

# Új tudományág az orvosi kutatás és klinikai gyakorlat középpontjában

## Gondolatok a bioinformatikáról

Falus András, Berke Tamás, Ortutay Csaba Péter, Suhai György

A NEW DISCIPLINE IN THE CENTER  
OF MEDICAL RESEARCH AND CLINICAL  
PRACTICE - THOUGHTS ON  
BIOINFORMATICS

Az orvosbiológiában forradalmi jelentőségű genomika egyik legfontosabb eleme a bioinformatika. A szerzők foglalkoznak a genomvariancia kérdéseivel és az összehasonlító genomika lehetőségeivel. Röviden szó esik az új gének azonosításáról, illetve a diagnosztikában kiemelt jelentőségű génexpressziós mintázatok elemző összehasonlításáról. A bioinformatika lehetővé teszi a patogének gyors azonosítását és különleges szerepet kap a molekuláris gyógyszertervezési eljárások között.

Bioinformatics is one of the most important element of genomic research revolutionising biomedical science. This review describes the phenomena of genomic variance and comparative genomics. Briefly, the review summarises the identification procedure of new genes and gene expression patterns highly important in diagnostics. Bioinformatic procedures make possible the rapid detection of pathogens and have principal role in molecular drug design technologies.

**genomika, bioinformatika, génlapka, SNP, génexpresszió, patogén, farmakogenomika**

**genomics, bioinformatics, gene chip, SNP, gene expression, pathogen, pharmacogenomics**

dr. Falus András (levelező szerző/correspondent): Semmelweis Egyetem, Genetikai, Sejt- és Immunbiológiai Intézet, Budapest/Semmelweis University, Department of Genetics, Cell- and Immunology; H-1089 Budapest, Nagyvárad tér 4.  
Berke Tamás: Genodia Molekuláris Diagnosztikai Kft., Budapest/Genodia Molecular Diagnostics  
Ortutay Csaba Péter: Mezőgazdasági Biotechnológiai Kutatóközpont, Bioinformatikai Kutatócsoport/  
Agricultural Biotechnology Center; Gödöllő  
Suhai György: Magyar Bioinformatikai Intézet Kft./Hungarian Bioinformatics Institute; Budapest

Érkezett: 2003. december 19. Elfogadva: 2004. január 13.

**A** genomika fogalmát viszonylag egyszerű meghatározni. Azonban amint különböző jelzőkkel illetjük a fogalmat – funkcionális genomika, orvosi genomika, immungenomika –, egyre részletesebb, speciális definíciót igényelhet a fogalommagyarázat. A genomika nem más, mint a gének (sok gén, akár az adott egyed összes génje) és azok funkcióinak tudománya. Ez magába foglalja a génszekvenciák meghatározását és a genomterképezést is.

Lényegéből fakadóan a bioinformatika egyidős a genomikával, hiszen létrejöttét a „biológiai tudományokban hirtelen és nagy intenzitással felgyülemlett információ” indokolta (1–3). Ez az információ pedig

nem más, mint genetikai információ, a genomok teljes vagy részleges meghatározásából származó nukleinsav és aminosav molekuláris szekvencia, a hozzá tartozó („annotált”) egyéb információval. A bioinformatika interdiszciplináris tudományág: a biológiai tudományok területén keletkezett információkat, adatokat korszerű informatikai – matematikai és számítástudományi – eszközökkel tárolják, dolgozzák fel, elemzik.

A klasszikus értelemben vett genomikához képest az orvosi genomika területe két szempontból is speciális. Egyrészt az emberi genom e tudományterület tárgya, tehát a humán genomika része, másrészt pedig olyan gének tanulmányozása a célja, amelyek emberi betegségek-

Az Orvosi genomika rovat szerkesztésében nyújtott segítségéért a szerkesztőség köszönetet mond dr. Falus Andrásnak.

kel hozhatók kapcsolatba, akár okozati tényezőként, akár betegségek terápiájának szóba jöhető eszközeként.

A bioinformatika – rendkívüli rugalmasságának köszönhetően – az orvosi genomikai kutatásokban is hamar igazolta létjogosultságát (4). A jól megfogalmazott informatikai igények kielégítésére specifikus bioinformatikai adathalmazokat, kereső- és elemző-rendszereket hozhattak létre. Az orvosi genomikai témák lényegéből fakadó intenzív változásokkal csak jól felkészült és az adott témában tájékozott szakember-gárda képes lépést tartani (5).

Régóta közismert, hogy egyes betegek jobban reagálnak a gyógyszeres beavatkozásra, mint mások, sőt, a hasonló betegségben szenvedő emberek közül egyeseknek még árthat is a másoknál hasznos terápiás beavatkozás. Miközben azokat a genetikai és molekuláris ösvényeket kutatjuk, amelyek a gyógyszeres beavatkozás hatékonyságát határozzák meg, a személyre szóló diagnózis és kezelés lehetősége is kezd elérhető közelségbe kerülni (6). E cél elérésének az a kulcsa, hogy egyre többet tudunk a humán genom változatosságáról, illetve a kezelésre való reagálást befolyásoló genetikai tényezőkről. A kutatás e területét farmakogenomikának hívják; a bioinformatika az a hajtóerő, amely elsősorban képes e szakterület fejlődését előrevinni. A bioinformatikának az a le-

egyszerűsített értelmezése, miszerint az a számítógépek, az internet, szoftverek és tárolt adatok együttes eszköze lenne a biomedicinálisan lényeges problémák megoldására, nem fejezi ki megfelelően azt a hatást, amit ezek a lehetőségek az orvostudományra gyakorolnak. A bioinformatika alkalmazása nélkülözhetetlen eleme nemcsak a farmakogenomikának, hanem általában a genomikának, a proteomikának, az újonnan felismert gének előrejelzésének, a fehérjék funkcionális predikciójának, illetve a kóros és normális szövetekben eltérő génkifejeződések kutatásának is. E technikák együttes használata ma már az orvosi kutatások legsikeresebb területének számít, és további fejlődésének is jelentős tényezője lesz a XXI. században. A bioinformatika gyökeresen új szemszögből foglalkozik az orvostudomány problémáival, és integrált megoldásokra nyújt reményt.

## Variációk a humán genomban: egyedi nukleotid-polimorfizmus

Ha személyre szabott gyógyszerkezelést kívánunk végezni a betegeknél, először is mérnünk és regisztrálnunk kell a populációkon belüli genetikai variációkat. Számítógép nélkül a funkcionálisan fontos genetikai variációk azonosítása rendkívül monoton és egyúttal lehetetlen volna, és a személyre szóló gyógyszerkezelés csak hosszú évek múlva lenne lehetséges. Elsősorban a Humán Genom Projekt következménye-

ként óriási számú SNP-t (single nucleotide polymorphism, egyedi nukleotid-polimorfizmus) – azaz egyetlen bázis stabil cseréjét, amely legalább egy populációban, 1%-nál nagyobb gyakorisággal fordul elő – írtak már le, így ez a sokféleség rendelkezésre áll az egyes emberek közötti különbségek megállapításához. Az SNP-k különösen akkor kapnak nagy jelentőséget, amikor a különböző betegek (ugyanazon) gyógyszeres kezelésre adott eltérő válaszait/reakcióit tanulmányozzák. Egy példa erre a tiopurintartalmú gyógyszerek alkalmazása sokféle betegség, többek között az akut lymphoblastos leukaemia és a krónikus gyulladáshoz vezető bélbetegség kezelésében. Ezeket a gyógyszereket a tiopurin-metiltransferáz enzim (TPMT) bontja le. Egy tanulmányban 24, akut lymphoblastos leukaemiában szenvedő és tiopurintartalmú gyógyszerrel kezelt gyermeket vizsgáltak; a gyermekek közül négyről megállapították, hogy a TPMT\*3A alléllra (az enzimet kódoló gén egyik példánya) nézve heterozigóta, amelyet két SNP jelenléte határoz meg; a 460. nukleotidpozícióban G (guanin) helyett A (adenin) van, a 719-es helyen pedig A helyett G található. Ez az allél a TPMT enzim hiányával áll kapcsolatban. A kezelés alatt mind a négy beteget életveszélyes májtoxikózis fenyegette, mivel szervezetükben a gyógyszer nem a tervezett ütemben, hanem annál jóval lassabban bomlott le. Így a terápiás gyógyszer-utánpótlás nem a szükséges gyógyszer szint fenntartását biztosította, hanem a szer toxikus szintet elérő felhalmozódását idézte elő. Ez a mérgezés elkerülhető, ha a betegeket szűrjük a TPMT-aktivitás szempontjából. A kis enzimaktivitású betegeknek akkor használnak a kezelést, amikor a tiopurintartalmú gyógyszer adagját csökkentik, míg a nagy enzimaktivitású betegeknél az emelt adag fejt ki hatását. Mára már a TPMT-aktivitás szűrése a klinika gyakorlat mindennapos módszerei közé tartozik. A beteg életének megmentése mellett a szűrés költségei bőven ellensúlyozzák a hosszabb kórházi tartózkodás költségeit (5).

Két évvel ezelőtt a Humán Genom Projekt keretében SNP-adatbázist hoztak létre és 1,4 millió SNP-t helyeztek el benne. Azóta ez az adatbázis megkétszereződött; ahogy a különböző populációk egyre több szekvenciaadata válik hozzáférhetővé, ez a szám valószínűleg nőni fog. Ezeknek az SNP-knek a többségét bioinformatikai algoritmusok segítségével tárták fel, amelyek a genom szekvenciáinak automatizált keresését teszik lehetővé. Ilyen megközelítés nélkül szinte lehetetlen volna alaposan megvizsgálni a humán genom szekvenciájának körülbelül 3 milliárd bázisának sorát. A bioinformatika abban tud segíteni, hogy kiszűrje azokat az SNP-eket, amelyeknél a legvalószínűbb, hogy gyógykezelési szempontból orvosilag érdekes következményt vonnak maguk után. Amennyiben az SNP-eket pontosan beillesztjük a humán genomba, meg tudjuk határozni, hogy melyikük okoz változást a fehérjék aminosavsorrendjében. A humán genom más, például egérgenommal való összehasonlítása során azokat az SNP-eket is azonosíthatjuk, amelyek a gének szabályozóterületein fordulnak elő.

A bioinformatika gyökeresen új szemszögből foglalkozik az orvostudomány problémáival, és integrált megoldásokra nyújt reményt.

A BIOINFORMATIKÁBAN HASZNÁLT KIFEJEZÉSEK

Kifejezés	Leírás
Bioinformatika	A biológiai információ tárolásával, előhívásával és elemzésével foglalkozó határtudomány.
BLAST	Bázissorrend-illesztés számítógépes eljárása. A homológia keresésére való elterjedt programrendszer.
cDNS	A messenger RNS-ből (mRNS) származtatott DNS-szekvencia.
EST	Expresszált szekvenciasorozat darabja az mRNS végéről.
Exon	Az az mRNS-t meghatározó DNS-szekvencia.
Filogenetika	A fajok és a bennük található gének közötti evolúciós kapcsolatok tudománya.
GenBank/ EMBL	Több millió egyedi szekvenciát és több milliárd DNS-bázispárt tartalmazó adatbázisok.
Genomika	Nagyszámú gén feladatát/funkcióját, kifejeződését/expresszióját, illetve kapcsolatát vizsgáló tudomány.
Homológia	Az az elképzelés/fogalom, hogy két gén közös őstől származik, illetve funkcionális szerepük azonos.
Intron	A gének az a része, amely az mRNS termelődése során kivágásra kerül.
Microarray/génlapka	Nagyon sok gén szimultán/egyidejű expressziója/kifejeződése, génlapnak (csipnek) is nevezik.
mRNS	Az exonokból álló nukleinsav-szekvencia (-sorrend), amely a transláció során fehérjévé „fordítódik” le.
Molekuláris epidemiológia	Járványtani vizsgálatok kivitelezése; a kórokozók genetikai állományának összehasonlító elemzésén alapul.
ORF	Nyitott leolvasókeret. Lényegében egy fehérjekódoló DNS-szekvencia (-sorrend).
Proteom	Egy szervezet, sejt stb. teljes fehérjemennyisége.
PSI-BLAST	Helyspecifikus BLAST. Távoli rokonságban lévő gének közötti homológia keresés programja.
RT-PCR	Reverz transzkriptáz-polimeráz lánreakció. A specifikus/fajlagos RNS-sokféleség szemikvantitatív mérése.
SNP	Nukleotid-polimorfizmus. Egy DNS-bázispár, amely a populáción belül eltér.
Transzkriptom	Egy sejtben, szervezetben stb. expresszált teljes mRNS-mennyiség.

**Összehasonlító genomika:  
következtetés a humán gének  
funkciójára homológia segítségével**

A különböző organizmusokban azonos funkciót ellátó és közös evolúciós eredetű fehérjéket homológoknak nevezik. Az ugyanazon információt kódoló génszekvenciákat is homológ szekvenciáknak nevezzük, és tárgyat képezik az összehasonlító genomikának (7).

Az összes állatfajhoz képest a *homo sapiens* igen nehéz genetikailag tanulmányozni. A hosszú generációs idő, a kis család (méret), az etnikai és etikai korlátok mind azt jelentik, hogy más fajokon könnyebb megismerni a gének funkcionális jelentőségét. Szerencsére az a tény, hogy a szekvenciák funkcionálisan fontos részei az evolúciós idő során megőrződtek, lehetővé teszi, hogy a homológia keresés módszerével a más fajokban fontos gének megfelelőjét az embernél is megtaláljuk. A homológia – vagy technikailag hasonlóság – keresése a bioinformatika egyik központi eszköze. Ha távolabbi rokonságban levő szekvenciákat keresünk, a homológia keresés jelenlegi egyik legalkalmasabb eszköze

a BLAST (basic local alignment similarity tool) vagy a PSI-BLAST program. Mindkét program hozzáférhető a Biotechnológiai Információ Nemzeti Központjának (National Center for Biotechnology Information) webhelyén – [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) –, és lehetővé teszi a keresést a legnagyobb nemzetközi szekvencia- (és bioinformatikai) adatbázisban (GenBank) (referencia: BLAST: Nucleic Acids Research). A homológia keresésnek a következő a logikája: ha a szekvenciában hasonló a párok lineáris illeszkedési sorrendje, arra következtethetünk, hogy a háromdimenziós szerkezet is azonos; ha ez a helyzet, a funkció is jó eséllyel hasonló lesz. Ha azt akarjuk megtudni, hogy konkrét fehérjék vagy biokémiai rendszerek – amelyek például az egérben, az ecetmuslicában-*Drosophilában* vagy az *Escherichia coliban* ismertek – jelen vannak-e az emberben, ma már e szekvenciák egyenként is összehasonlíthatók a jelenleg ismert 25-30 ezer emberi fehérjét kódoló génnel. Ez az „összehasonlító és szembeállító” módszer kulcsfontosságú számos emberi fehérje funkciója meghatározásának folyamatában, mert a hagyományos laboratóriumi technikákkal végzett meghatározás évekbe telne.

Ha különböző fajok homológ szekvenciáit sorba állítjuk, azonosítani tudjuk azokat a konzervált domén-szerkezeteket/struktúrákat, amelyek egy-egy konkrét fehérje működésére utalnak. Hasonlóképpen, különböző fajok homológ génje (5' irányban található) upstream régióinak összehasonlításával leírhatók lesznek a promoterrégiók, és azonosíthatók a potenciális transzkripciósfaktor-kötő helyek. Egérben, kutyában és emberben az emlőrák kialakulásában jelentőséget tulajdonított BRCA1 gén upstream régióinak összehasonlításával azonosították az Sp1, CREB és NF- $\kappa$ B transzkripciósfaktorok valószínű kötőhelyeit, bepillantást nyújtva abba, hogy ez a gén hogyan szabályozódik, mielőtt bármiféle valódi laboratóriumi kísérletet véghezvittek volna. Ez a munka is a „számítógép előtti”, *in silico* kutatás hatalmas távlataira utal; ennek kiváló hazai példája egy ortológ promotereket tartalmazó adatbázis, amelyben összehasonlíthatjuk különböző fajokban a hasonló szerkezetű promoterekkel rendelkező géneket, és ezzel új szabályozási régiókat kereshetünk.

## Új gének azonosítása

A bioinformatika fegyvertárának hathatós eszköze az orvosilag fontos génszekvenciák felfedezésének folyamata. Ahogy a kórházakban az antibiotikum-rezisztens baktériumtörzsek komoly fenyegetést jelentenek, óriási

a kényszer, hogy új antibakteriális anyagokat fedezzenek fel. Különösen nagy érdeklődést váltottak ki az antibakteriális peptidek, például a  $\beta$ -defenzinek; ezek természetes elemei a sok fajban megtalálható veleszületett immunitásnak. Az utóbbi időben a humán genom érzékeny bioinformatikai szűrésével a  $\beta$ -defenzin család 28 új tagját fedezték fel, miután a standard BLAST hasonlóságkeresés csak egy új defenzinszekvenciát tárt fel. Számos helyen végeznek kifinomult bioinformatikai technikákat abból a célból, hogy a humán genomban új antibakteriális géneket fedezzenek fel.

A kórokozók azonosítására alkalmazott gyorsabb, bioinformatika-alapú módszerek kimutatható klinikai előnnyel járhatnak.

## A génextpresszió mintázata mint diagnosztikai és prognosztikai eszköz

Diagnosztikai és prognosztikai eszközként gyakran vizsgálják a gének expressziójának/kifejeződésének különbségét (8). Az utóbbi időben a bioinformatika az új markerek azonosításának létfontosságú technológiájává vált, amellyel megjósolható egy betegség kimenetele. Egy vastagbélrákról szóló tanulmányban az Unigén adatbázist kutatták át olyan géneket keresve, amelyeknek több expresszált szekvenciamotívuma (EST, expressed sequence tag) állt kapcsolatban a rákos, mint a normális vastagbéliszövetrel. Harmincöt

jelölt nukleotidszakasz közül négyről mutatták ki, hogy jelentős mértékű, szignifikáns korrelációt mutat a hisztopatológiai paraméterekkel. Kvantitatív RT-PCR-rel (reverz transzkriptáz-polimeráz láncreakció) tovább jellemezték a kifejeződés mértékét; kimutatták, hogy a négy génből három gén expressziós (kifejeződési) szintje nagymértékben összefügg a rossz túlélési aránnyal, illetve mind a négy gén kapcsolatban állt metasztatikus betegséggel. Hasonló módszerek más betegségek esetén (is) könnyen alkalmazhatónak tűnnek.

EST-adatokból azonban nagyon nehéz mennyiségi és statisztikusan érvényes expressziós mintázatot kimutatni. Ezzel ellentétben az új betegségmarkerek azonosításában a DNS-génlapok (csipek) – magyarul génlapok – sokkal hatásosabb eszköznek bizonyulnak, bár ezeknél is jelentkeznek pontossággal és reprodukálhatósággal kapcsolatos problémák. A DNS-lapok lehetővé teszik a génextpressziós szintek elemzését jelzett DNS (akár mRNS-ből reverz transzkriptázzal szintetizált cDNS) felhasználásával. Ezeket olyan génlapokra viszik, amelyekeken üveg-, szilikon- vagy nejlonszerű felületen cDNS-ek vagy oligonukleotidok ezrei-tízezrei rendezett formában található (innen az angolban microarray-nek nevezett technológia neve). Egy pontban felvitt DNS-minta egy génnel képes kölcsönhatni. A hibridizálódás mértéke a génextpresszió szintjéről ad jelet. A bioinformatikai eszközök nélkülözhetetlenek a génlap- (csip/microarray) adatok elemzésében, mivel minden kísérletből egyedi adatpontok óriási mennyisége származik, és ezeket értelmezhető módon csoportosítani és elemezni kell. Ugyanilyen fontos azonban a génlapkísérletek tervezése – tehát a DNS-minták elhelyezési mintázatának kialakítása – közben a statisztikai megfontolások figyelembevétele. Ezek azonban rendkívül drágák, a kutatók költségvetése pedig korlátozott, az eredmények értékét viszont ronthatja a minták elégtelen számú (meg)ismétlése. Nincs az a mennyiségű multivariáns eljárást követő elemzés, amely javítana ezen a problémán.

E technológia fontosságát szemléltette nemrég a génextpressziós mintázat kutatása klinikai helyzetben, emlőkarcinómákban. Hierarchikus klaszterképzés (<http://genome-www4.stanford.edu/Microarray/-SDM/restech.html>) néven ismert csoportosítási módszerrel elemezték 85 génlapkísérletet, amely 78(-féle) tumort, három fibroadenómát és négy normális emlőszövetmintát képviselt. Ezzel a módszerrel a rákokat a következő csoportokba lehetett osztani: alaphám/basalis epithelszerű csoport, ERBB2-túlexpresszáldó csoport, illetve normális csoport. A túlélési analízis szignifikánsan különböző kimenetelt tárt fel a különböző csoportokba tartozó betegek számára, bizonyítva, hogy ez a technológia hatékony eszköz lesz a jövőben arra, hogy kiegészítse és gyarapítsa az emlőrák-osztályozásnak relatíve kevés hagyományos prognosztikai és prediktív markerét, így javulva a legmegfelelőbb klinikai kezelést. Mielőtt ez a technológia a mindennapi klinikai gyakorlatba kerülne, ezt és sok más kutatást meg kell majd ismételni a világ más központjaiban. Továbbá a tudósok és

## 1. TÁBLÁZAT

*Bioinformatika az interneten**Genomikai kutatóközpontok:*

- EBI: European Bioinformatics Institute in Hinxton, UK. ([www.ebi.ac.uk](http://www.ebi.ac.uk))
- EMBL: European Molecular Biology Laboratory. ([www.embl-heidelberg.de](http://www.embl-heidelberg.de))
- Sanger Centre: A major UK genome research centre. ([www.sanger.ac.uk](http://www.sanger.ac.uk))
- NCBI: National Center for Biotechnology Information USA. ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov))
- TIGR: The Institute for Genomic Research, USA. ([www.tigr.org](http://www.tigr.org))
- HGSC: Human Genome Sequencing Center at the Baylor College of Medicine. ([www.hgsc.bcm.tmc.edu](http://www.hgsc.bcm.tmc.edu))
- GVPN/Genethon: Centre de Recherche sur le Génome Humain. ([www.genethon.fr/php/index.php](http://www.genethon.fr/php/index.php))
- ExPASy: at the University of Geneva. ([us.expasy.org](http://us.expasy.org))
- HGMP: Human Genome Mapping Project, UK/MRC. ([www.hgmp.mrc.ac.uk](http://www.hgmp.mrc.ac.uk))
- The Jackson Laboratory. Scope: mammalian genetics. ([www.jax.org](http://www.jax.org))
- GenomeNet, Japan. Includes the KEGG metabolic database. ([www.genome.ad.jp](http://www.genome.ad.jp))
- BiSC: Bioinformatics Supercomputing Centre in Toronto. ([www.bioinfo.sickkids.on.ca](http://www.bioinfo.sickkids.on.ca))
- NCGR: National Center for Genome Resources, USA. ([www.ncgr.org](http://www.ncgr.org))
- SHGC: Stanford Human Genome Center. (<http://shgc.stanford.edu>)
- ANGIS: Australian National Genomic Information Service. ([www.angis.org.au/new](http://www.angis.org.au/new))
- Weizmann Institute, Israel. Bioinformatics and Biological Computing. (<http://bip.weizmann.ac.il>)

*Adatbázisok:*

- MGI: Mouse Genome Informatics. ([www.informatics.jax.org](http://www.informatics.jax.org))
- GenBank: Nemzetközi Genomikai Adatbázis. ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov))
- FSSP: Families of Structurally Similar Proteins. ([www.bioinfo.biocenter.helsinki.fi:8080/dali](http://www.bioinfo.biocenter.helsinki.fi:8080/dali))
- ISCB: International Society of Computational Biology. ([www.icsb.org](http://www.icsb.org))
- PDB: Protein Database, 3D and structure. Makromolekulák kísérletesen meghatározott háromdimenziós térszerkezetének adatbázisa. ([www.rcsb.org/pdb](http://www.rcsb.org/pdb))
- TRANSFAC: The transcription factor database. ([www.gene-regulation.com](http://www.gene-regulation.com))
- CSNDB: the Cell Signalling Networks Database. (<http://geo.nihs.go.jp/csndb>)
- MIPS: Munich Information Centre for Protein Sequences. ([www.mips.biochem.mpg.de](http://www.mips.biochem.mpg.de))
- Bodymap: Emberi és egér génexpressziós adatbázis, Tokyo. (<http://bodymap.ims.u-tokyo.ac.jp>)
- SHIGEN: SHared Information of GENetic resources, Japan. ([www.shigen.nig.ac.jp](http://www.shigen.nig.ac.jp))
- Kabat Antibodies – struktúrák és szekvenciák. ([www.rubic.rdg.ac.uk/%7Eandrew/bioinf.org/abs](http://www.rubic.rdg.ac.uk/%7Eandrew/bioinf.org/abs))
- ProDom: Protein Domain adatbázis. (<http://protein.toulouse.inra.fr/prodom/current/html/home.php>)
- GOLD: Genomes OnLine Database. (<http://wit.integratedgenomics.com/GOLD>)
- Cytokines: Információs központ a citokinekről (receptorai és azok háromdimenziós struktúrái) ([http://cmbi.bjmu.edu.cn/cmbidata/cgf/CGF\\_Database/cytweb](http://cmbi.bjmu.edu.cn/cmbidata/cgf/CGF_Database/cytweb))
- DOGS: Database Of Genome Sizes. ([www.cbs.dtu.dk/databases/DOGS](http://www.cbs.dtu.dk/databases/DOGS))
- Small RNA BCM adatbázis. (<http://mbr.bcm.tmc.edu/smallRNA>)

*Egyedi nukleotid polimorfizmus:*

- The SNP consortium. (<http://snp.cshl.org>)
- SIB: 21-es emberi kromoszóma cSNP (cDNA-n belüli SNP) adatbázisa. (<http://csnp.isb-sib.ch>)
- NCBI SNP adatbázis. ([www.ncbi.nlm.nih.gov/SNP](http://www.ncbi.nlm.nih.gov/SNP))
- Utah SNPs adatbázis. ([www.genome.utah.edu/genesnps](http://www.genome.utah.edu/genesnps))

*Alkalmazások, segédprogramok:*

- AAT: Analyzing and Annotating Tools. Génkeresés genomsekvenciákban. (<http://genome.cs.mtu.edu/aat.html>)
- Autogene integrált predikciós segédlet. ([www.bionet.nsc.ru/SRCG/AUG](http://www.bionet.nsc.ru/SRCG/AUG))
- BioEdit: Szekvenciaillesztéseket szerkesztő és analizáló program. ([www.mbio.ncsu.edu/BioEdit](http://www.mbio.ncsu.edu/BioEdit))
- CASSANDRA: Kódolószegmens felismerésére. (<http://hto-b.usc.edu/software/procrustes/cassandra>)
- CDS: Kódolórégiók keresésére. (<http://bioweb.pasteur.fr/seqanal/interfaces/cds-simple.html>)
- ClustalW: Szekvenciaillesztő program. ([www.ebi.ac.uk/clustalw](http://www.ebi.ac.uk/clustalw))
- Diogenes: Rövid genomsekvenciákban ORF keresése. ([www.cbc.umn.edu/diogenes](http://www.cbc.umn.edu/diogenes))
- DoubleScan: Fehérjekódoló gének összehasonlító *ab initio* predikciója egér és emberi DNS-ekben. ([www.sanger.ac.uk/Software/analysis/doublescan](http://www.sanger.ac.uk/Software/analysis/doublescan))
- GENEUIDER: Integrált számítási rendszer fehérjekódoló gének meghatározására. ([www.itba.mi.cnr.it/webgene](http://www.itba.mi.cnr.it/webgene))
- GeneDoc: Windows alapú szekvenciaanalizáló program. ([www.psc.edu/biomed/genedoc](http://www.psc.edu/biomed/genedoc))
- GeneID: Mesterséges intelligencia rendszer gerincesek génszerkezetének analizálására. (<http://igs-server.cnrs-mrs.fr/igs/banbury/genid-email.html>)
- IMGT: ImMunoGeneTics információs rendszer. (<http://imgt.cines.fr>)
- MetaGene: Egyszerre több génpredikciós eljárással vizsgálhat szimultán egy gént. (<http://rgd.mcw.edu/METAGENE>)
- ORPHEUS: szoftvercsomag génpredikcióra komplett baktériumgenomokban. (<http://pedant.gsf.de/orpheus>)
- PHYLIP: Filogenetikai programcsomag. ([www.evolution.genetics.washington.edu/phyliip.html](http://www.evolution.genetics.washington.edu/phyliip.html))
- PROMOTER: Átírás kezdőhelyeinek meghatározása gerincesek DNS-ében. ([www.cbs.dtu.dk/services/promoter](http://www.cbs.dtu.dk/services/promoter))
- SorFind: Véltegyedi kódolóexonok azonosítása és annotálása. ([www.rabbithutch.com](http://www.rabbithutch.com))
- tRNAscan: tRNA gének keresése a genomsekvenciákban. ([www.genetics.wustl.edu/eddy/tRNAscan-SE](http://www.genetics.wustl.edu/eddy/tRNAscan-SE))
- TWINSKAN: Adott szekvencia összehasonlító elemzése más organizmusok homológ szekvenciáival. (<http://genes.cs.wustl.edu>)
- Veil: Hangolható rejtett Markov-modell (HMM) gének keresésére eukarióták DNS-ében. ([www.tigr.org/%7Eesalzberg/veil.html](http://www.tigr.org/%7Eesalzberg/veil.html))
- Yeastgene: Egy ORF- vagy DNS-szekvencia-töredék YZ értékének kiszámítása *Saccharomyces cerevisiae* génjében. (<http://tubic.tju.edu.cn/cgi-bin/Yeastgene.cgi>)

*Folyóirat-gyűjtemények:*

- Nucleic Acids Research kutatási eredményei (<http://nar.oupjournals.org>)
- ENTREZ: MEDLINE és szekvencia adatbázisok együtt. ([www.ncbi.nlm.nih.gov/entrez](http://www.ncbi.nlm.nih.gov/entrez))
- PubMed Central: Tudományos folyóiratok gyűjteménye. ([www.pubmedcentral.nih.gov](http://www.pubmedcentral.nih.gov))
- MedMiner: GeneCards és PubMed együtt. (<http://discover.nci.nih.gov/textmining/filters.html>)

klinikusok képzése ezen a területen szintén kiemelkedő fontosságú lesz, mivel a génlapkísérletek tervezése és kivitelezése sok bonyolult tényezőtől áll (9).

## Patogének azonosítása

Bár a patogének azonosításában széleskörűen alkalmazzák a hagyományos mikrobiológiai módszereket, ezek lassúak, és nem mindig képesek detektálni az olyan mikrobákat, amelyek bonyolult tenyésztést igényelnek. Így a kórokozó azonosításáig a beteg specifikus kezelés nélkül marad. A kórokozók azonosítására alkalmazott gyorsabb, bioinformatika-alapú módszerek kimutatható klinikai előnnyel járhatnak. A tömegspektrometria proteomadatbázis-keresési algoritmusokkal párosítva közvetlen technikát nyújt a mikroorganizmusok azonosításához. Ezt a módszert sikeresen alkalmazták a *Helicobacter pylori*-baktériumtörzsek felismerésénél. A bioinformatika egyik alappilléret képviselő filogenetikai módszerek arra is felhasználhatók, hogy megalkossuk a szóban forgó kórokozó törzs „családfáját”. Ezáltal nyomon tudjuk követni földrajzi és/vagy időbeni terjedésüket helyi vagy akár világméretben is. Ilyen molekuláris epidemiológiai elemzéseket többször is végeztek már a HIV-vírus esetében. Filogenetikai összehasonlító elemzésekkel mutatták ki például egy ízben, hogy egy floridai fogorvos volt a HIV-fertőzés közös forrása látszólag egymással kapcsolatban nem álló betegek egy csoportjában. Egy másik esetben a bioinformatikai rendszerek segítségével sikerült kapcsolatba hozni egy észak-amerikai ételfertőzést a közel egy időben, Skandináviában zajlott, hasonló jellegű ételfertőzéssel. A molekuláris epidemiológiai vizsgálatok eredményeivel tudták a szakemberek a fertőzés forrását (szamóca) visszavezetni szerbiai eperföldekre. A génlapkatechnika és a bioinformatikai elemzés a fentiek mellett lehetővé teszi a különböző antibiotikum-rezisztenciájú törzsek gyors diagnosztikáját, illetve hagyományos mikrobiológiai eszközökkel nem elkülöníthető törzsek megkülönböztetését.

A bioinformatikának a mindennapi orvosi kezelések meghatározásában játszott szerepére egy másik, rendkívül hatásos példa a HIV-fertőzött betegek hatékony gyógyszerterápiájának megválasztása. Ez úgy történik, hogy molekuláris biológiai módszerekkel (RT-PCR és nukleinsav-szekvenálás) meghatározzák a betegben aktuálisan szaporodó vírustörzs gyógyszer-rezisztenciáját determináló géneket. Az így kapott genotípus-mintázatot – megfelelő algoritmusok segítségével – referencia-adatbázisban, ismert rezisztenciát eredményező szekvenciamintázatokkal hasonlítják össze. A klaszteranalízis eredményeként valószínűsíthető az adott vírustörzs különböző gyógyszer-kombinációkkal szemben kifejeződő ellenálló képessége. Ennek ismeretében a „próba szerencse” alapokra helyezett gyógyszerelés helyett célzott kezelés alkalmazható.

## Gyógyszertervezés a fehérjék térszerkezetének ismeretében

A bioinformatika önálló, igen dinamikusan fejlődő területe a strukturális bioinformatika; ez a fehérjék szerkezetének meghatározásával, a térszerkezet kialakulásának szabályszerűségével, a feltekeredés (folding) szimulálásával foglalkozik. Alapgondolata a következő: ha megértjük, hogy a fehérjék esetében az aminosav-szekvenciából hogyan következik a térszerkezet és abból a funkció, akkor képesek lehetünk új funkciójú enzimek kifejlesztésére, a hibás – betegséget okozó – enzimek kijavítására, a túlzott aktivitást mutató enzimek célzott kikapcsolására.

Mindezekhez kapcsolódóan csodálatos eredményeket láthatunk a gyógyszermolekulák tervezésében (rational drug design): a legismertebb példa egy inhibítormolekula kifejlesztése volt, ezt a HIV egyik proteázának blokkolására tervezték, az enzim háromdimenziós térszerkezetének ismeretében. A gyógyszerkutató cégek egyre gyakrabban alkalmazzák a bioinformatika módszereit azoknak a kombinációs hatóanyagkönyvtáraknak a megtervezésére, amellyel irányítottan kapcsolhatók ki enzimek. Ismert példa egy, a hippocampusban található ionsatorna célzott blokkolása ilyen módszerrel kifejlesztett gyógyszerrel. Ezzel kézzelfogható közelségbe került például az epilepsziás tünetek hosszú távú kezelése.

## Következtetések

Hippokratész óta *in vivo* technikákra támaszkodtak az orvosok a medicina problémáinak megfejtésében.

Pontosan 50 éve – a DNS-szerkezet tisztázásával – kezdődött a molekuláris biológia *in vitro* módszereinek alkalmazása, és ezek óriási lendületet adtak az orvosi kutatásnak. A laboratóriumi tudomány azonban közismerten lassú, munkaigényes és drága. Nem vitás, hogy a bioinformatika, amely számítógép-szimulációs eszközöket, elemző és kereső algoritmusokat, illetve nagyméretű adatbankokat használ, egyre több időt és energiát fog megtakarítani a laboratóriumban. S ami még fontosabb: a bioinformatika radikálisan új szemlélettel tekint az orvostudományban felmerülő kérdésekre; új módszereket nyújt nagyszámú, eltérő, de orvosiilag lényeges adat egyfajta komplex folyamattá alakításában, és integrált, rendezett megoldásokra ad reményt. Az elmúlt 3-4 évben számos kitűnő, szabadon hozzáférhető, felhasználóbarát hálóhelyet hoztak létre a világhálón (1. táblázat). Ezek egyenértékűek a molekuláris biológiai kitekkel, amelyek sokkal hatékonyabbá teszik a kutatást. Ugyanúgy, mint a kitek esetében is, fontos értenünk, hogy hogyan működnek ezek a bioinformatikai eszközök, és hogyan módosíthatók úgy, hogy a genetikai protokoll konkrét tudományos problémához igazodjon. Akik ezeket a rendszereket mágikus feketedobozokként kezelik, és nem néznek az alapbeállítás mögé, szinte biztosan hasznos információról maradnak le. Bár a bioinformatika biz-

tosítja az eszközöket, gondolkodnia a tudósak és a klinikusokak kell, a legtöbb esetben szorosan együtt. Akik azonban nem vesznek tudomást a bioin-

formatikáról, elkerülhetetlenül lemaradnak az orvosi kutatás, orvosbiológiai tudás, és végső soron a betegellátás élvonalától.

## IRODALOM

1. Barnes M, Gray I (eds.). Bioinformatics for geneticists. London: John Wiley & Sons; 2003.
2. Searls DB. The language of genes. *Nature* 2002;420:211-7.
3. Searls DB. Bioinformatics tools for whole genome. *Ann Rev Genomics and Human Genetics* 2000;1:251-79.
4. Bayat A. Bioinformatics. *Br Med J* 2002;324:1018-22.
5. Campbell AM, Heyer LJ. Discovering genomics, proteomics and bioinformatics. San Francisco: Benjamin Cummings; 2003. (Magyar fordítása hamarosan megjelenik a Medicina Könyvkiadó gondozásában.)
6. Lengauer T (ed.). Bioinformatics – from genomes to drugs. Weinheim: Wiley-VCH; 2002. Vols. I and II.
7. O'Brien SJ, et al. The promise of comparative genomics in mammals. *Science* 1999;286:458-62.
8. Sauter G, Simon R. Predictive molecular pathology. *New England J Med* 2002;347:1995-6.
9. Lynn David J, Lloyd AT, O'Farrelly C. Bioinformatics: implications for medical research and clinical practice. *Med Clin Exp* 2003;26:70-4.



## EGY TÖRTÉNETE MINDENKINEK VAN!

### EGY KIS HAZAI

Az ember szereti pozitív érzelmeit tárgyakkal, ajándékkal kifejezni. Ezek az érzések könnyen felébrednek, ha bajból, betegségből szabadulunk meg.

Milyen jó lenne, ha az orvos-beteg kapcsolat ezen megnyilvánulásai sohasem torzulnának! A téma most fókuszba került, és a média ennek minden formáját a hálapénz kifejezéssel illeti és az egyértelműen negatív fogalmak kategóriájába sorolja. A következő történet majd' harmincéves és a dolog más aspektusát világítja meg.

A sors kegyes volt hozzám, mert a belgyógyászat tanulását kiváló tanítómester irányításával működő orvoscsapat tagjaként kezdeztem. Tisztelve szeretett főnöküktől nem csupán a nagyviziteken tanultunk, hanem a szakmai ismereteken túl meríthettünk kulturális élményeiből, leckét kaphattunk történelemből és a napi politikából is.

A kollektíva tehát „jól együtt volt”, így a legnehezebb feladatokkal is könnyebben tudtunk megbirkózni. A munkánkkal járó nagy feszültséget időnként egymás cukkolásával igyekeztünk oldani.

Csapatunk alig ismerte a hálapénz fogalmát, ellenben a fiatal orvosok gyakorta kaptak „egy kis hazait”. Történt egyszer, hogy valaki felröppentette a hírt: ifjú kollégánk bizony semminek nem tud úgy örülni, mint a friss tojásnak. A gondjaira bízott három kórteremben dolgozó nővérek is be voltak avatva, és váltig állították: a doktor úr mást nem is fogad el, ne is próbálkozzon senki semmivel. Gyűlt is az adomány száz számra, s csak akkor fogott gyanút kollégánk, amikor már szinte lépni sem lehetett a dobozhegyektől a közös orvosi szobában. Az akciót azonban alig tudtuk leállítani. A nyolcágas kórtermek betegek szájáról szájra adták tovább a hírt az újaknak. E hiedelem macacsnak bizonyult, a betegek szívósan ragaszkodtak hozzá, és még hetek múlva is szíves mosollyal és kis tojásajándékkal köszönték meg a fáradozását.

A továbbiakat illetően az „ötletbörzén” elhangzott javaslatok közül az alábbi valósult meg. Mielőtt a tojásokkal „újjazdaggá” tett orvoskolléga megkezdte volna az adományok hazaszállítását, illetve szétosztását a rokon, baráti körben, hogy ne maradjon újabb tréfa nélkül, egyikünk a tojások közül többet megfőzött, majd az ily módon preparált tojásokat visszahelyezte a többi közé. Már csak azt sajnáltuk, hogy nem láthattuk a nyilvánvalóan meglepődött és tanácstalan felhasználók arcát. A történet végül is mosollyal zárult, miután az ok-okozati összefüggésekre fény derült.

Mindenki jól járt, mert a tréfa és játék az egészséges élet elixírje. Sajnos, manapság ez is egyre inkább hiánycikk! Csak a nosztalgia maradt iránta eleven.

**dr. Nemesánszky Elemér**