

ÁNGYÁN ANNAMÁRIA FRANCISKA – GÁSPÁRI ZOLTÁN

Új fehérjék a semmiből

Az egyes fajok közötti különbségek fő oka genetikai állományukban keresendő. Az egymással rokon csoportok genomjai között számos eltérés lehetséges, melyek közül az egyik legkézenfekvőbb a génkészletben rejlik. Bár napjainkban a szabályozó mechanizmusok szerepének vizsgálata is egyre inkább előtérbe kerül, a fehérjekódoló gének számát és főleg a kódolt fehérjék milyenségét érintő evolúciós változások szerepe megkerülhetetlen. Írásunkban röviden áttekintjük azokat a mechanizmusokat, amelyek a génkészlet megváltozásához vezethetnek, különös tekintettel egy régen felvetett, de csak az utóbbi években igazolt jelenségre, az ún. *de novo* géneletkezésre.

Mindenekelőtt röviden áttekintjük a fehérjék sejtbeli keletkezésének folyamatát. A DNS egy megfelelő szabályozó régiókkal rendelkező szakaszáról, azaz egy adott génről RNS-másolat készül a transzkripció (azaz átírás) során. Jelen tudásunk szerint a keletkező RNS-ek túlnyomó része úgynevezett nem kódoló RNS, amely pl. szabályozó vagy akár katalitikus szerepet is betölthet. A szabályozó RNS-ek igazi sokféleségének feltérképezése az utóbbi évek egyik legizgalmasabb új kutatási területe. A másik, jobban ismert csoport tagjai, a kódoló RNS-ek átírás és néhány tipikus módosítás után lefordítódnak fehérjékre a riboszómákon, ez a folyamat a transzláció. Ezen RNS-ek szokásos neve hírvívő (messenger) RNS, röviden mRNS, a név abból adódik, hogy ők viszik az információt DNS-től a fordítás helyére. Az mRNS-re jellemző két legfontosabb módosítás közül az egyik az ún. splicing, vagyis az intronnak nevezett szakaszok kivágása, ezek az mRNS-en belül elfoglalt helyüktől függetlenül nem vesznek részt a fehérje kódolásában. A másik módosítás az ún. poliadeniláció, amikor az adenin nevű bázis kb. 250 példányát tartalmazó szakasz kerül az mRNS végére. Mind a két módosítás tipikusan, de nem kizárólagosan jellemző az mRNS-molekulákra. A fordítás, vagyis az mRNS bázissorrendje és a fehérje aminosav-sorrendje közti megfeleltetés a genetikai kódszótár alapján történik, ami azt is jelenti, hogy az mRNS-molekulának – és így az őket meghatározó géneknek – rendelkezniük kell olyan szakasszal, amely a kódszótár alapján ér-

telmezhető és a szervezet számára hasznos, működőképes fehérje előállítását kódolja – nem lehetnek tehát akármilyenek. A fordítás megfelelő START és – mivel a transzláció bázishármasokhoz kötött – a hozzá képest meghatározott módon és tipikusan nem akármilyen közeli pozícióban elhelyezkedő STOP jel jelenlétéhez kötött.

A molekuláris biológusok előtt jól ismert géneletkezési forgatókönyvek már

korábban is fehérjét kódoló gén a köztük lévő DNS-szakasz kiesésével mintegy egybeolvad. Az így keletkező új gén nyilván mindkét „szülőjére” hasonlít, de egyikkel sem azonos. Természetesen mind a duplikációval, mind a fúzióval létrejött gének szelekció alatt állnak, azaz a káros változataikat hordozó egyedek evolúciós hátrányban, hasznos változataikkal rendelkezők pedig előnyben lesznek társaikhoz képest.



1. ábra. A fehérjekódoló gének működésének egyszerűsített vázlata. A megfelelő szabályozó régiókkal rendelkező DNS-szakasz átíródik mRNS-molekulává, amely jellegzetes módosításokon esik át, majd az általa hordozott információ a genetikai kódnak megfelelően lefordítódik fehérjeszekvenciára. (A sematikus rajzon a DNS- és a fehérjehosszak nem reálisak)

meglévő génekből indulnak ki [1,2]. Egy adott DNS-szakasz megduplázódásával egy gén új példánya jelenhet meg a genomban. Szerencsés körülmények között ez az új példány fennmaradhat a generációk során, és bázissorrendjének megváltozásával fokozatosan elveszítheti nagyfokú hasonlóságát az őt életre hívó eredeti szekvenciával. Ezek a változások természetesen a molekuláris funkció módosulásával is járhatnak, növelve ezzel az élőlényben megvalósulni képes biokémiai és/vagy szabályozási folyamatok sokféleségét. Maga a megduplázódás történhet a DNS szintjén, illetve átírt RNS-molekulák DNS-re való visszairásával és a visszaírt szakasz DNS-be integrálásával (retrotranszpozíció) is.

Egy másik lehetőség a korábban nem létező gén létrejöttére a fúzió, amikor két,

Ezen géneletkezési forgatókönyvek ismeretében némileg meglepő, hogy az összehasonlító genomikai kutatások egyre több olyan gént azonosítanak, amelyek „árvák” (orphan genes), azaz nem hasonlítanak egyetlen más ismert génhez sem. Az ilyen gének aránya egyes genomokban elérheti a 20%-ot is. Az árvaságnak, pontosabban rokonalanságnak több oka is lehetséges, például:

1. A gén duplikációval keletkezett, tehát valójában vannak rokonai, de az évmilliók során ezektől a rokonoktól annyira különbözővé vált, hogy a rokonság már nem ismerhető fel.

2. A gén duplikációval keletkezett, de rokonai mind vagy legalábbis nagyrészt „elvesztek” a vizsgált élőlényhez evolúciósan közeli organizmusokból, így látványosan közeli rokon nélkül maradt.

3. A génnek ténylegesen nincsenek fehérjét kódoló rokonai.

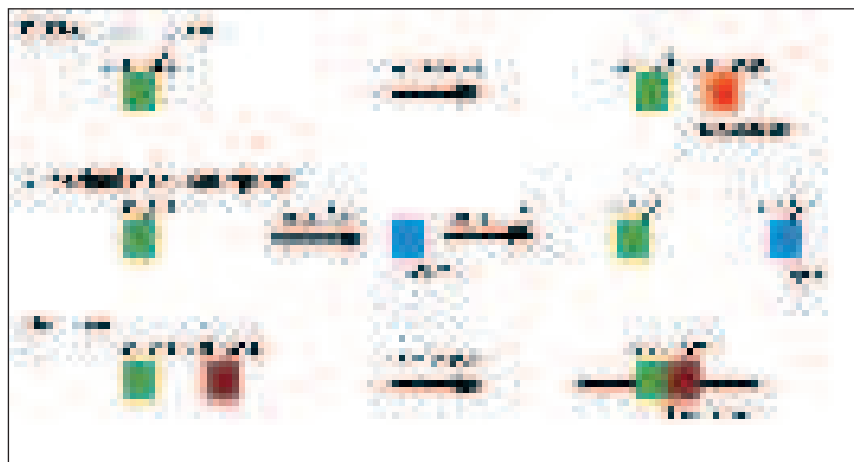
A 3. eset különösen érdekes, mert egy, az eddig ismertektől eltérő génkeletkezési mechanizmusra, az ún. *de novo* génképződésre utal. Ez a mechanizmus azt feltételezi, hogy egy olyan DNS-szakasz, amely korábban nem kódolt fehérjét, „hirtelen” szert tesz arra a képességre, hogy átíródjon, és RNS-átírata le is fordítódjon fehérjére. Ennek az első ránézésre majdnem lehetetlennek tűnő mechanizmusnak a lehetőségét már igen korán felvetették, a létezését azonban csak az utóbbi években sikerült megbízhatóan kimutatni. A bizonyítás ugyanis összetett folyamat: meg kell mutatni, hogy egy adott DNS-szakasz az egyik élőlényben aktív, míg annak közeli rokonaiban nem keletkezik róla fehérje. Az ehhez szükséges kísérletek pedig nem

meglétét feltételezi, azaz olyan DNS-szakaszokból indul ki, amelyek már többkevesebb hatékonysággal átíródnak valamilyen, természetesen nem fehérjekódoló RNS-molekulává. Ilyen módon az RNS-molekulának „csupán” a lefordításra kell alkalmassá válnia, ami a tankönyvi ismeretek szerint még mindig nagyon nehezen teljesíthető kritérium. Ugyanakkor több kísérlet is utal arra, hogy a transláció olyan esetekben is lejátszódhat, ahol hiányzik a specifikus START jel, vagyis az a bázishármas, ami a fehérjeszintézist általános esetben elindítja. Többek között egyes, ismétlődő bázishármasokat tartalmazó RNS-molekulák esetében kimutatták ezt a jelenséget. A jelenlegi kutatások arra utalnak, hogy a genom jóval nagyobb hányadáról történik transzkripció, mint korábban feltételezték, pl. az emberi gé-

megfelelő régiókban rokon fajok esetében nem, akkor egy potenciális *de novo* fehérjére bukkantunk, melynek létét természetesen mindenképpen érdemes kísérletileg ellenőrizni. A megfelelő régiók összehasonlításával több ilyen esetben azonosíthatók azok a mutációk, amelyek a funkcionális gén megjelenéséhez vezettek.

Természetesen nem mehetünk el amellett a kérdés mellett sem, hogy mire jók, egyáltalán hasznosak lehetnek-e a *de novo* génekről átíródó fehérjék (a továbbiakban: *de novo* fehérjék). Az ilyen fehérjék várhatóan nem keletkeznek nagy mennyiségben, hiszen a megfelelő szabályozó régiók feltehetőleg nem eleve a leghatékonyabb verzióban jelennek meg, amikor egyáltalán funkcionálissá válnak. A keletkező fehérjék evolúciósan akkor maradhatnak fenn, ha az élőlény számára előnyt jelentenek, de legalábbis nem károsak. Ha például toxikusak a sejtre nézve, vagy destruktív módon avatkoznak be egy szabályozó folyamatba megakadályozva két partner kölcsönhatását, akkor várhatóan a hordozó élőlény pusztulásához vezetnek. Abban a feltehetően kis valószínűségű, de nem lehetetlen esetben, ha valamilyen molekuláris folyamatot kedvezően befolyásolnak, továbbadhatnak a következő generációkba és elterjedhetnek, közben pedig a funkciójuk tökéletesedésének irányában meg is változhatnak további szelekció révén.

Most pedig lássunk néhány példát a közelmúltban azonosított *de novo* génekre! Az egyik legrészletesebb úttörő vizsgálatot Liping Wei és Qing-Rong Liu kutatócsoportjainak vezetésével több amerikai és kínai laboratórium összefogásával 2010-ben publikálták [4]. A kutatásban kifejezetten olyan gént kerestek, amely a nikotinfüggőség kialakulásában szerepet játszhat, és emberre specifikus *de novo* fehérjét kódol. Korábban publikált vizsgálatok újraelmzésével azonosították az FLJ33706 jelű gént, amely az akkori adatbázisokban nem kódoló RNS-t meghatározó szakaszként volt feltüntetve. A génről átíródó RNS hírvivő-RNS-ekre jellemző módon processzálódik (poliadeniláció és intronok kivágása is megtörténik), és megjelenik a fehérjetermék is. Utóbbit a kódoló régió alapján tervezett fehérjerészlet ellen termeltetett antitest segítségével mutatták ki. Egérben a fehérje nincs jelen, a további emlősszekvenciák elemzése alapján a gént tartalmazó régió csak méhlepényes emlősökben van meg, és a szekvencia megváltozásai fokozatosan, főemlősökben tették lehetővé az intronok kivágását, sőt az egyik intron kivágási helyei csak emberszabásúakban jelentek meg. A fehérjekódoló szakasz jelen formája pedig emberre specifikus, egyéb embersza-



2. ábra. Új gének keletkezésének néhány főbb mechanizmusa ([2] nyomán egyszerűsítve)

minden esetben végezhetőek el könnyen, különösen a keletkezett fehérjék kimutatása lehet nagyon munkaigényes, ezt nem is minden esetben végzik el a kutatók.

Hogyan lehetséges egy ilyen, látszólag lehetetlen esemény bekövetkezése? Ha sorra vesszük még egyszer az aktív gén működésének feltételeit, láthatjuk, hogy kellene olyan szabályozó régiók, amelyek az RNS-re való átíráshoz szükségesek, ezen túl pedig az átírt szakasznak jelentéssel kell bírnia a genetikai kódszótár alapján. Mindkét kritérium speciális kezdő- és végpontok meglétét feltételezi. Itt kell megjegyeznünk azt, hogy a szabályozó régiók nem kizárólagosan igen-nem kapcsolóként működnek: pontos szekvenciájuktól függően igen változó hatékonyságúak lehetnek. Ez azért fontos, mert ez a tulajdonság lehetővé teszi az ilyen régiók hasonló, kevésbé hatékony szekvenciákból való kialakulását és további optimalizálását.

A kutatók által javasolt részletes forgatókönyvek egyik típusa az első feltétel

nom nagyjából 75%-a átíródik valamilyen sejttypusban és/vagy életkorban [3]. Ez azt is jelenti, hogy kiváló „alapanyag” áll rendelkezésre a nem kódoló RNS-eket meghatározó DNS-szakaszokból fehérjekódoló gének keletkezésére.

Átíródó RNS hiányában a transzkripció feltételeinek is ki kell alakulniuk mutációk segítségével. A feltételezés az, hogy funkcionálishoz hasonló, de nem működő szabályozó régió már „rendelkezésre áll” az evolúció számára, hogy egy adott fajban egy vagy néhány mutáció segítségével kialakulhasson az aktív regulációs szakasz. Az átírt RNS ezek után az előző forgatókönyvben leírtak szerint fordítható le.

A *de novo* gének keresésének egyik legkézenfekvőbb módja a genomok számítógépes összehasonlító elemzése. Számos génpredikációs eljárás áll rendelkezésünkre, amely képes megjelölni egy DNS-szakaszról, hogy aktív fehérjekódoló génnek felel-e meg. Ha egy faj DNS-ének olyan helyén kapunk pozitív találatot, aminek

básúakban és főemlősökben több olyan eltérés is megtalálható, amely azt meg-
szakítja. Az elemzés rámutatott arra, hogy
a gén evolúciójában fontos szerepet ját-
szott bizonyos típusú ún. szétszórt ismét-
lődések (interspersed repeats) beilleszté-
se. Ezek olyan szekvenciák, amelyek a
genom számos helyén megtalálhatóak, és
bizonyos mechanizmusok révén másola-
taik a genomi DNS újabb pontjaira is be-
kerülhetnek. Az FLJ33706 gén esetében
ilyen – egyébként maguk fehérjét nem kó-
doló – ismétlődések „hozták magukkal”
a jelenlegi fehérjekódoló szakaszok egy
részét, valamint az intronok kivágásához
szükséges szekvenciareszleteket is. A *de
novo* gén kialakulásához tehát számos, a
genetikusok által jól ismert mechanizmus
(ismétlődések beillesztése, a szekvenciá-
ban egyes bázisok megváltozása vagy tör-
lődése) járult hozzá fokozatosan, így a *de
novo* jelleg nem hirtelen, a „semmitől”
történi megjelenést jelent, hanem megfe-
lelően egymás után következő elemi evolú-
ciós események eredménye.

A történet legérdekesebb része maradt a
végére: a génről átirított mRNS-ek szöve-
ti eloszlása egyértelműen azt mutatta, hogy az
által kódolt fehérje az emberi agyra spec-
ifikus. Az agyi előfordulást a fehérjeszín-
tű vizsgálatok is megerősítették. A fehérje
funkcionális szerepe ugyan nem tisztázott,
de azt sikerült kimutatni, hogy Alzheimer-
kóros agyi mintákban a fehérje megnöve-
kedett mennyiségben van jelen. A tanul-
mány tehát újabb részletet ad hozzá ahhoz
a mozaikhoz, amellyel az ember élővilág-
ban elfoglalt egyedi helyzetének gyökereit
próbáljuk megérteni.

Egy valamivel korábbi tanulmány há-
rom olyan gént azonosított, amelyek po-
tenciálisan emberre specifikus *de novo*
fehérjéket kódolnak [5]. A kutatók az em-
beri és a csimpánzfehérjék összehasonlító
elemzéséből indultak ki, és fokozatosan
szűkítették a fehérjék körét. Először kizár-
ták azokat a genomi régiókat, amelyeknek
megfelelő pozícióban a csimpánzgenom-
ban nem volt ismert szekvencia, mert ez
szekvenálási hibákból is adódhat, majd el-
vetették az összes olyan fehérjét, amelyek-
hez hasonló fehérje gondosabb elemzéssel
bármely más ismert organizmusban azo-
nosítható volt. Csak a több adatbázisban is
génként szereplő, RNS-re átiródozó szekven-
ciákat megtartva végül három gént talá-
tak, melyek potenciálisan emberi *de novo*
fehérjekódoló gének. A gének szerkeze-
te lényegesen egyszerűbb, mint a koráb-
bi példában, egyik fehérjekódoló régióját
sem szakítják meg intronok. Az emberi, a
csimpánz- és a makákószekvenciák össze-
vetése azt mutatta, hogy bár a megfelelő
DNS-szakasz nagyon hasonló a három faj-
ban, az emberihez hasonló fehérjekódoló

szakasszal a másik két faj nem rendelke-
zik. A kutatók kísérletileg is ellenőrizték
az emberi és a csimpánzének fehérjekó-
dolás szempontjából kritikus eltéréseit,
és megmutatták, hogy az ismert emberi
szekvenciaváltozatok egyike sem tartal-
maz a fehérjekódoló régiót megszakító
mutációt. Ezen felül a megfelelő adatbázi-
sok vizsgálatával a három gén által kódolt
fehérjékre jellemző darabokat azonosítot-
tak, megerősítve, hogy emberben fehérjét
kódolnak. A munka kritikusan megjegyzi,
hogy ez a módszer nem annyira megbíz-
ható, mint az előző példában a célfehérjé-
re kimutatására alkalmazott antitest-alapú
technika.

A harmadik bemutatott példában szin-
tén szisztematikusan kerestek embersza-
bású-specifikus *de novo* géneket [6]. A
rendelkezésre álló szekvenciaadatok alap-
ján megbecsülték az egyes emberi gé-
nek legkorábbi evolúciós megjelenésének
idejét a gerincesek között, majd további
szűrések után, melynek során többek kö-
zött feltétel volt, hogy a kísérleti adatok
szerint rézuszamjomban fehérjét nem kó-
doló régióról van szó, 24 potenciális *de
novo* gént azonosítottak. Az RNS-átíra-
tok szerkezetét kísérletesen ellenőrizték,
és megmutatták, hogy mindegyik esetben
a nem-főemlős genomok megfelelő régi-
óban nincs meg a megfelelő fehérjekó-
doló szakasz. A 24 gén közül 11 emberre
specifikus, 13 pedig emberben és csim-
pánzban is fehérjét kódol. A szerzők azt
találták, hogy az azonosított *de novo*
gén többségéről rézuszamjomban és/vagy
csimpánzban nem kódoló, de poliadenilált
RNS-molekulák íródtak át. Több esetben
is tapasztalták, hogy a fehérjekódoló gé-
nekről az RNS-átírás aktívabban történt,
mint a rokon faj megfelelő nem fehérjekó-
doló génje esetében. Érdekes, hogy az új
fehérjekódoló gének legtöbbször megint
csak az agyban fejeződött ki legnagyobb
mértékben. Noha ebben a tanulmányban a
potenciálisan keletkező fehérjék kísérletes
azonosítása nem történt meg, lényeges kö-
vetkeztetése, hogy a már aktíván átiródozó
régioók viszonylag könnyen adnak életet *de
novo* fehérjekódoló géneknek.

Írásunkban igyekeztünk rövid áttekin-
tést adni az újonnan keletkező fehérjék
egy érdekes osztályáról, mely egyre na-
gyobb figyelmet kap. Bár az egyik fő kér-
dés természetesen továbbra is az emberi
genom evolúciójának megértése, számos
más élőlényben is egyre több *de novo*
fehérjekódoló gént mutatnak ki. Kritikusnak
kell azonban lennünk több szempontból is:
egyrészt a funkcionális *de novo* gének két-
séget kizáró azonosítása alapos kísérleti el-
lenőrzést igényel, másrészt mivel egy igen
gyorsan népszerűvé váló, divatos témáról
van szó, nem feltétlenül könnyű megállni,

hogy ne akarjunk a kelleténél több helyre
is *de novo* géneket „belelátni”. Ezen gének
kétségtelenül szerepet játszanak az élővilág
evolúciójában, pontos azonosításuk, va-
lamint keletkezési mechanizmusuk és bi-
ológiai szerepük tisztázásának érdekében
azonban még hosszú évek kutatásai állnak
előttünk. ☛

Irodalom

- [1] Chen, S., Krinsky, B. H., and Long, M. *New genes as drivers of phenotypic evolution*. Nat. Rev. Genet. 7, (2013) 645-660.
- [2] Ranz, J. M. and Parsch, J. *Newly evolved genes: moving from comparative genomics to functional studies in model systems*. BioEssays 34, (2012) 477-483.
- [3] Djebali, S. et al. *Landscape of transcription in human cells*. Nature 489, (2012) 101-108.
- [4] Li, C.-Y. et al. *A human-specific de novo protein-coding gene associated with human brain functions*. PLoS Comp. Biol. 6, (2010) e1000734
- [5] Knowles, D. G., and McLysaght, A. *Recent de novo origin of human protein-coding genes*. Genome Res. 19, (2009) 1752-1759.
- [6] Xie, C. et al. *Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs*. PLoS Genetics 8, (2012) e1002942

Áprilisi számunkból

Tél Tamás: Örvenypöföktől a turbulenciáig

Dálya Gergely – Hanyecz Ottó – Szabó Róbert: Új feladat vár a bolygóvadásra
Kovács Zsófia–Pálffy József: Sávos vasérc: a rozsdamentes Föld talányos bányakincse

Lente Gábor: Metanolgazdaság – a jövő energiája?

Patkós András: Alkotás és hatás

Kiss L. László: Egy felfedezés története
Vásárhelyi Tamás: Herman Ottó, a tudománykommunikátor

K. A.: A Panama-csatorna

Olvasóink figyelmébe

Februári számunkban *Ujjfaludi László*-nak *A szépség rejtett dimenziói* című cikkében hibásan jelent meg az olvasók további tájékozódását elősegítő weboldal neve. A helyes cím:

<http://www.mystudios.com/artgallery/>