion Torrent/Proton Torrent
(Life Technologies)

Chemistry

- Similar to 454 technologies
- Template bead is in an emulsion bead that is loaded into a single well
- Sensor detects the release of a hydrogen ion as a nucleotide is inserted into the growing polynucleotide
- Differs from 454 which detects the release of pyrophosphate
    - Nucleotides sequentially washed across the chip
    - If nucleotide is inserted, H+ released and pH changes
    - Semi-conductor detects pH changes
    - pH changes indicates number of nucleotides inserted
        - no signal, that base not next in the sequence
        - 2x signal; two of the same base in order

Chip consists of three layers

- Microwells that hold DNA template
- Ion sensitive layer
- Ion detector
    - Small pH meter

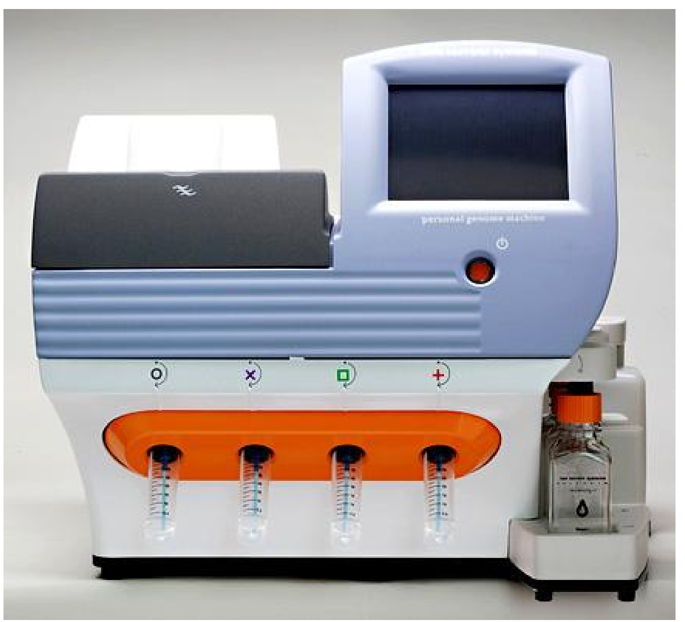


Semiconductor Sequencing Chips

Chip Types[1]	314	316	318	IP1/IP2/IP3*
# Wells per Chip	1,262,528	6,348,216	11,302,473	165 M/660M/1.2B
Volume, µL	7	30	30	55
# of Reads[1]	295,736	1,592,020	4,580,123	124-496,000,000
Yield/Q20, bases	24.6/**21.9** Mb	146.7/**122.5** Mb	600/**500** Mb	**10** / **60** / **480** Gb
Mean Read[1], bp	83	92	129	Up to 300
Longest Reads[1]	396	307	386	640
Run Time[1], Hrs	2.4	3.1	4.5	~4
Processing, Hrs[1]	0.3	2.0	4.5	Up to 8 hrs
Analysis[2], Hrs	12	18	30	Up to 1 day
Template Molecules	2.5×10^7	5×10^7	5×10^7	2.5×10^7
Cost per Run	$400	$500	$800	$1,000

Features of the Life Technologies Instruments

Feature	Ion Torrent PGM	Proton Torrent
Application	Small set of genes Small genomes Gene expression	Whole large genomes Transcriptomes
Throughput	10 Mb - 1 Gb	10 Gb
Read length	35-400 bp	200 bp
# Reads passing filter		60-80 million
Sequencing run time	1.5 hr (100 bp reads)	2-4 hr
Chips	• 314 (1 million wells, 10 Mb) • 316 (6 million wells, 100 Mb) • 318 (11 million wells, 1 Gb)	• Proton I (165 million wells; ***2 human exomes***) • Proton II (600 million wells; ***1 human genome***



ION TORRENT

PROTON TORRENT

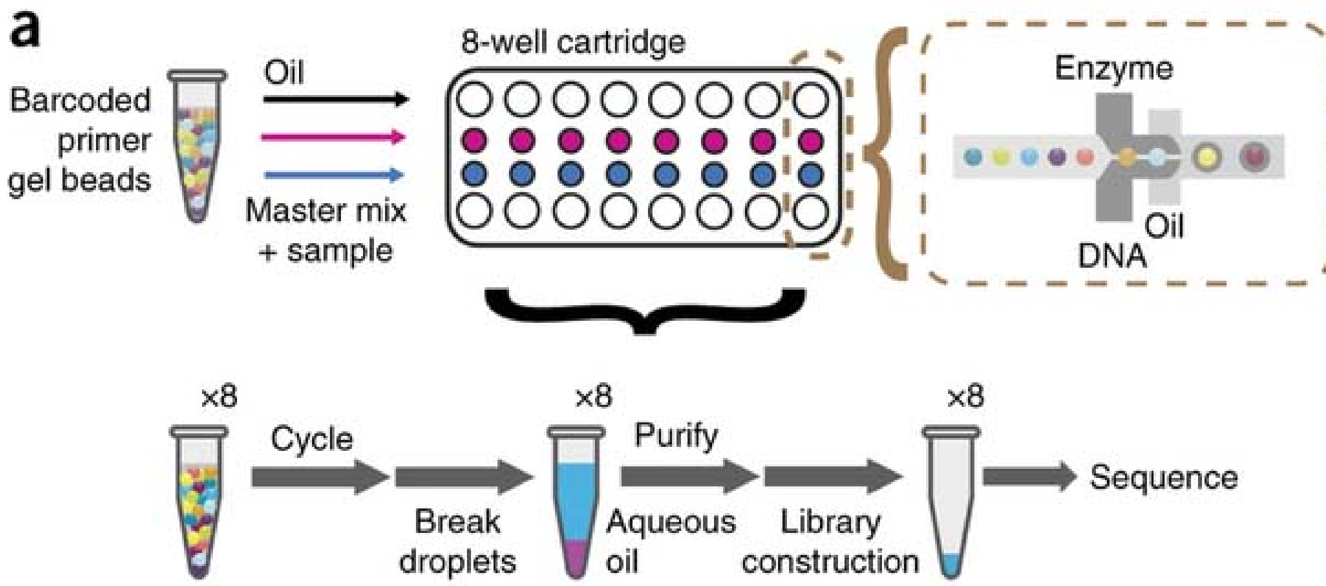
Alternative Library Preparation Protocols

10X Genomics
Synthetic Long Read Seqeuencing

Addresses assembly problems

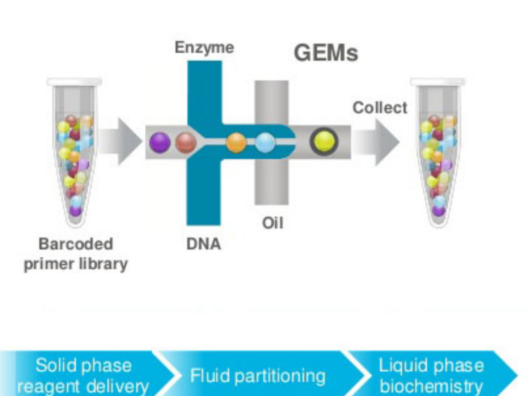
- Hard to link short reads (Illumina) into “long” contigs
- Longer contigs make final assembly of contigs into scaffolds easier
- Goal of 10X Genomics sequence library preparation:
    - Collect reads associated with a single genomic locations and assembly them into “artificial long reads”

Workflow



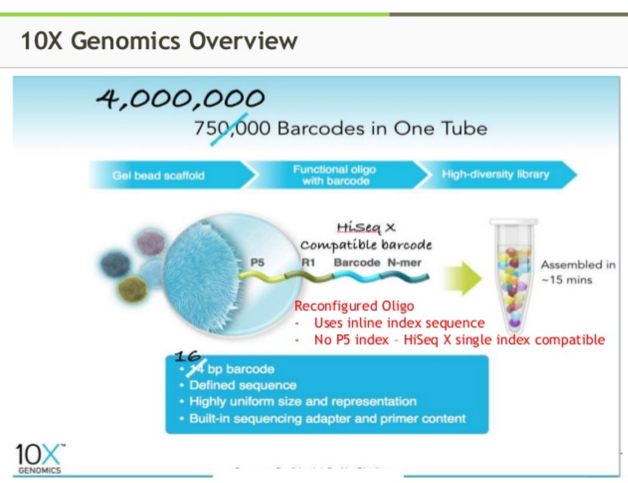
Methodology

- DNA fragment (up to ~100kb) is loaded into an emulsion bead similar to 454 technology
- Emulsion bead created by fluidic partioning device



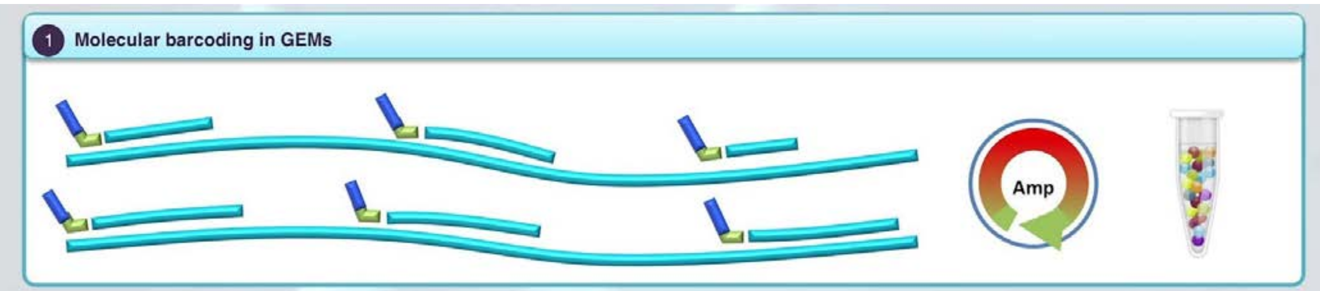
Emulsion micells (called GEM) contains

- Anchor bead
- DNA
- One 16bp barcode (out of ~4 million)
- Enzyme to fragment DNA??
- dNTPS



Isothermal Amplification

- Multiple fragments from a long fragment are amplified
- All contain the same barcode



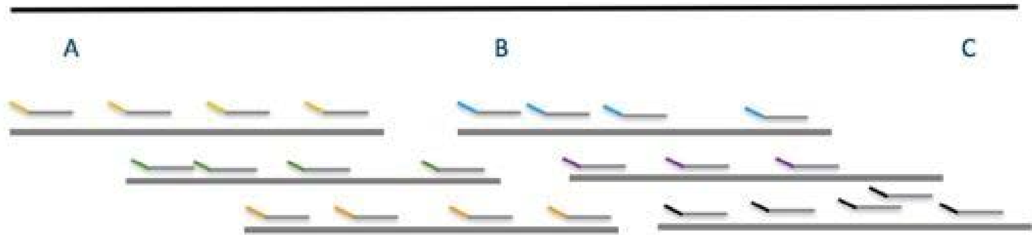
GEM pooling

- All GEM are pooled
- GEMs emulsion dissolved and all barcoded fragments pooled

Sequencing and assembly

- Illumina sequencing library developed with the barcoded fragments
- Library sequenced using Illumina technology
- Reads are portioned based on barcodes
- Contigs assembled from each barcode pool
- Contigs across barcode pools assembled into longer conting.

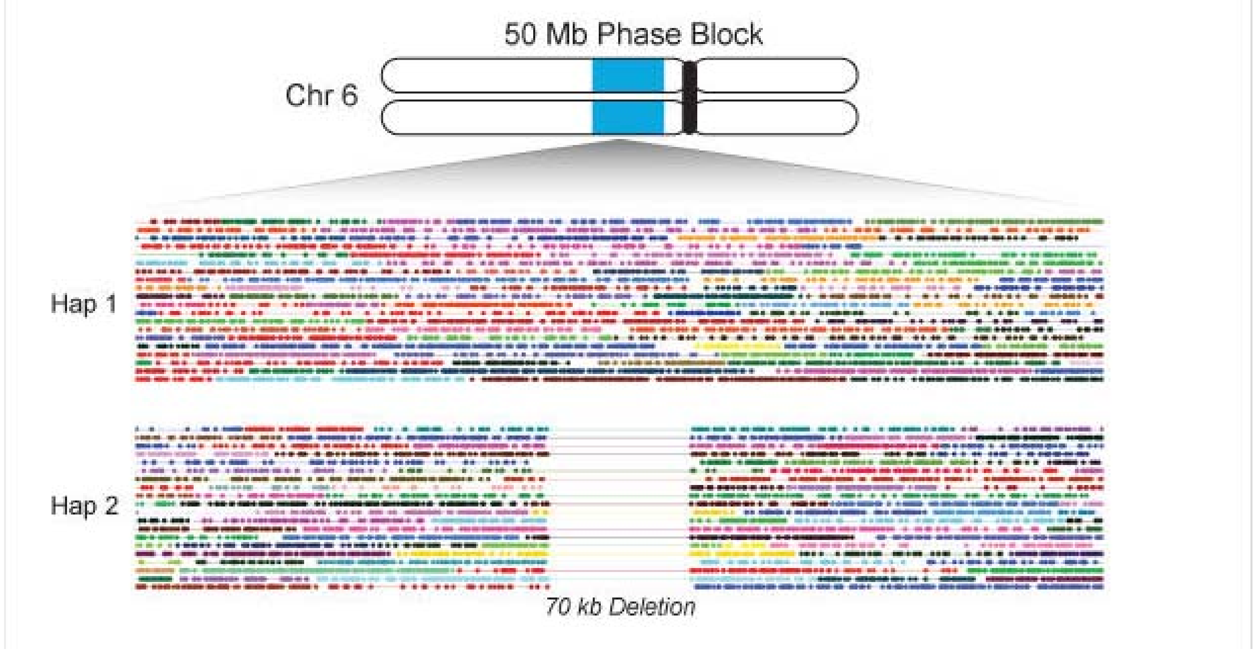
Assembled synthetic molecule (from 10X, Deanna Church)



- Note different colors on ends and that each original fragment is not completely sequenced.
- Multiple barcoded fragments pooled to create overlapping contig

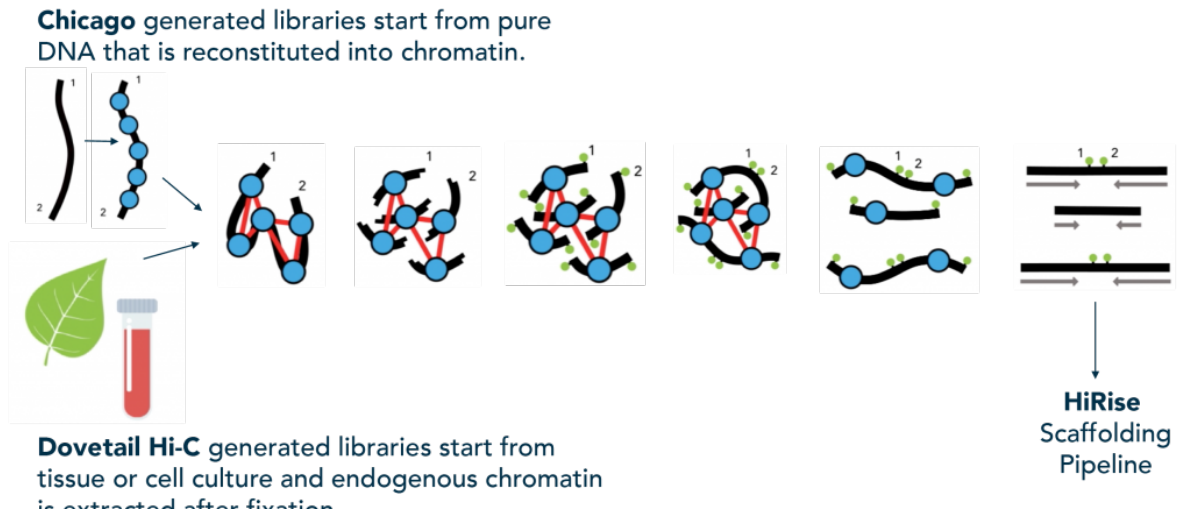
Variant Discovery

The two allelic regions of an individual differ by a deletion



Human Trio Analysis

Dovetail Genomics Sequencing



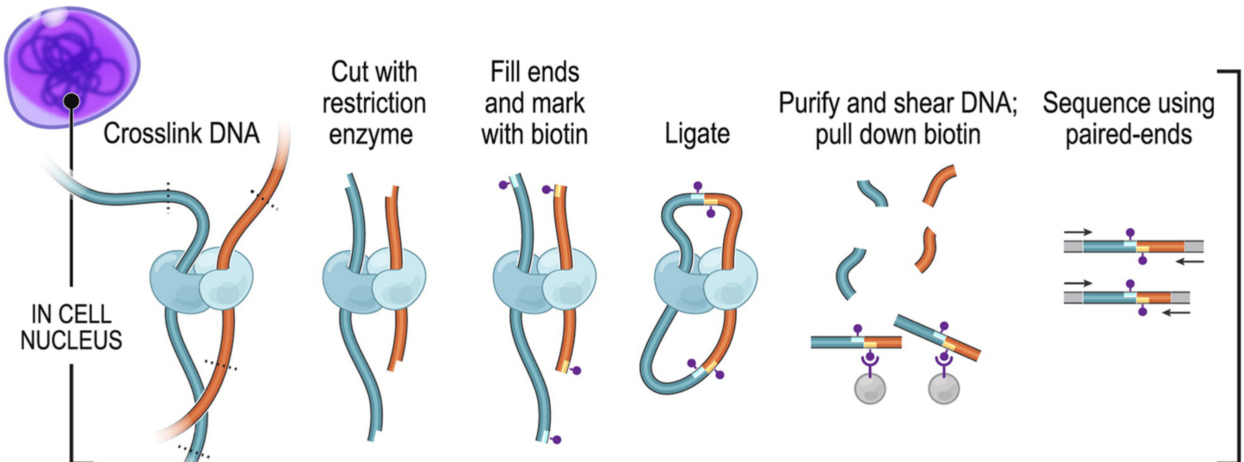
from: https://dovetailgenomics.com/technology/

Hi-C linking

- Based on links between natural interactions within a chromosome
    - Regions of the chromosome are associated via chromatin
    - Based on principal that DNA has a 3-D confirmation in the cell
    - 3D configuration occurs because controlling elements that regulate a gene's expression are not always immediately adjacent to coding region of the gene

Chicago

- An artificial linking procedure
- When used with Hi-C, the Hi-C derived relationships can be confirmed



from: http://science.sciencemag.org/content/326/5950/289/tab-pdf

Hi-C procedure

1. Crosslink the cells using formaldehyde to stick chromosomes together
2. Isolate “crosslinked” DNA bound with chromatin
3. Digest DNA with six-cutter restriction enzyme
4. Fill ends and add biotin to end
5. Ligate ends and pull down molecules with biotin procedure
6. Sequence pull down library using Illumina paired-end protocol

Assembly

- Long distance relationships can be used during assembly
- Distances between ends are typically >20Kb
- Data can be used in the final steps of assembly.

Illumina Paired End Sequencing

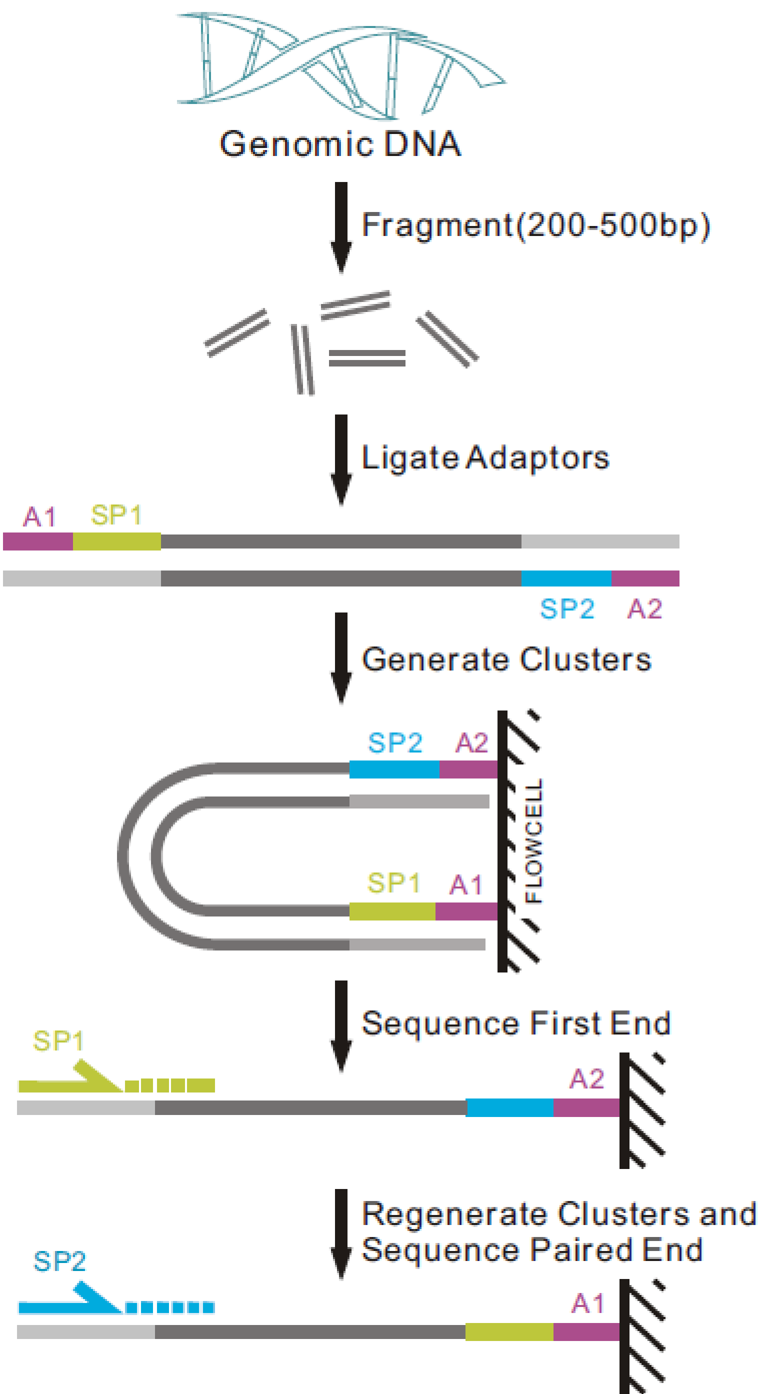
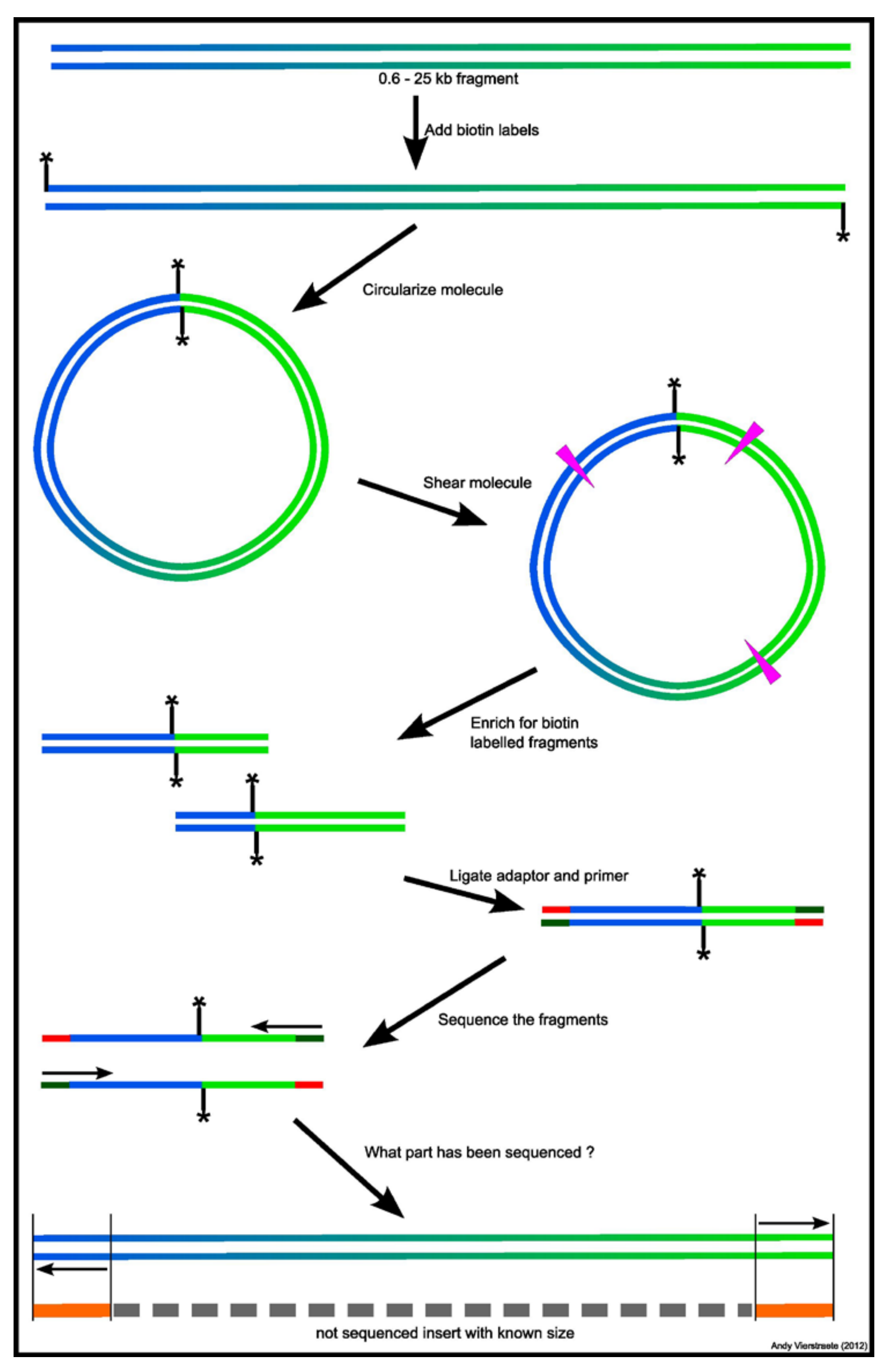


Figure 1-2-1 Pipeline of paired-end sequencing (www.illumina.com)

Illumina Mate Pair Library Sequencing



General Steps That Apply To ALL
Massively Parallel DNA Sequencing Systems

1. Isolate DNA

- Care is needed to ensure the DNA is of uniform high quality

2. Fractionate DNA into appropriate size for specific sequencing system

- Length will vary depending on the read length you will be generating

3. Amplify individual DNA fragments that will be sequenced

- This could be in a reaction emulsion bead (Roche 454) or reaction matrix (Illumina or Pacific Biological Science [PacBio])

4. Load DNA samples onto DNA sequencing matrix

- The matrix can be a solid chip with individual wells (Roche 454, PacBio) or a chip with sequencing oligonucleotides (Illumina)

5. Perform sequencing reactions

- Varies from system to system

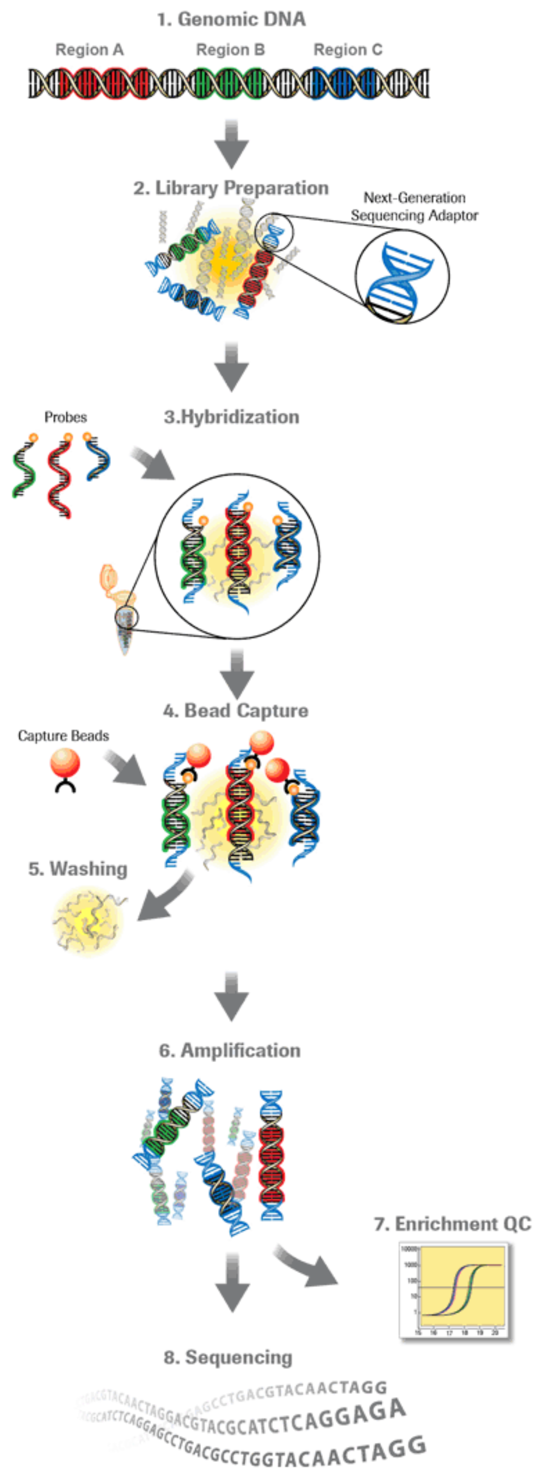
6. Collect DNA sequence data for each read

- Varies from system to system

Sequencing Only a Subset of the Genome

- Genes are the functional component of much of the genome
- They encode proteins that have the following functions:
  - Enzymatic
  - Regulatory
  - Structural
- Genes only make up a portion of the genome
  - Human: 3% of the length of the genome is gene sequence
    - 5'-UTR (untranslated regions)
    - Exons
    - Introns
    - 3'-UTR
- Many of the functional mutations are in the exon (coding) region of the genes
- Exons are only a portion of the gene sequence
  - They make up only a still smaller portion of the genome
  - Humans: 1% of the length of the genome is exons

- Goal of exon sequencing
  - Discover sequence changes just in the exon
  - Associate the sequence change with change in function



From: NimbleGen

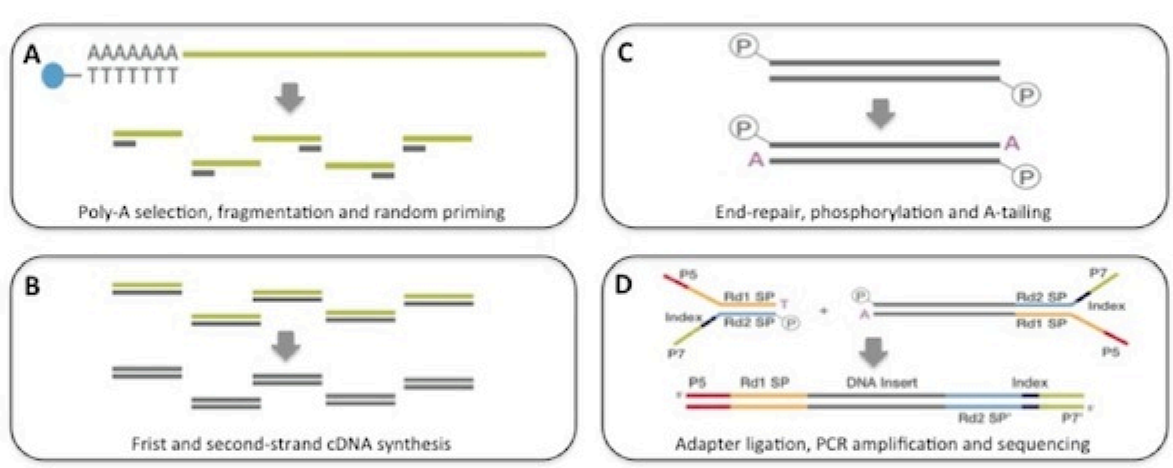
Sequence Capture Protocol (NimbleGen)

1. **Genomic DNA**: SeqCap EZ Oligo pool is made against target regions in the genome.
2. **Library Preparation**: Standard shot-gun sequencing library is made from genomic DNA.
3. **Hybridization**: The sequencing library is hybridized to the SeqCap EZ Oligo pool.
4. **Bead Capture**: Capture beads are used to pull down the complex of capture oligos and genomic DNA fragments.
5. **Washing**: Unbound fragments are removed by washing.
6. **Amplification**: Enriched fragment pool is amplified by PCR.
7. **Enrichment QC**: The success of enrichment is measured by qPCR at control loci.
8. **Sequencing-Ready DNA**: The end product is a sequencing library enriched for target regions, ready for high throughput sequencing.

Sequencing the Expressed Portion of the Genome

- Genes are expressed in a the following manners
    - Tissue-specific (where)
    - Temporal specific (when)
    - Quantitatively (how much)
- Transcriptomics
    - The study of gene expression
- Massively parallel sequencing has changed the study of the transcriptome
    - All the genes at a specific place or time can be accurately quantified
- Procedure
    - RNA-seq or massively parallel RNA sequencing
        - Very powerful
        - Can monitor expression of even rarely expressed genes

Illumina Tru-Seq RNA-seq protocol



Library prep begins from 100ng-1ug of Total RNA which is poly-A selected (A) with magnetic beads. Double-stranded cDNA (B) is phosphorylated and A-tailed (C) ready for adapter ligation. The library is PCR amplified (D) ready for clustering and sequencing.

RNA-seq procedure

1. Isolate RNA from target tissue
2. Select mRNA using poly-T primers
   - Based on principle that all mRNA in eukaryotes have a poly-tail
3. Perform first and second strand cDNA (copy DNA) synthesis to convert mRNA into cDNA
4. Prepare cDNA for sequencing by adding appropriate sequencing adaptors
5. Sequence the cDNA pool using a massively parallel technology
6. Align reads against a reference genome and quantify

Aligning RNA-seq Data to the Reference Sequence