

A GENOMKORSZAK BIOINFORMATIKÁJA

Patthy László

MTA SZBK Enzimológiai Intézet, Budapest – patthy@enzim.hu

Kevesebb, mint egy évtized telt el azóta, hogy az első önálló életre képes élőlény teljes genomszerkezetét, nukleinsav-szekvenciáját meghatározták, de az eltelt néhány év (és nagyszámú további genom-projekt) tapasztalata alapján biztosan állíthatjuk, hogy ez az esemény egy új korszak nyitányát jelentette, mely alapvető változásokat hozott az élettudományok minden területén.

A genomkorszak első néhány éve eredményeként több tucat baktérium (pl. *Haemophilus influenza*, *Mycoplasma genitalium*, *Escherichia coli*, *Bacillus subtilis*), számos Archaea (pl. *Archaeoglobus fulgidus*, *Methanococcus jannaschii*), néhány egysejtű eukarióta, (pl. *Saccharomyces cerevisiae*, *Encephalitozoon cuniculi*), egy virágos növény (*Arabidopsis thaliana*), két gerinctelen állat (egy fonálféreg, *Caenorhabditis elegans* és a harmatlégy, *Drosophila melanogaster*) és az ember (*Homo sapiens*) genomszerkezete vált ismertté. Így ma már elmondhatjuk, hogy a *genomszekvenálási programoknak* köszönhetően az élővilág legfontosabb, nagy csoportjainak mindegyikéről rendelkezünk genomikai ismeretekkel.

A genomkorszak egyik jellegzetességéről árulkodnak az egyes genom-projektek eredményeit közlő publikációk rendkívül hosszú szerzői listái: nagyszámú, sokféle szakértelmet reprezentáló kutatóból álló konzorciumok, vállalkozások munkája eredményeként születhettek csak meg a genomszekvenciák első összegzését bemutató közlemények (Adams és munkatársai, 2000; Lander és munkatársai, 2001; Venter és mun-

katársai, 2001; ezeket a közleményeket 200-300 szerző jegyzi). A különböző résztvevők együttműködésének nélkülözhetetlenségét tükrözi, hogy a közleményeket sokszor nem a szerzők hosszú listájával, hanem a konzorcium nevével jegyzik (lásd: *The Arabidopsis Genome Initiative*, 1999; *The C. elegans Sequencing Consortium*, 1998). Míg a 19. század végének, 20. század elejének kísérletes biológiájára a magányos tudós volt jellemző, úgy tűnik, a genomkorszakban a széles körű együttműködés válik általánossá.

A genomkorszak másik jellemző trendjének tekinthetjük, hogy a *biológiai adatgyűjtés tömegméréteiben történik*, iparszerűvé válik. Ipartörténeti hasonlattal azt is mondhatjuk, hogy a biológiai kutatás átlép a manuális korszakból a nagyüzemi korszakba. Ez a trend az olyan biológiai tulajdonságok vizsgálatánál érvényesült először, amelyek tömeges vizsgálatára alkalmas technológiák rendelkezésre álltak már a genom korszak kezdetén. A DNS-szekvenálás technológiai fejlettségének köszönhetően a genomszekvenálás nagyüzemi módszerekkel történhetett (valójában ez a tény tette lehetővé a genomkorszak gyors kiteljesedését). A DNS-szekvenálást végző laboratóriumok munkaszervezésüket, hatékonyságukat tekintve leginkább egy csúcstechnológiai üzemre emlékeztetnek.

A genom-projektek sikerének másik fő forrása az volt, hogy a nagyteljesítményű számítógépek kifejlesztése és elterjedése révén jelentős fejlődésnek indulhatott a *bioinformatika – a biológiai adatok számítógépes tárolásával, rendszerezésével, elemzésé-*

vel és értelmezésével foglalkozó tudomány. Jellegéből fakadóan a bioinformatika alkalmas volt a genomszekvenálásból származó nyers szekvencia-adattömeg feldolgozására, a genomszekvenciák rekonstrukciójára, az adatok tárolására és elemzésére.

A genomszekvencia meghatározása önmagában azonban csak annyit jelent, hogy ismerjük egy adott élőlény teljes örökítő anyagának, genomjának kémiai szerkezetét, de nem feltétlenül értjük a szekvencia biológiai jelentését. Bár tudjuk, hogy az adott élőlény működési programjának leírását a genom tartalmazza, ez a leírás egy számunkra még nem teljesen érthető nyelven íródott. Egy új tudományág, a *funkcionális genomika* célkitűzése az, hogy a genomszekvenciákban azonosítsa a géneket, azokhoz funkciókat rendeljen, és végső soron megfejtse azt, hogy a genomszekvencia hogyan kódolja az élőlény tulajdonságait.

Eszköztárát tekintve a funkcionális genomika az élettudományok valamennyi részterületének (molekuláris biológia, genetika, biokémia, szerkezeti biológia, sejtbiológia, élettan, bioinformatika, evolúciobiológia stb.) módszereit felhasználja, jellegét azonban a genomikai szemlélet határozza meg. A funkcionális genomika – a genomkorszak élettudománya – azonban több, mint ezen területek összege: a genom-szemléletnek megfelelően a biológiai adatokat, információkat tömegesen, a teljes genomra vonatkozóan igényli, ennek megfelelően olyan nagy teljesítményű (high-throughput) molekuláris biológiai, genetikai, biokémiai stb. módszerek kifejlesztését és elterjedését segítette (és segíti) elő, melyek alkalmasak a tömeges adatgyűjtésre.

Ebből a szempontból különleges helyet foglal el (jelentős helyzeti előnnyel indult) a bioinformatika, amely *par excellence* alkalmas a tömeges adatkezelésre és adatfeldolgozásra, és amely egyben a legkevesbé költséges funkcionális genomikai módszer. En-

nek köszönhető, hogy a bioinformatika a funkcionális genomika jelenleg legsikeresebb, legelterjedtebb és leggyorsabban fejlődő része. Egyre javuló megbízhatóságú bioinformatikai módszerek állnak rendelkezésünkre, melyek segítségével azonosítani lehet a genomszekvenciában a géneket, meg lehet jósolni a gének által kódolt mRNS-ek és fehérjék szekvenciáját, a fehérjék szekvenciájának összehasonlító elemzésével meg lehet jósolni azok valószínű biológiai funkcióját és szerepét, a fehérjék térszerkezetét, működési mechanizmusát stb. Jóllehet a számítógépen (*in silico*) végzett predikciók nem teszik feleslegessé az *in vitro* vagy *in vivo* kísérleti munkát, sokkal tervszerűbbé tehetik azt.

A genomszekvencia értelmezésének, a genom annotációnak első és alapvető lépése a gének bioinformatikai úton történő azonosítása. A *génazonosítás* még mindig megélő bizonytalanságait illusztrálhatjuk az emberi genomszekvencia példájával. Az emberi genom szekvenciáját meghatározó két független kutatócsoport (Celera, Human Genome Consortium) egyaránt arra a következtetésre jutott, hogy az emberi genom csak mintegy 30 000 fehérje-kódoló gént tartalmaz (Lander és munkatársai, 2001; Venter és munkatársai, 2001). A két csoport által megjósolt génkészlet összehasonlítása azonban jelentős eltéréseket mutatott (Hogenesch és munkatársai, 2001), jelezve hogy a két csoport által használt génpredikciós módszerek nem tökéletesen megbízhatóak. Az emberi genom szekvenciájának újabb, gazdagabb módszertani repertoárt alkalmazó elemzése arra utalnak, hogy az emberi génszám valójában magasabb lehet (Wright és munkatársai, 2001).

A bioinformatikai génpredikciós módszerek meglévő bizonytalanságai aláhúzzák a gének azonosítására szolgáló kísérleti módszerek fontosságát. Egy élőlény genomjában található aktív, transzkripcióra kerülő struk-

turális gének létét tükrözi a *transzkriptoma*, az adott genom génjeiről képződő RNS populáció. Egy adott élőlény transzkriptomájának vizsgálatával kísérletes bizonyítékot szerezhetünk olyan gének létrehozására vonatkozóan is, melyeket a rendelkezésre álló bioinformatikai módszerek nem detektálnak. A közelmúltban például Das és munkatársai (2001) RNS transzkriptumok érzékeny detektálására alkalmas módszerek segítségével tizenhét emberi szövetet vizsgáltak, és 5000-9000 olyan aktív emberi gén létét mutatták ki, melyet az eredeti génpredikciós program nem azonosított. (Ez a munka is amellet szol, hogy az emberi génszámot az első publikációk alulbecsülték.)

A genomszekvencia bioinformatikai úton történő értelmezésének következő lépése, hogy megjósolja az azonosított fehérjekódoló génekről képződő mRNS-ek szekvenciáját (és az azok által meghatározott fehérjék aminosav-sorrendjét). A mRNS szekvenciájának predikcióját bonyolítja az a tény, hogy az eukarióta élőlények génjeiben olyan szakaszok (intronok) találhatók, melyek a génről képződő elsődleges transzkriptumból egy érési folyamat (splicing) során kivágódnak, így hiányoznak az érett mRNS-ből. A predikció bizonytalanságának további forrása, hogy ugyanarról a génről többféle érési folyamat (alternative splicing) révén többféle mRNS képződhet attól függően, hogy milyen szövetben, sejttípusban, fejlődési vagy élettani állapotban kerül sor az adott gén kifejeződésére. A jelenlegi bioinformatikai módszerek még csak korlátozott megbízhatósággal képesek az alternatív transzkriptumok predikciójára, még kevésbé arra, hogy azok milyen arányban, milyen fiziológiás állapotban jelennek meg.

Jelenleg csak a transzkriptóma kísérletes vizsgálata ad ezekre a kérdésekre megbízható választ. Éppen ezért a funkcionális genomika egyik kulcsterülete a *transzkriptomika*, amely – a genom-szemléletnek megfelelően –

olyan, nagyteljesítményű módszereket fejlesztett ki és alkalmaz, amelyek egy-egy adott organizmus, szövet, sejttípus teljes transzkriptómájára vonatkozóan gyűjtenek párhuzamosan, nagytömegű kísérletes adatot. A mRNS-populációról készült cDNS-ek szekvenciavizsgálatával nyert adattömeg nemcsak a gének létrehozására vonatkozóan szolgáltat adatot, nemcsak azt tükrözi, hogy egy-egy génről milyen intenzitású transzkripció folyik, hanem azt is elárulja, hogy egy génről hányféle alternatív transzkriptum képződik, azoknak mi az aránya a különböző sejtekben stb. A transzkriptóma adatok és a genom-szekvencia adatok korreláció-vizsgálatától remélhető az, hogy a jövőben lényegesen megbízhatóbbá válnak a mRNS-predikció módszerei.

A mRNS szekvenciájának ismeretében nagy biztonsággal megjósolható az általa meghatározott fehérje aminosav-sorrendje, így a transzkriptóma (a transzkriptum-populáció) ismeretében az adott élőlény teljes fehérjekészletének, *proteómájának* elsődleges kémiai leírásáról is képet kapunk. Egy meghatározott mRNS-ről lefordított (transzlált) fehérje kémiai szerkezete azonban az átíródást követően sokféle reverzibilis és irreverzibilis átalakuláson mehet keresztül (pl. proteolitikus hasítás, foszforiláció, glikozilálás), és ezeknek az átalakulásoknak jelentős hatásuk lehet az adott fehérje aktivitására, stabilitására, sejten belüli lokalizációjára, kölcsönhatásaira stb. A rendelkezésre álló bioinformatikai eszközök ma már megbízhatóan képesek megjósolni az egyes fehérjék szubcelluláris (vagy sejten kívüli) lokalizációját, de bizonytalanabbak a különböző poszttranszlációs módosítások vagy fehérje-fehérje kölcsönhatások predikciójában. Sok esetben csak kísérleti módszerek segítségével győződhetünk meg arról, hogy egy – bioinformatikai módszerekkel megjósolt – poszttranszlációs módosítás valóban bekövetkezik-e. A reverzibilis módosítások esetén

eleve fenn áll az abból fakadó bizonytalanság, hogy az aminosav-szekvencia alapján csak a poszttranszlációs módosítás lehetőségét jósolhatjuk meg, azt azonban nem, hogy az mikor és milyen arányban következik be (hiszen azt a fehérjén kívüli tényezők határozzák meg). Ezeknek a kérdéseknek a megválaszolására a bioinformatika ma még nem képes, a proteoma szintjén gyűjtött kísérleti adatokra van szükség. Fontosságának megfelelően a funkcionális genomika egyik rendkívül dinamikus fejlődő területe a *proteomika*. A proteomikai vizsgálatok a genom-szemléletnek megfelelően egy-egy adott organizmus, szövettípus, sejttípus teljes fehérjekészletét jellemzik, különös tekintettel az egyes fehérjék poszttranszlációs módosításaira, sejten belüli lokalizációjának változásaira, más fehérjékkel, celluláris struktúrákkal való kölcsönhatására stb. A funkcionális genomika proteomikai igényeinek köszönhetően gyors módszertani fejlődésnek vagyunk tanúi a fehérje-elválasztástechnikák, fehérje-szerkezetvizsgálati módszerek (mikroszekvenálás, tömegspektrometria) és a fehérje-fehérje kölcsönhatások vizsgálata (kettős-hibrid rendszerek) területén.

A fehérjék jellegzetes, az aminosav-sorrendjük által meghatározott szerkezetet vesznek fel, a fehérjék biológiai aktivitása (és így a fehérjekódoló gének biológiai szerepe) a fehérjék térszerkezetéhez köthető. A funkcionális genomika egyik fontos területe – a *szerkezeti genomika* – foglalkozik a fehérjék térszerkezetének meghatározásával. Ezen a téren a bioinformatika kimagasló sikereket ért el az elmúlt néhány év során. A fehérjék aminosav-sorrendjének vizsgálata alapján ma már nagy megbízhatósággal meg lehet jósolni a fehérjék szerkezetének szerkezeti elemeit (alfa helix, beta lemez, transzmembrán helix stb.) és egyre növekvő biztonsággal és pontossággal meg lehet jósolni a fehérjék térszerkezetét. A fehérjék aminosav-sorrendjének illetve térszerkezeté-

nek összehasonlítására szolgáló bioinformatikai módszerek fejlődésével ugyanis nyilvánvalóvá vált, hogy az élővilág fehérjéi meglepően kisszámú (néhány ezer) alaptípusba sorolható fehérje-doménből épülnek fel, a legkülönbözőbb fehérjék ezek leszármazottai. A fehérjék térszerkezetének meghatározása azt mutatja, hogy az egy rokonsági körbe tartozó, homológ fehérje-domének térszerkezete rendkívül hasonló, így egy új családtag térszerkezetét modellezni lehet a homológia alapján. Ha tehát egy új fehérjét – aminosav-szekvenciájának bioinformatikai elemzése révén (molekuláris evolúciós vizsgálatok segítségével) – be tudunk sorolni egy olyan fehérje-domén családba, ahol legalább egy családtag térszerkezetét már ismerjük, akkor a homológia alapján modellezhetjük az általunk vizsgált fehérje szerkezetét is. A genom-projektek mintájára néhány éve elindított Szerkezeti Genomikai megaprojektek résztvevőinek célkitűzése, hogy minden fehérjecsalád egy-egy reprezentánsának térszerkezetét meghatározzák, röntgen kristallográfiás vagy NMR spektroszkópiai módszerekkel. Ily módon egy-egy genom teljes fehérjekészletének homológia modellezése válik lehetővé.

Mint a bevezetőben hangsúlyoztam, a funkcionális genomika fő célja a genom teljes funkcionális leírása. A gének azonosítása, az azok által meghatározott fehérjék szerkezetének megismerése megkönnyíti, hogy bioinformatikai eszközök segítségével megjósoljuk a fehérje-kódoló gén biológiai szerepét, a fehérje funkcióját, működési mechanizmusát, szerkezet-funkció összefüggéseit, kölcsönhatásait. A legelterjedtebb funkció-predikciós módszerek alapja az a tapasztalat, hogy az evolúció során a gének (és az általuk kódolt fehérjék) biológiai szerepe, funkciója általában folyamatos skálán változik, így a rokon gének és az általuk kódolt rokon fehérjék funkciójának különböző aspektusai is sok hasonlóságot mutatnak. Ha tehát egy

újonnan azonosított gén által kódolt, ismeretlen funkciójú fehérjéről – szekvenciájának molekuláris evolúciós vizsgálata alapján – megállapítjuk, hogy az rokonságban áll más, ismert funkciójú fehérjékkel, akkor ennek alapján kísérletesen ellenőrizhető jóslatokat tehetünk az új fehérje biológiai aktivitására, szerkezet-funkció összefüggéseire vonatkozóan. Ez a tény rendkívül vonzóvá tette a gének és fehérjék homológiájának kimutatására szolgáló bioinformatikai módszereket, így ez a terület a bioinformatika legrégebben művelt, legfejlettebb része. Az egyszerű homológia-vizsgálat alapján azonban keveset tudunk meg arról, hogy az adott fehérje milyen más fehérjékkel kölcsönhatásban fejt ki hatását. Erre a kérdésekre az *összehasonlító genomika* nyújthat további felvilágosítást. Ennek a megközelítésnek az a megfigyelés az alapja, hogy az egymással kölcsönhatásban lévő, funkcionálisan egymásra utalt fehérjék (pl. hormonok és receptorai, anyagszállítók különböző komponensei) evolúciója összefügg egymással, így a fehérjék kapcsolatrendszerük tükröződik evolúciós történetükben. A fehérjék közt kialakuló fizikai kölcsönhatások predikciója azonban ma még kevésbé nélkülözheti a kísérletes vizsgálatokból (pl. kettős hibrid vizsgálatokból) gyűjtött adatokat.

A fehérjék közötti fizikai és funkcionális kapcsolatok ismeretének hiányában nehéz megjósolni, hogy mi történik egy adott gén kiesése (vagy funkciójának megváltozása) esetén. Ma már tömegszűrésre alkalmas módszerek (gén-kiütési technikák, a gén kifejeződésének antiszenz oligonukleotidokkal történő gátlása, RNS interferencia jelenségét kiaknázó technikák) állnak rendelkezésünkre, melyek segítségével szisztematikusan vizsgálni lehet, hogy mi az élettani következménye annak, ha egy-egy gén funkciója kiesik. A különböző gén-inaktiválási vizsgálatok tapasztalata megerősíti, hogy a genomika korszakában az egyes gének és

fehérjetermékek funkciójának jellemzése nem lehet teljes, ha nem fordítunk kellő figyelmet kapcsolathálózatokra.

Érdekes módon az emberi genomban található gének számával kapcsolatos kezdeti várakozások és meglepő tapasztalatok exponálták leginkább azt, hogy a gének közötti kapcsolathálózat eddig viszonylag kevés figyelmet kapott. Még a közelmúltban is voltak olyan becslések, melyek az emberi genomban található fehérje-kódoló gének számát 120 000-150 000-re tették (Liang és munkatársai, 2000; Smaglik, 2000). A Celera és a Human Genome Consortium által becsült emberi génszám (kb. 30 000) ehhez képest meglepően alacsonynak tűnt, ha figyelembe vesszük, hogy a *C. elegans* fonálféreg genomja 19 000 gént tartalmaz. Az ember és fonálféreg közötti jelentős komplexitásbeli különbséget a viszonylag csekély génszámbeli különbséggel összevetve ma már világos, hogy a komplexitás a génszámnak nem lineáris függvénye (Claverie, 2001). Nyilvánvaló tehát, hogy a biológiai komplexitás a gének és géntermékek közötti kapcsolati hálózat összetettségének, a génexpressziós program komplexitásának lehet az eredménye (Adami és munkatársai, 2000; Claverie, 2001).

Jelenleg még nem állnak rendelkezésünkre olyan bioinformatikai módszerek, melyek lehetővé tennék a gének közötti kapcsolatok, a génexpresszió programjának predikcióját pusztán a genomszekvencia számítógépes elemzése alapján. A megfelelő bioinformatikai módszerek kifejlesztésének a korábbiakban az is lényeges akadály volt, hogy nem álltak rendelkezésre megbízható kvantitatív adatok arra vonatkozóan, hogy az egyes gének kifejeződése hogyan változik sejttípustól, élettani állapottól, fejlődési stádiumtól függően. Ma már nagy teljesítőképességű módszerek (pl. DNS-chip technikák) segítségével követni lehet, hogy egy élettani állapotváltozás hogyan hat az egyes gének kifejeződésére, az egyes gé-

nek kifejeződésének változása milyen program szerint történik. Ezekkel a módszerekkel fontos információt nyerünk arra vonatkozóan, hogy melyek azok a gének, melyek szerephez jutnak egy adott élettani folyamatban. A funkcionális genomika és a bioinformatika egyik legizgalmasabb feladata a gének expresszióját szabályozó régiók szisztematikus azonosítása, működésük törvényszerűségeinek tisztázása, hogy a jövőben a genomszekvencia számítógépes elemzésével lehessen megjósolni a gének kapcsolat-hálózatait.

A funkcionális genomika az élettudományok valamennyi részterületének módszereit felhasználja, azok integrálásának eredményeként született meg. A genom-projektek lényegesen hozzájárultak az élettani tudományok vizsgálati módszereinek gyors fejlődéséhez, és ez a fejlődés a nem-genomikai szemléletű élettani tudományterületekre is jelentős hatást fejt ki. A funkcionális genomika paradigmaváltást idézett elő a gyógyszeriparban is: míg a huszadik század során a gyógyszerfejlesztésben a kémia tudománya játszott meghatározó szerepet, mára egyre inkább a funkcionális genomikai megközelítés válik dominánssá. Míg a gyógyszerkutatás „klasszikus” korszakában a gyógyhatású vegyület előállítás sokszor évtizedekkel megelőzte hatásmechanizmusának tisztázását, a gyógyszer célpontjának azonosítását, az új paradigmában ez fordítva van. A funkcionális genomikán alapuló gyógyszerfejlesztésnél előbb kerül sor a gyógyszer-célpont azonosítására, gyógyszer-

célpontként való hitelesítésére, és ezt követi azoknak a vegyületeknek a kifejlesztése, amelyek a célpontra hatva a kívánt hatással rendelkeznek.

Magyarországon eddig csak áttételesen és töredékesen érvényesült a genomkorszak hatása. A funkcionális genomika igényeinek köszönhetően kifejlesztett korszerű műszerek némelyike már megtalálható az egyes kutatóhelyeken (ezzel segítve a molekuláris biológiai, genetikai, fejlődésbiológiai, diagnosztikai vizsgálatokat), genomikai programnak azonban ez aligha nevezhető. Jóllehet a funkcionális genomika műveléséhez szükséges szakértelem az élettudományok legtöbb területén (bioinformatika, molekuláris biológia, genetika, sejtbiológia stb.) nálunk is adott, ezek nem szerveződtek egybe egy-egy hazai dominanciájú funkcionális genomikai projektbe. A Magyar Tudományos Akadémia *Tudománypolitika Magyarországon* sorozatában a napokban közzétett Biológia című kiadvány (Friedrich, 2001) megállapítása még ma is érvényesnek tűnik: „hazánkban (eltérően valamennyi fejlett országtól) nincsen semmiféle humán (vagy bármily más) genom program. Természetesen az erre épülő nagyobb szabású funkcionális genomikai kutatás és gyógyszerkutatás is hányzik, így az ország sem profitalhat közvetlenül az ilyen kutatások ipari hasznosításából. Meggyőződésünk, hogy a funkcionális genomikát az ország jövője szempontjából fontos területek közé kell sorolni, és megfelelő támogatásban kell részesíteni.”

IRODALOM

- Adami, C., Ofria, C. és Collier T. C. (2000) Evolution of Biological Complexity. *Proc. Natl. Acad. Sci. USA.*, 97, 4463-4468.
- Adams, M. D. és munkatársai (2000) The Genome Sequence of *Drosophila melanogaster*. *Science*, 287, 2185-95.
- Claverie, J. M. (2000) Gene Number. What if there are only 30,000 Human Genes? *Science*, 291, 1255-1257

- Das, M., Burge, C. B., Park, E., Colinas, J. and Pelletier, J. (2001) Assessment of the Total Number of Human Transcription Units. *Genomics* 77, 71-8.
- Friedrich, P. (2001) Biológia, Magyar Tudományos Akadémia, Tudománypolitika Magyarországon, II. A diszciplínák művelése, Budapest.
- Hogenesch, J. B., Ching, K. A., Batalov, S., és munkatársai (2001) A Comparison of the Celera and Ensembl Predicted Gene Sets Reveals Little Overlap in Novel Genes, *Cell*, 106, 413 - 415.

- Lander, E. S. és mtai (2001) Initial Sequencing and Analysis of the Human Genome. *Nature*, 409, 860-921.
- Liang, F., Holt, I., Pertea, G., és munkatársai (2000) Gene Index Analysis of the Human Genome Estimates Approximately 120,000 Genes. *Nat. Genet.* 25, 239-240.
- Smaglik, P. (2000) Researchers Take a Gamble on the Human Genome. *Nature*, 405, 264.
- The *Arabidopsis* Genome Initiative (1999) Analysis of the Genome Sequence of the Flowering Plant *Arabidopsis thaliana*. *Nature*, 408, 796-815.
- The *C. elegans* Sequencing Consortium (1998) Genome Sequence of the Nematode *C. elegans*. A Platform for Investigating Biology, *Science*, 282, 2012-2018.
- Venter, J.C. és munkatársai (2001) The Sequence of the Human Genome. *Science*, 291, 1304-1351.
- Wright, F. A., Lemon, W. J., Zhao, W. D., és munkatársai (2001) A Draft Annotation and Overview of the Human Genome. *Genome Biol.* 2 0025.1-0025.18

