High Throughput Sequencing Revisited

Josh Granek

Amplicon Sequencing

PCR amplify and sequence a marker gene

	Marker Gene
Bacteria	16s rRNA
Fungi	18s or ITS rRNA
Archaea	16s rRNA
Protozoa	18s rRNA
Viruses	?????

Metagenomics

	What	Information	Analogy	Target Size	Cost	Discovery?
Amplicon	Marker Gene	Who is Present	Name	100bp - 1kb	Low	+/-
Shotgun Metagenome	Genomes	What Genes are Present	CV	100kb - 100Mb	High	++
Shotgun Metatranscriptome	All RNA	What Genes are Expressed	Twitter Feed	100kb - 100Mb	High	++

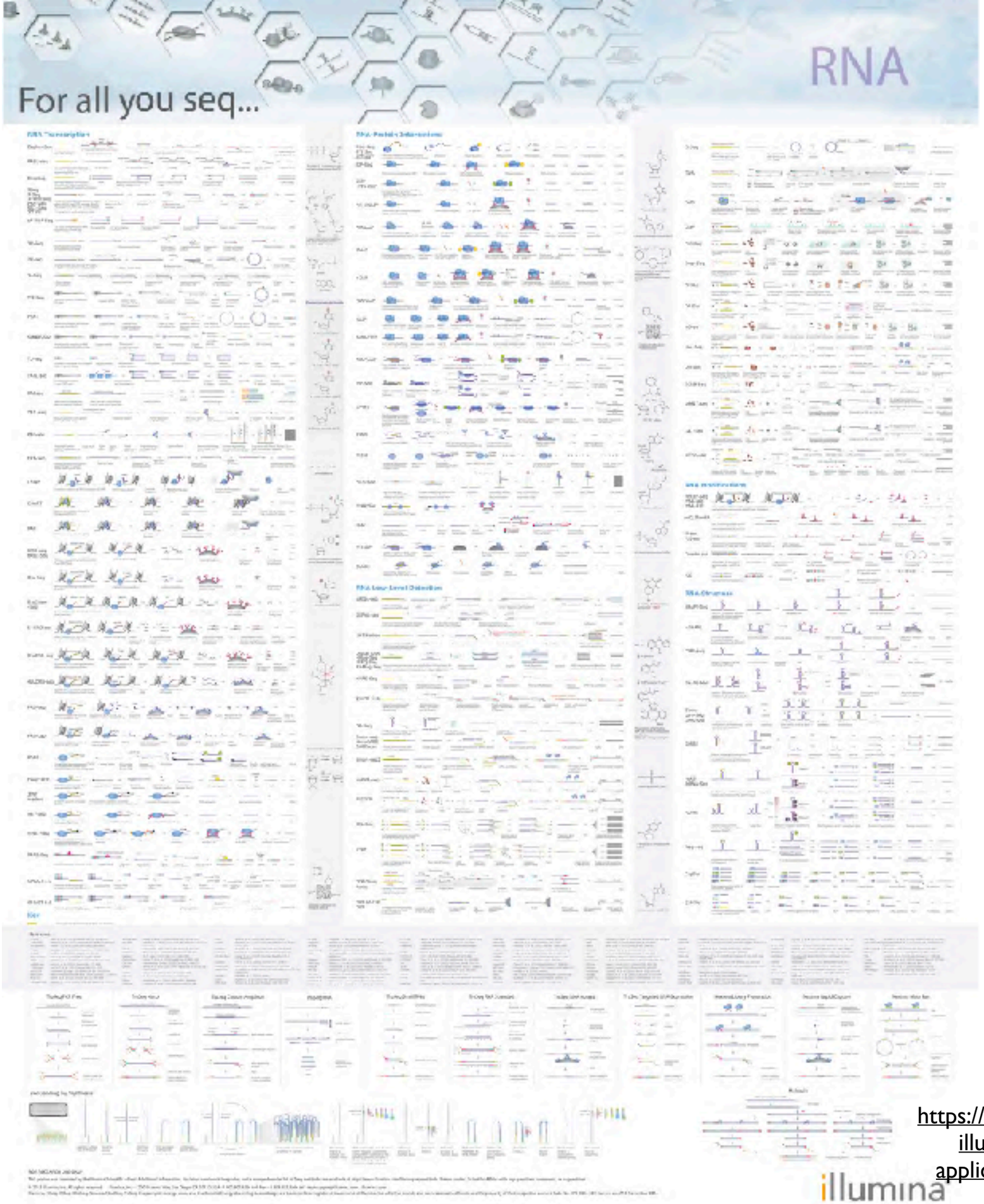
HTS Applications

- DNA-Seq
- RNA-Seq
- Amplicon Sequencing
- Many More
    - ChIP-Seq
    - Ribo-Seq
    - Hi-C
    - MethylC-Seq

https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/for-all-you-seq-dna.pdf

DNA-Seq

- De *Novo* Genome Sequencing
- Genotyping
  - GWAS
  - Genetic risk factors
- Mutation identification



https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/for-all-you-seq-rna.pdf

RNA-Seq

- Transcriptome: "Which genes are expressed in this sample?"
  - Differential Expression
  - Genome Annotation
- SNPs
- Gene Fusions

RNA-Seq

- Bulk RNA-Seq
- Single-Cell RNA-Seq (scRNA-Seq)

Amplicon Sequencing

- CRISPER Barcode Seq
- 16s rRNA

*-Seq Comparison

Method	Molecule	Target	Target Size (in humans)
DNA-Seq	DNA	Whole Genome	2×10^9 bp
RNA-Seq	RNA	Transcriptome	$<3 \times 10^7$ bp
Amplicon	DNA?	Target Region	10 - 10,000bp

HTS Applications

- DNA-Seq
- RNA-Seq
- Amplicon Sequencing
- Many More
    - ChIP-Seq
    - Ribo-Seq
    - Hi-C
    - MethylC-Seq

DNA Sequencing Technologies (Abridged)

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

DNA Sequencing Technologies (Abridged)

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

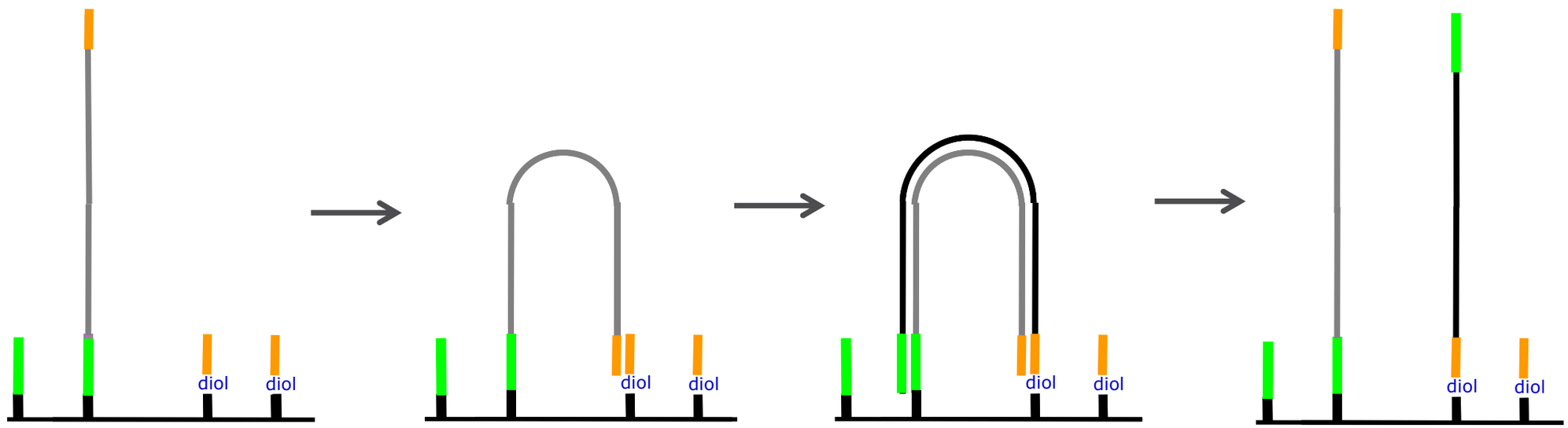
Comparing Technologies

Method	Read length	Accuracy	Reads per run	Max Output	Cost ($/Mb)	Pros	Cons
Sanger	400-900 bp	99.9%	1	900 bp	$2400	Longer reads.	Expensive. Low Output
Illumina	600 bp (300bp PE)	99.9%	$20x10^9$	6000 Gb	$0.01	High yield cost/base	Equipment expense. Short reads
PacBio	>10kb ave. >40kb max	99%	$5x10^5$	10 Gb	$0.08	Very long reads	Homopolymer errors. Moderate Output. Equipment expense.
Nanopore	>100 kb N50 >1Mb Max	92%	$1x10^6$	5 Gb	$0.10	Very long reads Portable Cheap Equipment	Homopolymer errors. Moderate Output.

https://en.wikipedia.org/wiki/DNA_sequencing
https://blog.genohub.com/2017/06/16/pacbio-vs-oxford-nanopore-sequencing/

Single Molecule Technologies

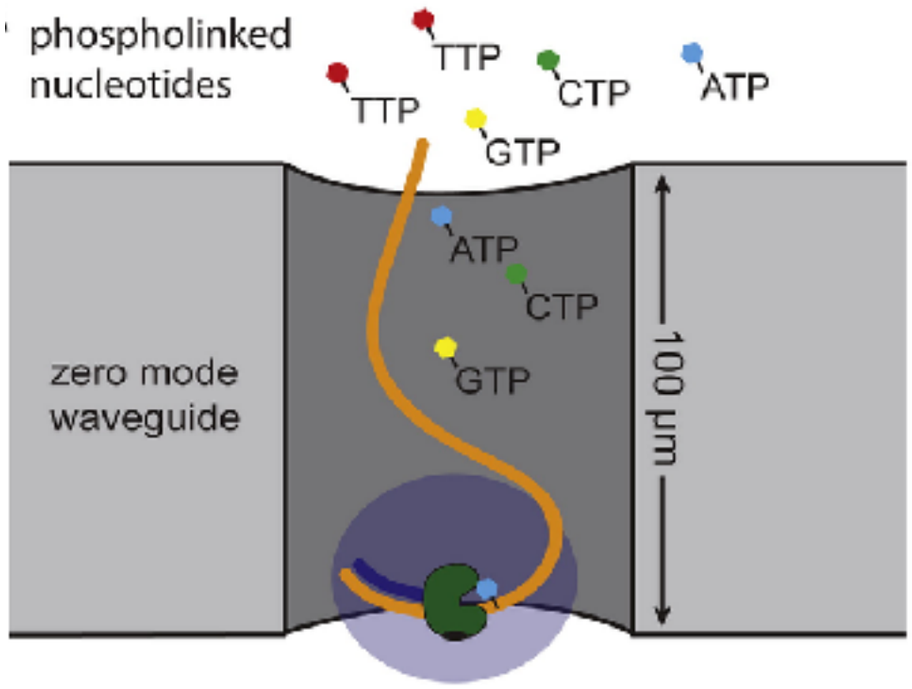
Cluster generation – hybridization and amplification

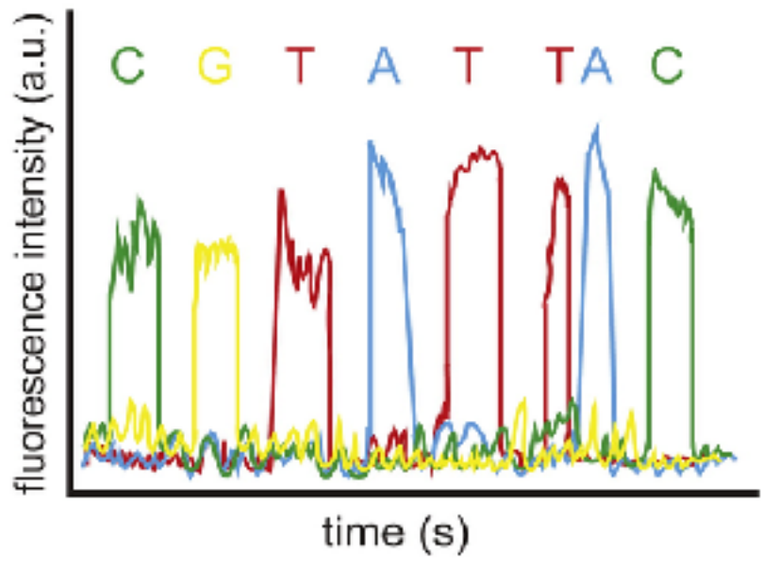


illumina

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

Pacific Biosciences

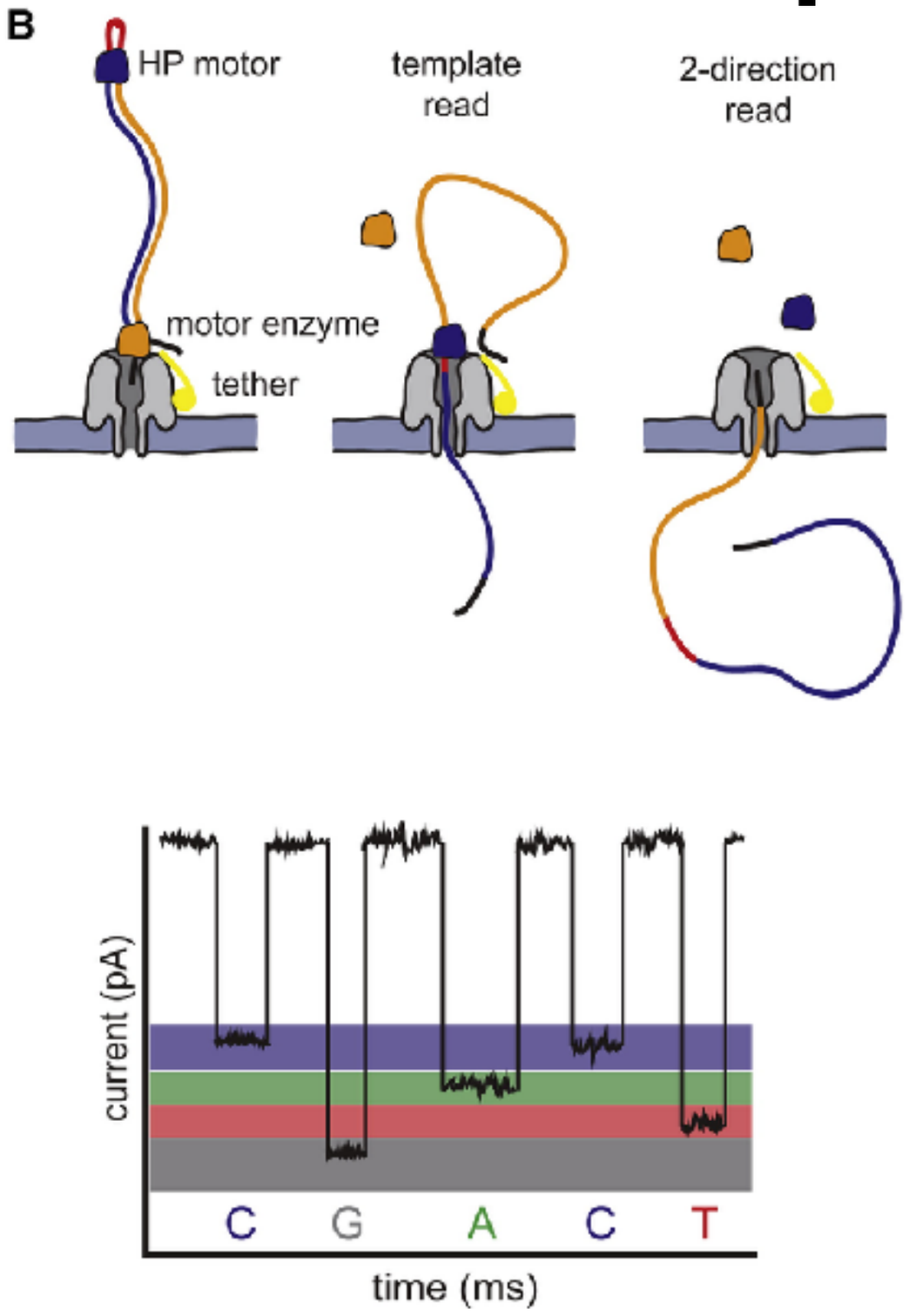




Reuter, et al 2015

1st Generation	2nd Generation	3rd Generation
Chemical (Maxim-Gilbert)	Pyrosequencing (454)	Single molecule real time (PacBio)
Chain Termination (Sanger)	Chain Termination (Illumina)	Nanopore sequencing (Oxford Nanopore)
Pyrosequencing	Sequencing by ligation (SOLiD sequencing)	
	Ion semiconductor (Ion Torrent)	

Oxford Nanopore



Reuter, et al 2015

Sequencers

https://www2.nanoporetech.com/images/product-page/MinIOn-Banner.jpg
http://www.gatc-biotech.com/en/gatc/sequencing-technologies/pacbio-rs-ii.html
http://www.dnavision.com/illumina.php

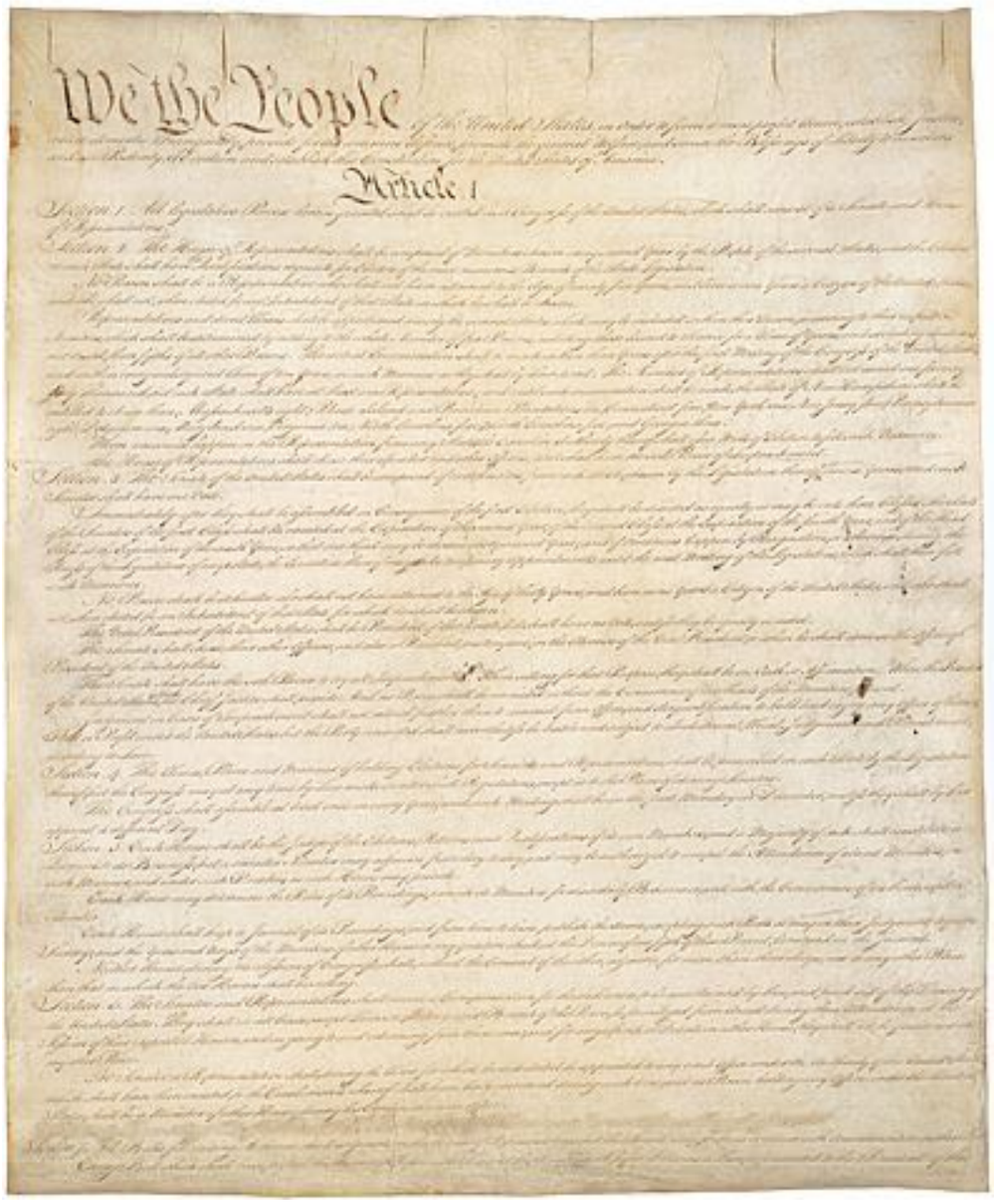
Why Long Reads?

- Structural Variation
  - Large Insertions or Deletions
    - Duplications
    - Translocations
- De Novo Genome Assembly
- Phasing
- Large Amplicons

Short Reads

```
e of the U
stablish J
Union, est
nited Stat
 to form a
rder to fo
e perfect
ion, estab
eople of t
 the Peopl
```

“Genome” Reference



Reference Based Mapping

We the People of the United States, in Order to form a more perfect Union, establish Justice, insur

e of the U
stablish J
Union, est
nited Stat
 to form a
rder to fo
e perfect
ion, estab
eople of t
 the Peopl

Reference Based Mapping

We the People of the United States, in Order to form a more perfect Union, establish Justice, insur

```
the Peopl
     eople of t
         e of the U
                   nited Stat
                                      rder to fo
                                           to form a
                                                        e perfect
                                                                  Union, est
                                                                    ion, estab
                                                                          stablish J
```

De Novo Assembly

Overlapping Random Fragments

rious disg

Age. "You

rinking Ag

uises of A

the portra

ugh the po

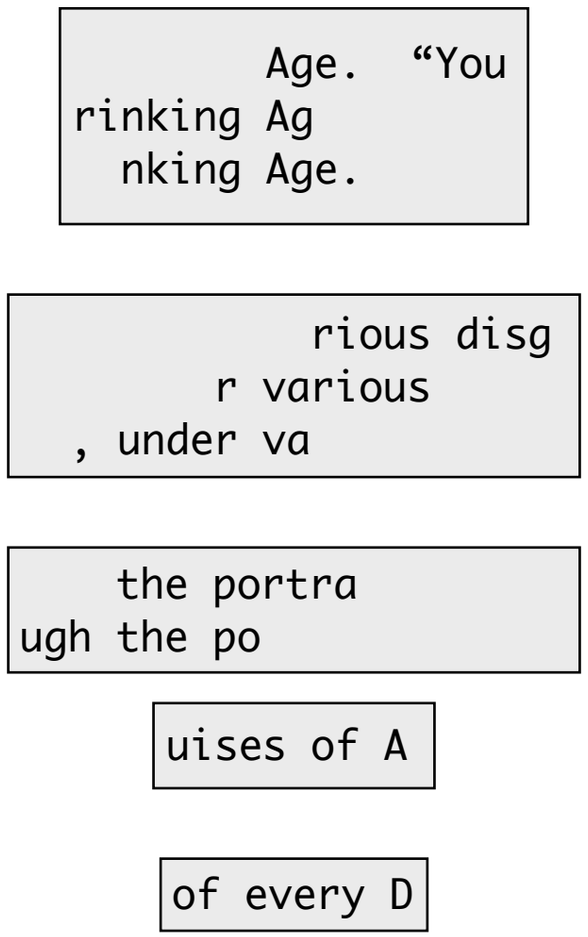
of every D

nking Age.

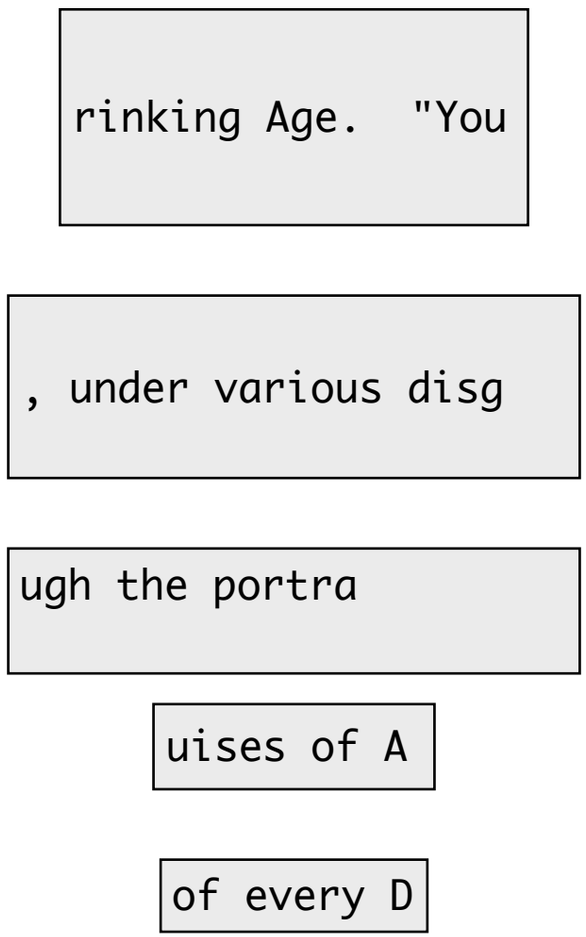
r various

, under va

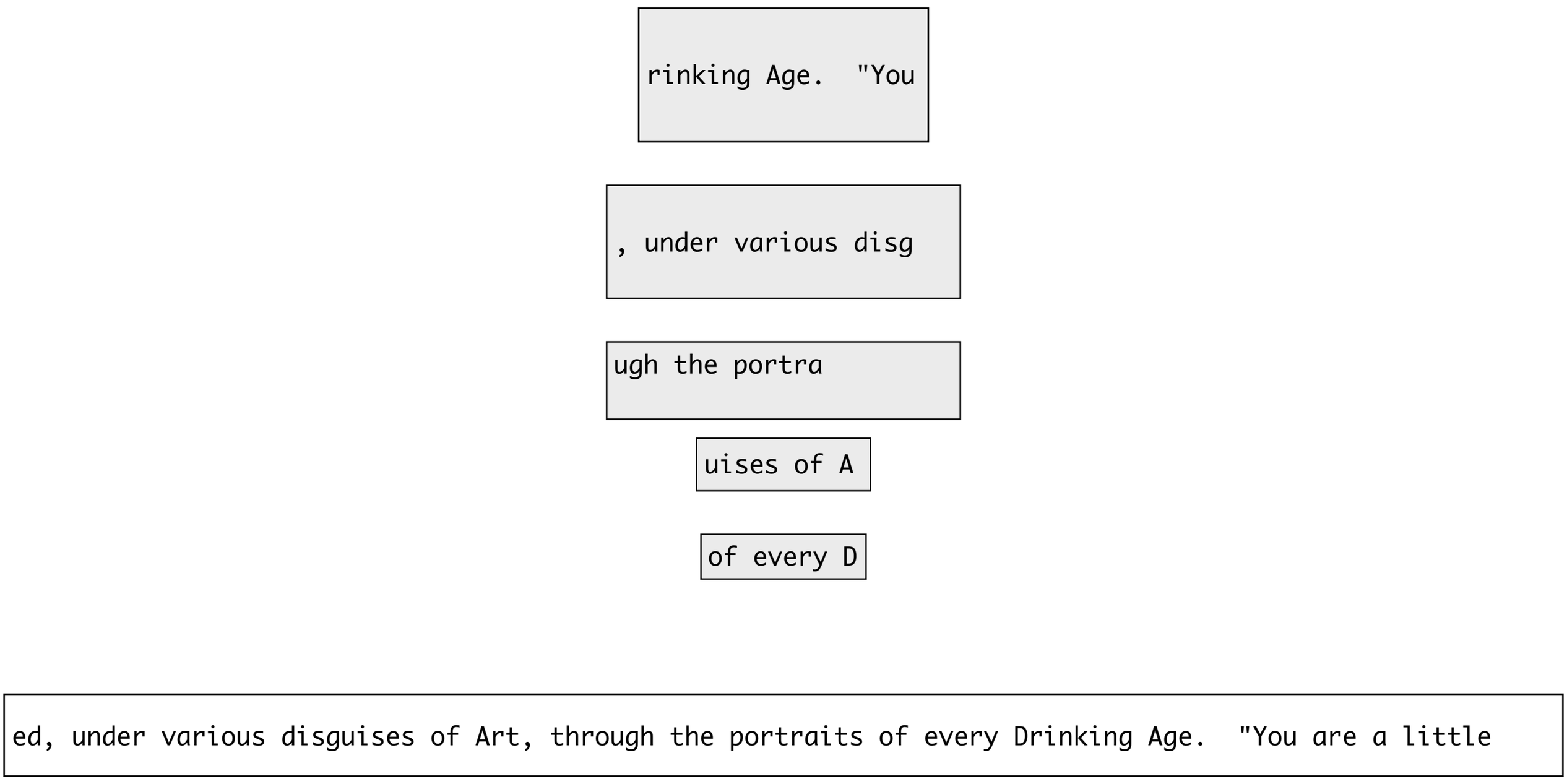
Assemble Contigs



Assemble Contigs



Assemble Contigs



More Reads

rious disg
Age."You
rinking Ag
uises of A
the portra
ugh the po
of every D
nking Age.
r various
, under va
Age."Yo
rough the
rinking Ag
ed, under
ugh the po
ry Drinkin
sguises of
 u are a li
"You are
, under va

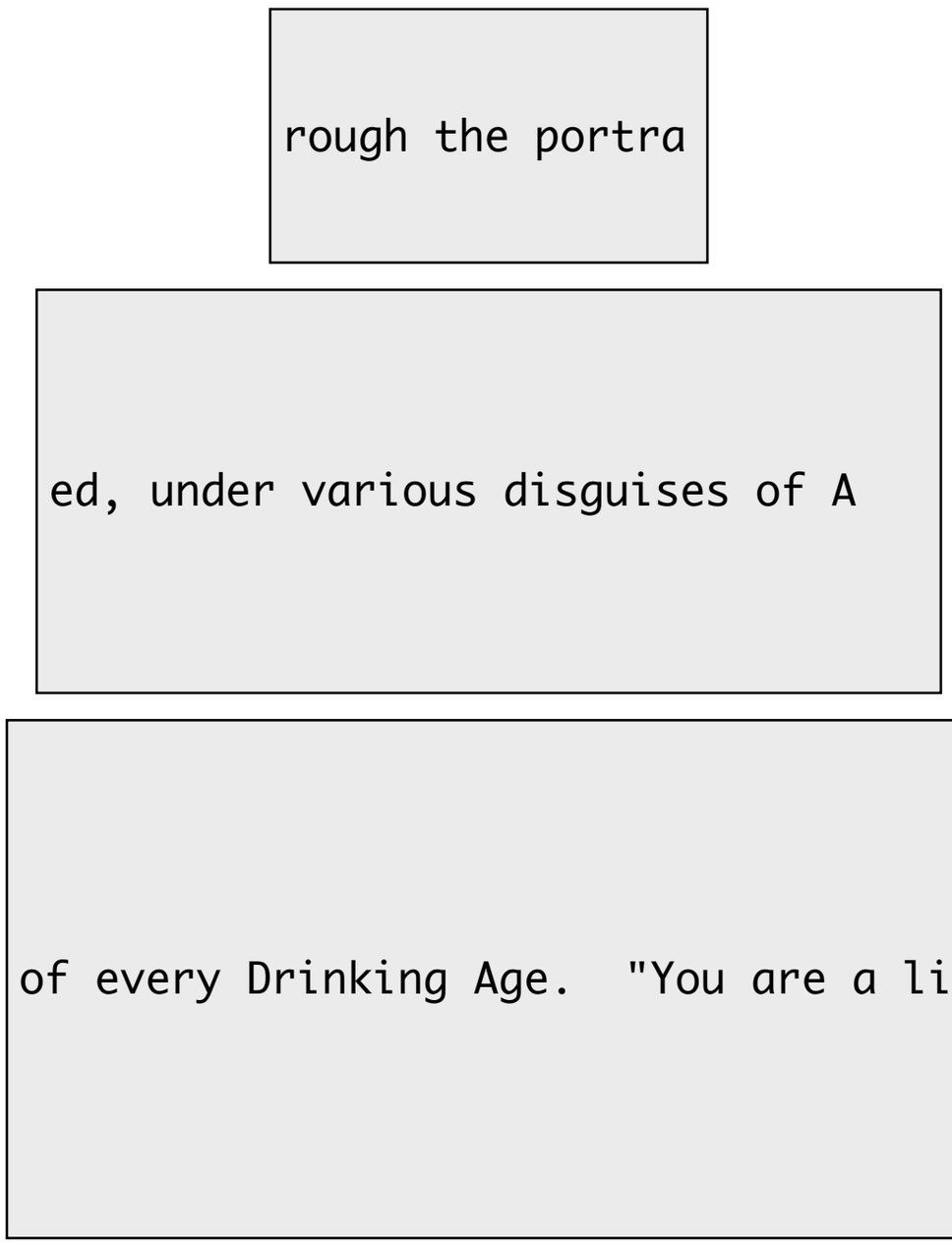
More Reads

```
rough the
  ugh the po
  ugh the po
      the portra
```

```
ed, under
  , under va
  , under va
       r various
           rious disg
                   sguises of
                     uises of A
```

```
of every D
     ry Drinkin
         rinking Ag
         rinking Ag
           nking Age.
                 Age.  "You
                 Age.  "Yo
                       "You are
                          u are a li
```

More Reads


rough the portra
ed, under various disguises of A
of every Drinking Age. "You are a li

More Reads

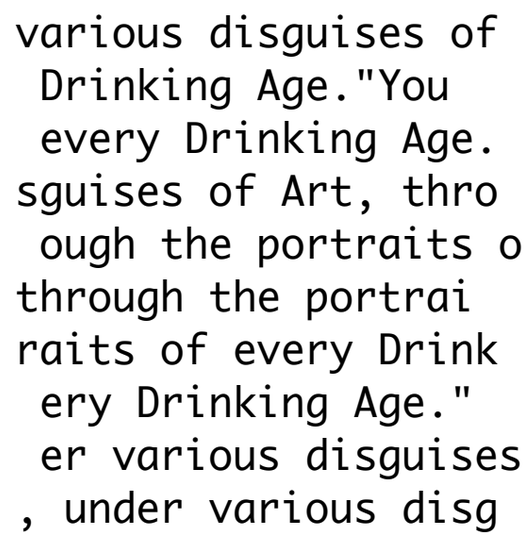
rough the portra

ed, under various disguises of A

of every Drinking Age. "You are a li

ed, under various disguises of Art, through the portraits of every Drinking Age. "You are a little

Longer Reads



Longer Reads

```
, under various disg
     er various disguises
        various disguises of
                  sguises of Art, thro
                                  through the portrai
                                     ough the portraits o
                                                 raits of every Drink
                                                          every Drinking Age.
                                                            ery Drinking Age.  "
                                                                Drinking Age.  "You
```

Longer Reads

```
, under various disguises of Art, through the portraits of every Drinking Age.  "You
```

```
ed, under various disguises of Art, through the portraits of every Drinking Age.  "You are a little
```

Fragmentation

"You
Age.
Art,
Drinking
a
are
disguises
ed,
every
little
of
of
portraits
the
through
under
various

Problem Sequences

- Repeats
  - Transposons
  - Centromeres
- Homologs
- Duplications

De novo “Reference”

ed, under various disguises of Art, through the portraits of every Drinking Age. "You are a little

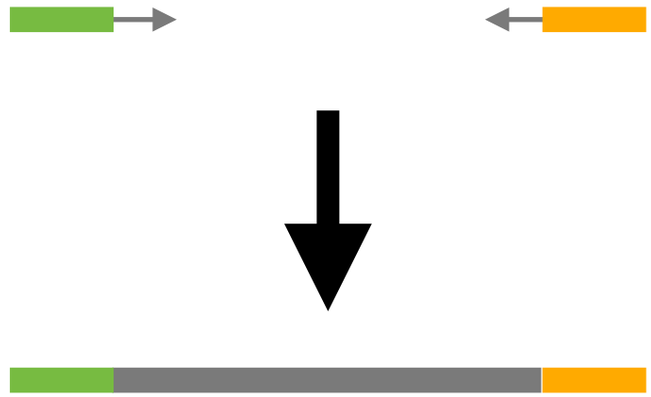
— *A Tale of Two Cities*

Library Preparation

Amplicon Library Prep

Purified DNA

PCR Amplification



DNA-Seq Library Prep

Amplicon-Seq

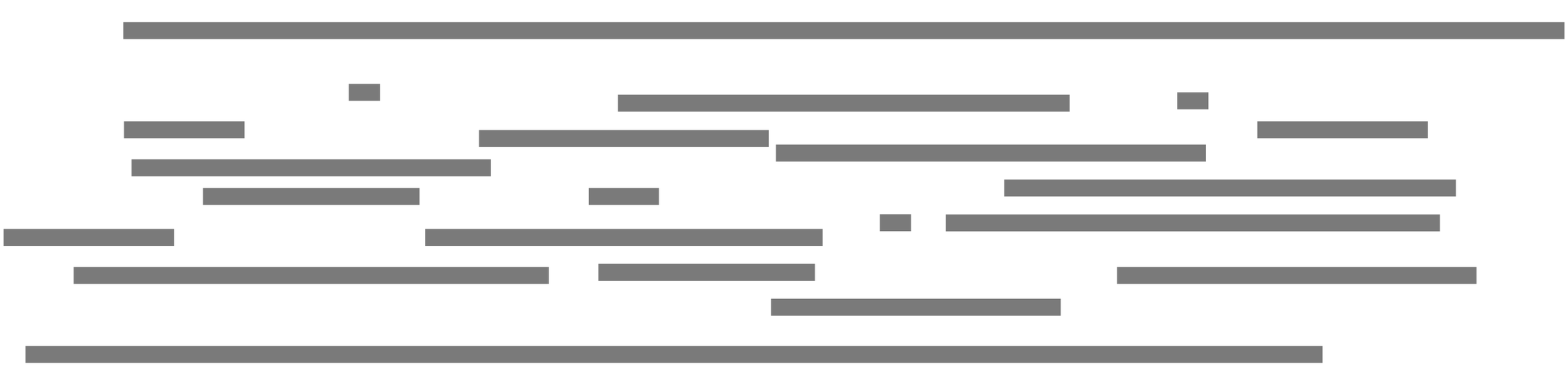
1. Purify DNA
2. PCR Amplify with Adapters

DNA-Seq

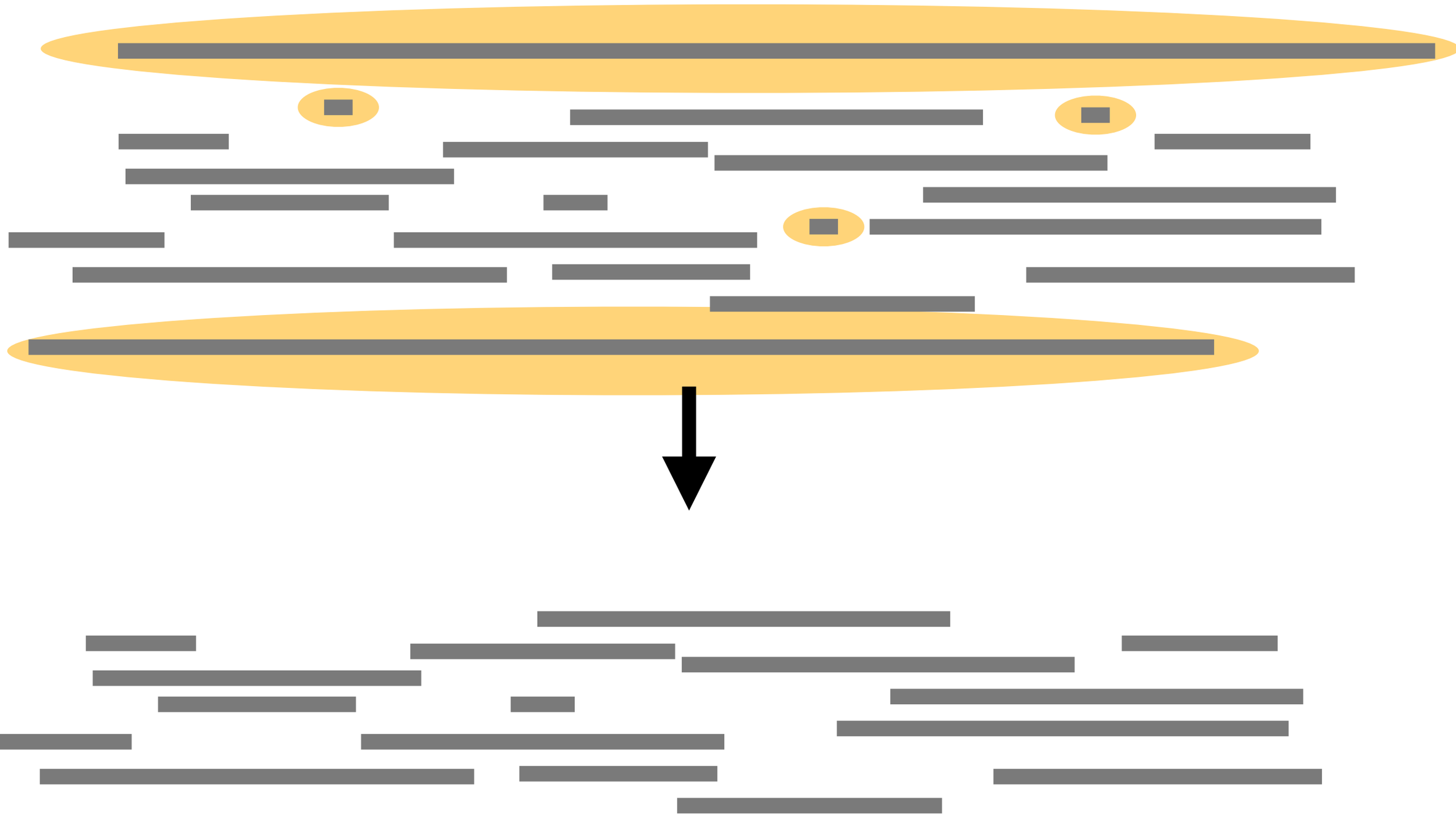
1. Purify DNA
2. Fragment
3. Size Select
4. Adapter Ligation

Purified DNA

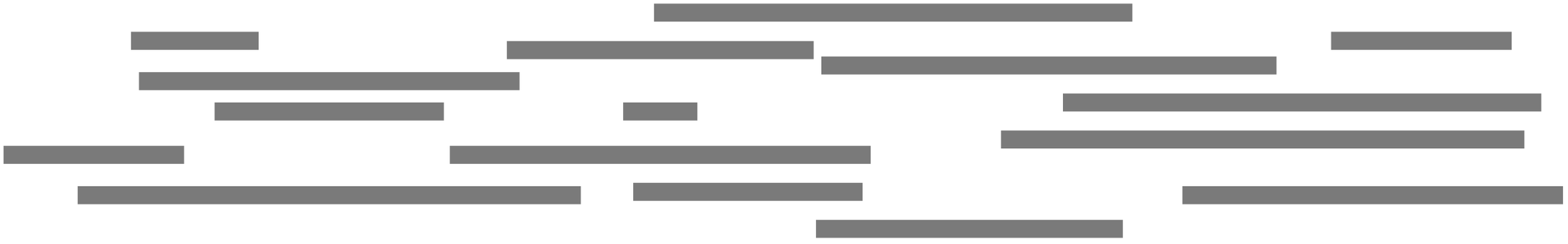
Fragmentation

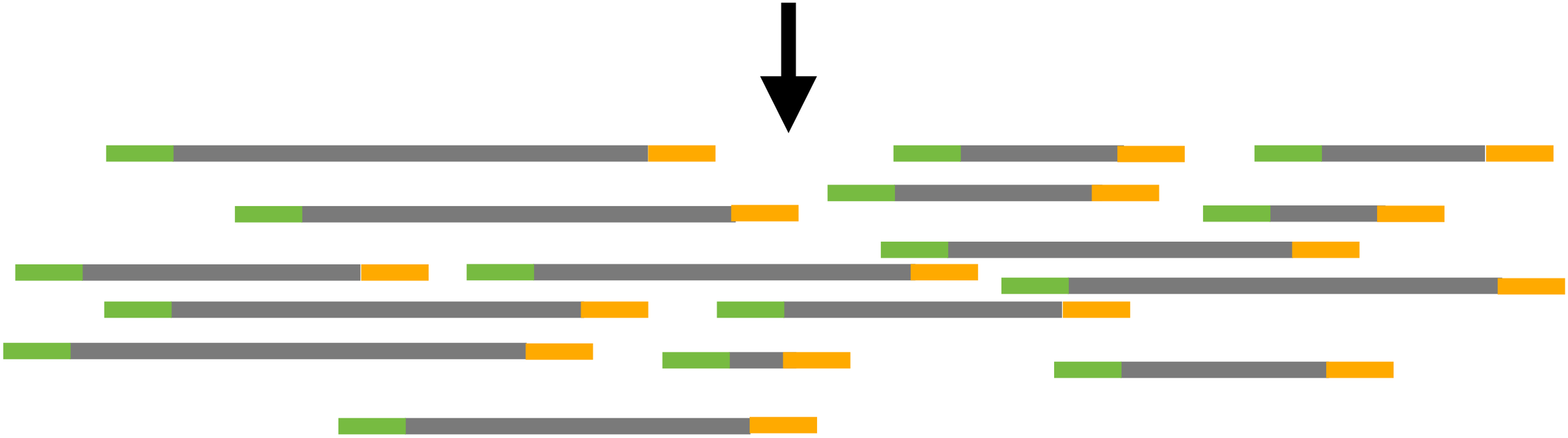


Size Selection



Adapter Ligation





RNA-Seq Library Prep

DNA-Seq

1. Purify DNA
2. Fragment
3. Size Select
4. Adapter Ligation

RNA-Seq

1. Purify RNA
2. Fragment
3. Size Select
4. Make DNA From RNA
5. Adapter Ligation

Sequencing Library

Amplicon Library

Shotgun Library