

# Sequencing Technologies

What is Next Generation Sequencing  
and  
What is it used for?

*Brandi Cantarel, PhD*  
*BICF*  
*08/01/2016*

# BICF GWAS Nano Courses

Bioinformatics GWAS Nano Course  
May 19th-20th 2016

Bioinformatics RNASeq Analysis using R  
December 6th-7th 2016

<http://www.utsouthwestern.edu/labs/bioinformatics/>

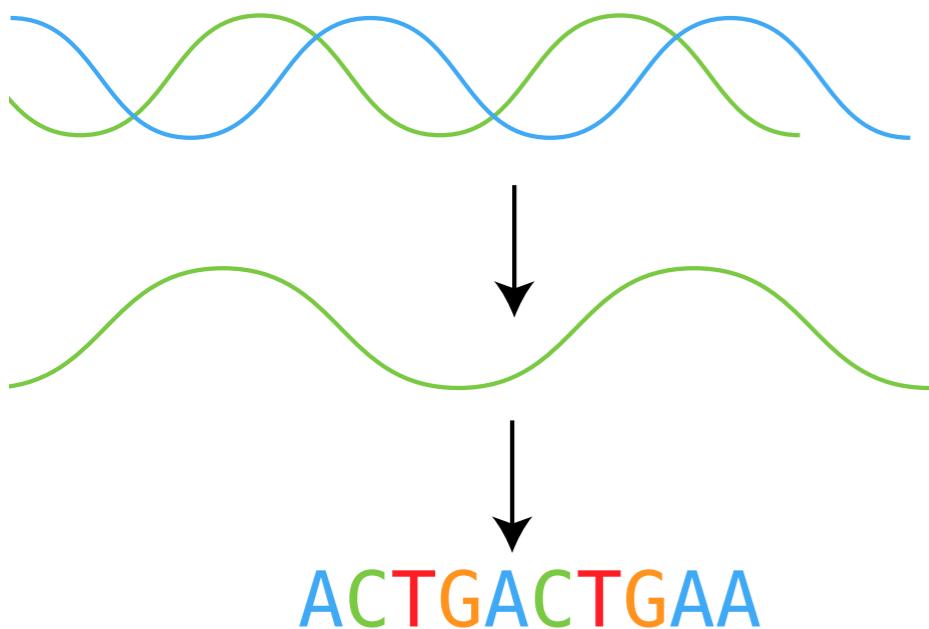
- Sequencing Methods
  - Sanger
  - Pyro
  - Solexa
  - SMRT
  - Field
- Sequencing Applications
  - DNA
  - RNA
  - Chromatin Sequencing
    - ChIP
    - ATAC
    - DNaseq
    - MNaseSeq
  - Clip
  - Ribo
  - Single-Cell
- Alternatives to Sequencing
  - Microarray
  - Nano string

# History of Sequence Similarity

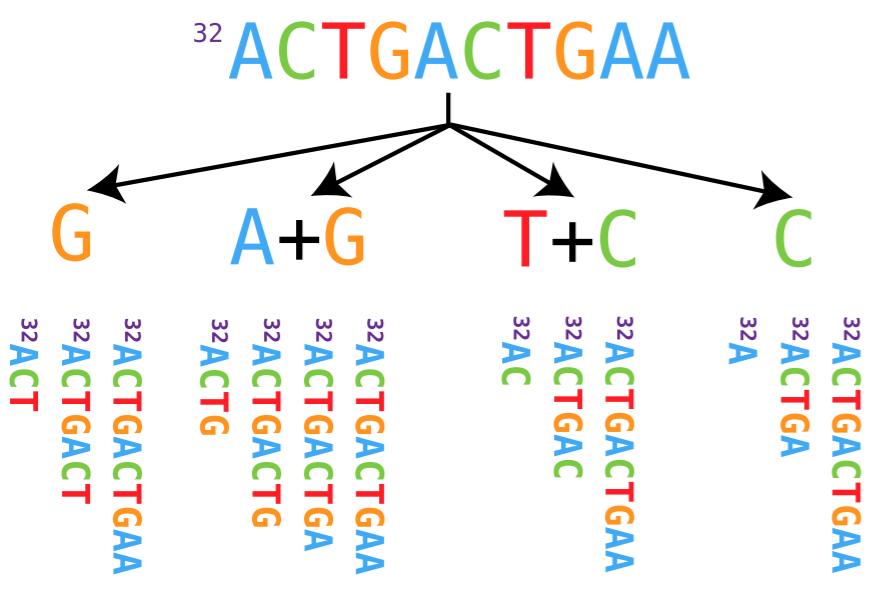
---

Watson and Crick solve the three-dimensional structure of DNA, working from crystallographic data produced by Rosalind Franklin and Maurice Wilkins	1953
Robert Holley and colleagues were able to produce the first whole nucleic acid sequence; Fred Sanger and colleagues developed a related technique based on the detection of radiolabelled partial-digestion fragments after two-dimensional fractionation	1965
Using Sanger's method, Walter Fiers' laboratory was able to produce the first complete protein-coding gene sequence	1972
Alan Maxam and Walter Gilbert developed an early DNA sequencing method using nucleic acid sequencing based on the detection of radiolabelled	1973
Frederick Sanger and colleagues develop a method of sequencing involving random termination	1977
Pål Nyrén and colleagues develop pyro-sequencing	1996
Shankar Balasubramanian and David Kleneman, and developed a sequencing method using reversible dye-terminators technology, and engineered polymerases.	1998
Stephen Turner and colleagues describe single molecule sequencing (SMRT)	2008
Oxford Nanopore introduces the MinION at AGBT	2012

# Maxim Gilbert



1) Obtain single stranded DNA

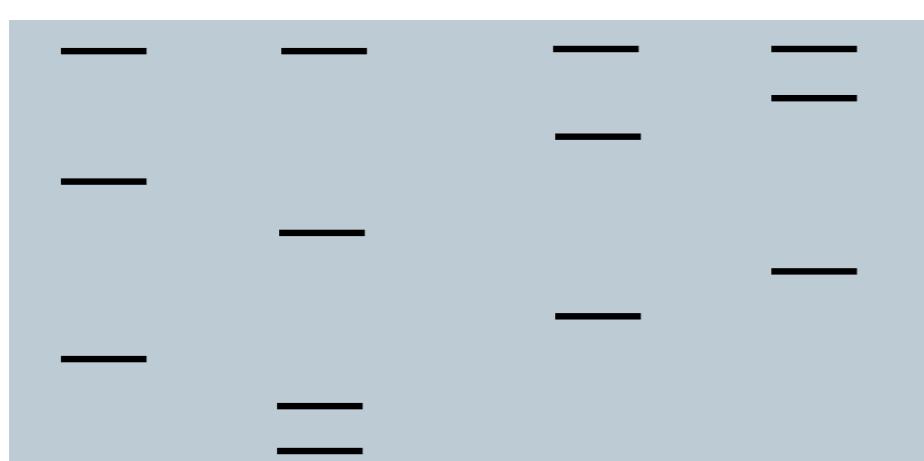


2) Add a  $^{32}\text{P}$  to 5' end

3) Cleave at specific nucleotides

4) Differently sized DNA strands

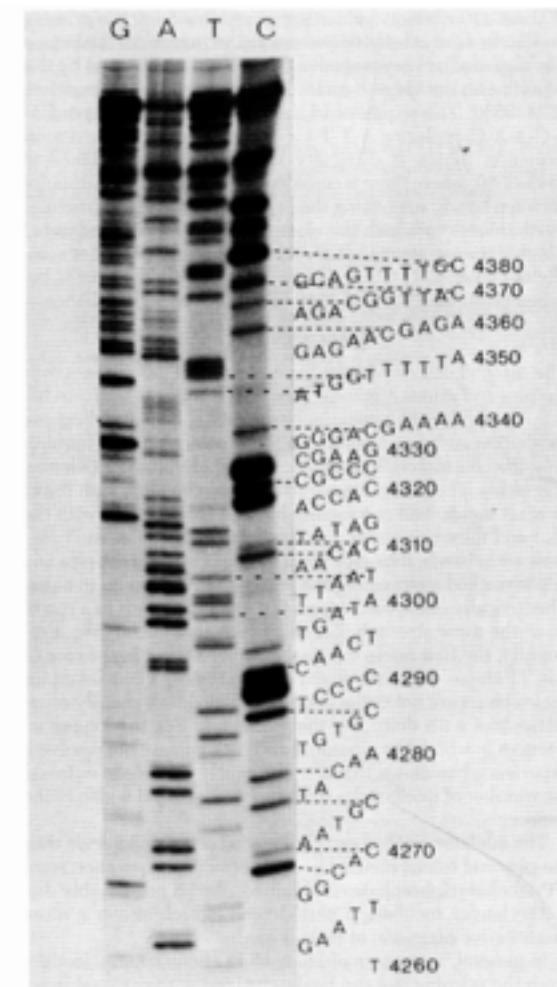
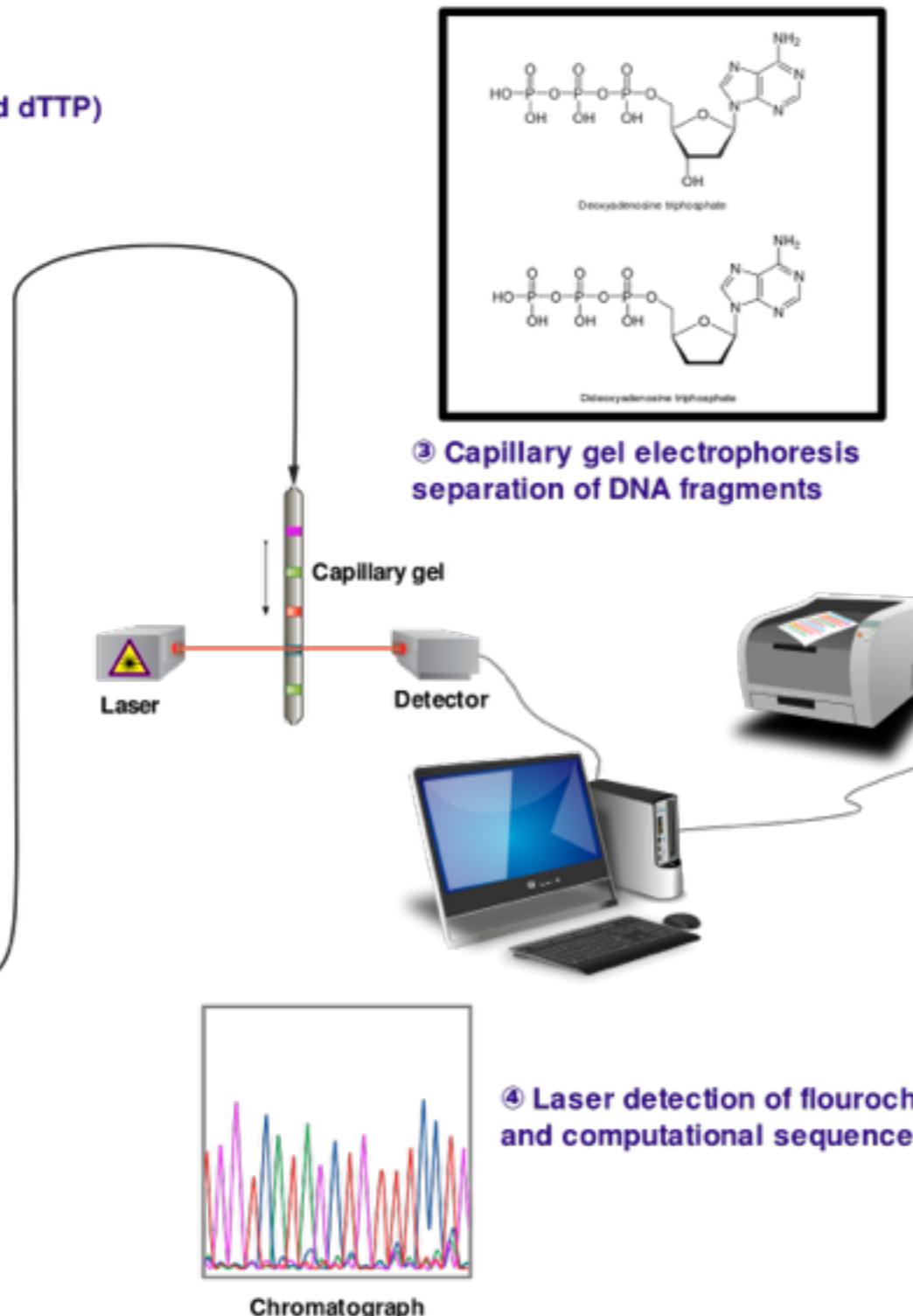
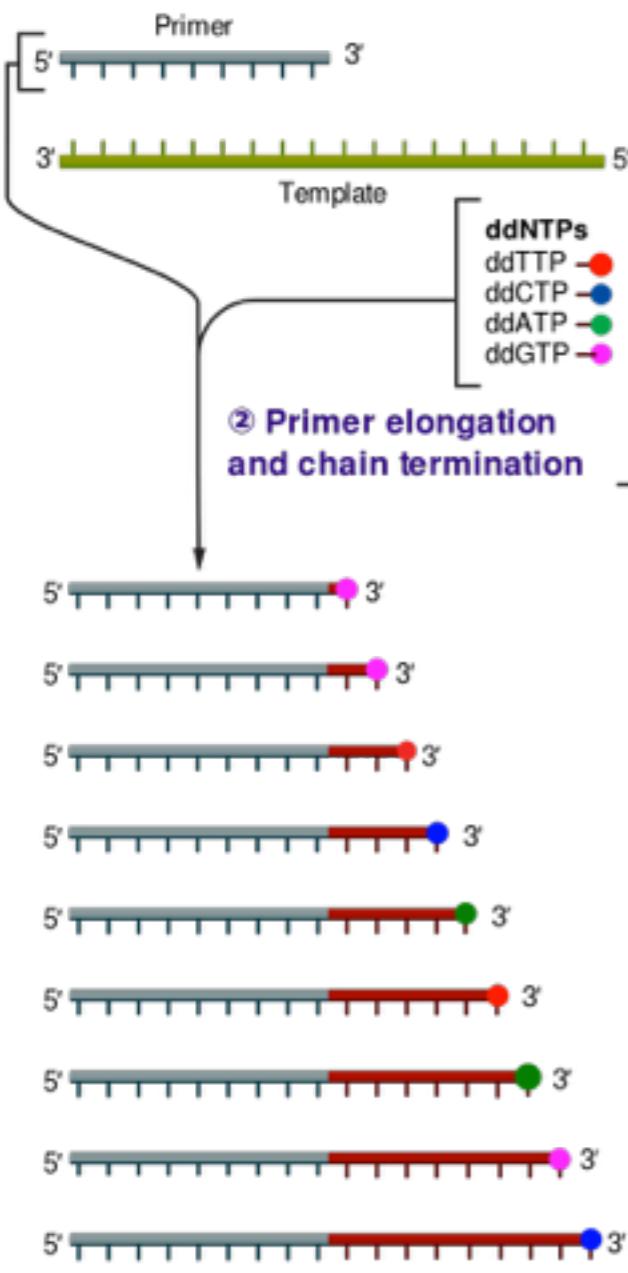
5) Electrophoresis through  
high resolution acrylamide gels



6) Deduce DNA sequence

# Sanger Sequencing

- ① Reaction mixture  
► Primer and DNA template ► DNA polymerase  
► ddNTPs with flourochromes ► dNTPs (dATP, dCTP, dGTP, and dTTP)



DNA Sequencing with chain-terminating inhibitors

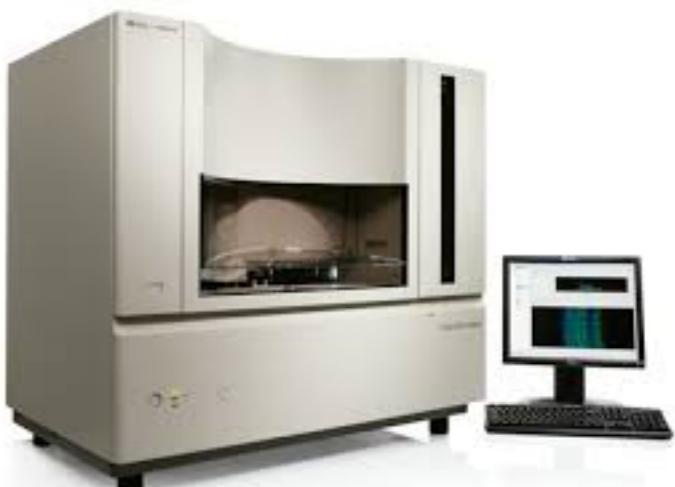
# Sanger Sequencing

- Sanger sequencing is still in many ways considered the “gold standard” for accuracy.
- Costs for sequencing samples in 96 well plates average about \$5 a read, or about \$10 per kb (1000 bases of DNA)



# When to use Sanger?

- Single gene
- 1-100 amplicon targets
- Microsatellite or STR analysis



# UTSW Sanger Core

## Sanger Sequencing Core

McDermott Center

RESOURCES ABOUT US



About Us

### > About Us

The McDermott Center Sequencing Core facility provides DNA Sequencing Services for UT Southwestern Medical Center Investigators. Samples are sequenced using Life Technologies® (LT) Dye Terminator 3.1 chemistry and 3730XL Genetic Analyzers.

We are located in room NA2.138 on the North Campus; we maintain a drop-off box in L5.268 on the South Campus.

### Facility Location

Eugene McDermott Center for Human Growth and Development  
Nancy B. and Jake L. Hamon Biomedical Research Building  
North Campus/NA2.132  
6000 Harry Hines Blvd.  
Dallas, TX 75390-8591  
Phone: 214-648-1894

### Staff Contacts

Ralf Kittler, Ph.D.

Director

Phone: 214-648-1661

[ralf.kittler@utsouthwestern.edu](mailto:ralf.kittler@utsouthwestern.edu)

Vanessa Schmid

Lab Manager

Phone: 214-648-5101

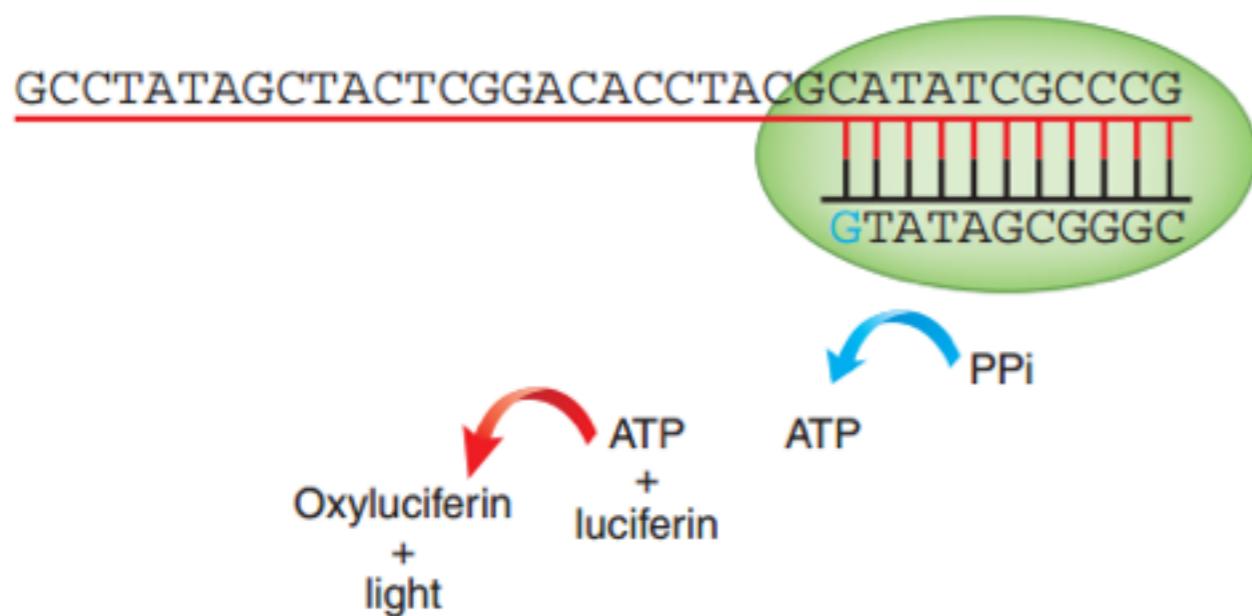
[vanessa.schmid@utsouthwestern.edu](mailto:vanessa.schmid@utsouthwestern.edu)

### Sanger Sequencing Staff

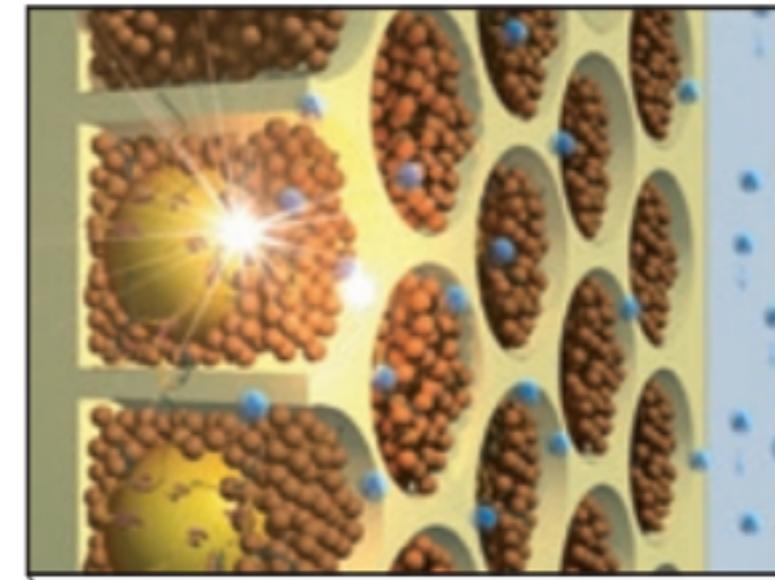


# Pyro-Sequencing

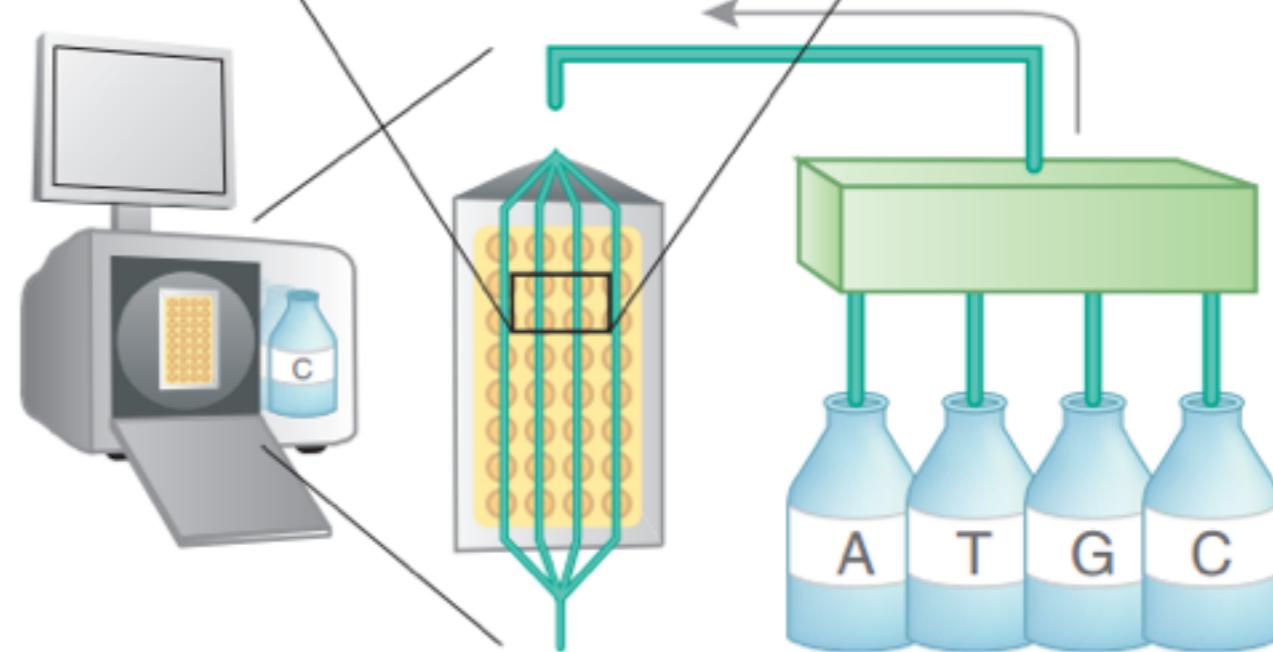
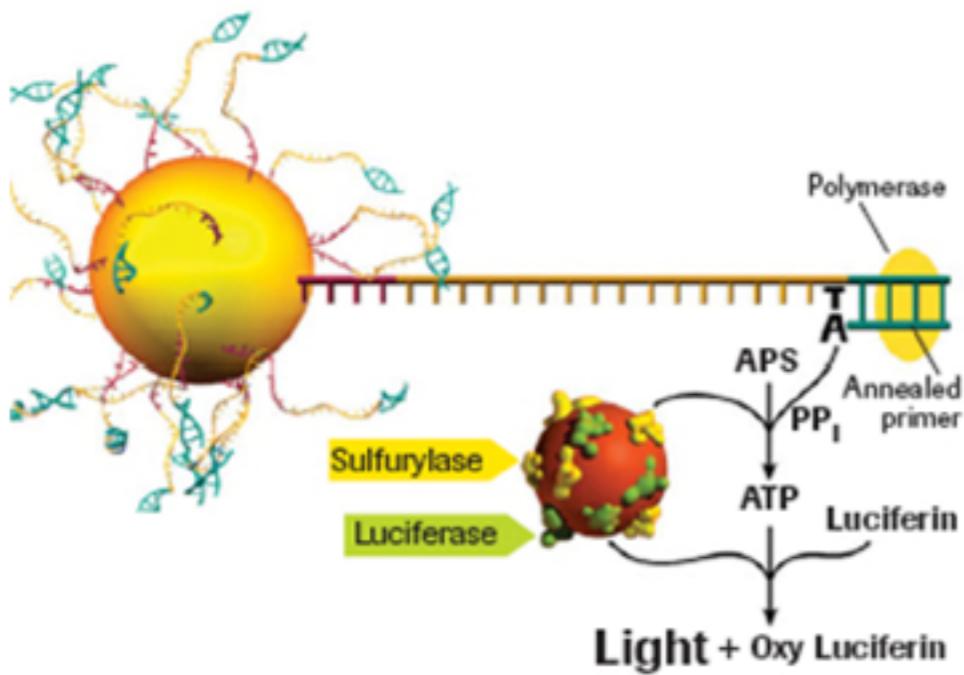
A



C



B

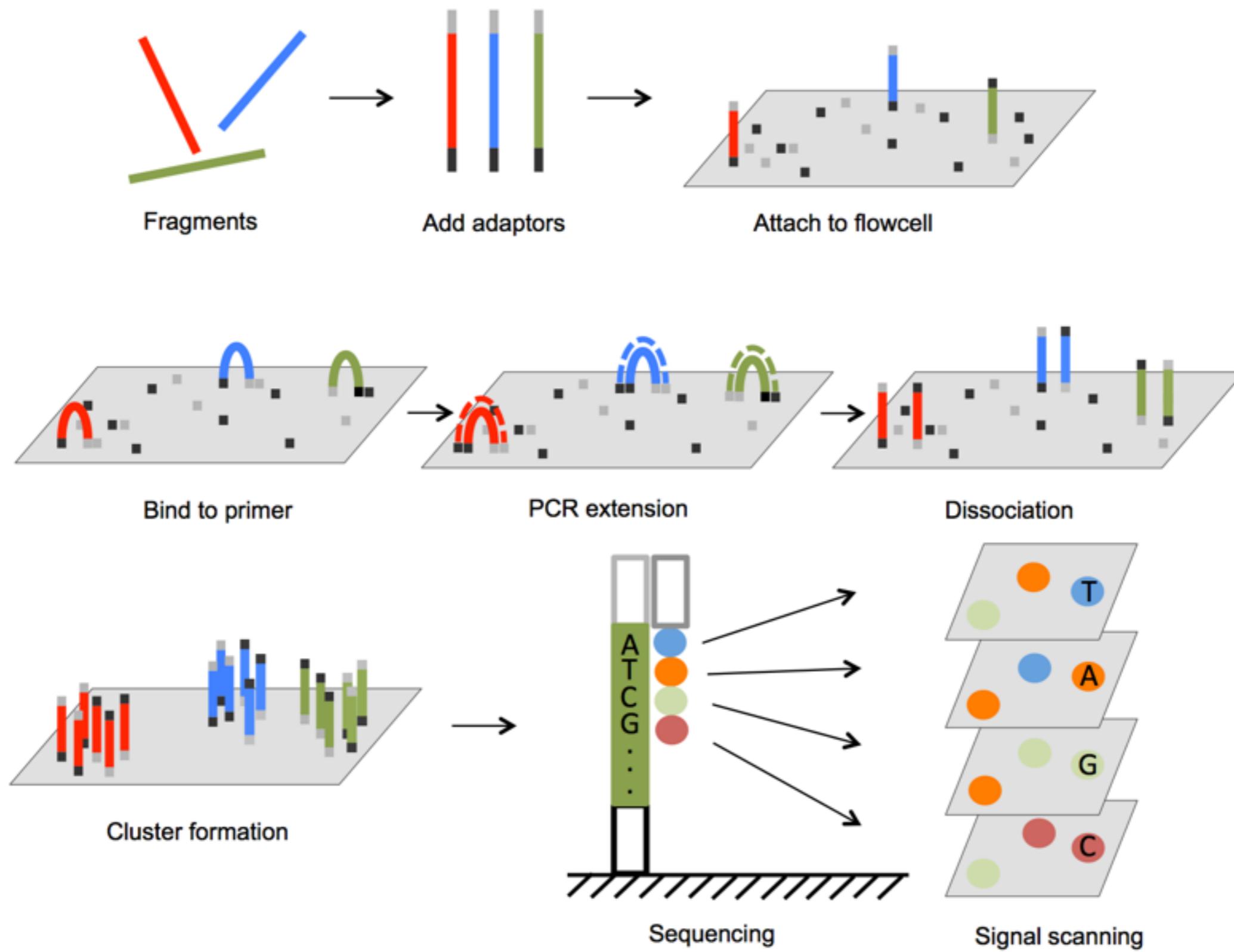


# Pyro-Sequencing

- Roche 454
  - Pyrosequencing (2004)
- Ion Torrent
  - 400bp reads
  - Inaccuracies accumulated in homopolymer regions
  - ~ \$0.63/Mbp — Hardware ~\$70K/machine
  - Low upfront and maintenance costs makes it attractive to independent labs



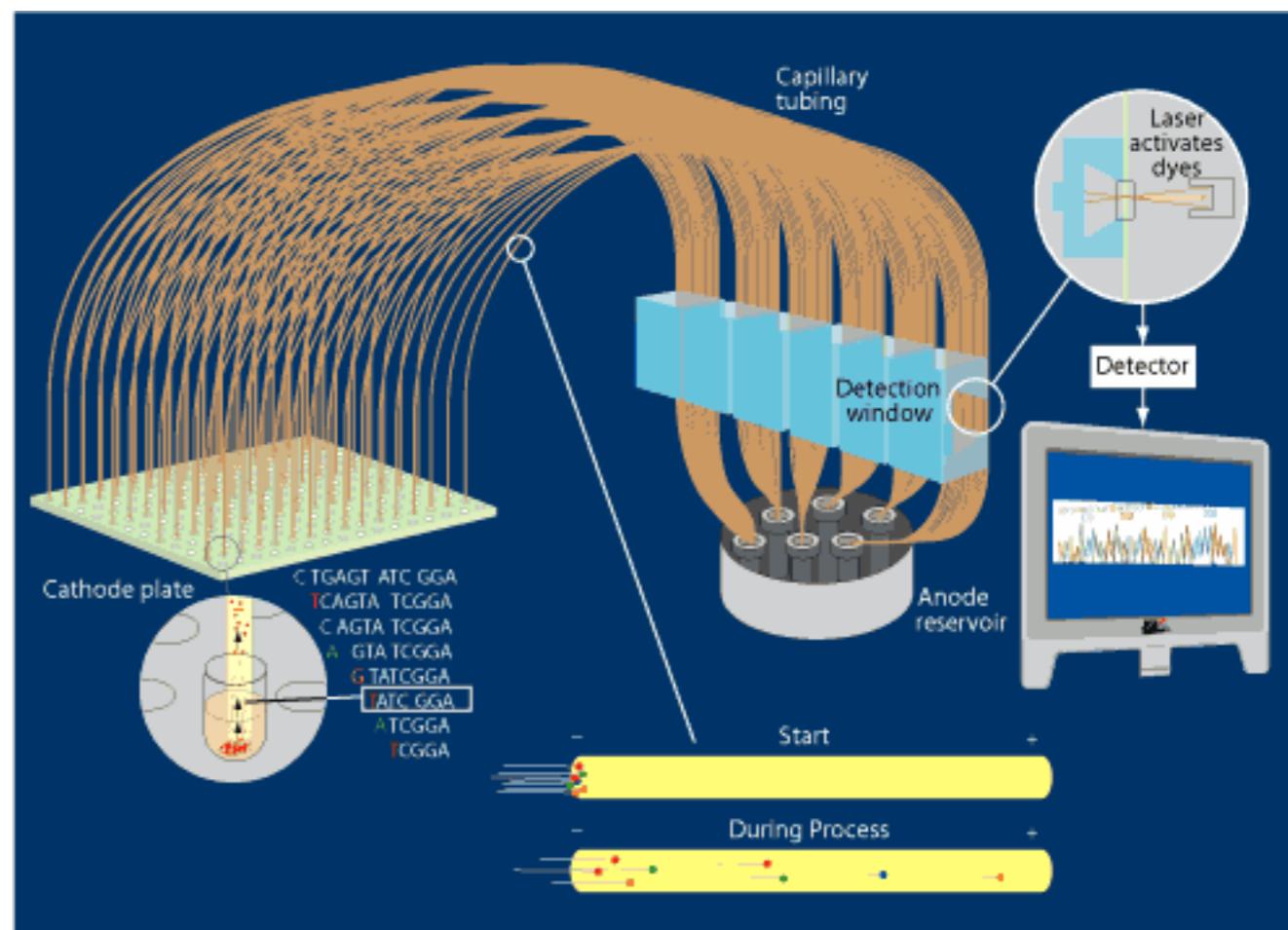
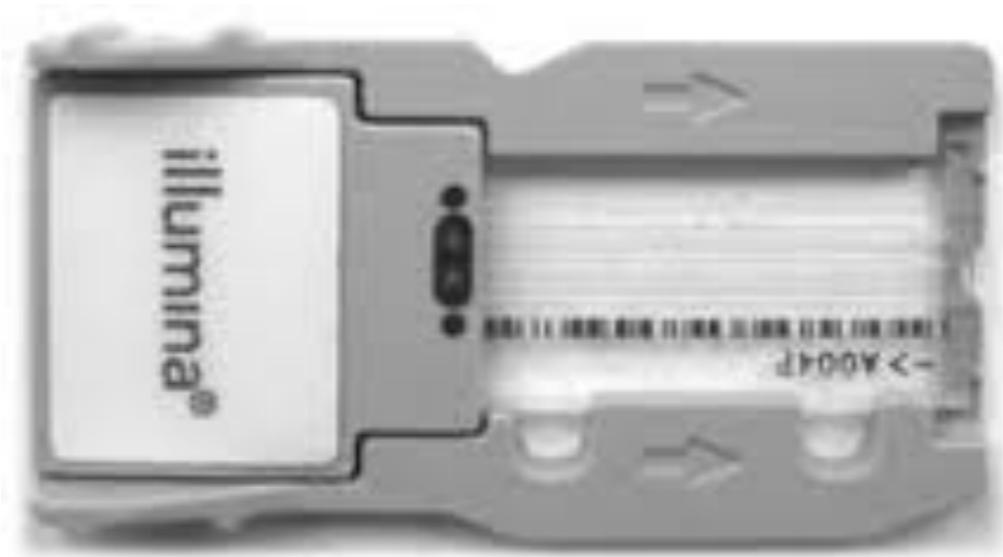
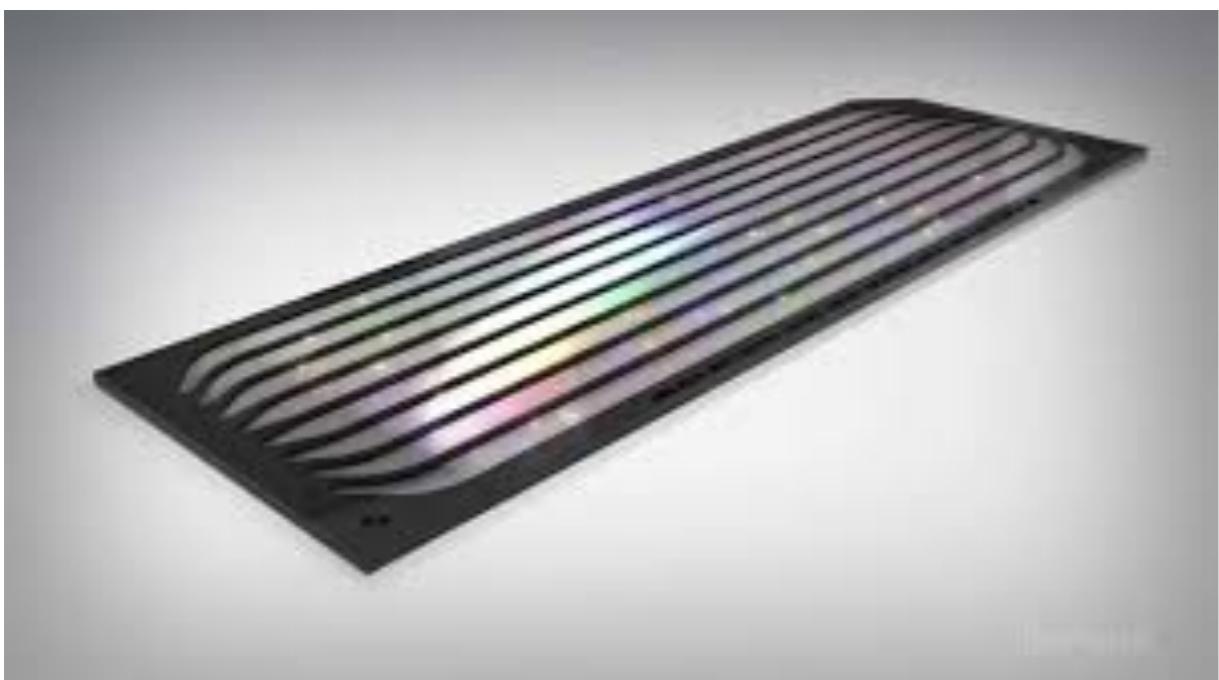
# Solexa-Sequencing



					
	MiniSeq System	MiSeq Series	NextSeq Series	HiSeq Series	HiSeq X Series*
<b>Key Methods</b>	Amplicon, targeted RNA, small RNA, and targeted gene panel sequencing.	Small genome, amplicon, and targeted gene panel sequencing.	Everyday exome, transcriptome, and targeted resequencing.	Production-scale genome, exome, transcriptome sequencing, and more.	Population- and production-scale whole-genome sequencing.
<b>Maximum Output</b>	7.5 Gb	15 Gb	120 Gb	1500 Gb	1800 Gb
<b>Maximum Reads per Run</b>	25 million	25 million <sup>†</sup>	400 million	5 billion	6 billion
<b>Maximum Read Length</b>	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp
<b>Run Time</b>	4–24 hours	4–55 hours	12–30 hours	<1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	<3 days
<b>Benchtop Sequencer</b>	Yes	Yes	Yes	No	No
<b>System Versions</b>	<ul style="list-style-type: none"> <li>• <a href="#">MiniSeq System</a> for low-throughput targeted DNA and RNA sequencing</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">MiSeq System</a> for targeted and small genome sequencing</li> <li>• <a href="#">MiSeq FGx System</a> for forensic genomics</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">NextSeq 500 System</a> for everyday genomics</li> <li>• <a href="#">NextSeq 550 System</a> for both sequencing and cytogenomic arrays</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">HiSeq 3000/HiSeq 4000 Systems</a> for production-scale genomics</li> <li>• <a href="#">HiSeq 2500 Systems</a> for large-scale genomics</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">HiSeq X Five System</a> for production-scale whole-genome sequencing</li> <li>• <a href="#">HiSeq X Ten System</a> for population-scale whole-genome sequencing</li> </ul>

\$1500 30x Human Genome (WGS)  
 \$750 100X Human Exome (WES)

# What is a flowcell or a lane?



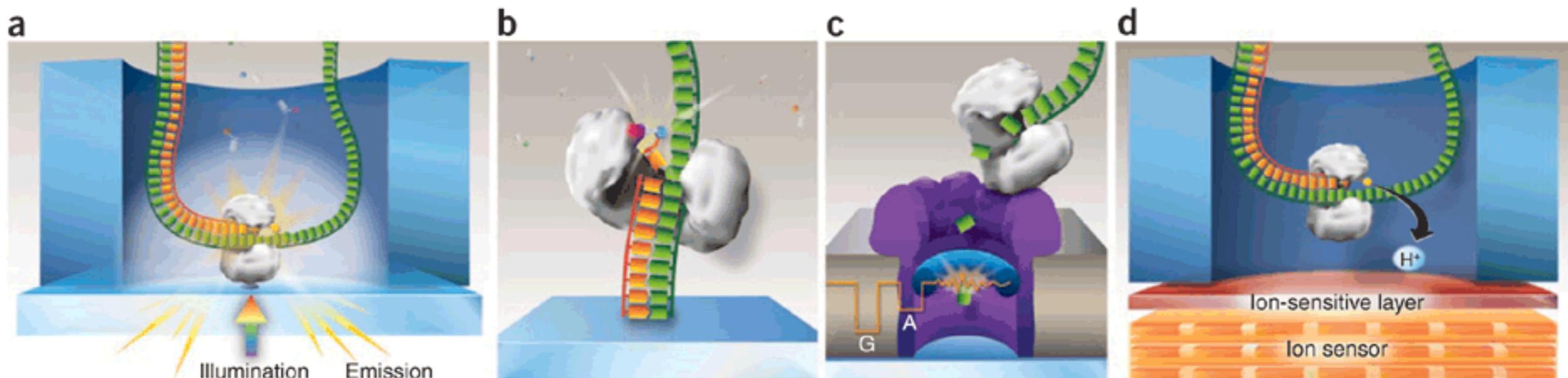
# When to use Illumina

- MiSeq
  - Micro biome 16S rRNA sequence
- Others
  - DNA
  - RNA
  - Chromatin Sequencing
    - ChIP
    - ATAC
    - DNaseq
    - MNaseSeq
  - Clip
  - Single-Cell
  - Microbiome

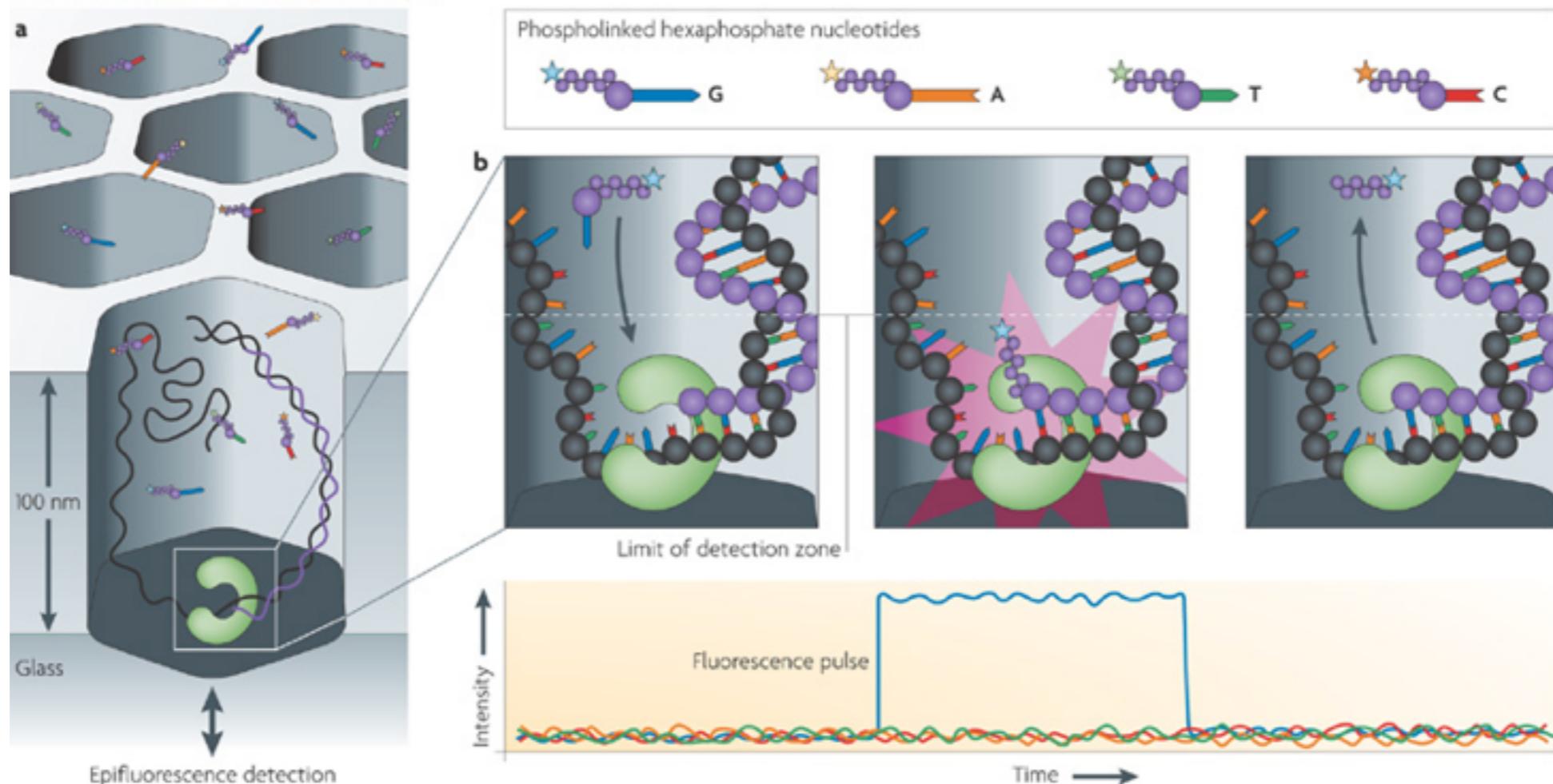
# UTSW NGS Cores

- McDermott Center
  - <http://www.utsouthwestern.edu/labs/next-generation-sequencing-core/>
- Center for Precision Genomics
  - <http://www.utsouthwestern.edu/research/core-facilities/genomics/index.html>
- CRI
  - <http://cri.utsw.edu/sequencing-facility-home/>

# SMRT-Sequencing



Pacific Biosciences — Real-time sequencing



# SMRT Sequencing

- Pacific Biosciences
  - Single Molecule Real Time (SMRT) Sequencing
  - Higher Error Rate than Illumina
  - Average read length > 1kb
  - ~\$2/Mbp



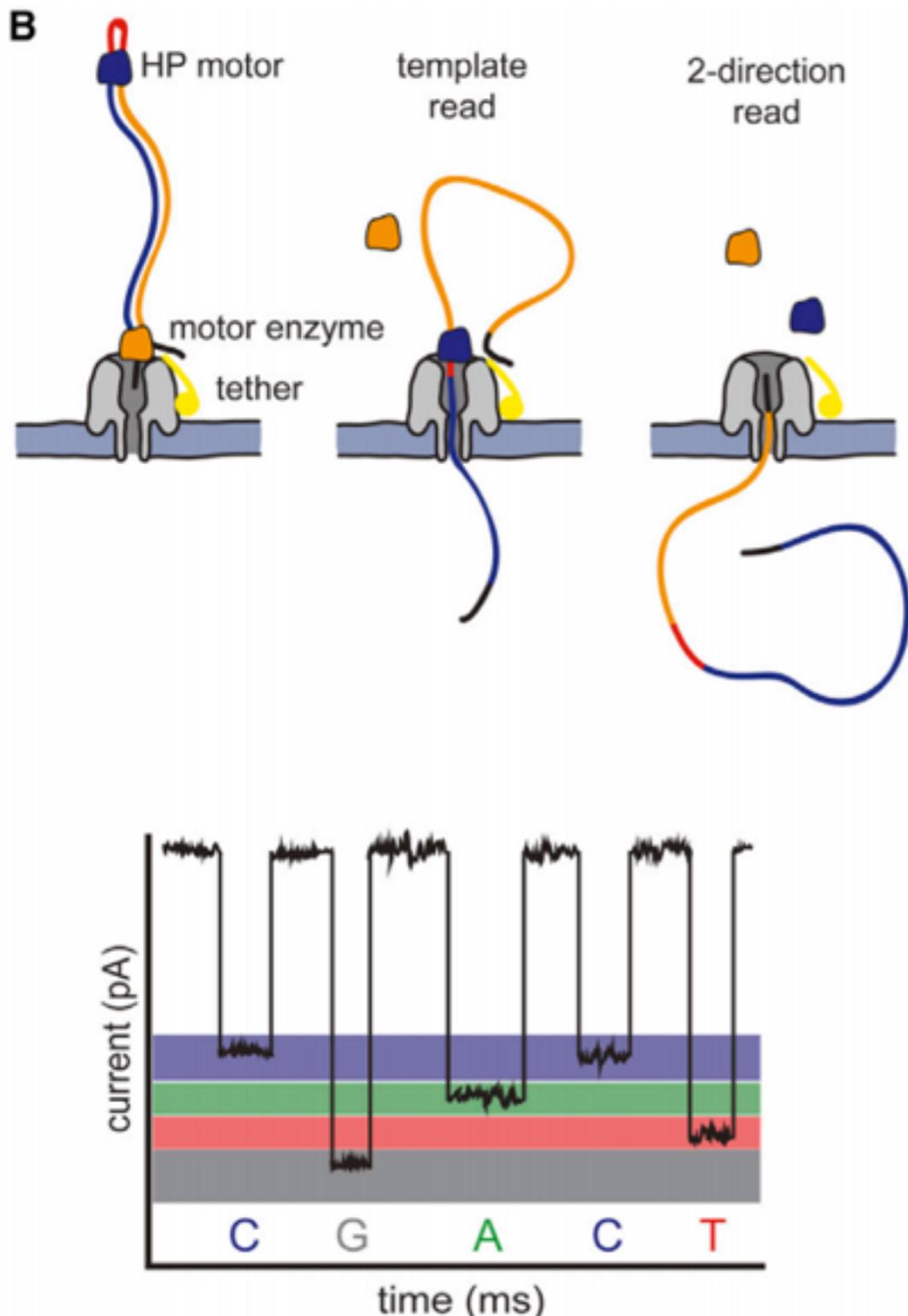
# When to use SMRT

- Great for Finishing Whole Genomes by Illumina/PacBio Hybrid Assembly
  - Bacterial Genomes can be sequenced in circular form
  - Eukaryotic genomes can be assembled into more complete chromosomes
- Ability to sequence through long repetitive regions that are hard to sequence by traditional methods
  - Great for Eukaryotic Genomes — particularly plant genomes
- Structural Variation Detection

# Where to get PacBio Sequencing

- <http://www.igs.umaryland.edu/resources/grc/index.php>
- <http://www.genome.arizona.edu/>
- <http://www.ncgr.org>
- <http://www.broadinstitute.org/scientific-community/science/platforms/broad-technology-labs/broad-technology-labs>

# Oxford Nanopore MinION



(B) Oxford Nanopore's sequencing strategy. DNA templates are ligated with two adaptors. The first adaptor is bound with a motor enzyme as well as a tether, whereas the second adaptor is a hairpin oligo that is bound by the HP motor protein. Changes in current that are induced as the nucleotides pass through the pore are used to discriminate bases. The library design allows sequencing of both strands of DNA from a single molecule (two-direction reads).

# Oxford Nanopore Minlon

- A “laptop powered” sequencing
- ~\$1000 for the run (1Gb)
- A single 18 hr run can produce >90 Mb of data from around 16,000 total reads, with read lengths between ~6 kb and >60 kb,
- error rates are high: insertion, deletion, and substitution rates of 4.9%, 7.8%, and 5.1%,
- Presently, it also has a very high run failure rate.



# When to use Nanopore

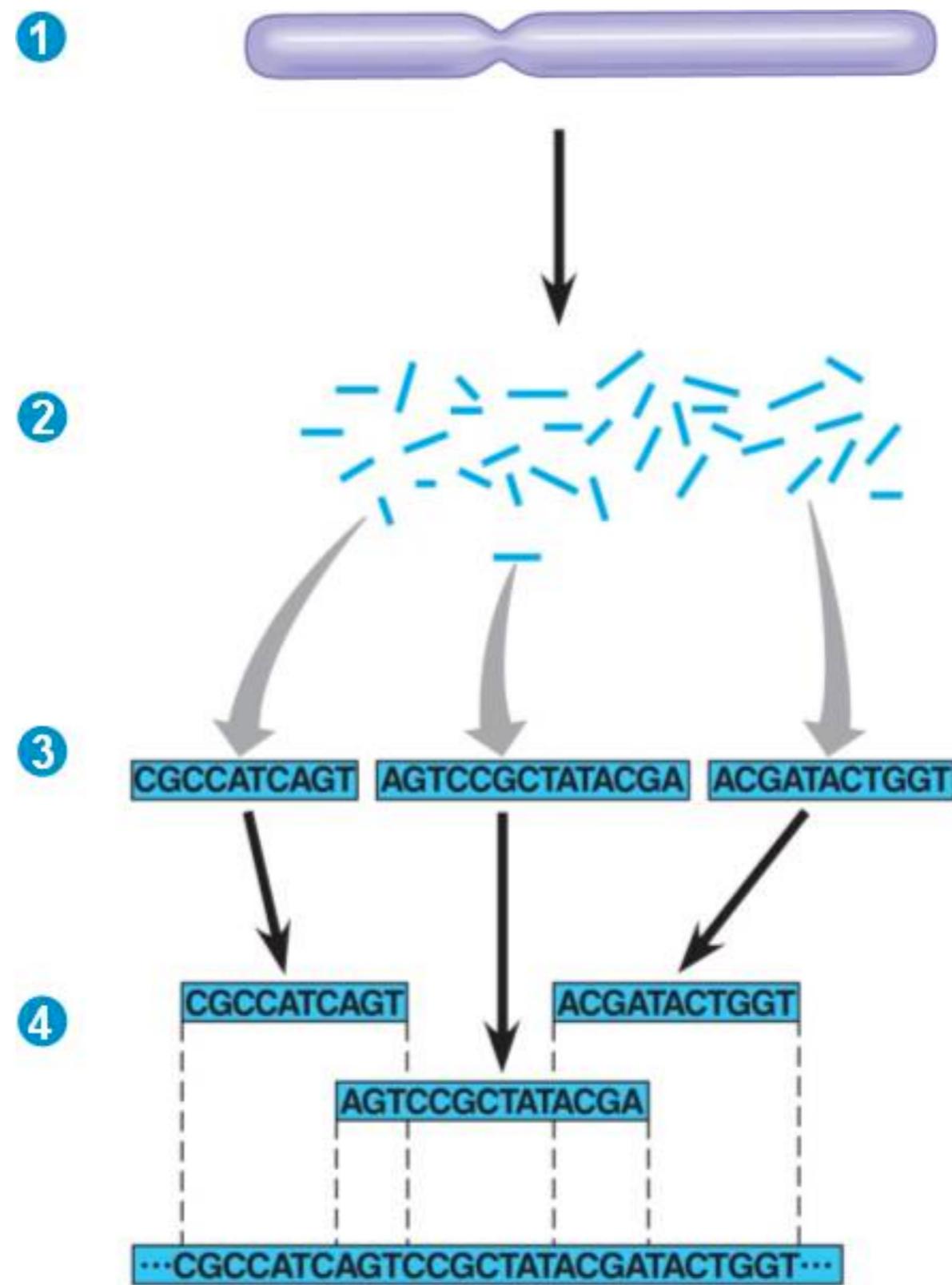
- Light weight and low power usage makes it interesting for “in the field” applications
  - Viral Outbreak
  - Bacterial Outbreak
  - Sampling in extreme environments
- Anyone can do it anywhere — no core needed

- Sequencing Methods
  - Sanger
  - Pyro
  - Solexa
  - SMRT
  - Field
- Sequencing Applications
  - DNA
  - RNA
  - Chromatin Sequencing
    - ChIP
    - ATAC
    - DNaseq
    - MNaseSeq
  - Clip
  - Ribo
  - Single-Cell
- Alternatives to Sequencing
  - Microarray
  - Nano string

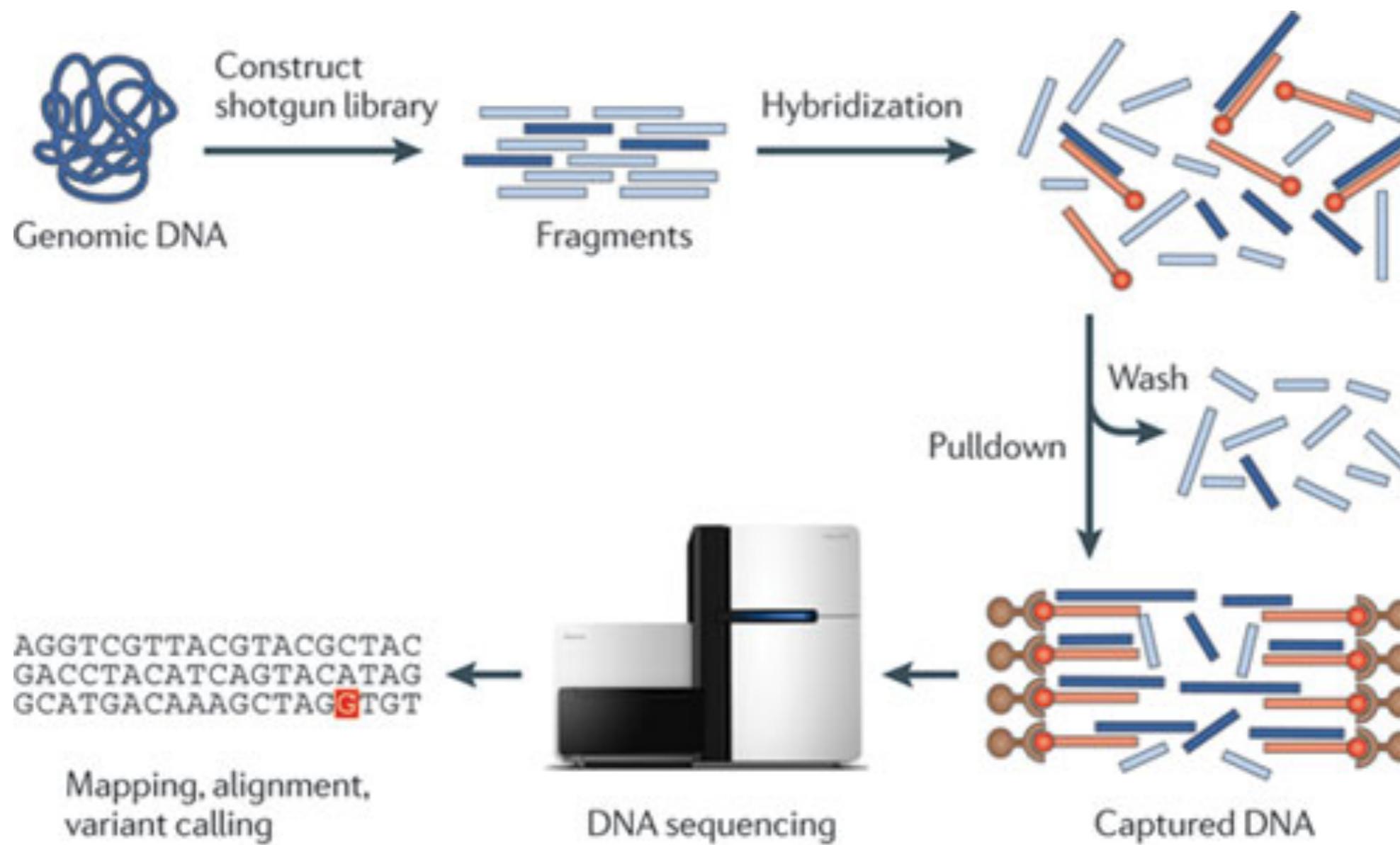
# DNASeq

- Whole Genome Shotgun (WGS)
  - Eukaryotic Genomes
  - Bacterial Genomes
  - Metagenomic Sequencing
- Targeted Sequencing
  - Whole Exome
  - Targeted Gene Panel
  - PCR Based Targeting

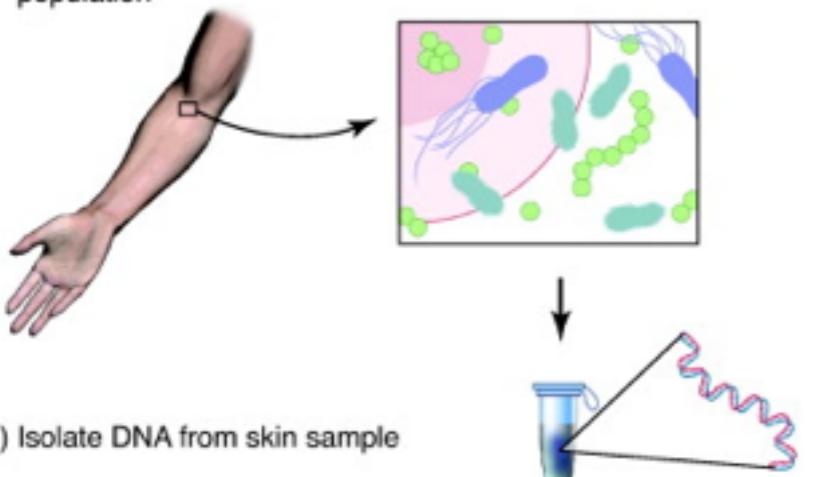
# Whole Genome Shotgun



# Targeted Capture



(i) Obtain superficial skin sample containing mixed bacterial population



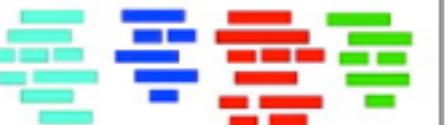
(ii) Isolate DNA from skin sample

(iii) Amplify bacterial 16S  
rRNA gene with primers  
encompassing variable  
regions of interest

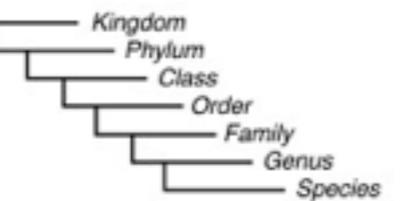
(iv) Sequence 16S rRNA genes

(v) Perform data processing, quality control and analysis of  
bacterial 16S rRNA sequences:

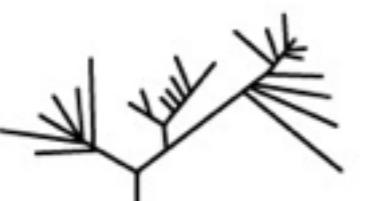
- Alignment of sequences



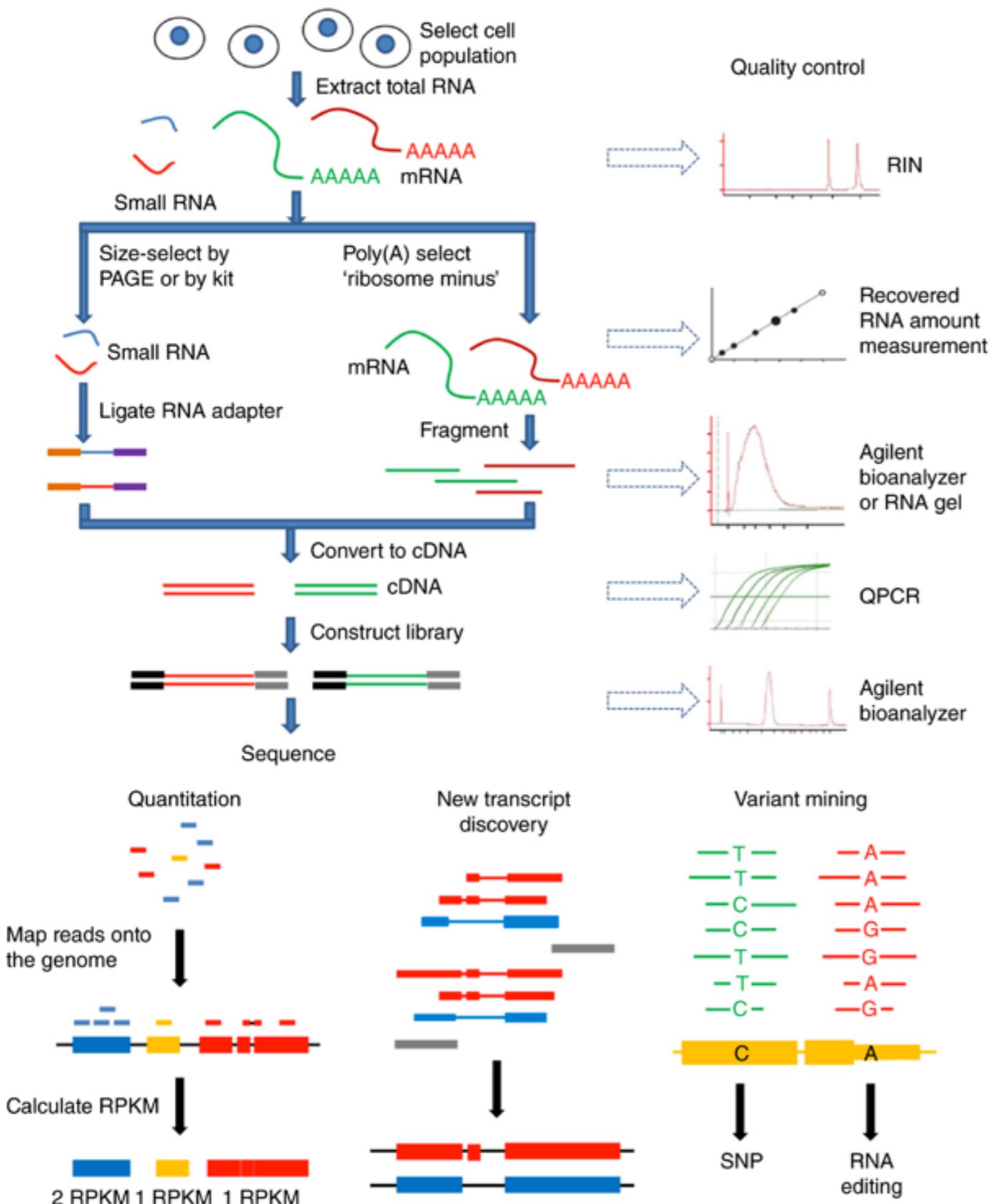
- Taxonomic classification using  
existing reference databases



- Community and phylogenetic  
analysis of sequences



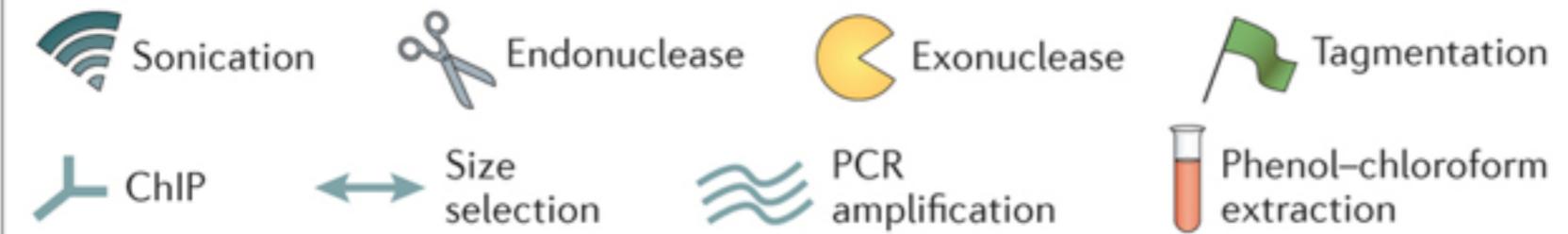
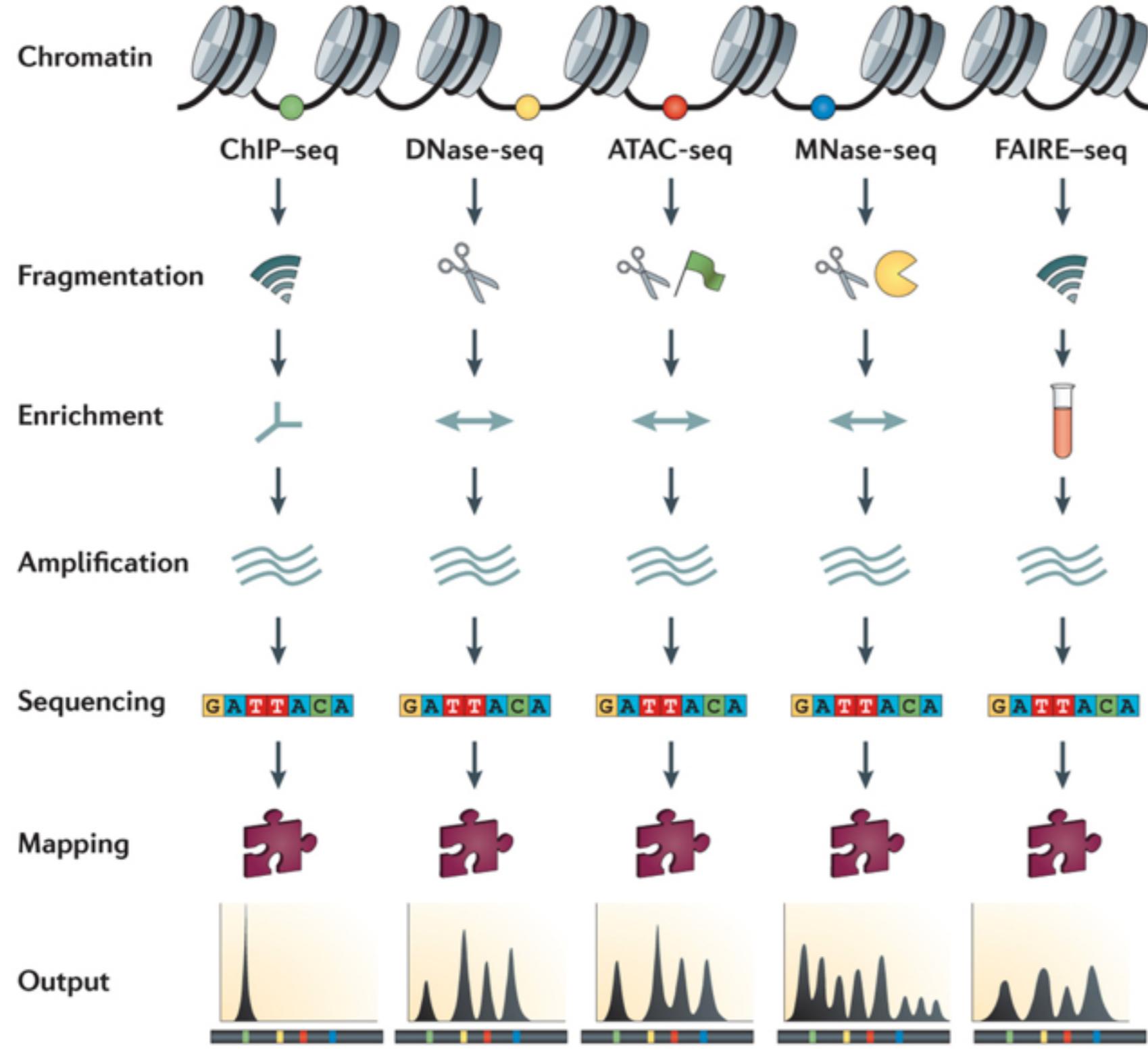
# PCR-Based Targeting



# RNASeq

# Chromatin Seq

- Chromatin Immune-Precipitation (ChIP) Seq
  - Identify regions of the genome that bind certain DNA binding proteins (using an antibody)
- DNase Seq
  - Identify accessible DNA regions by digestion with a DNase I
- Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-Seq)
  - Identify accessible DNA regions by cross linking nucleosomes and isolating non crosslink regions
- Assay for Transposase-Accessible Chromatin (ATAC Seq)
  - Uses a 2 part step for isolating open chromatic involving Tn5 insertion and PCR
- MNase
  - Identify unaccessible DNA regions by digestion of accessible regions using MNase



# Chromatin Seq

# CLIP-Seq

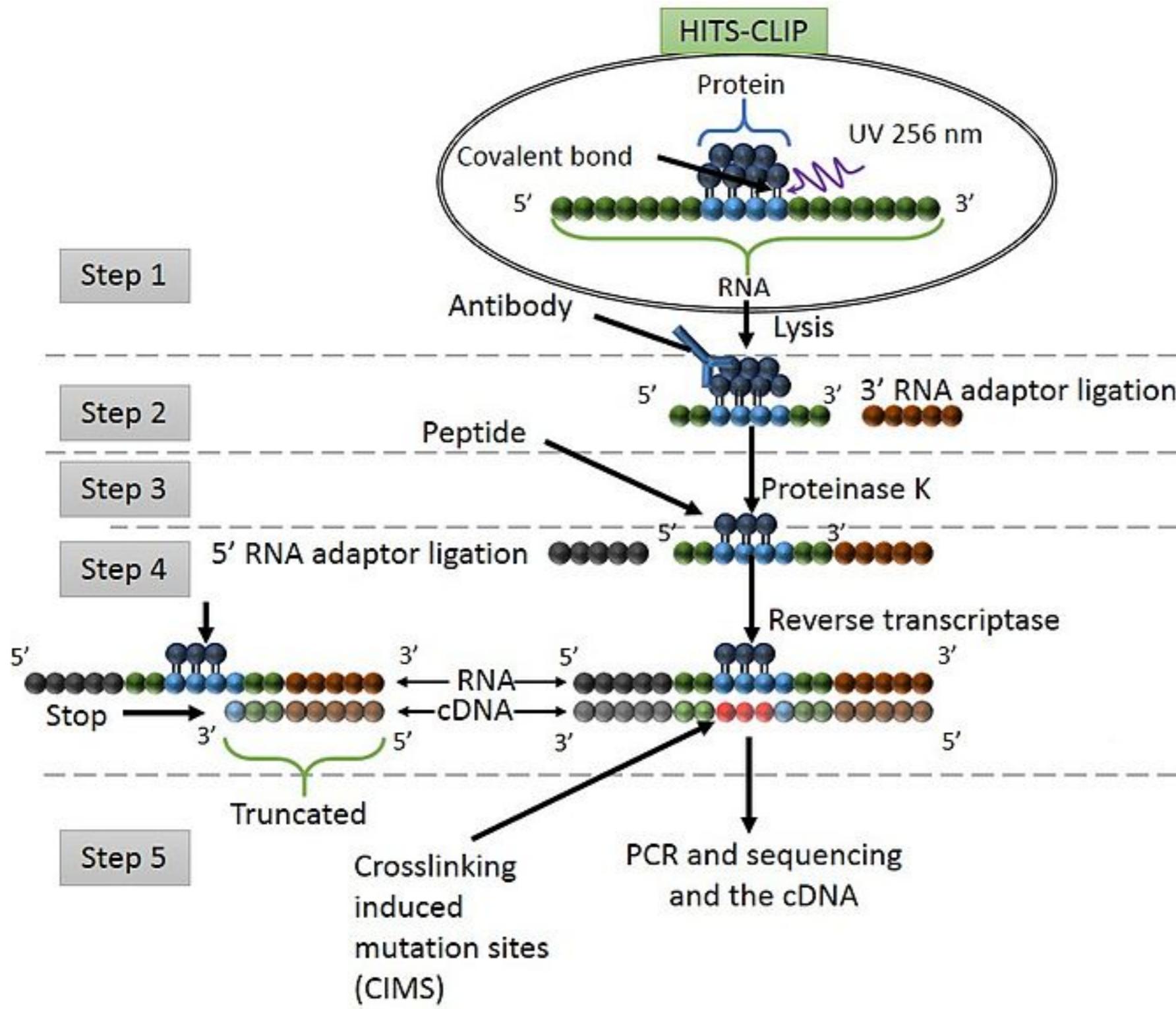


Figure 2: HITS-CLIP

**Step 1**

HITS-CLIP begins with the in-vivo cross-linking of RNA-protein complexes using ultraviolet light. The cell is lysed and the protein of interest is isolated using immunoprecipitation.

**Step 2**

Washing is performed to remove free RNA, and RNA adaptors are ligated at the 3' ends.

**Step 3**

Proteinase K digestion is performed. This leaves a peptide at the cross-link site that modifies the chemical structure of the nucleotide.

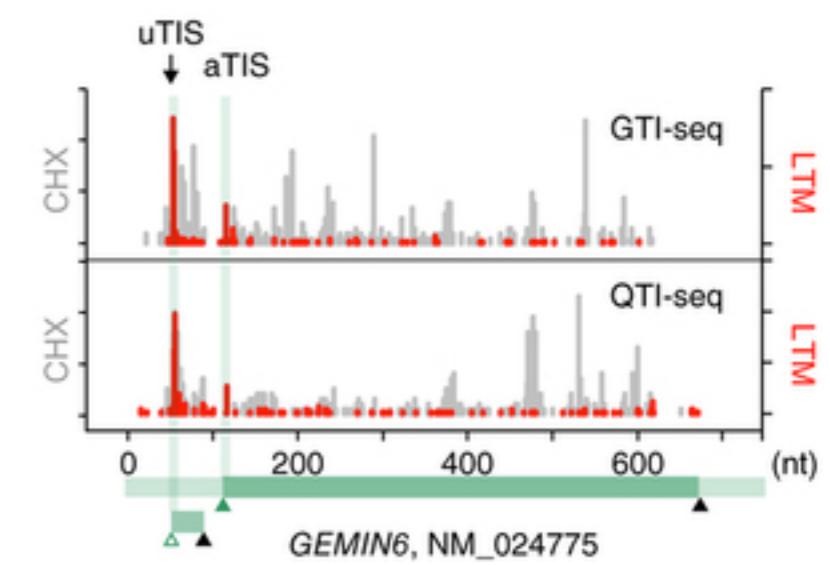
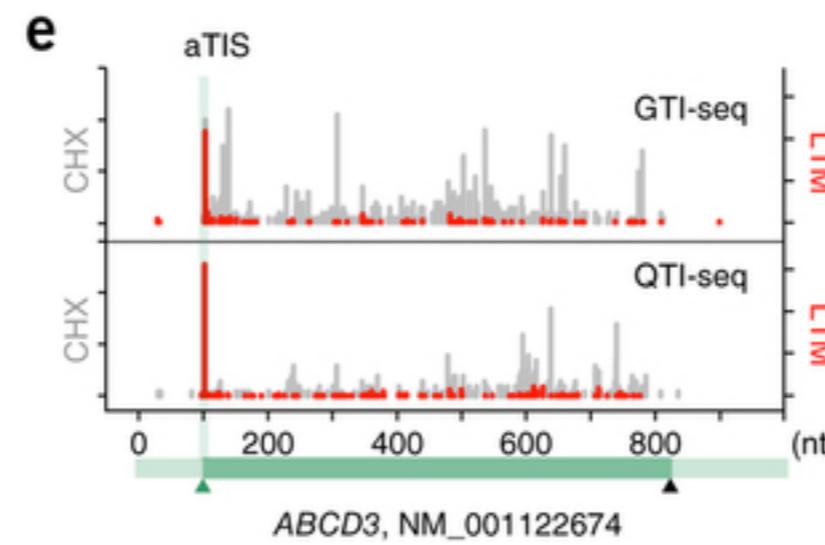
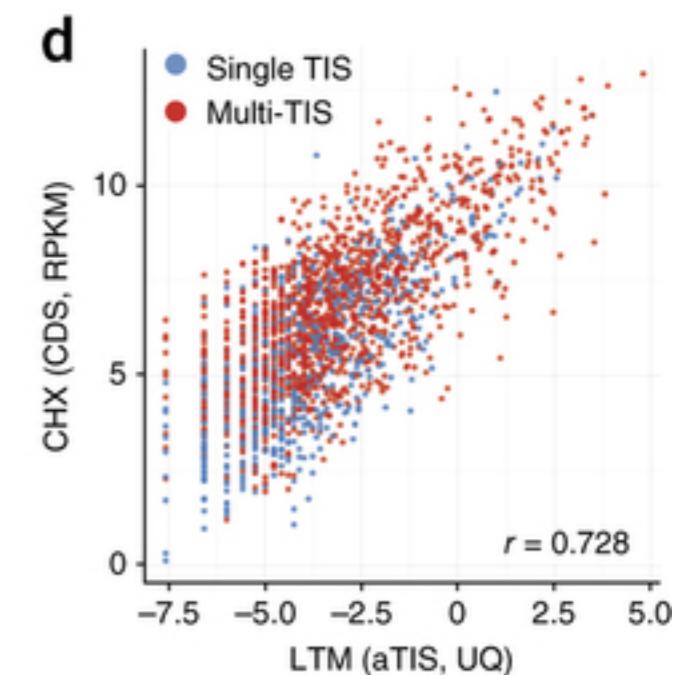
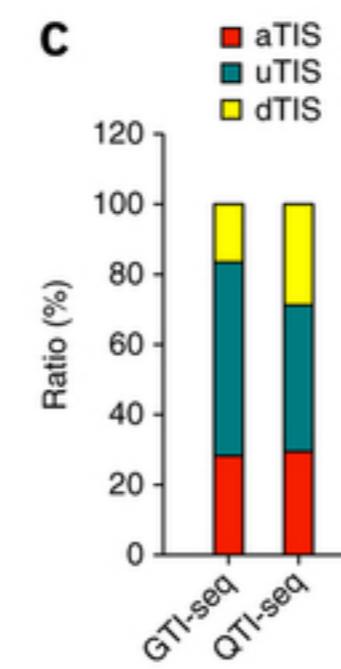
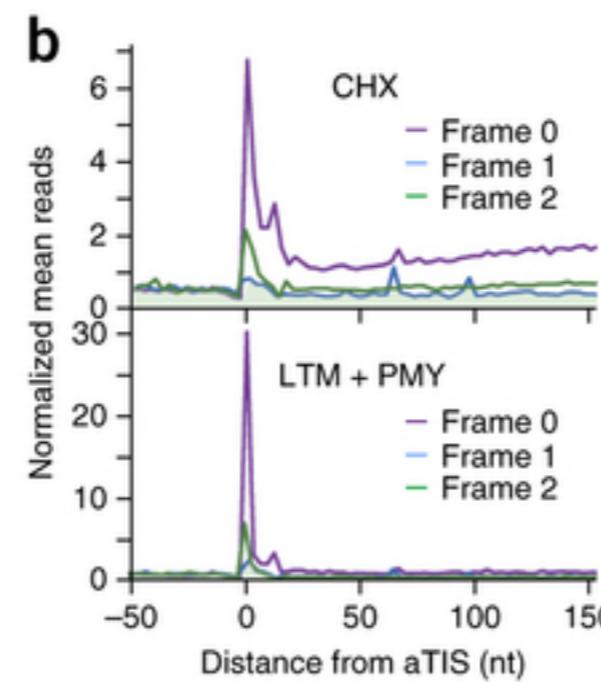
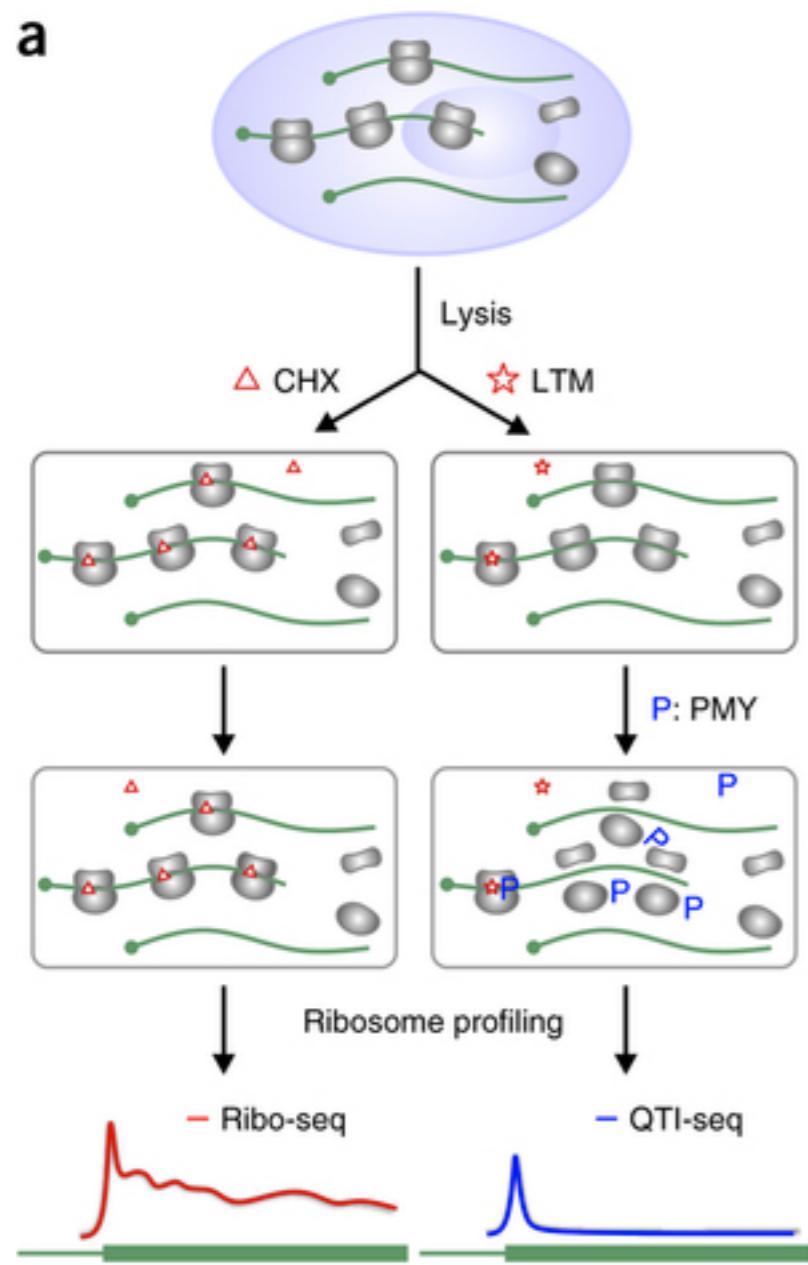
**Step 4**

5' RNA adaptors are ligated and cDNA is synthesized using reverse transcription.

**Step 5**

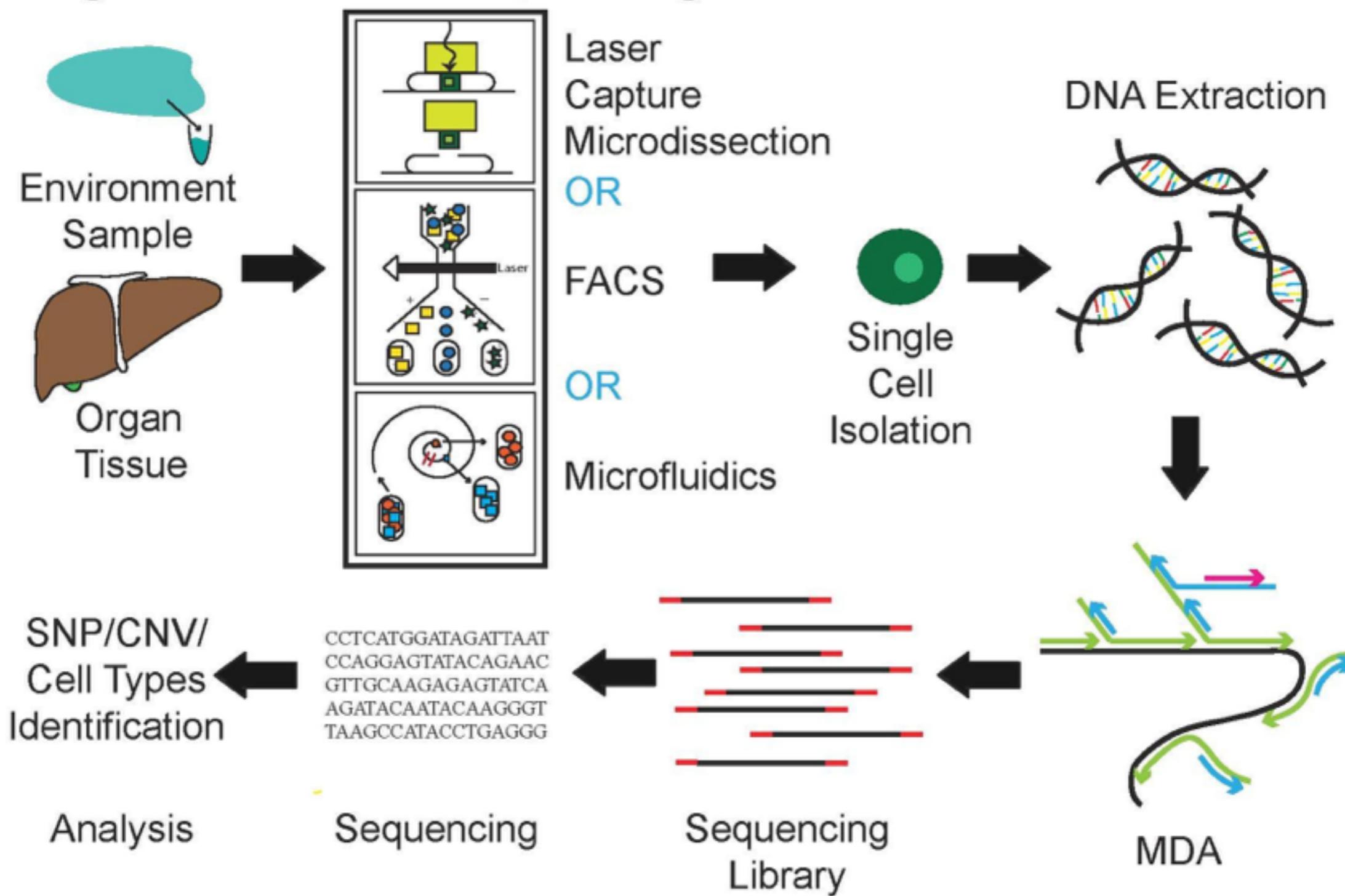
PCR and sequencing of the cDNA.

# Ribo Seq

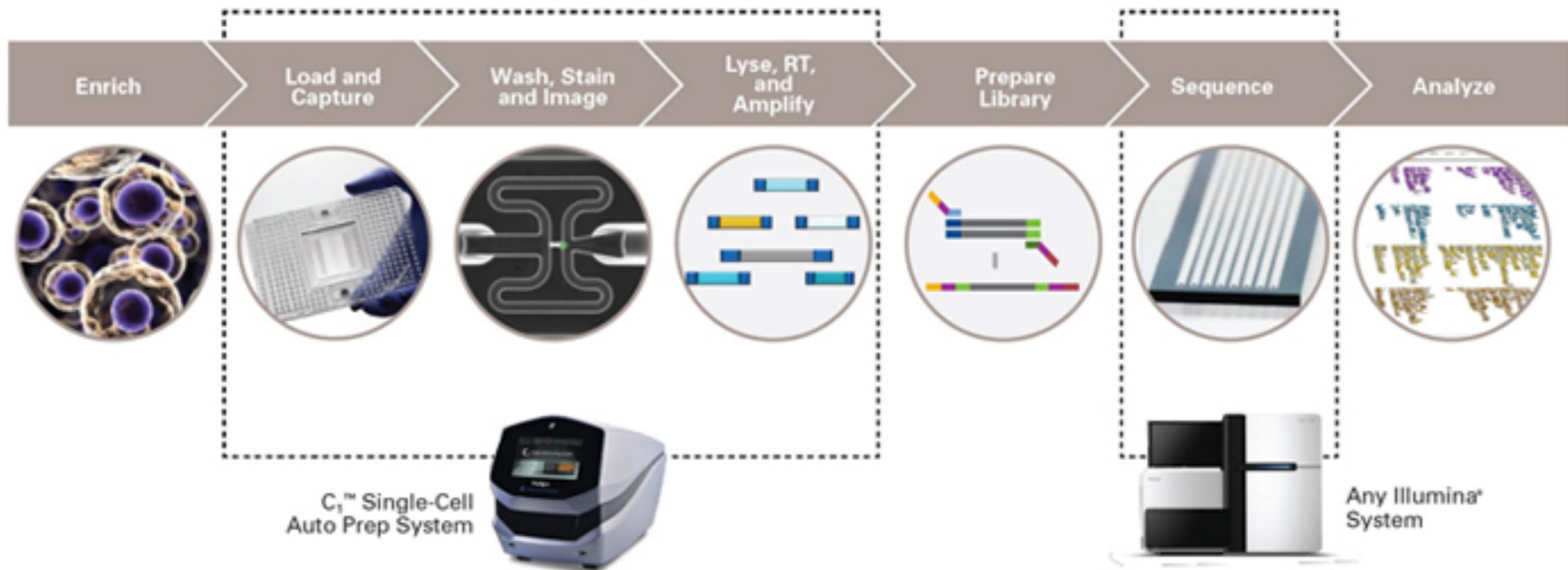


# Single Cell Seq

## Single Cell Genome Sequencing Workflow



# Fluidigm System



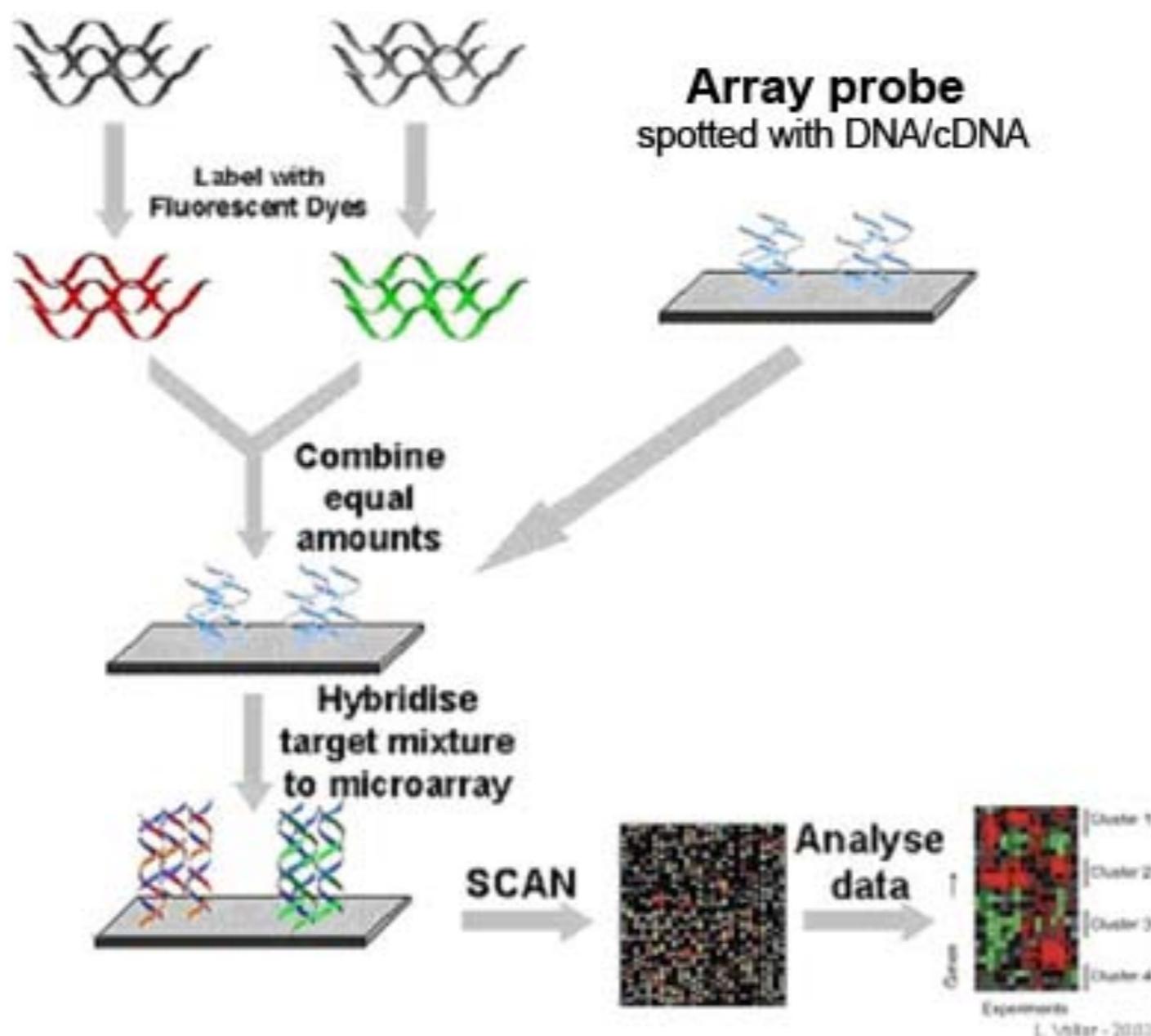
- Sequencing Methods
  - Sanger
  - Pyro
  - Solexa
  - SMRT
  - Field
- Sequencing Applications
  - DNA
  - RNA
  - Chromatin Sequencing
    - ChIP
    - ATAC
    - DNaseq
    - MNaseSeq
  - Clip
  - Ribo
  - Single-Cell
- Alternatives to Sequencing
  - Microarray
  - Nano string

Sequencing can still be very expensive, if you know what you are looking for.



# Microarrays

Target DNA/mRNA from sample



# DNA Microarrays

- aCGH
  - Array comparative genomic hybridization is a molecular cytogenetic technique for the detection of chromosomal copy number changes on a genome wide and high-resolution scale.
  - These arrays typically cost ~ \$100-\$200
  - Can be useful to find large copy number variation in many samples
  - More accurate in the determination of large copy number variation (CNV) than Exome or Targeted Sequencing
  - Therefore, this technique is complementary to Targeting DNA sequencing

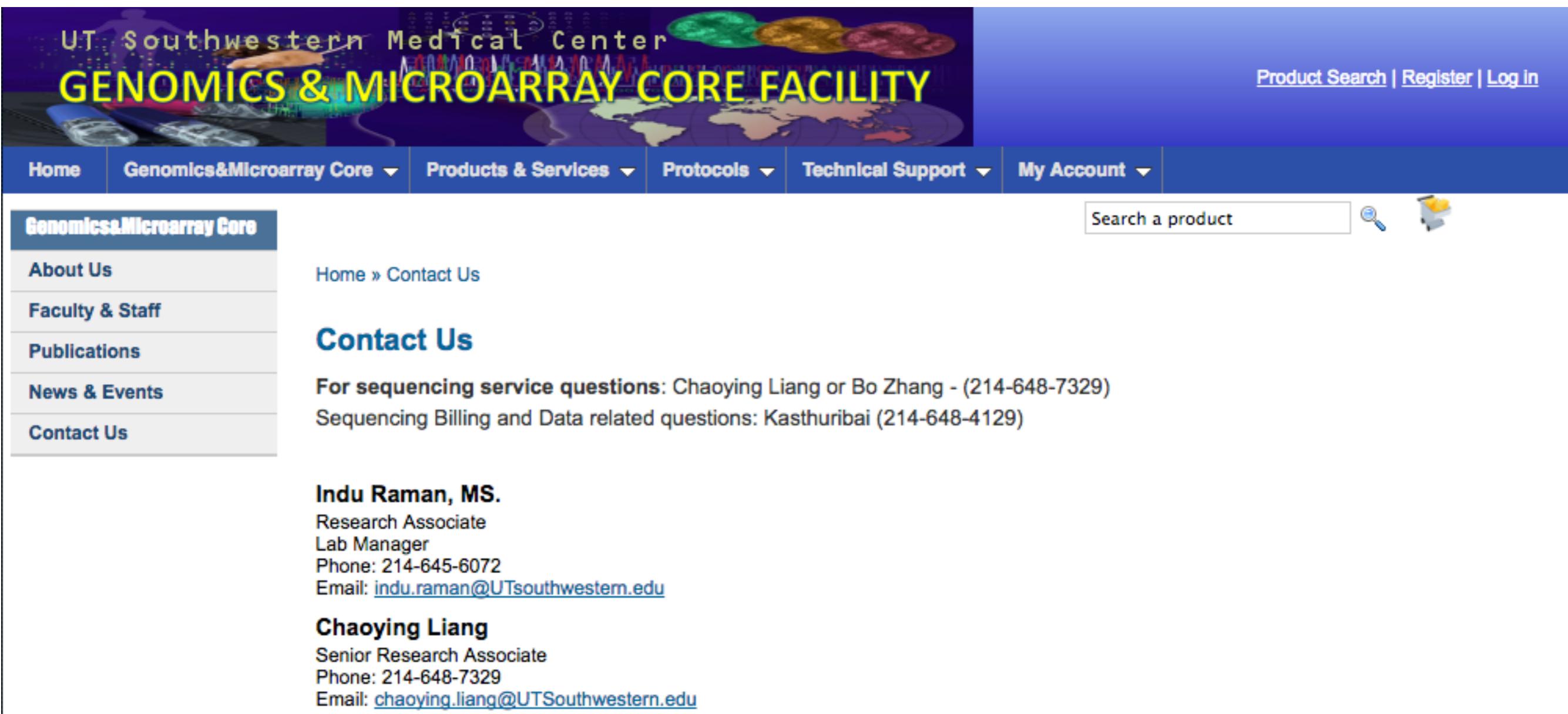
# DNA Microarrays

- Genotyping array
  - Typically cost \$100-200 compared to ~\$800 for WES (100X) or \$1500 for WGS (30-40X)
  - Used to detect polymorphisms within a population
  - Can be useful for SNP validation determined by exome sequence
  - Was used by the early human population studies to determine the ancestry and migration patterns of early humans
  - Can be customized to certain diseases (Immunochip)
- ChIP-chip
  - Similar to ChIP-Seq, ChIP-chip was an early method for detecting genetic regions bound by DNA binding proteins

# Gene Expression Microarrays

- Genotyping array
  - Typically cost \$300-400 (compared to ~\$700-\$800 for RNASeq)
  - Used to detect differences gene expression in many samples
  - Analysis is very easy

# Where to Get Microarray?



The screenshot shows the homepage of the UT Southwestern Medical Center Genomics & Microarray Core Facility. The header features a banner with a world map and the text "UT Southwestern Medical Center GENOMICS & MICROARRAY CORE FACILITY". On the right side of the header are links for "Product Search | Register | Log in". Below the header is a navigation bar with links for "Home", "Genomics&Microarray Core", "Products & Services", "Protocols", "Technical Support", and "My Account". A sidebar on the left contains links for "About Us", "Faculty & Staff", "Publications", "News & Events", and "Contact Us". The main content area has a "Search a product" input field with a magnifying glass icon. The "Contact Us" section includes contact information for Chaoying Liang and Bo Zhang, and profiles for Indu Raman and Chaoying Liang.

UT Southwestern Medical Center  
GENOMICS & MICROARRAY CORE FACILITY

Product Search | Register | Log in

Home | Genomics&Microarray Core | Products & Services | Protocols | Technical Support | My Account

Genomics&Microarray Core

About Us

Faculty & Staff

Publications

News & Events

Contact Us

Search a product

Home » Contact Us

## Contact Us

For sequencing service questions: Chaoying Liang or Bo Zhang - (214-648-7329)  
Sequencing Billing and Data related questions: Kasthuribai (214-648-4129)

**Indu Raman, MS.**  
Research Associate  
Lab Manager  
Phone: 214-645-6072  
Email: [indu.raman@UTsouthwestern.edu](mailto:indu.raman@UTsouthwestern.edu)

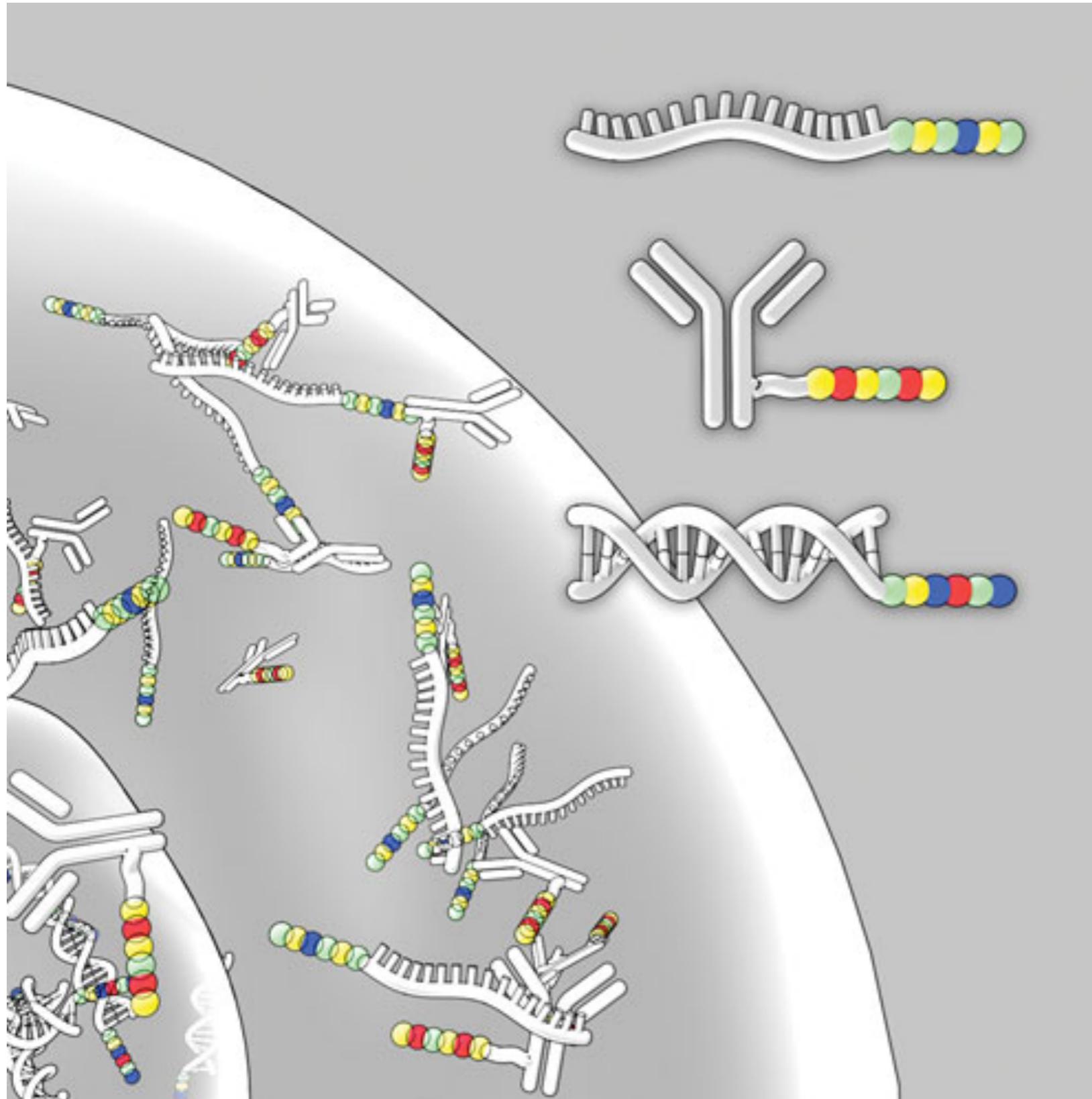
**Chaoying Liang**  
Senior Research Associate  
Phone: 214-648-7329  
Email: [chaoying.liang@UTSouthwestern.edu](mailto:chaoying.liang@UTSouthwestern.edu)

<https://microarray.swmed.edu/>

# Metagenomic Microarrays

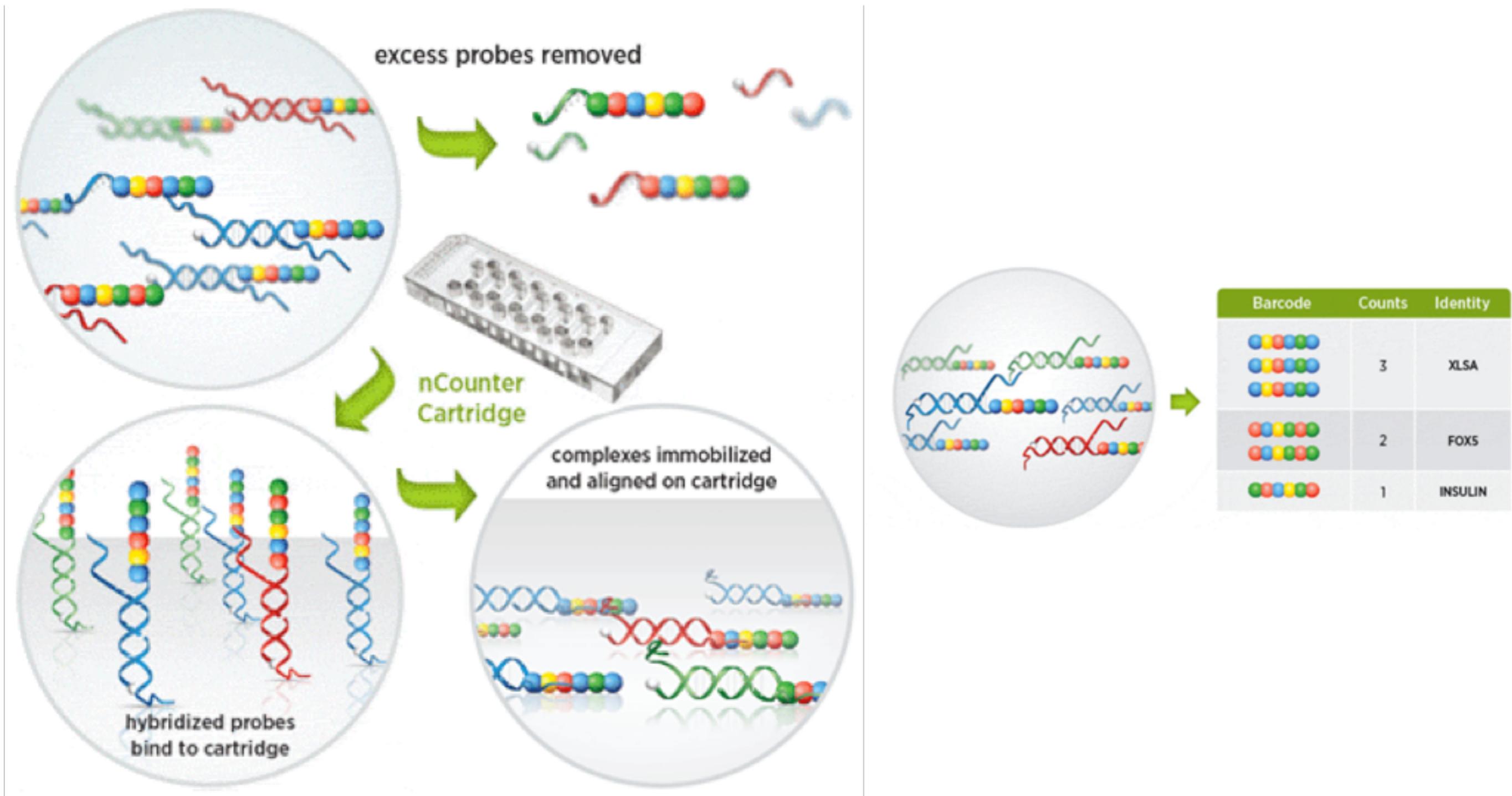
- Phylochip (Second Genome)
  - A microarray-based method that identifies and measures the relative abundance of more than 50,000 individual microbial taxa
  - With 1.2 million probes per chip, this ensures that measurements on important low abundance bacteria are not overwhelmed by commonplace, dominant microbial community members
  - Cheaper than 16S rRNA sequences
  - Analysis is very easy

# Nanostring



NanoString 3D Biology platform enables the discovery of up to 800 DNA, RNA, or protein targets from a single sample using fluorescent nucleic acid barcodes. DNA and RNA are detected by direct hybridization of oligo barcodes, and proteins are detected by barcode conjugated antibodies, which enables measurement of all three analytes from a single sample (1–2 FFPE slides or 150,000 PBMC or dissociated tumor cells) using a common optical reader.

# Nanostring



# Where to get Nanostring?

- BCM
  - <https://www.bcm.edu/research/advanced-technology-core-labs/lab-listing/genomic-and-rna-profiling-core/services/nanostring>
- MD Anderson

# Things to Consider

- How much funding do I have?
- Do I need the whole genome or transcriptome?
- Am I only interested in the variants/expression of a small set of pre-selected genes?
- Am I interested in rare known variants or low expressed genes?
- Is sequencing capable of accurately answering my question?

# Questions?