

Új generációs szekvenálás

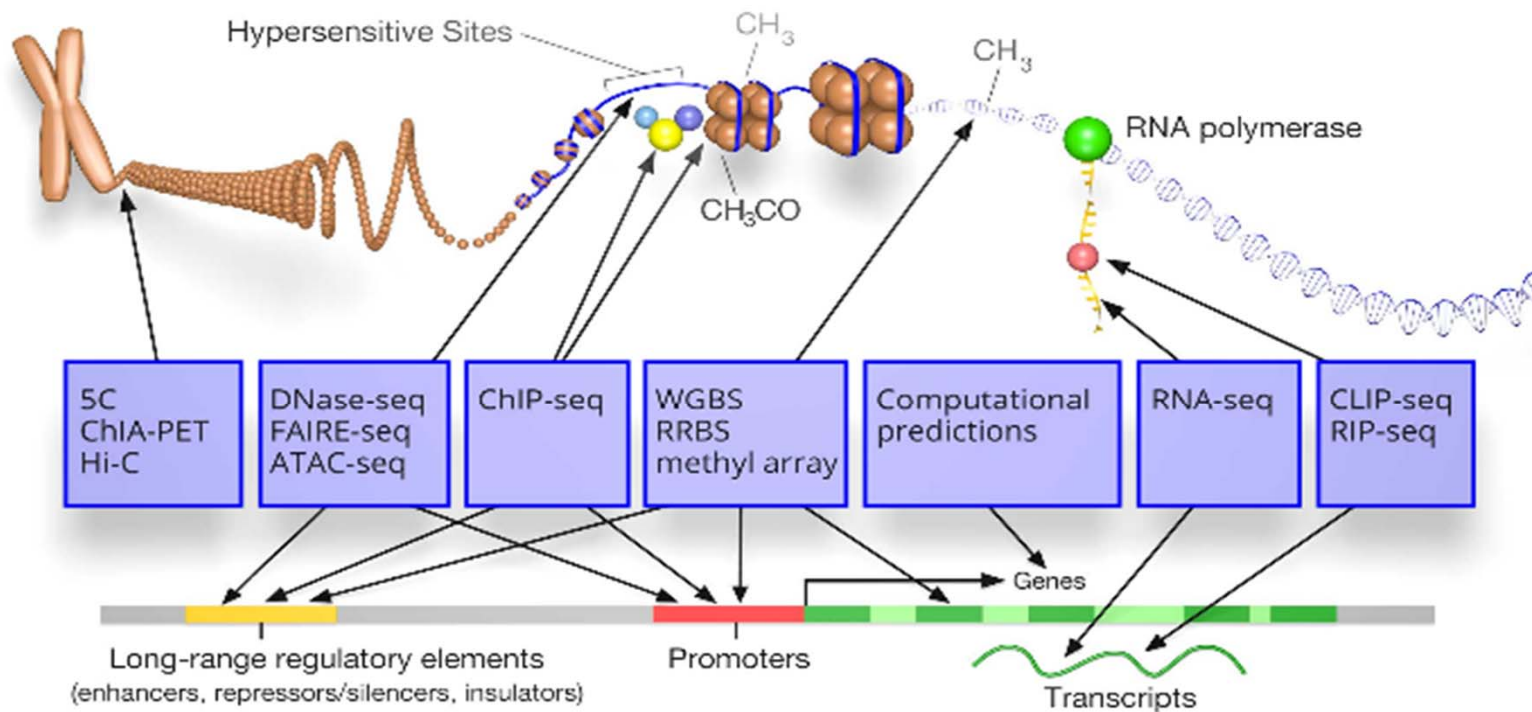
(Next Generation Sequencing)

Technikák

Alkalmazási területek

Bioinformatika, BME ABÉT, 2019 tavaszi félév

Előadó: Békési Angéla



Based on an image by Darryl Leja (NIH), Ian Dunham (EBI), Michael Pazin (NIH)

Mit és miért szekvenálunk?

DNS (genom) = az élő sejt programja
 genetikai kód = 4 betűs ABC-vel írt szöveg
 - és még sokkal több!!!

Nukleotid módosulások

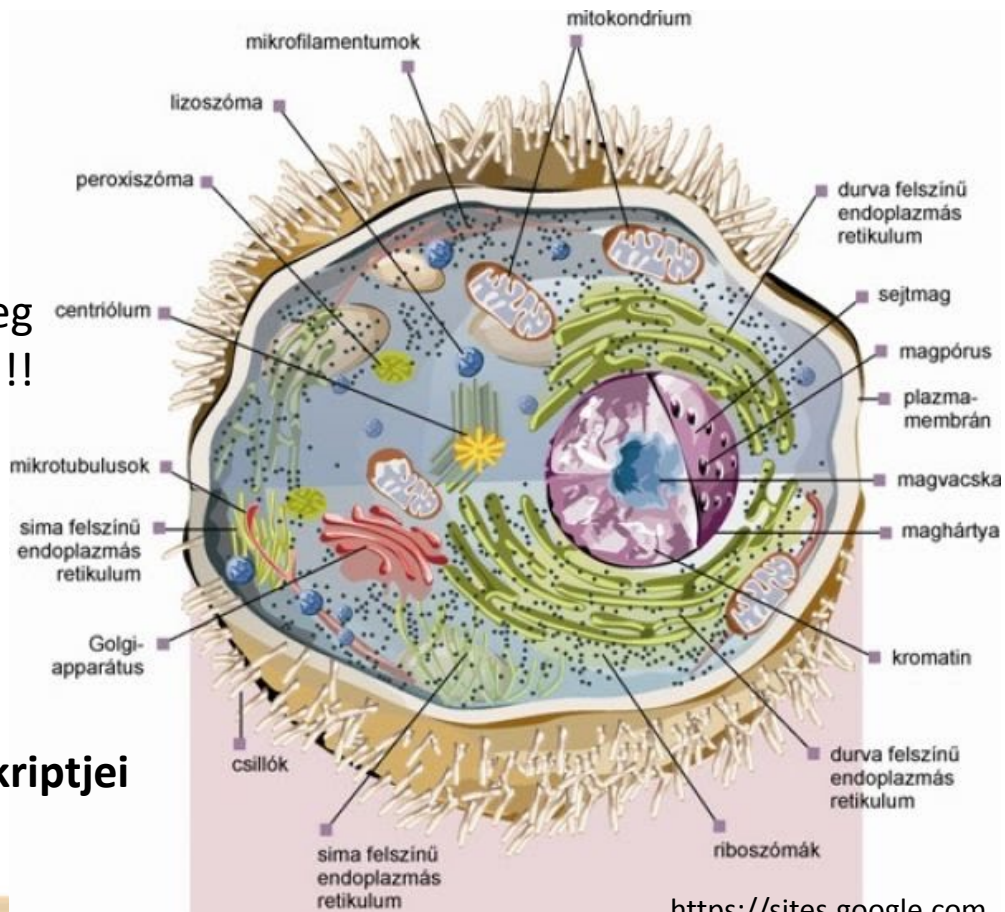
Kromatin szerkezet

Epigenetika

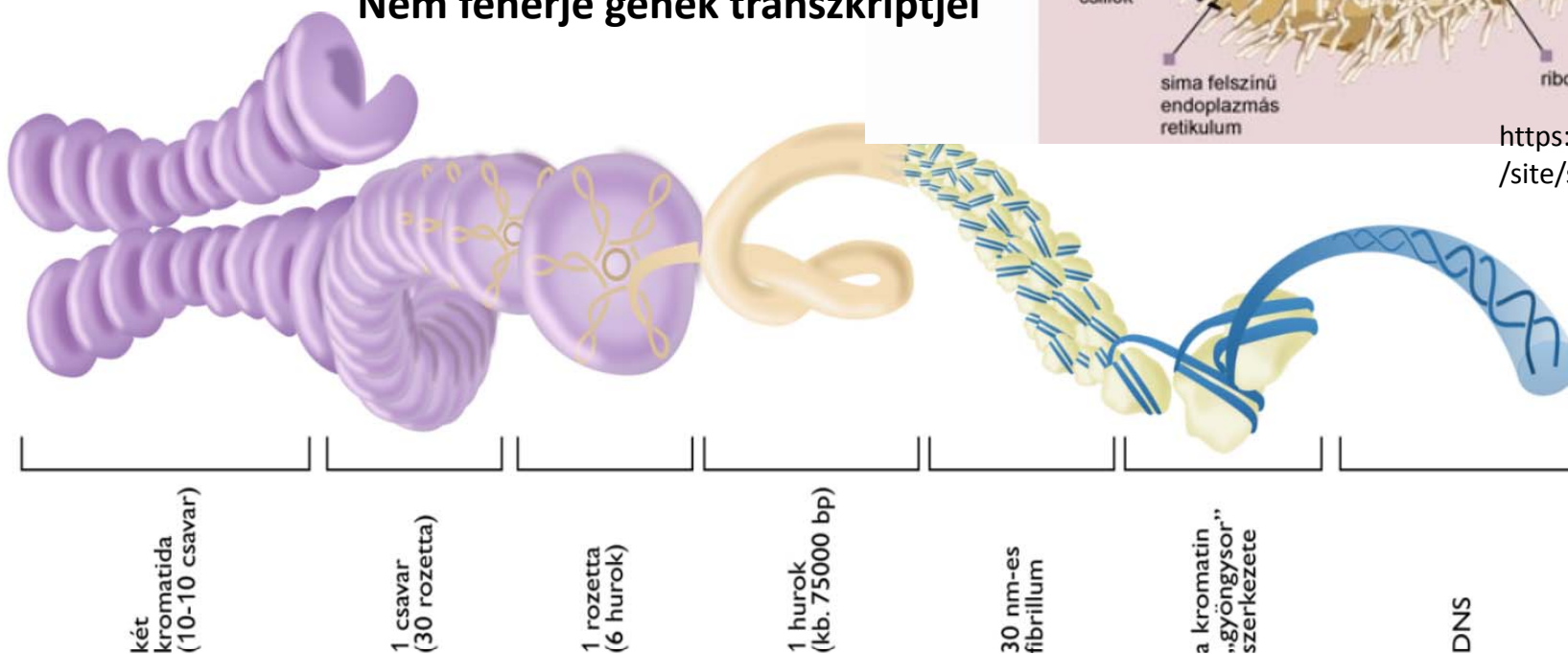
Transzkripció szabályozás

Poszt-transzkripció szabályozás

Nem fehérje gének transzkriptjei



<https://sites.google.com/site/steinnoemi/>



A poszt-genomi kor előzményei

A DNS örökítő anyag – kromoszómák, karyotipzálás 1930as évek

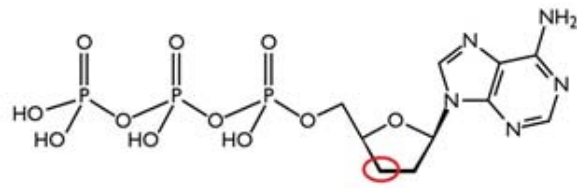
A DNS szerkezete – Watson és Crick 1953

DNS szekvenálás a Sanger féle didezoxi módszerrel 1977

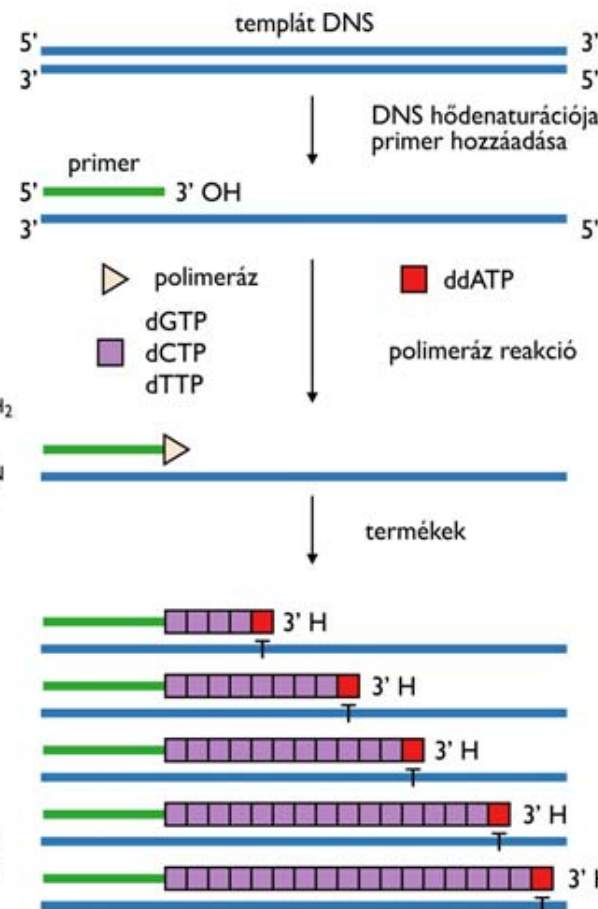


https://www.tankonyvtar.hu/hu/tartalom/tamop412A/2_011_0079_deak_alt_genetik

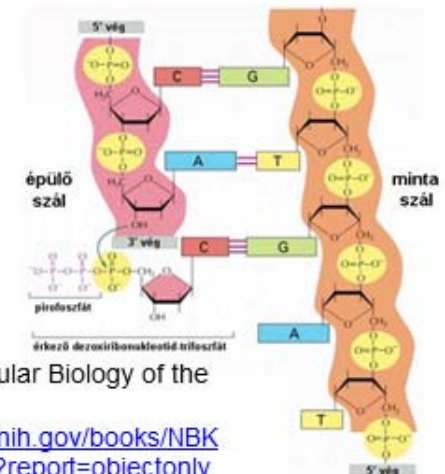
DNS-polimerizáció



ddATP: 2', 3'-didezoxi-ATP



<http://elte.prompt.hu/sites/default/files/tananyagok/ABiochemiaEsMolekularisBiologiaAlapjai/ch19s05.html>



Alberts et al.: Molecular Biology of the Cell, 4th edition - <http://www.ncbi.nlm.nih.gov/books/NBK26850/figure/A757/?report=objectonly>

A poszt-genomi kor előzményei

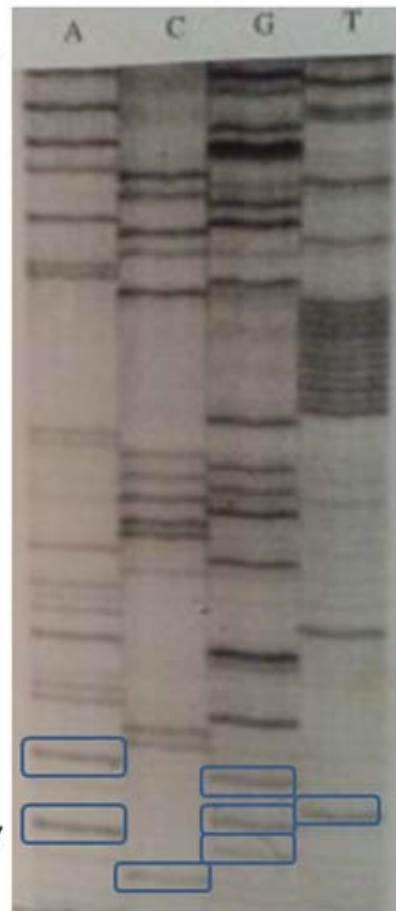
A DNS örökítő anyag – kromoszómák, karyotipzálás 1930as évek

A DNS szerkezete – Watson és Crick 1953

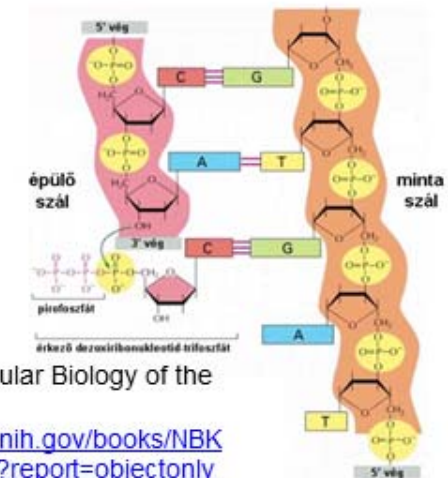
DNS szekvenálás a Sanger féle didezoxi módszerrel 1977

ddNTP a reakcióban

Futás iránya a gélben



DNS-polimerizáció



Alberts et al.: Molecular Biology of the Cell, 4th edition -
<http://www.ncbi.nlm.nih.gov/books/NBK26850/figure/A757/?report=objectonly>

A poszt-genomi kor előzményei

A DNS örökítő anyag – kromoszómák, karyotipzálás 1930as évek

A DNS szerkezete – Watson és Crick 1953

DNS szekvenálás a Sanger féle didezoxi módszerrel 1977

Automatizálás lehetősége:

4 féle fluorofórral jelölt ddNTP

Kapilláris gél-elektroforézis

Folyamatos kromatogramm



2001

humán genom projekt

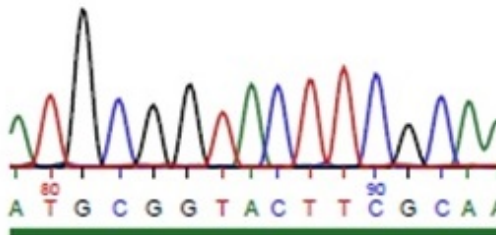
$3 \cdot 10^9$ bp – haploid genom

...Annotálás: ENCODE konzorcium

2004-től új generációs szekvenálási módszerek → 2012

1000 humán genom

...



A poszt-genomi kor lehetőségei és kihívásai

DNS szekvenálás – lineáris információ elolvasása

NGS technikák – exponenciálisan gyorsuló ütem és csökkenő ár

Alapvető változások a kísérletes tudományban is:

Fehérje tudomány:

-predikciós lehetőségek

-új fehérje azonosítási (MS) és klónozási stratégia (cDNS könyvtárak)

Rendszer szintű vizsgálódást célzó új tudományterületek: ...omika

genomika

transzkriptomika

proteomika

Trükkös kísérleti elrendezések (mit szekvenálunk!!!)

– intenzív annotáció, rendszerezés (ENCODE)

A mikroszkópos és molekuláris vizsgálatok összekötésének új lehetőségei:

citogenomika – a kromatin 3D szerkezetének feltérképezése

epigenomika – a DNS módosítások és a kromatin szerkezet jelentősége

Nagy áteresztőképességű szekvenálási lehetőségek fejlődése:

1. Microarrays / DNS chipek

Meghatározott szekvenciák detektálása, szemi-kvantitatív *expressziós adatok*,
klinikai jelentőség

2. Next Generation Sequencing (NGS)

(massively overlapping sequencing, deep sequencing, HTP-sequencing (HTS))

2A) second generation sequencing (2004-2010)

Short reads, paired-end vagy mate-pairs, PCR amplikonok ritkán egyedi molekulák

2B) third-generation sequencing (2010-től), *next-next generation, now-generation*

long-read – a problémás repetitív szakaszokra, szerkezeti variánsokra, humán genom bef-e

Egyedi molekula (**single molecule (SM)**) – nincs PCR-ből eredő mennyiségi eltolódás, hiba

Valós idejű detektálás (**real-time (RT)**) – gyorsabb, olcsóbb

Kémiai **jelölés nélküli** – kevesebb változó, kevesebb hiba, olcsóbb, gyorsabb

Single-cell sequencing

3. Jövendő lehetőségek

Érdekes próbálkozások alapvetően új megközelítések kidolgozására: pl.: alagútmikroszkópia, mechanikai azonosítás, tömegspektrometria (MS), elektronmikroszkópia (EM), röntgen foton-elektron mikroszkópia (XPS), RNAP, FRET és Raman spektroszkópia

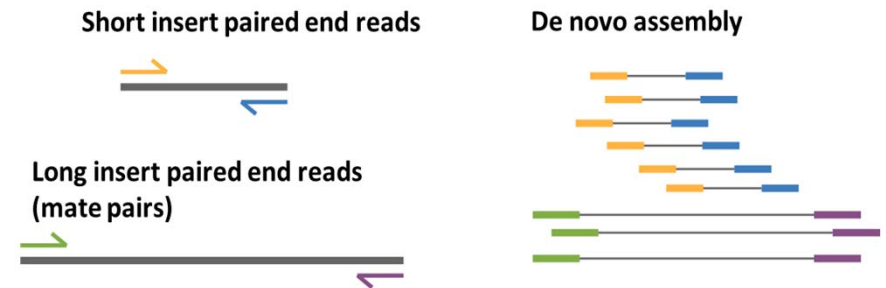
Közös lépések a különböző technikákban

I. Templát preparálás:

1) Egyedi DNS molekulából klonálisan felszaporított templát

DNS fragmentálás → fragmens vagy „mate-pair” könyvtár generálás: adapter ligálás → klonális amplifikáció:

- *szilárd fázison bridging PCR
- *emulzió PCR (EmPCR)
- *picotiter PCR
- *digital droplet PCR
- *in situ rolling circle amplification (RCA)



„Könnyebb” detektálás, de az ún. signal dephasing növeli zajt és a hibás leolvasások számát, ill. rövidíti a leolvasást.

2) Egyedi DNS molekula templát

3 stratégia az egyedi molekulák térben elkülönülő immobilizálására:

- 1) univerzális primer egyedi molekulái
- 2) az egyedi DNS molekulák
- 3) a polimeráz egyedi molekulái

Elkerüli a PCR-ből származó mennyiségi eltolódásokat, ill. hibákat, viszont a detektálás trükkösebb. Dupla nt beépülés egy ciklusban vagy sötét nt beépülés → deléciók

Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1), 31–46. <http://doi.org/10.1038/nrg2626>

WJ, Ansonge. (2015). Next Generation DNA Sequencing (II): Techniques, Applications. *Journal of Next Generation Sequencing & Applications*, 01(S1).

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351

Közös lépések a különböző technikákban

II: Szekvenálás:

Sequencing by synthesis (SBS) $\text{DNS/primer-3'-OH} + \text{dNTP} \rightarrow \text{DNS/primer-dNMP} + \text{PP}_i + \text{H}^+$
DNS polimeráz

1) Ciklikus reverzibilis termináció (cyclic reversible termination (CRT))

1 nt beépülés, mosás, leolvasás, hasítás, mosás
pl. Illumina

2) egy-nukleotid hozzáadás (single-nucleotide addition (SNA))

1. dNTP, beépülés, jeldetektálás, maradék dNTP elbontás, 2. ciklus: 2. dNTP...
pl. piroszekvenálás

3) Valós idejű szekvenálás (real-time sequencing (RTS))

nt beépülés, lehasadó fluorofór, jeldetektálás...
pl. PacBio Sequel

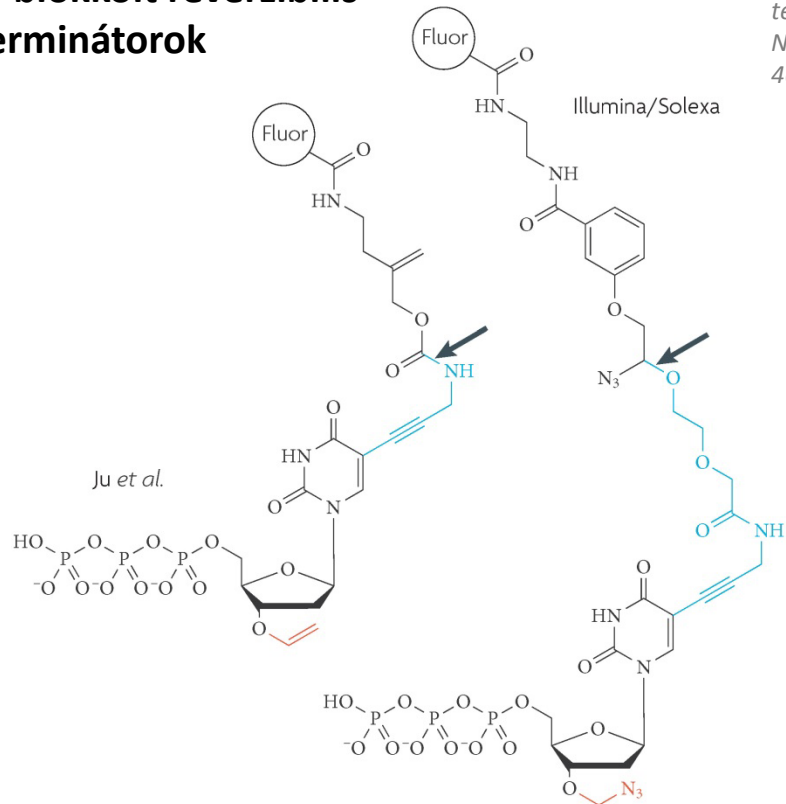
Sequencing by ligation (SBL) $\text{DNS}_1 \text{ 3'-OH} + \text{5'-PO}_4\text{-DNS}_2 \rightarrow \text{DNS}_{1-2}$
DNS ligáz

4) Ligálás által történő szekvenálás

1-2 bázisos fluorescens próba, hibridizálás, ligálás, mosás, leolvasás, mosás(hasítás, emésztés), regenerálás

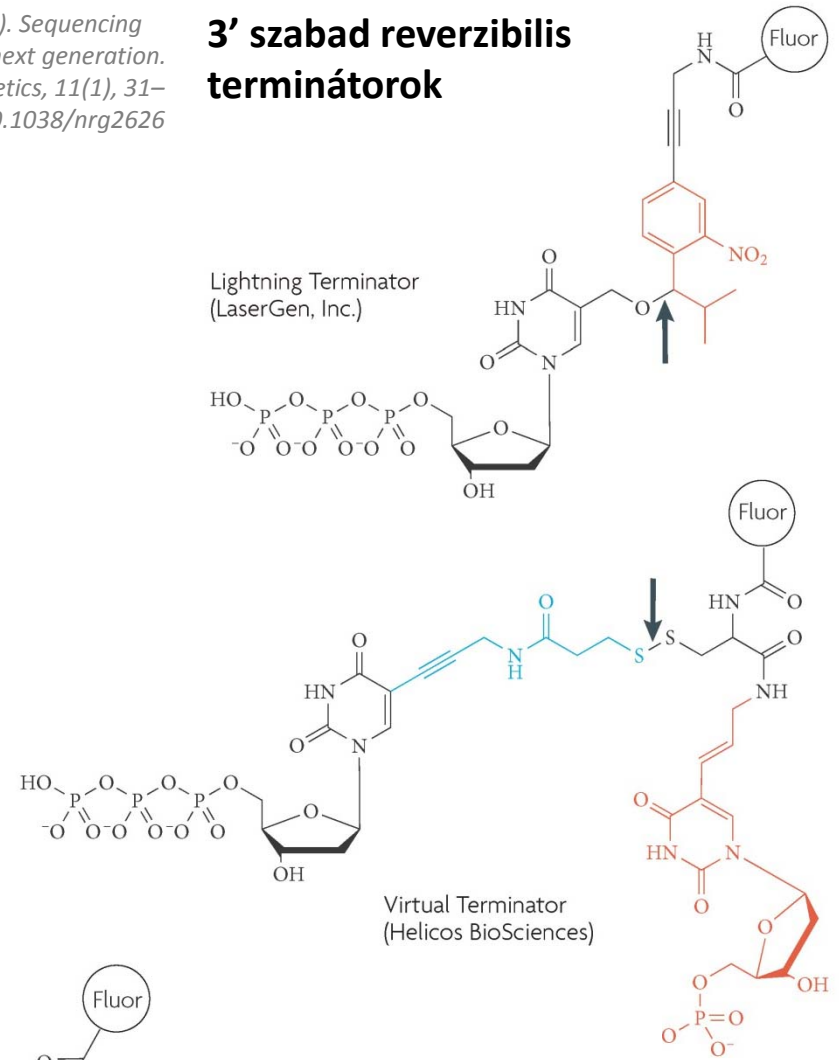
***Nanopóruson keresztül történő elektroforetikus áthaladás során**

3' blokkolt reverzibilis terminátorok

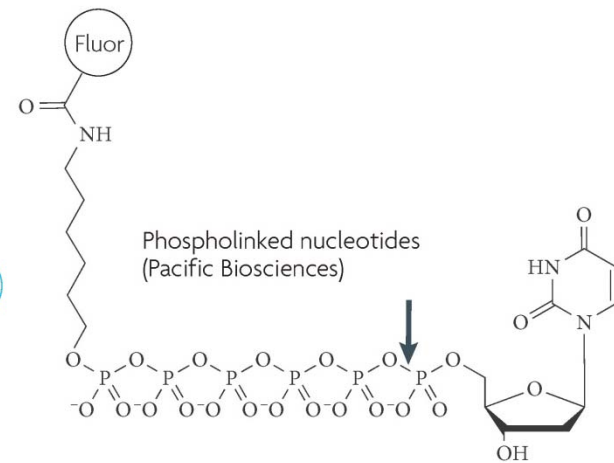
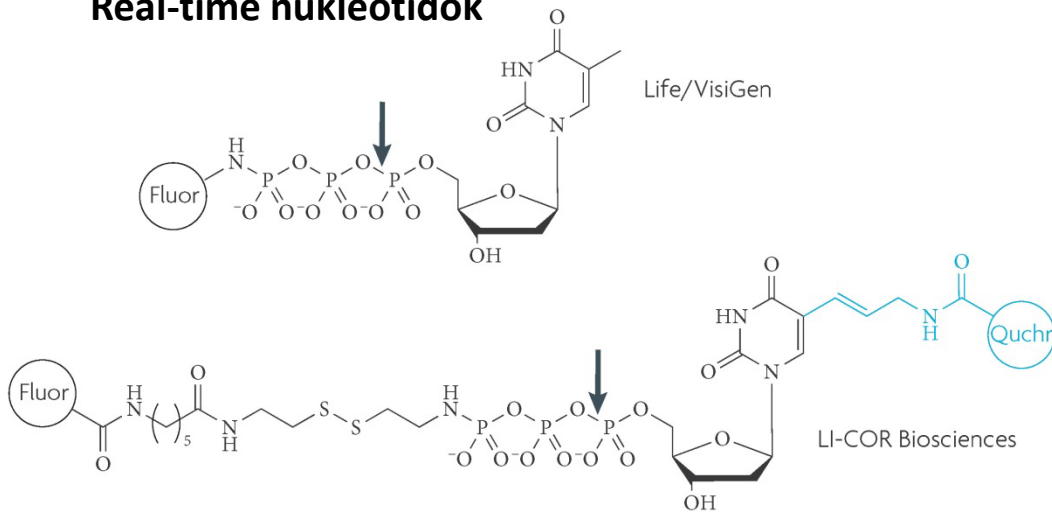


Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1), 31–46. <http://doi.org/10.1038/nrg2626>

3' szabad reverzibilis terminátorok



Real-time nukleotidok



Közös lépések a különböző technikákban

III. Detektálás, általában képalkotó eljárással (Imaging/detection):

Total internal reflection fluorescence (TIRF) – fluorescens jel pl. Illumina, Helicos

Epifluorescens mikroszkóp – fluorescens jel pl. polony szekvenálás, SMRT

CCD kamera (high-resolution charge-coupled device) – bioluminszcens jel pl. piroszekvenálás

PH-feszültség → szemikonduktor pl. ion torrent

Szemikonduktor alapú mikrochipek – áramerősség jel pl. nanopore technika

IV. Adatelemzés:

A módszerekről részletesen következő órákon lesz szó.

Az alkalmazott technika és a feltett biológiai kérdés nagyban meghatározó!!!

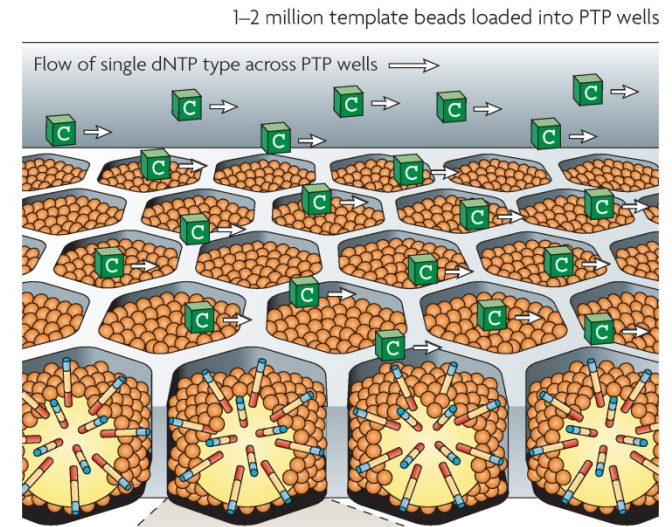
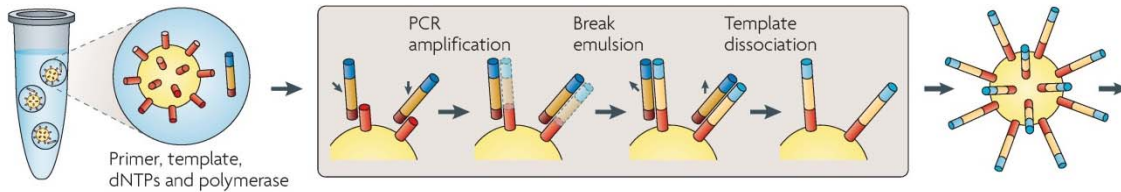
Piroszekvenálás (pyrosequencing):

Az első NGS módszer **454 Life technologies/ Roche**

- 1) DNS fragmentáció, adapter ligálás
- 2) Templát amplifikáció emulzió PCR-rel (EmPCR)

Marguiles et al (2005) "Genome sequencing in microfabricated high-density picolitre reactors" *Nature* 437, 376-380

a Roche/454, Life/APG, Polonator Emulsion PCR
One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion

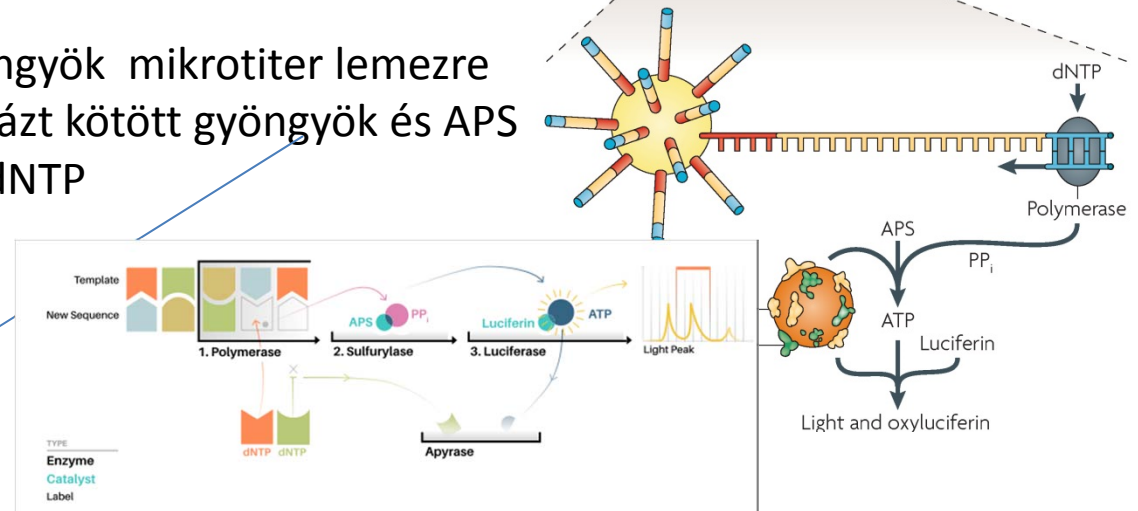


3) Szekvenálás:

- *amplifikált DNS-t tartalmazó gyöngyök mikrotiter lemezre
- *további ATP szulfurilázt és luciferázt kötött gyöngyök és APS
- *áramlási kamra, egy ciklus - egy dNTP
- *polimeráz reakció → pirofoszfát

4) Detektálás

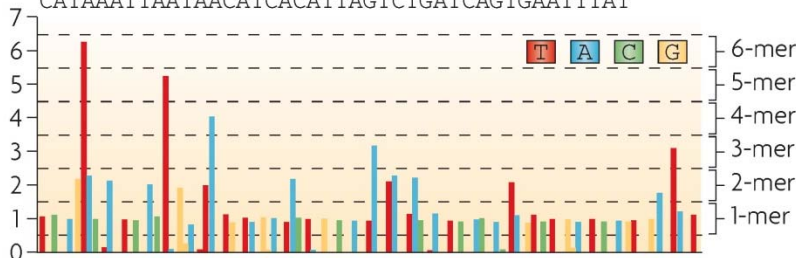
CCD kamera – biolumineszcens jel



d

Flowgram

TCAGGTTTTTTTAAACAATCAACTTTTTGGATTAAAATGTAGATAACTG
CATAAATTAATAACATCACATTAGTCTGATCAGTGAATTTAT



Több millió 200-400 bp read
Homopolimerek 6 nt-ig megbízhatóak
Leggyakoribb hiba az inzerció, aztán a delécio
2013: a Roche leállította
Qiagen - egyedi piroszekvenálás, kvantitatív...

Ion-torrent, Thermo Fisher Scientific

Hasonló elv, mint a pirosekvenálás:

~mintaelőkészítés

~mikrotiter lemez platform

~SBS, a dNTP-eket egyesével próbálja

A felszabaduló H^+ -okat detektálja a PP_i helyett

→ kihagyja a bonyolult enzim kaszkádot

→ más detektor: *licenz a DNA Electronics-tól*

A felszabaduló H^+ pH változást okoz, ez feszültség jellé konvertálódik, amit szemikonduktor chip detektál

Olcsóbb, rugalmasabb, egyszerűbb gyorsabb...

Több platform:

IonPGM: 5,5 millió read, ~400 b, 2Gb, 2-7h/futás

IonProtonSystem / IonPI chip: 80 millió read, ~200 b, 10 Gb, 2-4 h/futás

IonS5 / Ion540 chip: 80 millió read, ~400 b, 15 Gb, 2,5-5 h/futás

Kevés kiindulási DNS ~10 ng elegendő → mutáció analízis, gén expressziós profil

Kiegészítő kit-ek: target kiválasztáshoz, könyvtár preparáláshoz

Egyszerűsített adatelemzés. Ion Reporter Software and Server

Illumina/Solexa, SBS (sequence by synthesis) - 2006

1.) Könyvtár preparálás:

A DNS vagy cDNS random fragmentációja
 Adapter ligálás az 5' és 3' végeken
 (vagy egy lépéses „tagmentation”)
 PCR amplifikáció és gélből tisztítás

2.) Klaszter generálás:

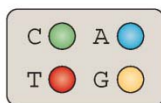
Áramlási cella felszínén immobilizált oligókkal
 Reverz komplementerei a könyvtár adaptoreinek
 Könyvtár betöltése hígán
 Klaszterek létrehozása szilárd fázisú amplifikációval (bridging PCR)

3.) Szekvenálás: 4-szín CRT

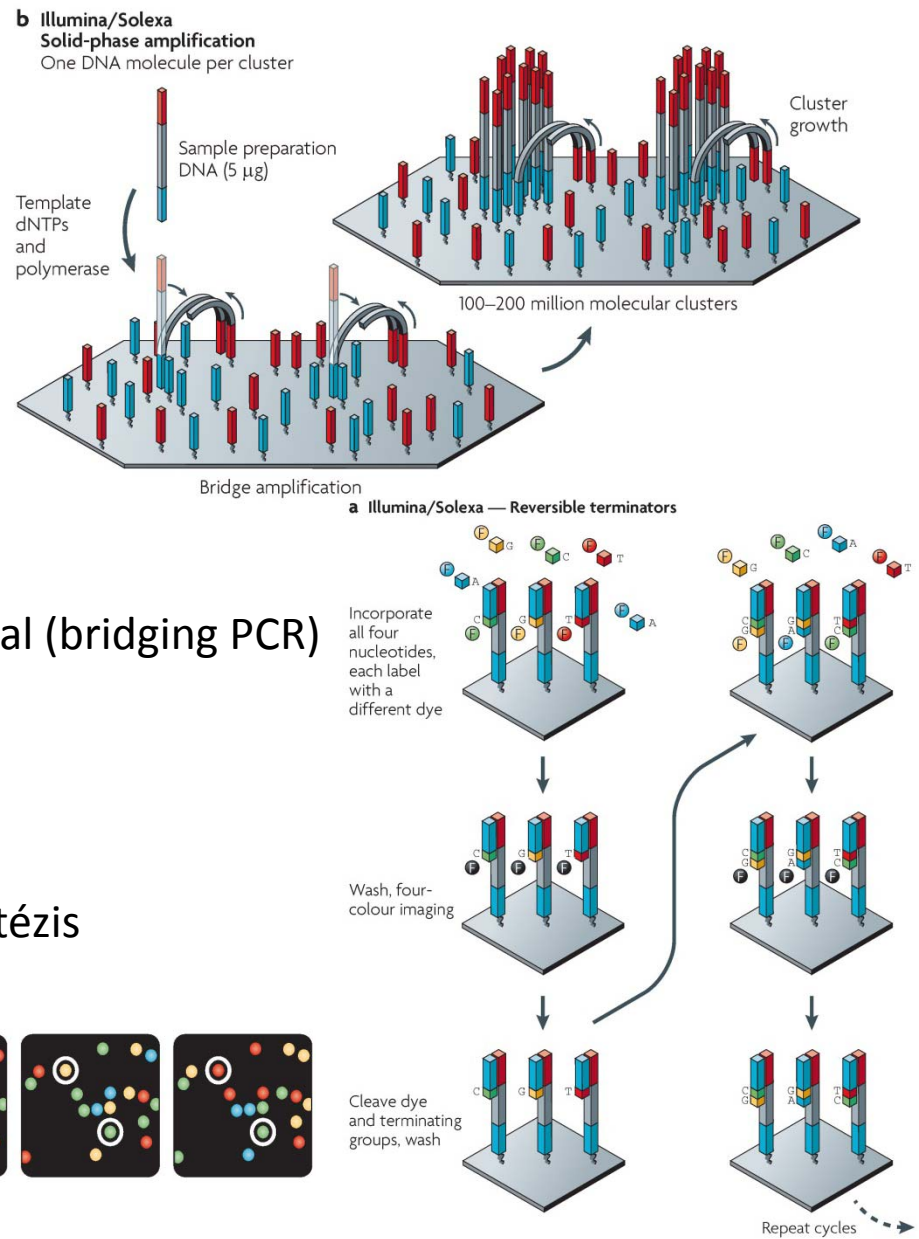
3'-O-azidomethyl reverzibilis terminátor
 TCEP: fluorofór hasítása és 3'-OH regenerálása
 Külön forward és denaturációt követő reverz szintézis
 →kontextustól független pontosság

4.) Detektálás

Total internal reflection fluorescence (TIRF)



Top: CATCGT
 Bottom: CCCCC



<https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html?langsel=/hu/>

Illumina dominálja/ta a piacot,
több szekvenátor az elmúlt 15 évben,
50-300 bp leolvasások,
Tökéletesített biokémia
0.001\$ / 1000 bázis 2016-ban

Illumina platformok

MiSeq: kis genomokra, amplikon vagy célzott génpanel szekvenálásra
25 millió leolvasás, 0.3-15 Gb, 1 áramlási cella, 2X300 bp leolvasás, 5-55 h

HiSeq: genomok, exomok, transzkriptomok szekvenálására

HiSeq 2500:

300 millió leolvasás, 10-300 Gb, 1-2 áramlási cella, 2X250 bp leolvasás, 7-60 h

HiSeq 4000: populáció skálájú teljes genom szekvenálásra is

2,5 milliárd leolvasás, 125-1500 Gb, 1-2 áramlási cella, 2X150 bp leolvasás, 7-60 h

Synthetic Long Reads (SLR)

Nagy DNS fragmenseket elválaszt emulzióban vagy mikrotiter lemezen

A partiókban külön-külön fragmentálja őket (~10 kb), és barcode-t ad hozzájuk

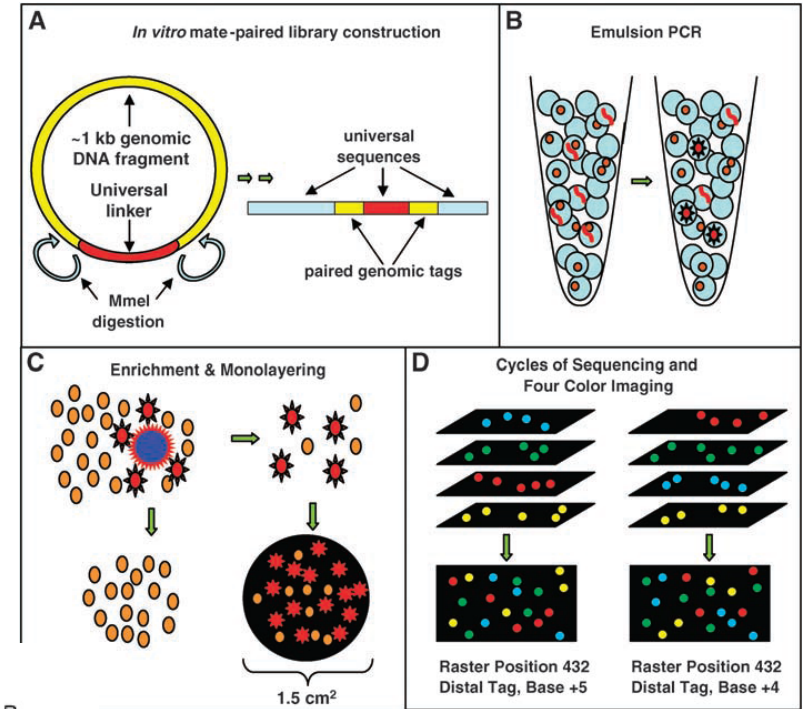
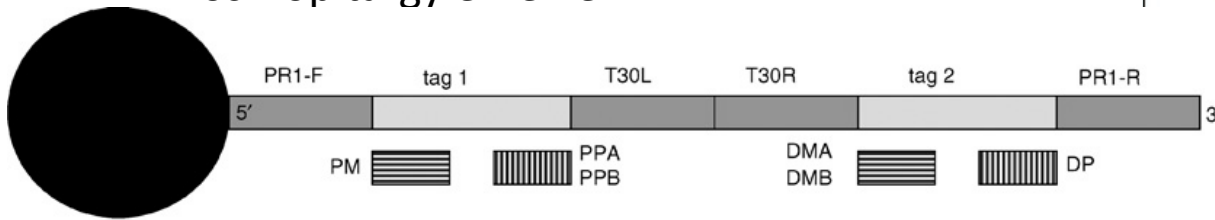
Ezután klasszikus HiSeq szekvenálás → a short read-eket bárkódok szerint először lokálisan szerelik össze, majd ezeket a „szintetikus hosszú szakaszokat” rendezik a referencia genomhoz

500 ng kiindulási DNS, csak 384 partió és index

Polony sequencing (Church's lab) -2006

Dover/Polonator instrument - 2009

- 1) DNS fragmentálás (~1kb)
- 2) paired end könyvtár gyártás: cirkularizálás T30 univ. linkerrel, cirkuláris PCR, emésztés Mmel enzimmal
→ mate-paired könyvtár, 135 bp
- 3) EmPCR primerrel való ligálás, EmPCR → könyvtár
- 4) A felszaporított fragmenseket kötő gyöngyök dúsítása, felvitel egy rétegben poliakrilamidban mikroszkóp tárgylemezre



Shendure, J., ... Church, G. M. (2005). Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science*, 309(5741), 1728–1732.

anchor primers

Anchor primer	Sequence	5' Phosphorylated?
PM	5'-AUCACCGACUGCCCA-3'	yes
PPA	5'-CGGCCGUACGUCCA-3'	no
PPB	5'-CGCCUUGGCCUCCGACT-3'	no
DMA	5'-AGUCGGAGGCCAAGG-3'	yes
DMB	5'-AGUUGGACGUACGGC-3'	yes
DP	5'-CCCGGGUCCUCAUUCUCT-3'	no

fluorescent nonamers



5'-Cy5-NNNNNNNT-3'

5'-Cy3-NNNNNNNA-3'

5'-Texas Red-NNNNNNNC-3'

5'-Cy5-N-Cy3-NNNNNNG-3'



5'-phos-TNNNNNNN-Cy5-3'

5'-phos-ANNNNNNNN-Cy3-3'

5'-phos-CNNNNNNN-Texas Red-3'

5'-phos-GNNNNNNN-Cy3-Cy5-3'

5) Szekvenálás – SBL

anchor primerek

degenerált 9mer próbák

egy pozíció fix – fluorescens jelölés

26 kombináció, 26 leolvasás

→ 2X13bp paired read

6) Detektálás: epifluorescens mikroszkóp

Porreca, G. J., ... Church, G. M. (2006). Polony DNA Sequencing. In *Current Protocols in Molecular Biology* (Vol. 76, p. 7.8.1-7.8.22). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Supported Oligonucleotide Ligation Detection (SOLiD) (SBL)

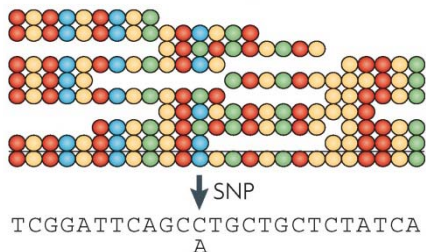
Thermo Fisher Scientific/Life Technologies/
Applied Biosystems/
Agencourt Personalized Genomics (APG)

- 1) DNA fragmentation, adapter ligation
- 2) EmPCR → fragmens könyvtár
- 3) Szekvenálás: SBL,
2-bázisos fluorescens póba
hasítható kötés az 5. és 6. nt közt
inozin univerzális bázis
- 4) Detektálás: fluorescens mikroszkóp

Two-base encoding: each target nucleotide is interrogated twice

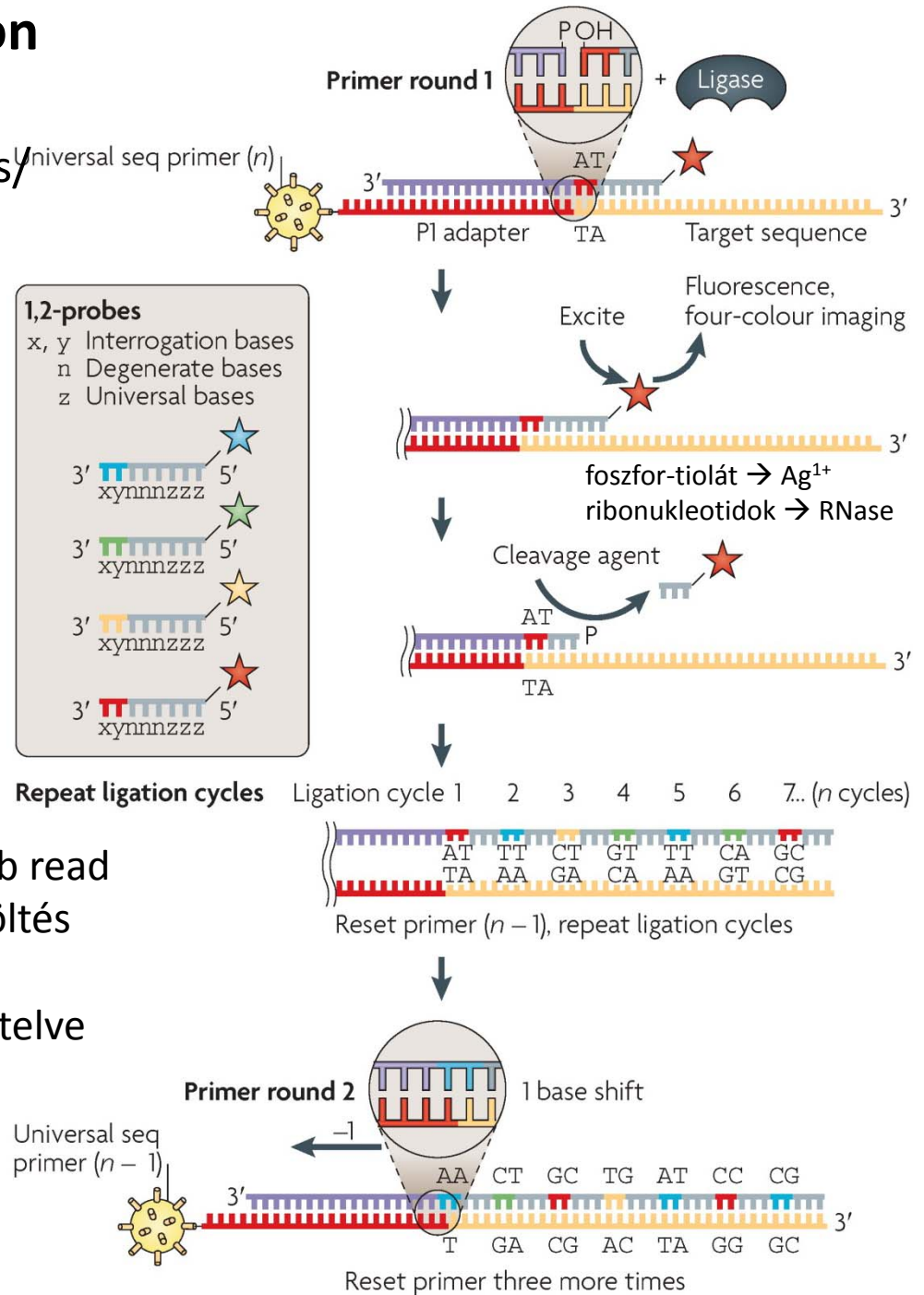
2nd base		Template sequence
A	A C G T	ATACAA GA
C	A C G T	CGCACT TC
G	A C G T	GCCTGG AG
T	A C G T	TATGTT CT

Alignment of colour-space reads to colour-space reference genome



100 millió-1 milliárd 35b read
1 futás, 5 primer ujrátöltés
→ 60 Gb
Minden nt kétszer tesztelve
→ 99.94% pontosság

Aluldetektálja a valós variánsokat.
Leggyakoribb hiba a bázis szubsztitúció.



Helicos BioSciences: HeliScope

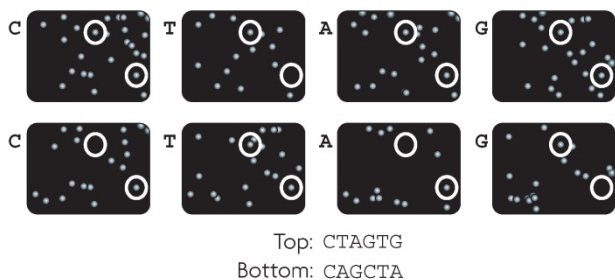
pioneer in Single Molecule Detection

- 1) DNS fragmentálás, adaptor ligálás
- 2) Immobilizálás ... felszínen:
 - a) univerzális primer egyedi molekulái
 - b) az egyedi DNS molekulák (*kétszeres leolvasás*)
- 3) Szekvenálás: „egy szín” – CRT

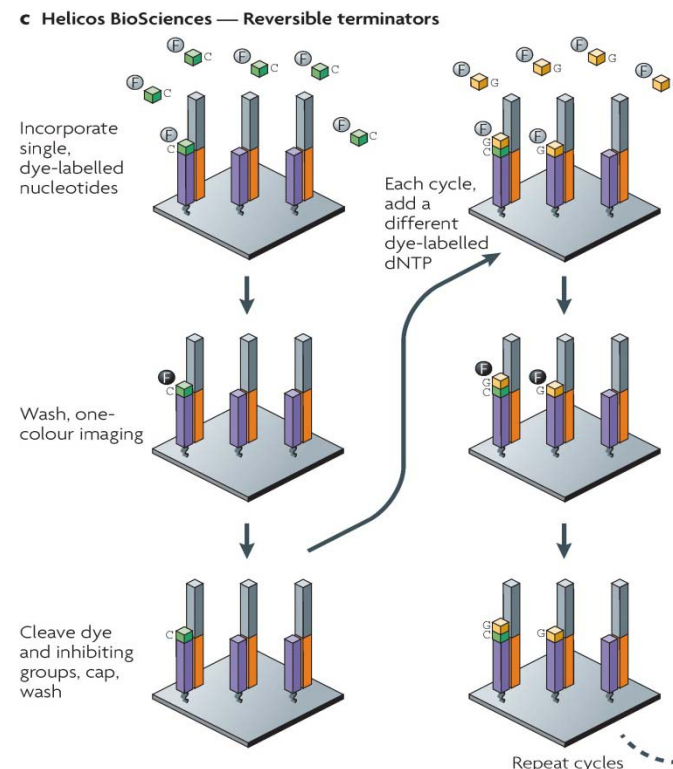
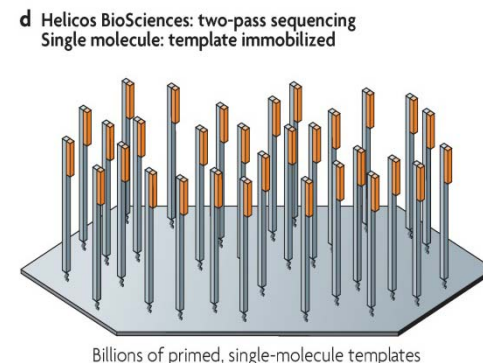
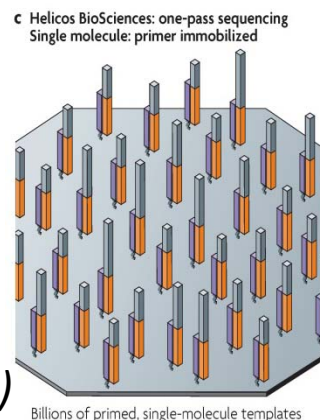
Virtuális terminátor: Cy5 fluorofór és gátló 2. nukleotid egy lépésben hasítható a beépült nt-ről

Leggyakoribb hiba: a homopolimer szakaszoknál deléció ← két nt beépülés egy ciklus alatt

- 4) Detektálás: egy csatornás TIRF



From a single HeliScope run using only 7 /50 channels,
 ~2.8 Gb of high-quality data in 8 days
 >25-base consensus reads with 0, 1 or 2 errors.
 >99% coverage of the *C. elegans* genome
 At regions >5-fold covered, the consensus accuracy was 99.999%

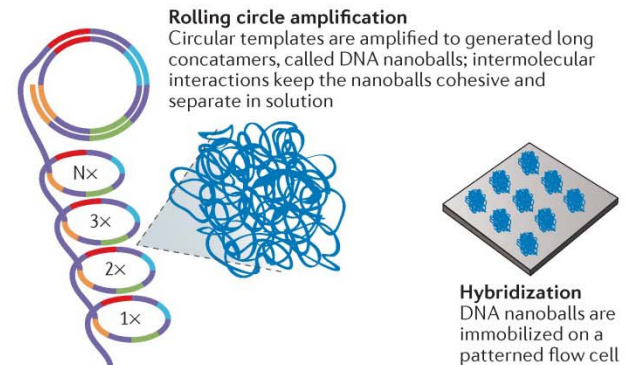
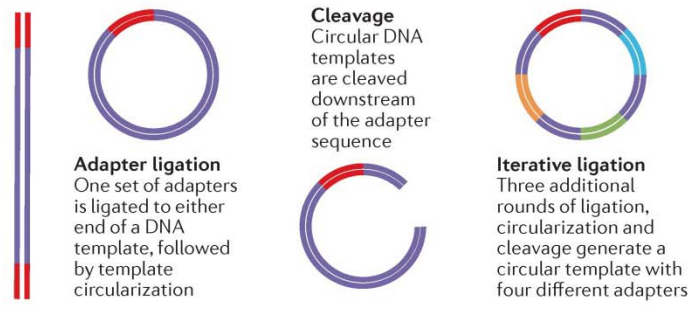


Stopped production → ...cont.
 2015: GenoCare instrument from
 Direct Genomics at BGI

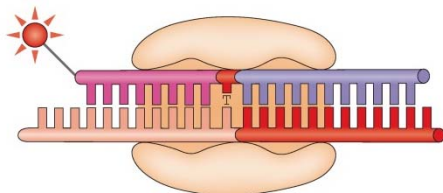
Complete Genomics – 2006 2013-tól BGI

- 1) Genomi DNS ~400 bp fragmensek
- 2) Rolling cycle amplification (RCA) → DNS „nanoballs”
- 3) Mintázott felületen elosztatva
- 4) Ismételt fluoreszcens hibridizációs próbák ligálása (SBL)...

d In-solution DNA nanoball generation (Complete Genomics (BGI))



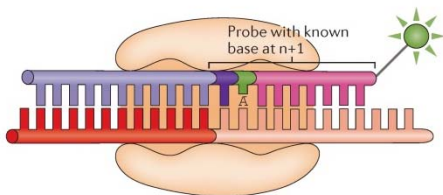
b Complete Genomics (BGI)



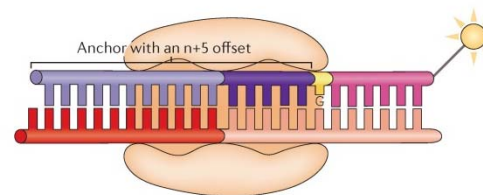
Single-base-encoded probes
A probe with a single known base and degenerate bases hybridizes to a template and is imaged



Reset
After each imaging step, both the probe and anchor are removed



Paired-end sequencing
Sequencing is performed for both the left and right sides of



Offset anchors
Subsequent rounds of hybridization and ligation use offset

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351.

Platformok:

Revelocity: mate pair seq: 300 b insert
2X28b reads, hiba arány 10^{-7}
BGISEQ-500: 200 Gb/futás,
rugalmasabb alkalmazhatóság, single cell seq.

A hosszú leolvasás egyértelmű előnyei:

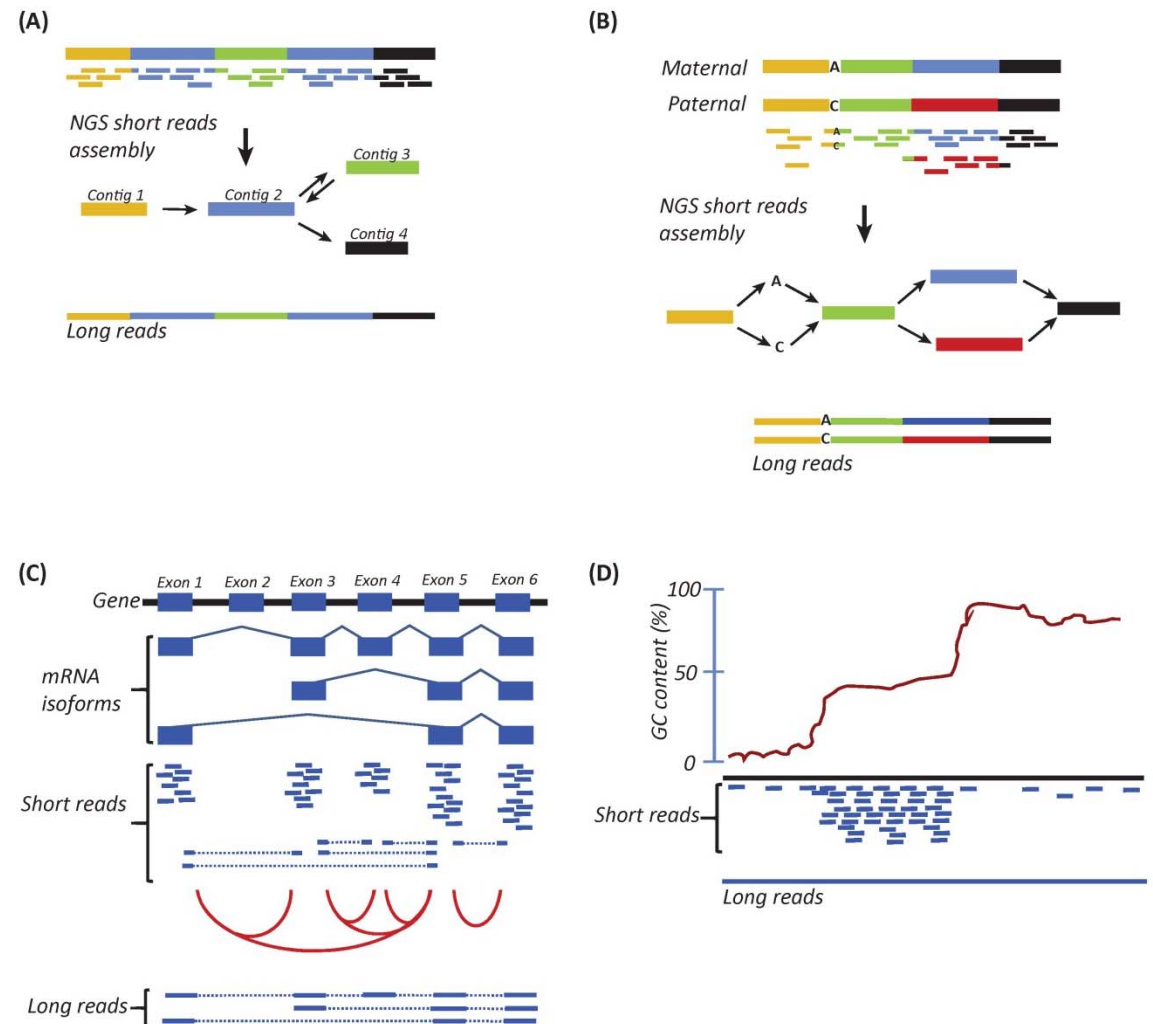
*de novo összeszerelés könnyebb és pontosabb repetitív régiókban is

*allélek közti különbség egyértelműbb (*genome phasing*)

*RNS izoformák elkülönítése
→ pontosabb expressziós adatok,
→ pontosabb génmodellek

*szélsőséges szekvencia környezetben a PCR okozta torzulások elkerülése

Comparison of the Performance of Next-Generation Sequencing (NGS) Short-Read and Long-Read Methods



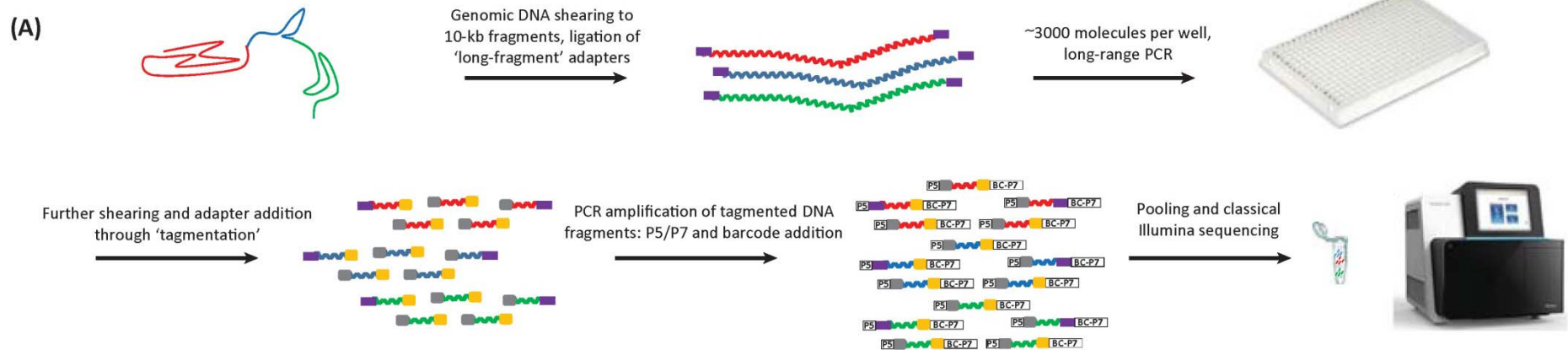
Synthetic long-read technológiák

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics*, 34(9), 666–681. <http://doi.org/10.1016/j.tig.2018.05.008>

BAC-by-BAC sequencing régebbi technológia

Átfedő bakterialis mesterséges kromoszóma (BAC) klónok végig a genomon
Ezek szekvenálása külön-külön

illumina SLR (ld. korábban)



10X Genomics, emulsion based system

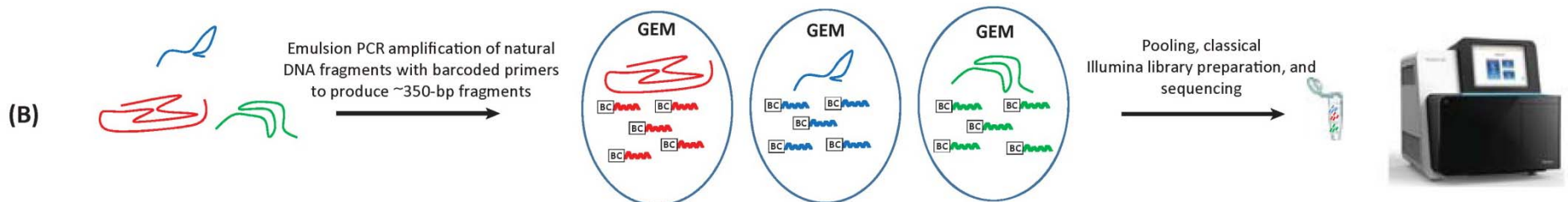
~1 ng kiindulási DNS, méretfüggetlen természetes fragmensek

GemCode 2015-ben:

akár 100 000 partíció és 737 000 különböző barcode

Chromium 2016-ban:

akár 1 millió partíció és 4 millió különböző barcode



Pacific Biosciences Single Molecule Real Time (SMRT) szekvenálás

- 1) DNS preparálás, fragmentálás
- 2) Polimeráz ($\phi 29$) egyedi molekulái immobilizálva **Zero-Mode Waveguide (ZMW)** platformon
- 3) Primerrel hibridizált DNS fragmensek egyedi molekulái kötődnek
- 4) Real-time nukleotidok (γ -foszfáton fluoreszcensen jelölt dNTP, bázison quencher) pM-nM

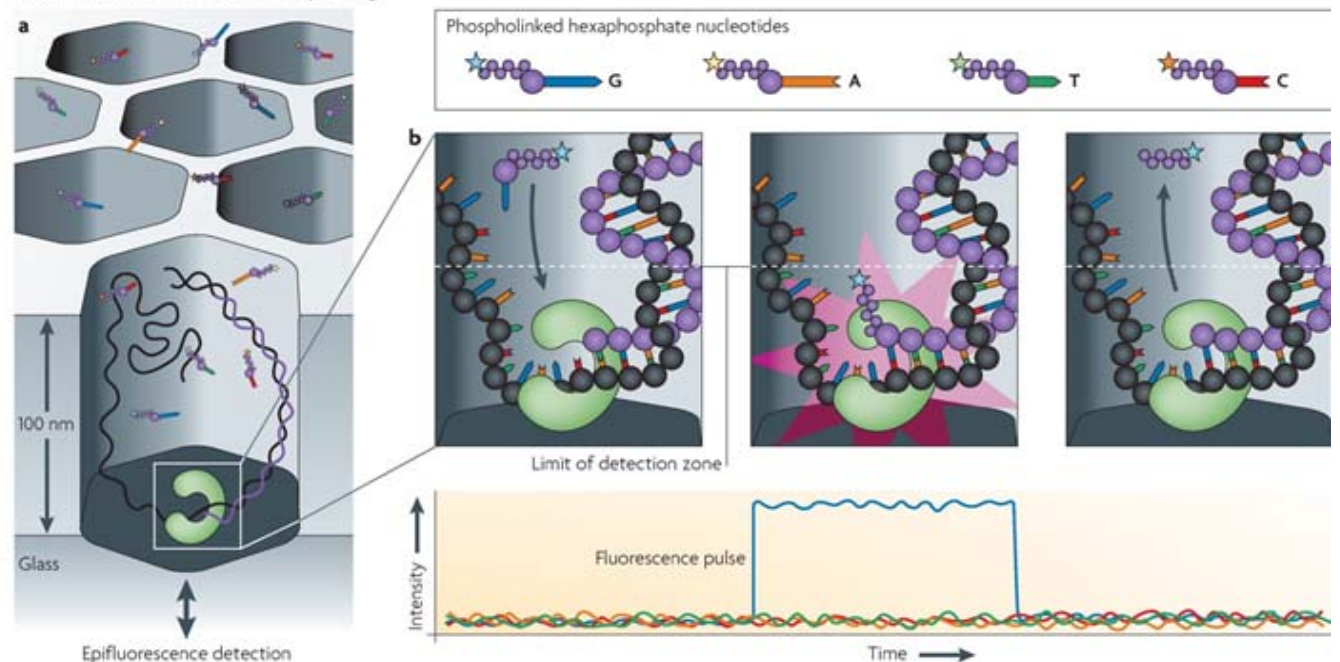
5) Detektálás:

epifluoreszcens mikroszkóp
4 szín, fluoreszcens
fluktuáció
Detektálási zóna a
polimeráz körül

Előnyei: SM, és RT, hosszú
read-ek, egyszerűbb kémia...

Eredendően nagy hiba arány
15X újraszekvenálás \rightarrow 99%
megbízhatóság

Pacific Biosciences — Real-time sequencing



2009: *E. coli* genom
38-szoros bázis lefedettség,
99.3% genom lefedettség
Pontosság >99.999%
Átlagos read: 964 bp

Pacific Biosciences Single Molecule Real Time (SMRT) szekvenálás

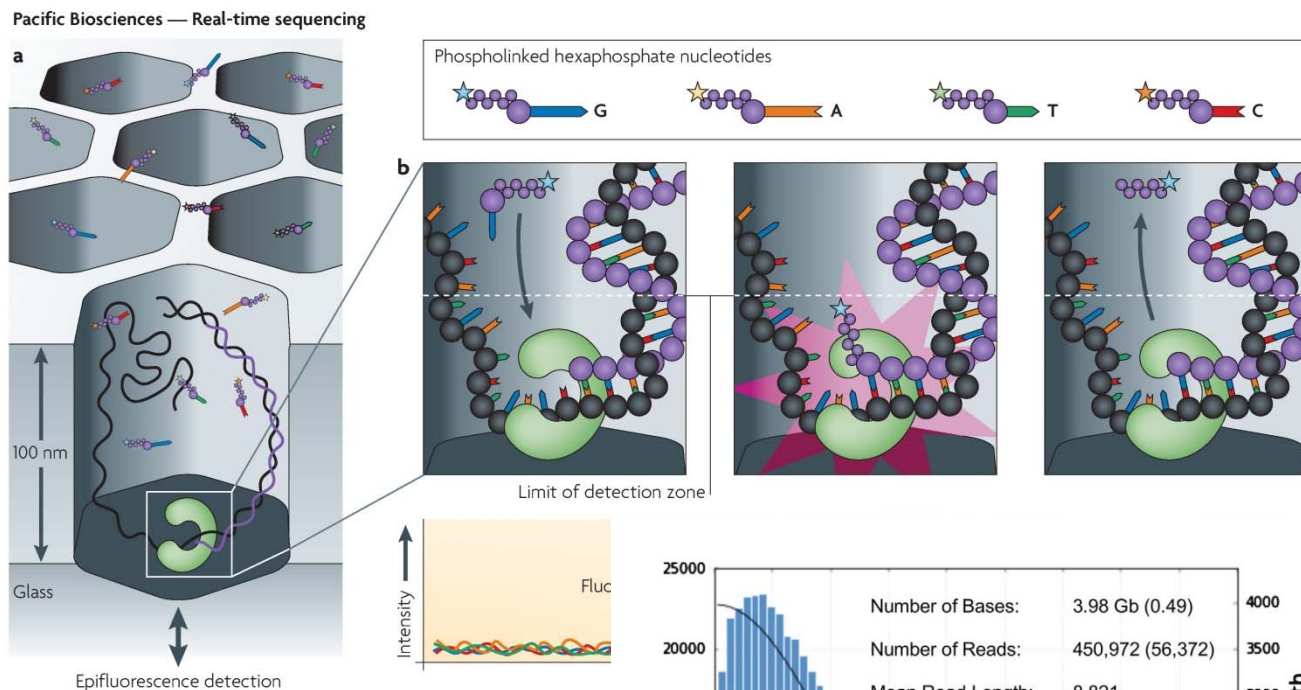
- 1) DNS preparálás, fragmentálás
- 2) Polimeráz ($\phi 29$) egyedi molekulái immobilizálva **Zero-Mode Waveguide (ZMW)** platformon
- 3) Primerrel hibridizált DNS fragmensek egyedi molekulái kötődnek
- 4) Real-time nukleotidok (γ -foszfáton fluoreszcensen jelölt dNTP, bázison quencher) pM-nM

5) Detektálás:

epifluoreszcens mikroszkóp
4 szín, fluoreszcens fluktuáció
Detektálási zóna a polimeráz körül

Előnyei: SM, és RT, hosszú read-ek, egyszerűbb kémia...

Eredendően nagy hiba arány
15X újraszekvenálás \rightarrow 99% megbízhatóság



2009: *E. coli* genom
38-szoros bázis lefedettség,
99.3% genom lefedettség
Pontosság >99.999%
Átlagos read: 964 bp

2014:
8 SMRT cella
P5 polimeráz
C3 kémia

PacBio platformok

PacBio RS II.

150 000 ZMW / SMRT cella
1 Gb / SMRT cella
1-16 cella / futás
0.5-6 h / SMRT cella

PacBio Sequel (2015)

(Roche-sal együtt fejlesztették klinikai alkalmazásokra is)

1 millió ZMW / SMRT cella
5-20 Gb / SMRT cella

Akár 30 kb átlagos read hosszúság

500 000 long SM read

>99% biztonság -- teljes genom szekvenálás

>99.999% biztonság -- RNS vagy amplicon szekvenálás

<1 nap

Sequel System II

hamarosan várható!!!

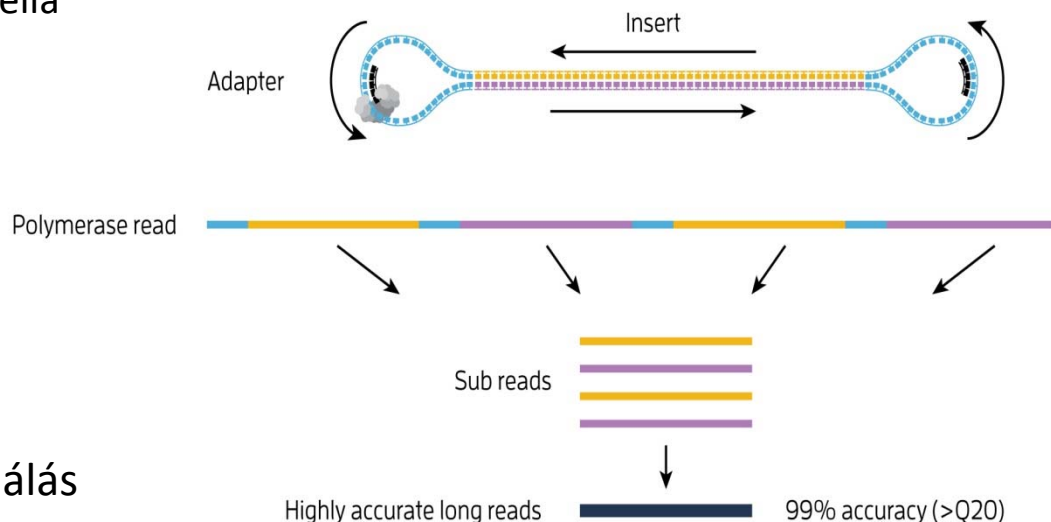
Epigenetikai módosítások közvetlen detektálása!!!

Pl. 6meA jellemzően hosszabb interpulse duration (IPD)

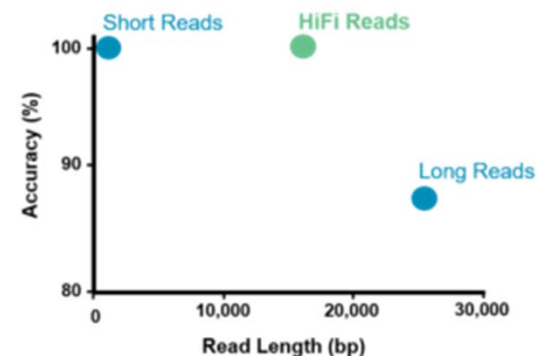
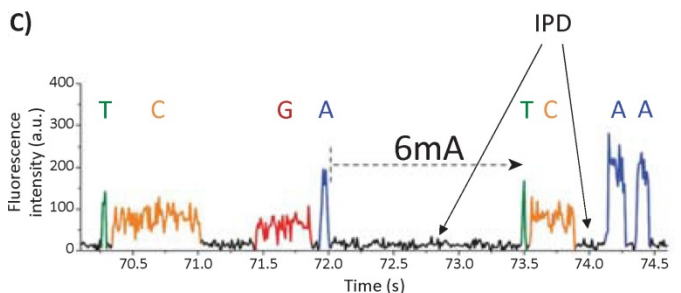
<https://www.pacb.com/smrt-science/smrt-sequencing/single-molecule-resolution/>

Tökéletesítések:

*újrászekvenálás (circular consensus sequencing (CCS))



*HiFi polimeráz



van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics*, 34(9), 666–681.

Nanopore szekvenálás

Oxford Nanopore Technologies (ONT)

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, 14(5), 265–279.

Teljesen újszerű elv:

Membránba ágyazott nanopóruson (porin fehérjék) keresztül, elektroforézis segítségével, „áthúzott” ssDNS vagy RNS bázisai különböző módon befolyásolják a póruson áthaladó áramerősséget

Fejlesztése:

The Nanopore Community

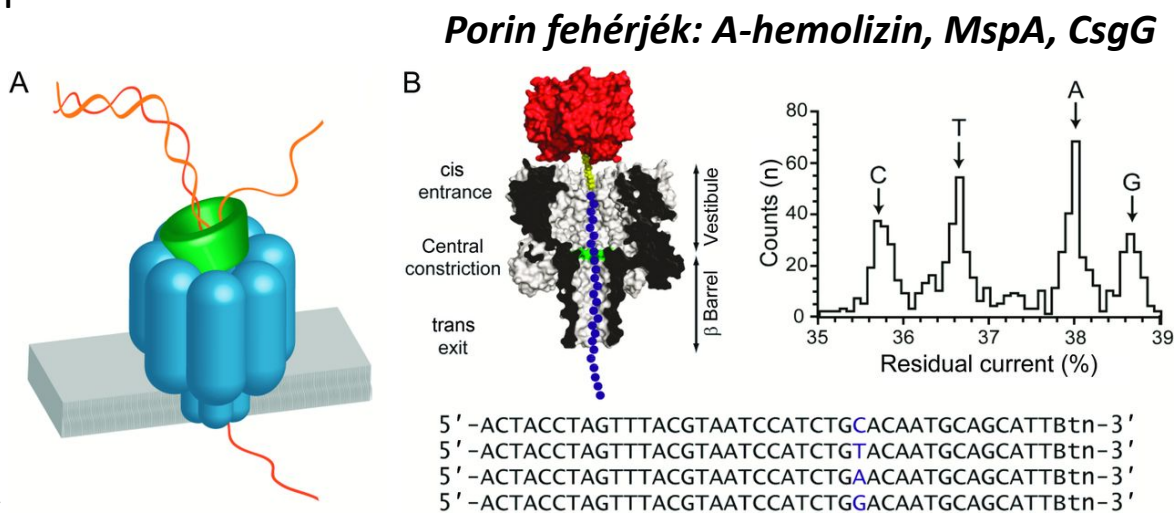
*MinION Analysis and Reference Consortium (MARC)

*MinION Early-access Program (MAP)

- 1) DNS izolálás, >30 kb fragmensek
 - 2) Könyvtár preparálás: vég javítás, két adaptor ligálása: vezető Y-adapter, hairpin adapter
Nincs fragmentálás, nincs amplifikáció!!! single-molecule és real-time
 - 3) A póruson való átjutás sebességét a kettős hélixet kinyitó motorfehérje szabályozza
 - 4) Valós idejű jel detektálás: ASIC mikrochip
- Bázis hívása: 5-6 bázisonként kapott jel profil elemzéséből HMM-segítségével

Előnyök: nagyon hosszú leolvasások (néhány 100 kb), kémiai módosítás nélkül, közvetlen leolvasás, akár RNS-t is!!!, gyors, kis eszköz igény szállítható, skálázható...

Hátrány: nagy hibaaarány...

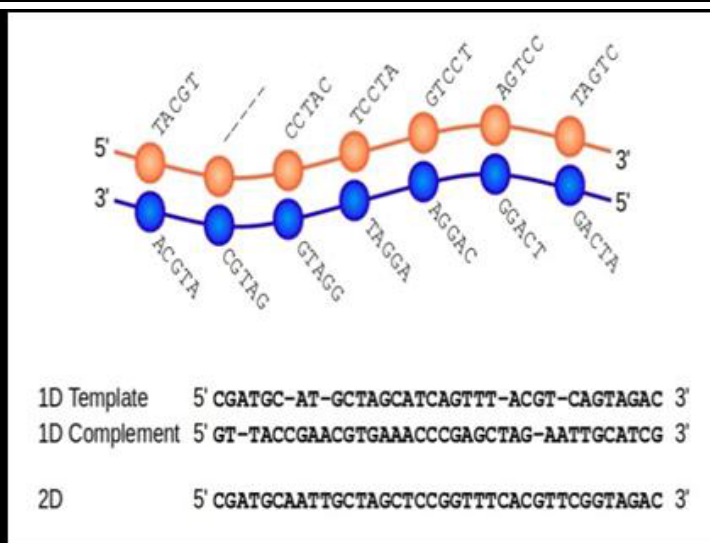
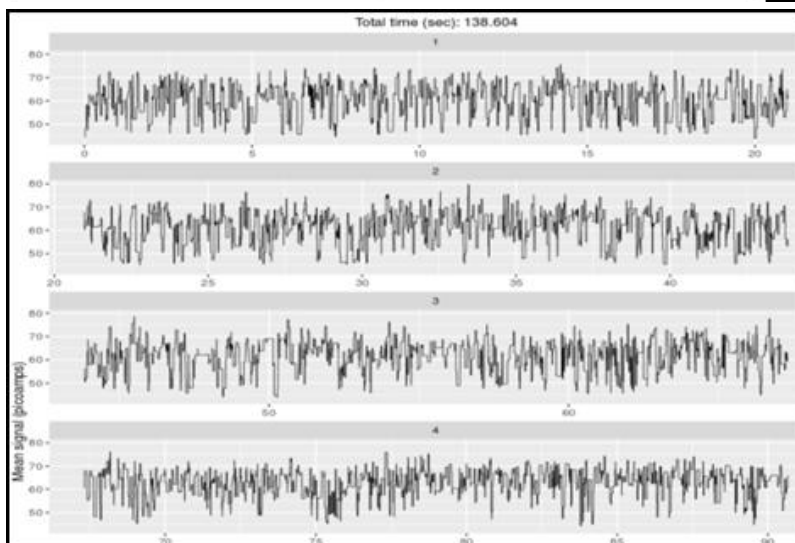
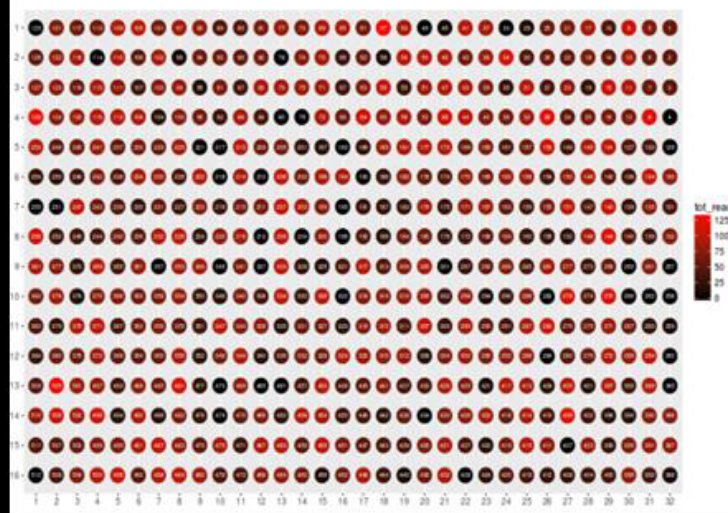
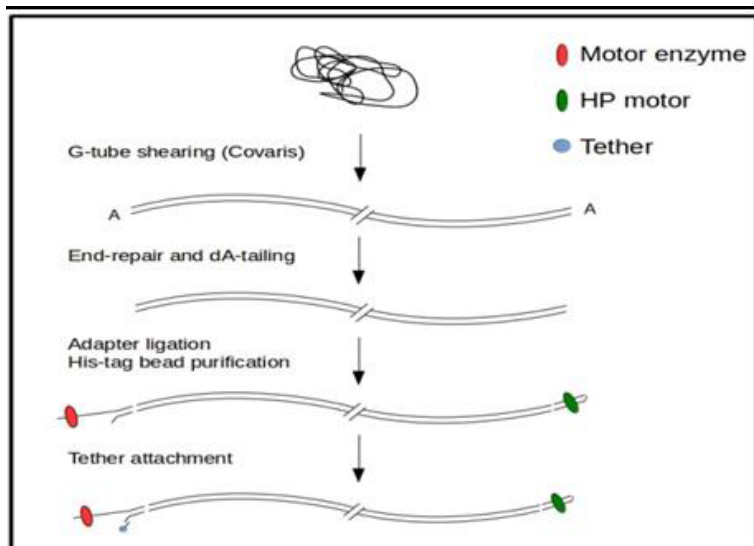


Minlon Oxford Nanopore, 2012 a zseb-szekvenátor...



Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, 14(5), 265–279.

512 nanopórus, mindegyik 4-szeresen kötve a microchip-hez



48h futás →
10-20 Gb,
Szimultán
adatelemzés

A 2D
leolvasás
növeli a
megbízható-
ságot

GridION

Multiple sequencing devices, one compute module
Use up to five MinION Flow Cells at a time
Benchtop processor capable of handling high data volumes in real time
Rapid, real-time applications such as *Read Until ...*

PromethION, 2015

High-throughput, high-sample number benchtop system
Modular: Up to 48 flow cells, each with up to 3,000 nanopore channels (total up to 144,000)
Flow cells may be run individually or concurrently
Same workflow as MinION at larger scale

Genia 2009-ben alakult, 2013-tól a **Roche** része szilárd nanopórus technológia fejlesztése

Előnyök: ellenállóbbak, kiszámíthatóbbak, könnyebben előállíthatóak

Szilárd fázisú membránok → túl vastagok (10-20 nm), 30-50 bázis egyszerre

Grafén nanopórus → túl vékony, a bázisnak csak 1/3-a van a csatornában egyszerre

Túl gyors a DNS áthaladása → külső felszínen immobilizált polimerázhoz kötött DNS lassúbb

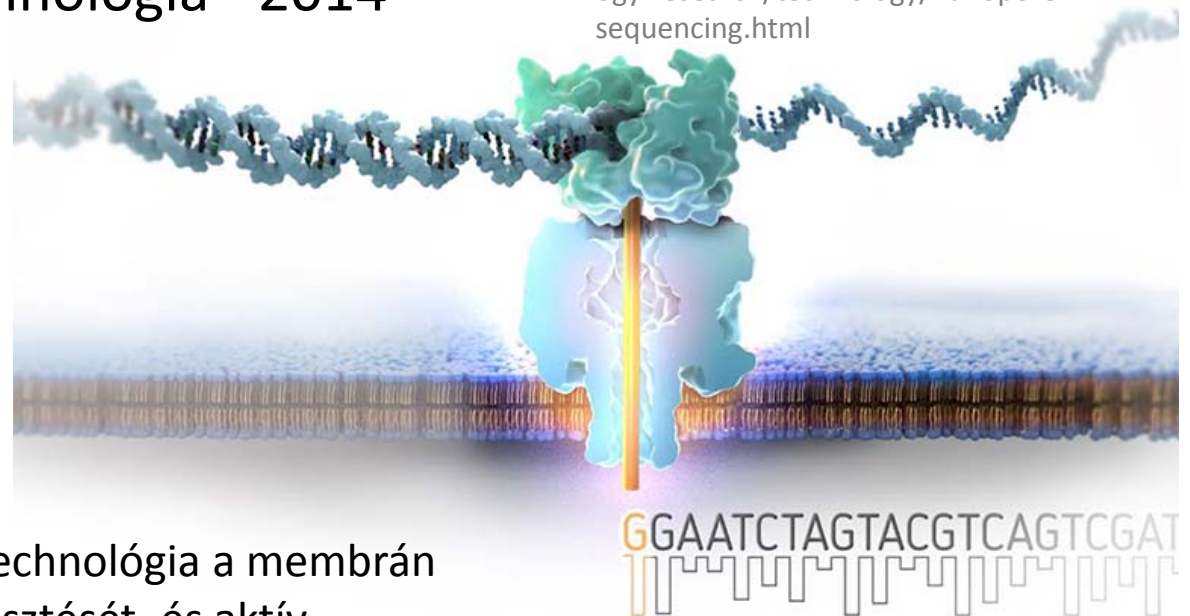
NanoTag sequencing technológia - 2014

- *Polimeráz immobilizálva a nanopórus közelében
 - *Real-time nukleotid: dNTP a γ -foszfáton jelölve egy nanoTag-gel
 - *A nukleotid beépüléskor lehasadó tag halad át a póruson, és ad jelet.
- Detektálás: mikrochip
(szemikonduktor integrált áramkör)

A beépített elektromos szenzor technológia a membrán összeszerelést, a nanopórus beillesztését, és aktív kontrollját is lehetővé teszi.

2015-ben ezzel a 100\$ / humán genom elérhetővé vált!!!

<https://sequencing.roche.com/en/technology-research/technology/nanopore-sequencing.html>



Egyéb genomikai kiegészítő technológiák:

Nanochannel genome mapping (optical mapping)

Irys System, BioNano Genomics

long >100 kb DNS,

szekvencia specifikus nicking endonukleáz reakció

in vitro DNS javítás fluoreszcens dNTP-vel

fluoreszcenciát detektál a nanocsatornán való áthaladáskor

→ nem nagyfelbontású szekvencia, hanem specifikus genomi helyek távolsága → optikai térkép, ami segíti a genom összeszerelést...

Chromosome Conformation Capture (Hi-C)

→ 3D szerveződésről információ:

1) kromoszóma kölcsönhatások,

2) contig-ok sorba rendezését és orientációját segíti

További cégek saját fejlesztésű szekvenátorai

Qiagen GeneReader

Agilent/Lasergen fejlesztés alatt: lightning terminátor

Jövőbeli lehetőségek: egészen új alapú technikák (pl.: MS, tunneling microscope, EM, XPS, FRET, Raman spektroszkópia, RNAP)

NGS adatok jellege, mérete – adatbázisok

NGS nyers adatok processzálása és elemzése
alkalmazási területtől függő – sok féle irány, sok
buktató, nagy számolási és tárhely kapacitás...

– Ligeti Balázs előadásai

NGS alkalmazási területei:

- 1) Teljes genom szekvenálás (**WGS**), teljes exom szekvenálás (**WES**), specifikus gén készlet szekvenálása (**targetted sequencing**) – **variáció analízis** („variation calling”)
 - **több adat**
 - **populáció genetika**
 - **klinikai jelentőség**
 - Single Nucleotide Polimorfism (SNP)
 - Copy Number Variation (CNV)
 - Structural Variations (SV)
- 2) **Metagenomika**: 1) baktérium flóra jellemzése..., 2) környezeti minták elemzése...
- 3) **RNA-seq** 1) Reverz transkripció (RT) és DNS szekvenálás vagy 2) direkt RNS szekvenálás
 - Expressziós szint – teljes transzkriptom analízis**
 - Riboszóma vagy poliszóma profil**
 - Reguláló RNS-ek jellemzése: miRNA, ncRNA,**
 - RNS szerkezet vizsgálat**
 - RNS-fehérje kölcsönhatás (CLIP-seq)**
- 4) Specifikus régiókra **dúsított minták analízise** / chromatin profiling – „**peak calling**”
 - ChIP-seq (transzkripciós faktor kötő helyek, hiszton módosítások...)
 - ATAC-seq
 - DNase-seq
 - DIP-seq (módosított bázisok) Bisulfite-Seq, MeDIP-Seq, excision-seq, dU-seq, U-DNA-seq
- 5) **3D kromatin szerkezet** leírása
 - Hi-C
 - ChIA-PET
 - gradient-seq to determine sonication resistant heterochromatin regions (srHCR)
- 6) **Antitest repertoire jellemzése** „clonotyping”
- 7) **Mitokondriális DNS vizsgálata**...

Teljes genom szekvenálás

HiSeq/Illumina, Ion Torrent/Thermo Fisher Sci.,
PromethION/Oxford Nanopore Technologies,
SMRT-seq PacBio

Humán referencia genom tökéletesítése – repetitív szakaszok

Csak **1.5% fehérje kódoló**

42% class I. transzpozabilis elem (**retrotranszpozonok maradványai**): LINE, SINE, LTR, ERV

2-3% class II transzpozabilis elem (**DNS transzpozonok**): α -szatelitok → higher order repeat (HOR)

>160 vakfolt az eukromatinban; a centromérák, telomerek repeat-jei is problémások

SMRT seq → 1Mb új szekvencia

→ több 10 000 SV megoldása

→ short tandem repeats (STPs) (pl. 2.25 kb tandem repeat of CGG)

Nanopore seq → kevesebb lyukat tud kitölteni, mint az SMRT

→ extrém hosszú leolvasás is: pl. Y kromoszóma centromere

60-szoros lefedettség, teljes hosszú leolvasás néhány 100 kb α -satelite repeat

Szerkezeti variációk – kromoszóma átrendeződések, CNV

csak teljes genom szekvenálással

short readok sokszor nem elégségesek

klinikai jelentőség is

Populáció genetika

Új model organizmusok teljes genomja – de novo assembly

összehasonlító genetika, evolúció

Teljes exom szekvenálás (WES)

1-2%-a a teljes genomnak

85%-a az ismert betegségokozó mutációknak

WES → mélyebb szekvenálás, több minta, kevesebb forrás → klinikai jelentőség

Mutáció/SNP analízis – „variation calling”

→ Populáció genomika

→ betegségek genetikai hátterének felderítése

→ tumorképződés és terápia

→ „driver mutations”

→ „drugable mutations” → precíziós, személyre szabott terápiák

→ „inherent tumor heterogeneity” (ITH)

→ vérplazmában cell-free DNA, circulating genomic subclones (cGSs)

A célzott szekvenálás (targeted seq) egyik fajtája:
az exonokat tartalmazó régiókat fel kell dúsítani...

Specifikus gén készlet szekvenálása

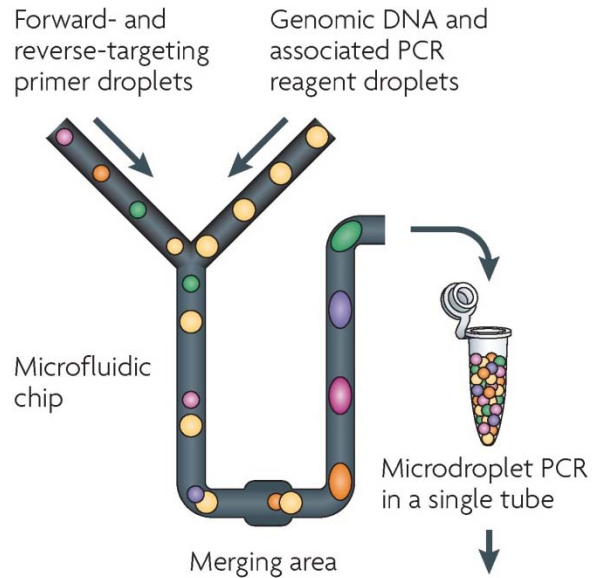
Targeted sequencing

*olcsóbb, gyorsabb, könnyebben kezelhető adat

*több minta

*„mélyebb”, nagyobb többszörös lefedettség

a Microdroplet PCR



RainDance Technologies,

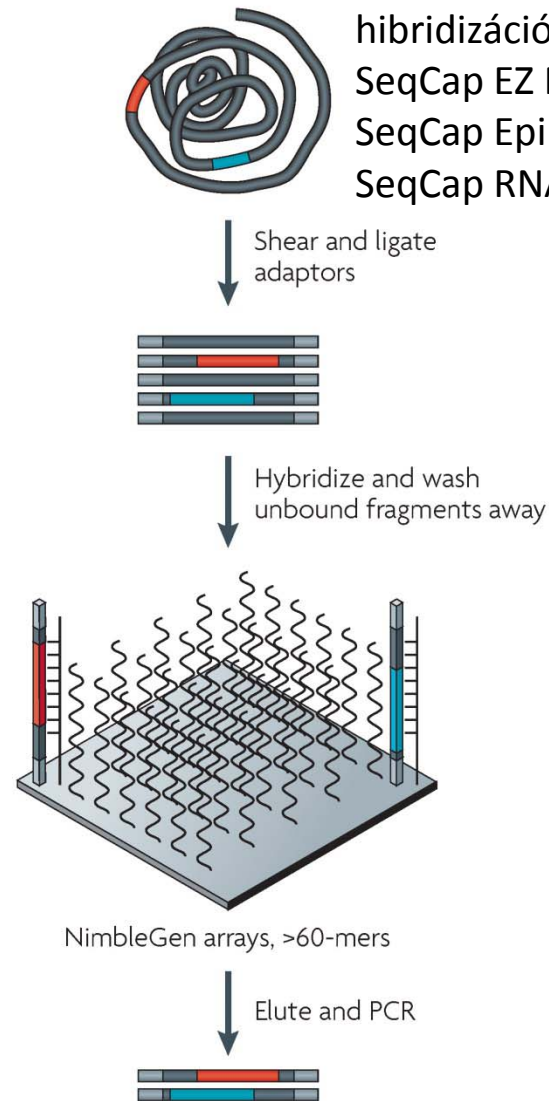
microdroplet PCR

3,976 különböző DNS szakasz szimultán felszaporítása

Ion AmpliSeq NGS Panels for Targeted Sequencing

Primer panels, Thermo Fisher Scientific

b Solid-phase capture



Roche/ NimbleGen SeqCap EZ Libraries:

hibridizációs oligonukleotid próbák → dúsítás

SeqCap EZ MedExome Target Enrichment Kit

SeqCap Epi System (metilációs targetek)

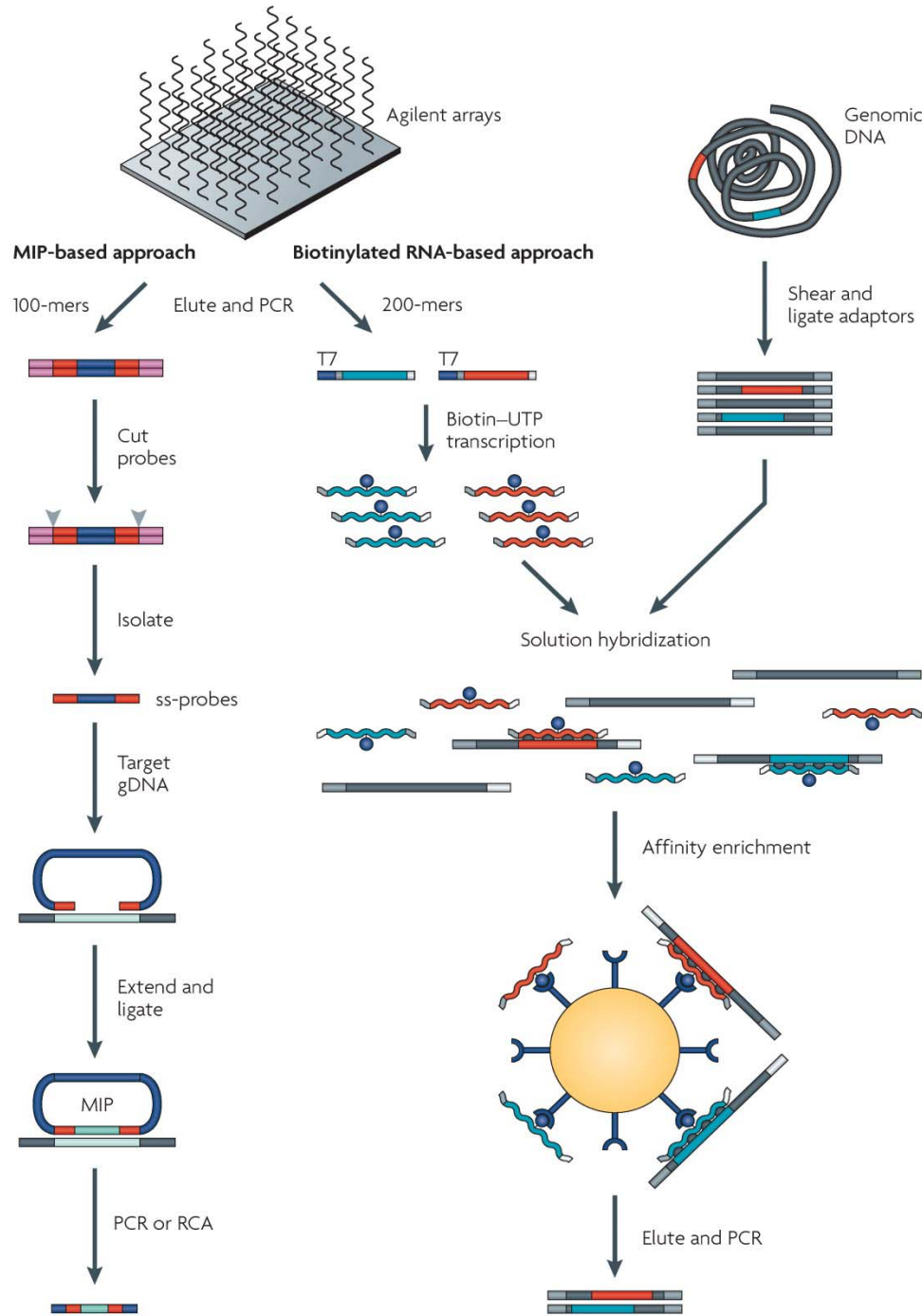
SeqCap RNA Enrichment Kits

Tewhey, R. et al. Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nature Biotech.* 27, 1025–1031 (2009).

Figures from: Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1), 31–46. <http://doi.org/10.1038/nrg2626>

c Solution-phase capture

molecular inversion probes (MIPs)



biotinylated RNA capture sequences

RNS szekvenálás

- 1) Reverz transzkripció és DNS szekvenálás: single read, szál-specifikus protokoll
- 2) Direkt RNS szekvenálás (nanopore seq., ONT)

Előnyök a microarray-hez képest:

- *Nem szorítkozik csak az ismert genomi szekvenciákra
- *Alacsony háttér zaj (referencia genomhoz rendezett szekvenciák)
- *kvantitatívabb

Alkalmazások:

- *transcriptional profiling (milyen szövetben, milyen mértékben, melyik izoforma expresszálódik)
- *SNP azonosítás
- *RNS editálás
- *differential gene expression analysis (pl. kezelés hatására megváltozott exp. szint)
- *új small, micro, és egyéb non-coding RNS-ek azonosítása

A teljes transzkriptom minden RNS-t tartalmaz:

mRNS, rRNS, tRNS, ncRNS, smallRNS, miRNS

Reverz transzkripció: univerzális RNS-tag-en keresztül

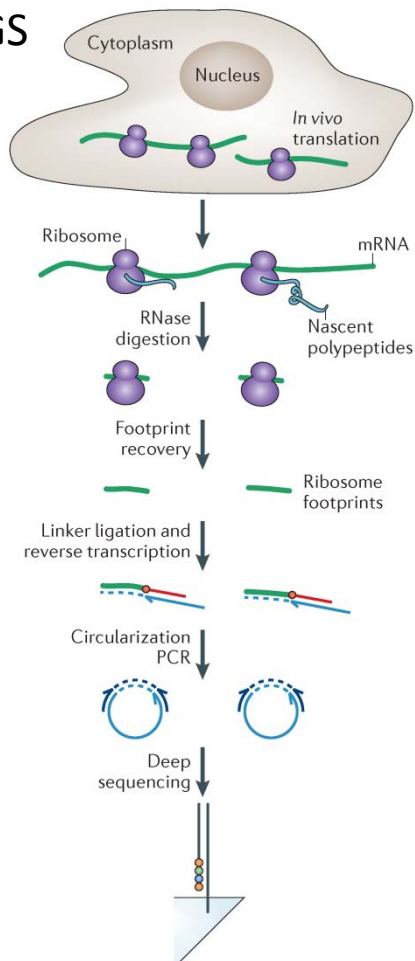
Régebben: random hexamer primerrel; oligodT-vel csak a polyA végű érett mRNS-ek

Bioinformatikai eszközök:

Sailfish, RSEM és BitSeq - expressziós szintek kvantifikálására
MISO - különböző izoformák expressziójának kvantifikálására

Riboszóma vagy poliszóma profil

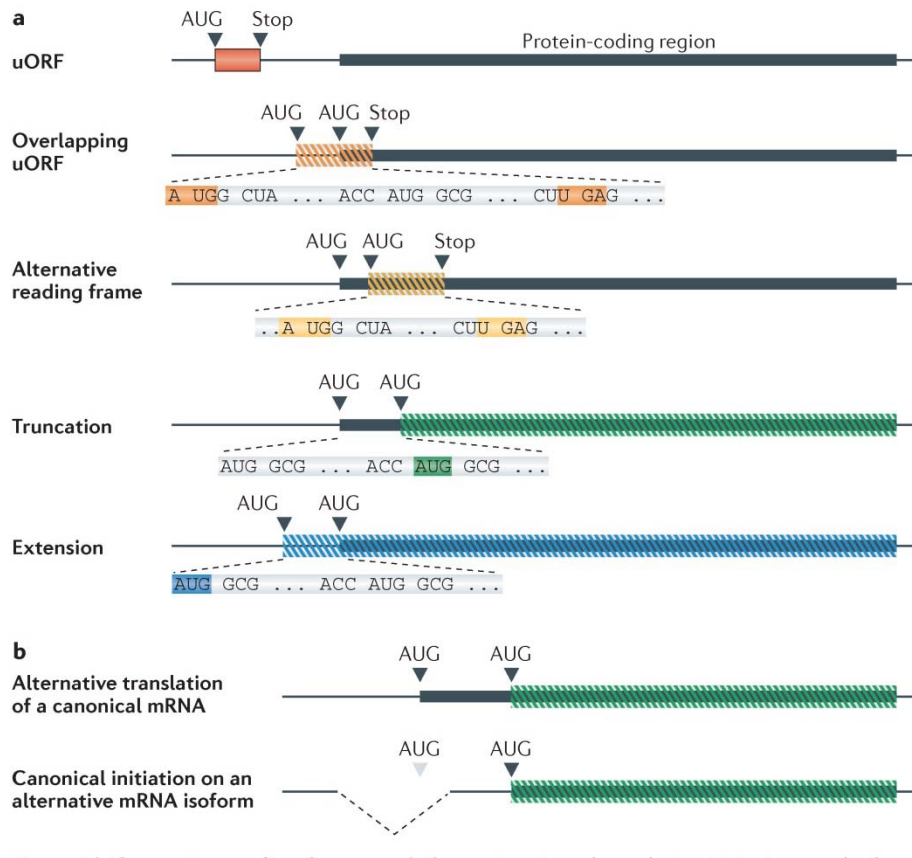
Riboszóma-RNS izolálás
 Fixálás, RNase emésztés
 Védett szakaszok visszanyerése linker ligálás
 Reverz transzkripció
 NGS



Előzőleg RNA footprinting

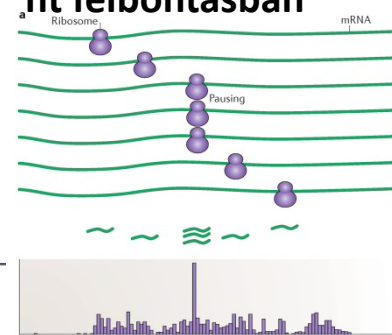
- problémás azonosítás
- poliszómák lenyomata nem egyértelmű
- kis átírási képesség

Új tudás a transzláció szabályozásról:



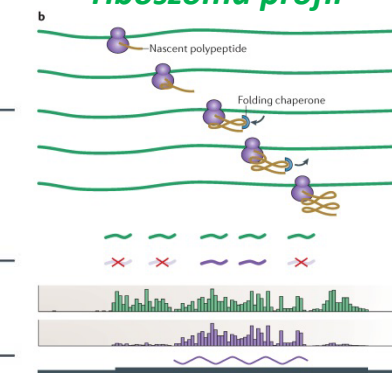
Ingolia, N. T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews Genetics*, 15(3), 205–213. <http://doi.org/10.1038/nrg3645>

A transzláció sebessége nt felbontásban



A fehérje folding mechanizmusa:

chaperonokra dúsított minta vs. eredeti riboszóma profil

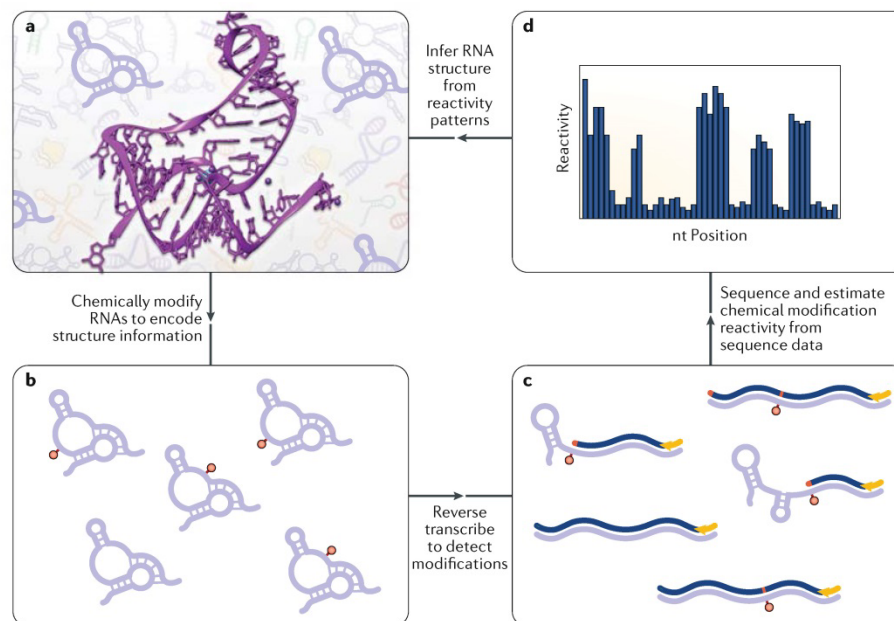


Nagy áteresztőképességű, nagyfelbontású RNS szerkezet vizsgálat

Szerkezeti RNS próbák (enzimatis vagy kémiai) és HTS kombinálása

Közös elv:

- 1) RNS folding
- 2) Szerkezet függő jelölés szerk. spec. próbakkal
- 3) Reverz transzkripció: a módosításnál megáll vagy mutáció épül be
- 4) Szekvenálás, leolvasások összerendezése -> akár egész transzkriptom analízise
- 5) Reaktivitás számolása minden nt-ra
- 6) 3D szerkezet modellezés



Strobel E.J. et al 2018 NatRevGenet

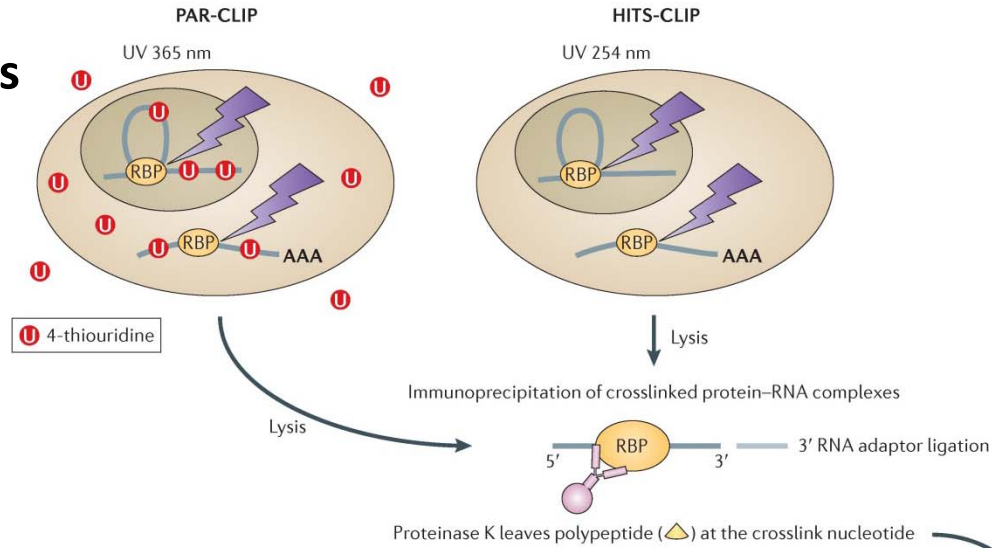
<https://doi.org/10.1038/s41576-018-0034-x>

Szerk. spec. próba	Detektálás elve	technikák
<i>RNáz szenzitivitás, ds- vagy ssRNS spec.</i>	RT mapping of fragment ends, read coverage	<i>PARS, FragSeq, dsRNA-seq</i>
<i>Nukleotid felbontású kémiai RNS próbák</i>	RT-stop, RT-mutate	<i>SHAPE-seq, DMS-seq, Mod-seq, CIRS-seq, Structure-seq, ChemModSeq, MAP-seq, SHAPE-MaP, RING-MaP, icSHAPE, MOCHA-seq, SHAPES, DMS-MaPseq</i>
<i>Ligálás alapú technikák</i>	mapping of ligated junctions	<i>RNA proximity ligation, LIGR-seq, PARIS, SPLASH</i>

RNS-fehérje kölcsönhatás

CLIP-seq

(Cross-linked immunoprecipitation)



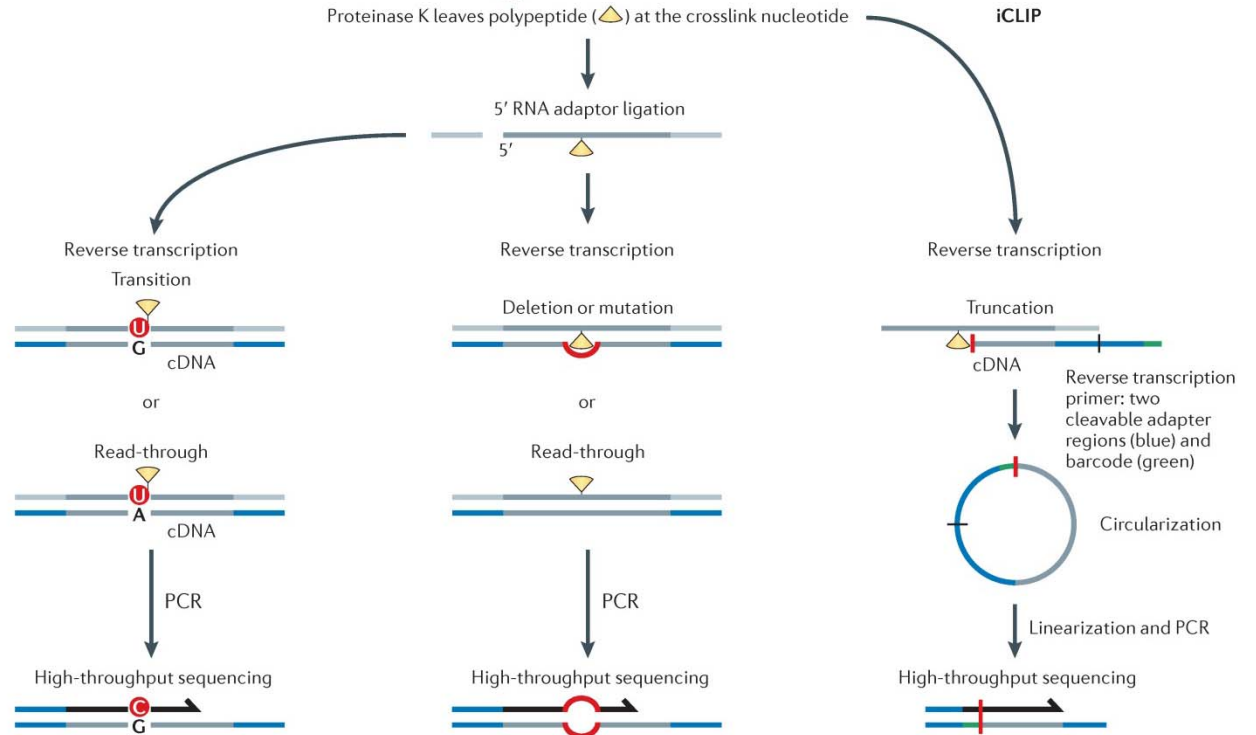
HITS-CLIP (CLIP-seq)

PAR-CLIP

Fotoaktiválható ribonukleozid:
4-tiouridin

iCLIP

Individual nucleotide
resolution CLIP



König, J., Zarnack, K., Luscombe, N. M., & Ule, J. (2012). Protein-RNA interactions: new genomic technologies and perspectives. *Nature Reviews Genetics*, 13(2), 77–83. <http://doi.org/10.1038/nrg3141>

Specifikus régiókra dúsított genomi DNS minták analízise

chromatin profiling

A genomi DNS szerkezete:

- 1) **Konstitutív heterokromatin:** kompakt, nincs transzkripció – pl. telomerek, (peri)centromérák – kromatin szerveződés a sejtmagban... a laminhálóhoz van kihorgonyozva
- 2) **Fakultatív heterokromatin:** sejttípus és –állapot függő kompakt kromatin
- 3) **Eukromatin:** a sejtmag belsejében lévő laza szerkezetű kromatin, transzkripció lehetséges:
 - 3A) represszált eukromatin (silencer vagy represszááló TF)
 - 3B) aktív eukromatin (aktiváló TF, coaktivátor, RNA pol)

Ezt a sokrétű szerkezetet lehet feltérképezni:

- Specifikus dúsításhoz kapcsolt NGS-sel - lefedettség: IP vs. Control – peak calling
- DNase szenzitivitás jellemzéssel
- Térben közeli régiók keresztkötéseit követő ligálással és NGS-sel – SV szerű események detektálása

Adatbázisok:

Cistrome, REMAP

ENCODE projekt

Encyclopedia of DNA Elements

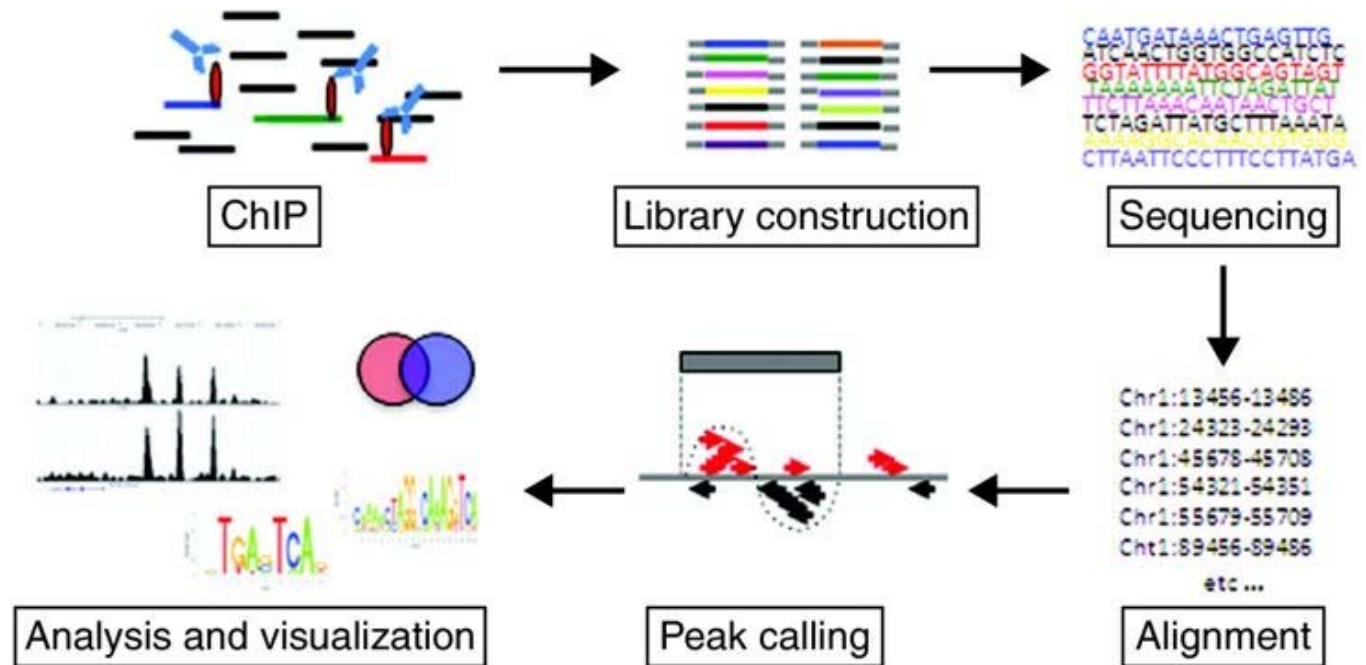
<https://www.encodeproject.org>

ChIP-seq

- 1) Kromatin izolálás, reverzibilis keresztkötés
- 2) Fragmentáció: szonikálás vagy enzimátikus
- 3) Immunoprecipitáció (IP) a kiválasztott kromatin kötő fehérjére (**transzkripciós faktorok, hiszton módosítások, RNS polimeráz, egyéb kromatin reguláló fehérje**), elúció proteináz K-val
- 4) NGS könyvtár generálás az IP előtti kontroll mintából és az IP-t mintából, szekvenálás

Lehetséges output-ok:
*konszenzus szekvencia a kötőhelyre (specifikus TF-ok esetén)

*profil átlapolása egyéb reguláló elemekkel, specifikus genomi régiókkal -> funkcionális következtetések



Hozáférhető DNS régiók feltérképezése eukromatin

ATAC-seq

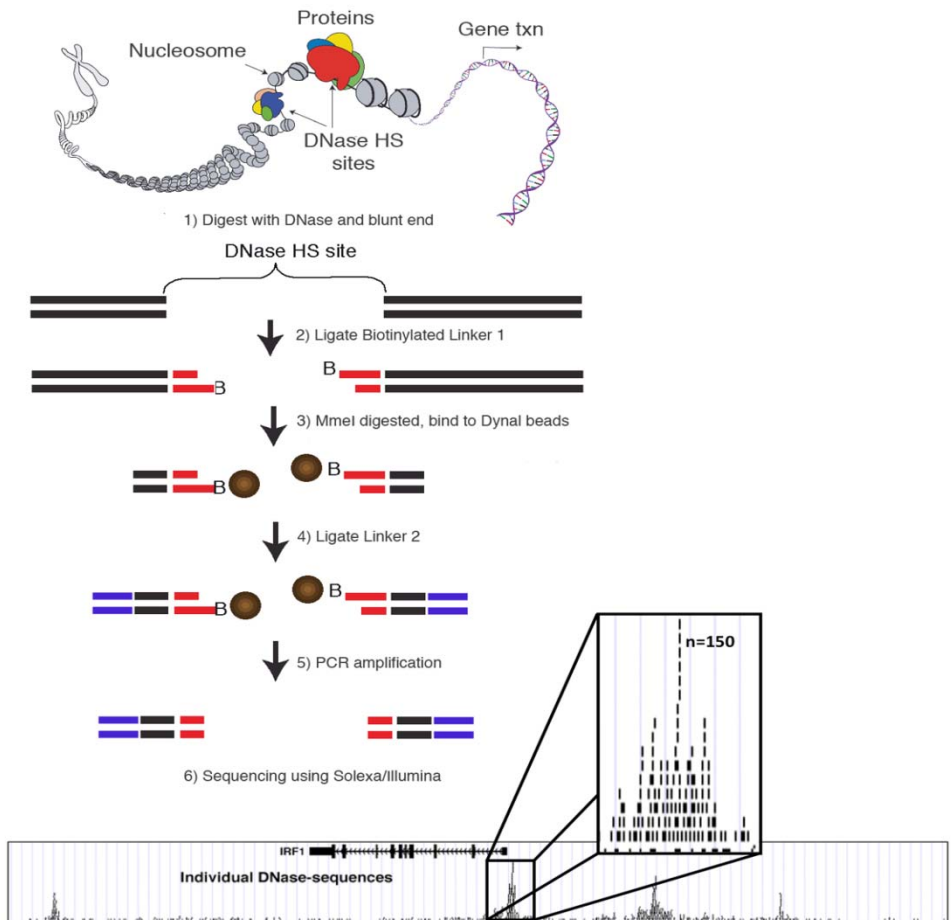
Hiperaktív transzpozáz → rövid DNS szakaszok átírása a nem védett helyekről
→ NGS

DNase-seq

Izolált sejtmagban a kromatin limitált emésztése DNase I enzimmal

- a DNase hiperszenzitív régiók degradálódnak
- A keletkező végek biotin-tag-gel jelölése, sztreptavidin gyöngyön immobilizálás
- Restriktációs emésztés (MmeI)
- Linker ligálás
- NGS

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).



Song, L., & Crawford, G. E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor Protocols*, 2010(2), pdb.prot5384.

3D kromatin szerkezet leírása:

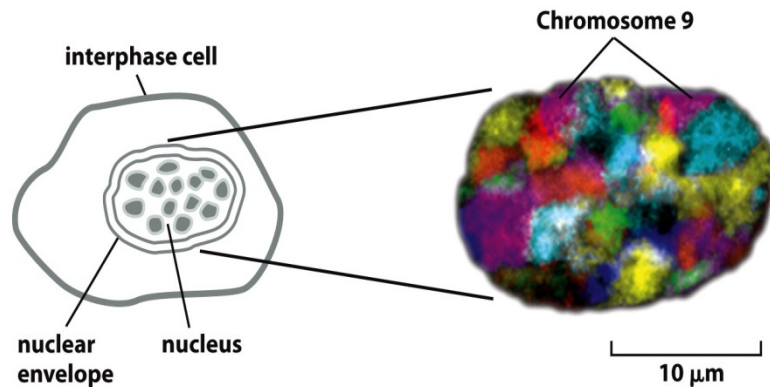
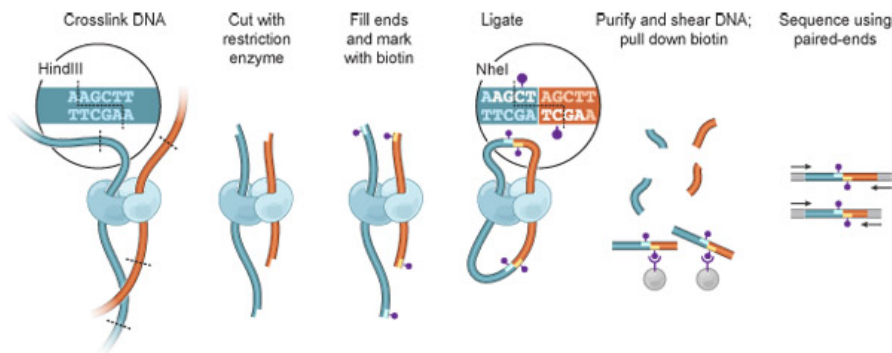


Figure 5-18 *Essential Cell Biology* (© Garland Science 2010)

Távoli enhancerek, silencerek
 Kromoszóma kölcsönhatások
 Kromatin átrendeződés: öregedés, szenescencia, lamin mutációk okozta szindrómák...

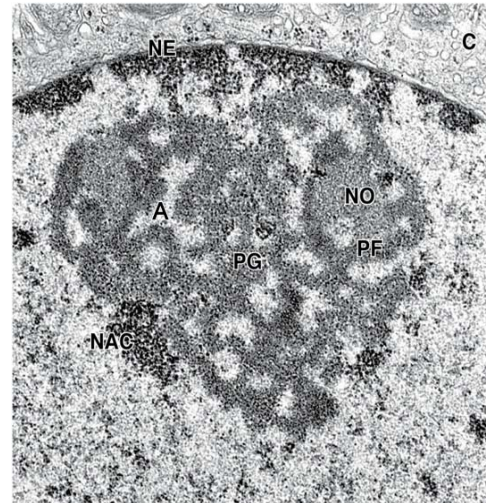
Módszerek:

Hi-C



van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., ... Lander, E. S. (2010). Hi-C: a method to study the three-dimensional architecture of genomes. *Journal of Visualized Experiments : JoVE*, (39).

A nukleólusz szerkezete (EM)



Cell-nucleus-mitochondrion-peroxisome_2018-VV
<http://semmelweis.hu>

ChIA-PETS Chromatin Interaction Analysis by Paired-End Tag Sequencing

- *Kromatin izolálás, keresztkötés
- *DNS emésztés az elérhető helyeken, végjavítás
- *Biotinilált linker ligálása a szabad végekhez
- *proximity ligation
- *Biotin pull down
- *NGS

Fullwood & Yijun, (2009). ChIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem.* 107(1); 30–39.

gradient-seq to determine sonication resistant heterochromatin regions (srHCR)

Becker, J. S., McCarthy, R. L., Sidoli, S., Donahue, G., Kaeding, K. E., He, Z., ... Zaret, K. S. (2017). Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. *Molecular Cell*, 68(6), 1023–1037.e15

A legfontosabb tudnivalók:

NGS: a közös lépések

Konkrét technológiák: Illumina, piroszekvenálás, ion torrent, SOLiD, BGI Nanoballs sequencing, PacBio SMRT, ONT nanopore seq,

Alkalmazások: WGS, WES, RNA-seq, ribosome profiling, RNS szerkezet, CHIP-seq, ATAC-seq, DNase-seq, 3D kromatin szerkezet

Köszönöm a figyelmet! 😊

