

Module 3

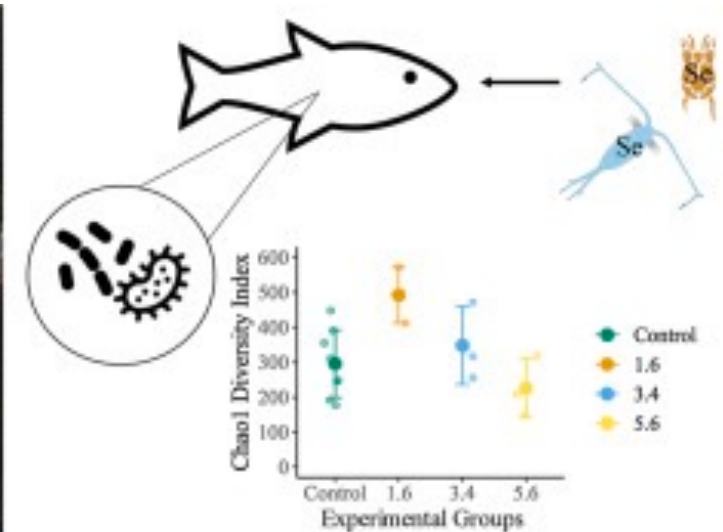
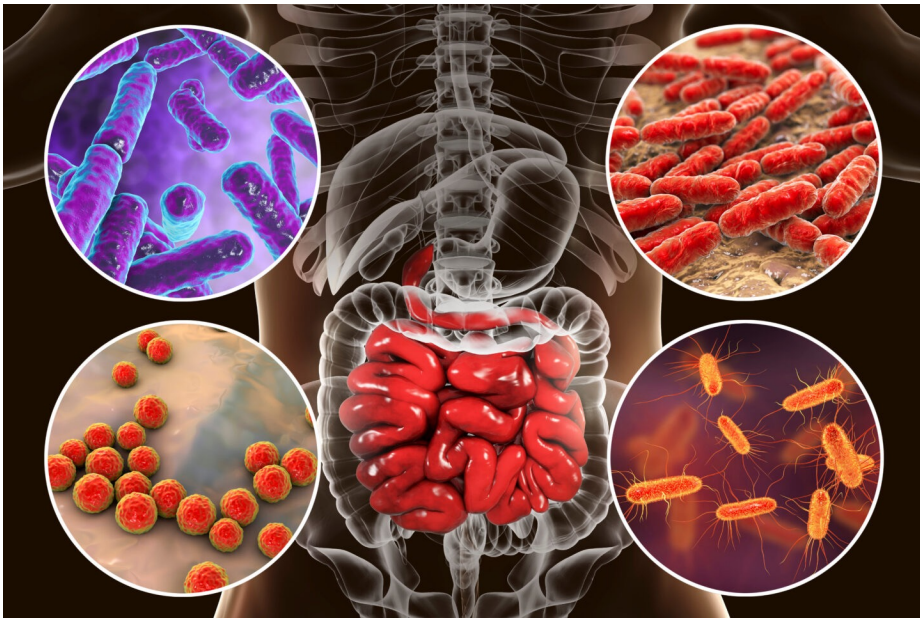
Measuring the microbiome using amplicon sequencing

Module Outcomes

1. Define what the microbiome is and why it is important
2. Describes the steps involved in Amplicon Sequencing
3. Explain what are the possible outputs (ie. files and data) from amplicon sequencing

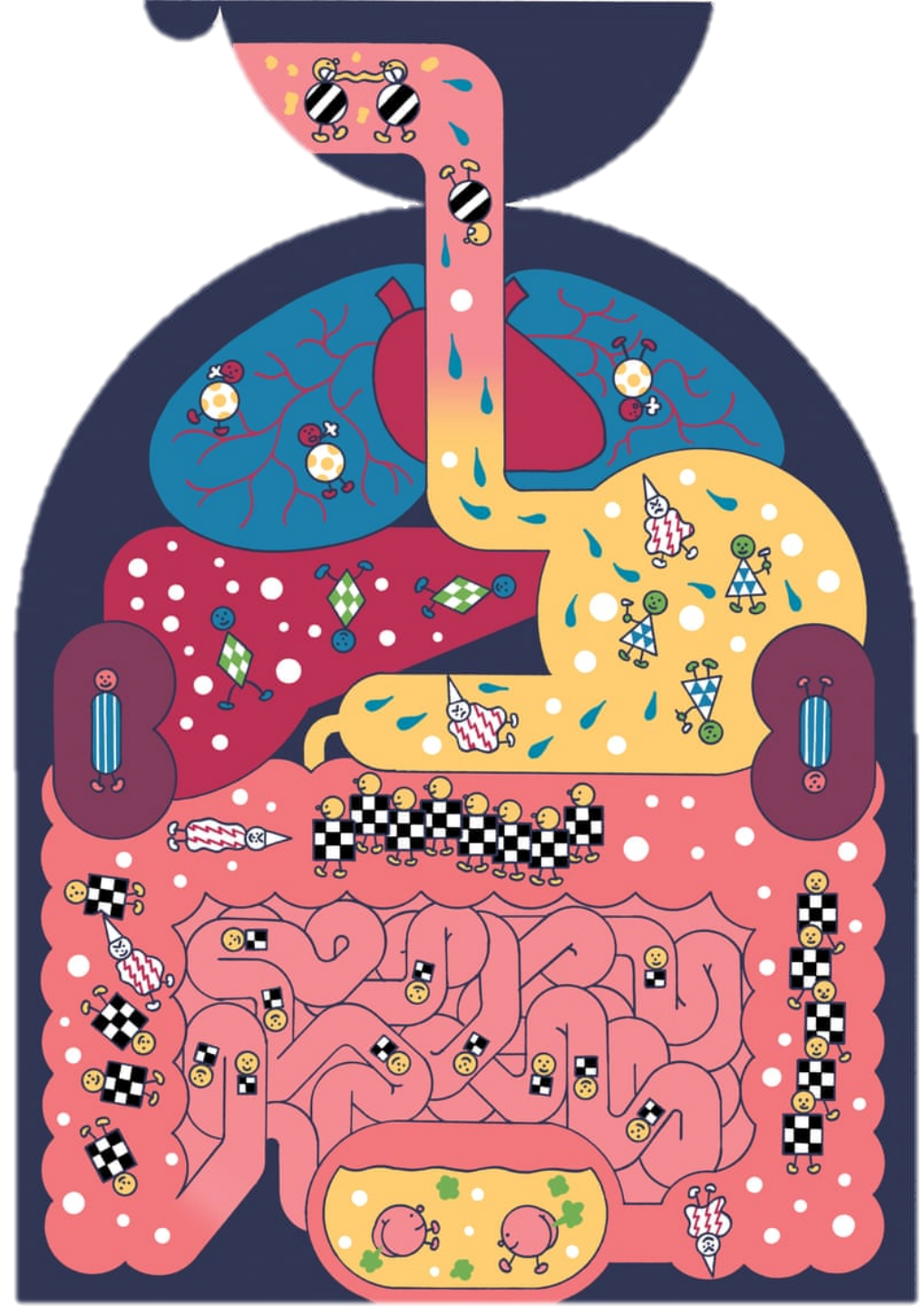
Microbiota

Community of microorganisms living in a certain environment



Microbiome


- Combined genetic material of all the microbes in the microbiota
- From this you can infer the different species and overall diversity of the community



NIH Human Microbiome Project


- Began in 2008 as an extension to the Human Genome Project
- Goal: to analyze the microbial communities in the 5 different parts of the body in over 300 individuals
 - Skin, mouth, nose, colon, and vagina
- Project lasted 5 years and cost \$150 million US dollars
- <https://www.hmpdacc.org/>

NIH Human Microbiome Project




Characterization of the microbiomes of healthy human subjects at five major body sites using 16S and metagenomic shotgun sequencing.

Enter HMP1



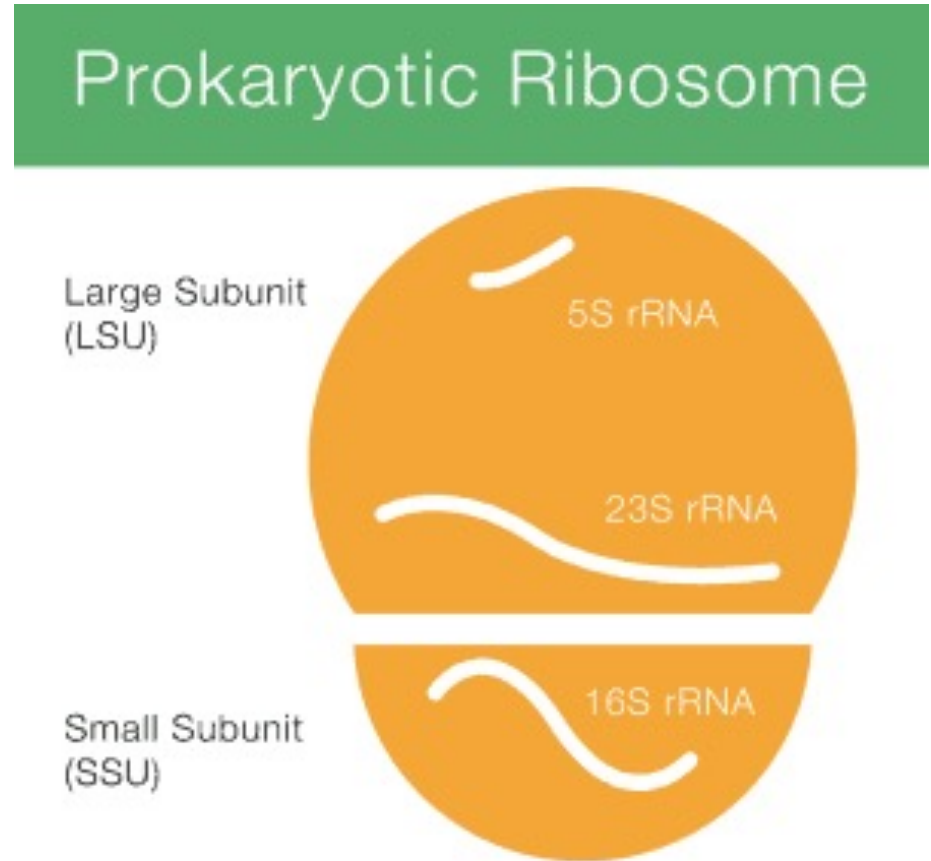
Characterization of microbiome and human host from three cohorts of microbiome-associated conditions, using multiple 'omics technologies.

Enter iHMP

Enter HMP Integrated Portal 

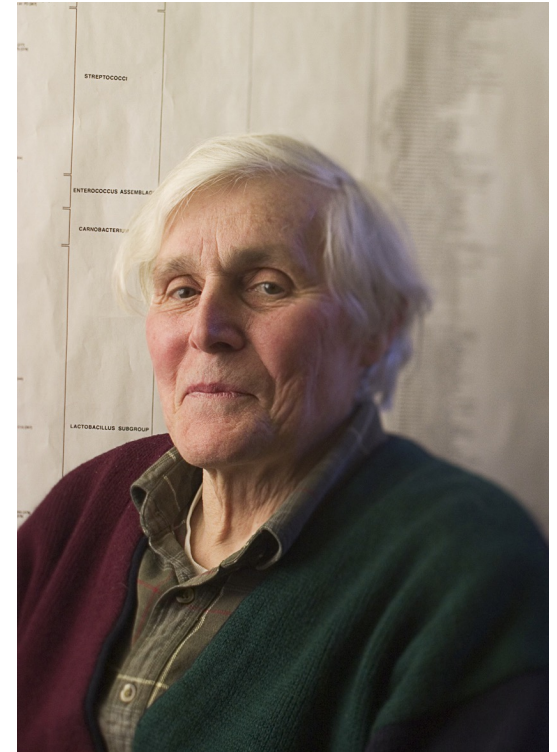
What is the 16S rRNA gene?

- Gene on the bacterial chromosome that encodes a rRNA molecule
- 16S rRNA is essential for the SSU of the ribosome needed for translation



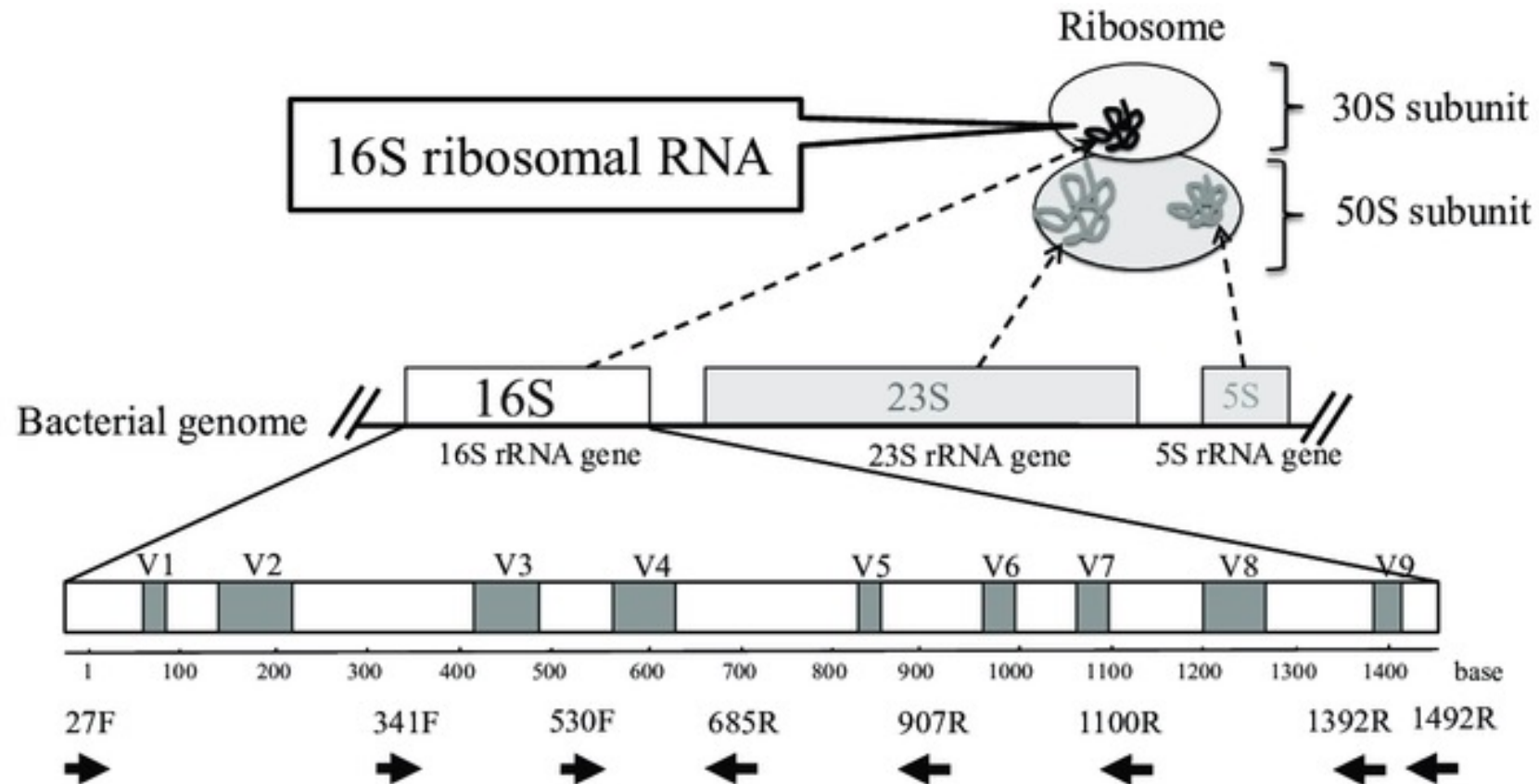
Why did we choose this gene?

1. This gene is conserved in almost all species of bacteria that we have cultured today
2. It is highly essential so it accrues changes slowly
3. But changes are still present enough to distinguish between different taxa



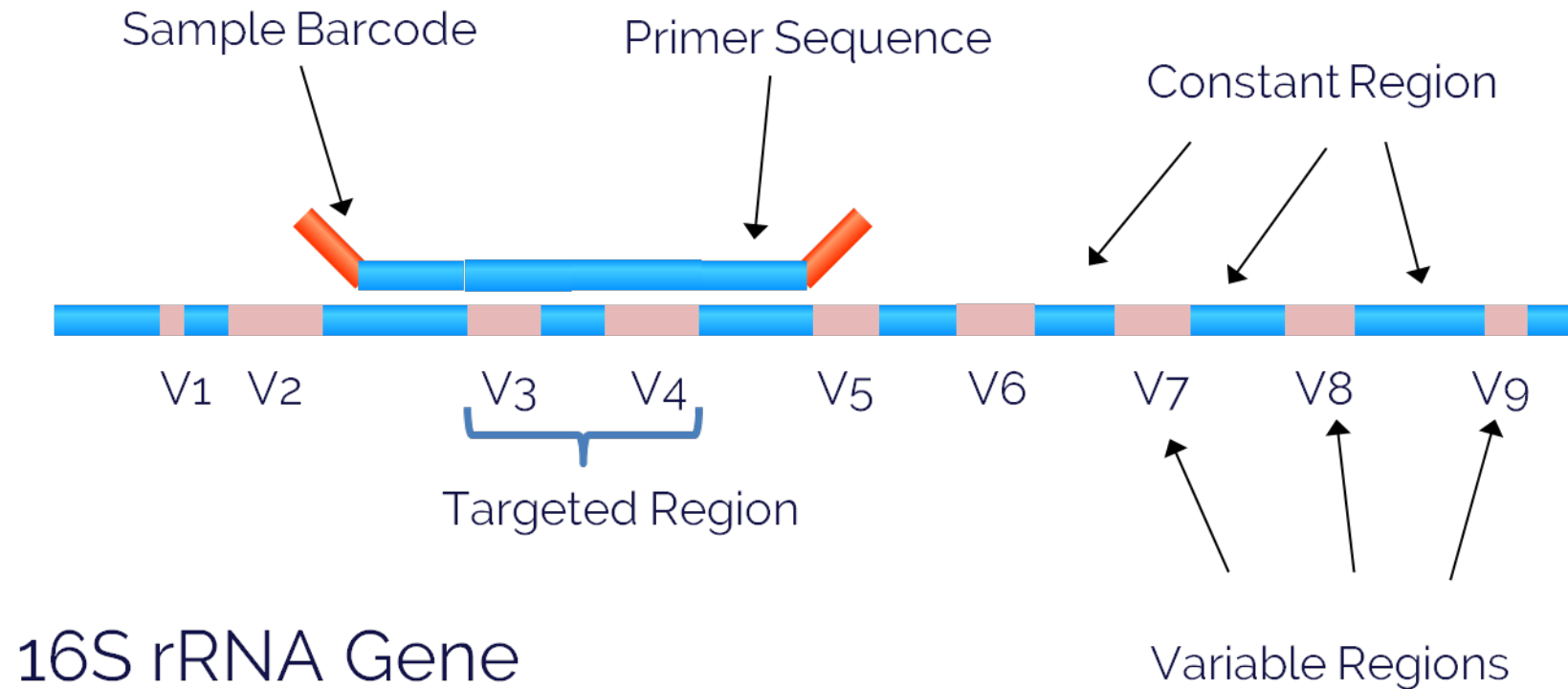
Carl Woese

16S rRNA

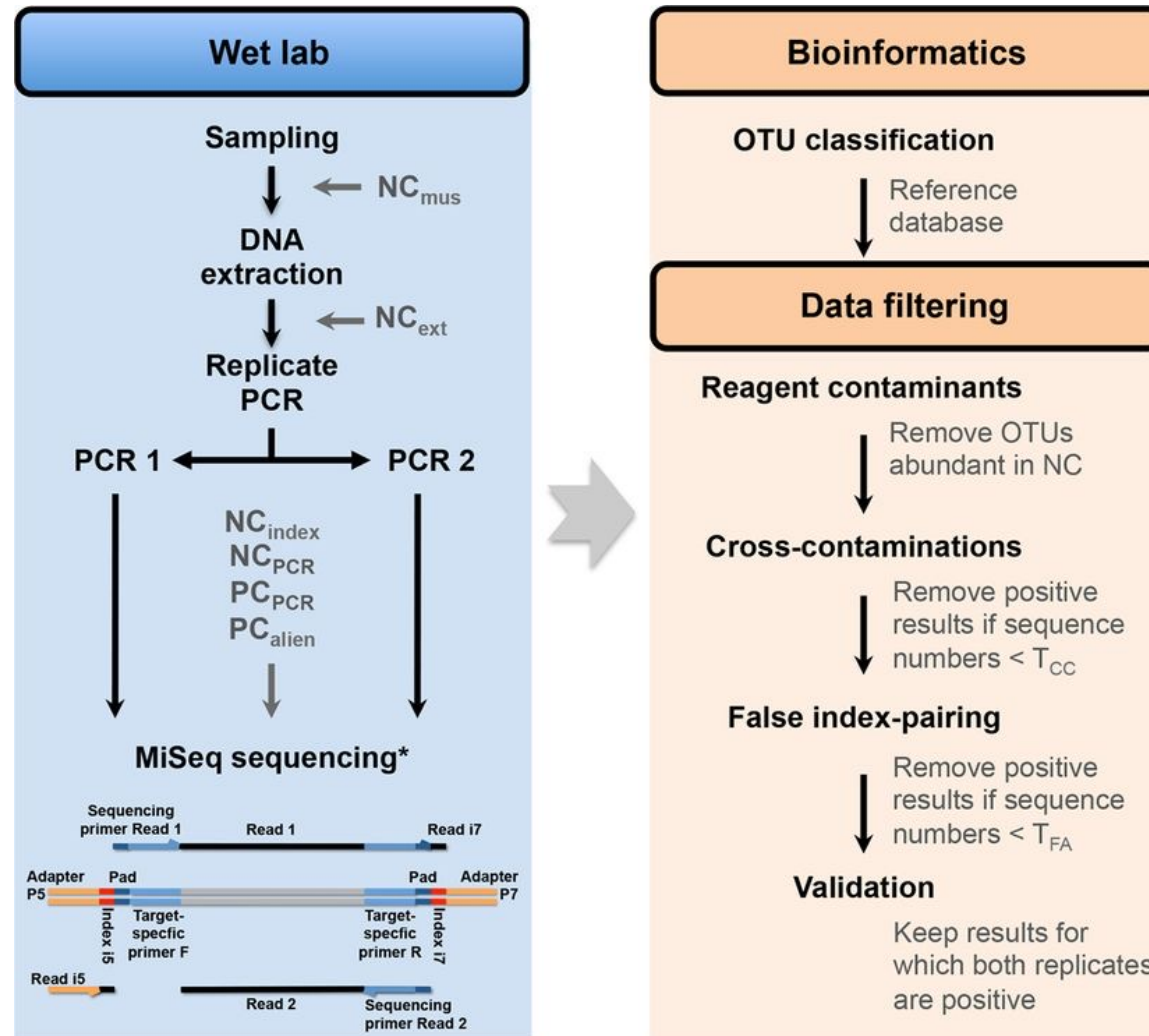


Advantage to PCR

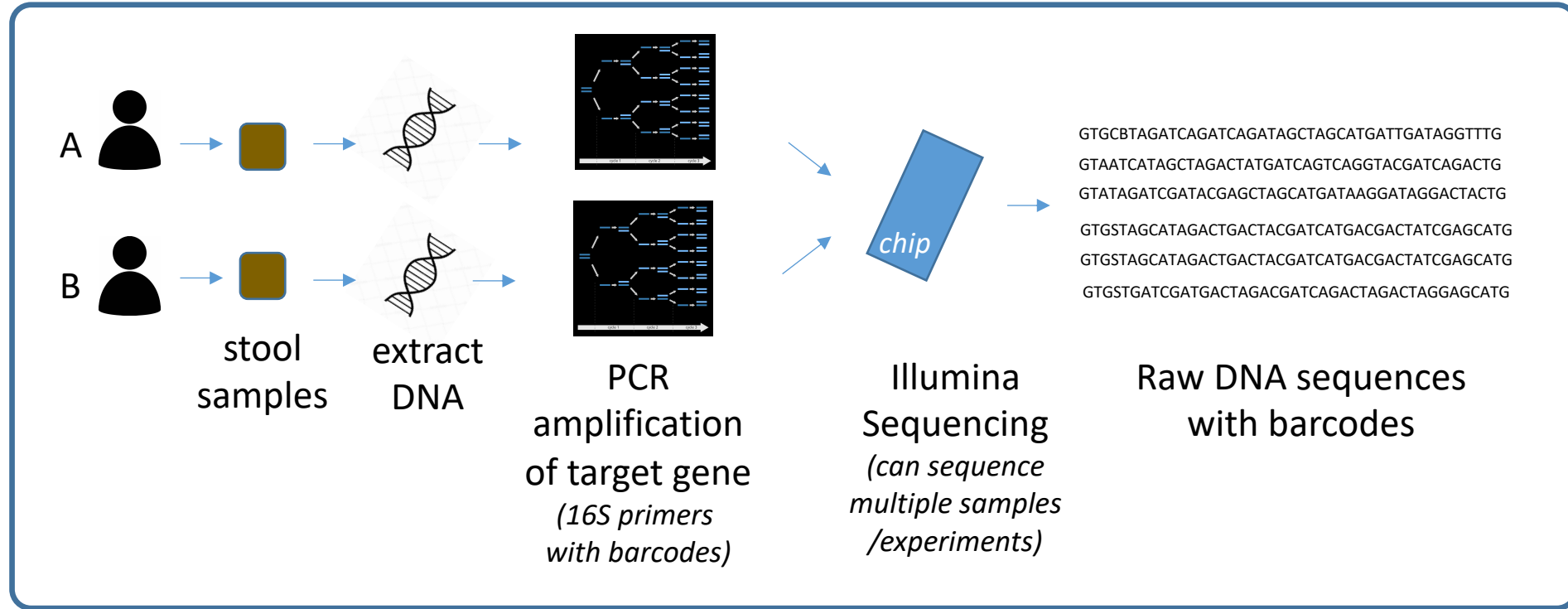
- Design primers that bind to the constant regions to help amplify the variable regions



16S Sequence Analysis Pipeline

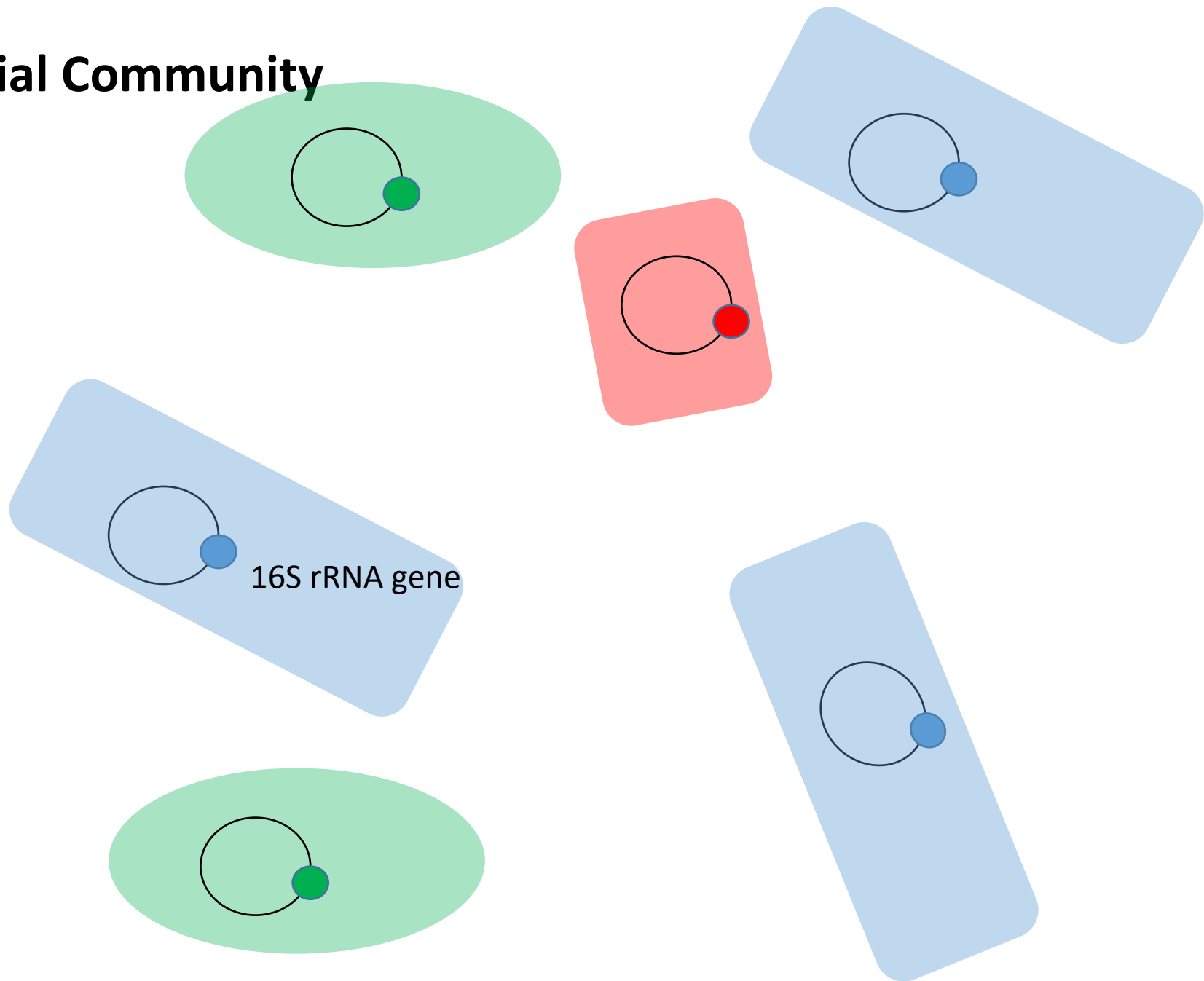


Generating the data

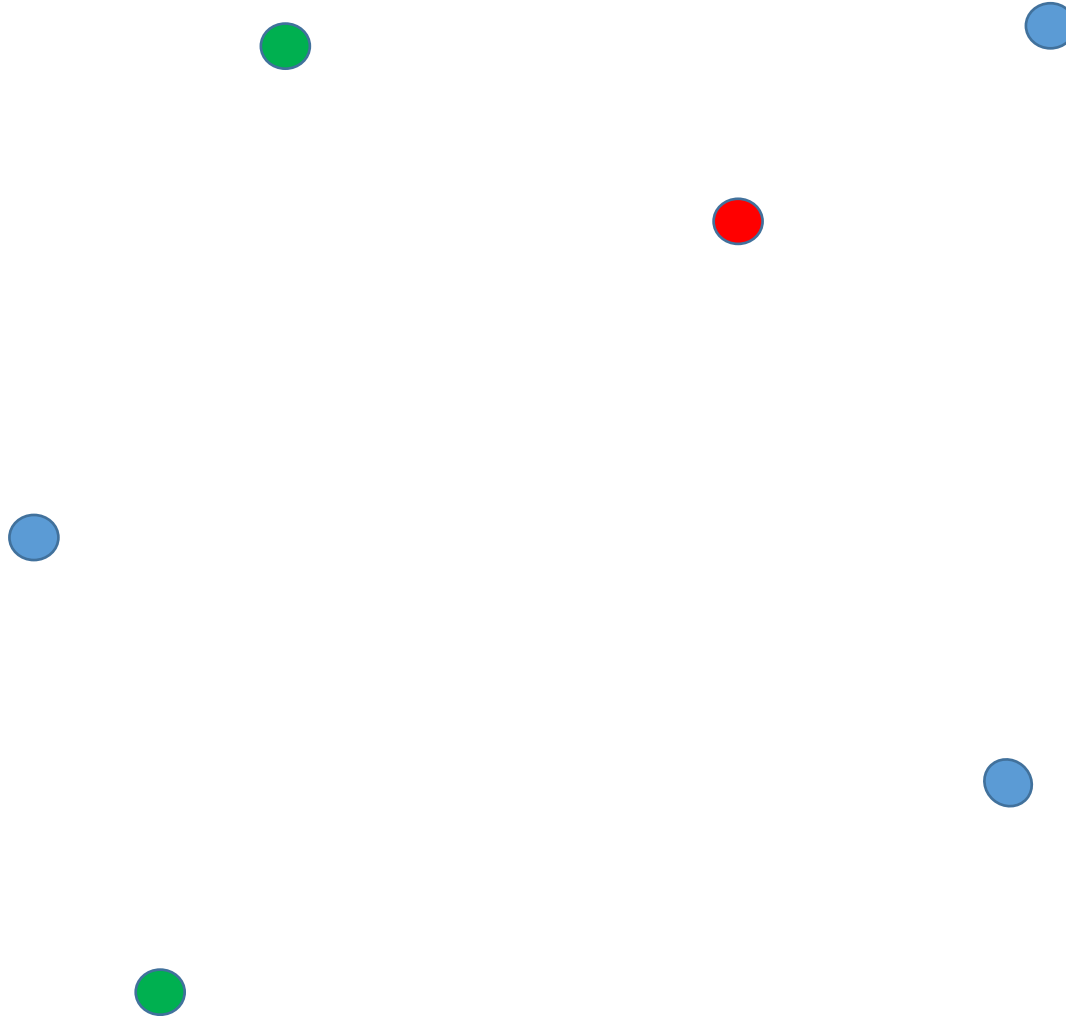


Your Microbial Community

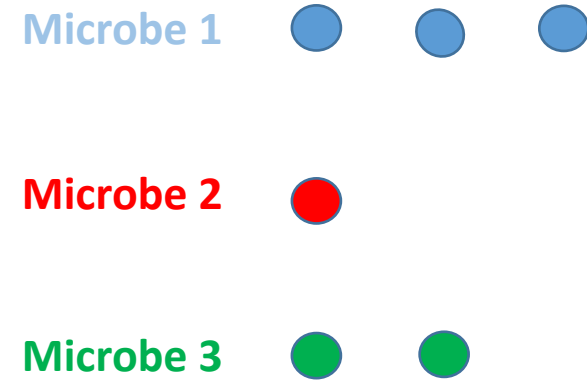
Sample 1



PCR Amplify 16S gene



High Throughput Sequencing



Microbial Communities Across Different Samples

	Microbe 1	Microbe 2	Microbe 3
Sample 1	3	1	2
Sample 2	1	0	5
Sample 3	0	1	2

16S rRNA gene

Constant regions



Microbe 1



Microbe 2



Microbe 3



16S rRNA gene

Constant regions



Microbe 1



Microbe 2



Microbe 3



16S rRNA gene

Constant regions

Variable regions



V1

V2

V3

V4

V5

V6

V7

V8

V9

Amplified
region

Microbe 1



V1

V2

V3

V4

V5

V6

V7

V8

V9

Microbe 2



V1

V2

V3

V4

V5

V6

V7

V8

V9

Microbe 3



V1

V2

V3

V4

V5

V6

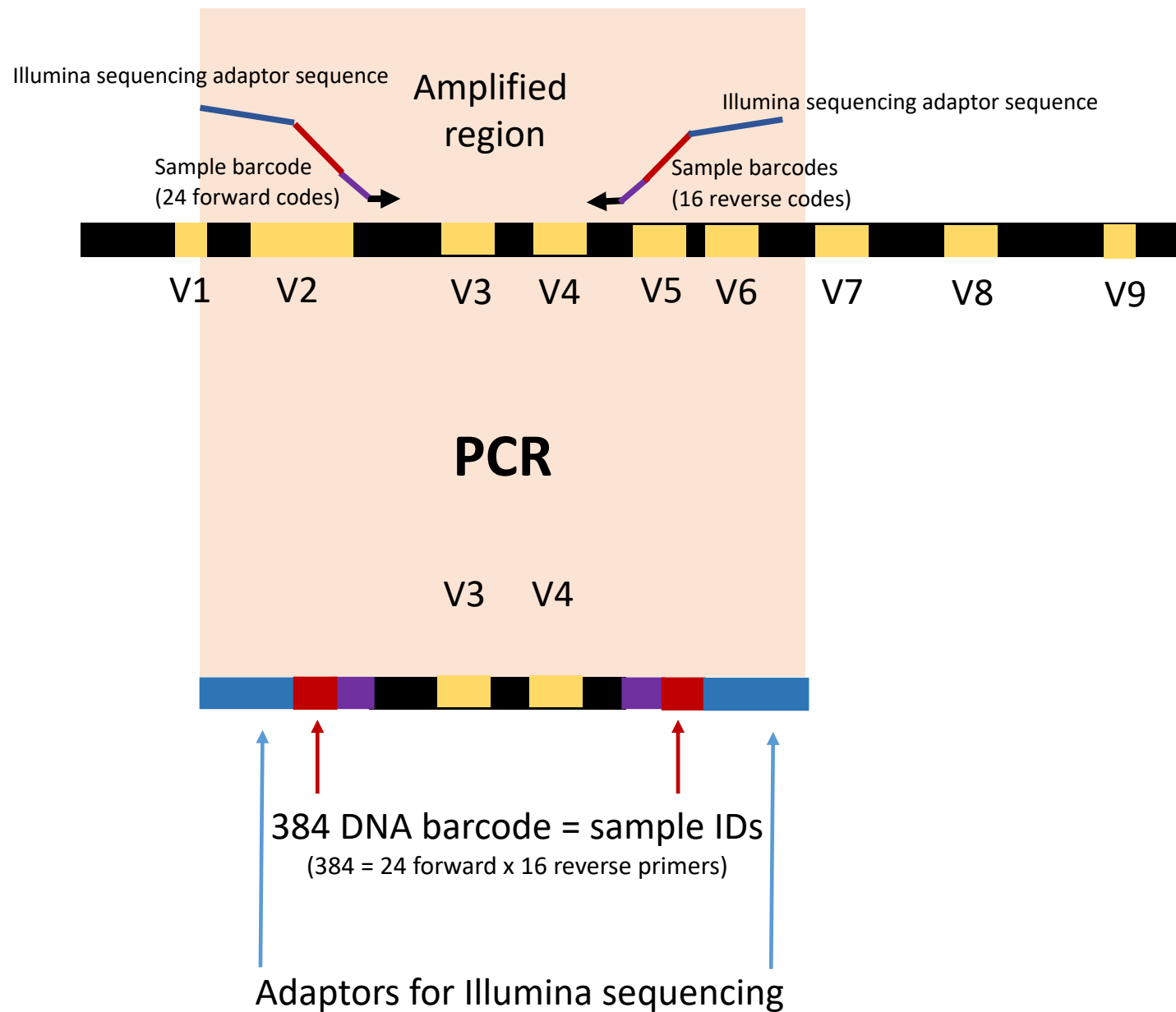
V7

V8

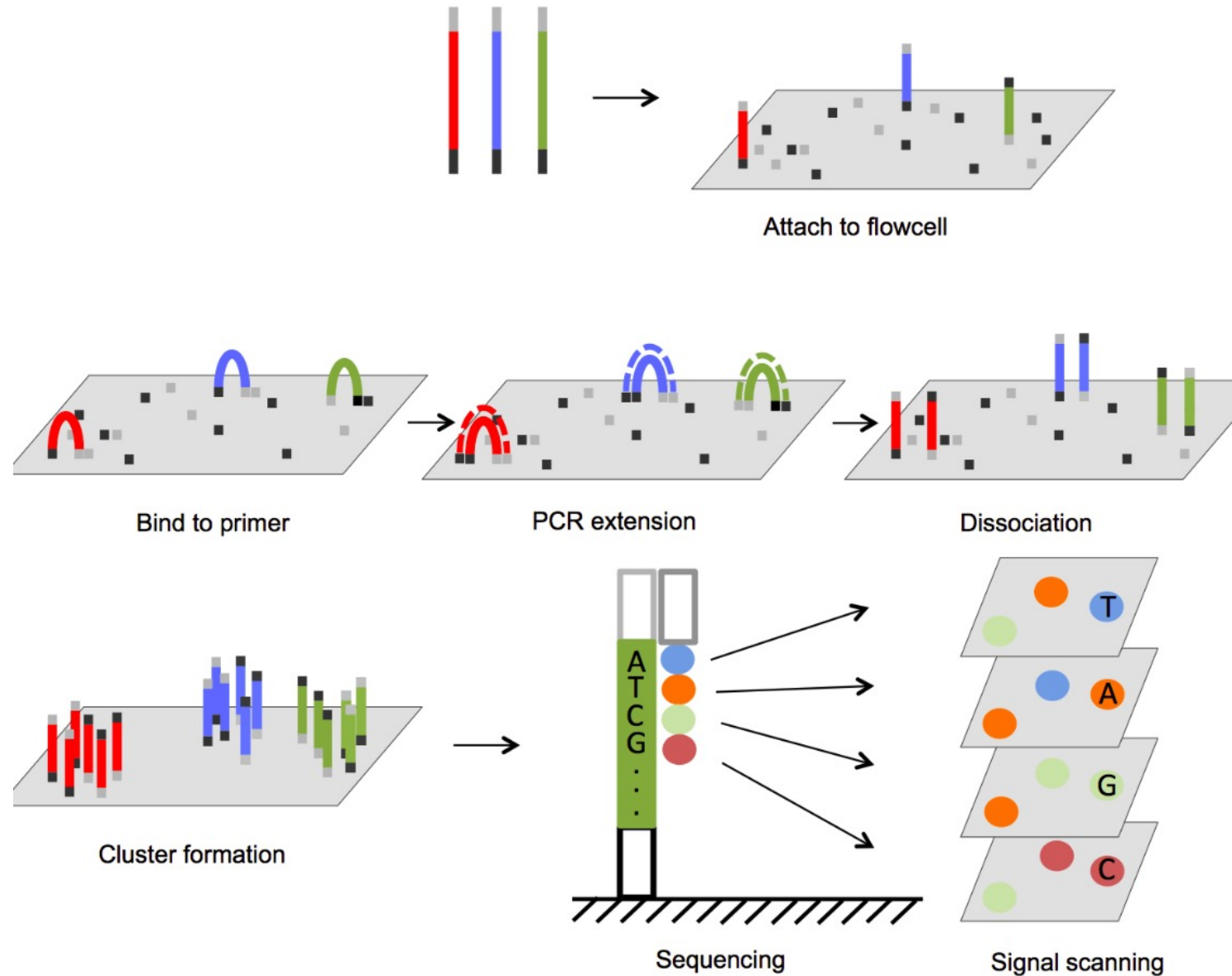
V9

Constant regions

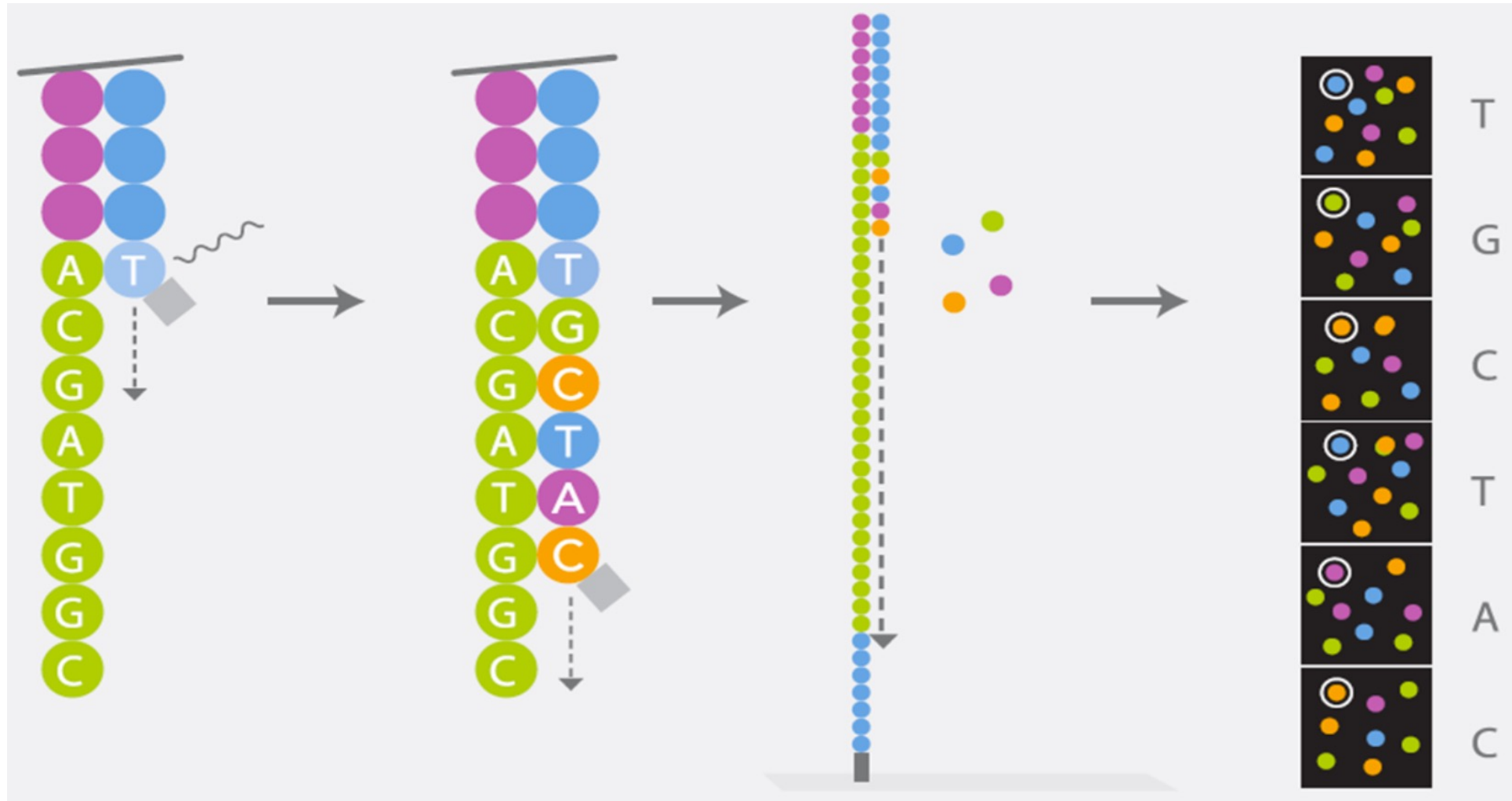
Variable regions



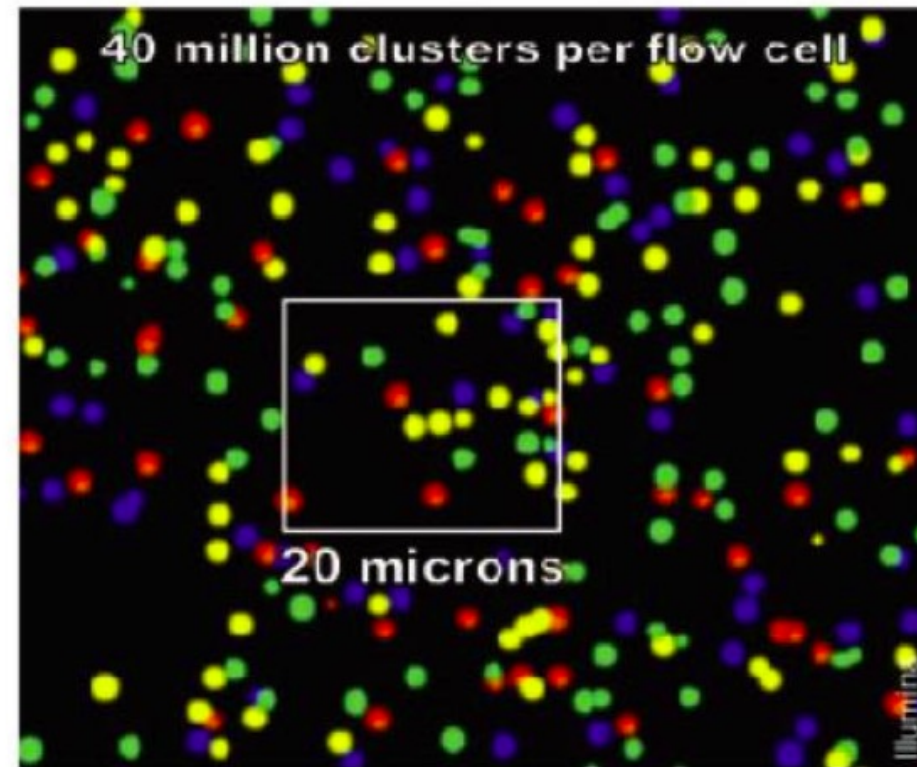
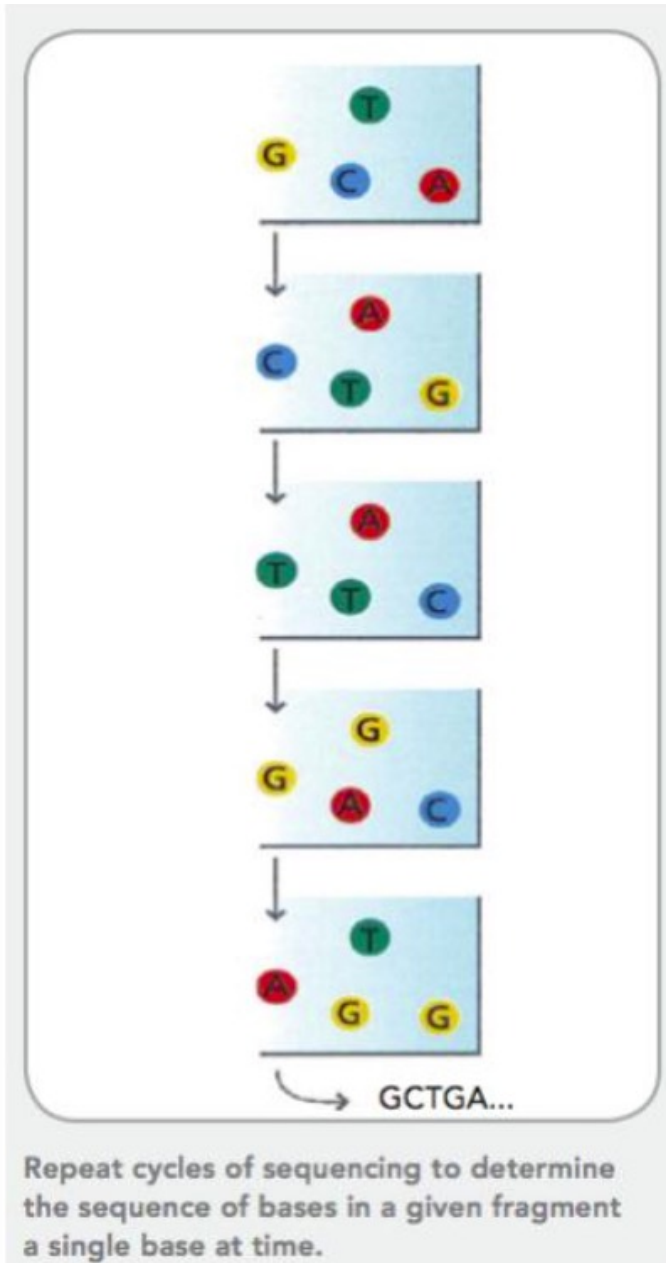
High Throughput Sequencing – Illumina MiSeq (overview)



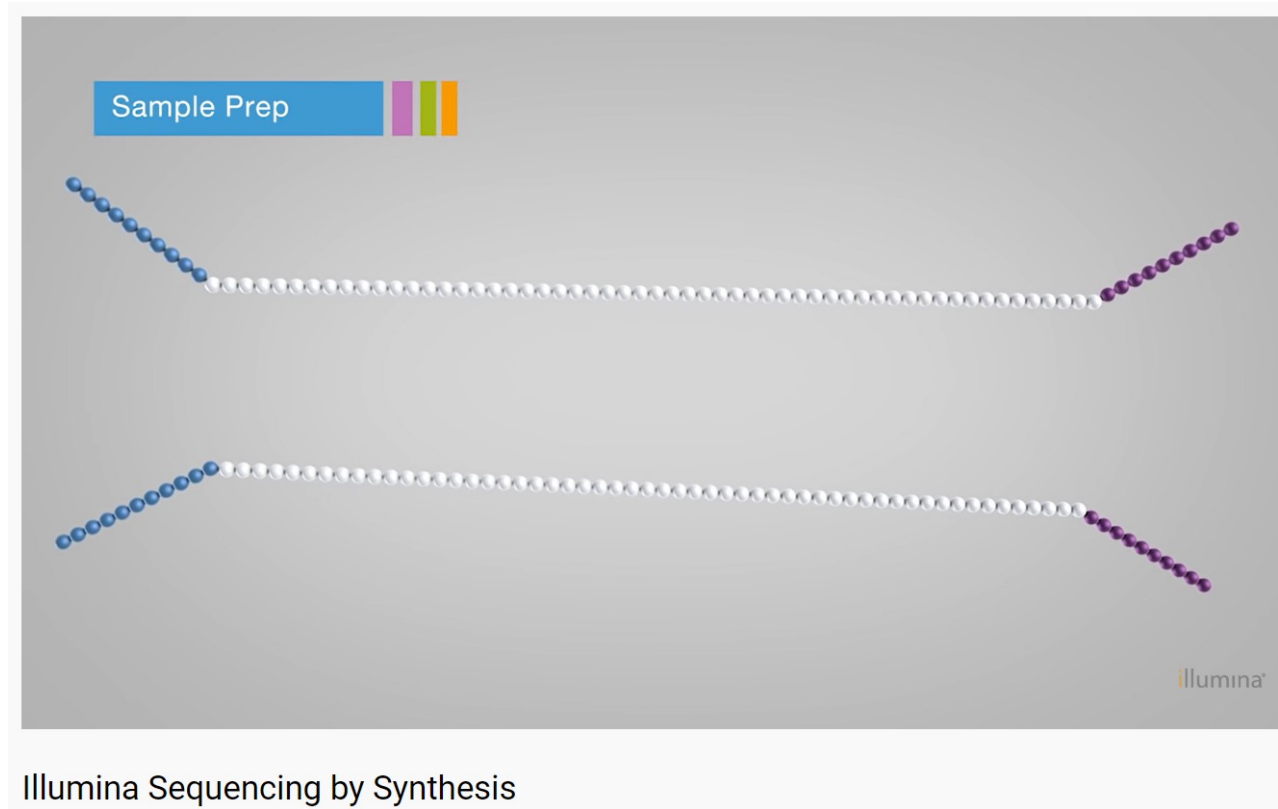
Illumina sequencing by synthesis



<https://www.cegat.de/web/wp-content/uploads/2015/10/sequencing-1.png>



Watch this video!



- <https://youtu.be/fCd6B5HRaZ8>

Sequence File Formats

Fasta file:

Record name;
starts with a ">"

DNA sequence

```
>sequence_1
TTTCCGGGGCACATAATCTTCAGCCGGGCGC
>sequence_1
TCAGCCGGGCCTTCAGCCGGGGCACATAATA
```

Qual file:

Record name;
starts with a ">"

Phred score

```
>sequence_1
57 67 59 61 59 61 60 57 64 52 56 54 56 62 57 58 54 55
65 65 60 57 62 54 53 60 61 62 53 57 49

>sequence_1
40 53 41 38 38 31 41 51 60 47 40 39 38 3 ..... 
```

fastq file:

Record name;
starts with a "@"

DNA sequence

Phred score

```
@sequence_1
TTTCCGGGGCACATAATCTTCAGCCGGGCGC
+
9C;=;<9@4868>9:67AA<9>65<=>591
@sequence_1
TCAGCCGGGCCTTCAGCCGGGGCACATAATA
+
(' %3 ( &&&% . . . . .
```

Phred Quality Score	Probability of incorrect base call	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%