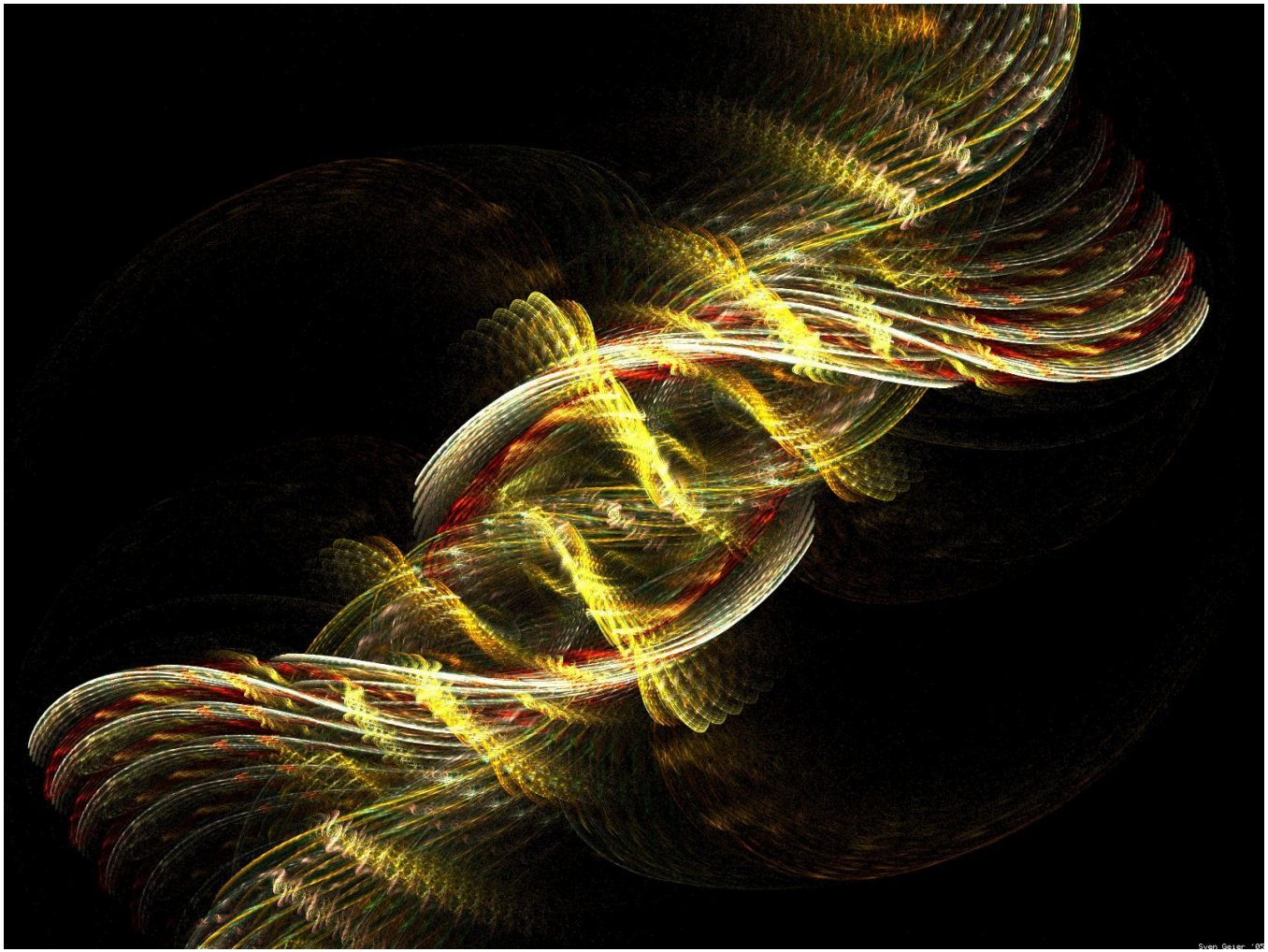
GENOMIKA

genomszekvenálási stratégiák



ELTE TTK Genetikai Tanszék

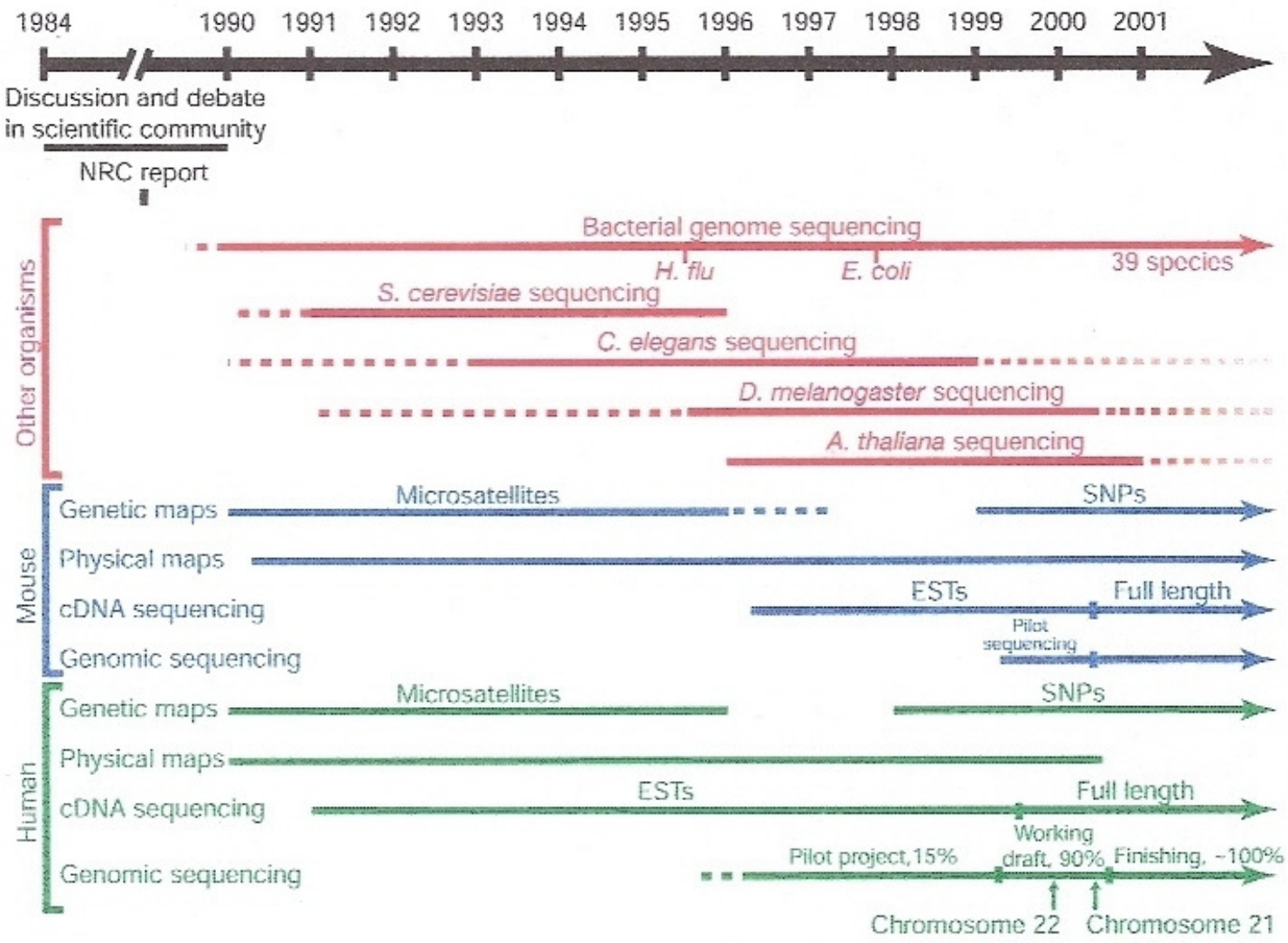
A genetikai információtartalom megismerése

- **I. negyed:** sejten belüli funkciók, kromoszómák (Miescher, Flemming, Mendel, Sutton, Morgan, stb.)
- **II. negyed:** az öröklődés és a sejtműködés molekuláris alapjai (DNS kettős hélix)
- **III. negyed:** az öröklődés és a sejtműködés biológiai mechanizmusai (transzkripció, transzláció, enzimek)
- **IV. negyed:** géntérképezés, gén- és genom szekvenálás, bioinformatika (**Genomika tudományág**)
- Az élő egyedtől a miniatűrön át a holisztikáig
- Genom szekvenálási projektek: **Genomika** és **Proteomika**

A humán genom szekvenálás első eredményei

- első gerinces genom, eukromatikus régió ~ 96 %-os fedettség
- nagymértékű variabilitás a különböző régiók, genetikai elemek és jellegek eloszlásában (pl. HOX klaszter – „repeat poor")
- ~ 30-40.000 gén, komplexitás és alternatív splicing
- komplex proteom, vertebrata-specifikus domén összeszerelés
- horizontális géntranszfer, transposable elemek inaktivációja
- kromoszóma szegmentális duplikációk (pericentromer, subtelomer)
- meiotikus mutációs ráta férfiakban és nőkben
- rekombinációs ráta eloszlás a kromoszóma karokon
- több mint 1 millió SNP, genome-wide linkage mapping

Genom projektek időskálán



Humán Genom Projekt - előzmények

- Első kezdeményezés: 1980-as évek eleje
- orvosbiológiai megközelítés, infrastruktúrális beruházás
- Folyamatban lévő genom szekvenálási projektek
- λ-fág, SV40, humán mitokondriális genom
- Genetikai és fizikai térképezések
- Botstein et al., 1980; Olson et al. és Sulston et al., 1986;
- DNS-szekvenálási technológia és bioinformatika
- shotgun sequencing, automated sequencing, ESTs, STSs,
- NRC Report 1988, US DOE, NIH,
- genetikai és fizikai térkép, parallel projektek modell organizmusokon, technológiai fejlesztések, bioetika

Universal Landmark

Sequence Tagged Site (STS) 1989

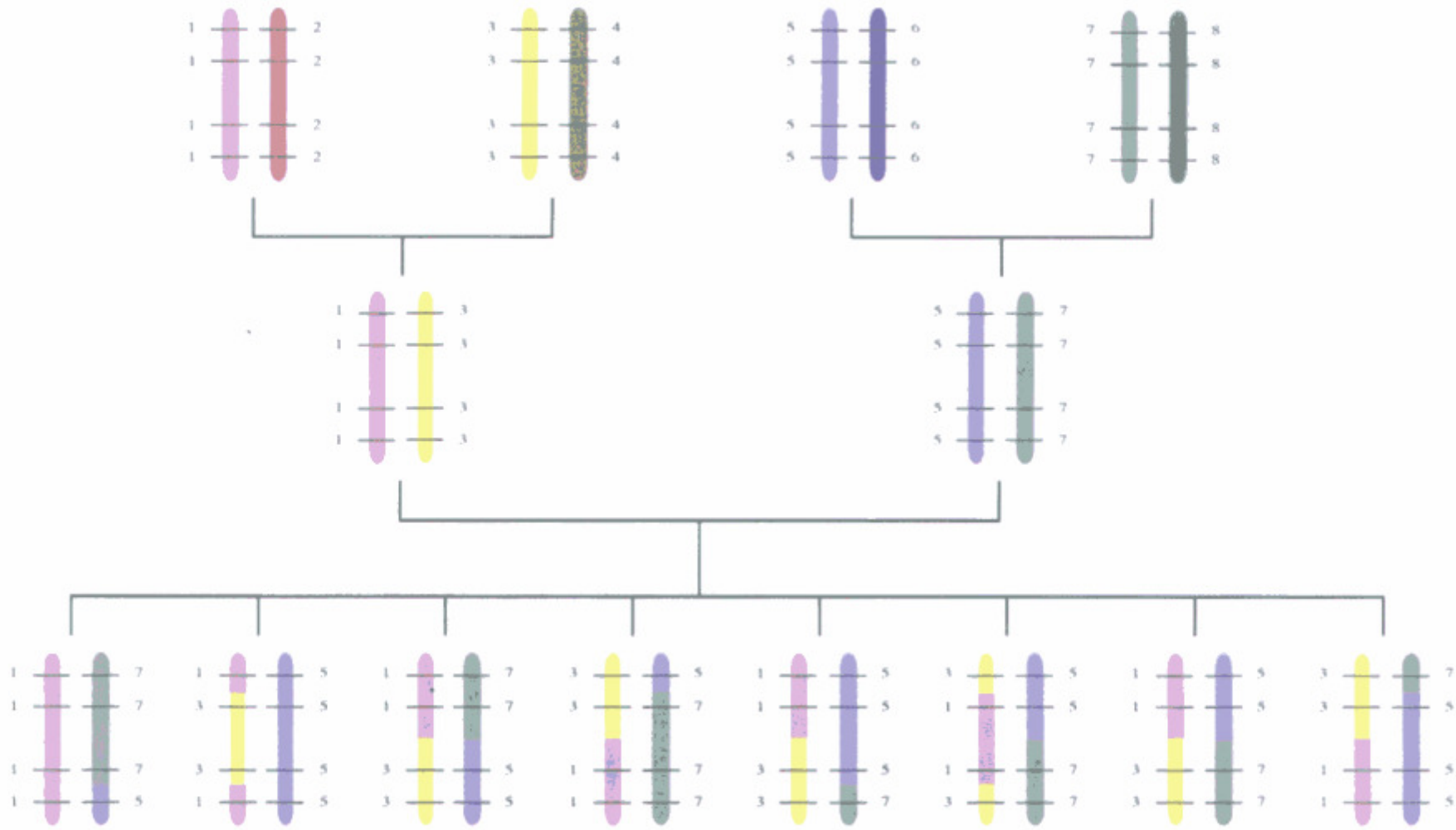
Replaces cloned DNA probe mapping landmarks with PCR assays.

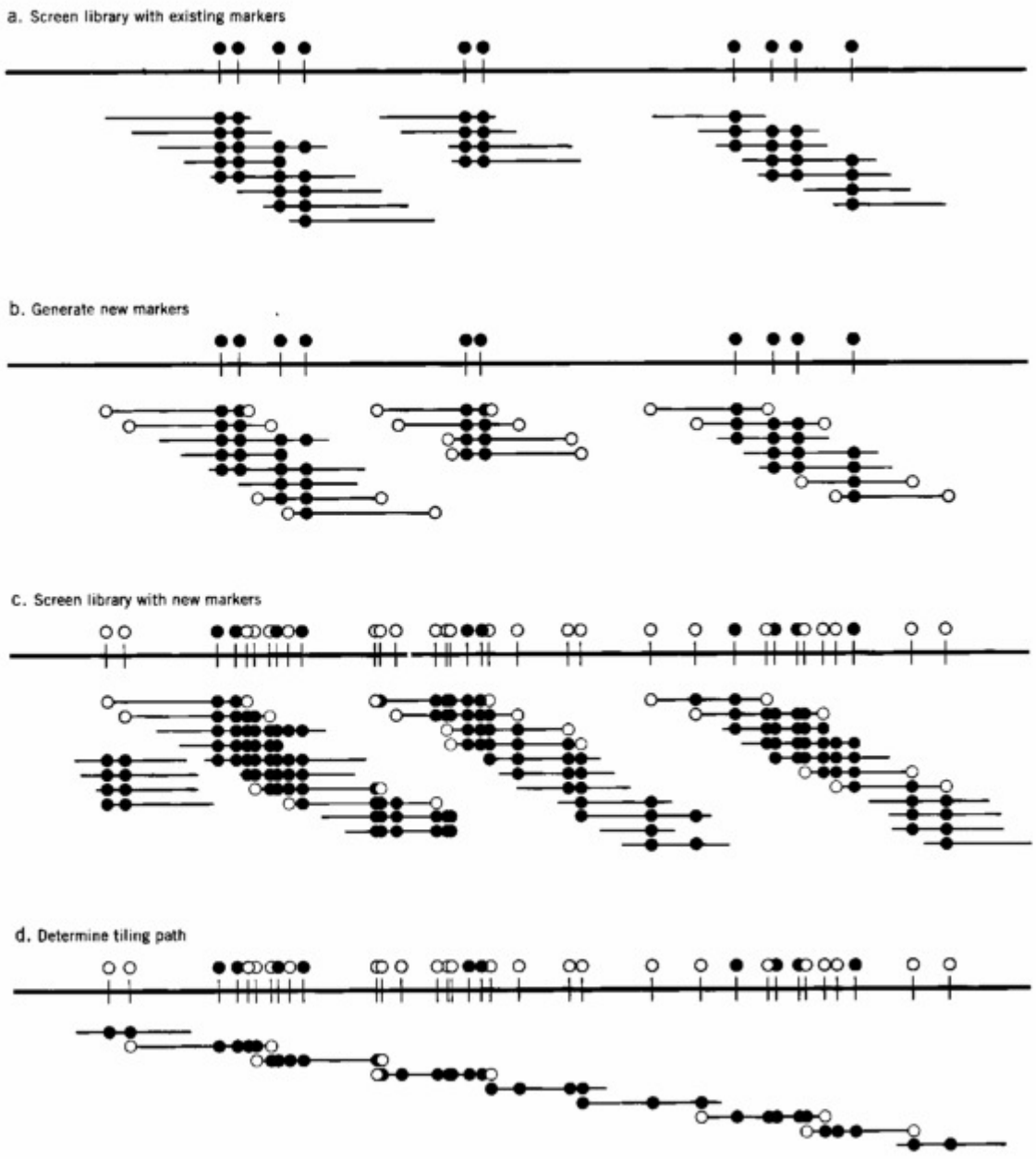
Each STS is uniquely described by a pair of oligonucleotides, a product size, and PCR reaction conditions. Can be stored and distributed electronically.

Enables merging of mapping data obtained from many labs using many different methods into a single consensus map of landmarks along a chromosome.

Eliminates the need for huge collections of cloned probe segments upon which prior maps depended.

Meiotic Breaks – Genetic Linkage Maps





Clone ends – Clone-based Physical Map

Humán Genom Projekt - célok

- A teljes emberi kromoszómális DNS szekvencia meghatározása
- Szekvencia adatbázisok kialakítása (bioinformatika)
- Az emberi genom összes génjének azonosítása és leírása (új gének, géntípusok meghatározása)
- DNS-szekvenálási technológia és adatfeldolgozás fejlesztése

Humán Genom Projekt
- résztvevők és módszerek

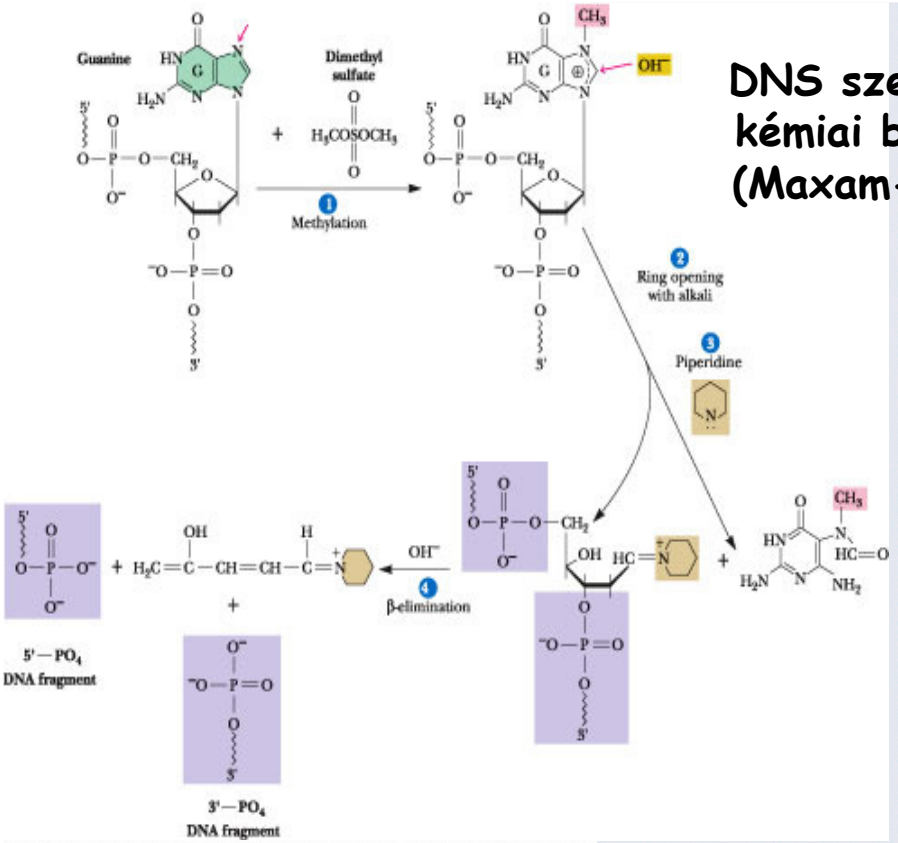
- **HUGO**: Human Genome Organization
- US DOE és NIH, UK MRC és WTSI, CEPH , FMDA, Japán, Európai Közösség (élesztő genom), Németország, Kína
- 1990-1995: genetikai és fizikai térképezés
- betegség gének, fizikai pontok fixálása, modell szervezetek
- large-scale sequencing: két fázisú „shotgun" szekvenálás
- 2001: draft genom szekvencia, 2003: teljes genom szekvencia

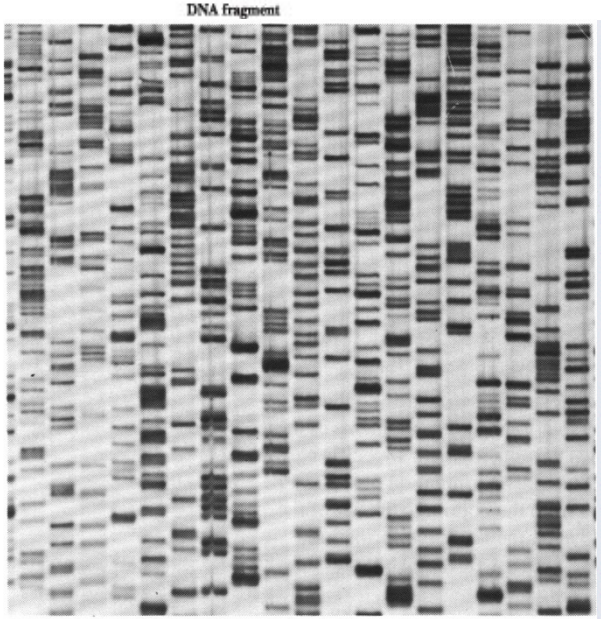
- **Celera Genomics**:
- Applied Biosystems., TIGR (C. Venter)
- 1998-2001: „whole genome shotgun"
- ABI PRISM 3700 DNA Analyzer

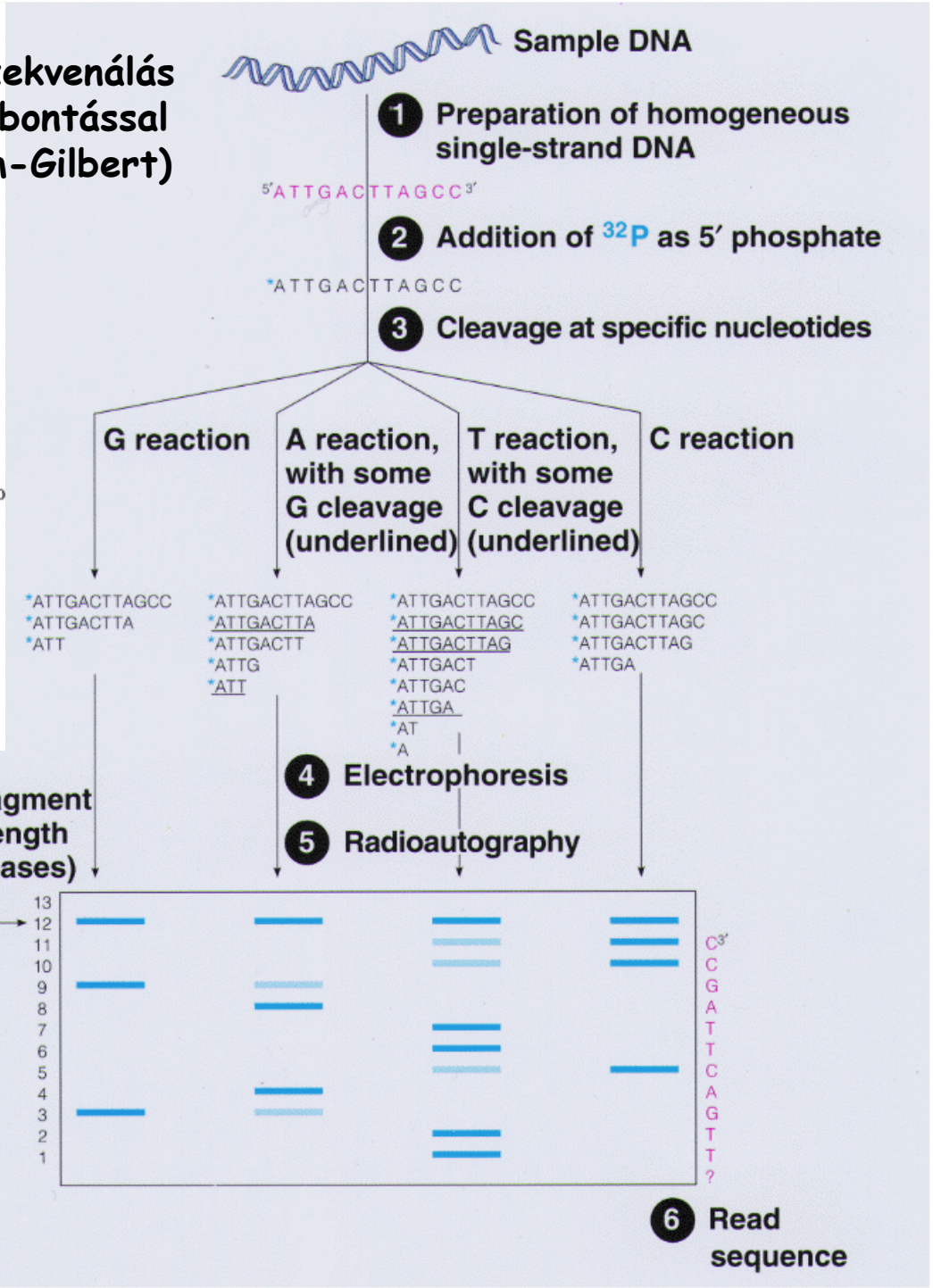
Technology speeds science. ABI sequencers at Venter Insitute, 2007.

Humán Genom Projekt

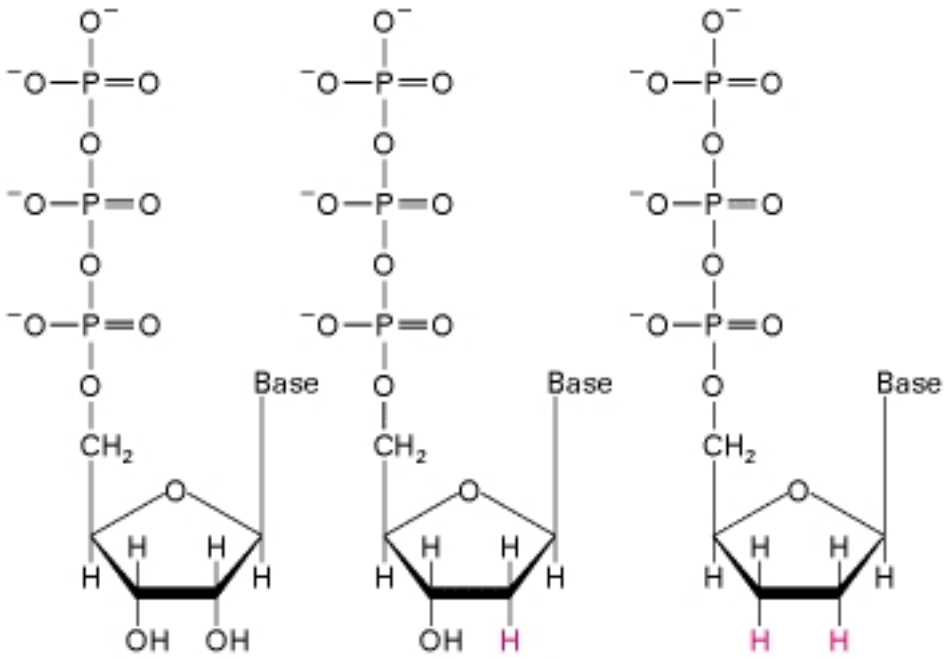
DNS szekvenálás kémiai bontással (Maxam-Gilbert)





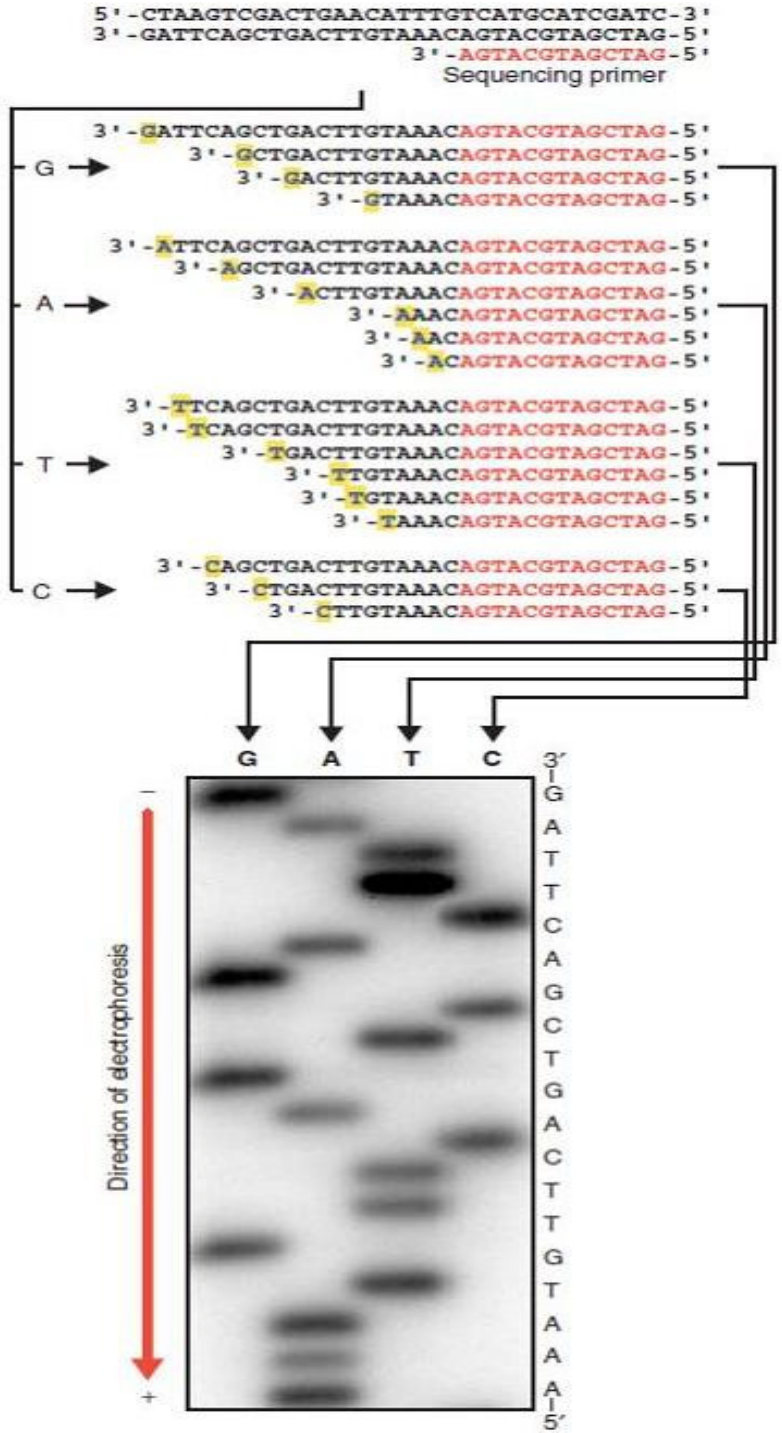


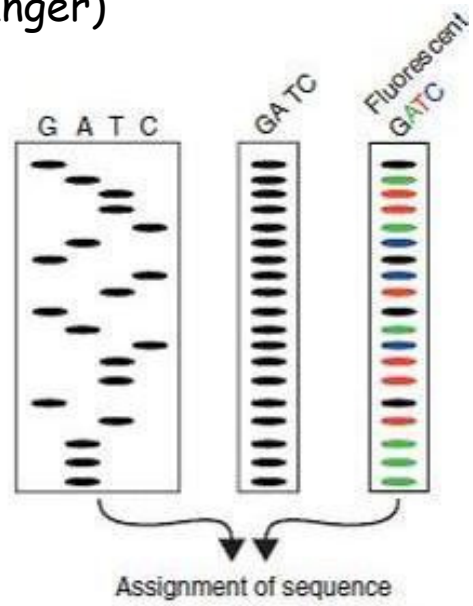
Sanger-szekvenálás

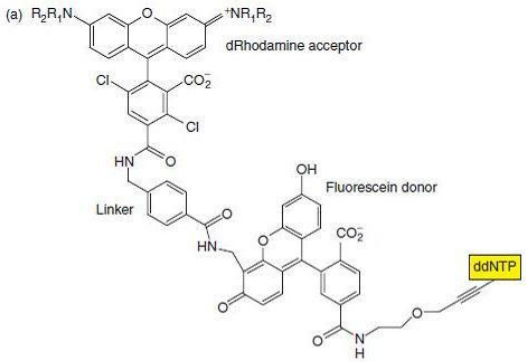


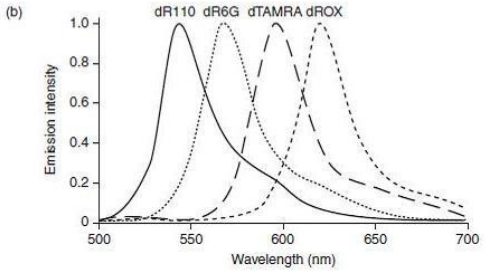
Molecular cell biology, Lodish

DNS szekvenálás lánctermináláss
(Sanger)

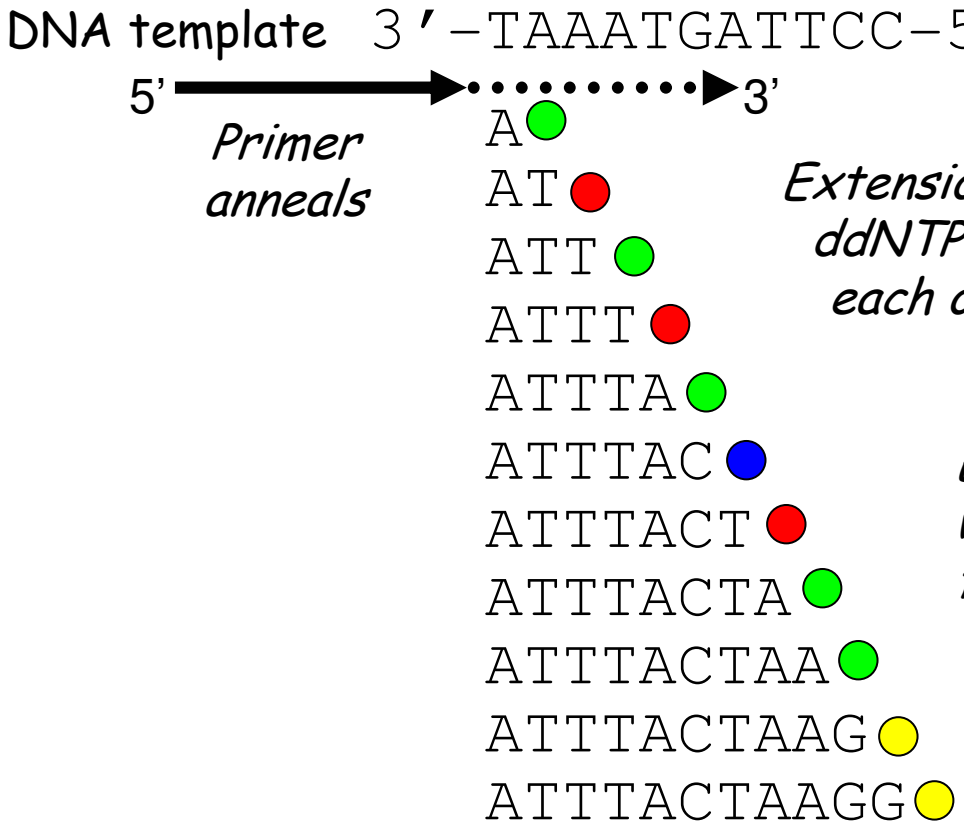






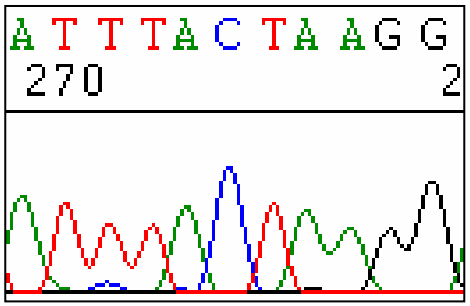


BigDye Terminator DNA Sequencing



Extension produces a series of ddNTP terminated products each one base different in length

Each ddNTP is labeled with a different color fluorescent dye



Sequence is read by noting peak color in electropherogram (possessing single base resolution)

Figure 10.5, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

DNS szekvenálás: Technológia és Bioinformatika

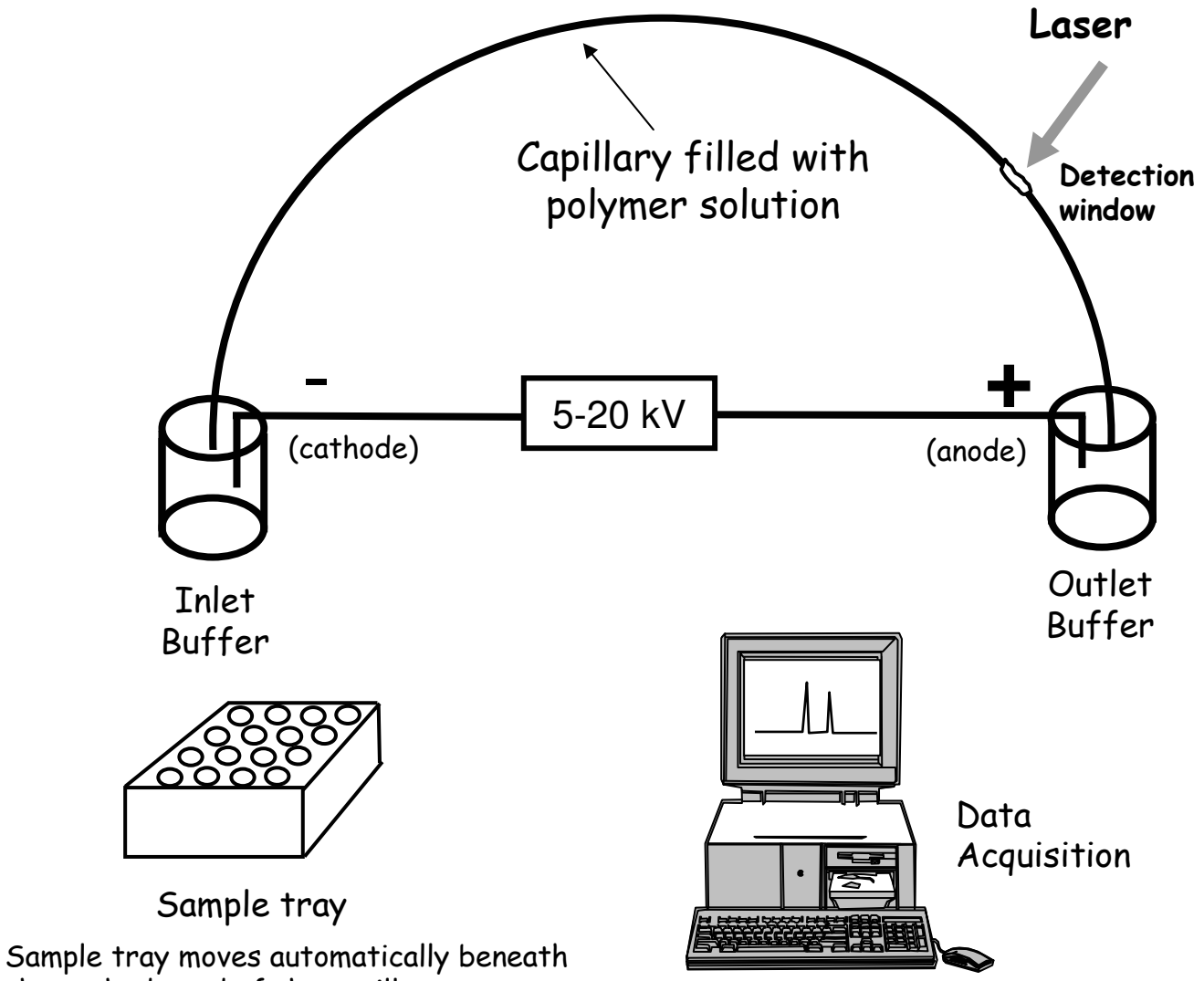


Figure 10.9, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Standard szekvenálás
650 bp leolvasás
2 h 30 min futási idő
– 16 kapilláris

1 nap – 100 000 bp

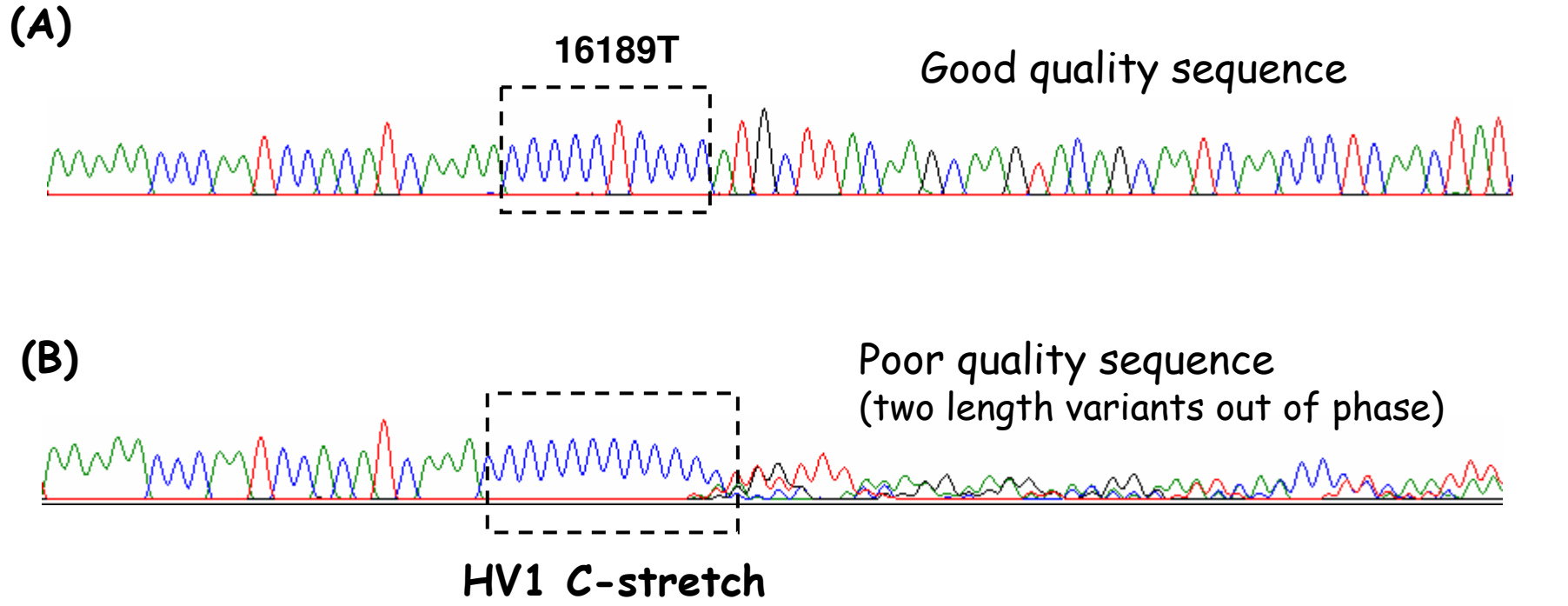


Figure 10.7, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

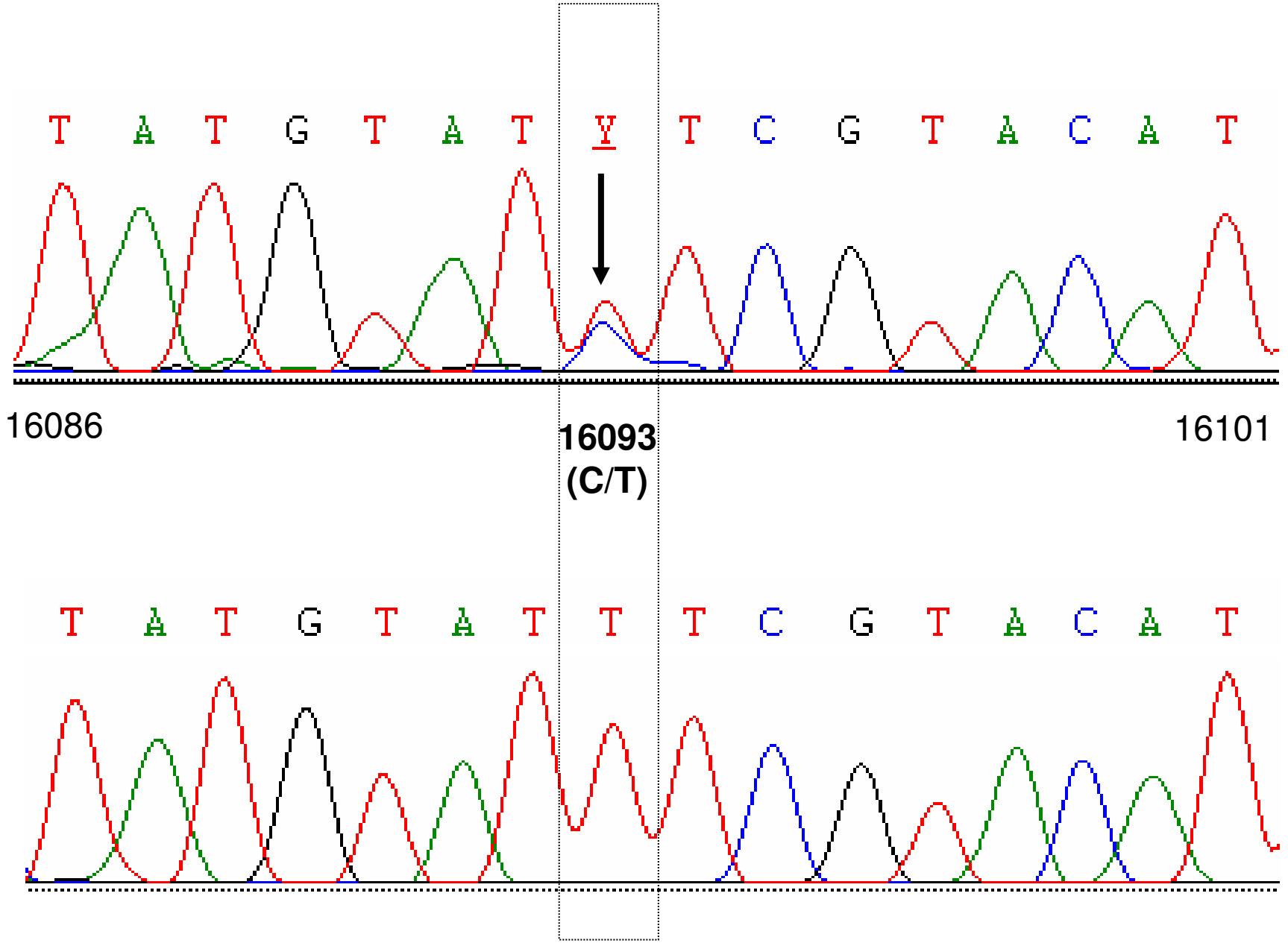


Figure 10.9, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

„shotgun" genom szekvenálási stratégiák

1. „chromosome walking"

2. Whole genome shotgun

3. Hierarchical shotgun

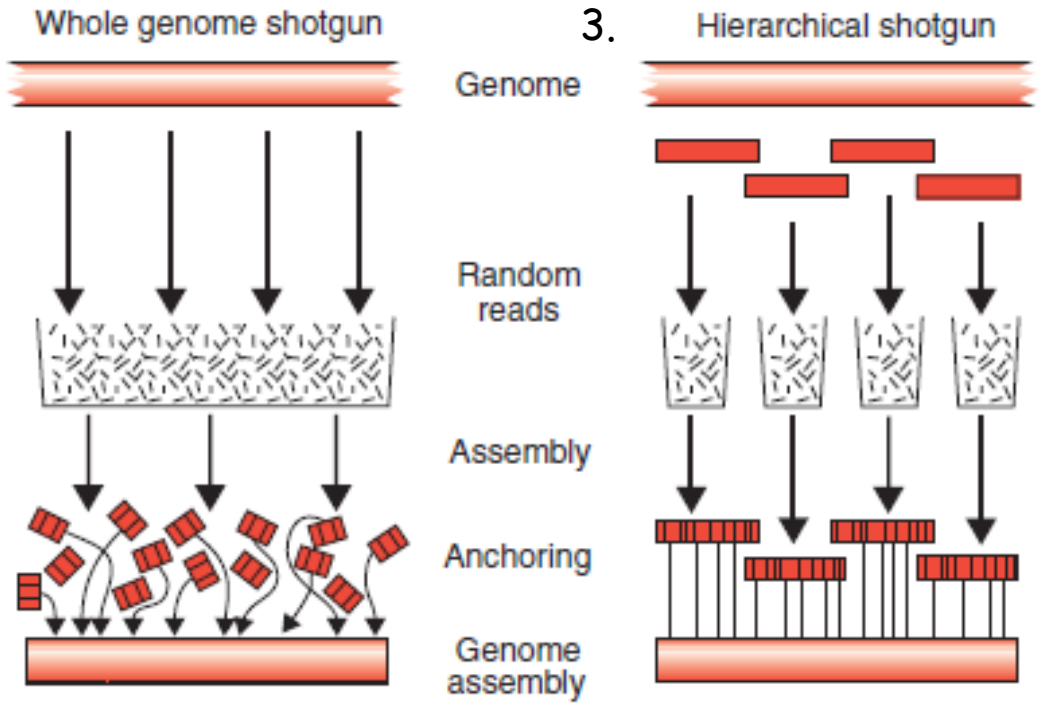


Figure 9.11. Assembling genomic data using the hierarchical and whole genome shotgun approaches. Adapted from Waterston, Lander and Sulston (2002), with permission

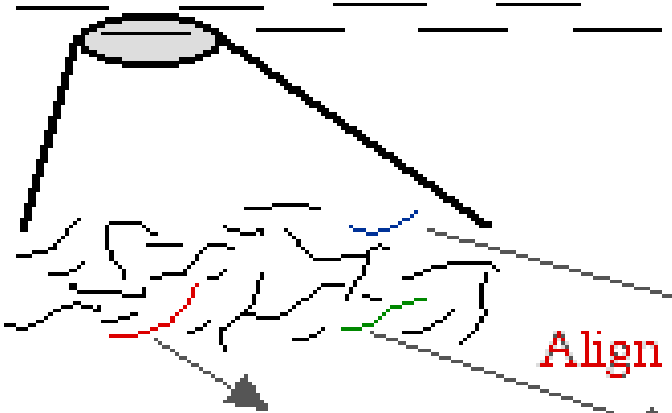
RJ Reece: Analysis of Genes and Genomes, 2004

Hierarchical Shotgun Sequencing Method

Genomic DNA

BAC Library

Create Contig Map



Sequence Each Contig
with Shotgun Approach

Align Contiguous Sequences

GCATTTCGAGTTACCTGGACAACCAGTG

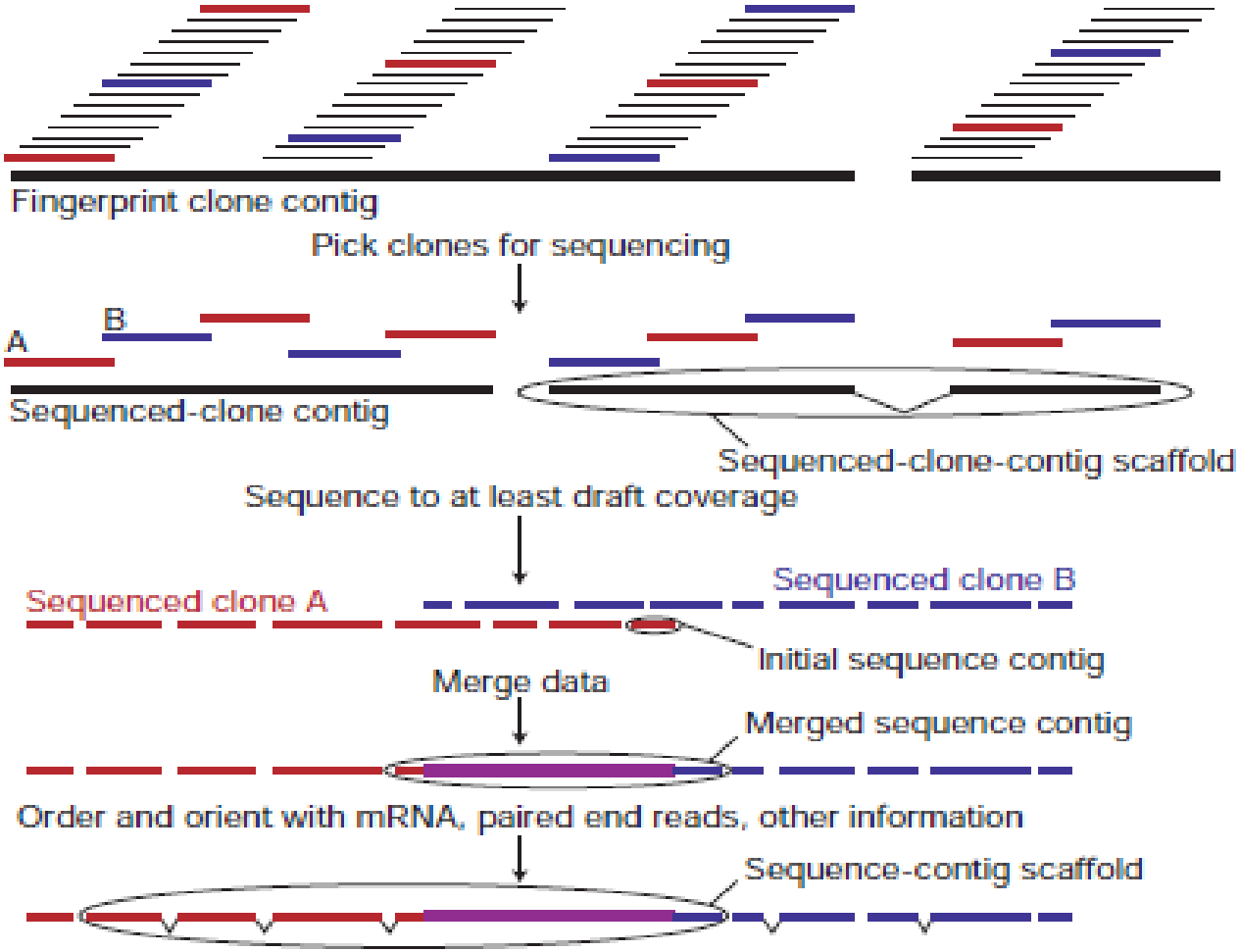
CCAGTGGTACTGAGGACGCAAGAGGCTTGA

GCTTGATTGGCCAATAATAGTATAT

GCATTTCGAGTTACCTGGACAACCAGTGGTACTGAGGACGCAAGAGGCTTGATTGGCCAATAATAGTATAT

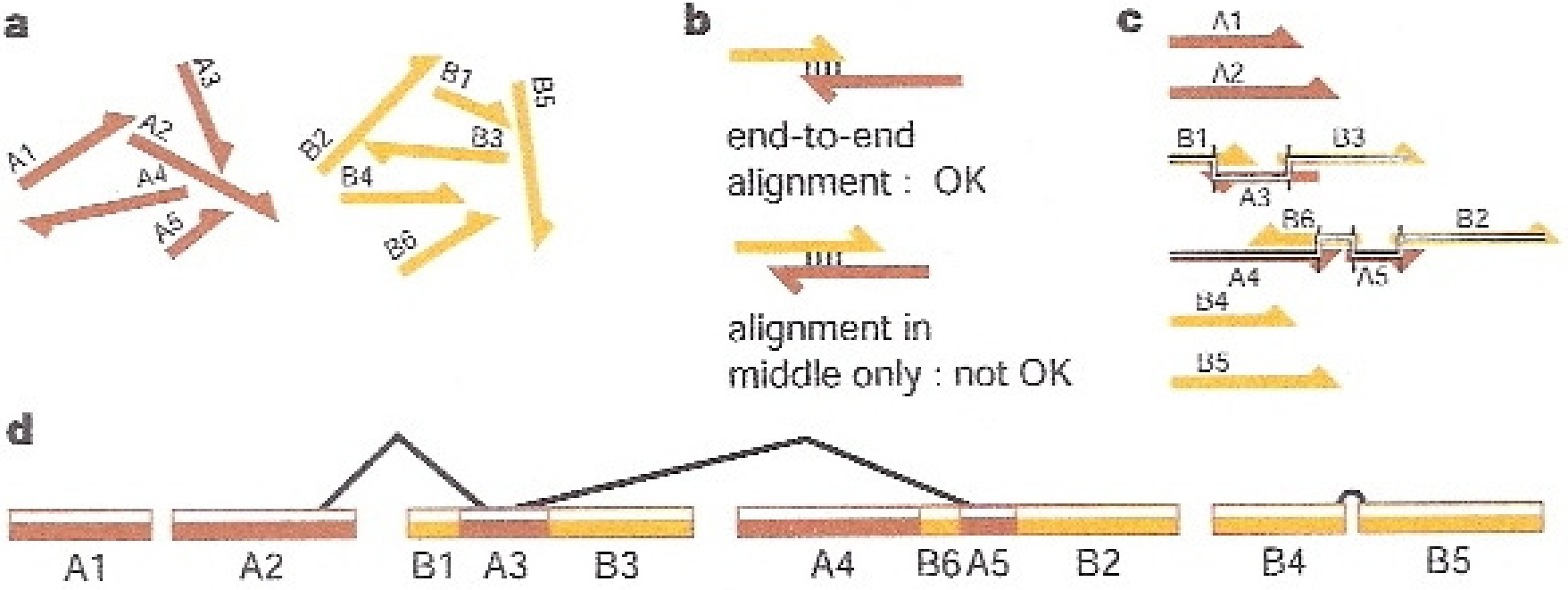
Generate Finished Sequence

Genom szekvenciaváz összeállítása



International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome, Nature 409, 860 (2001)

Genom szekvenciaváz összeállítása



International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome, Nature 409, 860 (2001)

Whole Genome Shotgun Sequencing Method

Genomic DNA

Sequence Each Fragment
with Shotgun Approach

GCATTTCGAGTTACCTGGACAACCAGTG

CCAGTGGTACTGAGGACGCAAGAGGCTTGA

GCTTGATTGGCCAATAATAGTATAT

Align Contiguous Sequences

GCATTTCGAGTTACCTGGACAACCAGTGGTACTGAGGACGCAAGAGGCTTGATTGGCCAATAATAGTATAT

Generate Finished Sequence

Teljes genom összeszerelés

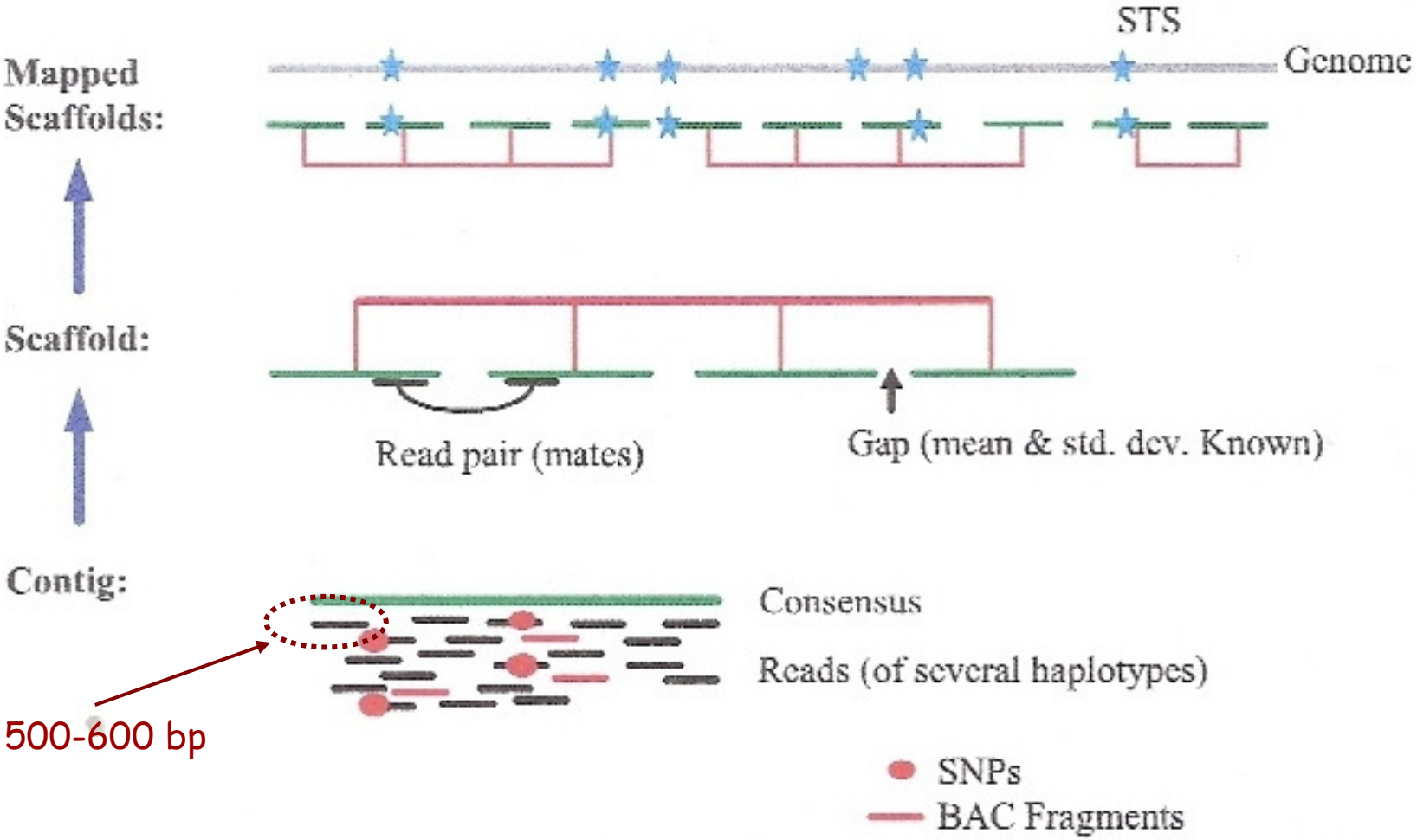


Fig. 3. Anatomy of whole-genome assembly. Overlapping shredded bactig fragments (red lines) and internally derived reads from five different individuals (black lines) are combined to produce a contig and a consensus sequence (green line). Contigs are connected into scaffolds (red) by using mate pair information. Scaffolds are then mapped to the genome (gray line) with STS (blue star) physical map information.

JC Venter, et al.: The Sequence of the Human Genome, Science 291, 1304 (2001)

STS genom térképezés

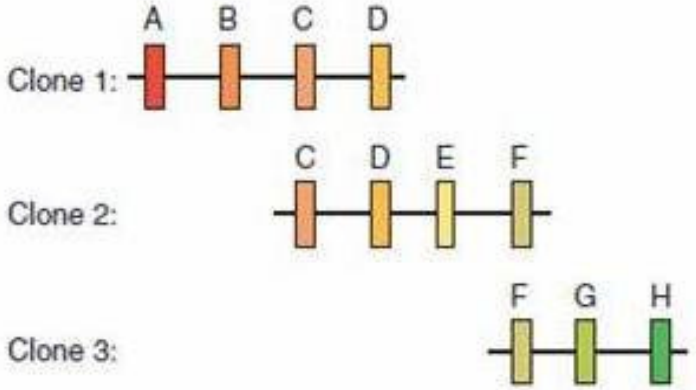


Figure 9.5. Aligning clones by STS mapping. Each clone contains several STSs. Clone 1 has four (A, B, C and D). Clone 2 also contains STSs C and D. Therefore clones 1 and 2 overlap with each other

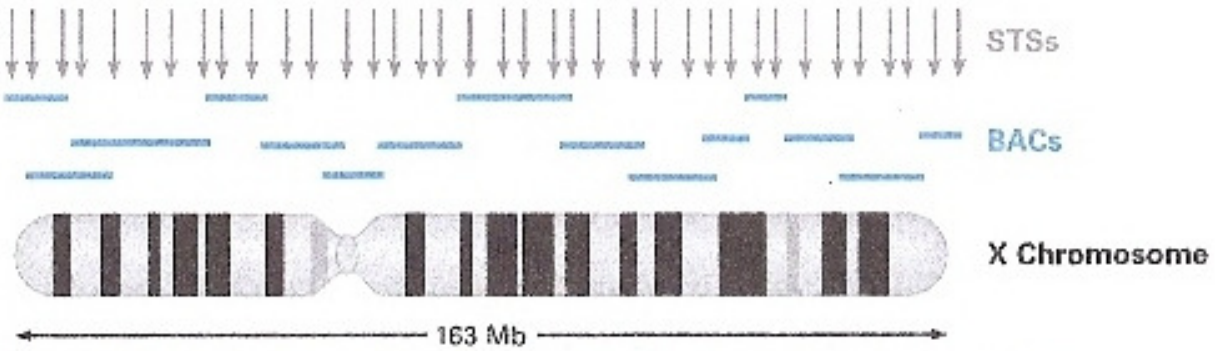


FIGURE 1.3 • Relationships of chromosomes to genome sequencing markers. The X chromosome is about 163 Mb in length. In this diagram, there are 16 overlapping BAC clones that span the entire length. In reality, 1,408 BACs were needed to span the X chromosome. Arrows (top) mark STSs scattered throughout the chromosome and on overlapping BACs.

Kromoszóma térképezés

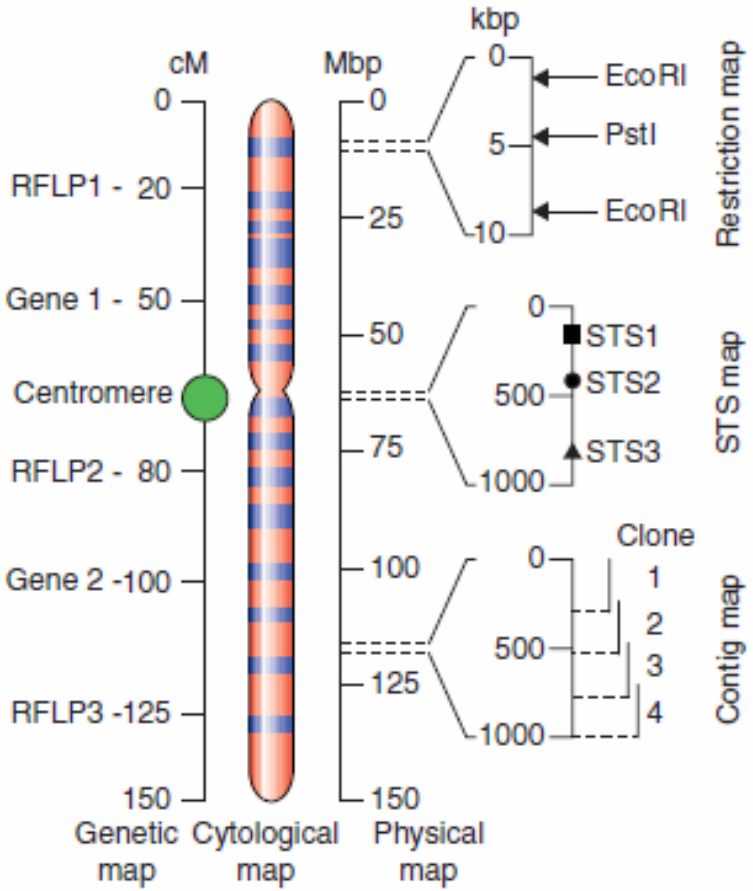


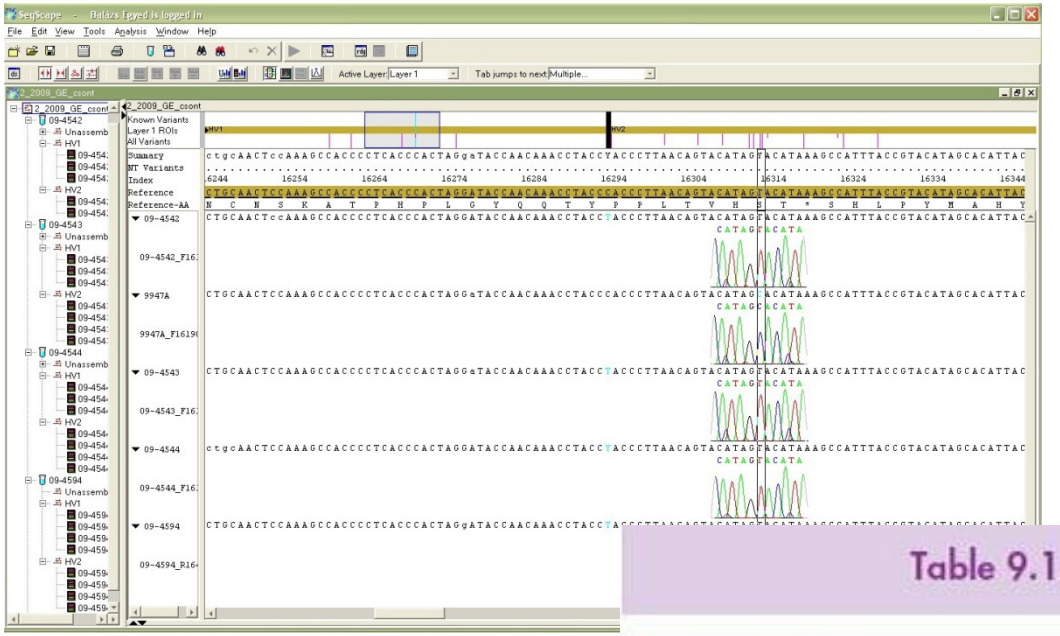
Figure 9.3. The different types of cytological, genetic and physical map of a chromosome. Genetic map distances are based on crossover frequencies and are measured in centiMorgans (cM), while physical distances are measured in megabase pairs (Mbp) or kilobase pairs (kbp)

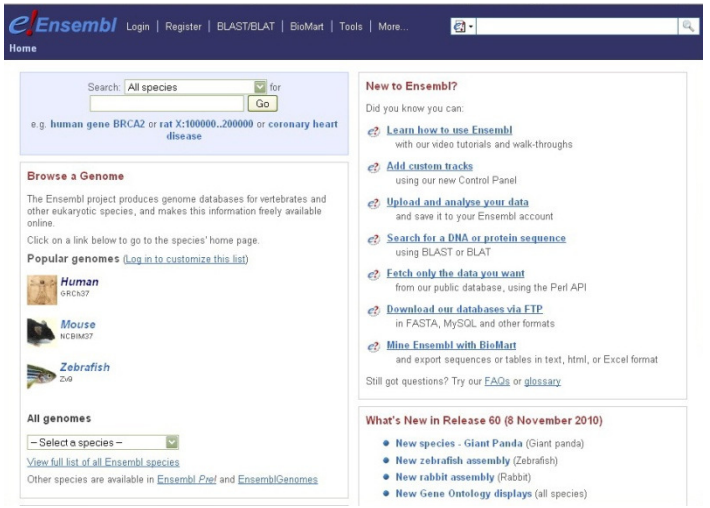
RJ Reece: Analysis of Genes and Genomes, 2004

Humán Genom Projekt - eredmények

- Tervezett 15 év helyett 2003-ban fejeződött be
- 2001: draft szekvencia publikálása (Science, Nature)
- Több személy genomjából nyert DNS szekvencia
- személyi DNS minták és sejtvonalak
- Megengedett hibaráta 1 / 10.000 (99,99 % pontosság)
- 4-5 X lefedettség, gapek lezárása (heterokromatin)
- Folytatódó genom projektek, annotáció, adatmegosztás:
- pl. Ensemble, Human Genome Diversity Project, stb.

Genom szekvenálás: Technológia és Bioinformatika





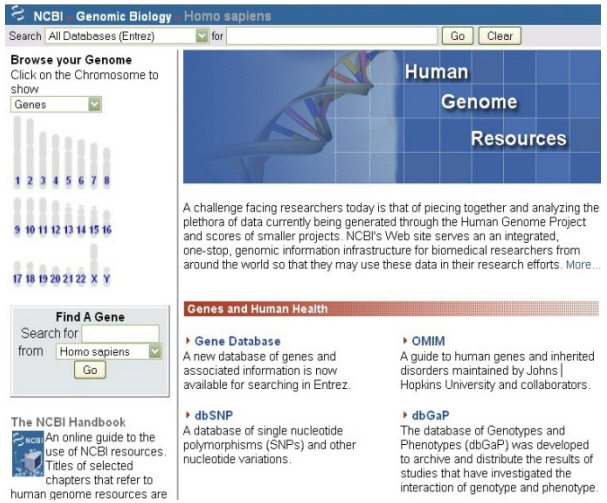


Table 9.1. Curated genome sequencing projects

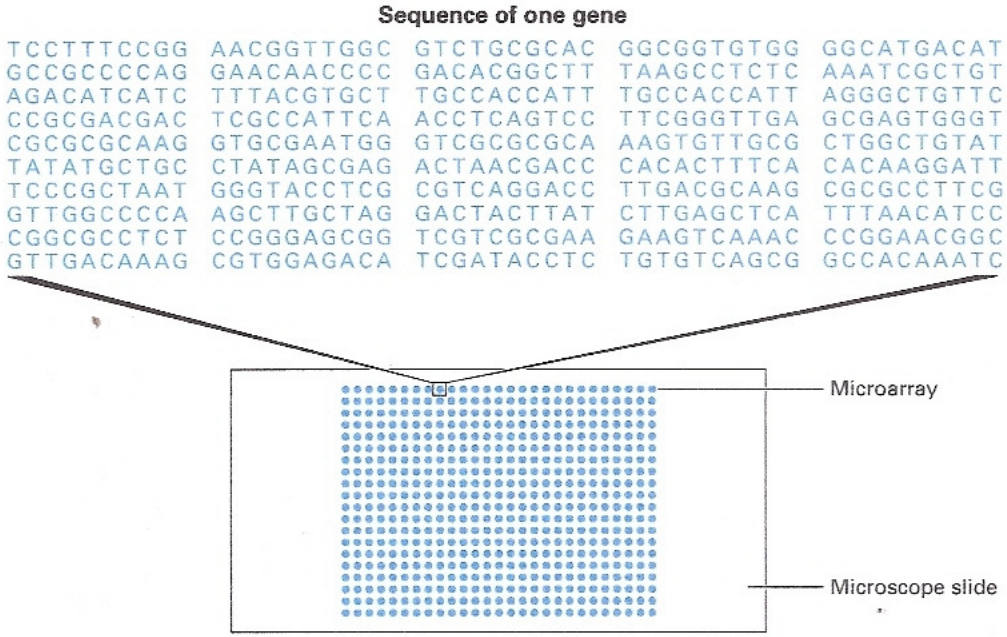
Organism (type)	Web site(s)
Escherichia coli (bacterium)	www.genome.wisc.edu
Bacillus subtilis (bacterium)	genolist.pasteur.fr/SubtiList
Saccharomyces cerevisiae (yeast)	genome-www.stanford.edu/Saccharomyces
Caenorhabditis elegans (nematode worm)	www.wormbase.org
Drosophila melanogaster (fruit fly)	flybase.bio.indiana.edu
Arabidopsis thaliana (plant)	www.arabidopsis.org
Mus musculus (mouse)	www.informatics.jax.org
Homo sapiens (human)	www.ncbi.nlm.nih.gov/genome/guide/human/

Genom szekvenálás: Technológia és Bioinformatika

- új algoritmusok és statisztikai eljárások az adatbázisokban rejlő információk, viszonyok, kapcsoltságok feltárására
- DNS és aminosav szekvencia-analízis, szekvencia-homológiák, protein domének és szerkezeti változatok
- a különböző típusú és eredetű információk menedzselése, az adatok kutatása és hozzáférhetősége (annotált genom szekvencia adatbázisok)


personal genome machine


HeliScope



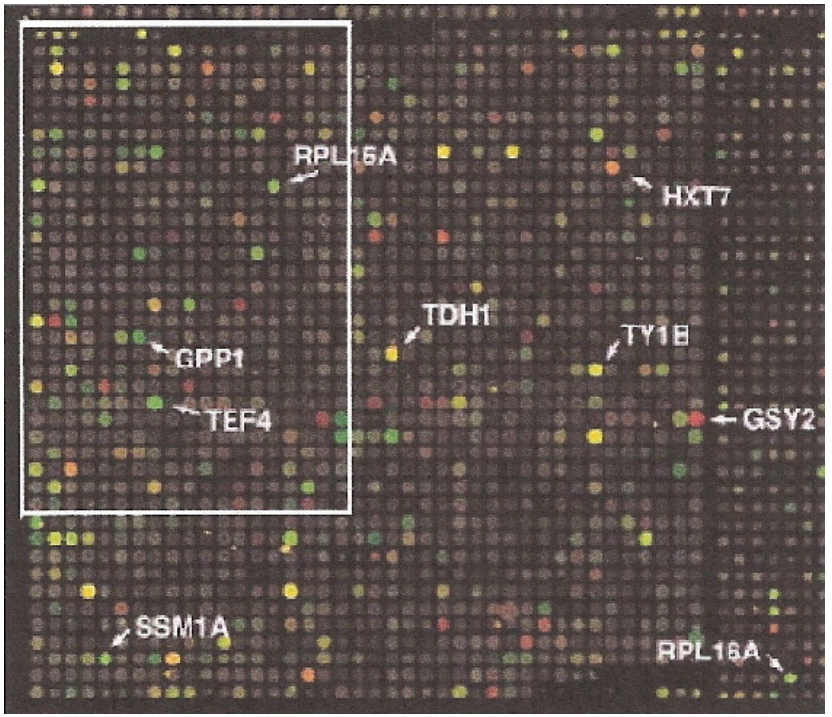
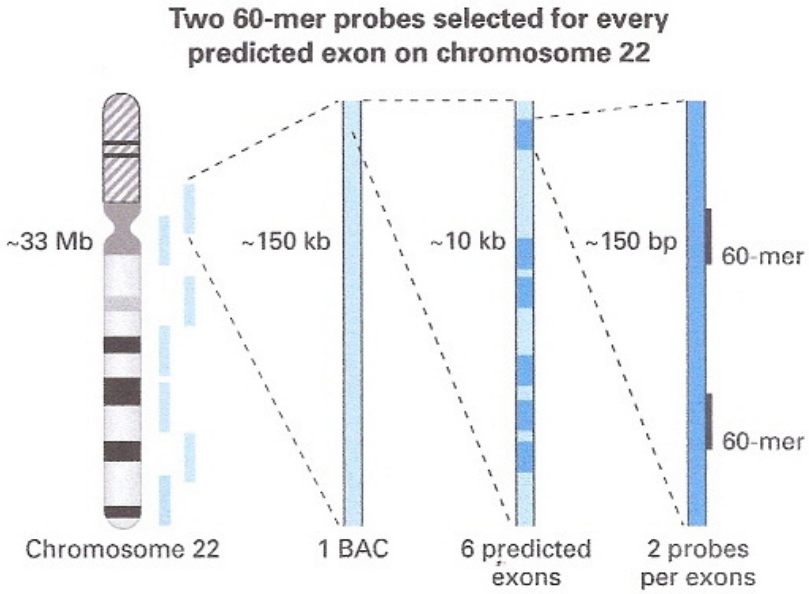
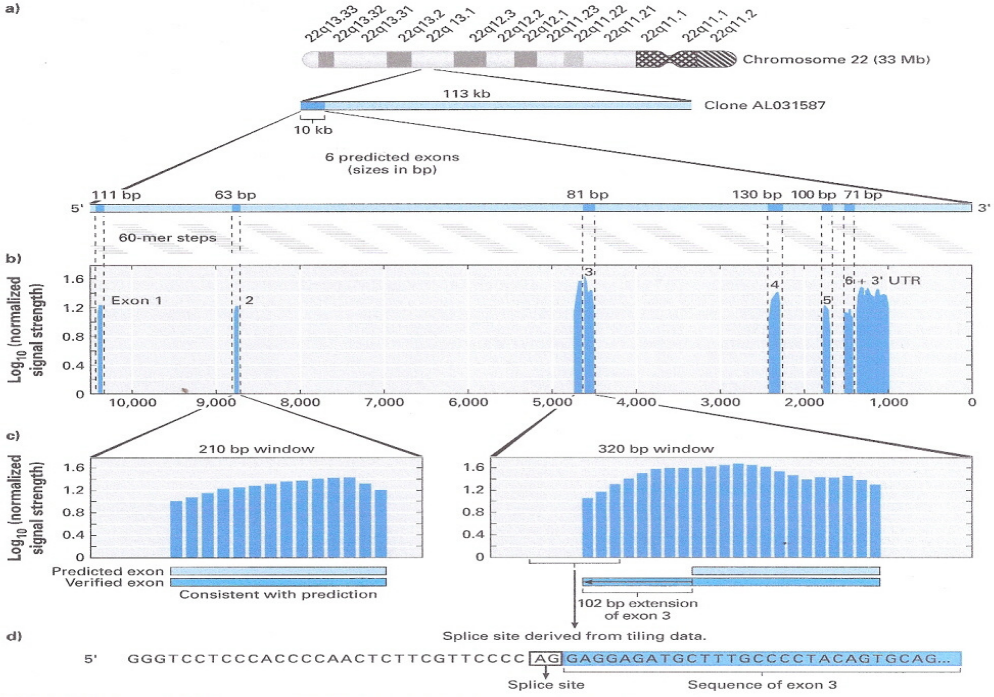
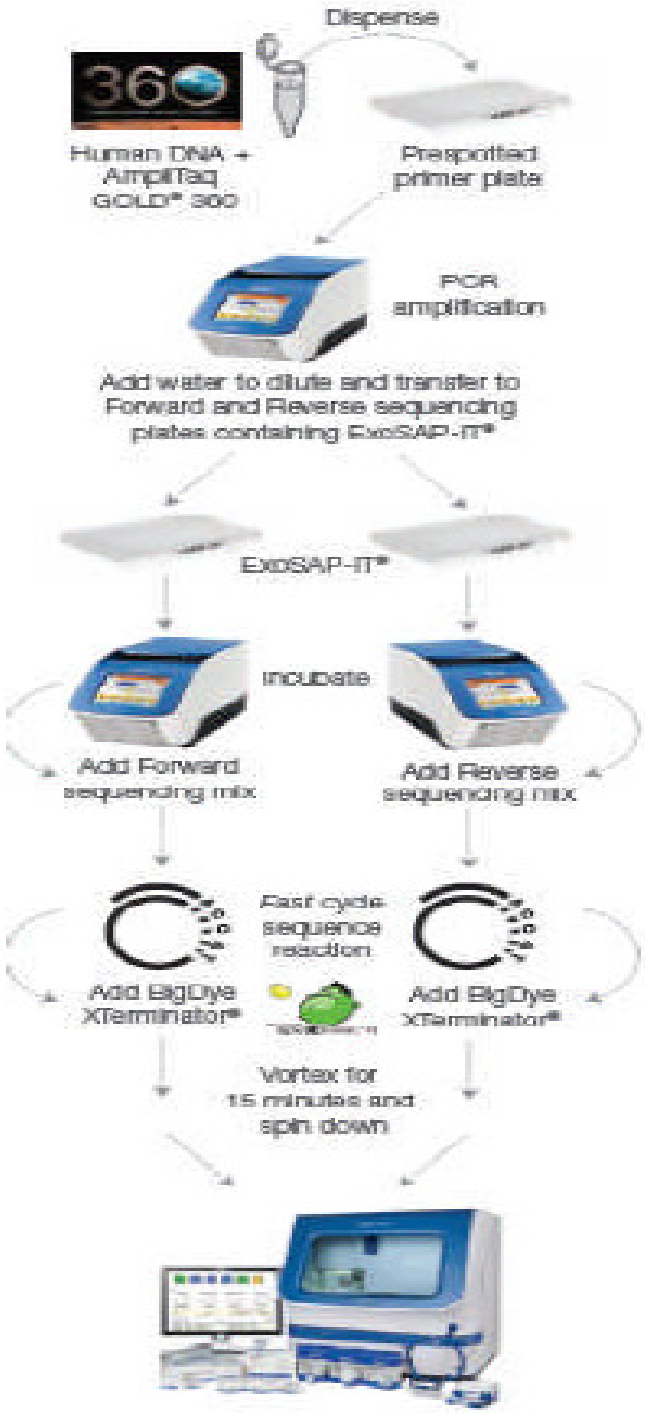


FIGURE 4.1 • DNA microarray. Each blue spot indicates the location of a PCR product on the glass slide. One particular spot has been chosen to illustrate the presence of one gene's sequence. On a real microarray, each spot is about 100 μm in diameter.







BRCA1 és BRCA2 gén resequencing

- mutációk diagnosztikai célú azonosítása

BRCA1 és BRCA2: 23 és 27 exon (80Kb)

Nincs elő screening: SSCP, DGGE, dHPLC, stb.

Egy minta – egy assay koncepció

Gyors, pontos, teljes lefedettséget ad

Nincs kereszt kontamináció

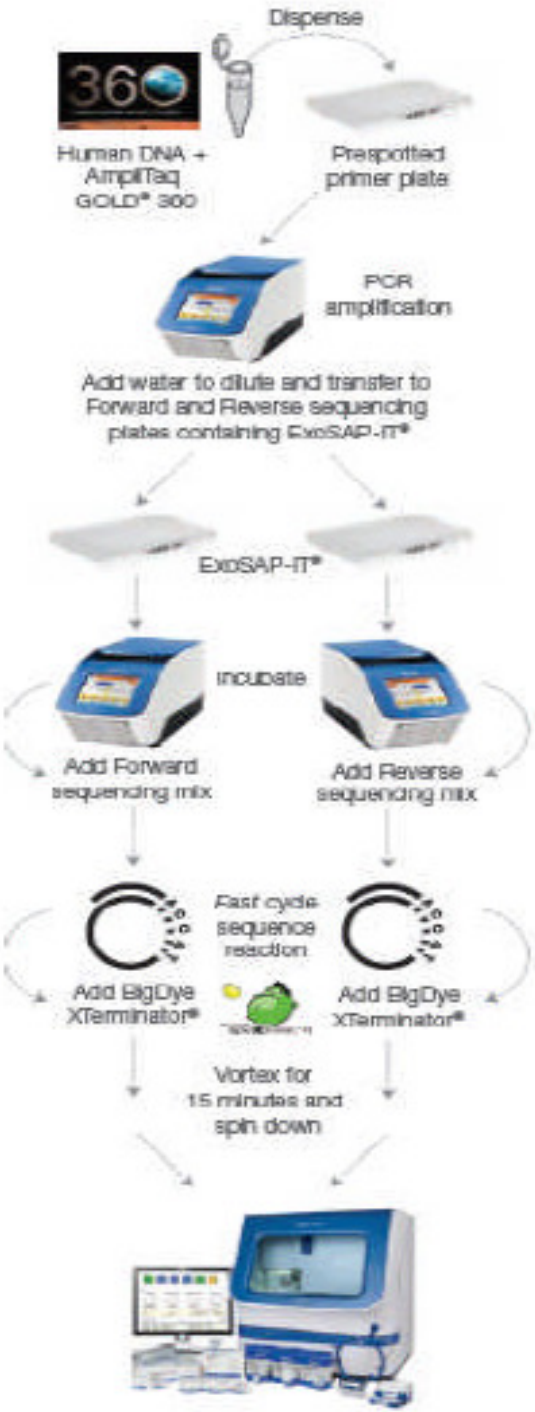
BRCA1 és BRCA2: 34 és 47 amplikon

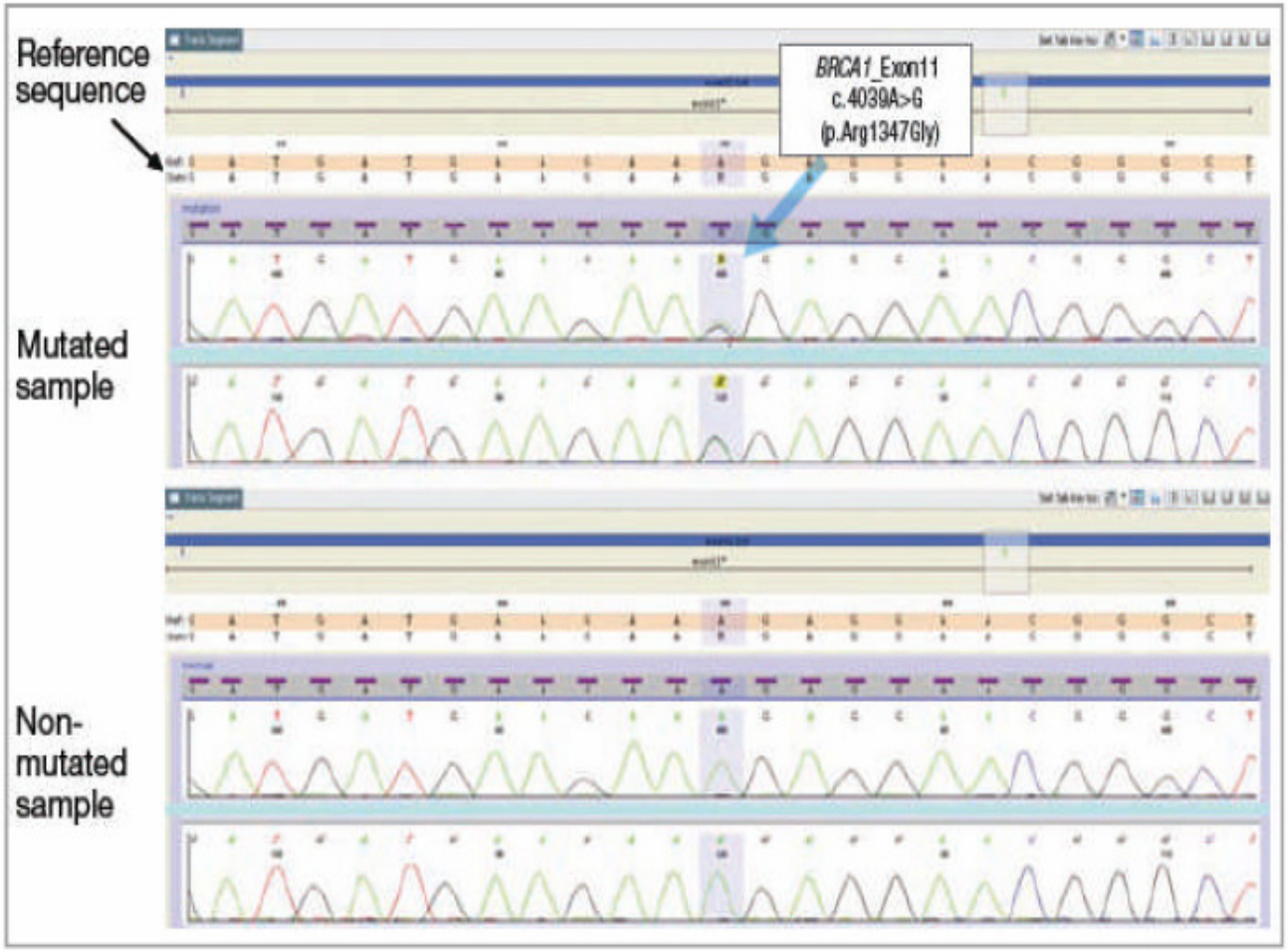
	1	2	3	4	5	6	7	8	9	10	11	12
A	Ex-1	Ex-10	Ex-11-8	Ex-15	Ex-23	Ex-1	Ex-10-1	Ex-11-5	Ex-11-13	Ex-14-2	Ex-22	MP-2
B	Ex-2	Ex-11-1	Ex-11-9	Ex-16	Ex-24	Ex-2	Ex-10-2	Ex-11-6	Ex-11-14	Ex-15	Ex-23	MP-3
C	Ex-3	Ex-11-2	Ex-11-10	Ex-17	MP-1	Ex-3	Ex-10-3	Ex-11-7	Ex-11-15	Ex-16	Ex-24	MP-4
D	Ex-5	Ex-11-3	Ex-11-11	Ex-18	MP-2	Ex-5	Ex-10-4	Ex-11-8	Ex-11-16	Ex-17	Ex-25	MP-5
E	Ex-6	Ex-11-4	Ex-11-12	Ex-19	MP-3	Ex-6	Ex-11-1	Ex-11-9	Ex-11-17	Ex-18	Ex-26	MP-6
F	Ex-7	Ex-11-5	Ex-12	Ex-20	MP-4	Ex-7	Ex-11-2	Ex-11-10	Ex-12	Ex-19	Ex-27-1	MP-7
G	Ex-8	Ex-11-6	Ex-13	Ex-21	MP-5	Ex-8	Ex-11-3	Ex-11-11	Ex-13	Ex-20	Ex-27-2	MP-8
H	Ex-9	Ex-11-7	Ex-14	Ex-22	MP-6	Ex-9	Ex-11-4	Ex-11-12	Ex-14-1	Ex-21	MP-1	MP-9

BRCA1 BRCA2 Multiplex no-template control

BRCA1 és BRCA2 gén resequencing

- mutációk diagnosztikai célú azonosítása





Next Generation Sequencing –

Massively Parallel Sequencing of clonally amplified (or single) DNA molecules

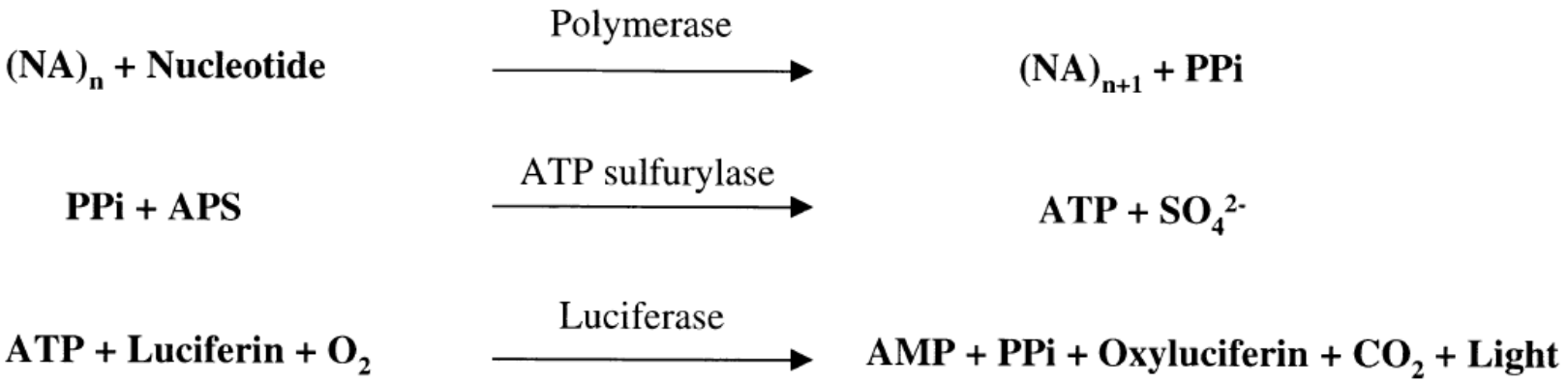
-Process millions of sequence reads in parallel
-Library preparation
-Specific adaptor oligos
-Little volume DNA template
-Produce shorter read lengths (35-400 bp)
-100 Mb to several Gb nucleotid sequence

Piroszekvenálás:
chemiluminescent detection of pyrophosphate

Enzimek:
Klenow fragment
ATP szulfuriláz
Luciferáz
Apiráz

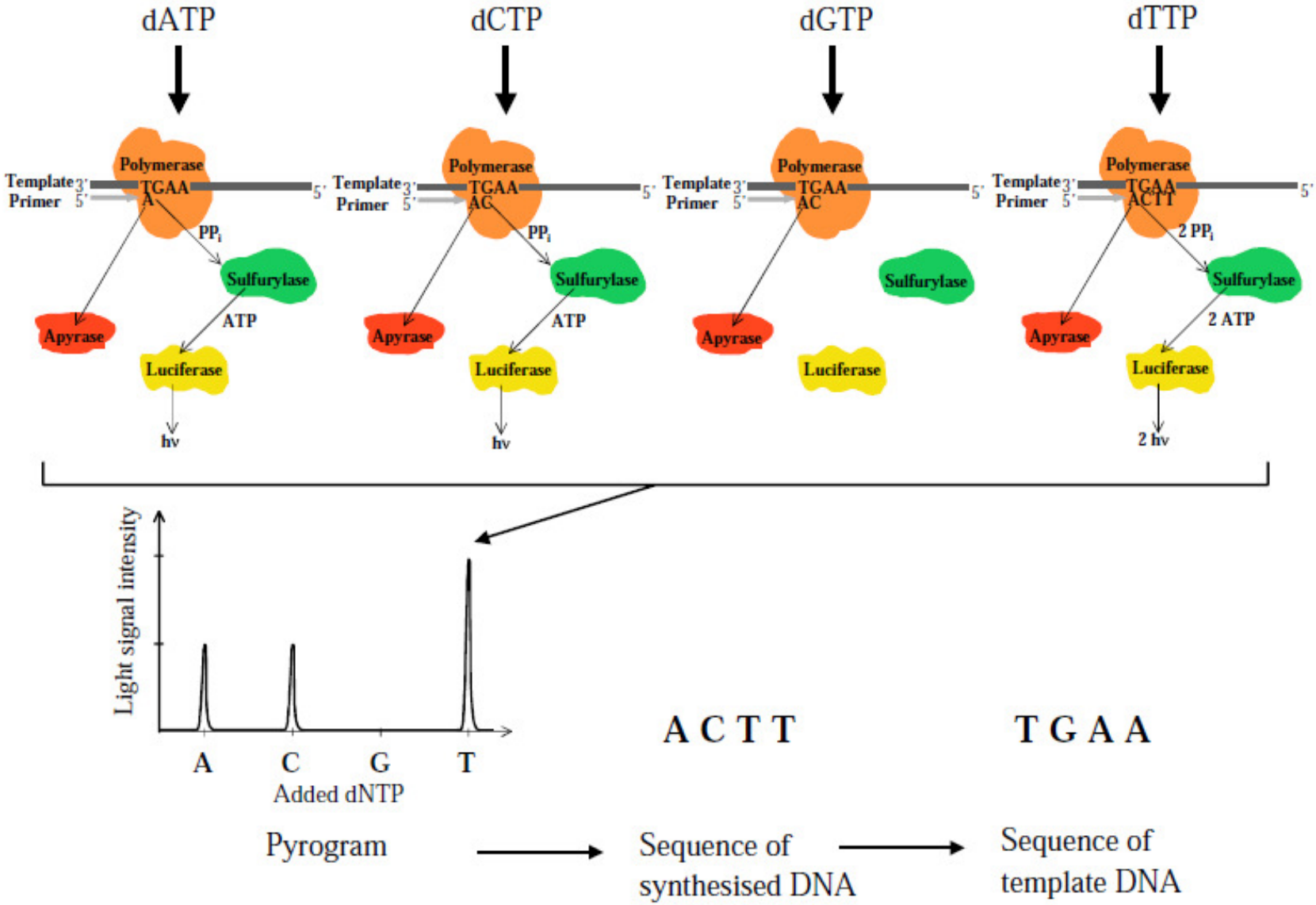
Vegyületek:
Adenozin-foszfoszulfát (APS)
D-luciferin
Templát
Primer

dNTP-k egymás után adva

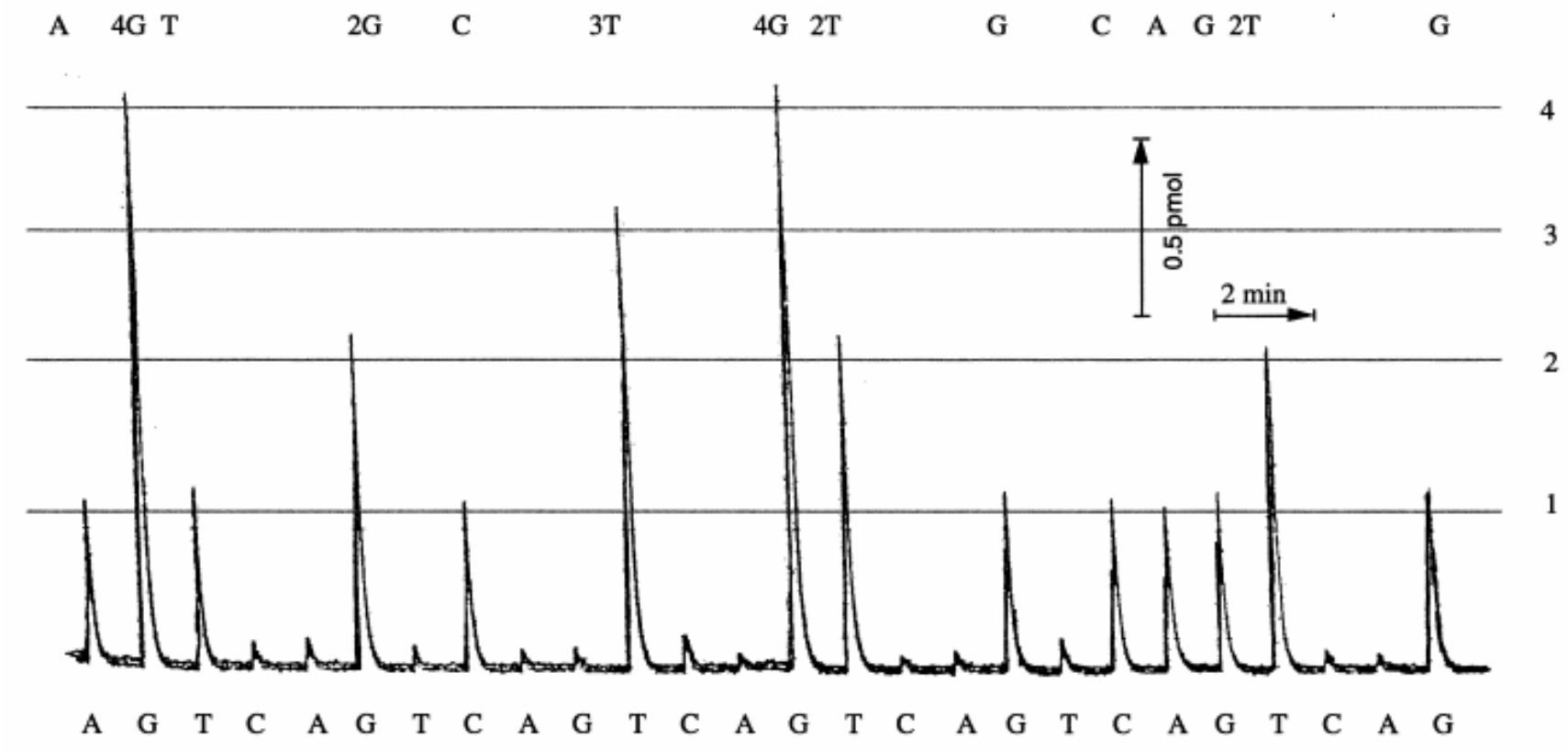


Ronaghi M (2001) Genom Res

Piroszekvenálás

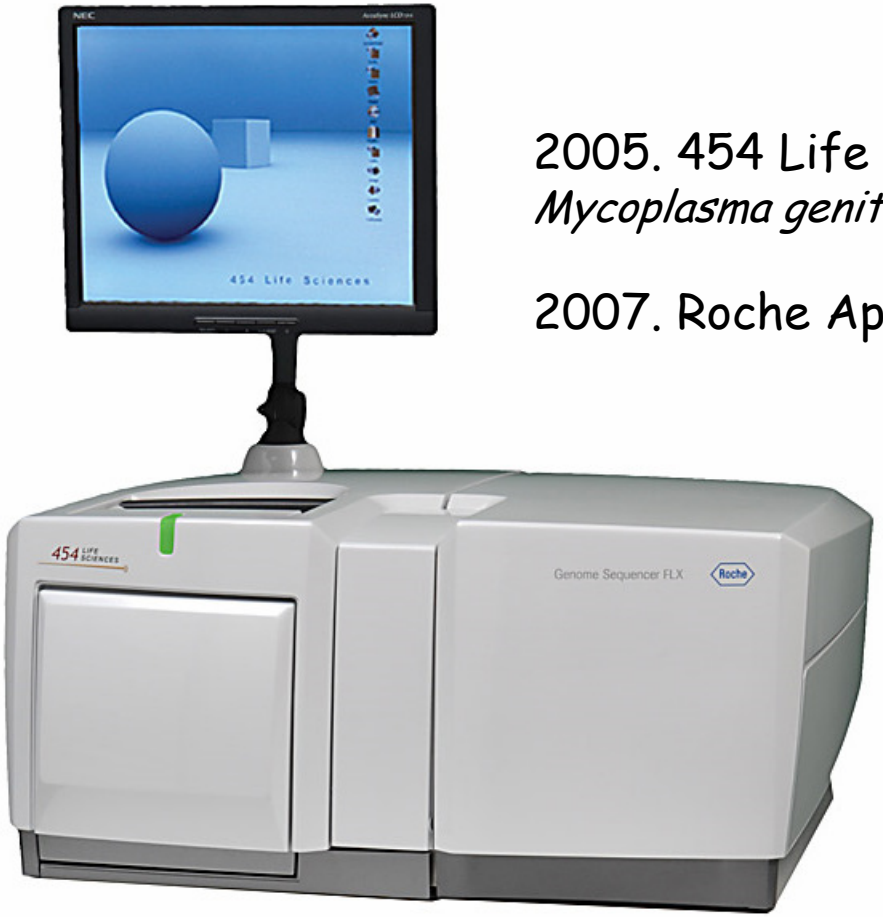


Ahmadian A (2006) Clin Chim Acta



Ronaghi M (2001) Genom Res

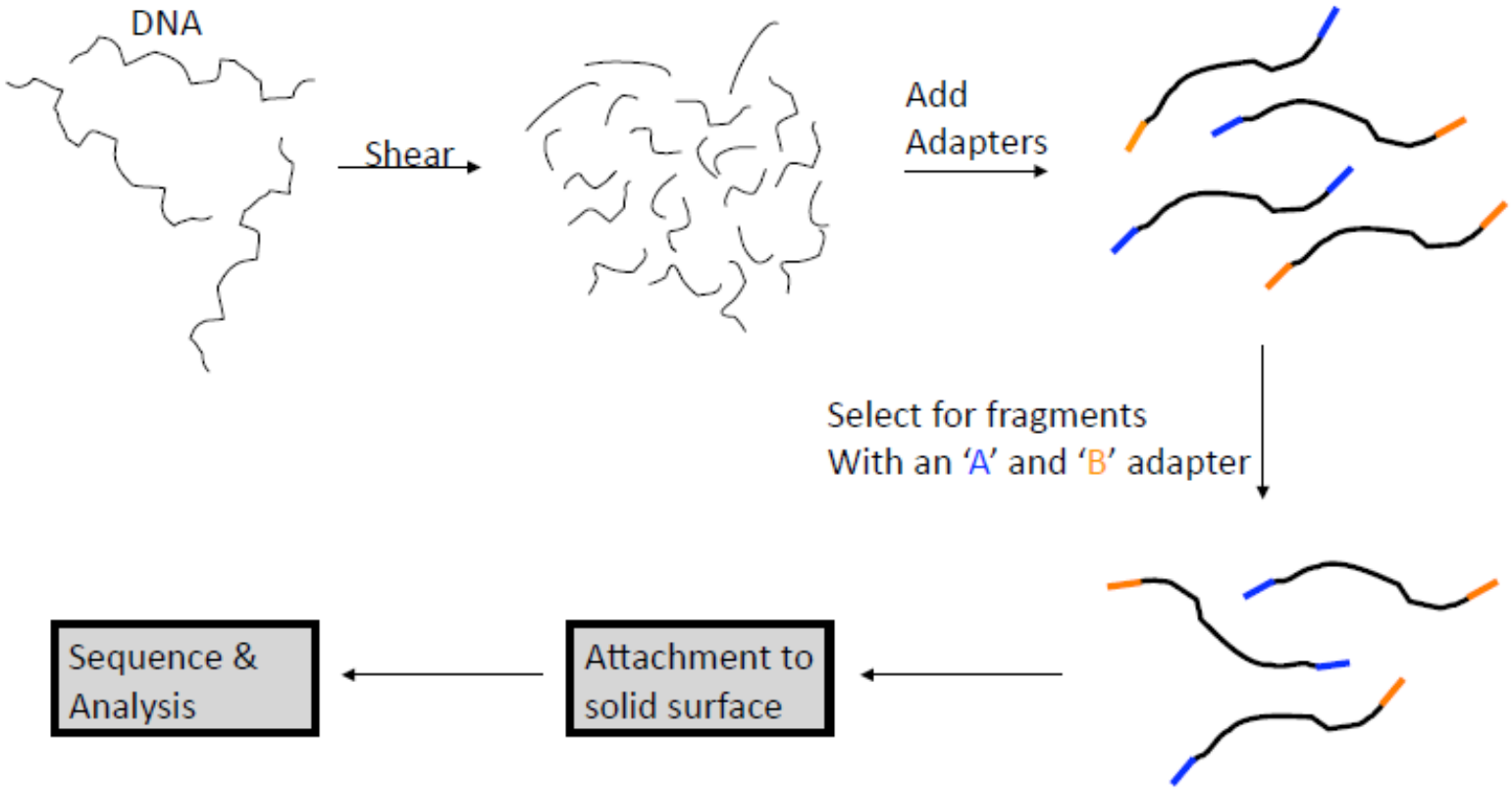
Roche/454 sequencing technology



2005. 454 Life Sciences cég fejlesztette (GS 20)
Mycoplasma genitalia 580 kb genom, 99.96% pontosság

2007. Roche Applied Science (GS FLX sorozat)

Minta előkészítés



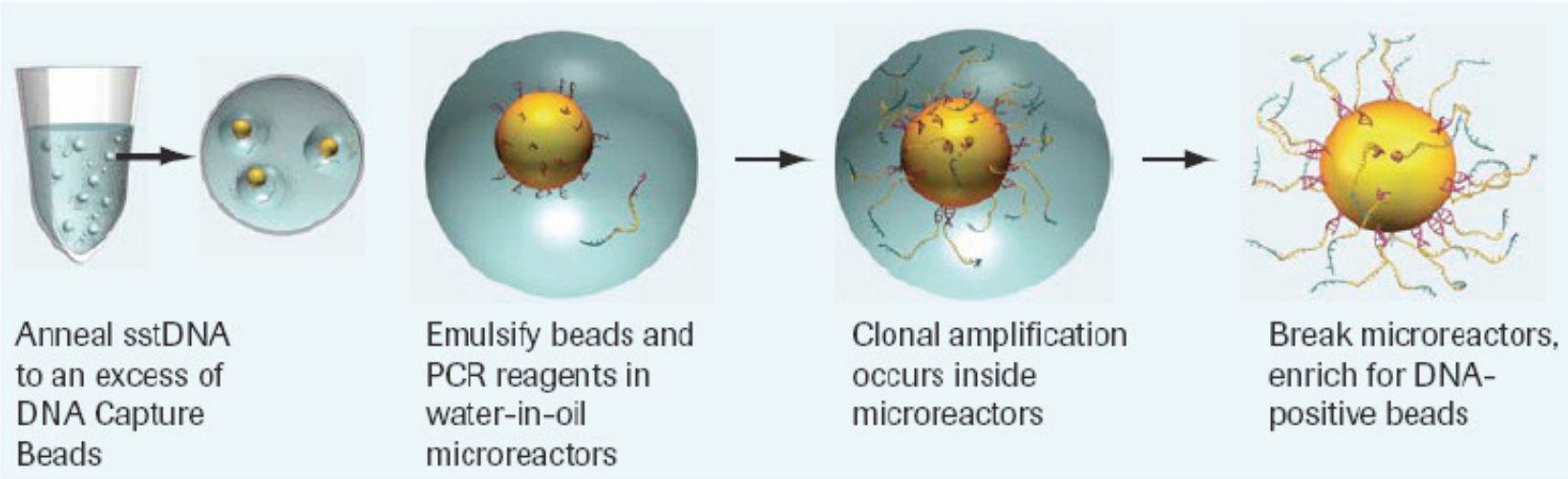
DNS fragmentumok törése (néhány 100 bp)
End-repair
Adapterek ligálás

Roche/454 sequencing technology
Klonális amplifikáció

Emulsion PCR

Mikroreaktorok
Víz-olaj emulzióban

Több millió kópiája egy fragmentumnak

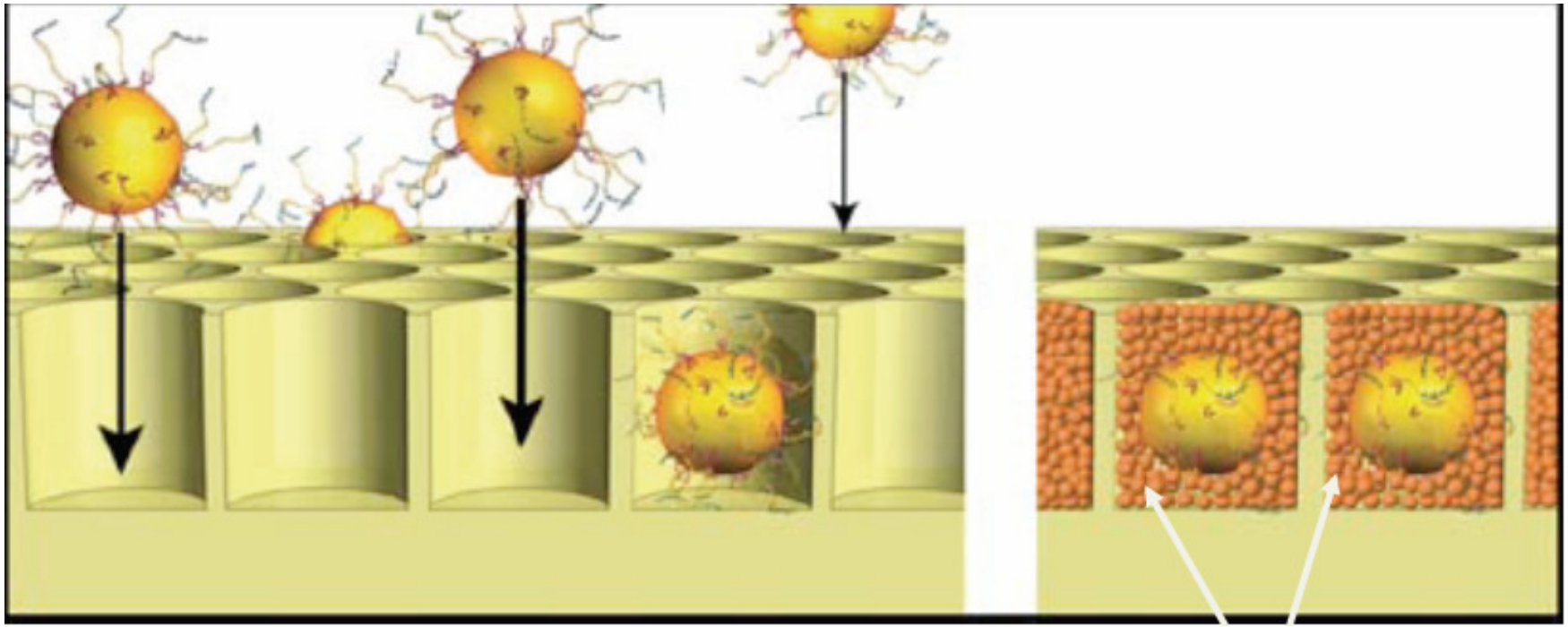


Each bubble in the emulsion will potentially contain a different fragment.

Roche/454 sequencing technology

Picotiter well plate betöltés

$3,4*10^6$ lyuk
Pikoliteres szekvenáló reakciók



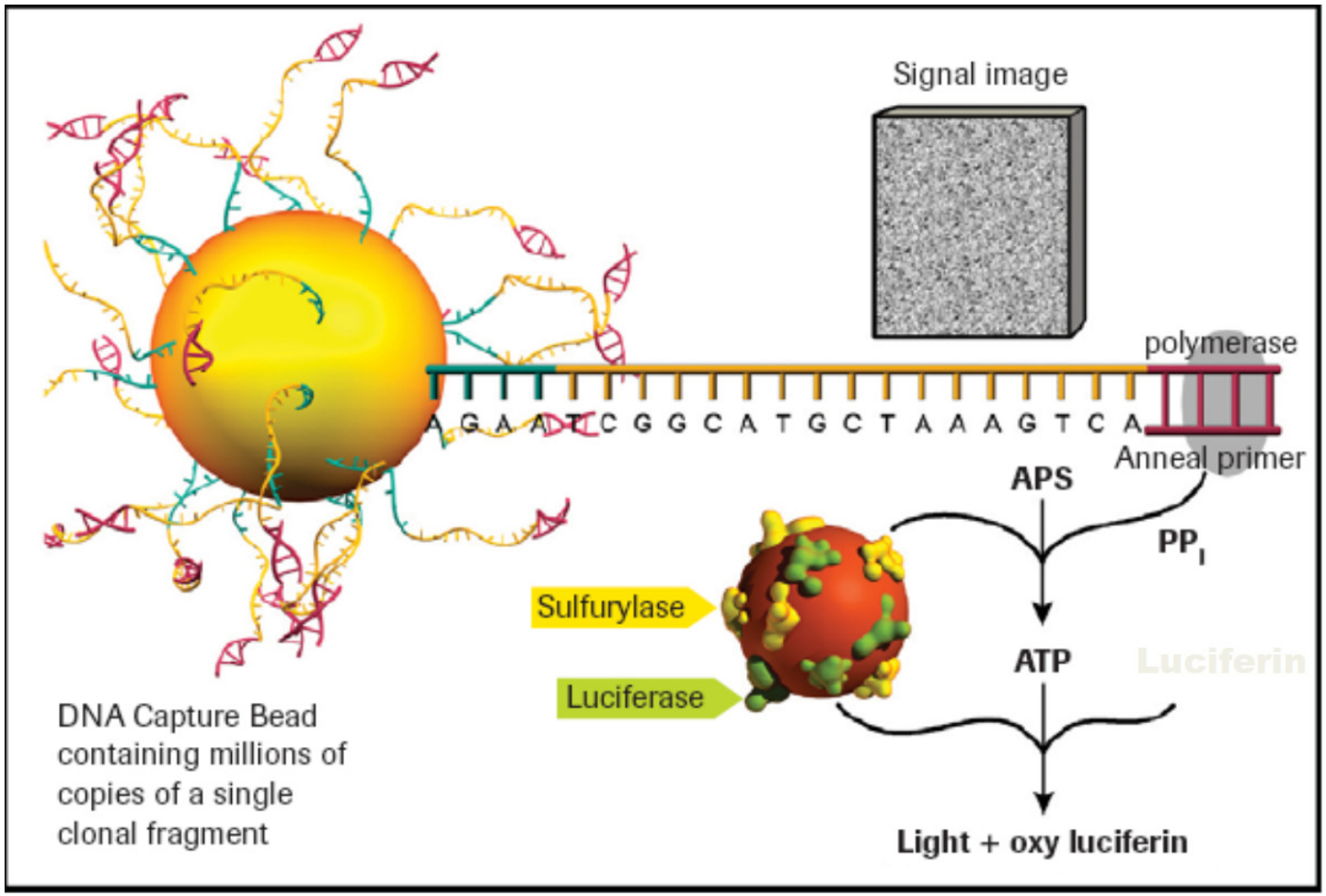
Instead of 96 reads/run, there are hundreds of thousands.

Packing beads and enzyme beads

Ansorge (2009) New biotechnology

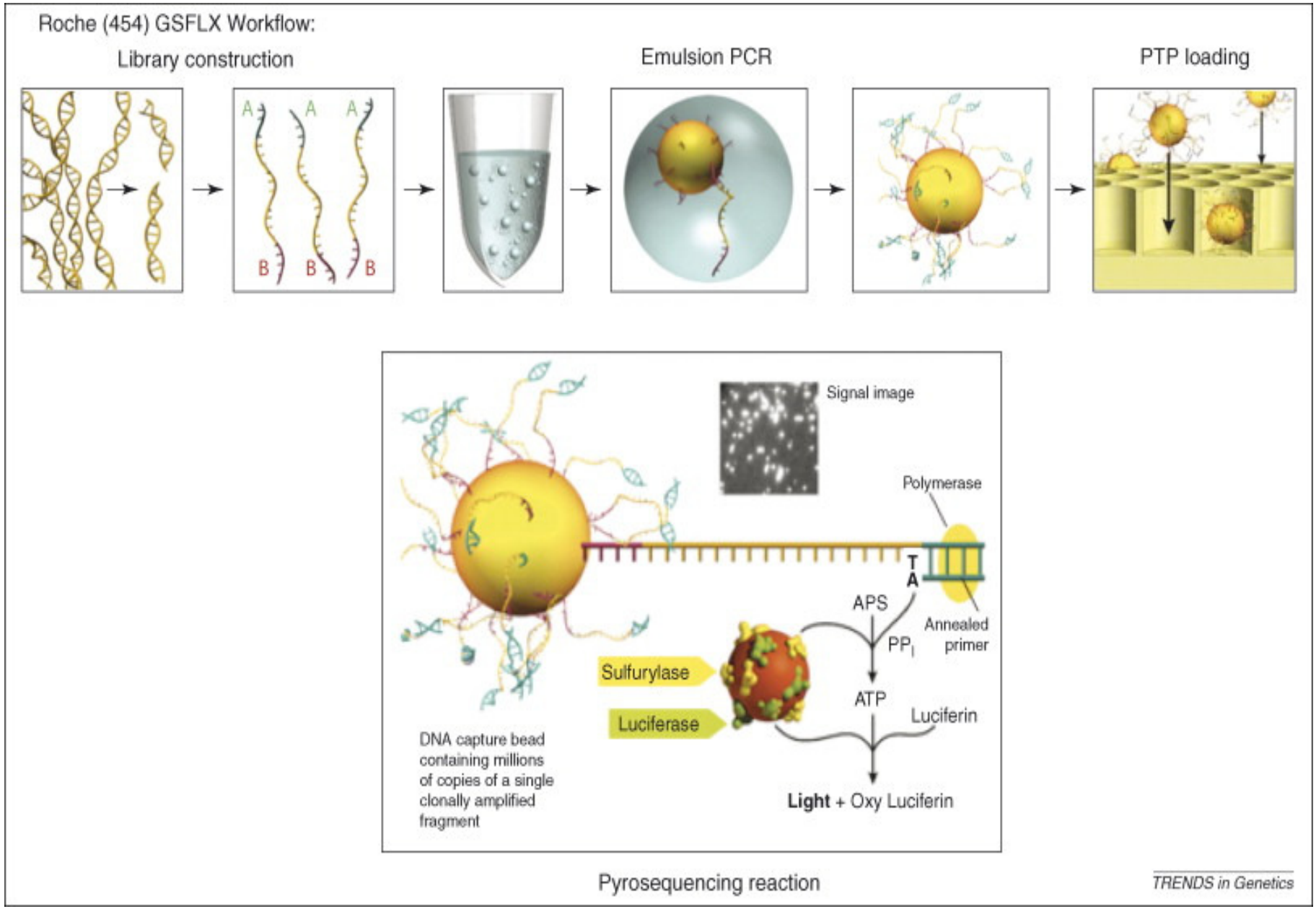
Roche/454 sequencing technology

Szekvencia meghatározás piroszekvenálással

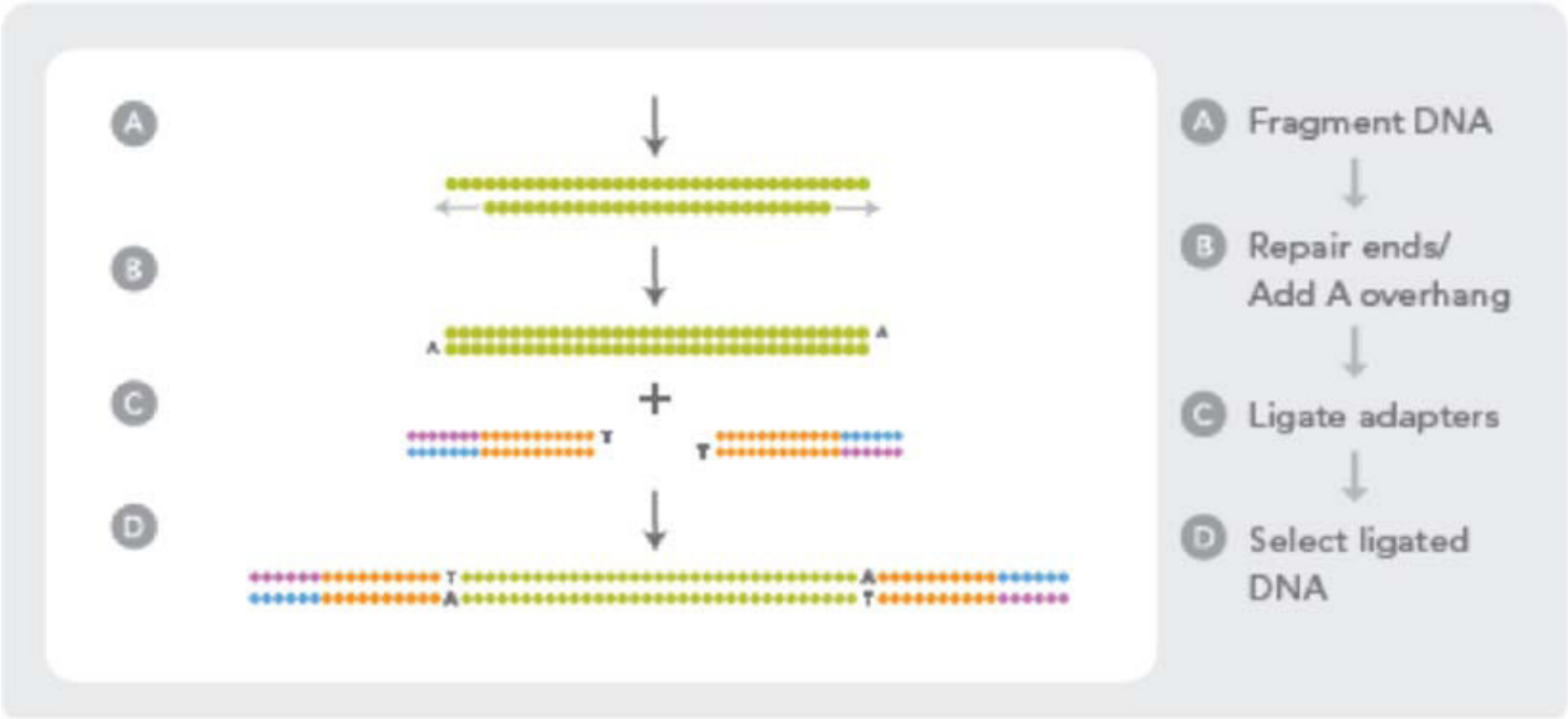


Ansorge (2009) New biotechnology

Next Generation Sequencing – Roche 454 platform



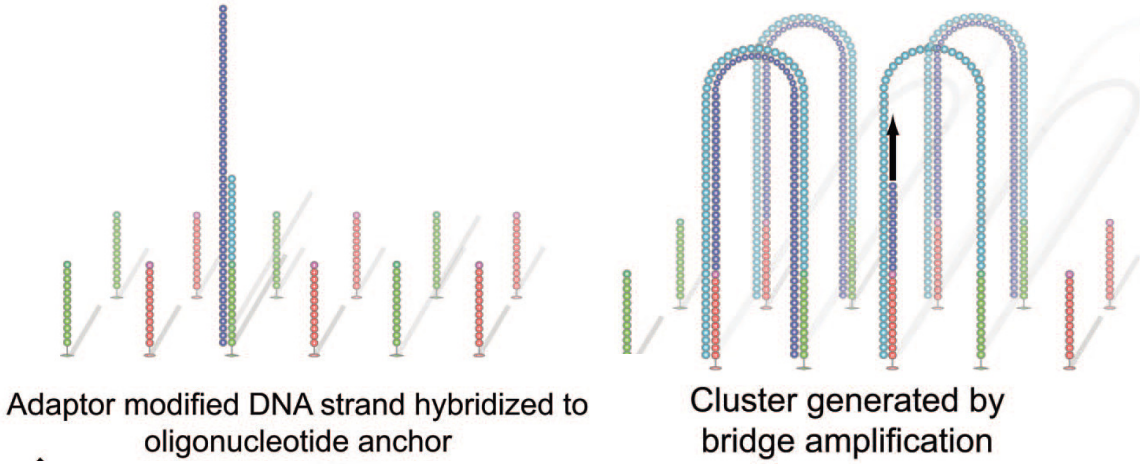
Illumina/Solexa sequencing
Minta előkészítés



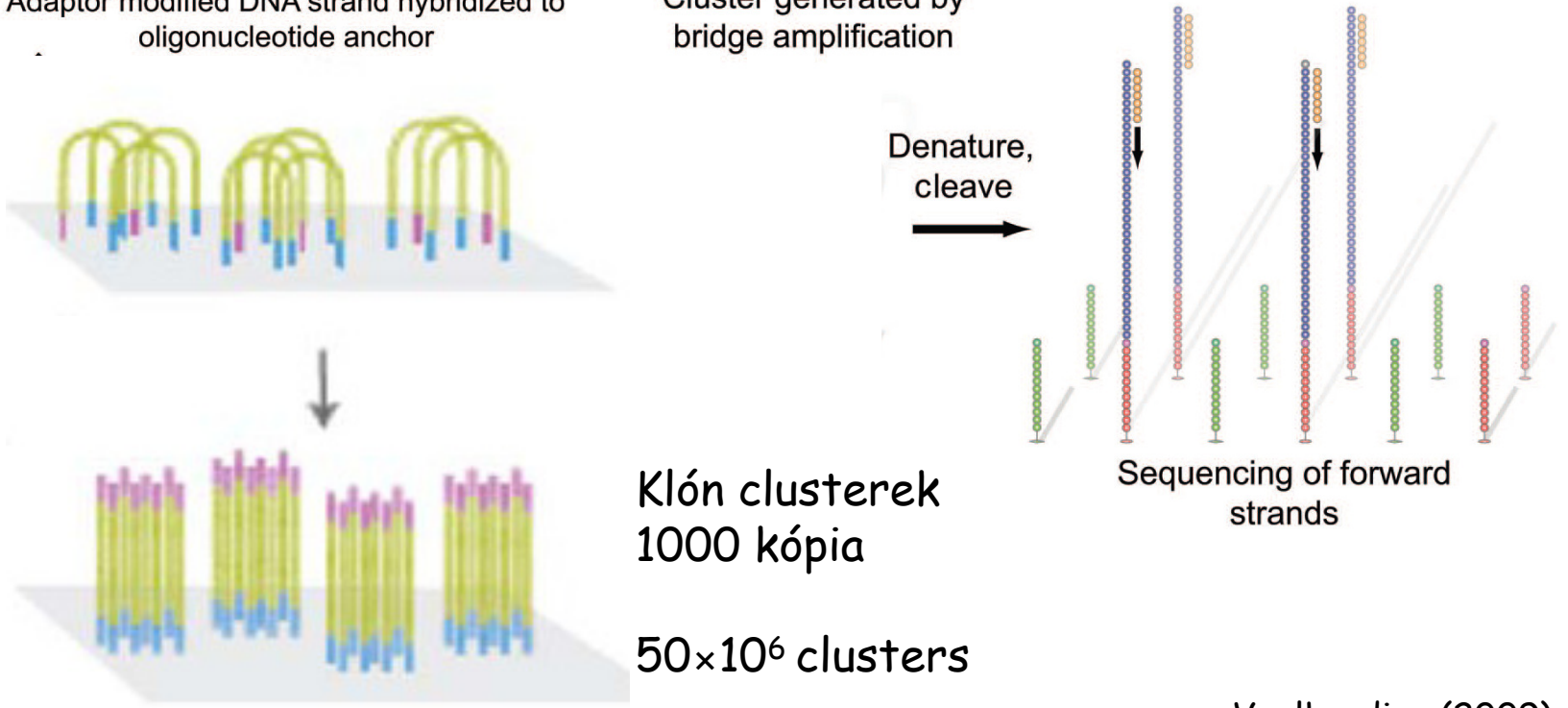
A) DNS fragmentálása (néhány 100 bp)
B) Végek javítása, A túlnyúló vég kialakítása
C) Adaptor ligálása (T túlnyúló vég)

Ansorge (2009) New biotechnology

Illumina/Solexa sequencing
Klonális amplifikáció



PCR reakció kihorgonyzott primerekkel
Bridge amplifikáció



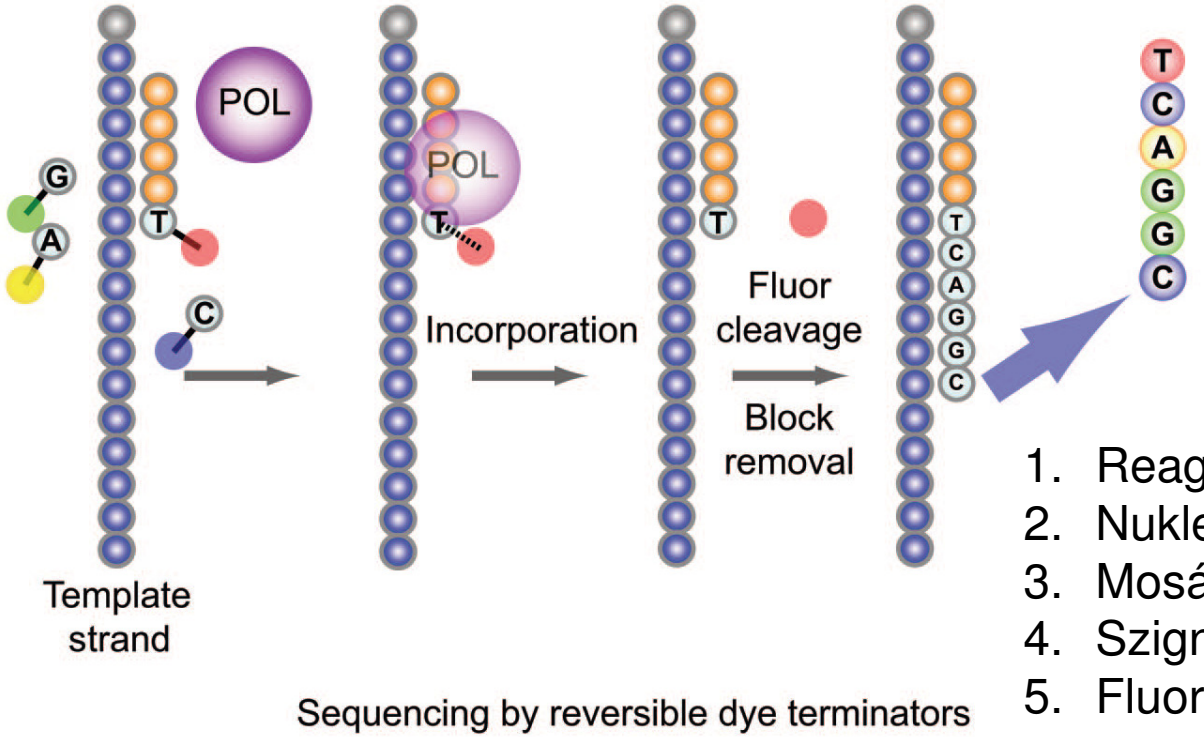
Klón clusterek
1000 kópia

50×10^6 clusters

Voelkerding (2009) Clin Chem

Illumina/Solexa sequencing

Szekvencia meghatározás szintézis során



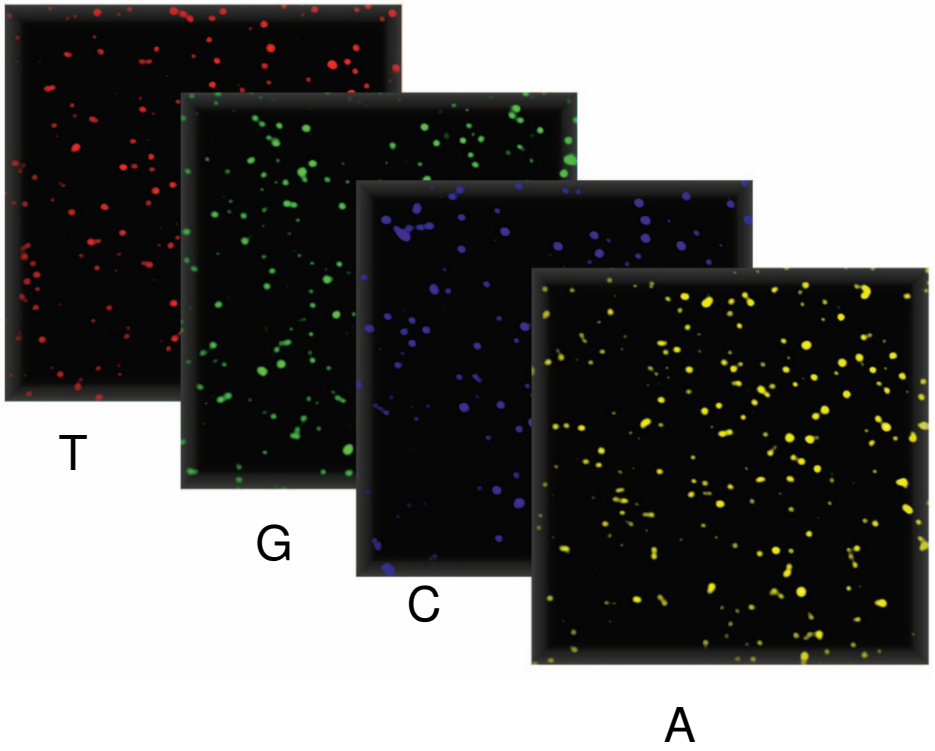
1. Reagensek hozzáadása
2. Nukleotid beépülése
3. Mosás
4. Szignál detekció
5. Fluorofór és a block eltávolítása

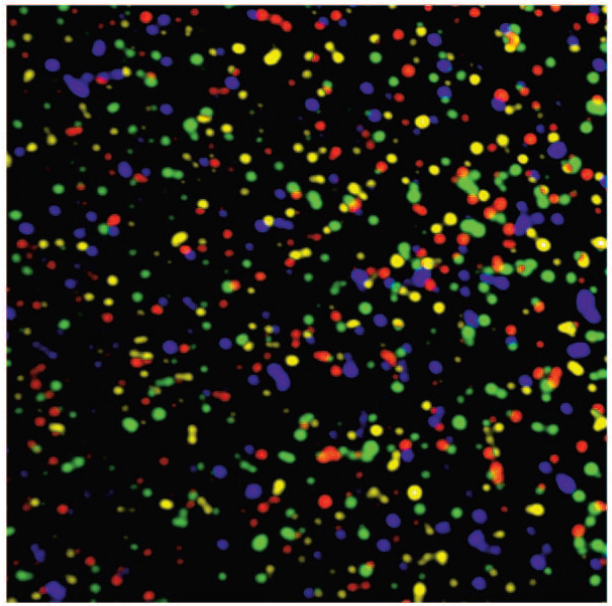
Fluoreszcensen jelölt reverzibilis terminátorok
Mind a 4 nukleotidot egyszerre adják a reakcióba

Voelkerding (2009) Clin Chem

Illumina/Solexa sequencing

A fluoreszcens szignál észlelése

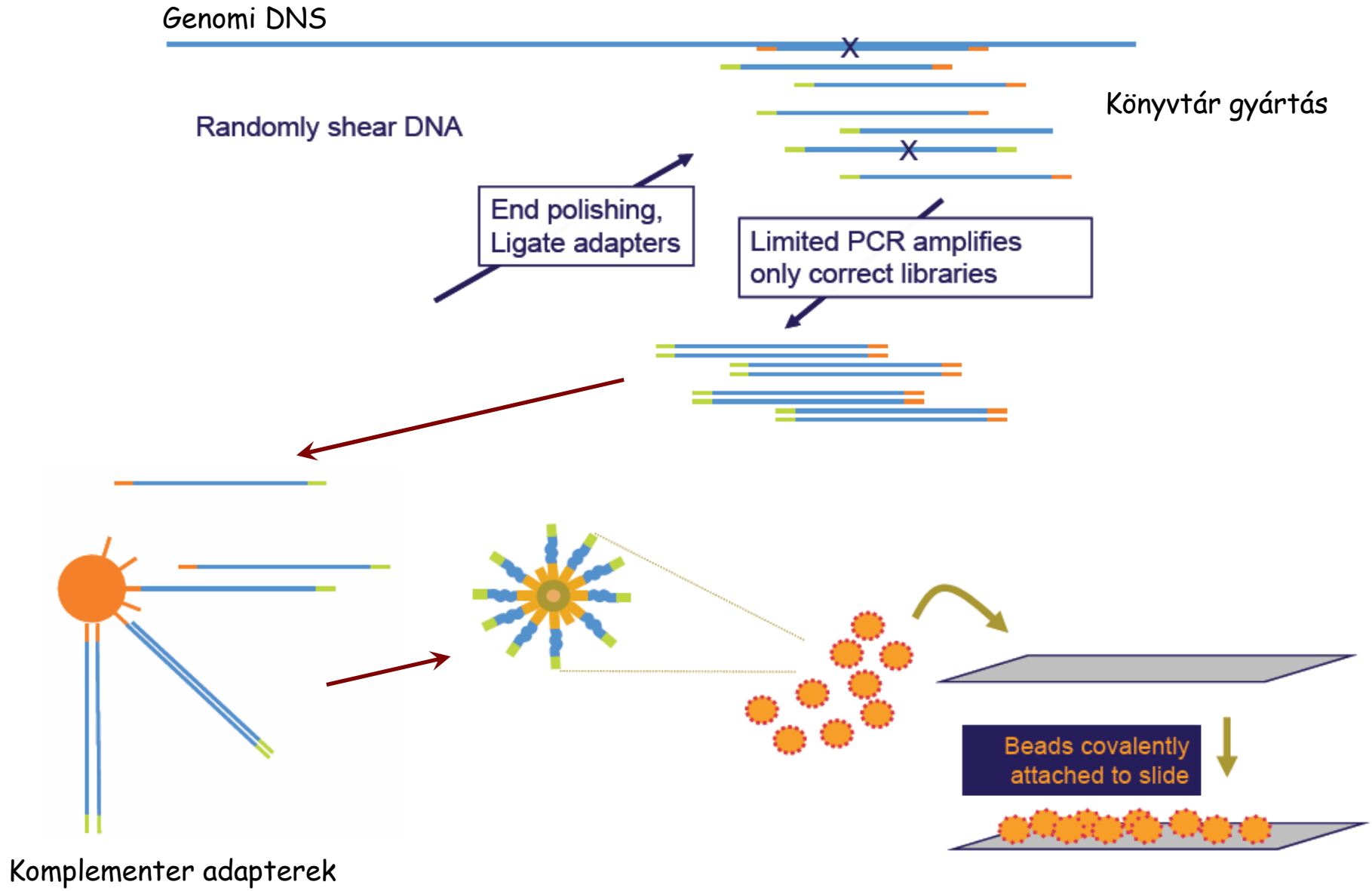




Voelkerding (2009) Clin Chem

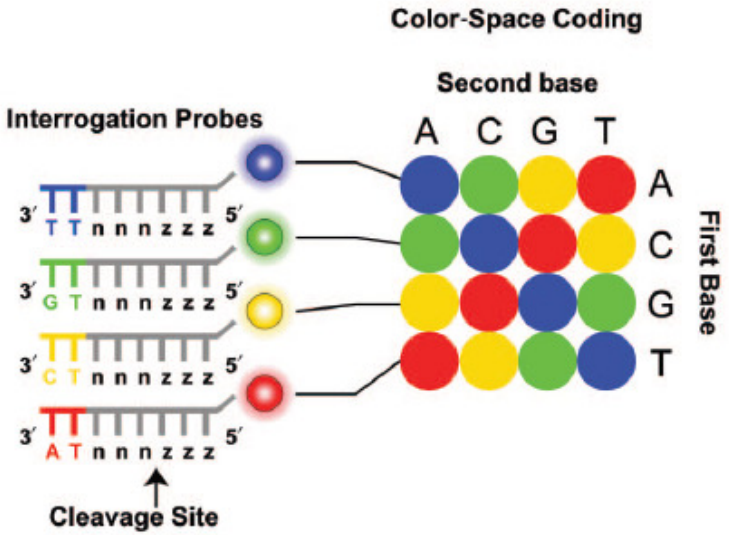
SOLID: Sequencing by Oligo Ligation and Detection

- Kémiai hasítás, amplifikálás és ligálás

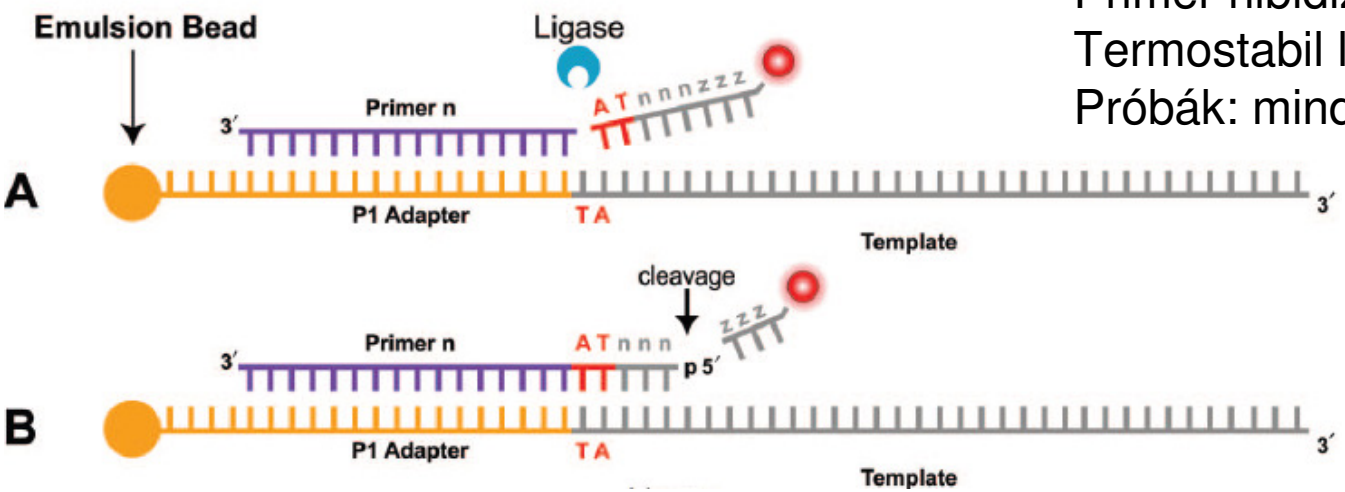


Applied Biosystems - SOLiD

Szekvencia meghatározás ligálással!



Próba
Octamer
2 próba specifikus bázis
3 degenerált bázis
3 univerzális
Fluoreszcens jelölő

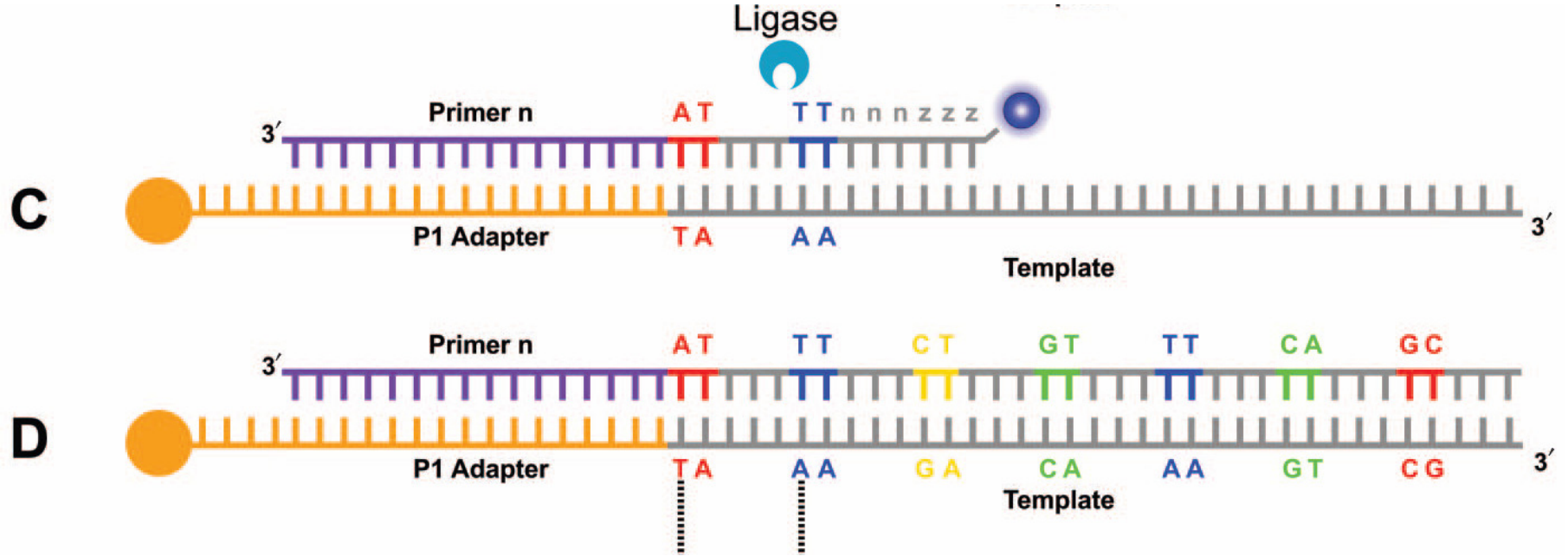


Primer hibidizálása az adapterhez
Termostabil ligáz
Próbák: mind a 16 féle kombináció

Ligálás
Mosás
Jel detektálása
Hasítás – 3 nukleotid

Voelkerding (2009) Clin Chem

Applied Biosystems - SOLiD

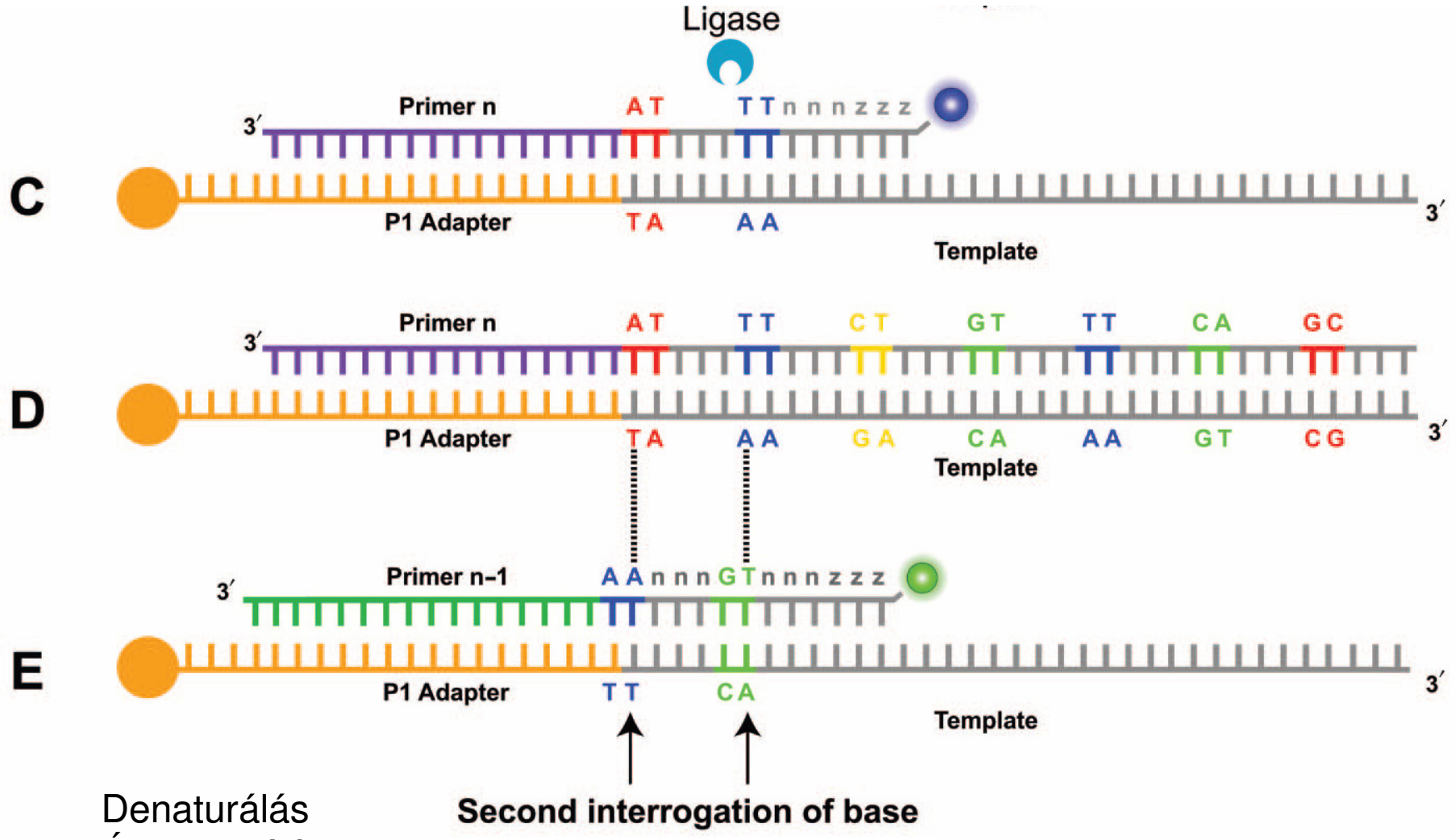


Újabb próba ligálása

7-szer zajlik le ez a ciklus

Voelkerding (2009) Clin Chem

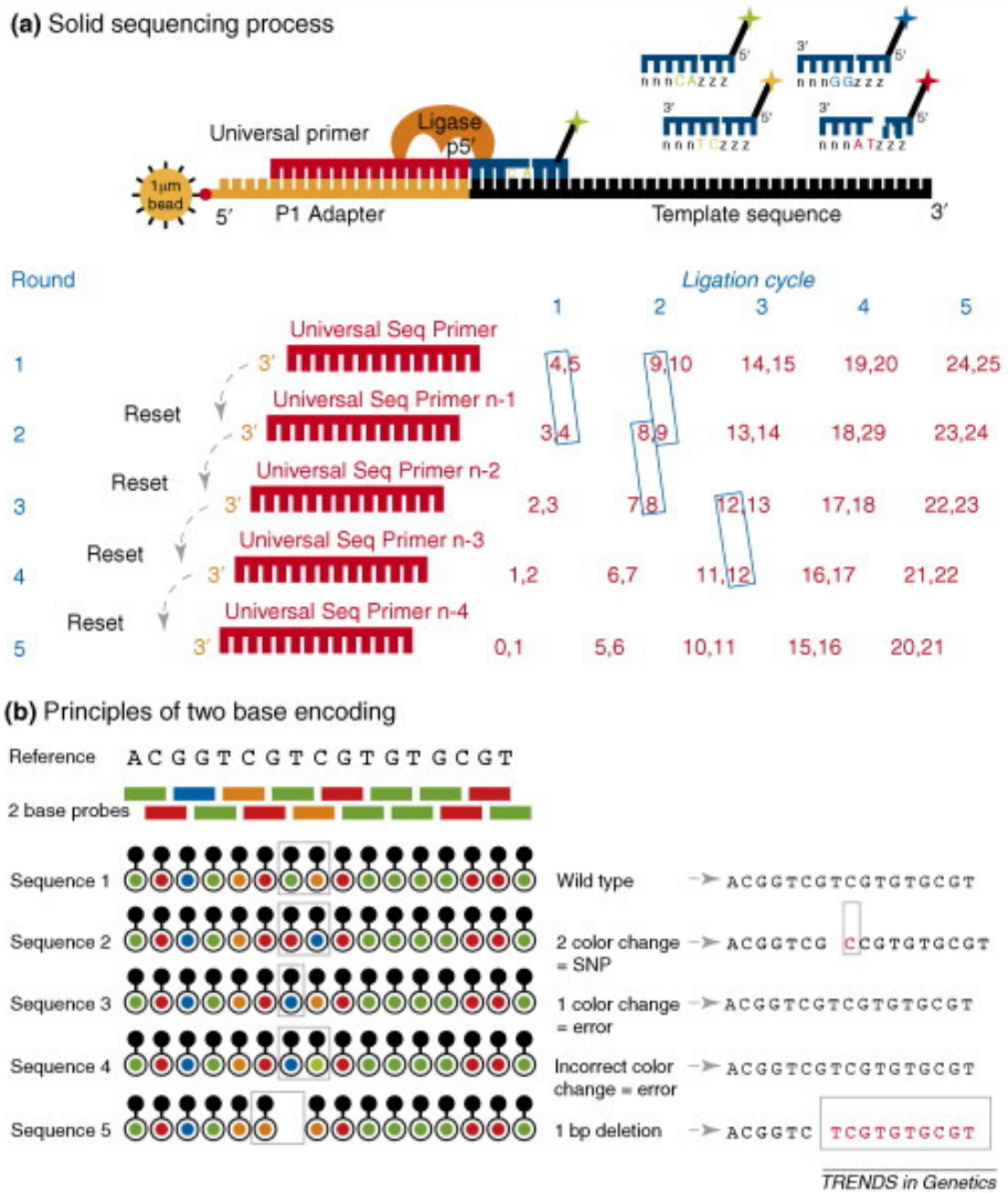
Applied Biosystems - SOLiD



Denaturálás
Új kör indítása n-1 adapter primerrel
5 kör

Minden nukleotidra 2* kérdez rá

Voelkerding (2009) Clin Chem


(a) Solid sequencing process
Universal primer
Ligase
p5′
1μm bead
5′
P1 Adapter
Template sequence
3′
Round
Ligation cycle
1
2
3
4
5
Universal Seq Primer
Universal Seq Primer n-1
Universal Seq Primer n-2
Universal Seq Primer n-3
Universal Seq Primer n-4
Reset
4,5
9,10
14,15
19,20
24,25
3,4
8,9
13,14
18,29
23,24
2,3
7,8
12,13
17,18
22,23
1,2
6,7
11,12
16,17
21,22
0,1
5,6
10,11
15,16
20,21
(b) Principles of two base encoding
Reference
A C G G T C G T C G T G T G C G T
2 base probes
Sequence 1
Sequence 2
Sequence 3
Sequence 4
Sequence 5
Wild type
ACGGTCGTCGTGTGCGT
2 color change = SNP
ACGGTCG CCGTGTGCGT
1 color change = error
ACGGTCGTCGTGTGCGT
Incorrect color change = error
ACGGTCGTCGTGTGCGT
1 bp deletion
ACGGTC TCGTGTGCGT
TRENDS in Genetics

Next Generation DNA Sequencing: SOLID

- Kémiai hasítás, amplifikálás és ligálás

Accuracy: 99.99 %

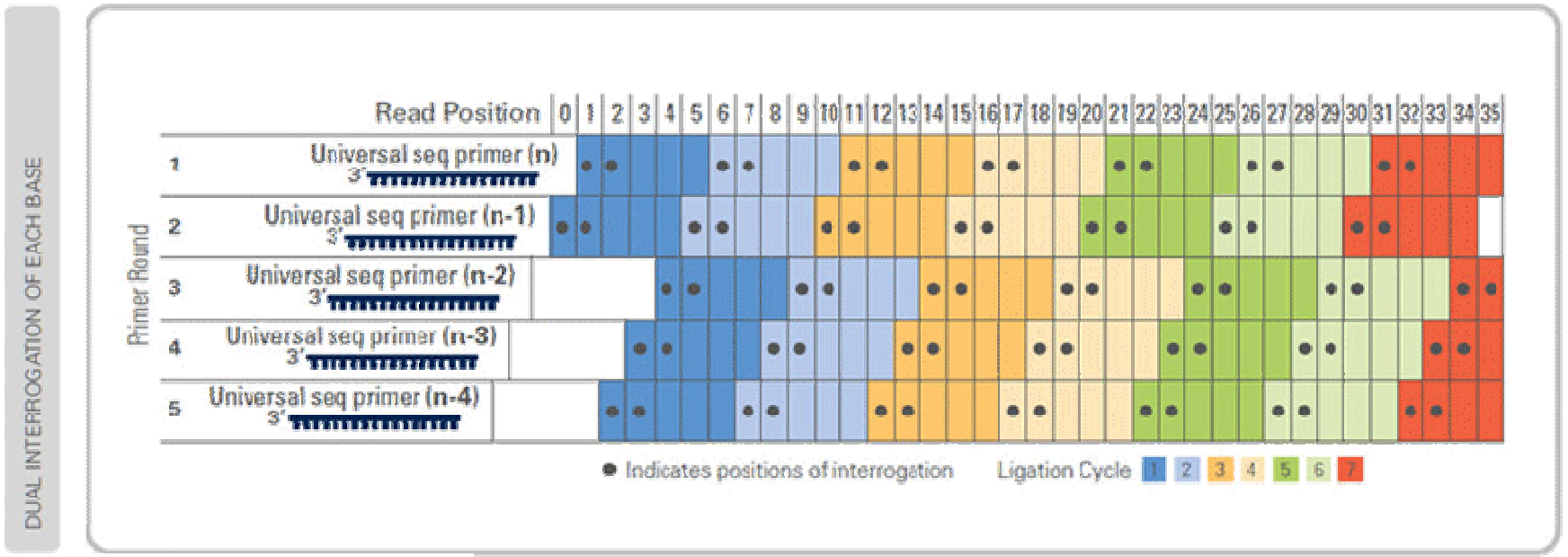


Table 2. AB SOLiD cycle number descriptions

Cycle number	Universal primer position	Base positions identified	Probe set[a]	Positions interrogated
1	n	4,5	NNNAA^NNN-fl	5,10,15,20,25
2	n-1	4,5	NNNAT^NNN-fl	4,9,14,19,24
3	n-2	4,5	NNNAC^NNN-fl	3,8,13,18,23
4	n	1,2	AANNN^NNN-fl	2,7,12,17,22
5	n-1	1,2	ATNNN^NNN-fl	1,6,11,16,21

a ^, position of cleavage on each 8mer, whereas fl indicates the position of the fluorescent group on the 8mer.

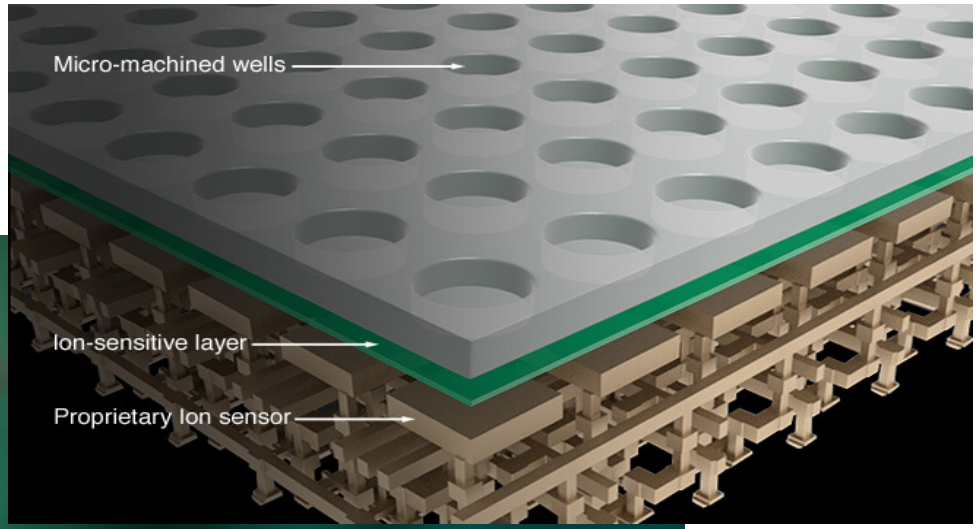
Table 1. Comparing metrics and performance of next-generation DNA sequencers

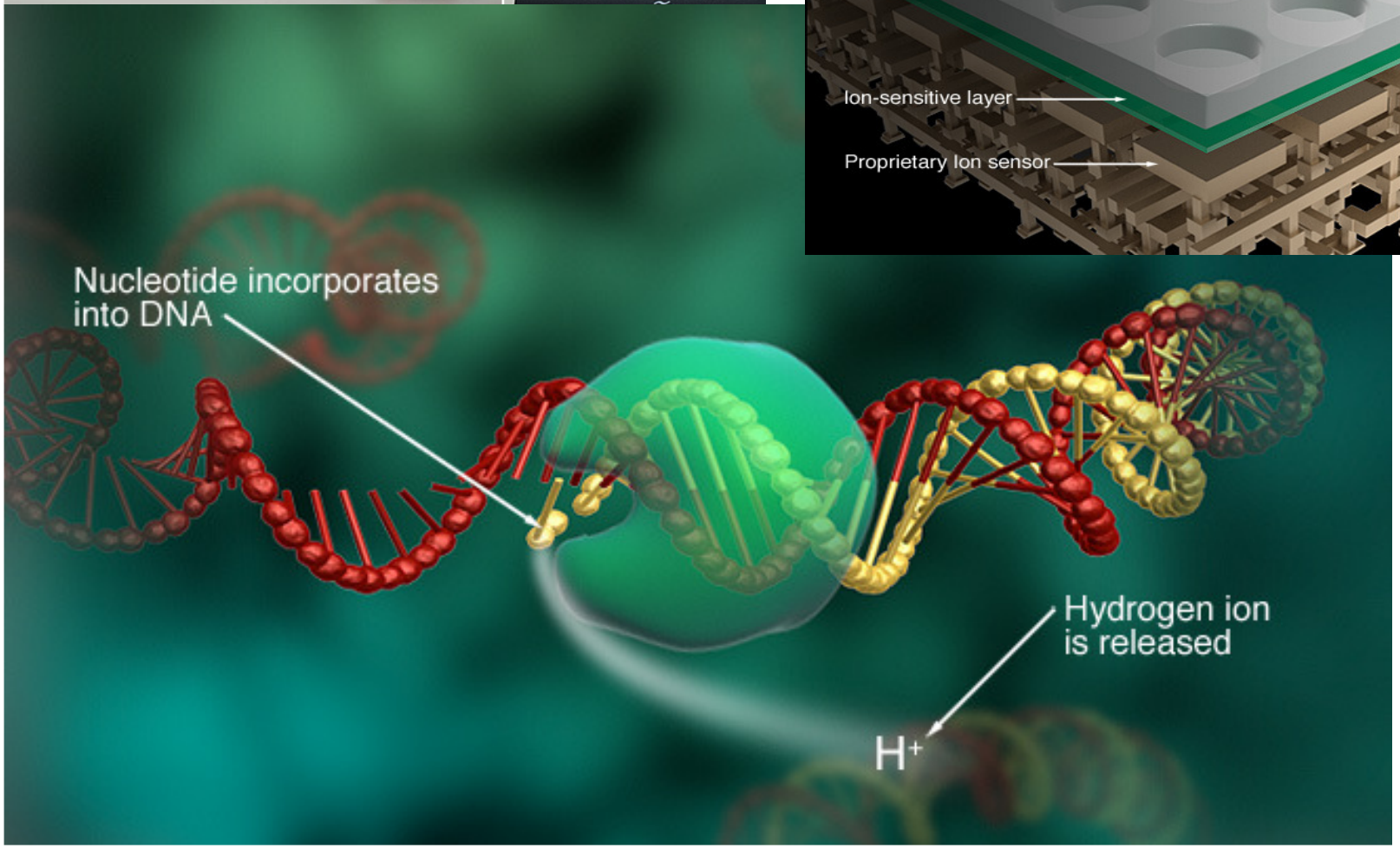
	Platform		
	Roche(454)	Illumina	SOLiD
Sequencing chemistry	Pyrosequencing	Polymerase-based sequencing-by-synthesis	Ligation-based sequencing
Amplification approach	Emulsion PCR	Bridge amplification	Emulsion PCR
Paired ends/separation	Yes/3 kb	yes/200 bp	Yes/3 kb
Mb/run	100 Mb	1300 Mb	3000 Mb
Time/run (paired ends)	7 h	4 days	5 days
Read length	250 bp	32–40 bp	35 bp
Cost per run (total direct[a])	$8439	$8950	$17 447
Cost per Mb	$84.39	$5.97	$5.81

a Total direct costs include the reagents and consumables, the labor, instrument amortization cost and the disc storage space required for data storage/access.

DNS szekvenálás félvezetőn: Ion Torrent

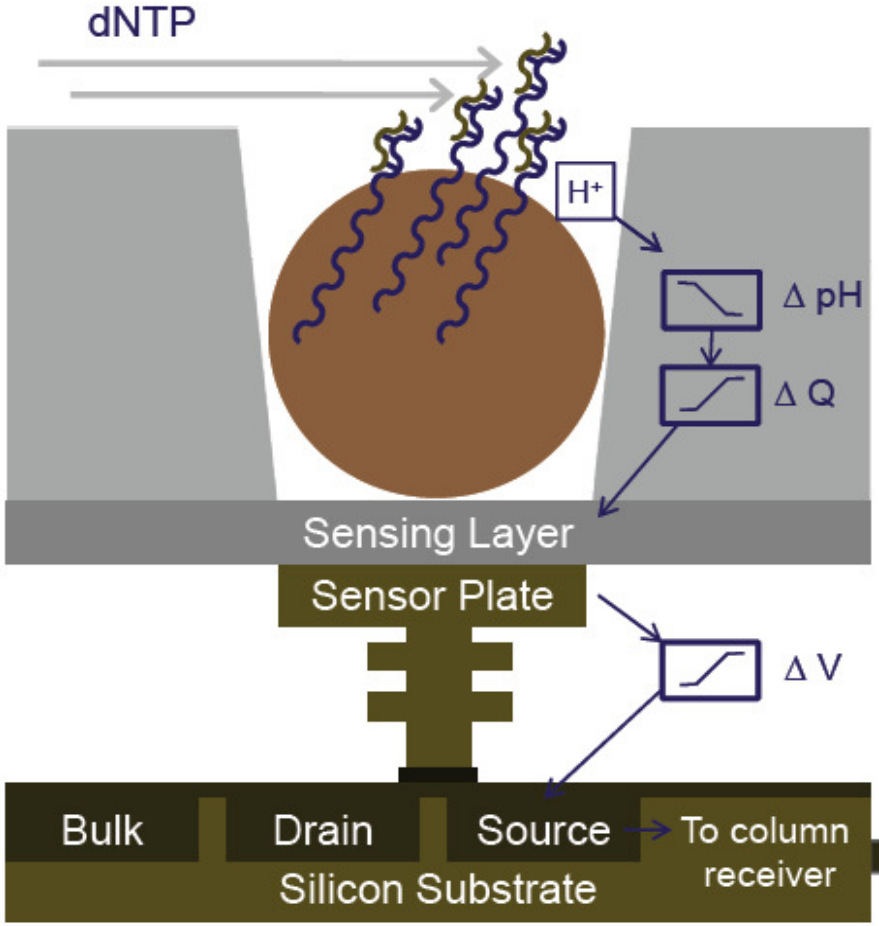
- hipergyors real-time szekvenálás





DNS szekvenálás félvezetőn: PGM

- hipergyors real-time szekvenálás

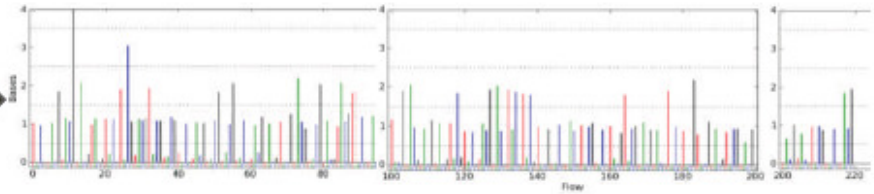


DNA → Ions → Sequence

- Nucleotides flow sequentially over Ion semiconductor chip
- One sensor per well per sequencing reaction
- Direct detection of natural DNA extension
- Millions of sequencing reactions per chip
- Fast cycle time, real time detection

Nincs PCR, fényextinkció, kamera, stb.

Helyette pH mérés mikrofluidokban



DNS szekvenálás félvezetőn: Ion Torrent

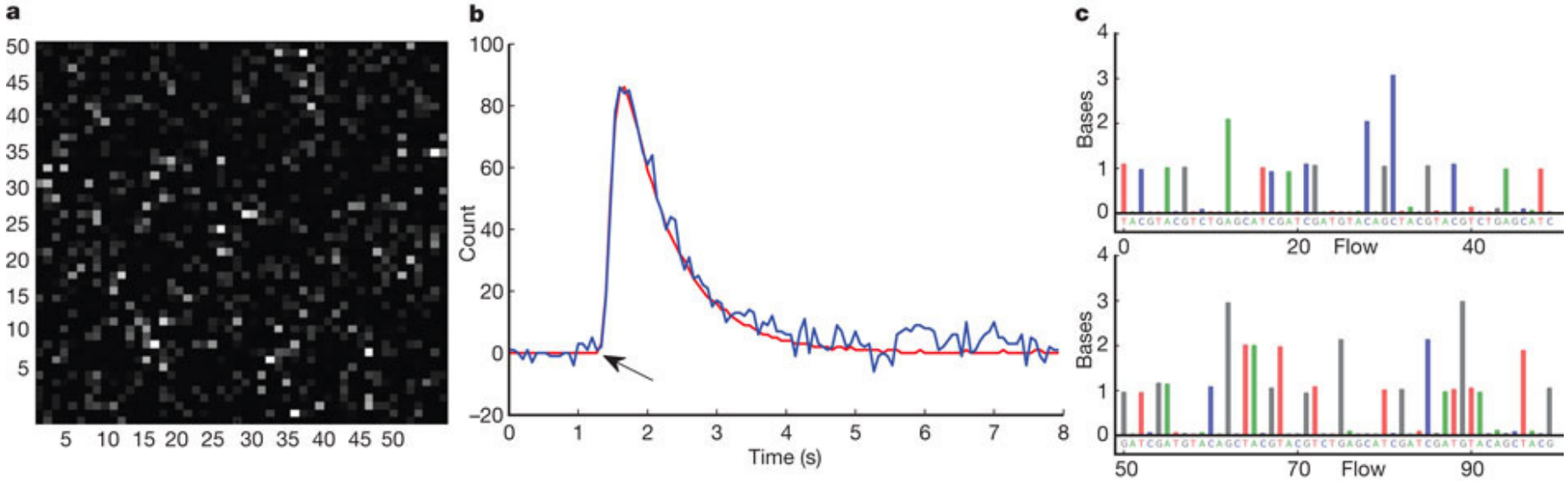


Table 1 | *Vibrio fisheri*, *E. coli*, *Rhodopseuomanas palustris* and *Homo sapiens*

	V. fisheri	*R. palustris*	*E. coli*	*E. coli*	*E. coli*	*H. sapiens*
GC content	38%	65%	51%	51%	51%	41%
Genome size	4.2 Mb	5.5 Mb	4.7 Mb	4.7 Mb	4.7 Mb	2.9 Gb
Number of runs x ion chip size	1 × 1.2 M	1 × 1.2 M	1 × 1.2 M	1 × 6.1 M	1 × 11 M	1,601 × 1.2 M 267 × 6.1 M 28 × 11.1 M
Fold coverage	6.2-fold	6.9-fold	11.3-fold	36.2-fold	58.4-fold	10.6-fold
Coverage	96.80%	99.64%	99.99%	100.00%	100.00%	99.21%
Reads ≥21 bases	261,313	444,750	507,198	1,852,931	2,594,031	366,623,578
Reads ≥50 bases	233,049	399,360	487,420	1,698,852	2,343,880	306,042,650
Reads ≥100 bases	156,391	160,726	400,743	1,012,918	1,779,237	139,624,090
Mapped bases	26.0 Mb	37.8 Mb	47.6 Mb	169.6 Mb	273.9 Mb	30.2 Gb

Coverage shows percentage of genome covered based on one or more reads mapping to each base of the reference genome. Reads align with 98% or greater accuracy.

JM Rothberg *et al. Nature* **475**, 348-352 (2011) doi:10.1038/nature10242

https://www.coursera.org/course/genomescience