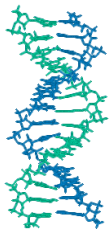
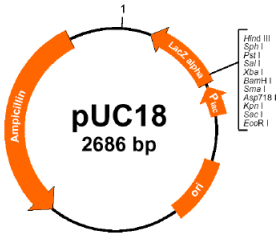
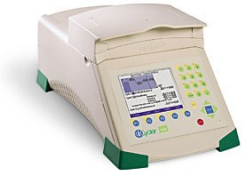
Molecular Biology



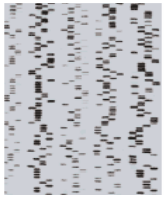
DNA structure and function



Recombinant DNA technology



DNA amplification



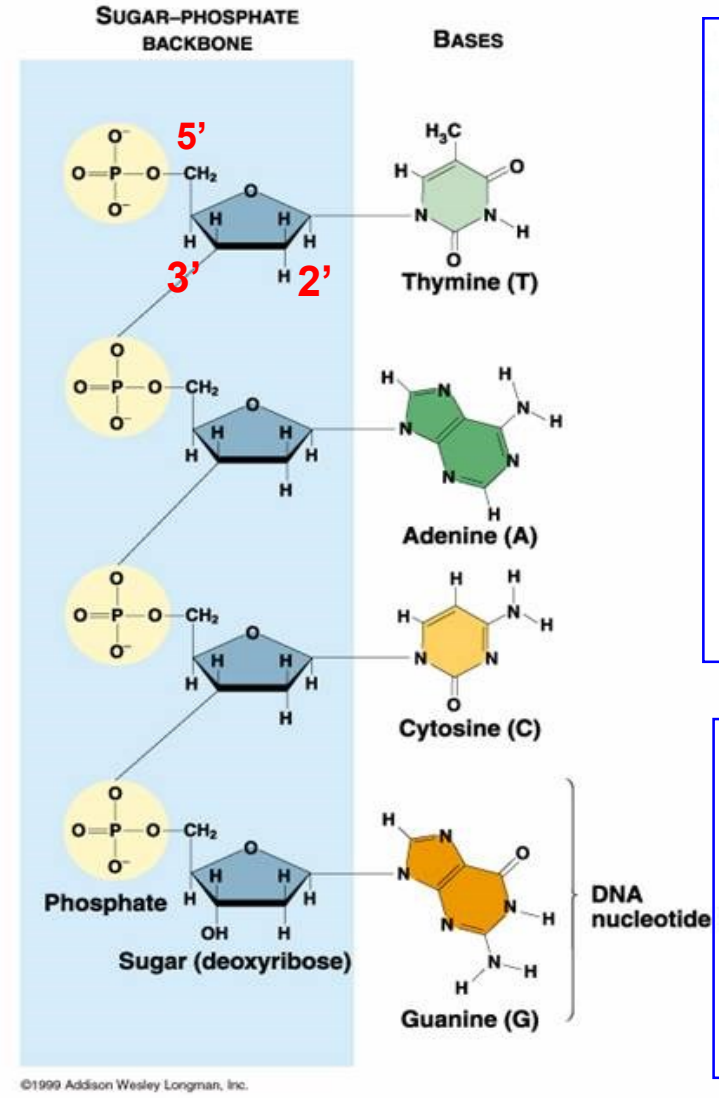
DNA sequencing

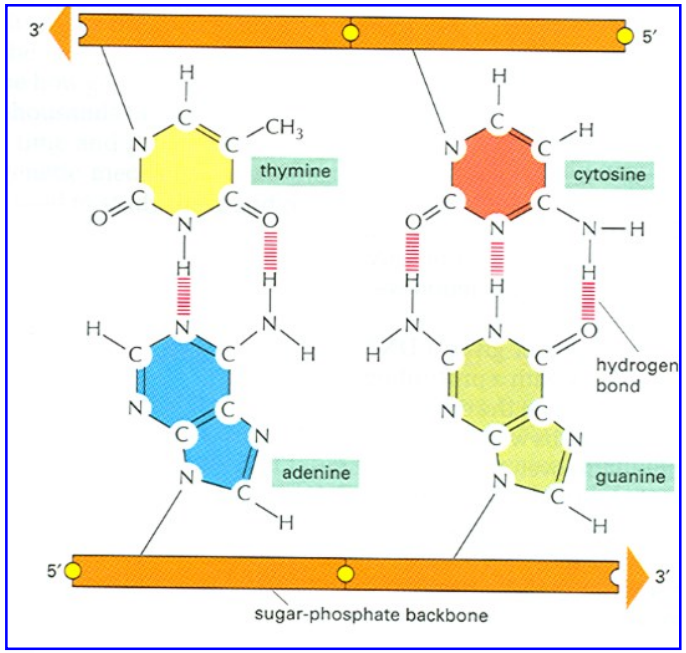
Arthur Günzl, PhD
Dept. of Genetics & Genome Sciences
University of Connecticut Health Center

Bioinformatics and Computational Biology Course (BME5800)

September 13, 2016

Deoxyribonucleic Acid (DNA)

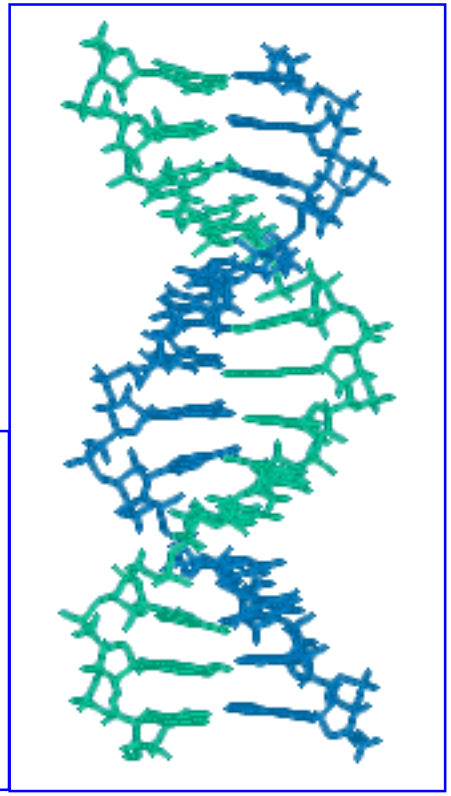




The haploid human genome consists of ~3,000,000,000 bp

The combined α-helical length is ~2 m

A human nucleus' diameter is ~10 μm, containing 46 chromosomes



From Alberts *et al.*, Molecular Biology of the Cell

Which one is the reverse primer ?

a. 5′-ACTGTTCCTCGAAGGGTT-3′

b. 5′-AACCCTTCGAGGAACAGT-3′

c. 5′-TGACAAGGAGCTTCCCAA-3′

d. 5′-GGTCCTGGAGAAAAGTCT-3′

e. None of the above

Recombinant DNA Technology

Recombination

Process in which DNA molecules are broken and the fragments are rejoined in new combinations

Recombinant DNA

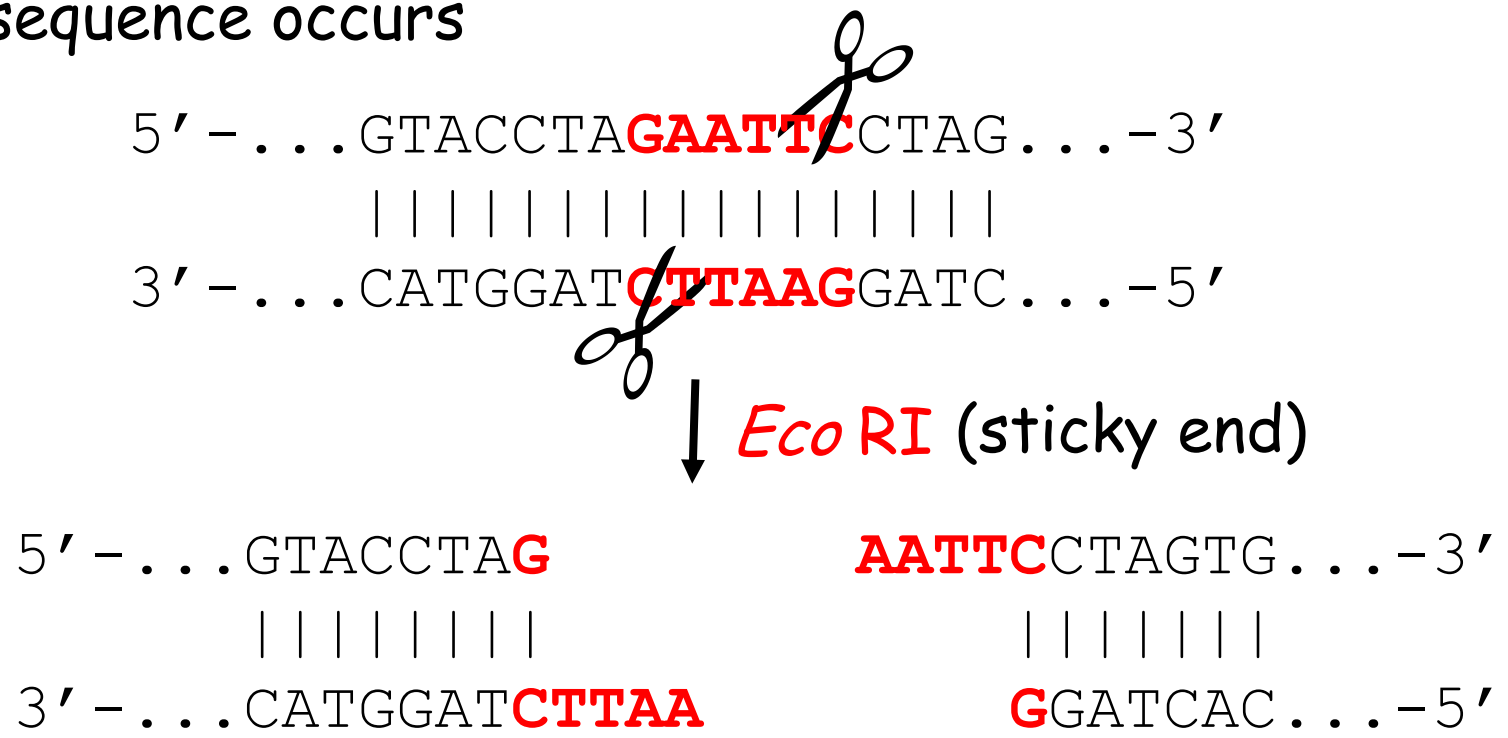
Any DNA molecule formed by joining DNA segments from different sources

Recombinant DNA Technology

Key Enzymes

Restriction enzyme

One of a large number of endonucleases that can cleave a DNA molecule at any site where a specific short nucleotide sequence occurs



RECOGNITION AND CLEAVAGE SITES OF SOME TYPE II RESTRICTION ENZYMES

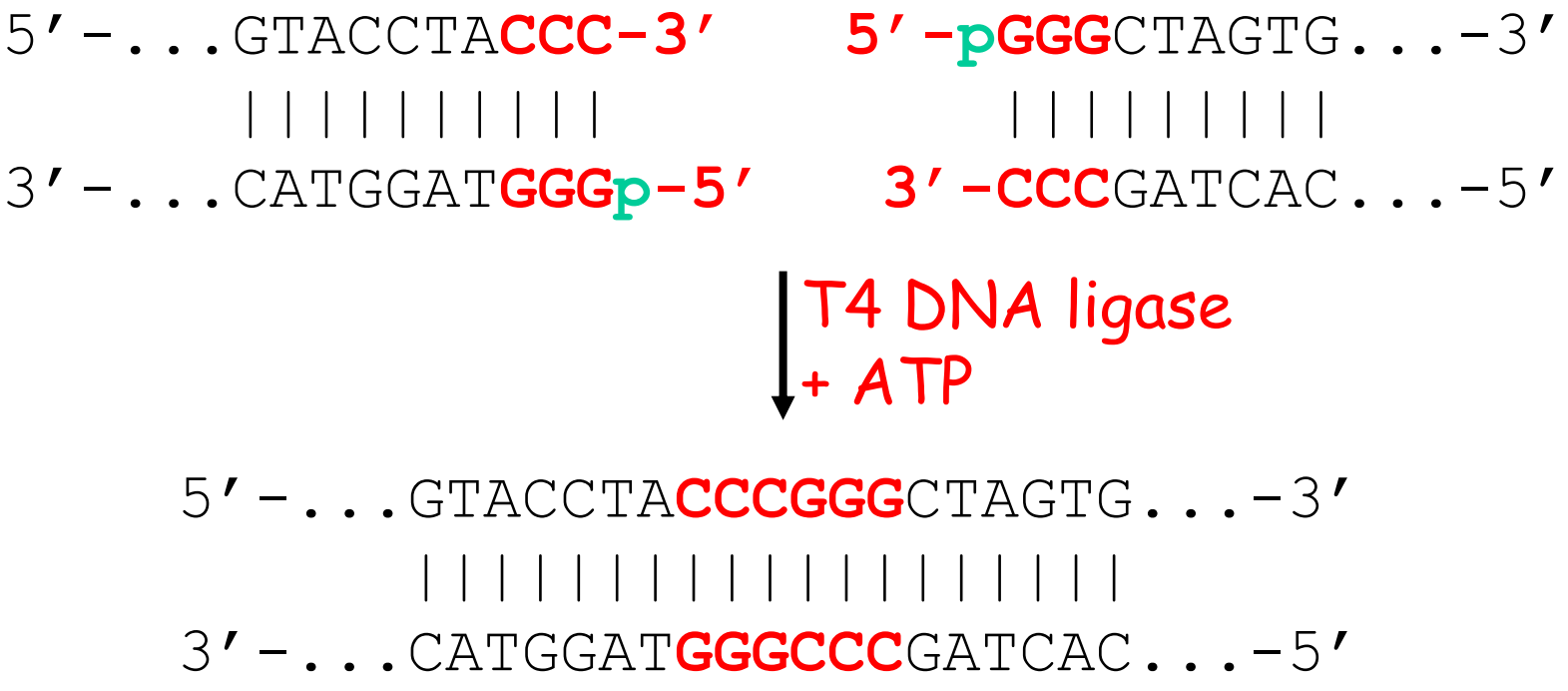
Enzyme	Recognition Sequence[a]	Microorganism
*Alu*I	AG↓C*T	*Arthrobacter luteus*
*Bam*HI	G↓GATC*C	*Bacillus amyloliquefaciens* H
*Bgl*I	GCCNNNN↓NGCC	*Bacillus globigii*
*Bgl*II	A↓GATCT	*Bacillus globigii*
*Eco*RI	G↓AA*TTC	*Escherichia coli* RY13
*Eco*RII	↓CC*(A/T)GG	*Escherichia coli* R245
*Eco*RV	GA*T↓ATC	*Eschericia coli* J62P7G74
*Hae*II	RGCGC↓Y	*Haemophilus aegyptius*
*Hae*III	GG↓C*C	*Haemophilus aegyptius*
*Hin*dIII	A*↓AGCTT	*Haemophilus influenzae* R_d
*Hpa*II	C↓C*GG	*Haemophilus parainfluenzae*
*Msp*I	C*C↓GG	*Moraxella* species
*Pst*I	CTGCA*↓G	*Providencia stuartii* 164
*Pvu*II	CAG↓C*TG	*Proteus vulgaris*
*Sal*I	G↓TCGAC	*Streptomyces albus* G
*Taq*I	T↓CGA*	*Thermus acuaticus*
*Xho*I	C↓TCGAG	*Xanthomonas holcicola*
*Sma*I	CCC↓GGG	*Serratia marcescens*

Recombinant DNA Technology

Key Enzymes

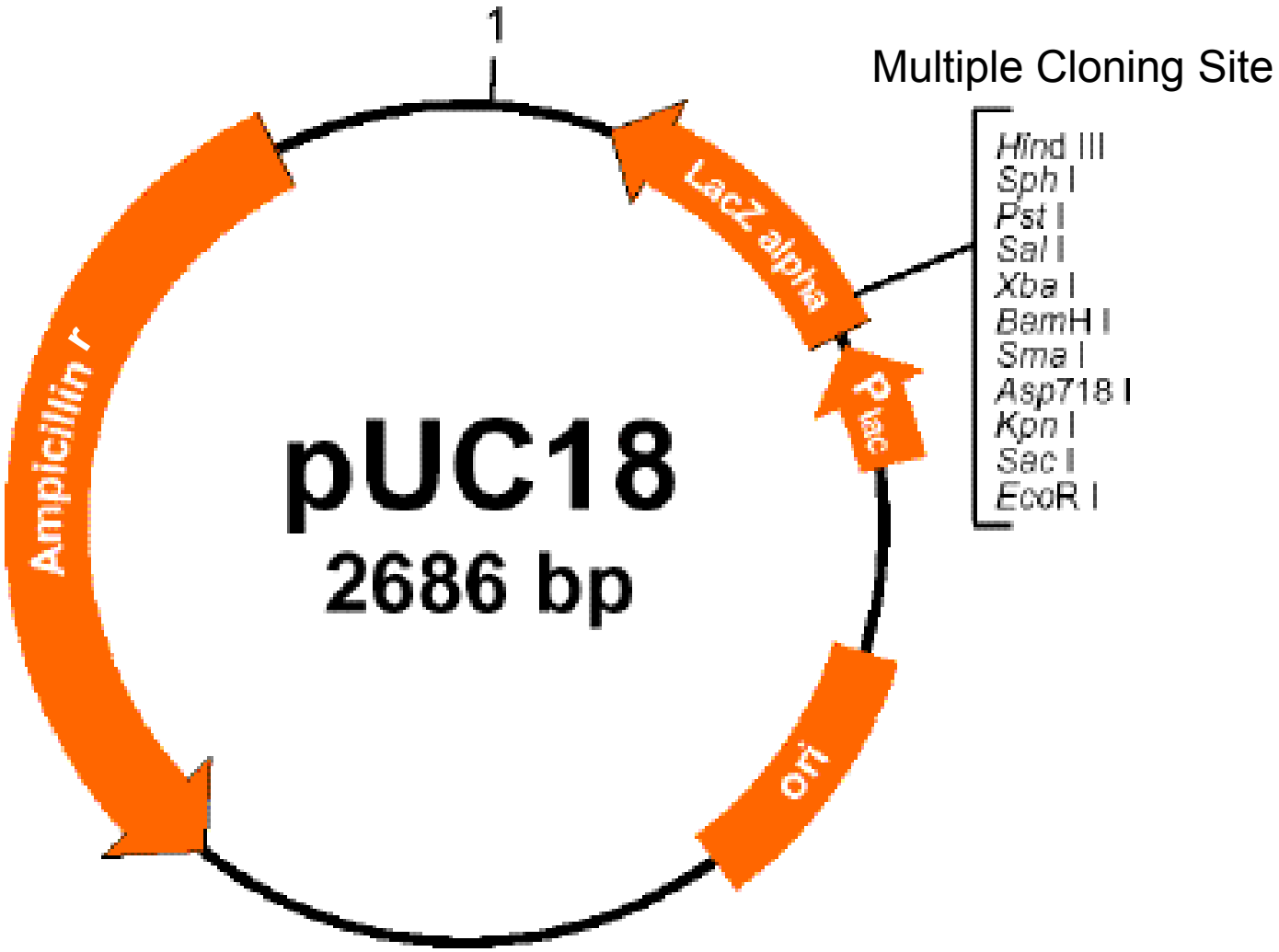
DNA ligase

Enzyme that joins the ends of two DNA strands together with a covalent bond to make a continuous DNA strand

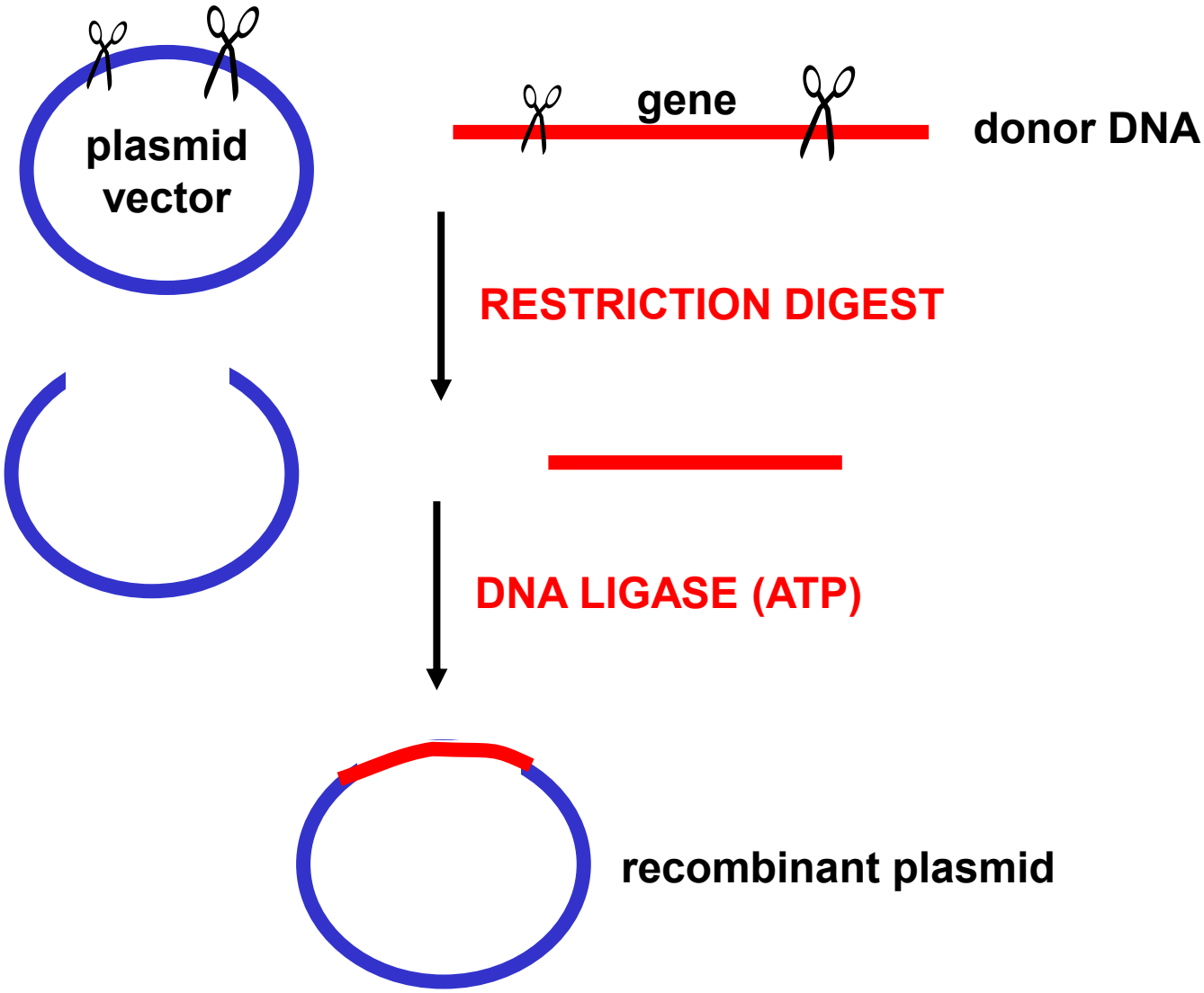


Recombinant DNA Technology

plasmid cloning vector



Gene cloning



Transformation of *E. coli*

The bacteria will not take up plasmid DNA from its environment. It has to be mistreated to do so, e.g. the membrane has to be damaged to allow for DNA entry.

E. coli is made competent for transformation by

$CaCl_2$ / heat shock

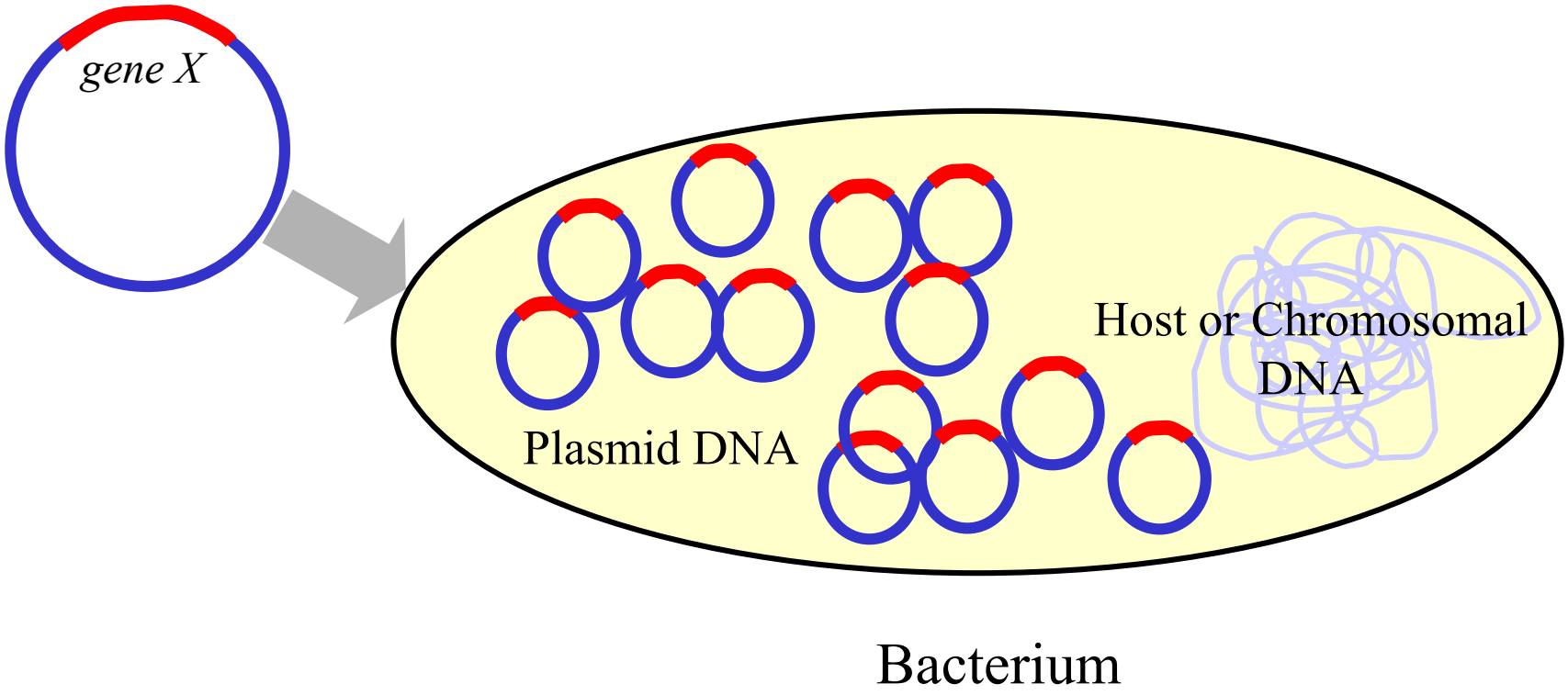
electroporation

Transformed *E. coli* are selected via expression of a selectable marker (Amp^r)

Plasmid DNA does not integrate into the bacterial genome

It replicates autonomously typically in high copy numbers

Due to its small size (~5 kbp) it can be separated from genomic DNA (~3,700 kbp) and efficiently purified.



How to obtain DNA fragments for cloning ?

(Note: Historic slide)

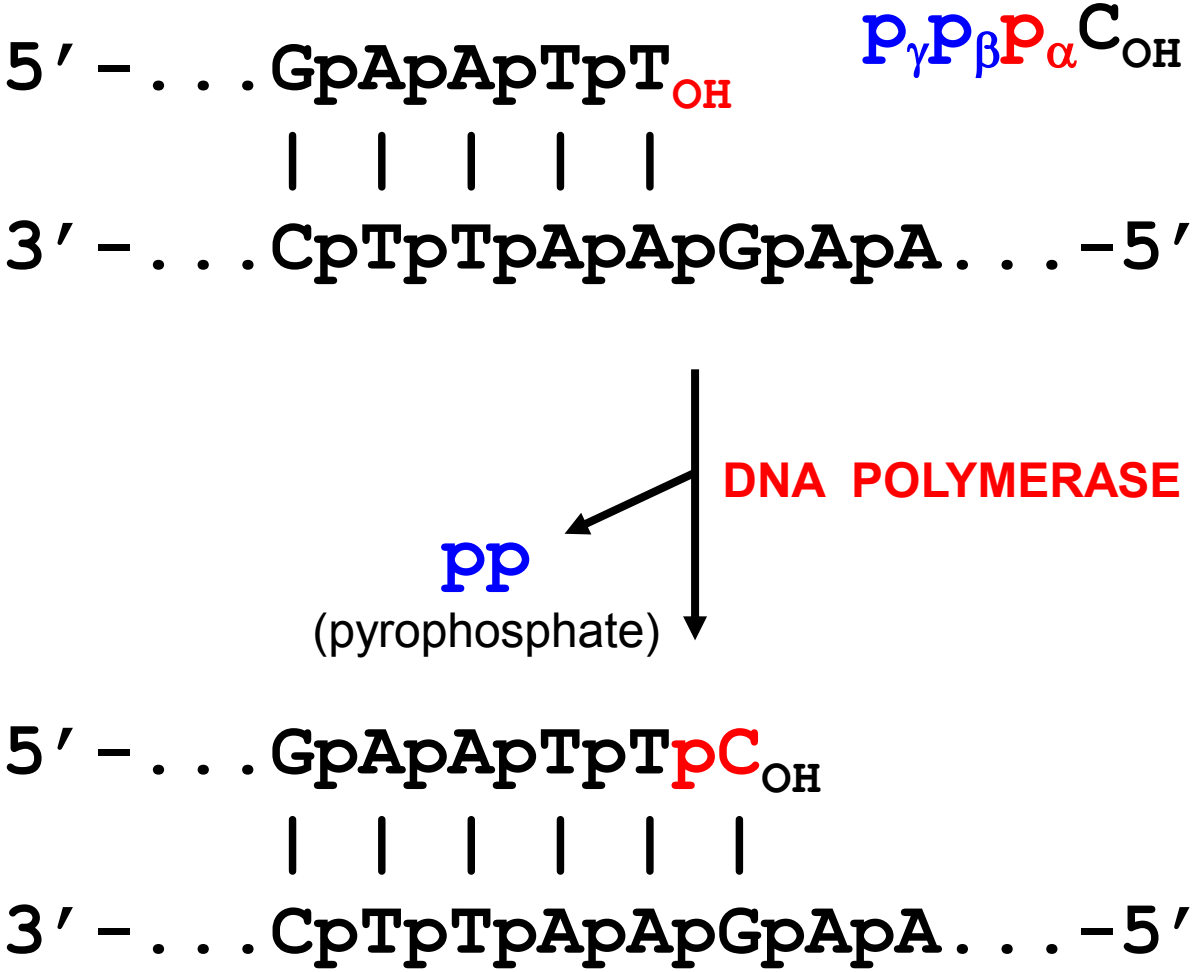
Libraries:

Genomic DNA – fragmentation (mechanical, sonication, enzymatic) –> cloning

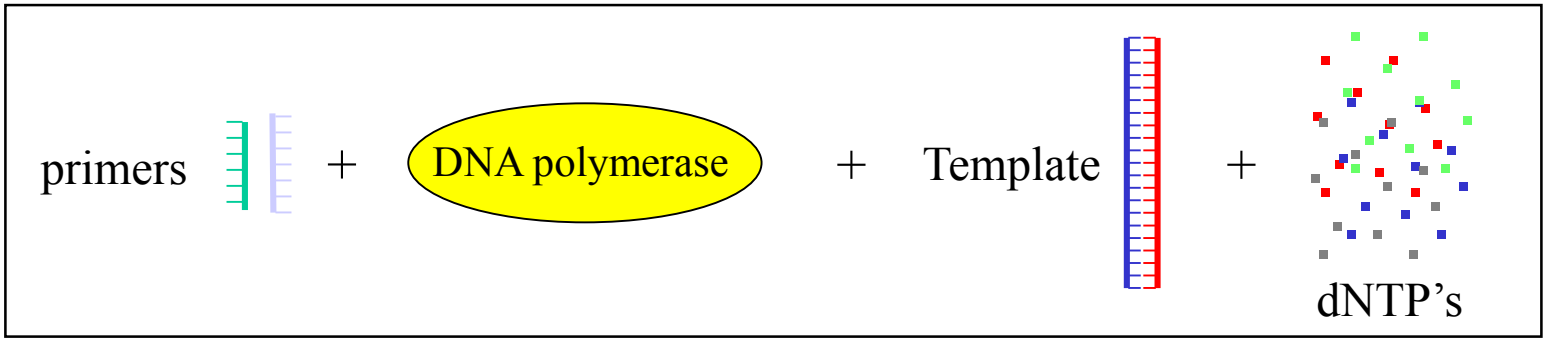
mRNA -> cDNA –> cloning

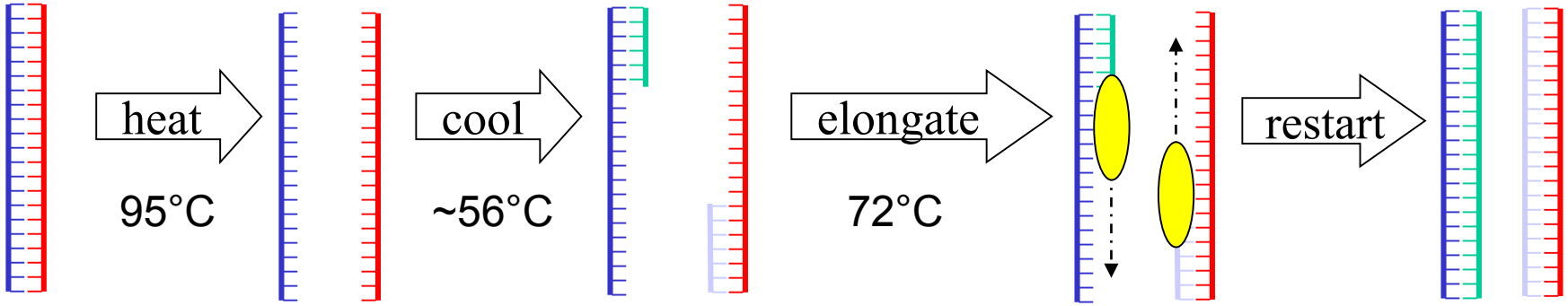
Nowadays, genome sequences are known or relatively easy to obtain. Thus, it is possible to amplify a DNA sequence of interest and either use it directly in an assay or clone it into a plasmid construct

DNA is synthesized by DNA polymerase



Amplification of DNA fragment of interest by the Polymerase Chain Reaction (PCR)





Heating denatures DNA (separates strands)

Primers anneal to complimentary strands

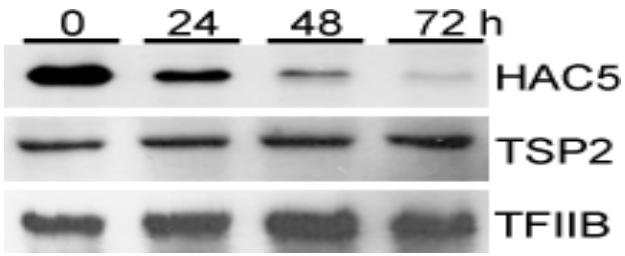
DNA polymerase elongates from primer annealing

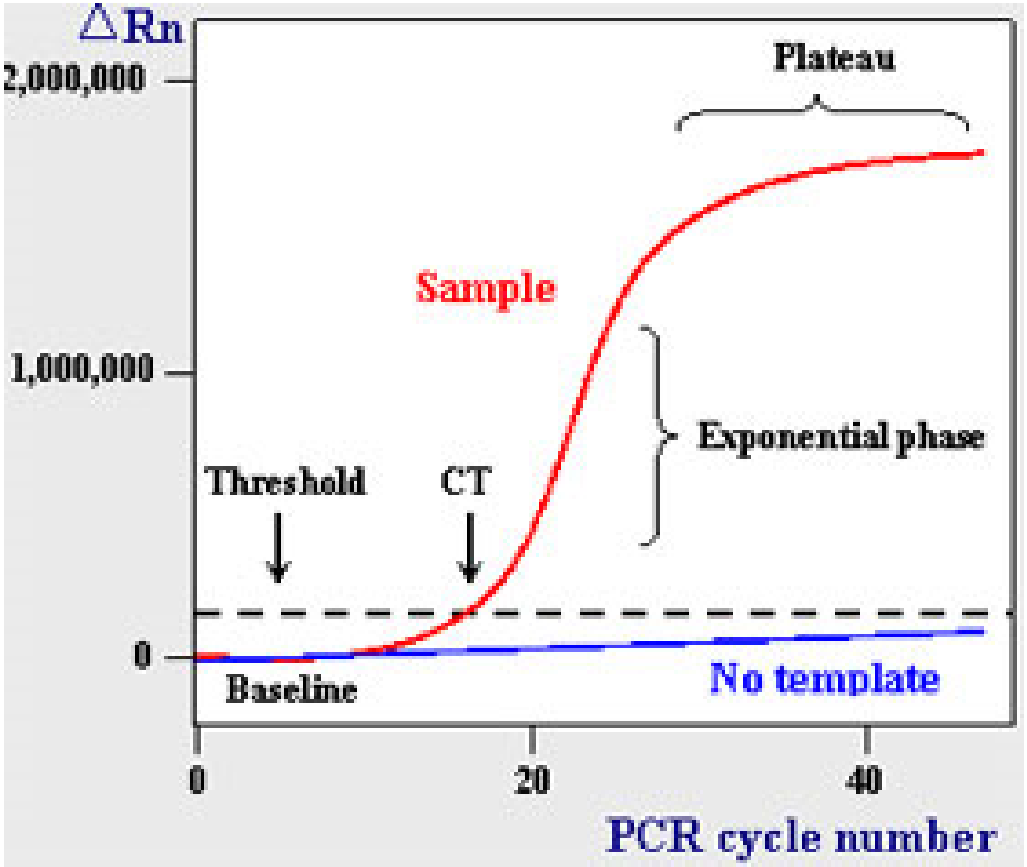
Multiple reaction cycles result in an exponential increase of specific amplification product

Kary B Mullis obtained the Nobel prize for Chemistry in 1993 for inventing PCR

[real-time] quantitative (q)PCR

Semi-quantitative (RT-)PCR

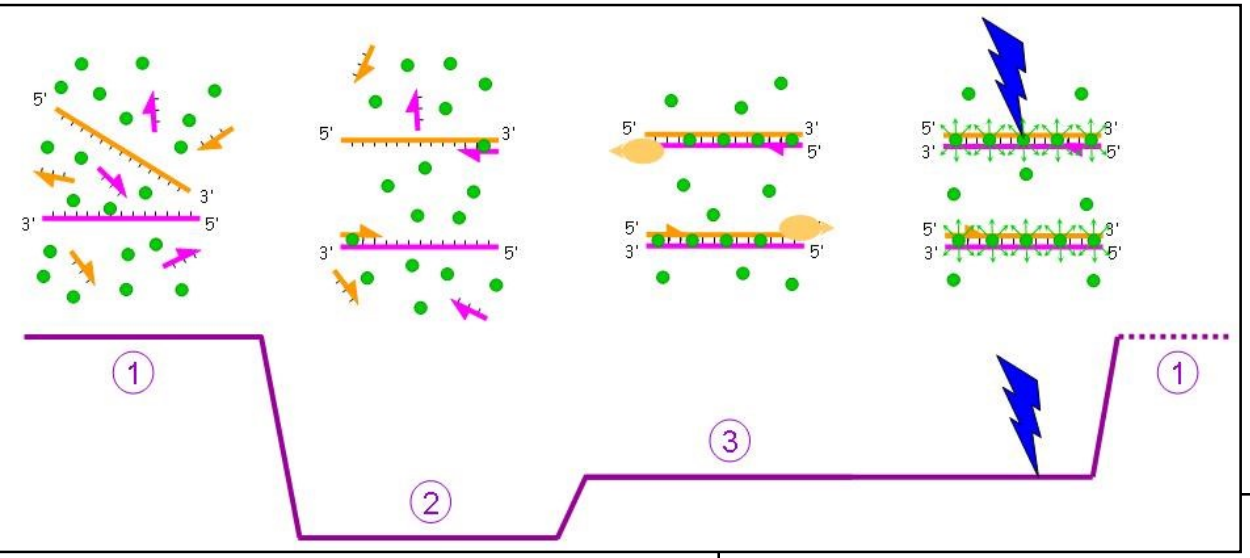




Ct (cycle threshold) is defined as the number of cycles required for the fluorescent signal to cross the threshold

qPCR

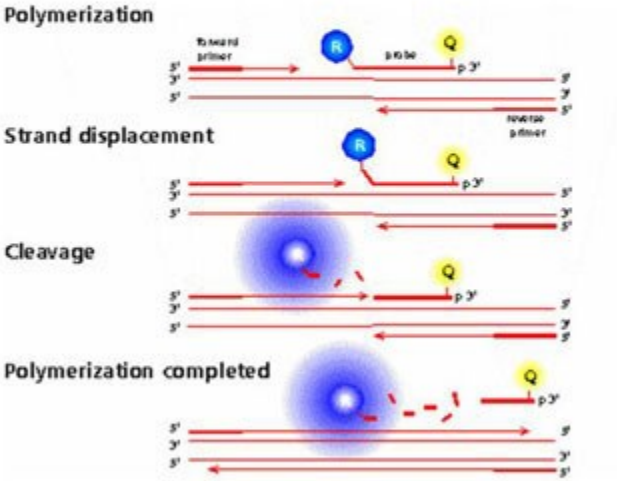
SYBR green

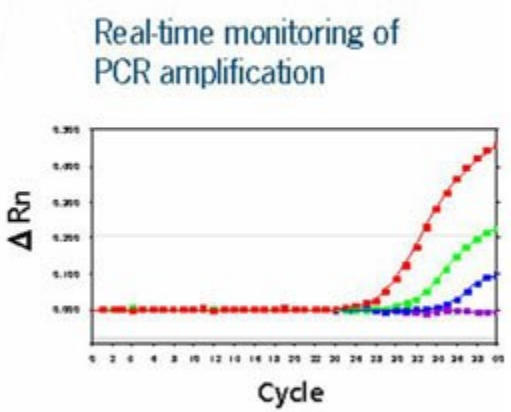


SYBR Green binds to DNA.
The resulting DNA-dye-complex absorbs blue light (λ_{max} = 497 nm) and emits green light (λ_{max} = 520 nm). The stain preferentially binds to double-stranded DNA

TaqMan system

The TaqMan probe principle relies on the 5′–3′ exonuclease activity of Taq polymerase to cleave a quenched dual-labeled probe that binds to the target sequence.
This method is highly specific and, due to the availability of different fluorophores, it can be applied to multichannel applications that measure more than one amplification in the same reaction.





DNA Sequencing

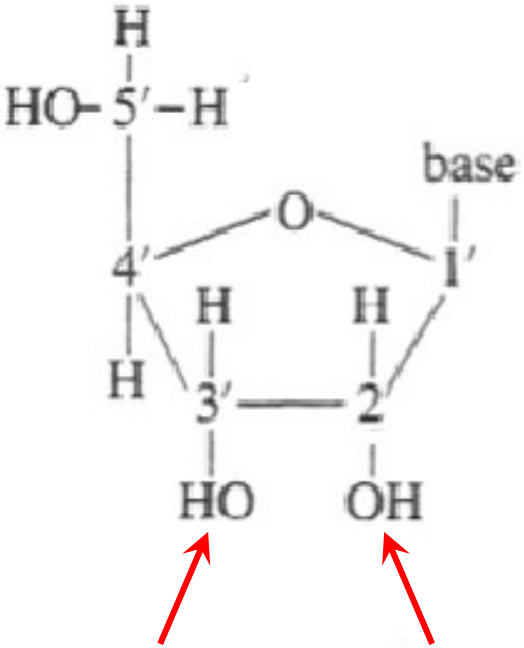
1st and "Next" generation sequencing (NGS) is predominantly based on the "chain termination" ("enzymatic", "dideoxy") method invented by Frederick Sanger *et al.* (1977, PNAS 74, pp 5463-5467).

The principle is based on dideoxy nucleotides in which the 3' hydroxyl group, which is essential for formation of the next phosphodiester bond, is replaced by a hydrogen.

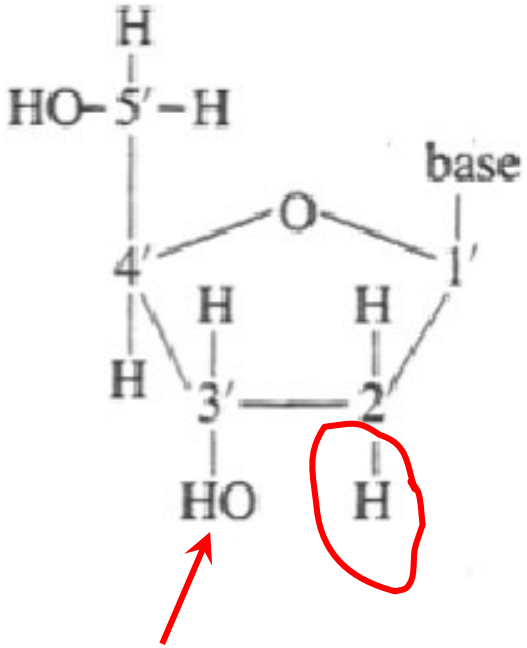
These nucleotides are so-called chain terminators because they prevent the incorporation of the next nucleotide and cause DNA synthesis to stop.

Frederick Sanger and Walter Gilbert obtained the Nobel prize for Chemistry in 1980 for ***Determination of base sequences in nucleic acids***

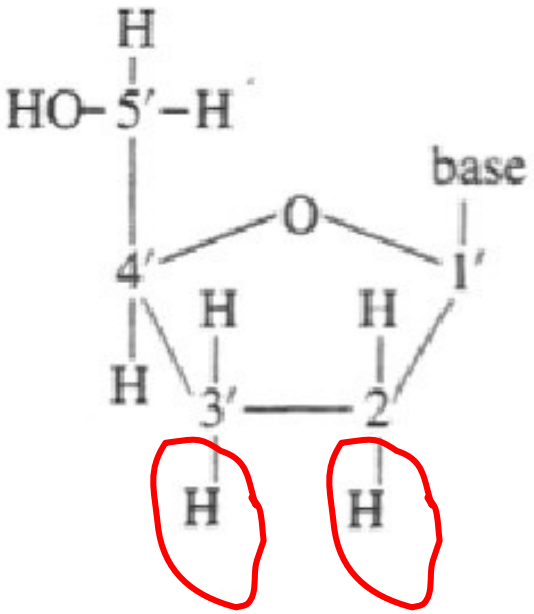
Ribose sugars



ribose

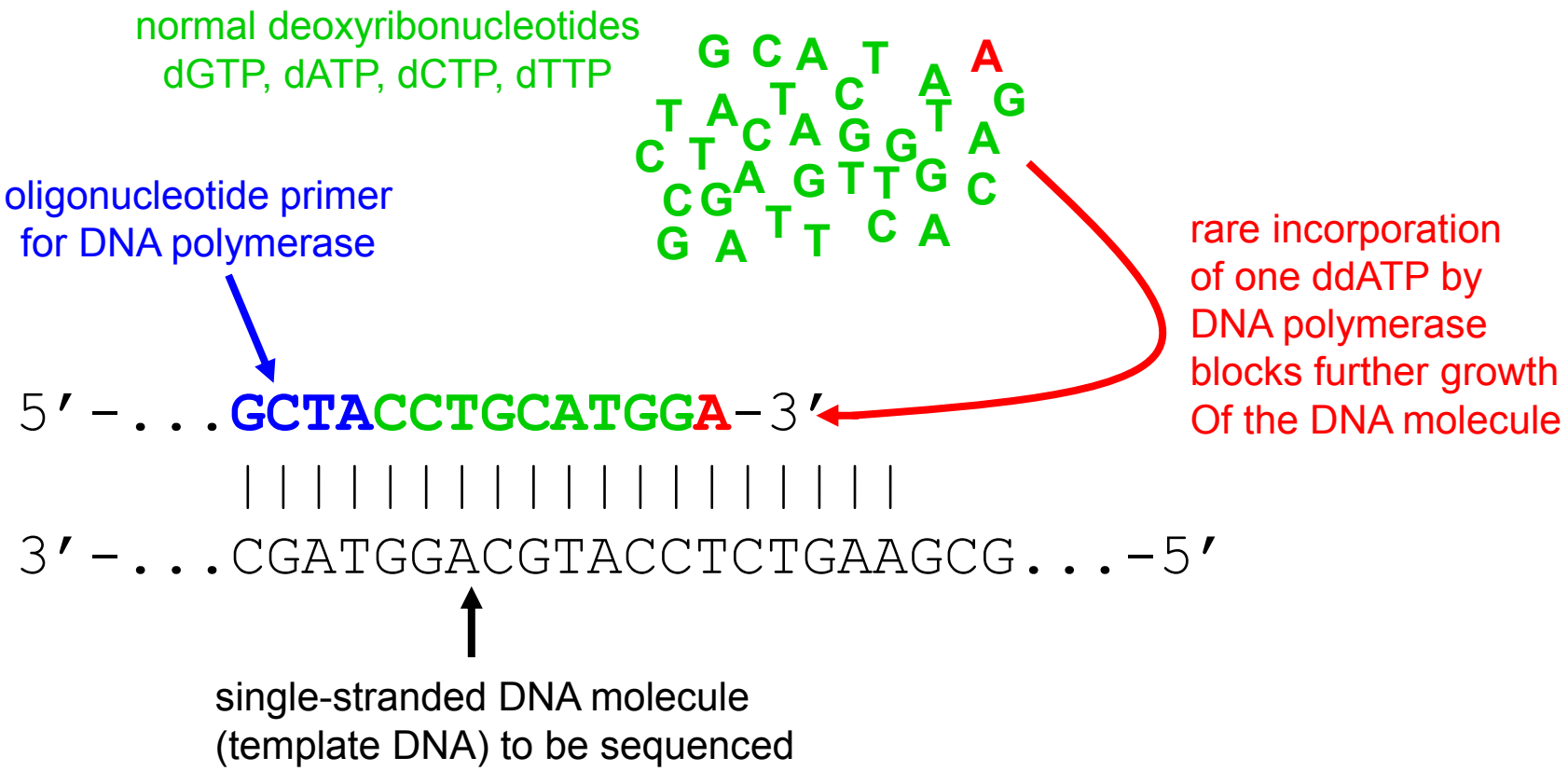


deoxyribose

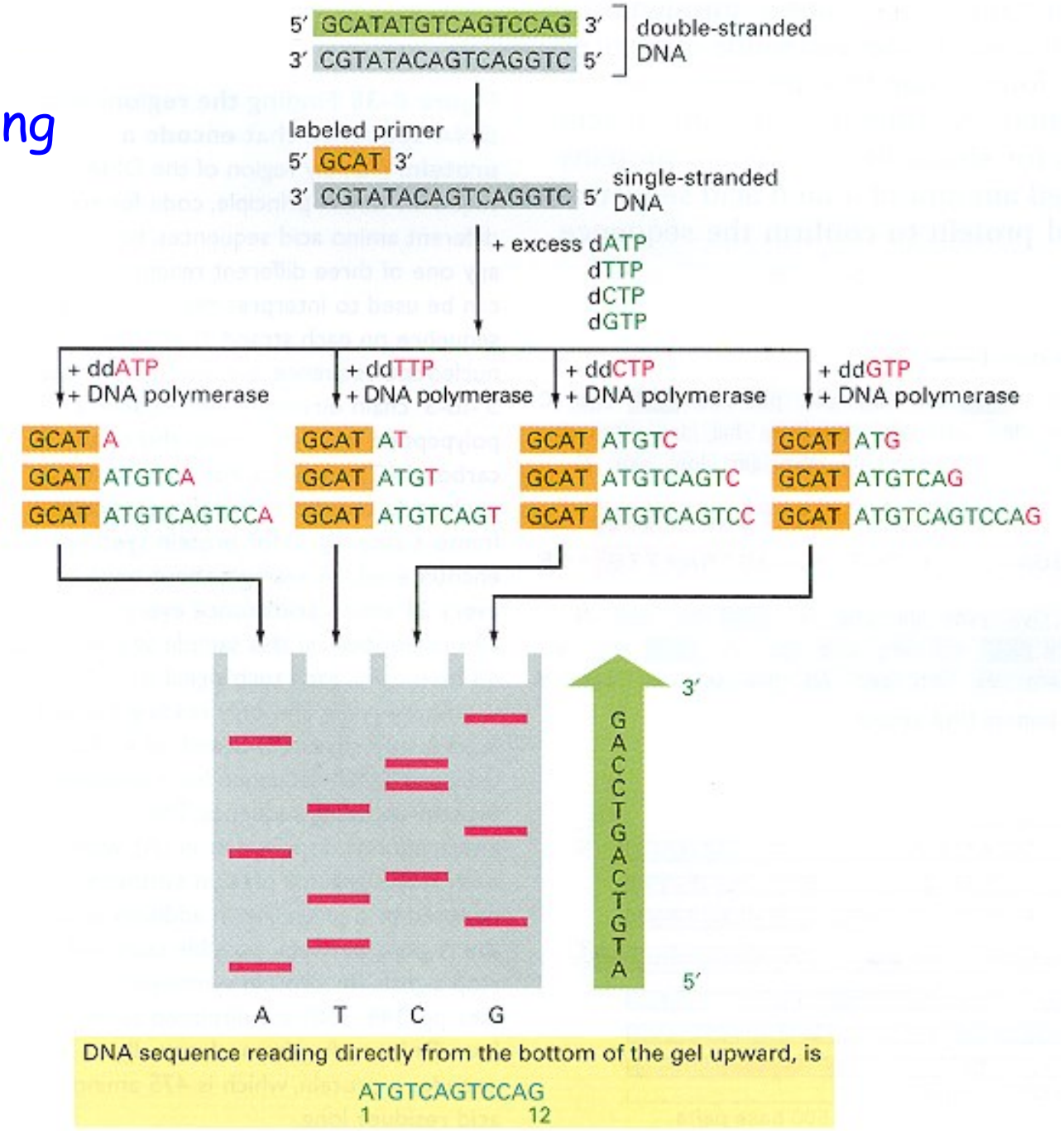


di-deoxyribose

DNA Sequencing

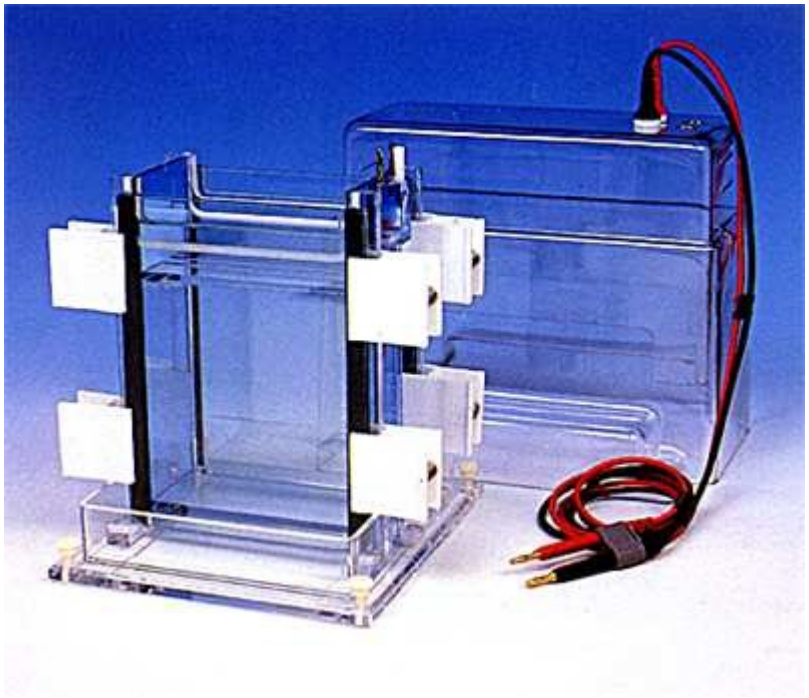


DNA Sequencing

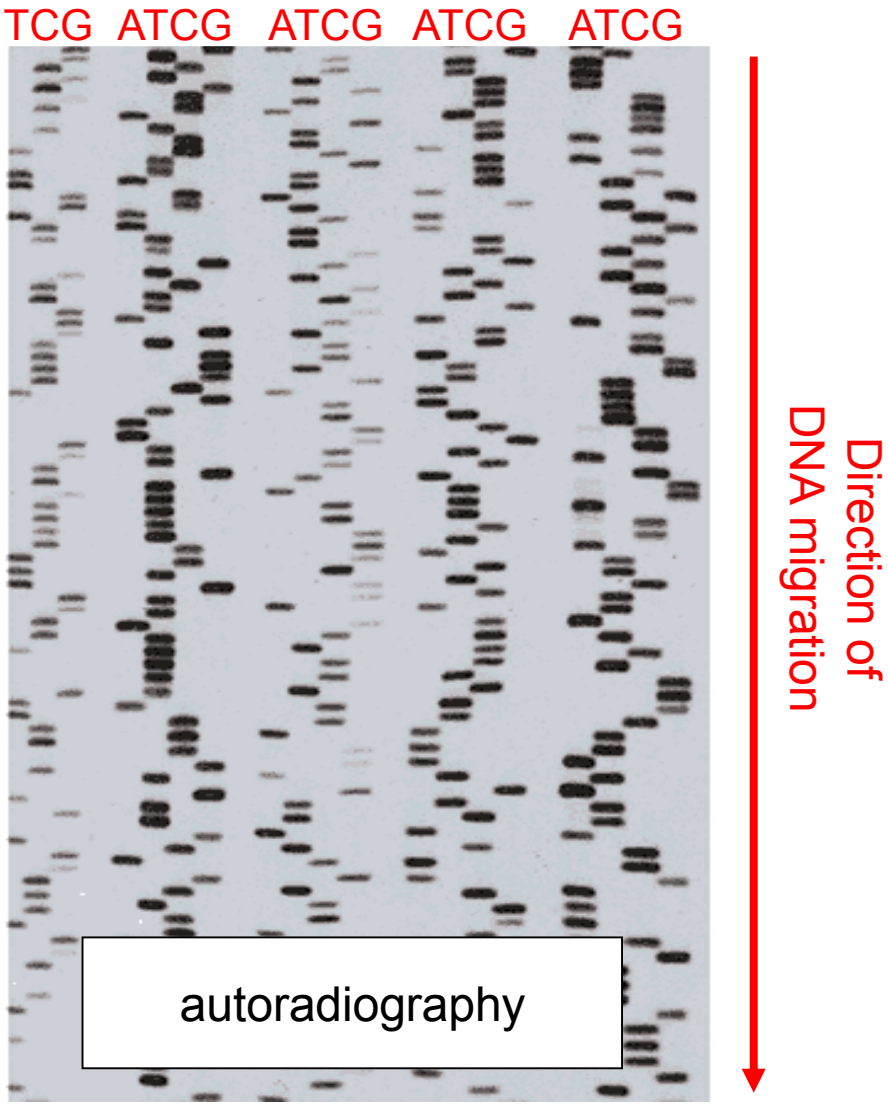


From Alberts *et al.*, Molecular Biology of the Cell, p 505

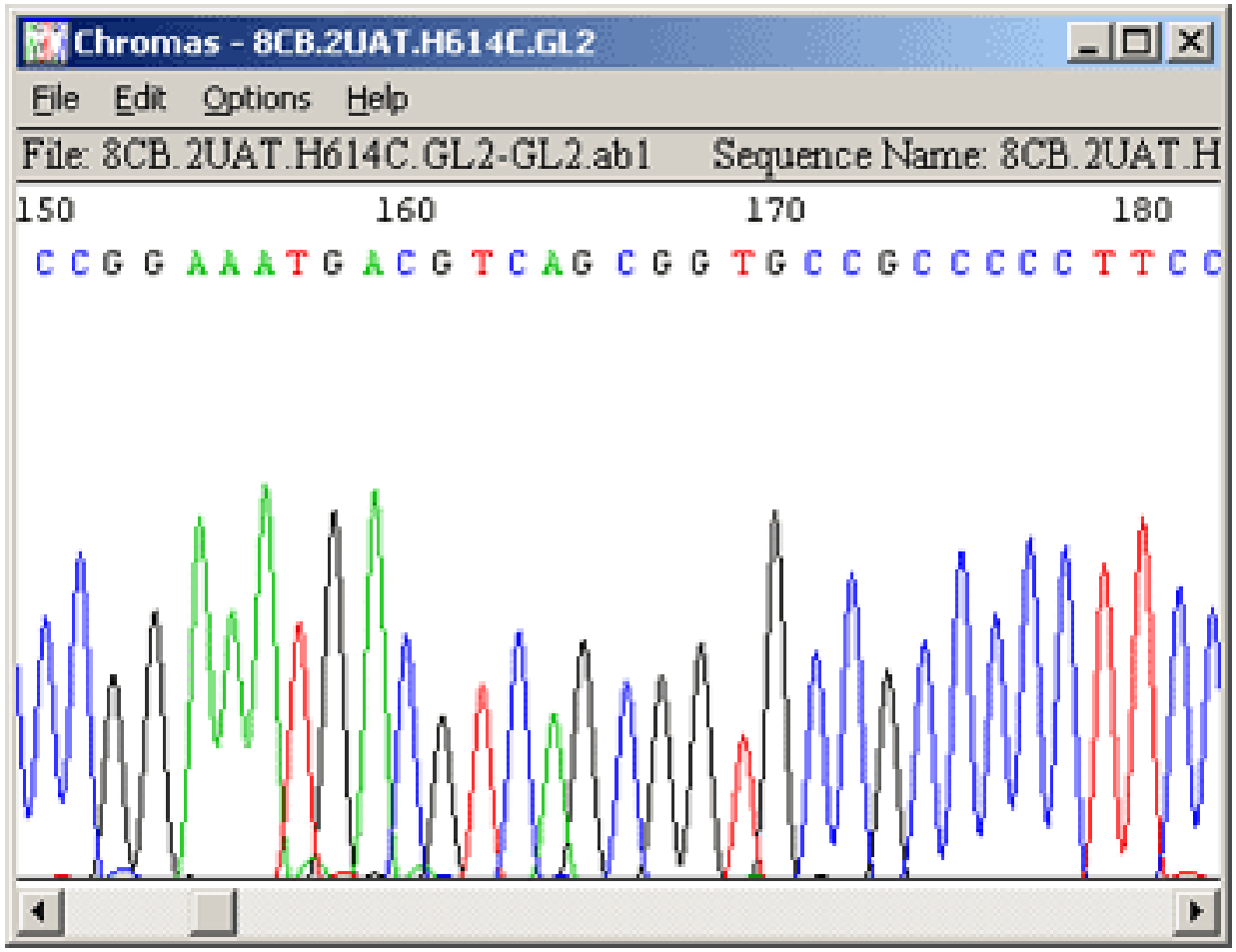
Polyacrylamide Gel Electrophoresis (PAGE)



PAGE apparatus



Modern Automatic Sequencing works in one reaction with fluorescent-labeled terminators



Next (2nd) generation sequencing technology:

454 sequencing (pyrosequencing)

illumina solexa sequencing

Advantages

fast (Hi-seq 4000: per 3.5 day run: 12 human genomes or 1.5 Tb of sequence data (~5 billion reads).

cost-effective (reduced cost per base)

without cloning (PCR integrated in the method)

Disadvantages

short reads (454: 1000 bp/read)/Solexa: up to 250 bp/read)

intense on bioinformatics

"next generation sequencing"

Sequencing the nuclear genome of the extinct woolly mammoth

Webb Miller[1], Daniela I. Drautz[1], Aakrosh Ratan[1], Barbara Pusey[1], Ji Qi[1], Arthur M. Lesk[1], Lynn P. Toms
Michael D. Packard[1], Fangqing Zhao[1], Andrei Sher[2]‡, Alexei Tikhonov[3], Brian Raney[4], Nick Patterson[5],
Kerstin Lindblad-Toh[5], Eric S. Lander[5], James R. Knight[6], Gerard P. Irzyk[6], Karin M. Fredrikson[7], Timothy T.
Sharon Sheridan[7], Tom Pringle[8] & Stephan C. Schuster[1]

Nature 2008, 456:387-392

A Draft Sequence of the Neandertal Genome

Richard E. Green,[1*†‡] Johannes Krause,[1†§] A
Udo Stenzel,[1†§] Martin Kircher,[1†§] Nick Patt
Markus Hsi-Yang Fritz,[4†] Nancy F. Hansen,[5†]
Jeffrey D. Jensen,[6†] Tomas Marques-Bonet,[7,1]
Hernán A. Burbano,[1†] Jeffrey M. Good,[1,8†] Rig
Barbara Höber,[1] Barbara Höffner,[1] Madlen S
Eric S. Lander,[2] Carsten Russ,[2] Nathaniel No
Christine Verna,[21] Pavao Rudan,[10] Dejana Br
Vladimir B. Doronichev,[12] Liubov V. Golovano
Javier Fortea,[14]¶ Antonio Rosas,[15] Ralf W. Schr
Daniel Falush,[19†] Ewan Birney,[4†] James C. Mul
Janet Kelso,[1†] Michael Lachmann,[1†] David Reich,[2,20*†] Svante Pääbo[1*†]

Neandertals, the closest evolutionary relatives of present-day humans, lived in large parts of Europe and western Asia before disappearing 30,000 years ago. We present a draft sequence of the Neandertal genome composed of more than 4 billion nucleotides from three individuals. Comparisons of the Neandertal genome to the genomes of five present-day humans from different parts of the world identify a number of genomic regions that may have been affected by positive selection in ancestral modern humans, including genes involved in metabolism and in cognitive and skeletal development. We show that Neandertals shared more genetic variants with present-day humans in Eurasia than with present-day humans in sub-Saharan Africa, suggesting that gene flow from Neandertals into the ancestors of non-Africans occurred before the divergence of Eurasian groups from each other.

Science 2010, 328:710-722

ARTICLE

doi:10.1038/nature13437

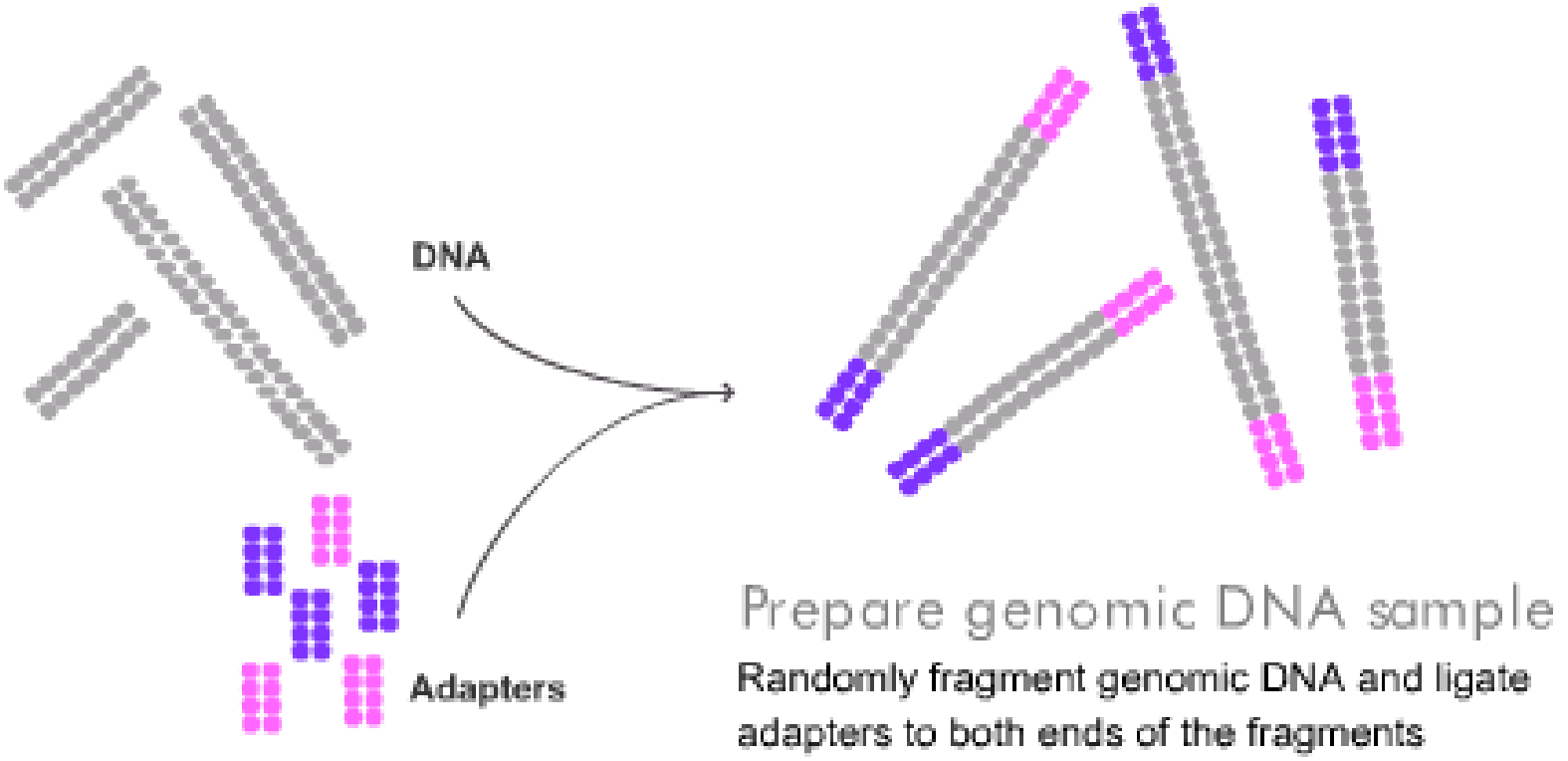
Single-cell RNA-seq reveals dynamic paracrine control of cellular variation

Alex K. Shalek[1,2,3*], Rahul Satija[3*], Joe Shuga[4*], John J. Trombetta[3], Dave Gennert[3], Diana Lu[3], Peilin Chen[4], Rona S. Gertner[1,2], Jellert T. Gaublomme[1,2], Nir Yosef[3], Schraga Schwartz[3], Brian Fowler[4], Suzanne Weaver[4], Jing Wang[4], Xiaohui Wang[4], Ruihua Ding[1,2], Raktima Raychowdhury[3], Nir Friedman[5], Nir Hacohen[3,6], Hongkun Park[1,2,3], Andrew P. May[4] & Aviv Regev[3,7]

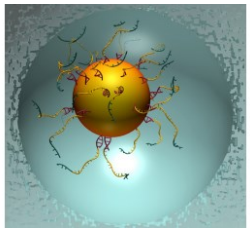
Nature 2014, 510:363-369

Solexa technology (http://www.illumina.com)

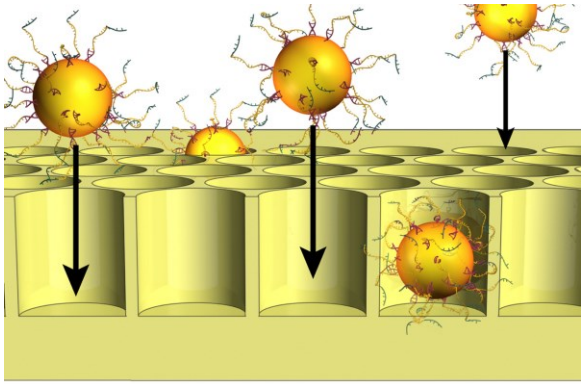
step 1 - prepare genomic DNA sample



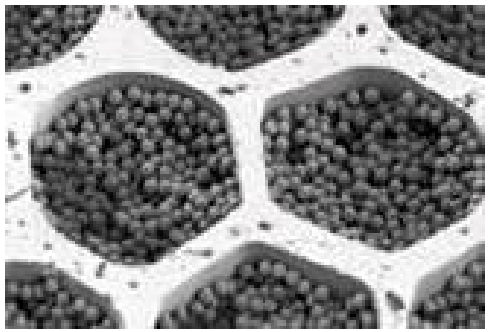
454 pyrosequencing



Attach DNA fragments to beads
Amplify the DNA

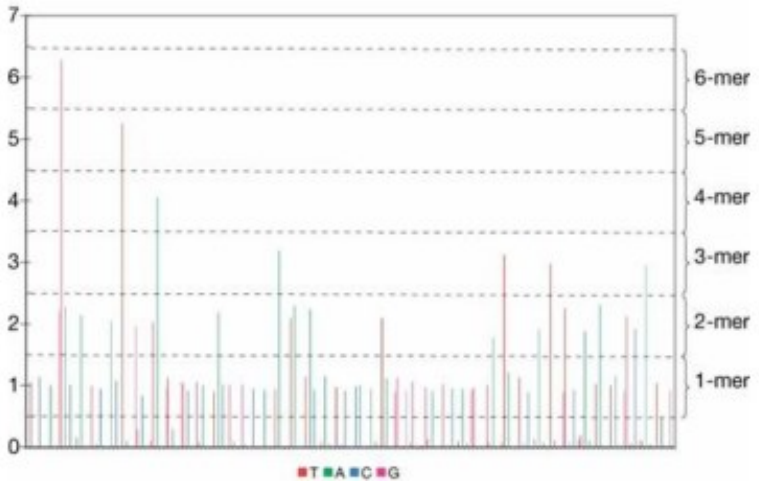


Place beads in microscopic wells
(44 μm in diameter / PicoTiterPlate)



A mix of bead-bound enzymes are also packed into the well.

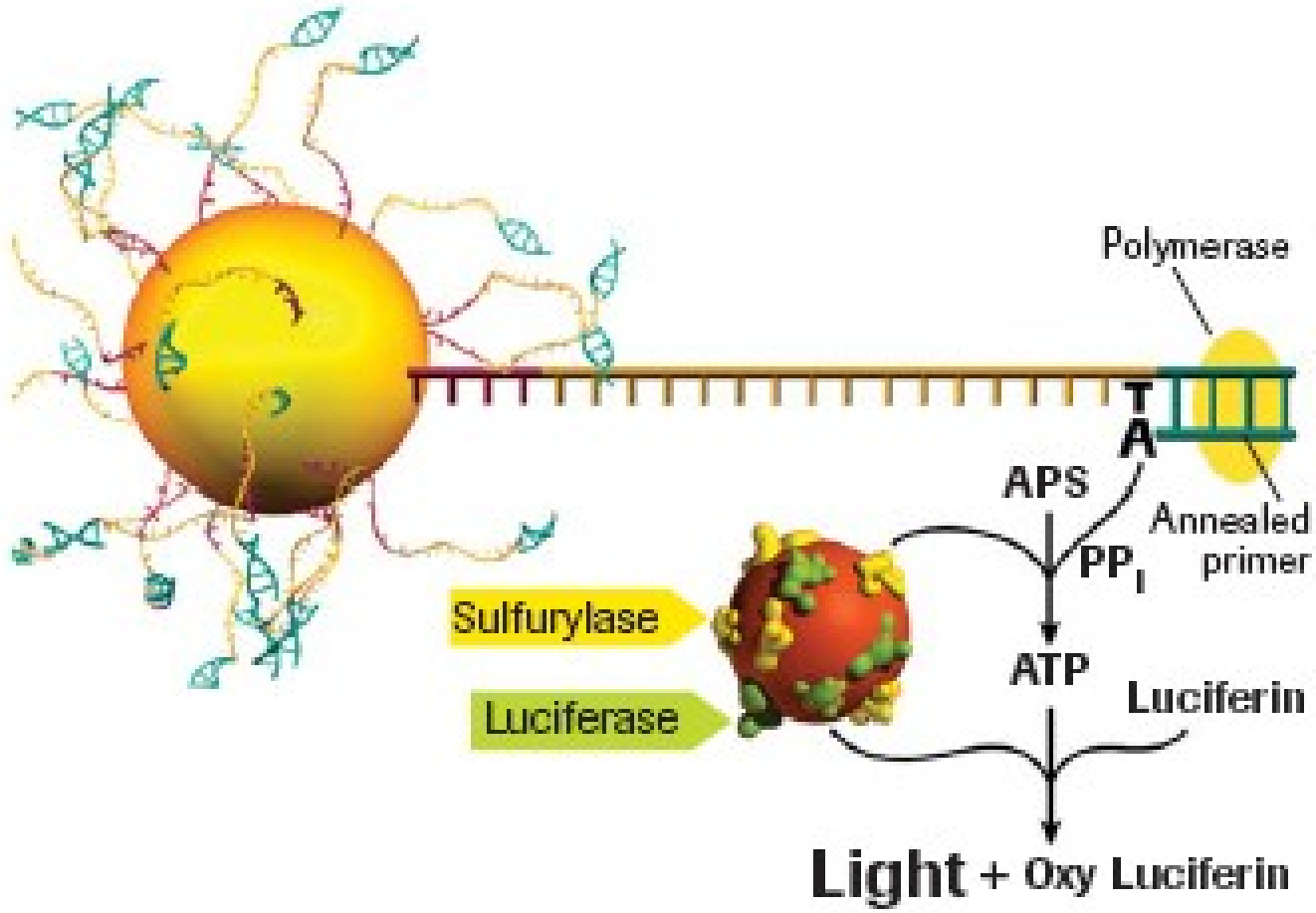
The GSflx can sequence 100 megabases of raw DNA sequence per 4.5-hour run



The plate is flowed successively with oligonucleotides and light flashes are measured

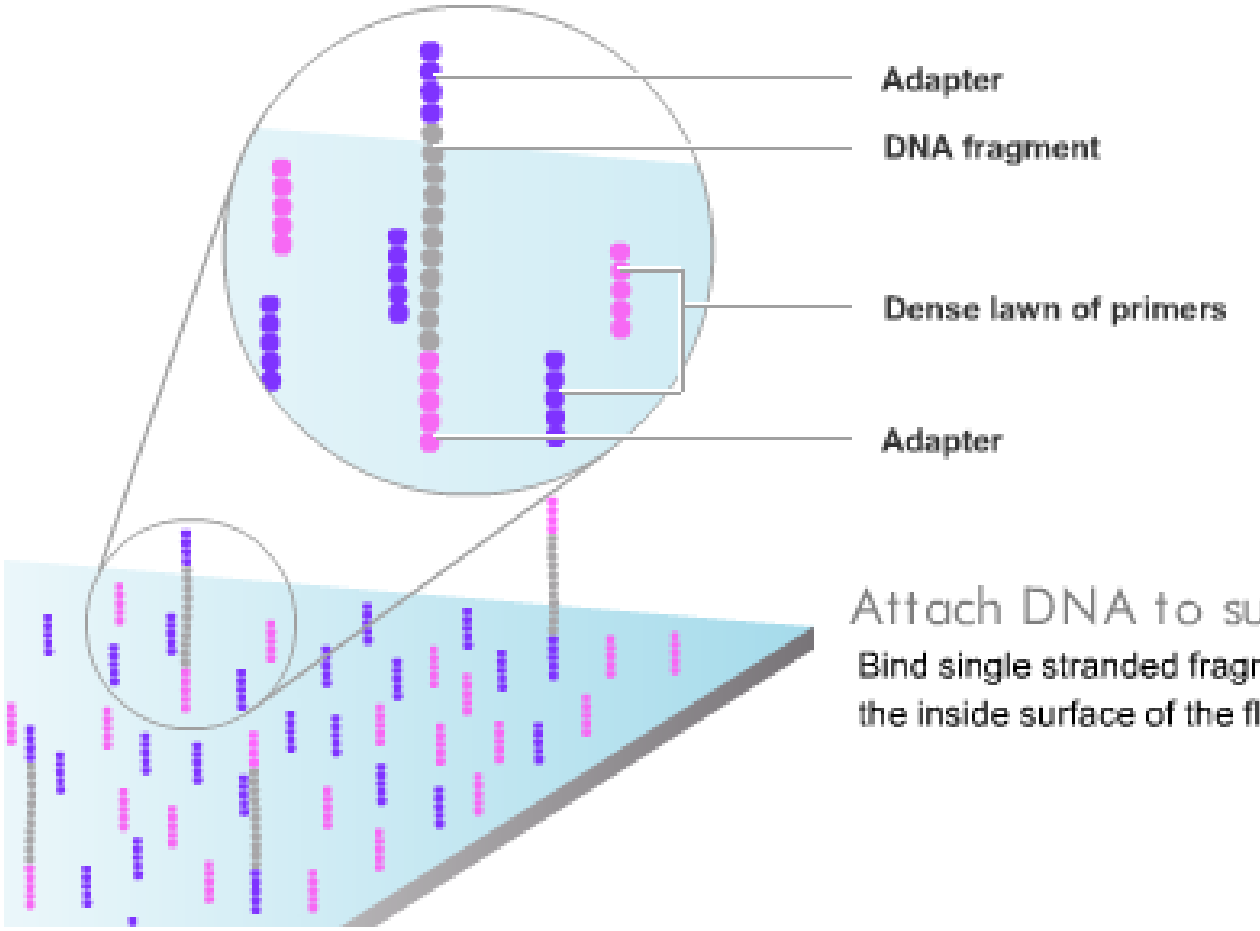
454 pyrosequencing

light signal



The solexa technology

step 2 - immobilize DNA on flow cell surface

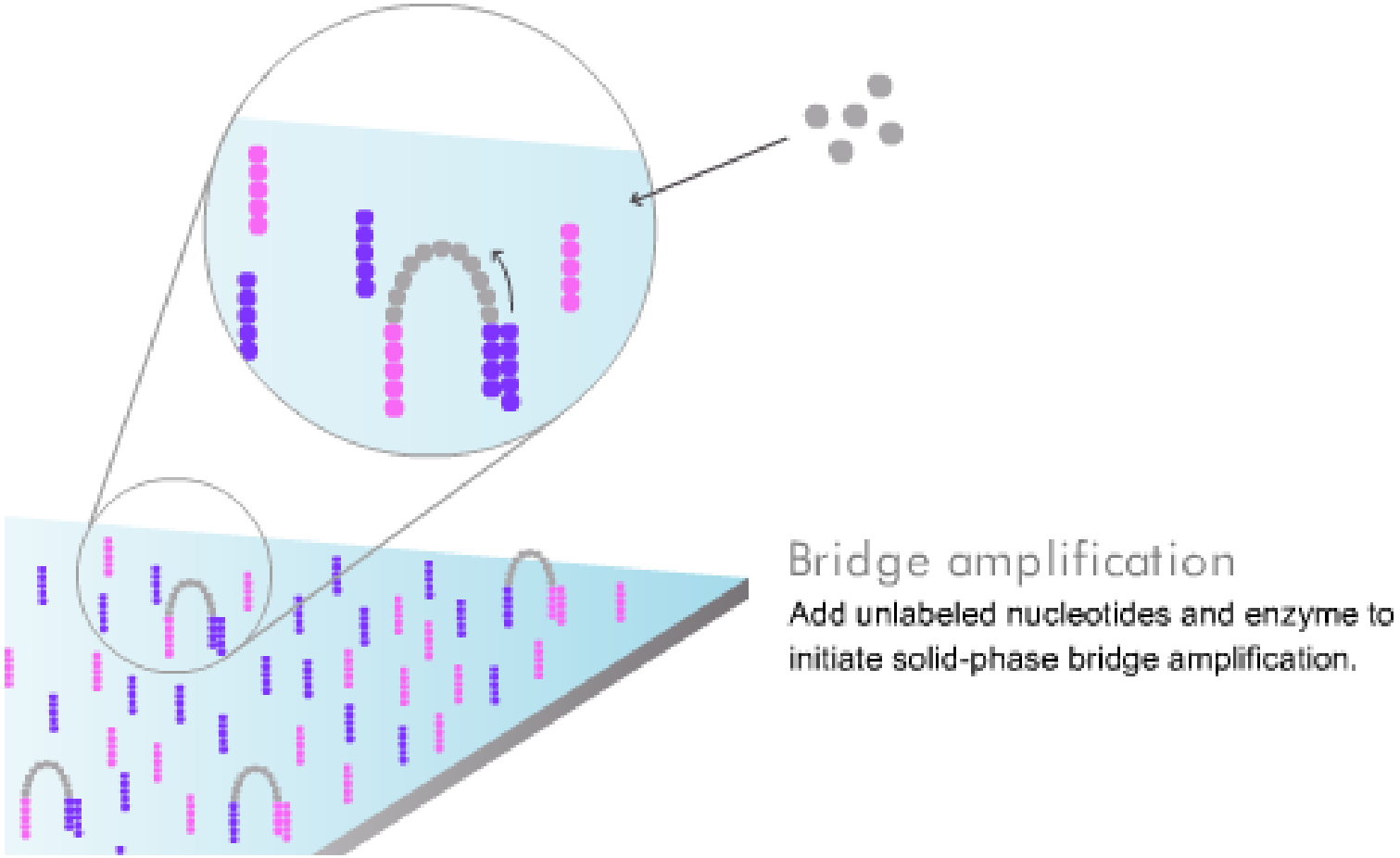


Attach DNA to surface

Bind single stranded fragments randomly to the inside surface of the flow cell channels.

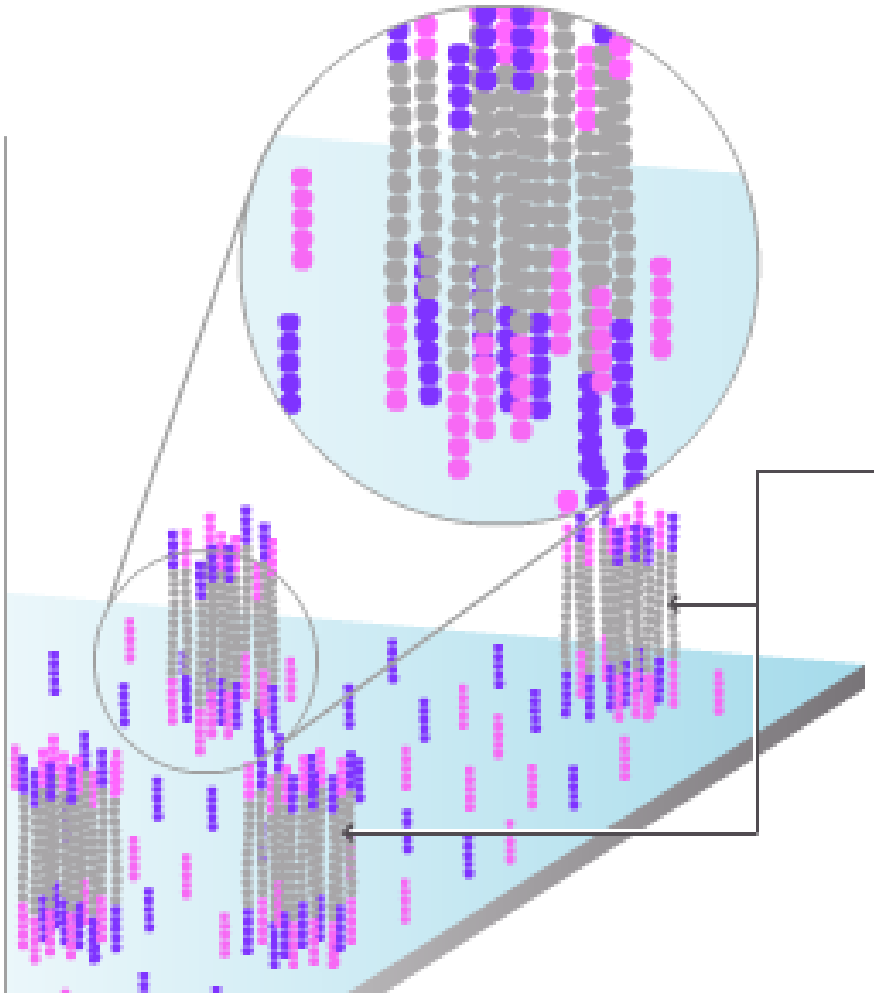
Solexa technology

step 3 - amplify DNA to immobilized clusters



Solexa technology

step 3 - amplify DNA to immobilized clusters

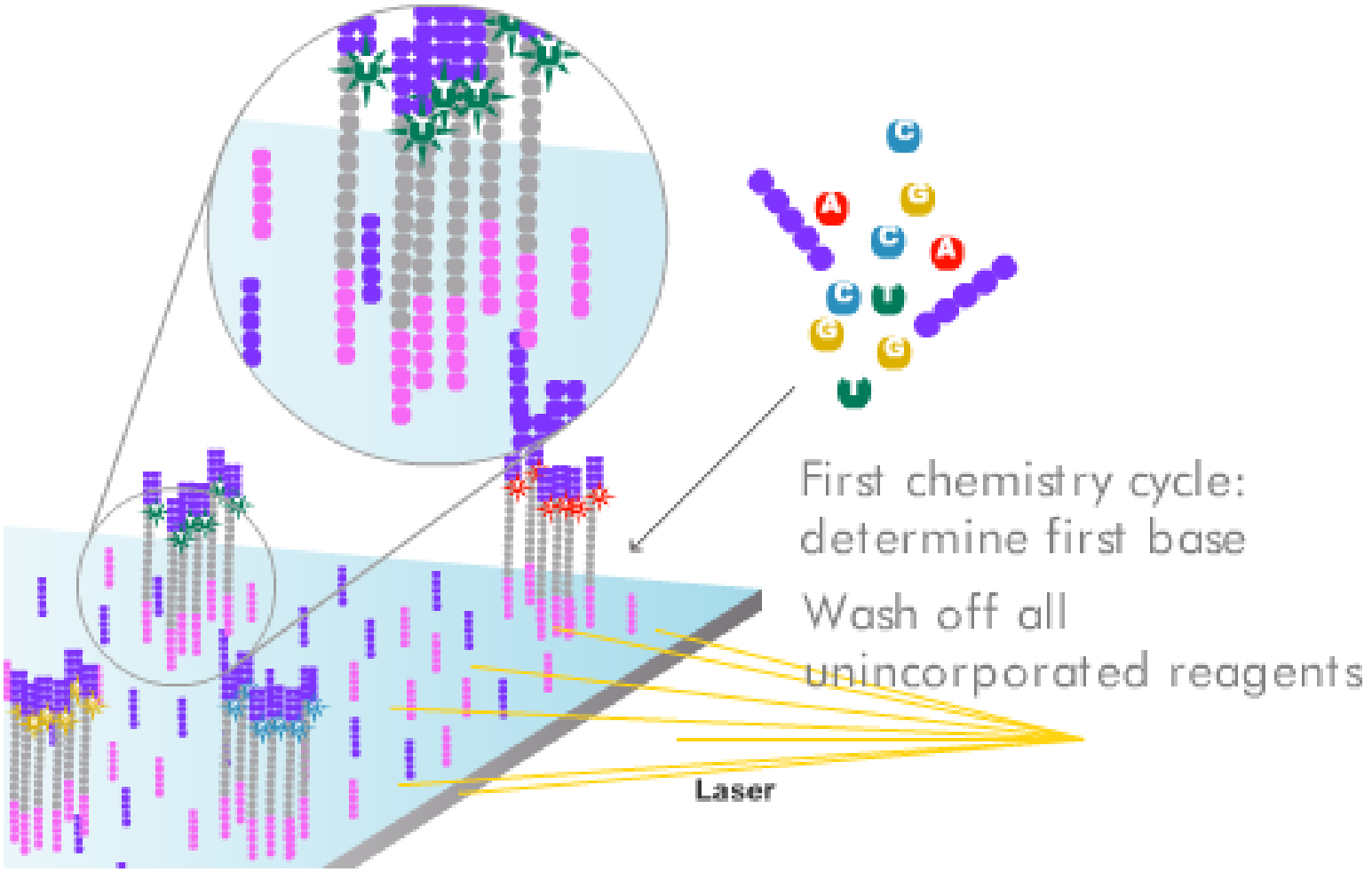


Completion of amplification

On completion, several million dense clusters of double stranded DNA are generated in each channel of the flow cell.

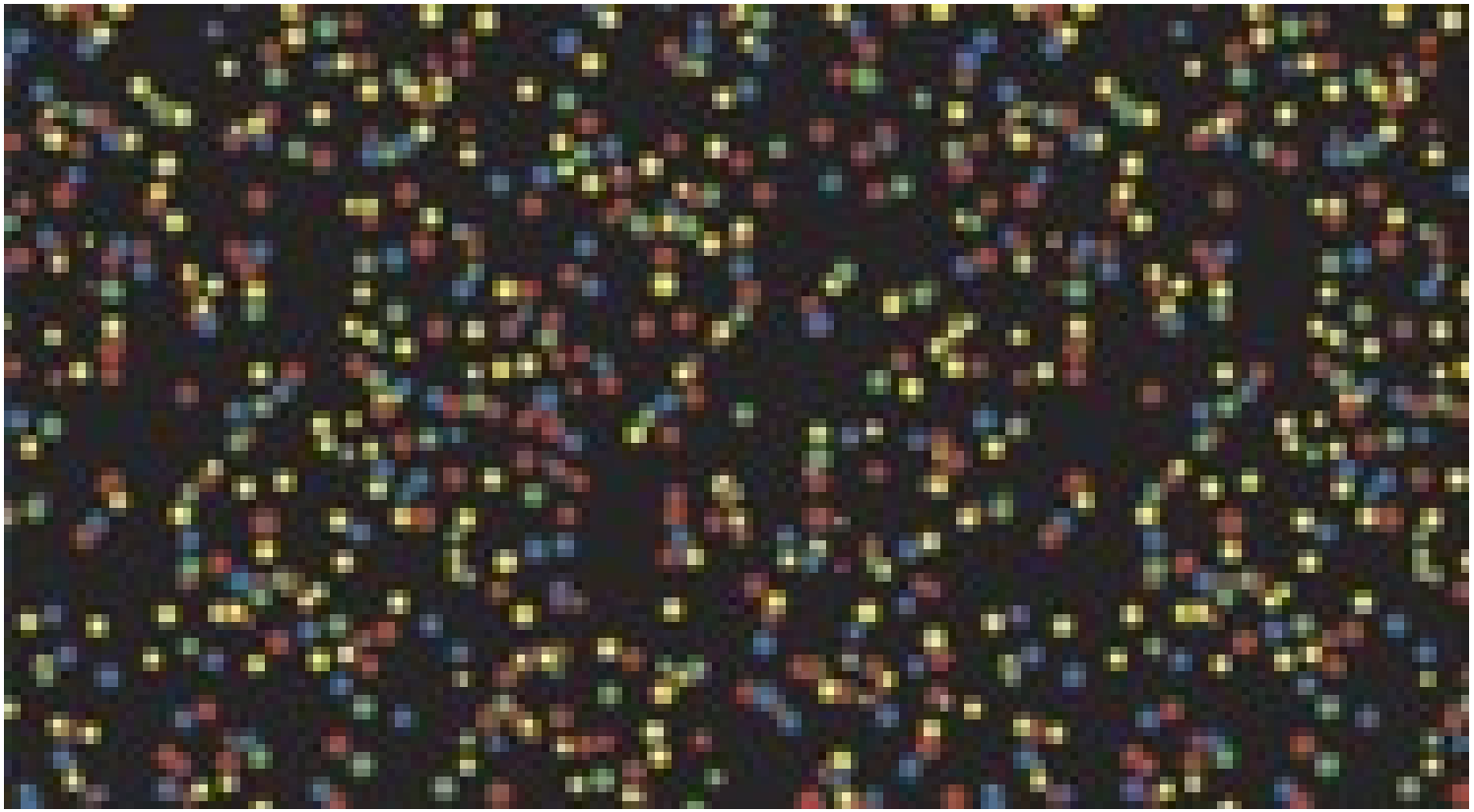
Solexa technology

step 4 - sequence one nucleotide at a time

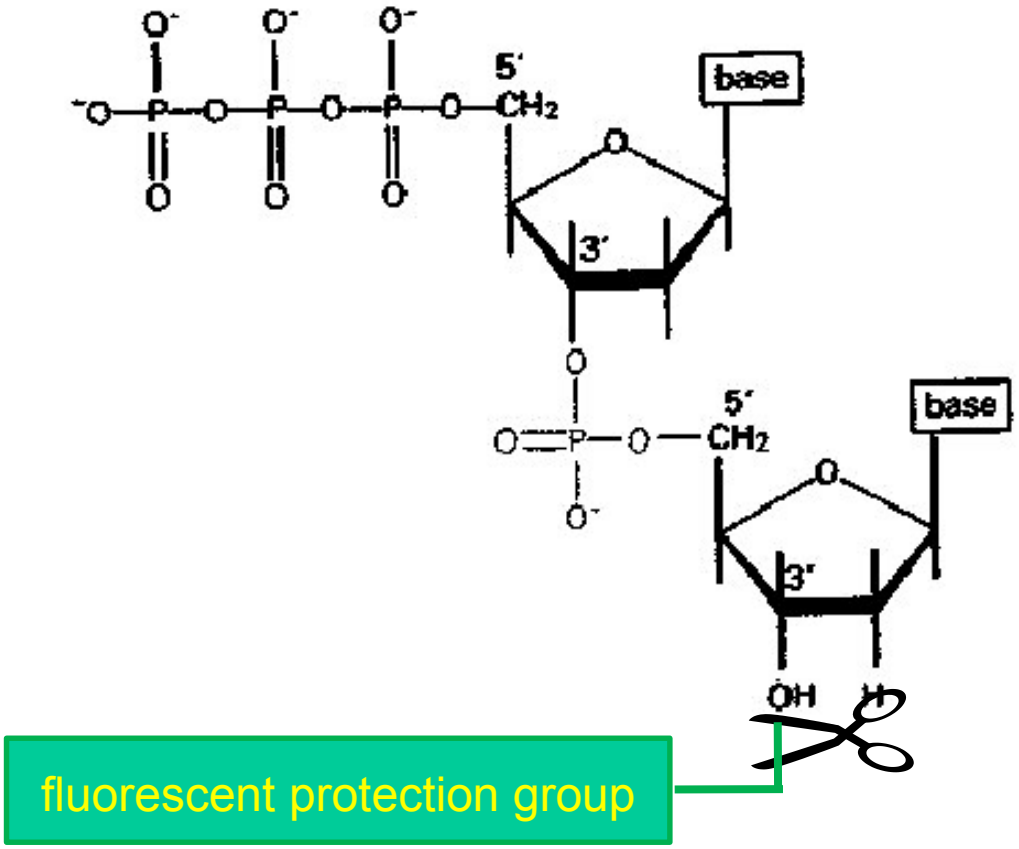


Solexa technology

step 4 - sequence one nucleotide at a time

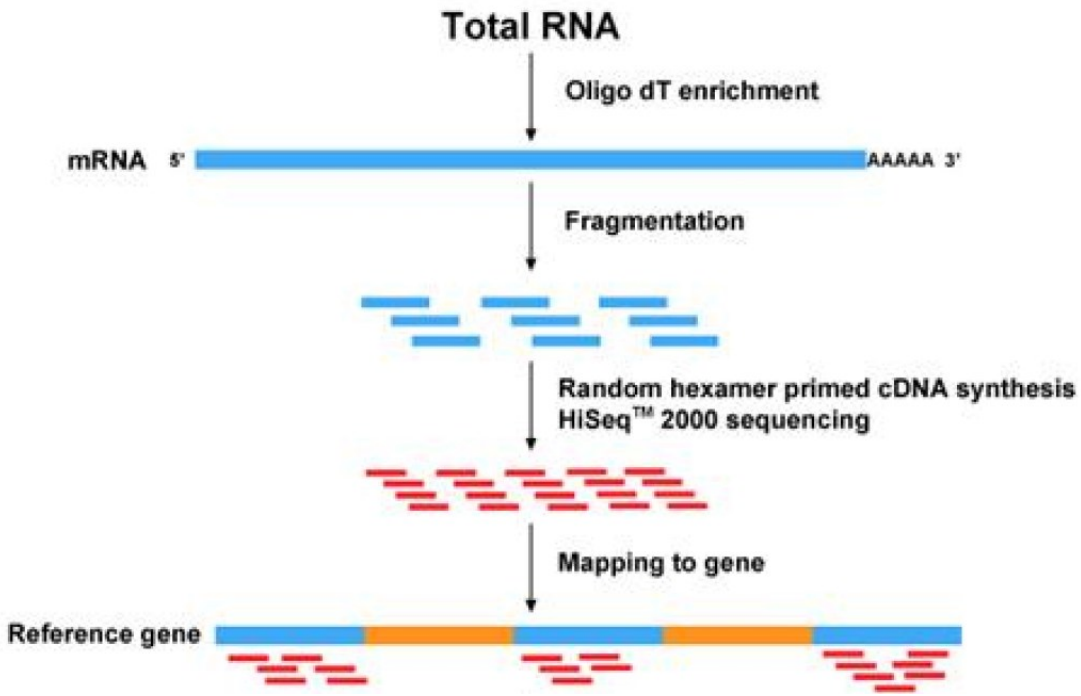


DNA polymerization requires the 3′ hydroxyl group



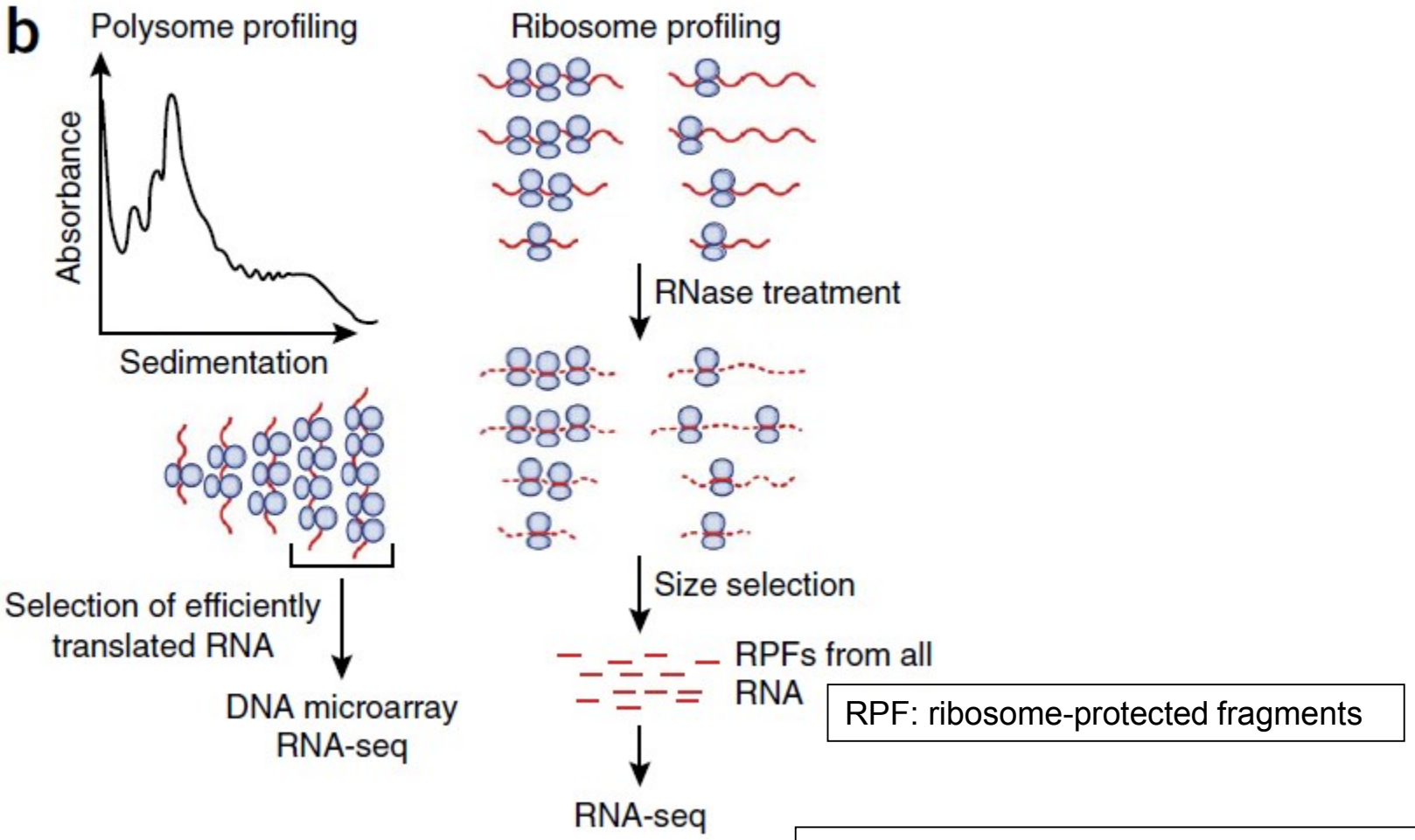
Next generation sequencing has enabled many innovative high throughput approaches

RNA-seq: Transcriptome; quantification of transcript levels, discovery of novel transcripts and transcript isoforms (alternative splicing).



Next generation sequencing has enabled many innovative high throughput approaches

Polysome and ribosome profiling: Translatome

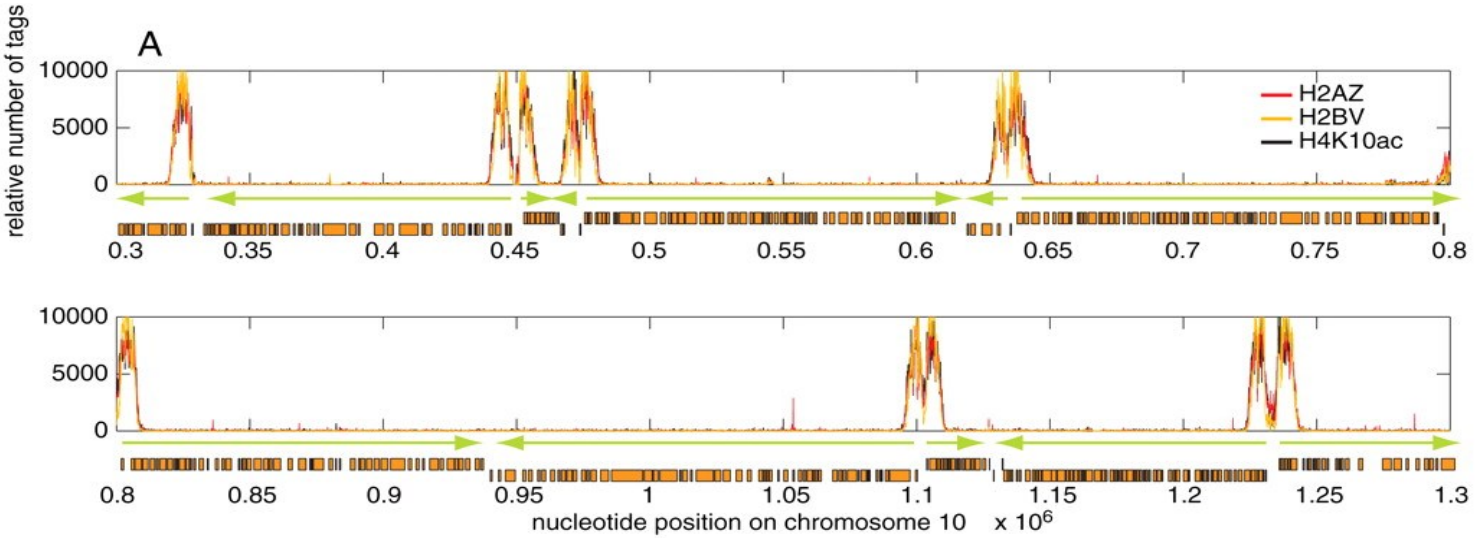


RPF: ribosome-protected fragments

From Piccirillo *et al*., 2014, *Nature Immunology*

Next generation sequencing has enabled many innovative high throughput approaches

Chromatin Immunoprecipitation (ChIP)-seq: Genome-wide binding sites of a histone, transcription factor, RNA polymerase, DNA repair enzyme, etc.



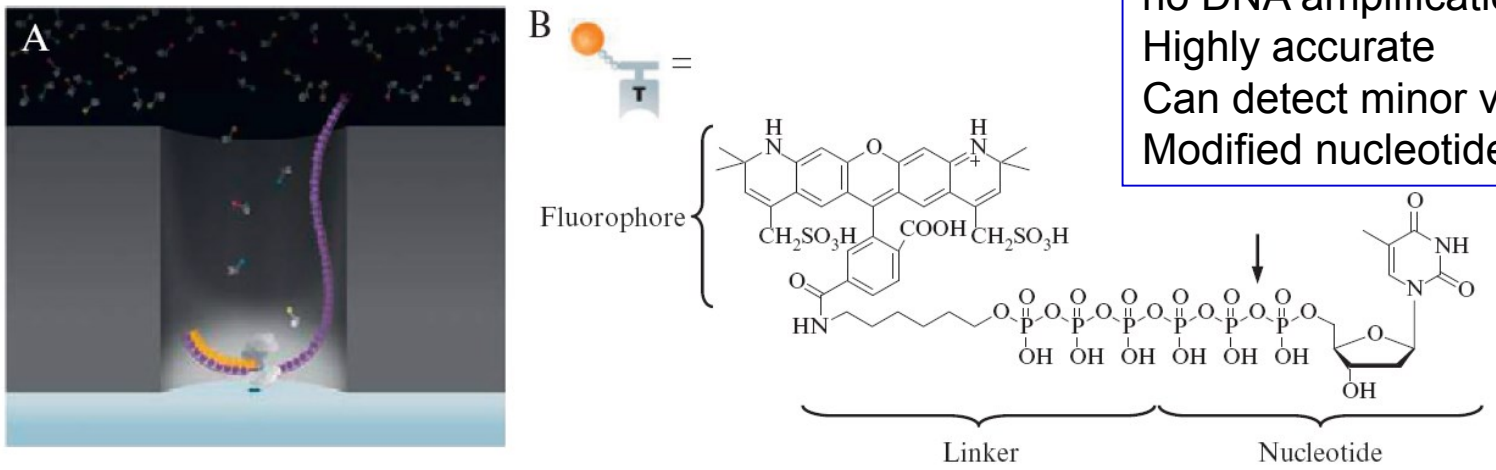
From Siegel *et al*., 2009, Genes & Dev

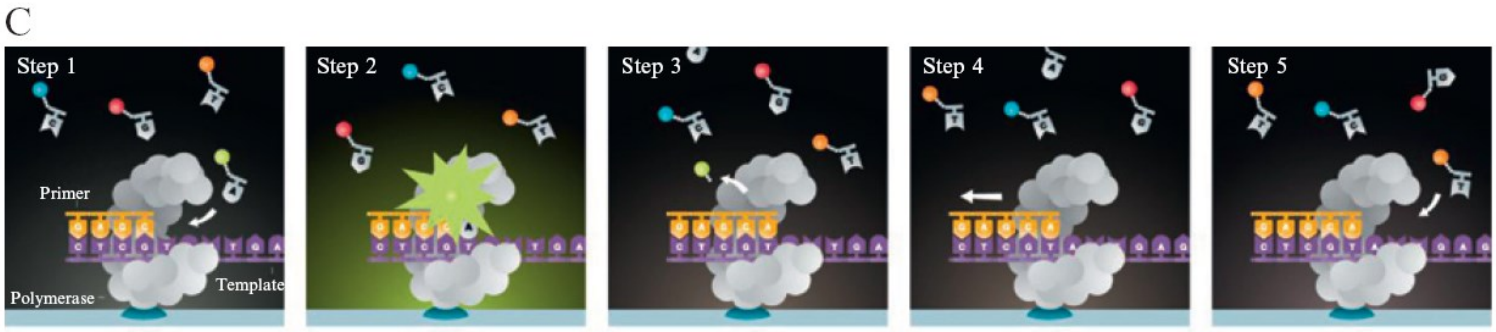
ChIP-seq revealed that open chromatin marks (modified histones and histone variants) identified regions of transcription initiation in *Trypanosoma brucei*

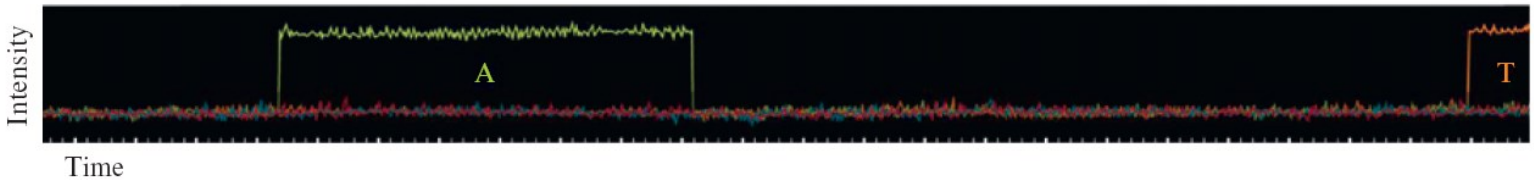
Third generation (single molecule) sequencing

Pacific Biosciences Single molecule real time (SMRT) technology

Reads: 3-5 kb, up to >20kb
no DNA amplification
Highly accurate
Can detect minor variants, <1%
Modified nucleotide detection







See video http://www.pacb.com/smrt-science/smrt-sequencing/

Nanopore sequencing

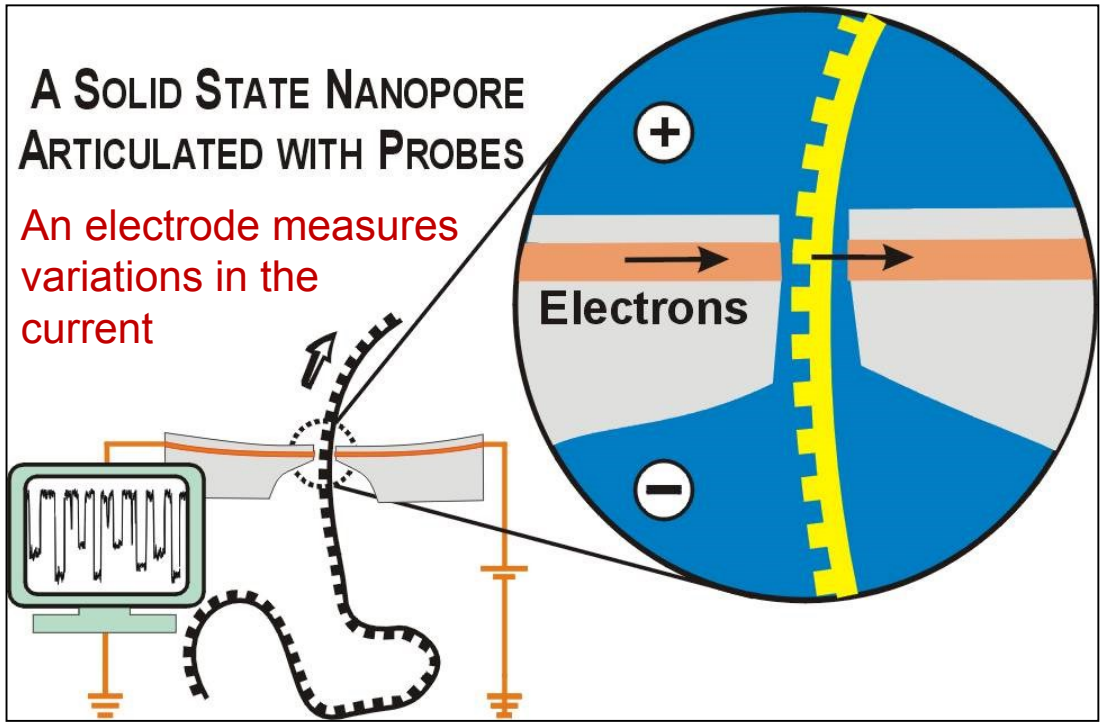
Bolisetty *et al. Genome Biology* (2015) 16:204
DOI 10.1186/s13059-015-0777-z

Genome Biology

METHOD Open Access

Determining exon connectivity in complex mRNAs by nanopore sequencing

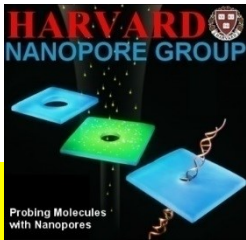
Mohan T. Bolisetty[1,2†], Gopinath Rajadinakaran[1†] and Brenton R. Graveley[1*]



No polymerization
No library preparation
No light detection

Sequence accuracy not high
Not yet enough reads for transcriptomes

Oxford Nanopore Technologies®



THE END

Any questions – contact me !

Arthur Gunzl

office R1144 (400 Farmington Ave)

phone 8878

E-mail gunzl@uchc.edu