

High Throughput Sequencing Sequence Data Generation and Analysis

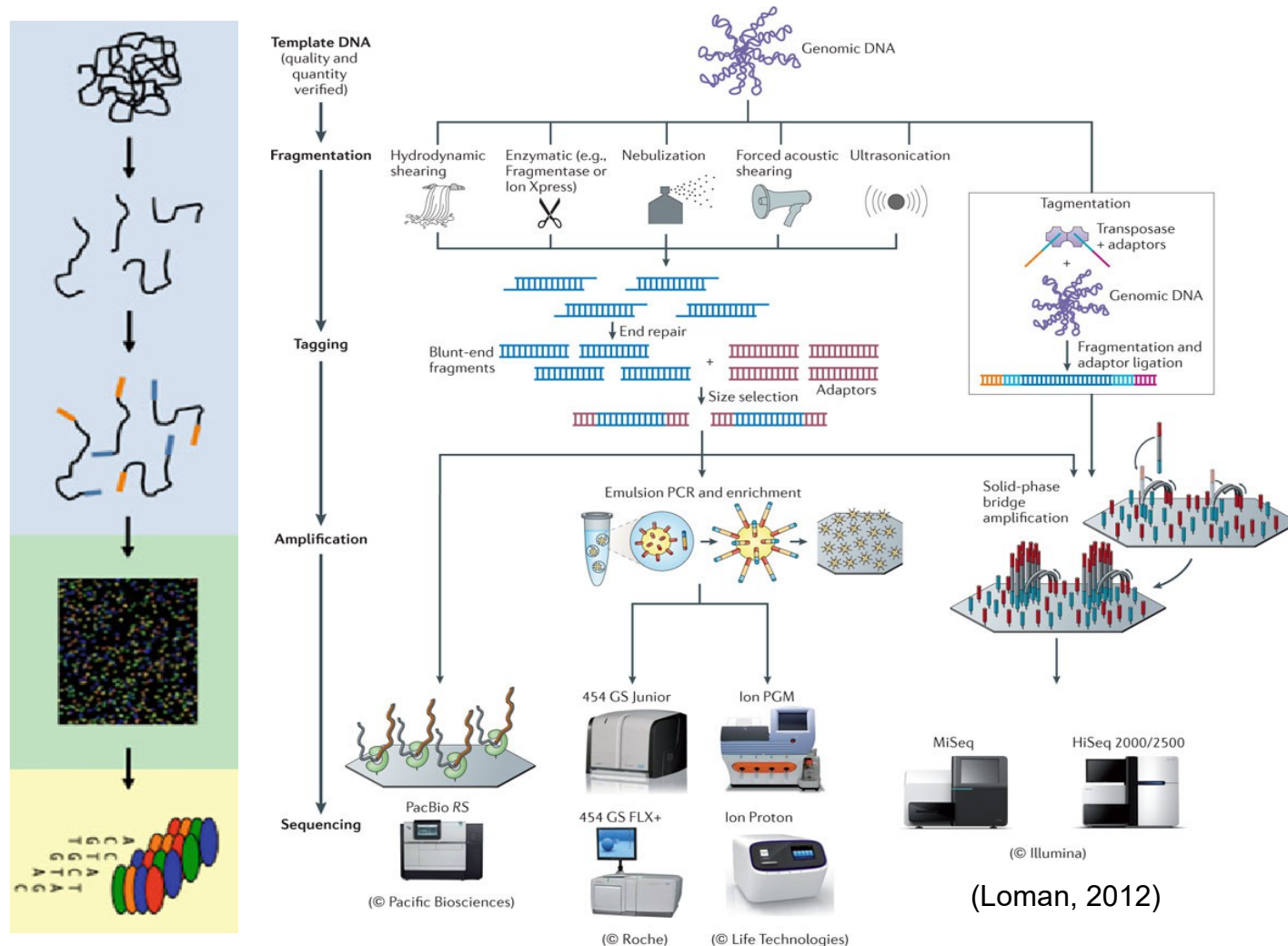
Daniel Sobral



FUNDAÇÃO CALOUSTE GULBENKIAN
Instituto Gulbenkian de Ciência

Bioinformatics and Computational Biology Unit

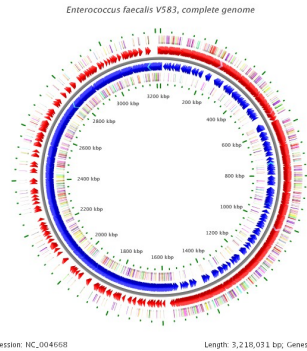
High-Throughput Sequencing Workflow



Nature Reviews | Microbiology

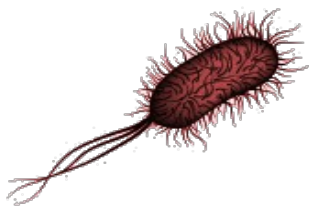
HTS Applications: Genomics Resequencing

A reference sequence already exists and is annotated



Often obtained using “traditional” Sanger (hard work!)

What are the mutations (if any) from reference causing the phenotype?

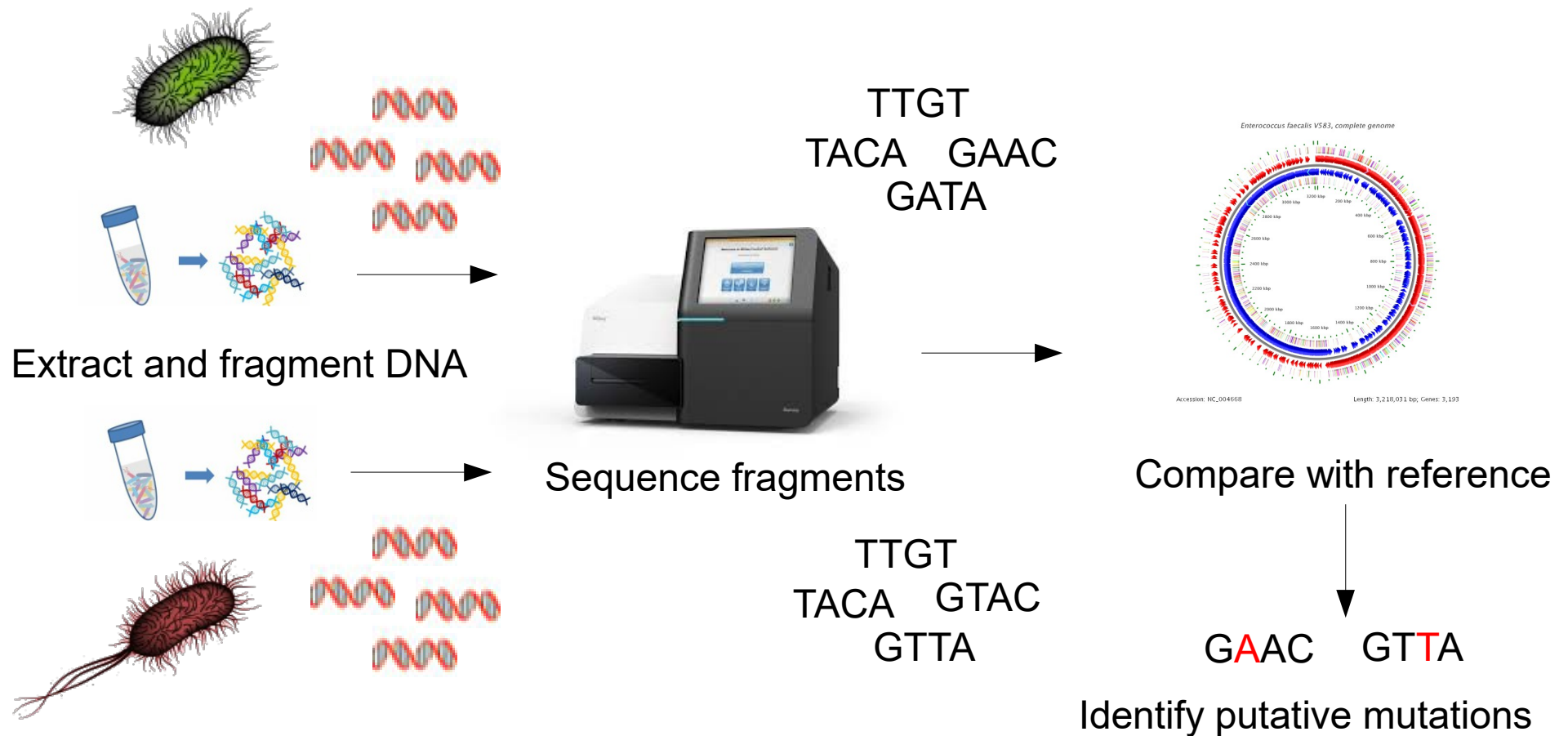


observed “mutants”
(with different phenotype)

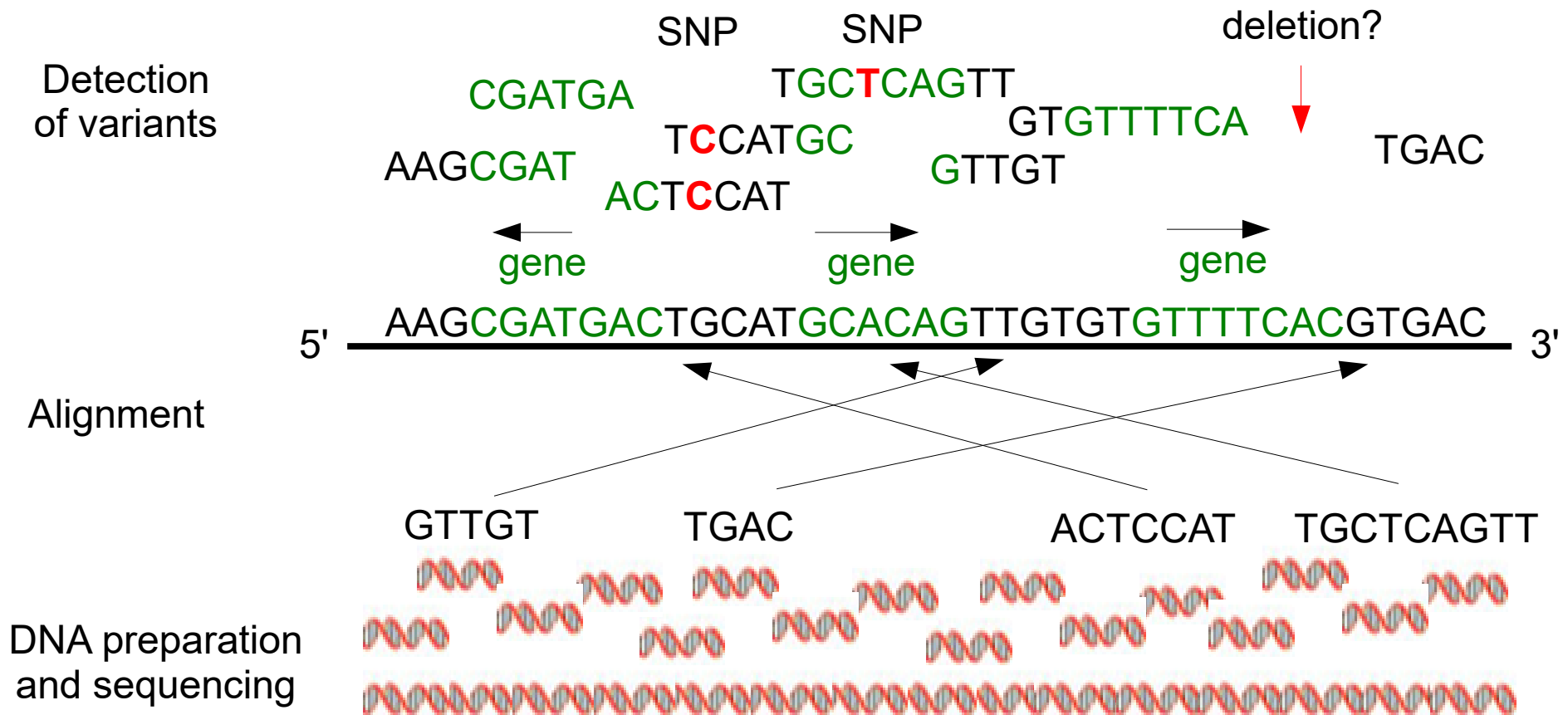


observed “mutants”
(with different phenotype)

HTS Applications: Genomics Resequencing Pipeline

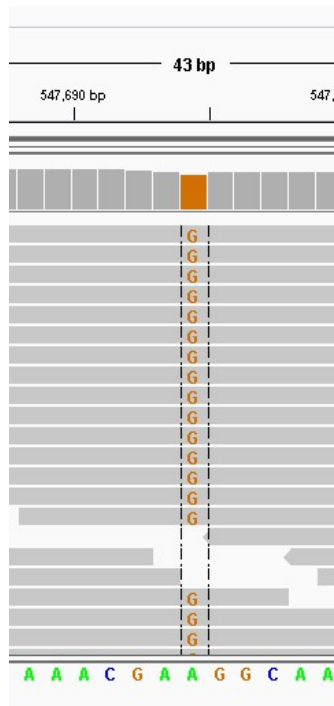


HTS Applications: Genomics Resequencing Pipeline

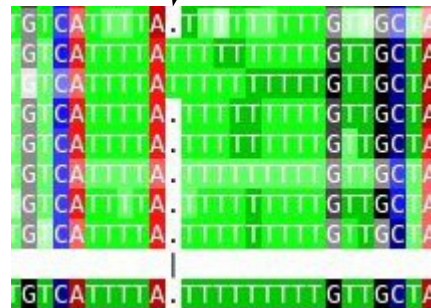


HTS Applications: Genomics Resequencing Pipeline

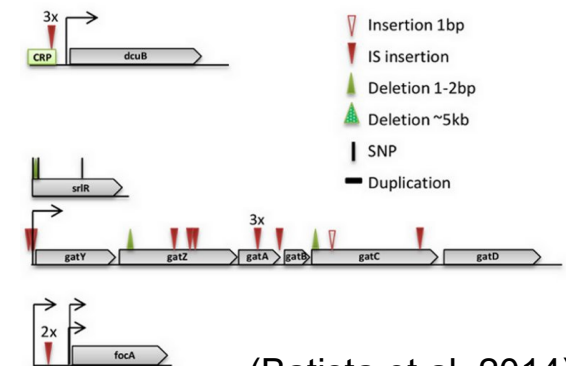
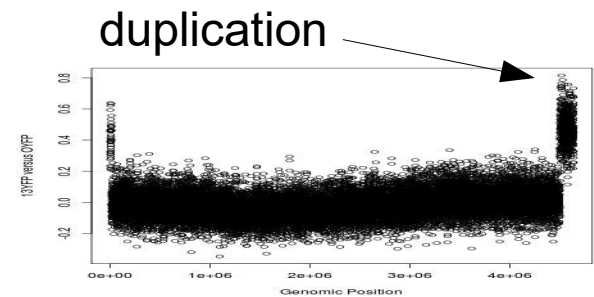
SNP finding is fairly “simple”, but other cases are not always so easy



True or artifact?

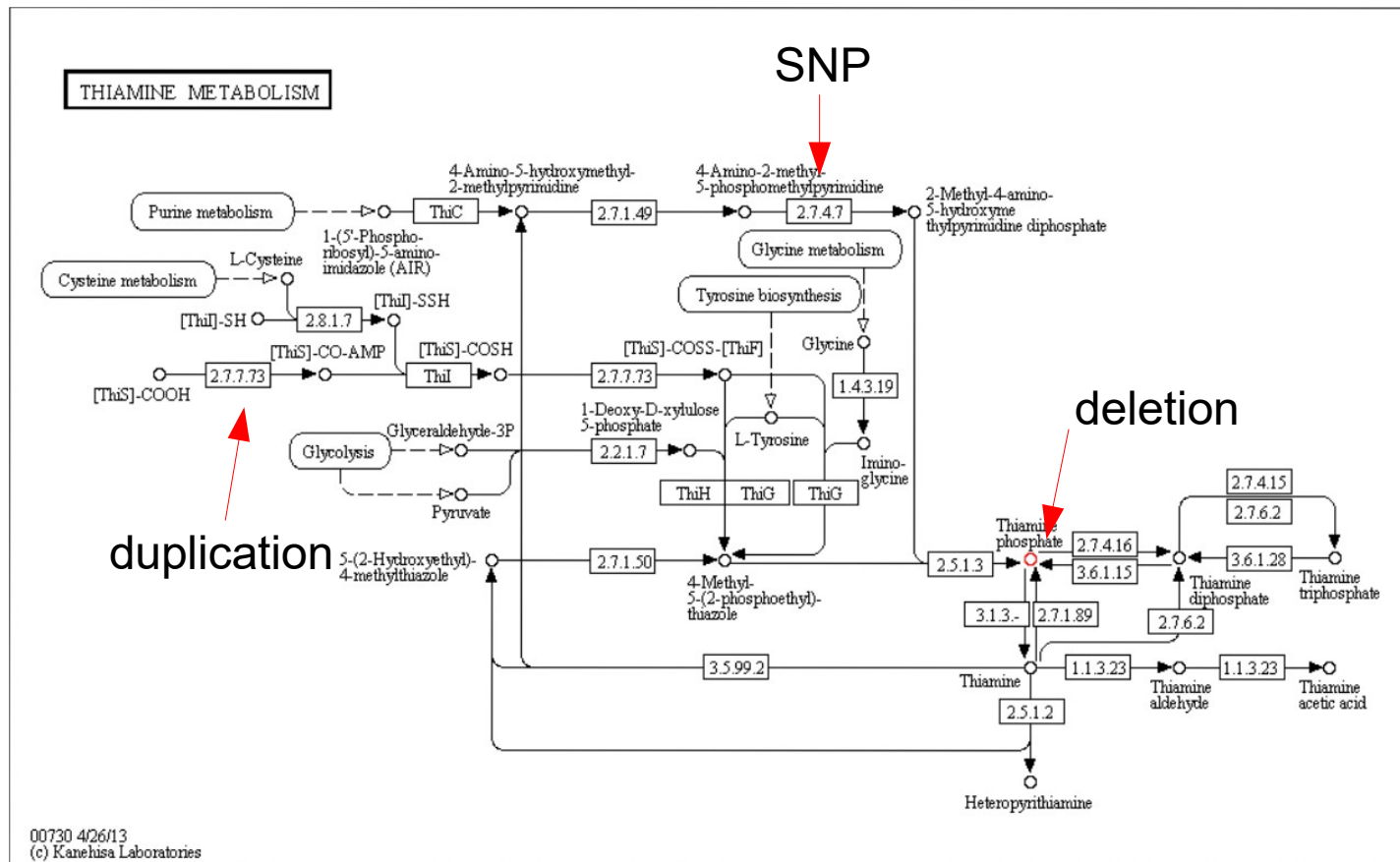


“complex” events most often ignored but in many cases important



(Batista et al, 2014)

HTS Applications: Genomics Resequencing post-analysis



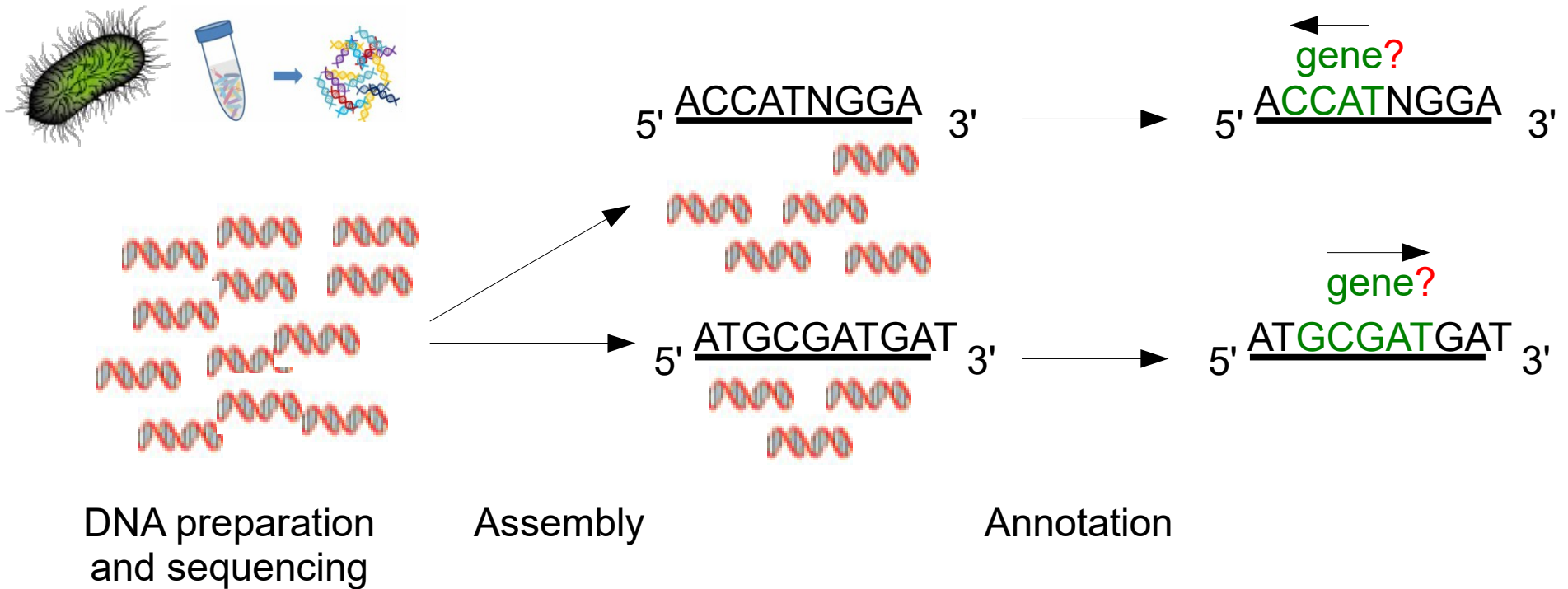
Variants
only make
sense in
context

How can we
be sure of
the effect of
the variant

qualitative
VS
quantitative
effect

HTS Applications: Genomics

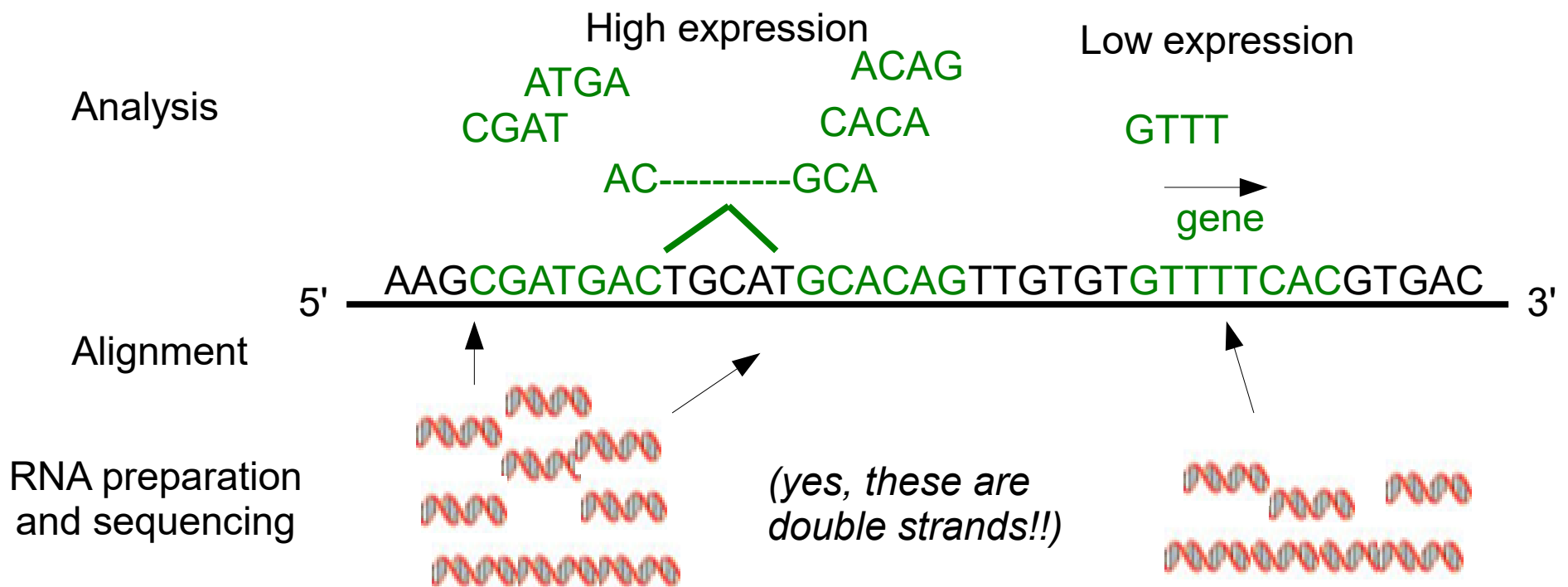
denovo Assembly of genomes



HTS Applications: Transcriptomics

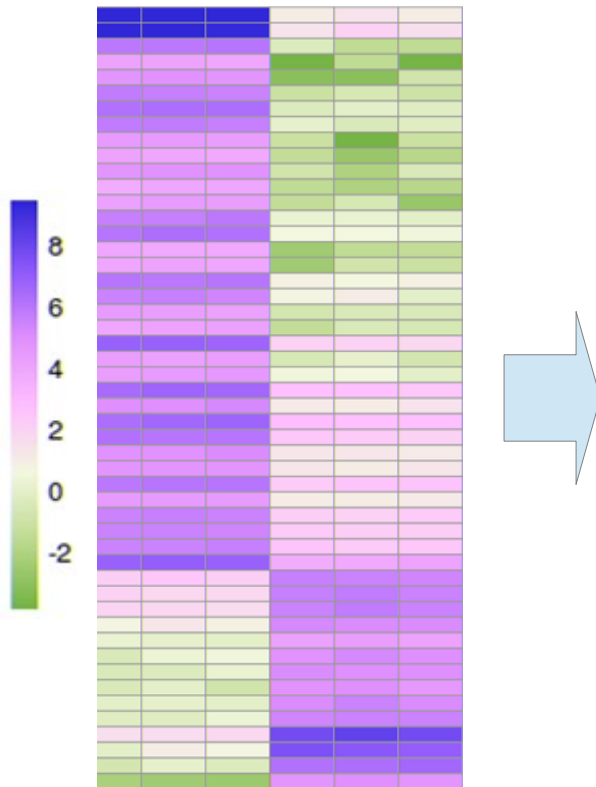
Sequencing your transcriptome

- Spliced alignment (eukaryotes)

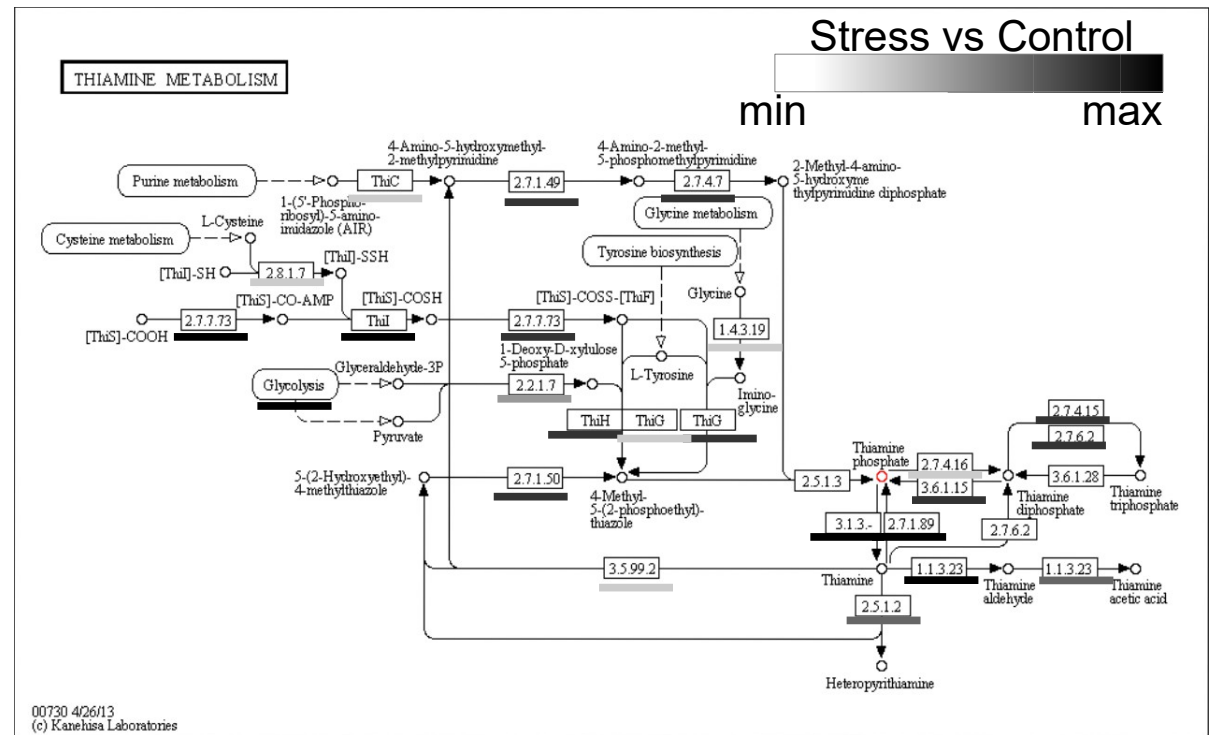


HTS Applications: Transcriptomics

Differential Gene Expression



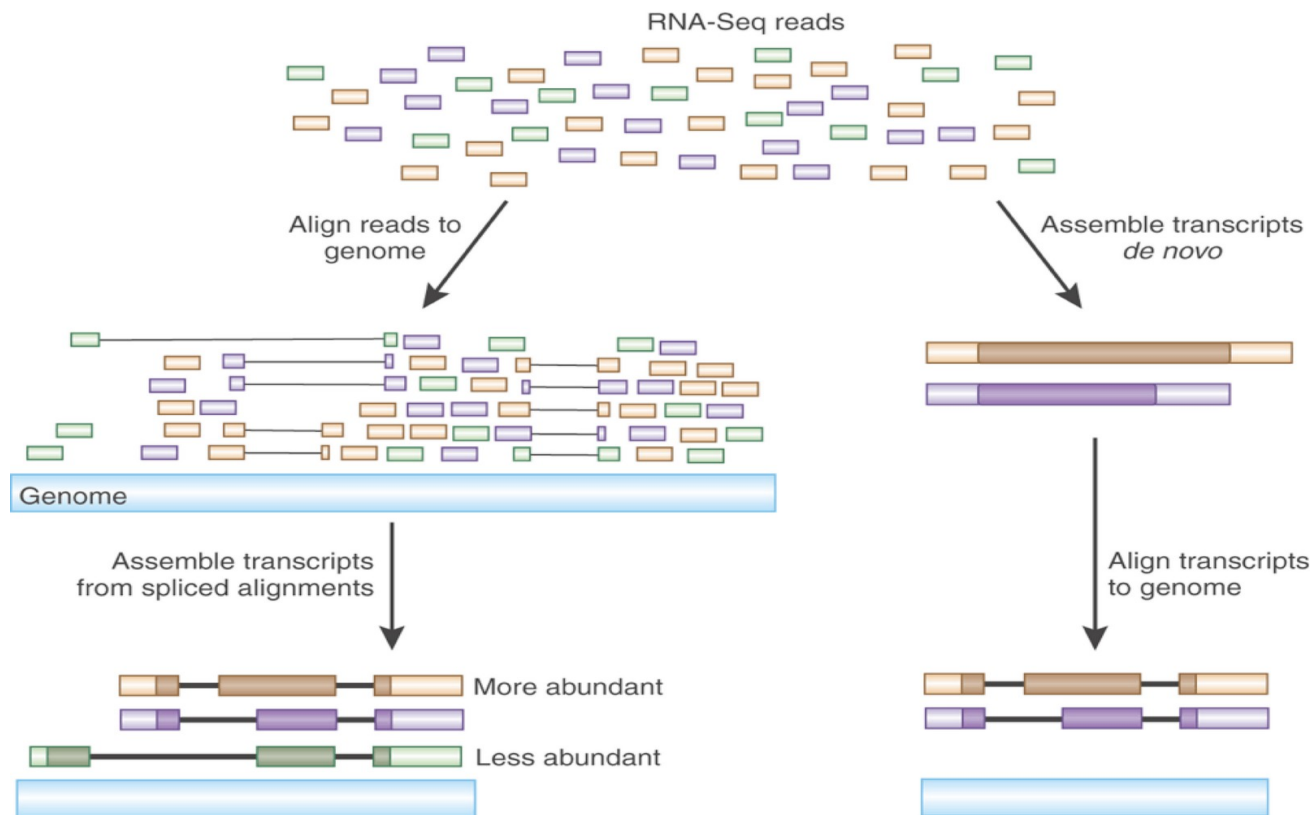
Differentially Expressed Genes



Fold Change of Gene Expression within a Pathway

HTS Applications: Transcriptomics

Discovering your transcriptome



Assembling transcripts *de novo* is even harder than the genome

Finding alternative Transcripts with RNA-Seq is a very active research area

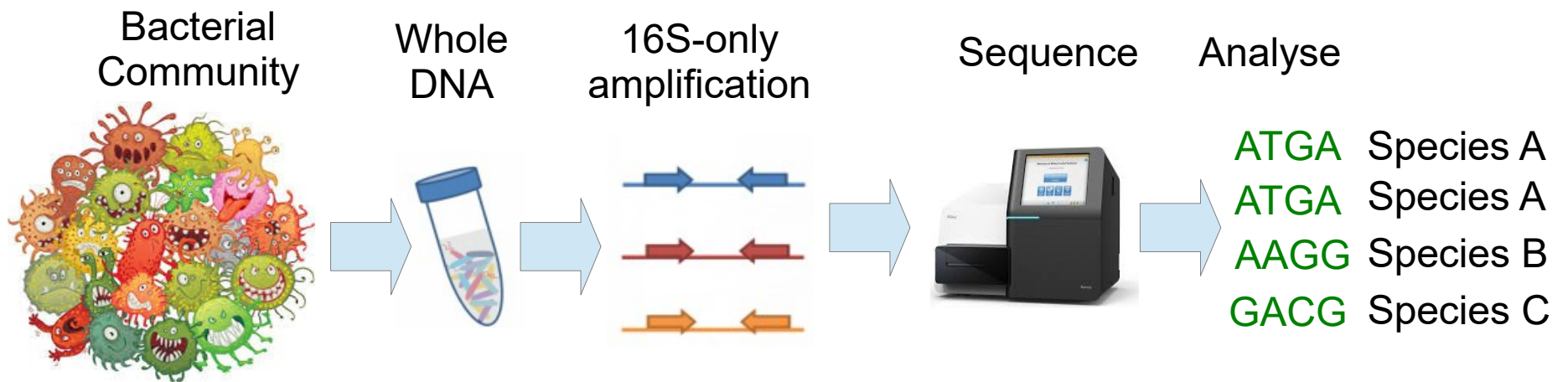
MUCH easier if genome is already available

You still need the painstaking work done before NGS

HTS Applications: Metagenomics

16S Metagenomics

- Estimate bacterial composition
 - Using only an universal marker gene (16S)



HTS Applications: Metagenomics

16S Metagenomics

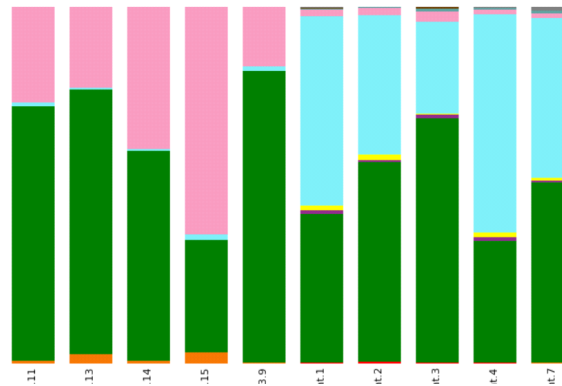
- Comparative bacterial composition



Normal



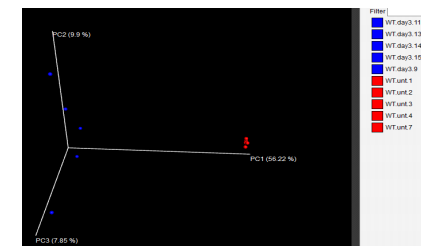
Antibiotic



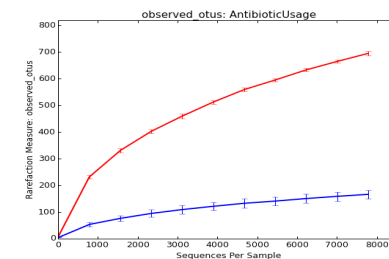
[View Table \(txt\)](#)

Legend	Taxonomy	%	WT.day3.11	WT.day3.13	WT.day3.14	WT.day3.15	WT.day3.9	WT.unt.1	WT.unt.2	WT.unt.3	WT.unt.4	WT.unt.7
Unassigned:Other		0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.4%	0.3%	0.3%	0.1%
k_Bacteria:Other		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
k_Bacteria:Actinobacteria		0.8%	0.7%	2.5%	0.8%	3.2%	0.4%	0.0%	0.1%	0.1%	0.0%	0.1%
k_Bacteria:Bacteroidetes		56.9%	71.5%	74.4%	58.8%	31.3%	81.7%	41.7%	56.1%	68.5%	34.0%	50.7%
k_Bacteria:Cyanobacteria		0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	1.0%	0.5%	1.0%	1.0%	0.3%
k_Bacteria:Deferribacteres		0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	1.4%	1.4%	0.3%	1.5%	0.8%
k_Bacteria:Firmicutes		22.9%	1.9%	0.4%	0.0%	1.6%	1.3%	53.0%	39.2%	25.7%	61.0%	44.7%
k_Bacteria:Proteobacteria		17.0%	26.7%	22.7%	39.8%	63.8%	16.7%	1.7%	2.0%	3.0%	1.5%	1.4%
k_Bacteria:TM7		0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	0.2%	0.6%	0.4%	0.8%
k_Bacteria:Tenericutes		0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.5%	0.0%	0.0%
k_Bacteria:Verrucomicrobia		0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.1%	0.2%	0.9%

Frequency of Taxa



Beta-Diversity
(comparative)

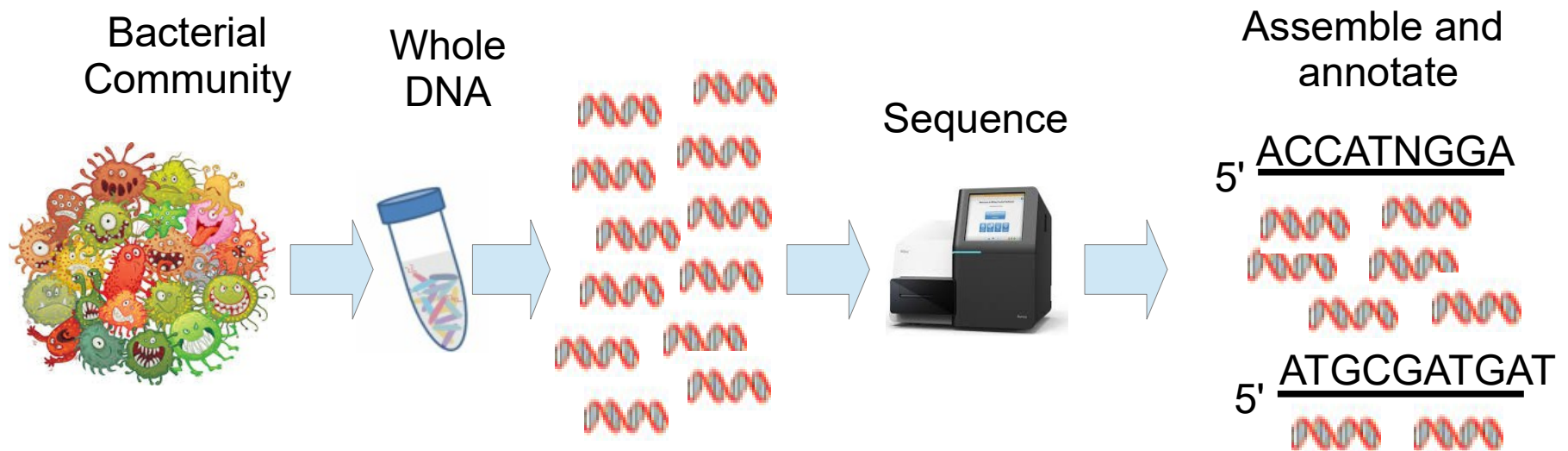


Alpha-Diversity
(total diversity)

HTS Applications: Metagenomics

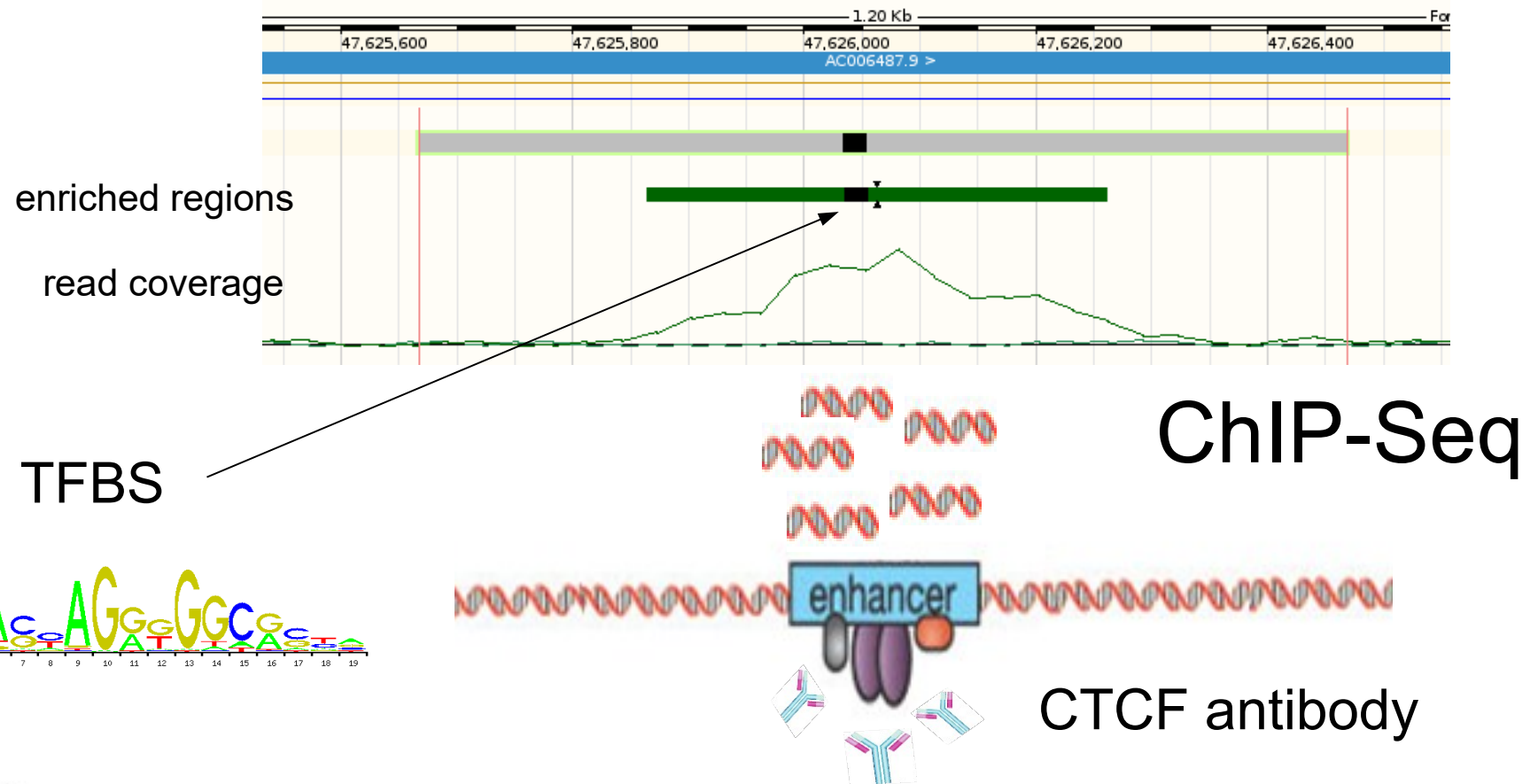
Whole Shotgun Metagenomics

- Similar to denovo genome assembly
 - Much more complex than a single species
 - Multiple species, at different frequencies



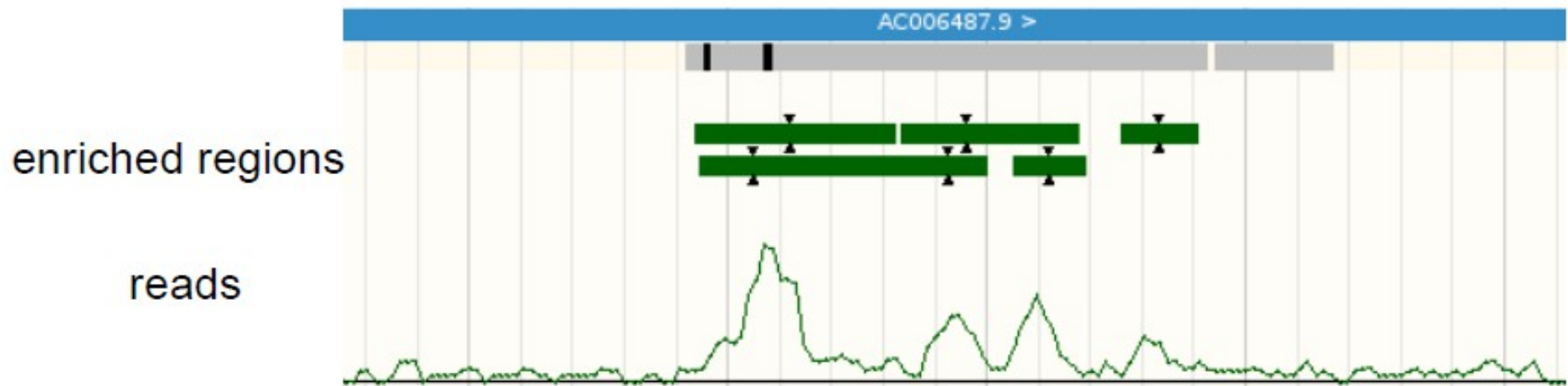
HTS Applications: Epigenomics

Targeted Sequencing



HTS Applications: Epigenomics

Targeted Sequencing



Dnase-Seq

