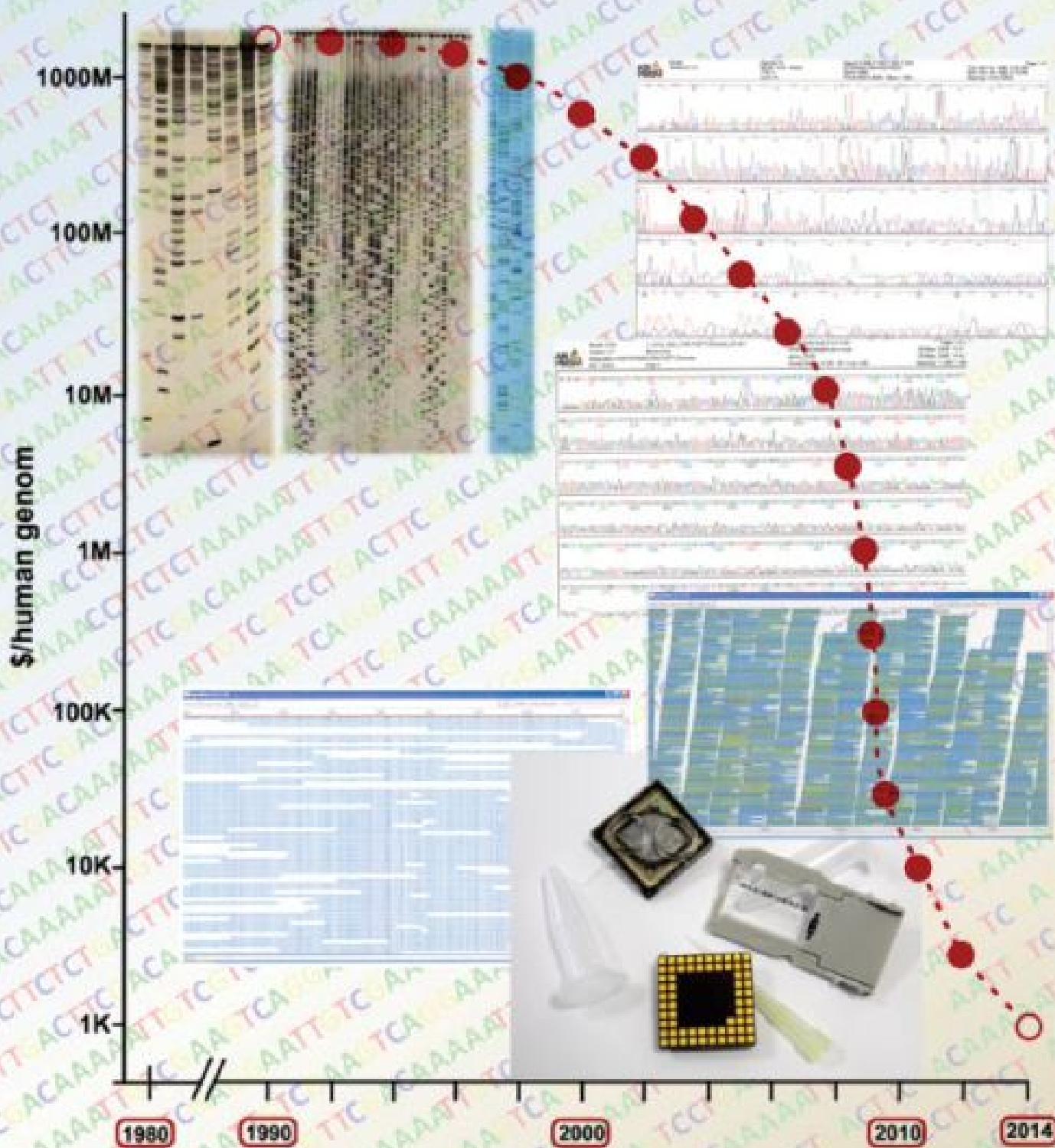


BIOKÉMIA

A Magyar Biokémiai Egyesület internetes folyóirata

XXXVIII. évfolyam 1. szám

2014. március



BIOKÉMIA

A Magyar Biokémiai Egyesület internetes folyóirata

Szerkesztőbizottság:

Bősze Szilvia, Erdődi Ferenc, Ifj. Gallyas Ferenc, Geiszt Miklós,
Kiricsi Mónika (titkár), Maksay Gábor, Nyitray László, Sarkadi Balázs,
Székács András, Szondy Zsuzsa

Főszerkesztő:

Szűcs Mária

szucs.maria@brc.mta.hu

Technikai szerkesztő:

Bérdi Péter

berdipeter@gmail.com

XXXVIII. ÉVFOLYAM 1. SZÁM

2014. március

TARTALOMJEGYZÉK

Címlapkép: Egy humán genomszekvenálás költségének változása 1990 és 2014 között (lásd Boros Imre írását 21. oldal)

AKIKRE BÜSZKÉK VAGYUNK

Kitüntetések, díjak 3.
Ifj. Gallyas Ferenc „Distinguished service award in cardiovascular science, medicine and surgery” díjat kapott 4.

REVIEW

Mones Letif, Fuxreiter Mónika: Fehérjeszerkezetektől a gyógyszertervezésig; gondolatok a 2013-as kémiai Nobel-díj kapcsán 8.
Boros Imre: Új DNS szekvenálási módszerek I. 20.

TUDOMÁNYOS KÖZLEMÉNY

Szakács Dávid és Beinrohr László: A XII-es faktor újrafelfedezése gyulladásközvetítőjeként: élettani szerepe és egy új kutatási projekt előzetes eredményei 32.

A 2013. ÉVBEN MEGJELENT KIEMELKEDŐ KÖZLEMÉNYEK LISTÁJA 44.

NAGY ELŐDEINK

Száz éve született Straub F. Brunó - Venetianer Pál, Patthy László és Mandl József írásai 51.
Elhunyt Bajusz Sándor - Medzihradzsky Kálmán 62.
Kucsman Árpád emléktábla avatás - Medzihradzskyné Schweiger Hedvig 65.

KONFERENCIA HÍREK

FEBS-EMBO jubileumi konferencia, Párizs, 2014 67.
Az MBKE 2014. évi vándorgyűlése, Debrecen 68.

FELHÍVÁSOK

Straub Örökség Alapítvány 69.
Bio-Science Kft. pályázat 70.

SZERKESZTŐI ROVAT

Új szerkesztőbizottsági tagok bemutatkozása 71.
Fórum újraindítása 72.

TUDOMÁNY ÉS MŰVÉSZET

Nyitrai Miklós: Mesés tájak, emberek – Afrika 73.

Kiadja a Magyar Biokémiai Egyesület

4012 Debrecen, Pf. 6. | <http://www.mbkegy.hu>

Felelős kiadó: Dr. Fésűs László | Az engedély száma: III/SZI/397/1977

HU ISSN 2060 8152 (Online) | HU ISSN 0133-8455 (Nyomtatott)

ÚJ DNS SZEKVENÁLÁSI MÓDSZEREK I.

Boros Imre

**SZTE Biokémiai és Molekuláris Biológiai Tanszék; SZBK Biokémiai Intézet
Szeged**

borosi@bio.u-szeged.hu

Gyakran elhangzó megállapítás, hogy valamely tudományterület vagy technológia forradalmi gyorsasággal fejlődik. Legtöbbször ez jóindulatú túlzás, a nukleinsav elsődleges szerkezetéről szóló ismeretek bővülésének jellemzésékor azonban bizonyosan nem az. DNS szekvenálási módszerek új nemzedékei jelennek meg és rendszeresek a szenzációs hírek új DNS vizsgálati módszerekről, valamint az alkalmazásukkal elért eredményekről. A technika terjedésének legutóbbi, média figyelmet (is) kapott hírei: A tömeges szekvenálás bevonul a napi klinikai gyakorlatba, miután az FDA első alkalommal engedélyt adott NGS (next generation sequencing) berendezés diagnosztikai célú alkalmazására (2013. november). Piacra kerül a DNS szekvenálásban „hangsebesség határt áttörő” készülék, amellyel a humán genomszekvencia meghatározás „demokratizálódik”, mert 1000 USD költségen megvalósítható (2014. január). Sikeres próbakísérleteket végeznek egyetlen molekula vizsgálatával szekvenciát leolvasó készülékkel, amelynek fejlesztői a másik álomhatárt, az 1000 USD készülék árat célozták meg (2014. február). Egy soha nem látott iramú technikai fejlődés lépései ezek, aminek eredményeként az egységnyi idő alatt nyerhető DNS szekvencia adat és az egységnyi adatmennyiség megszerzésének költsége öt nagyságrenddel változott néhány év során. Természetesen ellentétes irányba. A fejlődés, aminek üteme meghaladja a komputer számítási kapacitásának kétévenkénti duplázódását, a biokémia, elektronika, informatika, optika, nanotechnológia és számos más tudományterületek közös gyümölcse. Olyan gyors, hogy még elnevezésével is elmaradásban vagyunk. Egyszer a DNS szekvenálási módszerek következő generációjáról beszélünk, máskor már a DNS szekvenálási technikák második, harmadik, sőt negyedik nemzedékéről.

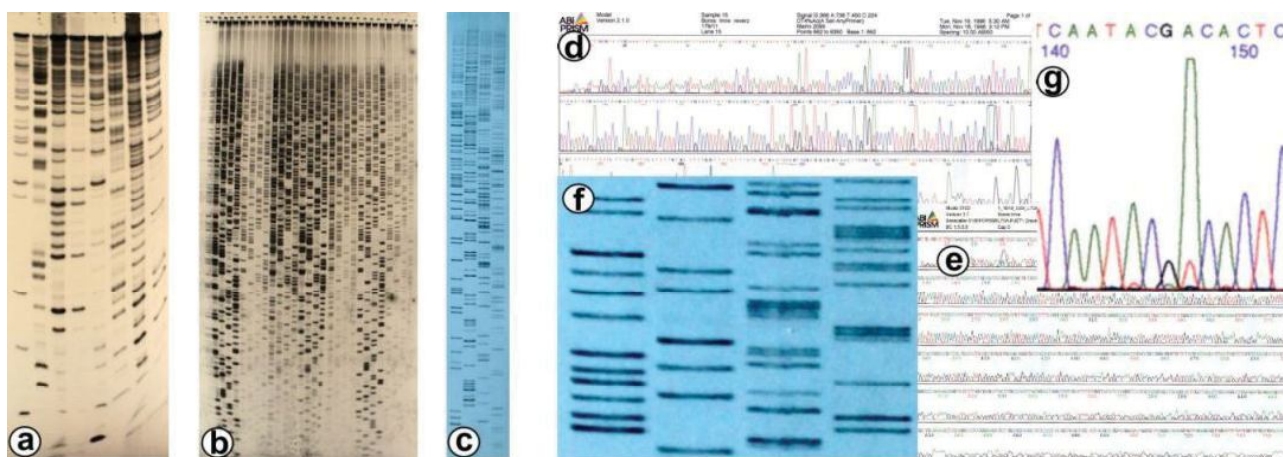
Az összefoglalóban áttekintem, hogy honnan indult és milyen irányba fejlődik a DNS szekvencia analízis. Személyes apropó az áttekintéshez, hogy a 70-es évek második felében közreműködtem, amikor az első hazai szekvenálási eredmények születtek, a 80-as évek végén az NIH-ban közelről hallottam a vitát arról, hogy a humán genomra szánt 3 milliárd dollár jó befektetés-e vagy kidobott pénz, 2013 végén pedig örömmel asszisztálhattam NGS készülék beüzemelésénél a vezetésemmel dolgozó munkacsoportban.

1. ábra (lásd címlap). Egy humán genomszekvenálás költségének változása 1990 és 2014 között. A genomszekvenálás gondolata a nyolcvanas évek végén került komolyan szóba, akkor a várható költséget 3 milliárd dollárra becsülték. A költségek csökkenése már a genomprojekt alatt elkezdődött, az NGS technikák terjedésével pedig zuhanásszerűvé vált. Várakozás szerint 2014-ben valóban teljesül az 1000 USD/genom cél.

Nukleinsavak primer szerkezetének meghatározásáról az első közlemények az 1960-as években jelentek meg. Ma meglepőnek tűnhet, hogy először RNS szekvenálásban született több eredmény. Ezt könnyen megérthetjük, tudva, hogy az abban az időben rendelkezésre álló módszerekkel kis DNS darabokat nem lehetett reprodukálhatóan előállítani és elkülöníteni, ellentétben a „kezelhető” méretű kisebb RNS molekulaféleségekkel. Az első teljes DNS genom, aminek nukleotidsorrendjét meghatározták, a θ X174 bakteriofágé volt. Az 5375bp szekvenciát Sanger és munkatársai közölték 1977-ben. Ez jelentős győzelem volt a DNS nukleotidsorrend-meghatározás akkor használt két módszerének versenyében. A Maxam és Gilbert által kidolgozott technikával kémiai reagenseket használva hoztak létre bázisspecifikus lánc töréseket radioaktív izotóppal jelzett DNS fragmentumban. A módszer fejlesztésével számos reakció típust használtak. Közös jellemzőjük volt, hogy a bázismódosítást báziseltávolítás, majd lánchasítás követte, és hogy a reakciót korlátozni kellett fragmentumonként egy vagy csak néhány lánc törés elérésére. A leggyakrabban használt reagensek és reakciók a következők voltak: dimetilszulfát (purinok metilálása), hidrazin (pirimidin nukleotidok bontása), piperidin (cukor-foszfát lánc hasítása). A négy (rendszerint G, A+C, T+C, C) elkülönített reakció termékeit 6M ureát tartalmazó akrilamid gélen választották el. A szekvenciát a gélről készült autoradiográfiáról lehetett leolvasni a fragmentumok elmozdulása alapján. A szekvenáláshoz viszonylag nagy mennyiségű (mikrogrammnyi) tisztított, csak egyik végen radioaktív jelet hordozó fragmentumok mintáin kellett reakciókat végezni. Ezekben esetenként radioaktív mintákat kellett forró vízfürdőben tartani, bepárolni, géleket beszárítani. Röviden: ma riasztónak hangzó manipulációk sorozatát kellett végezni. Egy-egy többnapos kísérlet eredménye néhányszor tíz, maximum pár száz nukleotidnyi leolvasható szekvencia volt.

A Sanger-féle módszer első változatát az előbbtől kissé korábban közölték. Ebben egyes-szálú DNS templáton végeztek korlátozott láncszintézist DNS polimeráz enzimmel. A szintézis négy mintában folyt. A négy reakcióban más-más DNS-alkotó nukleotid előfordulásakor történhetett lánctermináció. A módszer első változatában egy-egy dezoxinukleozid-trifoszfát limitált koncentrációban történő használatával értek el láncterminációt. Hamarosan tökéletesítették a módszert és az egyes mintákba kevert 2',3'-di-dezoxinukleozid-trifoszfát prekursorokkal

(ddTTP, ddCTP, ddATP, ddGTP) okoztak láncterminációt. Amikor egy ilyen nukleotid épül be - a normál dezoxinukleozid-trifoszfátokhoz viszonyított előfordulási aránya által meghatározott gyakorisággal, a polinukleotidláncban véletlenszerűen helyettesíti a megfelelő nukleotidípust -, 3'OH hiányában a szintézis nem folytatható. Az előzőhöz hasonlóan ebben a módszerben is a reakciókat a képződött fragmentumok méret szerinti elválasztása követte gélelektroforézissel, majd a gélről autográfia készítése és a filmről a szekvencia leolvasása alulról felfelé haladva.



2. ábra. A DNS szekvenálási módszerek első nemzedékét jellemző képek.

a, b, c) Autoradiogramok a hőskorból (1979-1993).

a) Maxam-Gilbert módszerrel készült szekvenálás. C, A+C, T+C és G specifikus reakciókból származó ^{32}P -jelölt fragmentumok 6M urea-20% akrilamid gélen elválasztva. A kémiai reakciók és a vastag gélek feloldóképességének korlátai miatt néhányszor tíz nukleotid szekvencia olvasható. (Csordás-Tóth Éva kísérlete).

b) Sanger-féle szekvenálás elektroforézisének autoradiogramja 1990-ből. A fragmentumok jelölése ebben az esetben is ^{32}P volt, a vékonyabb gél jobb felbontást eredményezett, de a béta bomlás nagy energiája miatt vastagok a csíkok, ami korlátozza a hosszú leolvasást.

c) A kisebb energiával bomló ^{35}S jelzés és ultravékony (0,2 mm), hűtött gélek használatával a gélelektroforézist használó láncterminációs módszer elérte teljesítése határát (1993).

d, e) Az első nemzedék felnőtt korát jelentették a gélelektroforézist, majd kapilláris elektroforézist használó szekvenáló automaták (1996 és 2004-ben ABI PRISM készülékkel készített szekvenálások). A négy eltérően fluoreszkáló láncterminátort használó készülékek szolgáltatták a referencia szekvenciák többsége is.

e, f) Az első nemzedékre jellemző gél- és kapilláris-elektroforézis csak korlátozottan teszi lehetővé a ritka szekvencia variánsok felismerését.

A két szekvenálási módszert csak rövid ideig használták egymás mellett. A Sanger-féle technika fejleszhető volt és hamarosan egyeduralkodóvá vált. Egyszerűbb a hozzá használható templát készítése, sőt ssDNS fág-vektorokat használva klónozáshoz közvetlenül izolálható a templát. Az oligonukleotid-szintézis fejlődése könnyen megvalósíthatóvá tette primerek szintetikus előállítását. Az enzimmel végzett reakció pedig kevesebb manipulációt követelt és megbízhatóbban működött, mint a bázis-specifikus kémiai lánctörések. A Sanger

módszer további fejlesztésének legfontosabb lépései voltak, hogy a nukleotidok megkülönböztetésére radioaktív jelzés helyett, amire előbb ^{32}P , később ^{35}S jelzést alkalmaztak, fluoreszkáló csoporttal jelzett prekursorokat kezdtek használni. A fluoreszkáló jelet előbb a primerre, majd a termináló nukleotidokra helyezték. Az áteresztőképesség növelésére a reakciókat és a poliakrilamid gélelektroforézist automatával végeztették, majd a gélelektroforézist kapilláris elektroforézissel váltották fel. A Sanger-féle szekvenálási módszer előnye az is, hogy alkalmas vektorba építve a szekvenálandó fragmentumokat ugyanolyan (univerzális) primerek használhatók tetszőleges DNS darabok nukleotidsorrendjének a meghatározásához. Ilyen esetekben az inszerciót határoló részek szolgálnak primer kapcsolódási helynek. A módszer utolsó változatában a négy láncterminációs szintézist egy reakcióban végzik hőstabil polimerázzal, spektrálisan megkülönböztethetően fluoreszkáló di-dezoxinukleozid-trifoszfátokat használva. A nukleotid-specifikusan terminálódott termékeket kapilláris-elektroforézissel választják el és lézerrel gerjesztve azonosítják őket. Ezzel a DNS szekvenálás technikája elérte azt a színvonalat, amivel 1997-2002 között sikerült meghatározni a legfontosabb modellorganizmusok (élesztő 1997, fonalféreg 1998, muslica 2000, lúdfű 2000, egér 2002) és a humán genom (2001) nukleotidsorrendjét. Az automatákban végzett Sanger-féle láncterminációs szekvenálások során 600-800 nukleotidos leolvasások végezhetők kis hibaszázalékkal. Erre a technikára, mint **első generációs szekvenálási módszer** hivatkoznak. Az ezt követő „**next-generation sequencing**” (NGS) technikák 2005-től kezdtek elterjedni. A módszerek egy részében a Sanger-féle technika tovább él, de vannak alapvetően új megközelítést alkalmazók is, amelyek képesek egyedi molekulák alapján nukleotidsorrendet meghatározni. E technikák – amelyeket egyes összefoglalók a szekvencia-meghatározási módszerek harmadik és negyedik nemzedékének említenek – rövid ismertetése előtt emlékezzünk meg Sanger munkásságáról és tekintsünk vissza a DNS szekvenálás hazai kezdetére is.

Frederick Sanger kétségtelenül a legjelentősebb hozzájárulást tette mind fehérjék, mind nukleinsavak tekintetében az elsődleges szerkezet meghatározására alkalmas módszerek kifejlesztésében. Sanger 2013 novemberében halt meg, 92 éves korában. Két alkalommal kapott Nobel-díjat, az egyiket a fehérje szekvenálási módszer (1958, inzulin szekvencia), a másikat a láncterminációs DNS szekvenálási módszer kifejlesztésért (1980).

Magyarországon *in vitro* DNS rekombinációs kísérleteket először az SZBK Biokémiai Intézetben, a Venetianer Pál vezetésével működő Nukleinsav

Csoportban végeztek a 70-es évek közepén. Ez volt világszerte is a technika születésének időszaka. Restriktív endonukleáz alkalmazás, rekombináns DNS előállítás és DNS nukleotidsorrend meghatározás is ekkor történt először hazánkban az SZBK 4. emeleti laboratóriumaiban. DNS szekvenálást Csordás-Tóth Éva végzett a Maxam-Gilbert módszerrel. Klónozott riboszomális RNS gén promoter és lambda fág kromoszómába épülési hely nukleotidsorrendjét határozta meg. Az eredmények két Nucleic Acids Research közleményben jelentek meg 1979 őszén. Jellemző, hogy akkor természetes volt egy-egy szekvenciárészlet meghatározásához készült autoradiogram képét a közlemény ábrájaként bemutatni.

A nukleinsav-szekvenálás legnagyobb publicitást kapott teljesítménye a humán genomszekvencia meghatározása volt, amit 2001-ben nyilvánítottak elkészültnek. Ekkorra ugyanis még csak az ún. *draft* szekvencia volt kész, ami számos lyukat tartalmazott. A *finished grade* szekvenciák minőségi jellemzői (lefedettség, hiba előfordulás) lényegesen jobbak. A humán genom szekvenálása közel három milliárd nukleotid sorrendjének a meghatározását jelentette. A projekt indításakor ekkora összegre becsülték (dollárban) a várható költségeket is. Az egész vállalkozás története érdekes és színes, de ennek az áttekintésnek nem témája¹. A jelen írás szempontjából legfontosabb, hogy eredményeként született meg az ún. humán genom referencia szekvencia.

A humán genomszekvencia teljes egészében első generációs szekvenálási technikával, négy szín fluoreszcens Sanger-féle láncterminációs módszerrel készült (a módszert nevezik ciklikus vagy reverzibilis láncterminációs módszernek is). Meghatározásában ABI PRISM 3700 (Applied Biosystems, Inc.) készülékek és továbbfejlesztett típusaik játszottak főszerepet. Egyidejűleg már lázas sebességgel folyt újabb szekvenáló módszerek fejlesztése és nem sokkal a humán genom szekvencia elkészülése után megjelentek és terjedni kezdtek az NGS módszerek. Az új módszereket több szempont szerint csoportosíthatjuk, bár nehéz osztályokba sorolni őket. A rendkívül gyors technikai fejlődés gyorsan devalválja a teljesítmény, sebesség, költség összevetéseket. A továbbiakban az NGS módszerek közös jellemzőit és a szekvenálás új nemzedékének legfontosabbnak vélt sajátosságait tekintem át. NGS megvalósítására alkalmas eszközökről és a módszerek specifikusabb jellemzőiről egy következő részben fogok rövid áttekintést adni.

¹ A témáról bővebben olvasható Venetianer Pál Az emberi genom című írásában, *Biokémia* XXXVII/3.2013. szeptember, 22-28. oldal (a szerkesztőbizottság megjegyzése).

Érdeemes tisztázni, hogyan definiálhatunk szekvenálókészülék és módszer nemzedéket. Erről nincs egységes vélemény a szakirodalomban.

A klasszikus Sanger-féle módszert követő szekvenálási technikák egyik legjelentősebb újdonsága, hogy fragmentumok halmazait tudják egyidejűleg kezelni. Sanger-féle szekvenáláshoz a DNS fragmentumokat klónozással vagy PCR-t alkalmazva felszaporítják. Az eljárás során, ha nem is tudják, hogy melyik DNS darab honnan származik, és minek felel meg, a manipuláció valamelyik lépésében mégis minden egyes DNS fragmentum fizikailag is elkülönített, ha kell, akár kézbe is vehető. Ehhez képest az NGS módszerek fragmentumok halmazán végeznek amplifikációt (ha szükség van a fragmentumok megsokszorozására) olyan elrendezésben, hogy egy-egy fragmentum minden másolata ugyanabban a térrészben (emulzió cseppben, szilárd felületen rögzítve), más fragmentumoktól elkülönítve helyezkedik el. Tekinthejtük ezt, a fragmentumokat egy halmazként kezelő, de a fragmentumokat felszaporítva vizsgáló technikát **második generációs** szekvenálási módszernek. Ilyet használ a kereskedelmi forgalomban beszerezhető készülékek többsége (Roche, SOLiD, Illumina, Ion Torrent). Ezekhez képest jelentős előrelépés az, amikor a szekvenciameghatározás nem felszaporított fragmentumokon, hanem egyedi DNS molekulák analízisével történik. Ez kiküszöböli az amplifikáció során fellépő hibákat. Már vannak olyan készülékek, amelyekkel egy molekula vizsgálata alapján lehet nukleotidsorrendet meghatározni (Pacific Biosciences PacBio, Oxford Nanopore MinION). Ezeket tekinthejtük a következő, **harmadik generációt** jelentő szekvenáló berendezéseknek. A negyedik generációt talán azok az eljárások képviselhetik, amelyek egyetlen molekula többszörös letapogatására lesznek képesek.

Az NGS módszerek mindegyikére jellemző, hogy az alkalmazás bonyolult, specifikus és költséges eszközt és értékelő software-t igényel. A nagy teljesítőképesség, a költséges készülék és a komoly kihívást jelentő értékelő informatika hármasa miatt ezért NGS készülékek a legtöbb esetben szakosodott vagy intézeti szolgáltató laboratóriumokban (core facility) működnek. A kisebb laboratóriumok számára is elérhető változatokként az utóbbi években kerültek piacra három cég készülékeinek bench-top változatai (Roche GS Junior, Life Technologies PGM és Illumina MiSeq). Ezek 100 ezer euró körüli vagy az alatti áron beszerezhetőek, ennek megfelelően egyedi laboratóriumokban is terjednek. Az Illumina MiSeq egy variánsa 2013 novemberében FDA engedélyt kapott klinikai alkalmazásra az USA-ban, így várhatóan nőni fog alkalmazása klinikai környezetben is.

Bár a különböző NGS módszerek lényegesen eltérnek, többé-kevésbé közös jellemzőjük a három következő munkafázis: (1) minta előkészítés (templát készítés), (2) szekvencia olvasás és rögzítés, (3) adatanalízis. A minták lehetnek amplifikációval előállított molekulahalmazok vagy egyedi DNS molekulák. Templát készítéséhez a mintából olyan fragmentum halmazt (könyvtár, library) hoznak létre, amelyik lehetővé teszi a fragmentumok elkülönített megsokszorozását, az egy-egy mintából származó fragmentumok későbbi azonosítását és segít a leolvasott nukleotidsorrendek összerendezésében. Ehhez adaptor és azonosító (barcode) részekből álló oligonukleotidokat kapcsolnak a fragmentumokhoz. Attól függően, hogy az oligonukleotidokat milyen hosszú DNS-darabokhoz, milyen trükköket használva ligálják, létrehozhatók specifikus könyvtárak, amelyek egymáshoz közelebbi vagy távolabbi helyzetű részek szekvenálását teszik lehetővé. Az ún. „mate-pair” technikával hosszabb fragmentumok végeiről nyerhető szekvencia adat, mert a könyvtár készítésekor előbb egy oligonukleotiddal körre zárják a fragmentumokat, majd feldarabolva őket, az oligonukleotidon át összekapcsolódott részekon végeznek szekvenálást. Ezzel az eredetileg egymástól távolabb fekvő részekről nyerhető szekvencia információ, ami segít a nehezen szekvenálható, ismétlődő részek áthidalásában. A hibridizációt és ligálást használó szekvenálási módszerekben is adaptor kapcsolási trükkökkel érik el azt, hogy egy-egy fragmentum több pontján is olvashatók legyenek rövid szekvencia részek.

Az adaptorok segítik a minta molekuláinak elkülönítését, mert velük kapcsolhatók a fragmentumok egyenként szilárd felülethez. Ezt az adaptorok és a szilárd felülethez rögzített komplementer oligonukleotidok közötti hibridizáció biztosítja. A fragmentumok sokszorozása PCR amplifikációval történik emulzióban vagy szilárd fázison rögzített templáton, és egy molekula több ezer-százezer másolatát eredményezi az emulzió egyetlen cseppjében, vagy a szilárd fázis közeli pontjain rögzítve. A szekvenálás alapulhat templát alapján végzett szintézisen, templát felszínen összeilleszkedő oligonukleotidok összekapcsolásán ligázzal és valami más elven, ami rendszerint a nukleotid bázisok eltérő méretét használja ki azonosításukra. A szintézist használó szekvenáló reakciók ciklikusan megfordíthatók láncterminációval (CRT), nukleotidok egyenként történő beépítésével (SNA), vagy folyamatos nukleotid beépítéseket valós időben követve történnek. Az adatrögzítés pedig leggyakrabban lumineszcencia jel, négy színű fluoreszcencia jel, protonkoncentráció vagy ionáram-változás mérése. Az egyes munkafázisok jellemzőire itt nem térek ki, mert azok eszköz-specifikusak.

A szekvenálás eredményével és minőségével kapcsolatos néhány fogalmat érdemes azonban áttekinteni. NGS módszerekkel a szerkezetre vonatkozó, minőségi jellegű adaton kívül, amit a nukleotidsorrend jelent, mennyiségi adatokat is nyerünk egy-egy nukleotidsorrend előfordulási gyakoriságáról a vizsgált mintában. A hagyományos technikával ez nem teljesül, mert az alacsony arányban jelenlevő szekvenciavariánsok nem észlelhetők gél- és kapilláris-elektroforézissel. Az NGS technikákban a könyvtár fragmentumai külön-külön szolgáltatnak szekvencia-adatot, ezért megszámlálhatók. Ha vannak ritka variánsok a mintában, akkor azok is kimutathatók, feltéve, hogy elég nagy volt a megszámlált fragmentumok száma, amit a szekvenálás mélysége („depth”) fejez ki. Természetesen feltétele a helyes eredménynek az is, hogy a könyvtár készítése során a fragmentumok eloszlási arányában ne történjen változás. Annak kifejezésére, hogy egy-egy ponton átlagosan hány leolvasás történt, a lefedettségi mutatót (coverage) is használják. Ez más esetekben azt jelzi, hogy a genom hány százalékáról nyertünk szekvencia adatot. A *coverage* tehát az előző értelemben a lefedés mélysége (vastagsága), a másodikban kiterjedése.

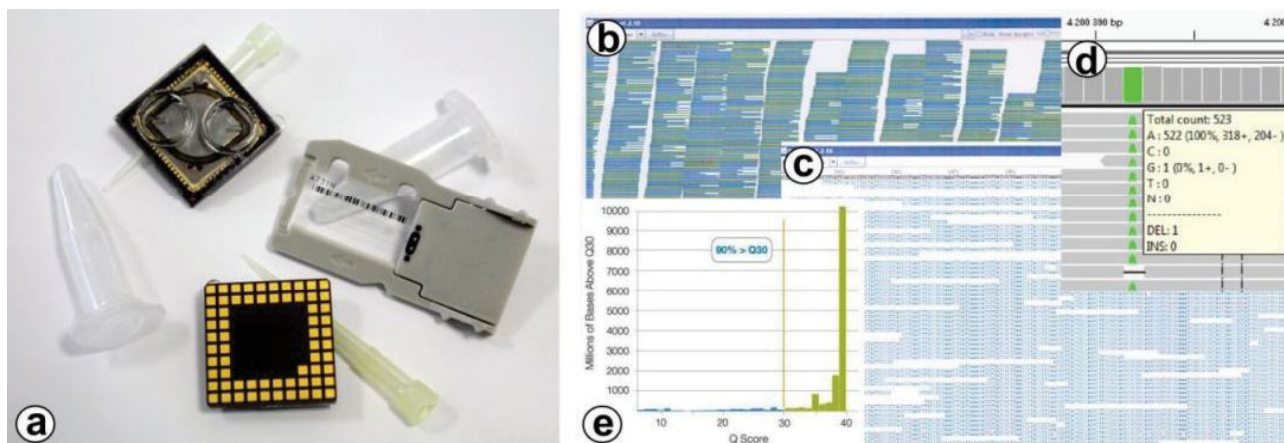
Amint egyes fragmentumainak mennyiségét kifejező információ alapkövetelmény transzkriptom analízisek és ChIP-seq alkalmazások során. Ezekben valójában ez az elsődlegesen keresett adat, a nukleotidsorrend a fragmentum azonosításához kell. Igény van mennyiségi adatokra más alkalmazások során is, pl. amikor szomatikus mutációk előfordulási gyakoriságát kell kideríteni. Ez gyakori kérdés a tumorok nagymértékű heterogenitása és változékonysága miatt, mert a mutációk előfordulásáról szóló információ fontos lehet célzott kezelési eljárások tervezéséhez. Szintén fontos a mennyiségi jellegű adat a metagenomikai alkalmazásokban, amikor egy adott mikroorganizmus-populáció összetevőinek mennyiségére a mintában jelenlevő szekvenciák arányai alapján következtetnek.

A kis arányban meglévő szekvencia-heterogenitás kimutathatósága a leolvasások számán kívül a meghatározási eljárás megbízhatóságának is függvénye. A megbízhatóság számszerűsítésére a *Q* érték szolgál. A Sanger-féle módszer hagyományos alkalmazásakor a szekvenáló reakciók és az adatrögzítés egymástól időben elválasztottak. Az egy szekvenálási reakcióból nyerhető információ mennyiség a gél felbontóképessége által meghatározott leolvasási hosszúság. A NGS eljárások többségében a szekvenálási reakció és az adatrögzítés ciklikusan ismétlődve történik. Ezért a jel és zaj arányának függvénye az, hogy az egymást követő reakciók hányszoros ismétlődés után eredményeznek még értékelhető adatot. Az ismétlések számának növelésével a jel-zaj arány

fokozatosan romlik. A NGS berendezések az érzékelt jelek nukleotidsorrendre fordításakor minden nukleotidra megbecsülik a leolvasás megbízhatóságát. A jel megfeleltetése nukleotidnak a bázis-azonosítás, „base call”. Ennek minősítéséhez a készülék programja értékeli a mért jel „jóságát”, összevetve a jelet jellemző értékeket (erősség, szélesség stb.) a már ismert szekvenciák leolvasása során nyert értékekkel. Az összevetés alapján minden leolvasott nukleotidhoz az olvasás megbízhatóságának valószínűségét tükröző számértéket rendel. Ebből származtatható a Q érték, ami a hibás azonosítás valószínűségével arányos. Q20 azt jelenti, hogy 100 alkalomból 1-szer várható hiba, Q30 esetén 1000-ben egy, Q40-nél 10.000 leolvasott nukleotidban egy tévedés várható. Tekintettel, hogy a leolvasások hossza rendszerint nem több mint néhány száz nukleotid, Q30 azt jelenti, hogy a leolvasások többsége hibátlan. Más szavakkal: Q30 értéknél a leolvasás pontossága 99,9%. Megjegyzendő, hogy természetesen Q értéket csak akkor lehet egy adott nukleotidhoz hozzárendelni, ha megtörtént a leolvasása.

A szekvenálás módszerétől függően különböző okokból származhatnak hibák. A hibák egy része eszköz, illetve módszer specifikus, más részük véletlen. Elmaradhat egy nukleotid érzékelése a fluoreszcens jel hiánya miatt, vagy egy védőcsoport eltávolítása nem történik meg, esetleg homopolimer részeknél nem megkülönböztethető az n és $n+1$ -es pozícióba történő beépülés. Általában az egy készülékre/módszerre jellemző hiba-spektrumot a hiba-modellben foglalják össze. A hibákat a készülékek programjai próbálják javítani. A programmal megvalósítható hibakorrekció sikere függ az egyes pozíciókban történt leolvasások számától, a szekvenálás mélységétől. Fontos észrevenni, hogy az amplifikációs lépéseket használó módszerek esetén a hibák korrekcióját tekintve lényeges különbséget jelent az, hogy melyik fázisban történt inszerció, delécio („indel”), helyettesítés vagy az arányokat megváltoztató szelektív sokszorozás. A könyvtár készítésekor fellépő hibák kiküszöbölésére más technikával létrehozott könyvtár, míg a rendszerhibák ellen más NGS eszközzel is végzett futtatás lehet hatásos.

Egy-egy szekvenáló projekt célja lehet *de novo* szekvencia-meghatározás, vagy a vizsgált minta már ismert szekvenciával történő összevetése, újraszekvenálása. A humán és a legfontosabb modell organizmus genomok *de novo* szekvenálása megtörtént a Sanger-féle módszerrel. A kapott eredmények képezik ezeknek az élőlényeknek a referencia-szekvenciáit (érdekes kérdés, hogy genomszekvenciák esetében mi a jó etalon). A *de novo* és az újraszekvenálás természetesen eltérő kihívást jelent, elsősorban az adatfeldolgozás tekintetében. Az előbbi esetében a szekvenciák összeillesztése, a lyukak befoltozása nehéz feladat.



3. ábra. Csoportkép a szekvenálási eljárások második nemzedékének jellemzőiről. a) Az NGS berendezésekben a nanotechnológia és elektronika ötvöződik módosított enzimek korábban elképzelhetetlennek vélt alkalmazásával. Két IonTorrent chip és egy MiSeq átfolyó cella, az NGS készülékek két bench-top változatának fogyóeszközei. Ezek az apró eszközök 3-5 millió fragmentum egyidejű szekvenálását és 3-15 Gb leolvasását teszik lehetővé. Az NGS eredménye a könyvtár minden fragmentumának sok-sok alkalommal leolvasott nukleotid sorrendje. Elrendezve a fragmentumokat (b), jól látható az egyes részek sokszoros lefedettsége, nagyobb nagyítást használva pedig (c), hogy a fragmentumok többségének olvasása hibátlan. A készülékek kiszolgáló programja már a futás alatt megbízhatósági értéket rendel minden leolvasott nukleotidhoz (e). A futtatás követő analízis pedig megadja az eltérések helyét, gyakoriságát és a két irányból tett leolvasások számát (d).

A lyukak megszüntethetősége elsősorban a leolvasások hosszától függ (read length: a nukleotidszám, ami a szekvenálásban résztvevő fragmentumokból átlagosan meghatározható). Ez a mai NGS készülékekkel végzett szekvenálások esetén széles tartományban, néhányszor tíz nukleotidtól (ilyeneket eredményeznek a ligálást használó módszerek) néhányszor ezer nukleotidig terjedhet. Meglepő talán, hogy az igazán hosszú leolvasásokra az egyedi molekulák vizsgálatával szekvenciát szolgáltató berendezések képesek.

Az NGS platformok további jellemzői között fontosak azok, amelyek megadják egy-egy készülék teljesítményét (capacity, output), a szekvenálási folyamat idejét (ez néhány órától több napig eltarthat), az egyidejűleg kezelhető minták számát (multiplexing vagy barcoding), és természetesen nem elhanyagolható a készülékek beszerzésének és működtetésének költsége. Az ezeket jellemző értékek gyorsan változnak. Az árat tekintve jellemzően képviselheti a szélsőértékeket két 2014 elején bemutatott készülék: az Illumina HiSeq X Ten rendszer - ami értékelői szerint valóban megvalósítja majd a régen kitűzött 1000 dollár per genom célt - listaára 10 millió dollár. Az Oxford Nanopore MinION a másik véglet. Ezt készítői a mindenki számára elérhető, olcsó eszköznek szánják, 1000 dollár körüli áron. A magas eszköz beszerzési árak persze nem mindenhol jelentenek gátat. A Pekingi Genomikai Intézetben (BGI) 2011-ben többek között

137 db HiSeq 2000 és 27 SOLiD készülék működött. Persze szükség is van a nagy kapacitásra az ott folyó 3M projekt teljesítéséhez, ami 1 millió faj, 1 millió humán genom, és 1 millió metagenom/mikrobiom meghatározását tűzte ki célként. Az alábbi lista a DNS szekvenálás fejlődésének legjelentősebb technikai/módszertani lépéseit tartalmazza. Az áttekintés második részében a 2005-től kibontakozó második szekvenáló nemzedék már széleskörűen használt vagy előrehaladott fejlesztési fázisban lévő módszereinek és készülékeinek jellemzőit összegezem röviden, az alkalmazott szekvenálási eljárás szerint csoportosítva őket.

1975-1977	Láncterminációs és kémiai DNS szekvenálási módszerek közlése.
1986	Első fluoreszcens jelzést használó automata szekvenáló készülék.
1999	Kapilláris szekvenáló készülékek bevezetése.
2005-től	NGS készülékek megjelenése.
2009-től	Bench-top NGS készülékek megjelenése.
2013	FDA engedély NGS alkalmazásra klinikai diagnosztikában.

Az összefoglaló a TÁMOP-4.2.4.A/2-11/1-2012-0001 Nemzeti Kiválóság Program című kiemelt projekt keretében nyert támogatással készült.

Irodalom

Az érdeklődő a szekvenálás korai és újabb módszereiről egyaránt könnyen találhat irodalmat. Az alábbi néhány közlemény közül néhány a korai időkből származik, a későbbi összefoglalók pedig referencia listájukkal is segíthetnek a további tájékozódásban.

Sanger, F., Coulson, A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Molec Biol*, **94**: 441-448.

Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*, **74**: 5463-5467.

Maxam, A.M., Gilbert, W. (1977) A new method for sequencing DNA. *Proc Natl Acad Sci USA*, **74**: 560-564.

Csordás-Tóth, E., Boros, I., Venetianer, P. (1979a) Nucleotide sequence of a secondary attachment site for bacteriophage lambda on the Escherichia coli chromosome. *Nucleic Acids Res*, **7**: 1335-1341.

Csordás-Tóth, E., Boros, I., Venetianer, P. (1979b) Structure of the promoter region for the rrnB gene in Escherichia coli. *Nucleic Acids Res*, **7**: 2189-2197.

Franca, L.T.C., Carrilho, E., Kist, T.B.L. (2002) A review of DNA sequencing techniques. *Q Rev Biophys*, **35**: 169–200.

Niedringhaus, T.P., Milanova, D., Kerby, M.B., Snyder, M.P., Barron, A.E. (2011) Landscape of Next-Generation Sequencing Technologies. *Anal Chem*, **83**: 4327–4341.

Mardis, E.R. (2013) Next-Generation Sequencing Platforms. *Annu Rev Anal. Chem*, **6**: 287–303.

Collins, F.S., Hamburg, M.A. (2013) First FDA authorization for next-generation sequencer. *N Engl J Med*, **369**: 2369–2371.



Boros Imre Miklós 1953-ban született Taktaharkányban. 1978-ben végzett a JATE biológus szakán. Ekkor már több éve diákkörös hallgatóként az SZBK-ban dolgozott Venetianer Pál csoportjában. 1985-ben szerzett kandidátusi fokozatot, 2000-ben MTA Doktora címet. Pályakezdésétől az SZBK Biokémiai Intézet munkatársa. 2002-től a Szegedi Egyetem egyetemi tanára, jelenleg a Biokémiai és Molekuláris Biológiai Tanszék és a Biológus Tanszékcsoport vezetője. Több alkalommal dolgozott hosszabb időn át az Egyesült Államokban (1986-89 NIH Cancer Institute, 1992-95 Case Western University, Medical School). Cancer Institute Oncology Research Faculty Development Award, Széchenyi Ösztöndíj, Szentágothai Ösztöndíj, Straub plakett, Ipolyi Arnold Díj elismeréseket kapott. Érdeklődése a génműködés transzkripció szinten megvalósuló szabályozása, az utóbbi tíz évben munkacsoportja a kromatinszerkezet szerepét és epigenetikai hatásait vizsgálja.