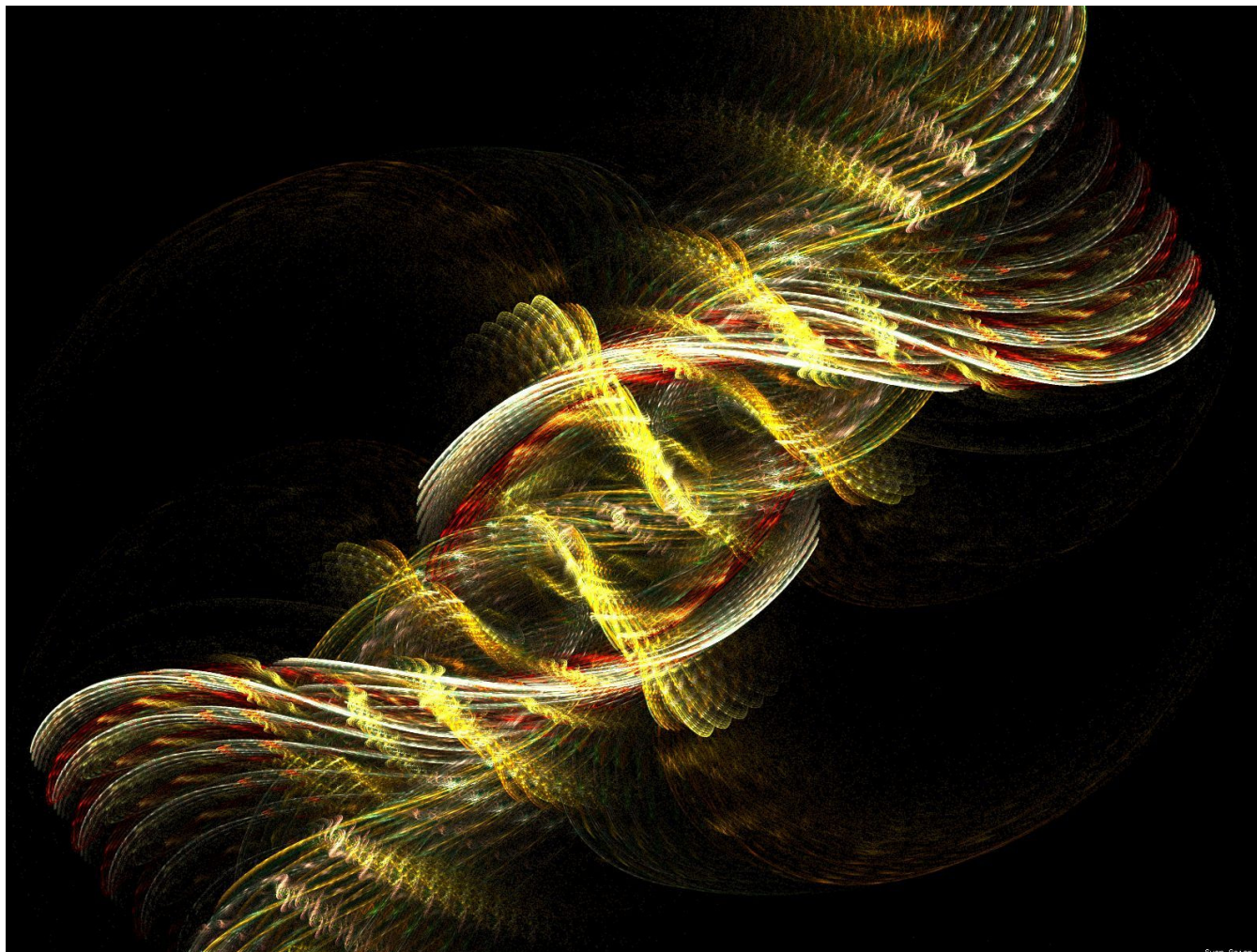


GENOMIKA

genomszekvenálási stratégiák



ELTE TTK Genetikai Tanszék

A genetikai információ tartalom megismerése

- **I. negyed:** sejten belüli funkciók, kromoszómák (Miescher, Flemming, Mendel, Sutton, Morgan, stb.)
- **II. negyed:** az öröklődés és a sejt működés molekuláris alapjai (DNS kettős hélix)
- **III. negyed:** az öröklődés és a sejt működés 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**

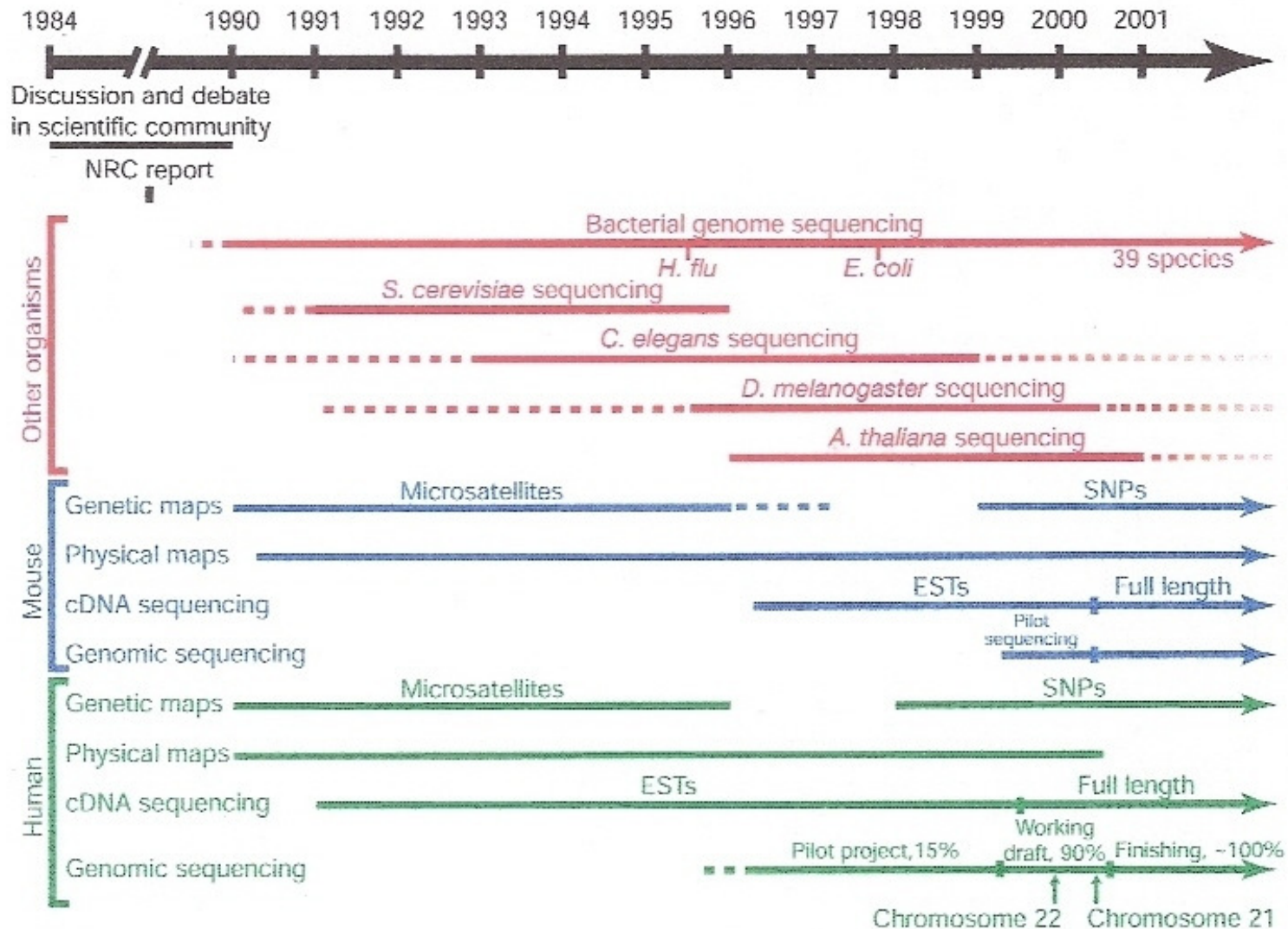
A humán genom szekvenálás első eredményei

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

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

Genom projektek időskálán



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

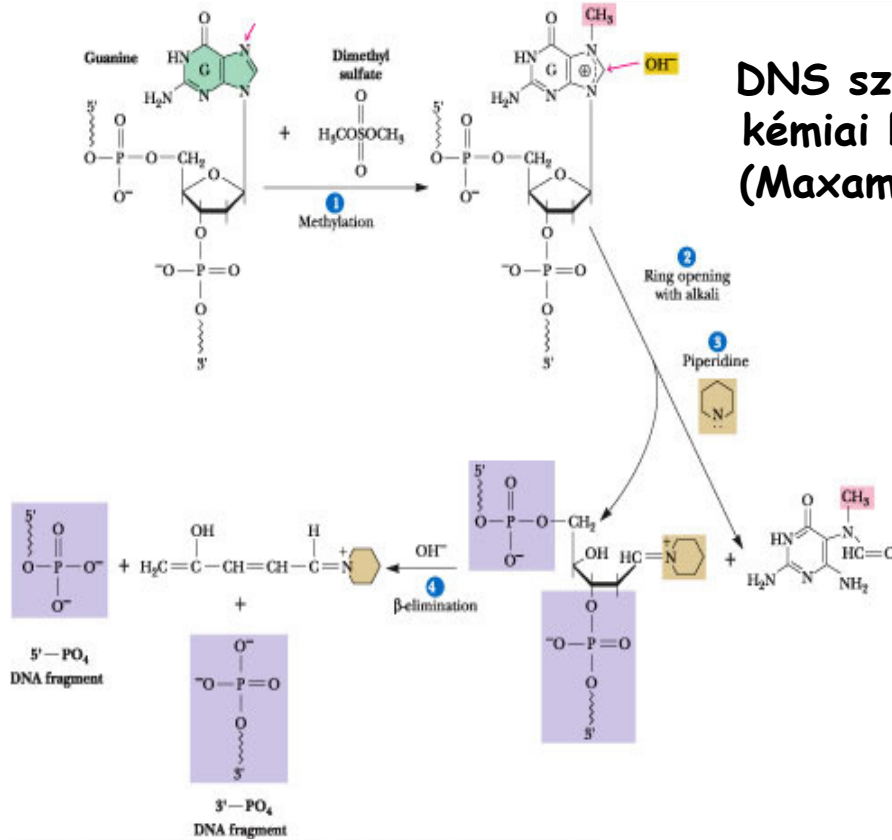


Technology speeds science. ABI sequencers at Venter Insitute, 2007.

Humán Genom Projekt



DNS szekvenálás kémiai bontással (Maxam-Gilbert)

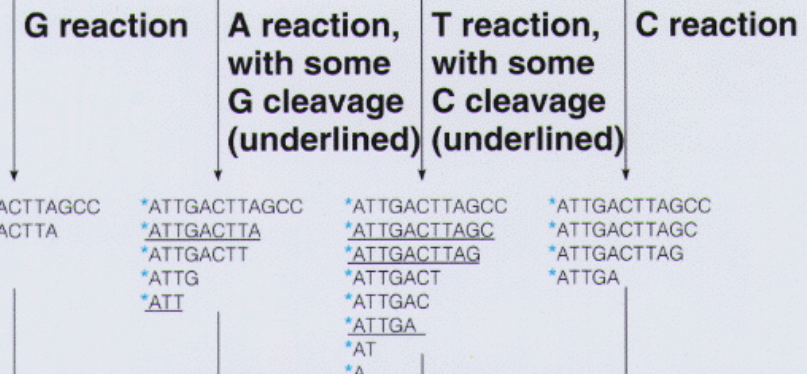


Sample DNA

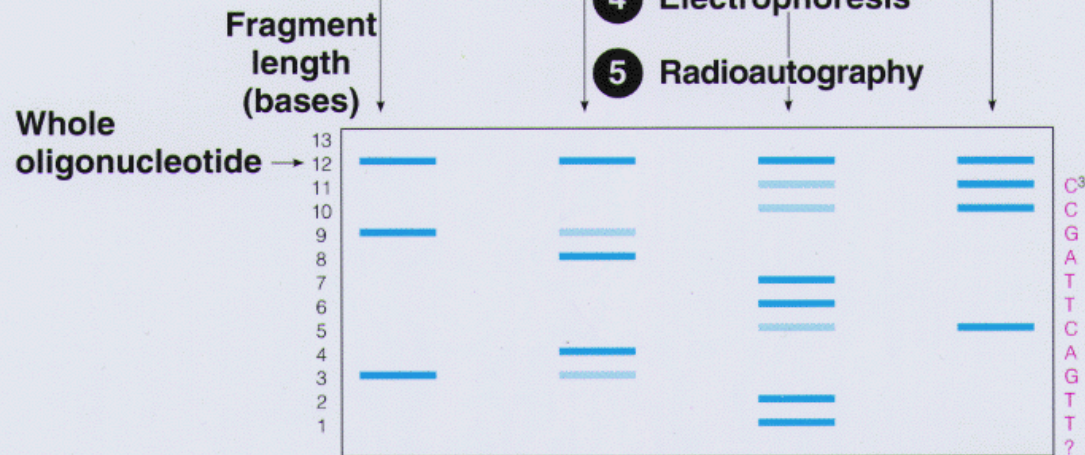
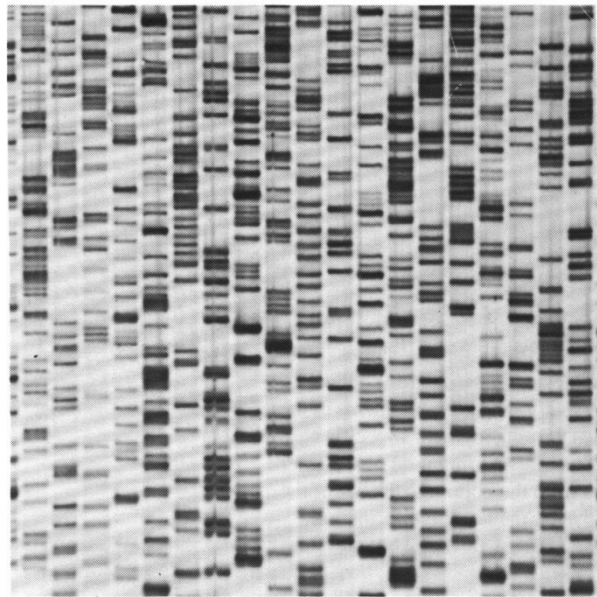
- 1** Preparation of homogeneous single-strand DNA
- 2** Addition of ³²P as 5' phosphate
- 3** Cleavage at specific nucleotides

5'ATTGACTTAGCC 3'

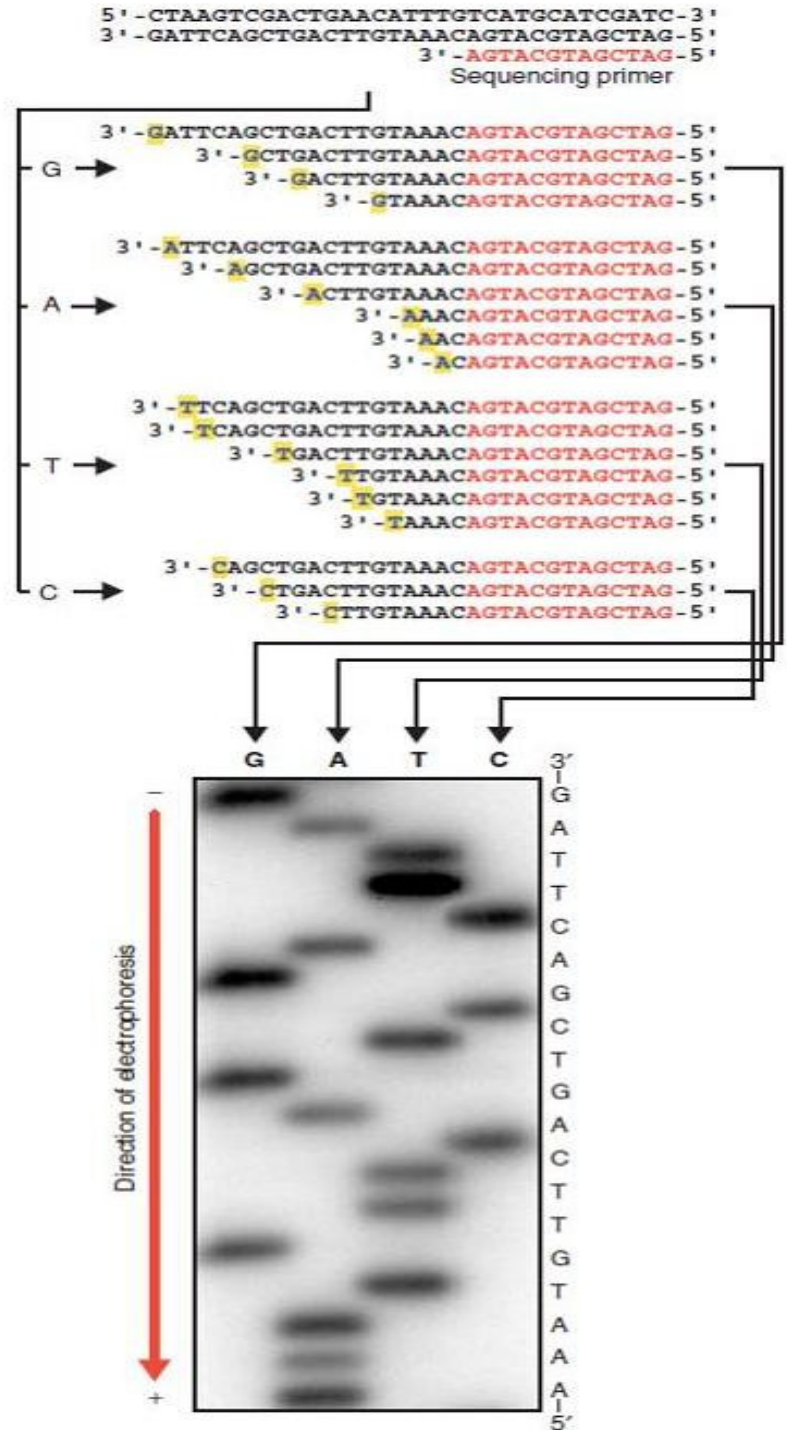
*ATTGACTTAGCC



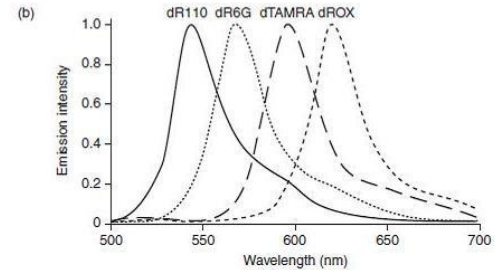
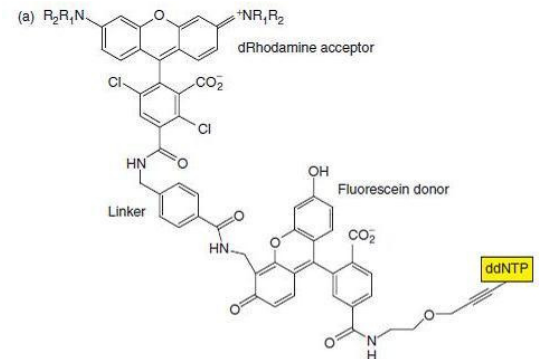
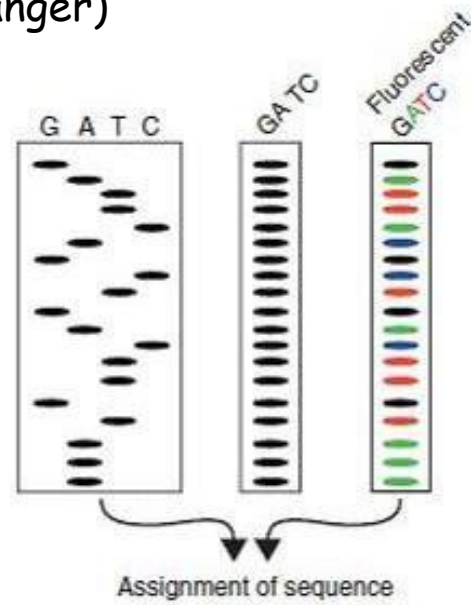
- 4** Electrophoresis
- 5** Radioautography



- 6** Read sequence



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 ●

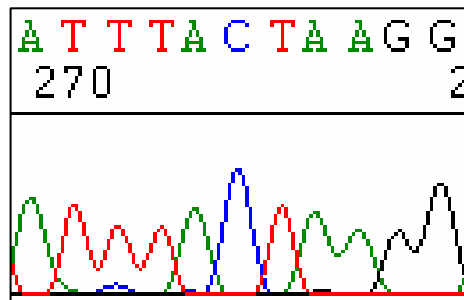
ATTTACTAA ●

ATTTACTAAG ●

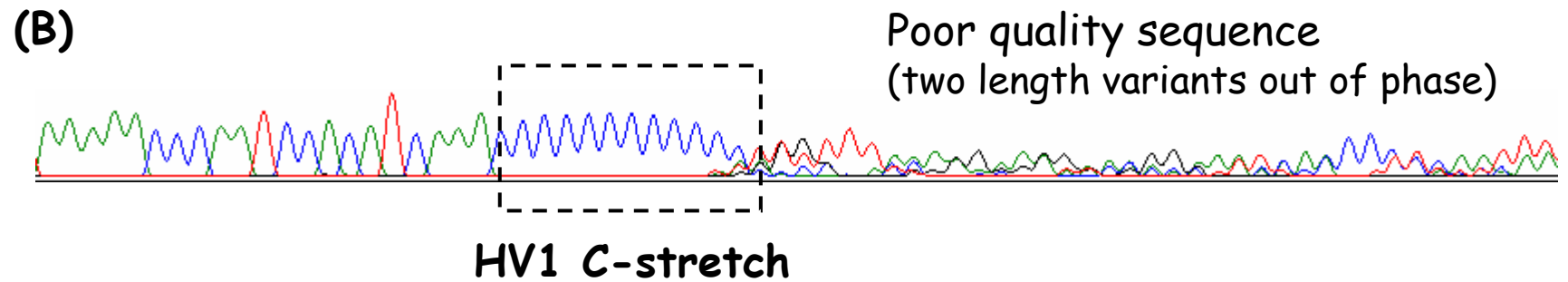
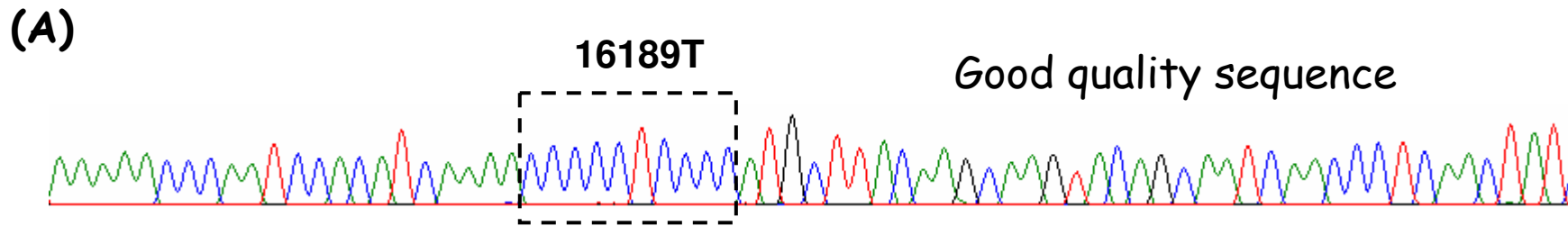
ATTTACTAAGG ●

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

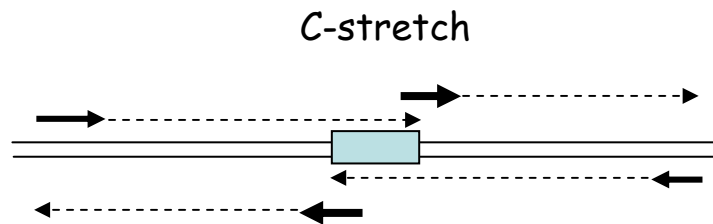
*Each ddNTP is labeled
with a different color
fluorescent dye*



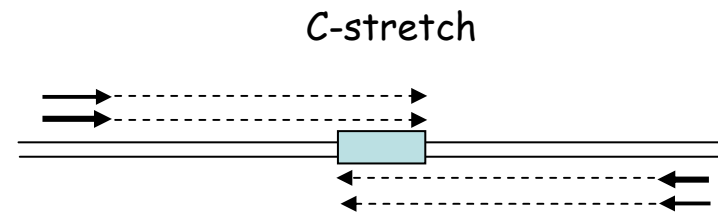
*Sequence is read by noting peak
color in electropherogram
(possessing single base resolution)*



(C) Primer strategies typically used with C-stretch containing samples



Use of internal primers



Double reactions from the same strand

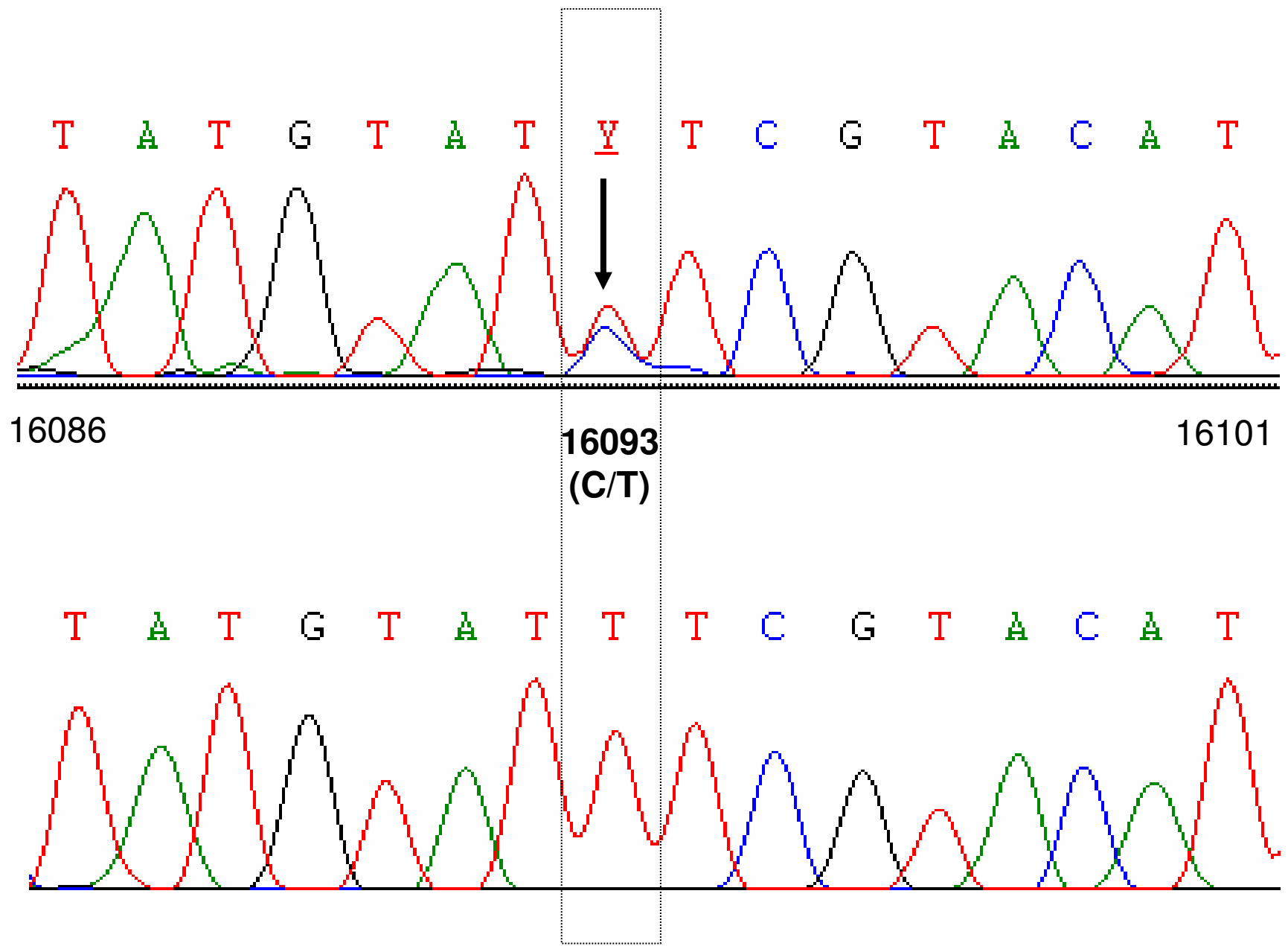


Figure 10.9, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

DNS szekvenálás: Technológia és Bioinformatika

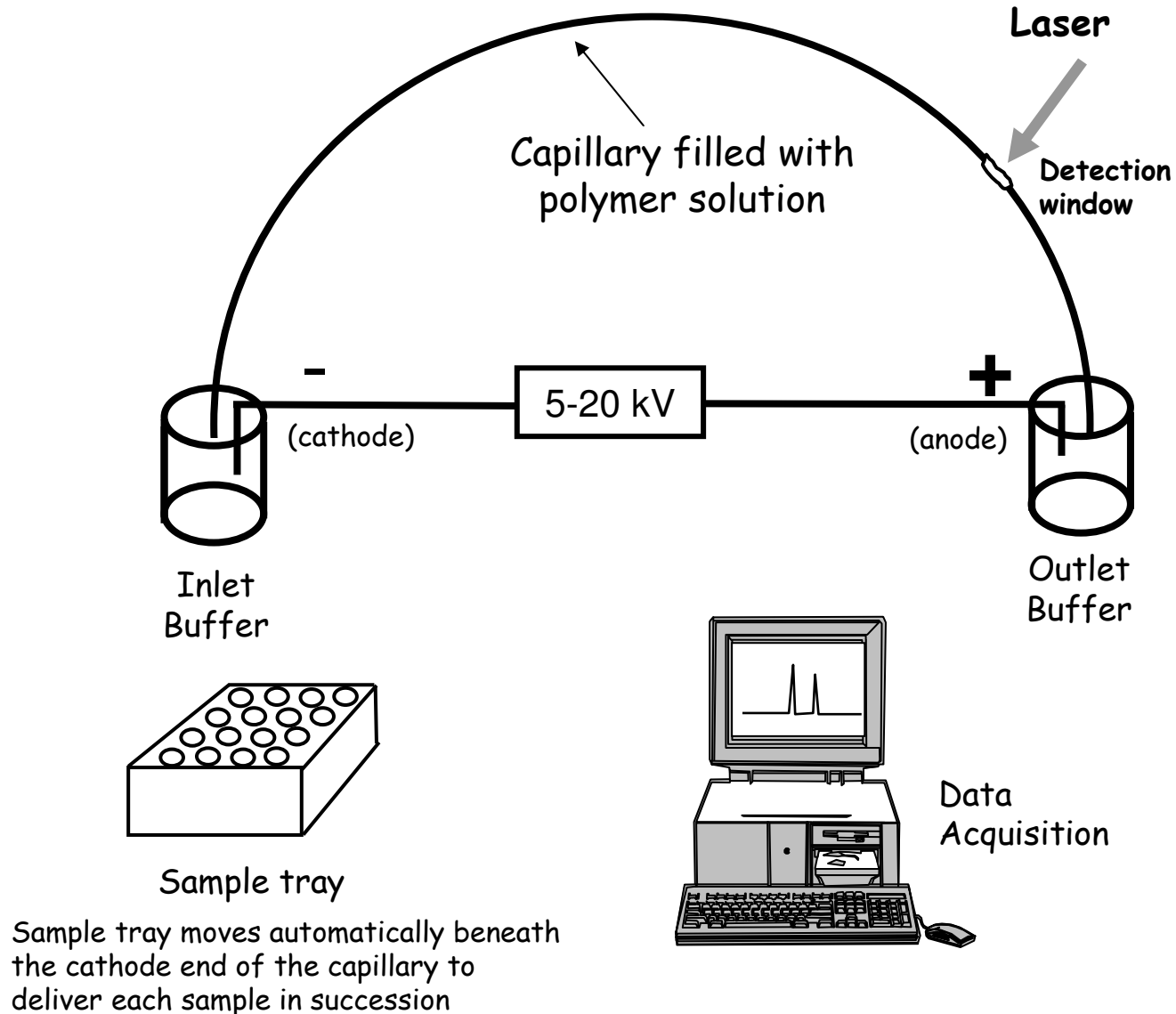
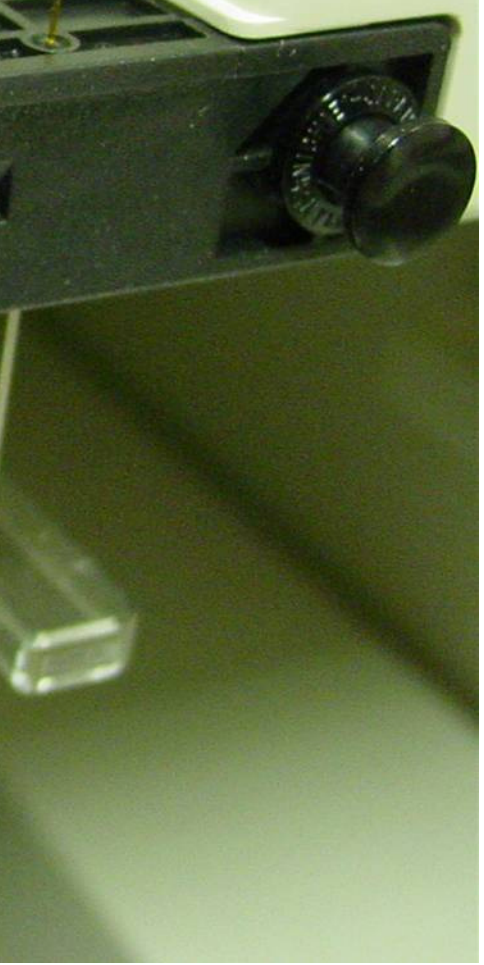
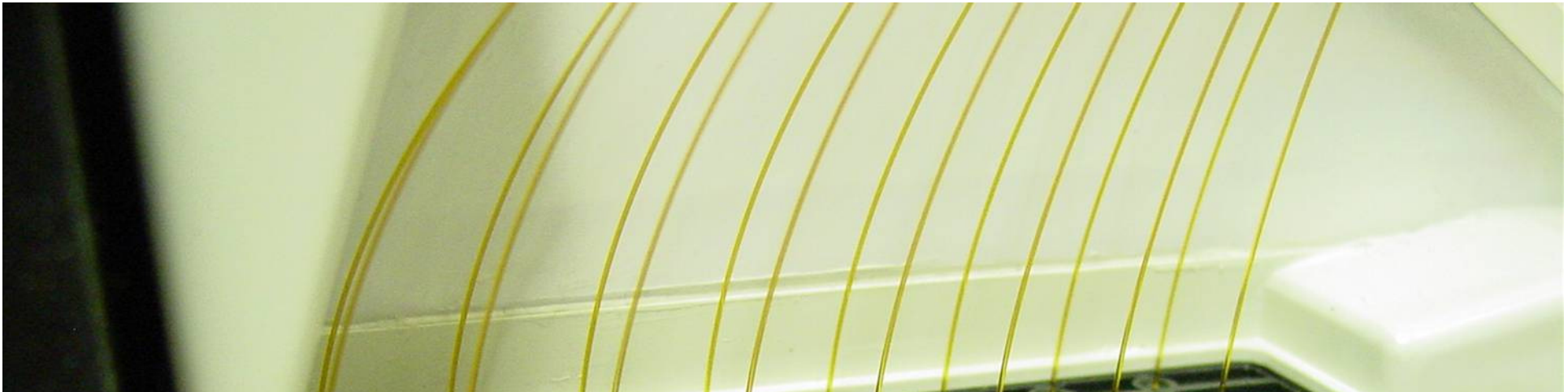


Figure 10.9, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press



„shotgun” genom szekvenálási stratégiák

1. „chromosome walking”

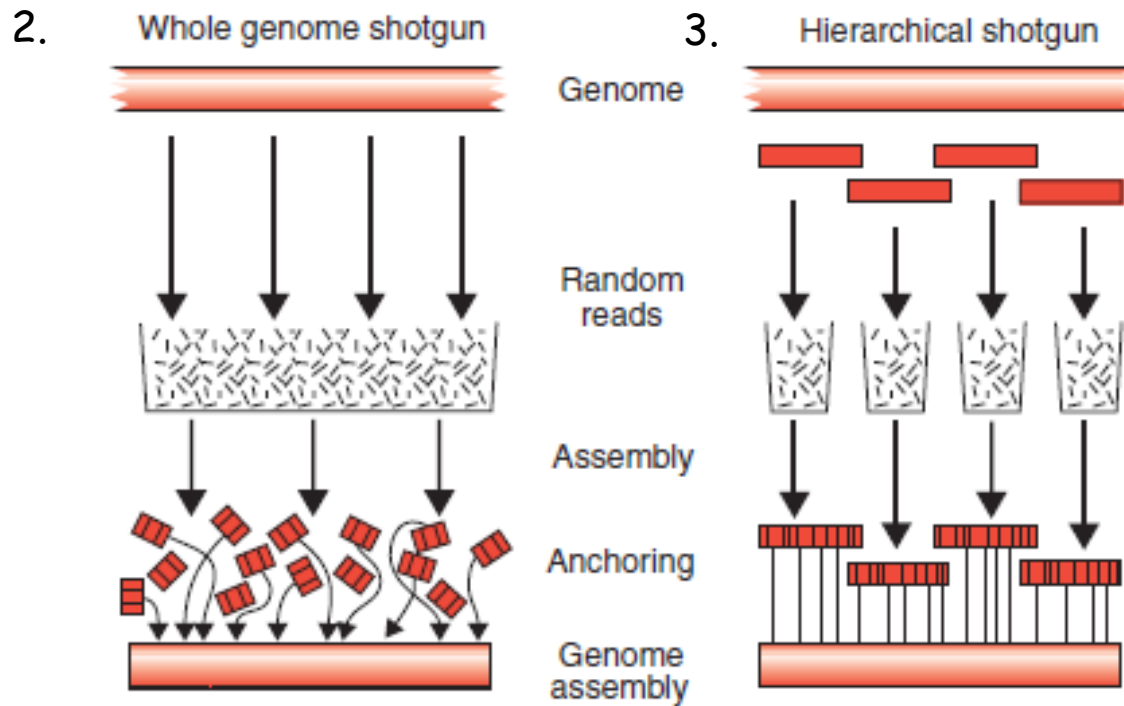


Figure 9.11. Assembling genomic data using the hierarchical and whole genome shotgun approaches. Adapted from Waterston, Lander and Sulston (2002), with permission

Hierarchical Shotgun Sequencing Method

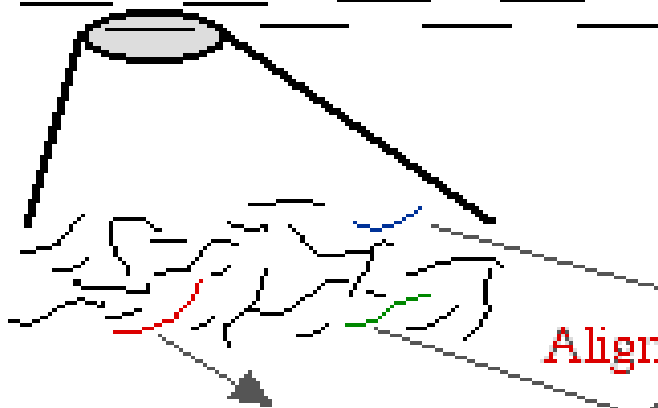


Genomic DNA



BAC Library

Create Contig Map



Sequence Each Contig
with Shotgun Approach

Align Contiguous Sequences

GCATTTGAGTTACCTGGACAACCAGTG

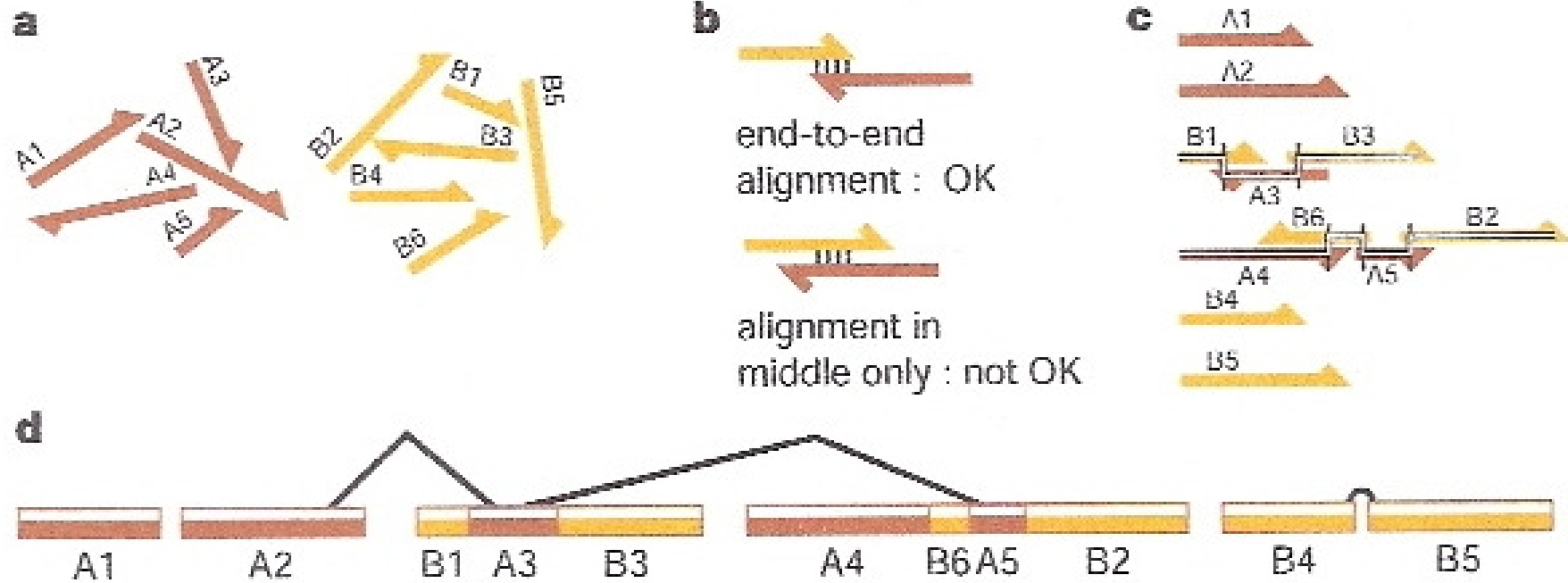
GCTTGATTGGCCAATAATAGTATAT

CCAGTGGTACTGAGGACGCCAAGAGGCTTGA

GCATTTGAGTTACCTGGACAACCAGTGGTACTGAGGACGCCAAGAGGCTTGATTGGCCAATAATAGTATAT

Generate Finished Sequence

Genom szekvenciaváz összeállítása

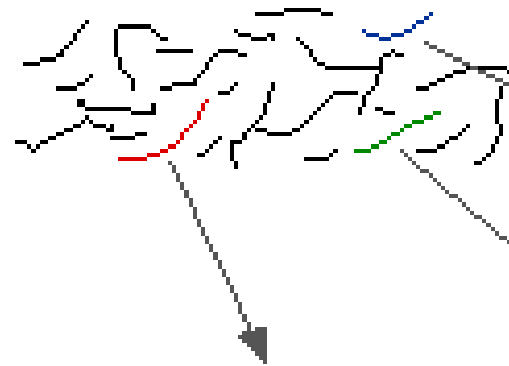


International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome, Nature 409, 860 (2001)

Whole Genome Shotgun Sequencing Method



Genomic DNA



Sequence Each Fragment
with Shotgun Approach

GCATTTGGAGTTACCTGGACAACCAAGTG

CCAGTGGTACTGAGGACGCCAAGAGGCTTGA

GCTTGATTGGCCAATAATAGTATAT

Align Contiguous Sequences

GCATTTGGAGTTACCTGGACAACCAAGTGGTACTGAGGACGCCAAGAGGCTTGGATTGGCCAATAATAGTATAT

Generate Finished Sequence

Teljes genom összeszerelés

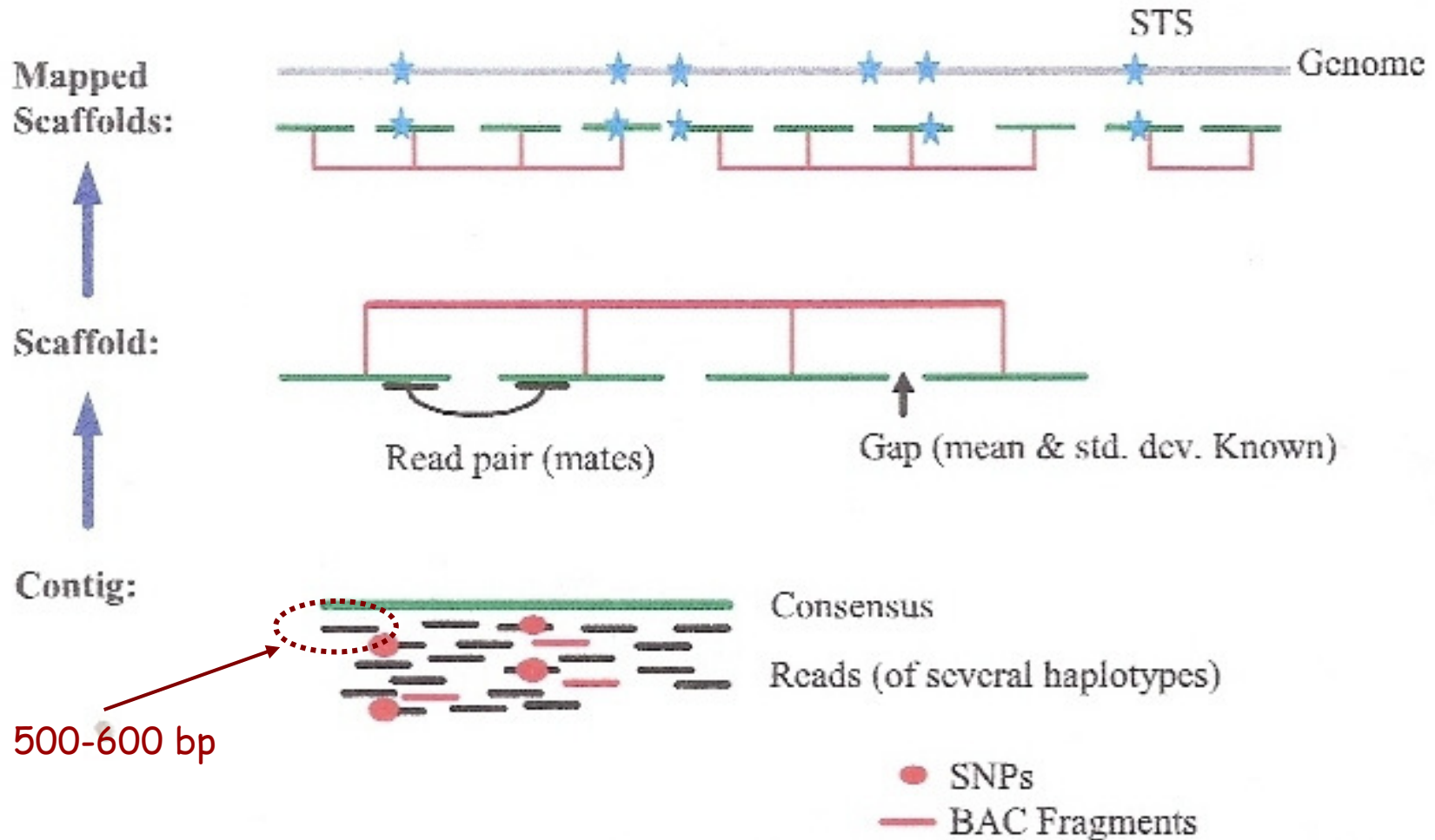


Fig. 3. Anatomy of whole-genome assembly. Overlapping shredded BAC fragments (red lines) and internally derived reads from five different individuals (black lines) are combined to produce a contig and a consensus sequence (green line). Contigs are connected into scaffolds (red) by using mate pair information.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

STS genom térképezés

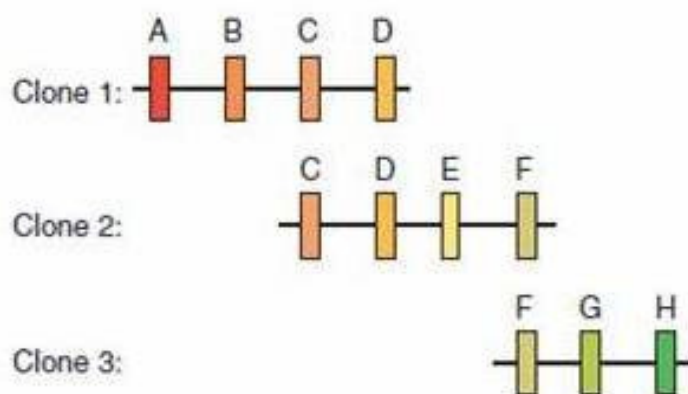


Figure 9.5. Aligning clones by STS mapping. Each clone contains several STSs. Clone 1 has four (A, B, C and D). Clone 2 also contains STSs C and D. Therefore clones 1 and 2 overlap with each other

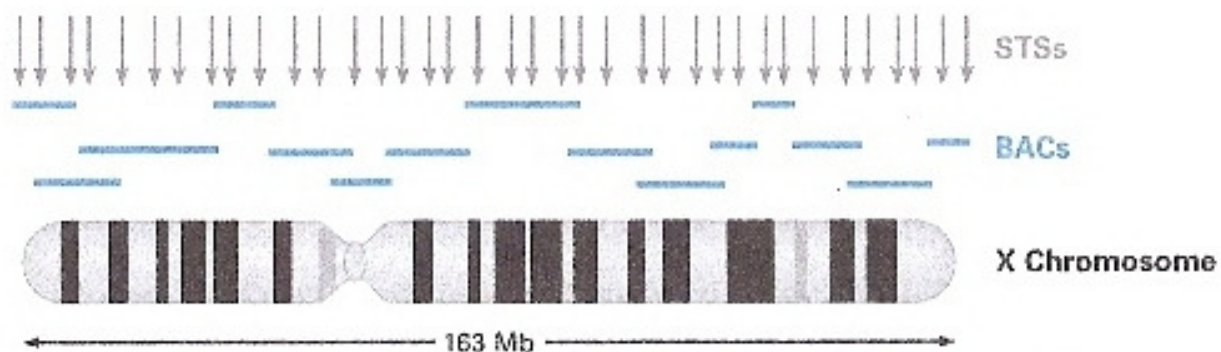


FIGURE 1.3 • Relationships of chromosomes to genome sequencing markers. The X chromosome is about 163 Mb in length. In this diagram, there are 16 overlapping BAC clones that span the entire length. In reality, 1,408 BACs were needed to span the X chromosome. Arrows (top) mark STSs scattered throughout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

Kromoszóma térképezés

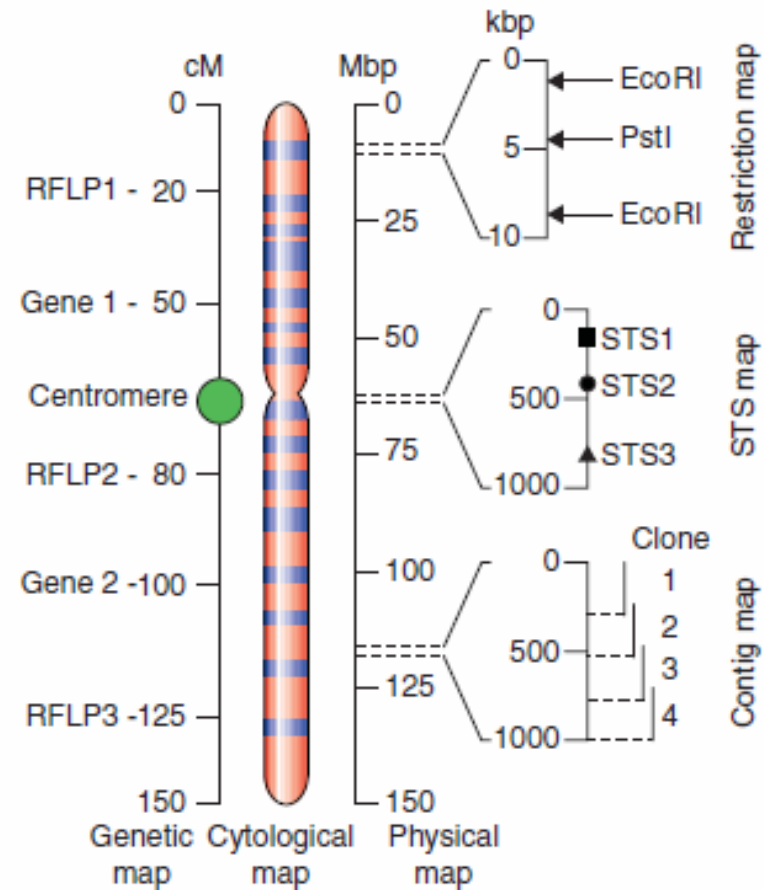
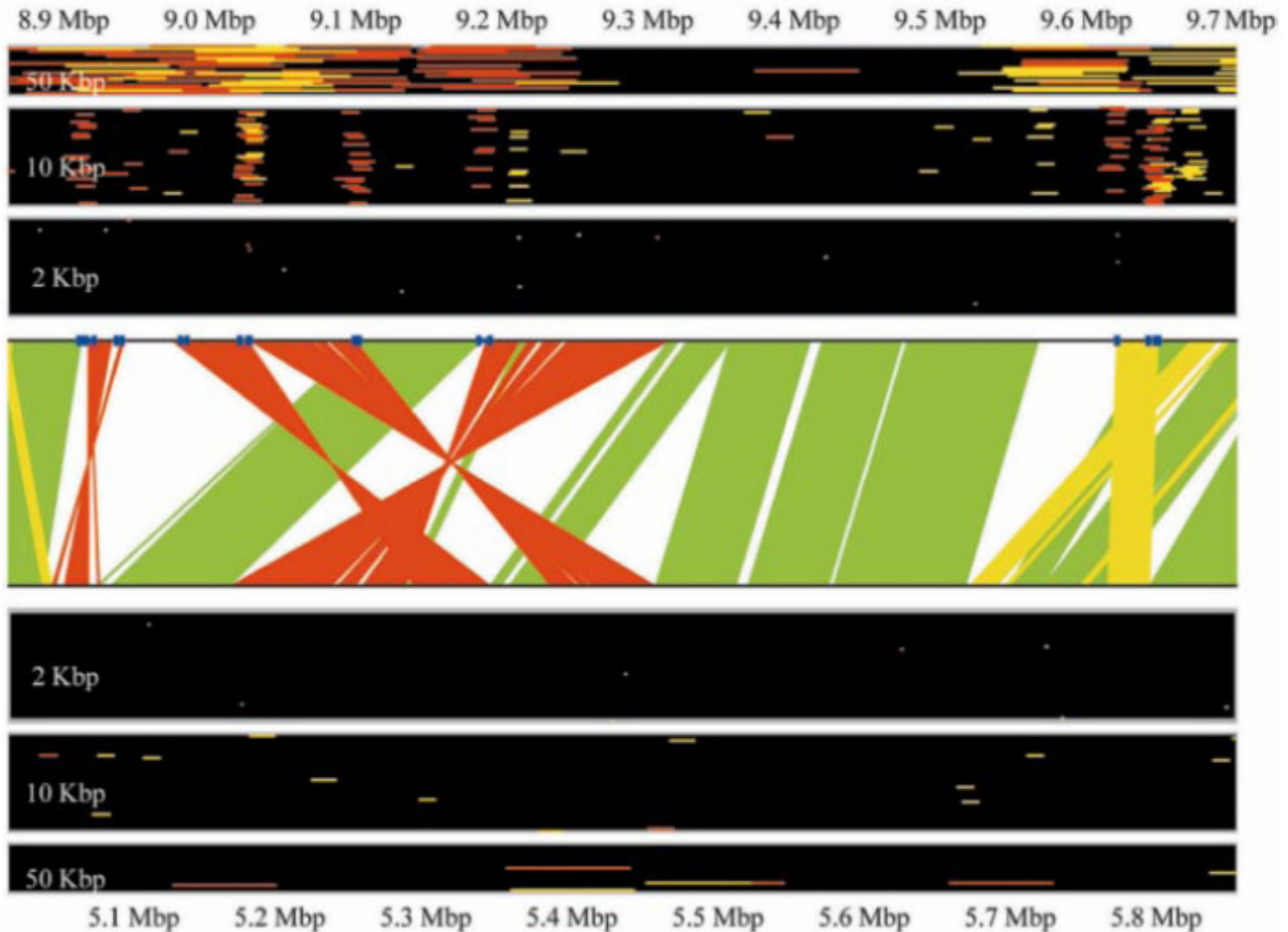


Figure 9.3. The different types of cytological, genetic and physical map of a chromosome. Genetic map distances are based on crossover frequencies and are measured in centiMorgans (cM), while physical distances are measured in megabase pairs (Mbp) or kilobase pairs (kbp)

Szekvencia térképek összehasonlítása



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

Kromoszóma fizikai térképek



Humán Genom Projekt - eredmények

- Tervezett 15 év helyett 2003-ban fejeződött be
 - 2000-2001: 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

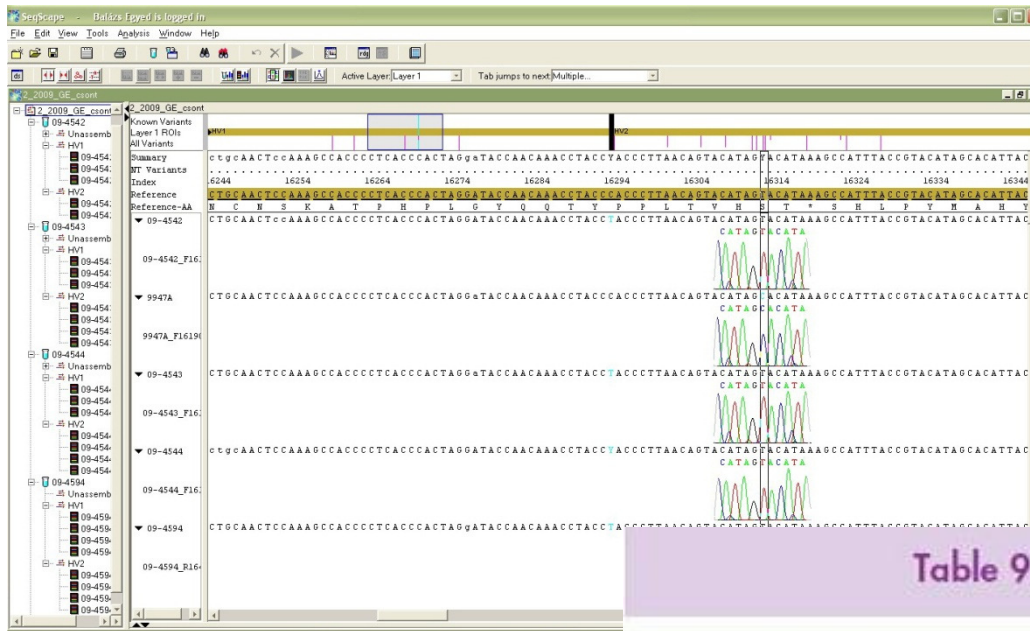


Table 9.1. Curated genome sequencing projects

Organism (type)	Web site(s)
<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)	<u>www.ncbi.nlm.nih.gov/genome/guide/human/</u>

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)



Sequence of one gene

```

TCCTTTCCGG AACGGTTGGC GTCTGCGCAC GCGGGTGTGG GGCATGACAT
GCCGCCCCAG GAACAACCCC GACACGGCTT TAAGCCTCTC AAATCGCTGT
AGACATCATC TTTACGTGCT TGCCACCATT TGCCACCATT AGGGCTGTTT
CCGCGACGAC TCGCCATTCA ACCTCAGTCC TTCGGGTTGA GCGAGTGGGT
CGCGCGCAAG GTGCGAATGG GTCGCGCGCA AAGTGTTCG CTGGCTGTAT
TATATGCTGC CTATAGCGAG ACTAACGACC CACACTTTCA CACAAGGATT
TCCCGCTAAT GGGTACCTCG CGTCAGGACC TTGACGCAAG CGCGCCTTCG
GTTGGCCCCA AGCTTGCTAG GACTACTTAT CTTGAGCTCA TTTAACATCC
CGGCGCCTCT CCGGGAGCGG TCGTGCAGAA GAAGTCAAAC CCGGAACGGC
GTTGACAAAG CGTGAGACA TCGATACCTC TGTGTCAGCG GCCACAAATC

```

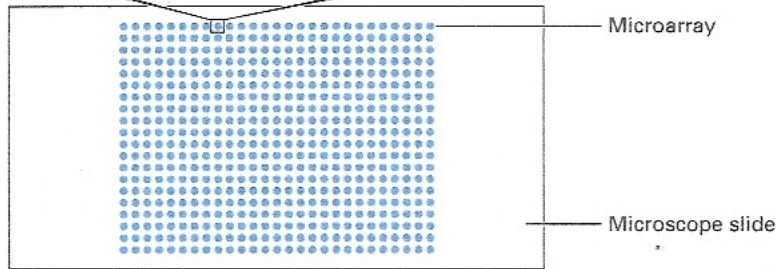
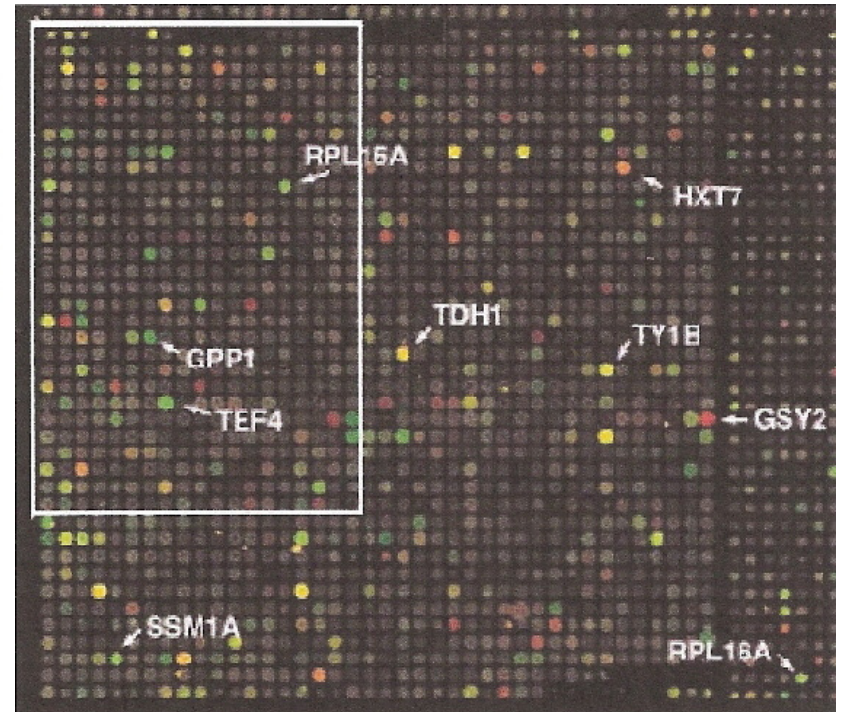
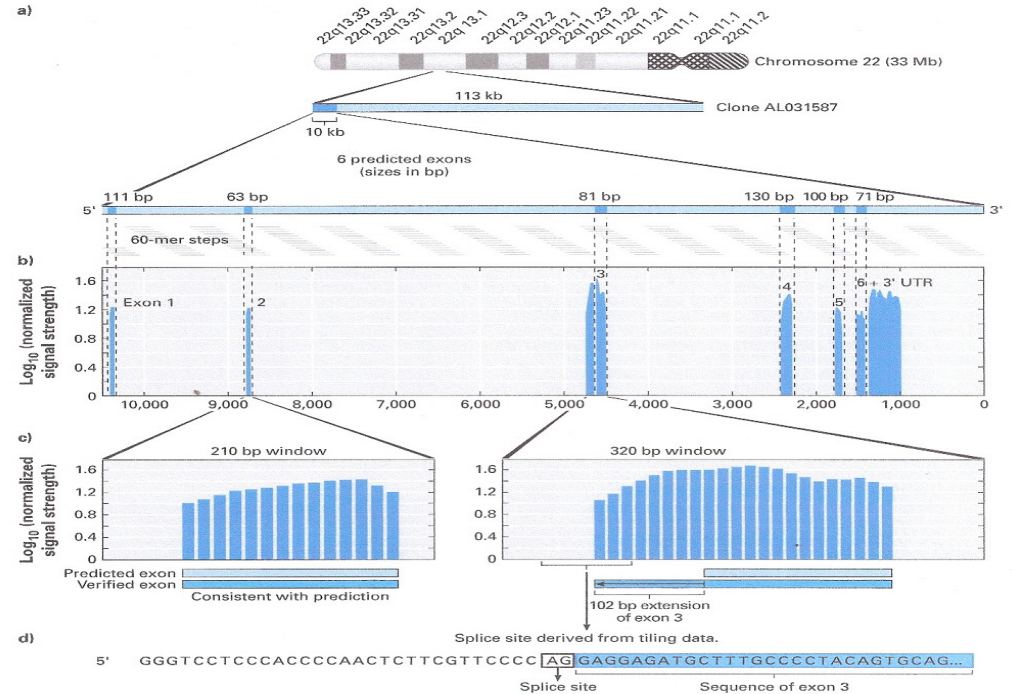
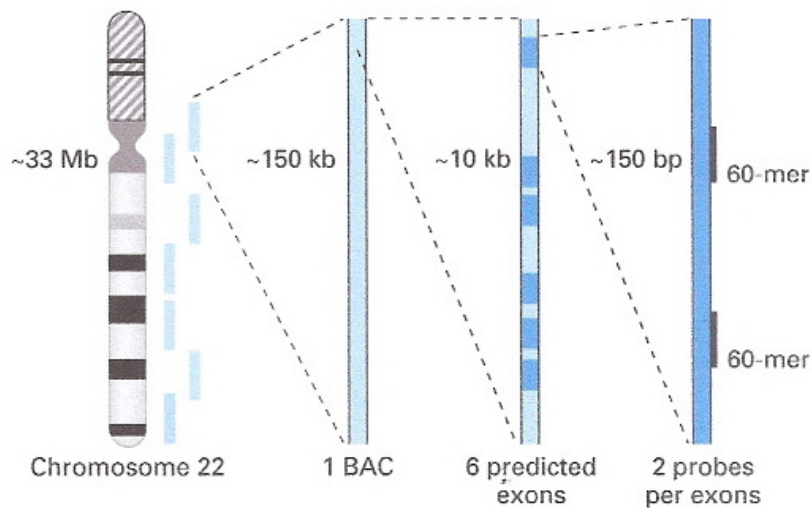
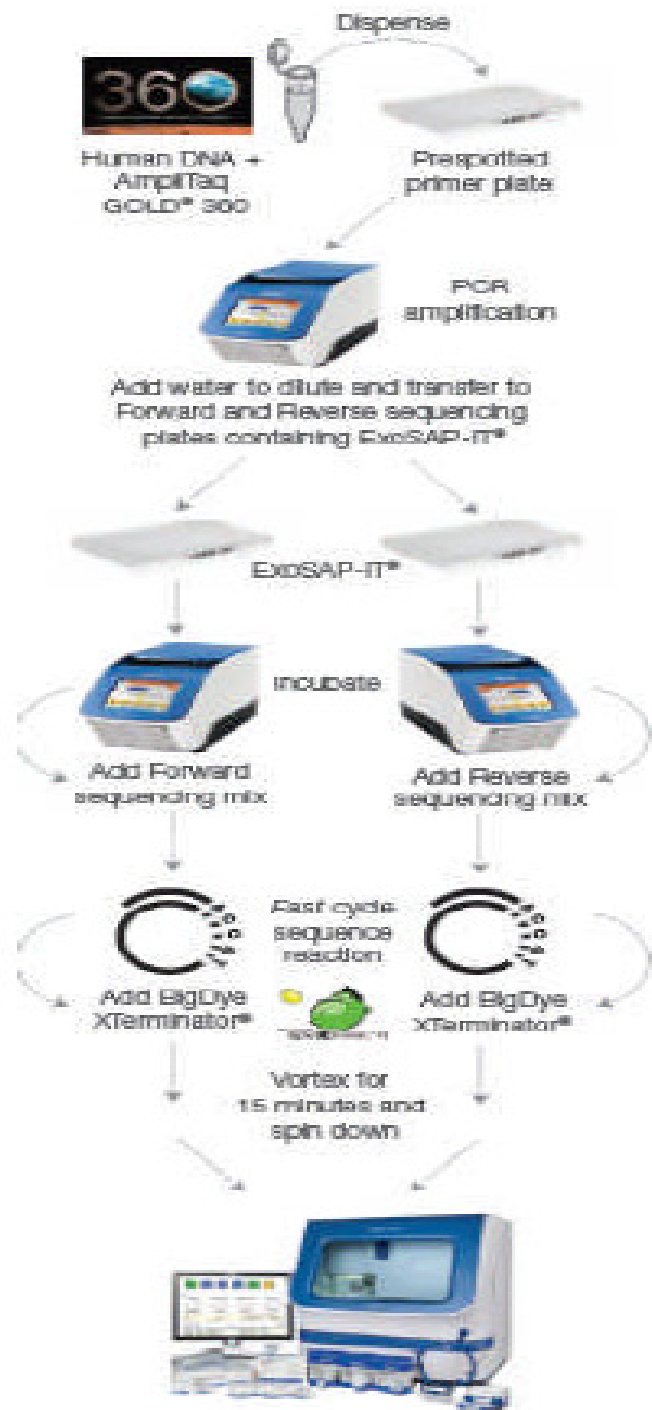


FIGURE 4.1 • DNA microarray. Each blue spot indicates the location of a PCR product on the glass slide. One particular spot has been chosen to illustrate the presence of one gene's sequence. On a real microarray, each spot is about 100 μm in diameter.



Two 60-mer probes selected for every predicted exon on chromosome 22





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

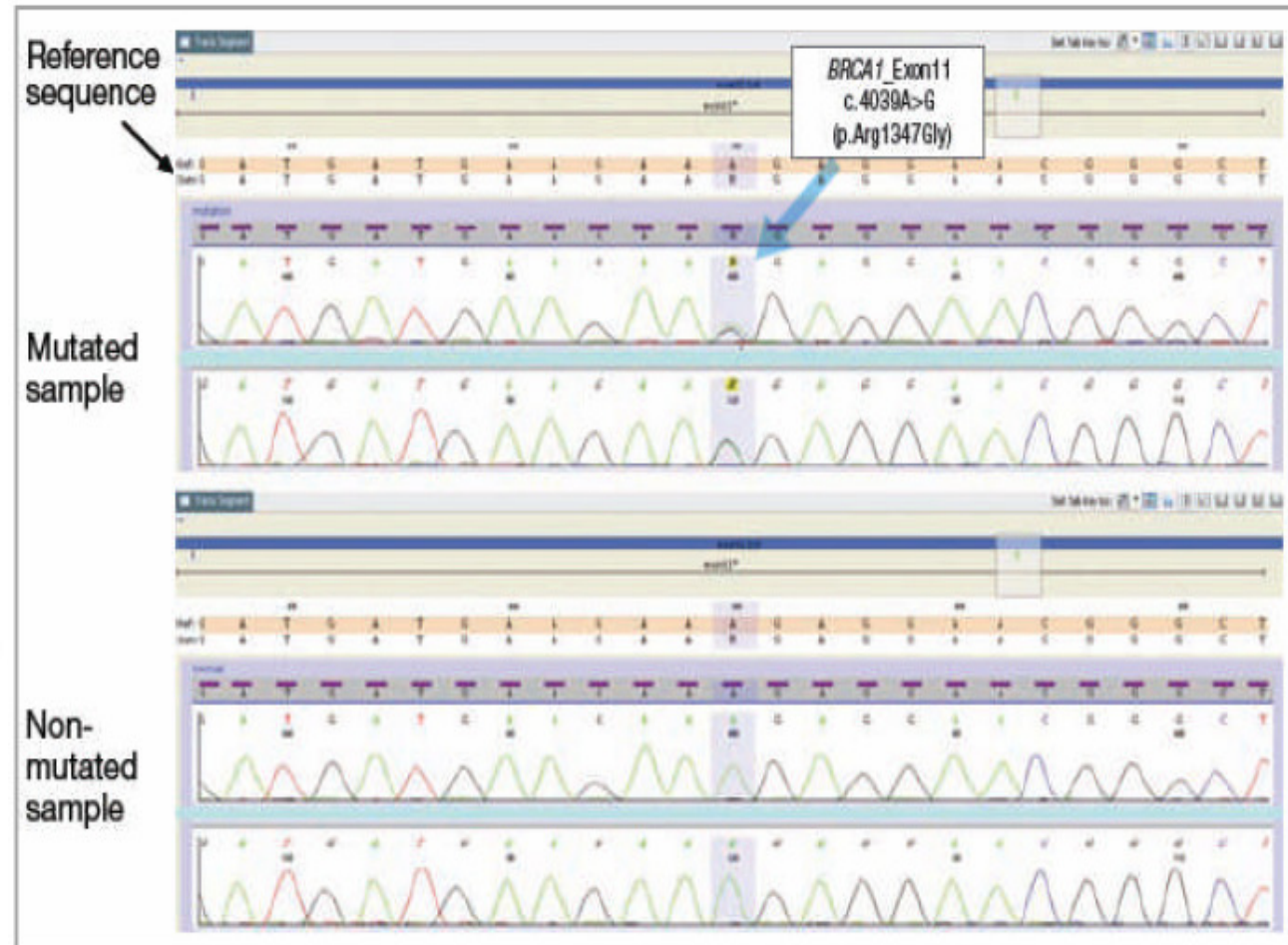
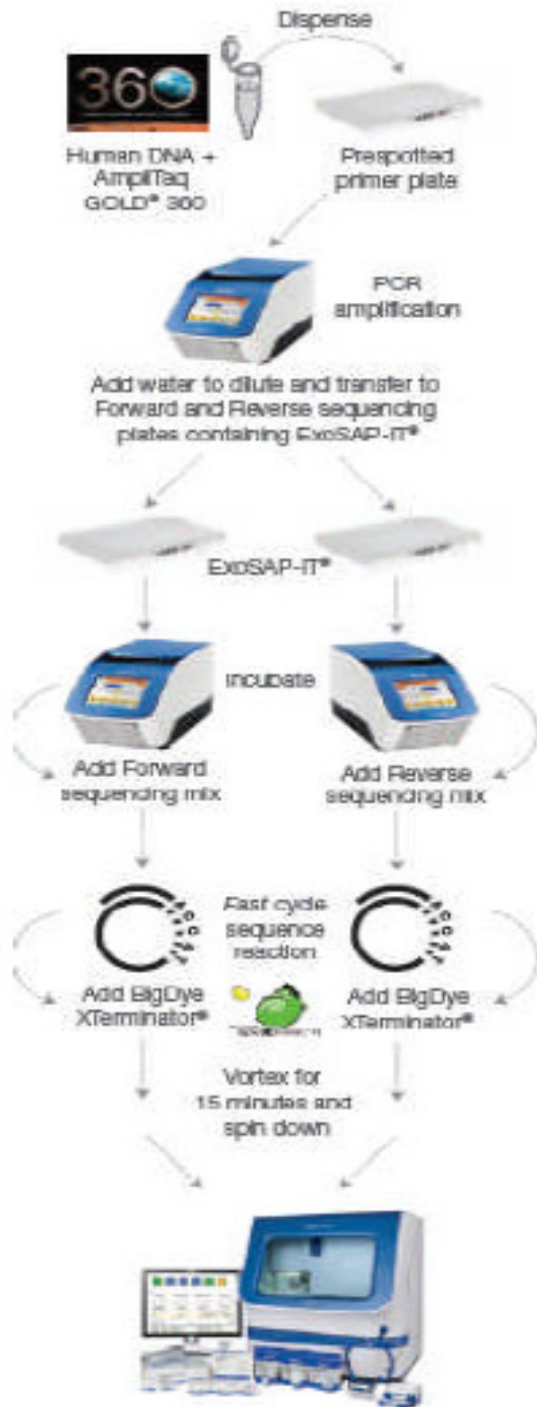
BRCA1 és BRCA2: 34 és 47 amplikon

	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-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 non-template control

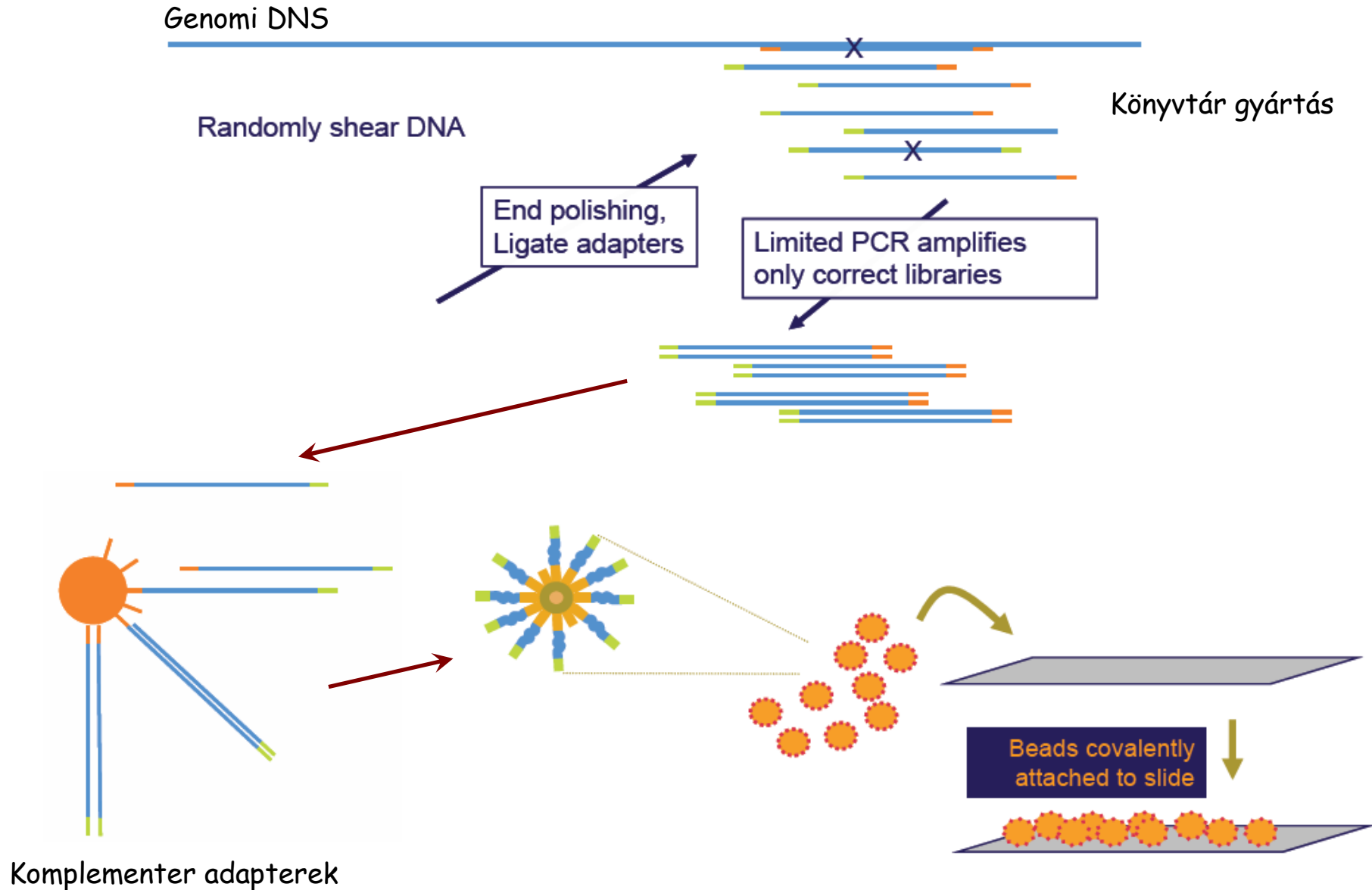
BRCA1 és BRCA2 gén resequencing

- mutációk diagnosztikai célú azonosítása



Next Generation DNA Sequencing: SOLID

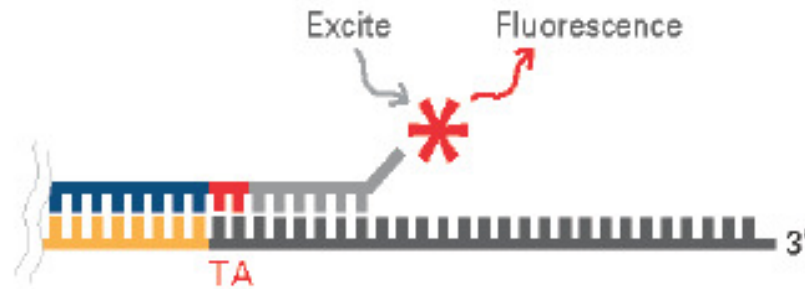
- Kémiai hasítás, amplifikálás és ligálás



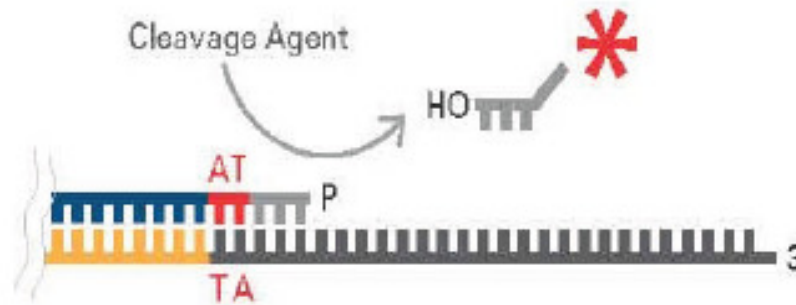
Next Generation DNA Sequencing: SOLID

- Kémiai hasítás, amplifikálás és ligálás

Ligation and Imaging



Cleavage



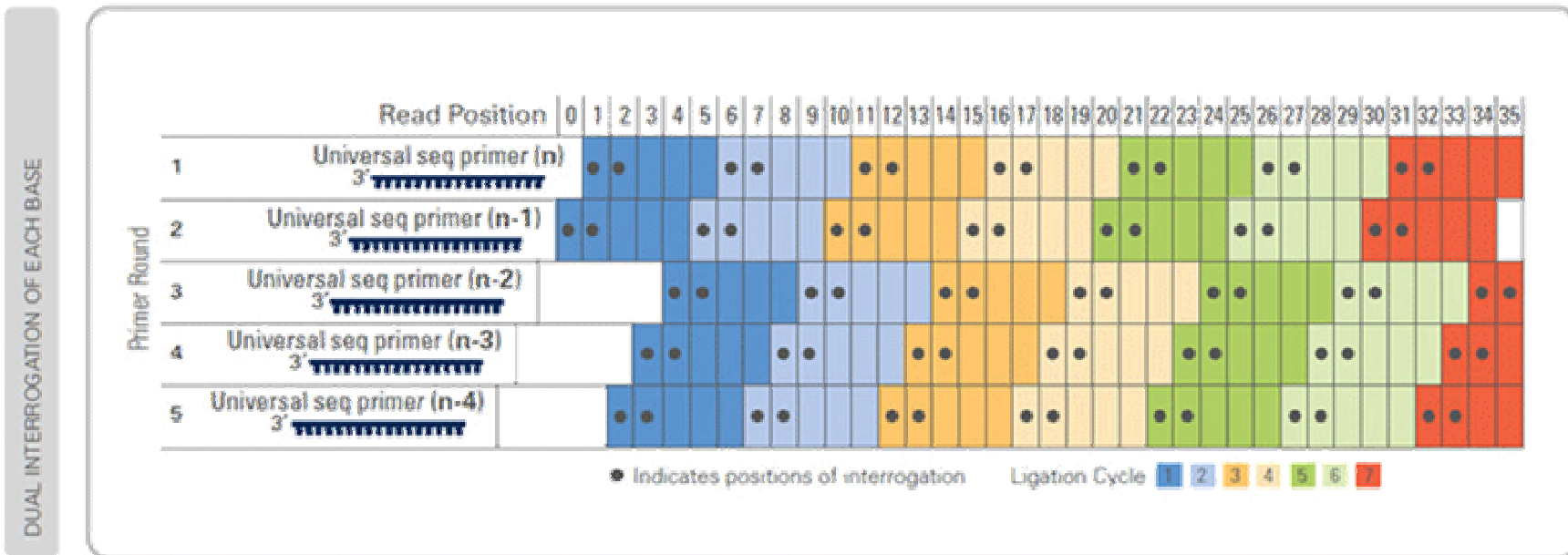
Repeat Ligation, Imaging, Cleavage

Ligation cycle 1 2 3 4 5 6 7 ... (n cycles)



Next Generation DNA Sequencing: SOLID

- Kémiai hasítás, amplifikálás és ligálás

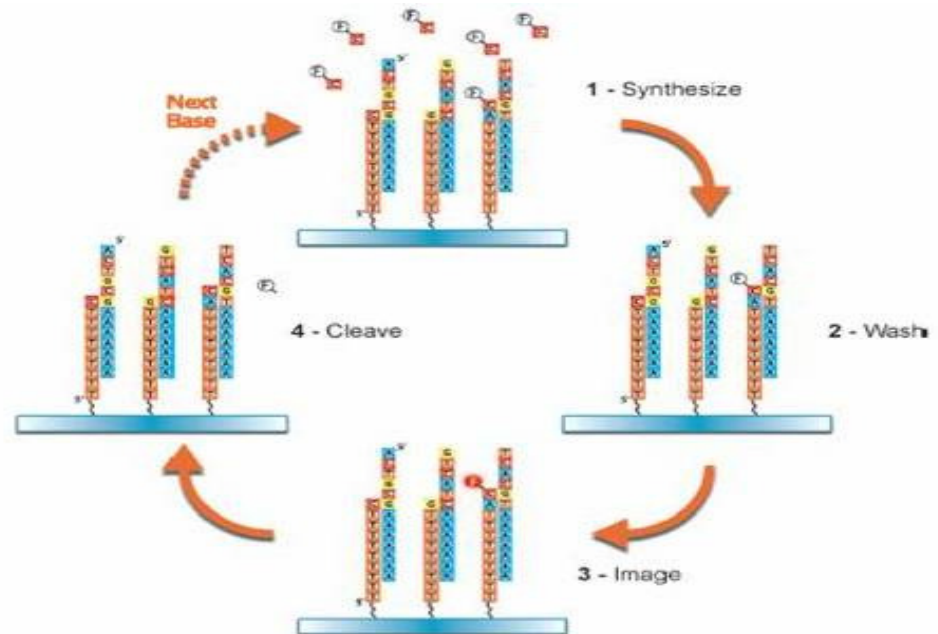
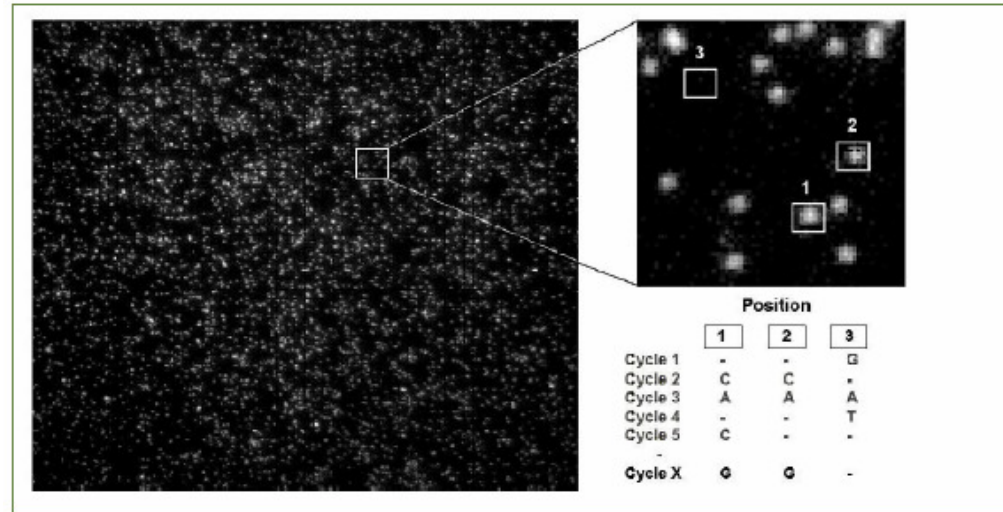


Ciklikus ligálás és primer annealing

Pontosság: 99.99 %

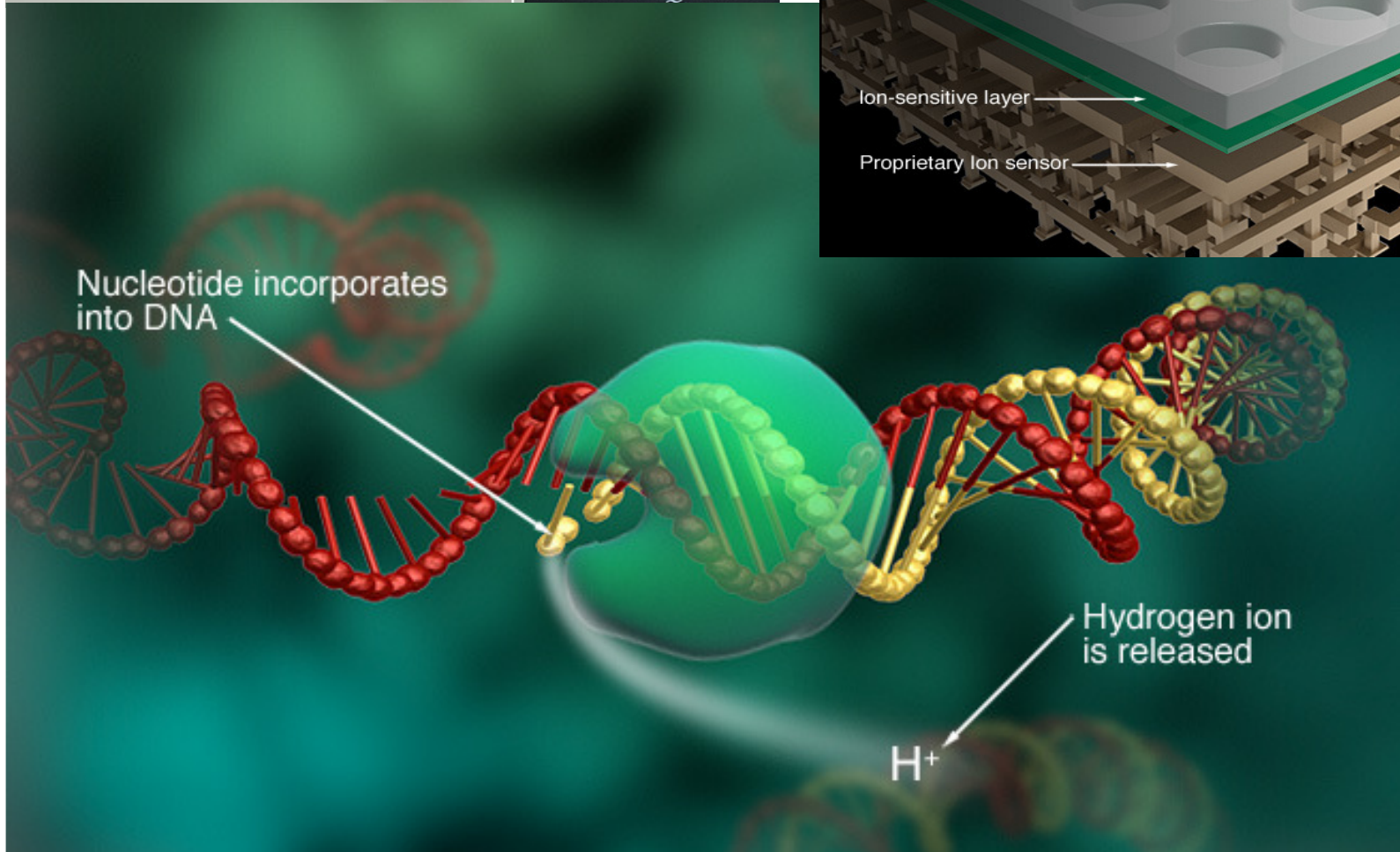
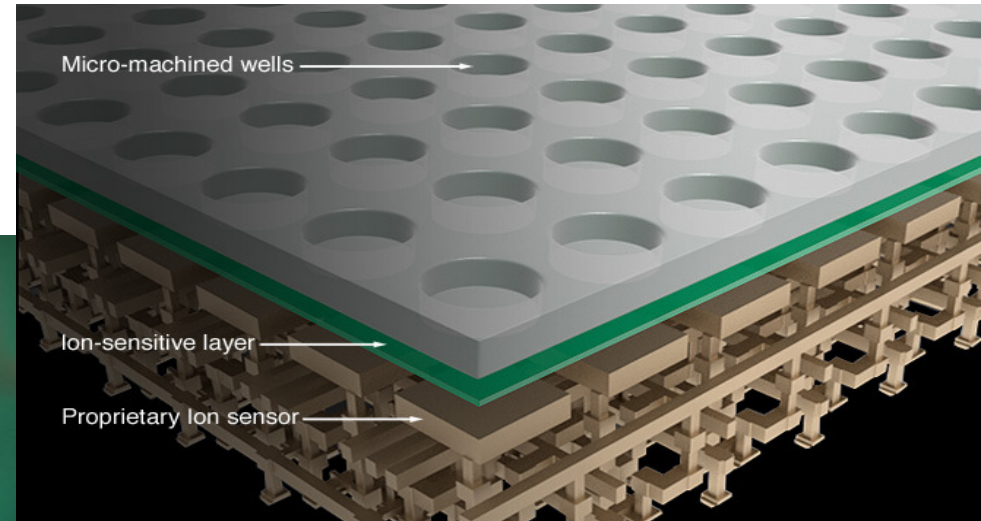
True Single Molecule DNA Sequencing: Helicos

- PCR amplifikálás nélkül



DNS szekvenálás félvezetőn: PGM

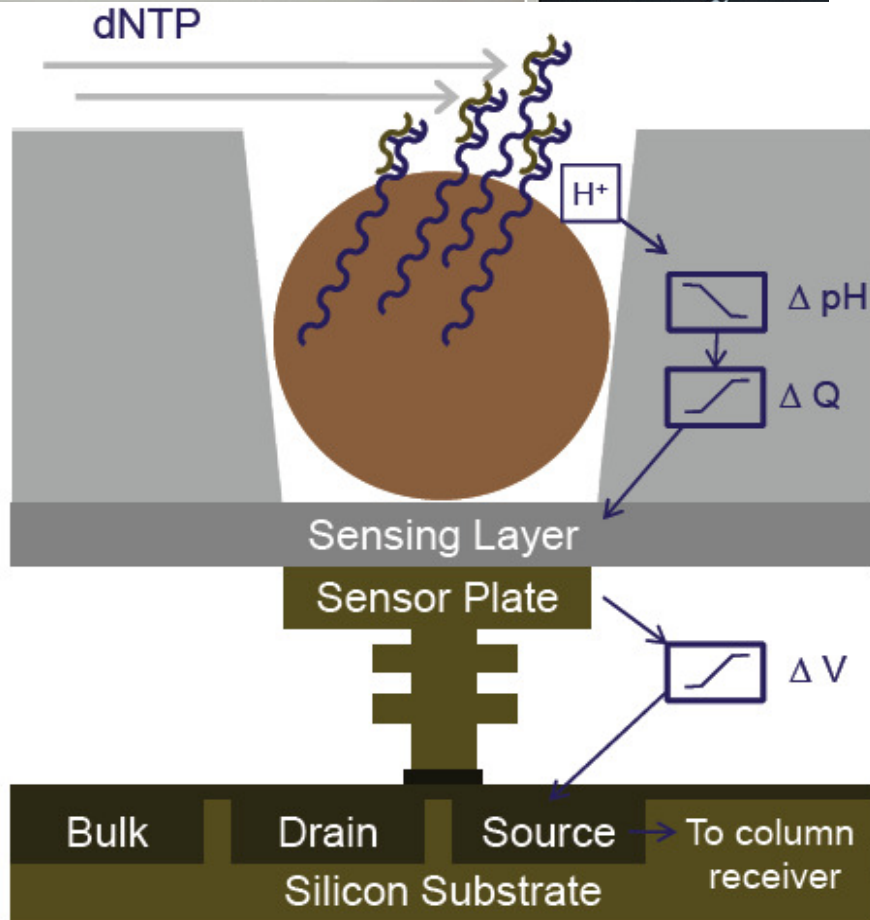
- hipergyors real-time szekvenálás





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

