

# Next Generation Sequencing

This slide deck will be obsolete in 5..4..

# Outline

- Some history
- Workflows for different platforms
- Applications

With liberal lifting from:

Andy Vierstraete, Ghent University

June 2012 slides at <http://users.ugent.be/~avierstr/nextgen/>

# Some history

1953 : Discovery of DNA structure by Watson and Crick

1973 : First sequence of 24 bp published

1977 : Sanger sequencing method published

1980 : Nobel Prize Wally Gilbert and Fred Sanger

1982 : Genbank started

1983 : Development of PCR

1987 : 1st automated sequencer : Applied Biosystems Prism 373

1996 : Capillary sequencer : ABI 310

1998 : Genome of *Caenorhabditis elegans* sequenced

2000 : Human genome sequenced

2005 : 1st 454 Life Sciences Next Generation Sequencing system : GS 20 System

2006 : 1st Solexa Next Generation Sequencer : Genome Analyzer

2007 : 1st Applied Biosystems Next Generation Sequencer : SOLiD

2009 : 1st Helicos single molecule sequencer : Helicos Genetic Analyser System

2011 : 1st Ion Torrent Next Generation Sequencer : PGM

2011 : 1st Pacific Biosciences single molecule sequencer : PacBio RS Systems

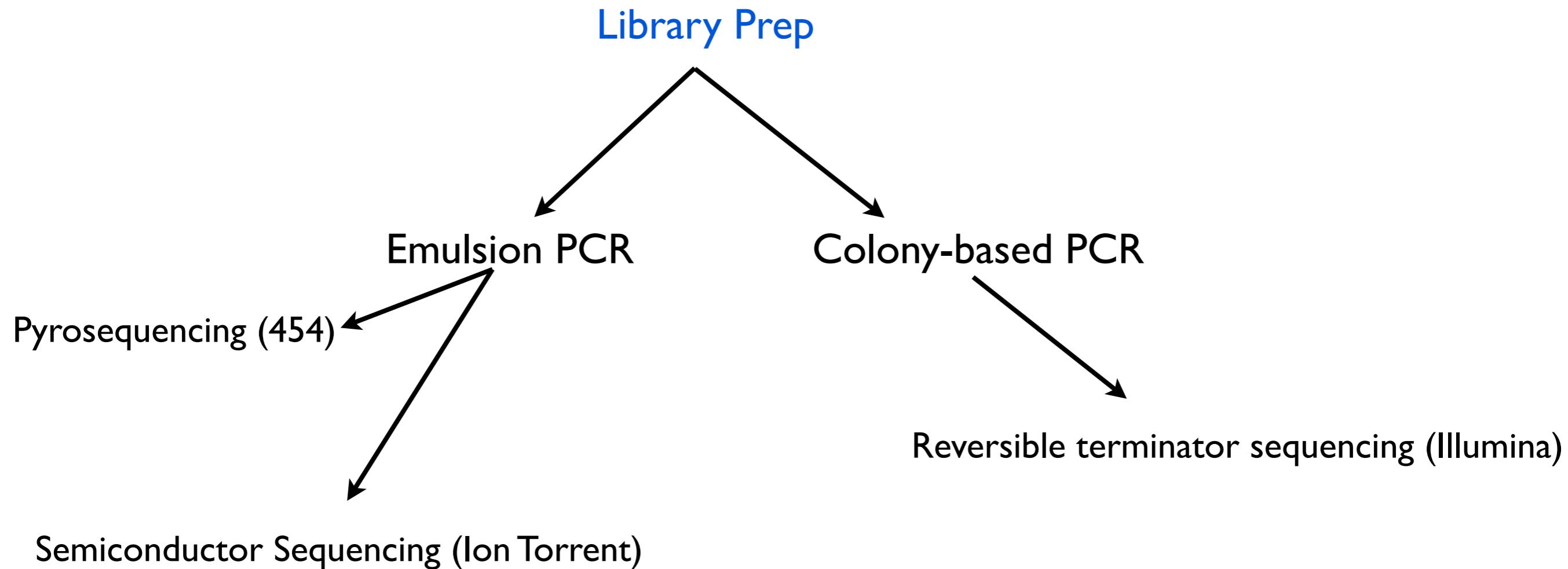
# Overview of Platforms

- 454 Sequencing (Roche)
  - ➔ GS Junior, GS FLX+
- Illumina (previously Solexa)
  - ➔ HiSeq, Genome analyzer IIx, MySeq
- Applied Biosystems (Life Technologies)
  - ➔ SOLiD 5500
- Ion Torrent (Life Technologies)
  - ➔ Personal Genome Machine, Proton
- Helicos
  - ➔ Helicos Genetic Analysis System
- Pacific Biosciences
  - ➔ PacBio RS
- Oxford Nanopore Technologies
  - ➔ GridION System MinION

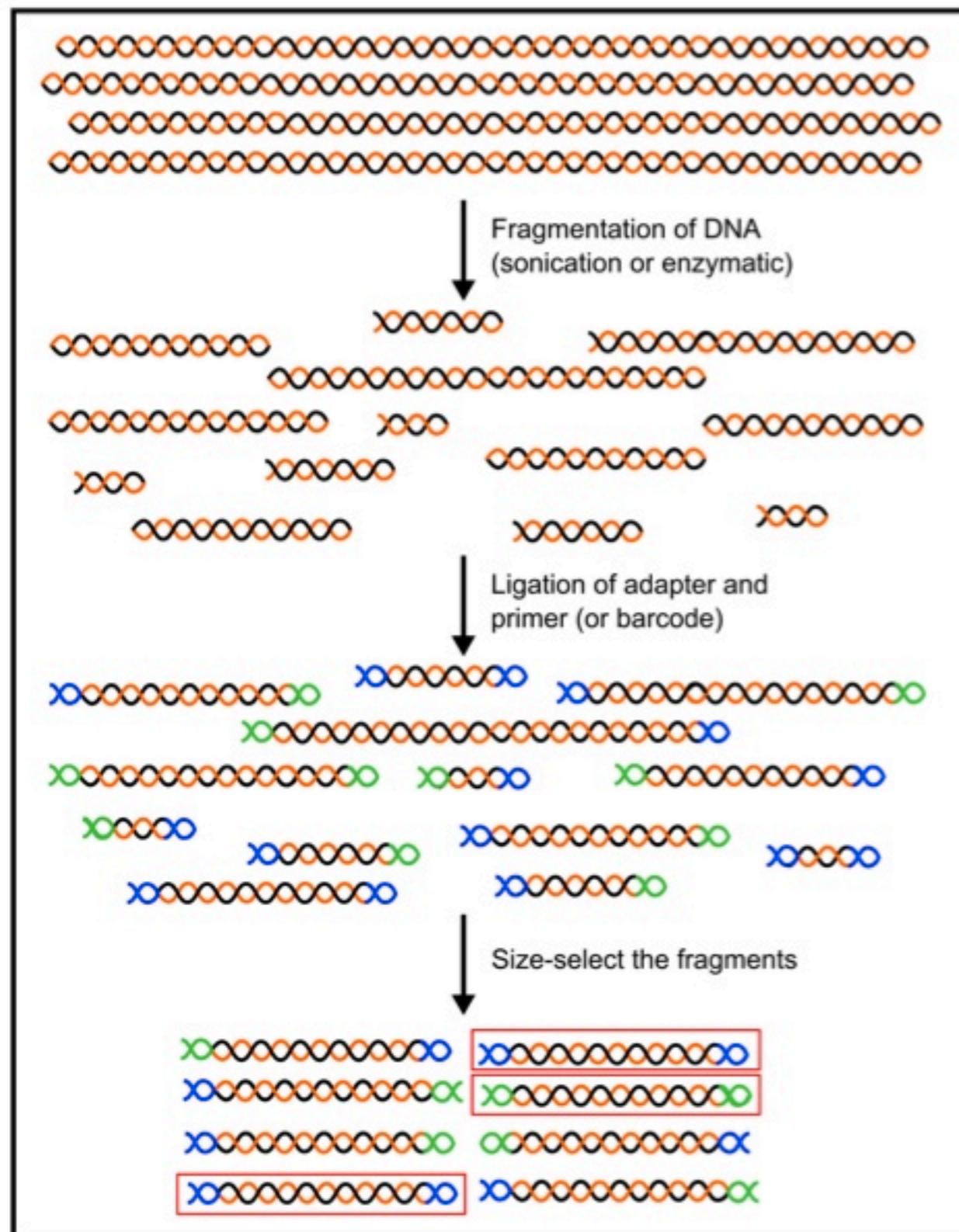
Amplified

Single-molecule

# Workflow

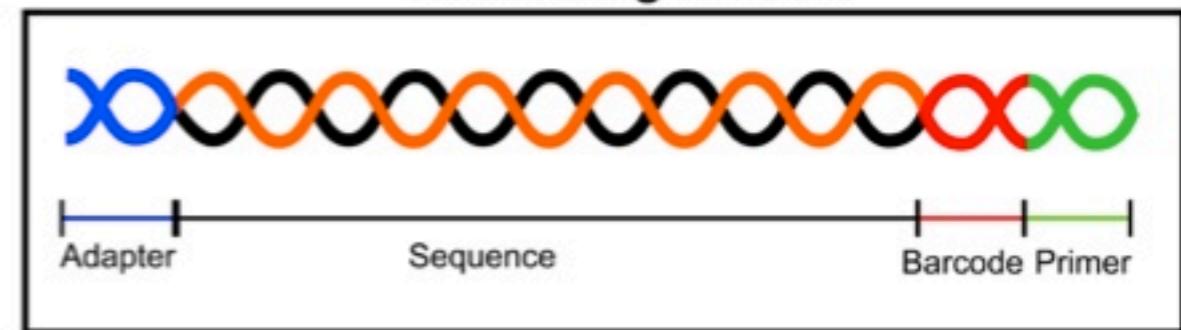


# Workflow: Library Prep

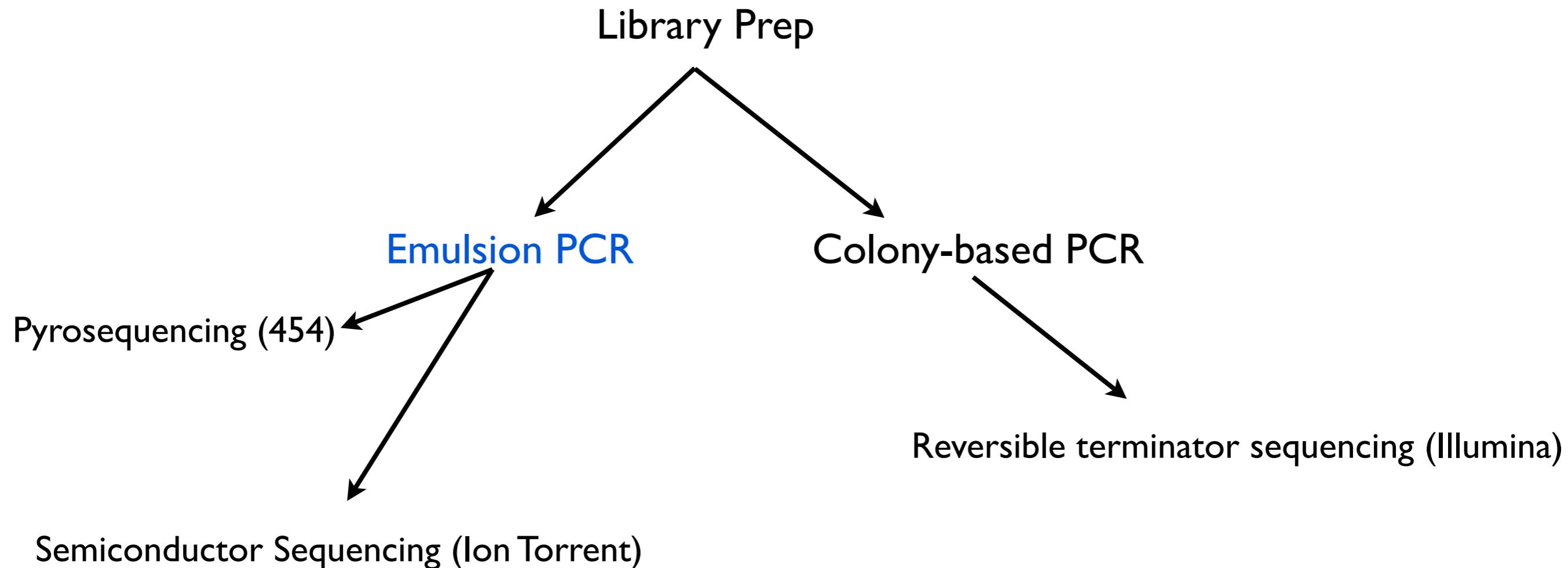


## Library preparation

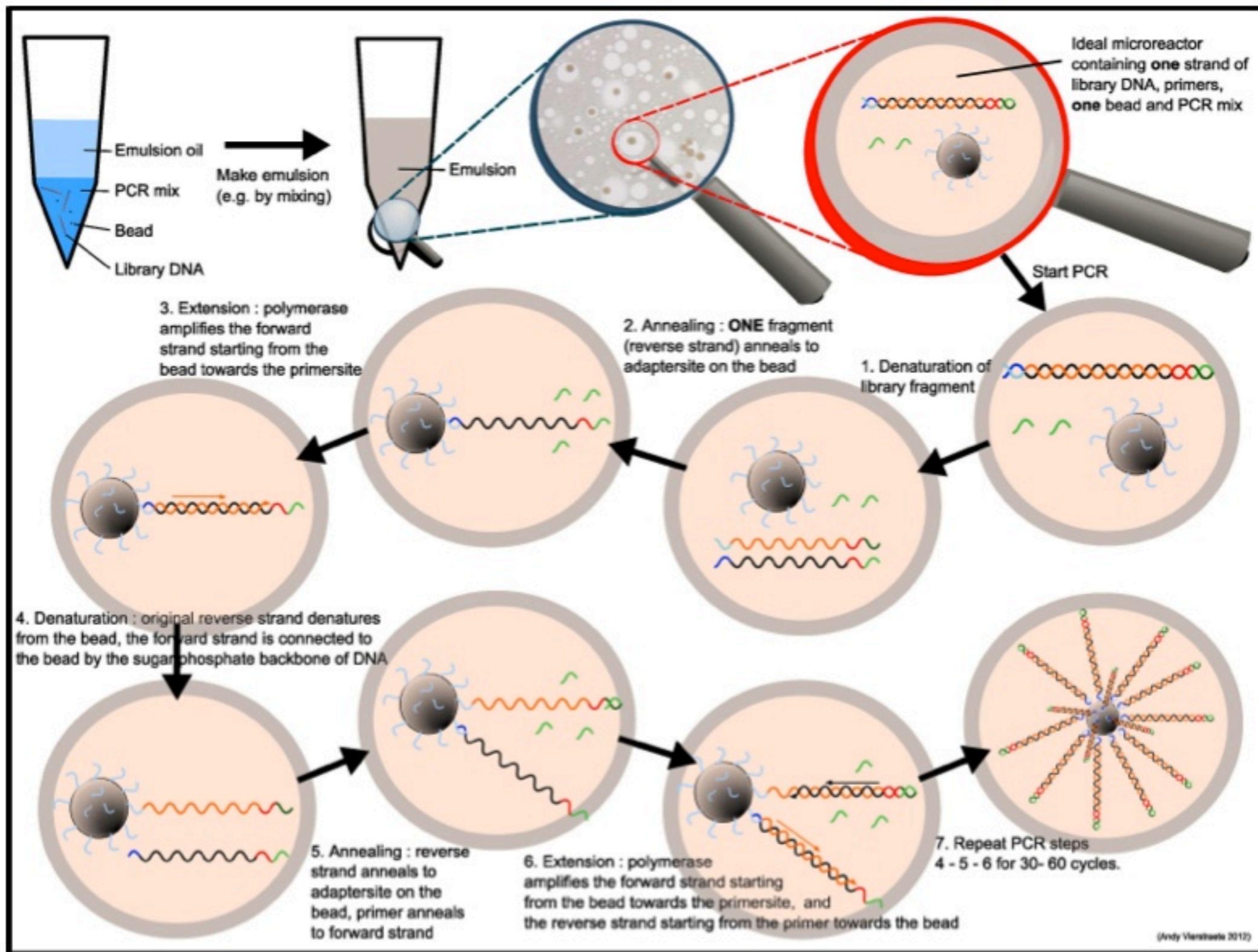
Good fragments :



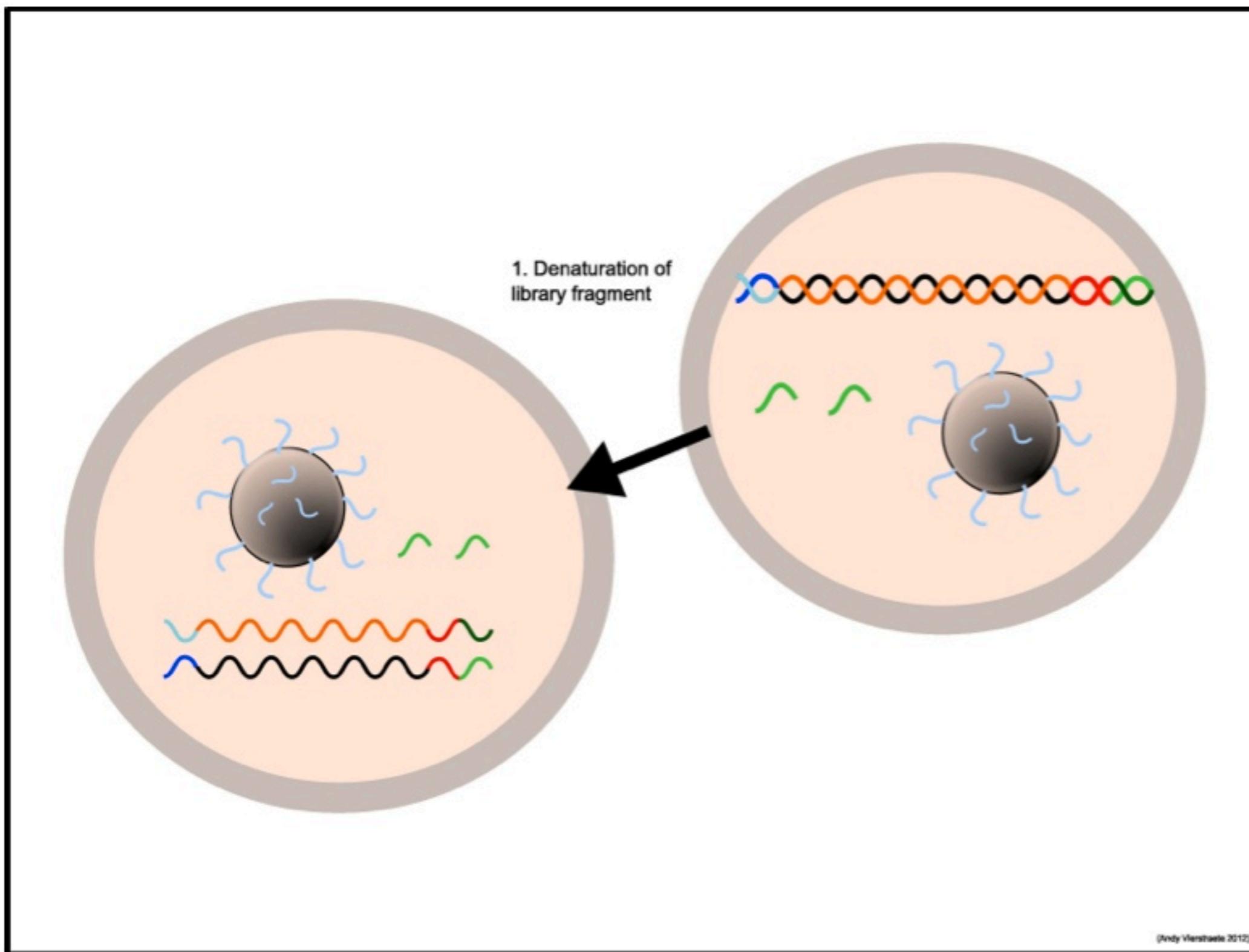
# Workflow



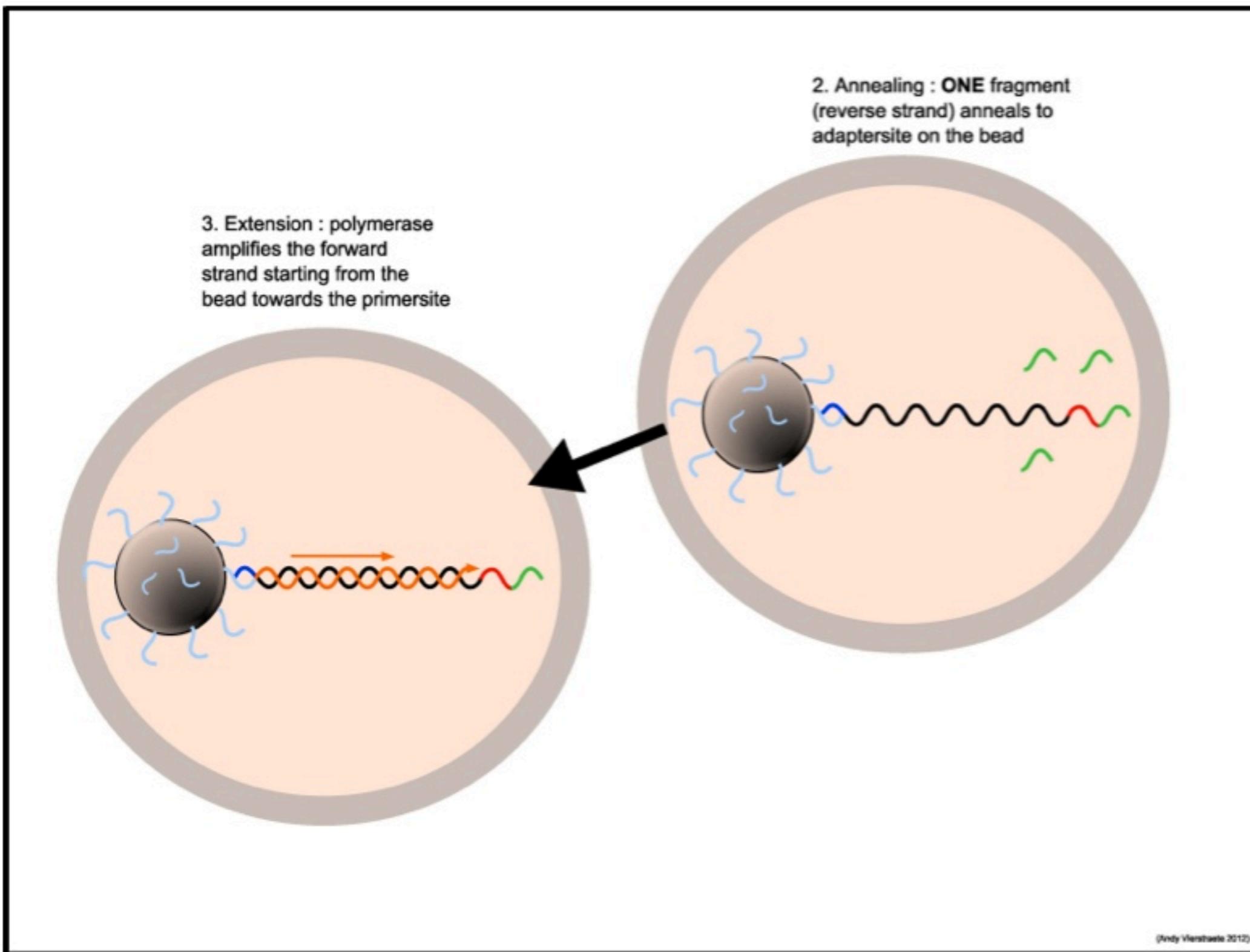
# Workflow: Emulsion PCR



# Workflow: Emulsion PCR

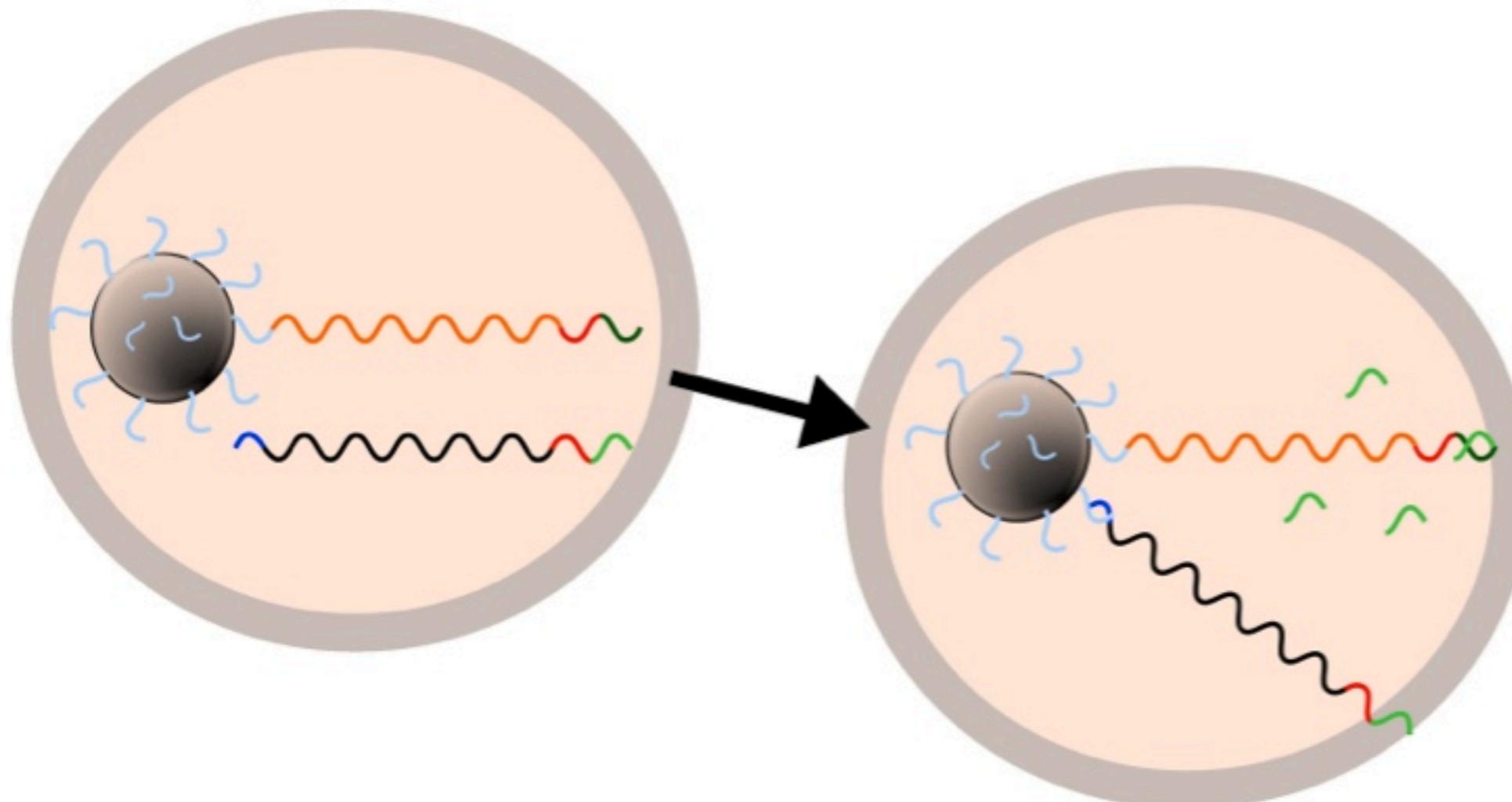


# Workflow: Emulsion PCR



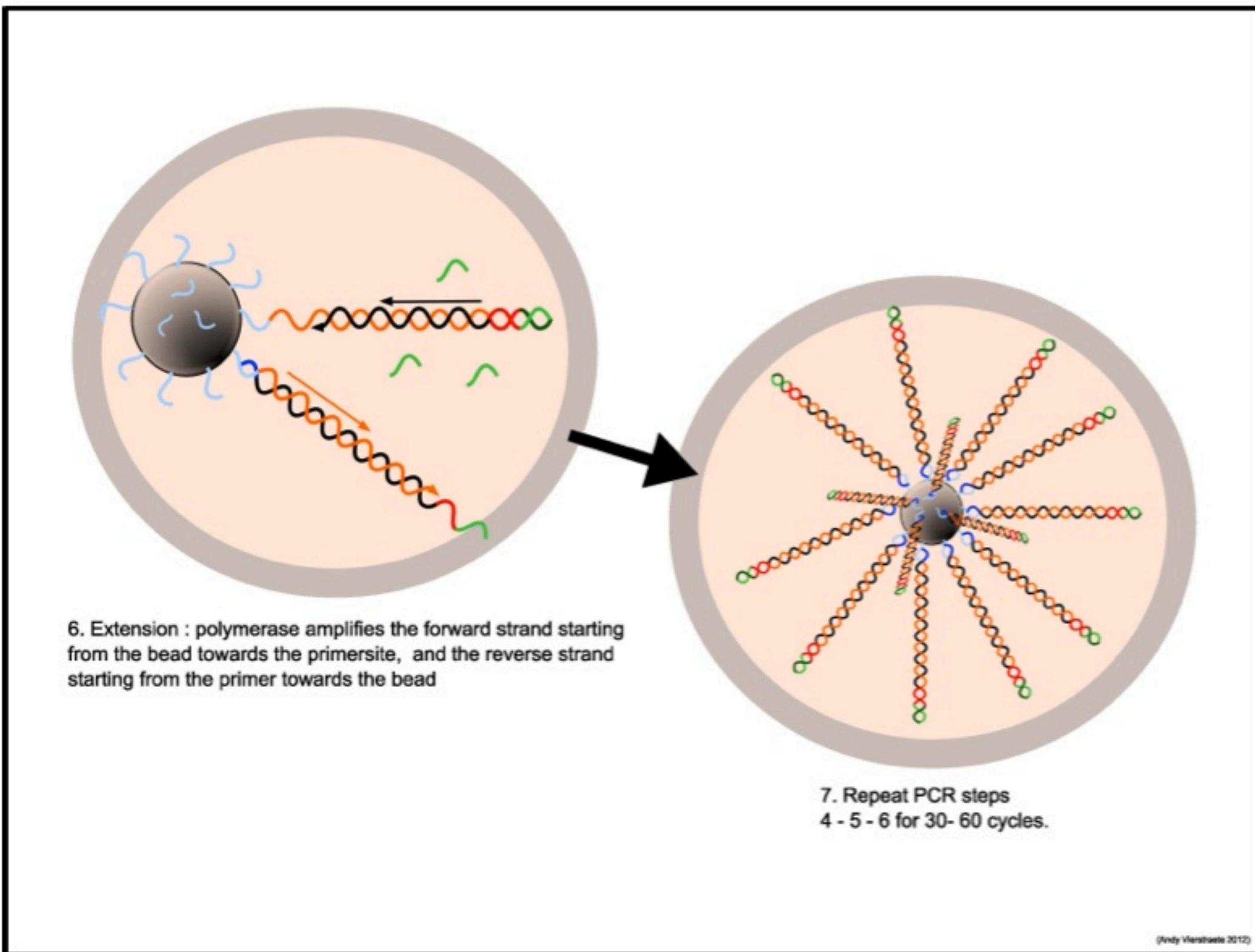
# Workflow: Emulsion PCR

4. Denaturation : original reverse strand denatures from the bead, the forward strand is connected to the bead by the sugar phosphate backbone of DNA



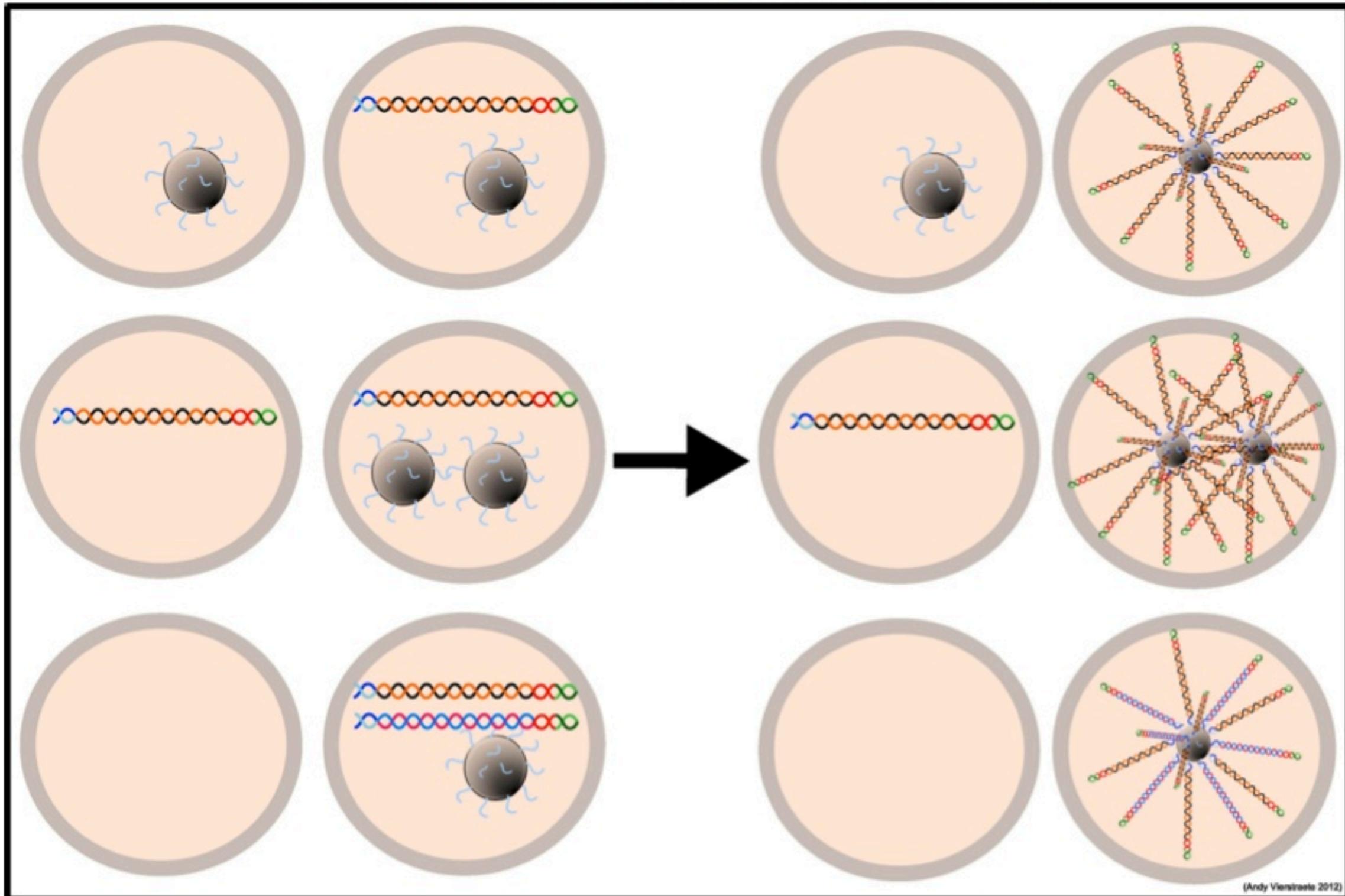
5. Annealing : reverse strand anneals to adaptersite on the bead, primer anneals to forward strand

# Workflow: Emulsion PCR

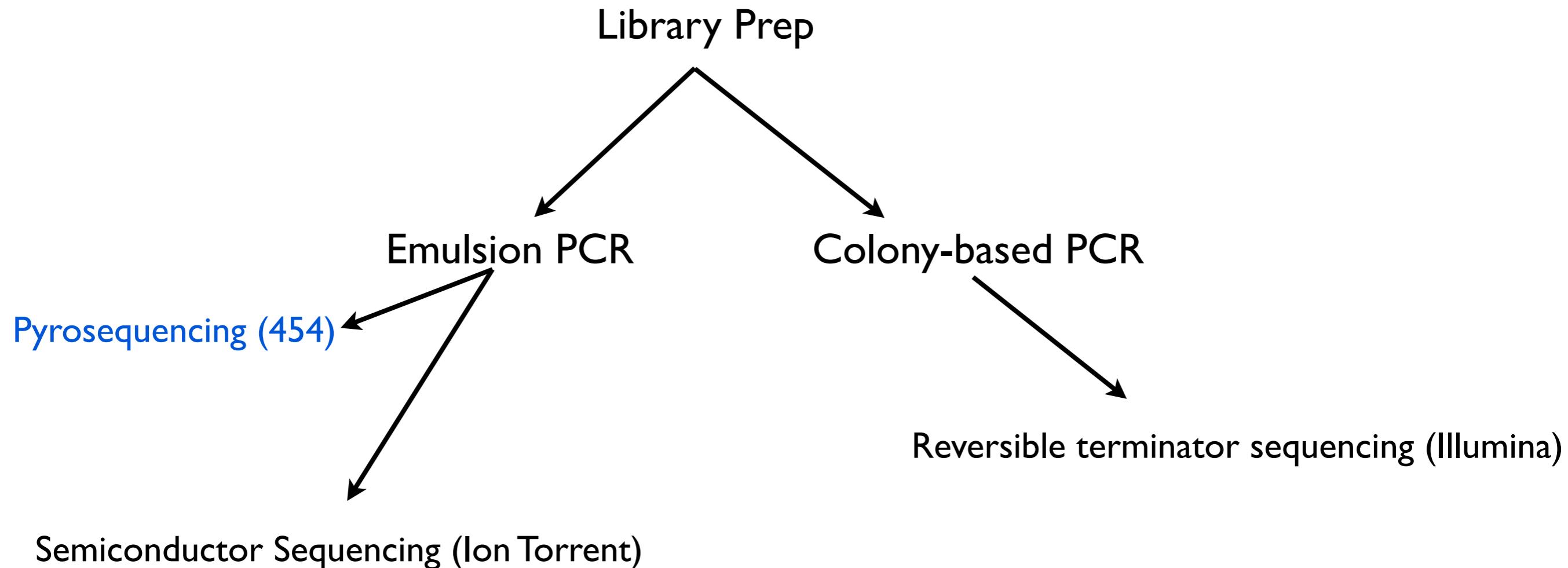


# Workflow: Emulsion PCR

different micro reactors : only 15 % are good ones



# Workflow



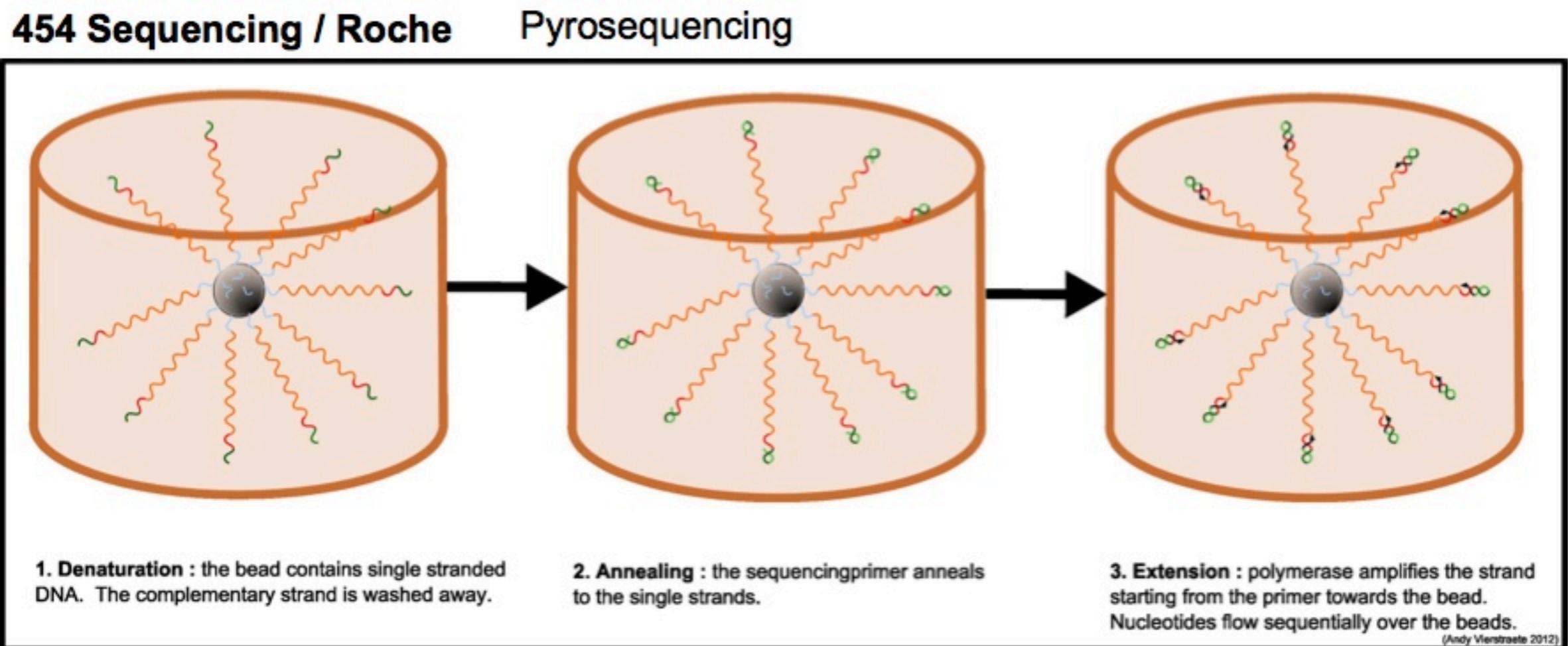
# Workflow: Pyrosequencing



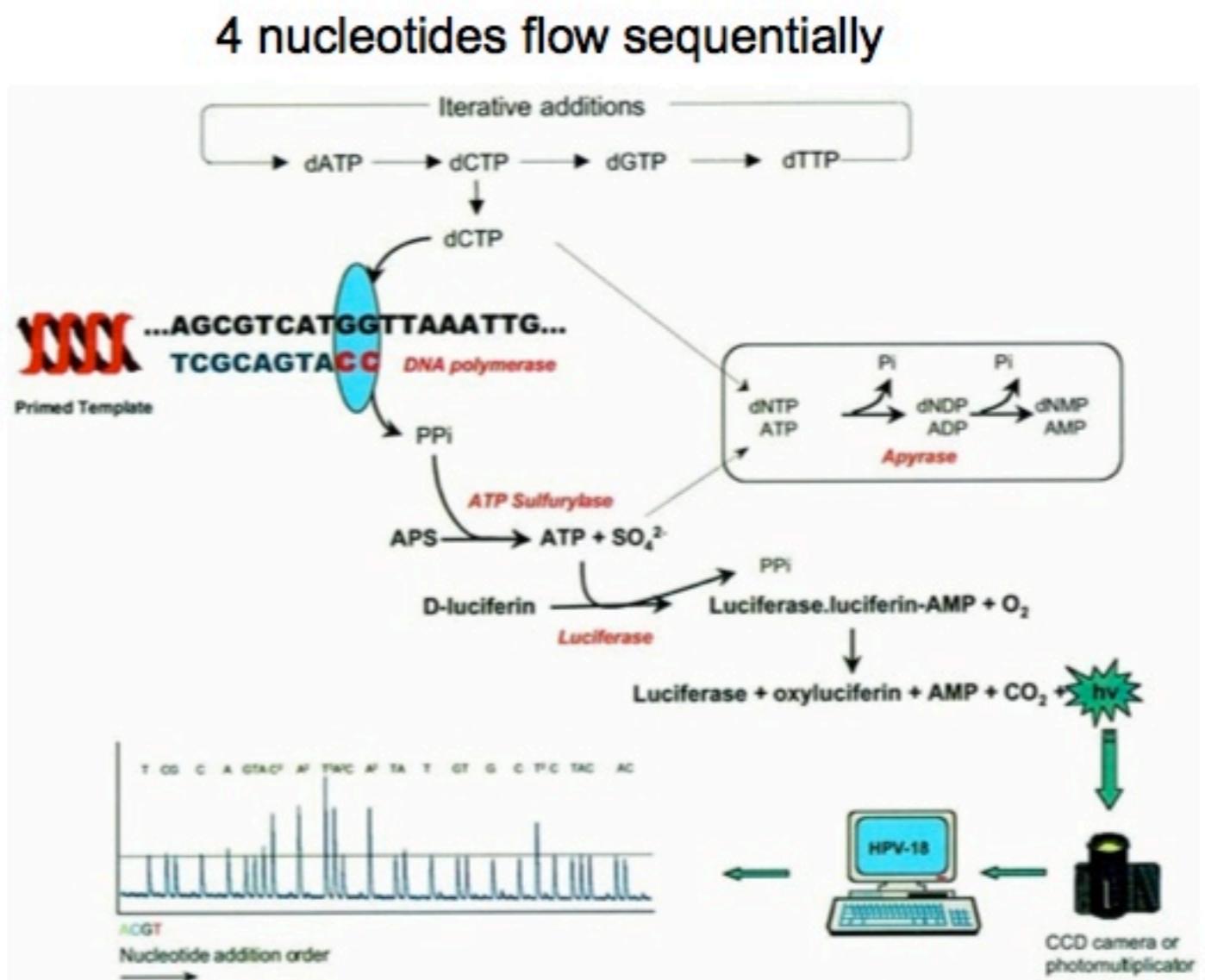
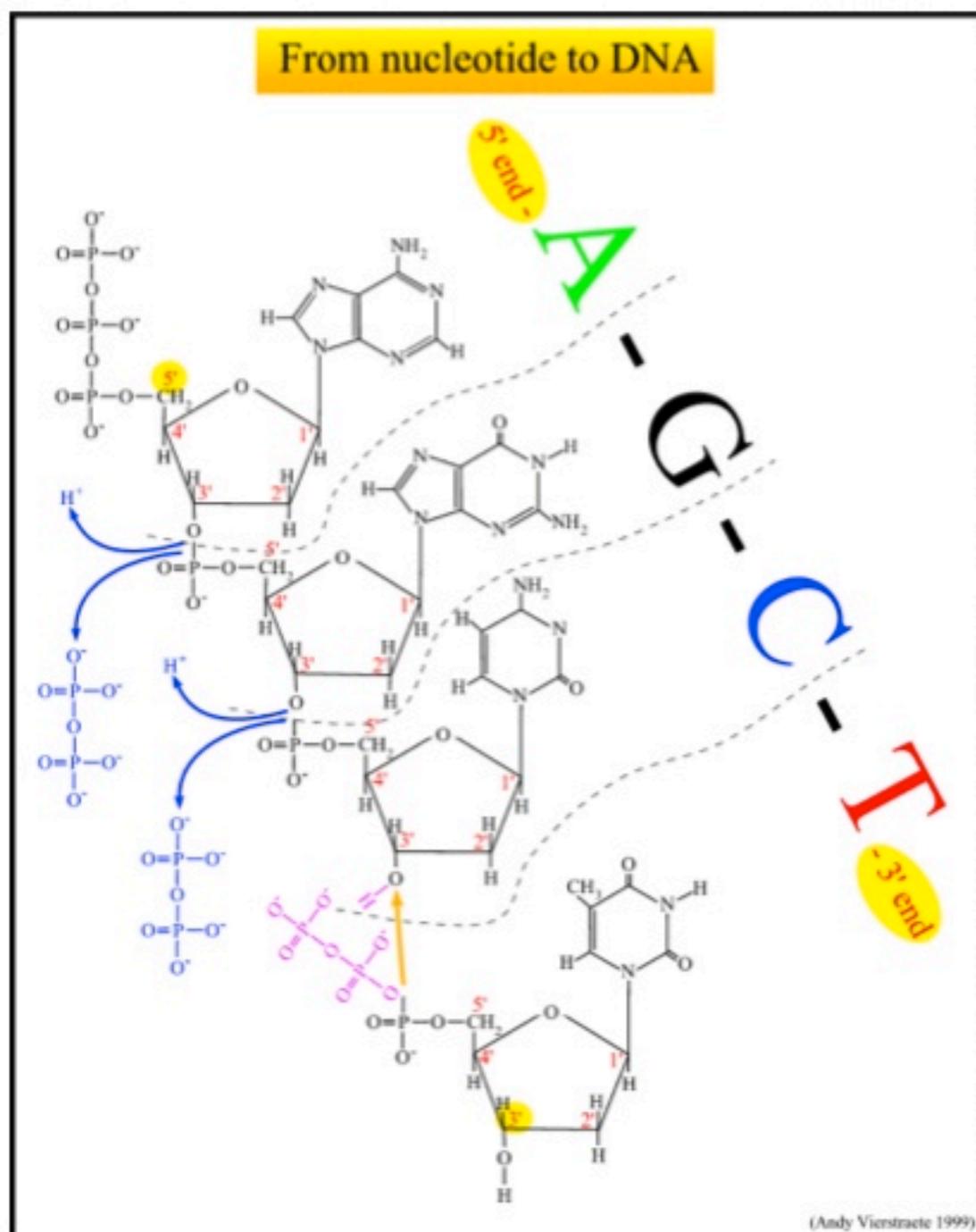
## 454 Sequencing / Roche

|                      | <b>GS Junior</b> | <b>GS FLX Titanium XL+</b> | <b>GS FLX Titanium XLR70</b> |
|----------------------|------------------|----------------------------|------------------------------|
| <b>Read Length</b>   | 400 bp           | 700 bp                     | 450 bp                       |
| <b>Throughput</b>    | 35 Mb            | 700 Mb                     | 450 Mb                       |
| <b>Reads per run</b> | 100,000          | 1,000,000                  | 1,000,000                    |
| <b>Accuracy</b>      | 99 %             | 99,997 %                   | 99,995 %                     |
| <b>Run Time</b>      | 10 hours         | 23 hours                   | 10 hours                     |

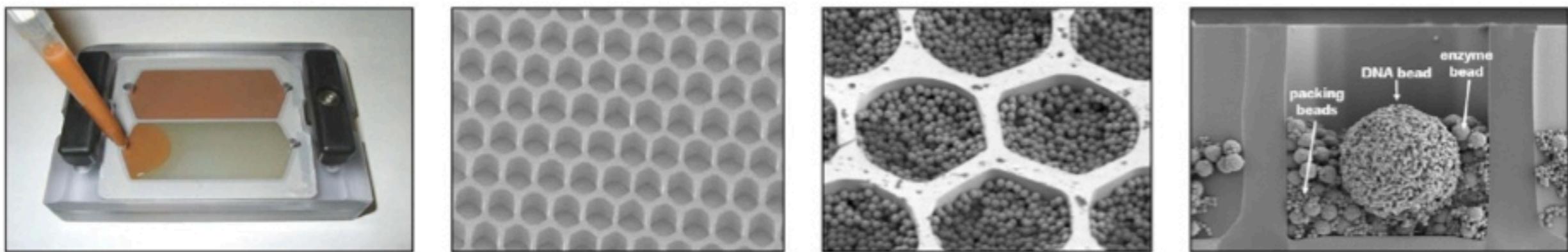
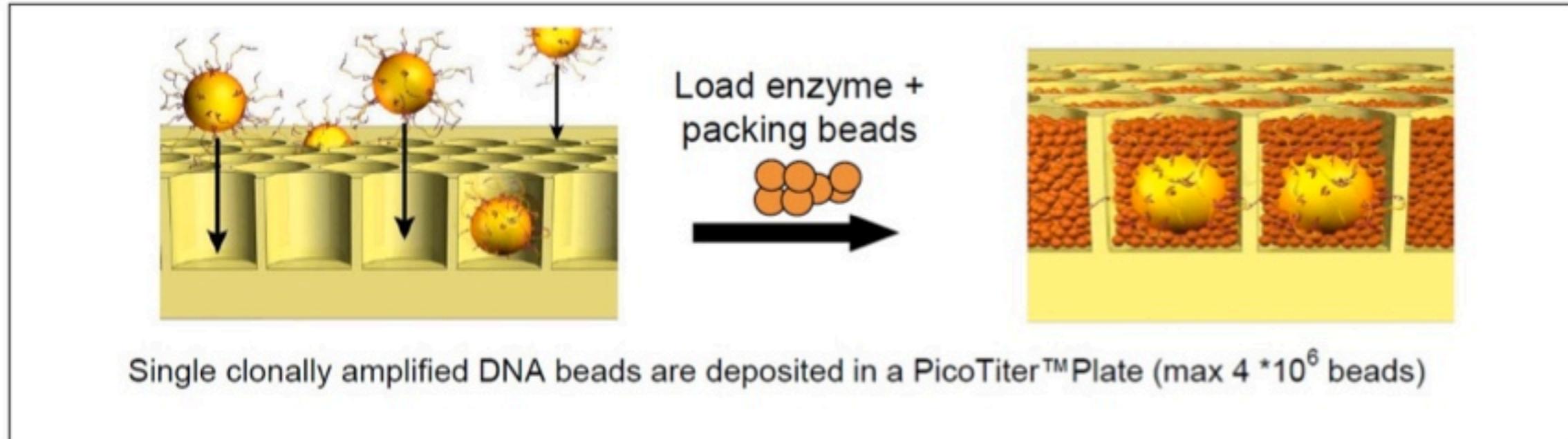
# Workflow: Pyrosequencing



# Workflow: Pyrosequencing

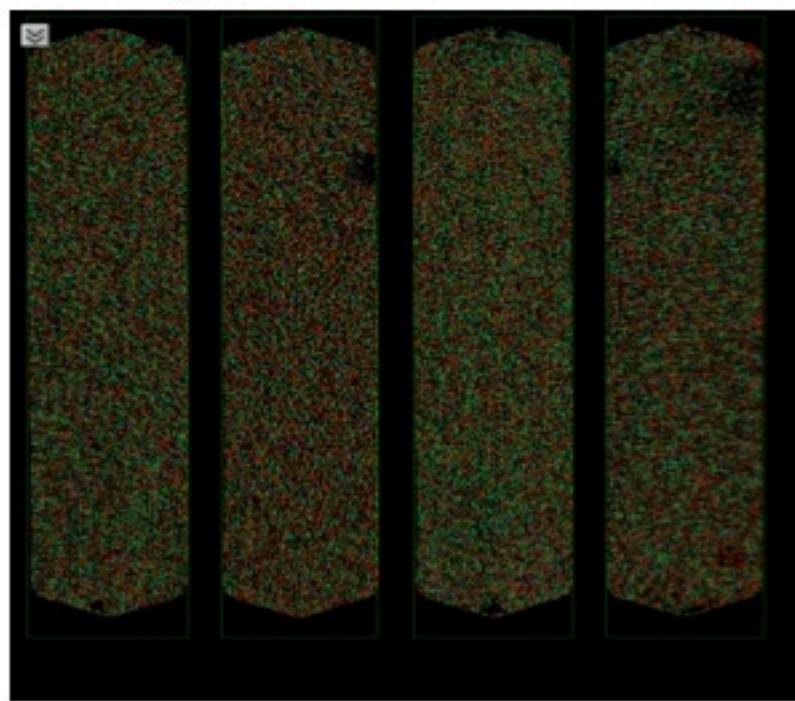


# Workflow: Pyrosequencing

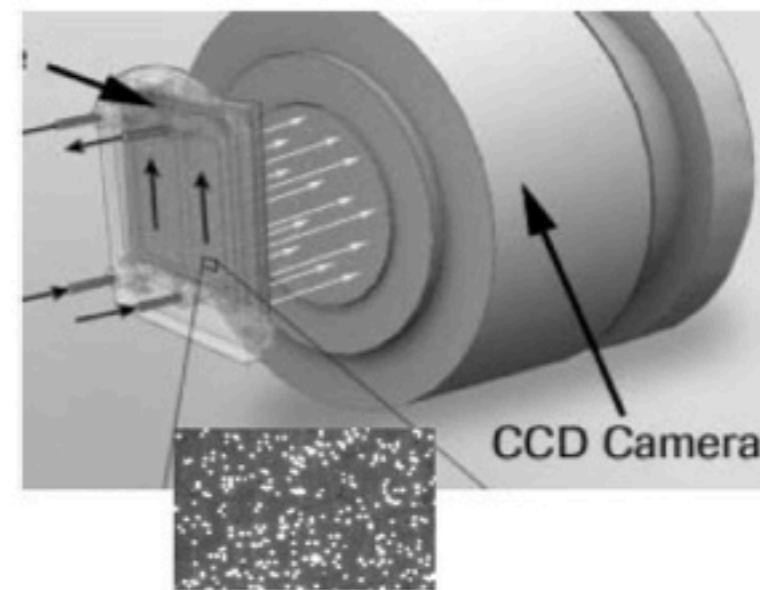


# Workflow: Pyrosequencing

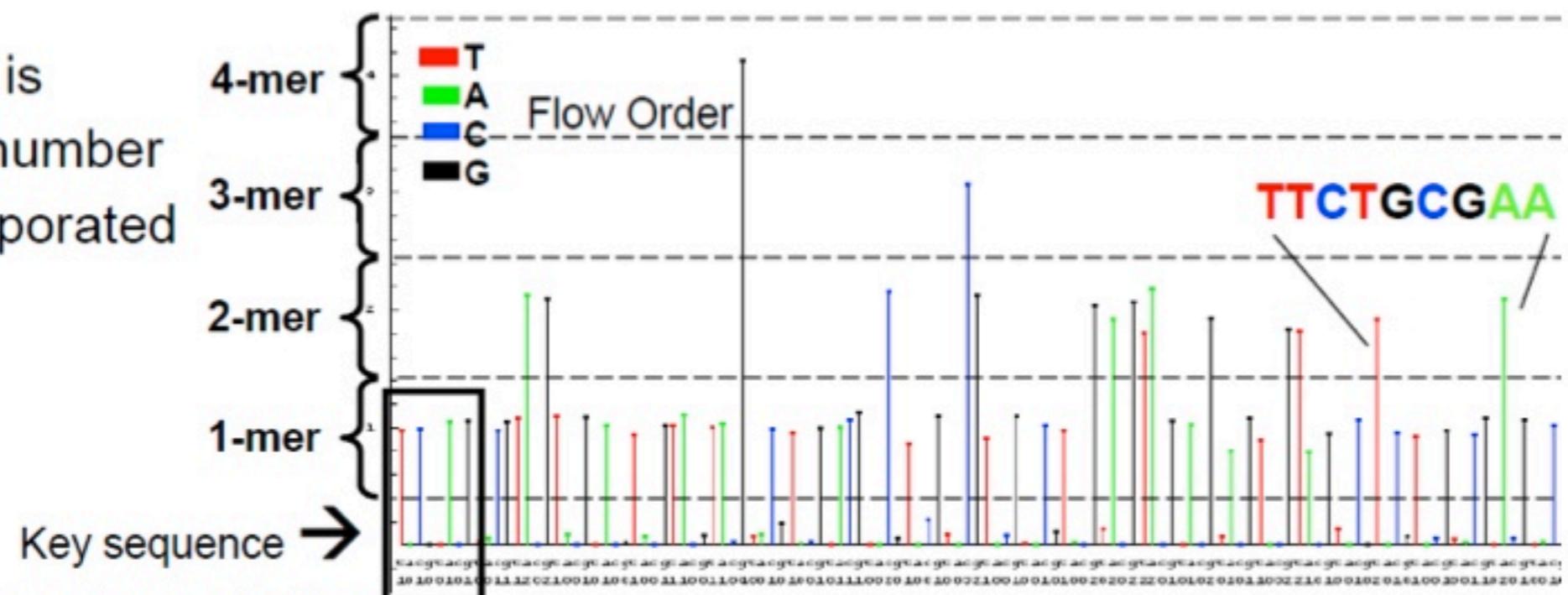
454 Sequencing / Roche



Pyrosequencing

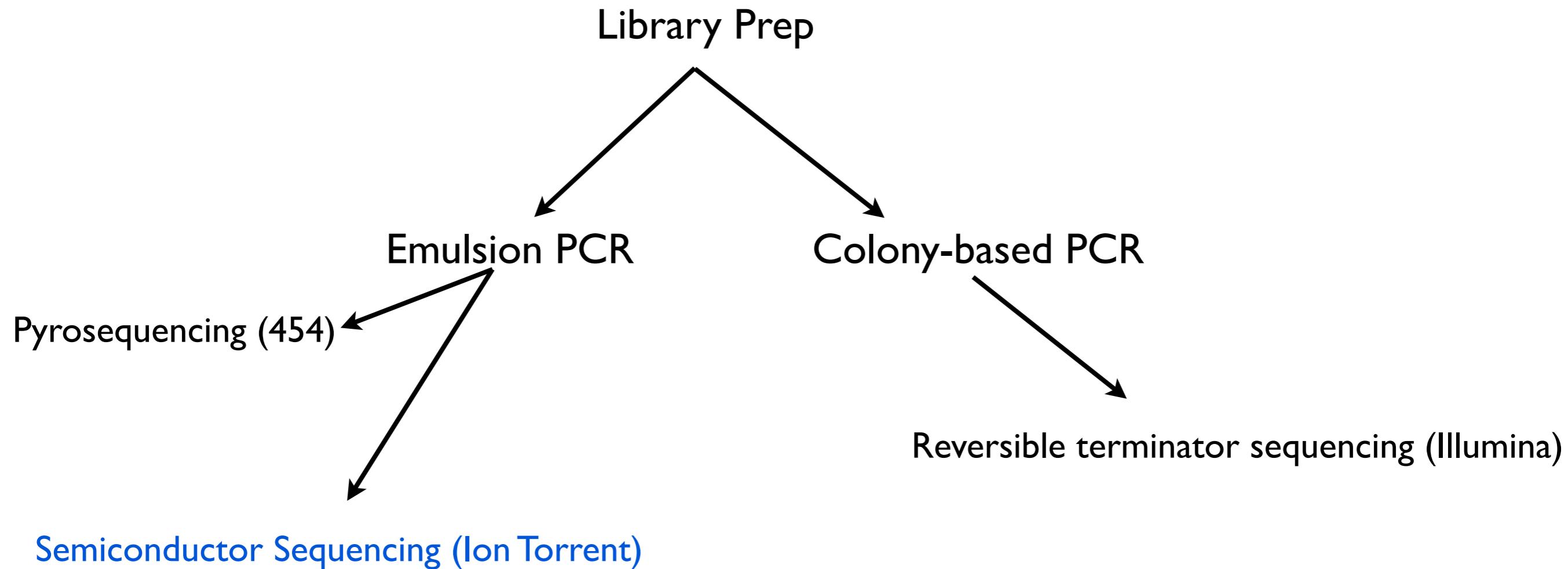


The signal strength is proportional to the number of nucleotides incorporated



TCAG for signal calibration and normalization

# Workflow



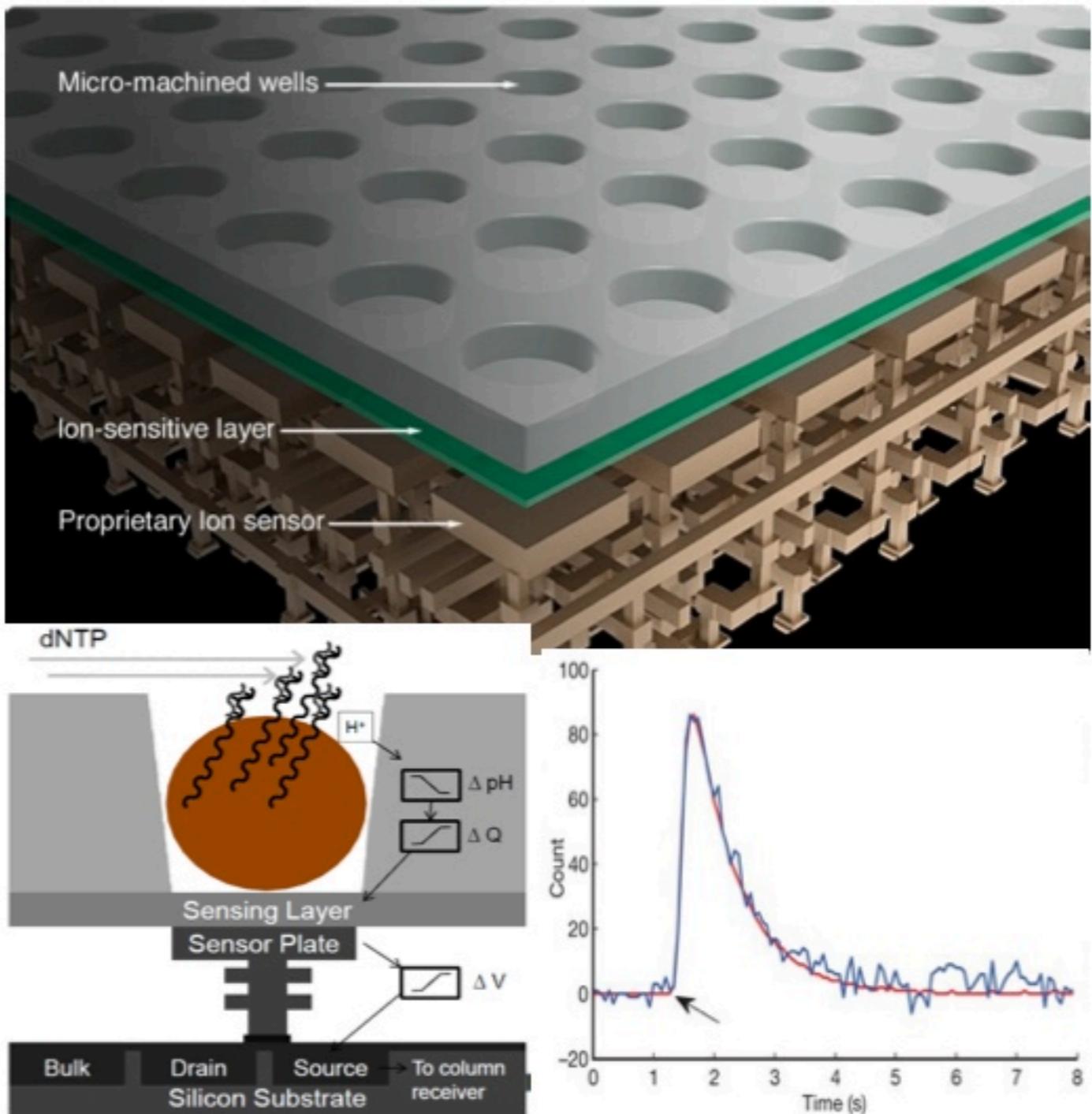
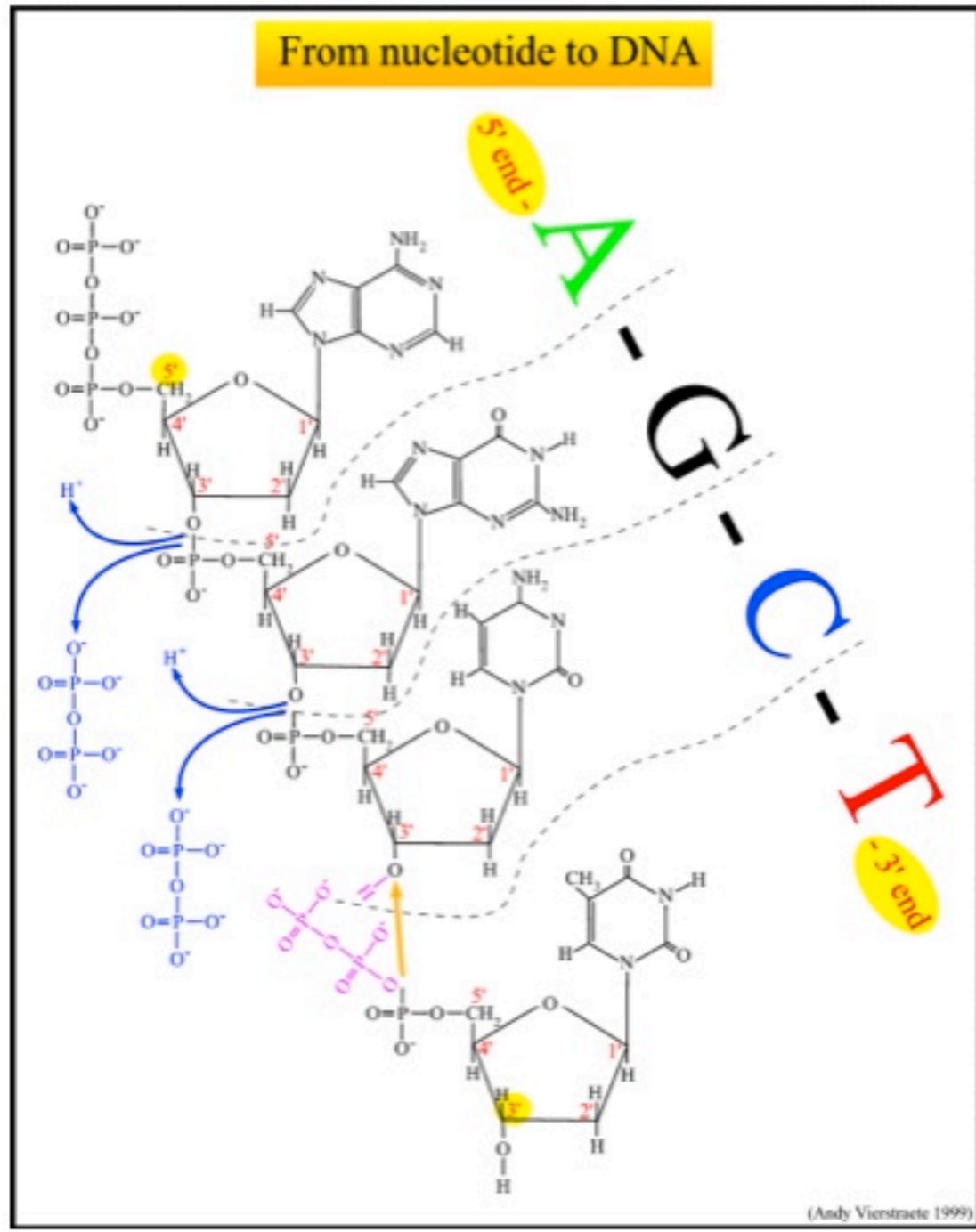
# Workflow: Semiconductor Sequencing

## **Ion Proton™ System performance specifications\* with Ion Proton™ I Chip at commercial launch**

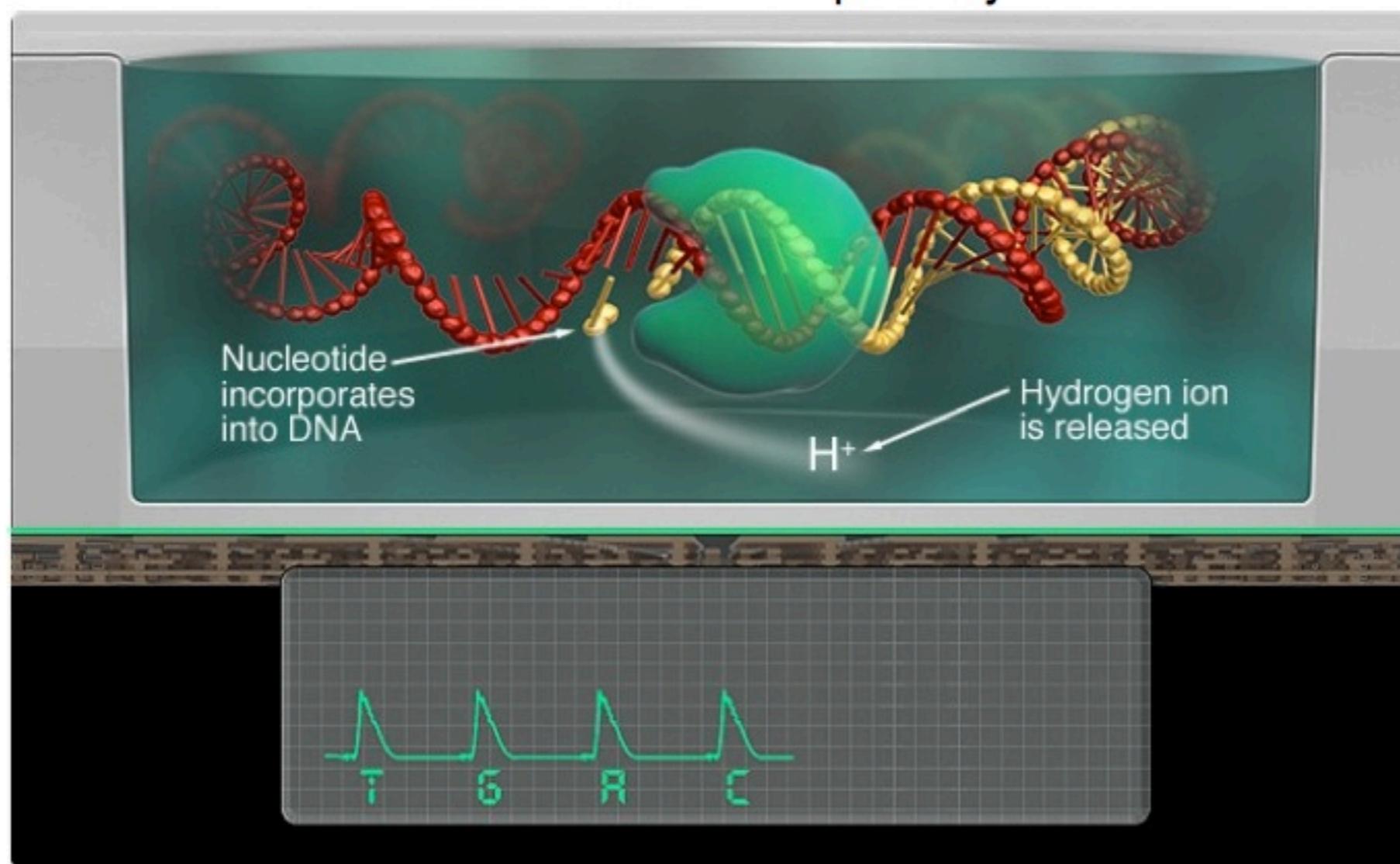
|                             |                                                                                                                                                                                                                                                         |                                                                                                                                        |  |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|--|
| Throughput                  | Up to 10 Gb<br><small>(Note: The Ion Proton™ II Chip will be available about six months after the Ion Proton™ I Chip. The Ion Proton™ II Chip will enable sample-to-variant analysis of a human genome in a single day, at up to 20x coverage.)</small> |                                                                                                                                        |  |
| Read length                 | Up to 200-base fragment reads                                                                                                                                                                                                                           |                                                                                                                                        |  |
| Sequencing run time         | As little as 2 hours for 100-base reads                                                                                                                                                                                                                 |                                                                                                                                        |  |
| Key applications            | Human-scale genome sequencing<br>Agricultural genome sequencing<br>Microbial genome sequencing<br>Exome sequencing<br>Sequencing sets of genes<br>Metagenomic sequencing                                                                                | Methylation sequencing<br>ChIP-Seq<br>Whole-transcriptome sequencing<br>Gene-expression<br>Small-RNA sequencing<br>MicroRNA sequencing |  |
| Available library solutions | Ion AmpliSeq™ Library Kit<br>Ion TargetSeq™ Exome Kit                                                                                                                                                                                                   | Ion Xpress™ Plus Fragment Library Kit<br>Ion Total RNA-Seq Kit                                                                         |  |
| Barcoding                   | 384 barcodes supported by Torrent Suite<br>96 off-the-shelf barcodes for DNA<br>16 off-the-shelf barcodes for RNA                                                                                                                                       |                                                                                                                                        |  |



# Workflow: Semiconductor Sequencing

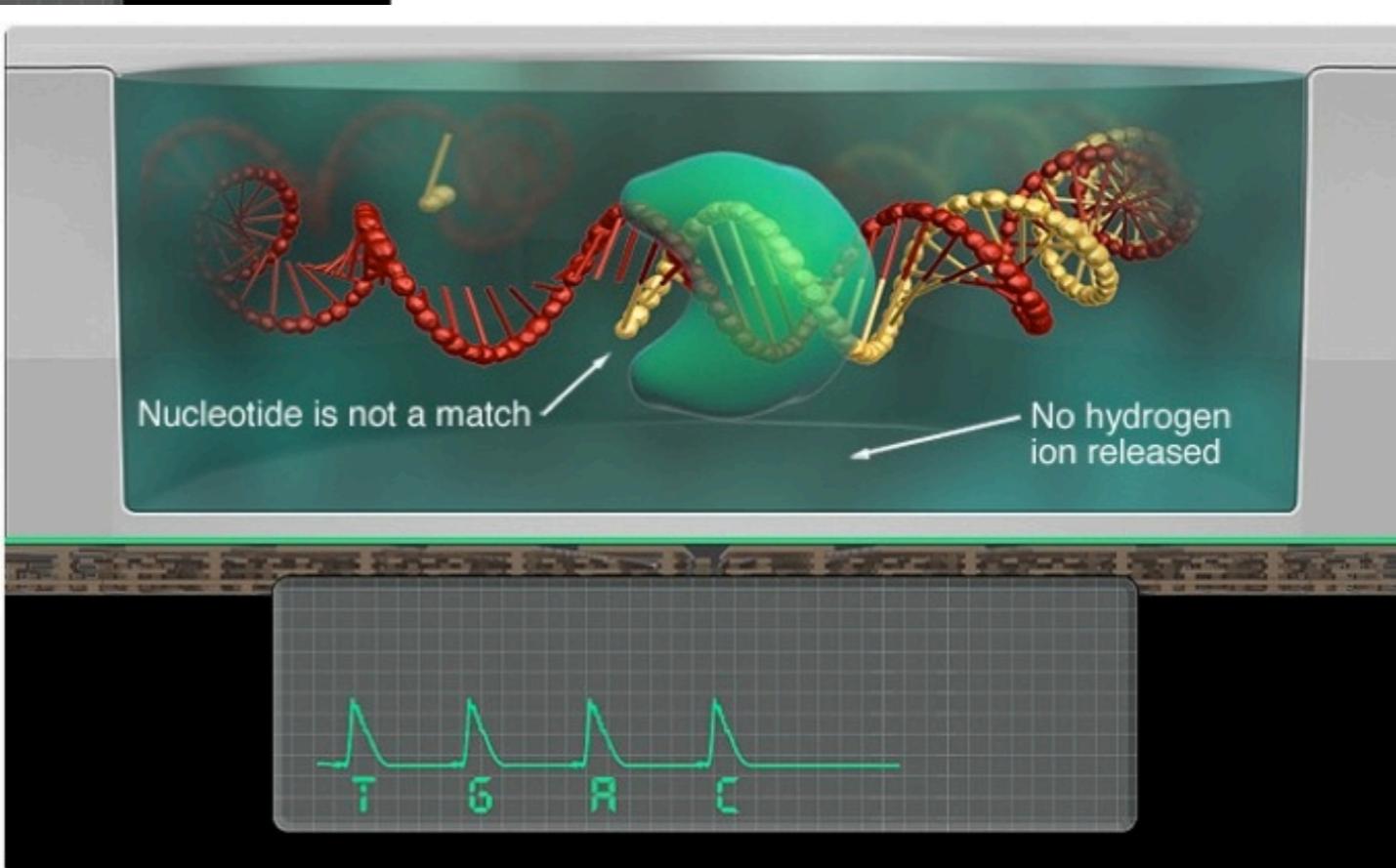
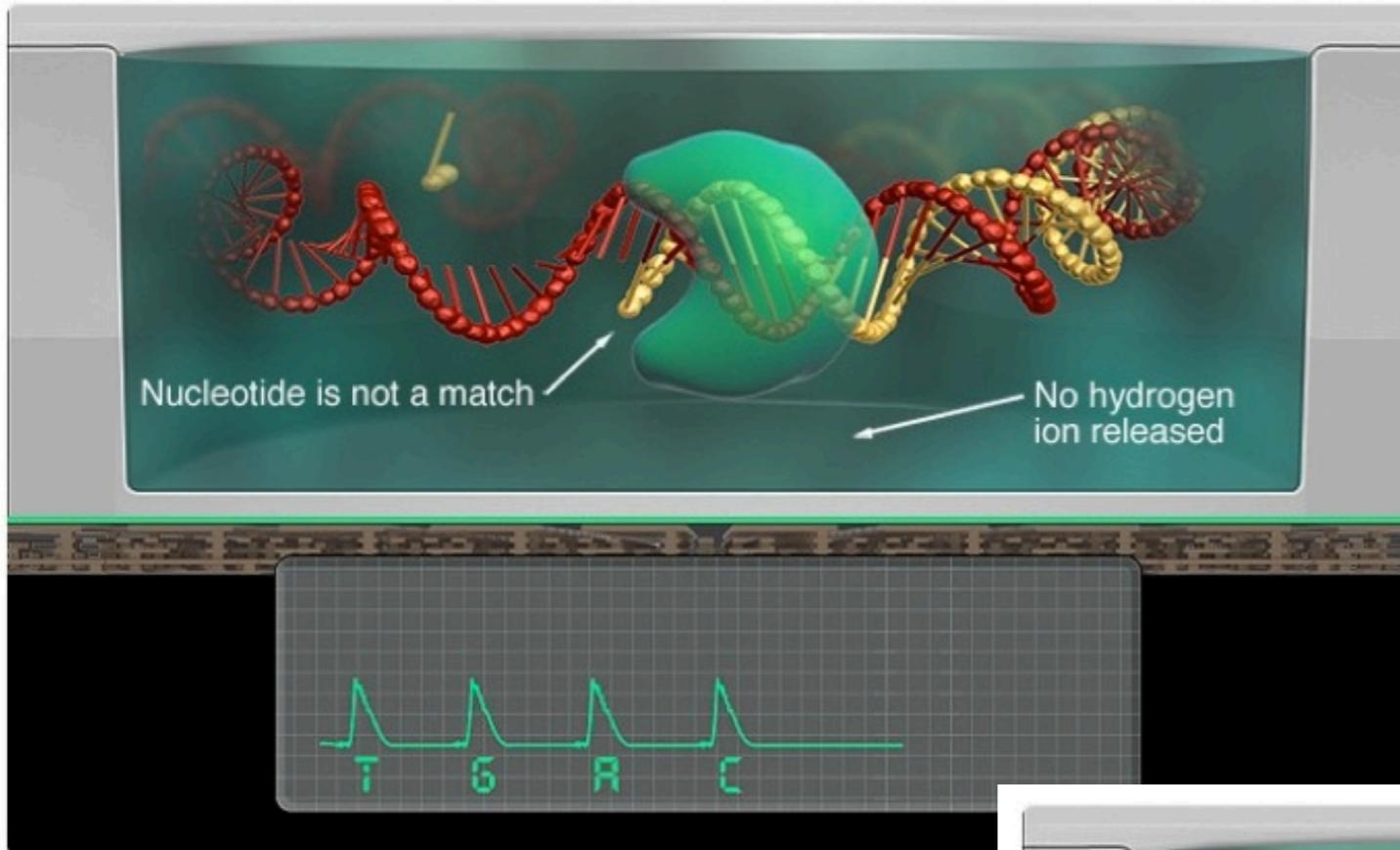


# Workflow: Semiconductor Sequencing

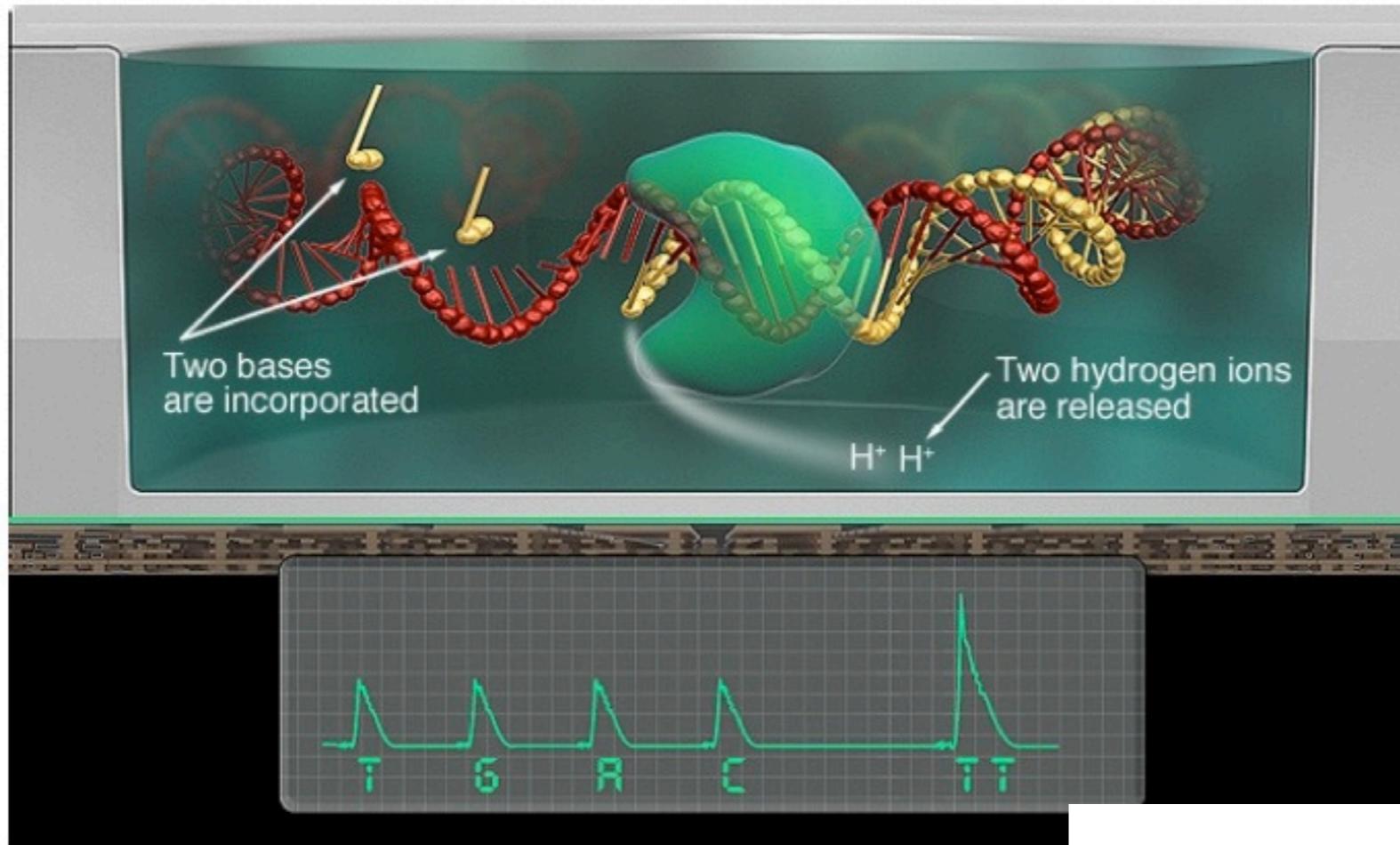


No camera, just a pH sensor

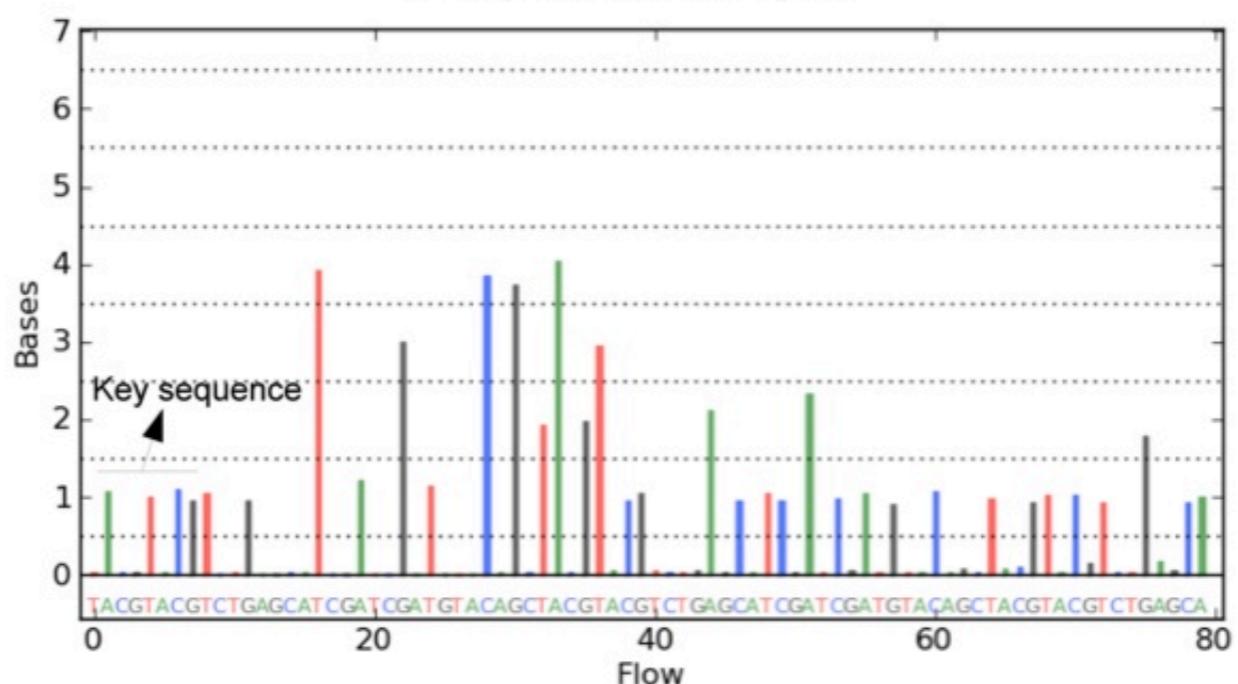
# Workflow: Semiconductor Sequencing



# Workflow: Semiconductor Sequencing

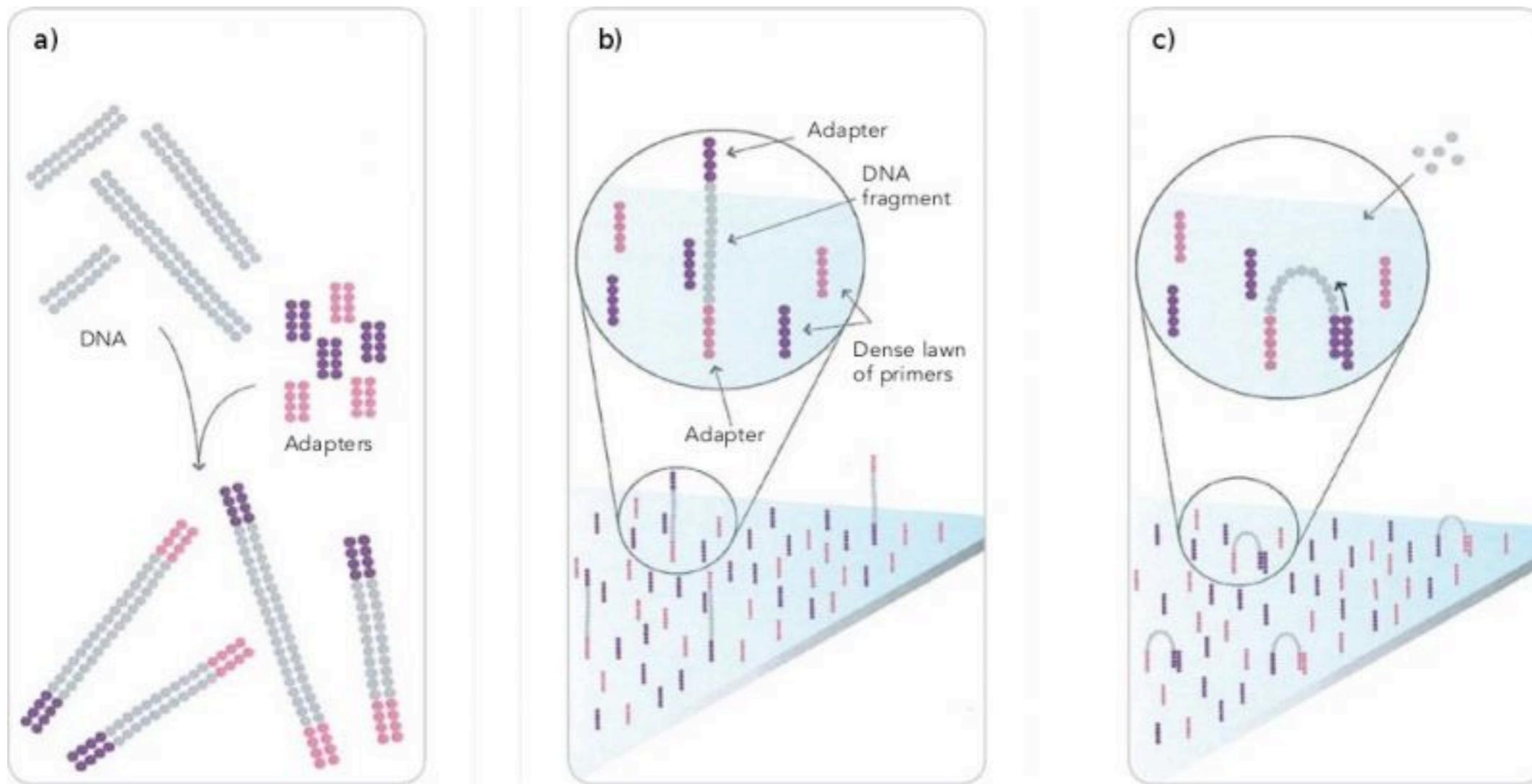


Average Corrected Ionogram



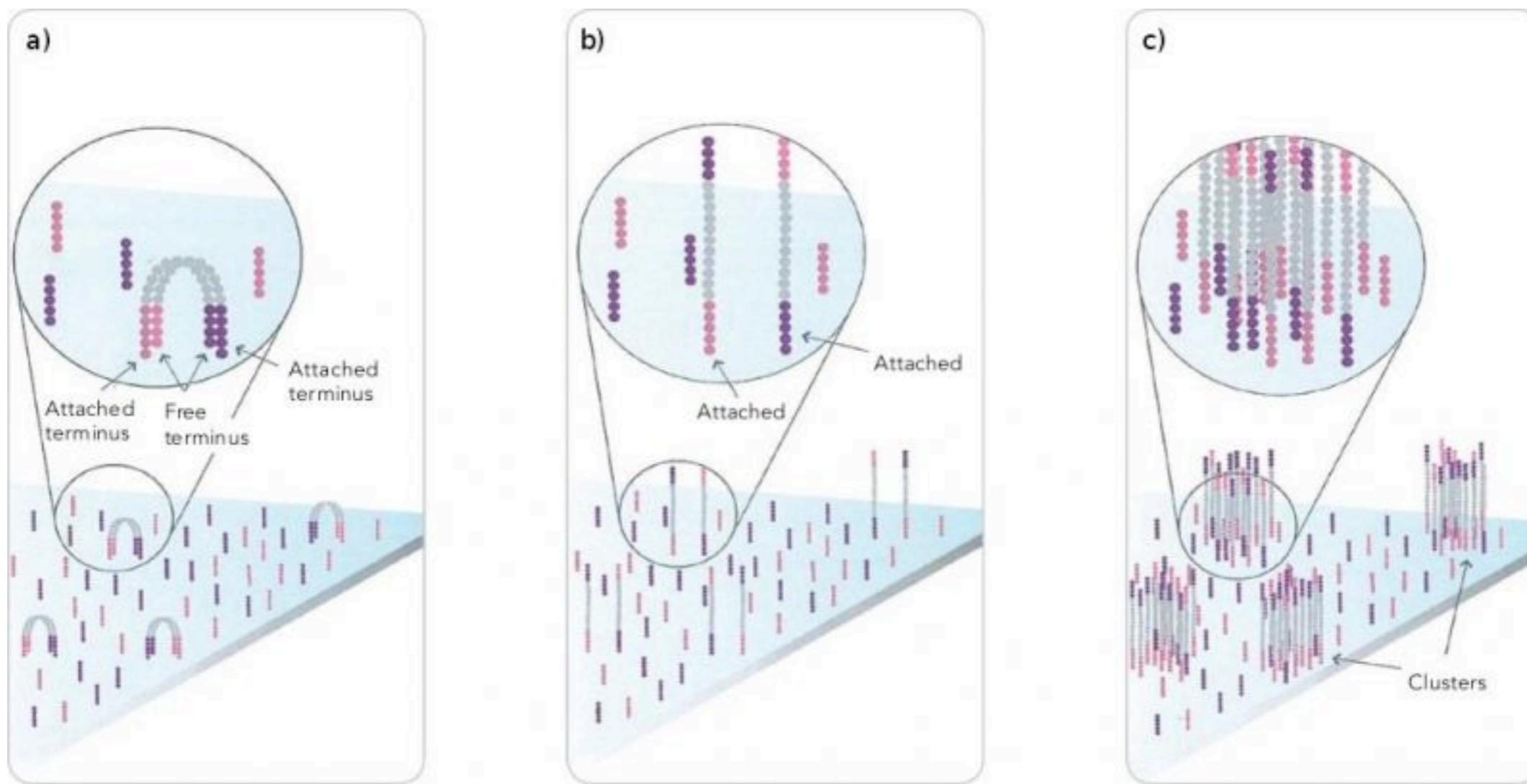
ATCGTGTTTAGGGTCCCCGGGGTT...

# Workflow: Colony-based PCR



- Adaptor sequences are attached to the ends of the fragments.
- The fragments are randomly distributed and attached across the surface.
- Bridge amplification is initiated and the fragments free adaptor end binds to a complementary sequence attached to the plate.

# Workflow: Colony-based PCR



- a) Bridge amplification is complete.
- b) The strands separate and the cycle is repeated.
- c) A cluster of fragments is formed after many iterations.

# Workflow: Reversible Terminator Seq

Illumina

HiSeq 2500 / 2000 / 1500 / 1000



HiScanSQ



Genome Analyzer IIx

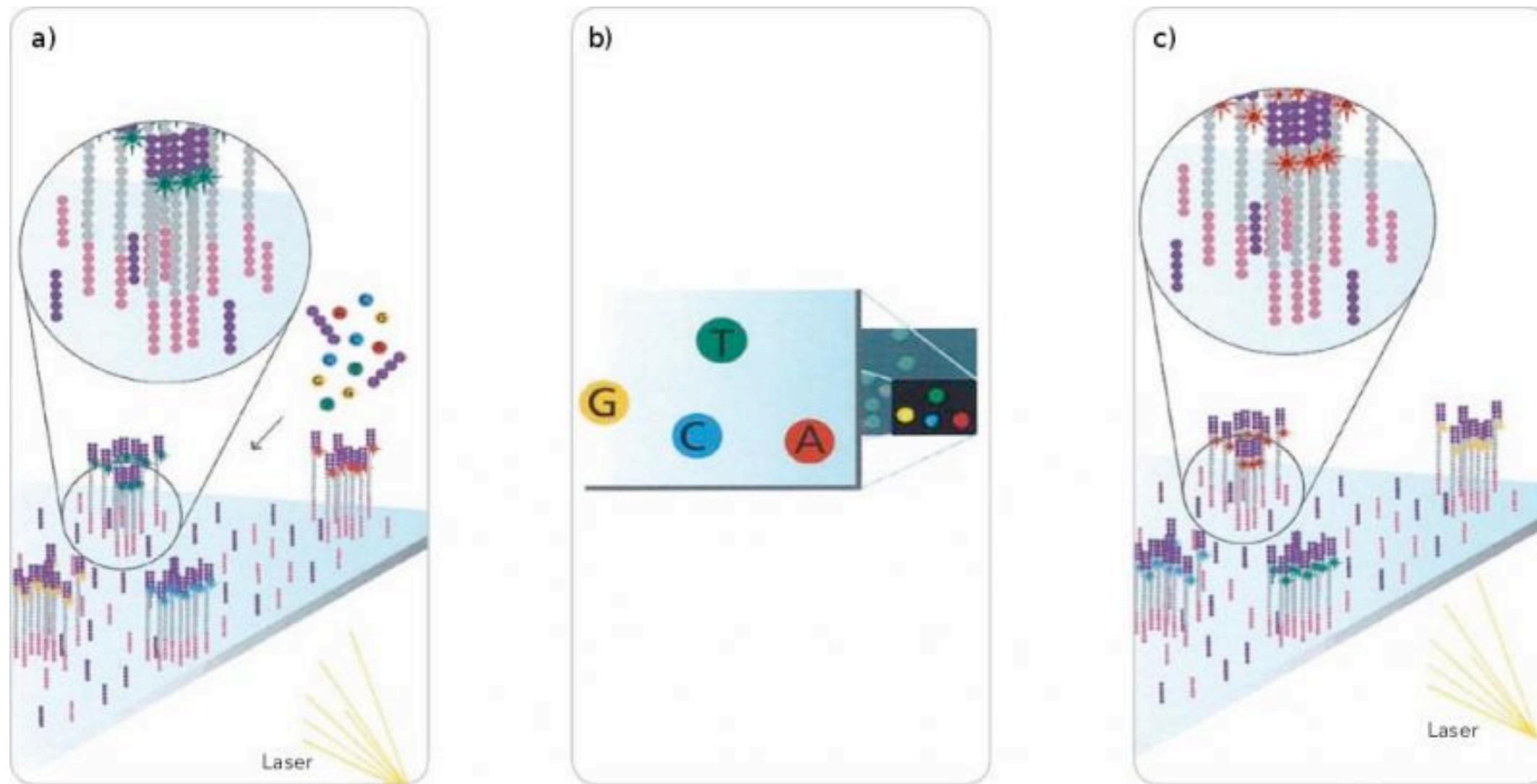


MiSeq



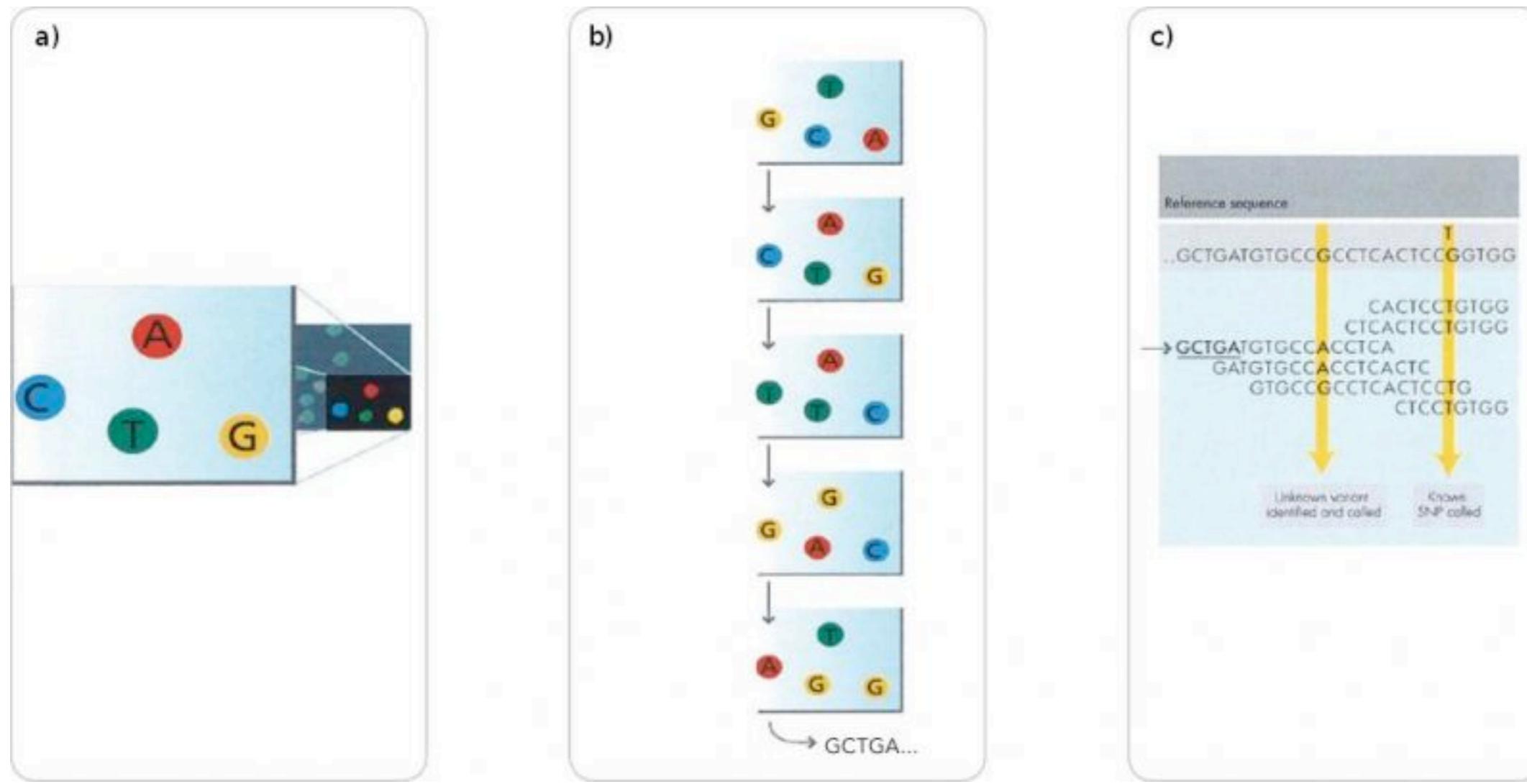
|                      | HiSeq         | HiScanSQ    | Genome Analyzer IIx | MiSeq       |
|----------------------|---------------|-------------|---------------------|-------------|
| <b>Read Length</b>   | 100 bp        | 100 bp      | 150 bp              | 250 bp      |
| <b>Throughput</b>    | 600 Gb        | 150 Gb      | 95 Gb               | 6 Gb        |
| <b>Reads per run</b> | 3,000,000,000 | 750,000,000 | 320,000,000         | 12,000,000  |
| <b>Accuracy</b>      | 99,9 %        | 99,9 %      | 99,9 %              | 99,9 %      |
| <b>Run Time</b>      | 11 days       | 8 days      | 14 days             | 20-35 hours |

# Workflow: Reversible Terminator Seq



- a) Sequencing-by-synthesis begins. A mix of all nucleotides and DNA polymerase is added to the surface. The nucleotides are not ordinary nucleotides but are also reversible terminators marked with removable fluorescent dyes so that when one nucleotide is attached, replication stops.
- b) A laser excites the fluorescent dyes and a camera captures which colors the clusters emit.
- c) The reversible terminators are removed and the cycle is repeated.

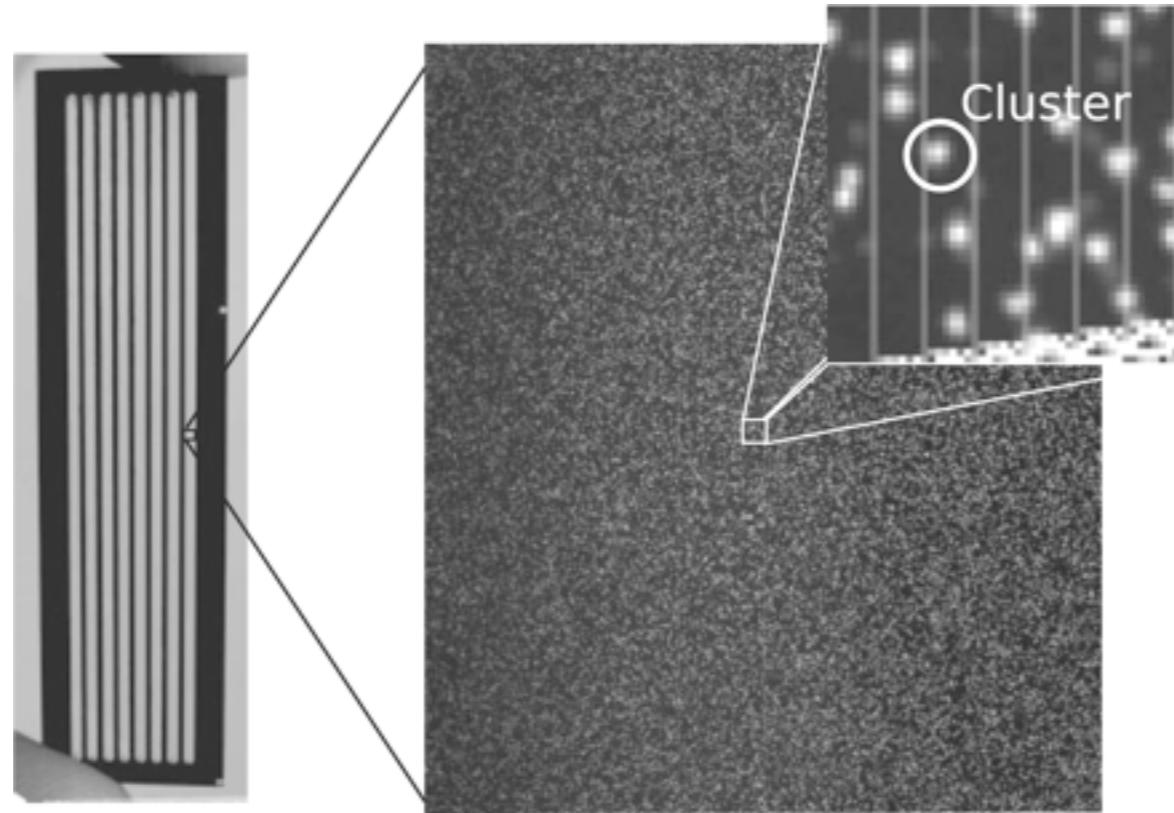
# Workflow: Reversible Terminator Seq



- a) The process is repeated and the second step is also recorded.
- b) The remaining steps are also recorded and a sequence can be read.
- c) The individual fragments are aligned to a reference genome, and differences can be detected.

# Workflow: Reversible Terminator Seq

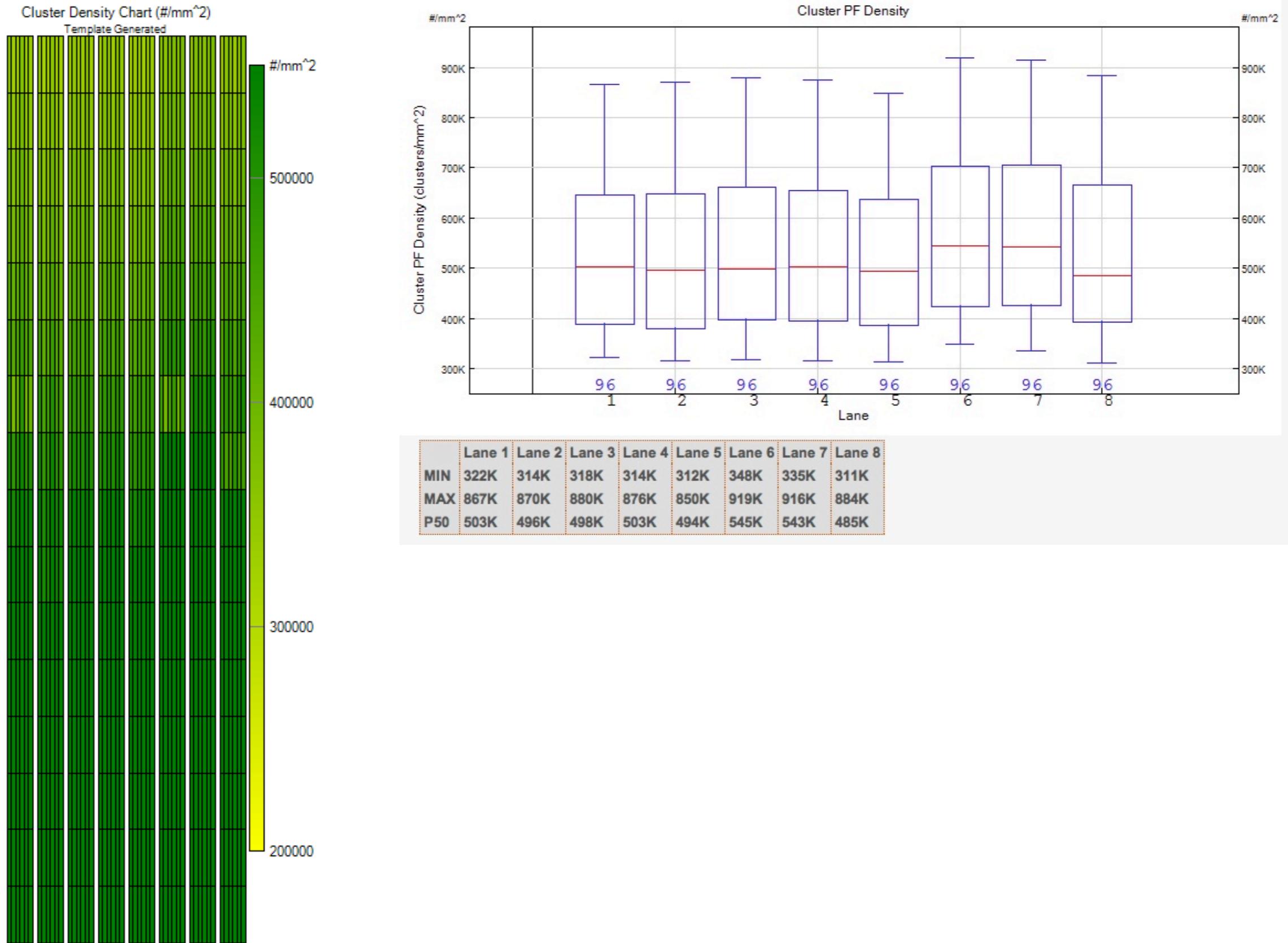
Each cycle results in a number of images, one for each color multiplied by the number required to cover the whole surface.



The images are then analyzed in two steps. First the clusters and their intensities are identified and written to intensity files, and then the bases are determined from the intensities.

Each base is assigned quality value ranging from 2 to 40, depending on the certainty in the basecalling.

# Workflow: Reversible Terminator Seq



# A Quick Comparison

- 454 has longer but fewer reads
- 454 is faster but more expensive
- 454 still has trouble resolving long homopolymeric repeats
- Ion Torrent might be the next big thing, but it will take time and investment for labs to move over to the new platform
- HiSeq currently dominates the deep sequencing market; the fast-but-less-deep market still up for grabs.

"Evaluation of base-call error, frameshift frequency, and contig length suggested that Illumina offered equivalent, if not better, assemblies than Roche 454."

Luo C, Tsementzi D, Kyripides N, Read T, Konstantinidis KT (2012) Direct Comparisons of Illumina vs. Roche 454 Sequencing Technologies on the Same Microbial Community DNA Sample. PLoS ONE 7(2): e30087. doi: 10.1371/journal.pone.0030087

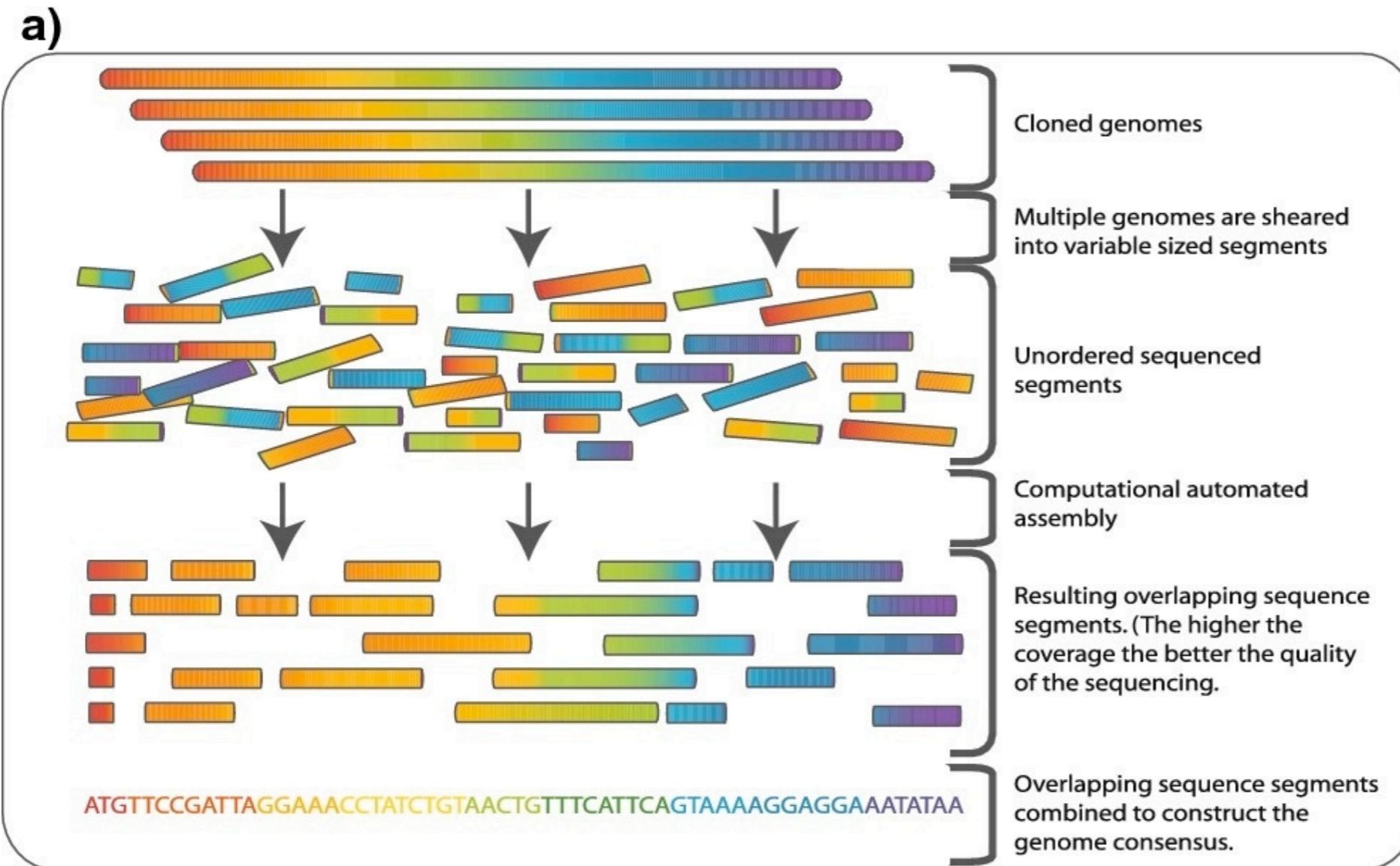
# Applications

Two primary classes: DNA or RNA

DNA:

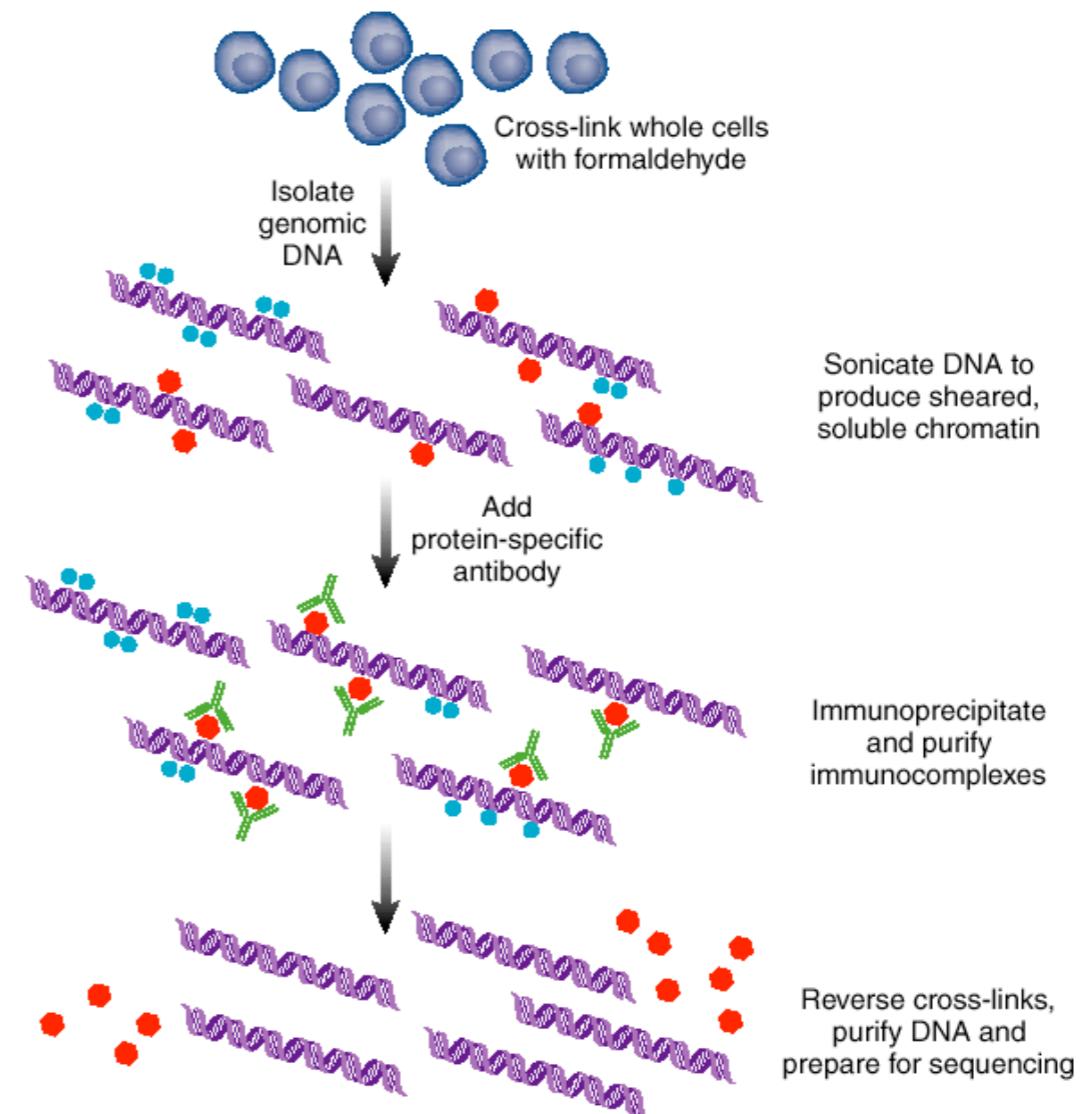
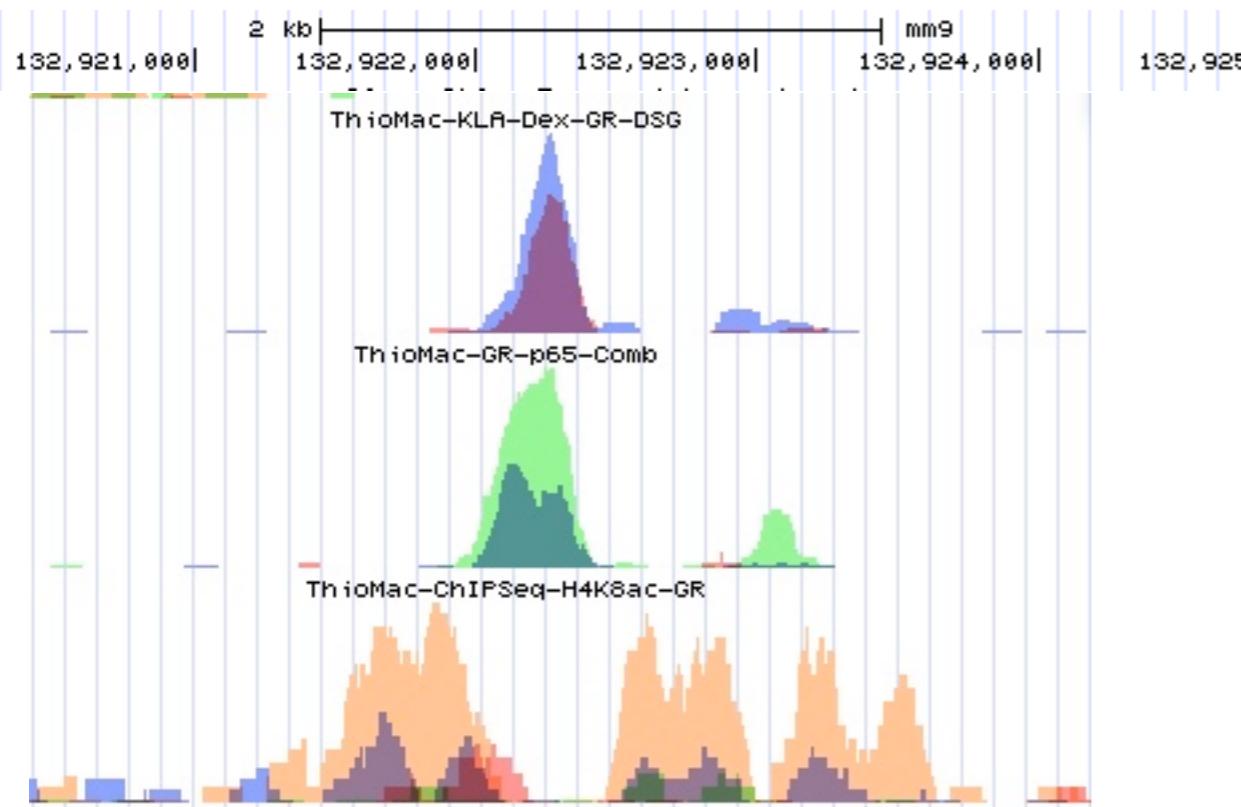
- Shotgun sequencing (whole genome, gDNA-seq)
- Exome sequencing
- Targeted mutation sequencing
- ChIP-seq
- ChIA-PET
- Methyl-seq (bisulfite, MeDIP)
- DNase-seq (DHS sequencing)
- MNase-seq
- Hi-C

# Whole-Genome Shotgun Sequencing



- De novo, re-sequencing, mutation discovery, metagenomics
- Big algorithmic need!
- <http://www.1000genomes.org>, <http://genome10k.soe.ucsc.edu>

# ChIP-seq



# ChIP-seq

Homer de novo Motif Results +

## Homer *de novo* Motif Results

[Known Motif Enrichment Results](#)

[Gene Ontology Enrichment Results](#)

If Homer is having trouble matching a motif to a known motif, try copy/pasting the matrix file into [STAMP](#)

More information on motif finding results: [HOMER](#) | [Description of Results](#) | [Tips](#)

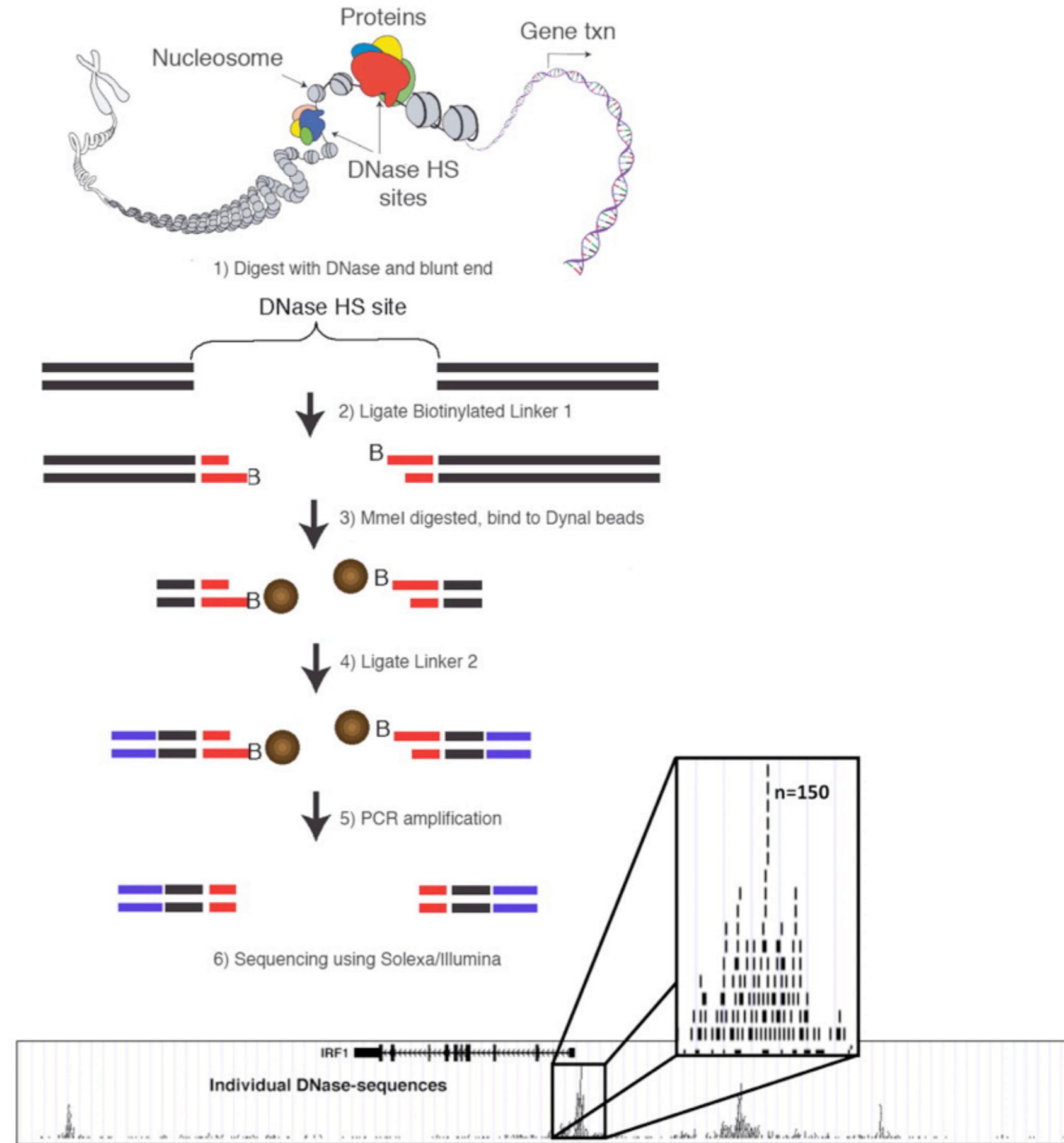
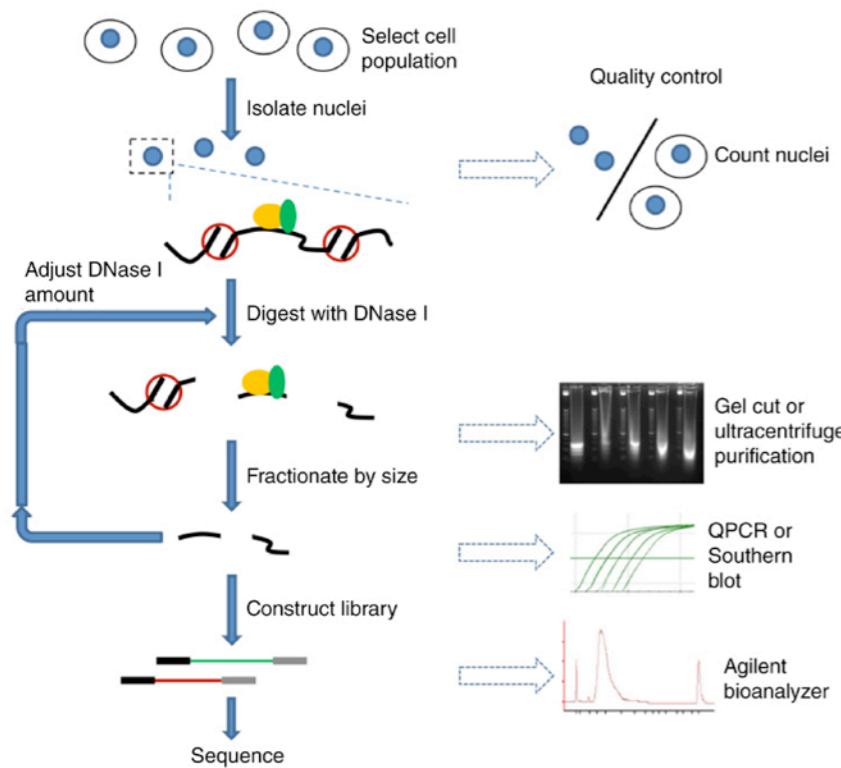
Total target sequences = 37301

Total background sequences = 35962

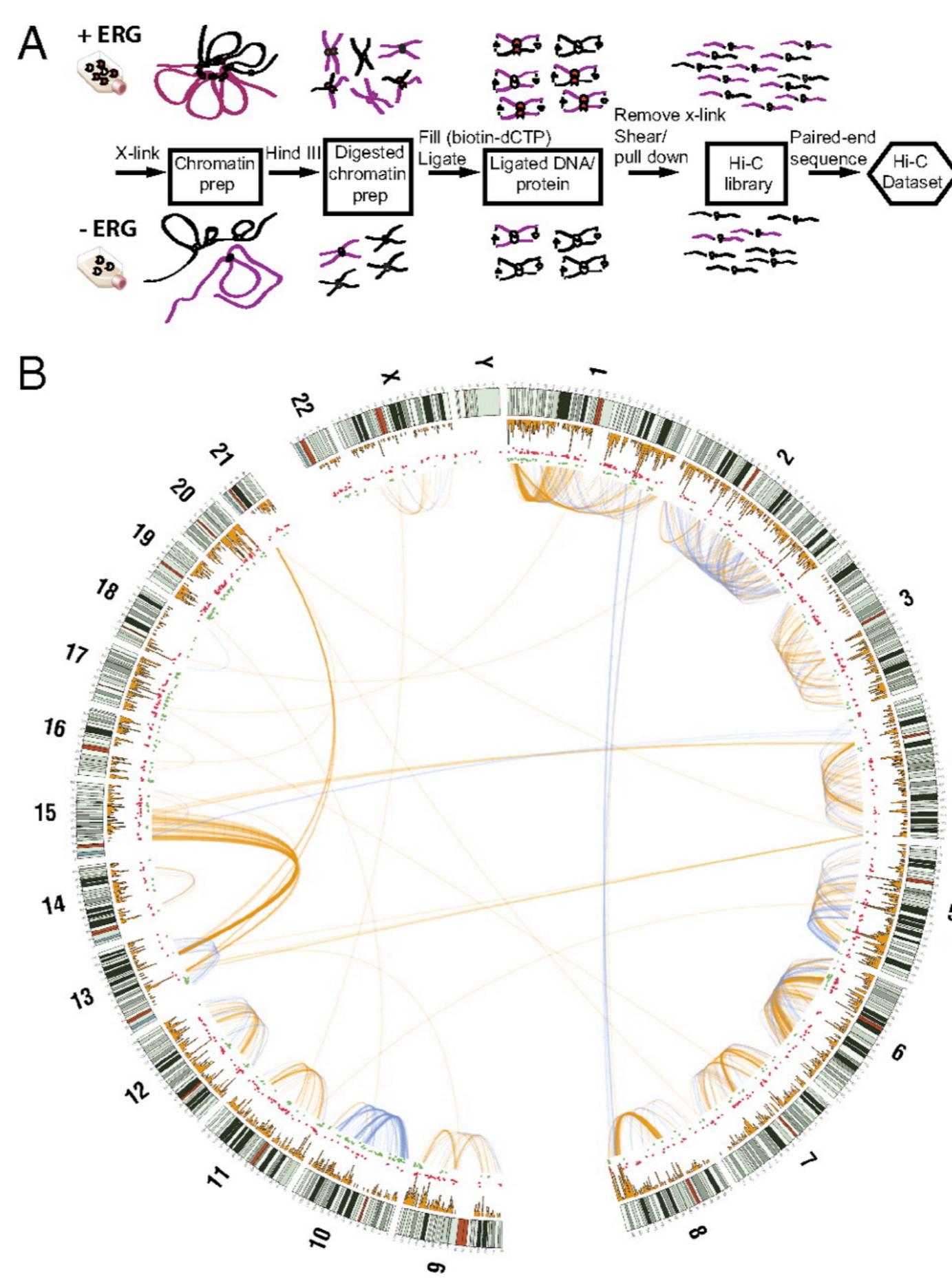
\* - possible false positive

| Rank | Motif                                                                                | P-value  | log P-value | % of Targets | % of Background | STD(Bg STD)        | Best Match/Details                                                                                                     | Motif File                          |
|------|--------------------------------------------------------------------------------------|----------|-------------|--------------|-----------------|--------------------|------------------------------------------------------------------------------------------------------------------------|-------------------------------------|
| 1    |    | 1e-12661 | -2.915e+04  | 70.91%       | 15.19%          | 40.5bp<br>(65.1bp) | Foxa2(Forkhead)/Liver-Foxa2-ChIP-Seq/Homer<br><a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>  | <a href="#">motif file (matrix)</a> |
| 2    |    | 1e-578   | -1.332e+03  | 27.14%       | 16.52%          | 54.0bp<br>(65.5bp) | NF1-halvesite(CTF)/LNCaP-NF1-ChIP-Seq/Homer<br><a href="#">More Information</a>   <a href="#">Similar Motifs Found</a> | <a href="#">motif file (matrix)</a> |
| 3    |  | 1e-384   | -8.860e+02  | 17.77%       | 10.53%          | 53.9bp<br>(62.1bp) | Unknown/Homeobox/Limb-p300-ChIP-Seq/Homer<br><a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>   | <a href="#">motif file (matrix)</a> |
| 4    |  | 1e-164   | -3.783e+02  | 3.17%        | 1.28%           | 52.2bp<br>(62.9bp) | PH0048.1_Hoxa13<br><a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>                             | <a href="#">motif file (matrix)</a> |
| 5    |   | 1e-151   | -3.485e+02  | 3.38%        | 1.47%           | 50.2bp<br>(65.4bp) | NF-E2(bZIP)/K562-NFE2-ChIP-Seq/Homer<br><a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>        | <a href="#">motif file (matrix)</a> |
| 6    |  | 1e-107   | -2.485e+02  | 1.21%        | 0.35%           | 56.3bp<br>(69.7bp) | CTCF(Zf)/CD4+-CTCF-ChIP-Seq/Homer<br><a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>           | <a href="#">motif file (matrix)</a> |
| 7    |  | 1e-72    | -1.671e+02  | 2.10%        | 1.02%           | 55.1bp<br>(58.5bp) | MA0029.1_Evi1<br><a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>                               | <a href="#">motif file (matrix)</a> |

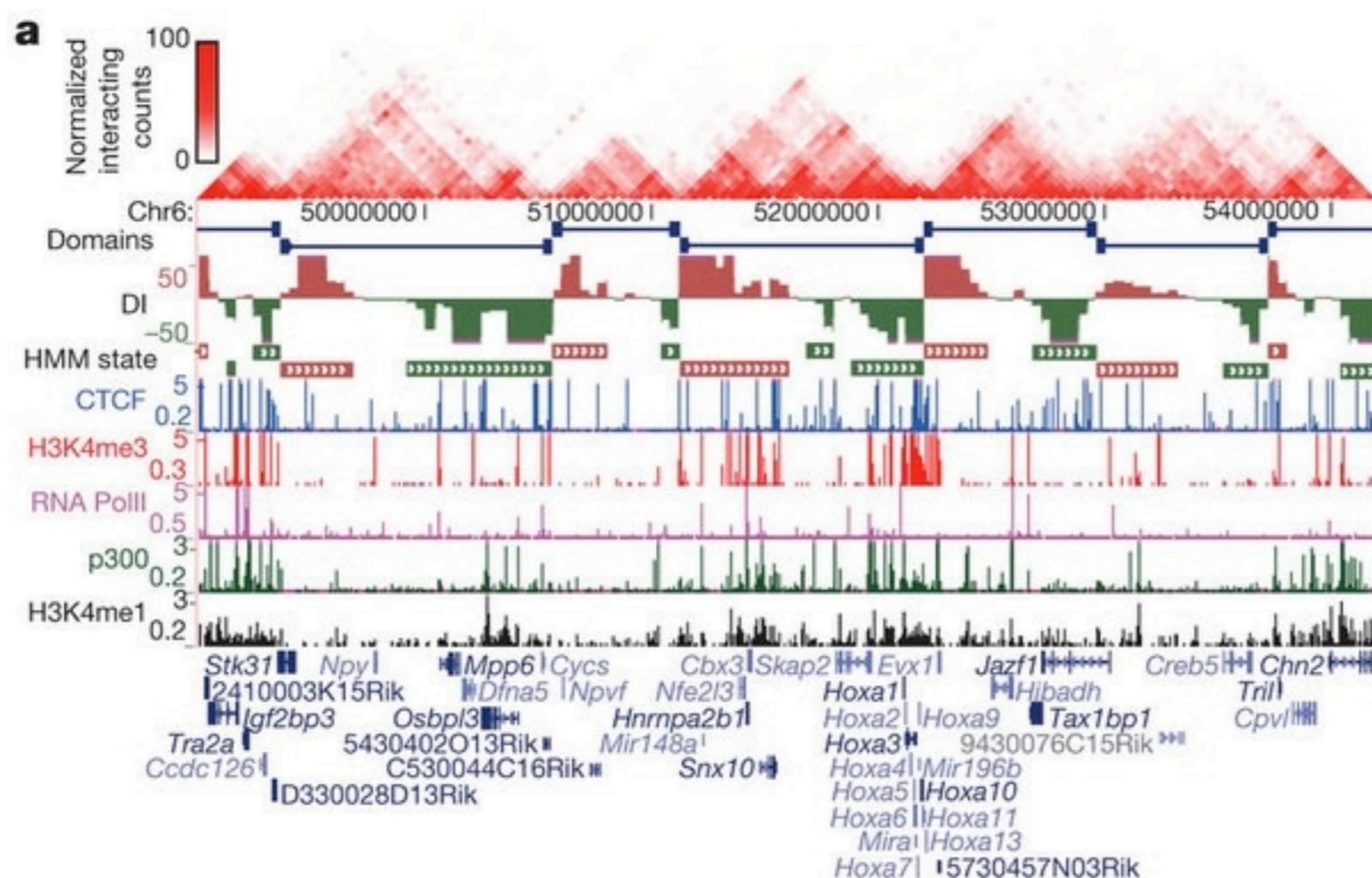
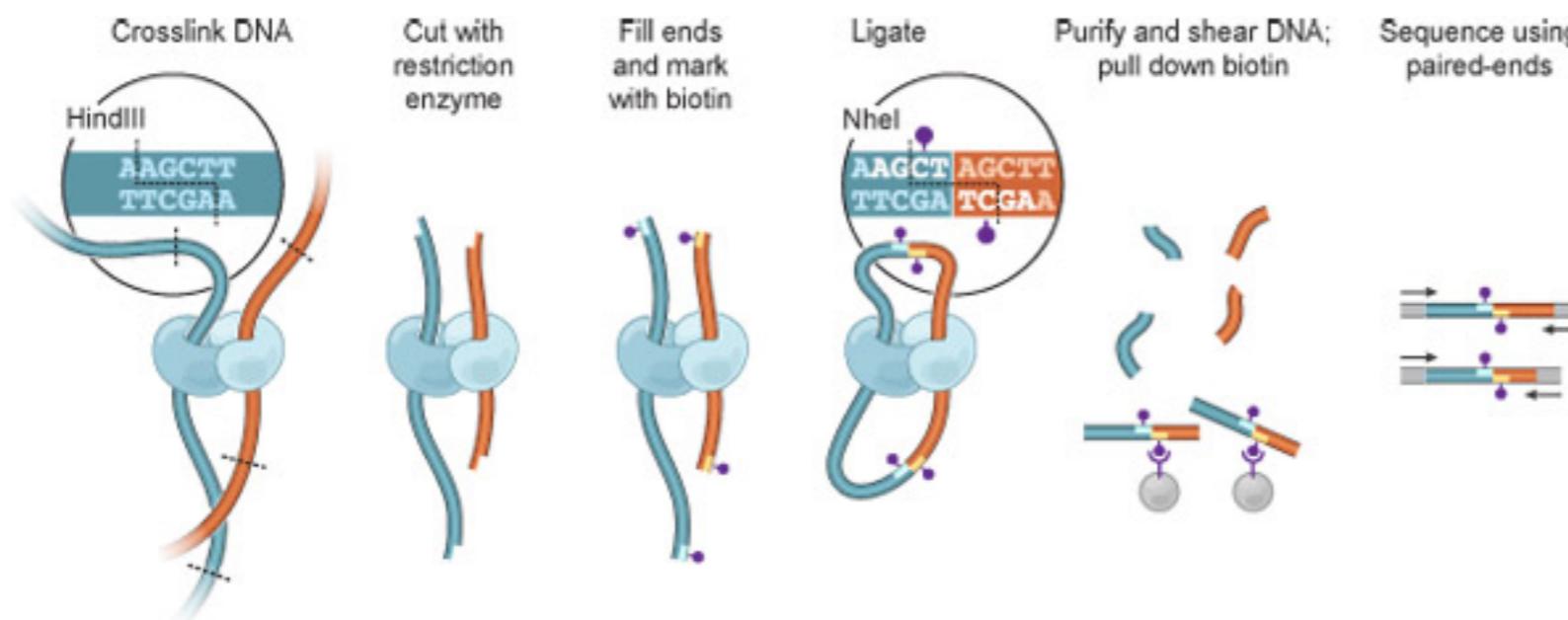
# DNAse-seq



# Hi-C



# Hi-C



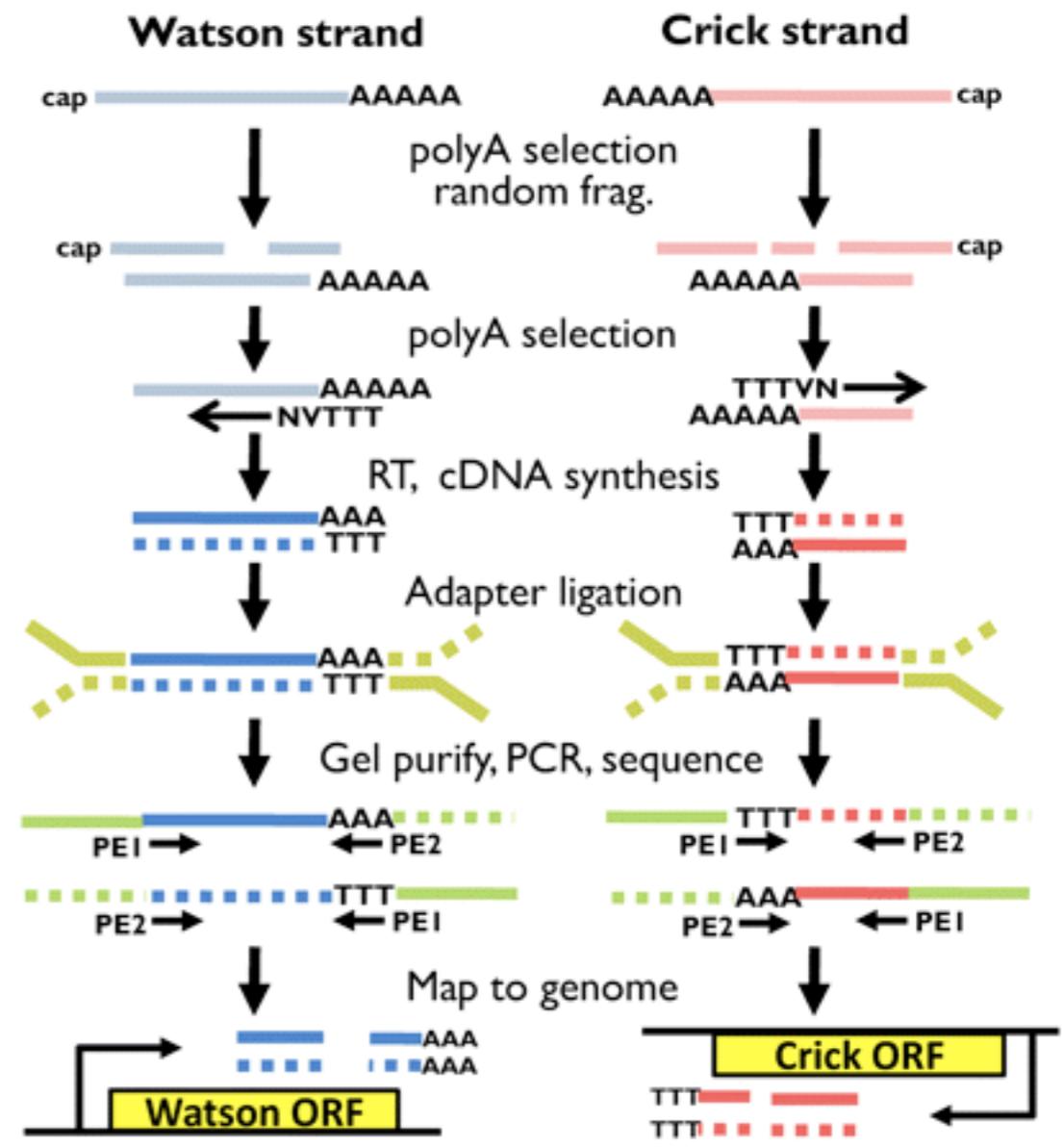
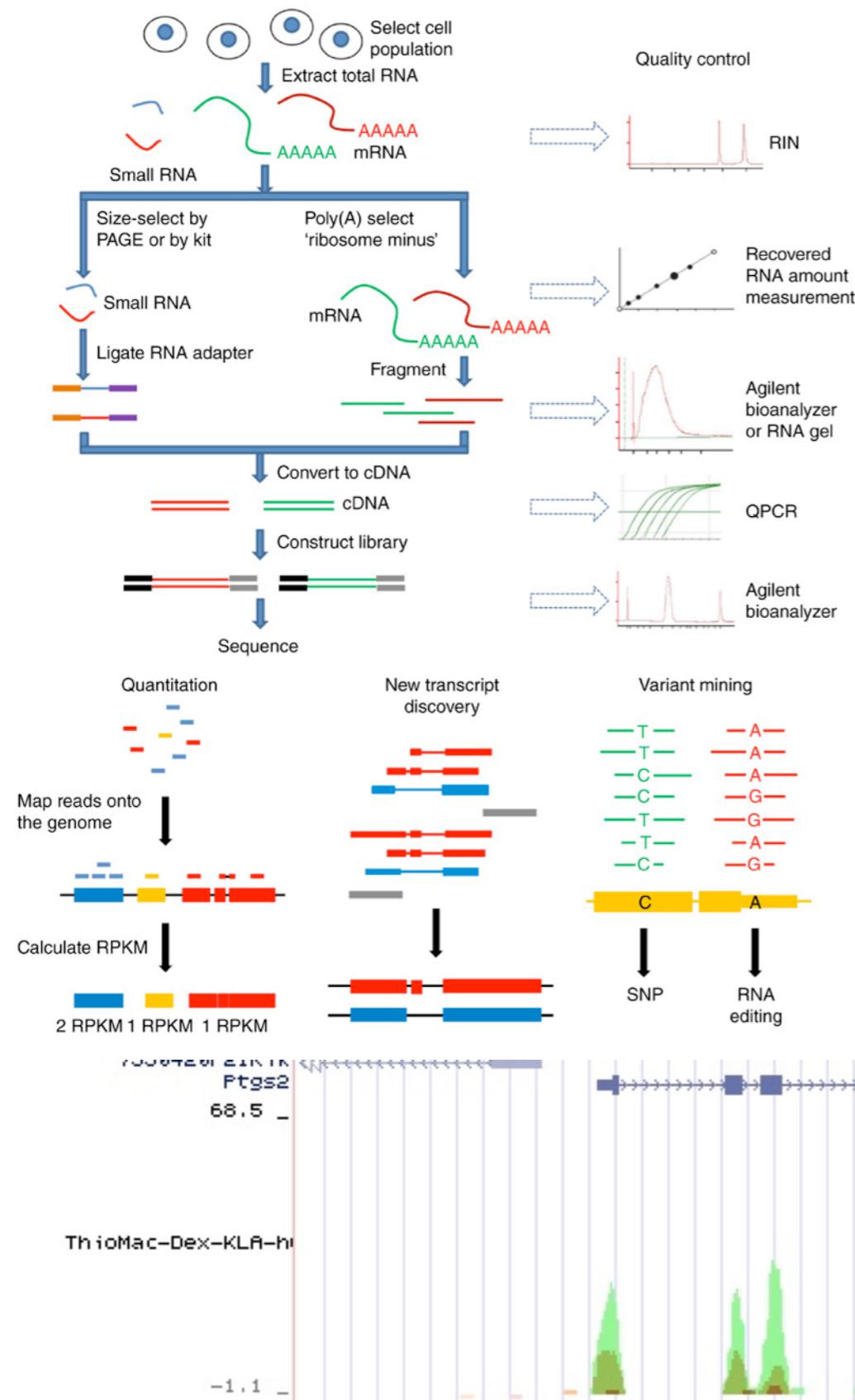
# Applications

Two primary classes: DNA or **RNA**

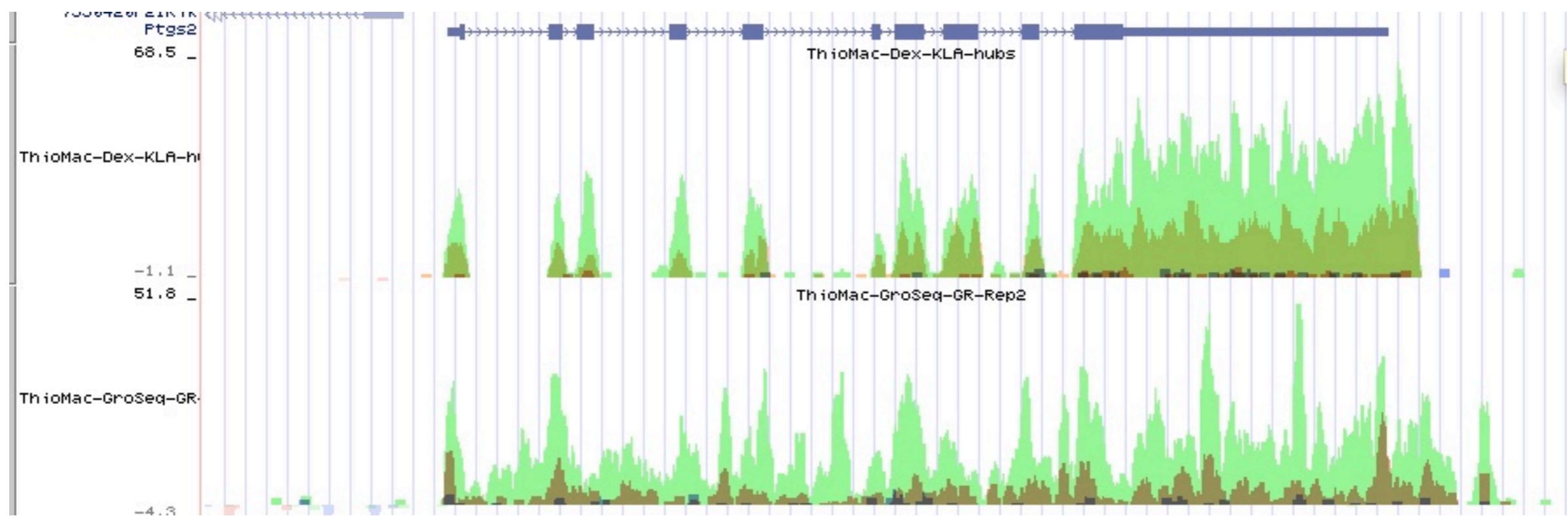
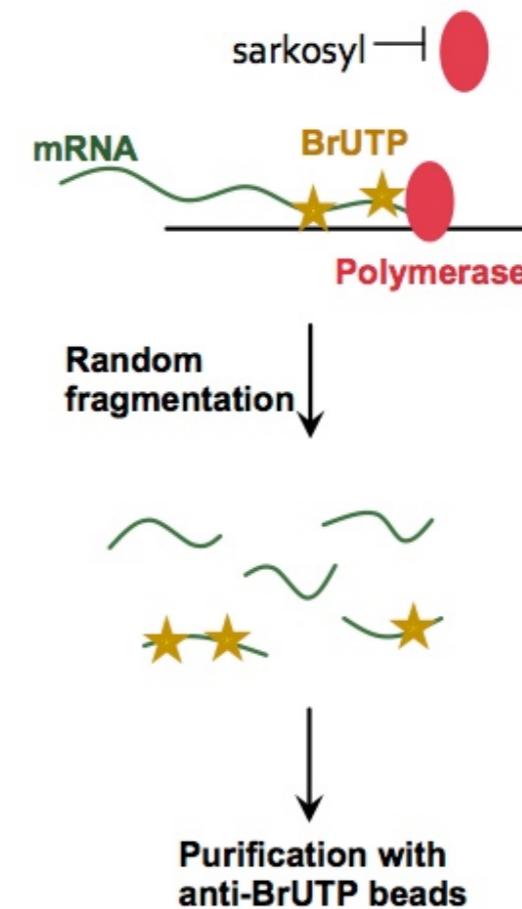
RNA:

- RNA-seq
- Ribo-seq
- GRO-seq
- miRNA-seq
- NET-seq
- CLIP-seq
- TSSa-seq
- CAGE

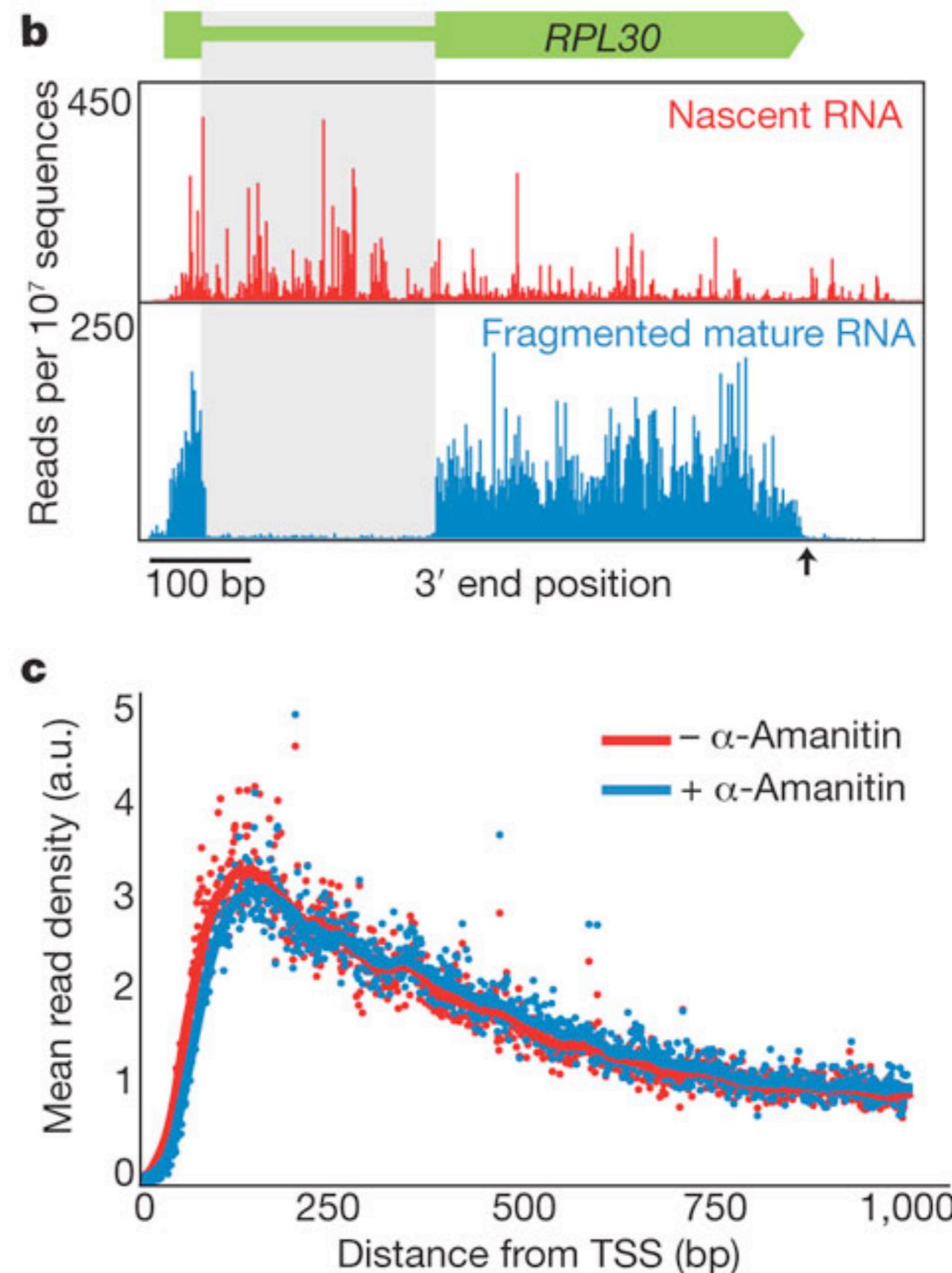
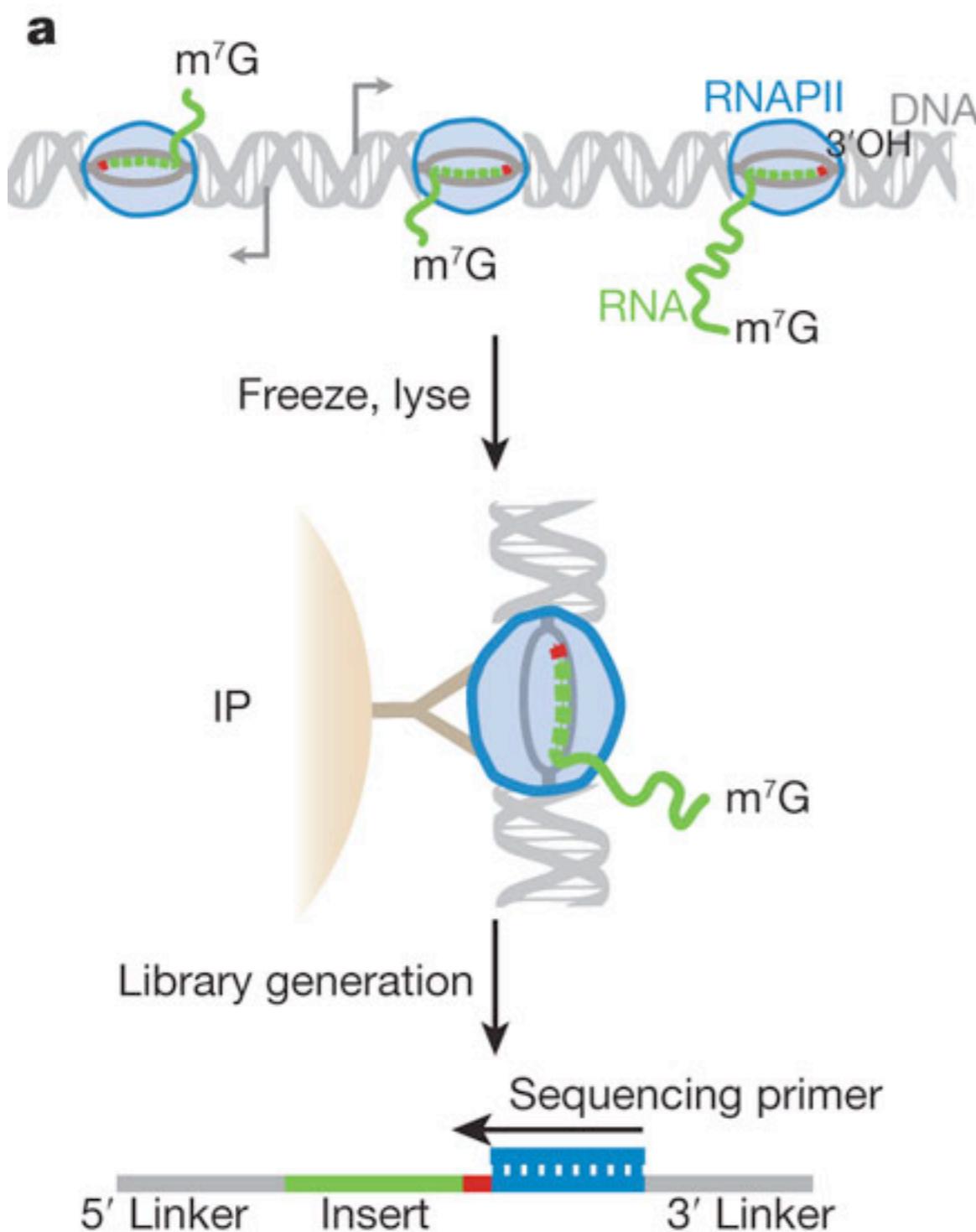
# RNA-seq



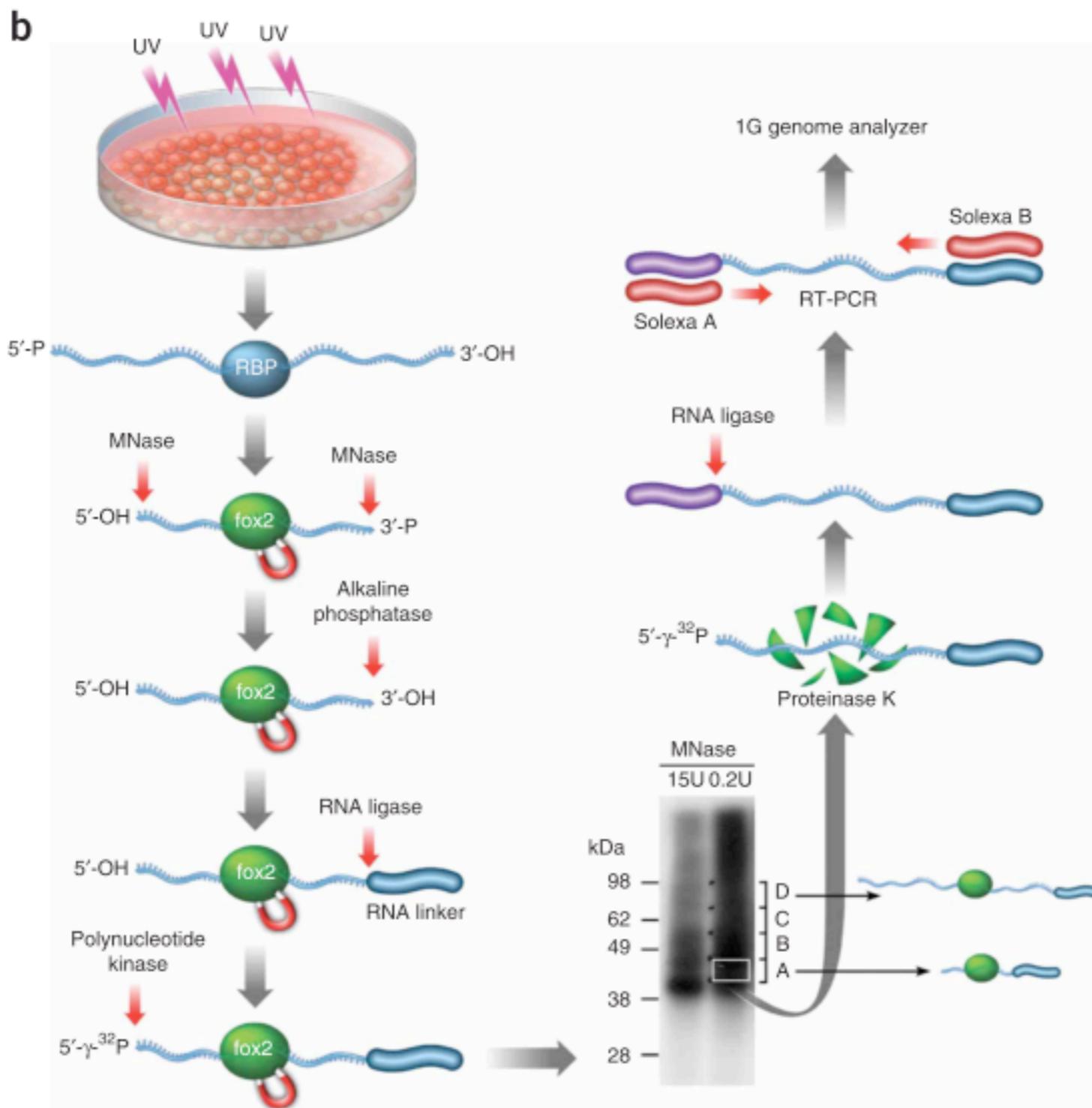
# GRO-seq



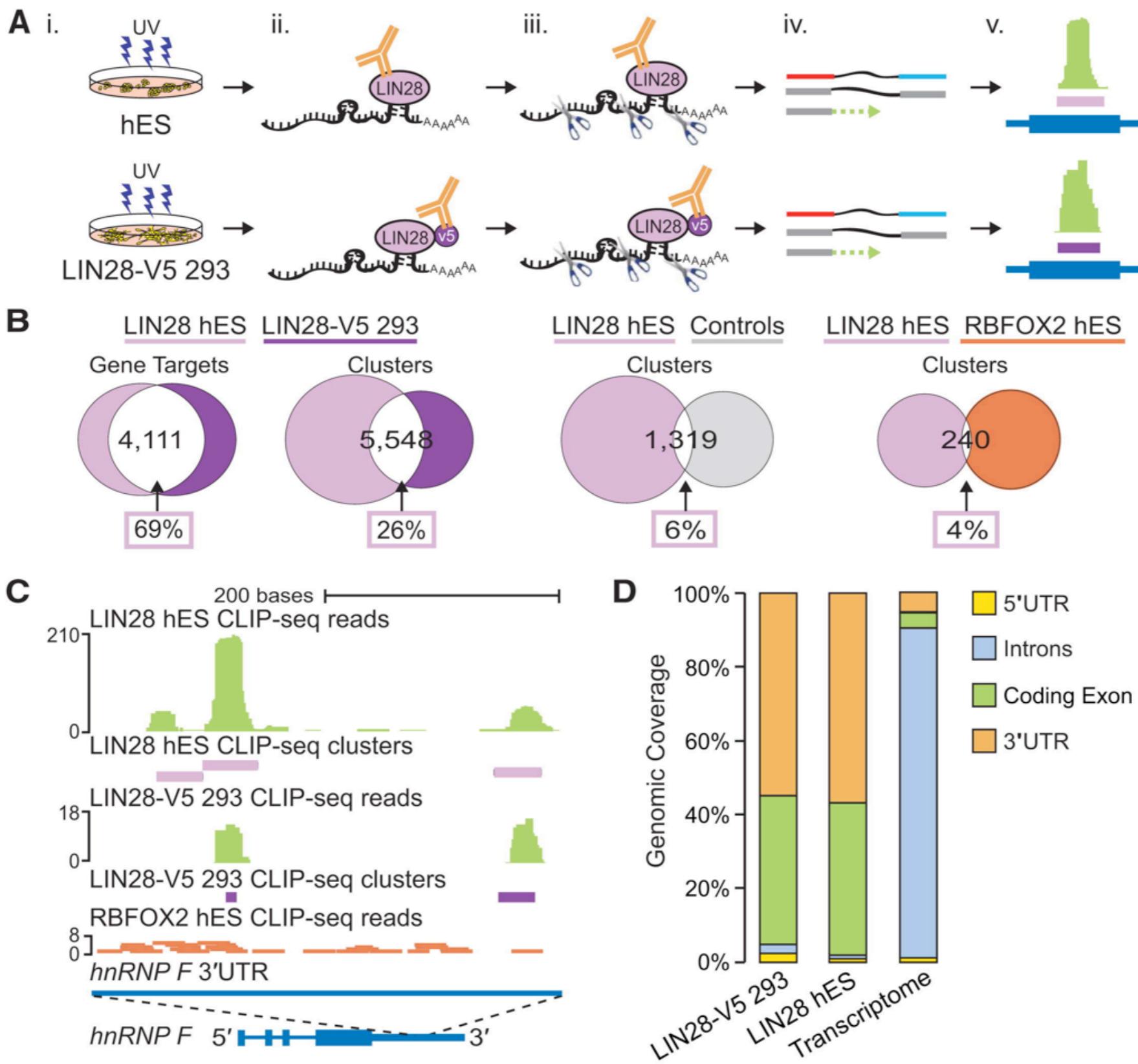
# NET-seq



# CLIP-seq



# CLIP-seq



# In conclusion

- There are many sequencing platforms available, and the technology is being very actively developed.
- Almost anything that results in fragments of DNA or RNA can be turned into a sequencing protocol.
- So much data, so little time!