

9. Genomika 1.

1. Bevezetés
2. Genomadatbázisok
3. Számítógépes funkcionális genomika
4. Szerkezeti genomika

Bevezetés

Új korszak hajnala

- Új korszak, paradigmaváltás, forradalom: a teljes genomok ismeretében a biológia adatokban gazdag tudománnyá válik. Új kutatási módszerek, új szemlélet.
- Hajtóerők:
 - ◆ Genomszekvenálási projektek (>60 teljes genom ismert, továbbiak vannak folyamatban)
 - ◆ Új kísérleti módszerek: nagy áteresztőképességű, párhuzamosított eljárások a biológiai minták jellemzésére: microarray technikák
- Bioinformatika (alapfeladata új szekvenciákhoz a funkció hozzárendelése):
 - ◆ Genomok előtti korszak: a bioinformatika csúcspontjai a homológián alapuló módszerek (BLAST, PSI-BLAST, felfűzés, stb.). Ez volt eddig ("klasszikus bioinformatika")
 - ◆ Genomok utáni korszak: egész sor új, nem homológián alapuló eljárás! **Új bioinformatika.**

Új tudományágak

Genomika

- **Genom:** egy élőlény v. sejt teljes gén-, ill. DNS-állománya.
- **Genomika:** a genom megismerése, ill. tanulmányozása, azaz a *teljes genetikai információ* felhasználása, szemben az egyes, kiválasztott gének vagy géncsoportok tanulmányozásával
- **Funkcionális genomika:** a génekhez a funkció hozzárendelése genomikai módszerekkel (számítógépes [*in silico*] és kísérleti)
- **Szerkezeti genomika:** a genomban kódolt fehérjék térszerkezetének kiderítése (számítógépes és kísérleti), ezek felhasználása (pl. a funkcionális genomikában)

Egyéb, teljes információval kapcsolatos fogalmak és tudományágak

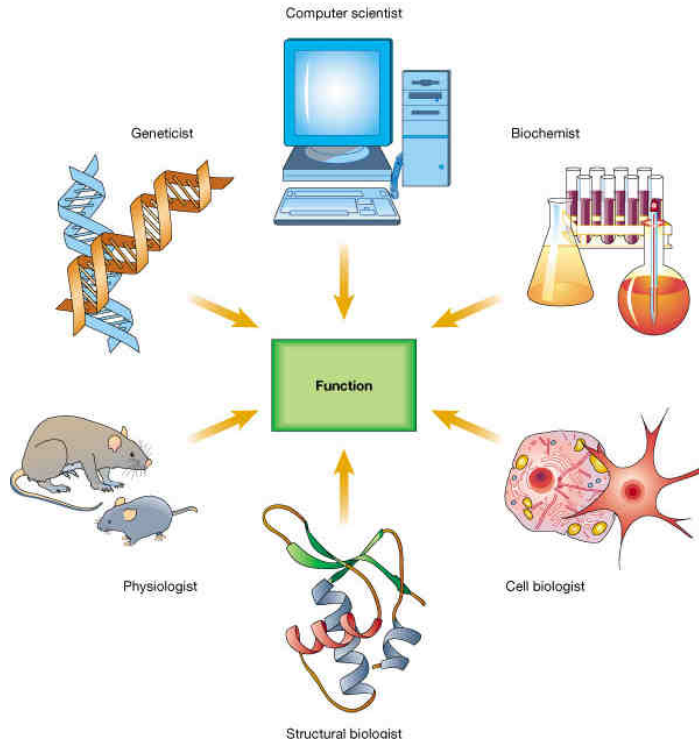
- **Proteom:** egy sejtben (annak adott állapotában) található, ill. expresszált fehérjék összessége
- **Proteomika:** a proteom kutatása, vizsgálata (főleg kísérleti)
- **Transzkriptom, transzkriptomika:** az mRNS-állomány, ill. vizsgálata
- **Metabolom, metabolomika:** az anyagcserehálózat, ill. vizsgálata
- ... stb., az "omikák forradalma"

Genomadatbázisok

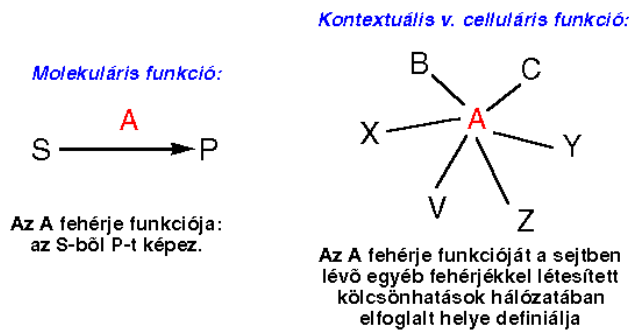
Ld. [Entrez Genome](#): Összes ismert genom, hierarchikus szervezésben (kromoszóma, térképek, gének, stb.)

Számítógépes funkcionális genomika

Mi a génfunkció?

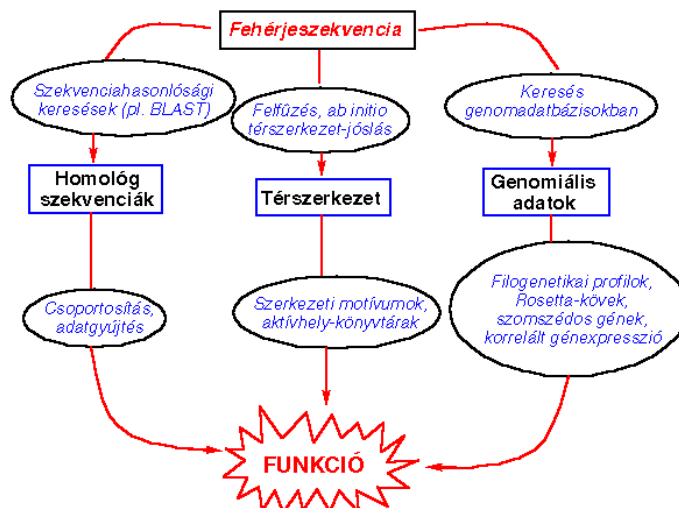


Számos szempont és megközelítés, ezek együttese adja a funkciót.



- A funkció klasszikus jelentése: a *molekuláris funkció* (pl. milyen reakciót katalizál vagy milyen más molekulát köt az adott fehérje)
- A funkció "posztgenomi", bővített jelentése: a *kontextuális v. celluláris funkció* (hol helyezkedik el az adott fehérje a sejt kölcsönhatásainak hálózatában)

Módszerek a funkcionális genomikában



- *Klasszikus bioinformatikai (homológián alapuló) módszerek:* homológia detektálása szekvencia alapján (BLAST, stb.) vagy gombolyfelismeréssel/felfűzéssel. Az ismert funkciójú homológ segítségével azonosítjuk a funkciót.
- *Szerkezeti genomikai módszer:* a kísérleti vagy jóslott szerkezet alapján azonosítjuk a funkciót (ld. később)
- *Nem (vagy csak kismértékben) homológián alapuló módszerek* (új bioinformatikai módszerek):
 - ◆ Tisztán számítógépes:
 - ◇ Filogenetikai profilok módszere
 - ◇ Rosetta–kő módszer
 - ◇ Szomszédos gének módszere
 - ◆ Kísérleti alapú, de számítógépesen kiértékelt:
 - ◇ Korrelált génexpressziók módszere (ld. később)

Tisztán számítógépes módszerek

Filogenetikai profilok módszere

- **Filogenetikai profil:** annak leírása, hogy egy adott gén mely organizmusokban található meg és melyekben nem (teljes genomok ismerete szükséges)
- Az azonos, v. nagyon hasonló (vagy éppen teljesen v. majdnem komplementer) filogenetikai profillal rendelkező gének között funkcionális kapcsolat valószínűsíthető, hiszen ez azt jelenti, hogy a szóban forgó gének mindenhol együtt fordulnak elő.

Példa:

Az **Org1** organizmusban találtunk 7 gént:

Gén1, Gén2, Gén3, Gén4, Gén5, Gén6, Gén7
(fehérjét kódoló gének)

Van még 3 organizmusunk: **Org2, Org3, Org4**, melyeknek a teljes genomja ismert.

A teljes genomokat megvizsgálva megállapítjuk, hogy az egyes organizmusok a következő géneket tartalmazzák:

Org2: Gén1, Gén2, Gén4, Gén5, Gén7
Org3: Gén2, Gén3, Gén5, Gén6, Gén7
Org4: Gén1, Gén3, Gén5, Gén6

Az adatok alapján elkészítjük a 7 gén **filogenetikai profilját**:

Gén	Org2	Org3	Org4
Gén1	1	0	1
Gén2	1	1	0
Gén3	0	1	1
Gén4	1	0	0
Gén5	1	1	1
Gén6	0	1	1
Gén7	1	1	0

(0, ill. 1 jelöli az adott gén hiányát, ill. jelenlétét az adott organizmusban.)

A táblázat alapján megállapíthatjuk, hogy:

Gén2 és Gén7 filogenetikai profilja azonos
Gén3 és Gén6 filogenetikai profilja azonos

Következtetés:

- Gén2 és Gén7 között funkcionális kapcsolat valószínűsíthető
- Gén3 és Gén6 között funkcionális kapcsolat valószínűsíthető

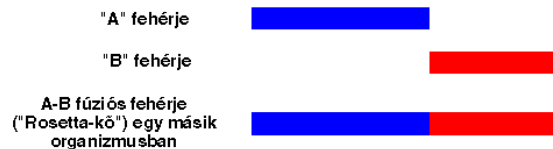
Minél több organizmus teljes genomja áll rendelkezésre a filogenetikai profilok elkészítéséhez, annál megbízhatóbb az eredmény.

Bonyodalmak: bizonyos evolúciós jelenségek megzavarják az elemzést:

- Génfunkciók redundanciája (ugyanarra a funkcióra sokszor több gén van)
- A gén felváltása egy másik génnel, ami nem ortológja az eredeti gén ortológjainak
- Horizontális géntranszfer (mikroorganizmusok közötti DNS–transzfer)
- Gének elvesztése egyes organizmusokban

Rosetta–kő módszer (doménfúziók módszere)

- Sok esetben valamely organizmus két különálló fehérjéje fúziós fehérjeként (egyetlen polipeptidlánccá összeolvadva) található meg valamely más organizmusban.
- Ha két fehérje megtalálható fúziós fehérjeként is, akkor közöttük funkcionális kapcsolat valószínűsíthető. (Számos példát ismerünk rá, hogy rokon funkciót betöltő fehérjék bizonyos organizmusokban fuzionálnak, mert közelségük előnyös a funkció szempontjából.)



Mi a Rosetta–kő?



Az ókori Egyiptomból, Rosetta városából származó kőlap, melyen ugyanaz a szöveg 3 különböző nyelven és írással (hieroglif, démotikus és görög) található meg. Ez tette lehetővé a hieroglifák megfejtését a görög változat alapján. A fúziós fehérjék egyfajta Rosetta–kövek: a bennük lévő, ismert funkciójú domén alapján a másik, ismeretlen funkciójú domén funkciójára lehet következtetni.

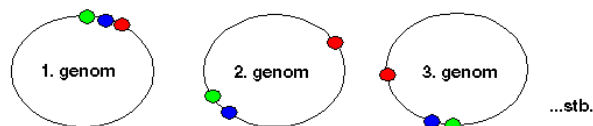
Bonyodalmak:

- Vannak "promiszkuita" domének, amelyek nagyon sok más fehérjével fuzionálnak

Szomszédos gének módszere

- Ha két gén az organizmusok nagy részében egymás mellett található a kromoszómán, akkor valószínűsíthetően funkcionális kapcsolat van közöttük
- Prokariótáknál gyakoriak az *operonok* (több, rokon funkciójú gén egymás után található, egy közös promotor alatt)
- Eukariótáknál az operonok ritkábbak, de a génszomszédság mégis jellemző

Három gén megfigyelt elhelyezkedése bakteriális kromoszómákon



Következtetés: ● és ● között funkcionális kapcsolat valószínűsíthető

Bonyodalmak:

- A szomszédság nem mindig jelent funkcionális kapcsolatot

Szerkezeti genomika

A szerkezeti genomika céljai

- Meghatározni a genomban kódolt összes fehérje térszerkezetét
- Térszerkezeti információ felhasználásával segíteni a funkciók azonosítását (ebben az értelemben a funkcionális genomika része)

A térszerkezetek meghatározása

- Klasszikus megközelítés: előbb azonosítjuk egy kiválasztott fehérje funkcióját, majd röntgenkristallográfiával vagy NMR-rel meghatározzuk a térszerkezetét
- Szerkezeti genomikai megközelítés: előbb meghatározzuk a térszerkezetet (lehetőleg az összes fehérjéét), aztán vizsgáljuk a funkciót (épp a térszerkezet segítségével is)

A fehérjeszerkezetek sokfélesége

- Hány különböző *fold* ("gomboly") van? Többféle becslés volt, 1000 és 100 000 között. Tegyük kb. 10 000-re.
- PDB-ben kb. 12 000 szerkezet van, de erősen redundáns, kb. 1000 gombolyt képviselnek. Az újonnan meghatározott szerkezetek többsége is már ismert gombolyhoz tartozik.
- Teljes genomokban lévő gének által kódolt fehérjéknek csak kb. 15–25%–a mutat homológiát már ismert térszerkezetű fehérjével.

Kísérleti szerkezeti genomika

- Szerkezeti genomika célja: a genomokból kiválasztani azokat a **célfehérjéket**, amelyeknek a térszerkezetét kísérletileg meghatározva az összes többi fehérje *homológiamodellezési távolságon belül* lesz (kb. 20% szekvenciaazonosság), így minden fehérje szerkezete homológiamodellezéssel megjósolható lesz.
- Membránfehérjék, nehezen kristályosítható fehérjék problémát jelentenek.
- Próbaprojektek elindultak, ld. pl. <http://proteome.bnl.gov/> (főleg élesztő).

Elméleti megközelítések

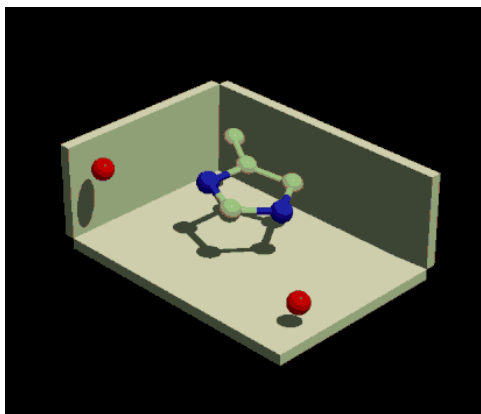
- Az ismert szerkezetű homológgal rendelkező szekvenciákra homológiamodell építhető.
- Pl. az élesztő kb. 6000 génjéből kb. 1000 mutat homológiát ismert szerkezettel, ezek homológiamodellezését elvégezték (MODELLER-rel).
- Hasonló projekt volt a SWISSPROT-ban lévő szekvenciákra: [3DCrunch](#)

A szerkezeti genomika mint a funkcionális genomika egyik módszere

- A szerkezet sokféleképpen segíthet a funkció azonosításában.
- A röntgenkristallográfia és a homológiamodellezés nagy felbontású, jó minőségű szerkezeteket ad, a felfűzési módszerek és az *ab initio* módszerek alacsony felbontású, gyengébb minőségű szerkezetet

Hogyan azonosítható a funkció a szerkezet alapján?

- **A fold (gomboly) alapján:** Ha az adott funkció és gomboly között egyértelmű kapcsolat áll fenn. Ez gyakran nem így van: ugyanazt a funkciót többféle gomboly is megvalósíthatja, és ugyanaz a gomboly többféle funkciót is megvalósíthat (pl. az alfa–béta hordó 16 különböző enzimen fordul elő).
- **Az aktív, ill. funkciós hely geometriája alapján:**
 - ◆ *Részletes funkcióshely–könyvtárak*, pl. [PROCAT](#). Pl. a szerin proteázok konszenzus szerinti aktív helye (Ser–His–Asp katalitikus triád):



Ezek azonosításához azonban nagy felbontású szerkezetre van szükség, gyenge minőségű szerkezetben nem azonosíthatóak.

- ◆ *Fuzzy Functional Forms (FFF)*: Olyan szerkezetimotívum-könyvtár, mely kevésbé egzaktul írja le a funkció helyet. Alkalmazható gyengébb felbontású szerkezeteknél is! ([GeneFormatics](#) cég erre jött létre)
 - Fejlesztés alatt lévő, kiforratlan módszerek!
-