

# A Humán genom projekt

Sasvári-Székely Mária

A Humán genom projekt fő célja a teljes emberi DNS-állomány szerkezetének feltárása, ami közvetlenül hárommilliárd bázispár szekvenciájának meghatározását jelenti. A program 1990-ben indult el, több ország állami támogatású kutatóintézeiben, az eredményeket mindenki számára hozzáférhetően tették közre a világhálón. 1998-ban indult el a Celera privát szektor hasonló célú programja. A két projekt 2001-ben, összehangoltan jelentette be a humán genom első, nyers példányának elkészítését, és a két projekt kutatói a Nature, illetve a Science egy teljes számában taglalták az első eredményekből levonható következtetéseket. Az eredmények azt mutatták, hogy a genetikai információnak mindössze igen kis része (kevesebb mint 1%-a) íródik át fehérjékre. A gének becsült száma meglepően alacsony (30-40 ezer; ez a szám a későbbiekben 40-60 ezerre nőtt). Mindezek alapján az emberi komplexitás gyökerét feltehetően nem a gének számában, hanem az információs egységek (domének) variabilitásában kell keresnünk. A humán genom teljes szekvenciájának megismerésével lezárul a „pregenomiális” korszak, és elkezdődik a „posztgenomiális” időszak, amelyben a genetikai információ értelmezése és a gének funkciójának megismerése mellett az emberek közti genetikai variabilitás feltérképezése is nagy hangsúlyt kap.

**humán genom projekt,  
genetikai polimorfizmusok, SNP, VNTR,  
humán gének, összefoglaló**

## THE HUMAN GENOME PROJECT

The main goal of the Human Genom Project is to reveal the complete structure of the human DNA, involving the sequence determination of the three billion basepairs. The program was launched in 1990, executed in laboratories of the public sector in several countries. Results of this project are made public on the Internet. In 1998 a private sector company, Celera, was founded with the same goal. In 2001 both projects announced the accomplishment of the first draft of the Human Genome. A whole issue of Nature and Science magazine was devoted to present the first results of their findings. According to the results less than one percent of the genetic information is transcribed to proteins. The estimated number of genes was also lower than expected (30-40 thousand, which number recently increased to 40-60 thousand). Based on these findings, human complexity should be ascribed to the variability of the information units (domains) rather than to the number of genes. Once the complete sequence of the humane genome is accomplished, the 'pre-genomic era' will be over and the 'post-genomic era' will begin, dealing with understanding the genetic information, annotating the genes and mapping the genetic variability of people.

**Human Genom Project,  
genetic polymorphism, SNP, VNTR,  
human genes, review**

dr. Sasvári-Székely Mária (levelezési cím/correspondent): Semmelweis Egyetem,  
Általános Orvostudományi Kar, Orvosi Vegytani, Molekuláris Biológiai és Patobiokémiai Intézet/  
Semmelweis University, Faculty of Medicine, Department of Medical Chemistry,  
Molecular Biology, and Pathobiochemistry; H-1444 Budapest, Pf. 260.  
E-mail: sas@puskin.sote.hu

Érkezett: 2002. december 11. Elfogadva: 2003. január 22.

**A** molekuláris biológia első nagy felfedezéseinek egyike a DNS szerkezetének megfejtése, a Watson- és Crick-féle kettős spirál modelljének megalkotása. Mi van megírva génjeinkben? Erre szeretnénk választ kapni.

## A Humán genom projekt

### Hárommilliárd betű megismerése

A Humán genom projekt (HGP) fő célkitűzése, hogy megismerjük e rendkívül hosszú molekulák teljes szerkezetét, ez egyben az ember öröklött információinak dekódolását jelenti. Az ember öröksége, a genom könyve hárommilliárd betűből áll, amely – ha minden oldalra ezer betűt nyomtatnának – háromezer, egyenként ezeroldalas könyv formájában jelenhetne meg. A könyv szövege nagyon sajátos lenne, hiszen a betűkészlet összesen négy jeltől áll: A, G, C és T (adenin, guanin, citozin, timin). Ezek a betűk ismétlődnek a milliányi oldalon, meghatározott sorrendben. Az első cél tehát ennek a hárommilliárd betűből álló „szövegnek” a megismerése, az emberi kromoszómák DNS-szekvenciájának a megfejtése. Ez volt a Humán genom projekt legfontosabb feladata, amely 2001-re 96%-ban teljesült (1). A DNS-szekvencia megismerésével lezárul a „pregenomikus” korszak, és az eredmény, ha nem is könyv formában, de internetes adatbázisokban mindenki rendelkezésére áll. A „posztgenomikus” korszak célja a genomialis szekvenciák annotációja, azaz annak megértése, hogy mit jelent ez a sok millió oldalt kitevő szöveg.

Nem bizonyult könnyű feladatnak a hárommilliárd betű megfejtése sem. Egy 1992-es becslés szerint – az akkori árak (1 USD/bázispár) és szekvenálási sebesség (száz ezer bázispár/év) adataiból kiindulva – harmincezer évre és hárommilliárd dollárra lenne szükség a hárommilliárdnyi bázispár meghatározásához. Hogyan sikerült mégis megfejtetni ezt a rendkívül fontos információt 2001-re? És mit tudunk meg ebből a hatalmas információból? Erről kívánok néhány gondolatot összefoglalni az alábbiakban. A kulcsfogalmak összefoglalása segíthet a gondolatmenet megértésében (1. táblázat).

### A hierarchikus módszer

A Humán genom projekt több állam által támogatott, nemzetközi projekt, amelyet 1990-ben indítottak el 15 éves időtartamra, hárommilliárd dollár támogatással. A Humán genom projekt fő támogatója az USA két hatalmas intézménye, a Department of Energy és a National Institutes of Health (NIH), de részt vett ebben a programban számos ország intézménye is, mint például az angol Wellcome Trust, továbbá francia, német, japán és kínai kutatóintézetek. A Humán genom projekt program fórumaként, a nemzetközi munka koordinálására, megalapították a Hugót (Human genome

### A KÖZLEMÉNYBEN ELŐFORDULÓ RÖVIDÍTÉSEK

BAC (bacterial arteficial chromosome): bakteriális mesterséges kromaszóma.  
 EST (expressed sequence tag): olyan betűsorozatokat jelent, amelyek megfelelnek egy-egy humán gén átírásra kerülő darabjának.  
 Hugo (Human genome organization): Humán genom szervezet.  
 LINE (long interspersed elements): hosszú betolakodó szekvenciák, körülbelül 6000 bázispár hosszúak, hozzávetőleg 8500 példányban fordulnak elő a humán genomban, azaz genomunk 20-25%-át foglalják el.  
 NIH: National Institutes of Health.  
 SINE (short interspersed repeats): rövid betolakodó idegen szekvenciák (100–300 bázispár hosszúak).  
 SNP (single nucleotide polymorphism): egypon-tos nukleotidvariáció.  
 STR (short tandem repeats): rövid egymás utáni ismétlődések.  
 Tiger (The Institute for Genomic Research): Genomikai Kutatóintézet.  
 VNTR (variable number of tandem repeats): változó számú egymás utáni ismétlések.

### 1. TÁBLÁZAT

*A Humán genom projekt témakörében használt kulcsfogalmak és rövid definíciójuk*

Bázispár: a DNS egy-egy betűpárral jelölt szerkezeti egysége.  
 DNS-szekvencia: a DNS-bázisok (A, G, C, T) egymást követő sorozatai.  
 Genom: egy ember teljes genetikai információja (körülbelül hárommilliárd bázispár).  
 Szekvenálás: a DNS-szekvencia meghatározása.  
 PCR: polimeráz láncreakció (polymerase chain reaction); egy adott DNS-részlet megsokszorozása.

organization). A program első igazgatója ugyanaz a Watson volt, aki megfejtette a DNS kettős spirált.

A Humán genom projekt fő stratégiája a genom hierarchikus lebontása mind kisebb és kisebb szerkezeti egységekre, és csupán ezt követte a szekvenálási munka (1). Sokan kritizálták a tervet a rendkívül időigényes hierarchikus rendszer kiépítése miatt. Miért kellett kiépíteni a hierarchikus rendszert?

A DNS-szekvenálás technikai kivitelezése maximum 500 betűs sorozatokban történik. Ezért a „végláthatatlan” DNS-molekulákat ilyen nagyságrendű darabokra kell bontani, és csupán ezeket a kisebb fragmenteket szekvenálják. A kapott nyers szekvenciákat azután összerakják, és így készül el a végleges szekvencia. A részek összerakása azonban nem könnyű. Hogy köny-

nyebben megértsük a probléma lényegét, térjünk vissza a könyvpéldánkra. Képzeld el, micsoda munka lenne összerakni a több ezer könyvnyi szöveget 300-500 betűs, oldalszámozás nélküli darabokból, különösen akkor, ha bizonyos szövegrészeket sokszorosan ismétlődnek! Mennyivel könnyebb a helyzet, ha a teljes anyagot először nagyobb darabokra osztjuk, majd azokat kisebb egységekre, és ezekben határozzuk meg a szövegrészeket, amelyekről így pontosan tudjuk, hogy hova valók.

Watson vezetése alatt a Humán genom projekt első célkitűzése az volt, hogy körülbelül 150 ezer bázispáronként meghatározzanak egy jelet – vagy más szóval markert –, ami ezt a darab információt a kromoszóma egy adott helyéhez köti. A markerek 100-200 betűből álló egyedi szekvenciasorozatokat, amelyek csak egyszer fordulnak elő a genomban. Ezt követte a genom feldarabolása körülbelül 150 000 bázispárból álló, részekben átfedő darabokra. Ezek a viszonylag nagy darabok mesterséges kromoszómaként kerültek be egy-egy baktériumklónba. A klónok összességét BAC-könyvtárnak (bacterial artificial chromosome) nevezték el. A BAC-könyvtár rendezése, sorba rakása a munka igen fontos része volt. A BAC-könyvtár rendezése során nemcsak az átfedő kromoszomális darabokat rakták sorba, hanem szelektáltak minden redundáns klónt. A rendezés eredményeképpen minden kromoszómához tartozik egy olyan BAC-klón-sorozat, amely ismert sorrendben tartalmazza a kromoszóma darabjait (1. ábra). A hierarchikus módszer elvét alkalmazva a BAC-klónok létrehozását egy-egy BAC-klón további feldarabolása követte. Az így keletkezett, kisebb (körülbelül 1500 bázispár), átfedő

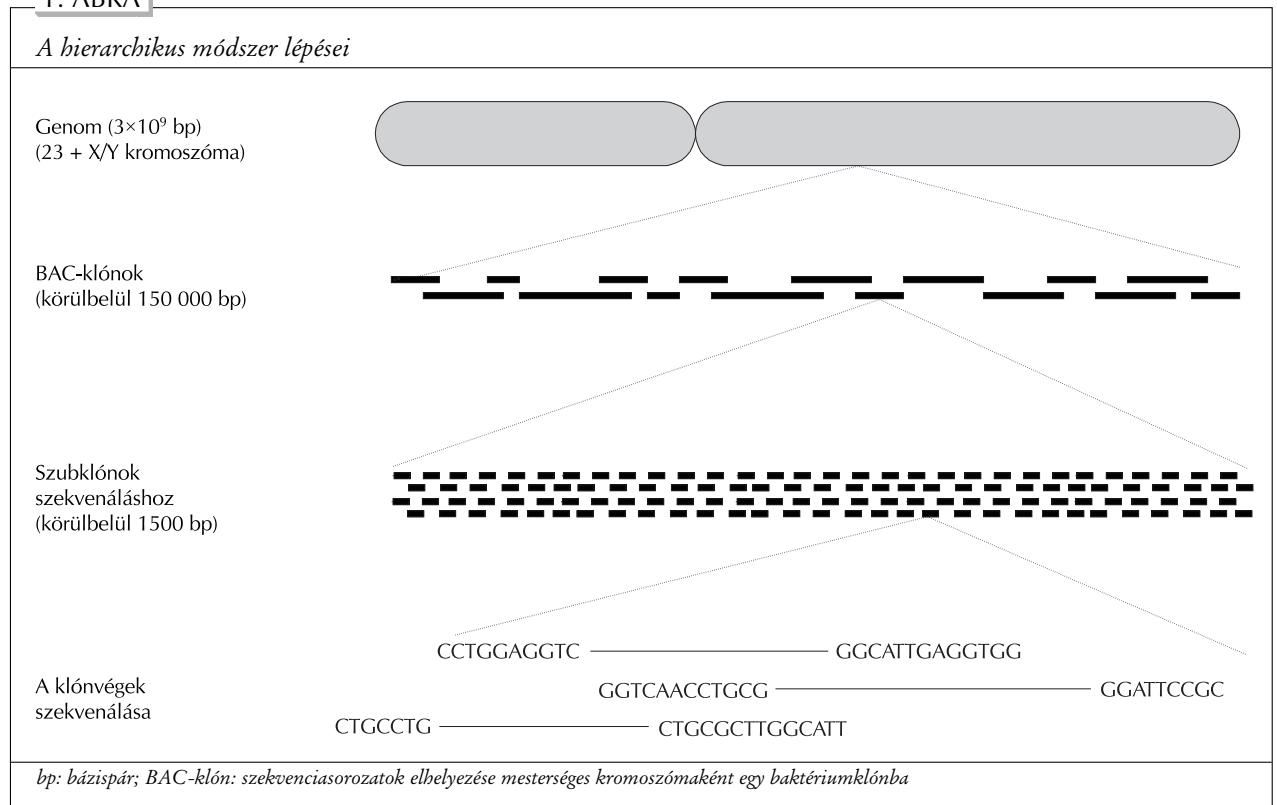
darabokat már közvetlenül szekvenálták (1. ábra). Érdemes megjegyezni, hogy elegendőnek bizonyult a szubklónok végeinek szekvenálása, nem volt szükség a teljes darab ismeretére. Ugyanis, ha elegendő átfedő darabunk van, akkor ez már lehetővé teszi a szekvencia összerakását (1. ábra).

A BAC-klónok rendezett könyvtára nemcsak a humán genom DNS-szekvenciájának megismeréséhez nyújtott óriási segítséget, hanem a további munkánál is igen fontosnak bizonyult, hiszen könnyen hozzáférhető formában tartalmazza az ember genetikai információját. Ma például, ha vizsgálni akarjuk kromoszomális állományunk egy meghatározott darabját, ezt bármikor megrendelhetjük a BAC-könyvtárakból.

Ez a munka a Humán genom projekt nemzetközi hálózatában folyt, amely megalapozta a projektközpontú hálózatok hasznosságát. Ez a kutatási stílus tükröződik vissza az új európai kutatási programokban is. A Humán genom projekt taglaboratóriumainak megállapodása szerint minden összerakott szekvenciát 24 órán belül mindenki számára hozzáférhetővé kellett tenni a HPG internetes adatbázisain keresztül, hogy elkerüljék a párhuzamos meghatározásokat. Reméljük, hogy ez az ésszerű kutatási mód is elterjed a jövőben, és egyre több szakinformációhoz férhetünk hozzá ingyenesen az interneten keresztül.

A humán genom megismerése során egyre inkább úgy tűnt, hogy a fehérjéket kódoló gének kis szigetekként helyezkednek el egy teljesen értelmetlennek tűnő információtengerben. Ezért hamarosan véleménykülönbség alakult ki azzal kapcsolatban, hogy miért nem csak a hasznos információt próbáljuk megismerni, mi-

1. ÁBRA



ért foglalkozunk a többi „informatikai szeméttel”? Venter hamarosan kidolgozott egy olyan rendszert, amellyel könnyen megtalálhatók a gének, és így gyorsan haladt a humán gének szekvenciájának meghatározása. A módszer alkalmazásakor a beazonosításhoz használandó betűsorozatokat nem a teljes humán genomból vette, hanem az átírt, fehérjekódoló információban, az m-RNS DNS-re írt másolatából, az úgynevezett cDNS-könyvtárból. Ezt elnevezte EST-nek (expressed sequence tag); ez olyan betűsorozatokat jelent, amelyek megfelelnek egy-egy humán gén átírásra kerülő darabjának. Először az agyi cDNS-könyvtár random primerrel történő PCR-amplifikációját készítette el. Az agyi cDNS-könyvtár tartalmazza mindazon gének információját, amelyek az agyban fejeződnek ki. Így az agyi EST-k megmutatták, hogy hol található a kromoszómákban agyifehérje-gének (2. ábra). Az első próbálkozások során 2375 EST-t szekvenáltak, majd megkeresték a szekvenciadarabok környezetében található géneket a Humán genom projekt adatbázisában. Az EST-k segítségével 17%-ban olyan géneket találtak meg, amelyek már előzőleg is szerepeltek az adatbázisokban; az EST-k 83%-a viszont addig ismeretlen gének helyét mutatta meg. Ez az információ igen értékes, eladható termék lett; megalapozta azt, hogy Venter kiváljon a Humán genom projektből, és megalapítsa saját cégét, a Celera-t.

## Verseny a genom megfejtéséért

Venter vezetésével – a magánszektor támogatására épülve – 1998-ban megalakult a Celera. A társaság célja az volt, hogy három év alatt elkészítik a teljes humán genom DNS-szekvencia-analízisét. Így az utolsó három évben óriási verseny indult meg a két intézmény, az állami támogatású Humán genom projekt és a magánszektorba tartozó Celera között. Hogyan történhetett meg, hogy ami 11 éves munkájába került a Humán genom projektnek, azt a Celera három év alatt készítette el? A válasz tulajdonképpen egyszerű: a DNS-szekvenálási technológiák óriási fejlődése gyorsította fel mindkét projekt munkáját az utolsó három évben. Ezen túlmenően a Celera más módszert alkalmazott a genom megfejtésére. A The Institute for Genomic

Research (Tiger) már korábban igen jelentős eredményeket ért el olyan komplex bioinformatikai rendszerek kidolgozásával, amelyek alkalmasak voltak hatalmas DNS-szekvencia-adatbázisok kezelésére. A módszert több százezer EST analízisére tudták alkalmazni, amellyel közel 30 000 humán gént azonosítottak. Ebből a komputer technológiából indult ki a Celera Humán genom projekt stratégiája is, amelynek lényege, hogy elhagyták a rendkívül időigényes hierarchikus rendszer kiépítését. Helyette közvetlenül alkalmazták az úgynevezett géppuskás feldarabolás (shotgun restriction digest method and sequencing) módszerét: a genomot egyből apró darabokra vágta fel, majd mindkét végét szekvenálták. A kapott szekvenciák összerakásához igen nagy teljesítményű bioinformatikai eszközöket használtak fel.

1998-ban született meg az első, nagy teljesítményű automata DNS-szekvenátor is, a Perkin-Elmer (Applied Biosystems) ABI PRISM 3700-as DNS-analizátora. A Celera egyrészt alkalmazta a Tiger tapasztalatait és programjait, másrészt az újonnan kifejlesztett automata DNS-analizátorokat, és így a Tiger eredeti kapacitásának hozzávetőlegesen ötvenszeresét tudta munkába állítani a humán genom megfejtésének érdekében. Az új gépek adta lehetőségek és a Celera kihívása a Humán genom projekt munkasebességét is megsokszorozta. Ugyanakkor a Celera felhasználta a HGP hierarchikus rendszerének eredményeit is, amelyeket a Humán genom projekt mindenki számára hozzáférhető internetes adatbázisokban tett közzé. A verseny helyébe hamarosan tárgyalások léptek, és ennek eredményeként mind az állami, mind a magánszektor ugyanabban az időben, 2001 februárjában jelentette meg eredményét a Nature (1), illetve a Science (2) egy teljes számában. Ezzel megszületett a humán DNS-szekvencia két példánya, amelynek első átfogó analízise igen izgalmas eredményeket hozott.

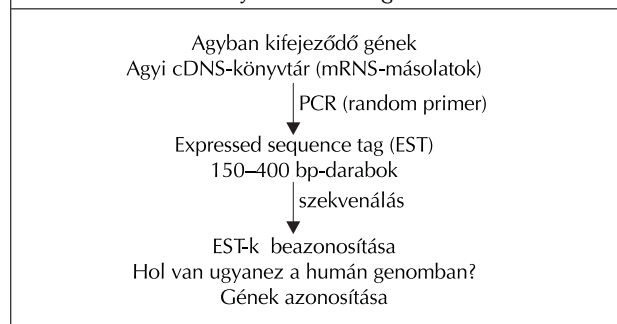
## Az eredmény

### A humán genom nyers szekvenciája

Mi a Humán genom projekt és a Celera vállalkozásának eredménye? Elsősorban egy közel hárommilliárd betűből álló sorozat, amely a Humán genom projekt esetében az interneten keresztül mindenki számára szabadon hozzáférhető (3), a Celera eredményei pedig bizonyos előfizetések mellett használhatók (4). Önmagában ez az információ azonban nem elegendő ahhoz, hogy felhasználhassuk a genomikai kutatásokban. Szükségesek még azok a bioinformatikai kezelési módok, amelyeknek jelentős része a Humán genom projekt honlapjain szintén szabadon használható (5). Ezek közé tartoznak például azok a programok, amelyekkel egy általunk meghatározott szekvencia azonosítható a genomban, vagy hasonló szekvenciák kereshetők (Blast). A Humán genom projekt genomiai szekvenciáit reprezentáló programok sokféleképpen felhasználhatók. Vizsgálható egy-egy kromoszóma, például a

## 2. ÁBRA

### Az EST készítése és felhasználása gének azonosítására



21-es kromoszómán elhelyezkedő gének tanulmányozása fontos információt adhat a 21-es triszómiát (Down-kór) vizsgáló kutatók számára. Kereshetünk az adatbankban a gén neve szerint is, nagy segítség az is, hogy a genomikai információ közvetlenül össze van kötve a PubMed adatbázisaival, azaz az adott génnel foglalkozó közleményekkel.

A humán genom jelenlegi példánya még nem a végleges forma, ez csupán a nyers szekvencia.

A humán genom jelenlegi példánya azonban még nem a végleges forma, ez csupán a nyers szekvencia. Miben különbözik ez a munkapéldány a végleges verziótól (6)? A hiányosságok kétfélek: egyrészt vannak olyan kisebb-nagyobb lyukak, ahol a szekvencia hiányzik.

Ezen túlmenően a megadott szekvenciák jelentős része még pontosításra szorul. A cél a 99,99%-os pontosság. A pontosság abban fejezhető ki, hogy az adott szekvenciárészletet hányszor határozták meg, illetve a meghatározások mennyire egyeznek. 2001 végére már három kromoszóma (20-as, 21-es és 22-es) végleges szekvenciája vált ismertté, a többi is gyors ütemben készül.

### Mit tudtunk meg?

A humán genom első nyers szekvenciájának közzétételére több szempontból is nagyon fontos eredmény volt (7). A DNS-szekvencia-meghatározás technikájának óriási fejlődése tette lehetővé, hogy elkészüljön ez a méreteiben is óriási munkapéldány. Az előzetesen vizsgált kisebb genomokhoz képest a humán genom körülbelül 25-ször nagyobb és hétszer több információt tartalmaz, mint az ezt megelőzően ismert összes genom együttvéve. Annak ellenére, hogy a nyers szekvencia véglegesítése még többéves munka lesz, a 2001-ben publikált eredmények is számos izgalmas konklúzió levonását tették lehetővé.

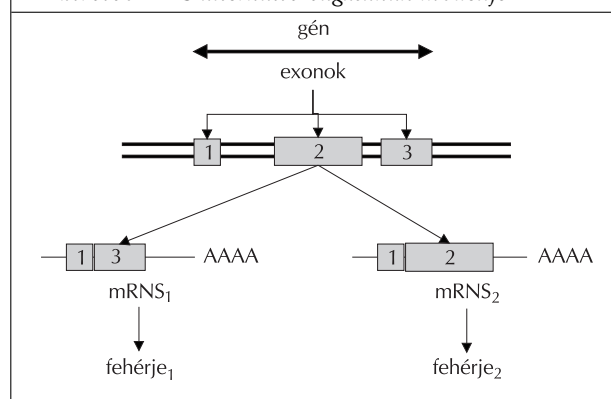
A legmeglepőbb eredmény talán az, hogy a gének számát 30-40 ezerre becsülték az első adatok alapján (ez az adat kissé növekvő tendenciát mutat, ma inkább 40-60 ezer génről beszélnek). Ez a szám jóval kisebb a vártnál, és mindössze kétszerese, mint amennyit a muslincában meghatároztak. Hol van hát a humán funkciók komplexitásának genetikai háttere? A komplexitás titka feltehetően nem a gének számában, hanem a gének által meghatározott fehérjék összerakási módjában van elrejtve (1). Feltehetően a gerincesekben, és különösen az emberben igen fontos szerepet játszik a DNS-ről képződő és a fehérjék szintézisét meghatározó

A gének száma jóval kisebb a vártnál.

hírvivő RNS (mRNS) alternatív vágása. Ennek eredményeként egyetlen gén többféle fehérjét is meg tud határozni (3. ábra). A variálható komplexitás másik aspektusa a fehérjék doménszerkezete. A domén olyan funkcionális egység a fehérje működésében, amely egy adott funkcióhoz kapcsolható. Egy adott domén többféle fehérjében is előfordulhat, ha ezeknek

### 3. ÁBRA

A hírvivő RNS alternatív vágásának modellje



a fehérjéknek van azonos jellegű funkciójuk. Az emberi fehérjék esetében körülbelül 1800 domént tételeznek fel; ez a szám közel kétszerese annak, amennyit az alacsonyabb rendűekben találtak. A doménszerkezet a fehérjék közti kölcsönhatások alapját képezi. Az egymással kapcsolódó fehérjék pedig bonyolult és magas szervezetségű információs hálózatok szerkezeti alapját képezik. Úgy képzelhetjük el tehát az emberi komplexitás molekuláris alapjait, mint egy igen flexibilis összerakó játékot, ahol igen nagy az elemek variációinak száma. Ehhez hozzájárul a gének exon/intron szerkezete is (az exon a fehérjét kódoló információ, az intron kivágódik az átírás során). A nyers szekvencia adatai alapján megállapították az emberi gének exonjainak (100-200 bázispár) és intronjainak (1000-4000 bázispár) átlagos hosszát, és azt, hogy egy emberi gén átlagosan 7-9 exonból áll. Mindezek alapján a ténylegesen fehérjében megjelenő információ a genom kevesebb, mint 1%-a.

A viszonylag kisszámú humán gén megkönnyíti a gének funkcionális azonosítását. A cél a humán gének és fehérjék teljes körű egymáshoz rendelése. Ez az eredmény feltehetően olyan mérföldkő lesz a biológiai és orvostudományok kutatásaiban, mint amilyen óriási hatást gyakorolt a kémiai periódusos rendszer elkészítése a kémiai tudományok fejlődésére. Tudjuk majd, hogy ezekből a szerkezeti egységekből, illetve ezek kombinációjából áll a testünk. Ezek közt kell keresnünk a gyógyszerek támadáspontjait, és a betegségek biológiai alapjait. Különösen fontossá válnak majd az *in silico* (kizárólag számítógép-analízisen alapuló) kutatások, amelyek bizonyos kutatási területeken már ma is hatalmas jelentőséget kapnak. Nem véletlen, hogy a Celera munkáját elsősorban gyógyszercegek támogatták, hiszen a gyógyszerkutatások céljaira rendkívül jól használhatók ezek az adatbázisok. Például ismerünk egy fehérjét, és azt kérdezzük a genomikai adatbázistól, hogy van-e hasonló, eddig még nem ismert fehérjét meghatározó gén? Így fedezték fel az Alzheimer-kór kialakulásában szerepet játszó preszenilin-2 fehérjét (8), vagy a teljes szerotoninreceptor (5-HT<sub>3A</sub>) egyik – eddig ismeretlen – alegységét (5-HT<sub>3B</sub>) (9). Ha ismert az a fehérjedomén, amelyre egy adott gyógyszer hat, akkor vizsgálható például az, hogy mely gének ren-

delkeznek e domén szekvenciájával. Az eredményül kapott gének, illetve fehérjék listája megadja nekünk az összes komponens, amire az adott gyógyszernek hatása vagy mellékhatása lehet. Ily módon jelentősen felgyorsítható a gyógyszerfejlesztési munka.

A humán genom nyers szekvenciájának megismerése arra is rámutatott, hogy génállományunk több mint 90%-a funkcionálisan semleges, nem tartalmaz átíró információ, és feltehetően nincs hatással a fenotípusra (1, 2). A hasznos információ tartalmazza az exonokat, a transzfer és riboszomális RNS-ek génjeit, továbbá a gén kifejeződésének (expressziójának) szabályozásáért felelős DNS-szakaszokat (promóter, enhancer régiók). Felmerül a kérdés, hogy mi a jelentősége a maradék DNS-nek? Ma még nem tudunk biztonsággal felelni erre a kérdésre, de feltehető, hogy ennek az információnak nincs közvetlen szerepe az életünkben. A humán genom mintegy 45%-át nagyrészt retrovírus eredetű, „parazita” szekvenciák (transzpozonok) sorozatai teszik ki. Ezek olyan ismétlődő elemek, amelyek önmagukat szaporították a törzsfajlás során. Idetartoznak a LINE (long interspersed elements) szekvenciák, amelyek körülbelül 6000 bázispár hosszúak és hozzávetőleg 8500 példányban fordulnak elő a humán genomban, azaz genomunk 20-25%-át foglalják el. Egy másik „betolakodó idegen” szekvencia a SINE (short interspersed repeats). Ezek jóval kisebbek, mint a LINE (100-300 bázispár), és önmagukban nem képesek szaporodni: ehhez szükségük van a LINE által kódolt reverz transzkriptázra. A SINE-k 1-2 millió példányban vannak jelen egy genomban, a genom 10-15%-át teszik ki. Lehet, hogy ezek között van olyan, ami mégiscsak összefügg az emberi gének működésével. Megállapították ugyanis, hogy az idetartozó Alu szekvencia majdnem minden gén környezetében előfordul, ennek szerepe azonban nem ismeretes. Vannak a genomunkban kiterjedt, egyszerű ismétlések is (például T betűk hosszú sorozata).

Az első eredmények alapján azt gondolhatjuk, hogy kisszámú génünk a retroviralis szekvenciák tengerében, szigetekként fordul elő. Az, hogy egy adott területen mennyi gén van, szintén igen változatos. Vannak igen aktív kromoszomális területek, míg máshol a gének sűrűsége alacsony. Érdekes módon a kromoszómavégek is tartalmazhatnak géneket, ilyen például a dopamin D4-es receptor génje, amely a kromoszóma csúcán, a telomer szekvenciák közvetlen közelében helyezkedik el. Egy ilyen lokalizáció feltehetően szerepet játszik a gén variabilitásában is. Az is érdekes, hogy igen sok a pszeudogén (egy aktív gén inaktív másolata). Jelenleg még tisztázatlan, hogy milyen következtetések vonhatók le egy adott gén funkciójával kapcsolatban a gén elhelyezkedése, környezete alapján.

## Kinek a genomját fejtettük meg?

A DNS-donorok mindkét projekt esetében önkéntesek és névtelenek voltak. A Celera két férfi és három nődonort választott a munkához (afrikai-amerikai,

ázsiai-kínai, spanyol-mexikói, valamint két kaukázusi donor), a genetikai anyagot immortalizált sejtvonalakba juttatták. Az, hogy aktuálisan melyik DNS-t dolgozták fel, kizárólag a technikai tényezőktől függött. Tehát igazából nem tudható, hogy kinek a DNS-e lett az első megismert genetikai adat. Bizonyos értelemben ilyen személy nem is létezik, hiszen a DNS-szekvencia a haploid kromoszómaállományra (plusz a mitochondriális DNS-re) vonatkozik, amelyből minden embernek két példánya van, azaz diploid, és a két példány részlegesen különbözik.

## Én más vagyok, mint te!

Következő izgalmas kérdésként az merült fel, hogy mennyire különbözik két ember genomja. Az összehasonlító szekvenciavizsgálatok eredményeként a 2001-es becslés szerint körülbelül minden ezredik DNS-betűnkben van egy különbség, ami a teljes genom vonatkozásában hozzávetőlegesen 2-3 millió betűnyi különbséget jelent (0,1%) (10). Figyelembe véve azonban, hogy e különbségeknek kevesebb mint 1%-a esik a gének kódolószakaszaiba, a tényleges különbség jóval kisebb. Így az egyes emberek közötti genetikai variabilitás forrása inkább néhány ezer betű. Természetesen nem mindegy, hogy hol megy végbe ez a betűcsere. Egyetlen bázispár megváltozása halálos kimenetelű mutációhoz vezethet, vagy rizikófaktoroként szerepelhet bizonyos rendellenességek kialakulásával kapcsolatban, de protektív szerepet is játszhat.

## A genetikai polimorfizmusok formái

Az emberek közti DNS-szekvencia-variabilitásokat polimorfizmusnak nevezzük. A polimorfizmus új kifejezés; jelzi, hogy a különbség teljesen semleges hatású is lehet. A betegségkókozó változatokat inkább mutációnak szokás nevezni. A mutációk ritkán – kevesebb mint a populáció 5%-ában – előforduló változatok. A mai kutatási tendencia arra utal, hogy a szekvenciavariációk listája gyorsabban gyűlik, mint ahogy az egyes változatoknak az emberi életminőségre való hatását fel tudnánk mérni. Ezért terjedt el a polimorfizmus elnevezés, jelezve, hogy a szekvenciális különbség hatását legtöbbször nem ismerjük.

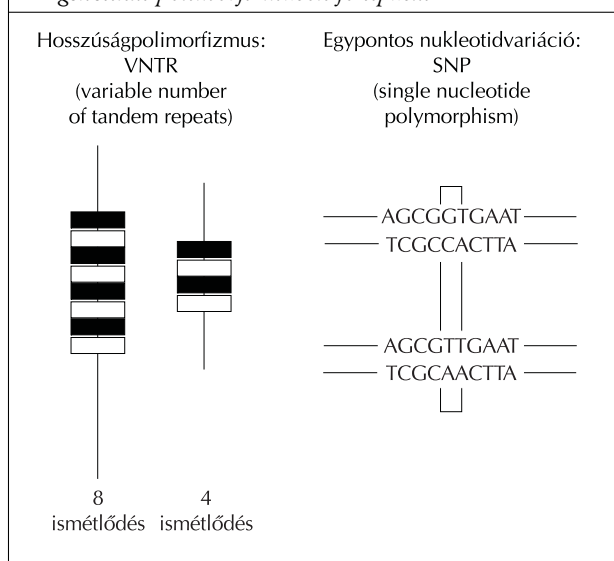
A polimorfizmusokat két fő csoportba szokás sorolni (4. ábra). Az egyik fő típus az egy pontos nukleotidvariáció (single nucleotide polymorphism, SNP). Az SNP azt jelenti, hogy a különböző eredetű humán szekvenciák egy adott pontban többféle variációban lé-

Úgy képzelhetjük el az emberi komplexitás molekuláris alapjait, mint egy flexibilis összerakó játékot, ahol igen nagy az elemek variációinak száma.

Az egyes emberek közötti genetikai variabilitás forrása néhány ezer „betű”.

4. ÁBRA

A genetikai polimorfizmusok fő típusai



tezhetnek. Annak ellenére, hogy a genetikai információ négy betűjéből maximum négyféle eset lehetséges, az SNP-k egy adott helyen legtöbbször csak kétfélek. A Humán genom projekt SNP-konzorciuma 2001 februárjában 1,4 millió SNP-t közölt (1); ez a szám az év végére megháromszorozódott (11). Az SNP-adatbázisoknak feltehetően óriási jelentősége lesz a genetikai meghatározottságú betegségek rizikófaktorainak feltérképezésében.

Az egyes emberek genomjai között található különbségek másik forrása a hosszúságpolimorfizmus. A genomra általában jellemző, hogy sok benne az ismétlődő szekvenciárészlet. Ez nem csak a „parazita” információra vonatkozik, mert a kódoló génekben is előfordulnak ismétlőszakaszok. Az eltérés az egyes emberek között az ismétlési számban lehet, ezáltal az adott szakasz hosszabb vagy rövidebb. Az ismétlődő

információ egységének hossza változó. Vannak nagyon rövid ismétlések (short tandem repeats, STR), ahol 1–5 betűből álló egység ismétlődik egymás után sokszor. Más esetben ez egy hosszabb információs egység. Ilyen például a dopamin D4-es receptor harmadik exonjában található, 48 betűből álló egység, amely 2–10-szer ismétlődhet, létrehozva a hosszú és a rövid formákat. Az ilyen hosszabb ismétléseket változó számú egymás utáni ismétlésnek (variable number of tandem repeats, VNTR) nevezik.

Mire használhatók a polimorf szekvenciák?

Mindenkit érdekel, hogy miért nem egyformák az emberek. Képességeink, betegségre való hajlamaink jelentős része öröklődik. Ezeknek az öröklődő vonásoknak az okait a polimorf gént változatok közt kell keresnünk (12).

Vannak olyan mutációk, amelyek monogénes öröklésmentűek, gyakran súlyos következményekkel járnak. Ezek közül a legsúlyosabbak a domináns öröklésmentű betegségek; idetartoznak a neurodegenerációs betegségek egyes formái (Huntington-kór, mendeli öröklésmentű Alzheimer-kór). Az ilyen mutációt homozigóta formában hordozó szülő, aki legtöbbször idősebb korában betegszik meg, valamennyi gyermekének átadja a betegséget. Ezekben az esetekben súlyos etikai problémát szül a genetikai diagnózis felállítása a gyermeknemzés előtt, hiszen a mai törvények szerint nem lehet megakadályozni a gyermekek nemzését. Kérdés, hogy a jövőben hogyan szabályozható etikai szempontból a genetikai tanácsadás?

Ugyanakkor a Humán genom projekt eredményei óriási mértékben hozzájárultak-hozzájárulnak a monogénes öröklődésű betegségekért felelős gének azonosításához. 2000-ben több mint harminc betegség esetén találták meg a betegségért felelős gént és annak mutációját.

Mindemellett a monogénes (mendeli) öröklésmentű betegségek ritkák. A nagy populációt érintő, öröklődő faktort tartalmazó népbetegségek (szív- és érrendszeri, mozgásszervi, neurológiai és pszichiátriai betegségek) hátterében nem egy, hanem sok gén áll. A komplex öröklődésű rizikófaktorok felkutatása nem egyszerű feladat, többféle stratégiát használnak. Ma egyre jobban terjed az úgynevezett kandidáns gén asszociációs vizsgálata. Ebben az esetben hipotetikus kiválasztunk egy vagy több gént, amelyet kandidáns gén(ek)nek nevezünk, és amelyről feltételezzük, hogy szerepet játszhat egy adott betegség kialakulásában. Ezután a kandidáns gén polimorf alléljainak gyakorisági értékeit hasonlítjuk össze különböző módszerekkel. Mérhetjük például az allél vagy géntfrekvenciák különbségeit az egészséges és a beteg populációban, vagy vizsgálhatjuk az egyes allélváltozatok preferenciális átadását is. Ezek a vizsgálatok ma az érdeklődés középpontjában állnak, azonban még igen sok ellentmondást tartalmaznak. Remélhető, hogy a humán genomális szekvencia és annak variabilitásának ismerete a komplex öröklődésű rendellenességek azonosításában is nagy segítséget nyújt majd.

Kitekintés

Az ember genetikai információanyagának megfejtése mérföldkő az emberi élet molekuláris alapjainak megismerésében. A projekt teljesítése a „pregenomikus” éra végét, egyben a „posztgenomikus” éra kezdetét jelzi, ahol a tudományos módszerek várhatóan meg fognak változni. Előtérbe kerül az *in silico* kutatás, az adatbányászat, a bioinformatika, és a jövőben feltehetően akkor végzünk hatékony munkát, ha fel tudjuk használni ennek a hihetetlenül nagy információnak a lehetőségeit (13). Folyamatosan állítják össze azokat az adatbázisokat, amelyek az összes emberi gént, illetve az összes humán fehérjét tartalmazzák. Ez az „orvosbiológiai periódusos rendszer” minden valószínűség szerint alapvető változásokat idéz majd elő a kutatási módszerek-

Az öröklődő vonások okait a polimorf gént változatok közt kell keresnünk.

ben: előtérbe kerülnek a sokfaktoros modelleket könnyen kezelő eljárások. Ezzel párhuzamosan a következő nagy feladat a proteomika keretében a humán genom által kódolt valamennyi fehérje szerkezeti és funkcionális feltérképezése lesz.

A következő lépések közé tartozik más fajok genetikai információjának megfejtése is. Elindult például a csimpánzgenom analízise: feltételezik, hogy az emberi genommal körülbelül 99%-ban egyezik, az újabb vizsgálatok azonban nagyobb különbségekre utalnak. Ugyanakkor sokat várnak a humán és a csimpánz genetikai állományának összehasonlító vizsgálatától, amelyek nemcsak az emberi kognitív funkciók genetikai alapjainak megértésében segíthetnek, de olyan kérdé-

sekre is feleletet adhatnak, hogy például miért rezisztensek a csimpánzok a HIV-vírusra.

Befejezés előtt áll az egérgenomprogram és a rizsgenomprogram. Számos más organizmus genetikai információtartalmának megfejtése is folyik. Az összehasonlító vizsgálatok valószínűleg még több adattal szolgálnak majd genetikai anyagunk organizációjának megértéséhez, és új távlatokat nyitnak a molekuláris evolúció kutatásában.

#### Köszönetnyilvánítás

A laboratóriumunkban folyó munkát az NKFP 0008/2002, az OMF 00215/2002, az OTKA-T035203 és az ETT-T30-002 témátámogatások fedezik.

## IRODALOM

1. International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921.
2. Venter JC, et al. The Sequence of the Human Genome. *Science* 2001;291:1304-51.
3. <http://www.ncbi.nlm.nih.gov/genome/seq/>
4. <http://www.celera.com>
5. Wolfsberg TG, McEntyre J, Schuler GD. Guide to the draft human genome. *Nature* 2001;409:824-6.
6. Aach J, Bulyk ML, Church GM, Comander J, Derti A, Shendure J. Computational comparison of two draft sequences of the human genome. *Nature* 2001;409:856-9.
7. Baltimore D. Our genome unveiled. *Nature* 2001;409:814-6.
8. Sherrington R, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* 1995;375:754-60.
9. Davies PA, et al. The 5-HT<sub>3B</sub> subunit is a major determinant of serotonin-receptor function. *Nature* 1999;397:359-63.
10. Nowotny P, Kwon JM, Goate AM. SNP analysis to dissect human traits. *Curr Opin Neurobiol* 2001;11(5):637-41.
11. <http://www.ncbi.nlm.nih.gov/SNP/>
12. Lai E. Application of SNP technologies in medicine: lessons learned and future challenges. *Genome Res* 2001;11(6):927-9.
13. Subramanian G, Adams MD, Venter JC, Broder S. Implications of the human genome for understanding human biology and medicine. *JAMA* 2001;286(18):2296-307.



## VIII. MÁJNAP

**Időpont:** 2003. május 9.

**Helyszín:** Semmelweis Egyetem, Budapest, Nagyvárad téri Elméleti Tömb, Díszterem

A program:

9.30: Kéri Ágnes: Növényi hatóanyag-tartalmú készítmények alkalmazása a máj- és epebetegségek kezelésében

10.00: Simon László: A *Helicobacter pylori*-fertőzés és az ulcusbetegség jelentősége idült májbetegségek mellett

10.30: Nemesánszky Elemér: Krónikus vírushepatitisek

11.00: Szlávik János: Védőoltások csökkent védekezőképességű felnőttek részére, különös tekintettel a májbetegségekre

11.30: Nebenführer László: Májbetegségeket kísérő bőrtünetek és bőrbetegségek

12.00–13.00: Ebédszünet

13.00: Nyirkos Péter: Gombás infekciók a szervezeten kívül és a szervezeten belül

13.30: Szalay Ferenc: Májbetegségek genetikai háttere (mi a hajlam?)

14.00: Káli Gábor: Hepaticus encephalopathia és májkóma

14.30: Budai József: Májbetegségek neurológiai tünetei

A tudományos programot a Magyar Gastroenterológiai Társaság támogatja, a Fővárosi Szent László Kórház Májbetegéért Alapítványa szervezi.

**Információ:** Tóth Andrea szervező, telefon/fax: (1) 212-2232, e-mail: newinst@hu.inter.net