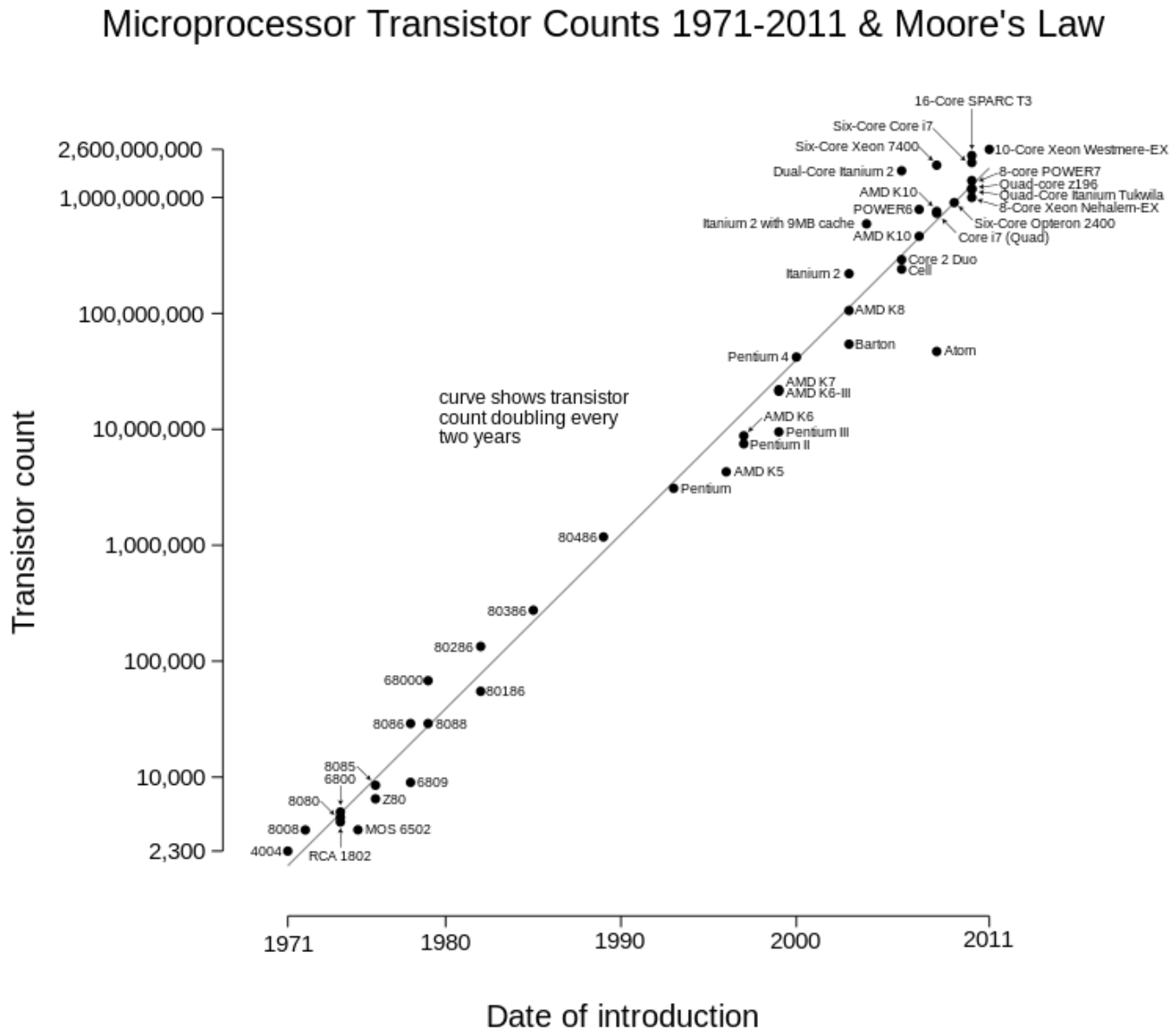
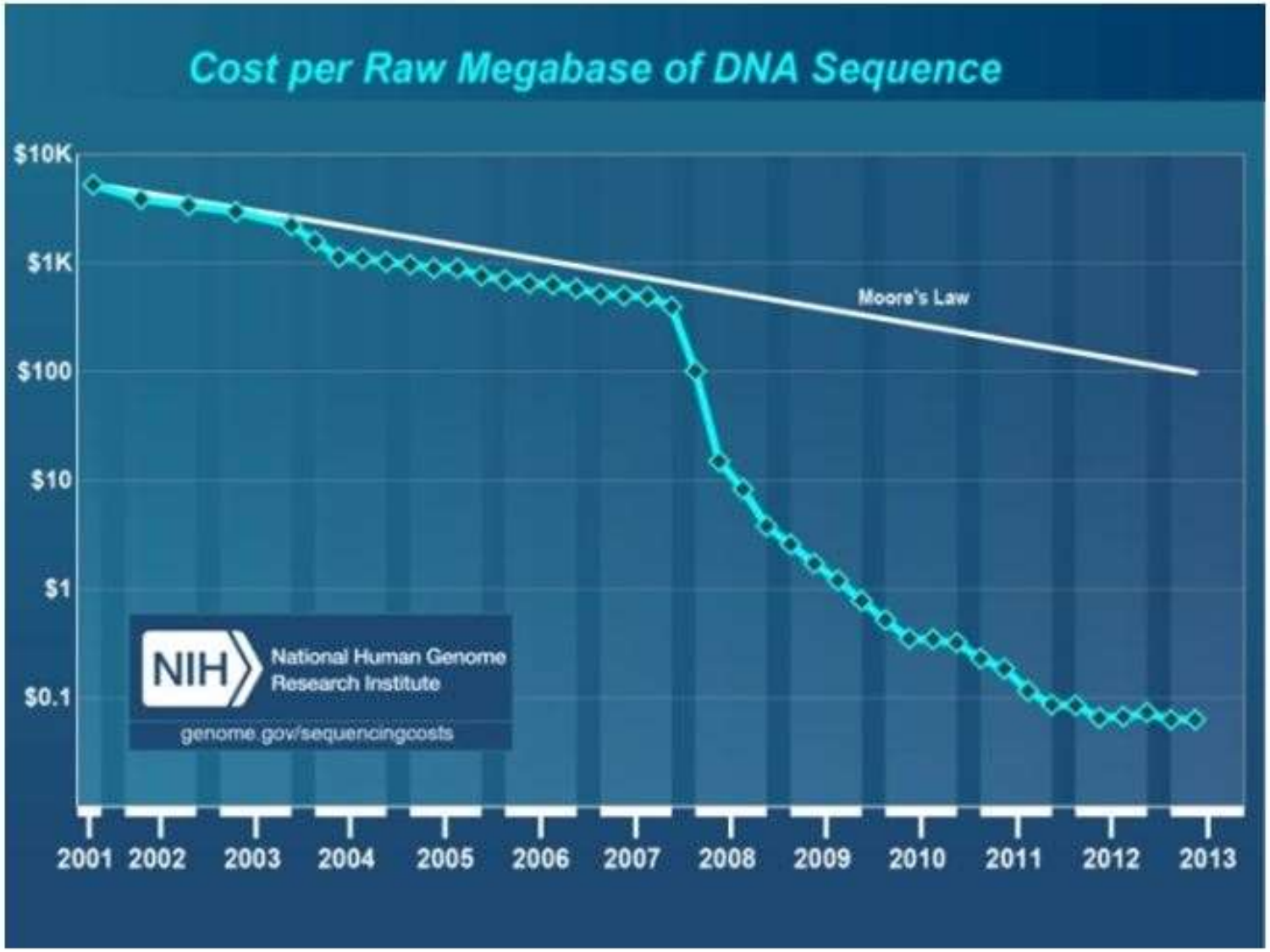
Moore’s Law

@ColauttiLab



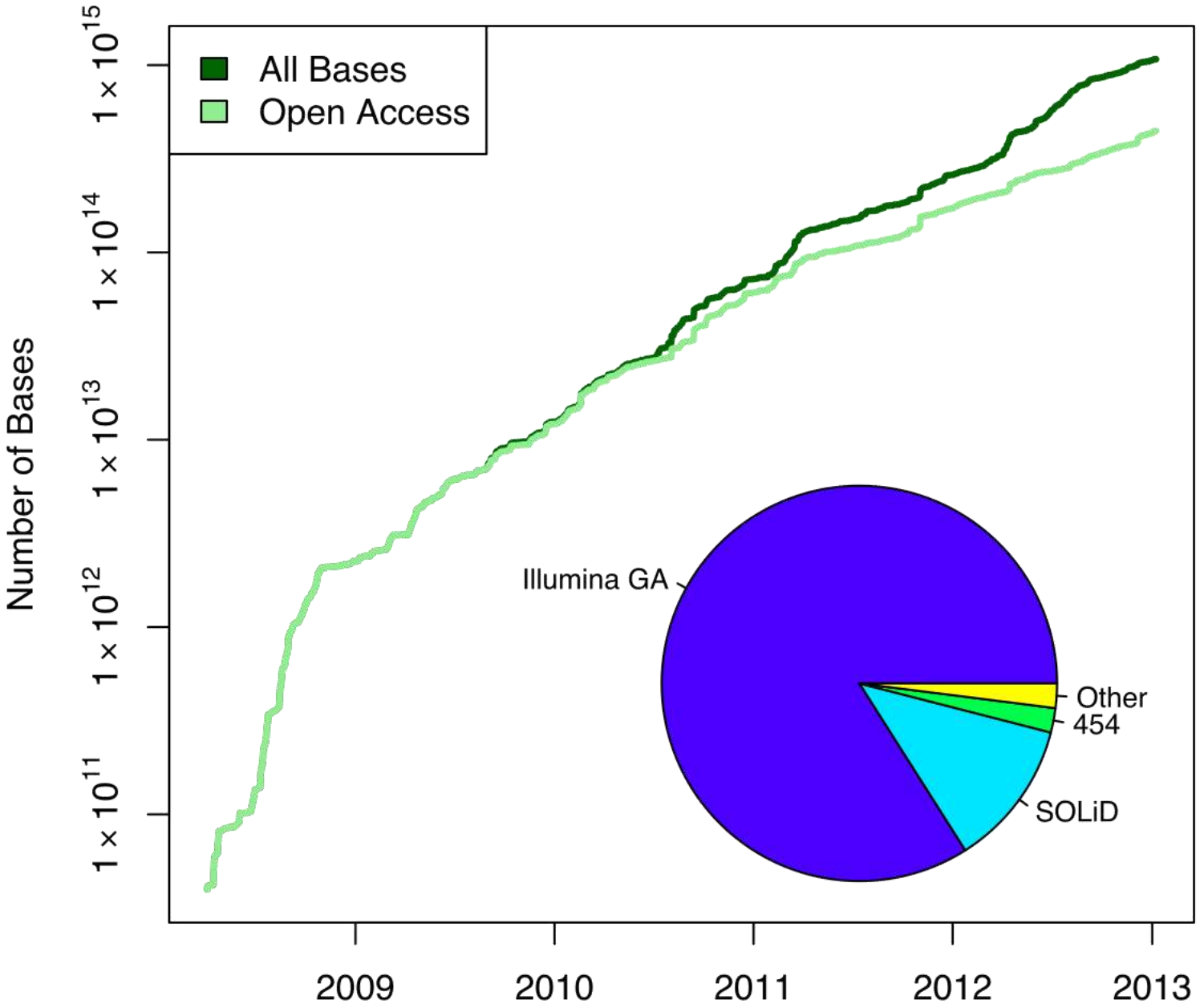
Moore's Law vs. Sequencing Technology

@ColauttiLab



Sequencing Read Archive (NCBI)

@ColauttiLab



Blue and green =
'Next Generation Sequencing'

2nd Gen: Sequencing Platforms

@ColauttiLab

MiSeq®

NextSeq® 500

HiSeq® 2500

HiSeq® 3000

Next Generation Sequencing
platforms from trusted names

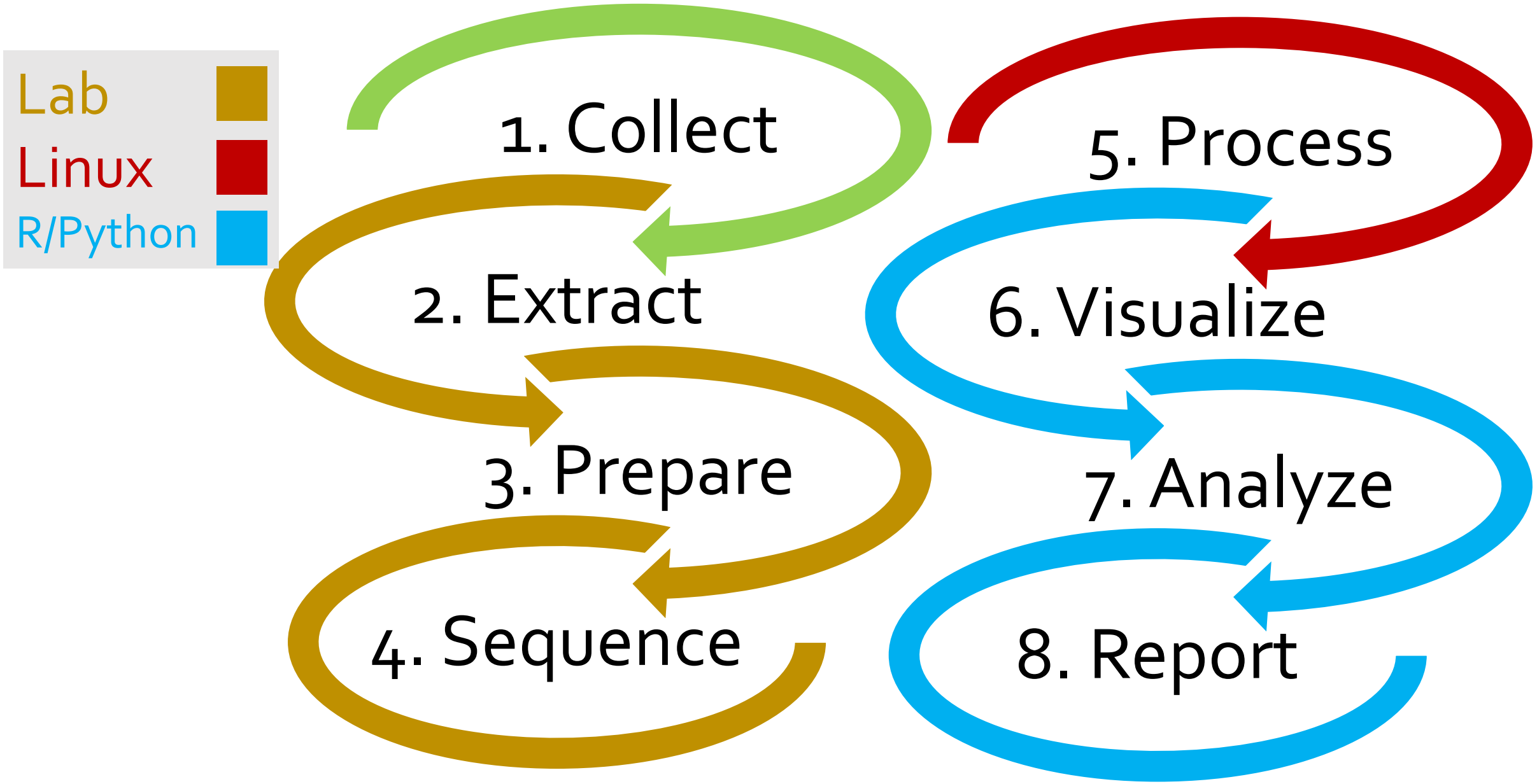
Ion Torrent™

PacBio RS II System

HiSeq® 4000

Next-Generation Sequencing: Typical Workflow

@ColauttiLab

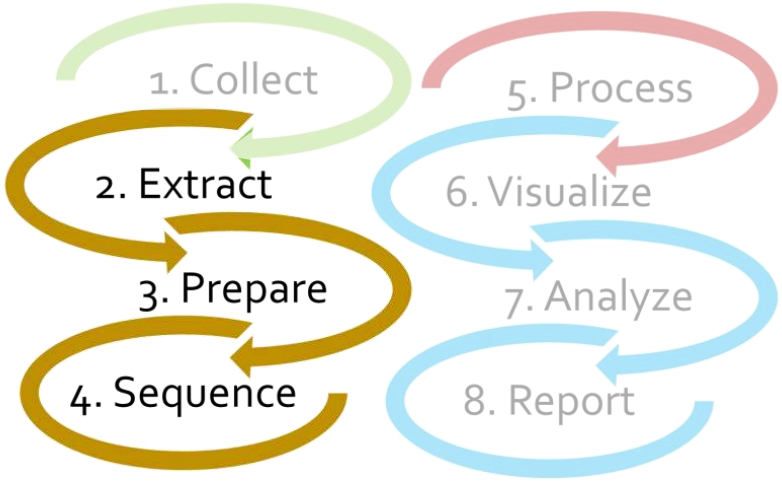


2nd Gen: Common Elements

@ColauttiLab

Sequencing Library Preparation

1. Extract & purify DNA*
2. Fragment to target size (75-750 bp)
3. Strand isolation
4. Clonal Amplification
5. Nucleotide detection



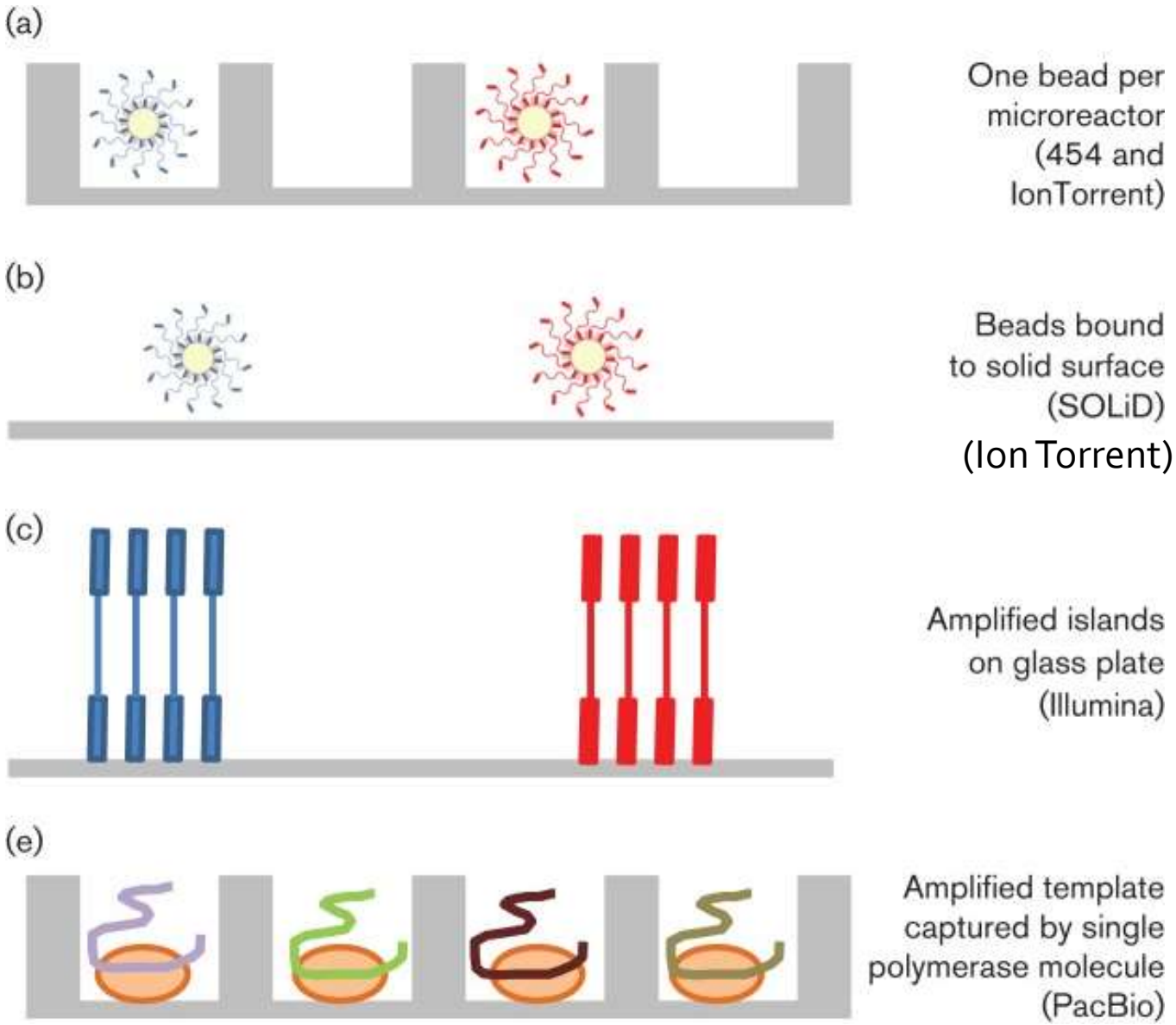
2. Fragment Sizes (longer list on website)

@ColauttiLab

Platform	Instrument	Mreads	Length (bp)	Gbp	Type
Illumina	NovaSeq 6000 S4	10000	300	3000	SR & PE
Illumina	NextSeq 500 High-Output	400	300	120	SR & PE
Illumina	HiSeq X	375	300	112.5	PE
Illumina	HiSeq High-Output v4	250	250	62.5	SR & PE
Illumina	MiSeq v3	25	600	15	SR & PE
Illumina	MiniSeq High-Output	25	300	7.5	SR & PE
Ion	Proton I	60	200	12	SR
Ion	PGM 318	4	400	1.6	SR
Ion	PGM 316	2	400	0.8	SR
Ion	PGM 314	0.4	400	0.16	SR
PacBio	PacBio Sequel	0.37	20000	7.4	SR
PacBio	PacBio RS II (P6)	0.055	15000	0.825	SR
Roche 454	GS FLX+ / FLX	0.7	700	0.49	SR
SOLiD	5500xl W	267	100	26.7	SR & PE

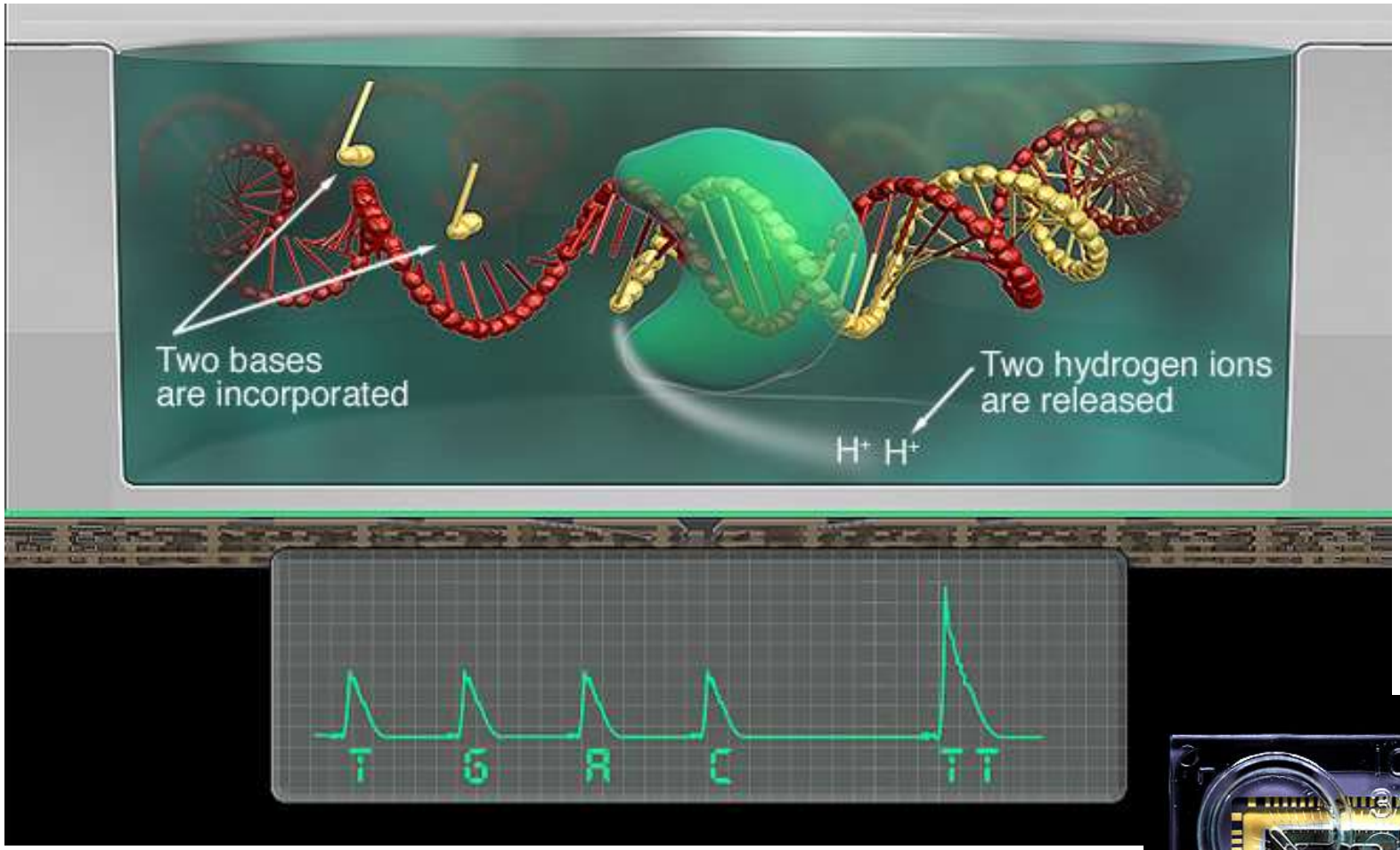
Sequence isolation (and cloning)

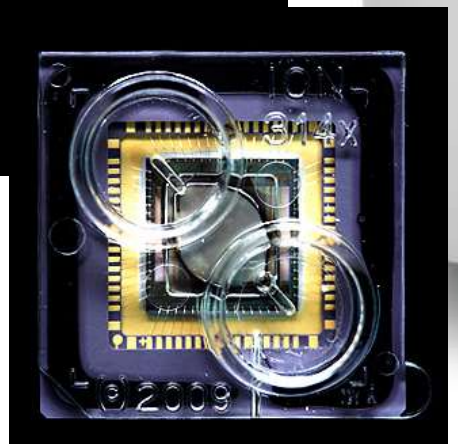
@ColauttiLab



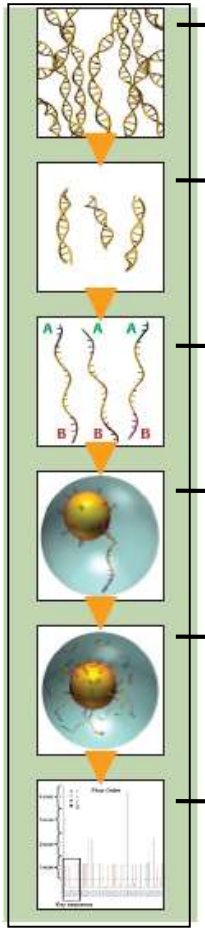
Ion Torrent

@ColauttiLab

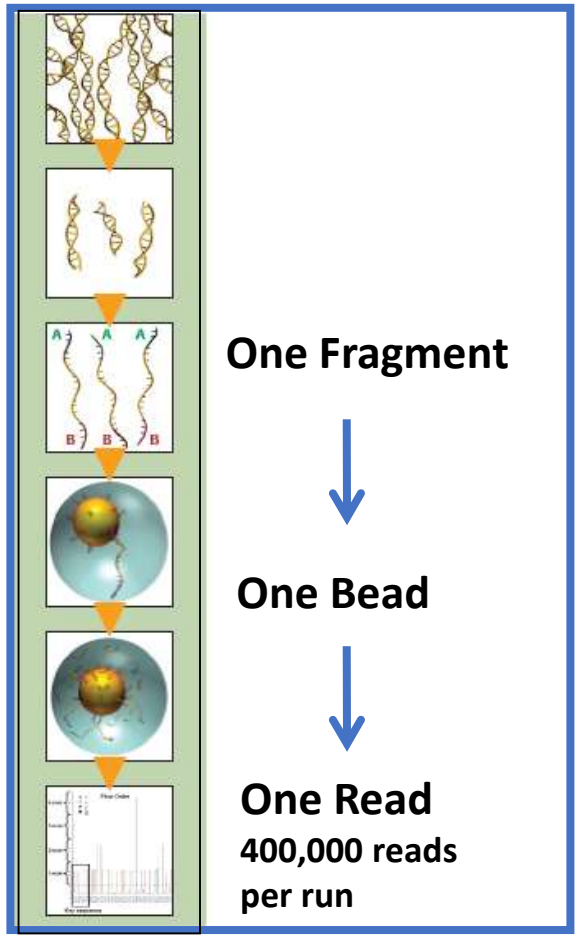




2nd Gen: 454 Sequencing (Roche; deprecated)



- Sample input: Genomic **DNA**, BACs, amplicons, cDNA
- Generation of small **DNA fragments** via nebulization
- Ligation of A/B-Adaptors flanking **single-stranded DNA fragments**
- Emulsification of beads and fragments in water-in-oil **microreactors**
- **Clonal amplification** of fragments bound to beads in microreactors
- **Sequencing** and **base calling**

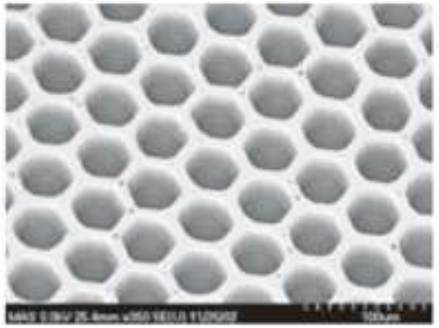


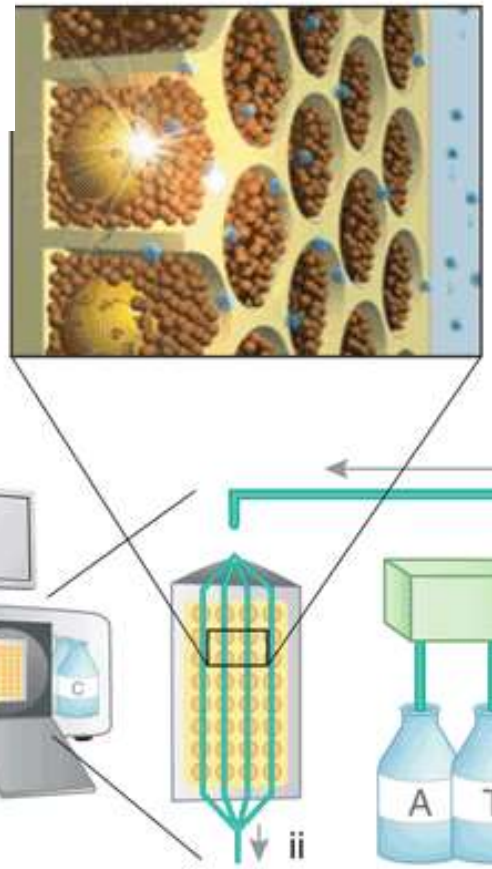
CSB2008 August 2008

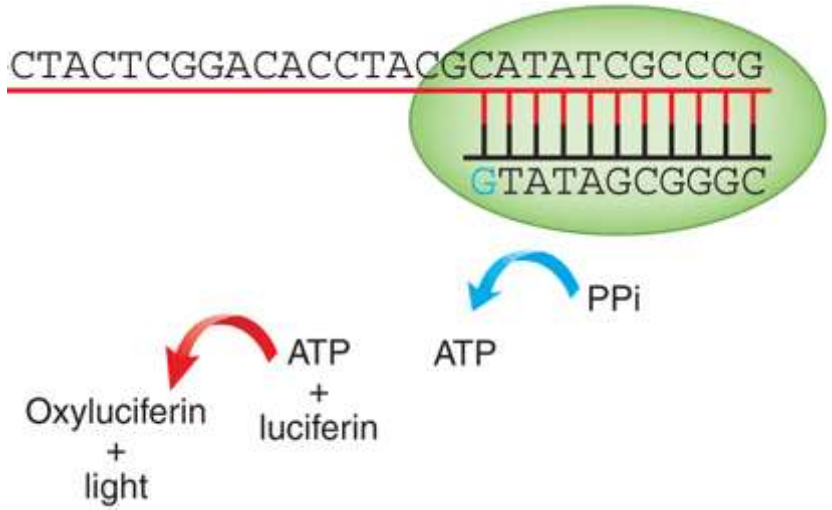
Rothberg & Leomon 2008

2nd Gen: 454 Sequencing (Roche; deprecated)

@ColauttiLab





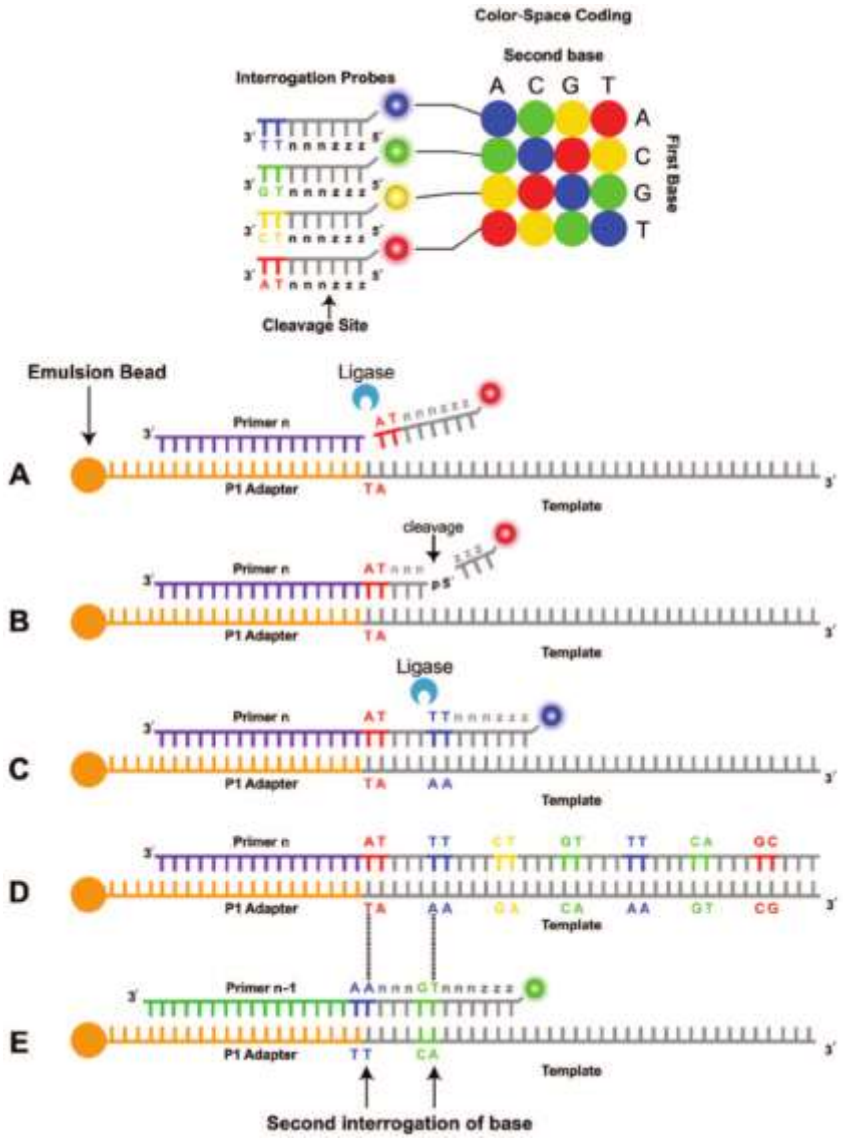


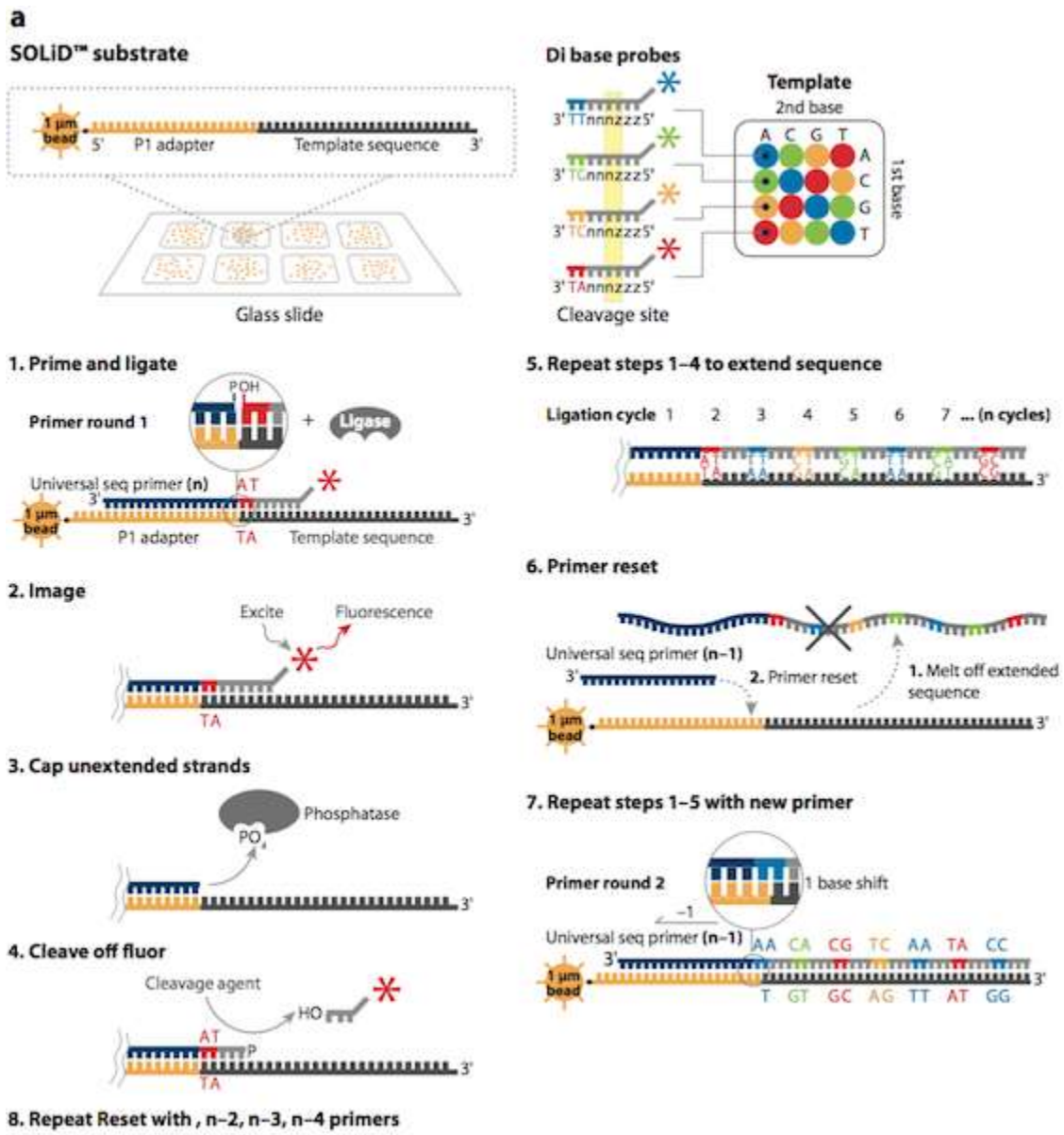
CSB2008 August 2008

Rothberg & Leomon 2008

SOLiD Sequencing (ABI)

@ColauttiLab

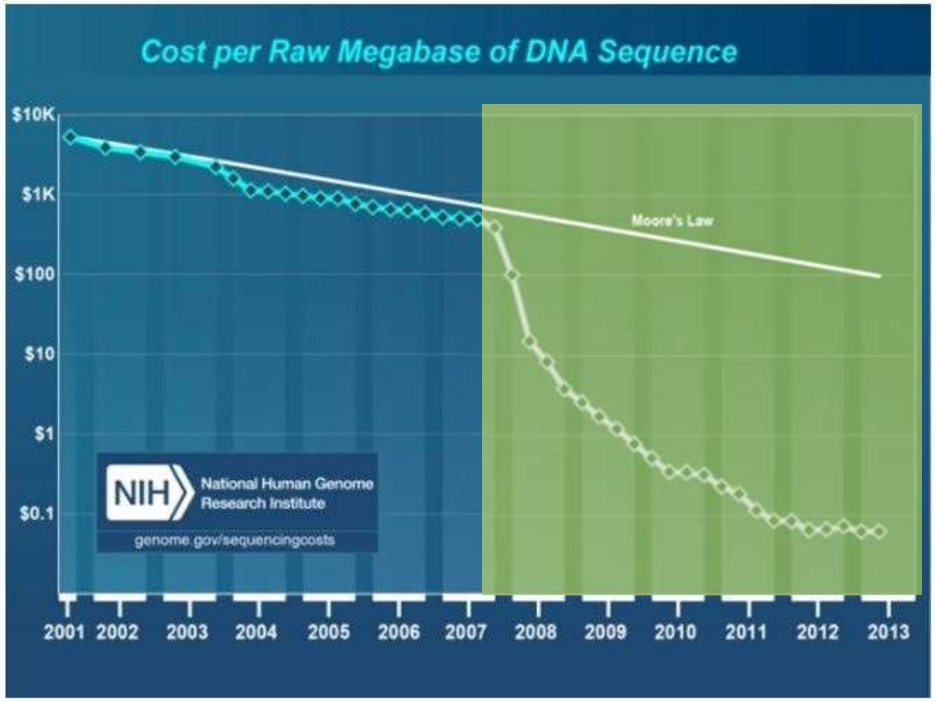




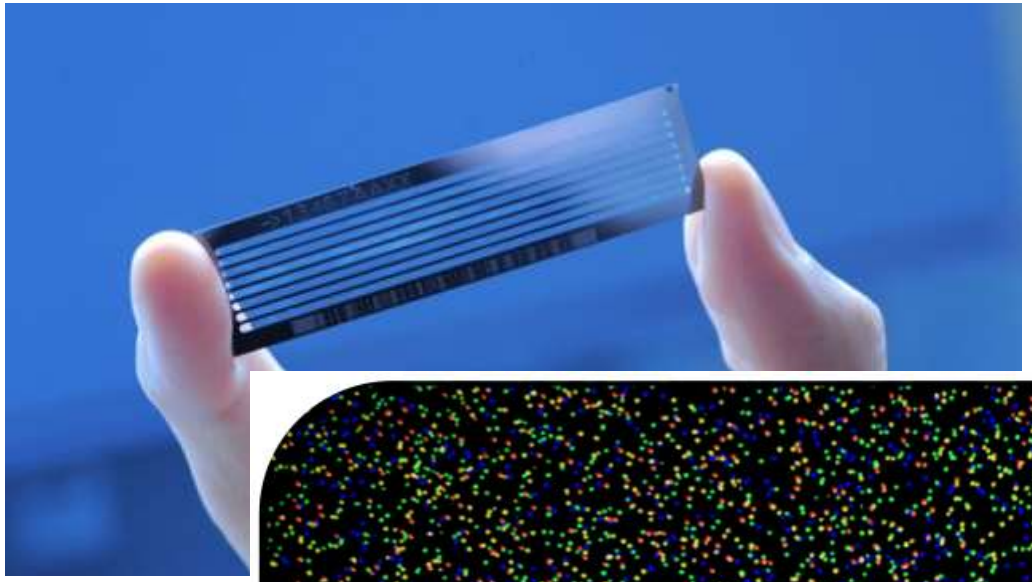
Illumina Devices

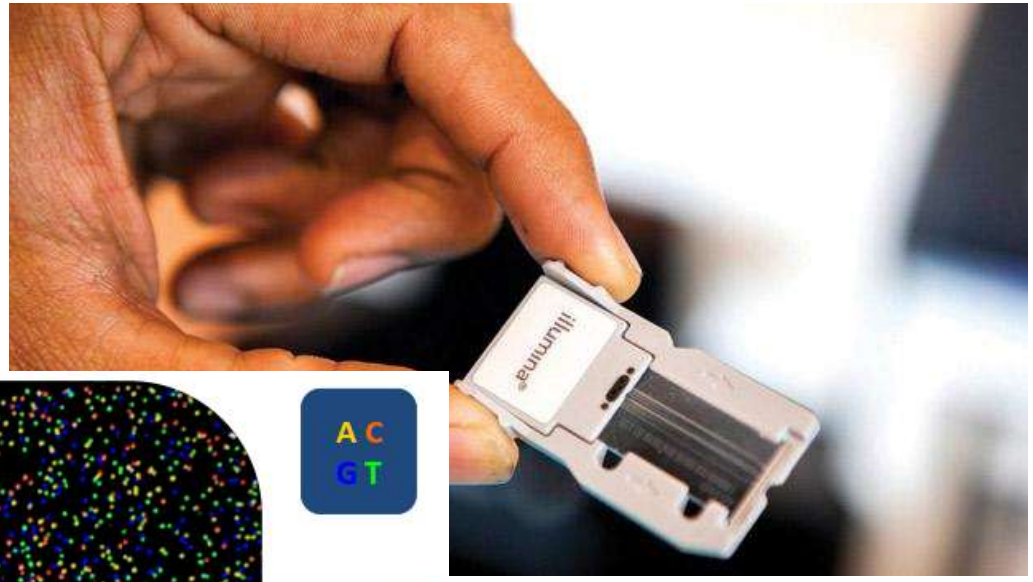
@ColauttiLab

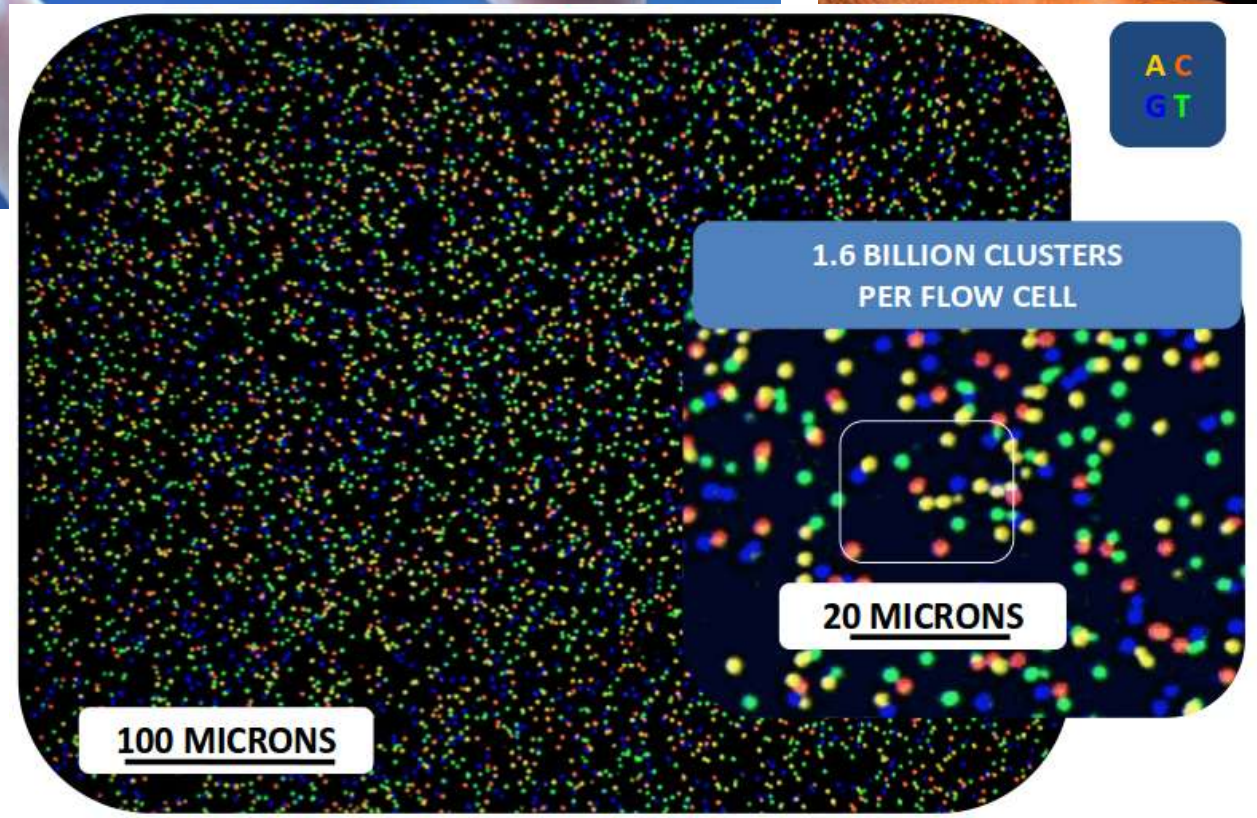
MiniSeq System



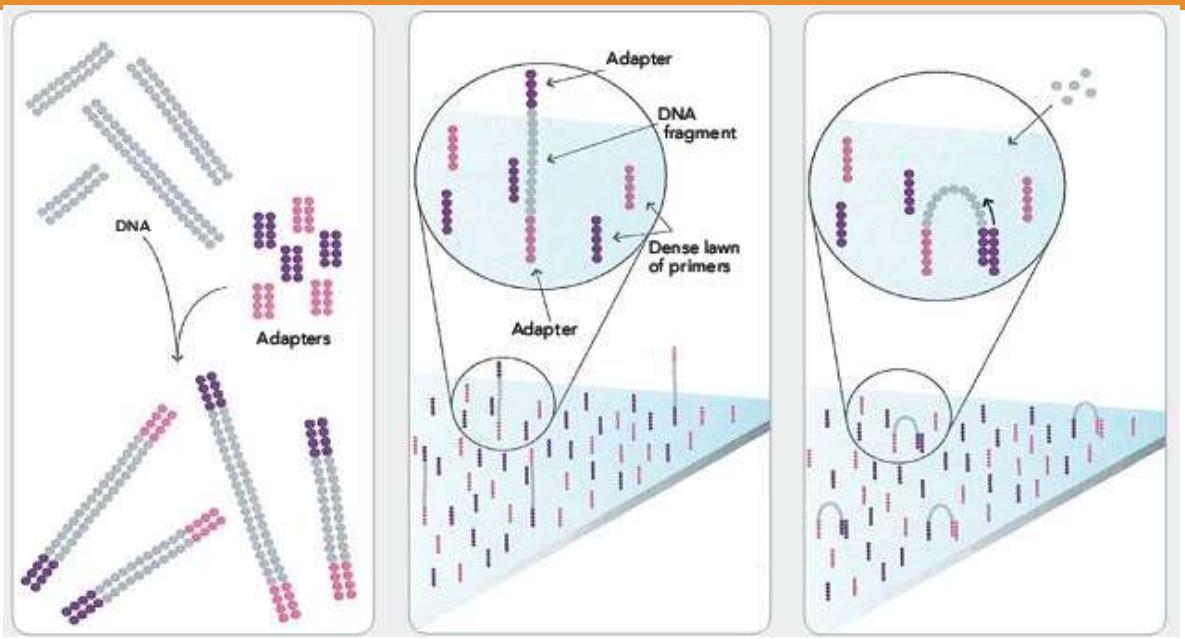
Illumina flow cells





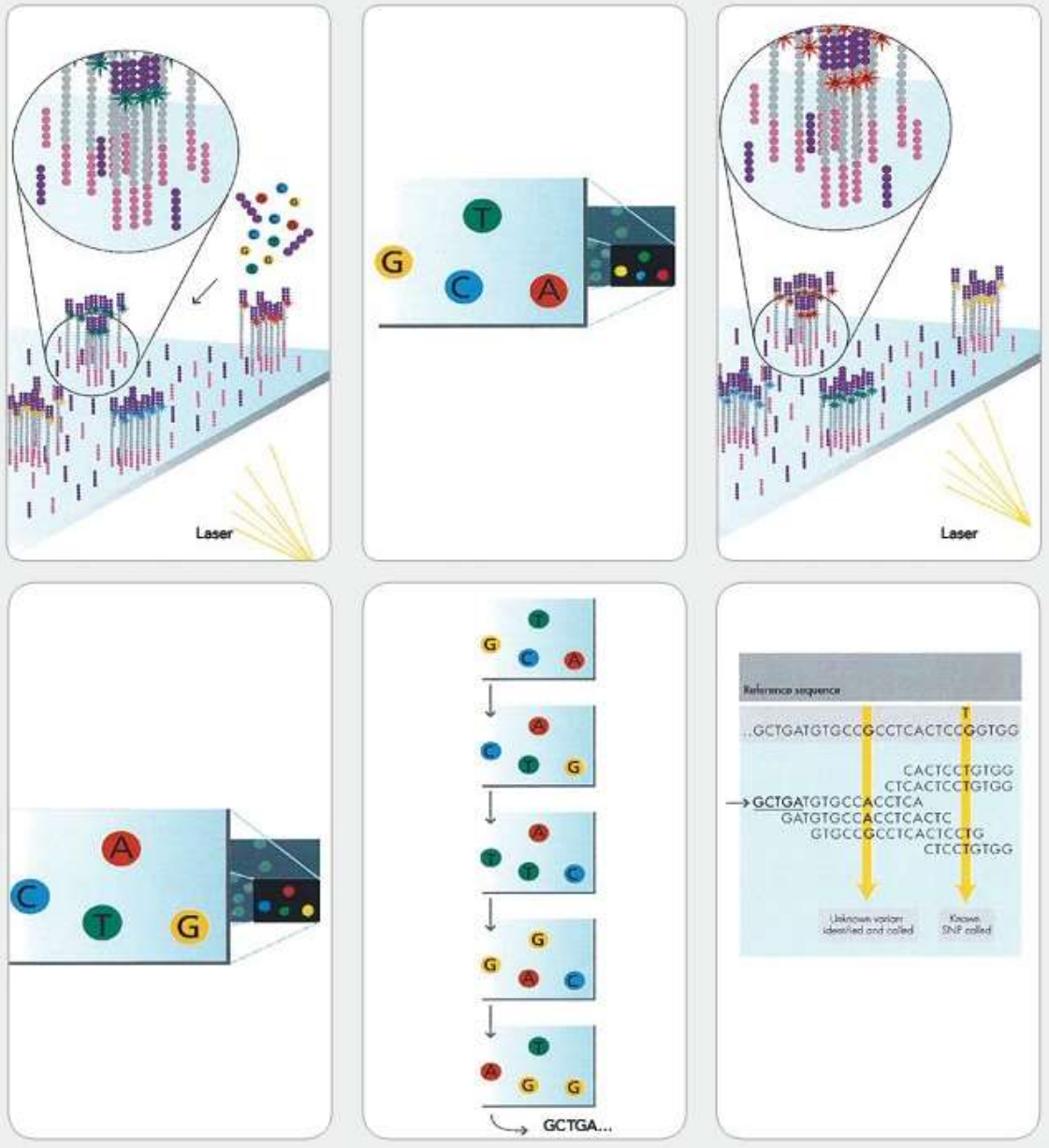


Illumina sequencing (formerly Solexa)



1. Prepare genomic DNA
2. Attach DNA to surface
3. Bridge amplification
4. Fragment become double stranded
5. Denature the double stranded molecules
6. Complete amplification

Illumina sequencing



7. Determine first base
8. Image first base
9. Determine second base
10. Image second base
11. Sequence reads over multiple cycles
12. Align data

Review

@ColauttiLab

Working in groups (15 mins):

- Stretch and divide into working groups
- Summarize sequencing-by-synthesis (SBS) with Illumina
- Review key concepts:
  1. How a flow cell works
  2. Contrast Sanger with SBS sequencing
- Try flowcharts or cartoons to simplify & summarize

BRAINSTORM:

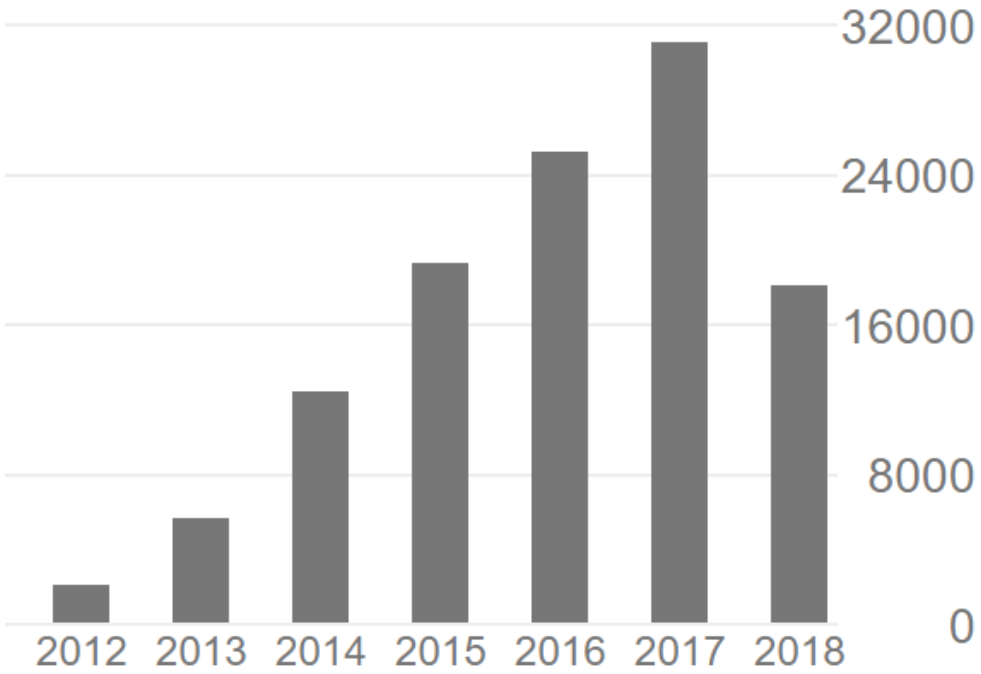
What are the main benefits & limitations of each technology?

Why is coding valuable for 2nd generation sequencing?

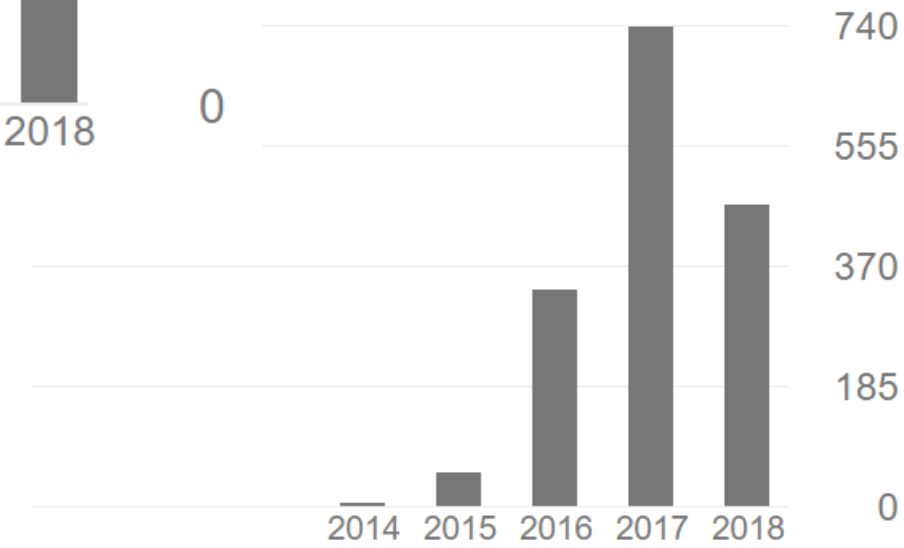
Is coding important?

@ColauttiLab

Web of Science (R Core Team)



Web of Science (R Studio)



R Introduction Tutorial

Field Excursion

@ColauttiLab

11:00

Groups 1 & 2: Boat Tour – Aquatic Sampling

Groups 3 & 4: QUBS Tour – Soil Sampling

12:00 LUNCH

13:00

Groups 3 & 4: Boat Tour – Aquatic Sampling

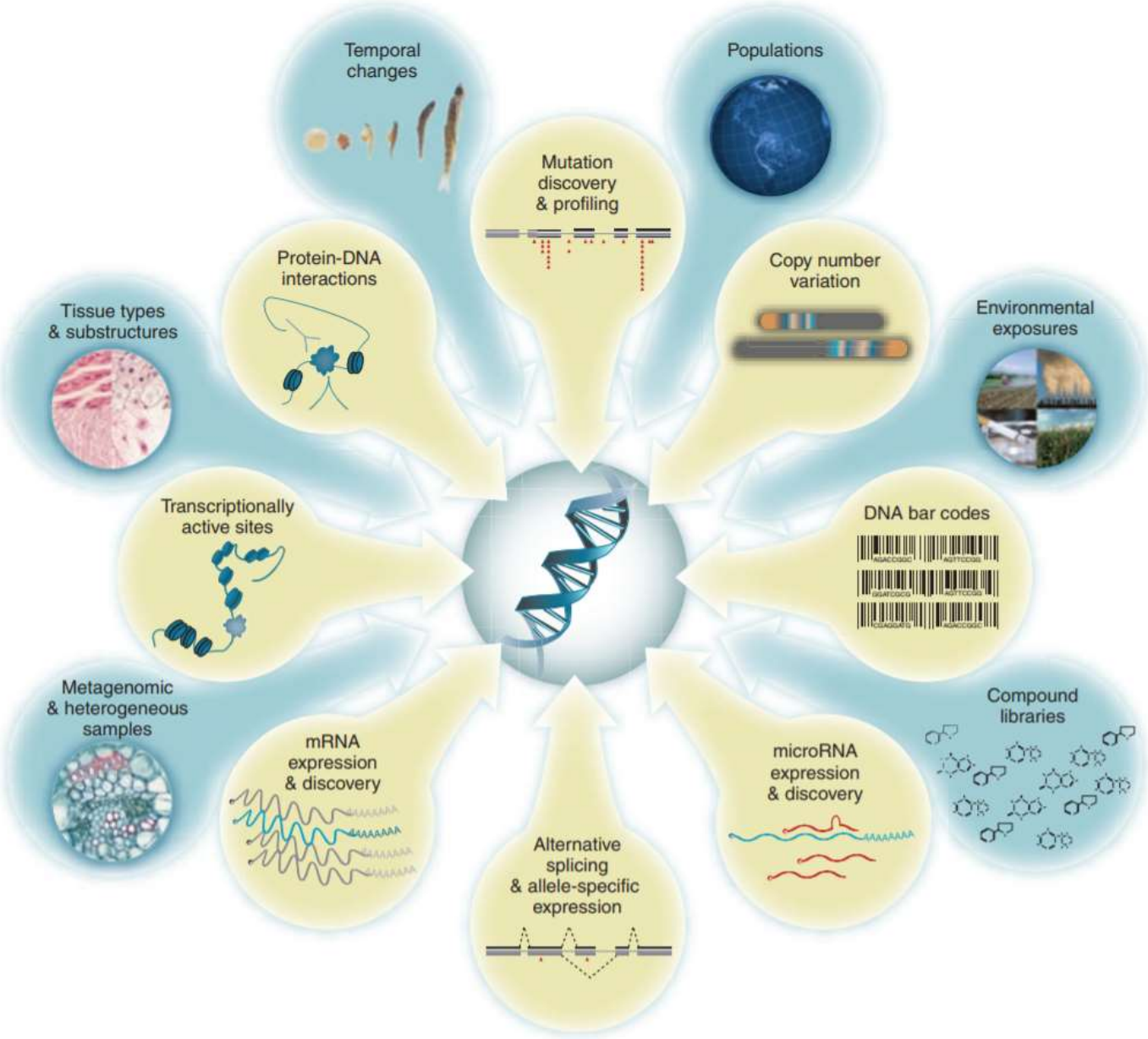
Groups 1 & 2: QUBS Tour – Soil Sampling

14:00 PRESENT

Apply your expertise in (eco)toxicology to formulate a question and design **a field sampling** protocol. *Consider using hand-drawn figs.* What are your major concerns or considerations?

Methods & Applications

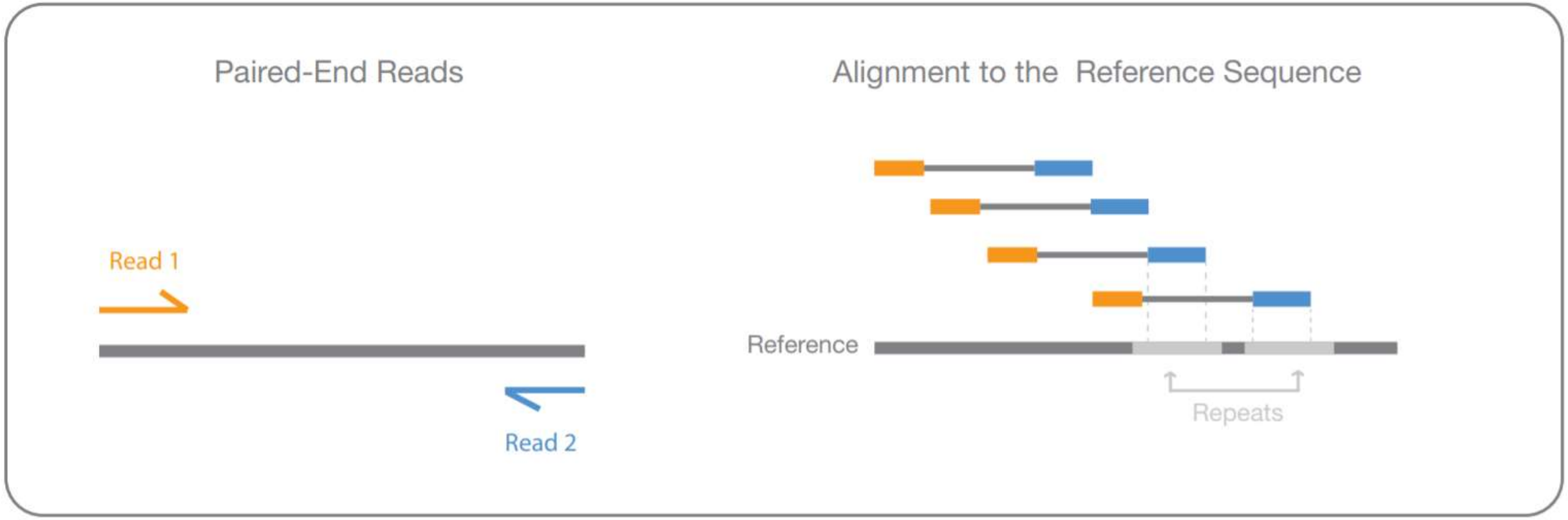
@ColauttiLab



Kahvejian et al. (2008)

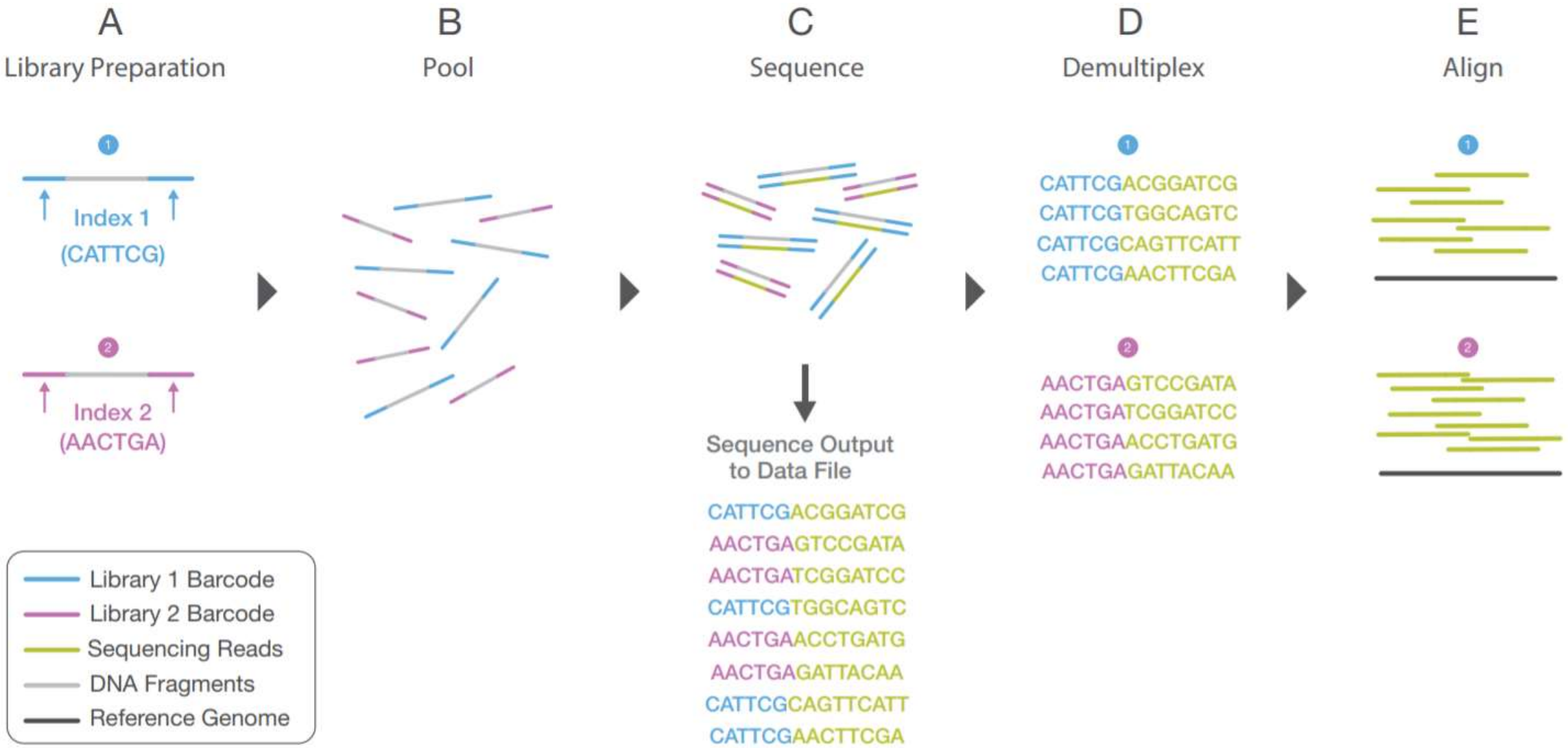
Single Read (SR) vs Paired-End (PE) Reads

@ColauttiLab

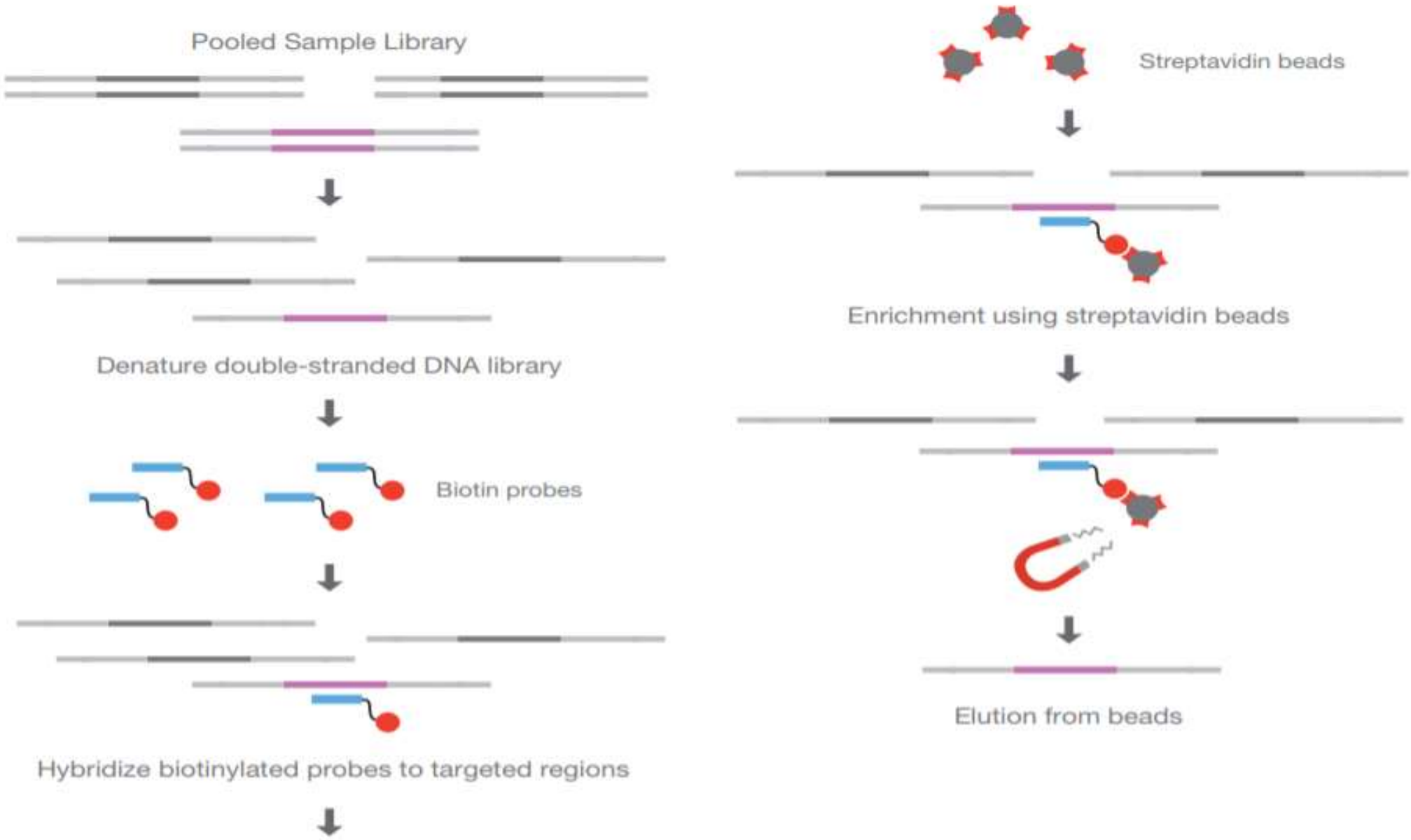


Multiplexing reduces per-sample costs

@ColauttiLab



Target Enrichment (Exome capture)



Other methods

@ColauttiLab

Genotype-by-sequencing

- e.g. RADSeq, POOLSeq
- Population genomics
- Association mapping (QTL, GWAS)

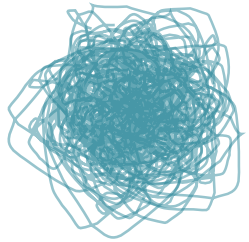
Epigenetics

- DNA methylation sites (Methy-Seq or Bisulfite sequencing)
- Protein-binding sites (CHIP-Seq)

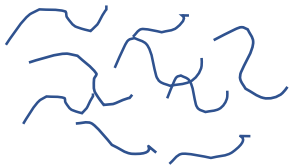
Whole Genome Shotgun Sequencing

@ColauttiLab

DIY Reference Genome Assembly (~$10,000)



Extract



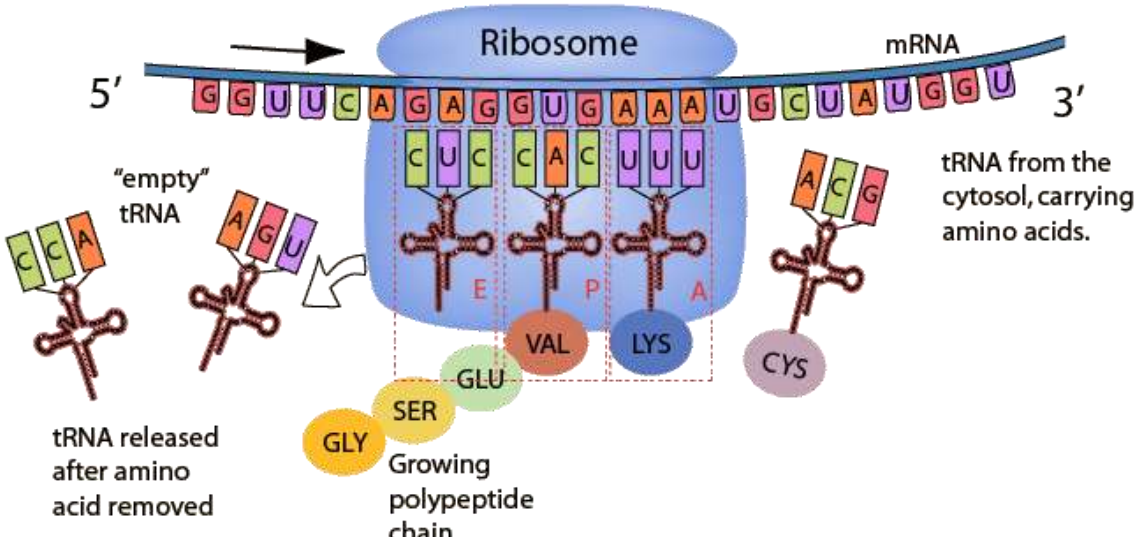
Fragment

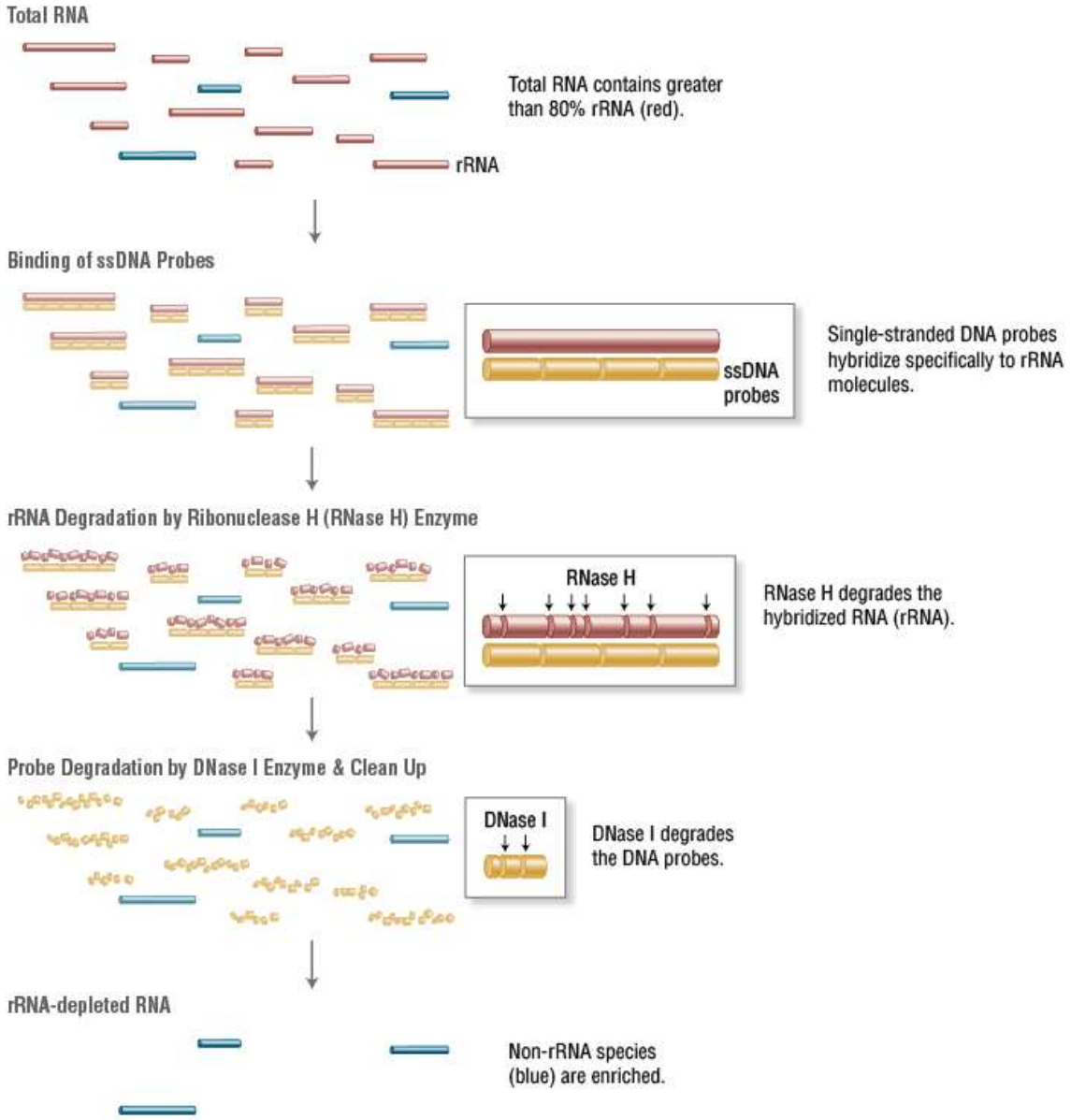
Sequence

Assemble

rRNA Depleted RNA sequencing

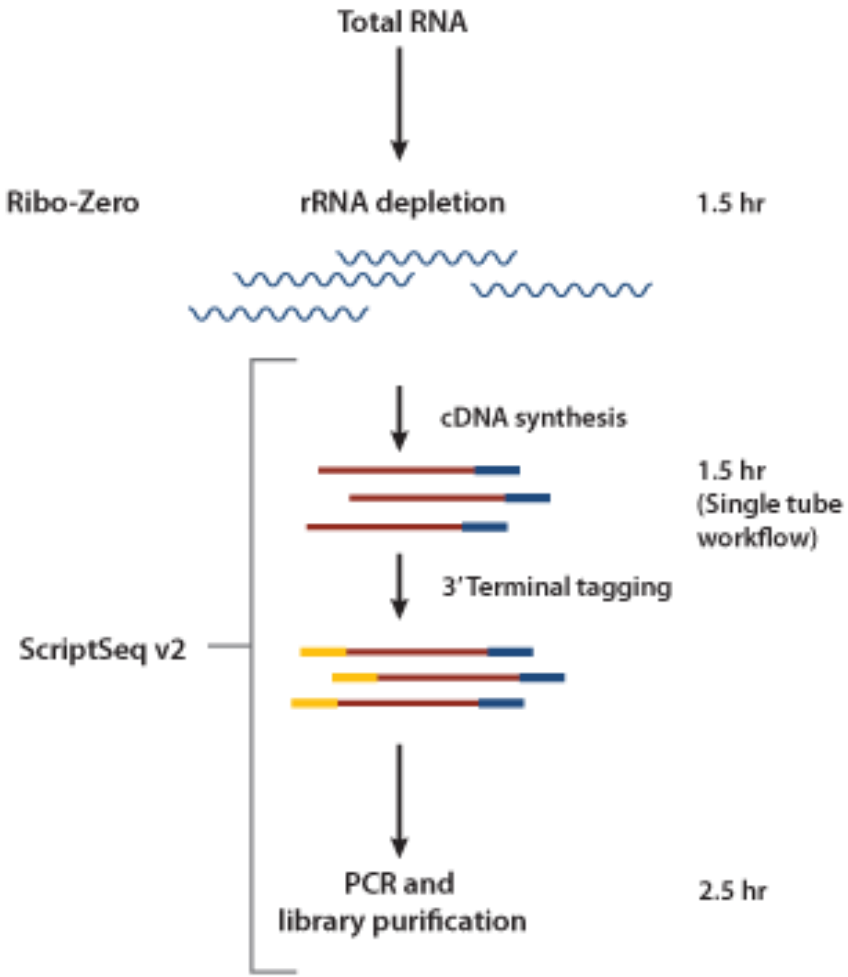
@ColauttiLab

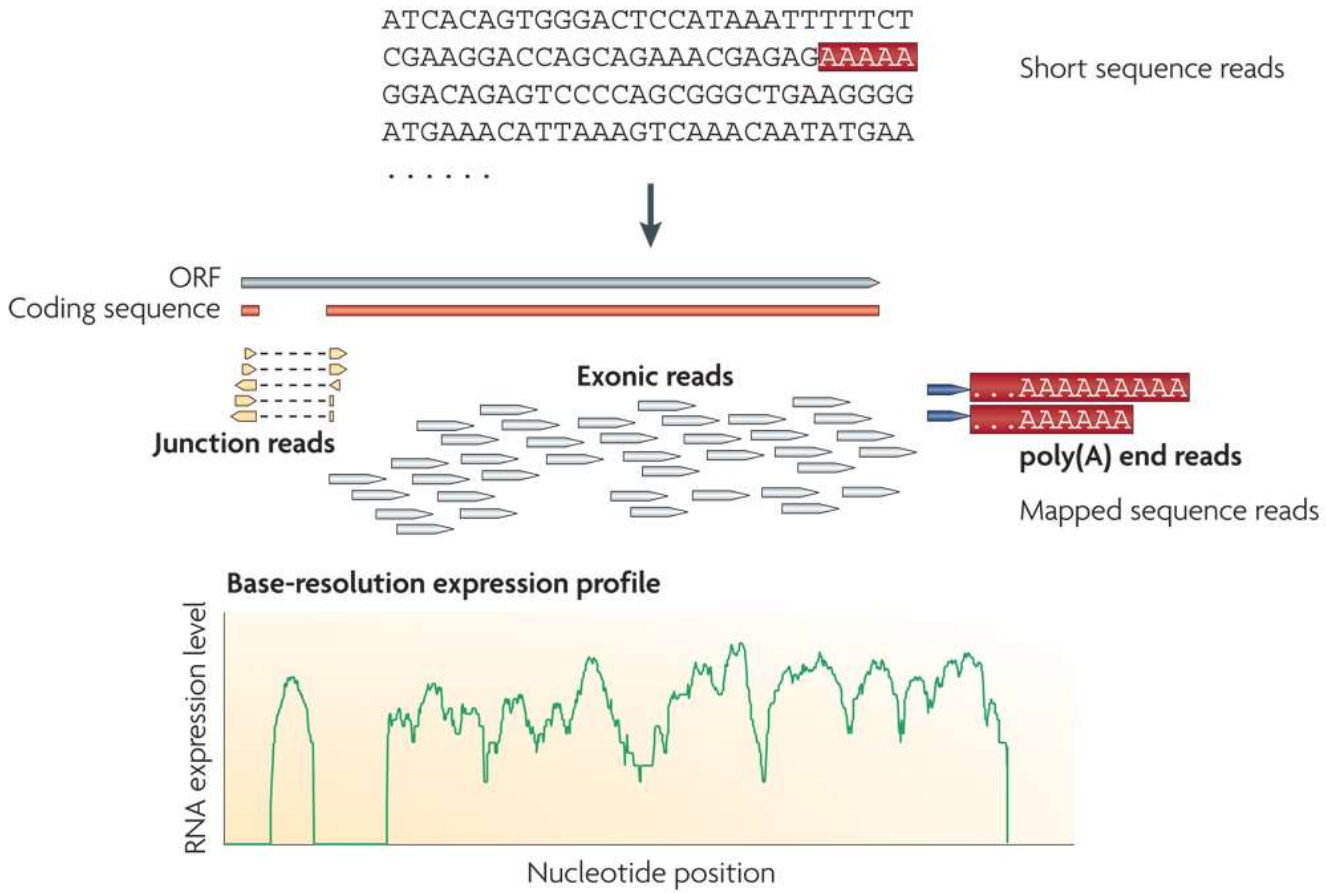




Most intracellular RNA
Is ribosomal RNA

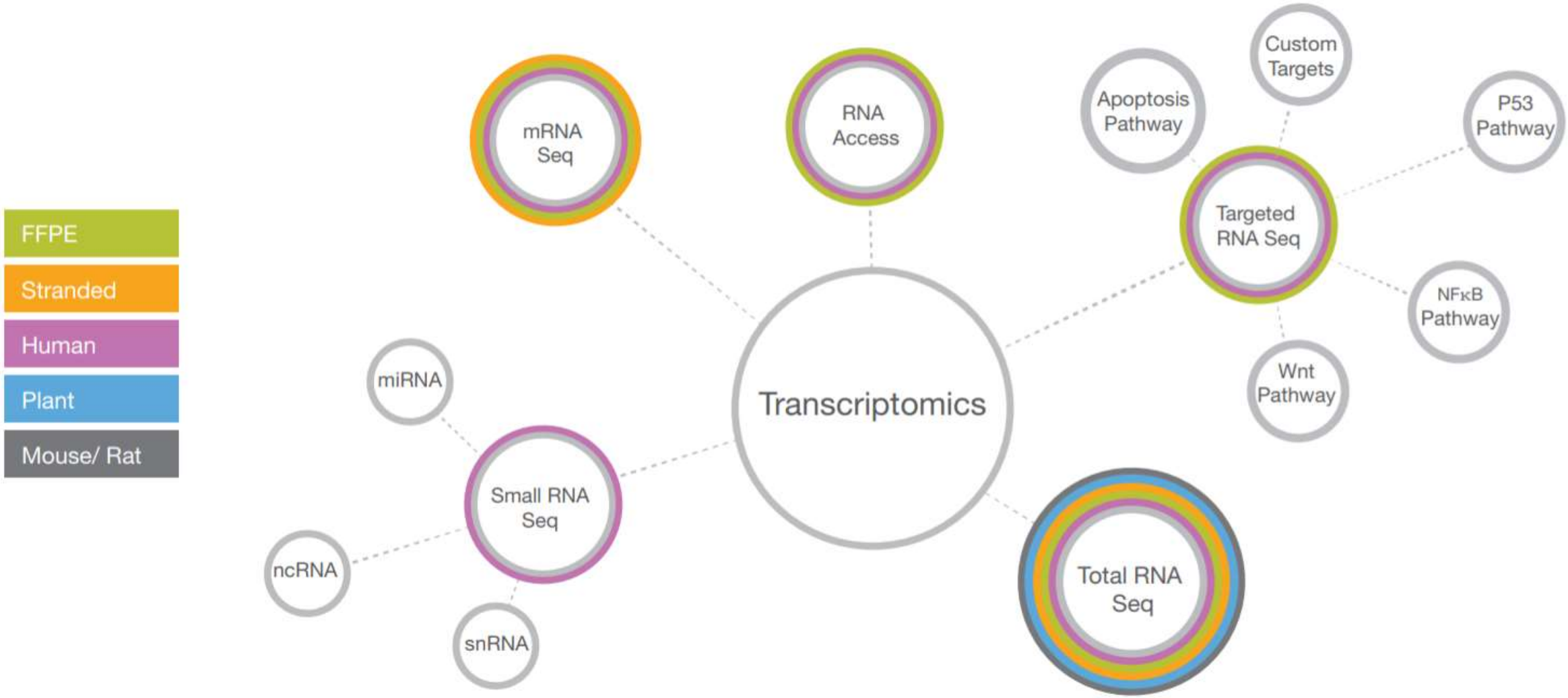
rRNA Depleted transcriptomics





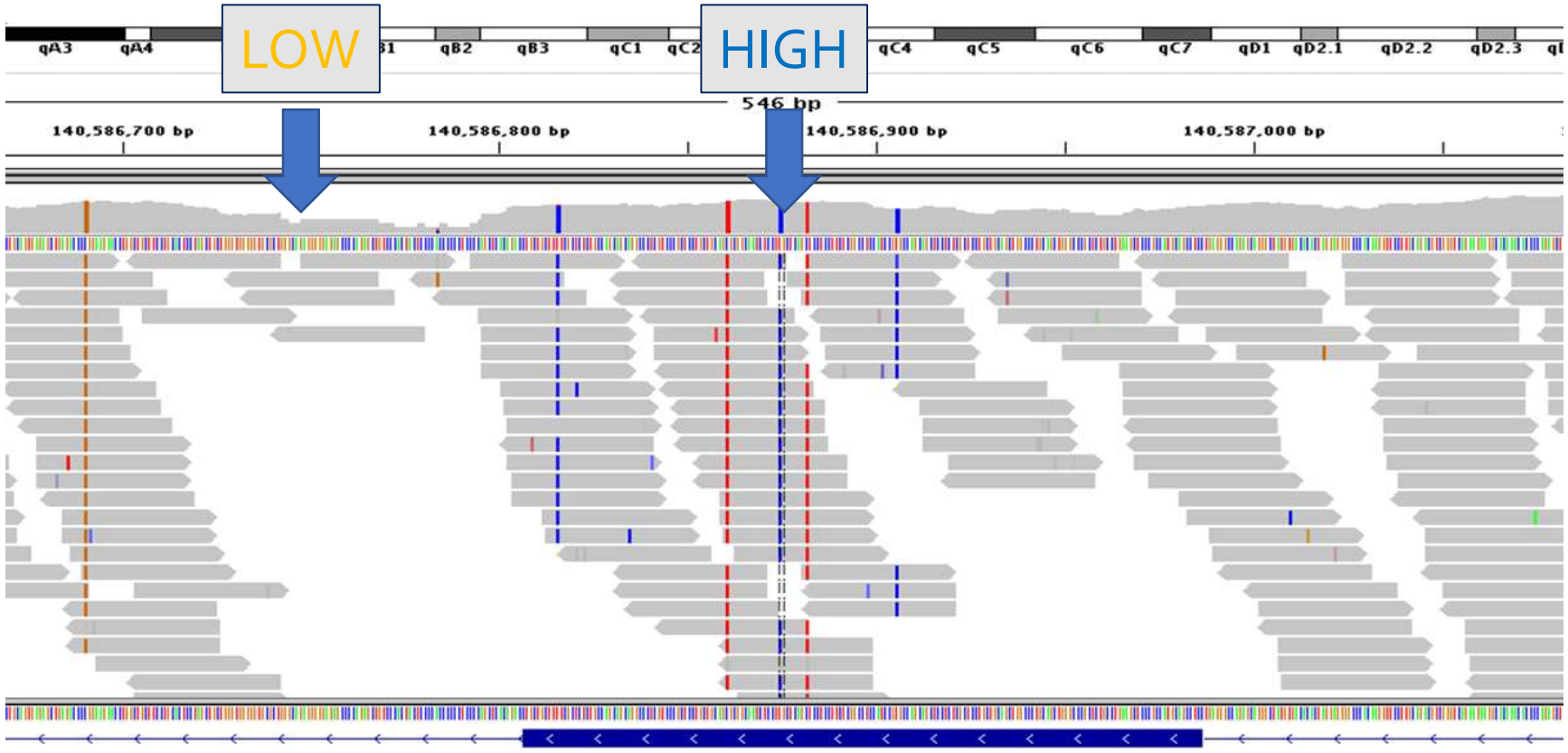
rRNA Depleted transcriptomics

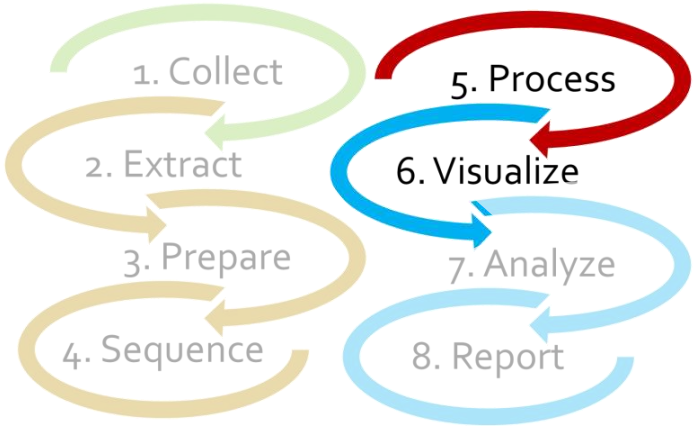
@ColauttiLab



Alignment coverage (e.g. RNA-Sequencing)

@ColauttiLab



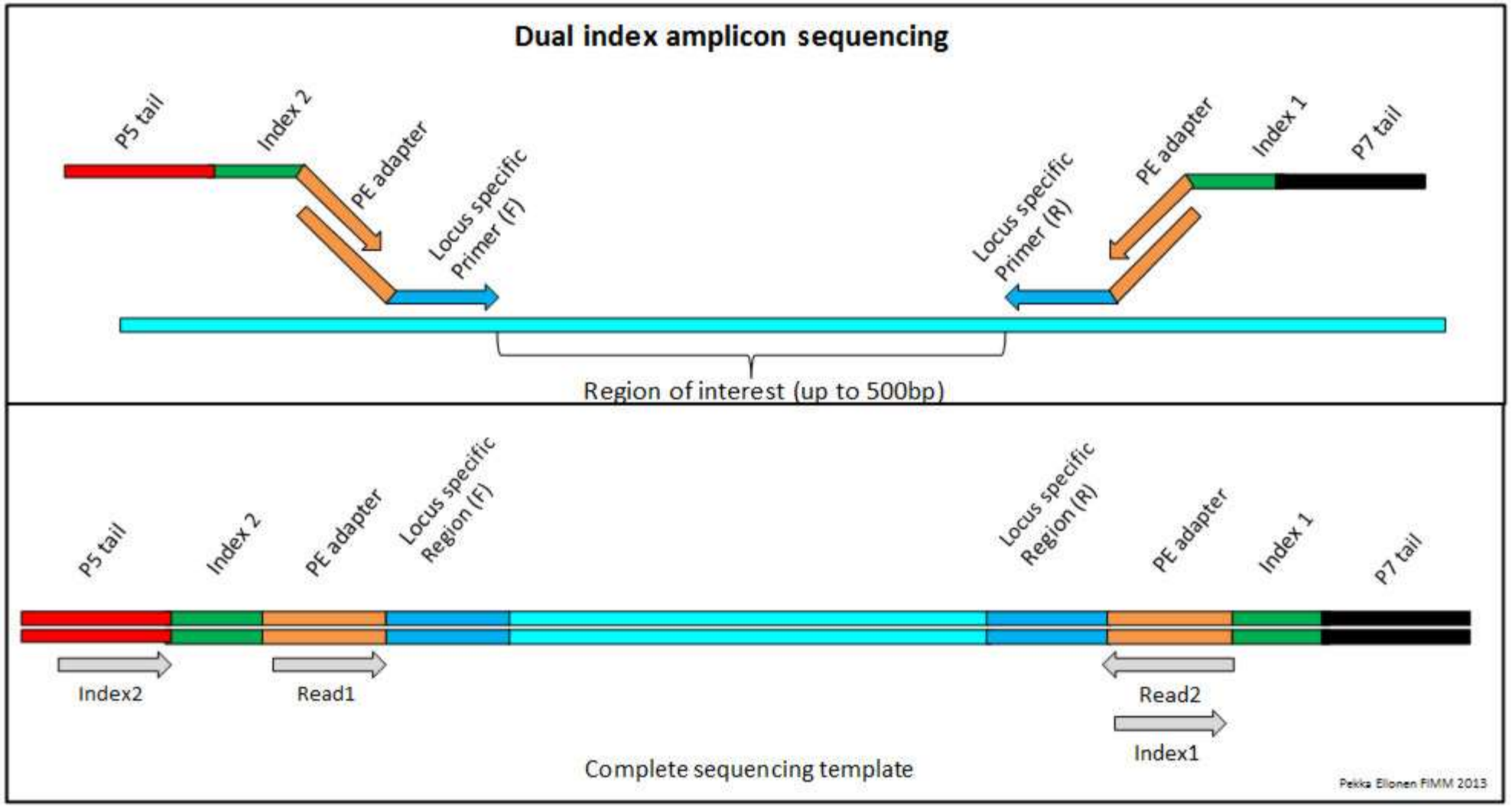


@ColauttiLab

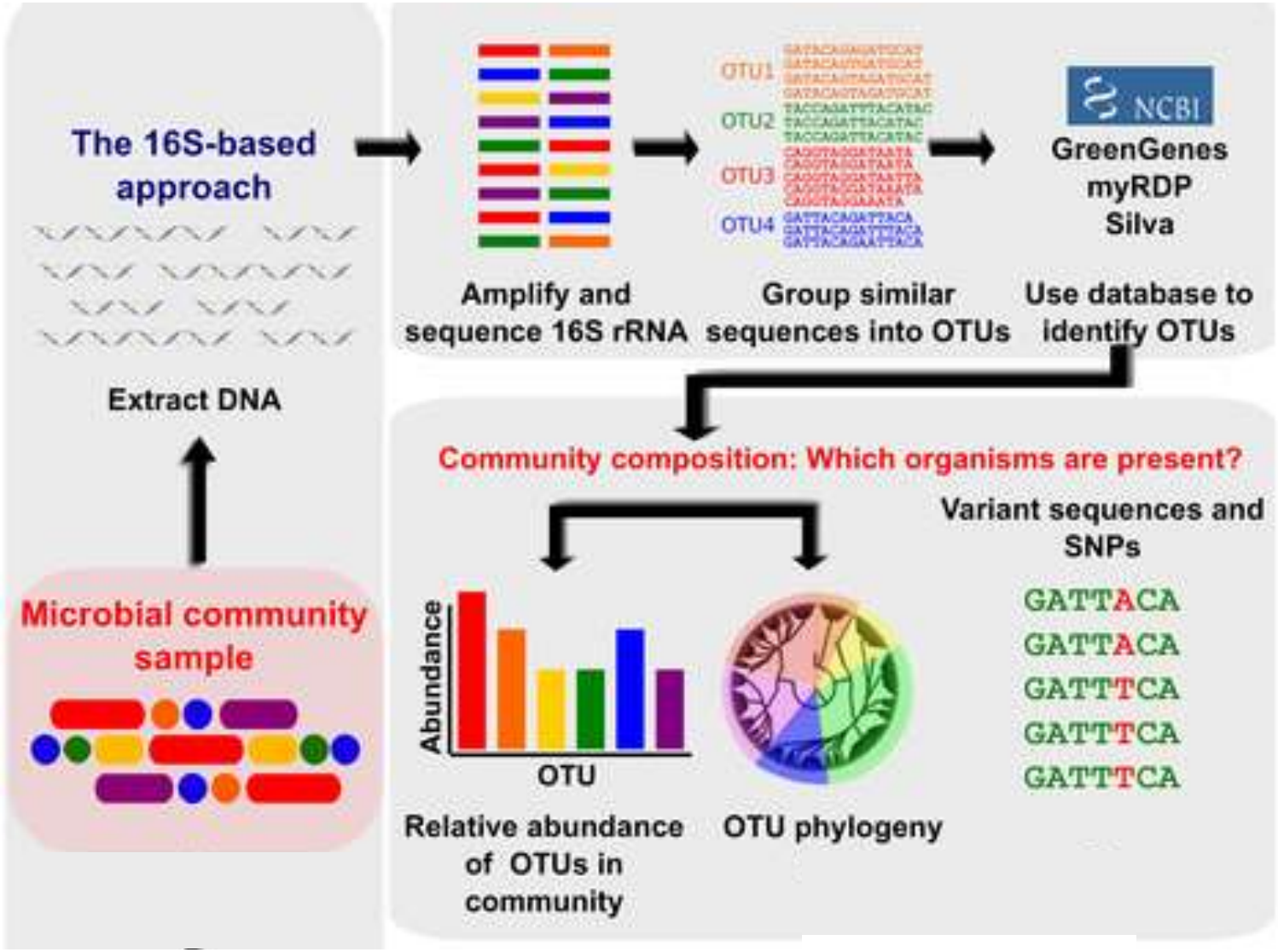
Transcriptome Tutorial

Amplicon sequencing (metabarcoding)

@ColauttiLab



Amplicon sequencing (metabarcoding)



OUT = Operational Taxonomic Unit

Review

@ColauttiLab

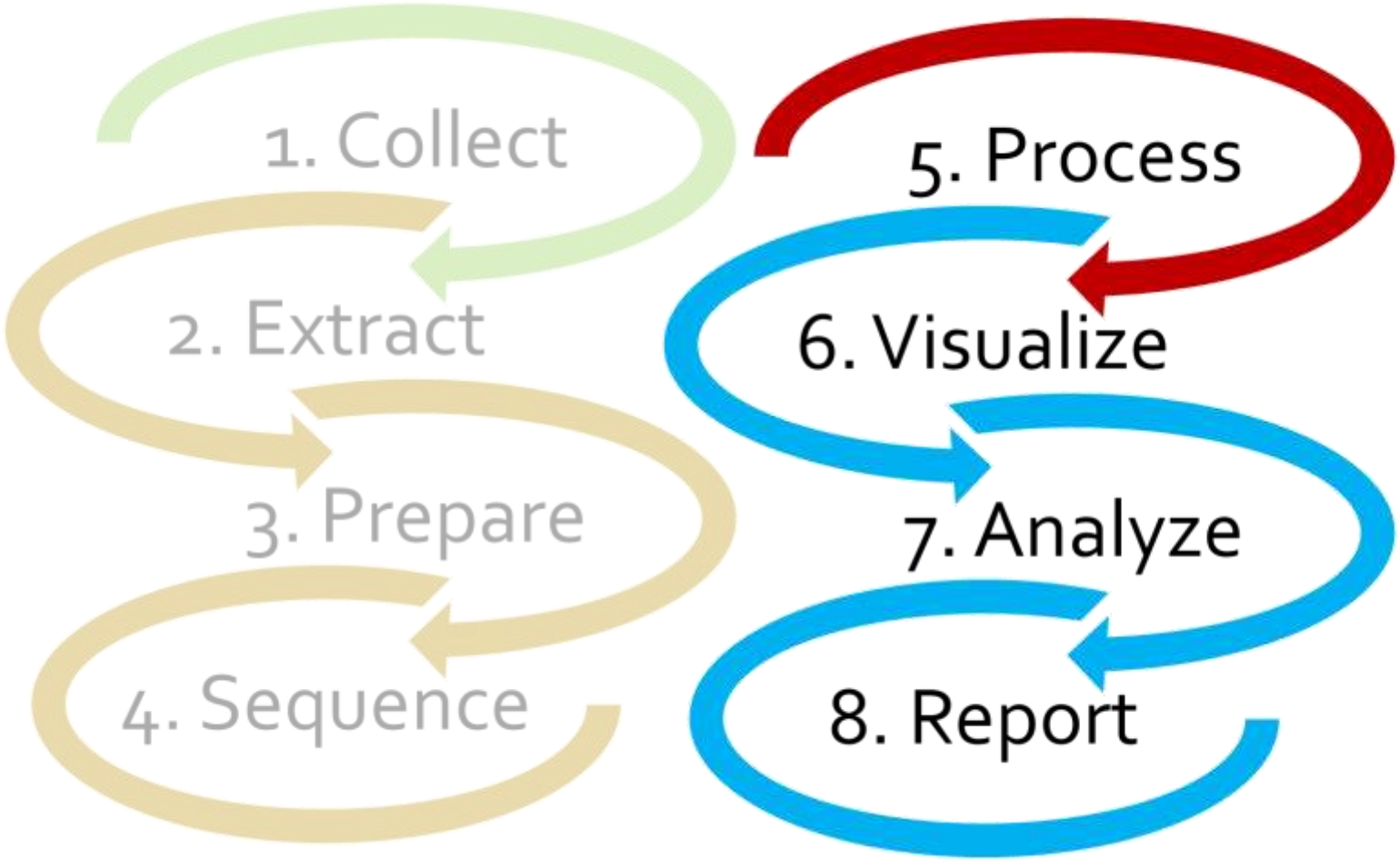
Working in groups of 2-4 (15 mins):

- Stretch and divide into working groups
- Review key concepts:
  1. SR vs PE sequencing
  2. Target enrichment
  3. Amplicon sequencing
  4. RNA sequencing
- Try flowcharts or cartoons to simplify & summarize

Which applications are relevant to (eco)toxicology? Explain.

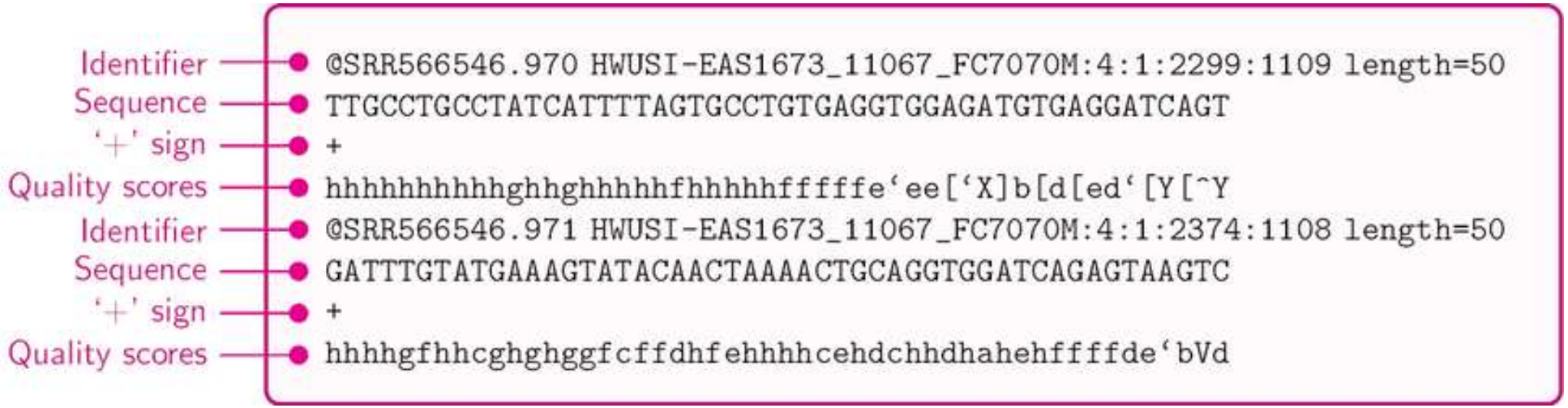
Bioinformatics: The **true cost** of NGS

@ColauttiLab



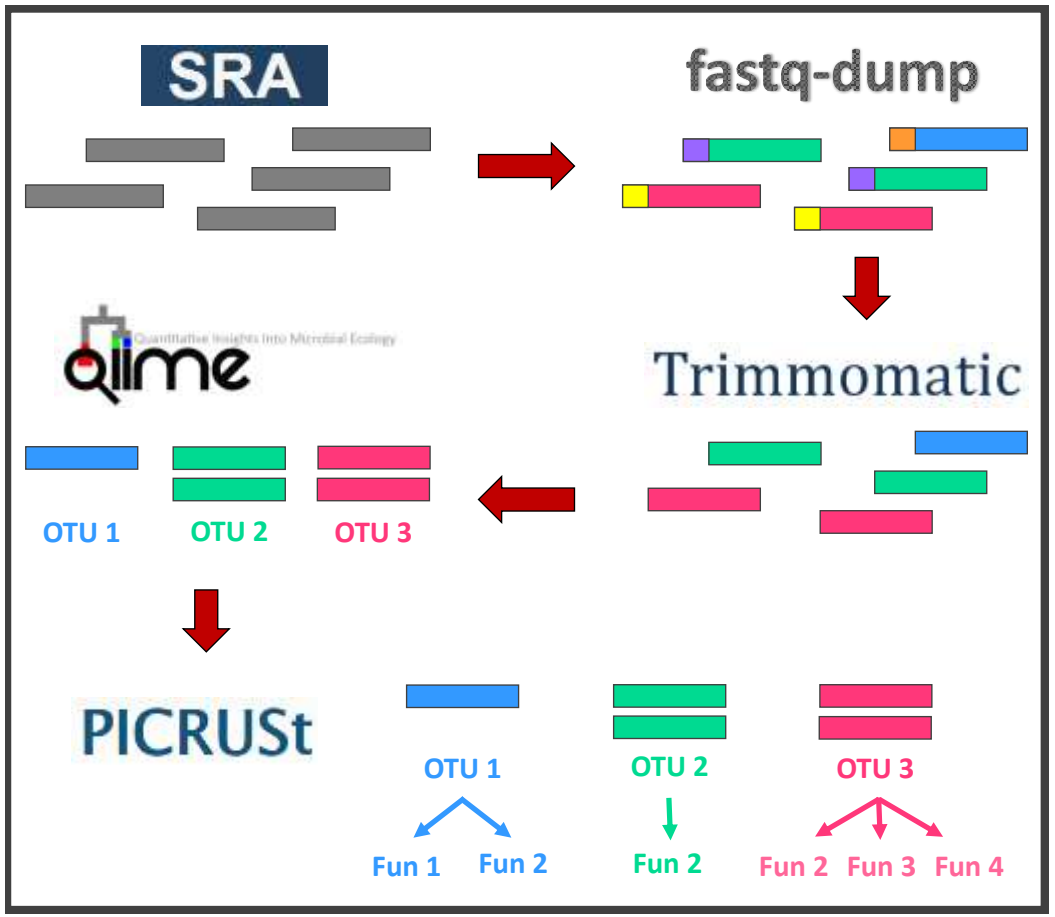
FASTQ file

@ColauttiLab

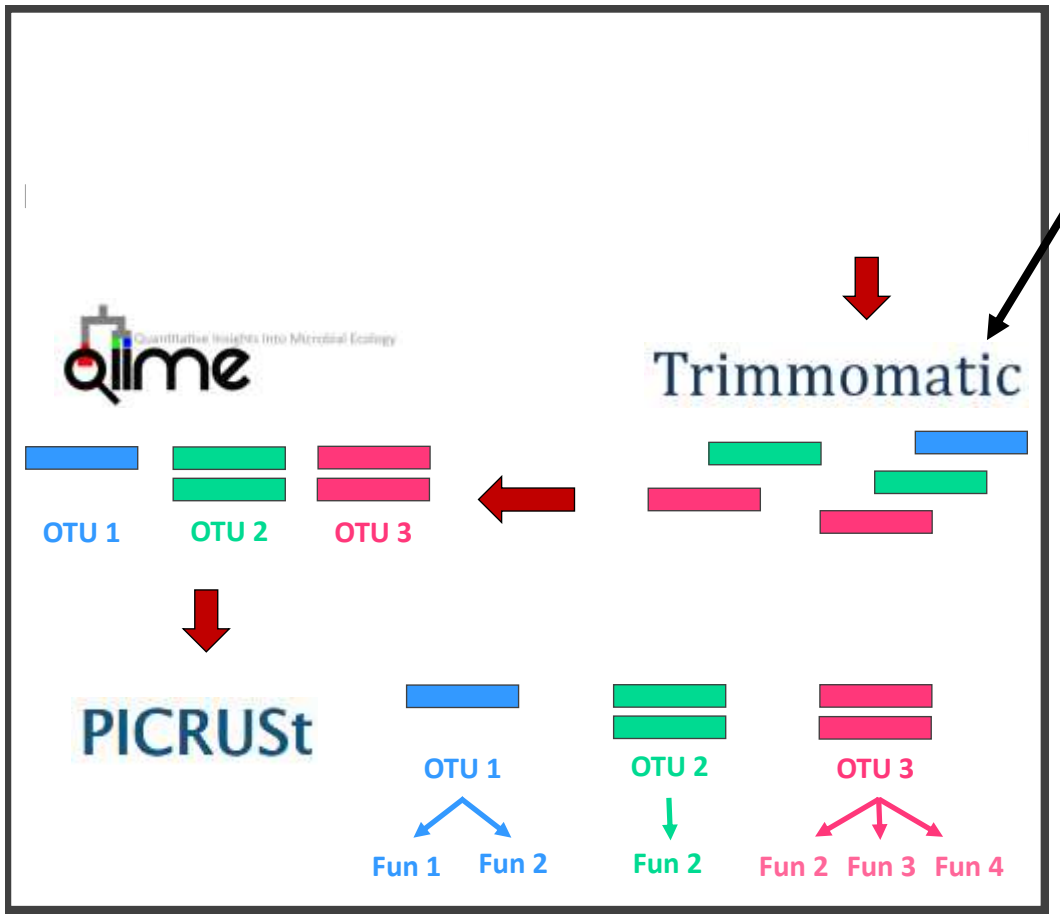


Example of a pipeline

@ColauttiLab



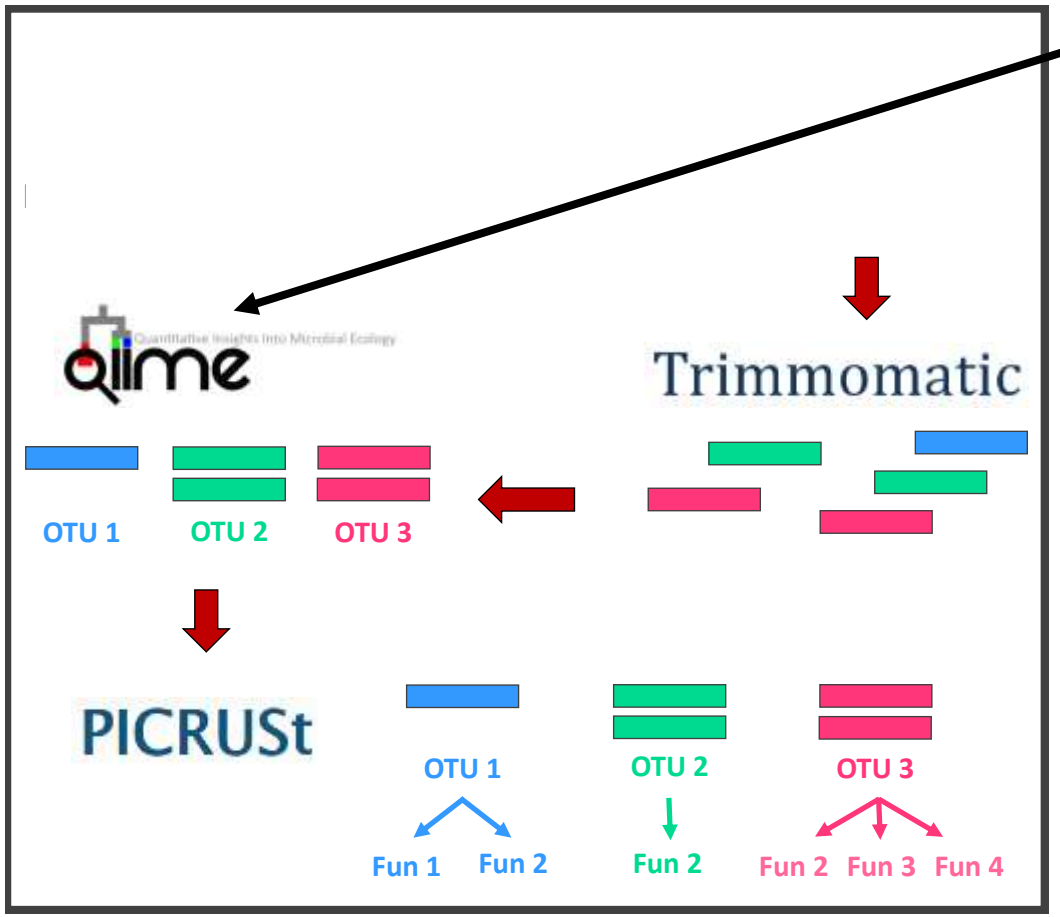
Example of a pipeline



The current trimming steps are:

- Cut adapters
- Cut bad parts of a read
  OR
- Cut the read to a specified length
- Drop bad reads

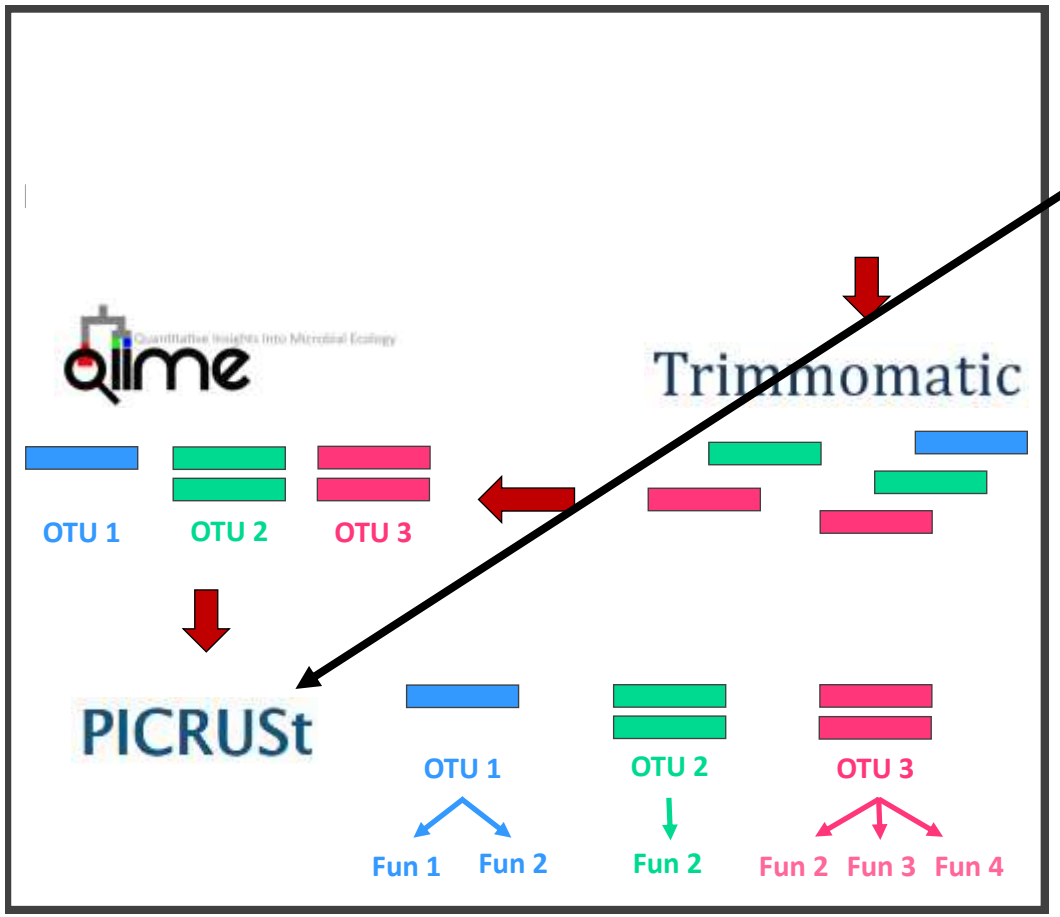
Example of a pipeline



Assign the reads to a "species"

Or OPERATIONAL TAXONOMIC UNIT

Example of a pipeline



A little extra:

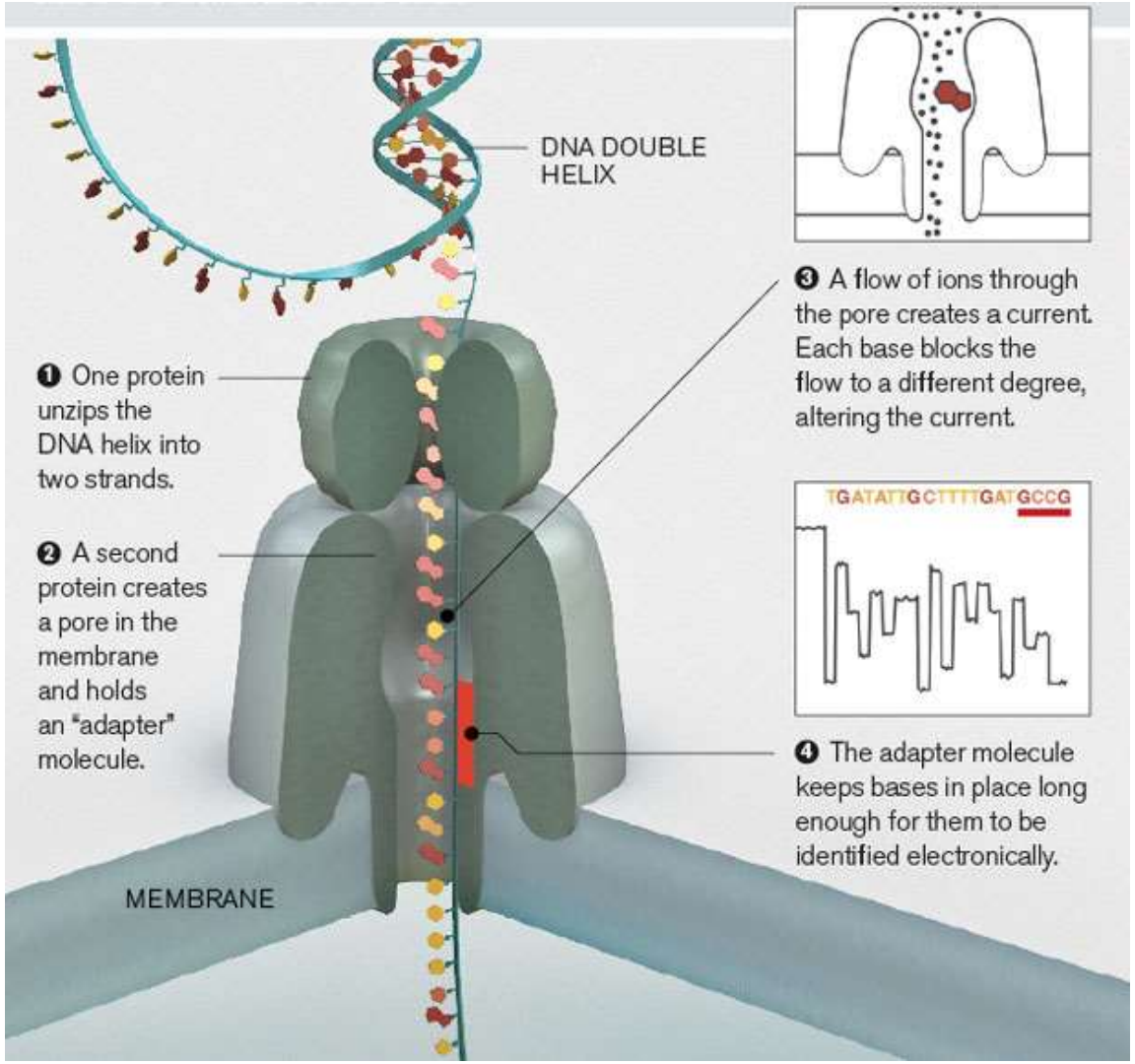
Assign a "function" to the OTU

@ColauttiLab

Metabarcoding Tutorial

3rd Gen: Nanopore Sequencing

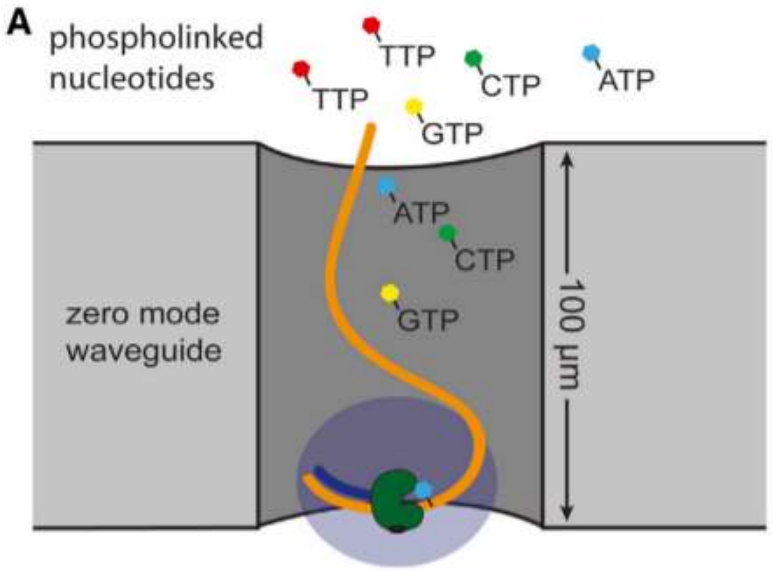
@ColauttiLab

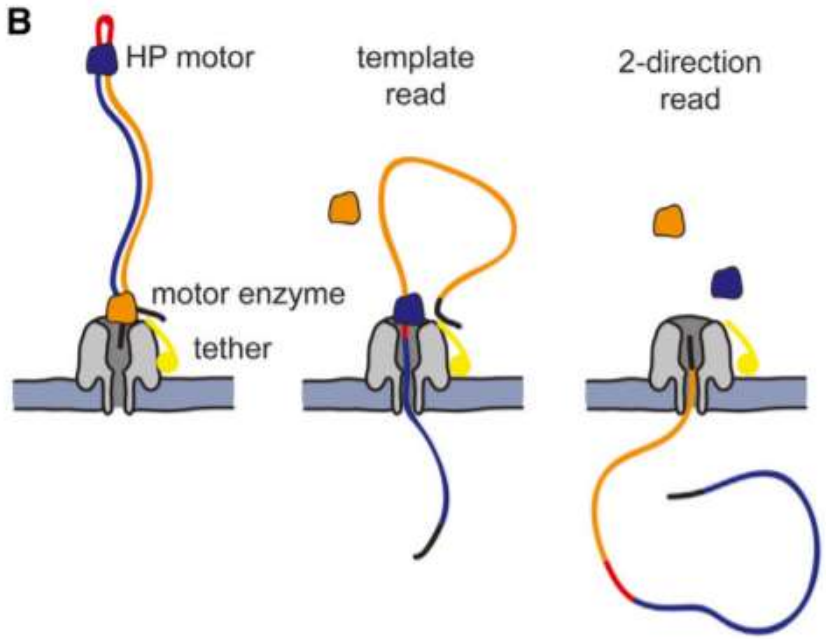


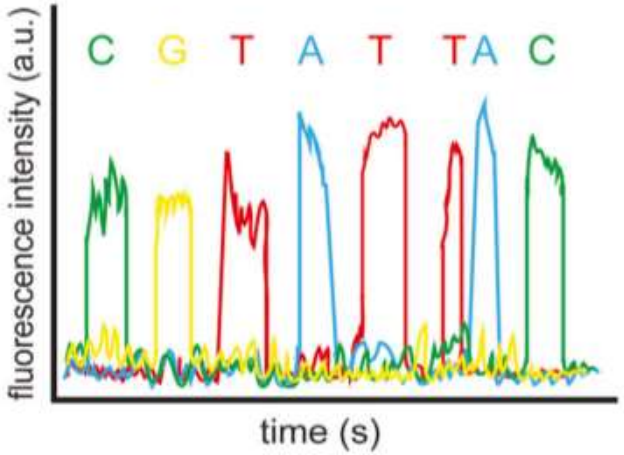
* Coming soon

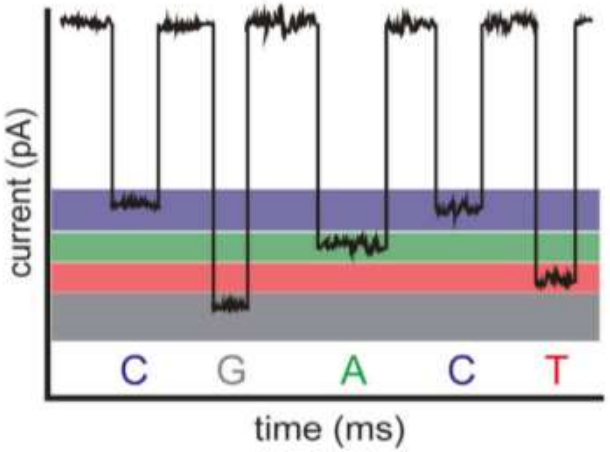
3rd Gen: Nanopore Sequencing

@ColauttiLab









Qian-long et al. (2014)

Nanopore Sequencing Comparison

@ColauttiLab

Platform	Instrument	Mreads	Length (bp)	Gbp	Type
Illumina	NovaSeq 6000 S4	10000	300	3000	SR & PE
Illumina	NextSeq 500 High-Output	400	300	120	SR & PE
Illumina	HiSeq X	375	300	112.5	PE
Illumina	HiSeq High-Output v4	250	250	62.5	SR & PE
Illumina	MiSeq v3	25	600	15	SR & PE
Illumina	MiniSeq High-Output	25	300	7.5	SR & PE
Oxford Nanopore	MinION		1M+	20	SR
Oxford Nanopore	PromethION			1000	SR
Ion	Proton I	60	200	12	SR
Ion	PGM 318	4	400	1.6	SR
Ion	PGM 316	2	400	0.8	SR
Ion	PGM 314	0.4	400	0.16	SR
PacBio	PacBio Sequel	0.37	20000	7.4	SR
PacBio	PacBio RS II (P6)	0.055	15000	0.825	SR
Roche 454	GS FLX+ / FLX	0.7	700	0.49	SR
SOLiD	5500xl W	267	100	26.7	SR & PE

Nanopore Demo