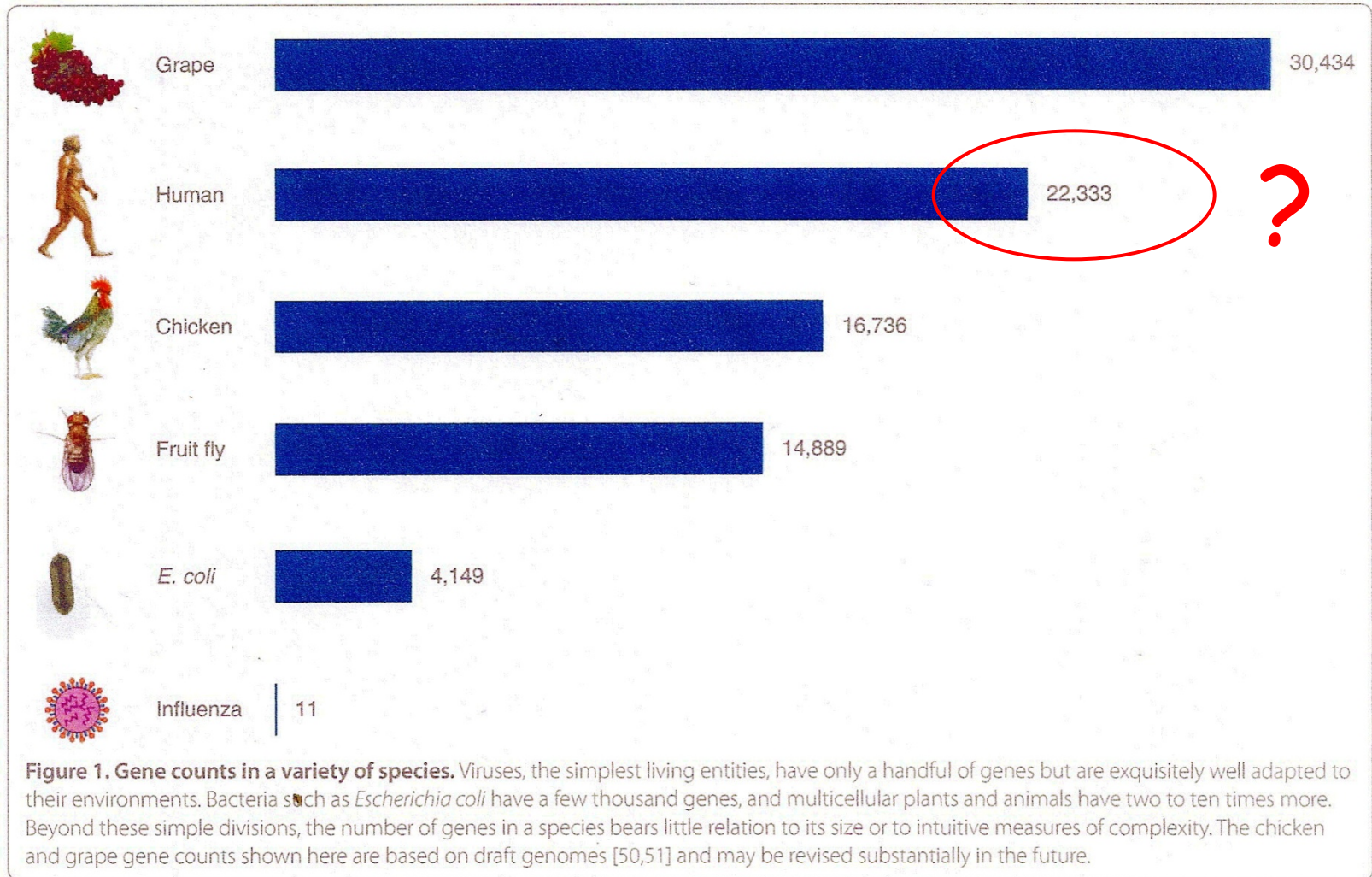


Genomika I.

Hogyan változtak az elképzelések a genom tartalmáról, a szerveződési komplexitás és génszám közti összefüggésről?

Genom szekvenálási stratégiák és új generációs genom szekvenálások (NGS)

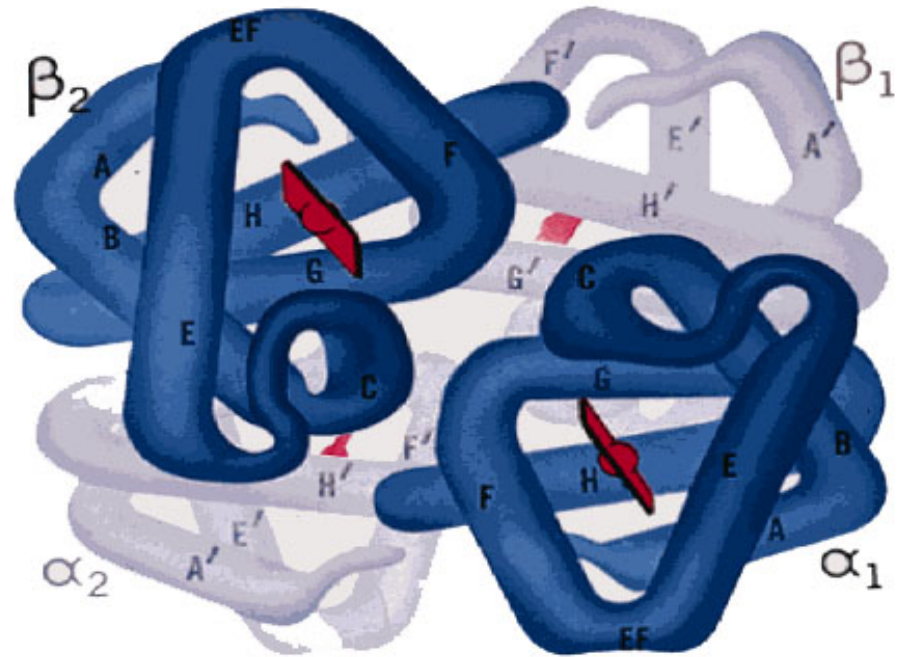
Humán genom: „... valahol a csirke és a szőlő között?”



Első becslések a genom méretéről és a gének számáról

1964: F. Vogel (Heidelberg)

- Hemoglobin α és β lánc
- leegyszerűsített feltevések
- Haploid genom: 3×10^9 bp
- Gének száma: 6,7 millió!



1990: NIH/DOE report on Human Genome Project

- becslés: 100 000 gén, az átlagos gén méret (30 000 bp) alapján

2001, Human Genome Project: csökkenő génszám, növekvő bizonytalanság

A génfogalom fejlődése: mit nevezünk ma egy génnek?

A Gén fogalmának jelentős átalakulása az elmúlt száz esztendőben:
protein/RNS kódolás, intron/exon fogalom, szabályozó funkciók, stb.

Gén

- 1950-es évektől
- a kódolási szabályok felismerése
- a DNS azon szakasza, amely egy transzkriptum (mRNS, rRNS, snRNS, tRNS, ...) átírásáért felelős genetikai információt tartalmazza

ORF

- *open reading frame* (nyitott leolvasási keret)
- 1990-es évektől (genomika korszaka)
- gének annotálása: bioinformatikai módszerekkel prediktálnak transzkriptumok átírását végző DNS szakaszokat (konzervált szekvencia motívumok alapján)

Egy gén a genetikai állományunk jól körülhatárolt szakasza, mely mRNS-ként átíródik és egy v. több fehérjét kódol. (pl. alternatív splicing: izoformák)

Human Genom Project (HUGO)

- Első kezdeményezés: 1980-as évek eleje
 - orvosbiológiai megközelítés, infrastruktúrális beruházások
- Folyamatban lévő genom szekvenálási projektek
 - λ -fág, SV40, humán mitokondriális genom
- Genetikai és fizikai térképezések
 - Botstein et al., 1980; 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 organismusokon, technológiai fejlesztések, bioetika

Universal Landmark

Sequence Tagged Site (STS) 1989

Replaces cloned DNA probe mapping landmarks with PCR assays.

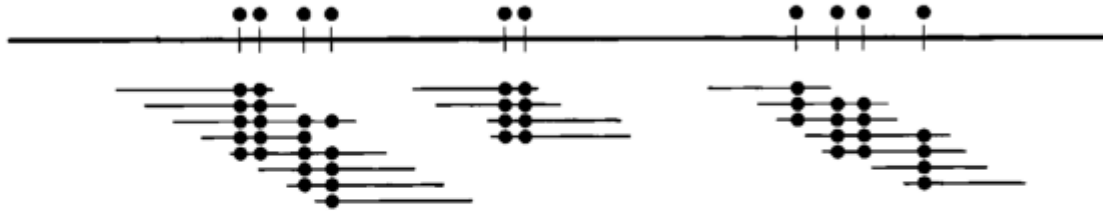
Each STS is uniquely described by a pair of oligonucleotides, a product size, and PCR reaction conditions. Can be stored and distributed electronically.

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

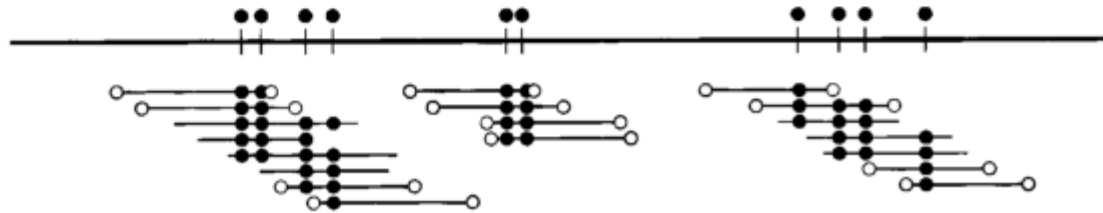
Eliminates the need for huge collections of cloned probe segments upon which prior maps depended.

Clone ends – Clone-based Physical Map

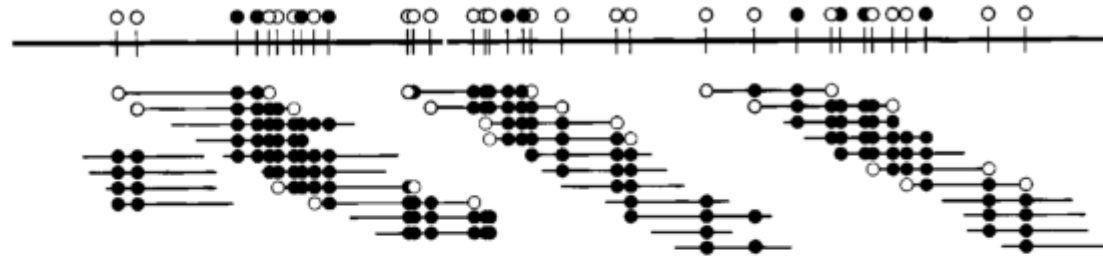
a. Screen library with existing markers



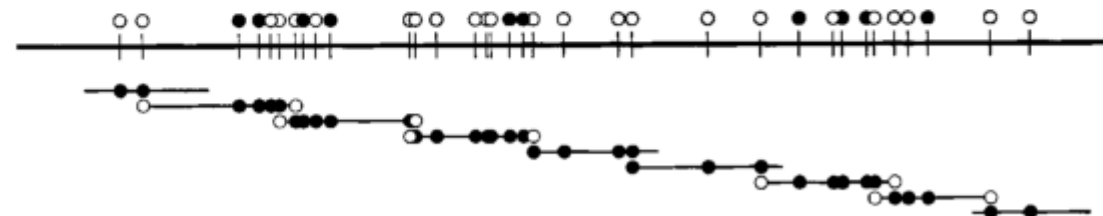
b. Generate new markers



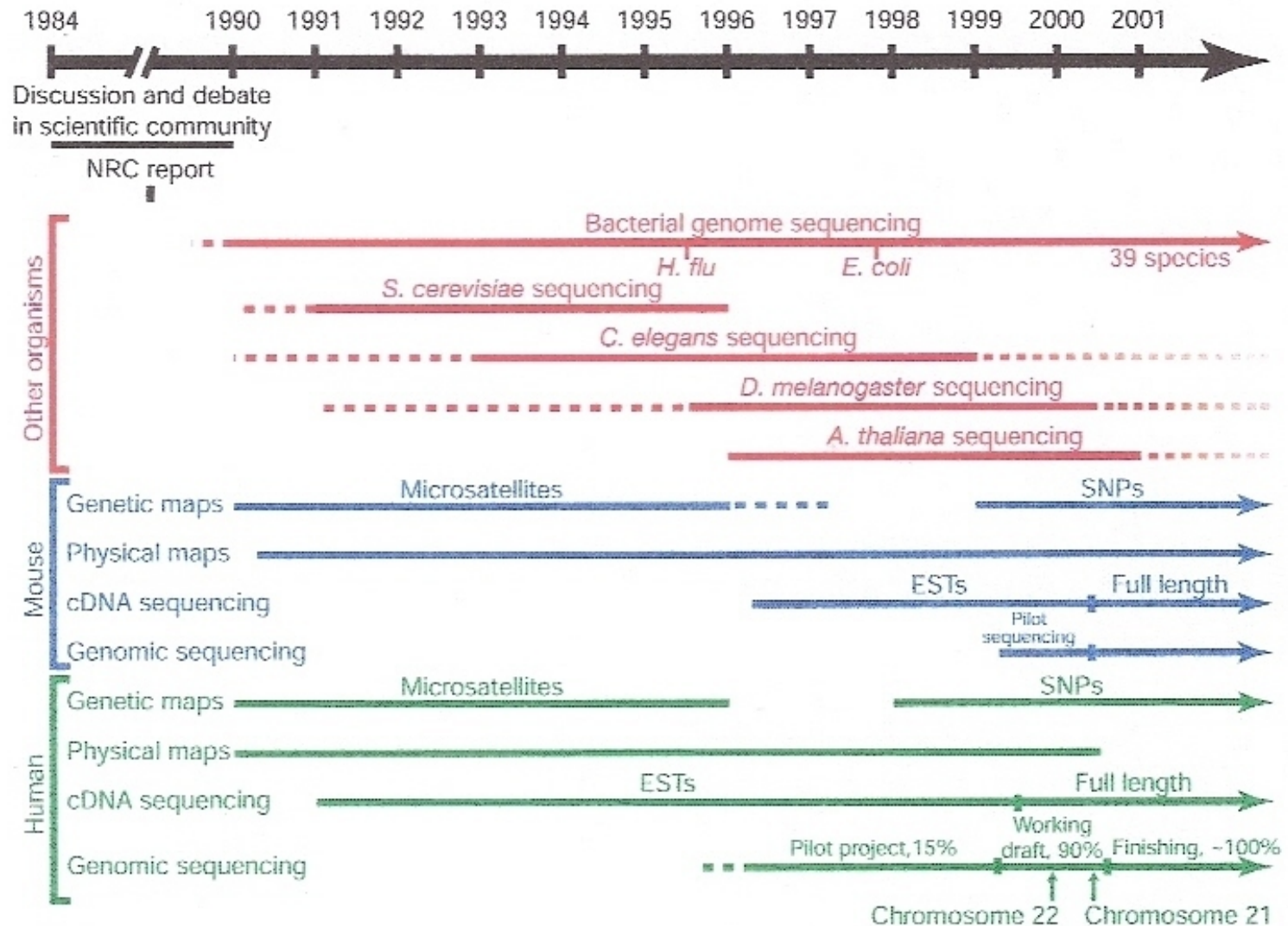
c. Screen library with new markers



d. Determine tiling path



Genom projektek időskálán - a „hőskor”



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.

„shotgun” genom szekvenálási stratégiák

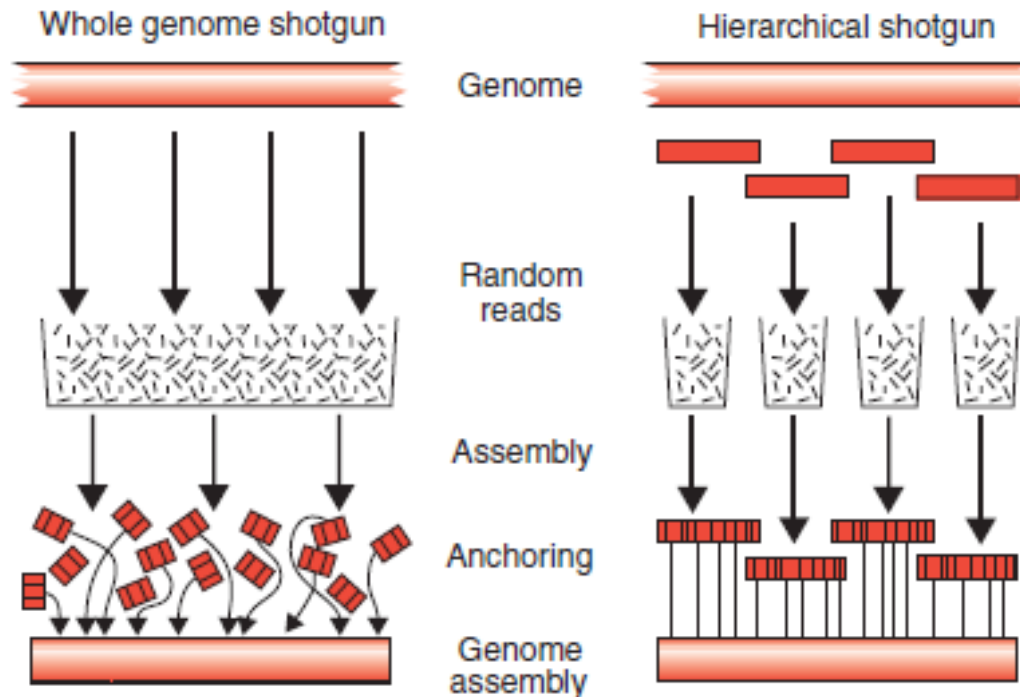
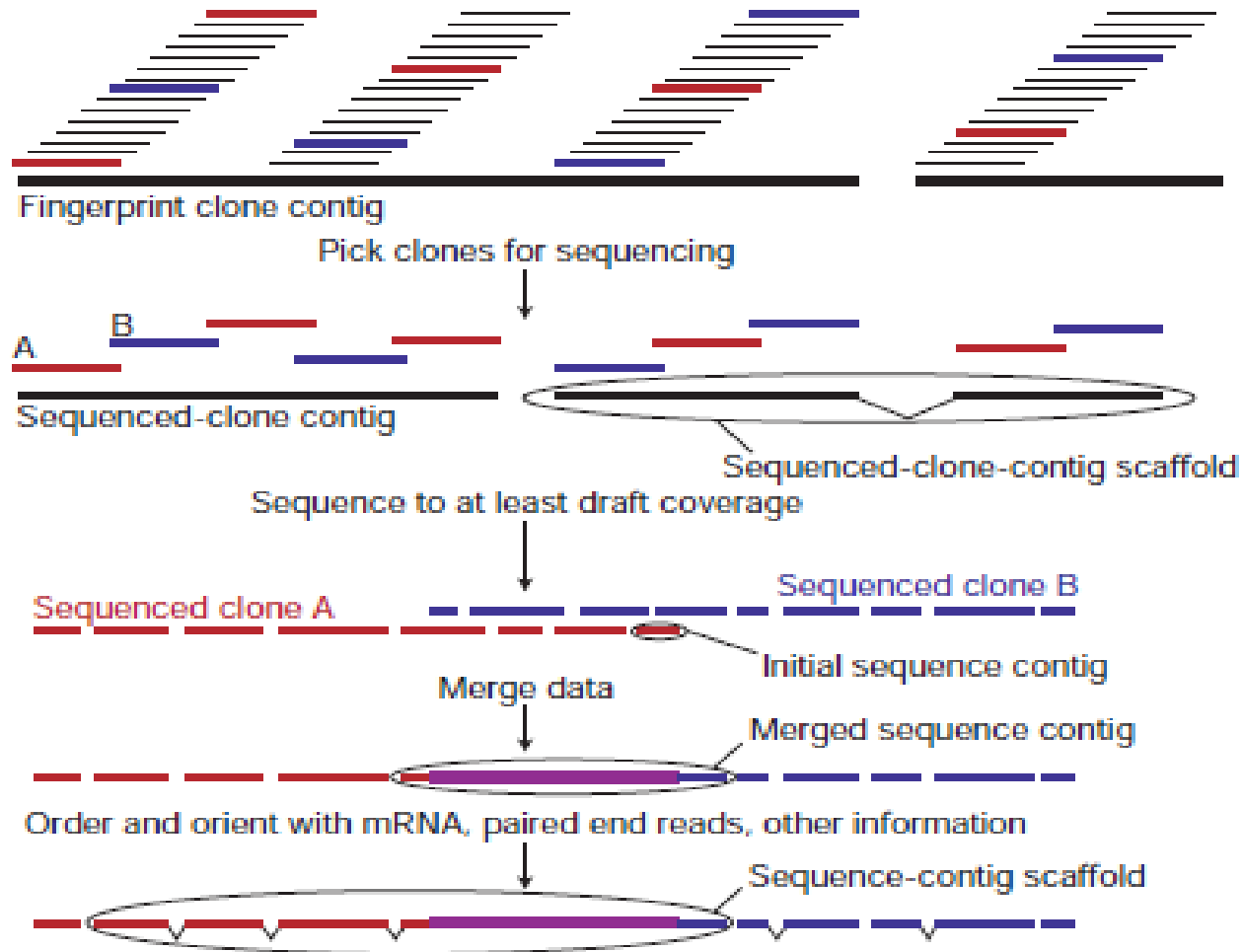


Figure 9.11. Assembling genomic data using the hierarchical and whole genome shotgun approaches. Adapted from Waterston, Lander and Sulston (2002), with permission

Genom szekvenciaváz összeállítása



Teljes genom összeszerelés

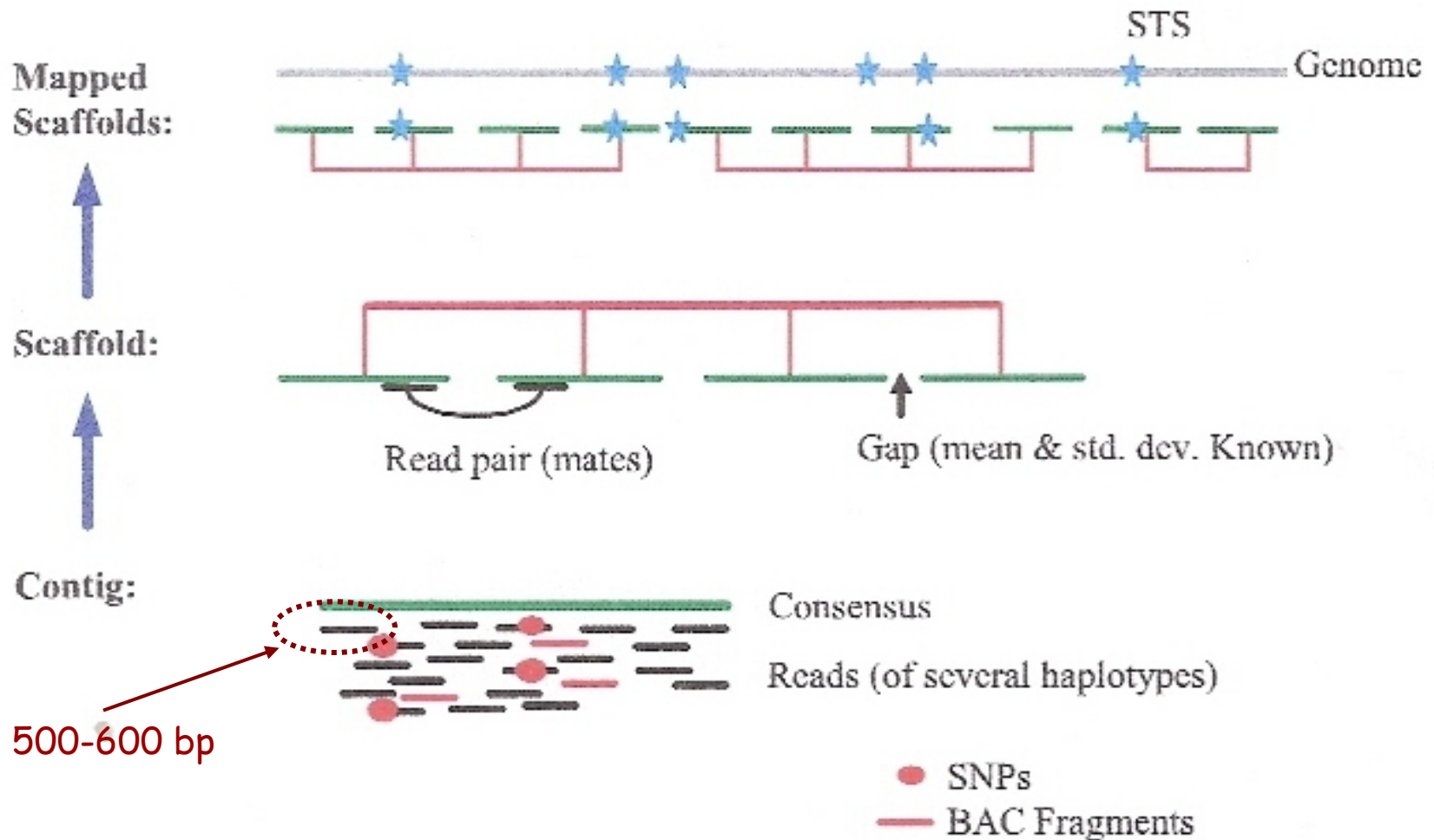


Fig. 3. Anatomy of whole-genome assembly. Overlapping shredded bactig fragments (red lines) and internally derived reads from five different individuals (black lines) are combined to produce a contig and a consensus sequence (green line). Contigs are connected into scaffolds (red) by using mate pair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

STS genom térképezés

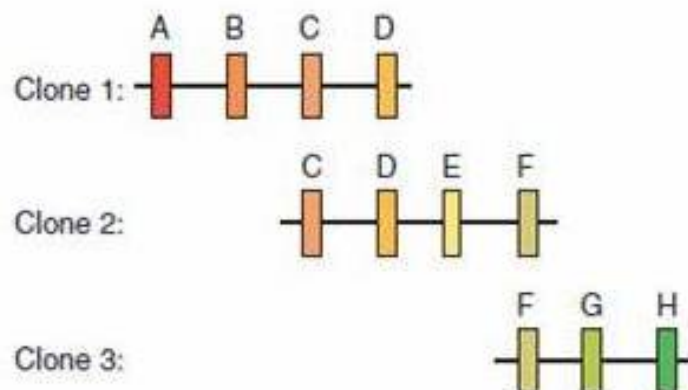


Figure 9.5. Aligning clones by STS mapping. Each clone contains several STSs. Clone 1 has four (A, B, C and D). Clone 2 also contains STSs C and D. Therefore clones 1 and 2 overlap with each other

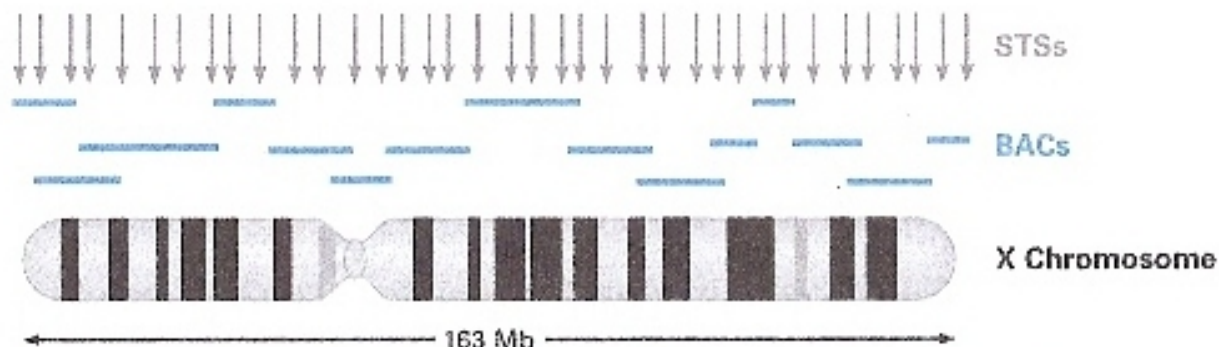


FIGURE 1.3 Relationships of chromosomes to genome sequencing markers. The X chromosome is about 163 Mb in length. In this diagram, there are 16 overlapping BAC clones that span the entire length. In reality, 1,408 BACs were needed to span the X chromosome. Arrows (top) mark STSs scattered throughout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

Humán Genom Projekt

Science

16 February 2001

Vol. 291 No. 5507
Pages 1145-1434 \$9

THE HUMAN GENOME



15 February 2001

nature

£5.45 €6.23 ¥154.0M\$ US\$16.00

www.nature.com

the human genome

Nuclear fission

Five-dimensional
energy landscapes

Seafloor spreading

The view from under
the Arctic ice

Career prospects

Sequence creates new
opportunities

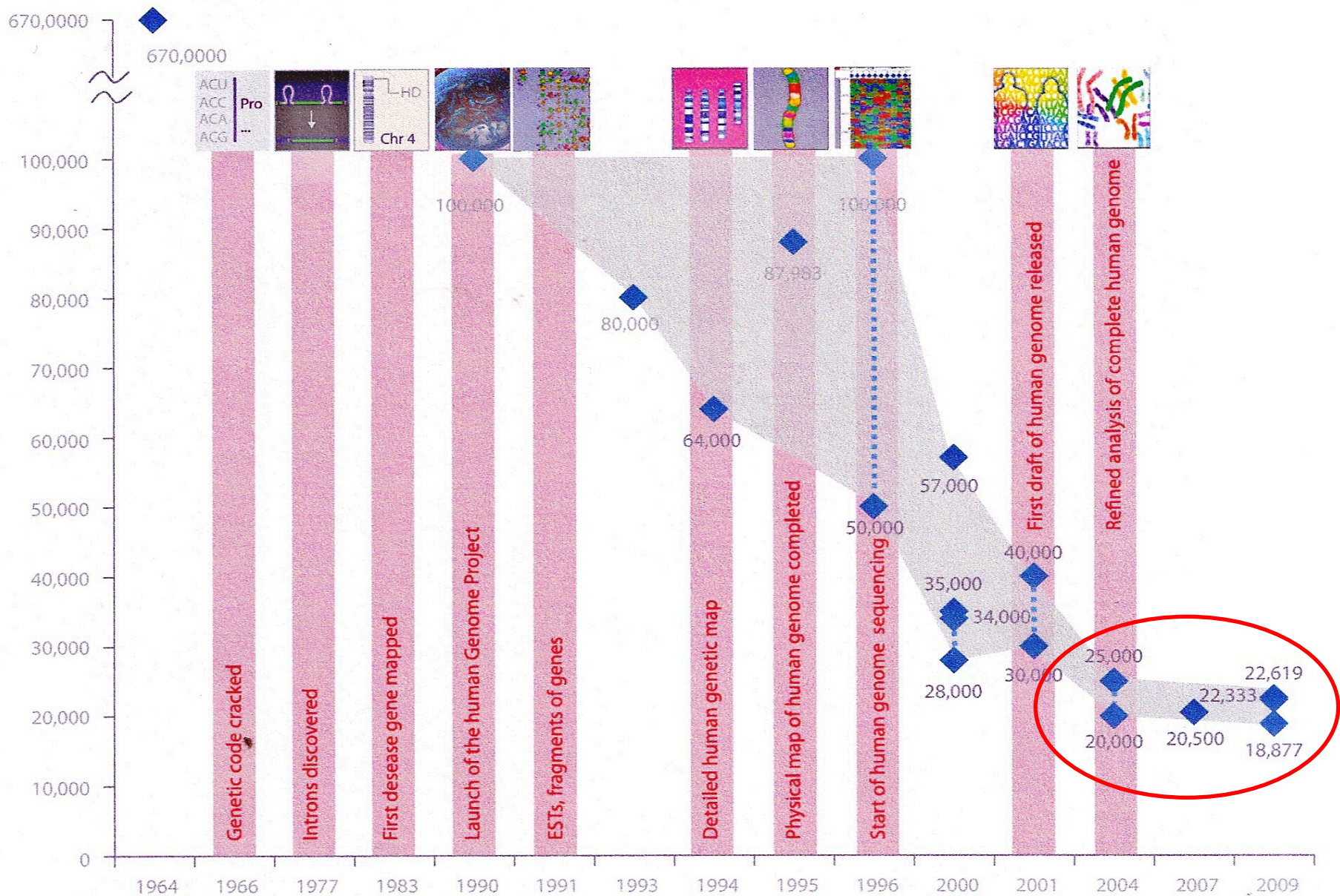


Figure 2. The trend of human gene number counts together with human genome-related milestones. Individual estimates of the human gene count are shown as blue diamonds. The range of estimates at different times is shown by the two vertical blue dotted lines. Note how this range has narrowed in recent years.

Hol tartunk most?

2001, Human Genome Consortium: 30 000 - 40 000 protein kódoló gén

Celera Consortium: 26 500 „erős” + 12 000 „gyenge” bizonyíték

2004, Human Genome Consortium: 20 000 - 25 000 gén

- kevesebb mint az Arabidopsis → szervezeti komplexitás?

2010, Ensembl: 22 619 / NCBI: 22 333 protein kódoló gén

CCDS: 18 173 (<http://www.ncbi.nlm.nih.gov/CCDS/CcidsBrowse.cgi>)

fals pozitívak: retrotranszpozonok, pszeudogének, „orphan” DNS

2016.09.12.: CCDS: 18 892 genes > 1 CCDS ID: 7 576

Új gének és gén átrendeződések

- CGH analízisek: rokon fajok között kb. azonos génszám
- *de novo* gén keletkezés: génduplikáció és specializáció
- *de novo* eredet: új humán gének? (Knowles and McLysaght, 2009)
- génszám eltérések egyének között: segmental duplications
- large-scale copy number polymorphisms (CNVs)
- emberi „pángenom”: változatok rasszok, csoportok között.

(Li R, et al., 2010, Nat Biotechnol, 28:57-63)

- kb. 40 Mb új szekvencia, + 1,3 %

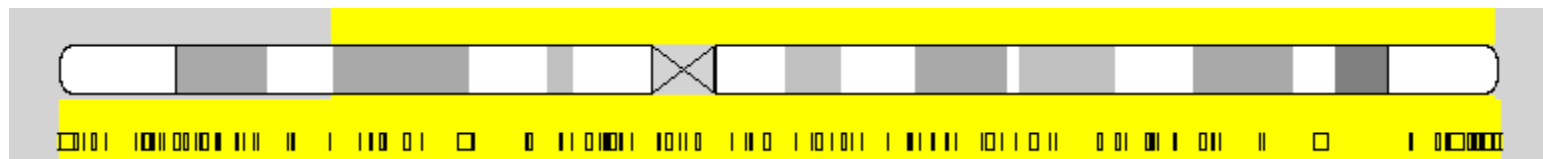
Copy Number Variation (CNV)

Kópia-szám variabilitás

A diploid szervezetek alapesetben minden génből két másolatot hordoznak (homológ párok). Az emberi genom vizsgálata során felismerték, hogy hosszú (ált. több Kb vagy Mb) DNS szakaszok előfordulhatnak kettőnél több példányban is. Ezeket copy-number variation (CNV)-nak nevezték el. Az egyes egyének között a CNV mintázat különböző lehet.

Kb. 300 emberen végzett vizsgálatban 1447 CNV-t mutató genomikus szakaszt azonosítottak, ez kb. a genom 12%-át fedti le.

Sikerült néhány CNV-t betegségekkel kapcsolatba hozni. Pl. a prosztataraák betegség az UGT2B17 gén kópia szám változataival hozták kapcsolatba. Vagy a HIV fertőzéssel szembeni ellenálló képesség a CCL3L1 gén több mint két példányával kapcsolatos.



Az ember 20. kromoszómáján kimutatott CNV-k helyzete és kiterjedése

Table 1. Novel human protein-coding genes and supporting evidence.

Gene name	Ensembl ID	Length (codons)	Longest chimp ORF ^a	Expression support and tissue ^b	Primate shared disablers ^c	Other major sequence differences	Presence of enabler in other human complete genome sequences ^d	HapMap SNPs
<i>CLLU1</i>	ENSG00000205056	121	42	EST/cDNA: Blood (<u>AJ845165</u> , <u>AJ845166</u>); UniGene: Blood, embryonic tissue, eye, lymph, lymph node, muscle, pharynx, tonsil (Hs.339918)	1-bp indel ^e	Macaque: 4- and 1-bp indels	Sequence available and enabler conserved in all	1 syn.; 1 nonsyn.
<i>C22orf45</i>	ENSG00000178803	159	87 (25 amino acids align with human sequence)	EST/cDNA: Kidney, other (<u>AX747284</u> , <u>AK091970</u> , <u>DA635985</u>); ArrayExpress: Sperm, lung (E-GEOD-6872, E-GEOD-3020)	Premature stop codon	Chimp: 1-bp indel; Macaque: lacks ATG start codon; 4-bp indel	Reverse strand is available and conserved in Venter	1 nonsyn.
<i>DNAH10OS</i>	ENSG00000204626	163	90 (75 amino acids align with human sequence)	EST/cDNA: Hippocampus (<u>AK127211</u>); UniGene: Blood, embryonic tissue, eye, lymph, lymph node, muscle, pharynx, tonsil (Hs.339918)	10-bp indel	Chimp: 2- and 1-bp indels; Macaque: lacks ATG start codon; 13-, 8-, 1-, and 1-bp indels	Reverse strand is available and conserved in Venter, Watson and HuAA	1 syn.; 1 nonsyn.

^aLength in codons of longest in-frame (alignable) ORF starting from any ATG in the region.

^bType of data/database is listed followed by tissue information with database identifiers in parentheses. Underlined accession numbers are full-length, spliced cDNA.

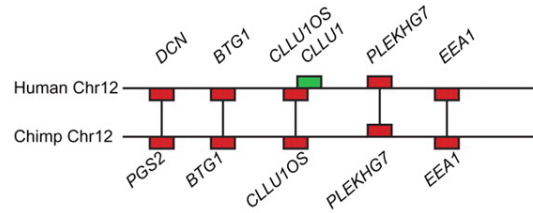
^cShared disablers are sequence differences shared by chimp, gorilla, orangutan, gibbon, and macaque that eliminate the capacity to produce a protein similar to the human protein.

^dIndependently sequenced whole genomes: Venter, Watson, HuAA, HuBB, HuCC, HuDD, and HuFF. All data are listed where available.

^eNot shared with orangutan.

Sequence changes in the origin of *CLLU1* from noncoding DNA. (A) Region of conserved synteny between human and chimp chromosomes 12.

A



B

Start

Human
Chimpanzee
Macaque

```
GTTTGGAGG - - - ATGTTCAAC AAATGCTCCTTTCA TTCTCTATTTACAGACCTGCCGCA
GTTTGGAGG - - - ATGTTCAAT AAATGCTGCTTTCA CTCTCTATTTACAGACCTGCCGCA
GTTTGGAGG - - - ATGCTCAAT AAATGCTCCTTTCA TTCTCTCATTACA AACTGCCGCA
```

Human
Chimpanzee
Macaque

```
GACAATTC TGCTAGCAGCC TTTGTGCTATTATCTGTTTTCTAAAC TTAGTAATTGAGTGT
GACAATTC TGCTAGCAGCC TTTGTGCTATTATCTGTTTTCTAAAC TTAGTAATTGAGTGT
GACAATTC TGCTAGCAGCC TTTGTGCTATTATCTGTTTTCTAAAC TTAGTAATTGAGTGT
```

Human
Chimpanzee
Macaque

```
GATCTGGAGACTAA - CTCTGAAATA AAATAAGCTGATTATTTATTTATTTTCTCAAAACAA
GATCTGGAGACTAAACTCTGAAATA AAATAAGCTGATTATTTATTTATTTTCTCAAAACAA
TATCTGGAGACTAAACTCTGAAATA AAATAAGCTGATTATTTATTTATTTTCTCAAAACAA
```

Human
Chimpanzee
Macaque

```
CAGAATACGATTTAGCAAATTACTTCTTAAGATAT TATTTTACATTTCTATATTTCTCCTA
CAGAATACGATTTAGCAAATTACTTCTTAAGATACTTATTTTACATTTCTATATTTCTCCTA
CAGAATA TGATTTAGCAAATTACTTCTTAAGATAT TATTTTGCAC TTCTATATTTCTCCTA
```

Human
Chimpanzee
Macaque

```
CCCTGAGTTGATGTGTGAGC AATATGTCAC TTTCAT AAAGCCAGGTATACA - - - TTATG
CCCTGAGTTGATGTGTGAGC CGTATGTCAC TTTCAT AAAGCCAGGTATACA - - - TTATG
CCCTGAGTTGATGTGTGAGC AATATGTCAC TTTCAC AAAGCCAGGTATATATACATTACG
```

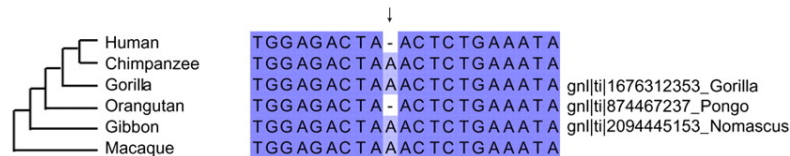
Human
Chimpanzee
Macaque

```
GACAGGTAAGTAAAAAACATATTTATTTATTTCTACGTTTTGTCCAAAAATTTTAAATTTCT
GACAGGTAAGTAAAAAACATATTTATTTATTTCTACGTTTTGTCCAAAAATTTTAAATTTCT
GACAGGTAAGTAAAAAACATATTTATTTATTTCTACGTTTTGTCCAAAAATTTTAAATTTCT
```

Human
Chimpanzee
Macaque

```
AACTGTTGCGCGTGTGTTGGTAA - - - TGTA AAACAAACTCAGTACA
AACTGTTGCGCGTGTGTTGGTAA - - - TGTA AAACAAACTCAGTACA
AACTGTTGCGCATGTGTTGGTAA - - - CGTA AAACAAACTCAGTACG
```

C



Knowles D G , McLysaght A Genome Res. 2009;19:1752-1759



TABLE 3.1 Approximate fractional composition of the human genome

TYPE OF DNA	FRACTION
Coding exons	0.008
Internal introns	0.308
5' Untranslated regions	
Exons	0.045
Introns	0.002
3' Untranslated regions	
Exons	0.006
Introns	0.001
Intergenic DNA	0.683
Conserved noncoding DNA	0.016
Pseudogenes	0.007
Mobile genetic elements	0.446

Note: Derived from various references given in the text. Intergenic DNA is all DNA except coding exons and internal introns. The fractions do not sum to one because mobile elements, pseudogenes, and transcription factor binding sites reside in introns, UTRs, and/or intergenic DNA.

TABLE 3.2 Haploid genome size, number of protein-coding genes, and average number of nucleotides per gene for some well-characterized eukaryotic genomes

	GENOME SIZE (MB)	GENE NUMBER	KILOBASES/GENE		
			TOTAL	CODING	NON-CODING
Unicellular species					
<i>Encephalitozoon cuniculi</i>	2.90	1997	1.45	1.01	0.44
<i>Saccharomyces cerevisiae</i>	12.05	6213	1.94	1.44	0.50
<i>Schizosaccharomyces pombe</i>	13.80	4824	2.86	1.43	1.43
<i>Cyanidioschyzon merolae</i>	16.52	5331	3.10	1.55	1.55
<i>Cryptococcus neoformans</i>	19.05	6572	2.89	1.62	1.27
<i>Plasmodium falciparum</i>	22.85	5268	4.34	2.29	2.05
<i>Entamoeba histolytica</i>	23.75	9938	2.39	1.14	1.25
<i>Leishmania major</i>	33.60	8600	3.91	2.15	1.76
<i>Thalassiosira pseudonana</i>	34.50	11242	3.07	0.99	2.08
<i>Trypanosoma</i> spp.	39.20	10000	3.92	1.96	1.96
Oligocellular species					
<i>Ustilago maydis</i>	19.68	6572	2.99	1.84	1.15
<i>Aspergillus nidulans</i>	30.07	9541	3.15	1.57	1.58
<i>Dictyostelium discoideum</i>	34.00	9000	3.78	2.45	1.33
<i>Neurospora crassa</i>	38.64	10082	3.83	1.44	2.39
Land plants					
<i>Arabidopsis thaliana</i>	125.00	25498	4.90	1.80	3.10
<i>Oryza sativa</i>	466.00	60256	7.73	1.18	6.55
<i>Lotus japonicus</i>	472.00	26000	18.15	1.35	16.80
Animals					
<i>Caenorhabditis elegans</i>	100.26	21200	4.73	1.25	3.48
<i>Drosophila melanogaster</i>	137.00	16000	8.56	1.66	6.90
<i>Ciona intestinalis</i>	156.00	16000	9.75	0.95	8.80
<i>Anopheles gambiae</i>	278.00	13683	20.32	1.64	18.68
<i>Fugu rubripes</i>	365.00	38000	9.61	0.93	8.68
<i>Bombyx mori</i>	428.70	18510	23.16	1.66	21.50
<i>Gallus gallus</i>	1050.00	21500	48.84	1.44	47.40
<i>Mus musculus</i>	2500.00	24000	83.33	1.30	82.03
<i>Homo sapiens</i>	2900.00	24000	96.67	1.33	95.36

Source: Lynch 2006a.

Gének száma

vs.

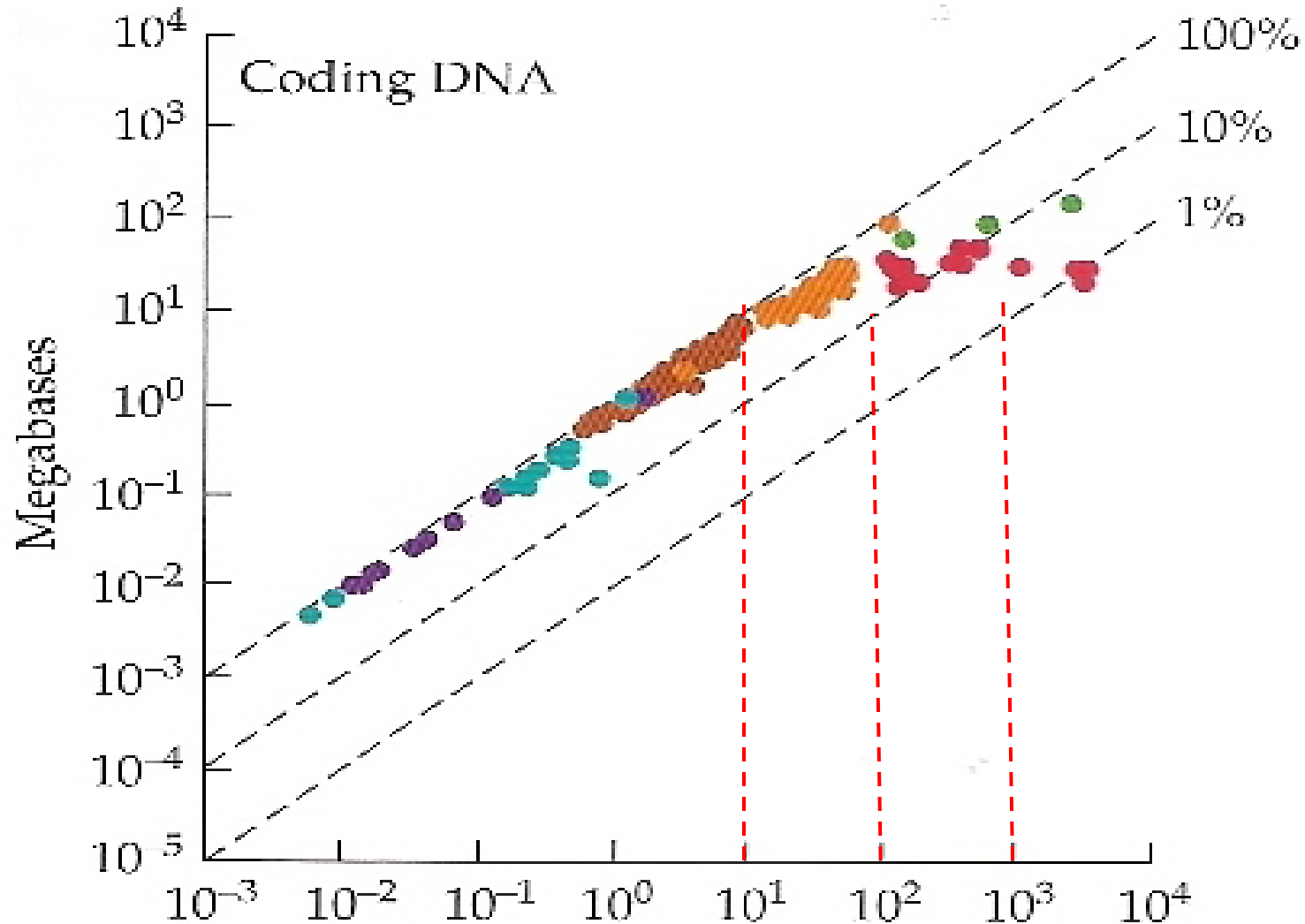
Kódoló szekvenciák
hossza

Genom méret

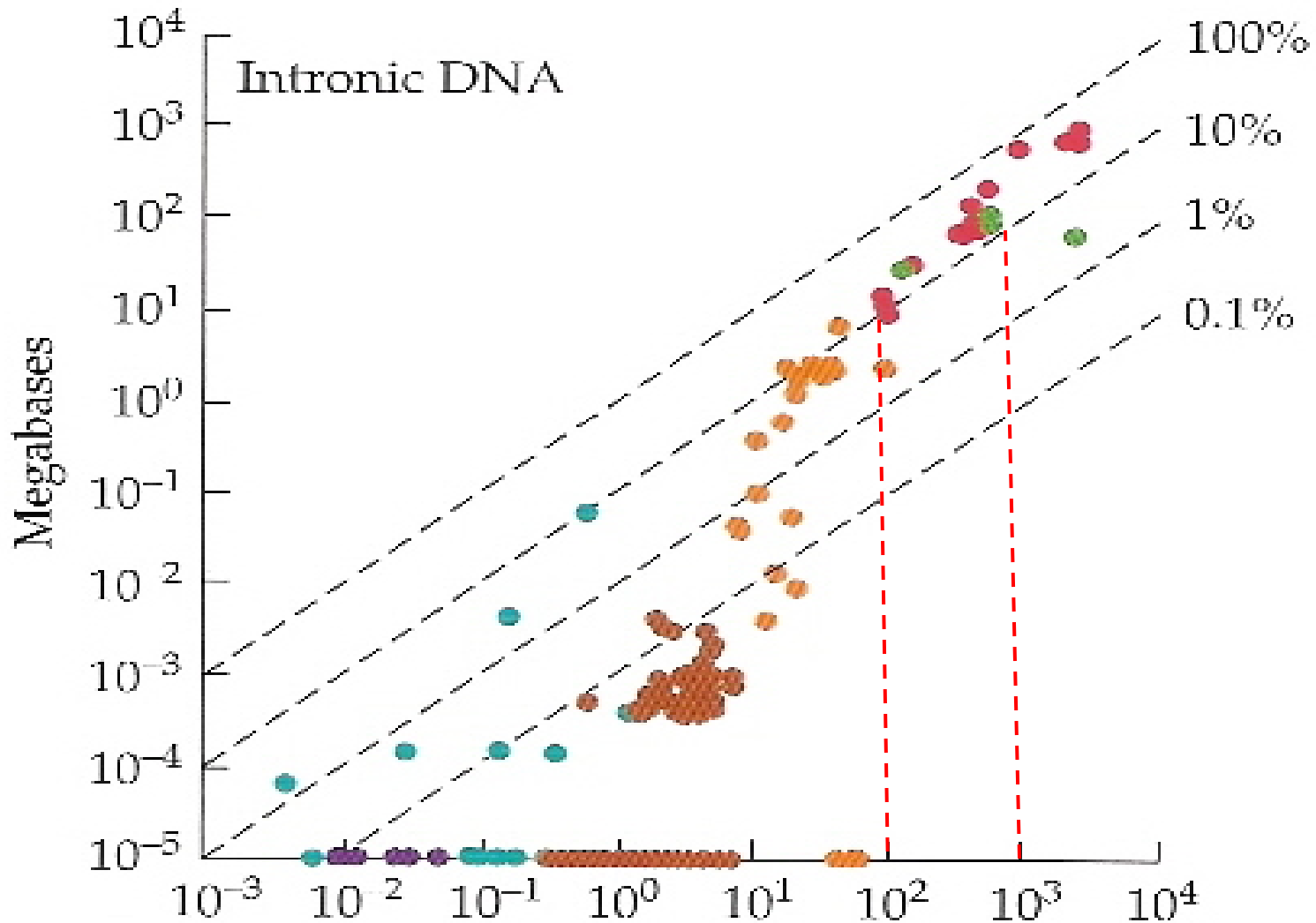
vs.

Nem-kódoló
szekvenciák hossza

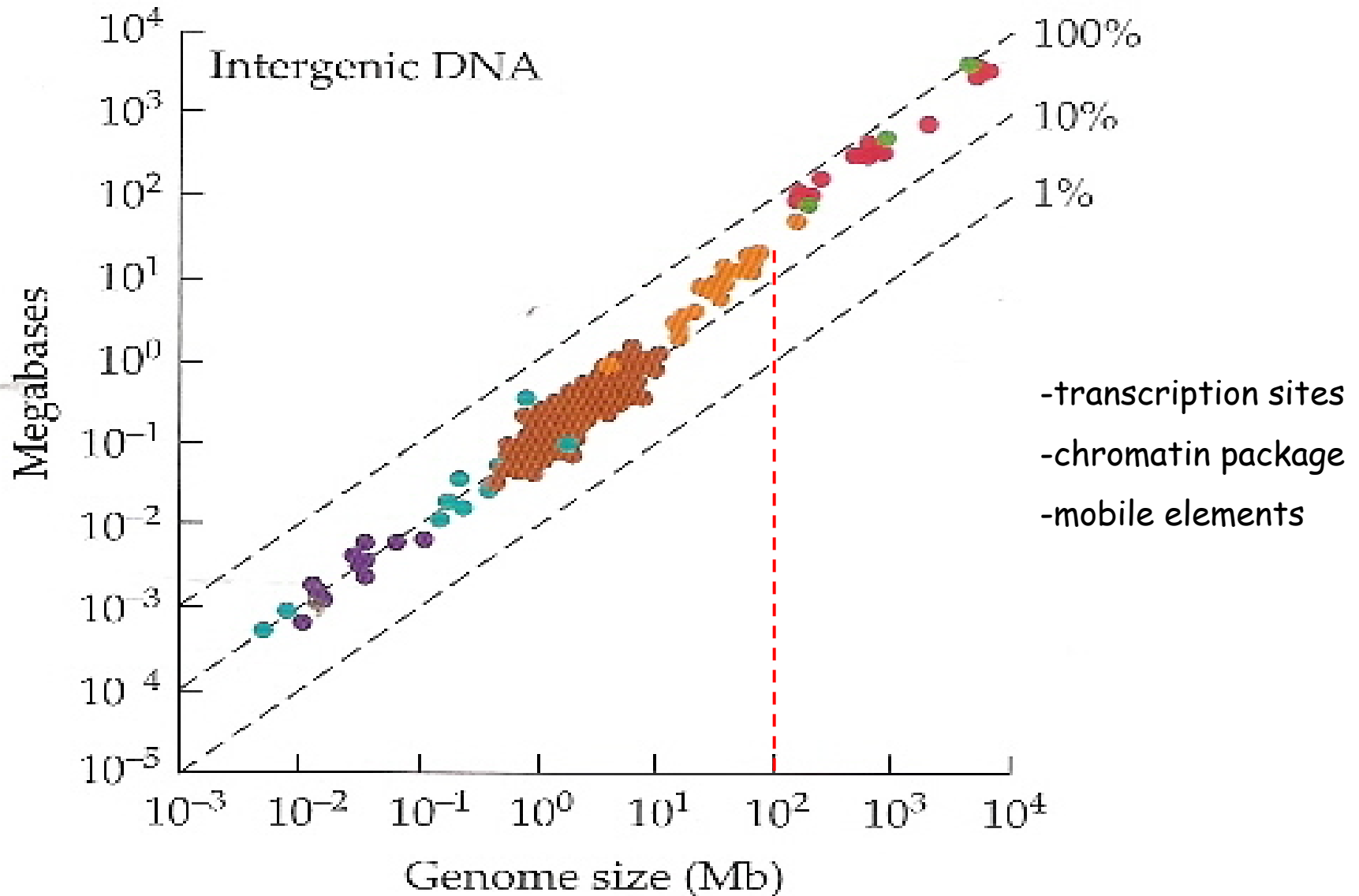
Genom méret vs. kódoló szekvenciák



Genom méret vs. intronok



Genom méret vs. intergénikus DNS



Genom méret és szerkezeti komplexitás

- WGC: rekurrens mutációk teljes egyedi genomok összehasonlítása révén (pl. Y-kromoszóma), divergens kromoszómák
- Prokarióta: 350-8000 gén, 0.5 - 9 Mb genom
- Multicelluláris Eukarióta: > 13.000 gén, > 100 Mb genom
- Noncoding DNA expanzió (intronok, mobilis elemek, pseudogének)
- Organizmus mérete vs. sejttípusok száma - pozitív korreláció
- Génszám / genom méret vs. multicellularitás / szerkezeti komplexitás

Van korreláció? Nem a genom mérettől v. génszámtól függ, hanem ahogy a gének működnek (transzkripciós szabályozás, alternatív splicing, stb.)

Genom szekvenálás: Technológia és Bioinformatika

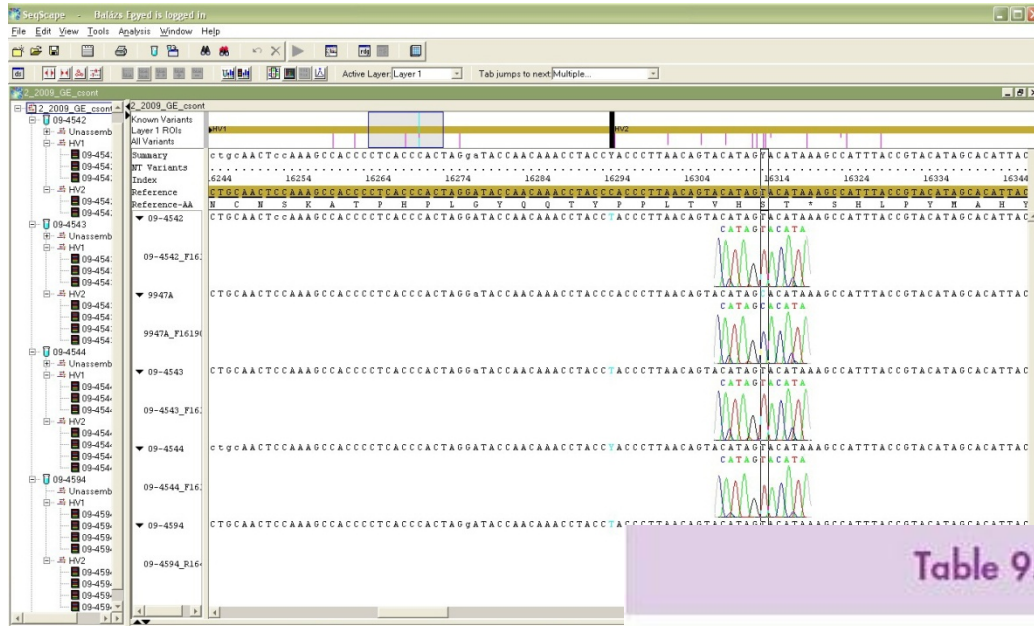


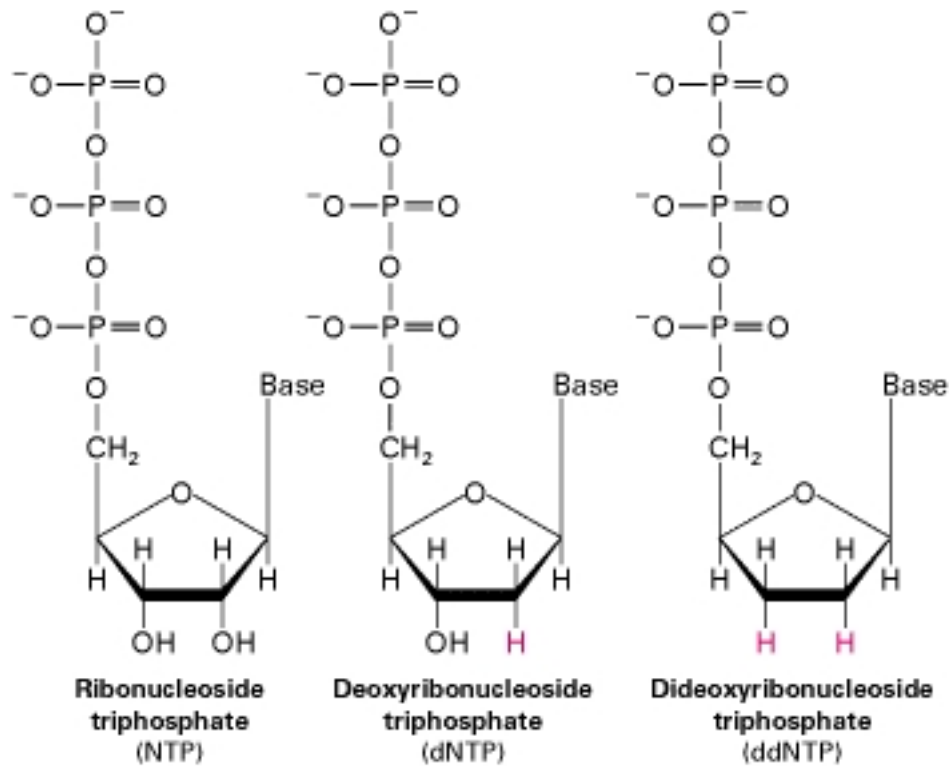
Table 9.1. Curated genome sequencing projects

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

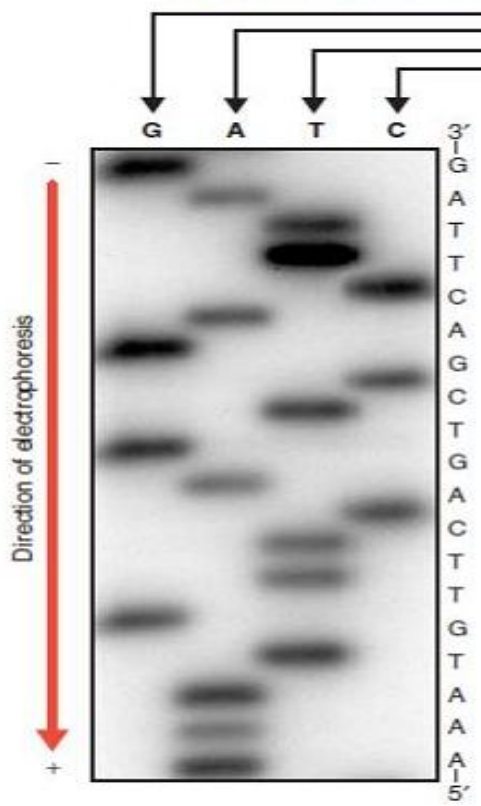
Genom szekvenálás: Technológia és Bioinformatika

- 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)

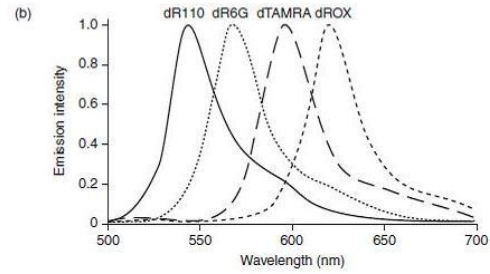
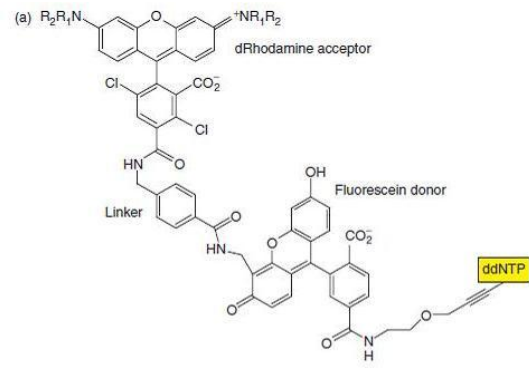
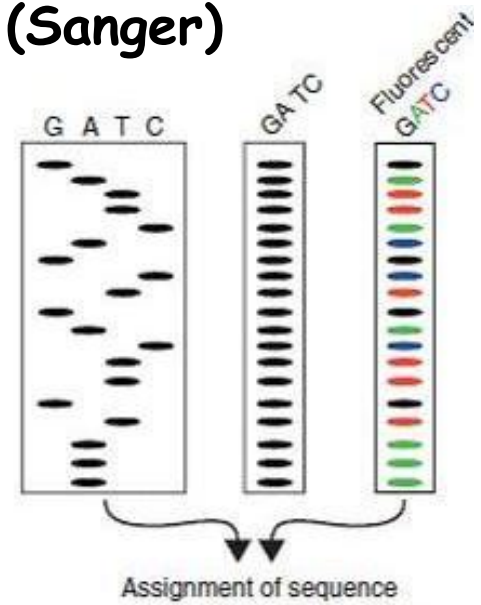
Sanger-szekvenálás



5' - CTAAGTCGACTGAACATTGTCAATGCATCGATC - 3'
 3' - GATTCAGCTGACTTGTAAACAGTACGCTAGCTAG - 5'
 3' - AGTACGCTAGCTAG - 5'
 Sequencing primer





DNS szekvenálás láncterminálással (Sanger)



BigDye Terminator DNA Sequencing

DNA template 3' - TAAATGATTCC - 5'

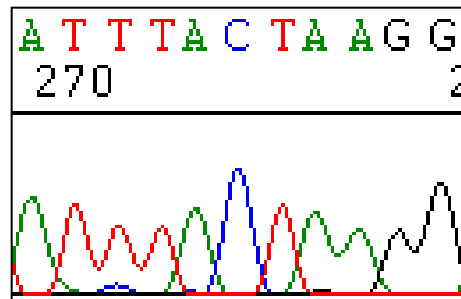
5'   3'

*Primer
anneals*

- A ●
- AT ●
- ATT ●
- ATTT ●
- ATTTA ●
- ATTTAC ●
- ATTTACT ●
- ATTTACTA ●
- ATTTACTAA ●
- ATTTACTAAG ●
- ATTTACTAAGG ●

*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)*

DNS szekvenálás: Technológia és Bioinformatika

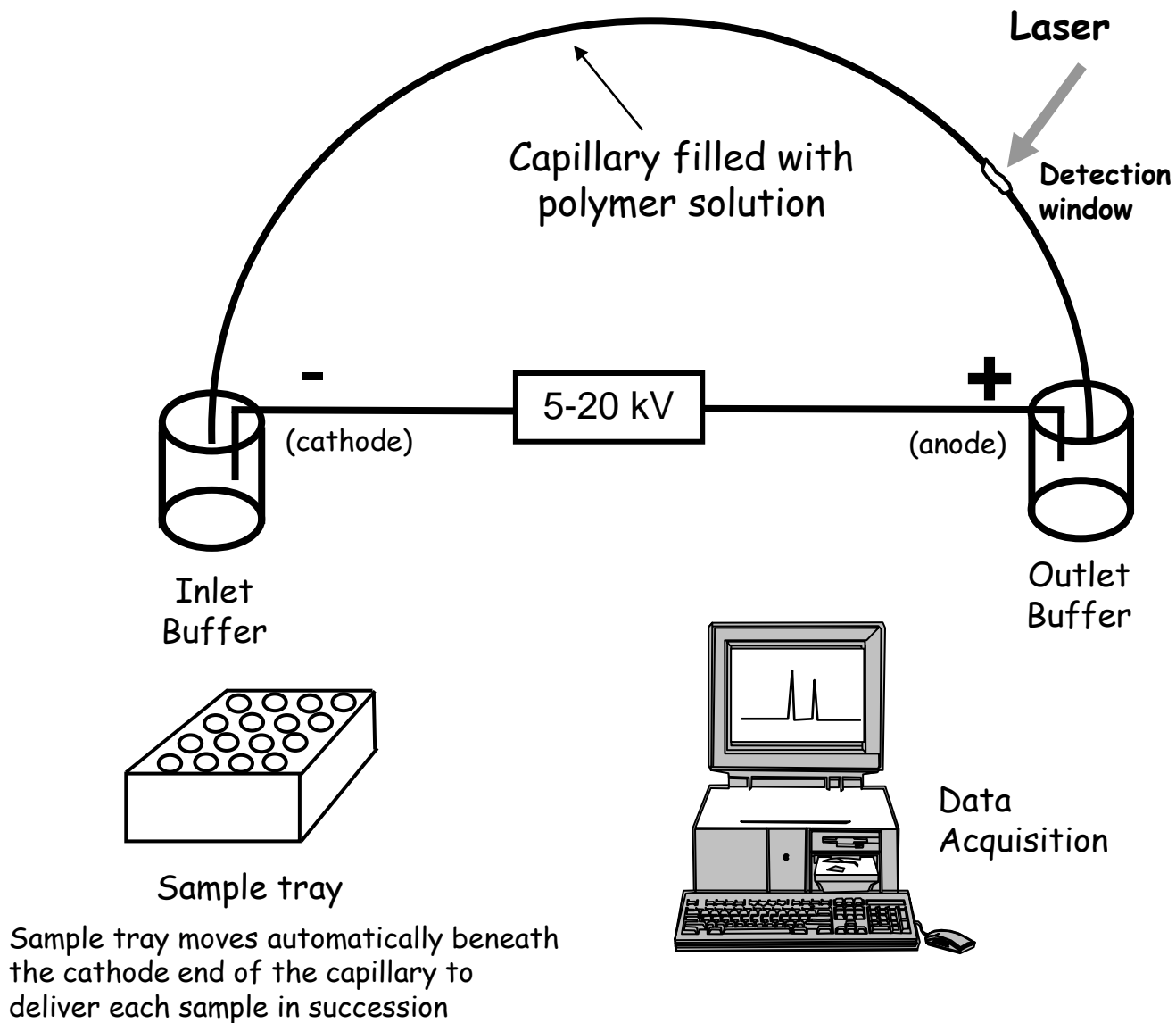


Figure 10.9, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

BRCA1 és BRCA2 gén resequencing

- mutációk diagnosztikai célú azonosítása

BRCA1 és BRCA2: 23 és 27 exon (80Kb)

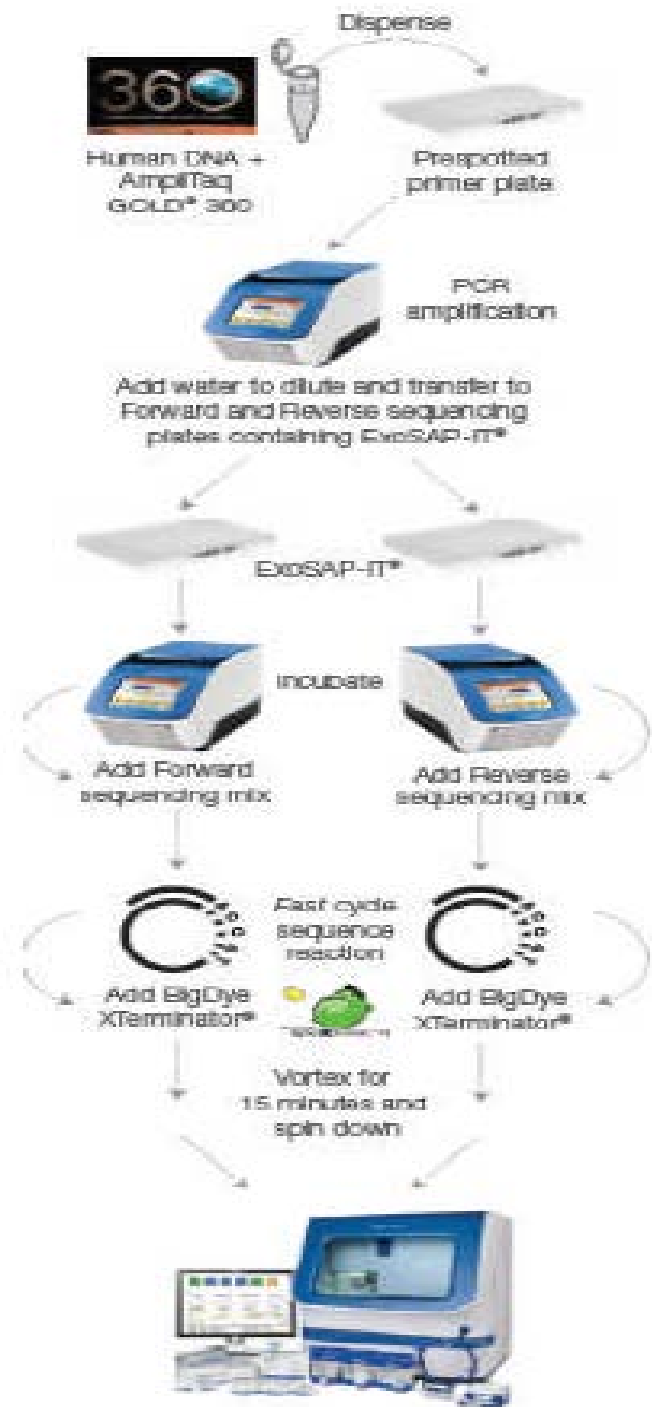
Nincs elő screening: SSCP, DGGE, dHPLC, stb.

Egy minta - egy assay koncepció

Gyors, pontos, teljes lefedettséget ad

Nincs kereszt kontamináció

BRCA1 és BRCA2: 34 és 47 amplikon

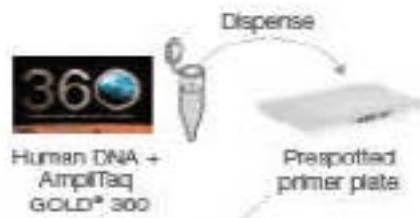


	1	2	3	4	5	6	7	8	9	10	11	12
A	Ex-1	Ex-10	Ex-11-8	Ex-15	Ex-23	Ex-1	Ex-10-1	Ex-11-5	Ex-11-13	Ex-14-2	Ex-27	MP-2
B	Ex-2	Ex-11-1	Ex-11-9	Ex-16	Ex-34	Ex-2	Ex-10-2	Ex-11-6	Ex-11-14	Ex-15	Ex-23	MP-3
C	Ex-3	Ex-11-2	Ex-11-10	Ex-17	MP-1	Ex-3	Ex-10-3	Ex-11-7	Ex-11-15	Ex-16	Ex-24	MP-4
D	Ex-5	Ex-11-3	Ex-11-11	Ex-18	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-7	MP-9

■ BRCA1 ■ BRCA2 ■ Multiplex non-templated control

BRCA1 és BRCA2 gén resequencing

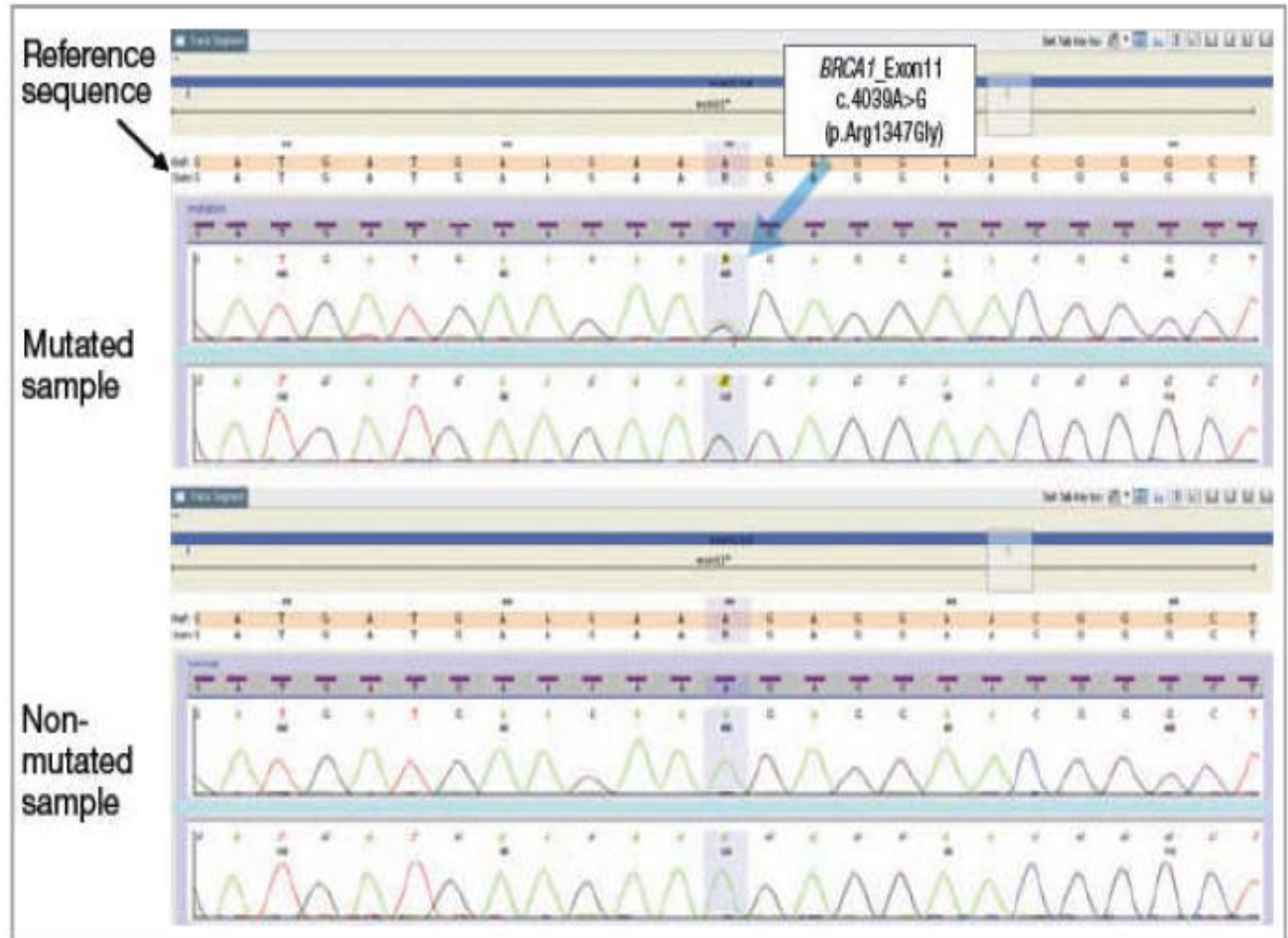
- mutációk diagnosztikai célú azonosítása



Add water to dilute and transfer to Forward and Reverse sequencing plates containing ExoSAP-IT®



Vortex for 15 minutes and spin down



Piroszekvenálás:

chemiluminescent detection of pyrophosphate

Enzimek:

Klenow fragment

ATP szulfuriláz

Luciferáz

Apiráz

Vegyületek:

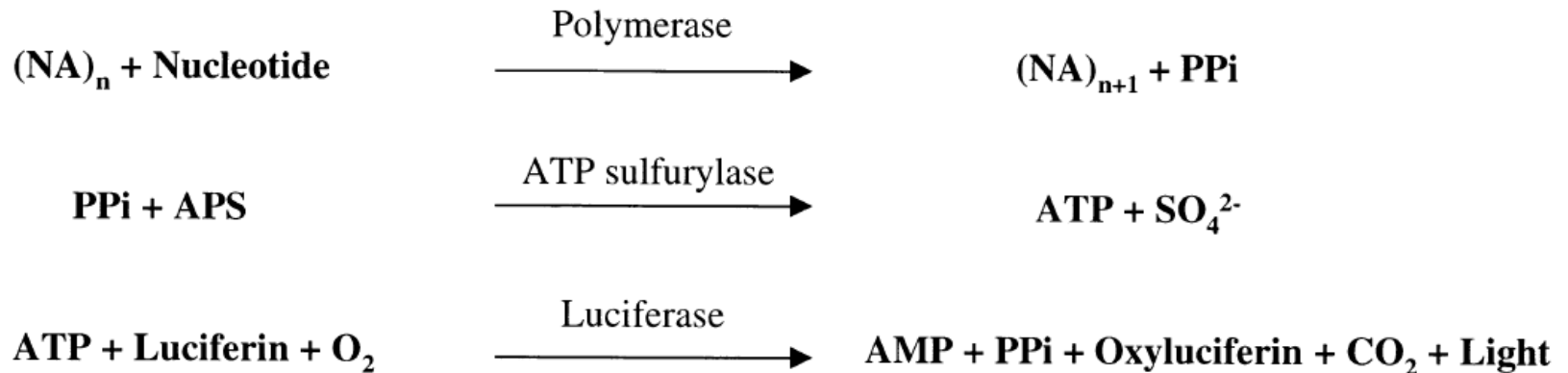
Adenozin-foszfoszulfát (APS)

D-luciferin

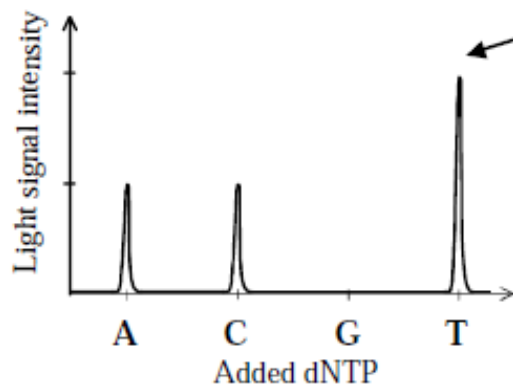
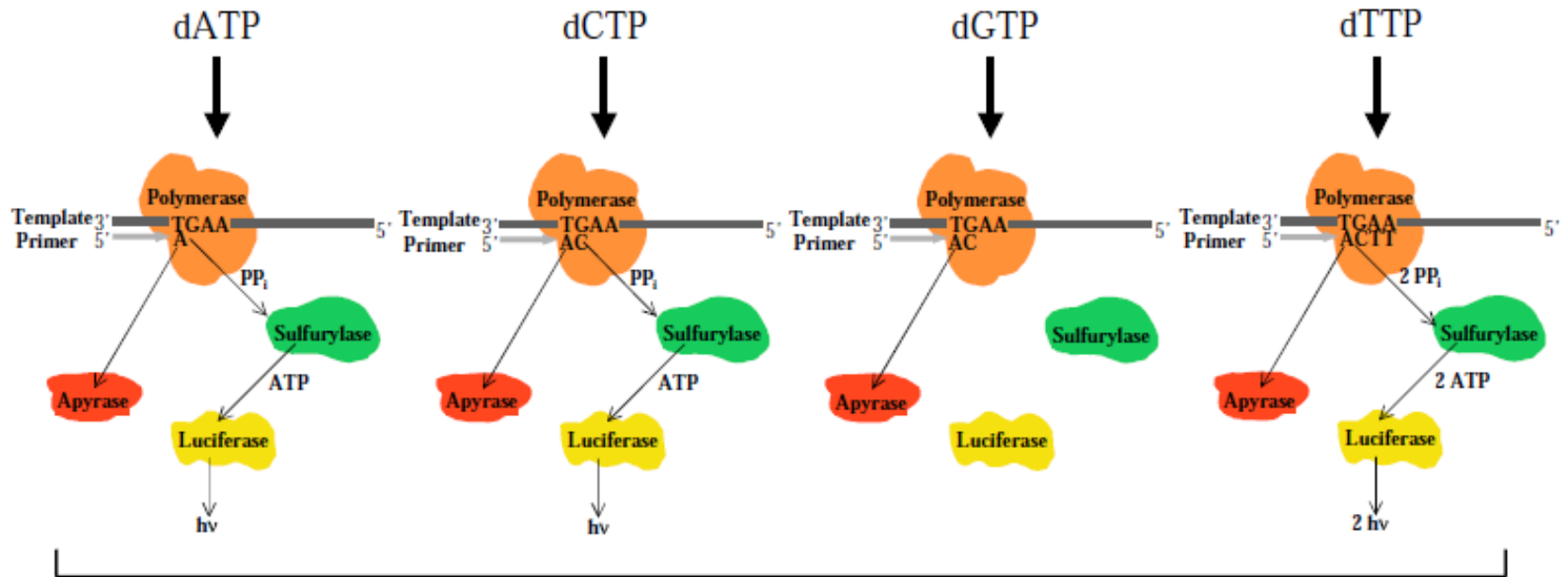
DNS Templát

Primer

dNTP-k egymás után adva



Piroszekvenálás



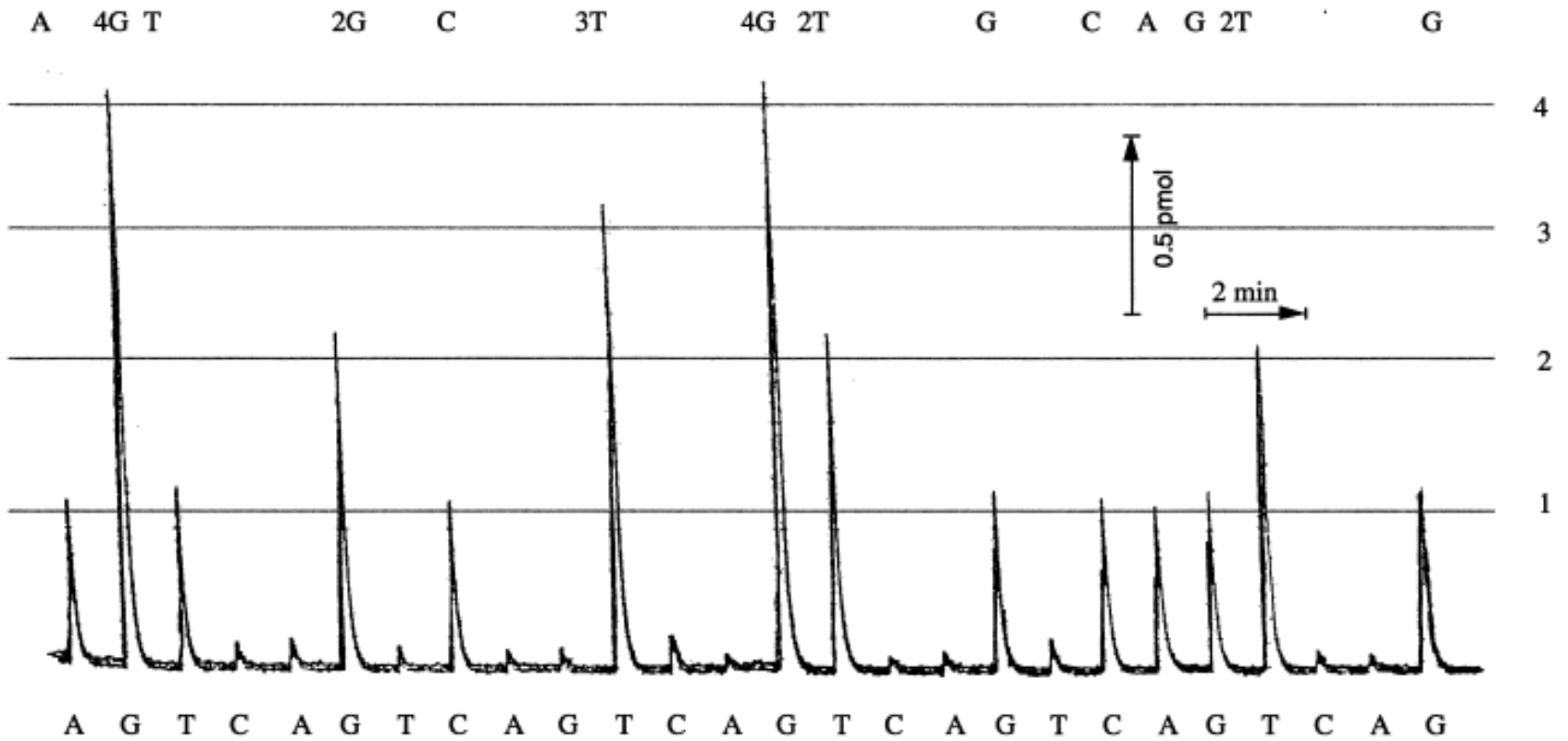
Pyrogram

A C T T

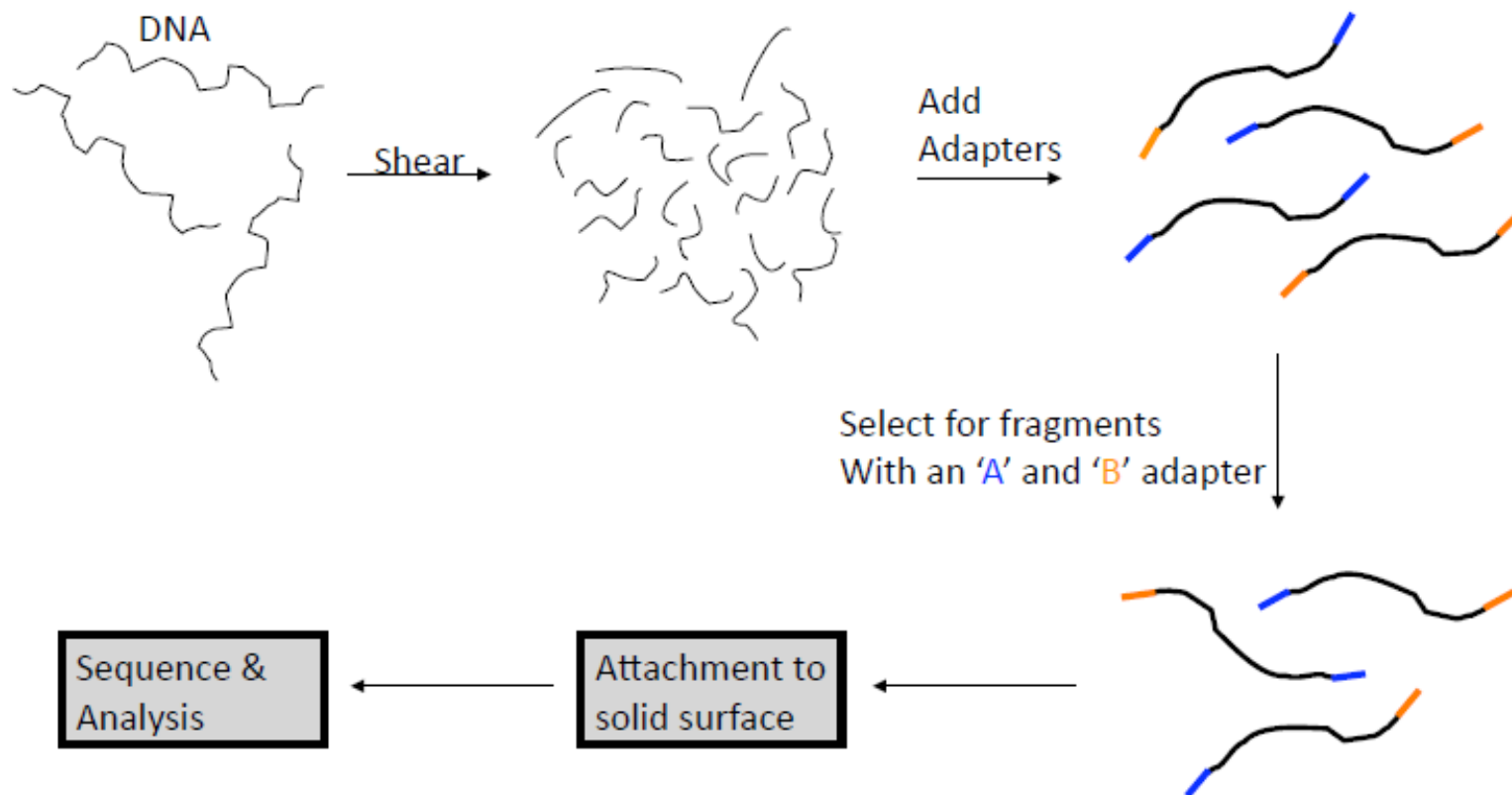
Sequence of synthesised DNA

T G A A

Sequence of template DNA



Minta előkészítés genomszekvenáláshoz



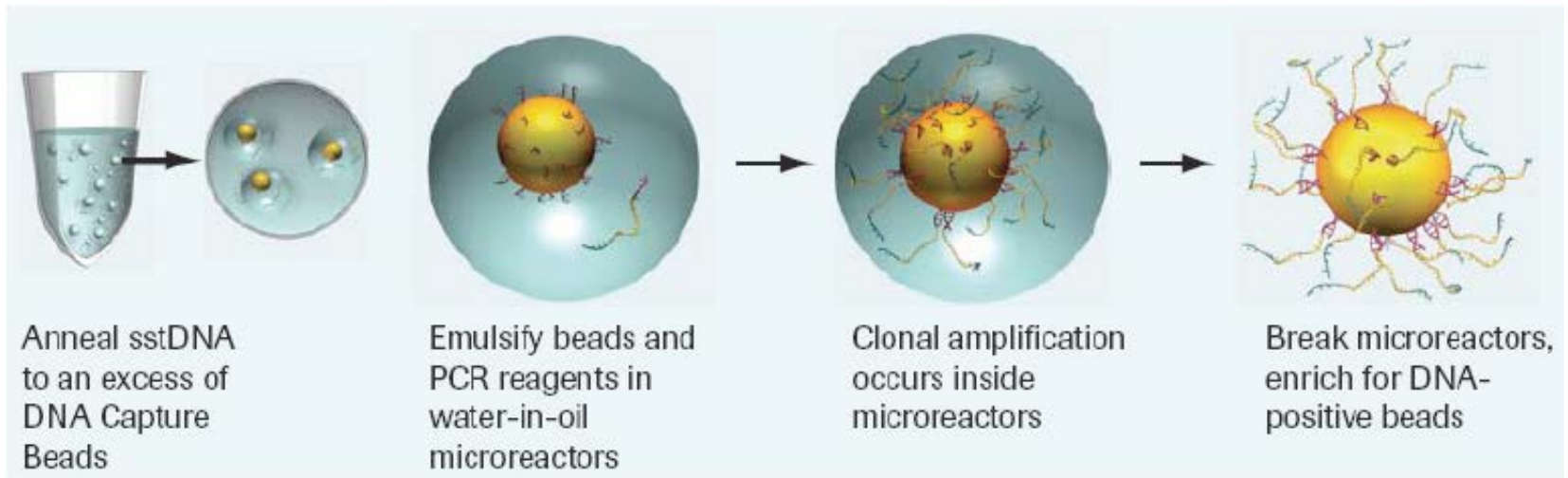
DNS fragmentumok törése (néhány 100 bp)
End-repair
Adapterek ligálás

Klonális amplifikáció

Emulsion PCR

Mikroreaktorok
Víz-olaj emulzióban

Több millió kópiája egy fragmentumnak



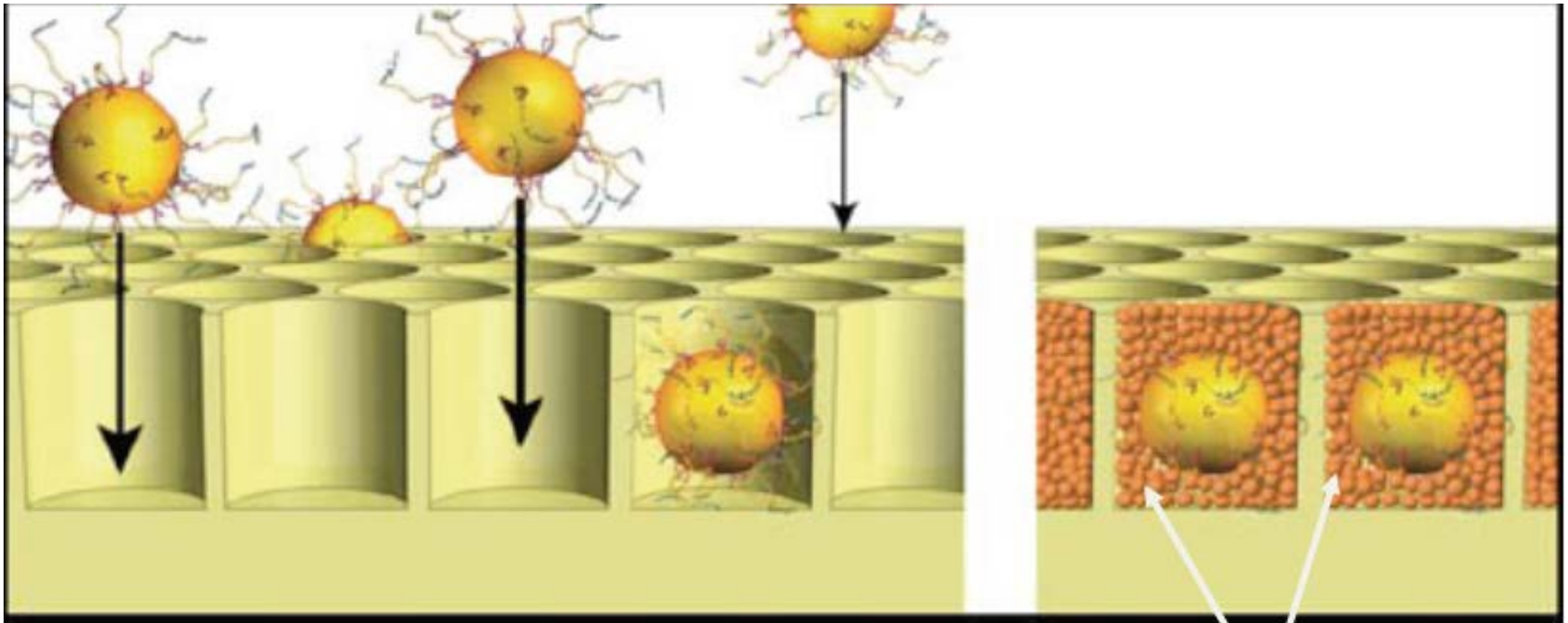
Each bubble in the emulsion will potentially contain a different fragment.

Roche/454 sequencing technology

Picotiter well plate betöltés

$3,4 \cdot 10^6$ lyuk

Pikoliteres szekvenáló reakciók

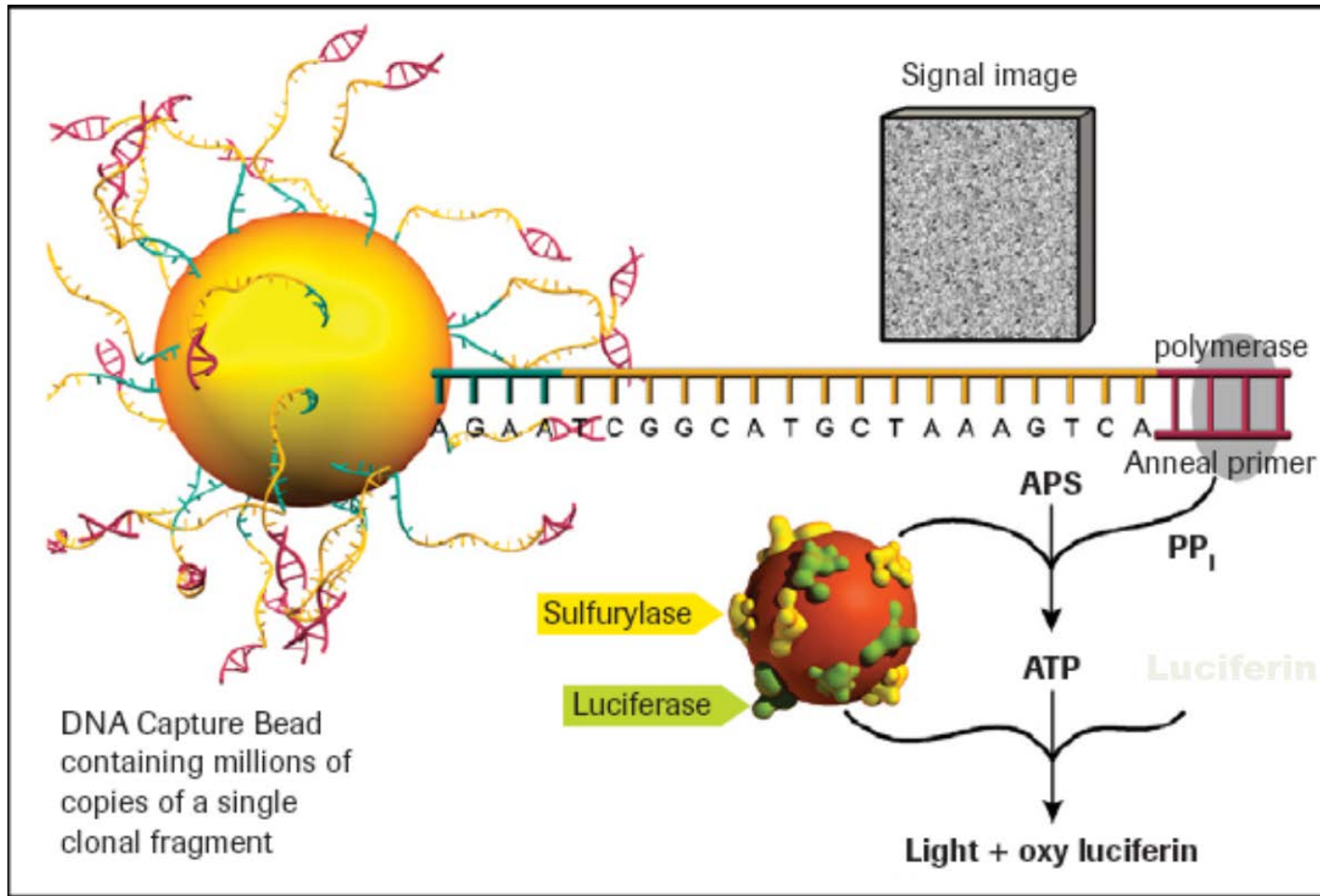


Instead of 96 reads/run, there are hundreds of thousands.

Packing beads and enzyme beads

Roche/454 sequencing technology

Szekvencia meghatározás piroszekvenálással



Roche/454 sequencing technology

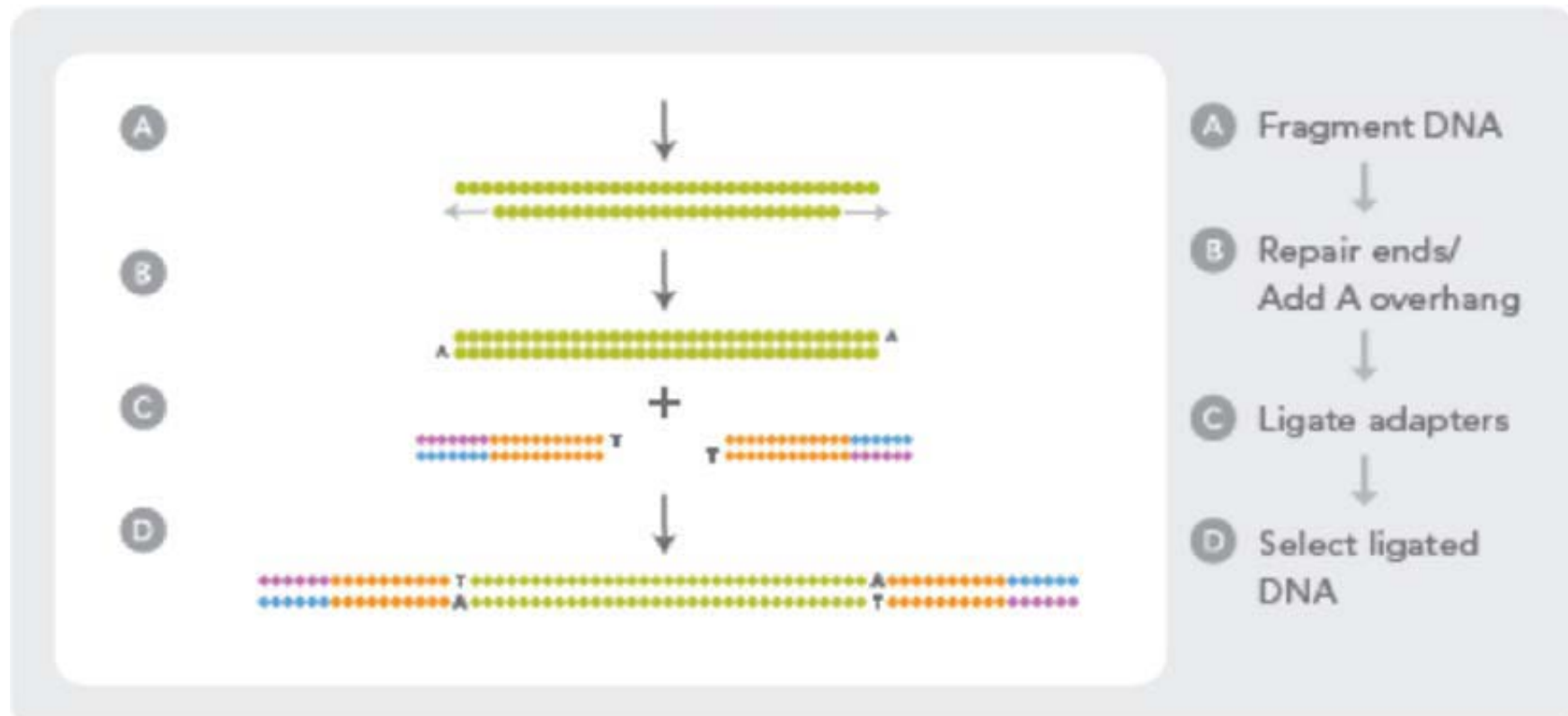


2005. 454 Life Sciences cég fejlesztette ki:
Mycoplasma genitalia - 580 069 bp

2007. Roche Applied Science megvette a céget:
GS FLX series, Junior,...
 $3,4 \times 10^6$ well picotiter plate

Illumina/Solexa sequencing

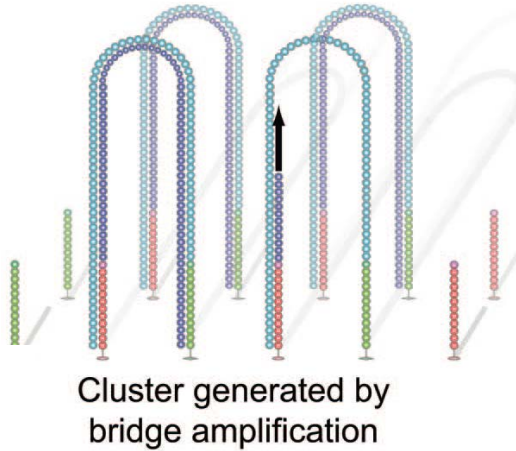
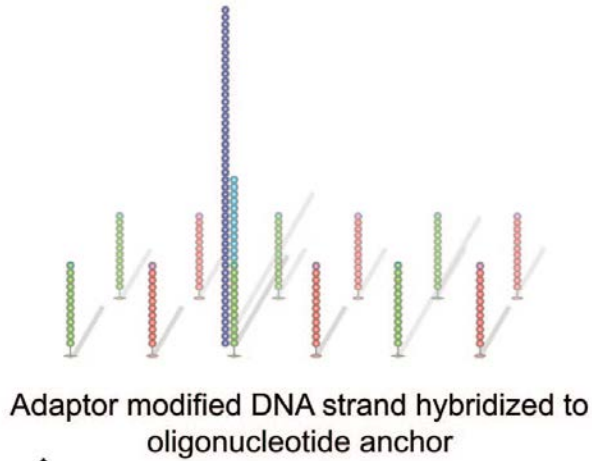
Minta előkészítés



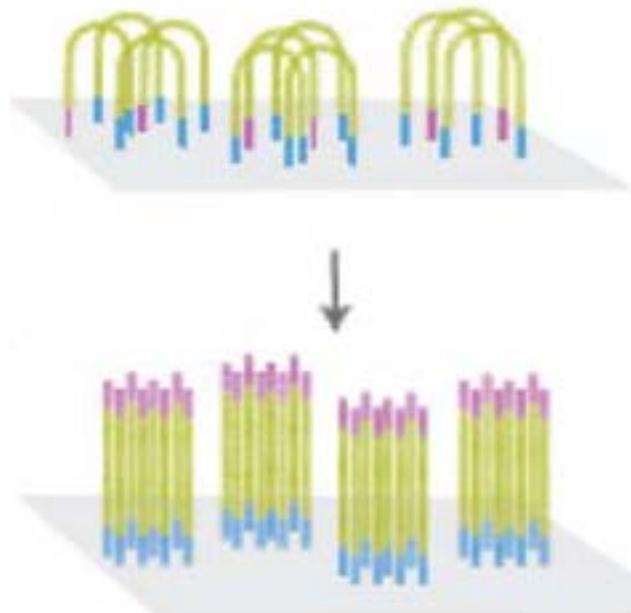
- A) DNS fragmentálása (néhány 100 bp)
- B) Végék javítása, A túlnyúló vég kialakítása
- C) Adaptor ligálása (T túlnyúló vég)

Illumina/Solexa sequencing

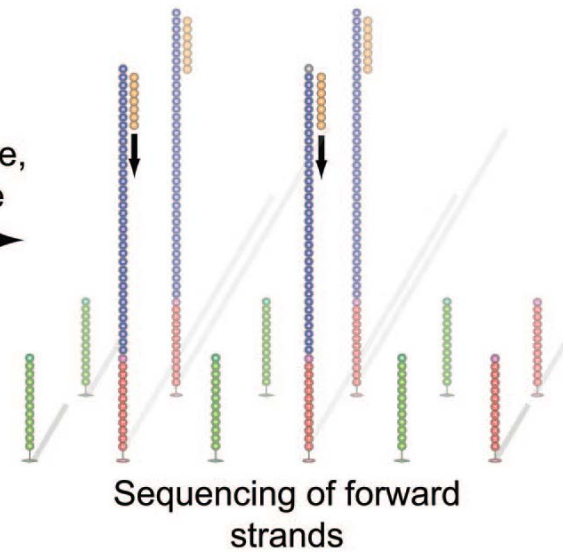
Klonális amplifikáció



PCR reakció kihorgonyozott primerekkel
Bridge amplifikáció

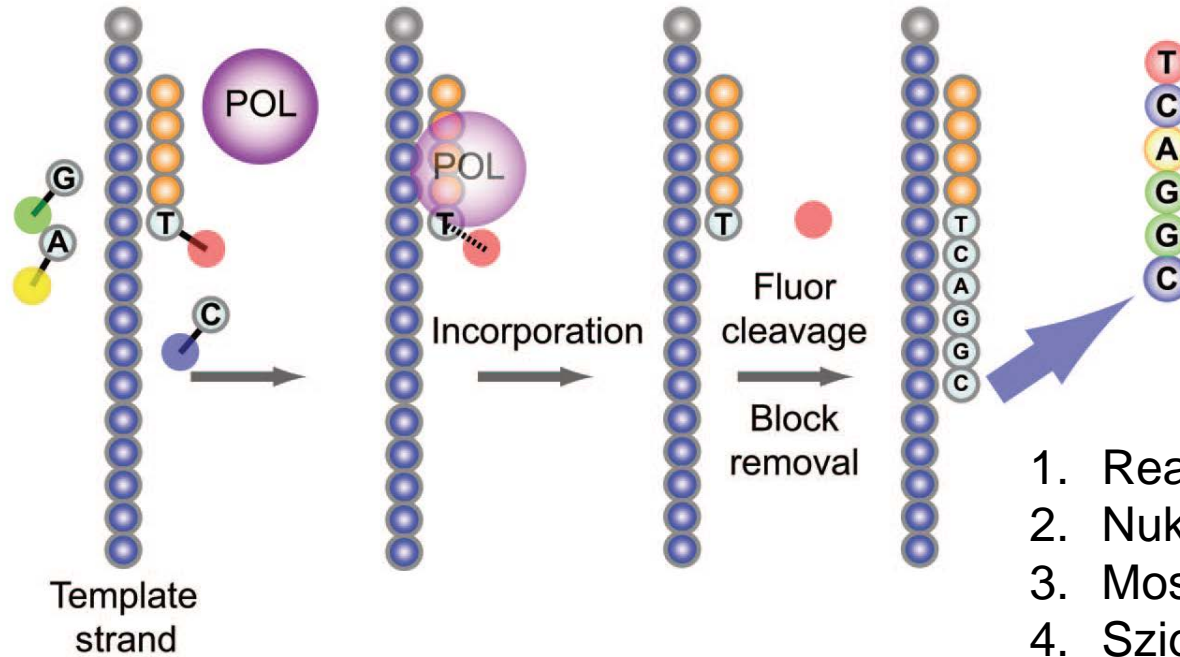


Denature,
cleave



Illumina/Solexa sequencing

Szekvencia meghatározás szintézis során



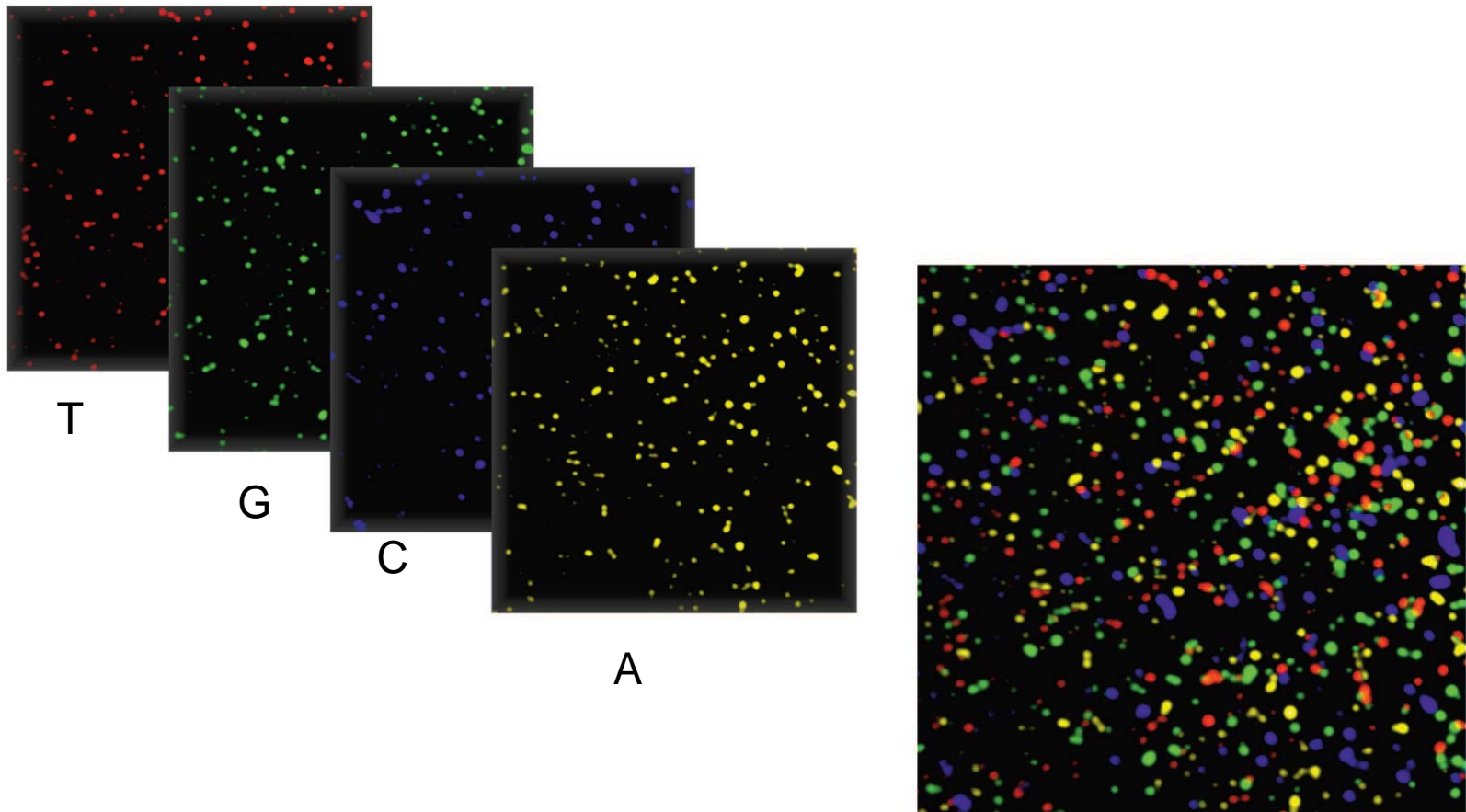
Sequencing by reversible dye terminators

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

Illumina/Solexa sequencing

A fluoreszcens szignál észlelése



SOLID: Sequencing by Oligo Ligation and Detection

- Kémiai hasítás, amplifikálás és ligálás

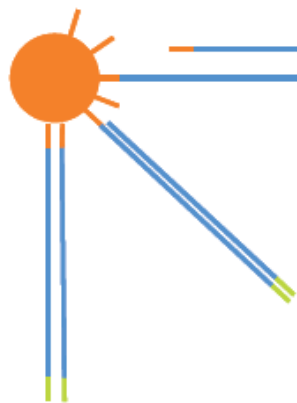
Genomi DNS

Randomly shear DNA

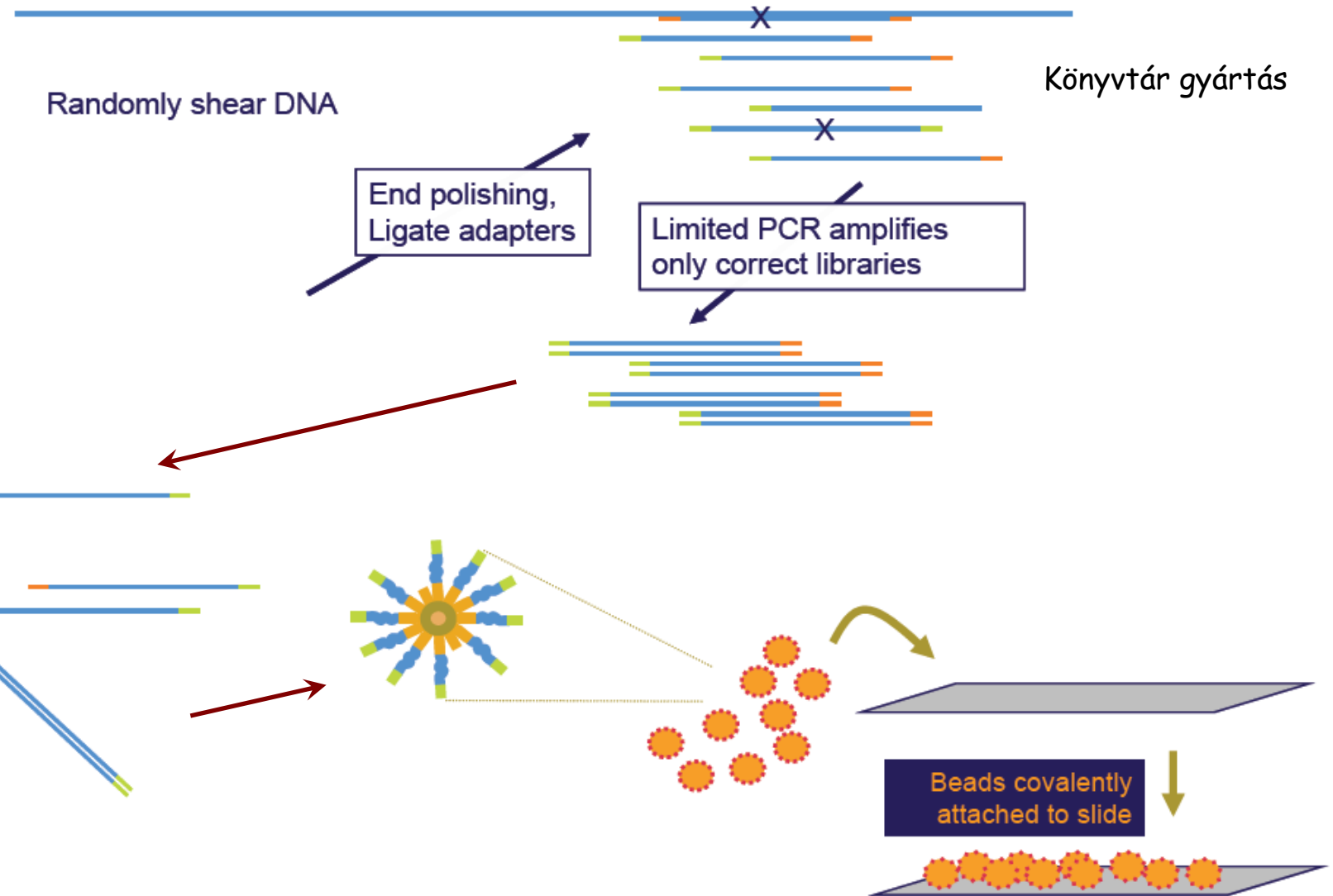
Könyvtár gyártás

End polishing,
Ligate adapters

Limited PCR amplifies
only correct libraries

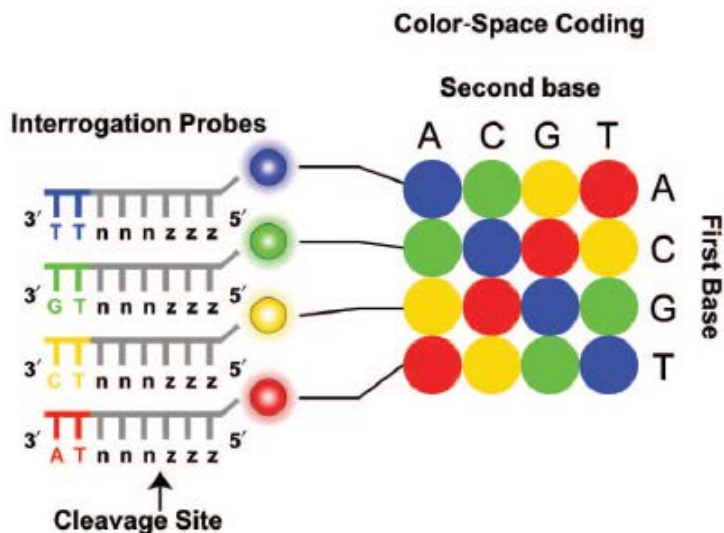


Komplementer adapterek



Applied Biosystems - SOLiD

Szekvencia meghatározás ligálással!



Próba

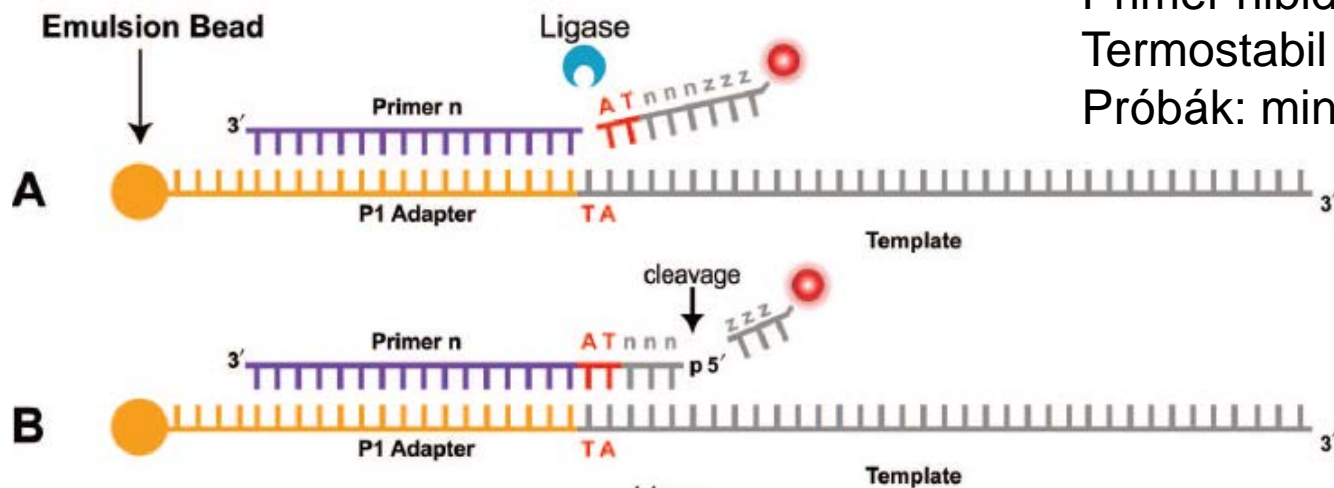
Octamer

2 próba specifikus bázis

3 degenerált bázis

3 univerzális

Fluoreszcens jelölő



Primer hibidizálása az adapterhez

Termosztabil ligáz

Próbák: mind a 16 féle kombináció

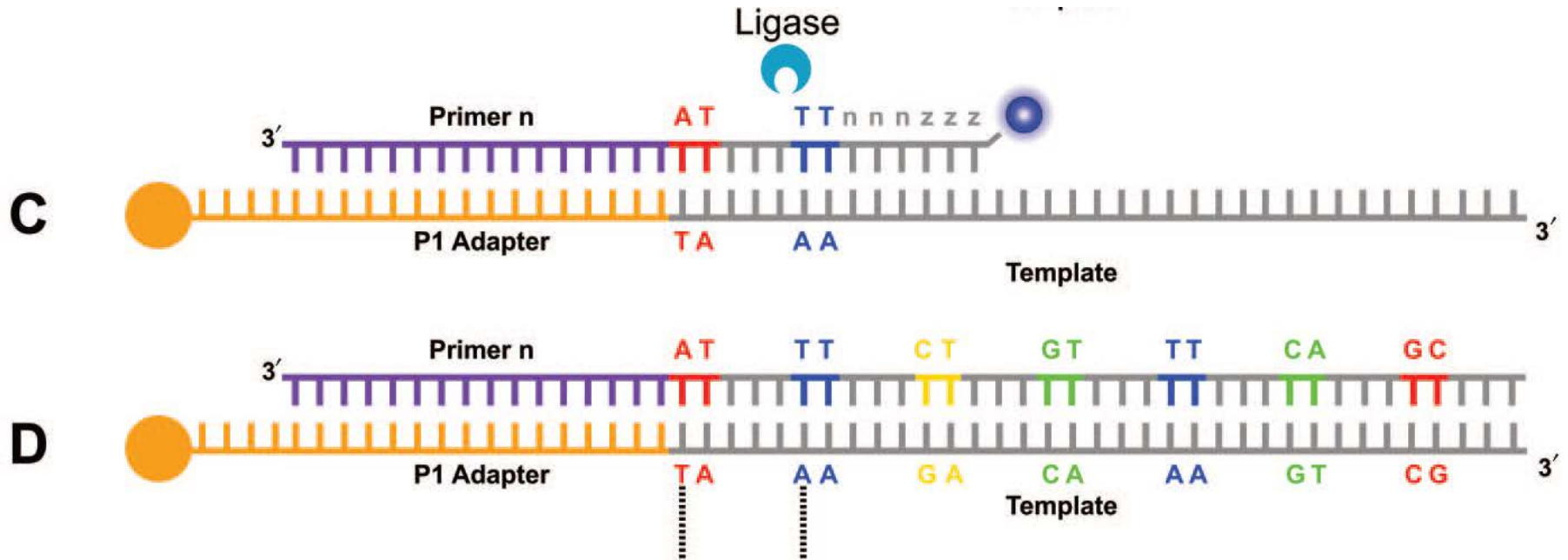
Ligálás

Mosás

Jel detektálása

Hasítás – 3 nukleotid

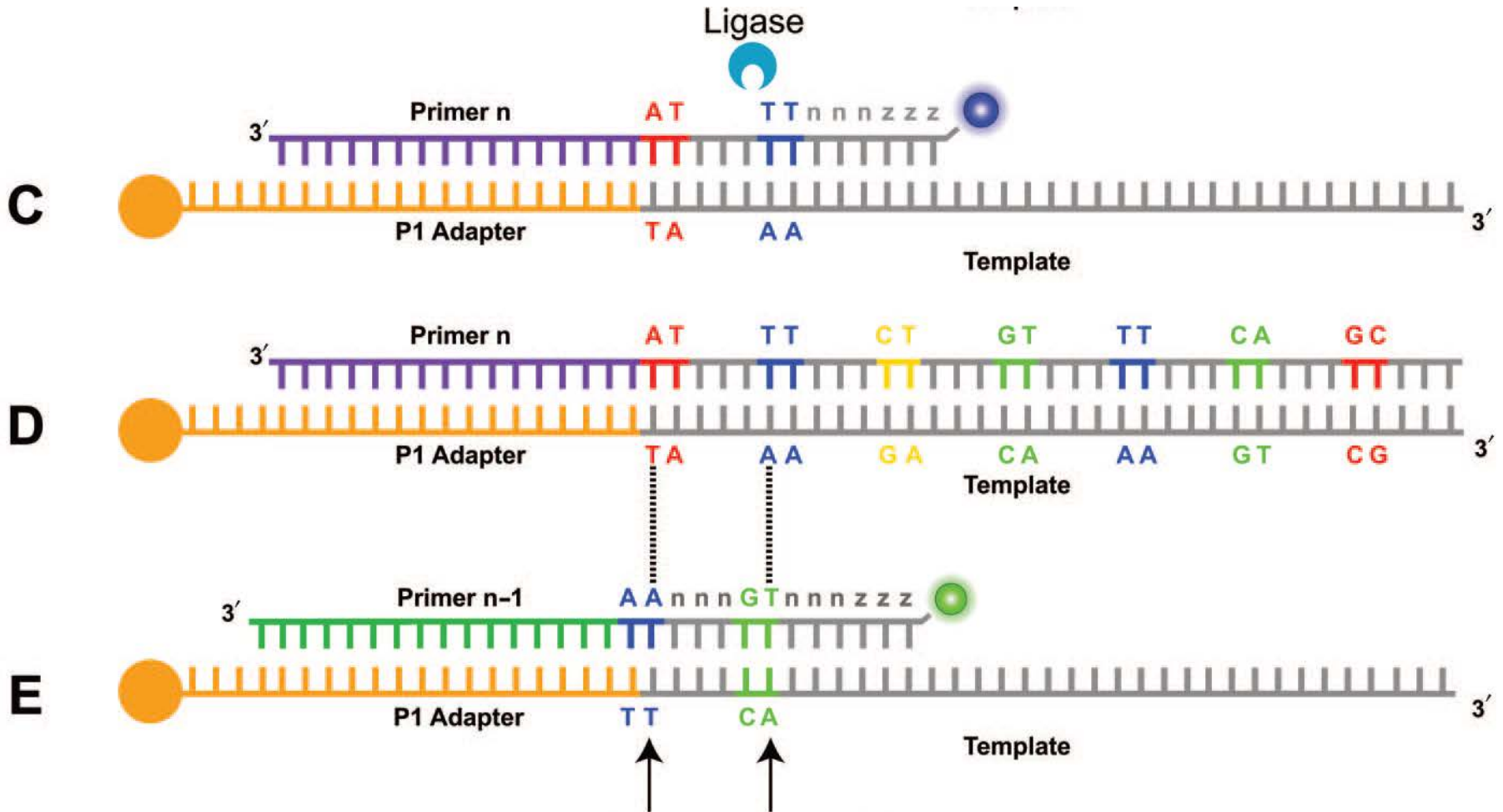
Applied Biosystems - SOLiD



Újabb próba ligálása

7-szer zajlik le ez a ciklus

Applied Biosystems - SOLiD



Denaturálás

Second interrogation of base

Új kör indítása n-1 adapter primerrel

Minden nukleotidra 2* kérdez rá

5 kör

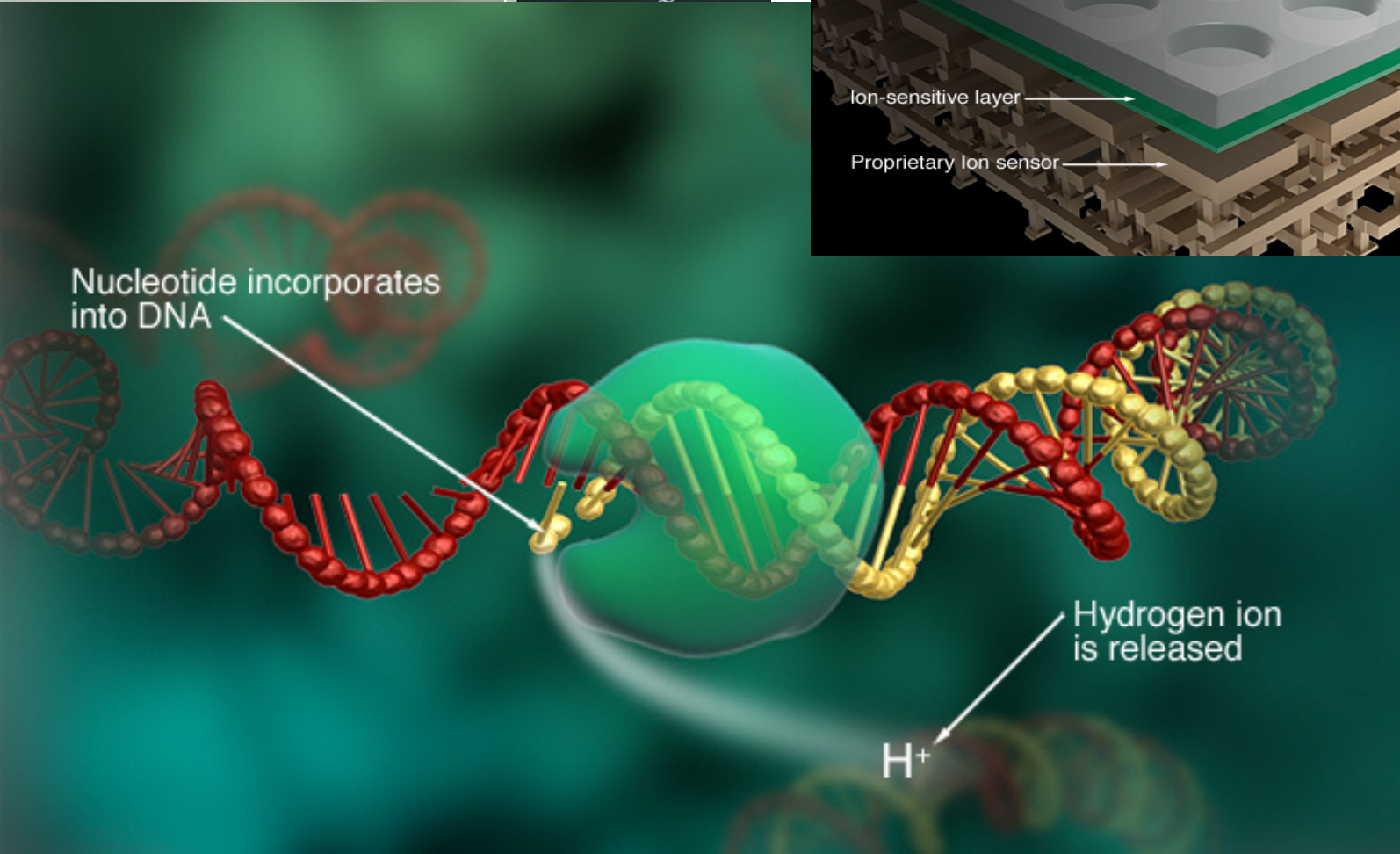
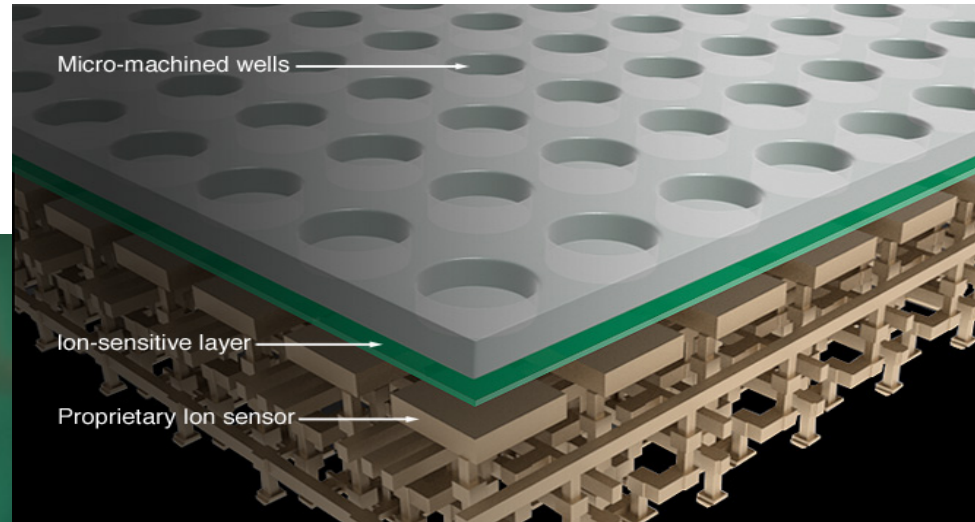
Table 1. Comparing metrics and performance of next-generation DNA sequencers

	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.

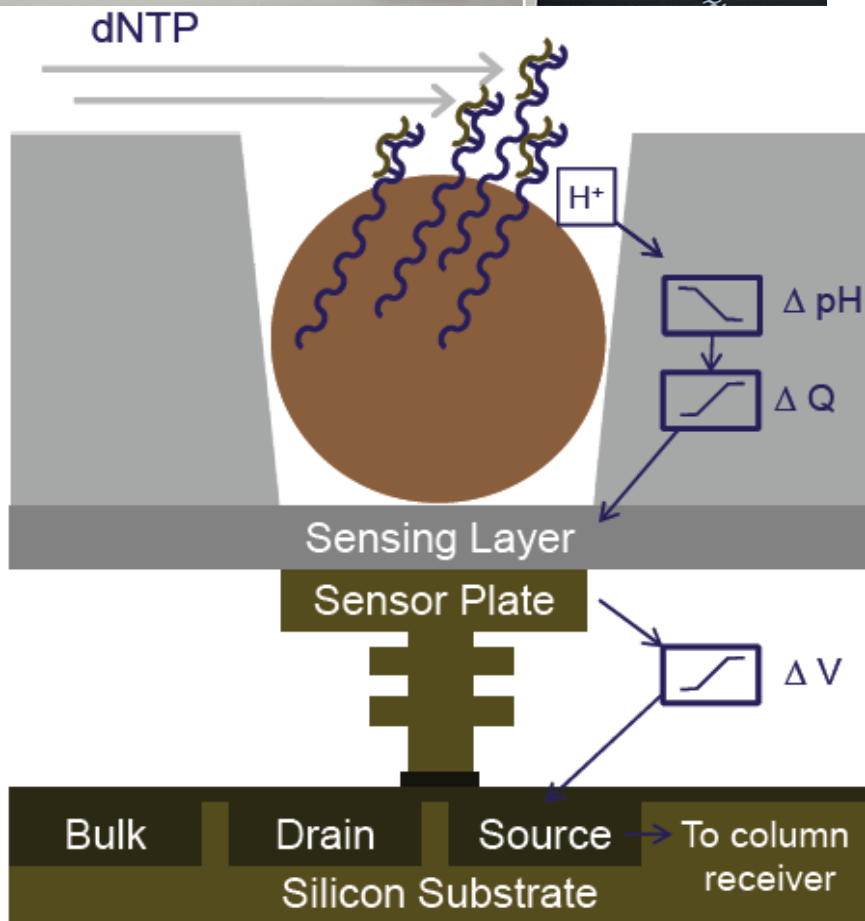
DNS szekvenálás félvezetőn: Ion Torrent

- hipergyors real-time szekvenálás



DNS szekvenálás félvezetőn: PGM

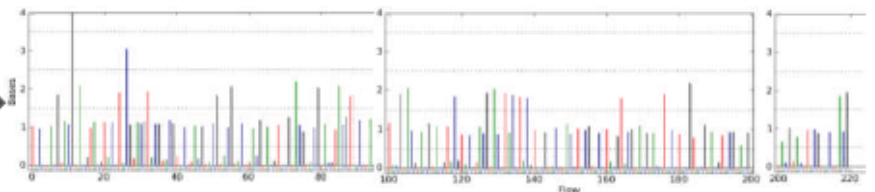
- hipergyors real-time szekvenálás



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

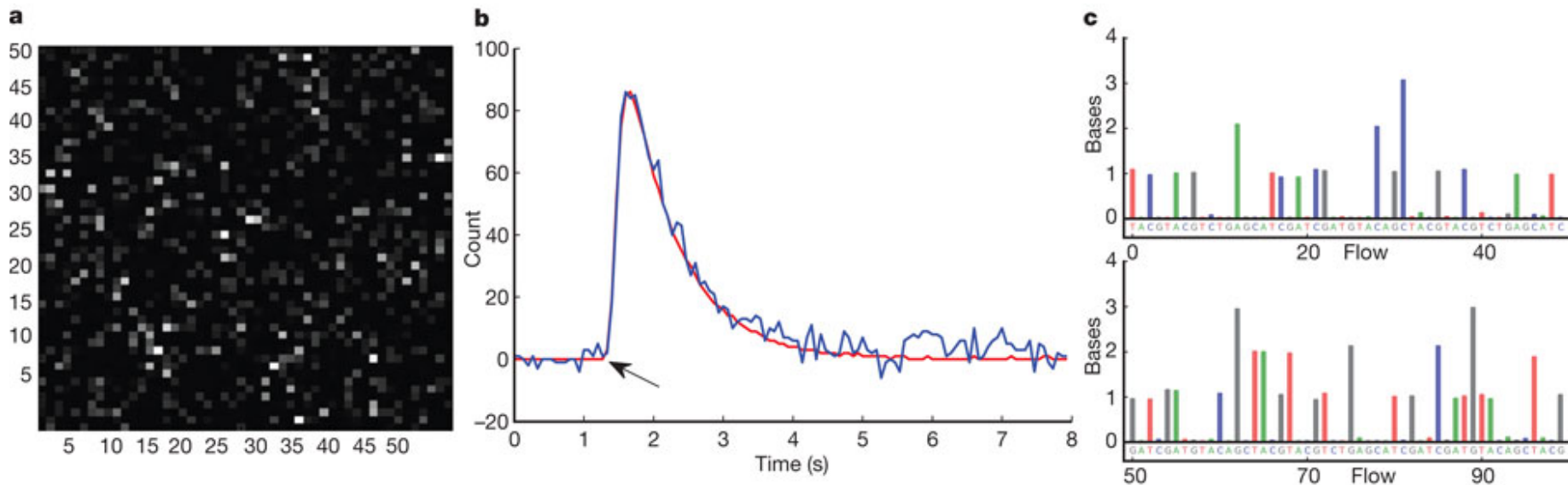


Table 1 | *Vibrio fischeri*, *E. coli*, *Rhodopseudomonas palustris* and *Homo sapiens*

	<i>V. fischeri</i>	<i>R. palustris</i>	<i>E. coli</i>	<i>E. coli</i>	<i>E. coli</i>	<i>H. sapiens</i>
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.