

Jelátviteli útvonalak rendszerszintű adatbázisa és elemzése: új fehérjefunkciók, jelátviteli keresztbeszélgetések és gyógyszercélpont-jelöltek meghatározása

Doktori értekezés

Korcsmáros Tamás

Témavezetők:

Dr. Csermely Péter egyetemi tanár
Semmelweis Egyetem, Orvosi Vegytani Intézet

Dr. Vellai Tibor egyetemi docens
Eötvös Loránd Tudomány Egyetem, Genetikai Tanszék

Dr. Papp Balázs tudományos főmunkatárs
Magyar Tudományos Akadémia, Szegedi Biológiai Központ, Biokémiai Intézet

Biológus Doktori Iskola
Szegedi Tudományegyetem, Természettudományi és Informatikai Kar



Szeged
2011

Tartalomjegyzék

TARTALOMJEGYZÉK	2
RÖVIDÍTÉSEK JEGYZÉKE	4
I. BEVEZETÉS	5
I.1. ÁLTALÁNOS BEVEZETŐ	5
I.2. A JELÁTVITELI ÚTVONALAK KUTATÁSÁNAK IRODALMI HÁTTERE	7
I.2.1. Biokémiai és proteomikai megközelítés.....	8
I.2.2. Genetikai és genomikai megközelítés	10
I.2.3. Sejtbiológiai megközelítés.....	13
I.2.4. Bioinformatikai és rendszerbiológiai megközelítés.....	14
I.3. ÖSSZEFOGLALÁS	27
I.4. A JELÁTVITELI HÁLÓZATOK	29
II. CÉLKITŰZÉSEK.....	32
III. MÓDSZEREK	34
III.1. EGY JELÁTVITELI ADATBÁZIS KÉSZÍTÉSE: SIGNALINK.....	34
III.1.1. Az adatbázis készítésének forrásai.....	34
III.1.2. Az adatbázis készítésének módszertana	39
III.2. MÁS JELÁTVITELI ADATBÁZISOKKAL VALÓ ÖSSZEHASONLÍTÁS	47
III.2.1. Az összehasonlítás háttere.....	47
III.2.2. A KEGG adatbázissal való összehasonlítás.....	48
III.2.3. A Reactome adatbázissal való összehasonlítás	49
III.2.4. A NetPath adatbázissal való összehasonlítás	49
III.3. AZ ÚTVONAL-HÁLÓZATOK ÁBRÁZOLÁSA	50
III.4. AZ ADATBÁZISON ELVÉGZETT ELEMZÉSEK	51
III.4.1. Génfunkció predikciók.....	51
III.4.2. A keresztbeszélgetések expressziós vizsgálata	54
III.4.3. Gyógyszercélpont-jelöltek meghatározása	56
IV. EREDMÉNYEK	57
IV.1. JELÁTVITELI ÚTVONALAK EGYSÉGES GYŰJTEMÉNYE	57
IV.1.1. A <i>SignaLink</i> adatbázis minőségi tulajdonságai.....	57
IV.1.2. A <i>SignaLink</i> adatbázis mennyiségi tulajdonságai	58
IV.2. A SIGNALINK ADATBÁZIS ÖSSZEHASONLÍTÁSA MÁS ADATBÁZISOKKAL	61
IV.3. A SIGNALINK ADATBÁZIS ÚTVONALAINAK ÉS HÁLÓZATÁNAK ÁBRÁZOLÁSAI	64
IV.3.1. Az útvonalak részletes ábrázolása	65
IV.3.2. A jószolt komponensek ortológia-alapú ábrázolása	68
IV.3.3. Az útvonal-hálózatok ábrázolása.....	68
IV.4. A SIGNALINK ADATBÁZISON VÉGZETT VIZSGÁLATOK EREDMÉNYEI	70
IV.4.1. Új génfunkciók azonosítása és elemzése	71
IV.4.2. Összehasonlító vizsgálatok	76
IV.4.3. Keresztbeszélgetések expressziós vizsgálata	84
IV.4.4. Gyógyszercélpont-jelöltek meghatározása.....	87

V. AZ EREDMÉNYEK MEGVITATÁSA	91
V.1. A SIGNALINK ADATBÁZIS.....	91
V.1.1. A <i>Signalink</i> adatbázis előnyei.....	91
V.1.2 A <i>Signalink</i> adatbázis hátrányai.....	93
V.2. A SIGNALINK ADATBÁZISON VÉGZETT VIZSGÁLATOK.....	94
V.2.1. Új génfunkciók azonosítása	94
V.2.2. Az útvonalak és a keresztbeszélgetések összehasonlítása.....	94
V.2.3. Új gyógyszercélpont-jelöltek azonosítása.....	97
V.3. A SIGNALINK ADATBÁZISON VÉGEZHETŐ TOVÁBBI VIZSGÁLATOK	98
V. ÖSSZEFOGLALÁS.....	101
VI. SUMMARY	103
IRODALOMJEGYZÉK.....	105
SAJÁT PUBLIKÁCIÓK JEGYZÉKE.....	121
A DISSZERTÁCIÓHOZ KAPCSOLÓDÓ KÖZLEMÉNYEK	121
A DISSZERTÁCIÓTÓL FÜGGETLEN KÖZLEMÉNYEK	121
ÖSSZESÍTÉS	122
KÖSZÖNETNYILVÁNÍTÁS	123

Rövidítések jegyzéke

- AP – Affinitás-tisztítás (*affinity purification*)
- C. elegans* – *Caenorhabditis elegans*
- CGC – *Cancer Gene Census* adatbázis
- COSMIC – *Catalogue of Somatic Mutations in Cancer* adatbázis
- D. melanogaster* – *Drosophila melanogaster*
- DNS – dezoxiribonukleinsav
- EGF – epidermális növekedési faktor (*Epidermal Growth Factor*)
- GO – *Gene Ontology*
- H. sapiens* – *Homo sapiens*
- Hh – Hedgehog
- HTP – nagy áteresztő képességű (*high-throughput*)
- IGF – Inzulinhoz hasonló növekedési faktor (*Insulin-like Growth Factor*)
- Ins – Inzulin
- JAK/STAT – Janus aktivált kináz / Jelátviteli és transzkripció aktivátor (*Janus Activating Kinase / Signal Transducer and Activator of Transcription*)
- MAPK – mitogén aktivált fehérje kináz (*Mitogen Activated Protein Kinase*)
- MS – Tömegspektrometria (*mass spectrometry*)
- NHR – Nukleáris hormonreceptor (*Nuclear Hormone Receptor*)
- RNS – ribonukleinsav
- RNSi – RNS interferencia
- TF – transzkripció faktor
- TGF- β – Transzformáló növekedési faktor (*Transforming Growth Factor beta*)
- WNT – Wingless és int fehérjéhez hasonló
- Y2H – élesztő kettős-hibrid (*yeast-2-hybrid*)

I. Bevezetés

I.1. Általános bevezető

A jelátviteli útvonalak (más néven szignáltranszdukciós pályák) alapvető szerepet játszanak számos sejtteni folyamat szabályozásában, mint amilyen a sejtosztódás, differenciáció, sejtpusztulás és anyagcsere. Továbbá fontos szerepük van az immun- és hormonrendszer működésének, valamint a stressz-adaptáció és az öregedési folyamat szabályozásában. A jelátviteli rendszerek orvosi biológiai fontosságát jelzi, hogy hibás működésük sokféle rendszerszintű betegség (pl. rák, cukorbetegség, neurodegeneratív elváltozások) kiváltó oka lehet.

A jelátviteli útvonalak sejten kívülről érkező jeleket érzékelnek és továbbítanak a sejttag felé. E kívülről érkező jelek (ligandumok) specifikus receptorokhoz kötődnek, majd jelerősítő és jeltovábbító molekulák segítségével az általuk közvetített biológiai információ a sejttagba jut, ahol specifikus génexpressziós mintázatot generál. Érdekes módon a jelátviteli pályák száma (típusa) viszonylag alacsony, és egy-egy pályát csupán néhány – max. 10-20 – fehérje alkotja. Ez látszólagos ellentmondásban áll a jelátviteli útvonalak által meghatározott sejtípusok sokféleségével. Az elmúlt évtized kutatásai rámutattak arra, hogy a jelátviteli útvonalak nem önállóak, hanem számos keresztbeszélgetést (ún. *signaling cross-talk*) alkotnak. Ezért ma már a jelátviteli pályák helyett inkább jelátvitel hálózatokról beszélünk, amelyeket rendszerszemléleti alapon, hálózatos módszerekkel érdemes vizsgálni.

A jelátviteli útvonalakat alkotó fehérjékről és azok működéséről ma már nagy mennyiségű adat áll rendelkezésre (Bader és mtsai, 2006). Többféle olyan adatbázis létezik, amely ezeket az információkat tartalmazza és rendszerezi. Az útvonal adatbázisok egy része speciális, meghatározott szempontok és vizsgálati célok alapján készültek, így rendszerszintű elemzésre vagy összehasonlításra általában nem alkalmasak (pl.: KEGG (Ogata és mtsai, 1999)). Az adatbázisok másik típusát általános céllal hozták létre; nagyrészt az elmúlt években ún. nagy áteresztőképességű (*high-throughput*; *HTP*) módszerekkel generált adatokat tartalmaz. A HTP vizsgálatok

molekuláris kapcsolatok ezreit képesek kimutatni, így a hálózatkutatások során gyakran használják ezeket az adatbázisokat. Ugyanakkor jelentős részük a jelátvitel szempontjából fontos (extracelluláris, membránkött, illetve sejtmagi) fehérjéket módszertani okokból nem tartalmazzák. További módszertani probléma ezen adatbázisoknál, hogy általában irányítatlan kapcsolatokat tartalmaznak. E tulajdonságok miatt a kézi gyűjtéssel készült adatbázisok nélkülözhetetlenek. A kézi gyűjtésű jelátviteli adatbázisok általában kevesebb információt tartalmaznak, de az adatok részletesebbek és megbízhatóbbak. A legtöbb ilyen manuális jelátviteli adatbázis azonban pontos útvonal-definíció és szabványosított gyűjtési feltételek nélkül készült. Ennek egyik következménye az, hogy még egy adott adatbázison belül sem lehet (vagy szabad) az ott lévő útvonalakat összehasonlítani vagy a közöttük lévő kapcsolatokat vizsgálni.

Ezeket a szempontokat figyelembe véve kidolgoztunk egy új eljárást és egy számítógépes programot a jelátviteli útvonalak egységes adatbázisának elkészítésére. Az eljárás a kísérletes adatok pontosan meghatározott kézi gyűjtésén alapul. Az így elkészített adatbázist SignaLinknek neveztük el, és három többsejtű élőlény (a *Caenorhabditis elegans* fonálféreg, a *Drosophila melanogaster* gyümölcsleány és az ember *Homo sapiens*) fontosabb jelátviteli útvonalainak hálózatát tartalmazza. A hálózatok irányított kapcsolatokat tartalmaznak, és mindegyik kapcsolatnál feltüntettük az azt leíró irodalmi hivatkozásokat. Az adatbázis 444 *C. elegans*, 211 *D. melanogaster* és 525 emberi jelátviteli fehérjét tartalmaz, amelyek között 237, 233, illetve 991 jelátviteli kapcsolatot találtunk. A korábban publikált útvonal-adatbázisokkal végzett összehasonlításunk rámutatott a SignaLink adatbázis előnyeire és hátrányaira. Az adatbázisban lévő útvonal hálózatokat a különböző vizsgálatoknak megfelelően ábrázoltuk, és interaktív formában elérhetővé tettük. A vizsgált három élőlény integrált jelátviteli hálózatának összehasonlítása számos új, kísérletesen még nem igazolt jelátviteli kapcsolatot és funkciót jósolt meg. Összesen 88 *C. elegans*, 92 *D. melanogaster* és 73 emberi fehérje új jelátviteli funkcióját javasoltuk. A jelátviteli rendszerek felépítésének megismerése céljából összehasonlítottuk a három faj jelátviteli útvonalait és a közöttük lévő kapcsolatokat, majd minőségi és mennyiségi alapon elemeztük az adatokat. Expressziós mintázatok felhasználásával megvizsgáltuk a

keresztbeszélgetések szövet-specifikus tulajdonságait. Az eredményeknek széleskörű gyakorlati vonatkozása is van, mivel az összehasonlító bioinformatikai vizsgálatok segítségével új gyógyszer-célpont-jelölteket sikerült azonosítanunk.

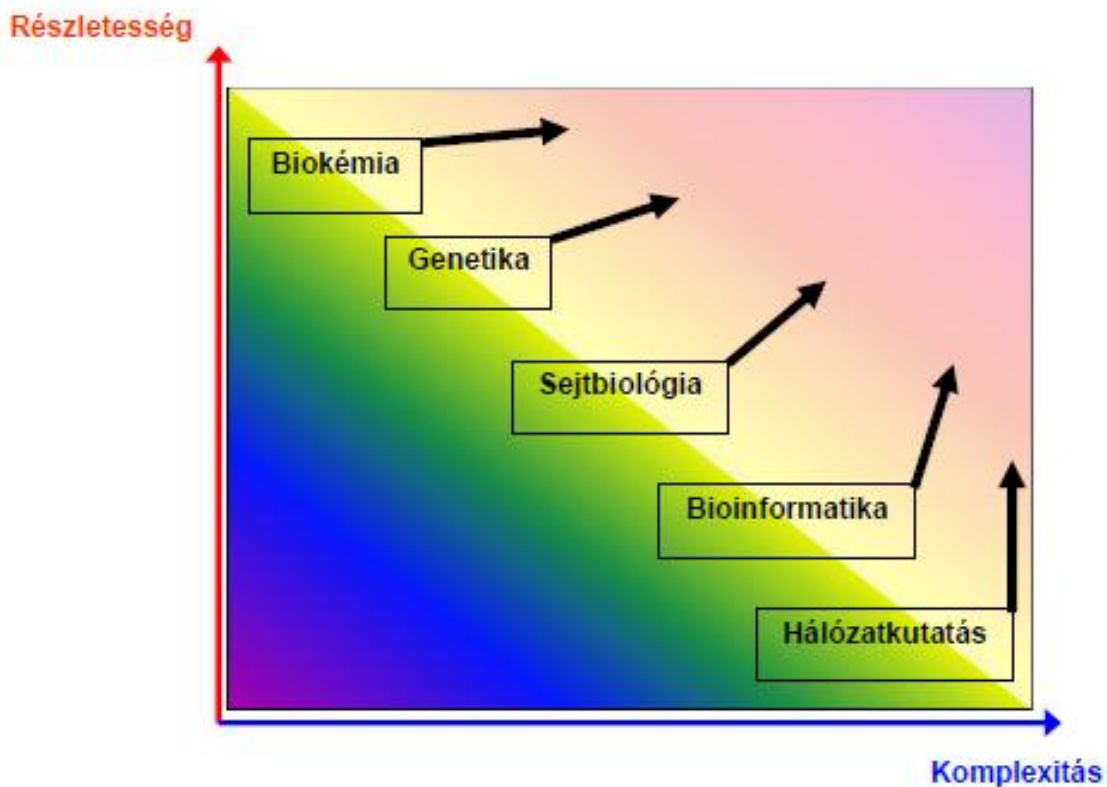
Disszertációmban csak olyan vizsgálatok szerepelnek, amelyek elvégzésében közvetlenül vettem részt. Az általam önállóan végzett vizsgálatokat egyes szám első személyben, a kollégáimmal és TDK hallgatóimmal közösen végzett vizsgálatokat többes szám első személyben írom. Az „Összefoglalás” részben felsorolom az önállóan és az együttműködésben végzett vizsgálatokat.

1.2. A jelátviteli útvonalak kutatásának irodalmi háttere

A jelátviteli útvonalak vizsgálata többféle biológiai diszciplína tárgya. A biokémiai, genetikai és sejtbiológiai kutatások évtizedek óta vizsgálják a jelátviteli rendszerek szerkezetét és funkcióját, míg a genomikai, a proteomikai és a rendszerbiológiai kutatások eredményei csak az utóbbi évtizedben jelentek meg. A rendszerbiológia három igen különböző részterületet foglal magába: a bioinformatikai adatfeldolgozást, a rendszer-szintű modellezést és a hálózatkutatást. Ezek a megközelítések többféle szinten vizsgálják a jelátviteli rendszereket. Az alacsonyabb komplexitású, részletesebb vizsgálatoktól (biokémia) a komplexebb, de az összetevőkről viszonylag kevés információt feltáró megközelítésekig (rendszerbiológia, hálózatkutatás) egyaránt nyújt értékes adatokat (**1. ábra**).

A jelátviteli kérdéseket vizsgáló egyes megközelítési módok között hasonlóságokat és különbségeket is találhatunk. A megközelítési módokat az alábbi szempontok szerint mutatom be:

- Mi a megközelítés tárgya?
- Milyen vizsgálati módszert használ a megközelítés?
- Milyen eredményekhez vezet a megközelítés a jelátviteli rendszerről?
- Melyek a főbb céljai az adott megközelítési módnak?

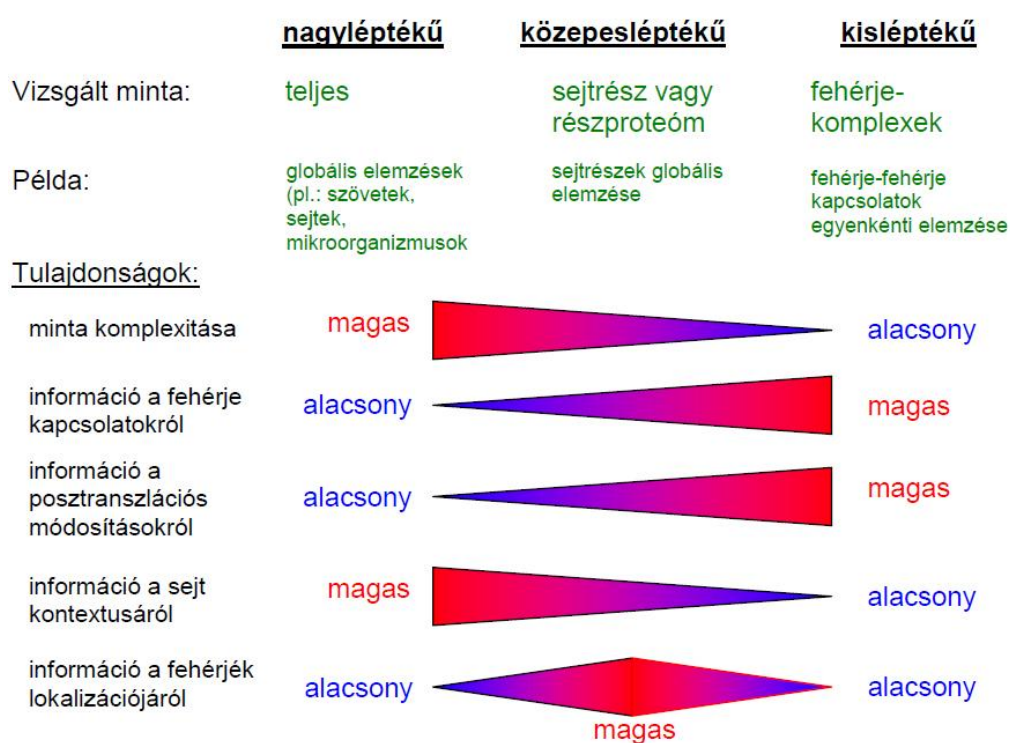


1. ábra A jelátvitelt vizsgáló megközelítések összehasonlító skálája. Az ábrán a részletesség és a komplexitás függvényében látható a jelátvitel vizsgálatában központi szerepet betöltő öt diszciplína. Ha a jelenlegi tudásunk szintjét úgy határozzuk meg, hogy a biokémia a legrészletgazdagabb, de legkevésbé komplexebb, míg a hálózat kutatás a legkomplexebb, de legkevésbé részletes eredményeket produkál, valamint a többi megközelítés közöttük helyezkedik el, akkor egy vonalba rajzolhatjuk őket. Ez alatt ismerjük a jelátviteli rendszer működését. Az egyes diszciplínák mind haladnak a teljes megismerés felé (fekete nyilak). A két szélsőséget nézve, a biokémia egyre inkább a komplexitás, míg a hálózat kutatás a részletgazdagság felé halad.

1.2.1. Biokémiai és proteomikai megközelítés

A biokémiai megközelítés alapvető tárgya a jelátvitelben résztvevő komponensek (elsősorban fehérjék) szerkezetének és működésének a megismerése. Forrásai a kis léptékű (*small-scale*) szerkezeti biológiai és fehérje-dinamikai vizsgálatok, valamint újabban a nagyobb léptékű proteomikai kísérletek. A biokémiai megközelítések során szerkezeti vizsgálatokat (pl.: röntgen-krisztallográfia, mágneses magrezonancia (NMR)) alkalmaznak, illetve a fehérjék működését mérő enzimkinetikai méréseket. A proteomikai vizsgálatok célja a fehérjék dinamikai állapotának megismerése, a fehérje-fehérje kapcsolatok és az ezek hatására bekövetkező poszttranszlációs módosítások nagyléptékű detektálása. A proteomikai megközelítések egyre többet használt módszere az affinitás-tisztítást követő tömegspektrometria (AP+MS), amely a fehérjekomplexek vizsgálatára igen alkalmas eljárás (Xia és mtsai,

2004). A 2D gélelektroforézis tömegspektroszkópiával történő kombinálása a fehérje-fehérje kapcsolatok által végbement változásokat (pl. foszforiláció) tudja detektálni (Zhu és Snyder, 2002). Hasonló célokat szolgál az immunprecipitációt követő kromatográfia és tandem tömegspektrometria (LC-ESI-MS/MS) (Ewing és mtsai, 2007). Értékes eredményeket adnak az élesztő kettős-hibrid (Y2H) és ennek továbbfejlesztett technikái (pl.: ubikvitin-Y2H, MAPPIT (*Mammalian Protein-Protein Interaction Trap*)), valamint a protein *chip* módszerek (Cutler, 2003; Huang, 2004; Venkatesan és mtsai, 2009; Xia és mtsai, 2004). Ezek a technikák nagyszámú fehérje kapcsolatról adnak információt. Egy nagy léptékű proteomikai vizsgálatban az egér májsejtjeiben lévő összes fehérje foszforilációs állapotait határozták meg tömegspektrometriai módszerekkel (Villen és mtsai, 2007). **(2. ábra)** A proteomika mellett fontos megemlíteni a lipidomikát is, amely hasonló vizsgálatokkal és szerkezeti megközelítésekkel a lipidek tulajdonságait vizsgálja (Balogh és mtsai, 2010; Ivanova és mtsai, 2009). Jelenleg még nem áll rendelkezésre megfelelő mennyiségű adat, hogy felhasználjuk a lipidomikát a jelátvitel kutatásában. Feltehetően azonban ez a megközelítés pár éven belül számos új eredményekkel fog szolgálni ezen a területen is.



2. ábra A proteomikai megközelítések kiterjedésének (skálájának) és fontosabb tulajdonságainak összehasonlító ábrája. (Schubert és mtsai, 2006)

A biokémia szerkezeti alapú megközelítéssel él, amelynek tárgyai lehetnek fehérjekomplexek, fehérjék vagy újabban egyre inkább a fehérjéket felépítő funkcionális egységek, a domének. A fehérjedomének vizsgálata egyre célszerűbb megközelítési mód: a fehérjék moduláris felépítése következtében a domének kapcsolatainak, evolúciójának vizsgálata logikusabban és egyszerűbben kivitelezhető, mint ha egész fehérjéket vizsgálnának (Bhattacharyya és mtsai, 2006). A jelátvitel egy nagymértékben szabályozott folyamat. A szabályozást a biokémiai megközelítés a résztvevő elemek szintjén vizsgálja. A szabályozó domének fontos szerepet töltenek be a fehérje-fehérje kapcsolatok kialakításában (Pawson és Nash, 2003), illetve az effektor funkcióval rendelkező fehérjék irányításában, útvonalba rendezésében (Bhattacharyya és mtsai, 2006). Mark Gerstein neves fehérjehálózat-kutató egy újszerű, hálózatos megközelítéssel bizonyította a domének számának szerepét a fehérjehálózatok dinamikai működésében (Kim és mtsai, 2006). A biokémiai szemlélet a résztvevő elemek, fehérjék szerkezetének felderítése révén kívánja megismerni a jelátviteli útvonalak felépítését és működését.

A biokémia és a proteomika hosszú távú célja a jelátvitelben résztvevő összes komponens szerkezetének, működésének megismerése, továbbá az ezek között kialakuló közvetlen kapcsolatok molekuláris és kinetikai felderítése.

1.2.2. Genetikai és genomikai megközelítés

A genetika - szemben a biokémiával - főleg funkcionális oldalról közelíti meg a jelátviteli rendszereket. Megközelítésében konkrét elemek (gének, fehérjék) hiányának vagy túlműködésének fenotípusát vizsgálja. Ennek a megközelítésnek a rendszerszintű vizsgálata a genomika. A genomikai vizsgálatok a teljes genomi szekvenciák ismeretéből indulnak ki (Blencowe és mtsai, 2009; Zhu és Snyder, 2002).

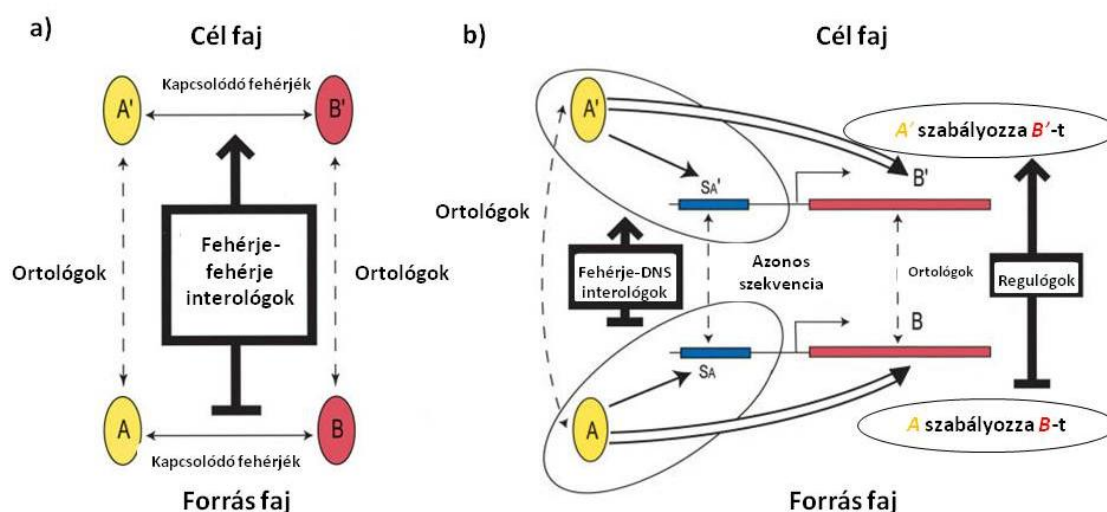
A klasszikus genetika mutáns élőlények generálásán és *in vivo* vizsgálatán alapszik. Az egyszeres mutánsok kombinálásával (episztázis vagy kettős mutáns elemzés) lehetőség nyílik a genetikai interakciók feltárására, amelyek gyakran nem közvetlen (fizikai), hanem indirekt (szabályozási) kapcsolatokat mutatnak. A genetikai megközelítés tehát funkcionális kapcsolatokból épít működési rendszert, hálózatot.

A nagyszámú genetikai analízis integrált vizsgálata és összehasonlítása már rendszerszemléletű megközelítést igényel. A funkcionális kapcsolatokat érdemes összevetni a közvetlen fehérje-fehérje kapcsolatokkal (Kelley és Ideker, 2005). Két neves *Drosophila* kutató (Friedman és Perrimon) egy átfogó tanulmányban kifejtik, hogy a klasszikus genetikának és a molekuláris biológiának mekkora lendületet adott az elmúlt évtizedben kidolgozott RNSi (RNS interferencia) technika (Friedman és Perrimon, 2007). Az RNSi technika az eukarióta sejtekben jelenlévő mRNS gátló mechanizmust használja fel. Kétszálú RNS-ekből rövid, egyszálú siRNS-ek vagy miRNS-ek keletkezhetnek, amelyek specifikusan képesek egy adott gén funkcióját „csendesíteni”. Az RNSi technika igen értékes módszernek bizonyult a gén és géntermék jelátviteli útvonalakban betöltött szerepének meghatározásában (Friedman és Perrimon, 2007).

Az egyszeres mutánsok elemzése „lineáris”, különálló jelátviteli útvonalakat vizsgál. Eközben egyre több adat bizonyítja, hogy az önálló jelátviteli pályák számos ponton keresztbeszélgetnek egymással (*signaling cross-talk*), így egy sokkal bonyolultabb jelátviteli hálózatot alkotnak (Friedman és Perrimon, 2007). A genetikusok az útvonal komponenseket a betöltött funkciók (fenotípusok) alapján határozzák meg. Ezért a genetikai megközelítésben két útvonal közötti kapcsolat (keresztbeszélgetés) két különálló funkció között fellépő (funkcionális) kapcsolatnak tekinthető. Érdekes módon, a biokémikusok ezeket az útvonalak közötti kapcsolatokat gyakran nem is tekintik keresztbeszélgetésnek, mivel az összekötött útvonalak szerkezetileg és evolúciósan hasonlóak.

A genomika sok (potenciálisan egy genomban található összes) gén működését vizsgálja. Ebben a hagyományos genetikai kísérleti módszerek és a bioinformatikai szekvencia elemző alkalmazások nyújtanak segítséget. A genetika és a bioinformatika összefonódását a funkcionális genomikai kutatásokban érhetjük leginkább tetten. Ezen kutatások korábbi kísérletek eredményeit (például funkciók vagy szekvenciák meghatározása) hasonlítják össze az evolúciós történések azonosítására, illetve új információkat jósolnak az összehasonlítások eredményeként (Ruvkun és Hobert, 1998). Új funkciókat például szekvencia hasonlóság alapján lehet jósolni: egy ismert tulajdonság átvitelével az egyik fajból a másikba. Ezáltal szerkezeti (pl.: doménösszetétel), expressziós vagy fizikai kapcsolatok jósolhatók (Salgado és mtsai,

2008; Storm és Sonnhammer, 2003; Yellaboina és mtsai, 2008). A kapcsolatok jóslására Vidal kutatócsoportja kidolgozta az ún. interológ koncepciót, amelyben fehérje-fehérje kapcsolatokat jósolnak két olyan fehérje között, amelynek ortológjai egy másik fajban ismert módon kapcsolódnak (Yu és mtsai, 2004). Később azt találták, hogy az interológok kevésbé konzerváltak, mint maguk az ortológok (azaz a kapcsolatok gyorsabban változnak az evolúció során, mint maguk a fehérjék) (Persico és mtsai, 2005). Az interológ kapcsolatok a *high-throughput* (HTP) vizsgálatokban kapott kapcsolatokhoz képest is több hamis pozitív kapcsolatot tartalmaznak, azaz kevésbé megbízhatóak (Cusick és mtsai, 2009). Mégis, az interológok pontos definíciója és az, hogy a HTP eredmények megbízhatóságát könnyebben megbecsülhetővé tették (Jensen és mtsai, 2009; Li és mtsai, 2008) hozzájárult ahhoz, hogy elterjedt megközelítés legyen a funkcionális genomikában (Brown és Jurisica, 2005; Huang és mtsai, 2004; Kemmer és mtsai, 2005; Persico és mtsai, 2005). Az interológ koncepciót dolgozta tovább Gerstein és csoportja a fehérjék és a DNS közötti regulációs kapcsolatok jóslására (regulógok), ahol egy transzkripciós faktor és egy DNS szakasz (transzkripciós faktor kötőhelye) kapcsolatát jósolták hasonló ortológiai-alapú módszerrel (Yu és mtsai, 2004). Egy újabb megközelítés a betegség fenotípus információ fajok közötti (fenológ) átvitelét javasolta (McGary és mtsai, 2010).



3. ábra Kapcsolatok jóslása ortológia alapján. a) Két ismert kapcsolódó fehérjepár alapján jósolható az ortológ fehérjepár kapcsolata, amelyet fehérje-fehérje interológoknak neveznek. **b)** A fehérje-fehérje interológokhoz hasonlóan fehérje-DNS interológokat is lehet jósolni, amelyek alapján szabályozási kapcsolatokat is lehet jósolni. Ezeket regulógoknak nevezték el. (Yu és mtsai, 2004)

A genetika és ezen belül a genomika is azon összetevőket tudja vizsgálni, amelyeknek a kapcsolatai ismertek, illetve rendelkeznek detektálható funkcióval vagy expresszióval. A genomikai megközelítésmód fő célja a genomban található összes gén funkciójának és funkcionális kölcsönhatásának megismerése. A modern bioinformatika ehhez a célhoz járul hozzá úgy, hogy egyre több gén funkcióját képes jósolni és felderíteni.

I.2.3. Sejtbiológiai megközelítés

A sejtbiológia a jelátviteli útvonalak sejten belüli elhelyezkedésével és az útvonalak által szabályozott biológiai folyamatok működésével foglalkozik. A sejtbiológiai kutatások gyakran más tudományterületek eredményein alapulnak. Ezek főleg biokémiai eredmények, például egy folyamat reakciójának részletes ismerete (Weng és mtsai, 1999), a résztvevők koncentrációja, sejten belüli grádiense (Levchenko, 2003), vagy a résztvevők szerkezeti tulajdonsága (Hlavacek és mtsai, 2003). A sejtbiológiai kutatások fókuszában az egyes összetevők sejten belüli lokalizációja és a sejtek dinamikai állapotai vannak (Weng és mtsai, 1999). Hasonlóan a genomikához, a sejtbiológia területén is megjelent a bioinformatika, amely a vizsgálatok során keletkezett sok adat átlátható kezelésével támogatja a kísérletes kutatókat. (Weng és mtsai, 1999). A bioinformatika a tárolt adatok elemzésében és *in silico* modellezésében is egy fontos eszköz. Mivel ez nagyban átfed a rendszerbiológiával, így ezekről később részletesebben lesz szó.

A sejtbiológusok általában nem az egész sejtet és nem az összes sejttani folyamatot vizsgálják, hanem kisebb (kezelhetőbb) egységeket. Gyakran csak egy konkrét sejt-kompartimentumot vagy jelátviteli folyamatot vizsgálnak, illetve ezek időbeli változásait különböző stimulusok hatására (White és Anderson, 2005). A legtöbb sejtbiológiai kutatás figyelembe veszi a keresztbeszélgetések együttes tér és időbeni jelenlétét. Azaz kizárnak olyan lehetséges kapcsolatokat, amelyek az eltérő expresszió miatt a valóságban nem következnek be. Így bár csökken a keresztkapcsolatok száma, de a megbízhatóságuk nő (Levchenko, 2003). A sejten zajló folyamatok pontos megismerése során keletkezett adatokat adatbázisokba rendezik, amelyekből modelleket lehet készíteni. Ezek a modellek segíthetnek átláthatóvá tenni a sejtekben zajló jelátviteli folyamatokat.

1.2.4. Bioinformatikai és rendszerbiológiai megközelítés

A bioinformatika és a rendszerbiológia (*systems biology*) számítástechnikai eszközökkel vizsgálja a korábban már bemutatott diszciplínák, a biokémia, a proteomika, a genetika, a genomika és a sejtbológia adatait. A bioinformatika a biológiai adatok minőségi és mennyiségi növekedésével és a számítástechnika fejlődésével párhuzamosan alakult ki. A rendszerbiológia – felhasználva a bioinformatikai úton elemzett és adatbázisokban tárolt információkat – különböző felbontású és komplexitású modellek segítségével vizsgálja a sejtteni folyamatok felépítését és működését. A bioinformatikai és rendszerbiológiai megközelítések jelátvitellel kapcsolatos tárgyalását érdemes három részre bontani: (1) a nagyméretű adatbázisok készítésére és kezelésére, amely a bioinformatika egyik fő területe; (2) ezen adatok biológiai értelmezésére, a rendszerszintű folyamatok modellezésére és szimulációk végzésére; végezetül (3) a nagy jelátviteli rendszerek szerkezetét és dinamikáját más komplex rendszerekkel közvetlenül összehasonlítható módon vizsgáló hálózat kutatásra.

1.2.4.1. Bioinformatikai adatbázisok

Az adatbázisok vagy közvetlen kísérleti eredmények alapján keletkeznek vagy adatbázisok egymásra, egymáshoz illesztéséből jönnek létre. A jelátvitelben fontos adatbázisok gyűjteménye Lindvall összefoglaló cikkében található meg (Lindvall és mtsai, 2003), illetve a PathGuide nevű internetes gyűjteményben (<http://pathguide.org>; (Bader és mtsai, 2006). Ezek közül a legismertebbeket és a legjellemzőbbeket az **1. táblázat** emeli ki. A jelátviteli kapcsolatokat bemutató adatbázisok listája található Ideker és Lauffenburger hálózat-kutatók összefoglaló cikkében (Ideker és Lauffenburger, 2003).

A bioinformatikai megközelítések másik trendje az előzőekben bemutatott adatbázisok értelmezése. Az ide tartozó módszerek alkalmasak arra, hogy biológiai információt nyerjünk a bonyolult adatokból (Lindvall és mtsai, 2003; Neves és Iyengar, 2002). A bioinformatikai elemzés megközelítéseit mutatja be a **2. táblázat**.

1. táblázat A jelátviteli adatbázisok és jellemzőik. Ebben és a következő táblázatokban a „negatívumok” általában nem az adott cikkekre, vagy az abban bemutatott eredményekre vonatkoznak, hanem arra, hogy az integratív és rendszerszintű megismerésre törekedő megközelítések szempontjából a cikk mely részei nem megfelelőek. (A referenciaként feltüntetett első szerzők ABC-sorrendjében.)

Jelátvitellel kapcsolatos adatbázisok			
Elnevezés	Referencia	Pozitívum	Negatívum
REACTOME: molekuláris reakció gyűjtemény, útvonalanként csoportosítva	(Joshi-Tope és mtsai, 2005)	<ul style="list-style-type: none"> Sok faj adatai Minden reakciónak (kapcsolatnak) és elemnek saját úrlapja van Manuálisan készített, hivatkozásokkal ellátott gyűjtemény 	<ul style="list-style-type: none"> Nem hálózatos megközelítés, csak reakciónként vizsgálhatunk egy útvonalat Nem jelátvitel centrikus Sok útvonal nem szerepel benne Nem egységes: önálló, saját meghatározáson alapuló reakciócsoportokat tartalmaz, amelyek között léteznek átfedések (ezeket azonban nem jelzik)
NetPath: Fehérjeinterakciók jelátviteli útvonal adatbázisa	(Kandasamy és mtsai, 2010)	<ul style="list-style-type: none"> A kapcsolatokhoz hivatkozás is van rendelve 	<ul style="list-style-type: none"> Adatok csak az emberből Csak a rákban és az immun-folyamatokban fontos útvonalak Irányítatlan kapcsolatok
TRANSPATH: Útvonalgyűjtemény	(Krull és mtsai, 2006)	<ul style="list-style-type: none"> Jelátviteli útvonalak összetevőit és kapcsolatait tartalmazza Manuális gyűjtés Nemcsak fehérjetermészetű molekulákat ábrázol 	<ul style="list-style-type: none"> A teljes adatbázishoz csak anyagi hozzájárulással lehet hozzáférni
KEGG: sok folyamatot, köztük jelátviteli útvonalakat tartalmazó gyűjtemény	(Ogata és mtsai, 1999)	<ul style="list-style-type: none"> Nagy adatbázis Manuálisan készített Az egyes elemeknek külön adatlapja van Több faj útvonalait tartalmazza külön-külön 	<ul style="list-style-type: none"> Funkcionális vagy nem pontosan meghatározott szempontok alapján lettek rendszerezve az útvonalak A kapcsolatokról nincs információ Szubjektív gyűjtéssel készült

1. táblázat (folytatás)

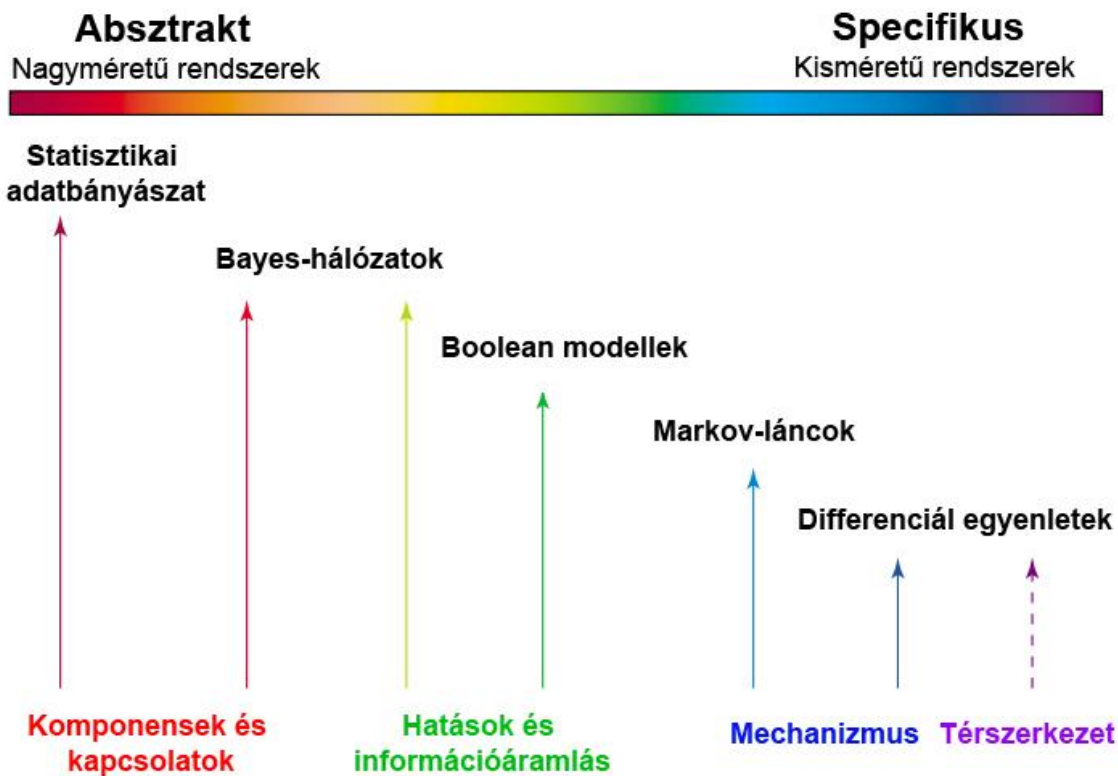
Elnevezés	Referencia	Pozitívum	Negatívum
Cell Circuits: sejthálózat-modelleket összesítő gyűjtemény és eszköz	(Mak és mtsai, 2007)	<ul style="list-style-type: none"> • Az eddigi hálózatos eredményeket gyűjti össze cikkekből és adatbázisokból, és lehetővé teszi ezen modellek együttes vizsgálatát és ábrázolását • Jelátviteli útvonal-centrikus adatbázis • Több faj adatait tartalmazza 	<ul style="list-style-type: none"> • Még kevés adatbázis/cikk van feltöltve • Csak egy integráló eszköz, eredményeket, elemzéseket nem közöl
BIOCARTA: Jelátviteli útvonalgyűjtemény	http://biocarta.com	<ul style="list-style-type: none"> • Sok útvonal adatait tartalmazza • Irányított kapcsolatok 	<ul style="list-style-type: none"> • Funkcionális vagy nem pontosan meghatározott szempontok alapján lettek rendszerezve az útvonalak • A kapcsolatokról nincs információ • Szubjektív gyűjtéssel készült
NCI Nature – Pathway Interaction Database: Jelátviteli útvonalgyűjtemény	http://pid.nci.nih.gov	<ul style="list-style-type: none"> • Sok útvonal adatait tartalmazza • Irányított kapcsolatok • A kapcsolatok-hoz van hivatkozás is 	<ul style="list-style-type: none"> • Csak emberi adatok • Szubjektív gyűjtéssel készült • Funkcionális vagy nem pontosan meghatározott szempontok alapján lettek rendszerezve az útvonalak
Science Signaling – STKE: Jelátviteli útvonal gyűjtemény	http://sciencesignaling.org	<ul style="list-style-type: none"> • Sok útvonal adatait tartalmazza • Irányított kapcsolatok • Sok kapcsolathoz van hivatkozás is 	<ul style="list-style-type: none"> • Funkcionális vagy nem pontosan meghatározott szempontok alapján lettek rendszerezve az útvonalak • Szubjektív gyűjtéssel készült

2. táblázat A bioinformatikai elemzések jellemzői. (A referenciaként feltüntetett első szerzők ABC-sorrendjében.)

Bioinformatikai elemzések megközelítései			
Fontosabb jellemző	Referencia	Pozitívum	Negatívum
A sejtes hálózatok kvantifikálására létrehozott kezelő és tároló rendszer	(Campagne és mtsai, 2004)	<ul style="list-style-type: none"> • Sejt- és útvonal független • Pontos hálózatdinamikai vizsgálatokat tesz lehetővé 	<ul style="list-style-type: none"> • Nem áll rendelkezésre még elegendő adat a használatához, még kisebb rendszeren sem
Jelátviteli útvonal adatok és miRNS adatok együttes vizsgálata	(Cui és mtsai, 2006)	<ul style="list-style-type: none"> • Két különböző rendszer integratív elemzése 	<ul style="list-style-type: none"> • Nagyon speciális jelátviteli útvonal-adatok felhasználása
Lineáris útvonalak azonosítása fehérje-fehérje interakciók hálózatokban	(Huffner és mtsai, 2007)	<ul style="list-style-type: none"> • Meglévő hálózatos információk segítségével jelátviteli információk generálása 	<ul style="list-style-type: none"> • Lineáris útvonalakat állít elő
Kinázok és szubsztrátjaik szisztematikus keresése	(Linding és mtsai, 2007)	<ul style="list-style-type: none"> • Kísérletes adatok és bioinformatikai módszerek összekapcsolása • A kapott eredmények részletes elemzése 	<ul style="list-style-type: none"> • Csak kinázok
Általános keretrendszer felállítása a jelátviteli hálózatok vizsgálatára	(Papin és Palsson, 2004)	<ul style="list-style-type: none"> • Sztöchiometria és reakció alapú jelátviteli vizsgálat és útvonal definiálás 	<ul style="list-style-type: none"> • Kis rendszeren végezhető csak el az elemzés
Fehérje-fehérje interakciók és expressziós mintázatok alapján a jelátviteli hálózat rekonstruálása	(Steffen és mtsai, 2002)	<ul style="list-style-type: none"> • Integratív megközelítés • Meglévő információk segítségével hálózati információk generálása 	<ul style="list-style-type: none"> • Csak élesztőn tesztelték • Csak egy útvonalra
Bioinformatikai munka a rendelkezésre álló bizonyítékok felhasználásáról fehérje funkciók jóslására	(Xiong és mtsai, 2006)	<ul style="list-style-type: none"> • Teljes genomra készült • Még ismeretlen szekvenciákhoz funkció párosítás különböző kísérletes bizonyítékokat tartalmazó adatbázisok alapján 	<ul style="list-style-type: none"> • Csak élesztőn tesztelték • A fehérjéknek csak a Gene Ontology adatbázisban lévő tulajdonságait adja meg a program

1.2.4.2. Rendszerbiológiai modellezések

A rendszerbiológiai modellezés megközelítésének kiterjedését mutatja be a **4. ábra**. Itt jól látható, hogy minél nagyobb a vizsgált rendszer, annál egyszerűbb (absztraktabb) modelleket tudunk csak készíteni, viszont ezekről általánosabb megállapításokat tehetünk. Szélsőséges esetben szinte már csak egyszerű hálózatokról beszélünk, ahol pusztán a rendszer elemeit és a közöttük lévő kapcsolatok létét látjuk. A kiterjedés túloldalán az egzakt matematikai modellekkel, differenciálegyenletekkel leírt specifikus és kisebb rendszerek modellezését találjuk. Általánosságban az mondható, hogy az egyszerűsítő megközelítés inkább minőségi, míg a kisebb rendszereket egzakt módon vizsgáló kvantitatív eredményeket ad (Ideker és Lauffenburger, 2003). Amikor rendelkezünk a céljainknak megfelelő modellel, akkor különböző szimulációkat hajthatunk végre működésük tesztelésére (Neves és Iyengar, 2002). Később a szimulációkkal kapott eredményeket validálhatjuk kísérletekkel vagy a kísérleti eredmények illeszkedését vizsgálhatjuk a modellhez szimulációs eljárásokkal (Bundy és mtsai, 2007). Egyes kísérleti eredményeket megerősíthetnek szimulációk. Például a költséges vagy kockázatos kísérletek tervezést elősegíthetik az előzetes szimulációk. Ha valamilyen technikai nehézség miatt egy adott kontroll kísérlet nem végezhető el, azokat helyettesítheti egy, az adott kísérleti rendszerről készült szimuláció. A jelátvitel működésével kapcsolatos reprezentatív modellezési megközelítéseket tartalmazza a **3. táblázat**.



4. ábra A rendszerbiológiai modellek kiterjedésének összehasonlító ábrája. A színskála bal oldalán az absztrakt, nagyméretű rendszerek modellezése látható, míg a jobb oldalon a specifikus, kisméretű, de részletes modellek. Balról jobbra haladva négy csoportot különböztethetünk meg. A legabsztraktabb (differenciálegyenleteken és Markov-láncon alapuló) modellek csak a rendszeralkotó elemeket és az azok között lévő kapcsolatokat veszik figyelembe. Ezeket követik a kisebb, de bonyolultabb rendszerek működését (pl.: az információ áramlást) modellező Boolean modellek. Ezután nagyobb, de egyszerűbb rendszerek egyszerűbb algoritmusok általi modellezése található (Bayes-hálózatok). Az ábra bal oldalán a kis rendszerek speciális, statisztikai vizsgálata helyezkedik el. (Ideker és Lauffenburger, 2003).

3. táblázat A bioinformatikai modellezések és jellemzőik. (A referenciaként feltüntetett első szerzők ABC-sorrendjében.)

Bioinformatikai modellezések megközelítései a jelátvitel kutatásban			
Fontosabb jellemző	Referencia	Pozitívum	Negatívum
Irányítás alapú matematikai szemlélet fontossága a jelátviteli hálózatok működésének megértésében	(Araujo és Liotta, 2006)	<ul style="list-style-type: none"> • Hálózatszemlélet 	<ul style="list-style-type: none"> • Egyszerű, kis rendszeren végezhető csak el az elemzés
Pusztán az útvonalakra nem jellemző, emergens jelátviteli hálózati tulajdonságok kimutatása differenciálegyenletekkel	(Bhalla és Iyengar, 1999)	<ul style="list-style-type: none"> • Kísérleten alapuló adatok • Több útvonal (4) együttes vizsgálata 	<ul style="list-style-type: none"> • Nagyobb rendszeren nem vagy csak nehezen alkalmazható • Az útvonalak funkcionális alapon lettek definiálva
Matematikai modell valós jelátviteli hálózat felderítésére funkcionális alapon	(Kholodenko és mtsai, 2002)	<ul style="list-style-type: none"> • Kvantifikálni szeretné a jelátvitelt • Funkcionális alapon csoportokra bontja a hálózatot 	<ul style="list-style-type: none"> • Útvonalspecifikus • Sok becsült paraméter szükséges hozzá • Nagyobb rendszeren nem, vagy csak nehezen alkalmazható
A jelterjedés specifikusságának és hűségének a vizsgálata	(Komarova és mtsai, 2005)	<ul style="list-style-type: none"> • Két fontos jelátviteli rendszerszintű tulajdonság vizsgálata • Az útvonalak metszetében lévő, un. közös elemek szerepének elemzése 	<ul style="list-style-type: none"> • Elméleti, kevés kísérletes példa
A jelátvitel és a fluxus-egyensúly analízis (FBA) összekapcsolása egy specifikus példán	(Krauss és Brand, 2000)	<ul style="list-style-type: none"> • Kísérletes munka • Kvantifikálhatóvá teszi a jelátvitelt 	<ul style="list-style-type: none"> • Sejt- és útvonalspecifikus • Roppant kis rendszeren nézték • Nagyobb rendszeren nem, vagy csak nagyon nehezen alkalmazható
Differenciálegyenletek a sejtciklus dinamikai modellezésére	(Tyson és mtsai, 2003)	<ul style="list-style-type: none"> • Pozitív és negatív visszacsatolások pontos hatása, működése 	<ul style="list-style-type: none"> • Sok becsült paraméter szükséges hozzá • Nagyobb rendszeren nem, vagy csak nehezen alkalmazható
A visszacsatolások rendszerszintű vizsgálata a jelátvitelben	(Wolkenhauer és mtsai, 2005)	<ul style="list-style-type: none"> • Rendszerszintű megközelítés • A visszacsatolások részletes matematikai elemzése 	<ul style="list-style-type: none"> • Csak MAPK útvonalra

1.2.4.3. Hálózatos és rendszerszintű megközelítések

A hálózatos megközelítés csak a biológiában újkeletű. A szociológusok csaknem száz éve, a fizikusok pedig két évtizede alkalmazzák a hálózatos módszereket. A biológiában először az ökológia területén használták a hálózattudományt, és csak miután az utóbbi pár évben a rendelkezésre álló adatok mennyisége és minősége elérte a szükséges szintet, akkor jelent meg a biokémia, genetika, sejtbiológia területén és így a jelátvitel kutatásában is (Barabasi és Oltvai, 2004; Csermely, 2006). A hálózatos megközelítés lényege, hogy a komplex rendszert, mint komponensek (pontok) és kapcsolatok (élek) összefüggő, akár dinamikusan változó együttesét veszi alapul. A komponensek közötti kapcsolatoknak lehet iránya, ami a jelátvitel vizsgálatához nélkülözhetetlen. Emellett egyes kapcsolatoknak lehet erőssége és súlya, amit a gyakoriság vagy a bizonyítékok mennyisége alapján rendelhetünk élekhez. A hálózatos módszerek egy igen dinamikusan fejlődő területe a hálózatos csoportok, modulok meghatározása (Adamcsek és mtsai, 2006; Kovacs és mtsai, 2010).

Léteznek olyan megközelítések, amelyek nem egy adott rendszer működésének a megértésével foglalkoznak, hanem a rendszer csoportra jellemző (ún. emergens) tulajdonságait vizsgálják. Az emergens tulajdonságok csak a rendszer egészének ismeretében értelmezhetőek. Ilyen emergens tulajdonság például az adott rendszer robusztusságának, szenzitivitásának vagy stabilitási állapotainak vizsgálata (Neves és Iyengar, 2002). A rendszer komplexitásának megismerését a sejtbiológiában az egy helyen vagy egy időben működő jelátviteli egységekkel próbálják átláthatóbbá tenni (White és Anderson, 2005). A bioinformatikában és a rendszerbiológiában az átláthatóságot az adatok csoportosításával, a résztvevő folyamatok, útvonalak modularizálásával próbálják elérni (Neves és Iyengar, 2002). Az adatokat csoportosító eljárások egy része az adatbányászat módszerével él, míg egyre inkább terjednek ezen a téren is a hálózatos megoldások.

A hálózatos megközelítésekkel kapcsolatos módszereket három csoportba lehet osztani (Xia és mtsai, 2004). Az első kettőbe a forrásadatok összeállítása és kiegészítése tartozik, amelyekről korábban már volt szó, míg a harmadik kifejezetten a topológiai vizsgálatokról szól. Az első módszertani csoportba (forrásadatok összeállítása) azok a kísérletes módszerek tartoznak, amelyek nagy áteresztő képességűek. Ezek a nagy áteresztő képességű eljárások nagymennyiségű kapcsolati információkat állítanak elő, a

proteomikai vizsgálatok kapcsán már részletesen felsoroltuk ezeket. A hálózatos vizsgálatokhoz alkalmazott különböző kapcsolattípusok irány és forrás szerinti csoportosítása látható a **4. táblázatban**. A következő csoportba (forrásadatok kiegészítése) azok a bioinformatikai és kísérletes megközelítések tartoznak, amelyek szekvencia, szerkezet vagy akár hálózatos topológia alapján jósolják a kapcsolatokat. A harmadik csoportba (topológiai vizsgálatok) tartoznak azok a programok, amelyek a hálózatok topológiáját, csoportjait határozzák meg, illetve azokat hasonlítják össze és elemzik. Az utóbbi években egy negyedik módszertani csoportról is beszélhetünk. Ide azok az újabb hálózatdinamikai megközelítések tartoznak, amelyek például a játékelmélet hálózatos megközelítése felé is haladnak (pl.: (Kiss és mtsai, 2009; Kovacs és mtsai, 2005)).

4. táblázat A hálózati kapcsolattípusok irány és forrás szerinti csoportosítása. (Beyer és mtsai, 2007).

A kapcsolat iránya	A kapcsolat forrása	
	Fizikai (fehérje-fehérje interakció)	Funkcionális (genetikai interakció)
Irányított	<ul style="list-style-type: none"> Fehérje-gén Fehérje-RNS Fehérje-fehérje (<i>protein chip</i>) Receptor-ligandum 	<ul style="list-style-type: none"> Episztázis elemzés Gén kiütés/csendesítés (deléciós mutánsok, RNSi)
Irányítatlan	<ul style="list-style-type: none"> Fehérje-fehérje (TAP–MS, Y2H) Gén-gén (ko-reguláció) DNS–DNS (kromoszóma-kapcsolatok) 	<ul style="list-style-type: none"> Kettős-letális vizsgálat Kombinatorikus RNSi

A hálózatos vizsgálatok megközelítése két irányból, fentről lefelé vagy lentől felfelé valósul meg. A hagyományos megközelítés (biokémia) lentől felfelé építkezik, és a rendszer összetevői felől halad a teljes rendszer felé (Stevens, 2004). Ezzel szemben a fentről lefelé megközelítés az újabb, amely a rendszer megismerése felől halad a rendszeralkotók felé. A hálózatos, fentről lefelé történő megközelítés lényege, hogy először csak kevés információt tudunk a rendszerről, és ez alapján végzünk elemzéseket (Bray, 2003). Erre a molekuláris szintű hálózatok jó példák. Ugyanakkor, ha túl sok információt akarunk a hálózatos megközelítéssel vizsgálni, különböző problémákba ütközünk: minél részletesebben ismerjük a hálózatot, annál kisebb

rendszer tudunk csak vizsgálni (Bornholdt, 2005). Természetesen e vizsgálat optimuma folyamatosan javul a számítástechnika fejlődésével. A jelátvitel hálózatos vizsgálata pontosan ezt a problémát veti fel, ugyanis a metabolikus hálózat mellett ez az egyik legjobban ismert molekuláris hálózat. Ennél a megközelítésnél a kapcsolatok irányai mindenképp előkerülnek, viszont az összetevőkről didaktikai és tudománytörténeti okokból torz és nem objektív képünk van. A jelátvitel hálózatos ábrázolásához, rekonstrukciójához a vizsgálati szintet kell először definiálni, és ahhoz objektíven hűnek kell maradni. Vizsgálni lehet a hálózati topológia függvényében az egyes pontokat (fehérjéket), pontcsoportokat (fehérjekomplexek), útvonalakat, teljes kapcsolati hálózatokat. Mindehhez három különböző forrástípust szoktak használni a pontokról, a kapcsolatokról és a hálózat működéséről (Papin és mtsai, 2005). A jelátvitel rekonstrukciójának hálózatos és integrált megközelítésének példáiról szól az **5. táblázat** és a **6. táblázat**, ahol bemutatom az útvonalak elemzésének és egy sejt jelátvitelének hálózatos és az integrált megközelítéseit, valamint ezek előnyeit és hátrányait.

5. táblázat Az útvonalak hálózatos elemzésének jellemzői. (A referenciaként feltüntetett első szerzők ABC-sorrendjében.)

Útvonalak elemzésének hálózatos megközelítései			
Fontosabb jellemző	Referencia	Pozitívum	Negatívum
Az emlős sejtek TGF-β útvonalának fehérje-fehérje kapcsolatainak felderítése	(Barrios-Rodiles és mtsai, 2005)	<ul style="list-style-type: none"> • Kísérletes munka • Hálózatszemplélet • Az útvonal beillesztése a sejtés hálózatba • Eredetileg nem a TGF-β útvonalhoz rendelt fehérjék és kapcsolatok azonosítása 	<ul style="list-style-type: none"> • Sejt- és útvonalspecifikus • TGF-β centrikus megközelítés • Csak egyfajta, ráadásul élesztő kettős-hibrid módszerrel vizsgálták a kapcsolatokat
Összefoglaló munka egy kanonikus útvonal (TGF-β) hálózatba illesztéséről	(Lutz és Knaus, 2002)	<ul style="list-style-type: none"> • Részletes útvonalelemzéssel bizonyítja, hogy az útvonal nem önállóan, hanem egy komplex rendszer részeként működik és áll szabályozás alatt 	<ul style="list-style-type: none"> • Csak egy útvonalra nézték • A kereszt-beszélgetéseket hasonlóan problémás definícióval meghatározott útvonalakhoz nézték
A Toll-receptor család jelátviteli hálózatának elkészítése	(Oda és Kitano, 2006)	<ul style="list-style-type: none"> • Kísérletes munka • Hálózat- és rendszerszemlélet • Hálózatábrázolás 	<ul style="list-style-type: none"> • Sejt- és jelspecifikus
Biokémiailag és hálózatos szempontból is speciális, fontos fehérjék azonosítása az inzulin jelátviteli útvonal szabályozásában	(Taniguchi és mtsai, 2006)	<ul style="list-style-type: none"> • Kísérletes bizonyítékok alapján • Biokémiailag új és célravezető tulajdonság-együttes meghatározása 	<ul style="list-style-type: none"> • Csak egy útvonalra nézték, és ebben a formában nehezen általánosítható a definíció hálózatos vizsgálatra
A <i>C. elegans</i> TGF-β jelátviteli rendszerének vizsgálata igen sok kapcsolat-felderítő eljárás segítségével	(Tewari és mtsai, 2004)	<ul style="list-style-type: none"> • Kísérletes munka • Erőteljes hálózatszemplélet • Különböző modern technikai eredmények kombinálása • Bővíthető kapcsolati oldalról 	<ul style="list-style-type: none"> • Útvonal-, illetve rendszerspecifikus • A résztvevő elemek néha előre meghatározottak • Nagyobb rendszeren nehezen és költségesen alkalmazható

6. táblázat A jelátviteli rendszerek hálózatos megközelítése, fontosabb jellemzői. (A referenciaként feltüntetett első szerzők ABC-sorrendjében.)

Adott sejtek jelátvitelének hálózatos megközelítései			
Fontosabb jellemző	Referencia	Pozitívum	Negatívum
Rákos sejtek jelátviteli hálózatának globális vizsgálata	(Cui és mtsai, 2007)	<ul style="list-style-type: none"> Hálózatos megközelítés Nagyméretű adatsor A rákkal kapcsolatos adatok (mutáció, metilációs) integrálása 	<ul style="list-style-type: none"> Nem pontos útvonal definíció
Jelátviteli útvonalak globális illesztése	(Diao és mtsai, 2007)	<ul style="list-style-type: none"> Hálózatos megközelítés Nagyméretű adatsor 	<ul style="list-style-type: none"> KEGG útvonaladatok (az összeillesztésük nem mindig megfelelő – ld. később) Csak ember
Egy adott szerv (a <i>C. elegans</i> vulva) fejlődését irányító rendszer számítógépes vizsgálata	(Fisher és mtsai, 2005)	<ul style="list-style-type: none"> Kísérleten alapuló adatok Dinamikus modell készítése 	<ul style="list-style-type: none"> Szerv- és útvonalspecifikus
Összefoglaló a visszacsatolások szerepéről a jelátviteli folyamatokban	(Freeman, 2000)	<ul style="list-style-type: none"> Átfogó tanulmány Sok kísérletes példával 	<ul style="list-style-type: none"> Folyamat és nem hálózatcentrikus megközelítés
Sejtkultúrán végzett jel és jelválasz meghatározás ellentétes hatású citokinekre	(Gaudet és mtsai, 2005)	<ul style="list-style-type: none"> Kísérletes munka Nagyméretű keletkezett adatsor 	<ul style="list-style-type: none"> Sejt- és jelspecifikus
A <i>C. elegans</i> korai embrionális fejlődésében résztvevő ún. molekuláris gépezetek szisztematikus felderítése	(Gunsalus és mtsai, 2005)	<ul style="list-style-type: none"> Kísérletes munka Erőteljes hálózatszemlélet Különböző modern technikai eredmények kombinálása Bővíthető kapcsolati oldalról Útvonalfüggetlen megközelítés 	<ul style="list-style-type: none"> Sejt- és rendszerspecifikus
Egy adott szerv (a <i>C. elegans</i> vulva) fejlődését irányító rendszer felderítése	(Inoue és mtsai, 2005)	<ul style="list-style-type: none"> Kísérletes munka Útvonalfüggetlen megközelítés 	<ul style="list-style-type: none"> Szervspecifikus

6. táblázat (folytatás)

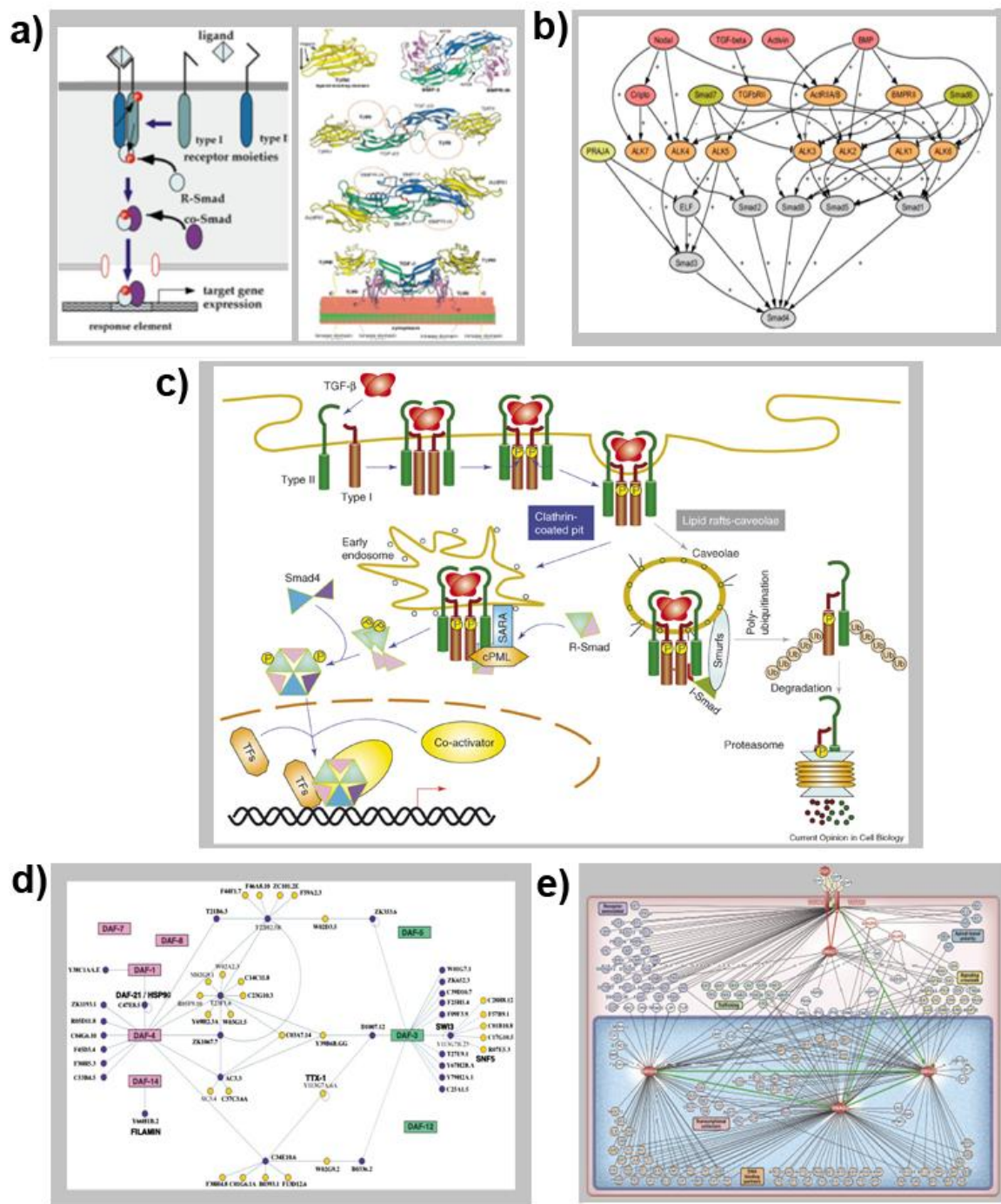
Fontosabb jellemző	Referencia	Pozitívum	Negatívum
Esszenciális résztvevők azonosítása a növényi melléksejt egyik útvonalában	(Li és mtsai, 2006)	<ul style="list-style-type: none"> • Kísérletes munka • Hálózatszemlélet • Robusztusság vizsgálat • Dinamikus modell készítése 	<ul style="list-style-type: none"> • Sejt-, jel- és útvonalspecifikus • Kis rendszeren nézték
Makrofág sejtkultúrán végzett teljes jelátviteli hálózat vizsgálat	(Natarajan és mtsai, 2006)	<ul style="list-style-type: none"> • Kísérletes munka • Hálózatszemlélet • Keresztbeszélgetés felderítése top-down megközelítéssel 	<ul style="list-style-type: none"> • Sejt- és jelspecifikus
Makrofág sejtkultúrán végzett, teljes molekuláris hálózat felmérés	(Oda és mtsai, 2004)	<ul style="list-style-type: none"> • Kísérletes munka • Hálózatszemlélet • Sok kapcsolat felderítése, sok fehérje között 	<ul style="list-style-type: none"> • Sejt- és szignálspecifikus • Nem publikálták Pubmed által jegyzet lapban
A vizsgált sejt összes fehérjéit vizsgálva tömegspektrometriá-val azonosították azok foszforilációs helyeit, és meghatározták a reakciókat térben és időben	(Olsen és mtsai, 2006)	<ul style="list-style-type: none"> • Kísérletes munka • Nagyméretű adatsor • Útvonal független megközelítés 	<ul style="list-style-type: none"> • Sejtspecifikus eredmények (ráadásul rákos sejtvonalon készültek) • Nem hálózatos (kapcsolat orientált) megközelítés • Csak foszforiláción alapuló reakciókat vizsgáltak
T-sejtek foszforilációs hálózatának kísérletes és hálózatos vizsgálata	(Stoevesandt és mtsai, 2007)	<ul style="list-style-type: none"> • Kísérletes munka • Hálózatszemlélet • Sejtenkénti vizsgálat 	<ul style="list-style-type: none"> • Csak foszforilációs reakciókat tartalmazó útvonalak • Sejtspecifikus

A bioinformatikai adatbázisok célja a rendelkezésre álló biológiai adatok testreszabott együttes kezelése, értelmezése. A rendszerbiológia célja az adatok alapján a valósághoz minél közelebb álló modellek generálása és szimulációk elvégzése, akár egyes jól ismert jelátviteli utakról, akár ezek együtteséről van szó. A hálózatos kutatásokhoz szükséges, hogy minél több adat álljon rendelkezésre, amelyek alapján a biológiai és egyéb komplex rendszerek hálózatos felépítéséről és dinamikai működéséről tudunk meg rendszerszintű információkat.

1.3. Összefoglalás

A bemutatott tudományos megközelítésekre általánosságban jellemző, hogy egymásra épülő vizsgálati szinteken nézik az adott rendszer felépítését és működését, valamint ennek megfelelő módszerekkel állítanak elő eredményeket, modelleket, adatbázisokat vagy szimulációkat. Az előzőekben ismertetett kutatási megközelítéseket az **5. ábra** mutatja be a TGF- β útvonal példáján. A jelátviteli megközelítési példák alapján a következő megállapításokat tehetjük:

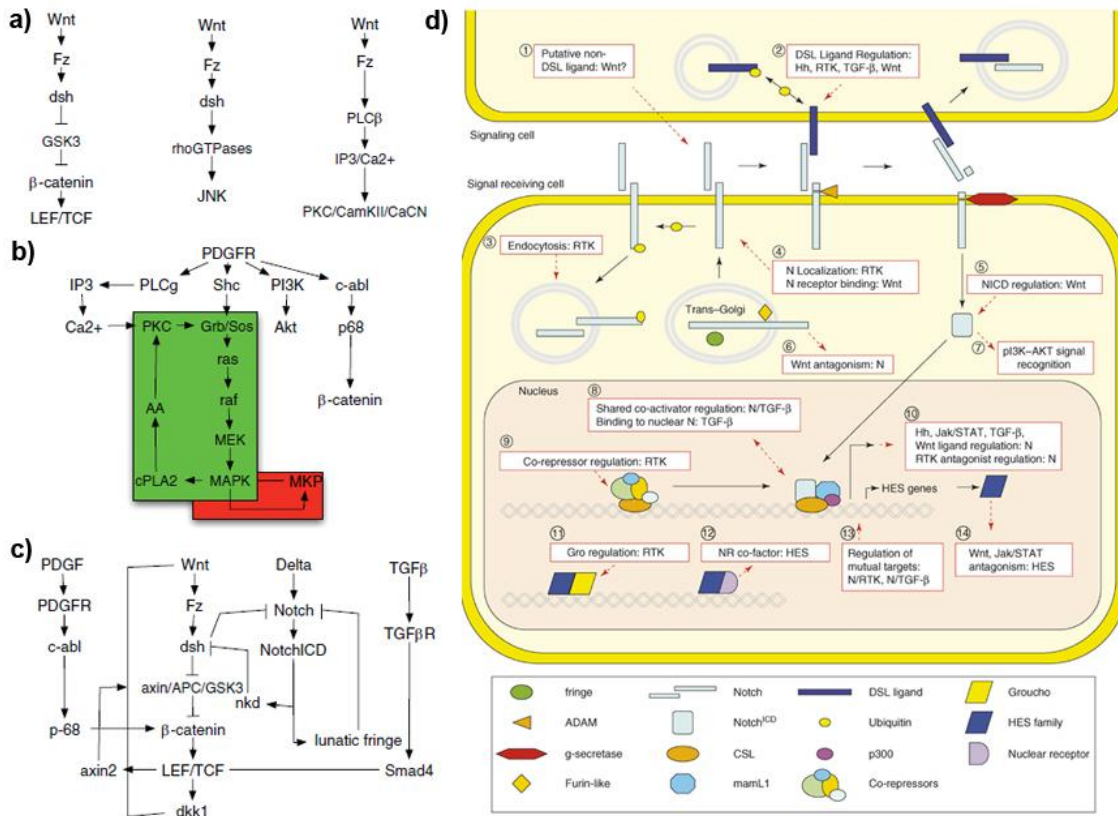
- Egyre több tanulmány kísérleteken alapuló hálózatos megközelítéssel él.
- Egyre több rendelkezésre álló adatot használnak fel, hogy nagyobb és biztosabb modellt készítsenek.
- A hálózat működésének és (kvantifikált alapokon nyugvó) dinamikájának a megértése egyre centrálisabb kérdés.
- Minden megközelítés valamiben specifikus (pl. a vizsgált rendszerben, kapcsolatokban) és egy-két paraméterében meg általános (pl. útvonal-függetlenség, résztvevő elemek).
- Sok színvonalas, nagyobb eredményeket produkáló munka negatívuma, hogy a jelátviteli hálózat csak egy részére vonatkozik. A teljes útvonalhálózattal nem készült még átfogó vizsgálat.
- Összességében megállapítható, hogy túl sok a specifikus megközelítés. Mind általános szabályokat, mind megbízható következtetéseket nehéz ilyen eredményekből levonni. A specifikus megközelítéseknek óriási hátránya, hogy nehezen vethetőek össze, hiszen az eredmények kísérlet- és rendszerspecifikusak.



5. ábra A jelátviteli útvonalak kutatási megközelítései a TGF- β útvonal példáján. **a)** A biokémiai megközelítés az útvonal összetevők szerkezeti megközelítésére fókuszál. **b)** A genetikai megközelítés a kapcsolódó fehérjék funkcionális szerepére a kapcsolat hatására (serkentő vagy gátló) fókuszál. **c)** A sejtszintű megközelítés során a jelátvitelhez köthető egyéb folyamatok (pl.: endocitózis, proteasomális degradáció) is megjelennek. **d)** A bioinformatikai megközelítések esetén különböző forrásokból integrálják a genetikai és biokémiai adatokat. **e)** A hálózatos megközelítéseknél nagyon sok komponens van jelen, amelyek között csak a kapcsolat megléte ismert, további részletek nem. (Itoh és ten Dijke, 2007; Kitisin és mtsai, 2007; Raftery és Sutherland, 1999; Tewari és mtsai, 2004)

I.4. A jelátviteli hálózatok

A jelátviteli útvonalak kutatása nemrégiben egy komoly paradigmaváltáson esett át. Egy évtizeddel ezelőtt még a fejlett eukarióták esetében is a jelátviteli útvonalakat önálló, egymástól jól elkülönülő lineáris molekuláris (genetikai vagy fehérje) kaszkádokként kezelték. Az előző fejezetben bemutatásra került módszerek fejlődésének köszönhetően egyre több jelátviteli keresztbeszélgetést (ún. *signaling cross-talk*) írtak le az egyes útvonalak között. Ennek következtében ma már a jelátviteli rendszereket egységes jelátviteli hálózatként kezelik, amelynek a sűrűn kapcsolt útvonalak a főbb építőkövei (**6. ábra**) (Kestler és mtsai, 2008; Papin és mtsai, 2005).



6. ábra A jelátviteli útvonalak és hálózatok felépítése, a keresztbeszélgetések jelentősége. **a)** A WNT ligandum által aktivált lineáris elrendezésű jelátviteli útvonalak. **b)** Pozitív és negatív visszacsatolási hurkok egy növekedési faktor (PDGF) által aktivált MAPK kaszkádban. A pozitív visszacsatolás zölddel, a negatív pirossal van jelölve. Itt az is látható, hogy az aktiválás hatására nemcsak a MAPK, hanem egyéb útvonalak felé is haladhat a jel. **c)** A PDGF, a WNT, a Notch és a TGF-β jelátviteli útvonalak közötti keresztbeszélgetések által alkotott jelátviteli hálózat. **d)** A humán Notch útvonal és az ezzel kapcsolatban álló további útvonalak. Ezen a példán látható, hogy a Notch útvonal szinte minden komponensére hatnak más útvonalak (piros dobozból kimenő piros nyilak), illetve az, hogy a Notch útvonal is több másik útvonalra hat (piros dobozba menő fekete nyilak). A kétirányú piros nyilak jelölik az oda-vissza hatás. **a-c)** (Kestler és mtsai, 2008) **d)** (Hurlbut és mtsai, 2007)

Annak ellenére, hogy a hálózatos elképzelés ma már általánosan elterjedt, a jelátviteli útvonalak definíciója csak keveset változott. Léteznek szerkezeti, funkcionális, szövet-, és betegség-specifikus szempontok alapján meghatározott útvonalak. Ennek következtében a keresztbeszélgetések kutatása is ugyanezen szempontok szerint különbözik. Vannak olyan vizsgálatok, amelyek a keresztbeszélgetések szerepét egy-egy sejtsors, sejttípus, illetve egy vagy több útvonal szemszögéből vizsgálják. A **7. táblázatban** néhány példát mutatok ezekre a megközelítésekre.

7. táblázat. Példák a keresztbeszélgetéseket vizsgáló megközelítésekre, a vizsgált rendszer komplexitásának sorrendjében. A vizsgálatok kiterjedése egy darab útvonal keresztbeszélgetéseitől egészen a keresztbeszélgetések sejtek közötti összekapcsoló szerepéig terjednek.

A vizsgálat tárgya	Hivatkozás
Egy útvonal keresztbeszélgetéseinek modellezése	(McClean és mtsai, 2007)
Egy útvonal keresztbeszélgetései egészséges sejtekben	(Guo és Wang, 2009) (Hurlbut és mtsai, 2007) (Lutz és Knaus, 2002)
Egy útvonal keresztbeszélgetései őssejtekben	(Katoh és Katoh, 2007)
Két útvonal keresztbeszélgetései	(Borisov és mtsai, 2009; Wang és mtsai, 2009)
Több útvonal keresztbeszélgetéseinek modellezése	(Dumont és mtsai, 2001)
Több útvonal keresztbeszélgetései egy adott sejttípusban	(Fraser és Germain, 2009)
Bizonyos útvonalak keresztbeszélgetései egy adott szövettípusban	(Natarajan és mtsai, 2006)
Több útvonal együttműködése őssejtekben	(Blank és mtsai, 2008)
Több útvonal együttműködése egy szerv fejlődésében	(Fisher és mtsai, 2007; Robinson, 2007; Sternberg, 2005)
Több útvonal keresztbeszélgetései egy szervben	(Yan és mtsai, 2004)
Több útvonal keresztbeszélgetései a rák fejlődésében	(Katoh, 2007)
Több útvonal együttműködése a rákos és az egészséges őssejtek differenciálódásában	(Dreesen és Brivanlou, 2007)
Több (9) útvonal keresztbeszélgetései egy általános fehérjehálózatban	(Lu és mtsai, 2007)
Nagyon sok (580) útvonal keresztbeszélgetései egy általános fehérjehálózatban	(Li és mtsai, 2008)
Két útvonal keresztbeszélgetéseinek szerepe a sejtek közötti kommunikációban egy adott szervben	(Boswell és mtsai, 2008)

Mivel az egyes útvonalak által továbbítható jelek száma és kombinációja véges és korlátozott, az útvonalak közötti keresztbeszélgetések jelentősen képesek növelni a lehetséges fenotípusok és továbbítható információk számát. A keresztbeszélgetések képesek megnövelni a jelátviteli útvonalak által közvetített jelek kombinációs képességét azáltal, hogy új bemenet/kimenet lehetőségeket teremtenek. A nagyobb kombinációs képesség mellett a keresztbeszélgetések jelenléte a robusztusság egyik fontos alappillére, a redundanciát is növelik a rendszerben, ezáltal a rendszer adaptációs képességét is javítja (Taniguchi és mtsai, 2006). Ugyanakkor egy-egy új kapcsolat két útvonal között komoly szabályozási kérdéseket is felvetett. A keresztbeszélgetések szabályozása több szinten és mechanizmussal valósulhat meg. Az eddig feltárt, leggyakoribb megoldások során a keresztbeszélgetések szabályozása állványféhérjéken keresztül, visszacsatolási (*feed-back*) hurkok segítségével, kinetikai szigeteléssel vagy tér- és időbeli expressziós mintázatok segítségével valósulhat meg (Behar és mtsai, 2007; Bhattacharyya és mtsai, 2006; Freeman, 2000; Kholodenko, 2006).

A keresztbeszélgetések vizsgálatához pontos útvonal és útvonalhatár definícióra van szükség. Gerstein és munkatársai egy 2007-es tanulmányukban összefoglalták az ezekkel kapcsolatos problémákat. Rámutattak arra, hogy a különböző rendszereken, különböző céllal készített útvonalak együttes vizsgálata nem megfelelő a keresztbeszélgetések vizsgálatához (Lu és mtsai, 2007). Bauer-Mehren és munkatársai pedig arról készítettek egy összefoglalót, hogy ehhez a megváltozott szemlélethez és a keresztbeszélgetések rendszerszintű vizsgálatához új típusú útvonal-adatbázisokra van szükség (Bauer-Mehren és mtsai, 2009).

II. Célkitűzések

Az alapvető célom egy új és egységes jelátviteli adatbázis létrehozása volt a korábbiakban bemutatott paradigmaváltás miatt. Célom volt továbbá, hogy minél teljesebb körűen felhasználjam a már ismertetett módszereket és eredményeket, valamint ezek olyan módosításait, amelyek megfelelnek az új, hálózatos és integratív megközelítéseknek. Egy olyan jelátviteli útvonal adatbázis elkészítését tűztem ki célul:

- amely hiánypótló adatbázisként tud szolgálni a modern jelátvitel kutatáshoz és megfelel a legmodernebb hálózatos szemléletnek,
- amely objektíven rendszerezett útvonalakra épül és általánosítható eredményeket ad (a későbbi specifikus vizsgálatokhoz),
- amely a felhasznált diszciplínák (biokémia, genetika, bioinformatika, stb.) mindegyikének új eredményeket és lehetőségeket ad.

Egy ilyen adatbázis elkészítése, ábrázolása, ellenőrzése és összehasonlítása a már létező adatbázisokkal új módszerek kidolgozását is igényli. Ezért szintén célul tűztem ki:

- olyan élőlények és jelátviteli útvonalak kiválasztását, amelyek jelátviteli adatai megfelelő minőségben és mennyiségben állnak rendelkezésre,
- egy olyan gyűjtési módszer kidolgozását, amely összehasonlítható, reprodukálható és objektív módon biztosítja a jelátviteli adatokat adatbázisba kerülését,
- egy olyan keretrendszer kifejlesztését, amely alkalmas felhasználóbarát módon az adatok gyűjtésére, ellenőrzésére és a létrejött jelátviteli hálózat megjelenítésére,
- több olyan újszerű ábrázolási mód kidolgozását, amely a kapott eredmények megjelenítését a leginkább biztosítja,
- más jelátviteli útvonal adatbázisok részletes vizsgálatát, hogy lehetőség legyen egy minél pontosabb és átfogóbb összehasonlításra a létrehozott Signalink adatbázissal.

Az adatbázis elkészítése mellett célul tűztem ki olyan vizsgálatok elvégzését a létrehozott jelátviteli útvonal adatbázison, amelyekre korábban nem nyílt lehetőség. Ezért célom volt annak illusztrálása és bizonyítása, hogy a létrehozott adatbázis:

- rendszerszintű összehasonlító és elemző vizsgálatok elvégzésére alkalmas,
- segítségével a különböző fajok jelátviteli útvonalainak összehasonlítása során új fehérjék és funkciók azonosíthatóak,
- alkalmazásával az összes útvonalat egyszerre lehet vizsgálni, így elemezhetővé válnak a jelátviteli útvonalak átfedésében lévő (ún. többútvonalas, *multi-pathway*) fehérjék,
- az elméleti kutatások mellett az alkalmazott kutatás és fejlesztés (pl. rák- és gyógyszerkutatás) számára is hasznosítható eredményeket ad.

III. Módszerek

III.1. Egy jelátviteli adatbázis készítése: *Signalink*

III.1.1. Az adatbázis készítésének forrásai

III.1.1.1. A vizsgálandó élőlények kiválasztása

A jelátviteli és egyéb biológiai hálózatokat legtöbbször az élesztőn vizsgálták, ugyanis eukarióták közül erről az élőlényről tudjuk a legtöbbet, és egyszerűen végezhető rajta kísérletek. Azonban mivel az élesztő egysejtű, a fejlett jelátvitelnek csak kis része található meg benne. Többsejtű állatok hálózatainak vizsgálata még kezdeti stádiumban van, de elég sok adattal rendelkezünk már most is ahhoz, hogy rajtuk nézzük a jelátvitel működését. A *Drosophila melanogaster* régóta a genetikusok egyik modellállata, a korai embrionális fejlődés és a szervek kialakulása kapcsán gyakran vizsgálják jelátviteli hálózatát. A *Caenorhabditis elegans* az elmúlt három-négy évtizedben lett a leginkább használt modellállat, főleg a szervek kialakulását és egyszerű idegrendszerének működését lehet rajta vizsgálni, valamint a jelátviteli útvonalak közötti keresztbeszélgetések felderítéséhez is tökéletes. A gerincesek esetén is felmerült néhány lehetőség (karmosbéka, egér, patkány és ember). A rendelkezésre álló adatok típusai, mennyisége és az eredmények orvostudományi alkalmazhatósága alapján az emberi jelátviteli hálózat mellett döntöttem, mivel a sejtkultúrákban igen sok, más élőlényekben csak közvetetten kimutatott kapcsolatot publikáltak.

III.1.1.2. A vizsgálandó útvonalak kiválasztása

Alapvető probléma a jelátvitel kutatásában az útvonalak osztályozása, típusainak és számának a meghatározása. Nagyon szigorú, főleg biokémiai elvek szerint csak 7 útvonalról beszélhetünk (Pires-daSilva és Sommer, 2003). Funkcionálisabb, nemcsak az egyedfejlődést figyelembe vevő megközelítések esetén 17 (Gerhart, 1999), míg a rendszerszemlélettől igen távol álló csoportosítások 40 vagy akár 50 útvonalról is beszélnek (pl.: Ogata és mtsai, 1999). Az utóbbiak és általánosságban a legtöbb megközelítés alkalmatlan mindenfajta hálózatos ábrázolásra, rendszervizsgálatra, mivel az útvonalakat nem egy platformon, hanem külön-külön megjelenített, tárgyalt vagy

kísérletesen vizsgált, lineáris útvonalként kezelik. Az ilyen megközelítések alaphibája, hogy egy fehérje sok helyen is szerepelhet, és az egyes ábrákat összekötő kapcsolatok (keresztbeszélgetések) elvesznek, vagy nem kerülnek megjelenítésre. E fenti két problémát akartam megoldani, amikor egy gyökeresen új megközelítést kezdtünk el tervezni.

A SignaLink adatbázis építése során nyolc, kanonikusnak nevezett útvonalat használtunk, amelyek mind az egyedfejlődés, mind a felnőtt élet során funkcióval rendelkeznek. Ezek az EGF/MAPK, TGF- β , Wntless/WNT, Hedgehog (Hh), inzulin (INS/IGF), JAK/STAT, Notch és a nukleáris hormon receptor útvonalak. Ez a biokémiailag és evolúciósan is indokolt útvonalválasztás a tirozin-kináz alapú EGF-IGF (receptor-tirozinkináz, RTK) útvonal-család szétbontását leszámítva megegyezik Pires-daSilva és Sommer osztályozásával (Pires-daSilva és Sommer, 2003). A **8. táblázatban** röviden bemutatom ezen útvonalak főbb jellemzőit.

8. táblázat A Signalink adatbázisban lévő 8 útvonal főbb funkciói, jellemző reakciótípusai, valamint az útvonal indításául szolgáló ligandumok és az végső komponensként szolgáló transzkripciós faktorok (TF) fontosabb tulajdonságai.

Útvonal neve	Funkció	Ligandum típusa	Jellemző reakciótípus	Jellemző TF aktiválás	Kombinációs képesség foka és oka
EGF/MAPK	Növekedés, osztódás, túlélés, differenciáció, sejsors meghatározás	Testfolyadékban keringő növekedési faktorok	Receptor TyrKináz, szabad TyrKináz és Ser/ThrKináz	Kináz bemegy a sejtmagba, vagy a citoplazmában serkent egy TF-et	Sok kináz, ezek különböző foszforilációs helyeken tudnak hatni, különböző sorrendekben. Az adapter és állványfehérjék fontosak
Ins/IGF	Sejt-metabolizmus, szabályozás (öregedés)	Testfolyadékban keringő hormon	Receptor TyrKináz és szabad TyrKináz	Kináz gátol egy TF-et a sejtmagban	Különböző affinitás és expressziós tulajdonságú izoformákon keresztül dinamikus rendszer
JAK/STAT	Sejtosztódás, fejlődés, differenciáció, apoptózis, természetes és adaptív immun-funkciók	Testfolyadékban keringő citokinek	TyrKináz, specifikus dimerizációk	A dimer bejut a sejtmagba, és ott specifikus TF-et aktivál	Óriási kombinációs képesség a főbb ligandumok, JAK és STAT elemek kombinálódásával, valamint a homo- és heterodimerek miatt.
TGF-β	Sejsors és osztódás, differenciáció, apoptózis, sejt felismerés	Testfolyadékban keringő citokinek	Receptor Ser/ThrKináz, komplexképzés, fehérjelebontás	Komplex-képződést követően sejtmagba jutás és közvetlen TF hatás	Nagyfokú kombinációs képesség a ligandumok és a receptorok szintjén, majd kisebb fokú a receptorokat követő elemeknél
Hedgehog	Sejsors és osztódás, differenciáció és polaritás meghatározás, szervfejlődés	Lipidtesthez kötött rövid és hosszútávú ható ligandumok	Komplexképzés, (részleges) fehérjelebontás, Ser/ThrKináz	Jel hatására egy fehérje be tud jutni a sejtmagba, ahol transzkripciót aktivál Ha nincs jel, akkor ugyanez a fehérje részleges proteolízist követően bejut a sejtmagba, és több TF-et gátol	Kisebb fokú kombinációs képesség
WNT	Sejsors és polaritás meghatározás, szervfejlődés	Testfolyadékban keringő vagy sejthez kötött, vagy lipidtesthez kötött növekedési faktorok	Komplexképzés, fehérjelebontás, Ser/ThrKináz	Jel hatására egy fehérje be tud jutni a sejtmagba, ahol több TF-et aktivál	Nagyfokú kombinációs képesség a ligandumok és a receptorok szintjén
Notch	Sejsors és osztódás, differenciáció, sejthalál, szervfejlődés	Közvetlen sejt-sejt kapcsolaton keresztül	Integrált receptor-mediátor- transzkripciós faktor fehérje, proteolízis	Protelízist követően sejtmagba jutás és közvetlen TF hatás	Kisebb fokú kombinációs képesség
NHR	Fejlődés és növekedés, szaporodás, nem-meghatározás, metabolizmus szabályozás	Testfolyadékban keringő, lipiddoldékony szteroidok, hormonok	Sejttagon belüli integrált receptor-transzkripciós faktor fehérje	Közvetlen TF hatás	Kisebb fokú kombinációs képesség

III.1.1.3. A vizsgálandó kapcsolatok típusának meghatározása

Mivel a biokémia és a genetika különböző megközelítéssel és céllal vizsgálja a jelátvitelt (szerkezeti, illetve funkcionális), a kapcsolatokat is ez alapján definiálja. Emiatt a biokémikusok csak direkt kapcsolatokat írnak le, ahol pontosan lehet tudni, hogy két adott fehérje közvetlen kapcsolódik. Ezzel szemben a genetikai kapcsolatok gyakran indirekt kapcsolatok, ahol két adott gén (esetleg géntermék) funkcionális egymásra hatását definiálják kapcsolatnak. A genetikai kapcsolatok lehetnek közvetlen kapcsolatok is, de pusztán a genetikai tesztek alapján ezt nem mondhatjuk ki. Ha meg tudjuk állapítani, hogy melyik fehérje melyikre hat közvetlenül, például enzimatikus módosítással, vagy a két gén terméke között meg tudjuk mondani, hogy melyik gén funkciója melyik gén funkciójára hat, akkor irányt is adhatunk a kapcsolathoz. További, más vizsgálatokkal, információkkal az irányított közvetlen vagy közvetett kapcsolathoz egymásra hatási információt is rendelhetünk, vagyis azt, hogy a kapcsolat serkentő, vagy gátló hatást közvetít.

Fontos itt felhívni a figyelmet arra, hogy a fenti eltérés a biokémiai és genetikai megközelítésben, konkrétan a jelátvitel vizsgálatánál, olyan jelentősebb problémát eredményezhet, hogy egy genetikai indirekt kapcsolat lehet, hogy nem direkt illetve indirekt fehérje-fehérje kapcsolatokon, hanem transzkripción keresztül bekövetkező fehérje-gén-fehérje kapcsolatokon alapul. Azaz nem is a jelátviteli útvonal alkotó fehérjéi között bekövetkező kapcsolatról van szó, hanem a teljes jelátviteli rendszer elemei közötti hatásról. Erre a megkülönböztetésre mind a biokémiában, mind a korrekt hálózatos vizsgálatok során feltétlenül szükség van, míg a pusztán funkcionális kapcsolatokkal dolgozó genetika jogosan él az elhanyagolás lehetőségével. A probléma ott van, hogy gyakran a biokémiai és hálózatos vizsgálatoknál nem veszik figyelembe az előbb leírtakat, és így téves eredmények születnek.

Mind a biokémiai, mind a genetikai kutatások során előfordul, hogy a kapcsolatokhoz nem tudunk irányt rendelni (pl.: az élesztő kettős-hibrid eljárások vagy géninterakciót prediktáló adatbázisok), így pusztán a kapcsolat létéről van információnk, és irányítatlan közvetlen vagy közvetett kapcsolatokról beszélhetünk.

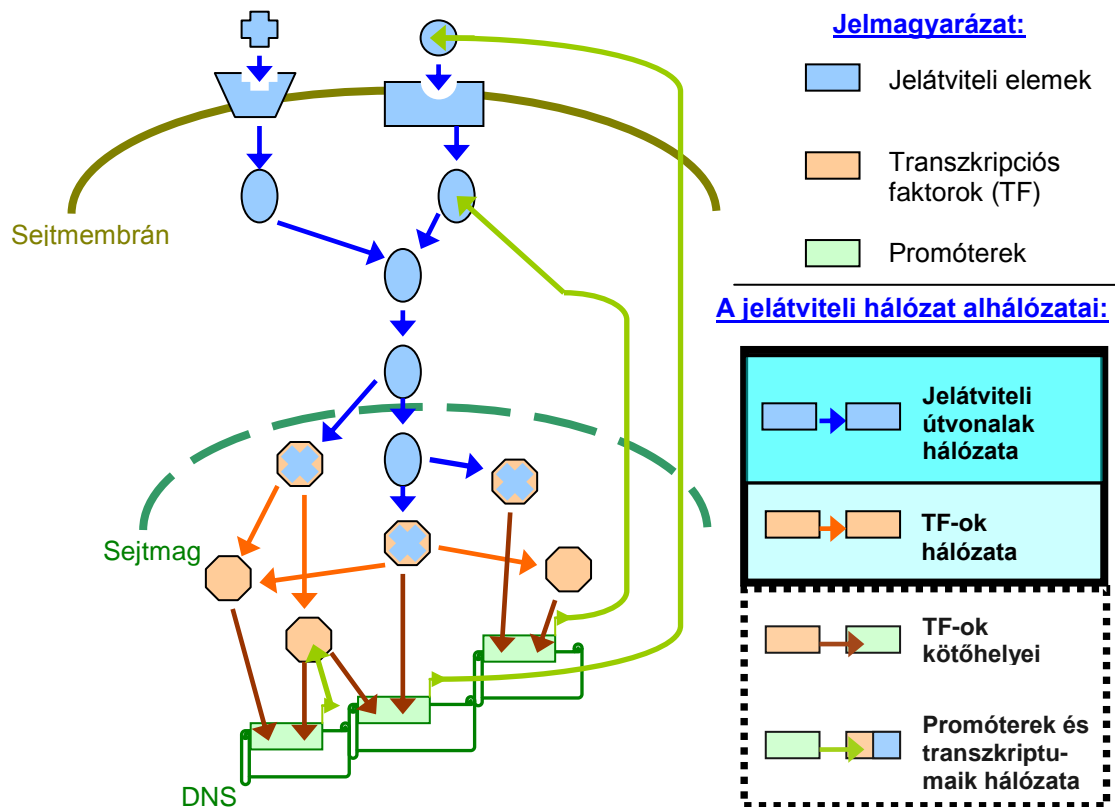
Tehát négy kapcsolat típusról beszélhetünk: irányított és irányítatlan valamint mindkettőn belül direkt és indirekt kapcsolatokról. Az irányított kapcsolat esetén mindkét altípushoz serkentő vagy gátló funkciók is rendelhetőek. A jelátvitelt vizsgáló

megközelítések általában nem szokták így definiálni vagy egyszerre használni ezeket a kapcsolatokat. Ennek következtében „zavaros” hálózatokat kapunk, amelyek elemzésével (főleg a direkt-indirekt kapcsolatok megkülönböztetésének hiánya miatt) téves következtetéseket vonhatunk le. Ugyanakkor megfelelő definíció mellett fontos minden fent említett kapcsolattípust felhasználni, különös tekintettel, a legtöbb információval rendelkező, direkt kapcsolatokra.

III.1.1.4. A vizsgálandó jelátviteli szakasz kiválasztása

Mint azt a Bevezető részben említettem, a jelátviteli útvonalak határait gyakran nem definiálják. A megközelítés típusa és a konkrét munka függvénye, hogy a jelátvitelben résztvevő elemek közül melyeket vesznek számításba, és ezek között milyen kapcsolatokat vizsgálnak. Ez a gyakorlat több szempontból is veszélyes és hibás. Egyrészt, a kapott eredmények nehezen általánosíthatóak, nem összeilleszthetőek vagy összehasonlíthatóak. Másrészt, ha a vizsgált alrendszer kiválasztása nem objektíven történt, akkor az abból levont bármilyen következtetés sem biztos, hogy megállja a helyét. Végül, ha rendszerszemlélettel akarunk jelátviteli hálózatot vizsgálni, akkor rendelkezünk kell első lépésként a rendszer, valamint annak résztvevőinek, kapcsolatainak és alrendszereinek korrekt és objektív alapokon nyugvó meghatározásával.

Az **7. ábra** mutatja be az általunk vizsgált négyfajta jelátviteli alhálózatot. Az első a jelátviteli útvonalak hálózata, a második az ezeket követő transzkripciós faktorok hálózata, a harmadik a transzkripciós faktorok és DNS-en lévő kötőhelyeik hálózata, végül a negyedik a génekből keletkező fehérjék hálózata. Jelen munkában az első hálózattal, a jelátviteli útvonalak hálózatával foglalkozom.



7. ábra. A jelátviteli hálózat általam csoportosított négy része. Az első alhálózat a jelátviteli útvonalak hálózata, amely a ligandtól a transzkripciós faktorig tart. A következő alhálózat a transzkripciós faktorok hálózata. Utóbbi két alhálózat természetesen átfed. Ezt követi a transzkripciós faktorok és a DNS-en lévő kötőhelyeik alhálózata, majd a megfelelő génekről a transzkriptumok alhálózata, amely visszacsatol az első két alhálózat elemeire. A disszertáció alapvetően az első alhálózattal, a jelátviteli útvonalak hálózatával foglalkozik, de tartalmaz transzkripciós faktorokat is.

III.1.2. Az adatbázis készítésének módszertana

Az adatbázis készítésének módszertana a (Korcsmaros és mtsai, 2010) publikáció 44 oldalas mellékleteként került részletes leírásra és közlésre.

III.1.2.1. Az alkalmazandó keretrendszer kiválasztása

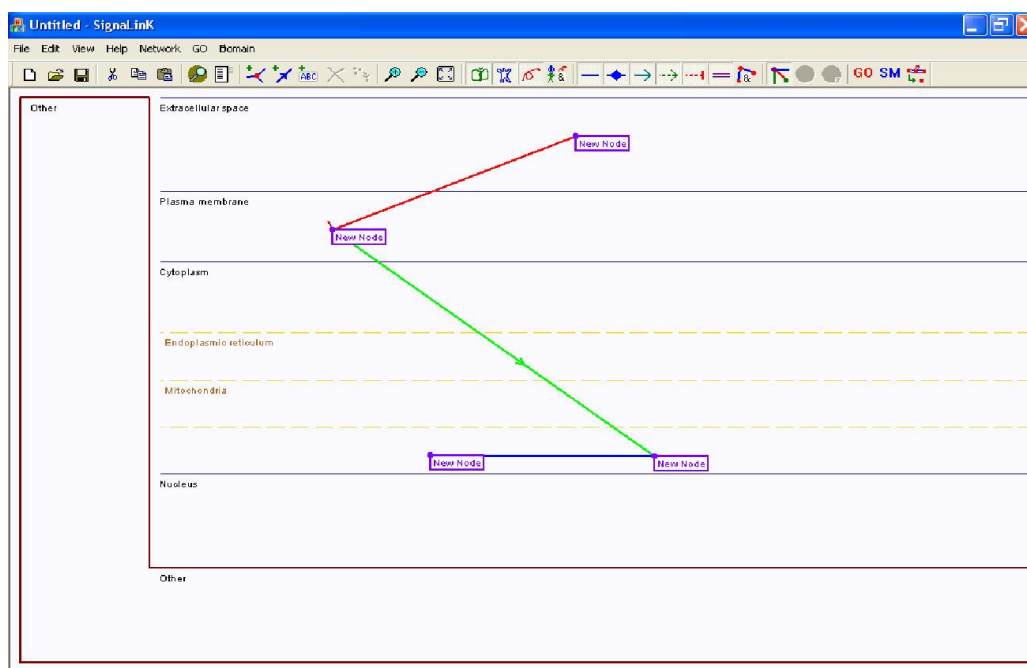
Az előző fejezetben láthattuk, hogy a célul kitűzött feladat több szempontból is új, mivel egyszerre több faj jelátviteli hálózatát kívánjuk elkészíteni egységes szabályokkal és összehasonlító céllal. Több kapcsolattípust szeretnénk figyelembe venni, és csak pontosan meghatározott tulajdonságokat illesztünk a rendszerbe.

Annak érdekében, hogy mindezen feladatokat elvégezhessük szükség volt egy munkafelületre, vagy más néven keretrendszerre. A keretrendszer készítése tehát egy

integrálási folyamatot igényel, a már létező eszközök és programok hasznos és jó tulajdonságainak az összegyűjtését (ld. a 1.2.4-es bioinformatikai megközelítéseket bemutató fejezetet). Következésképpen a célkitűzés meghatározása után az első lépésünk egy megfelelő keretrendszer (program) készítése volt, ezt követte a hálózat építése, majd később az elemzés.

III.1.2.2. A keretrendszer készítése és a hálózatépítés menete

Egy olyan keretrendszerre volt szükség, amely rendelkezik a célkitűzésben megfogalmazott funkciókkal, és amely szükség esetén bővíthető is. Ebből a célból Szalay Máté informatikus kollégám elkészítette a Signalink-programot, amely egy C++ nyelven írt hálózatábrázoló és -építő program (a program, szemben a Signalink adatbázissal, hálózattal kis "l"-lel irandó, a megkülönböztetés miatt). A program platformja a sejt egyszerűsített ábrázolása, amely különböző sejtttereket különít el. A program platformján pontok hozhatók létre, amelyeket összeköthetünk kapcsolatokkal (8. ábra).



8. ábra A Signalink program alapfelülete. A pontokat és kapcsolataikat a sejt különböző térségeiben (extracelluláris tér, sejtmembrán, citoplazma, endoplazmatikus retikulum, mitokondrium és a sejtmag) helyezhetjük el. Az eszköztár lehetővé teszi a megjelenítés során a fajokénti és a kapcsolattípusonkénti szűrést, valamint a színezést.

A három élőlény jelátviteli útvonal-hálózatát külön-külön, *C. elegans*, *D. melanogaster* és humán sorrendben építettem fel. Miután felépítettem a 8 útvonalat *C. elegans*-ban, elkészítettem a gyümöcslégy, és az ember útvonalait. Az adatbázis először 2006 decemberében készült el, majd az adatok 2008 májusában – azonos módszerek szerint – lettek frissítve. Az adatbázis készítése négy alapvető szakaszból állt:

1. Útvonal-specifikus összefoglalók és adatbázisok (honlapok) keresése.
2. Ezek alapján útvonal-fehérjék és ortológjaik beillesztése egy adott faj adott útvonalaiba.
3. További irodalmi keresések alapján a beillesztett fehérjék kapcsolatait leíró kísérletes cikkek keresése (valamint az így talált kapcsolódó jelátviteli fehérjék adatainak beillesztése.)
4. A beillesztett jelátviteli fehérjék további jelátviteli tulajdonságainak (pl.: jelátviteli pozíció, funkció) gyűjtése más adatbázisokból.

A következő alfejezetekben ebben a sorrendben mutatom be részletesen az egyes szakaszokat.

III.1.2.3. A hálózatépítés forrása és módszertana – az irodalom

A Signalink adatbázis készítésének első lépéseként minden útvonalhoz megkerestem az adott faj adott útvonalairól elérhető összefoglaló jellegű forrásokat (review-k és honlapok). Általában a kézi gyűjtéssel készülő útvonal adatbázisok útvonalanként 5-15 összefoglaló cikket szoktak felhasználni. A Signalink készítése során összesen 170 összefoglaló cikket használtam fel, ami útvonalanként átlagosan 20 cikket jelent. Az összes felhasznált irodalmi adat megtalálható a <http://signalink.org> weboldalon. Az adatbázis készítése során felhasznált internetes adatbázisokat és alkalmazásokat a **9. táblázat** tartalmazza.

9. táblázat A Signalink adatbázis készítése során felhasznált internetes adatbázisok és alkalmazások nevei és elérhetőségei.

Az adatbázis vagy alkalmazás neve	Internetes címe
BioCarta	http://www.biocarta.com
BioMART	http://www.biomart.org
Cell Signaling Pathway Diagrams	http://www.cellsignal.com/reference/pathway
DAVID	http://david.abcc.ncifcrf.gov
Drosophila Interaction Database	http://proteome.wayne.edu/PIMdb.html
DrugBank	http://redpoll.pharmacy.ualberta.ca/drugbank
Ensembl	http://www.ensembl.org
FlyBase	http://www.flybase.org
GAD	http://geneticassociationdb.nih.gov
Hedgehog Pathway Gene List	http://hedgehog.sfsu.edu/genes
Human Protein Reference Database	http://www.hprd.org
KEGG	http://www.kegg.jp
NetPath	http://www.netpath.org
Notch homepage (Greenwald lab)	http://cpmcnet.columbia.edu/dept/greenwaldlab
OMIM	http://www.ncbi.nlm.nih.gov/omim
Oncomine	http://www.oncomine.org
Orthodisease	http://orthodisease.sbc.su.se
Pathguide: The Pathway Resource List	http://www.pathguide.org
PharmGKB	http://www.pharmgkb.org
Protein Identifier Cross-Reference Service (PICR)	http://www.ebi.ac.uk/Tools/picr
Reactome	http://www.reactome.org
Synergizer	http://llama.med.harvard.edu/synergizer/translate
WormBase	http://www.wormbase.org
WormBook	http://www.wormbook.org
The Wnt Homepage (Nusse lab)	http://www-leland.stanford.edu/~rnusse/wntwindow.html

III.1.2.4. A hálózatépítés forrása és módszertana – a pontok

Hogy milyen gének (fehérjék) kerüljenek az adatbázisba és melyik útvonalhoz, azt az útvonalakról szóló összefoglaló cikkek, valamint az egyes útvonalakkal foglalkozó laboratóriumok weblapjai (ld. **9. táblázat**) határozták meg. Így az útvonalakhoz leginkább értő szakemberek információja alapján kerültek az egyes fehérjék egy adott útvonalhoz. További útvonal információval szolgált egy fehérjéről a Wormbase-ben és a Flybase-ben feltüntetett kapcsolati partnerek listája (Harris és mtsai, 2010; Wilson és mtsai, 2008), valamint két kapcsolatkereső program (az iHOP és a Chilobot) által megtalált jelátviteli szomszéd fehérjék (Chen és Sharp, 2004; Fernandez és mtsai, 2007). A **9. ábra** bemutatja a Signalink program pont űrlapját.

The screenshot shows a software window titled 'Signalink' with a close button in the top right corner. The window contains three main sections for different species, each with a checked checkbox on the left:

- Drosophila:** Includes input fields for 'name', 'Code', and 'SwissProtPrimaryAC'. Below these are 'Domains' and 'Go terms' fields with 'SMART' and 'GO' buttons respectively. There are also two 'Essentiality' checkboxes. Two dropdown menus for 'Topological property' and 'Functional property' are set to 'unknown', with corresponding '2' versions below.
- Human:** Identical layout to the Drosophila section.
- C. elegans:** Identical layout to the Drosophila section.

Below the species sections, there are additional fields:

- Family:** A dropdown menu currently showing '<other>'. Below it is another empty dropdown.
- Location:** A dropdown menu showing 'plasma membrane'. Below it is a checkbox for 'Second location' followed by another '<other>' dropdown.
- Description:** A large, empty text area with a vertical scrollbar on the right.

At the bottom of the window, there are 'OK' and 'Cancel' buttons, and two checkboxes for 'Drug target' and 'Disease gene'.

9. ábra A Signalink program pont űrlapja az adatbázis tulajdonságok megadásához. A pont űrlap megfelelő mezőiben lehetősége van a felhasználónak megadni az egyes fajok génjeinek azonosítóit, biokémiai és jelátviteli tulajdonságait, valamint jelölni az orvosi szerepüket (pl.: gyógyszercélpont). A program figyelmeztet, ha már létező génnevet szeretnénk újra bevinni.

A fehérjék beillesztése az adatbázisba az alábbi lépéseket tartalmazta:

- Fajspecifikus azonosítók (ORF vagy Ensembl azonosító) kikeresése és beillesztése fajspecifikus adatbázisokból:
 - *Wormbase* (verzió: 191) (Rogers és mtsai, 2008),
 - *Flybase* (verzió: 2008.6) (Wilson és mtsai, 2008),
 - *Ensembl* (verzió: 49) (Flicek és mtsai, 2008).

- Általános fehérjeazonosítók az *UniProt* adatbázisból (verzió: 87) kerültek beillesztésre (Boutet és mtsai, 2007). Ha egy fehérjéhez több UniProt azonosító tartozott, akkor az az azonosító került beillesztésre, amelyhez a leghosszabb aminosav-szekvencia tartozott.
- A fehérjék szinonimáit összefoglaló cikkekből (pl.: (Kyriakis és Avruch, 2001)) és az *iHOP* adatbázis „*synonym*” rovatából származtak (2008 májusi verzió) (Fernandez és mtsai, 2007).
- A fehérjeazonosítók átalakítására két internetes szolgáltatás a *Protein Identifier Cross-Reference Service* (Cote és mtsai, 2007) és a Synergizer (Berriz és Roth, 2008) szolgált.
- Minden beillesztett fehérje után a megfelelő ortológokat is beillesztettem a másik két fajba. Ehhez az ortológ-csoportokat tartalmazó InParanoid adatbázist (verzió: 6.1) használtam (Berglund és mtsai, 2008).
- A beillesztés után minden fehérje egy vagy több útvonalhoz lett rendelve, majd ezen belül egy-egy útvonal-régióhoz (gerinc vagy mellék):
 - Egy fehérjét akkor lett a gerinc-régióba sorolva, ha az adott fehérje esszenciális az útvonalon folyó jel továbbításában, és rendelkezik legalább egy, az útvonalra jellemző biokémiai tulajdonsággal (pl.: szerin/treonin aktivitás).
 - Egy fehérjét akkor lett a mellék- vagy más néven perifériás-régióba sorolva, ha nem tartozik a gerinc régióba, de az ott lévő fehérjék szabályozásában részt vesz.
- Az egyes fehérjék a jelátviteli topológiai pozíció szerint is besorolásra kerültek. Egy fehérje maximum két ilyen pozíciót kaphatott:
 - ligandum: olyan fehérje, amely elindítja az adott útvonal jelét
 - receptor: olyan fehérje, amely közvetlenül a ligandumtól kap jelet
 - mediátor: olyan fehérje, amely közvetíti a jelet a receptor(ok) és a transzkripciós faktor(ok) között
 - kofaktor: olyan fehérje, amely módosítja vagy szabályozza más fehérjék működését (sok ilyen fehérje az útvonal mellék-régiójának a tagja)
 - transzkripciós faktor: olyan fehérje, amely (1) a kapott jelet továbbítja egy másik transzkripciós faktornak (TF), vagy (2) komplexet képez más

TF fehérjékkal, vagy (3) a DNS megfelelő promóteréhez kötve közvetlenül szabályozza a transzkripciót

- egyéb fehérje: olyan fehérje, amely nem jelátviteli funkciót tölt be (pl.: sejtmozgás, transzport, membrán-folyamatok) (ezen fehérjék jelentős része kofaktor is egyben)
- ismeretlen: olyan fehérje, amelynek jelátviteli szerepe az adott útvonalban nem ismert

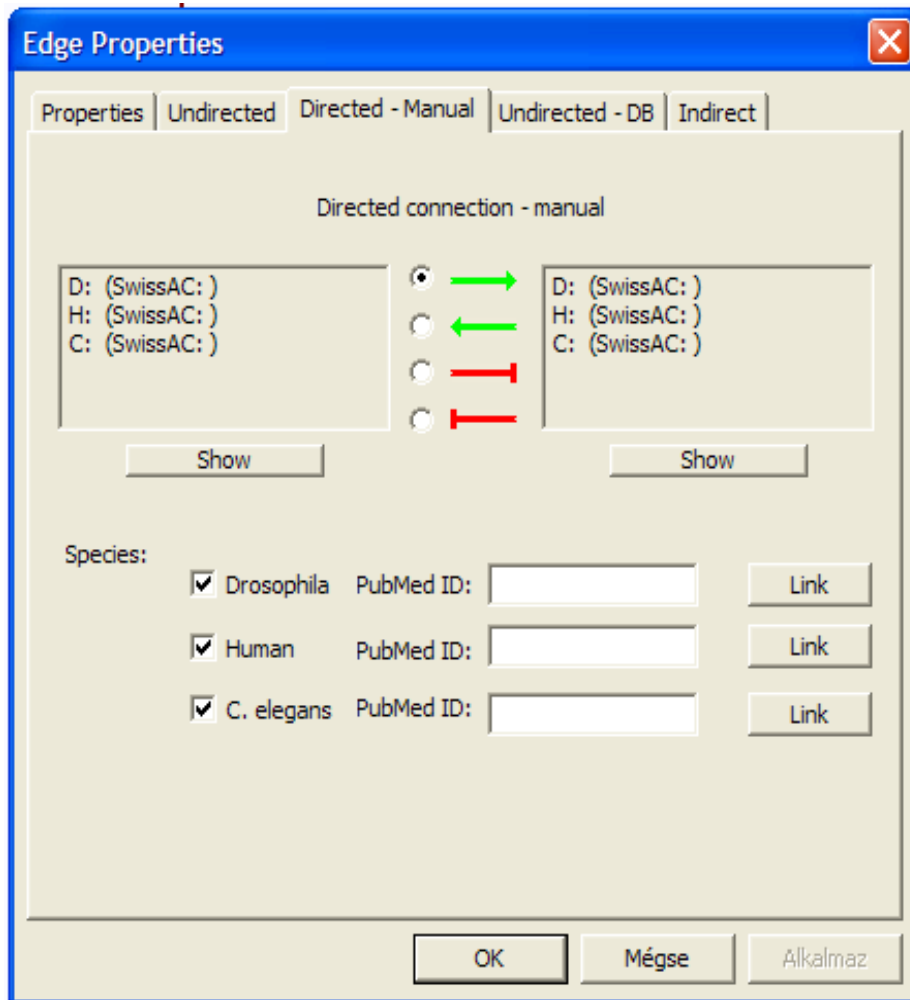
III.1.2.5. A hálózatépítés forrása és módszertana – a kapcsolatok

Mint ahogy azt a III.1.1.3. fejezetében ismertettem, a jelátviteli hálózatok esetében négyfajta kapcsolattípusról (irányítatlan, irányított, ezeken belül pedig közvetlen vagy közvetett) beszélhetünk. A négyfajta kapcsolattípus közül a Signalinkben csak irányított (közvetlen vagy közvetett) kapcsolatok szerepelnek. Pontosan meg van határozva hogy mely fehérje melyik másik fehérjére hat. Az **10. ábra** bemutatja a Signalink program kapcsolatainak űrlapját.

Az irányított kapcsolatokról a bizonyítékok alapvetően konkrét kísérletes cikkekből és csak ritkán összefoglalókból származnak. Ezeket a cikkeket összefoglaló cikkek, fajok webserverei, útvonal-honlapok és a már említett kapcsolatkereső programok segítségével találtam meg. Minden esetben a kapcsolat beillesztése során feltüntetésre került a bizonyítékul szolgáló cikk Pubmed azonosítóját.

A beillesztett kapcsolatok elsődleges irodalmi adatokon alapulnak, amelyek a következő forrásokból származtak:

- összefoglaló (*review*) publikációk
- faj-specifikus adatbázisok (Flybase, Wormbase) és UniProt (Boutet és mtsai, 2007; Harris és mtsai, 2010; Wilson és mtsai, 2008)
- iHOP, ChiliBot kapcsolatkereső-programok, valamint a Pubmeden végzett keresések (Chen és Sharp, 2004; Fernandez és mtsai, 2007)



10. ábra A Signalink program kapcsolatainak űrlapja az adatbázis tulajdonságok megadásához. A fenti fülekkel lehet kiválasztani a megfelelő kapcsolattípus űrlapját, és megadni az adott kapcsolat irodalmi hivatkozását.

A kapcsolatok beillesztése az adatbázisba az alábbi lépéseket, tulajdonságokat tartalmazta:

- Minden beillesztett kapcsolat irányított.
- A kapcsolat hivatkozása a legkorábbi kísérletes cikkekre hivatkozik.
- Minden összefoglaló esetében és a kísérletes cikkek 70 százalékában a teljes cikket vizsgáltuk.

- A kapcsolatok kísérletes bizonyítéka két részre oszlik:
 - direkt vagy közvetlen bizonyíték: biokémiai bizonyíték támasztja alá, hogy a két fehérje közvetlenül kapcsolódik
 - indirekt vagy közvetett bizonyíték: nincs bizonyíték közvetlen biokémiai meg határozott kapcsolatra, de a kísérletes bizonyítékok alapján ez valószínűsíthető (pl.: mRNS vagy fehérje mennyiség korrelációja, enzimaktivitás változása vagy a kapcsolódó fehérjék domén-vizsgálata)
- Egy kapcsolat hatása lehet serkentő/aktiváló vagy gátló. Az indirekt kapcsolatok esetében a ++ és a -- korrelációt tekintettük serkentőnek, míg a +- és -+ kapcsolattípust gátlónak.
- A Signalinkben azok a kapcsolatok, amelyek azonos fehérjepárok között vannak, de kétirányúak, külön kapcsolatként kerültek beillesztésre.
- A fehérje lista összegyűjtése után két további, ellenőrző lépés elvégzése:
 - fehérjelista ellenőrzése (minden fehérje kapcsolatainak újraellenőrzése; minden fehérje ahhoz az útvonalhoz lett rendelve, amelyhez irányított kapcsolata volt),
 - az ortológként beillesztett fehérje kapcsolatok ellenőrzése.

III.2. Más jelátviteli adatbázisokkal való összehasonlítás

III.2.1. Az összehasonlítás háttere

A kézi gyűjtéssel készült munkák esetében a gyűjtési módszert és a keletkezett adatbázist validálni kell. Mivel a Signalink készítése során a gyűjtési módszer kézi keresések alapján és nem algoritmusok által történt, ezért a Signalink minőségéről hasonló módon készült adatbázisokkal történő validálás a megfelelő.

A jelátviteli útvonalakat, adatokat tartalmazó adatbázisok száma folyamatosan növekszik (Bader és Mtsai, 2006). Ugyanakkor a legtöbb ilyen adatbázis kevesebb mint 200 molekulát tartalmaz, és sejt- vagy szövetspecifikusak. Több nagyobb adatbázis pedig csak bizonyos útvonalkomponenseket (pl.: kinázokat) tartalmaz. Azok között az adatbázisok között, amelyek ingyenesek, több mint 200 molekulát és minden útvonalkomponenst tartalmaznak, nincs olyan, amely hasonló célból és módszerrel jött létre, mint a Signalink. Mivel nincs megfelelő *standard*, amivel validálni lehetne a

Signalink adatbázist, három ismert és sokat használt adatbázis humán adataival hasonlítottuk össze: a KEGG, Reactome és NetPath adatbázisokkal (Joshi-Tope és mtsai, 2005; Kandasamy és mtsai, 2010; Ogata és mtsai, 1999). Referenciaként ezt a három adatbázist egymással is összehasonlítottuk. Az összehasonlítást Farkas Illés biofizikus hálózatkutatóval közösen végeztem. A teljes összehasonlítás a (Korcsmaros és mtsai, 2010) publikáció mellékleteként került közlésre.

III.2.2. A KEGG adatbázissal való összehasonlítás

A KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (Ogata és mtsai, 1999) nagyon sok útvonal adatát tartalmazza, köztük több jelátviteli és endokrin útvonalat is. A Signalinkben lévő 8 útvonalból a KEGG 7-et tartalmaz, ezek a KEGGben használt elnevezésekkel a következők: MAPK, WNT, Notch, Hedgehog, TGF-beta, Jak-STAT és Insulin.

A 7 KEGG útvonal fájlt PSI-MI xml-formátumban a következő forrásból töltöttük le: <ftp://ftp.genome.jp/pub/kegg/pathway/organisms/hsa>. Az egyes útvonal-fájlok 489 fehérjét tartalmaztak, amelyek ortológia és egyéb funkcionális szempontok alapján 285 fehérjecsoportra lettek felosztva. Az egyes fehérjék közötti kapcsolatokról nem találtunk információt, csak a csoportok közötti kapcsolatokról (210 irányított kapcsolat). Az összehasonlíthatóság érdekében fehérje-fehérje hálózatot kellett gyártanunk ebből az adatsorból. Erre a fehérjekomplexek hálózatos vizsgálata során alkalmazott mátrix- és spoke-modellek adnak lehetőséget (Orchard és mtsai, 2007). Az elemzés során „KEGG mátrix”-nak neveztük azt a modellt, ahol két irányítottan kapcsolódó csoport minden tagját irányítottan összekötöttük. „KEGG spoke”-modellnek pedig azt, ahol a forrás csoportnak csak az egyik tagját kötöttük össze a célcsoport egyik tagjával. (A fehérje Swiss-Prot azonosítóját ABC-sorrendbe rendeztük, majd az elsőt használtuk a „spoke”-modellhez; amennyiben csak TrEMBL azonosító állt rendelkezésre, akkor az ABC-ben utolsót használtuk.)

III.2.3. A Reactome adatbázissal való összehasonlítás

A Reactome adatbázisban 5 útvonal egyezik a SignaLinkben lévő 8 útvonallal (Joshi-Tope és mtsai, 2005), ezek neve: EGF, IGF/Ins, Notch, TGF és WNT. A <http://www.reactome.org/download/current> webcímről töltöttük le ezen útvonalak adatait (*uniprot_2_pathways.stid.txt* és *homo_sapiens.interactions.txt*) 2009 áprilisában. Az első adatfájl (*uniprot_2_pathways.stid.txt*) tartalmazta a fehérje listákat. Az összehasonlításhoz az alábbi útvonalnév párosításokat használtuk (SignaLink név és zárójelben a Reactome elnevezés): EGF/MAPK útvonal („Signaling by EGFR”), IGF/Ins útvonal („Signaling by Insulin receptor”), Notch útvonal („Signaling by Notch”), TGF útvonal („Signaling by TGF beta”) és a WNT útvonal („Signaling by Wnt”). A másik fájl (*homo_sapiens.interactions.txt*) tartalmazta a kapcsolati adatokat, amelyek olyan kapcsolatokat is tartalmaztak, amely két útvonal elemét kötötték össze (keresztbeszélgetés). Ez a fájl az 5 útvonalra különböző kapcsolattípusokat tartalmazott, mindegyik 1-1 fehérjét köt össze. A kapcsolattípusok közül kettő fehérjekomplextagsági adatokból lett származtatva. Az összehasonlíthatás során – mivel a SignaLink csak valódi, fehérje-fehérje reakciókat tartalmaz – nem megfelelő ezeket a kapcsolatokat felhasználni. Mivel azonban a legtöbb vizsgálatban a Reactome adatait egységesen szokták vizsgálni, az összehasonlítást elvégeztük a fehérjekomplexadatokkal együtt és azok nélkül is.

III.2.4. A NetPath adatbázissal való összehasonlítás

A NetPath adatbázis eredetileg a „Human Protein Reference Database” részeként jelent meg (Keshava Prasad és mtsai, 2009), nemrégiben azonban önálló cikként is publikálták (Kandasamy és mtsai, 2010). A korábbi két útvonal-adatbázishoz képest a NetPath más szerkezetű és jóval kisebb. A NetPath 20 útvonalhoz tartalmaz fehérjéket és ezek kapcsolatait. Érdekes módon – noha ezt nem írják sehhol – ebben az adatbázisban egy fehérje több útvonalhoz is tartozhat. A NetPath-ban 5 olyan útvonal van, amely összehasonlítható a SignaLinkkel, ezek elnevezése a NetPath-ban: EGFR1, Hedgehog, Notch, TGF beta Receptor és WNT (a többi 15 útvonal immunfolyamatokra vagy ráksejtekre jellemző). Az útvonaladatokat egy egységes PSI-MI formátumú XML fájlból értük el a következő webcímről: http://www.netpath.org/data/psimi_2

A NetPath-kapcsolatok nem egy-egy fehérje között vannak (nem binárisak), hanem kapcsolódó fehérje listák kerültek feltüntetésre. Ezeket a kapcsolódó fehérje listákat KEGG-nél már alkalmazott mátrix-moddal elemeztük (Orchard és mtsai, 2007) és a saját-kapcsolatokat (*self-loop*) nem vettük figyelembe. Mivel a NetPath mátrix modellje esetében minden fehérjéhez csak 2-3 kapcsolat tartozott, úgy ítéltük meg, hogy ez megfelelően pontos módszer a NetPath útvonal-hálózatának létrehozására.

III.3. Az útvonal-hálózatok ábrázolása

A Signalink adatbázisban lévő útvonalakat és keresztbeszélgetés hálózatot a hagyományos hálózatábrázolással és új (általunk kidolgozott) megközelítéssel is ábrázoltuk. Erre az adatok értelmezése és illusztrálása miatt volt szükség. Összesen 7 különböző ábrázolási módot alkalmaztunk a vizsgálatok során, amelyek a hálózat jelátviteli, ortológiai és keresztbeszélgetés tulajdonságain alapultak. Ezt a 7 megközelítést a hálózat elrendezése alapján 2 fő csoportba lehetett sorolni: hagyományos és útvonalalapú (keresztbeszélgetés-megjelenítés) szerintire. Előbbibe három, utóbbiba négy ábrázolási mód tartozik.

A hagyományos hálózatábrázolási elrendezés során kétfajta jelátviteli tulajdonság (útvonal és útvonal pozíció) ábrázolásához a Cytoscape nevű vizualizáló programot és az ehhez tartozó Golorize programot használtam Fazekas Dávid biológus hallgató segítségével (Cline és mtsai, 2007; Garcia és mtsai, 2007). Az ortológiai információk ábrázolásához egy interaktív, ortológia-információkat is tartalmazó és webes kezelőfelülettel rendelkező programot használtam, amelyet terveim alapján FLASH-nyelven Palotai Robin informatikus kollégám készített.

A keresztbeszélgetés-hálózathoz négyfajta megjelenítést készítettünk. Az egyik az útvonalak és jelátviteli pozíciók közötti keresztbeszélgetés-hálózatot ábrázolja egy Python és OpenGL program segítségével, amelyet Szuromi Gábor informatikus készített. A másik ugyanennél az elrendezésnél az ortológia információkat egy folyamatos színskálával jelzi. A harmadik ábrázolási megközelítés során egy Python program segítségével kapcsolati mátrixot rajzoltunk az útvonalak között, amelyekhez expressziós adatokat illesztettünk a III.4.2. fejezetben leírtak szerint. Ezt a programot és az ábrákat Fazekas Dávid készítette. Végül azokat az ún. többútvonalas (*multi-pathway*) fehérjéket ábrázoltam, amelyek több útvonalnak is tagjai. Ebben az ábrázolási

elrendezésben – amelyet Dr. Farkas Illés hálózatkutató javasolt – azt jelenítettük meg, hogy hány azonos fehérje van az egyes útvonalak között. Az egyezést kapcsolattal jelöltem két útvonal között. Utóbbihoz a már említett Cytoscape programot használtam (Cline és mtsai, 2007).

A hálózat pontjainak és a közöttük lévő kapcsolatok méretének, vastagságnak minden esetben informatív funkciója van, a jelenlévő fehérjék vagy fehérje-fehérje kapcsolatok számát tudjuk így jelezni. Míg a színezésnek egyes ábrákon csak illusztráló szerepe van. Ez a munka a (Korcsmaros és mtsai, 2010) és a (Korcsmaros és mtsai, 2011) publikációkban került közlésre.

III.4. Az adatbázison elvégzett elemzések

III.4.1. Génfunkció predikciók

A fajok közötti génrokonságok (ortológia) alapján lehetőség van génfunkciók jóslására a vizsgált fajok között. Például, ha az egyik fajban ismert egy fehérje funkciója, és ennek a fehérjének a másik fajban van ortológja, amelynek nem ismerjük a funkcióját, feltételezhető, hogy (a szekvenciahasonlóság alapján) az utóbbi hasonló molekuláris funkciót tölt be. Az ilyen ún. annotáció áthelyezéseket (transzfereket) fehérje-fehérje és szabályozási kapcsolatok, illetve fenotípusok esetében már leírtak (McGary és mtsai, 2010; Yu és mtsai, 2004). A SignaLink adatbázis adatszerkezete és gyűjtési módszere lehetővé teszi, hogy hasonló predikciókat tegyünk az egyes fehérjék jelátviteli funkcióira (Korcsmaros és mtsai, 2010).

III.4.1.1. Lehetséges jelátviteli fehérjék azonosítása ortológia alapján

Az ortológia kapcsolatok felderítését általában szekvencia alapú megközelítésekkel végzik, gyakran hálózatos kapcsolati információkkal kiegészítve (Bandyopadhyay és mtsai, 2006; Koonin, 2005). Például a PathBLAST nevű eszköz sorbarendezett fehérjéket vagy utakat illeszt fajok között a fehérjék szekvenciahasonlósága alapján (Kelley és mtsai, 2004). A COG adatbázisban (*Clusters of Orthologous Groups*) az ortológiai csoportokat kétirányú (reciprok) BLAST vizsgálattal határozzák meg legalább három faj alapján (Altschul és mtsai, 1997; Tatusov és mtsai, 2000). Még nagyobb érzékenységet érnek el azok a módszerek, amelyek nem csak a legnagyobb BLAST érték, hanem egy BLAST értéktartomány

alapján határozzák meg az ortológ csoportokat. Ilyen módszer a Signalink készítése során is alkalmazott InParanoid is (Berglund és mtsai, 2008). Az InParanoid két részre osztja az ortológokat: (1) olyan homológ szekvenciájú fehérjék, amelyek a fajképződést megelőző duplikáció eredményeként jöttek létre, és nem valószínű, hogy egyezik funkciójuk (ezeket *out paralog*-oknak nevezik); és (2) olyan homológ szekvenciájú fehérjék, amelyek a fajképződést követően duplikáció eredményeként jöttek létre, és így valószínű, hogy egyezik a funkciójuk (ezeket *in paralog*-oknak nevezi). Az InParanoid két faj összehasonlítása során az illesztett fehérjék BLAST értékeinek teljes tartományát használja (nemcsak a maximum értékeket) (Berglund és mtsai, 2008). Minden egyes ortológiai-csoport tagjához egy értéket rendel (InParanoid érték), hogy mennyire lehet funkcionális ortológ az adott fehérje (Berglund és mtsai, 2008).

Ahogy az a III.1.2.3.-as fejezetben szerepelt, a Signalink adatbázis készítése során is már alkalmaztuk az InParanoid adatbázist arra, hogy egy-egy beillesztett, ismert jelátviteli fehérje szintén ismert jelátviteli ortológját is megtaláljam és beillesszem. Hasonló módszert alkalmazva olyan fehérjéket tudtam azonosítani, amelyeknek nem ismert a jelátviteli szerepe (nincs kapcsolata más jelátviteli fehérjével), de van olyan ortológja, amelynek ismert a jelátviteli szerepe. Hasonlóan a Signalink adatbázis készítéséhez, ez alkalommal is csak azokat az ortológokat vettem figyelembe, amelyek InParanoid értéke 0,3 felett van. Ezt követően két interneten elérhető fehérje-fehérje kapcsolat-keresővel, az iHOP és a ChiliBot programokkal ellenőriztem, hogy a prediktált jelátviteli fehérjéknek valóban nincs ismert jelátviteli kapcsolata (Chen és Sharp, 2004; Fernandez és mtsai, 2007). Az ellenőrzésben Rovó Petra, Spiró Zoltán és Zsákai Lilian TDK hallgatók vettek részt.

III.4.1.2. Ortológia alapú útvonal annotáció jóslás

Mindhárom vizsgált faj (*C. elegans*, *D. melanogaster* és *H. sapiens*) esetében felsoroltuk azokat a fehérjéket, amelyeknek nincs jelátviteli kapcsolatuk, de legalább egy másik fajban ismert olyan ortológjuk, amely a vizsgált 8 útvonal valamelyikében funkcióval rendelkezik. Hasonlóan a funkcionális ortológia koncepcióhoz (Bandyopadhyay és mtsai, 2006), minden ilyen fehérjénél feltételeztem, hogy útvonal annotációt (jelátviteli szerepet) jósolhatjuk (áthelyezhetjük) a fajok között. Más szóval

ezen fehérjék esetében azt az útvonal tagságot jósoltam, amely az ortológjáról ismert valamelyik másik fajban.

III.4.1.3. A prediktált fehérjék újdonságvizsgálata

Annak érdekében, hogy a jósolt fehérjék új jelátviteli szerepét bizonyítsam, két vizsgálatot végeztem el. Megvizsgáltam az irodalmat fél-automata módszerekkel már létező annotációk után kutatva, valamint összehasonlítottam predikcióinkat korábban már publikált predikciókkal.

A jósolt annotációk osztályozását öt interneten elérhető keresőprogram segítségével végeztem el (Pubmed, UniProt, GO, iHOP és Chilibot keresők; (Ashburner és mtsai, 2000; Boutet és mtsai, 2007; Chen és Sharp, 2004; Fernandez és mtsai, 2007). A három fajban a prediktált szignológokat 5 csoportba osztottam az alapján, hogy mennyire ismertek tulajdonságaik az irodalomban: (1) az irodalom nem említ ortológia információt vagy biokémiai funkciót; (2) ismert az ortológiai információ, de a biokémiai funkció nem; (3) nincs ismert ortológiai információ, de a biokémiai funkció ismert; (4) ismert az ortológiai információ és a biokémiai funkció is; (5) ismert az ortológiai információ, a biokémiai funkció is és egy útvonal esetében már leírták a szerepét az adott fehérjének. Ez az osztályozás az ismeretlenebb (1) felől halad az ismertebb (5) felé. Ugyanakkor még az 5. kategóriába sorolt fehérjék esetében is lehetett új útvonal annotációt jósolni.

A következő vizsgálat során interológokat (ortológia alapján jósolt kapcsolatokat) alkalmaztam annak érdekében, hogy megvizsgáljam az ortológ predikciók újdonságát. A Signalink adatbázis alapesetben nem tartalmaz ilyen jósolt kapcsolatokat, így ezeket a meglévő információk alapján kellett meghatároznom úgy, hogy interológ kapcsolattal összekötöttük azokat a fehérjéket, amelyek ortológjai egy másik fajban ismert módon kapcsolatban vannak egymással. Csak azokat az interológokat vizsgáltuk, amelyek legalább egyik tagja jósolt fehérje. Ezeket az interológokat hasonlítottam össze, három fajspecifikus interológ adatbázissal: a WI8 nevű féreg, a DroID nevű gyümölcslelgy és a HomoMINT nevű emberi interológokat jósoló adatbázissal (Persico és mtsai, 2005; Simonis és mtsai, 2009; Yu és mtsai, 2008). Ez az összehasonlítás kimutatta a jósolt fehérjék jelenlétét a már korábban publikált ortológia alapján jósolt adatbázisokban.

III.4.2. A keresztbeszélgetések expressziós vizsgálata

A keresztbeszélgetés expressziós vizsgálatmódszertana a (Korcsmaros és mtsai, 2010) publikáció mellékleteként került közlésre.

III.4.2.1. A szövet-specifikus keresztbeszélgetés expressziós vizsgálatának módszertana

A keresztbeszélgetések szöveti megjelenésének vizsgálata céljából az emberi fehérjék és kapcsolatok esetében szöveti expressziós adatokat használtam fel. Két nagy adatbázis létezik, amely a legtöbb emberi fehérjét kódoló mRNS expresszióját tartalmazza: egyik a GNF Atlas (Su és mtsai, 2004), a másik az eGenetics (Kelso és mtsai, 2003). Mindkét adatbázis elérhető az EMSEMBL-BioMART oldalról (Flicek és mtsai, 2008). A GNF Atlas néhány nagy kísérletben kimért expressziókat tartalmaz 46 humán szövet, szerv és sejtvonal esetében, míg az eGenetics adatsor sok kisméretű vizsgálat (7016 cDNS könyvtár és 104 SAGE könyvtár) adatait tartalmazza 50 humán szövet és szerv esetében. A legmagasabb expressziót tartalmazó 19 szövettípusnál a GNF Atlas átlagosan 251,6, míg az eGenetics 242,4 SignaLinkben is előforduló fehérjét tartalmaz. Mivel az eGenetics adatsor könnyebben felhasználható volt, és nem volt jelentős adatmennyiségbeli különbség a GNF Atlas-hoz képest, ezért a vizsgálataimat csak az eGenetics adatosorral végeztem.

Az eGenetics adatsor minden fehérjéhez felsorolja azokat a szövet- és szervtípusokat, amelyekben az adott mRNS expresszálódik (van/nincs bináris adatsor). A SignaLinkben lévő 525 humán fehérje mind szerepel az eGenetics adatsorban, de közülük csak 323 fehérje expressziójáról van adat. Az **10. táblázatban** felsorolom azokat a szövet-, és szervtípusokat, amelyekben a legtöbb SignaLinkben is szereplő fehérje szerepel.

A keresztbeszélgetés expresszió-vizsgálat során megnéztem, hogy mennyire szelektív az egyes keresztbeszélgetésekben résztvevő fehérjék expressziója a különböző szövetekben. Annak érdekében, hogy statisztikailag megfelelő állítást tudjak megfogalmazni, több szövettípust együttesen vizsgáltam. Ugyanakkor, ha minden szövettípusból használtam volna adatokat, azok a különbségek, amelyeket ki akartam mutatni, összemosódtak volna. Ezért kiválasztottam öt különböző szövetet az alábbi 3

szempont alapján: (1) minél több SignaLink fehérje mRNS-expresszióját tartalmazza; (2) az adott szövet szövettanilag minél inkább homogénebb legyen; (3) lehetőleg különböző eredetű (ekto-, mezo-, entoderma) és funkciójú legyenek. Az **10. táblázatban** szereplő listából választottam ki az alábbi 5 szövet- illetve szervtípust: kardiovaszkuláris, végbél, hámszövet, máj és izom. A **11. táblázat** felsorolja, hogy átlagosan ebben az 5 szövetben az egyes útvonalak fehérjéinek hány százaléka expresszálódik. Ezen értékek ismeretében megállapíthatjuk, hogy az 5 kiválasztott szövet és az expressziós adatsor megfelelő ahhoz, hogy a keresztbeszélgetések szöveti expresszióját megvizsgáljuk.

10. táblázat Az expresszálódó SignaLink fehérjéket kódoló mRNS-ek száma az egyes szövet- és szervtípusok szerint az eGenetics adatbázisból. A táblázat legelső sorában az összesített érték szerepel („*anatómiai rendszer*”). Szürke háttérrel az az 5 szövettípus van kiemelve, amelyet a későbbi keresztbeszélgetés vizsgálatok során használtunk.

eGenetics adatbázis szövet/szerv név	Expresszálódó SignaLink fehérje-számok	Az 525 emberi SignaLink fehérje százalékos megoszlása az egyes szövet-, és szervtípusokban
anatómiai rendszer (<i>minden szövet és szerv</i>)	323	61,8%
légzőrendszer	303	57,9%
férfi ivarrendszer	294	56,2%
női ivarrendszer	293	56,0%
központi idegrendszer	288	55,1%
kötőszövet	264	50,5%
végbél	261	49,9%
endokrin-rendszer	249	47,6%
kiválasztórendszer	247	47,2%
hámszövet	246	47,0%
látási szervek	238	45,5%
vázrendszer	237	45,3%
máj (és epe)	236	45,1%
limforetikuláris	235	44,9%
kardiovaszkuláris	230	44,0%
méh	220	42,1%
hematológiai rendszer	212	40,5%
perifériás idegrendszer	190	36,3%
izom	188	36,0%
mell	174	33,2%

11. táblázat Az egyes útvonal fehérjék átlagos expressziója a kiválasztott 5 szövetben.

Útvonal név	Az útvonalfehérjék átlagos expressziója a kiválasztott 5 szövettípusban
Notch	48,6%
IGF	45,9%
TGF-β	44,2%
EGF/MAPK	42,3%
NHR	38,8%
Hh	34,2%
WNT	33,3%
JAK/STAT	25,8%

III.4.2.2. Betegség-specifikus keresztbeszélgetések expressziós módszertana

A keresztbeszélgetések betegségben betöltött szerepét egy gyakori és homogén szöveti jellemzőkkel rendelkező ráktípus, a hepatocelluláris karcinóma példáján vizsgáltam meg. Ehhez két hepatocelluláris karcinóma expressziós vizsgálat (Chen és mtsai, 2002; Wurmbach és mtsai, 2007) adatait töltöttem le az Oncomine nevű rákos expressziós adatokat tartalmazó internetes adatbázisból (Rhodes és mtsai, 2007). A keresztbeszélgetésben résztvevő fehérjék expresszióváltozását néztem az egészséges máj expressziós adataihoz képest. Annak a fehérjének a mennyiségét tekintettem változónak, amelynek legalább az egyik májrakos expressziója az egészséges expresszióval összehasonlítva a t-próba teszt p-értéke 0,05 alatt van.

III.4.3. Gyógyszercélpont-jelöltek meghatározása

A gyógyszercélpont-jelöltek azonosításához abból indultam ki, hogy a jelenleg forgalomban lévő gyógyszercélpontokra milyen tulajdonságok a leginkább jellemzőek (Gao és mtsai, 2008; Spiro és mtsai, 2008; Yildirim és mtsai, 2007). A fehérjékről ezeket a tulajdonságokat a DAVID nevű internetes szolgáltatás segítségével gyűjtöttem össze (Dennis, Jr. és mtsai, 2003). A betegségekkel kapcsolatos annotációkat az OMIM, GAD és Orthodisease adatbázisokból (Amberger és mtsai, 2009; Becker és mtsai, 2004; O'Brien és mtsai, 2004), a doménszerkezeti információkat az InterPRO adatbázisból (Hunter és mtsai, 2009), valamint a molekuláris funkció és sejttér elhelyezkedési adatokat a GO adatbázisból (Harris és mtsai, 2004) töltöttem le. A jelenleg ismert és használt gyógyszercélpontok listáját a DrugBank adatbázisból töltöttem le (Wishart, 2008), és további elemzésükhöz pedig a PharmaGKB adatbázis internetes felületét használtam (<http://www.pharmgkb.org>; (Medina és mtsai, 2010)).

IV. Eredmények

A módszertani rész első fejezetében bemutatott egységes gyűjtési módszerrel létrehoztam egy új jelátviteli útvonal adatbázist, amelyet SignalLinknek nevezünk el. Az adatbázis a <http://signalink.org> weboldalon érhető el (ingyenes). Az alábbi 4 fejezetben ezzel az adatbázissal foglalkozom. Először bemutatom az adatbázis minőségi és mennyiségi tulajdonságait, majd részletesen összehasonlítom a SignalLinket három ismert útvonal-adatbázissal. Ezt követően a SignalLink adatbázis ábrázolási lehetőségeit és megoldásait, végül az adatbázison elvégzett elemzések eredményeit mutatom be.

IV.1. Jelátviteli útvonalak egységes gyűjteménye

A SignalLink adatbázis és ennek részletes elemzése, összehasonlítása a (Korcsmaros és mtsai, 2010) publikációban és annak 44 oldalas mellékleteként került közlésre.

IV.1.1. A SignalLink adatbázis minőségi tulajdonságai

A fonálféreg *Caenorhabditis elegans*, a gyümölcslégy *Drosophila melanogaster* és az ember főbb jelátviteli útvonalainak egységes gyűjteményét készítettem el. A sok elérhető osztályozási rendszer közül (Bader és mtsai, 2006) Pires-daSilva és Sommer összefoglalója alapján választottuk ki azt a nyolc jelátviteli útvonalat, amelyet az adatbázisba rendeztünk. Ezek az útvonalak – EGF/MAPK (*Epidermal Growth Factor/Mitogen-Activated Protein Kinase*), TGF- β (*Transforming Growth Factor-beta*), *Wingless/WNT*, *Hedgehog* (Hh), inzulin/IGF-1 (*Insulin-like Growth Factor-1*), JAK/STAT (*Janus Activating Kinase / Signal Transducer and Activator of Transcription*), Notch és a nukleáris hormon receptor útvonalak – központi jelentőségűek az egyedfejlődés és a felnőtt élet során egyaránt (Pires-daSilva és Sommer, 2003). Ezeknek az útvonalaknak a megkülönböztetése biokémiailag és evolúciósan is indokolt, mivel az egyes útvonalakra különböző biokémiai mechanizmusok jellemzőek, amelyek egy útvonalon belül hasonló evolúciós eredettel rendelkeznek. Az adatbázis készítése során egységes szabályok szerint kézzel gyűjtöttem a genetikai és fizikai kapcsolatokat bizonyító adatokat. A SignalLink adatbázisban a fehérjék és a kapcsolatok szöveti jellemzők nélkül szerepelnek, így a létrejött hálózat lehetséges kapcsolati gyűjteménynek is nevezhető. A jelátviteli

komponensekre vonatkozó szövet- és betegség-specifikus jellemzőket expressziós mintázatok hozzáadásával lehet a jelátviteli hálózathoz rendelni (ld. IV.4.3.-as fejezet).

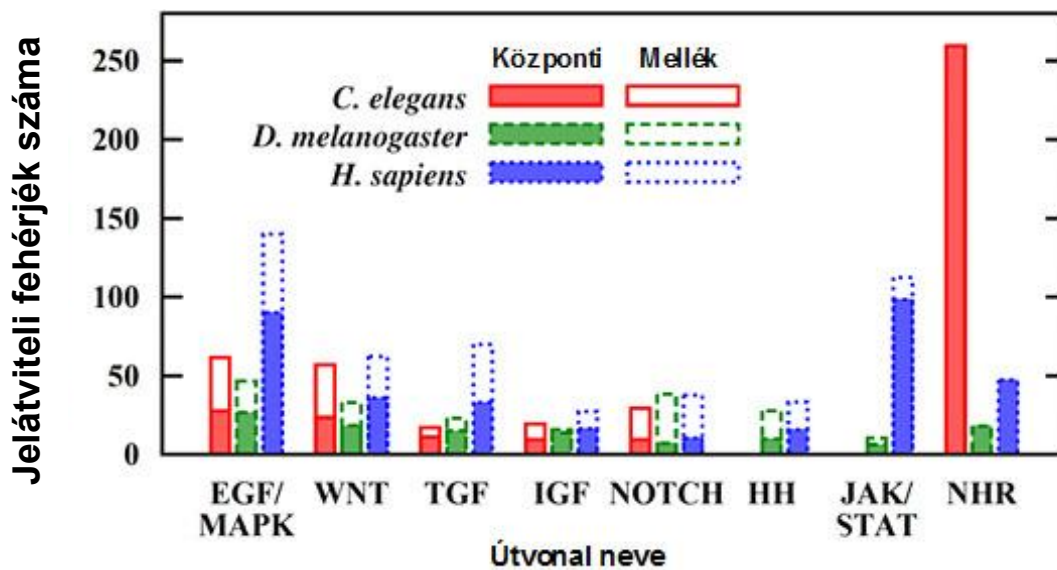
A SignaLink jelátviteli útvonal gyűjteménynek 5 olyan főbb tulajdonsága van, amelyek együttes jelenléte megkülönbözteti a többi útvonal-adatbázistól:

- A teljes adatbázis egységes gyűjtési módszerek alapján készült (minden faj, minden útvonala esetén).
- Az útvonalak biokémiai mechanizmusok és evolúciós rokonság alapján definiáltak, és tartalmazzák a sejtekben előforduló legtöbb jelátviteli mechanizmusokat (pl.: kinázok, peptidázok).
- Egy adott útvonal komponens több útvonalhoz is tartozhat (ezeket többútvonalas fehérjéknek neveztük).
- Egy jelátviteli útvonal komponens három tulajdonsággal rendelkezik:
 - mely útvonal(ak) tagja
 - az útvonal(ak)on belül milyen régió tagja (pl.: központi régió, amelybe a jel terjedése szempontjából esszenciális fehérjék tartoznak, vagy mellék régió, amelyek a központi régiót szabályozzák)
 - az útvonalakon belül milyen jelátviteli topológia csoport tagja (útvonalanként maximum 2, az alábbiak közül: ligandum, receptor, mediátor, kofaktor, transzkripciós faktor, egyéb fehérje)
- A kapcsolatok irányítottak, tartalmazzák a kapcsolat hatás-információját (aktiválás, gátlás), valamint rendelkeznek az eredeti kísérletes irodalmi forrás Pubmed azonosítójával.

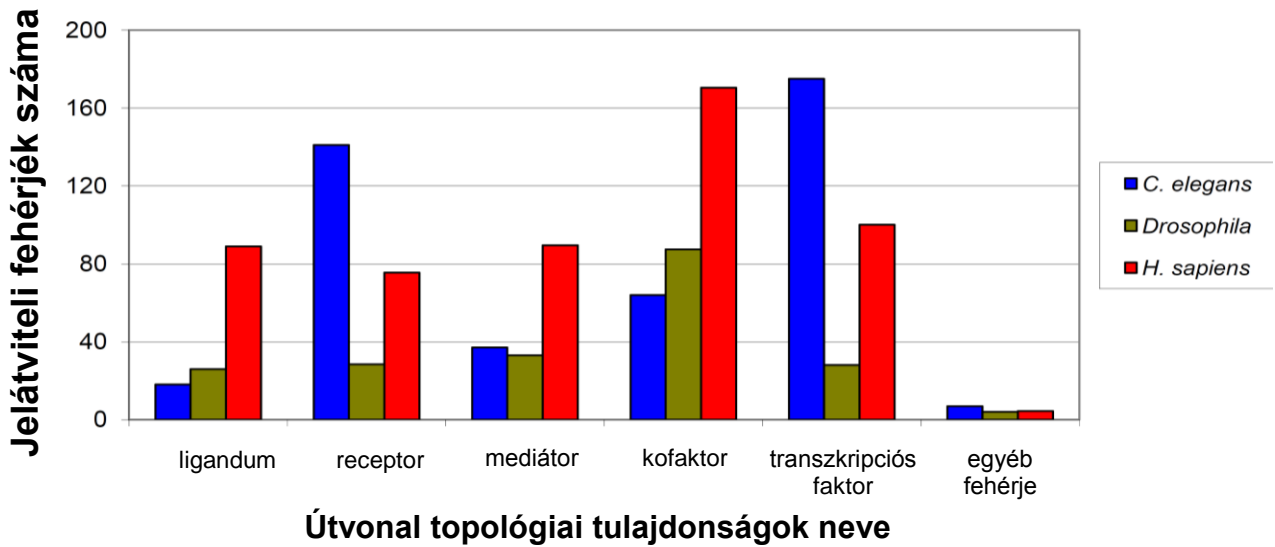
IV.1.2. A SignaLink adatbázis mennyiségi tulajdonságai

A SignaLink 2010-ben publikált verziója (amely gyűjtése 2008-ban lezárult) 442 *C. elegans* fehérjét, 211 *D. melanogaster* fehérjét és 525 emberi fehérjét tartalmaz, amelyek között rendre 237, 233 és 991 kapcsolat található. Mivel egy fehérje csak egyszer szerepelhet az útvonal gyűjteményben, de több útvonalhoz is tartozhat, nem triviális a fehérjék útvonalankénti számának meghatározása. Ennek érdekében Farkas Illéssel súlyoztuk a fehérjék számát útvonalanként, illetve a jelátviteli topológiai csoport (pl.: receptor, mediátor) szerint. Ha például egy fehérje n jelátviteli útvonalnak a tagja, és ezek az útvonalak az EGF/MAPK és TGF- β , akkor $n=2$. Ebben az esetben mindkét

útvonalban a fehérje $1/n$ részben van jelen. Hasonlóan, ha egy fehérjéhez m darab jelátviteli topológia tulajdonság rendelhető (pl.: EGF/MAPK receptor, TGF- β receptor és TGF- β mediátor, akkor $m=3$, tehát a fehérje $1/m$ részben van jelen az egyes jelátviteli topológiai csoportokban. A **11. ábra** az egyes komponensek útvonalakba történő csoportosítását, míg a **12. ábra** a jelátviteli topológiai tulajdonság szerinti megoszlásukat mutatja.

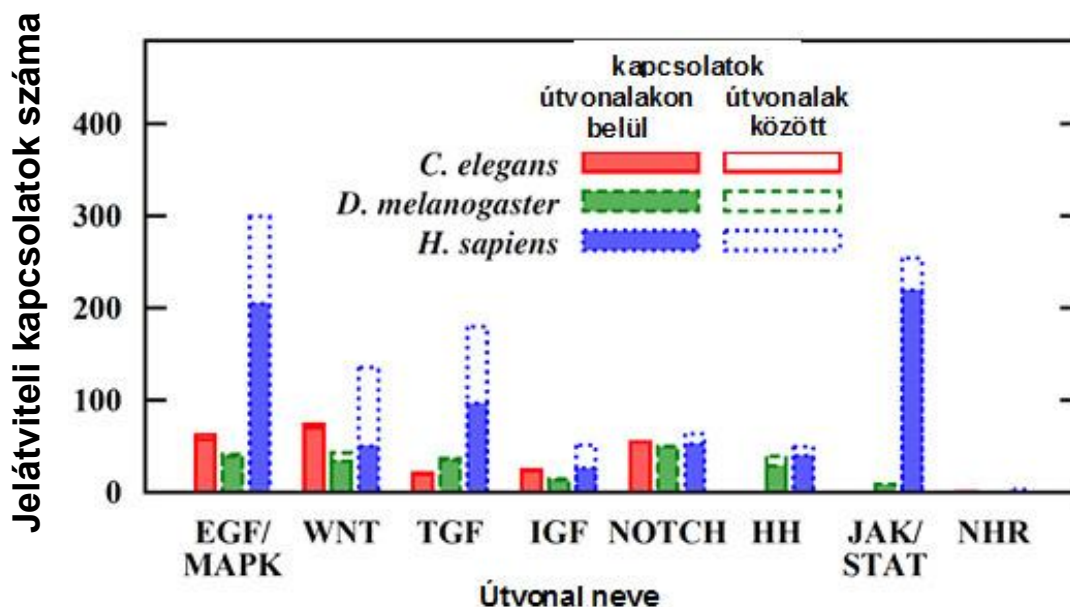


11. ábra Súlyozott fehérjeszámok az egyes útvonalakban fajonként. A központi (gerinc) és mellék (perifériás) jelátviteli útvonal fehérjék teli és üres oszlopokkal vannak jelölve. A fehérjék súlyozott száma figyelembe veszi, hogy egy fehérje több útvonalnak is tagja lehet, és ezzel az útvonalszámmal osztja az adott fehérje számát. A fajok és az útvonalak közötti különbségek részletes elemzése a IV.4.2. fejezetben kerül bemutatásra.



12. ábra Súlyozott fehérjeszámok az egyes fajokban a jelátviteli topológiai tulajdonságok szerint csoportosítva. A fehérjék súlyozott száma figyelembe veszi, hogy egy fehérjére több topológiai tulajdonság is jellemző lehet, és ezzel a tulajdonság-számmal osztja az adott fehérje számát. A fajok és az topológiai csoportok közötti különbségek részletes elemzése a IV.4.2. fejezetben kerül bemutatásra.

A jelátviteli kapcsolatokat külön határoztuk meg az útvonalakon belül és között, valamint a jelátviteli topológiai csoportok között. A kapcsolatok az adatbázisban nem súlyozottak (pl.: erősség vagy valószínűség szerint), de hasonlóan a fehérjék számolásához, a kapcsolatokat is súlyoztuk annak alapján, hogy hány útvonalat vagy jelátviteli topológiai csoportot kötnek össze. Például ha két kapcsolódó fehérje (A és B) n és m darab útvonalnak a tagja, akkor a kapcsolataik $1/(n*m)$ súllyal járulnak hozzá minden egyes útvonal közötti kapcsolathoz (keresztbeszélgetéshez). A SignaLinkben a legtöbb jelátviteli kapcsolat egyirányú ($A \rightarrow B$, $B \rightarrow A$), de megkülönböztettünk kétirányú kapcsolatokat is ($A \rightarrow B$, $B \rightarrow A$). Az útvonalakon belüli és közötti kapcsolatok számát a **13. ábra** mutatja be.



13. ábra. Súlyozott jelátviteli kapcsolatok száma az egyes fajokban útvonalanként. A kapcsolatok súlyozott száma figyelembe veszi, ha a kapcsolódó fehérjéparok több különböző útvonalban is szerepelhetnek. Az ábrán megkülönböztetjük az útvonalakon belüli kapcsolatokat (teli szín) és az útvonalak közötti kapcsolatokat, keresztbeszélgetéseket (fehér színű oszlopok). A fajok és az útvonalak közötti különbségek részletes elemzése a IV.4.2. fejezetben kerül bemutatásra.

IV.2. A SignaLink adatbázis összehasonlítása más adatbázisokkal

Mivel jelenleg nincs elérhető *standard* jelátviteli útvonal adatbázis, amivel validálni lehetne a SignaLink adatbázist, három ismert és sokat használt adatbázis humán adataival hasonlítottuk össze: a KEGG, Reactome és NetPath adatbázisokkal (Joshi-Tope és mtsai, 2005; Kandasamy és mtsai, 2010; Ogata és mtsai, 1999). A páronkénti összehasonlítások részletes eredményei megtalálhatóak a (Korcsmaros és mtsai, 2010) publikáció mellékletében. Ebben a fejezetben a négy útvonal-adatbázis egymással történt összehasonlítását mutatom be. Ez az összehasonlítás kiemeli az adatbázisok fontosabb minőségi tulajdonságait, különös tekintettel a SignaLinkben szereplő tulajdonságok meglétére vagy hiányára (12. táblázat).

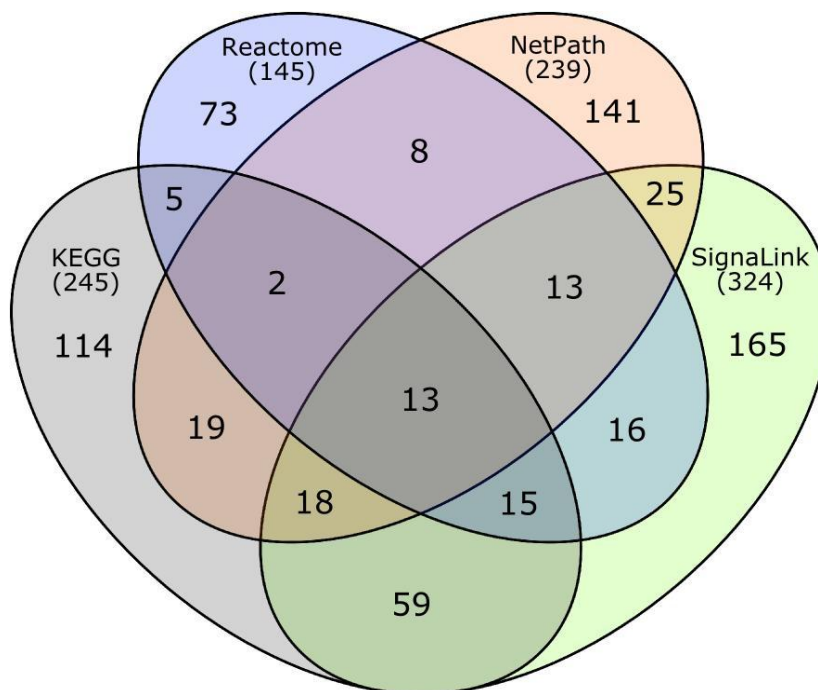
12. táblázat A SignaLink és három másik jelátviteli útvonal-adatbázis (KEGG, a Reactome és NetPath) tulajdonságainak összehasonlítása. A táblázat a SignaLinktől való eltéréseket mutatja. A legfontosabb minőségi különbségek szürke háttérrel szerepelnek (2008-as állapot alapján).

	KEGG	Reactome	NetPath	SignaLink
A jelátviteli fehérjék forrása	Szelektív gyűjtés az irodalomból	Reakció alapú kézi gyűjtés	Szelektív kézi gyűjtés a HPRD interakciós adatbázisból	Kézi gyűjtés útvonal-review-k és kísérletes irodalmi források alapján
A kapcsolatok forrása	Csak review-k	Pubmed ID és reakció részletek minden egyes kapcsolathoz	Pubmed ID és reakció részletek minden egyes kapcsolathoz	Pubmed ID minden egyes kapcsolathoz
Az útvonalak száma	Sok, alapvetően emberben (nincs NHR)	10, de nem minden fajban (nincs Hh, NHR, JAK/STAT)	20, csak emberben (nincs WNT, IGF, NHR, JAK/STAT)	8, három fajban (EGF/MAPK, WNT, TGF- β , Notch, IGF, Hh, NHR, JAK/STAT)
Útvonalak definíciója	Nem elérhető	Nem elérhető; Egységes gyűjtési módszer	Rák vagy immun-útvonalak	Biokémiailag definiált útvonalak, fontosak a fejlődésben; Egységes gyűjtési módszer
Az útvonalak közös platformon és a kereszt-beszélgetések elemezhetőek	Nem lehetséges	Lehetséges, de nincs globális útvonalnézet	Nem lehetséges	Lehetséges

A páronkénti összehasonlítások során azt láttuk, hogy az egyes útvonalak méretei jelentősen különböztek a SignaLinkben és a vizsgált 3 másik útvonal-adatbázisban. A **13. táblázat** összehasonlítja a 4 útvonal-adatbázis fontosabb mennyiségi tulajdonságait. Mind a 4 adatbázisban 4 közös útvonal szerepelt, az EGF/MAPK (MAPK, EGFR1), Notch, TGF- β és a WNT. Így ezen útvonalak fehérje és kapcsolatszámát tudtuk összehasonlítani a négy adatbázis között (**14. ábra**).

13. táblázat A SignaLink, a KEGG, a Reactome és NetPath útvonal-adatbázis mennyiségi tulajdonságainak összehasonlítása. A KEGG, Reactome és NetPath esetében azokat az útvonalakat összehasonlítottuk, amelyek a SignaLinkben lévő 8 útvonal valamelyikével megegyeztek (pl.: az EGFR, EGFR1 és MAPK útvonalakat mind az EGF/MAPK útvonallal vettettük össze). A KEGG esetében a „mátrix” adatsort, a Reactome esetében a komplexeken belüli kapcsolatok nélküli adatsort néztük (ld. III.2.2-3. és IV.2.1-2.). * Zárójelben az összehasonlított útvonalak SignaLinkre jellemző értékei láthatóak.

	KEGG	Reactome	NetPath	SignaLink
Útvonalak	MAPK, Insulin, Hedgehog, JAK/STAT, Notch, TGF-β, WNT	EGFR, Insulin receptor, Notch, TGF, WNT	EGFR1, Hedgehog, Notch, TGF, WNT	EGF/MAPK, IGF, Hedgehog, JAK/STAT, Notch, NHR, TGF-β, WNT
Fehérje és kapcsolat darabszámok*				
Fehérjék	429 (483)	120 (348)	362 (355)	525
Kapcsolatok	1.502 (990)	682 (689)	457 (701)	991
<i>ebből keresztbeszélgetés</i>	225 (228)	79 (226)	60 (171)	300
Felhasznált cikkek száma	73	166	351	941



14. ábra. A KEGG, Reactome, NetPath és SignaLink 4 közös útvonalának az összehasonlítása egy Venn-diagramon. Az egyes adatbázisok nevei alatt zárójelben szerepel, hogy összesen hány fehérje van az EGF/MAPK, Notch, TGF-β és WNT útvonalakban. (Csak az UniProt/SwissProt adatbázisban is szereplő fehérjék lettek összehasonlítva, az UniProt/TrEMBL adatbázisban lévő, kísérletesen még nem bizonyított fehérjék nem.)

Az adatbázisok összehasonlítása során az alábbi konklúziókat tudjuk levonni a Signalink adatbázisról:

- Pontosan definiált útvonalakat tartalmaz.
- A letölthető fájlok struktúrája és információi alapján lehetséges az útvonalak egységes, rendszerszintű vizsgálata, beleértve a keresztbeszélgetések vizsgálatát is.
- Az elemzett útvonalakból a legtöbb fehérjét és kapcsolatot tartalmazza.
- A legtöbb keresztbeszélgetésben résztvevő fehérjét és keresztbeszélgetést tartalmazza.
- A legnagyobb az átfedése a másik három útvonal-adatbázissal.
- A legtöbb irodalmi forrást tartalmazza.
- Minimálisan tartalmaz izoformákat.
- Alacsony azon fehérjék száma, amelyek kísérletesen nem lettek még bizonyítva (UniProt/TrEMBL-ben listázott fehérjék).

IV.3. A Signalink adatbázis útvonalainak és hálózatának ábrázolásai

Egy-egy útvonal vagy hálózat megismeréséhez fontos az értelmezést támogató ábrázolás is. A hálózatokat többfajta módon lehetséges ábrázolni, és mindig az adott felhasználási cél határozza meg, hogy melyik a megfelelő. A Signalink adatbázisban lévő útvonalakat és hálózatot a hagyományos hálózatelrendezések mellett útvonal-alapú, ún. keresztbeszélgetés-hálózat ábrázolással is megjelenítettük. Ezekre az ábrázolásokra az adatbázis illusztrálása mellett az adatok és elemzések értelmezése miatt is szükség volt. Az ábrák a (Korcsmaros és mtsai, 2010) és a (Korcsmaros és mtsai, 2011) publikációkban kerültek közlésre.

Összesen 7 különböző ábrázolási módot alkalmaztam a vizsgálataim során, amelyek a hálózat jelátviteli, ortológia és keresztbeszélgetés tulajdonságain alapultak. Ezt a 7 megközelítést a résztvevő elemek és a közöttük lévő kapcsolatok alapján 4 csoportba lehet sorolni. Az egyik ilyen csoport a fehérjék közötti hálózatok ábrázolására, a másik három a keresztbeszélgetések megjelenítésére fókuszál:

1. Hagyományos fehérje-fehérje alapú hálózat-elrendezést alkalmaztam 3 esetben, amelyek között a legfontosabb különbség a színezés volt. A fehérjéket és a közöttük lévő kapcsolatokat a jelátviteli tulajdonságaik (útvonaltagság vagy pozíció), illetve az ortológia információjuk alapján színeztem. Itt a 3 vizsgált faj 8-8 útvonalát együttesen és külön-külön is ábrázoltam. Minden így készített hálózatkép dinamikusan *zoomolható* formában elérhető a <http://signalink.org> honlapon.
2. Az egyik keresztbeszélgetés vizsgálatához olyan ábrázolási elrendezést alkalmaztam, amelynél a pontok az útvonalak és a közöttük lévő kapcsolatok irányított keresztbeszélgetések. Ebben az elrendezésben – hasonlóan az előző megközelítéshez – a pontokat és a kapcsolatokat akár a jelátviteli, akár az ortológia információjuk alapján lehet színezni.
3. Egy másik keresztbeszélgetés vizsgálat során kapcsolati mátrixot készítettünk, ahol a kapcsolódó útvonalak közötti keresztbeszélgetések találhatóak a mátrix egyes celláiban (a főátlóban pedig az útvonalakon belüli kapcsolatok). Ennél a megjelenítésnél az expresszió fokának megfelelően színeztük a kapcsolatokat.
4. Végül, a több útvonalban is szerepet játszó többútvonalas fehérjéket ábrázoltuk úgy, hogy ezeket útvonalanként „összejtettük”, és az egyezést az útvonalak közötti kapcsolattal jelöltük.

IV.3.1. Az útvonalak részletes ábrázolása

Az útvonalak és a fehérjék részletes, mechanisztikus ábrázolásához egy olyan ábrázolási módot választottam, ahol a fehérjék tulajdonságai szerint színeztem a pontokat. Ez alapján kétfajta színezést alkalmaztam. Az egyikben az útvonalak és az útvonalrégiók (gerinc, mellék) szerint, a másikon a jelátviteli pozíció csoport (ligandum, receptor, stb.) szerint színeztem a fehérjéket (pontok). Az útvonal régiók esetében minden útvonalat máshogy színeztem, és a gerinc elemeket az adott útvonal színének sötétebb, a mellék elemeket világosabb tónusú színével jelöltem. Ha egy pont (fehérje) több régióhoz vagy pozíció csoporthoz tartozott, akkor ez egy tortadiagramhoz hasonló színezéssel került jelölésre. A **15. ábra** mutatja az alkalmazott színekódokat. Az emberi jelátviteli hálózat példáján látható, hogy a fehérjék útvonalankénti, illetve útvonal pozíciónkénti színezése szerint hogyan néz ki a jelátviteli hálózat (**16. ábra**).

a)

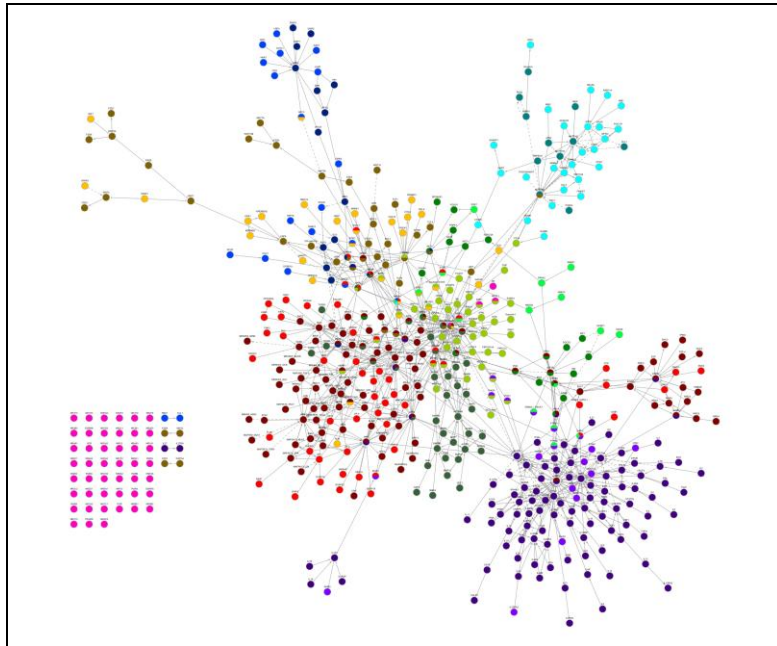
	gerinc	mellék
EGF/MAPK	dark red	red
WNT	olive green	orange
TGF- β	green	light green
IGF	dark green	yellow-green
Notch	teal	cyan
Hedgehog	dark blue	blue
JAK/STAT	purple	light blue
NHR	pink	

b)

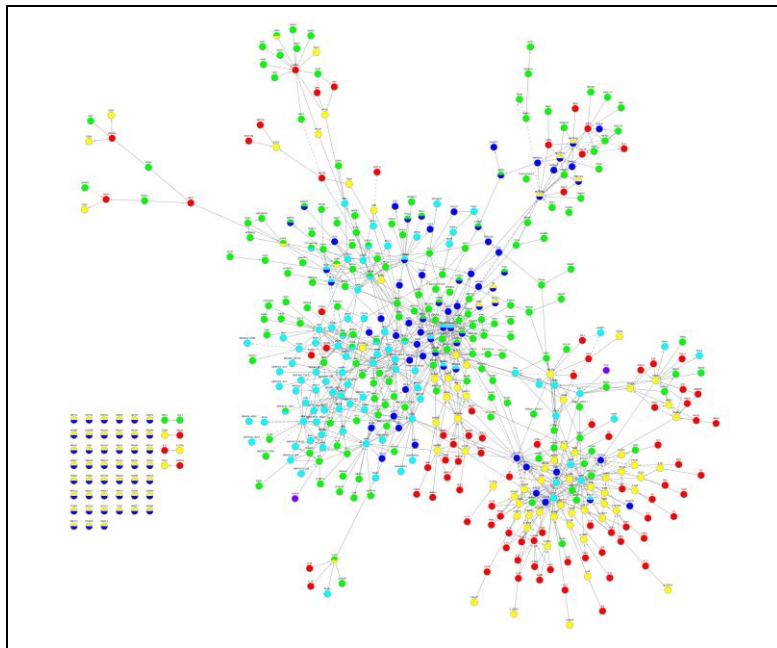
ligandum	red
receptor	yellow
kofaktor	light green
mediátor	cyan
transzkripciós faktor	blue
egyéb fehérje	purple

15. ábra A SignaLink részletes útvonal ábrázolásánál alkalmazott színek. a) Az útvonalrégiók szerinti színezés. b) A jelátviteli topológia pozíció csoport szerinti színezés.

a)



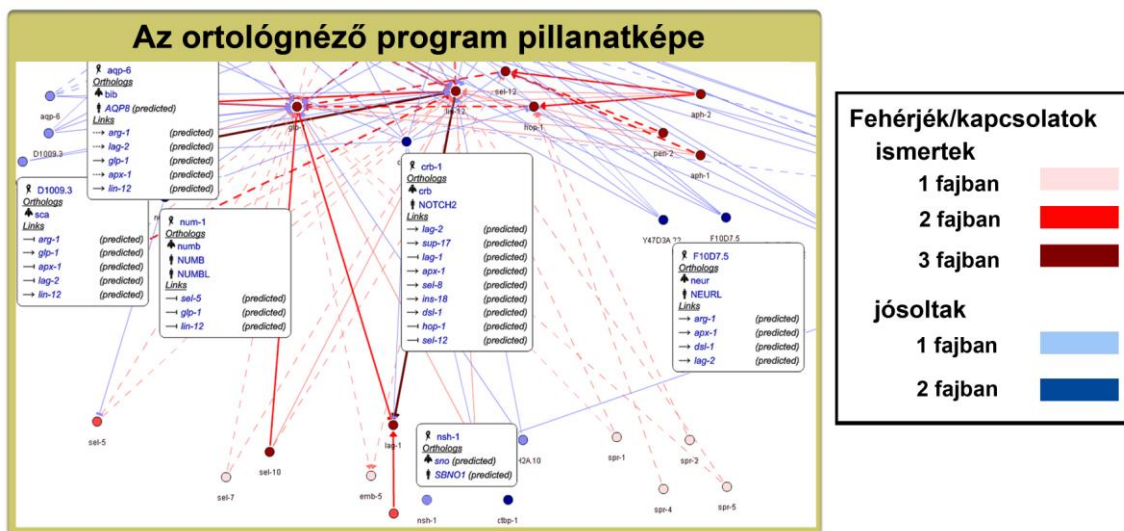
b)



16. ábra. A SignaLink jelátviteli útvonalfehérjék hálózata emberben. **a)** A fehérjék színezése útvonalak és útvonalrégiók szerint. **b)** A fehérjék színezése az útvonalbeli topológia pozíciók szerint. A részletes színekód a **15. ábra** látható.

IV.3.2. A jószolt komponensek ortológia-alapú ábrázolása

Annak ellenére, hogy ma már igen sok hálózatábrázolási eszköz létezik (Gehlenborg és mtsai, 2010), ezek közül csak igen kevés képes egyszerre ábrázolni az ismert és a jószolt fehérjéket, kapcsolatokat. Ennek érdekében Palotai Robin informatikussal kifejlesztettünk egy ortológia hálózatnéző programot (<http://signalink.org/signalog>). Ezzel a nézőprogrammal ábrázolhatóak az útvonaltagság jóslások (ld. IV.4.1. fejezet), valamint a felhasználónak lehetősége van útvonalanként vagy együttesen megvizsgálni az egyes fajok jelátviteli hálózatait. A színezés az ortológia foka szerint történik, külön jelölve az ismert és a jószolt fehérjéket, illetve kapcsolatokat (**17. ábra**). Az általunk kifejlesztett nézőprogram különlegessége, hogy bármikor átválthatunk a fajok között, miközben a nézet nem változik, csak a fehérjék elnevezése és a színezés. Az interaktív elemzést elősegítendő a fehérjék fajspecifikus adatbázisokhoz, a kapcsolatok Pubmed azonosítókhöz vannak hiperlinkelve, így a további adatok egy kattintásra elérhetőek.

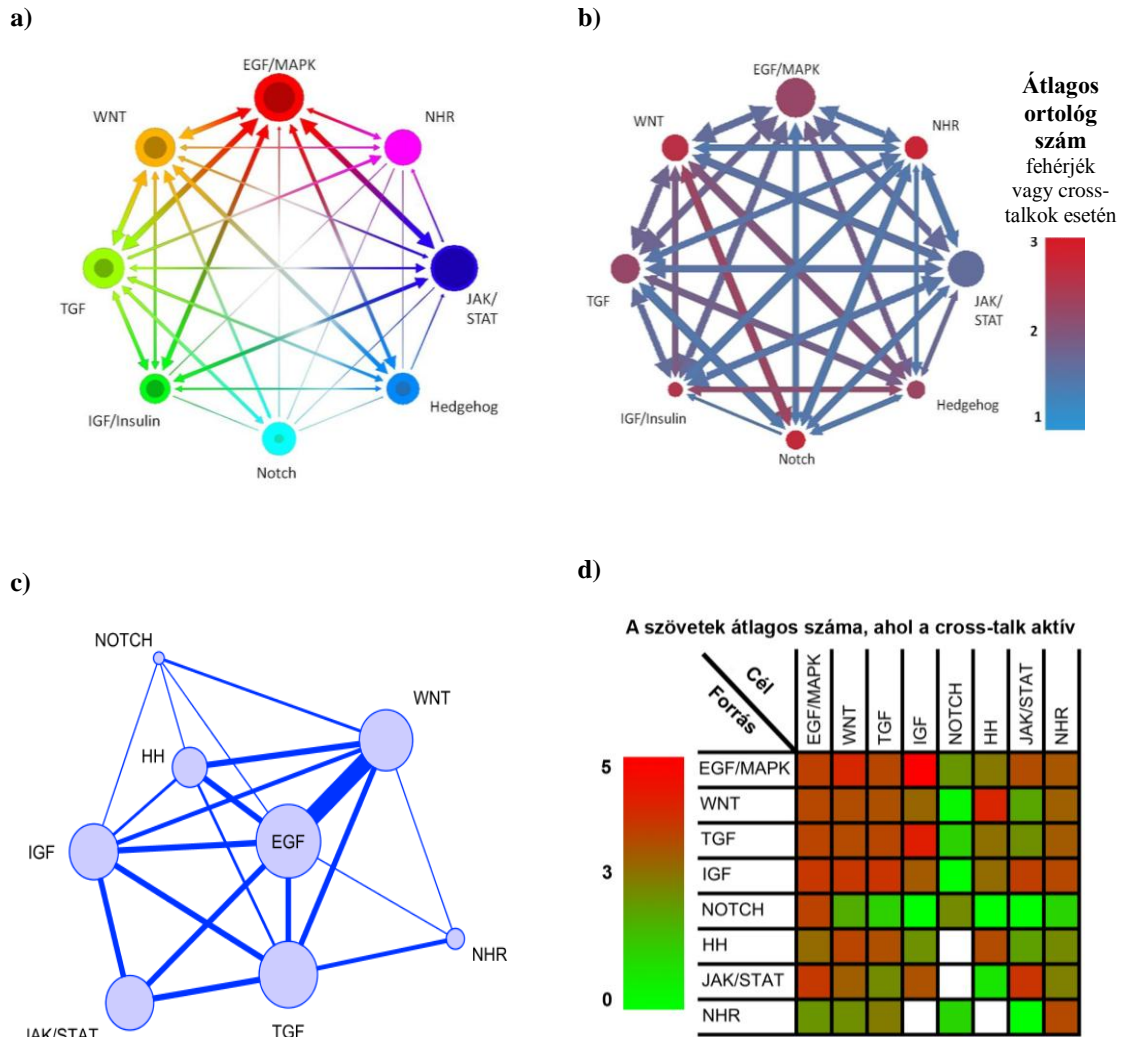


17. ábra Az ortológnező program többszörösen kombinált képernyőnézete a Notch útvonalról. Az ismert fehérjék és kapcsolatok jelenlétét más vizsgált fajokban a piros árnyalataival, míg a jószolt fehérjéket és kapcsolatokat a kék árnyalataival jelöltük.

IV.3.3. Az útvonal-hálózatok ábrázolása

Az útvonal hálózatok ábrázolása a Signalink adatbázis elemzése kapcsán végzett keresztbeszélgetés vizsgálatok és a többútvonalas fehérjék azonosítása miatt került kidolgozásra (részletesen ld. IV.4.2. fejezet). Ezen ábrázolási elrendezések és

megjelenítési módok összevetése céljából a **18. ábra** 1-1 típuspéldát mutat az általunk kidolgozott útvonal, illetve keresztbeszélgetés-ábrázolások közül. Ebben a fejezetben a megjelenítési típusokat elemezem, míg az ábrák tartalmi elemzése az ezt követő fejezetekben kerül kifejtésre.



18. ábra A SignaLink adatbázis útvonalalapú, keresztbeszélgetés megjelenítési típusai az emberi útvonal keresztbeszélgetés hálózatának esetén. **a)** Útvonalanként színezve. **b)** Az ortológia foka alapján színezve. **c)** Az azonosított többútvonalas fehérjék hálózata, útvonalanként összeajtva a fehérjéket, az útvonalak közötti kapcsolatok közös fehérjét jelentenek. **d)** Az útvonal kapcsolatok szöveti aktivitásának vizsgálata céljából készített mátrix.

A **18. ábra** a-b) részén az egyes pontok útvonalakat reprezentálnak, a kapcsolatok pedig keresztbeszélgetéseket. A pontok mérete arányos az adott útvonalhoz tartozó fehérjék számával, a kapcsolatok vastagsága az adott útvonalpár között futó fehérje-fehérje kapcsolatok számával arányos. A kapcsolatok nyilainak mérete pedig az adott irányba menő kapcsolatok számával arányos. A színezés az **a)** ábrarészen az útvonal típusa szerint történt, hasonlóan a 16. ábra, míg a **b)** ábrarészen a jobb oldalon lévő skála szerint színeztük a hálózatot. Ez a skála az ortológia fokát, tehát a rokon fehérje vagy a rokon fehérjék közötti kapcsolatok átlagos számát jelzi az adott útvonalban vagy keresztbeszélgetésben. A **18. ábra c)** ábrarészen a korábbiakhoz képest más megközelítést láthatunk. Itt a pontok csak a több útvonalba is tartozó, többútvonalas fehérjéket jelzik, útvonalanként csoportosítva. A pontok mérete az adott útvonalban lévő többútvonalas fehérjék számával arányos, míg a közöttük lévő kapcsolat vastagsága a közös fehérjék számával arányos. A **18. ábra d)** ábrarészen útvonal-mátrix elrendezésben az egyes kapcsolatokban résztvevő fehérjék expressziójának fokát jelenítjük meg egy színskála segítségével. A kapcsolat-mátrix elrendezéséből fakadóan a főátlóban az útvonalon belüli kapcsolatok, az átló alatt és felett az irányított keresztbeszélgetések becsült expressziója látható. (Ezen az ábrán nem egy, hanem öt szövet átlagos expressziója látható.)

IV.4. A SignaLink adatbázison végzett vizsgálatok eredményei

A létrehozott SignaLink adatbázis olyan – más adatbázisokban nem elérhető – minőségi és mennyiségi tulajdonságokkal rendelkezik, amelyek lehetővé teszik új, rendszerszintű elemzések elvégzését. Négy ilyen elemzést végeztünk a SignaLink adatbázis segítségével:

1. Új génfunkciók azonosítása és elemzése.
2. A jelátviteli útvonalak fajkon belüli és fajok közötti összehasonlítása.
3. A keresztbeszélgetések szövet-, és betegség-specifikus elemzése.
4. Új gyógyszercélpont-jelöltek azonosítása.

Az alábbi fejezetekben ezen vizsgálatok eredményeit mutatom be.

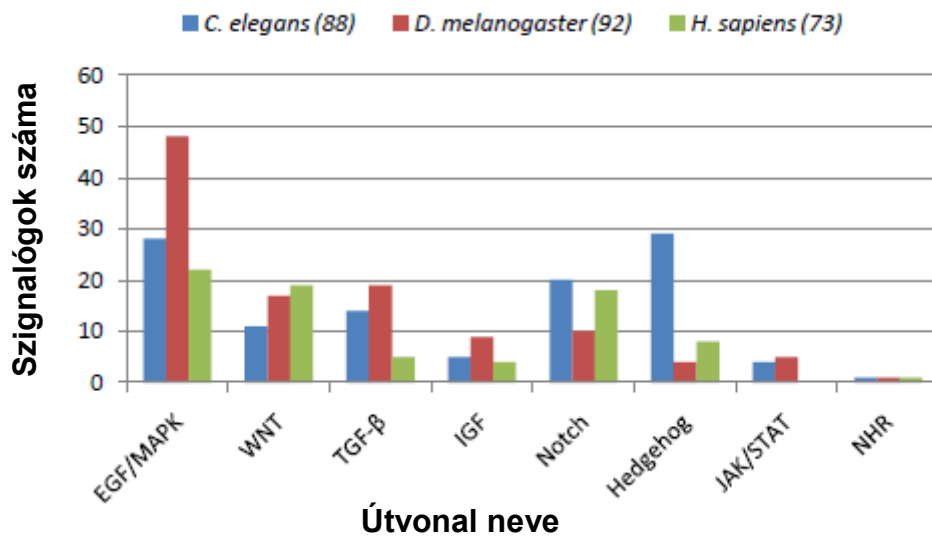
IV.4.1. Új génfunkciók azonosítása és elemzése

Tekintettel arra, hogy a SignalLink adatbázis készítésekor az egyes fajok fehérjéit egységes szabályok szerint gyűjtöttem, lehetőség nyílt a fajok közötti összehasonlítások elvégzésére. Ezen összehasonlítások alkalmasak lehetnek ún. ortológia alapú génfunkciók jóslására, és így új jelátviteli komponensek azonosítására, elemzésére. Az itt bemutatásra kerülő eredmények a (Korcsmaros és mtsai, 2011) publikációban kerültek közlésre.

IV.4.1.1. Új jelátviteli komponensek jóslása és elemzése

A SignalLink adatbázis fajok közötti összehasonlítása során sikerült új jelátviteli komponenseket azonosítanom. Összesen 88 *C. elegans*, 92 *D. melanogaster* és 73 emberi fehérjét találtam, amelyek jelátviteli szerepe eddig nem volt ismert, de valamelyik másik fajban létezik olyan ortológjuk, amely jelátviteli funkcióval rendelkezik. Feltételeztem, hogy ez a 253 fehérje ugyanabban az útvonalban funkcionál, mint az ismert ortológja. Az ortológiai tulajdonság alapján azonosított új jelátviteli fehérjéket szignológoknak neveztem el. A 253 fehérjéhez 301 útvonaltagságot jósoltam, mivel 39 szignológot több útvonalba is jósoltam. A szignológok teljes listája megtalálható a <http://signalink.org> honlapon, és interaktívan elemezhetőek a <http://signalink.org/signalog> honlapon.

A szignológok jellemzése céljából összehasonlítottam a jósolt útvonaltagságaikat (**19. ábra**). A szignológok jelentős része az EGF/MAPK útvonalba tartozik (a szignológok 25%-a a férgekben, 42,5%-a a gyümölcslegyekben és 28,6%-a az emberben). Ez az eredmény egybevág azzal, hogy az EGF/MAPK útvonalnak van a legtöbb fehérjéje (ld. IV.1.2. és (Korcsmaros és mtsai, 2010)). A *C. elegans*-ban körülbelül ugyanennyi szignológ lett a Hedgehog útvonalba jósolva (29 fehérje; 25,9%). Ez azért érdekes, mert a féregben a Hedgehog útvonal nem aktív, mivel az evolúció során bizonyos Hedgehog-gének eltűntek a féreg genomjából (Pires-daSilva és Sommer, 2003). Ugyanakkor az eredményeink és az irodalmi adatok azt mutatják, hogy a Hedgehog útvonal fontos elemei jelen vannak ebben az élőlényben is (Korcsmaros és mtsai, 2010; Pires-daSilva és Sommer, 2003). Az EGF/MAPK mellett legyekben a szignológok a WNT és TGF- β útvonalakban, az emberben a WNT, Notch és Hedgehog útvonalakban vannak leginkább.



19. ábra A SignaLink adatbázis fajok közötti összevetése során jóslott, ún. szignálóg fehérjék útvonalankénti megoszlása az egyes fajok között. A fajnevek után a zárójelben az adott fajban talált szignálógok száma szerepel. Mivel 39 szignálóg több útvonalba is tartozik, az egyes útvonalakban lévő szignálóg számot összeadva több szignálógot is kaphatunk, mint az adott fajban talált összes szignálóg száma.

Ezután megvizsgáltam az ortológia alapú útvonaltagság átvitelének koncepcióját. Mindhárom vizsgált faj esetében a SignaLink ismert jelátviteli fehérjéit az ortológiai információk alapján (InParanoid 6.1.-es verzió; (Berglund és mtsai, 2008)) illesztettük. Az ortológgal rendelkező fehérjék aránya férgekben 93,4%, gyümölcslégyben 81,6%, míg emberben 64,8% volt. Ezen ortológok jelátviteli szerepét megvizsgálva azt találtam, hogy rendre 83,2%, 67,5% és 82,6% olyan ortológ fehérje létezik, amelynek van ismert jelátviteli szerepe. Azaz a három fajban átlagosan a fehérjék 77,8% rendelkezik olyan ortológgal, amely jelátviteli funkciót tölt be legalább egy másik fajban. Átlagosan a jelátviteli fehérjék 67,8%-a rendelkezett legalább egy olyan ortológgal, amely ugyanabban az útvonalban tölt be funkciót. Tehát az ortológia alapú jelátviteli szerepjöslés releváns koncepció a jelátviteli funkciók azonosítása szempontjából.

IV.4.1.2. A szignálóg-jóslások újdonságvizsgálata

Génfunkció jóslásokkal kapcsolatos vizsgálatoknál három kérdést kell megvizsgálni:

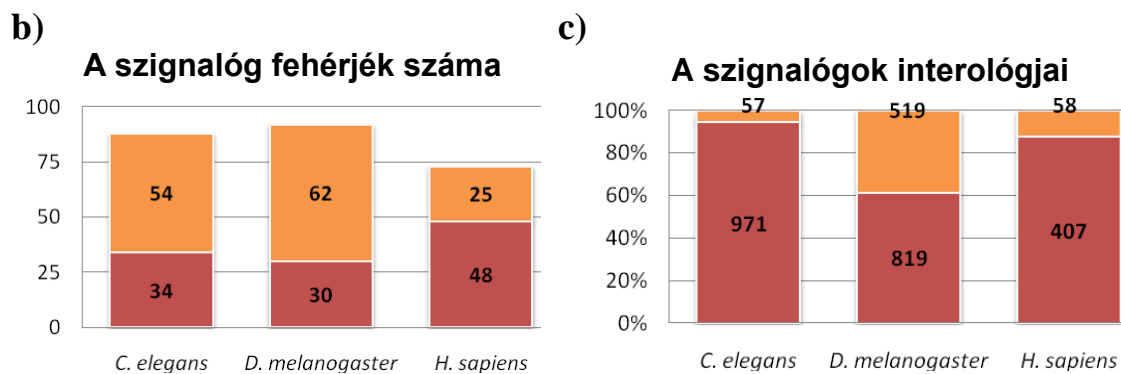
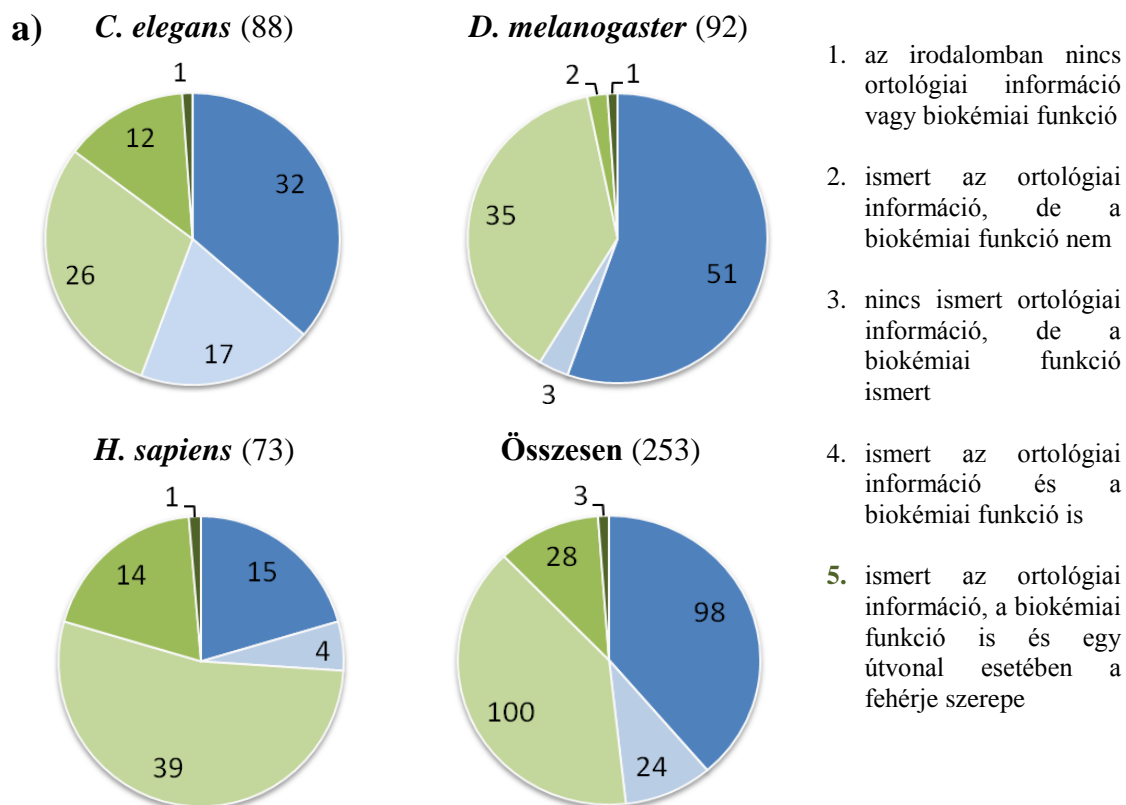
- 1) A jóslott tulajdonság valóban új-e?
- 2) A jóslott komponensekről készült-e már valamilyen jóslás?
- 3) A jóslott tulajdonság megfelelő-e?

Az első két kérdést a szignálógok újdonságvizsgálatával kívántam megválaszolni, míg a harmadik kérdésre a következő alfejezetben térek ki.

Annak érdekében, hogy megállapítsam, hogy a szignálógok mennyire tekinthetők új jelátviteli komponensnek, 5 csoportba osztottam ezeket a fehérjéket. A szignálógokat az alapján csoportosítottam a 3 fajban, hogy mennyire ismertek a tulajdonságaik az irodalomban (ld. III.4.1.3. alfejezet). Az osztályozás az ismeretlenebb (1) felől haladt az ismertebb (5) felé (**20. ábra**). Mindhárom fajban csak 1-1 olyan fehérjét találtam, amelynek már volt ismert jelátviteli szerepe. Ezeket a fehérjéket nem zártam ki a szignálógok listájából, mivel ezekhez is tudtam további új útvonalakat jósolni. Nem meglepő módon azt találtam, hogy az irodalomban főleg az emberi szignálóg fehérjék az ismertebbek: 74 százalékukról rendelkezünk valamilyen adattal. Ezzel szemben *C. elegans*-ban és *D. melanogaster*-ben a szignálóg fehérjék több mint a fele (55,7%, ill. 58,7%) nincs karakterizálva. Mindezen vizsgálatok alapján megállapítható, hogy a szignálóg fehérjék valóban új jelátviteli komponenseknek tekinthetők.

Ezt követően azt vizsgáltam meg, hogy a jóslott komponensek mennyire vannak jelen hasonló módszerekkel készült predikciókban. Ehhez az összehasonlítóshoz ortológia alapján jóslott, ún. interológ kapcsolatokat használtam fel (Yu és mtsai, 2004). Összevettem a Signalink adatbázisból készített interológokat azokkal a már publikált jóslásokkal, amelyeket három fajspecifikus interológ adatbázis (WI8, DroID, HomoMINT) tartalmaz (Persico és mtsai, 2005; Simonis és mtsai, 2009; Yu és mtsai, 2008). Csak azokat az interológokat vizsgáltam, amelyekben legalább egy szignálóg fehérje részt vesz. A feregben, a gyümölcsleányban és az emberben rendre 34, 30 és 48 olyan szignálógot találtam, amely csak a Signalinkben fordul elő. Ez azt jelenti, hogy nagy azon fehérjék aránya, amelyeket ortológia alapú módszerekkel eddig nem vizsgáltak (**20. ábra**). A Signalinkben és a három interológ adatbázisban összesen

1028, 1338 és 465 interológot találtam rendre a féregben, a gyümölcsleányban és az emberben. Az átfedés a SignaLink és a fajspecifikus adatbázisok között igen alacsony volt: 5,5% a féregben, 38,8% a gyümölcsleányban és 12,5% az emberben (**20. ábra**). Ezek a közös interológok értelmezhetőek úgy, mint már ismert, publikált ortológia alapú predikciók. Mivel kevés ilyen interológ van, feltételezhetjük, hogy a legtöbb szignológ jóslás új, eddig nem publikált információkat tartalmaz. A csak a SignaLinkben jelen lévő interológok magas száma az alábbi 3 ok miatt valószínű: (1) a SignaLink készítése során alkalmazott egységes gyűjtési módszer a 3 faj esetében; (2) a jelátviteli fehérjék alul reprezentációja a fajspecifikus interológ adatbázisokban (5-5% a féregben és a gyümölcsleányban és 2,8% az emberben); (3) az emberi interológ adatbázisban (HomoMINT) alkalmazott szigorú megbízhatósági szűrő. Összességében megállapíthatjuk, hogy a szignológok nagy része még nem szerepelt ortológ alapú prediktáló vizsgálatban, valamint a SignaLinkhez hasonló, egységesen gyűjtött adatsorok elősegíthetik az ortológiai alapú predikciókat.



Jelen van
■ csak a SignaLinkben ■ a SignaLink és a 3 interológ adatbázis metszetében

20. ábra A szignológ fehérjék újdonságvizsgálata. a) A szignológok osztályozása az irodalomban ismert tulajdonságaik alapján. b-c) A szignológ és az interológ (ortológ kapcsolat) adatok összehasonlítása a SignaLink és három fajspecifikus interológ adatbázis (WI8, DroID, HomoMINT) felhasználásával. Készült a (Korcsmaros és mtsai, 2011) publikáció nyomán.

Végül azt vizsgáltam meg, hogy a jóslott tulajdonságok mennyire vannak jelen a már létező útvonal-adatbázisokban. Ebből a célból összevettem a szignológok listáját és jóslott útvonaltagságait a KEGG és a Reactome adatbázissal (Joshi-Tope és mtsai, 2005; Ogata és mtsai, 1999). Érdeemes megjegyezni, hogy a KEGGben mindhárom vizsgált faj útvonaladatai szerepelnek, míg a Reactome-ben csak az emberi adatok. A feregben, gyümölcsleányban és emberben 23, 19 és 35 olyan szignológot találtam, amely a KEGGben már szerepelt, azaz 20, 26 és 47 százaléka a szignológoknak már szerepelt ebben a jelátviteli útvonal adatbázisban. A három fajból azonban csak 11, 5 és 15 fehérje (átlagosan 13,3%) volt a KEGGben azokhoz az útvonalakhoz rendelve, amelyhez jósltam. A Reactome az emberi szignológok közül 16 fehérjét (22%) tartalmazott, de csak egyet abban az útvonalban, amelyhez jósltam. Tehát az összehasonlításhoz használt fajok és adatbázisok függvényében a szignológok 20-47 százaléka van jelen a KEGGben vagy a Reactome-ban. Ugyanakkor a jóslott útvonaltagságok nagytöbbsége (86,7%-a) új, még nem szerepelt egyik adatbázisban sem.

IV.4.1.3. A szignológ-jóslások kísérletes validációja

A szignológok útvonaltagság jóslását kísérletesen hat Notch útvonalhoz jóslott *C. elegans* gén esetében vizsgáltuk meg, és bizonyítottuk be. Ez a hat gén: *aqp-6*, *nsh-1*, *num-1*, *crb-1*, *D1009.3*, *F10D7.5*. Azt találtuk, hogy ezek a gének genetikai interakciót mutatnak a *lin-12* génnel, amely a Notch receptort kódolja. Tehát sikerült kísérletesen validálnunk mind a hat gén jóslott Notch útvonaltagságát. Mivel ezeket a kísérleteket Rovó Petra TDK hallgató végezte, így részletesebben nem kerülnek a disszertációban bemutatásra. Az eredmények a (Korcsmaros és mtsai, 2011) publikációban kerültek közzlésre.

IV.4.2. Összehasonlító vizsgálatok

A SignalLink egységes gyűjtési módszerek mellett jött létre: mindhárom vizsgált faj útvonalainak fehérjéi és kapcsolatai azonos szempontok alapján kerültek be az adatbázisba. Ennek köszönhetően egy olyan útvonal-adatbázis jött létre, amelyben a jelátviteli útvonalakat, az azokat összekötő többútvonalas fehérjéket, és magukat a keresztbeszélgetéseket hasonlíthatjuk össze akár egy-egy fajon belül, akár a három faj

között. Az itt bemutatásra kerülő eredmények a (Korcsmaros és mtsai, 2010) publikációban kerültek közlésre.

IV.4.2.1. Az útvonalak összehasonlító elemzése

Ebben az alfejezetben az útvonal, útvonal-régió és útvonal-pozíció számadatait elemzem először a fajokon belül, majd a fajok között összehasonlítva.

Mindhárom fajban a 8 útvonalból csak néhány tartalmaz a többihez képest jelentősen több fehérjét. Az összes fehérje közül 26-38% között van azoknak a száma, amelyek az EGF/MAPK és a WNT útvonalakban vesznek részt. A többi útvonal közül a féregben az NHR, a gyümölcslégyben a Hedgehog és a Notch, az emberben a TGF- β és a JAK/STAT útvonalban van még jelentős mennyiségű fehérje. Összességében minden egyes fajban a fehérjék 68-85%-a vesz részt ezekben az útvonalakban, és az összes többútvonalas fehérje 56-70%-a van az EGF/MAPK, TGF- β és WNT útvonalakban.

A fajokon belüli útvonal régió összehasonlítás során azt láthatjuk, hogy a *C. elegans*-ban a legtöbb útvonal esetében a fehérjék egyenlő arányban vannak jelen az útvonalak központi (gerinc) és mellék (perifériás) régióiban. Ez alól a Notch és az NHR útvonalak a kivételek. Előbbiben a központi, utóbbiban a mellék régió van túlsúlyban. Ezzel szemben a másik két fajban a központi és a mellék régiójú fehérjék aránya minden útvonal esetében 1,5.

A fajokon belüli útvonal pozíciók összehasonlítása rámutatott arra, hogy mindhárom fajban a legtöbb mediátor pozíciójú jelátviteli fehérje az EGF/MAPK és az IGF útvonalakban van. Tehát itt van a többihez képest sok olyan fehérje, amely a receptorok és a transzkripciós faktorok között továbbítja a jelet. Ezek jelentős része kináz funkcióval rendelkezik. Féregben a transzkripciós faktorok a legnépesebb pozíciójú csoport (39%-kal). A másik két fajban pedig a kofaktorok (32 és 42%-kal). Emberben a ligandumok és receptoraik főleg a JAK/STAT útvonalban alkotnak jelentős méretű jelátviteli pozíció-csoportot.

A fajok közötti útvonal régió összehasonlítás szerint az ember EGF/MAPK útvonalának központi régiója jelentősen több fehérjét tartalmaz a másik két fajhoz képest (a nagyszámú párhuzamos MAPK kaszkádok miatt). Ugyanezt a jelenséget láthatjuk az emberi JAK/STAT útvonal központi régiója esetén is. Itt a nagyszámú receptor és ligandum növeli meg ennek a régióknak a méretét. A Notch útvonalnál

mindhárom faj esetén megfigyelhető, hogy a szabályozó, mellék régió nagyobb méretű, mint a központi. A többi útvonal esetében a régiók aránya teljesen egyforma.

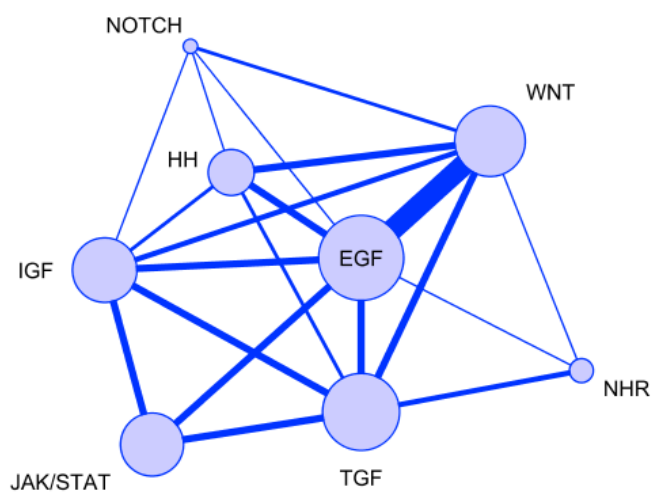
A fajok közötti útvonalméret különbség általában a környezeti különbségekre vezethető vissza, amelyhez az adott élőlények adaptálódtak. Például abban a környezetben, amelyben a *C. elegans* él, a lipofil ligandumok könnyen el tudnak jutni a féreg sejtjeibe és sejtmagjába. Ezért nem meglepő, hogy a féreg NHR útvonala nagyméretű, a jelátviteli fehérjék 58%-a idetartozik. Az NHR útvonal miatt tapasztaltuk, hogy a féregben sok a transzkripciós faktor, hiszen a nukleáris hormonreceptorok mind egyben transzkripciós faktorok is. Mivel az ember sokkal komplexebb élőlény, a sejtjei nagyrésze nem közvetlenül érintkezik a környezet ingereivel (ligandumjaival), így más típusú útvonalakra van szükség ezen jelek továbbításához. Ezért az emberben a sejteket érő jelek jelentős részét például a JAK/STAT útvonal közvetíti, amely a másik két élőlényhez képest igen nagyméretű (emberben a vizsgált jelátviteli fehérjék 21%-a idetartozik). Bár a JAK/STAT útvonal alsóbb (*downstream*) részén nincs olyan sok fehérje, a fő változatosságot és jelkombinációs képességet a nagyszámú ligandumok és receptorok alkotják az útvonal felsőbb részén.

IV.4.2.2. A többútvonalas fehérjék azonosítása és összehasonlító elemzése

Azokat a fehérjéket, amelyek nemcsak egy, hanem több útvonalra is hatnak, azaz keresztbeszélgetésben vesznek részt, többútvonalas fehérjéknek neveztük el. (Azokat a fehérjéket, amelyek több útvonal felől kapják a jelet, ebben a vizsgálatban nem tekintettük többútvonalas fehérjének.) Ezzel a definícióval *C. elegans*-ban, *D. melanogaster*-ben és emberben rendre, 6, 12 és 62 többútvonalas fehérjét azonosítottunk. Az emberi útvonalakban a többútvonalas fehérjék aránya, amely több útvonalban is funkcionál 5%-tól (Notch útvonal) 46%-ig (IGF) terjedt. Több olyan fehérjét is találtam, amely több útvonalnak volt központi eleme. Ilyen például az AXIN nevű állvány (*scaffold*) fehérje és a GSK3 kináz. Mindkét fehérjéről készült már összefoglaló tanulmány annak kapcsán, hogy több útvonalban is szerepük van (Frame és Cohen, 2001; Luo és Lin, 2004).

Érdekes módon csak az EGF/MAPK útvonal az, amely olyan többútvonalas fehérjéket tartalmaz, amelyek minden másik útvonalra hatnak. A skála túloldalán a

Notch és a JAK/STAT útvonalak állnak, amelyek többútvonalas fehérjéi csak 2-3 másik útvonalra hatnak (**21. ábra**). Fontos megjegyezni, hogy a 62 többútvonalas fehérje 45%-át (28 fehérjét) leírták már, hogy valamilyen betegségben nem megfelelően működik. Ezen betegséghez köthető fehérjék aránya az összes jelátviteli fehérjét nézve csak 25,5% (165 a 646-ból), míg az összes ismert fehérjét nézve csak 20% (3.929 fehérje a 19.534-ből). Azaz a többútvonalas fehérjék között szignifikánsan gyakoribbak a betegséghez köthető fehérjék (khi négyzet próba, $p < 0,001$).



21. ábra A SignaLink emberi jelátviteli útvonalaiban azonosított többútvonalas fehérjék hálózata, útvonalanként összeajtve a fehérjéket. Minden pont mérete arányos az adott útvonalba tartozó többútvonalas fehérjék számával, míg az útvonalak között lévő vonalak a közös többútvonalas fehérjék számával arányosak. Tehát ebben az ábrázolásban egy vonal nem fizikai kapcsolatot jelent két fehérje között, hanem közös fehérjét két útvonal között.

Az egyes útvonalakban jelenlévő többútvonalas fehérjék száma: EGF/MAPK (41), TGF- β (29), WNT (22), IGF (17), JAK/STAT (16), Hh (9), NHR (3), Notch (2)

IV.4.2.3. A keresztbeszélgetések összehasonlító elemzése az egyes fajokban

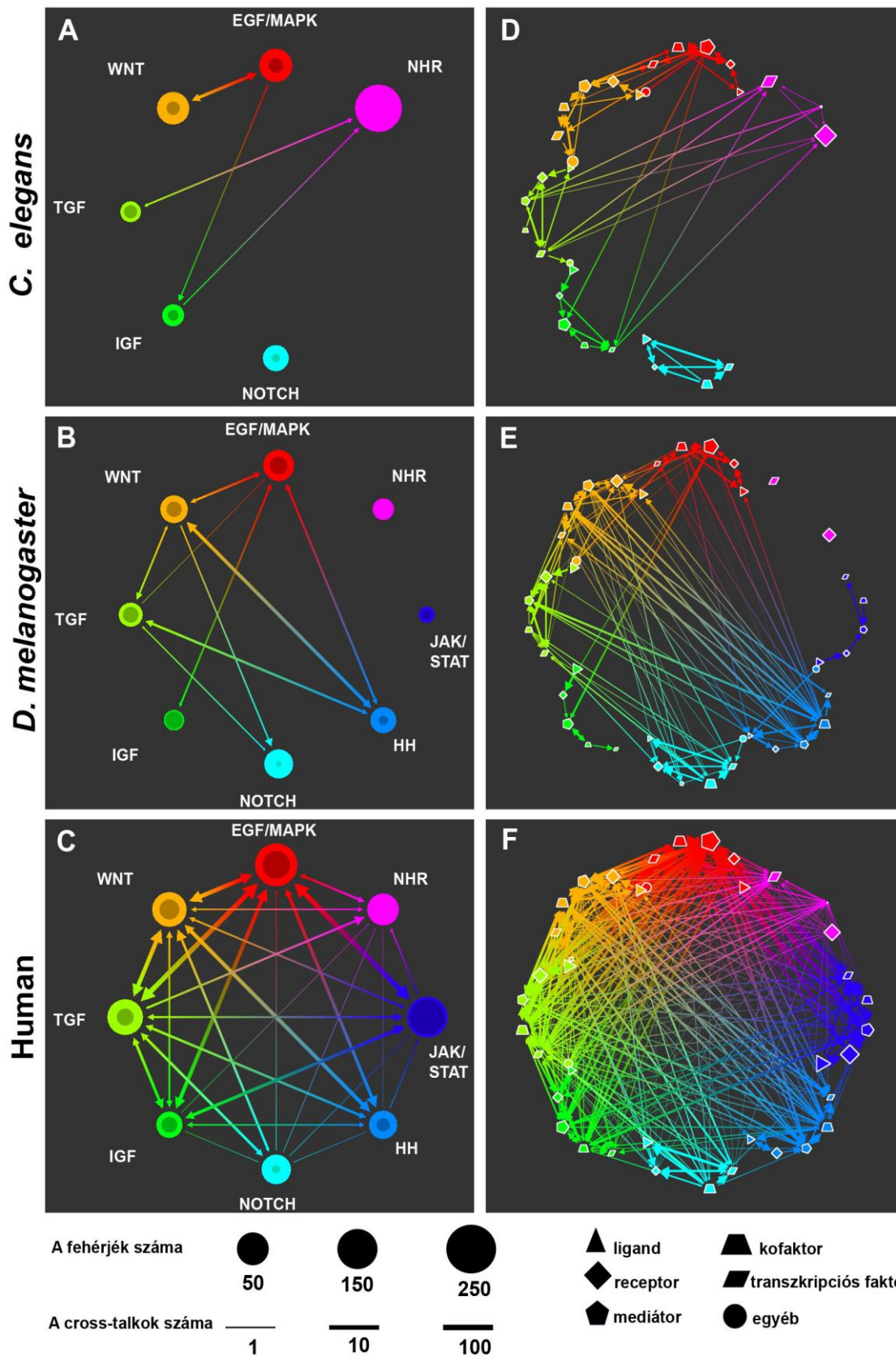
A SignaLink az első olyan nagyméretű jelátviteli útvonal-hálózat, amelyben lehetőség van a keresztbeszélgetések rendszerszintű összehasonlítására. Erre a keresztbeszélgetések nagy száma és az egységes gyűjtés mellett a keresztbeszélgetések irányának feltűntetése is lehetőséget ad. Ebben az alfejezetben a keresztbeszélgetések az egyes fajokon belül, a következő alfejezetben a fajok között kerülnek összehasonlításra.

A részletes összehasonlító elemzések mellett fontosnak éreztem az ún. útvonal-hálózatok ábrázolását is. Az útvonal-hálózatok jellemzője, hogy a hálózat pontjai nem fehérjék, hanem útvonalak vagy jelátviteli pozíció-csoportok. Az ilyen ábrázolási mód segítségével a jelátviteli hálózatot az útvonalak szintjén tudjuk vizsgálni, kiemelve az útvonalak közötti keresztbeszélgetéseket (**22. ábra**).

Cross-talk

útvonak között

jelátviteli topológiai csoportok között



22. ábra (előző oldal) A SignaLink jelátviteli útvonalainak és a közöttük lévő irányított keresztbeszélgetések (cross-talkok) útvonal alapú ábrázolása. A pontok mérete arányos az adott jelátviteli útvonal vagy pozíció csoport méretével, míg a vonalak és nyilak vastagsága az adott irányba tartó keresztbeszélgetések összesített számával (logaritmikus skálán). Az egyes útvonalak központi részei sötétebb színnel, a mellék régiók világosabb színnel vannak jelölve. **a-c)** Az útvonalak közötti keresztbeszélgetés-hálózat. Emberben minden útvonal kapcsolatban van a másikkal. **d-f)** A jelátviteli pozíció-csoportok közötti keresztbeszélgetések hálózata. Emberben szinte minden jelátviteli pozíció csoport részt vesz a keresztbeszélgetésekben, míg a másik két fajban főleg csak a kofaktorok és a mediátorok működnek így. Készült a (Korcsmaros és mtsai, 2010) publikáció nyomán.

A *C. elegans*-ban a nyolcból csak hat útvonal aktív, mivel a JAK/STAT és a Hedgehog útvonal az evolúció során visszafejlődött (Pires-daSilva és Sommer, 2003). A kevesebb számú útvonal mellett ráadásul ritka a keresztbeszélgetés-hálózat, hiszen a lehetséges 30 keresztbeszélgetés típusból csak 5 van jelen (ld. **22. ábra**). Ezek közül a legtöbb keresztbeszélgetés az EGF/MAPK és WNT útvonal között van, amelyek méretüknél fogva is a legtöbb fehérjét tartalmazó útvonalak. Érdekes módon a keresztbeszélgetések összesített iránya egy-egy útvonal között néha különbözik, azaz néhány útvonal kapcsolat aszimmetrikus. Az EGF/MAPK-ból az IGF útvonal felé 1,5 keresztbeszélgetés van, míg az ellenkező irányba egy sem.

A *D. melanogaster*-ben mind a nyolc vizsgált útvonal aktív, de az NHR és a JAK/STAT útvonalak izoláltak, nem rendelkeznek keresztbeszélgetésekkel. Nem számítva az NHR és JAK/STAT útvonalakat, a gyümölcslégy keresztbeszélgetés-hálózata sűrű, a 30 lehetséges keresztbeszélgetés típusból 16 jelen van. A legtöbb keresztbeszélgetés a WNT és Hedgehog útvonalak között van, ráadásul ez pont aszimmetrikus keresztbeszélgetés, mivel a Hedgehog útvonalból a WNT útvonalba 6,9 keresztbeszélgetés, az ellenkező irányban pedig 3,1 keresztbeszélgetés van.

Emberben mind a nyolc útvonal aktív, és mind össze van kötve egymással. Mivel az útvonalak között lévő 28 kapcsolat irányított, összesen 56 különböző keresztbeszélgetés típus lehetséges. Ezen 56 lehetséges kapcsolattípusból mindössze 4 nincs jelen. A legtöbb keresztbeszélgetés az EGF/MAPK útvonalat köti össze a WNT, TGF- β , IGF és JAK/STAT útvonalakkal. Szemben a másik két fajjal, emberben az NHR útvonal minden más útvonallal kapcsolatban van, amelynek fő oka a transzkripció faktorok kapcsolódásai a sejtmagban. Emberben jellegzetes aszimmetrikus keresztbeszélgetés az EGF/MAPK és TGF- β közötti kommunikáció. Ebben az esetben

sokkal több (összesen 160) keresztbeszélgetés megy a TGF- β útvonalból az EGF/MAPK-ba, mint az ellenkező irányba (összesen 104).

Statisztikai módszerrel (Wald-féle megbízhatósági vizsgálat) Böde Csaba biostatistikussal megnéztük, hogy emberben melyik keresztbeszélgetés gyakoribb. 95%-os megbízhatósági szintnél mondható, hogy egy útvonalpár között lévő kapcsolat gyakoribb, mint a többi. A **14. táblázat** mutatja a relatív gyakoriságokat az egyes útvonalpárok között. A lehetséges összes 56 keresztbeszélgetés átlagos relatív gyakorisága 1,79%, és a Wald megbízhatósági tartomány felső határa 7,27%. Azaz az ennél nagyobb relatív gyakoriságú keresztbeszélgetések szignifikánsan gyakoribbak, mint a többi keresztbeszélgetés. Ez alapján az EGF/MAPK és a TGF- β útvonal 3 keresztbeszélgetése kifejezetten gyakoribb, beleértve a két útvonal közötti kétirányú keresztbeszélgetéseket.

14. táblázat Az irányított keresztbeszélgetések gyakoriságának statisztikai vizsgálata. Wald-féle megbízhatósági értékek az útvonalpárok kapcsolódásának vizsgálata során. A táblázatban csak a 10 legnagyobb értéket adó irányított keresztbeszélgetést soroltuk fel. Azok a keresztbeszélgetések, amelyek relatív gyakorisága 7,27% felett van, statisztikailag szignifikánsnak tekinthetők.

Irányított keresztbeszélgetés útvonalpárok között		Kapcsolatok száma	Relatív gyakoriság	
IGF	EGF/MAPK	36	3,10%	
EGF/MAPK	IGF	50	4,31%	
EGF/MAPK	WNT	53	4,57%	
TGF-β	WNT	57	4,91%	
WNT	EGF/MAPK	65	5,60%	
JAK/STAT	EGF/MAPK	83	7,16%	
EGF/MAPK	JAK/STAT	90	7,76%	Statisztikailag szignifikánsan gyakoribb keresztbeszélgetések
WNT	TGF-β	90	7,76%	
TGF-β	EGF/MAPK	104	8,97%	
EGF/MAPK	TGF-β	160	13,79%	

IV.4.2.4. A keresztbeszélgetések összehasonlító elemzése a fajok között és szerepük a jelátviteli hálózat komplexitásában

Miután összehasonlítottuk az egyes fajokban az útvonalak közötti keresztbeszélgetéseket, érdemes megnéznünk ezek hasonlóságait és különbségeit a fajok között is. Ez a vizsgálat azért érdekes és fontos, mert három olyan faj jelátviteli

keresztbeszélgetés hálózatát hasonlítjuk össze, amelyek szervezeti szinten jelentősen különböznek egymástól, de a jelátviteli útvonalaik típusai és az azokat alkotó fehérjék száma igen hasonló. Felmerül tehát a keresztbeszélgetések lehetséges szerepe a biológiai változatosság/morfológiai komplexitás kialakításában.

C. elegans-ban a 8 vizsgált útvonalból 6 aktív, és ezek közül a Notch útvonal nincs kapcsolatban más útvonalakkal (nem rendelkezik keresztbeszélgetésekkel). A féregben az útvonalak között csak kevés (5 darab) keresztbeszélgetést találunk (a lehetséges $6 \times 5 = 30$ -ból). *D. melanogaster*ben mind a 8 útvonal aktív, de az NHR és a JAK/STAT útvonal izolált (nincs keresztbeszélgetésük). A 6 aktív útvonal között a gyümölcslégyben már több mint a lehetséges keresztbeszélgetések fele, 16 keresztbeszélgetés típus van jelen. Ezzel szemben az emberben már mind a 8 útvonal aktív, és az összes lehetséges keresztbeszélgetés is jelen van, azaz minden útvonal kapcsolatban van a másikkal (22/a-c. ábra). Ha megnézzük a jelátviteli pozíció csoportokat (22/d-f. ábra), látható, hogy *C. elegans*-ban keresztbeszélgetés főleg receptorok, mediátorok és transzkripciós faktorok szintjén történik. Gyümölcslégyben és emberben bármelyik topológia csoport között találunk keresztbeszélgetést.

A keresztbeszélgetések nagy és növekvő száma a már meglévő fehérjék hatékony felhasználását is jelzi: az evolúció során meglévő útvonalfehérjék alakultak át úgy, hogy képessé váltak egy másik útvonal fehérjéjével is kapcsolatba lépni (Bhattacharyya és mtsai, 2006). A keresztbeszélgetés típusok növekedése mellett a jelátviteli komplexitás másik fontos indikátora maguknak a keresztbeszélgetések számának az aránya az összes jelátviteli kapcsolathoz képest. Míg a férgek keresztbeszélgetései az összes jelátviteli kapcsolat 4,6%-át alkotják, addig ez a szám a gyümölcslégyben már 10,5%, emberben pedig 30,3%. Azaz az emberben jelenlévő jelátviteli kapcsolatok egyharmada útvonalak közötti kapcsolat. Érdekes módon, a keresztbeszélgetések számának növekedése nincs összefüggésben a fajok összfehérje számával, amely a féregben, gyümölcslégyben és emberben rendre 20.100, 13.800 és 23.000. Szintén nincs korreláció az adott fajjal kapcsolatos cikkek számával sem (3.889, 11.367 és 214.193 cikk a Pubmedben 2010 áprilisában).

IV.4.3. Keresztbeszélgetések expressziós vizsgálata

Láthattuk, hogyan növekszik a keresztbeszélgetések száma és típusa főleg → gyümölcslegy → ember sorrendben, és ez hogyan növeli a jelátvitel kombinációs képességét. Ugyanakkor ez komoly szabályozási kérdéseket is felvethet. Az ismert keresztbeszélgetéseket szabályozó mechanizmusok közül (állványfehérjék, visszacsatolási */feed-back/* hurkok, kinetikai szigetelés, térbeli és időbeli expressziós különbségek) a keresztbeszélgetések szövet- és betegség-specifikus jelenlétét, expresszióját vizsgáltuk meg részletesen.

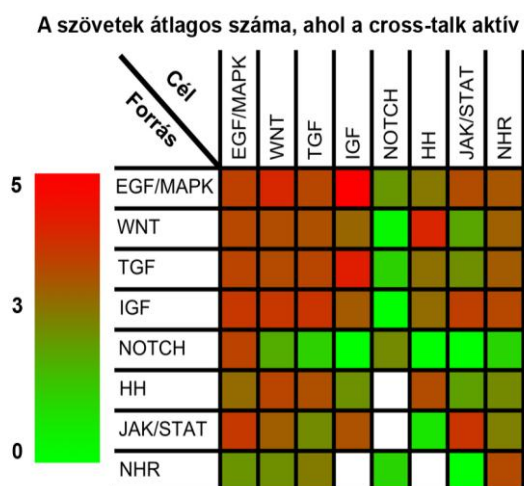
IV.4.3.1. Szövetspecifikus keresztbeszélgetés-expresszió

A jelátviteli útvonalak közötti kapcsolatok, hasonlóan a többi fehérje-fehérje kapcsolathoz nem általánosan vannak jelen minden szövetben. Annak érdekében, hogy a keresztbeszélgetések szövetspecifikus tulajdonságát felderítsük, mRNS expressziós mintázatokat használtunk fel (ld. III.4.2.1. alfejezet). Az itt bemutatásra kerülő eredmények a (Korcsmaros és mtsai, 2010) publikációban kerültek közzé.

Aktívnak tekintetem egy adott szövetben egy kapcsolatot, ha a kapcsolódó fehérjéjárt kódoló mindkét mRNS expresszálódik az adott szövetben. Ez egy közelítés, mivel feltételezhető, hogy azon fehérjék, amelyek mRNS-e expresszálódik, azok jelen vannak. Természetesen ez nem jelenti azt, hogy az adott mRNS biztosan transzlálódik, hiszen a mikroRNS-ek képesek ezt gátolni, vagy a frissen keletkező fehérjét az ubiquitin-proteaszóma rendszer le tudja bontani. Ami biztos, hogy azon fehérjék, amelyek mRNS-e nem expresszálódik, biztosan nem lehetnek jelen, tehát a kapcsolataik nem lehetnek aktívak. A Módszertan részben leírt megfontolások és eljárás alapján 5 szövet (végbél, izomszövet, hámszövet, máj és kardiovaszkuláris) együttes expresszióját vizsgáltam a Signalink adatsoron.

A vizsgálatom során minden egyes útvonalpár esetén megnéztem, hogy az összes Signalinkben szereplő kapcsolatból hány van jelen a vizsgált 5 szövetben, és ez utóbbit átlagoltam. Ezt követően útvonalpáronként összegeztem az egyes fehérje-fehérje kapcsolatra kapott átlagos szöveti jelenléteket. Az eredmény ábrázolása céljából egy irányított kapcsolatokat bemutató kapcsolat-mátrixot készítettünk, ahol az egyes kapcsolatok aktivitását (jelenlétét) egy színskálával jeleztük (**23. ábra**). A kapcsolat-mátrix lehetőséget biztosít arra, hogy azonosítsunk olyan aktív kapcsolatokat is, ahol

különbséget látunk két útvonalpár irányított kapcsolatában (aszimmetrikus keresztbeszélgetések). Azonban jelentős különbséget sehol sem találtam az egyes keresztbeszélgetés típusok között. A mátrix elrendezése miatt a mátrix főátlójában kiegészítő információkat láthatunk, mivel itt nem útvonalak közötti, hanem az egyes útvonalakon belüli kapcsolatok szerepelnek. Érdekes módon a Notch útvonalat leszámítva minden útvonal jelen van a legtöbb vizsgált szövetben, és a keresztbeszélgetéseikben jelentős különbségeket találtam. Ez alapján két csoportra osztottam az útvonalpárokat, amelyek az átlagtól eltérő aktivitással rendelkeztek. Az alábbi nagyobb útvonalak keresztbeszélgetései az átlagnál kevésbé voltak aktívak (**23. ábra**, zöld színnel jelölt kockák): EGF/MAPK-IGF, IGF-JAK/STAT, EGF/MAPK-JAK/STAT, IGF- TGF- β és IGF-WNT. Az EGF/MAPK-IGF keresztbeszélgetést példaként vizsgálva az irodalmi adatok is megerősítik ezen keresztbeszélgetések alulreprezentációját a vizsgált szövetekben, mivel ezekről csak 2 publikációt találtam (Roudabush és mtsai, 2000; Vardy és mtsai, 1995). Három nagyobb és több kisebb útvonal keresztbeszélgetései ezzel szemben az átlagnál gyakrabban aktívak (**23. ábra**, piros színnel jelölt kockák): EGF/MAPK-NHR, NHR-TGF- β és NHR-WNT. Utóbbiak részletesebb vizsgálata rámutatott arra, hogy ennek egyik oka a SMAD3 fehérje, amely több útvonalban is jelen van (többútvonalas fehérje) és aktív keresztbeszélgetése van a legtöbb vizsgált szövetben. A SMAD3 a fibrogenézis folyamatának egyik fontos eleme, így a vizsgált szövetekben gyakran expresszálódik (Flanders, 2004; Roberts és mtsai, 2001).

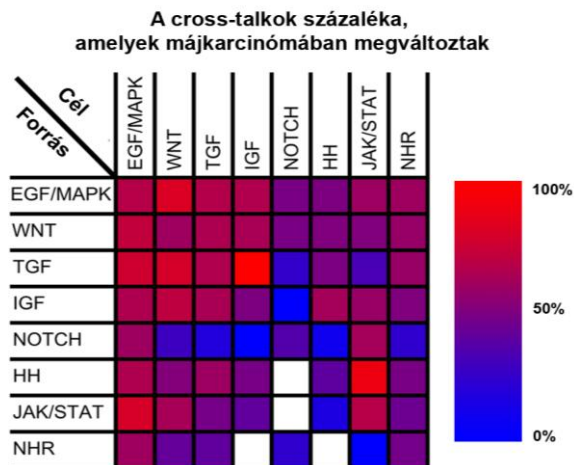


23. ábra A keresztbeszélgetések (cross-talkok) szöveti aktivitásának vizsgálata. Színskála jelöli a szövetek átlagos számát, ahol a keresztbeszélgetés aktív. A piroshoz közeli szín jelzi, hogy az adott keresztbeszélgetés gyakori mind az 5 vizsgált szövetben, míg a zöldhöz közeli szín a ritka, vagy egyáltalán nem jellemző szöveti megjelenést mutatja. A kapcsolat-mátrix elrendezéséből fakadólag a mátrix főátlójában nem keresztbeszélgetések, hanem az egyes útvonalakon belüli kapcsolatok aktivitása látható. A fehérrel jelölt kocka azt jelzi, hogy a Signalink nem tartalmaz keresztbeszélgetést abban az irányban.

IV.4.3.2. Betegség-specifikus keresztbeszélgetés-expresszió

A fehérje-fehérje és a jelátviteli kapcsolatokra nemcsak az jellemző, hogy jelentős részük szövetspecifikus, hanem az is, hogy betegség (pl.: rák), esetén ezek megváltoznak (Stelling és mtsai, 2004). Mivel ma már a rákosodási folyamatot rendszerszintű változásnak és rendszerbetegségnek tekintik (Hornberg és mtsai, 2006), úgy döntöttem, hogy megvizsgálom, hogy az útvonalak kapcsolatai mennyire változnak meg rákban. Ennek érdekében egy gyakran előforduló és homogén szöveti tulajdonságokkal rendelkező ráktípus, a hepatocelluláris karcinóma expressziós mintázatát hasonlítottam az egészséges máj expressziós mintázatához a III.4.2.2. alfejezetben leírtak szerint. Megváltozottak tekintetem egy kapcsolatot, ha az Oncomine adatbázisból letöltött elemzés szerint szignifikánsan különbözött a kapcsolódó fehérjepár valamelyikének az expressziója az egészségestől. Nem különböztettem meg, hogy a fehérjét kódoló mRNS jobban vagy kevésbé expresszáldott, csak azt néztem, hogy megváltozik-e. Ezt követően – hasonlóan a szöveti vizsgálathoz – útvonalpáronként összegeztem az egyes fehérje-fehérje kapcsolatra kapott átlagos változási értékeket. Az eredményt a 23. ábra hasonló kapcsolat-mátrix segítségével ábrázoltuk, ahol az egyes kapcsolatok megváltozását egy százalékos színskálával jeleztem (24. ábra).

24. ábra A keresztbeszélgetések májkarcinóma esetén bekövetkező expresszió-változásának vizsgálata. A színskálán a kapcsolatok százalékos számát jelöltem, ahol a keresztbeszélgetés megváltozott. A piros közeli szín azt mutatja, hogy a keresztbeszélgetések jelentős része megváltozott májkarcinómában, míg a kék közeli szín azt jelzi, hogy ritkán, vagy egyáltalán nem változik meg. A kapcsolat-mátrix elrendezéséből fakadólag a mátrix főátlójában nem keresztbeszélgetések, hanem az egyes útvonalakon belüli kapcsolatok megváltozása látható. A fehérrel jelölt kocka azt jelzi, hogy a SignaLink nem tartalmaz keresztbeszélgetést abban az irányban.



Azt találtam, hogy a 8 vizsgált útvonalból három (a WNT, az NHR és a JAK/STAT) fehérjéinek mindössze a 30%-a esetében változik meg az expresszió májkarcinómában, míg a másik 5 útvonal esetében ez a szám 50% körül volt. Az EGF/MAPK útvonal keresztbeszélgetései jelentősen megváltoztak a májkarcinómában,

hasonlóan további három útvonal (WNT, TGF- β és IGF) keresztbeszélgetéseire. Ezzel szemben a Notch és az NHR útvonalak keresztbeszélgetései nem változtak meg jelentősen májkarcinómában.

IV.4.4. Gyógyszercélpont-jelöltek meghatározása

A jelátviteli fehérjék között felülreprezentáltak az emberi betegséghez köthető gének (Sakharkar és mtsai, 2007), és ezen fehérjék a legintenzívebben kutatott gyógyszercélpontok (Chaudhuri és Chant, 2005). Ráadásul újabban már hálózatos módszerekkel is vizsgálják a gyógyszercélponti tulajdonságaikat (Berger és Iyengar, 2009; Korcsmaros és mtsai, 2007; Spiro és mtsai, 2008). Több tanulmány a következő 4 tulajdonságot azonosította, amelyek leginkább alkalmassá tesznek egy fehérjét arra, hogy gyógyszercélpont legyen: (1) betegséghez köthető gén kódolja, (2) a plazmamembránban helyezkedik el, (3) enzimatis funkcióval és (4) kináz doménnel rendelkezik (Fabbro és mtsai, 2002; Gao és mtsai, 2008; Yildirim és mtsai, 2007). Felhasználva a SignaLink adatbázist és az eddigi eredményeinket elemeztem a SignaLinkben szereplő és jószolt fehérjéket, hogy új gyógyszercélpont jelölteket határozzunk meg. Két fehérjelistán néztük az említett négy gyógyszercélponti tulajdonság jelenlétét az ember esetében: (1) A jószolt, szignológ fehérjék listája (2) A keresztbeszélgetések kialakításában fontos fehérjék listája. Az itt bemutatásra kerülő eredmények zöme a (Korcsmaros és mtsai, 2010) és a (Korcsmaros és mtsai, 2011) publikációkban kerültek közzé.

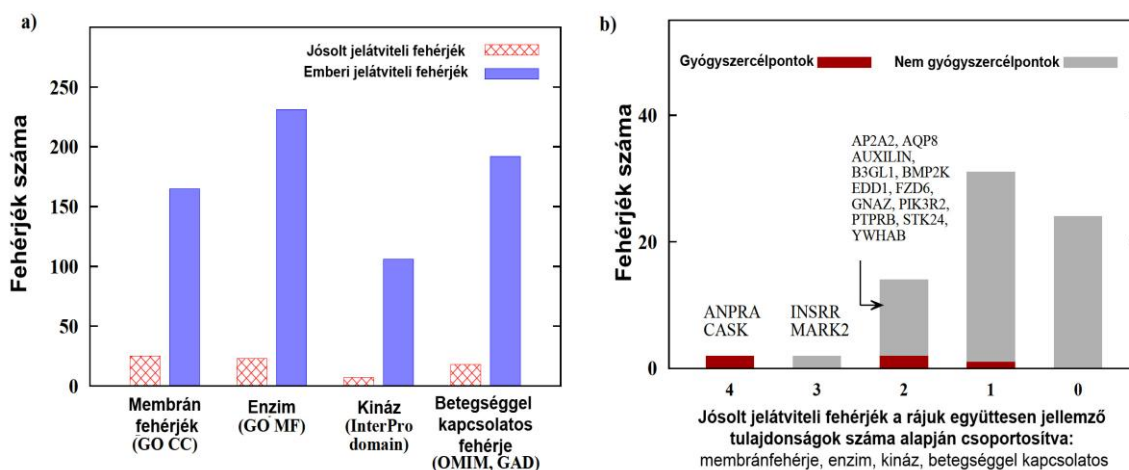
IV.4.4.1. A szignológok elemzése során azonosított gyógyszercélpont-jelöltek

A szignológok, olyan fehérjék, amelyeknek jószoltam a jelátviteli útvonalbeli funkcióját az ortológjuk más fajban betöltött és ismert funkciója alapján. Az emberben 73 ilyen szignológ fehérjét találtam. Ezek közül mindössze öt az, amely már gyógyszercélpont, tehát a fennmaradó 68 fehérje részletesebb vizsgálata rámutathat új gyógyszercélpont-jelöltekre. Sőt, azon öt új jelátviteli fehérje esetében, amely már ismert gyógyszercélpont, egy ilyen vizsgálat rámutathat a lehetséges mellékhatások okaira, illetve a gyógyszer hatásmechanizmus részleteire is.

Az elemzésem során azt néztem, hogy a 73 szignológ fehérjéből hányra igaz a 4 gyógyszercélponti tulajdonság (**25. ábra**). Azonosítottam 2 fehérjét (ANPRA és

CASK), amelyekre mind a 4 tulajdonság jellemző volt. Kiderült, hogy az ANPRA és a CASK már ismert gyógyszercélpont, ami erősíti a 4 tulajdonságon alapuló osztályozási rendszert. Fontos viszont megjegyezni, hogy egyik fehérjét se írták még le jelátviteli fehérjeként. Az ANPRA a kardiovaszkuláris homeosztázisban vesz részt, a CASK egy multidomén állványfehérje, ami a szinapszisok melletti ioncsatornák összeszerelésében vesz részt (Boutet és mtsai, 2007). Ezen felül most az ANPRA-t a WNT, a CASK-ot az EGF/MAPK útvonalhoz tudtam rendelni.

Ezt követően két olyan fehérjét találtam, amelyre 3 tulajdonság volt jellemző a négyből, de egyik sem ismert gyógyszercélpont vagy betegség gén. Ez a két fehérje az INSRR és a MARK2. Előbbit szekvenciahasonlóság alapján már korábban prediktálták az IGF/inzulin útvonalhoz (Hanze és mtsai, 1999), most ezt a predikciót erősítettem meg egy rendszerszintű vizsgálattal. A MARK2-nek az epitélium fejlődésében és a graft kilökődésben van ismert szerepe (Boutet és mtsai, 2007). A szignológ vizsgálattal a MARK2-t két jelátviteli útvonalba is jósoltam, az EGF/MAPK és a WNT útvonalakba. Érdeemes megjegyezni, hogy azok a fehérjék, amelyek egynél több útvonalban is részt vesznek, fontosak lehetnek a gyógyszercélpontok azonosítása során (Hopkins, 2008; Korcsmaros és mtsai, 2007; Korcsmaros és mtsai, 2010; Spiro és mtsai, 2008).



25. ábra A jósolt jelátviteli fehérjék (szignológok) gyógyszercélponti jelentőségének elemzése. a)

Négy, a már ismert gyógyszercélpontokra leginkább jellemző tulajdonság elemzése a jósolt jelátviteli fehérjék között (kontrollként a Signalink adatbázisban szereplő összes emberi jelátviteli fehérje szerepel). Az egyes tulajdonságok forrása zárójelben szerepel. **b)** A 4 tulajdonság együttes jelenléte alapján csoportosított jósolt jelátviteli fehérjék bemutatása. Külön színnel jelöltem a már ismert gyógyszercélpontokat és azokat, amelyek még nem ismertek. Azon fehérjéknek, amelyek a leginkább esélyes új gyógyszercélpontok-jelöltek, illetve amelyek már ismert gyógyszercélpontok, a nevét is feltüntettem az ábrán.

Végül 12 olyan fehérjét találtam, amelyre a négyből két tulajdonság volt jellemző. Ezek közül egyik fehérje sem volt ismert gyógyszercélpont, és három közülük (GNAZ, PTPRB és STK24) membránban található enzim. Ez azért fontos, mert ez a kombináció a leggyakoribb a jelenlegi gyógyszercélpontok között (Gao és mtsai, 2008).

IV.4.4.2. A keresztbeszélgetések elemzése során azonosított gyógyszercélpont-jelöltek

A Signalink adatbázis elemzése során azonosítottam 62 olyan emberi keresztbeszélgetésben résztvevő fehérjét, amelyeknek egynél több útvonalban van szerepük. A 62 többútvonalas fehérje közül 21 (33,8%) már ismert gyógyszercélpont, ami jelentősen nagyobb arány, mint amit az összes jelátviteli fehérje között (16,9%; 89/525) vagy az összes ismert emberi fehérje között (8,2%; 1.610/ 19.534) láthatunk. Következésképp a maradék 41 többútvonalas fehérje releváns, új gyógyszercélpont jelölt lehet. Az elemzésem során ezt a 62 fehérjét kiegészítettem azokkal, amelyeket a rákos keresztbeszélgetés expresszió vizsgálat során (IV.4.3.2. alfejezet) azonosítottunk, mint megváltozott expresszióval rendelkező fehérjét. Összesen 244 ilyen, májkarcinómában megváltozott expressziójú fehérjét találtunk, amelyek közül 28 már szerepelt a 62 többútvonalas fehérje között. Így összesen 278, az útvonalak összekötésében jelentős fehérjét vizsgáltam meg, hogy mennyire alkalmas gyógyszercélpontok.

Hasonlóan a szignológoknál elvégzett vizsgálathoz, azt néztem, hogy a 278 fehérjéből hányra igaz a 4 gyógyszercélponti tulajdonság. Nyolc olyan fehérjét találtunk, amelyre mind a 4 tulajdonság jellemző, és ezek közül 4 (EGFR, ErbB2, IGF1R és INSR) már ismert, mint gyógyszercélpont. Érdekes megjegyezni, hogy ezek mind a patológiában igen ismert receptor-tirozinkinázok, amely alátámasztja az osztályozási rendszerünket. A másik négy keresztbeszélgetésben résztvevő fehérje, amelyre igaz a négy gyógyszercélpontokra jellemző tulajdonság, de még nem gyógyszercélpontok az IRAK1, MAP3K13/LZK/MLK, ROR2 és TGFBR1 fehérjék. Ezek közül a ROR2 esszenciális a csontfejlődésben (Liu és mtsai, 2007), míg a másik három fehérje a gyulladási folyamatokban vesz részt (Klein és mtsai, 2001). Kiemelendő, hogy nemrég publikálták, hogy kísérletes vizsgálatok szerint a ROR2 az oszteosarkóma kezelésnél hatékony célpont lehetne (Morioka és mtsai, 2009). Az

IRAK1, a TGFBR1 és az MLK által érintett útvonalakra több gyulladáscsökkentő gyógyszer (köztük az imiquimod, a dexametazon és az L-arginin) is hat, de ezek aspecifikus hatású kezelések (Klein és mtsai, 2001). Mindezek a már létező megközelítések alátámasztják a javasolt gyógyszercélpont-jelöltek relevanciáját.

V. Az eredmények megvitatása

V.1. A *Signalink* adatbázis

A létrehozott *Signalink* adatbázis egy új és átfogó jelátviteli hálózati adatforrás, amely három többsejtű állat, a fonálféreg *C. elegans*, a gyümölcsleány *D. melanogaster* és az ember útvonalainak kommunikációs térképét tartalmazza. A *Signalink* készítése során egységes és pontosan dokumentált, kézi gyűjtési módszert alkalmaztam. Ennek eredményeként a már létező, hasonló forrásokhoz képest mind minőségileg, mind mennyiségileg jobb adatbázist sikerült létrehozni. Valamint az általunk kidolgozott adatszerkezet lehetővé tette a jelátviteli hálózat rendszerszintű vizsgálatait.

V.1.1. A *Signalink* adatbázis előnyei

A rendszerszintű vizsgálatokhoz megfelelő méretű, igen gyakran használt adatbázisok, az ún. nagy-áteresztőképességű (*high-throughput*) adatforrások. Ezek azonban nem alkalmasak a jelátviteli hálózat átfogó vizsgálatára, mivel technikai okokból, a sejten kívüli térben lévő ligandumokat, a membránhoz kötött receptorokat és a sejtmagban lévő transzkripciós faktorokat csak alacsony számban tartalmazzák. Így a jelátviteli hálózat vizsgálatához egy másik típusú adatbázist, a kézzel készített adatbázisokat érdemes használni. Ezen adatforrásokkal kapcsolatban egy nemrég megjelent elemzés (Cusick és mtsai, 2009) megállapította, hogy: (1) a kiválasztott kísérleti módszerek miatt gyakran torzítottak; (2) ritkán vannak pontosan definiálva a gyűjtési céljaik; (3) nehéz megállapítani a teljességüket és a megbízhatóságukat. Ezeket a kézi gyűjtéssel szembeni jogos kritikákat, és az igényt egy megbízható, nagyméretű jelátviteli adatbázis létrehozására tartottam szem előtt, amikor kialakítottam a *Signalink* adatbázis gyűjtési módszerét.

A *Signalink* adatainak beviteléhez egy nagyon szigorú kézi gyűjtési protokollt dolgoztam ki. Ennek alapját útvonal-szakértők által készített összefoglalók (*review-k*) adják, amelyek segítségével illesztettem be a megbízható, kísérletesen kimért és publikált kapcsolati információkat. Ezáltal a jelenlegi ismereteinkhez képest igen teljes, és megbízható adatbázist kaptunk. A gyűjtésre került útvonalak meghatározásánál

fontos – és más forrásoknál gyakran nem alkalmazott – cél volt, hogy csak biokémiaiailag jól definiált útvonalakat válasszak ki. Ennek a különbségnek egyik példája a SignaLinkben az EGF/MAPK útvonal, amely az EGF ligandum és a terminális MAPK közötti, evolúciós és biokémiai szempontok alapján is indokolt útvonal fehérjéit és azok kapcsolatait tartalmazza. Ugyanakkor több adatbázis ezt az útvonalat sok (al)útvonalra osztja fel (pl.: EGFR, RAS, p38, JNK, ERK, ASK). Ennek a szigorú útvonal meghatározásnak az is a következménye, hogy lecsökkenti az adatbázisban szereplő útvonalak teljes számát. Ugyanakkor úgy gondolom, hogy a megfelelően azonosított és csoportosított útvonalak, elkerülendő a mesterséges csoportosításokat, sokkal fontosabb indikátorai egy adatbázis jóságának, mint a benne szereplő útvonalak száma.

Minden hálózati adatbázisra igaz, hogy az élő szervezetekben jelenlévő kapcsolatoknak csak egy kis részét tartalmazzák, mivel igen sok kapcsolatot, beleértve a jelátviteli fehérjék kapcsolatait is, még nem fedeztek fel (Venkatesan és mtsai, 2009). Ebből következik, hogy az adatbázisok készítői megpróbálnak minél több kapcsolatot összegyűjteni. Azonban a SignaLink készítése során alkalmazott protokollban a megbízhatóság fontosabb szempont volt, mint az adatbázis méret. Ezért úgy gondoljuk, hogy a SignaLink kevés hamis pozitív kapcsolatot tartalmaz.

A folyamatosan növekvő számú jelátviteli adatbázisok (ld. (Bader és mtsai, 2006)) jelentős része esetleges, és kevesebb mint 200 molekulát tartalmaz. Továbbá gyakran limitálva van egy adott útvonalkomponensre, például kinázokra, illetve ingyenesen hozzáférhető akadémiai használatra. Még azon kevés adatbázis között, amelyek megfelelnek ezen kritériumoknak (több mint 200 fehérje, minden útvonalkomponens, ingyenes hozzáférés) sem találunk olyat, amely a SignaLink céljaival és módszertanával összehasonlítható lenne. Azaz nincs „*gold standard*”. Hogy képet kapjunk arról, hogy milyen előnyökkel rendelkezik a SignaLink a többi hasonló útvonal-adatbázishoz képest, a III.2 és IV.2 fejezetekben leírtak szerint összehasonlítottuk a KEGG, Reactome és NetPath adatforrásokkal. Mindezen összehasonlításaink eredményeként megállapítottuk, hogy a SignaLink (1) pontosan definiált útvonalakat tartalmaz; (2) a legtöbb fehérjét és kapcsolatot tartalmazza az összehasonlított útvonalak között; (3) a legtöbb többútvas fehérjét és keresztbeszélgetést tartalmazza; (4) a legnagyobb az átfedése a másik három útvonal-adatbázissal; (5) a legtöbb irodalmi forrást tartalmazza.

Az elmúlt évtizedben megváltozott a jelátviteli útvonalakról alkotott képünk, és ma már ezeket az útvonalakat keresztbeszélgetésekkel sűrűn kapcsolt hálózatnak tekintik. Ezért a SignaLink adatbázis legfontosabb előnye, hogy a keresztbeszélgetéseket és keresztbeszélgetésekben résztvevő fehérjéket a többi adatbázishoz képest magasabb számban tartalmazza, valamint gyűjtési módszere lehetővé teszi az útvonalak egységes, rendszerszintű vizsgálatát. Mindezen tulajdonságoknak köszönhetően a SignaLink adatbázis megfelel a jelenlegi kihívásoknak, és több rendszerszintű megközelítés számára is hatékony forrás tud lenni.

V.1.2 A SignaLink adatbázis hátrányai

Annak ellenére, hogy mindent megtettünk a SignaLink készítése során, hogy egy jó minőségű, megbízható adatforrás legyen, természetesen ennek az adatbázisnak is vannak korlátai. Például a SignaLink nem tartalmazza az összes jelátviteli útvonalat. Még a jól definiálható útvonalak közül is csak nyolc, a többihez képest kiemelten fontosnak ítélt útvonalat tartalmaz. Továbbá nem tartalmazza az összes ismert jelátviteli fehérjét sem. Csak azokat a fehérjéket tartalmazza, amelyek szerepét a kiválasztott nyolc útvonal valamelyikében kísérletesen igazolták, és rendelkeznek irányított kapcsolattal egy másik fehérjéhez. Bár a SignaLink gyűjtése során nagyszámú összefoglaló és kísérletes publikáció használtam fel, az adatbázis mostani verziója 2008 májusában lett lezárva, így az azóta megjelent eredményeket nem tartalmazza. Terveink szerint 2011-től kezdve minden második év júliusában frissítjük az adatbázist. Ezen frissítések során hasonló módszerekkel meghatározott további útvonalakkal bővíthet az adatbázis, valamint megbízható további adatbázisok integrálásával nőhet az adatbázisban lévő fehérjék közötti kapcsolatok száma.

A SignaLinkben szövet-specifikus jellemzők nélkül szerepelnek a fehérjék és a kapcsolatok. Ez egyrészt előny, mert a többi adatbázis általában csak egy-egy szövet adatait tartalmazza. Másrészt hátrány is, mivel olyan kapcsolatokat, csoportosulásokat is megjelenít, amelyek az élő szervezetben nem, vagy csak nagyon kivételes esetben jönnek létre. Annak érdekében, hogy a SignaLink adatait dinamikusan, szövet-, vagy betegségspecifikusan lehessen vizsgálni, szükség van a III.4.2. és az IV.4.3. fejezetekben leírtak szerint specifikus expressziós mintázatok illesztésére. (Ennek felhasználóbarát és könnyen kezelhető megoldásáról az V.3. fejezetben írok.)

Összességében az ismertett korlátok ellenére a SignaLink adatbázis több előnnyel rendelkezik, mint hátránnyal. Emellett az eddigiekben bemutatott tulajdonságai lehetővé teszik olyan elemzések elvégzését, amelyekre a korábbi adatbázisokkal nem, vagy csak nagyon limitált módon volt lehetőség: (1) szövet-, vagy betegség-specifikus keresztbeszélgetések összehasonlítása; (2) több útvonalhoz tartozó (többútvonalas) fehérjék megtalálása; (3) a gyógyszerkutatás számára fontos (kerülendő vagy támadandó) új fehérjék felsorolása.

V.2. A SignaLink adatbázison végzett vizsgálatok

V.2.1. Új génfunkciók azonosítása

Az ún. poszt-genomi korszakban új eszközökre és módszerekre van szükség, hogy a már publikált genomi információkat fel tudjuk használni a sejttani folyamatok megértéséhez. Az egyes fajokban jelenlévő jelátviteli útvonalakról a tudásunk korántsem teljes. Ha különböző fajok ismert jelátviteli rendszereit összehasonlítjuk, akkor a még nem felfedezett, „fehér-foltokról” információkat tudhatunk meg. Ennek megfelelően összevettem két modellorganizmus és az ember jelátviteli fehérjéinek a jelátviteli útvonalakhoz való eddigi hozzárendelését. Ha találtam olyan, kevésbé ismert fehérjét, amelynek a másik fajban lévő ortológjának ismert volt már a jelátviteli szerepe, akkor ez alapján predikáltam a kevésbé ismert fehérje lehetséges útvonaltagságát. Az ilyen fehérjéket szignalogoknak neveztem el. Természetesen az összehasonlításhoz használt fajok és útvonalak számának növelésével ez a módszer tovább bővíthető. Úgy gondoljuk, hogy az általunk jóslott, ún. szignalog fehérjék ígéretes forrásai lehetnek a jelátvitellel kapcsolatos funkcionális hipotéziseknek, és elősegíthetik a kísérletek tervezését és értékelését egyaránt. Az új jelátviteli komponensek azonosítása, nemcsak a jelátvitellel foglalkozó szakterületekre lehet hatással: Ha jobban ismerjük a gének lehetséges funkcióit, jobban meg tudjuk érteni a nem várt mutáns fenotípusokat is. Különösen akkor, ha a mutációt szenvedett szignalog génnek a jóslott útvonalához hasonló mutáns fenotípusa van.

V.2.2. Az útvonalak és a keresztbeszélgetések összehasonlítása

A SignaLink adatbázis gyűjtési módszere és adatszerkezete lehetővé tette a jelátviteli útvonalak rendszerszintű vizsgálatát és összehasonlítását. Mivel mindhárom

vizsgált fajban mind a 8 útvonal azonos gyűjtési szabályok mellett került az adatbázisba, az útvonalak méretét, fehérjéinek típusait össze tudtuk hasonlítani. Az útvonalak összehasonlítása rámutatott az útvonalak központi és mellék régióinak egymáshoz viszonyított arány különbségeire mind a fajokon belül, mind pedig a fajok között. Ezeknél az összehasonlításoknál az egyes útvonalak topológiai pozícióiban megmutatózó (pl. ligandum, receptor) közötti különbségek is jól láthatóak voltak. A bemutatott összehasonlítások leíró jellegűek, de a jelátviteli rendszer evolúciójának és működésének megértése szempontjából alapvetőeknek tekinthetők. Jól megfigyelhető volt, hogy az egyes útvonalak különböző stratégiák szerint alakultak ki, illetve működnek. Például az emberben a WNT, az IGF, a Hedgehog és TGF- β útvonalak központi és mellék régióinak aránya egyforma, míg más útvonalak esetében ez jelentősen különbözik (a Notch útvonalnál a mellék, míg az EGF/MAPK és a JAK/STAT útvonalnál a központi régió aránya volt jelentősen nagyobb). Az NHR útvonal – főleg a SignaLink adatbázis gyűjtési elvei miatt – szinte csak központi elemeket tartalmazott. Ezen különbségekre a jelátviteli topológiai pozíciók eloszlása ad magyarázatot. A jelátviteli topológiai pozíciók eloszlásának különbségei mind a fajokon belül, mind a fajok között megfigyelhetőek voltak. A WNT vagy a JAK/STAT útvonal esetében egy-egy ligandum vagy receptorcsalád nagyszámú jelenléte többfajta bemeneti (input) lehetőséget teremt. Az EGF/MAPK útvonal esetében a főleg a központi régióban lévő mediátorok nagy száma volt megfigyelhető, feltehetően a három szintből álló MAPK-kaszkádot alkotó fehérje-családok jelenléte miatt. A Notch útvonalban a már említett mellék régió főleg a jelterjedésben közvetlen részt nem vevő, de azt szabályozó, kofaktor fehérjék miatt nagyméretű.

A SignaLink adatbázis készítésének egyik alapvető célja volt a keresztbeszélgetések rendszerszintű vizsgálatának lehetővé tétele. Az útvonalalapú jelátviteli hálózatmegjelenítés kiemelte a keresztbeszélgetések jelenlétét az útvonalak között. Ebben a fajta ábrázolásban az egyes keresztbeszélgetés típusok minőségi és mennyiségi különbségeit is meg tudtuk figyelni. Mivel a SignaLink minden fehérje-fehérje kapcsolata irányított, a kapcsolatokból összevont útvonalak közötti keresztbeszélgetések is irányítottak. Így az útvonalalapú hálózatmegjelenítésnél az egyes útvonalpárok közötti keresztbeszélgetések irányának szimmetrikussága is láthatóvá vált. A fajok közötti összehasonlítás kimutatott 2 olyan keresztbeszélgetést

(EGF/MAPK-WNT és EGF/MAPK-IGF), amely mindhárom fajban jelen volt, és jelentős mennyiségű kapcsolatot tartalmazott. Sikertelenül aszimmetrikus keresztbeszélgetéseket is azonosítani, például az emberi EGF/MAPK és TGF- β útvonalpár közötti kapcsolat főleg az EGF/MAPK útvonal irányába tartalmaz több irányított kapcsolatot. A fajok közötti, rendszerszintű összehasonlítás során csak az emberben találtuk azt, hogy minden útvonal kapcsolatban van a másikkal.

Mivel a SignalLink eredeti adatait szövetspecifikusság nélkül gyűjtöttük, felmerült, hogy az egyes szövetekben lévő expresszióknak milyen hatása van a keresztbeszélgetések megjelenésére. Öt egészséges és egy rákos szövet expressziójának felhasználásával vizsgáltuk meg a keresztbeszélgetések szövetspecifikus jelenlétét. Sikertelenül három különböző keresztbeszélgetés mintázattípust azonosítani az útvonalak között:

- Az EGF/MAPK útvonal kapcsolatai a legtöbb vizsgált szövetben expresszálódnak, és ezek a kapcsolatok a májkarcinómában megváltoznak.
- A JAK/STAT útvonal kapcsolatai szintén gyakran expresszálódnak a szövetekben, de csak kissé változnak meg májkarcinómában.
- A Notch útvonal kapcsolatai ritkán vannak jelen a szövetekben, és nem is nagyon változnak meg májkarcinómában.

Irodalmi adatok alapján megvizsgáltuk, hogy ezen jellemzők kialakulásának milyen okai lehetnek. Azt találtuk, hogy az EGF/MAPK keresztbeszélgetéseinek erőteljes jelenléte annak a következménye, hogy ez az útvonal, főleg a MAPK kaszkádok miatt, központi jelentőségű a jelátviteli hálózatokban (Manning és mtsai, 2002). Bizonyított az is, hogy hibás működése több ráktípus kialakulásához is vezethet (Amit és mtsai, 2007). A JAK/STAT útvonalhoz tartozó receptorok (pl.: citokin receptorok) az immunsejtekre jellemző jelátviteli útvonalakban jelentősek. Ezen receptorok jelentős része több szövetben is jelen van (Haan és mtsai, 2006). A Notch útvonalnak főleg a fejlődő szövetekben van szerepe, ami csökken a már differenciált szöveteknél. Valószínűleg emiatt találtuk úgy, hogy a Notch útvonal kevésbé van jelen a vizsgált szövetekben (Bolos és mtsai, 2007). Ugyanakkor a Notch útvonalnak fontos szerepe lehet a májrák kialakulásában: Egy nemrég megjelent tanulmány rámutatott arra, hogy a Notch útvonal képes gátolni a WNT onkogén útvonalat, azonban májrákban

a Notch útvonal fehérjéinek az expressziója csökkent mértékű (Wang és mtsai, 2009). Tehát az irodalmi adatok megerősítik az általunk azonosított három útvonaljellemzőt.

V.2.3. Új gyógyszercélpont-jelöltek azonosítása

A hálózat-elemzések a rendszerszintű adatforrások felhasználásával jelentősen hozzájárulhatnak új gyógyszercélpontok felfedezéséhez (Farkas és mtsai, 2011; Hopkins, 2008; Korcsmaros és mtsai, 2007; Korcsmaros és mtsai, 2010; Spiro és mtsai, 2008). A SignaLink segítségével lehetőségem volt gyógyszercélpont-jelölteket azonosítani a jelátviteli hálózatban. Ehhez két különböző elemzés eredményeként létrejött listát alkalmaztam: (1) a jósvolt jelátviteli fehérjék (szignológok) listáját; (2) a keresztbeszélgetés elemzésben kapott ún. többútvonalas fehérjék és rákban megváltozó keresztbeszélgetésekben résztvevő fehérjék listáját. Ezek közül néhány fehérje megfelelően specifikus és alkalmazható célpontként, de néhány túl központi vagy aspecifikus.

A szignológok elemzésével rámutattam arra, hogy az eddig nem ismert jelátviteli funkcióval rendelkező fehérjék között sok lehetséges új gyógyszercélpont van. A tulajdonságaik vizsgálatával sikerült 14 szignológ fehérjére leszűkítenünk az ígéretes jelöltek listáját, amelyek új terápiás célpontok (útvonalak új blokkolási helyei) is lehetnek. Továbbá sikerült öt, már jelenleg is alkalmazott gyógyszercélpontnak útvonal-annotációt rendelnem, amivel mellékhatások és egyéb gyógyászati tulajdonságok elemzése vált lehetővé.

A többútvonalas fehérjék és a rákban megváltozó keresztbeszélgetésekben résztvevő fehérjék elemzése során 4 ígéretes célpont-jelöltet azonosítottam. Közülük egyről, a ROR2-ről a vizsgálataimmal párhuzamosan jelent meg egy közlemény, ahol kemoterápiás célpontnak javasolják. A másik 3 gyógyszercélpont-jelölt pedig már eddig is nem specifikus célpontja volt több, gyulladás elleni hatóanyagnak. Ugyanezen elemzés során további 35, kevésbé ígéretes, de speciális kezelés esetén hatékony célpontot is találtam. Fontos megjegyezni, hogy a többútvonalas fehérjék célpontként való használata – definíciójukból fakadólag – aspecifikus. Következésképp a többútvonalas fehérjék listája nemcsak ígéretes célpontokat, hanem bizonyos esetekben kerülendőeket is tartalmaz. Ugyanakkor rendszerszintű károsító hatást (pl.:

kemoterápiás kezelés során) ezen fehérjék gátlásával könnyebben lehet elérni, mint a sima útvonal fehérjék célzása esetén.

Egy-egy útvonal közötti összekötetés gyakran csak finomhangolja az egész hálózatot, vagy olyan új funkciót hoz létre, amely például a rákossejtet ellenállóbbá teszi (Kitano, 2007). Az ilyen keresztbeszélgetésben résztvevő fehérjék gátlása (ha az expressziójuk rákban emelkedett), vagy serkentése (ha az expressziójuk rákban csökkent) elősegítheti az adott rákos sejt elpusztítását. Fontos azt is megjegyezni, hogy néhány keresztbeszélgetésben résztvevő fehérje részleges blokkolása nagyobb hatással lehet az egész sejt jelátviteli rendszerére, mint egy útvonal központi elemének teljes blokkolása (Csermely és mtsai, 2005; Korcsmaros és mtsai, 2007).

A gyógyszercélpontok megfelelő meghatározásához a jelátviteli hálózat minél teljesebb körű ismerete szükséges. Ezáltal elkerülhetőek a visszacsatolási-hurkokban résztvevő fehérjék célzásai, amelyek ellenkező hatást érnek el (Sergina és mtsai, 2007). Ezért a SignaLink és a hozzá hasonló adatbázisok illetve elemzések sok kapcsolattal és jelátviteli komponenssel szolgálhatnak a modern gyógyszercélpont-kereséshez.

V.3. A SignaLink adatbázison végezhető további vizsgálatok

A SignaLink adatbázison végzett elemzések rámutattak arra, hogy egy rendszerszintű szempontok alapján, egységes gyűjtési szabályokkal felépített útvonal-adatbázissal olyan kérdésekre tudunk választ találni, amelyet az eddigi forrásokkal nem lehetett volna megválaszolni. Az elemzésekben érintett témakörök mellett a SignaLink adatait további megközelítések során is fel lehet használni.

A jelátviteli rendszer matematikai modellezése során fontos, hogy minél megbízhatóbb és részletesebb hálózat álljon a rendelkezésünkre (Papin és mtsai, 2005). A SignaLink adatbázis könnyen kezelhető, egységes szerkezete, amely statikus adatokat tartalmaz könnyen integrálható sztöchiometriai és expressziós adatokkal. A reakcióállandók ismeretében pedig az egyes útvonalak dinamikája válhat modellezhetővé numerikus vagy differenciálegyenlet alapú megközelítésekkel. Mivel a SignaLink irányított kapcsolatokat tartalmaz, sokkal alkalmasabb forrás a perturbáció elemzések elvégzésére, mint az irányítatlan kapcsolatokat tartalmazó nagy áteresztő képességű (*high-throughput*) adatokat tartalmazó adatbázisok.

A szintetikus biológia (*synthetic biology*) célja a sejtek mesterséges megváltoztatása, új feladatok létrehozása a metabolikus vagy a jelátviteli rendszer mesterséges áthuzalozása révén (Bhattacharyya és mtsai, 2006; Friedland és mtsai, 2009). A jelátviteli rendszeren ezeket a módosításokat állványfehérjék vagy visszacsatolási-hurkokban lévő fehérjék megváltoztatásával lehet elérni. Ehhez pedig szükség van a jelátviteli rendszer felépítésének minél részletesebb ismeretére, különös tekintettel a keresztbeszélgetésekben résztvevő fehérjékre.

Az összehasonlító evolúciós vizsgálatok az egyedfejlődés és a testfelépítés különbségeit illetve konzerváltságát vizsgálja. Mivel ezek kialakításában a regulációs és egyéb kapcsolatok fontosak (Levine és Tjian, 2003; Szathmary és mtsai, 2001), és a jelátviteli rendszer kiemelt szereppel bír, szükség van olyan forrásra, amely megfelelő mennyiségű és minőségű jelátviteli kapcsolatot tartalmaz. A SignaLinkben egységesen gyűjtött információk vannak három, testfelépítésében különböző, de jelátviteli komponenseinek számában és típusában igen hasonló élőlényből. Ennek köszönhetően a SignaLink alkalmas lehet evolúciós hipotézisek vizsgálatára akár az egyes útvonalak evolúciója során bekövetkezett változásokkal, akár rendszerszintű kérdésekkel, például a keresztbeszélgetések és az ún. kritikus pontok (Taniguchi és mtsai, 2006) evolúciós szerepével kapcsolatban.

Az elméleti kutatások mellett a SignaLink adatbázis a kísérlettervezések és kiértékelések során is hasznos forrás lehet. Az elmúlt évben Farkas Illéssel együttműködésben kifejlesztettem egy ingyenesen elérhető alkalmazást, a PathwayLinkert, amely – többek között – a SignaLink adatain alapul, és a kísérlettervezések és kiértékelésekben segíti a kutatókat. A PathwayLinker (<http://PathwayLinker.org>) felhasználva a sejt ismert fehérje-fehérje hálózatát, a beadott fehérjékről képes megmondani és ábrázolni, hogy milyen távol vannak egy-egy jelátviteli útvonaltól. Ez azért nagyon fontos, mert a kísérletek során gyakran úgy módosítanak 1-1 fehérjét, hogy annak nem tervezett globális vagy meglepő hatása van a rendszerre, amely megnehezíti a kísérletek lebonyolítását vagy értékelését. Az is kevésbé ismert, de gyakori probléma, hogy a kísérlet elvégzéséhez egy olyan anyagot vagy mutáns hátteret alkalmaznak a kutatók, amely a kísérlet elvégezhetősége szempontjából fontos, de a detektált eredményt nem megfelelően torzítja (Farkas és mtsai, *beadv.*).

Egy másik alkalmazás a Fazekas Dáviddal közösen kifejlesztett ExViz program, amely a Signalink keresztbeszélgetés adatait használja a keresztbeszélgetések expressziójának vizualizációjá. (A béta verzió ingyenesen elérhető a <http://netbiol.elte.hu/exviz> weboldalról.) Az ExViz program a Signalink elemzése során már bemutatott keresztbeszélgetés expressziós kapcsolatmátrix kezelőfelülete. Segítségével a legtöbb emberi vagy egér expressziós adatok betölthetőek. Az ExViz színskálán jelöli az útvonal-kapcsolatok szövet-, vagy betegség-specifikus jelenlétét vagy differenciál expresszióját a Signalink hálózatában. A program több lehetőséget is ad a felhasználónak, hogy a vizsgálati beállításokat módosíthassa, így mindenki a saját céljai szerint tudja alakítani. Az ExViz a Signalink hálózati adatait felhasználva hálózatképet is készít, és az expressziós értékek szerint színezi a pontokat.

Az itt bemutatott lehetőségek mellett további megközelítések és alkalmazások is profitálhatnak a Signalink adatbázis adatainak, elemzésének és ábrázolási megoldásainak felhasználásával. Bízom benne, hogy a jövőben mind az alap, mind az alkalmazott kutatások során egy hasznos forrás és eszköz lesz!

V. Összefoglalás

A különböző tudományterületek integrálása, és az eredmények hálózatos módszerekkel történő vizsgálata a jelátviteli útvonalak kutatásának egy új irányzata. A bemutatott munkában öt terület, a biokémia, a genetika, a sejtbiológia, a bioinformatika és rendszerbiológia eredményeit és módszereit integráltam, hogy három többsejtű élőlény (a *Caenorhabditis elegans*, a *Drosophila melanogaster* és az ember) jelátviteli útvonalainak hálózatát felépítsem egy erre a célra készített programkörnyezetben. A létrehozott jelátviteli adatbázis a felhasznált területek számára új eredményekkel tud szolgálni. Ezen eredményeket az adott diszciplína a hagyományos eszköztárával nehezen tudta volna elérni, mivel a jelátvitel rendszerszintű vizsgálatához szükséges az integratív megközelítés.

A korábbi nézőpontokkal ellentétben a jelátviteli útvonalakat ma már nem önálló, lineáris útvonalaknak tekintik, hanem keresztbeszélgetések által erősen keresztkapcsolt hálózatnak. Ez a paradigmaváltás szükségessé tette új, rendszerszintű megközelítések (gyűjtési és elemzési technikák) kidolgozását. Annak érdekében, hogy ennek megfeleljünk, létrehoztam egy jelátviteli útvonal adatbázist, amelyet Signalinknek neveztem el. A Signalink három többsejtű élőlény, a fonálféreg *C. elegans*, a gyümölcsleány *D. melanogaster* és az ember 8, egyedfejlődésben és a felnőtt életben is fontos jelátviteli útvonalait tartalmazza, egységes gyűjtési szabályok alapján. Az elkészített adatbázist összehasonlítottuk a már meglévő útvonal-adatbázisokkal, és azt találtuk, hogy több minőségi és mennyiségi szempont szerint is jobb, sőt, hiánypótló. Kidolgoztunk több új megjelenítési lehetőséget a jelátviteli hálózatok ábrázolására. A <http://Signalink.org> honlapon interaktív, felhasználóbarát módon ezek az ábrázolási módok megtalálhatóak.

A Signalink egységes gyűjtésének és adatszerkezetének köszönhetően lehetőségünk nyílt az eredeti céljainkban megfogalmazott rendszerszintű vizsgálatokra. Ennek eredményeként sikerült azonosítanom 253 új jelátviteli komponens (szignológot), és igazolnom ezek újdonságát. A Signalink adatait felhasználva összehasonlítottam a jelátviteli útvonalakat és az azok között lévő keresztbeszélgetéseket minőségi és mennyiségi szempontból. Miután azt találtam, hogy emberben mind a nyolc vizsgált útvonal egymással kapcsolatban van, megvizsgáltam a

keresztbeszélgetések szövet-specifikus jelenlétét is. Egészséges szövetekből és hepatocelluláris karcinómából származó minták felhasználásával bemutattam, hogy a SignaLink statikus adatait expressziós mintázatokkal integrálva lehet specifikus vizsgálatokat végezni. Ezen vizsgálat során három útvonal-expresszió típust találtam. Megvizsgálva a szignológok és a keresztbeszélgetésekben fontos fehérjék tulajdonságait, sikerült két olyan listát (*short list*) készíteni, amely az ígéretes, új gyógyszer-célpont-jelölteket tartalmazza.

Végezetül megállapítható, hogy egy rendszerszintű vizsgálatokra alkalmas, keresztbeszélgetéseket és többútvonalas fehérjéket, azok irányított kapcsolatait tartalmazó adatbázis jött létre. A létrejött adatbázis alkalmas arra, hogy elméleti és alkalmazott kutatásokat végezzünk rajta, hiszen a jelátvitel evolúciójától, modellezésétől, a jelátvitel felépítésén át, a betegségek hatásainak és gyógyszerek lehetséges új célpontjainak azonosítása is lehetséges a SignaLink alkalmazásával. Ezen megközelítések egy részét a disszertáció „Eredmények” részében mutattam be, és a jövőben elvégezhető vizsgálatok közül néhányat a „Az eredmények megvitatása” részben soroltam fel.

A doktori munkám során kapott új tudományos eredményeim a következők:

1. Egységes gyűjtési szabályrendszer alapján létrehoztam a SignaLink adatbázist, hogy megalapozzam a jelátvitel rendszerszintű és összehasonlítható elemzését.
2. Összeállítottam három többsejtű faj, a *C. elegans*, a *D. melanogaster* és a *H. sapiens* 8 jelátviteli útvonalának hálózatát, és 253 új jelátviteli komponens (szignológot) azonosítottam.
3. Megvizsgálva a szignológok és a keresztbeszélgetésekben fontos fehérjék tulajdonságait, sikerült kiemelnem 30 ígéretes, új gyógyszer-célpont-jelöltet.

A gyűjtési szabályrendszert témavezetőimmel és Farkas Illéssel közösen alakítottuk ki. Az adatbázisban lévő adatokat én gyűjtöttem, az elemzéseket és az ábrázolásokat Farkas Illéssel együtt terveztük, és Fazekas Dávid TDK hallgatómmal közösen végeztük el. A szignológok azonosítását és az összehasonlításokat egyedül végeztem. A gyógyszer-célponti elemzéseket Farkas Illéssel közösen végeztük.

VI. Summary

Integrating existing disciplines and applying network analysis is a novel trend for signal transduction research. In the presented work, I integrated the methods and results of 5 disciplines, biochemistry, genetics, cell biology, bioinformatics, and systems biology to create a signaling pathway network in a designated program environment for 3 metazoans (*Caenorhabditis elegans*, *Drosophila melanogaster*, and human). I performed manual curation with uniform rules, using bioinformatical sources and scientific publications. The developed network database could give new results and possibilities for the mentioned 5 disciplines. These possibilities cannot be achieved by the classical methods of these disciplines as they require integrative approach and systems level examination of the signaling system.

In contrast to earlier views, intracellular signaling is now viewed as a set of intertwined pathways forming a single signaling network. This paradigm shift calls for novel experimental, curation, and network modeling techniques. Accordingly, I developed a signaling pathway database, called Signalink. Signalink contains the signaling pathways of 3 metazoans, the nematode *C. elegans*, the fruit fly *D. melanogaster* and humans. These pathways are important in both development and adulthood, and they were curated based on uniform curation rules. We compared Signalink with 3 existing pathway databases. Signalink was found to be better both in quantitative and qualitative properties. We created several novel methods to visualize signaling networks. All these visualized networks can be interactively accessed in a user-friendly website at <http://Signalink.org>.

Thanks to the uniform curation rules and data structure of Signalink, we had the opportunity to perform system-level examination on the signaling network. I identified 253 novel signaling proteins, called signalogs, and verified their novelty. Based on the pathway data of Signalink, I compared the pathways and their cross-talks, and found that only in humans every pathway can cross-talk. I found that in humans, cross-talk expression is tissue and disease-specific, which underscores its importance in development and medicine. I could identify 3 cross-talk expression types. Examination of signalogs, multi-pathway proteins, and proteins important in cross-talk of cancer cells, allowed us to create a short list of possible novel drug targets.

Finally, I conclude that the newly developed database is capable for system-level analyses, as well as to examine cross-talks and multi-pathway proteins. The database can be further used for theoretical and applied research (e.g., evolution of signaling networks, modeling of drug effects). Some approaches have been shown in the result chapter of the Thesis, while I listed some other possibilities in the Discussion chapter.

The major scientific results of my PhD work are the following:

1. Based on uniform curation rules, I developed the Signalink database to allow system-level and comparative analyses on the signaling network.
2. I compared the signaling pathway networks of 3 metazoan species, *C. elegans*, *D. melanogaster*, and *H. sapiens*, and identified 253 novel signaling components (signalogs).
3. With the property analysis on signalogs and cross-talking proteins, I could highlight 30 promising novel drug target candidates.

The curation protocol was created with my mentors and Illés Farkas. The curation was done by me, the analyses and visualization was plant in partnership with Illés Farkas, and was done with Dávid Fazekas. I alone identified the signalogs an compared the networks. The drug target relevance analyses were done by me and Illés Farkas.

Irodalomjegyzék

1. Adamcsek B, Palla G, Farkas IJ, Derenyi I, Vicsek T (2006) CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* **22**: 1021–1023
2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402
3. Amberger J, Bocchini CA, Scott AF, Hamosh A (2009) McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* **37**: D793–D796
4. Amit I, Wides R, Yarden Y (2007) Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. *Mol Syst Biol* **3**: 151
5. Araujo RP és Liotta LA (2006) A control theoretic paradigm for cell signaling networks: a simple complexity for a sensitive robustness. *Curr Opin Chem Biol* **10**: 81–87
6. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29
7. Bader GD, Cary MP, Sander C (2006) Pathguide: a pathway resource list. *Nucleic Acids Res* **34**: D504–D506
8. Balogh G, Peter M, Liebisch G, Horvath I, Torok Z, Nagy E, Maslyanko A, Benko S, Schmitz G, Harwood JL, Vigh L (2010) Lipidomics reveals membrane lipid remodelling and release of potential lipid mediators during early stress responses in a murine melanoma cell line. *Biochim Biophys Acta* **1801**: 1036–1047
9. Bandyopadhyay S, Sharan R, Ideker T (2006) Systematic identification of functional orthologs based on protein network comparison. *Genome Res* **16**: 428–435
10. Barabasi AL és Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**: 101–113
11. Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, Shinjo F, Liu Y, Dembowy J, Taylor IW, Luga V, Przulj N, Robinson M, Suzuki H, Hayashizaki Y, Jurisica I, Wrana JL (2005) High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* **307**: 1621–1625

12. Bauer-Mehren A, Furlong LI, Sanz F (2009) Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol* **5**: 290
13. Becker KG, Barnes KC, Bright TJ, Wang SA (2004) The genetic association database. *Nat Genet* **36**: 431–432
14. Behar M, Dohlman HG, Elston TC (2007) Kinetic insulation as an effective mechanism for achieving pathway specificity in intracellular signaling networks. *Proc Natl Acad Sci U S A* **104**: 16146–16151
15. Berger SI és Iyengar R (2009) Network analyses in systems pharmacology. *Bioinformatics* **25**: 2466–2472
16. Berglund AC, Sjolund E, Ostlund G, Sonnhammer EL (2008) InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res* **36**: D263–D266
17. Berriz GF és Roth FP (2008) The Synergizer service for translating gene, protein and other biological identifiers. *Bioinformatics* **24**: 2272–2273
18. Beyer A, Bandyopadhyay S, Ideker T (2007) Integrating physical and genetic maps: from genomes to interaction networks. *Nat Rev Genet* **8**: 699–710
19. Bhalla US és Iyengar R (1999) Emergent properties of networks of biological signaling pathways. *Science* **283**: 381–387
20. Bhattacharyya RP, Remenyi A, Yeh BJ, Lim WA (2006) Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu Rev Biochem* **75**: 655–680
21. Blank U, Karlsson G, Karlsson S (2008) Signaling pathways governing stem-cell fate. *Blood* **111**: 492–503
22. Blencowe BJ, Ahmad S, Lee LJ (2009) Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. *Genes Dev* **23**: 1379–1386
23. Bolos V, Grego-Bessa J, de la Pompa JL (2007) Notch signaling in development and cancer. *Endocr Rev* **28**: 339–363
24. Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, Kaimachnikov NP, Timmer J, Hoek JB, Kholodenko BN (2009) Systems-level interactions between insulin-EGF networks amplify mitogenic signaling. *Mol Syst Biol* **5**: 256
25. Bornholdt S (2005) Systems biology. Less is more in modeling large genetic networks. *Science* **310**: 449–451

26. Boswell BA, Lein PJ, Musil LS (2008) Cross-talk between fibroblast growth factor and bone morphogenetic proteins regulates gap junction-mediated intercellular communication in lens cells. *Mol Biol Cell* **19**: 2631–2641
27. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A (2007) UniProtKB/Swiss-Prot: The Manually Annotated Section of the UniProt KnowledgeBase. *Methods Mol Biol* **406**: 89–112
28. Bray D (2003) Molecular networks: the top-down view. *Science* **301**: 1864–1865
29. Brown KR és Jurisica I (2005) Online predicted human interaction database. *Bioinformatics* **21**: 2076–2082
30. Bundy JG, Papp B, Harmston R, Browne RA, Clayson EM, Burton N, Reece RJ, Oliver SG, Brindle KM (2007) Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling. *Genome Res* **17**: 510–519
31. Campagne F, Neves S, Chang CW, Skrabanek L, Ram PT, Iyengar R, Weinstein H (2004) Quantitative information management for the biochemical computation of cellular networks. *Sci STKE* **2004**: 111
32. Chaudhuri A és Chant J (2005) Protein-interaction mapping in search of effective drug targets. *Bioessays* **27**: 958–969
33. Chen H és Sharp BM (2004) Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinformatics* **5**: 147
34. Chen X, Cheung ST, So S, Fan ST, Barry C, Higgins J, Lai KM, Ji J, Dudoit S, Ng IO, Van De RM, Botstein D, Brown PO (2002) Gene expression patterns in human liver cancers. *Mol Biol Cell* **13**: 1929–1939
35. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2**: 2366–2382
36. Cote RG, Jones P, Martens L, Kerrien S, Reisinger F, Lin Q, Leinonen R, Apweiler R, Hermjakob H (2007) The Protein Identifier Cross-Referencing (PICR) service: reconciling protein identifiers across multiple source databases. *BMC Bioinformatics* **8**: 401
37. Csermely P (2006) Weak links: Stabilizers of Complex Systems from Proteins to Social Networks. pp 392

38. Csermely P, Agoston V, Pongor S (2005) The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol Sci* **26**: 178–182
39. Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O'Connor-McCourt M, Purisima EO, Wang E (2007) A map of human cancer signaling. *Mol Syst Biol* **3**: 152
40. Cui Q, Yu Z, Purisima EO, Wang E (2006) Principles of microRNA regulation of a human cellular signaling network. *Mol Syst Biol* **2**: 46
41. Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, Simonis N, Rual JF, Borick H, Braun P, Dreze M, Vandenhoute J, Galli M, Yazaki J, Hill DE, Ecker JR, Roth FP, Vidal M (2009) Literature-curated protein interaction datasets. *Nat Methods* **6**: 39–46
42. Cutler P (2003) Protein arrays: the current state-of-the-art. *Proteomics* **3**: 3–18
43. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**: 3
44. Diao Y, Li M, Feng Z, Yin J, Pan Y (2007) The community structure of human cellular signaling network. *J Theor Biol* **247**: 608–615
45. Dreesen O és Brivanlou AH (2007) Signaling pathways in cancer and embryonic stem cells. *Stem Cell Rev* **3**: 7–17
46. Dumont JE, Pecasse F, Maenhaut C (2001) Crosstalk and specificity in signalling. Are we crosstalking ourselves into general confusion? *Cell Signal* **13**: 457–463
47. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y, Vasilescu J, Abu-Farha M, Lambert JP, Duewel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* **3**: 89
48. Fabbro D, Ruetz S, Buchdunger E, Cowan-Jacob SW, Fendrich G, Liebetanz J, Mestan J, O'Reilly T, Traxler P, Chaudhuri B, Fretz H, Zimmermann J, Meyer T, Caravatti G, Furet P, Manley PW (2002) Protein kinases as targets for anticancer agents: from inhibitors to useful drugs. *Pharmacol Ther* **93**: 79–98
49. Farkas IJ, Korcsmaros T, Kovacs IA, Mihalik A, Palotai R, Simko GI, Szalay KZ, Szalay-Beko M, Vellai T, Wang S, Csermely P (2011) Network-based tools for the identification of novel drug targets. *Sci Signal* **4**: t3

50. Fernandez JM, Hoffmann R, Valencia A (2007) iHOP web services. *Nucleic Acids Res* **35**: W21–W26
51. Fisher J, Piterman N, Hajnal A, Henzinger TA (2007) Predictive modeling of signaling crosstalk during *C. elegans* vulval development. *PLoS Comput Biol* **3**: e92
52. Fisher J, Piterman N, Hubbard EJ, Stern MJ, Harel D (2005) Computational insights into *Caenorhabditis elegans* vulval development. *Proc Natl Acad Sci U S A* **102**: 1951–1956
53. Flanders KC (2004) Smad3 as a mediator of the fibrotic response. *Int J Exp Pathol* **85**: 47–64
54. Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Eyre T, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Holland R, Howe KL, Howe K, Johnson N, Jenkinson A, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Prlic A, Rice S, Rios D, Schuster M, Sealy I, Slater G, Smedley D, Spudich G, Trevanion S, Vilella AJ, Vogel J, White S, Wood M, Birney E, Cox T, Curwen V, Durbin R, Fernandez-Suarez XM, Herrero J, Hubbard TJ, Kasprzyk A, Proctor G, Smith J, Ureta-Vidal A, Searle S (2008) Ensembl 2008. *Nucleic Acids Res* **36**: D707–D714
55. Frame S és Cohen P (2001) GSK3 takes centre stage more than 20 years after its discovery. *Biochem J* **359**: 1–16
56. Fraser ID és Germain RN (2009) Navigating the network: signaling cross-talk in hematopoietic cells. *Nat Immunol* **10**: 327–331
57. Freeman M (2000) Feedback control of intercellular signalling in development. *Nature* **408**: 313–319
58. Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ (2009) Synthetic gene networks that count. *Science* **324**: 1199–1202
59. Friedman A és Perrimon N (2007) Genetic screening for signal transduction in the era of network biology. *Cell* **128**: 225–231
60. Gao Z, Li H, Zhang H, Liu X, Kang L, Luo X, Zhu W, Chen K, Wang X, Jiang H (2008) PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics* **9**: 104
61. Garcia O, Saveanu C, Cline M, Fromont-Racine M, Jacquier A, Schwikowski B, Aittokallio T (2007) GOLORize: a Cytoscape plug-in for network visualization with Gene Ontology-based layout and coloring. *Bioinformatics* **23**: 394–396

62. Gaudet S, Janes KA, Albeck JG, Pace EA, Lauffenburger DA, Sorger PK (2005) A compendium of signals and responses triggered by prodeath and prosurvival cytokines. *Mol Cell Proteomics* **4**: 1569–1590
63. Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, Gavin AC (2010) Visualization of omics data for systems biology. *Nat Methods* **7**: S56–S68
64. Gerhart J (1999) 1998 Warkany lecture: signaling pathways in development. *Teratology* **60**: 226–239
65. Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, Berriz GF, Bertin N, Huang J, Chuang LS, Li N, Mani R, Hyman AA, Sonnichsen B, Echeverri CJ, Roth FP, Vidal M, Piano F (2005) Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436**: 861–865
66. Guo X és Wang XF (2009) Signaling cross-talk between TGF-beta/BMP and other pathways. *Cell Res* **19**: 71–88
67. Haan C, Kreis S, Margue C, Behrmann I (2006) Jaks and cytokine receptors--an intimate relationship. *Biochem Pharmacol* **72**: 1538–1546
68. Hanze J, Berthold A, Klammt J, Gallaher B, Siebler T, Kratzsch J, Elmlinger M, Kiess W (1999) Cloning and sequencing of the complete cDNA encoding the human insulin receptor related receptor. *Horm Metab Res* **31**: 77–79
69. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la CN, Tonellato P, Jaiswal P, Seigfried T, White R (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**: D258–D261
70. Harris TW, Antoshechkin I, Bieri T, Blasiar D, Chan J, Chen WJ, de la CN, Davis P, Duesbury M, Fang R, Fernandes J, Han M, Kishore R, Lee R, Muller HM, Nakamura C, Ozersky P, Petcherski A, Rangarajan A, Rogers A, Schindelman G, Schwarz EM, Tuli MA, Van Auken K, Wang D, Wang X, Williams G, Yook K, Durbin R, Stein LD, Spieth J, Sternberg PW (2010) WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res* **38**: D463–D467

71. Hlavacek WS, Faeder JR, Blinov ML, Perelson AS, Goldstein B (2003) The complexity of complexes in signal transduction. *Biotechnol Bioeng* **84**: 783–794
72. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* **4**: 682–690
73. Hornberg JJ, Bruggeman FJ, Westerhoff HV, Lankelma J (2006) Cancer: a Systems Biology disease. *Biosystems* **83**: 81–90
74. Huang S (2004) Back to the biology in systems biology: what can we learn from biomolecular networks? *Brief Funct Genomic Proteomic* **2**: 279–297
75. Huang TW, Tien AC, Huang WS, Lee YC, Peng CL, Tseng HH, Kao CY, Huang CY (2004) POINT: a database for the prediction of protein-protein interactions based on the orthologous interactome. *Bioinformatics* **20**: 3273–3276
76. Huffner F, Wernicke S, Zichner T (2007) Faspad: fast signaling pathway detection. *Bioinformatics* **23**: 1708–1709
77. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**: D211–D215
78. Hurlbut GD, Kankel MW, Lake RJ, Artavanis-Tsakonas S (2007) Crossing paths with Notch in the hyper-network. *Curr Opin Cell Biol* **19**: 166–175
79. Ideker T és Lauffenburger D (2003) Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol* **21**: 255–262
80. Inoue T, Wang M, Ririe TO, Fernandes JS, Sternberg PW (2005) Transcriptional network underlying *Caenorhabditis elegans* vulval development. *Proc Natl Acad Sci U S A* **102**: 4972–4977
81. Itoh S és ten Dijke P (2007) Negative regulation of TGF-beta receptor/Smad signal transduction. *Curr Opin Cell Biol* **19**: 176–184
82. Ivanova PT, Milne SB, Myers DS, Brown HA (2009) Lipidomics: a mass spectrometry based systems level analysis of cellular lipids. *Curr Opin Chem Biol* **13**: 526–531
83. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C (2009) STRING 8--a global

- view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**: D412–D416
84. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* **33**: D428–D432
 85. Kandasamy K, Mohan S, Raju R, Keerthikumar S, Kumar GS, Venugopal AK, Telikicherla D, Navarro DJ, Mathivanan S, Pecquet C, Gollapudi SK, Tattikota SG, Mohan S, Padhukasahasram H, Subbannayya Y, Goel R, Jacob HK, Zhong J, Sekhar R, Nanjappa V, Balakrishnan L, Subbaiah R, Ramachandra YL, Rahiman A, Keshava Prasad TS, Lin JX, Houtman JC, Desiderio S, Renault JC, Constantinescu S, Ohara O, Hirano T, Kubo M, Singh S, Khatri P, Draghici S, Bader GD, Sander C, Leonard WJ, Pandey A (2010) NetPath: a public resource of curated signal transduction pathways. *Genome Biol* **11**: R3
 86. Katoh M (2007) Networking of WNT, FGF, Notch, BMP, and Hedgehog signaling pathways during carcinogenesis. *Stem Cell Rev* **3**: 30–38
 87. Katoh M és Katoh M (2007) WNT signaling pathway and stem cell signaling network. *Clin Cancer Res* **13**: 4042–4045
 88. Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T (2004) PathBLAST: a tool for alignment of protein interaction networks. *Nucleic Acids Res* **32**: W83–W88
 89. Kelley R és Ideker T (2005) Systematic interpretation of genetic interactions using protein networks. *Nat Biotechnol* **23**: 561–566
 90. Kelso J, Visagie J, Theiler G, Christoffels A, Bardien S, Smedley D, Otgaar D, Greyling G, Jongeneel CV, McCarthy MI, Hide T, Hide W (2003) eVOC: a controlled vocabulary for unifying gene expression data. *Genome Res* **13**: 1222–1230
 91. Kemmer D, Huang Y, Shah SP, Lim J, Brumm J, Yuen MM, Ling J, Xu T, Wasserman WW, Ouellette BF (2005) Ulysses - an application for the projection of molecular interactions across species. *Genome Biol* **6**: R106
 92. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadrans S, Chaerkady R, Pandey A (2009) Human Protein Reference Database--2009 update. *Nucleic Acids Res* **37**: D767–D772
 93. Kestler HA, Wawra C, Kracher B, Kuhl M (2008) Network modeling of signal transduction: establishing the global view. *Bioessays* **30**: 1110–1125

94. Kholodenko BN (2006) Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol* **7**: 165–176
95. Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, Hoek JB (2002) Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc Natl Acad Sci U S A* **99**: 12841–12846
96. Kim PM, Lu LJ, Xia Y, Gerstein MB (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* **314**: 1938–1941
97. Kiss HJ, Mihalik A, Nanasi T, Ory B, Spiro Z, Soti C, Csermely P (2009) Ageing as a price of cooperation and complexity: self-organization of complex systems causes the gradual deterioration of constituent networks. *Bioessays* **31**: 651–664
98. Kitano H (2007) Towards a theory of biological robustness. *Mol Syst Biol* **3**: 137
99. Kitisin K, Saha T, Blake T, Golestaneh N, Deng M, Kim C, Tang Y, Shetty K, Mishra B, Mishra L (2007) Tgf-Beta signaling in development. *Sci STKE* **2007**: cm1
100. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, Hewett M, Lin Z, Liu Y, Liu S, Oliver DE, Rubin DL, Shafa F, Stuart JM, Altman RB (2001) Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. *Pharmacogenomics J* **1**: 167–170
101. Komarova NL, Zou X, Nie Q, Bardwell L (2005) A theoretical framework for specificity in cell signaling. *Mol Syst Biol* **1**: 23
102. Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet* **39**: 309–338
103. Korcsmaros T, Farkas IJ, Szalay MS, Rovo P, Fazekas D, Spiro Z, Bode C, Lenti K, Vellai T, Csermely P (2010) Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics* **26**: 2042–2050
104. Korcsmaros T, Szalay MS, Bode C, Kovacs IA, Csermely P (2007) How to design multi-target drugs: Target-search options in cellular networks. *Exp Op Drug Discovery* **2**: 799–808
105. Korcsmaros T, Szalay MS, Rovo P, Palotai R, Fazekas D, Lenti K, Farkas IJ, Csermely P, Vellai T (2011) Signalogs: orthology-based identification of novel signaling pathway components in three metazoans. *PLoS One* **6**: e19240
106. Kovacs IA, Palotai R, Szalay MS, Csermely P (2010) Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. *PLoS One* **5**: e12528

107. Kovacs IA, Szalay MS, Csermely P (2005) Water and molecular chaperones act as weak links of protein folding networks: energy landscape and punctuated equilibrium changes point towards a game theory of proteins. *FEBS Lett* **579**: 2254–2260
108. Krauss S és Brand MD (2000) Quantitation of signal transduction. *FASEB J* **14**: 2581–2588
109. Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, Kel-Margoulis O, Wingender E (2006) TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res* **34**: D546–D551
110. Kyriakis JM és Avruch J (2001) Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. *Physiol Rev* **81**: 807–869
111. Levchenko A (2003) Dynamical and integrative cell signaling: challenges for the new biology. *Biotechnol Bioeng* **84**: 773–782
112. Levine M és Tjian R (2003) Transcription regulation and animal diversity. *Nature* **424**: 147–151
113. Li S, Assmann SM, Albert R (2006) Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. *PLoS Biol* **4**: e312
114. Li Y, Agarwal P, Rajagopalan D (2008) A global pathway crosstalk network. *Bioinformatics* **24**: 1442–1447
115. Linding R, Jensen LJ, Ostheimer GJ, van Vugt MA, Jorgensen C, Miron IM, Diella F, Colwill K, Taylor L, Elder K, Metalnikov P, Nguyen V, Pasculescu A, Jin J, Park JG, Samson LD, Woodgett JR, Russell RB, Bork P, Yaffe MB, Pawson T (2007) Systematic discovery of in vivo phosphorylation networks. *Cell* **129**: 1415–1426
116. Lindvall JM, Blomberg KE, Smith CI (2003) In silico tools for signal transduction research. *Brief Bioinform* **4**: 315–324
117. Liu Y, Ross JF, Bodine PV, Billiard J (2007) Homodimerization of Ror2 tyrosine kinase receptor induces 14-3-3(beta) phosphorylation and promotes osteoblast differentiation and bone formation. *Mol Endocrinol* **21**: 3050–3061
118. Lu LJ, Sboner A, Huang YJ, Lu HX, Gianoulis TA, Yip KY, Kim PM, Montelione GT, Gerstein MB (2007) Comparing classical pathways and modern networks: towards the development of an edge ontology. *Trends Biochem Sci* **32**: 320–331

119. Luo W és Lin SC (2004) Axin: a master scaffold for multiple signaling pathways. *Neurosignals* **13**: 99–113
120. Lutz M és Knaus P (2002) Integration of the TGF-beta pathway into the cellular signalling network. *Cell Signal* **14**: 977–988
121. Mak HC, Daly M, Gruebel B, Ideker T (2007) CellCircuits: a database of protein network models. *Nucleic Acids Res* **35**: D538–D545
122. Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* **27**: 514–520
123. McClean MN, Mody A, Broach JR, Ramanathan S (2007) Cross-talk and decision making in MAP kinase pathways. *Nat Genet* **39**: 409–414
124. McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, Marcotte EM (2010) Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci U S A* **107**: 6544–6549
125. Medina MW, Sangkuhl K, Klein TE, Altman RB (2010) PharmGKB: very important pharmacogene - HMGCR. *Pharmacogenet Genomics* **2**:98-101
126. Morioka K, Tanikawa C, Ochi K, Daigo Y, Katagiri T, Kawano H, Kawaguchi H, Myoui A, Yoshikawa H, Naka N, Araki N, Kudawara I, Ieguchi M, Nakamura K, Nakamura Y, Matsuda K (2009) Orphan receptor tyrosine kinase ROR2 as a potential therapeutic target for osteosarcoma. *Cancer Sci* **100**: 1227–1233
127. Natarajan M, Lin KM, Hsueh RC, Sternweis PC, Ranganathan R (2006) A global analysis of cross-talk in a mammalian cellular signalling network. *Nat Cell Biol* **8**: 571–580
128. Neves SR és Iyengar R (2002) Modeling of signaling networks. *Bioessays* **24**: 1110–1117
129. O'Brien KP, Westerlund I, Sonnhammer EL (2004) OrthoDisease: a database of human disease orthologs. *Hum Mutat* **24**: 112–119
130. Oda K és Kitano H (2006) A comprehensive map of the toll-like receptor signaling network. *Mol Syst Biol* **2**: 15
131. Oda T, Kimura T, Matsuoka Y, Funahashi A, Muramatsu M, Kitano H (2004) Molecular Interaction Map of a Macrophage. *AfCS Research Reports* **2**: 1–12
132. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **27**: 29–34
133. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**: 635–648

134. Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stumpflen V, Ceol A, Chatr-aryamontri A, Armstrong J, Woollard P, Salama JJ, Moore S, Wojcik J, Bader GD, Vidal M, Cusick ME, Gerstein M, Gavin AC, Superti-Furga G, Greenblatt J, Bader J, Uetz P, Tyers M, Legrain P, Fields S, Mulder N, Gilson M, Niepmann M, Burgoon L, De Las RJ, Prieto C, Perreau VM, Hogue C, Mewes HW, Apweiler R, Xenarios I, Eisenberg D, Cesareni G, Hermjakob H (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat Biotechnol* **25**: 894–898
135. Papin JA, Hunter T, Palsson BO, Subramaniam S (2005) Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol* **6**: 99–111
136. Papin JA és Palsson BO (2004) Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. *J Theor Biol* **227**: 283–297
137. Pawson T és Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* **300**: 445–452
138. Persico M, Ceol A, Gavrilu C, Hoffmann R, Florio A, Cesareni G (2005) HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinformatics* **6** Suppl 4: S21
139. Pires-daSilva A és Sommer RJ (2003) The evolution of signalling pathways in animal development. *Nat Rev Genet* **4**: 39–49
140. Raftery LA és Sutherland DJ (1999) TGF-beta family signal transduction in *Drosophila* development: from Mad to Smads. *Dev Biol* **210**: 251–268
141. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincaid-Beal C, Kulkarni P, Varambally S, Ghosh D, Chinnaiyan AM (2007) Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**: 166–180
142. Roberts AB, Piek E, Bottinger EP, Ashcroft G, Mitchell JB, Flanders KC (2001) Is Smad3 a major player in signal transduction pathways leading to fibrogenesis? *Chest* **120**: 43S–47S
143. Robinson GW (2007) Cooperation of signalling pathways in embryonic mammary gland development. *Nat Rev Genet* **8**: 963–972
144. Rogers A, Antoshechkin I, Bieri T, Blasiar D, Bastiani C, Canaran P, Chan J, Chen WJ, Davis P, Fernandes J, Fiedler TJ, Han M, Harris TW, Kishore R, Lee R, McKay S, Muller HM, Nakamura C, Ozersky P, Petcherski A, Schindelman G, Schwarz EM, Spooner W, Tuli MA, Van Auken K, Wang D,

- Wang X, Williams G, Yook K, Durbin R, Stein LD, Spieth J, Sternberg PW (2008) WormBase 2007. *Nucleic Acids Res* **36**: D612–D617
145. Roudabush FL, Pierce KL, Maudsley S, Khan KD, Luttrell LM (2000) Transactivation of the EGF receptor mediates IGF-1-stimulated shc phosphorylation and ERK1/2 activation in COS-7 cells. *J Biol Chem* **275**: 22583–22589
 146. Ruvkun G és Hobert O (1998) The taxonomy of developmental control in *Caenorhabditis elegans*. *Science* **282**: 2033–2041
 147. Sakharkar MK, Sakharkar KR, Pervaiz S (2007) Druggability of human disease genes. *Int J Biochem Cell Biol* **39**: 1156–1164
 148. Salgado D, Gimenez G, Coulier F, Marcelle C (2008) COMPARE, a multi-organism system for cross-species data comparison and transfer of information. *Bioinformatics* **24**: 447–449
 149. Schubert P, Hoffman MD, Sniatynski MJ, Kast J (2006) Advances in the analysis of dynamic protein complexes by proteomics and data processing. *Anal Bioanal Chem* **386**: 482–493
 150. Sergina NV, Rausch M, Wang D, Blair J, Hann B, Shokat KM, Moasser MM (2007) Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature* **445**: 437–441
 151. Simonis N, Rual JF, Carvunis AR, Tasan M, Lemmens I, Hirozane-Kishikawa T, Hao T, Sahalie JM, Venkatesan K, Gebreab F, Cevik S, Klitgord N, Fan C, Braun P, Li N, Ayivi-Guedehoussou N, Dann E, Bertin N, Szeto D, Dricot A, Yildirim MA, Lin C, de Smet AS, Kao HL, Simon C, Smolyar A, Ahn JS, Tewari M, Boxem M, Milstein S, Yu H, Dreze M, Vandenhoute J, Gunsalus KC, Cusick ME, Hill DE, Tavernier J, Roth FP, Vidal M (2009) Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. *Nat Methods* **6**: 47–54
 152. Spiro Z, Kovacs IA, Csermely P (2008) Drug-therapy networks and the prediction of novel drug targets. *J Biol* **7**: 20
 153. Steffen M, Petti A, Aach J, D'haeseleer P, Church G (2002) Automated modelling of signal transduction networks. *BMC Bioinformatics* **3**: 34
 154. Stelling J, Sauer U, Szallasi Z, Doyle FJ, III, Doyle J (2004) Robustness of cellular functions. *Cell* **118**: 675–685
 155. Sternberg PW (2005) Vulval development. *WormBook* 1–28 <http://wormbook.org>
 156. Stevens CF (2004) Systems biology versus molecular biology. *Curr Biol* **14**: R51–R52

157. Stoevesandt O, Kohler K, Wolf S, Andre T, Hummel W, Brock R (2007) A network analysis of changes in molecular interactions in cellular signaling. *Mol Cell Proteomics* **6**: 503–513
158. Storm CE és Sonnhammer EL (2003) Comprehensive analysis of orthologous protein domains using the HOPS database. *Genome Res* **13**: 2353–2362
159. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101**: 6062–6067
160. Szathmary E, Jordan F, Pal C (2001) Molecular biology and evolution. Can genes explain biological complexity? *Science* **292**: 1315–1316
161. Taniguchi CM, Emanuelli B, Kahn CR (2006) Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Biol* **7**: 85–96
162. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33–36
163. Tewari M, Hu PJ, Ahn JS, Ayivi-Guedehoussou N, Vidalain PO, Li S, Milstein S, Armstrong CM, Boxem M, Butler MD, Busiguina S, Rual JF, Ibarrola N, Chaklos ST, Bertin N, Vaglio P, Edgley ML, King KV, Albert PS, Vandenhoute J, Pandey A, Riddle DL, Ruvkun G, Vidal M (2004) Systematic interactome mapping and genetic perturbation analysis of a *C. elegans* TGF-beta signaling network. *Mol Cell* **13**: 469–482
164. Tyson JJ, Chen KC, Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* **15**: 221–231
165. Vardy DA, Kari C, Lazarus GS, Jensen PJ, Zilberstein A, Plowman GD, Rodeck U (1995) Induction of autocrine epidermal growth factor receptor ligands in human keratinocytes by insulin/insulin-like growth factor-1. *J Cell Physiol* **163**: 257–265
166. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, Yildirim MA, Simonis N, Heinzmann K, Gebreab F, Sahalie JM, Cevik S, Simon C, de Smet AS, Dann E, Smolyar A, Vinayagam A, Yu H, Szeto D, Borick H, Dricot A, Klitgord N, Murray RR, Lin C, Lalowski M, Timm J, Rau K, Boone C, Braun P, Cusick ME, Roth FP, Hill DE, Tavernier J, Wanker EE, Barabasi AL, Vidal M (2009) An empirical framework for binary interactome mapping. *Nat Methods* **6**: 83–90
167. Villen J, Beausoleil SA, Gerber SA, Gygi SP (2007) Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A* **104**: 1488–1493

168. Wang CC, Cirit M, Haugh JM (2009) PI3K-dependent cross-talk interactions converge with Ras as quantifiable inputs integrated by Erk. *Mol Syst Biol* **5**: 246
169. Weng G, Bhalla US, Iyengar R (1999) Complexity in biological signaling systems. *Science* **284**: 92–96
170. White MA és Anderson RG (2005) Signaling networks in living cells. *Annu Rev Pharmacol Toxicol* **45**: 587–603
171. Wilson RJ, Goodman JL, Strelets VB (2008) FlyBase: integration and improvements to query tools. *Nucleic Acids Res* **36**: D588–D593
172. Wishart DS (2008) DrugBank and its relevance to pharmacogenomics. *Pharmacogenomics* **9**: 1155–1162
173. Wolkenhauer O, Sreenath SN, Wellstead P, Ullah M, Cho KH (2005) A systems- and signal-oriented approach to intracellular dynamics. *Biochem Soc Trans* **33**: 507–515
174. Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, Fiel I, Thung S, Mazzaferro V, Bruix J, Bottinger E, Friedman S, Waxman S, Llovet JM (2007) Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* **45**: 938–947
175. Xia Y, Yu H, Jansen R, Seringhaus M, Baxter S, Greenbaum D, Zhao H, Gerstein M (2004) Analyzing cellular biochemistry in terms of molecular networks. *Annu Rev Biochem* **73**: 1051–1087
176. Xiong J, Rayner S, Luo K, Li Y, Chen S (2006) Genome wide prediction of protein function via a generic knowledge discovery approach based on evidence integration. *BMC Bioinformatics* **7**: 268
177. Yan SJ, Gu Y, Li WX, Fleming RJ (2004) Multiple signaling pathways and a selector protein sequentially regulate *Drosophila* wing development. *Development* **131**: 285–298
178. Yellaboina S, Dudekula DB, Ko MS (2008) Prediction of evolutionarily conserved interologs in *Mus musculus*. *BMC Genomics* **9**: 465
179. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M (2007) Drug-target network. *Nat Biotechnol* **25**: 1119–1126
180. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, Bertin N, Chung S, Vidal M, Gerstein M (2004) Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res* **14**: 1107–1118
181. Yu J, Pacifico S, Liu G, Finley RL, Jr. (2008) DroID: the *Drosophila* Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC Genomics* **9**: 461

182. Zhu H és Snyder M (2002) "Omic" approaches for unraveling signaling networks. *Curr Opin Cell Biol* **14**: 173–179

Saját publikációk jegyzéke

A disszertációhoz kapcsolódó közlemények

1. **Korcsmaros T ***, Farkas IJ *, Szalay MS, Rovó P, Fazekas D, Spiro Z, Böde C, Lenti K, Vellai T, Csermely P (2010) Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics* **26**:2042-2050
IF: 4,9 **Független idézetek száma: 3**
2. **Korcsmaros T ***, Szalay MS *, Rovó P, Palotai R, Fazekas D, Lenti K, Farkas I J, Csermely P, Vellai T (2011) Signalogs: orthology-based identification of novel signaling pathway components in three metazoans. *PLoS ONE* **6**(5), e19240
IF: 4,4

A disszertációtól független közlemények

1. Nardai G, **Korcsmaros T**, Csermely P (2002) Reduction of the endoplasmic reticulum accompanies the oxidative damage of diabetes mellitus, In: *Redox regulation* (eds.: A. Pompella, G. Banhegyi and M. Wellman-Rousseau), NATO Science Series, **I/347**, 281-289
2. Nardai G, **Korcsmaros T**, Papp E, Csermely P (2003) Reduction of the endoplasmic reticulum accompanies the oxidative damage of diabetes mellitus. *Biofactors* **17**, 259-267
IF: 1,9 **Független idézetek száma: 12**
3. Papp E, **Korcsmaros T**, Nardai G, Csermely P (2004) Changes of cellular redox homeostasis and protein folding in diabetes, In: *Cellular dysfunction in atherosclerosis and diabetes - Reports from bench to bedside* (eds.: M. Simionescu, A. Sima, D. Popov), Plenum Press, 228-235
4. Nardai G, Stadler K, Papp E, **Korcsmaros T**, Jakus J, Csermely P (2005) Diabetic changes in the redox status of the microsomal protein folding machinery. *Biochemical and Biophysical Research Communications* **334**, 787-795
IF: 3,0 **Független idézetek száma: 14**
5. Nardai G, Papp E, **Korcsmaros T**, Stadler K, Jakus J, Csermely P (2005) Possible links between metabolism and oxidative protein folding. Consequences of a diabetes study, In: *Redox regulation* (eds.: A. Pompella, G. Banhegyi and M. Wellman-Rousseau), NATO Science Series, **363**, 101-109
6. Papp E, Nardai G, Sreedhar AS, **Korcsmaros T**, Csermely P (2005) Effects of unfolded protein accumulation on the redox state of the endoplasmic reticulum, In: *Redox regulation* (eds.: A. Pompella, G. Banhegyi and M. Wellman-Rousseau), NATO Science Series, **363**, 111-119

7. Papp E, Szaraz P, **Korcsmaros T**, Csermely P (2006) Changes of endoplasmic reticulum chaperone complexes, redox state, and impaired protein disulfide reductase activity in misfolding alfa-1-antitrypsin transgenic mice. *FASEB Journal* **20** (7): 1018-20
IF: 6,7 **Független idézetek száma: 12**
8. **Korcsmaros T**, Kovacs IA, Szalay MS, Csermely P (2006) Molecular chaperones: The modular evolution of cellular networks. *Journal of Bioscience* **32** (3): 441-446
IF: 1,0 **Független idézetek száma: 15**
9. Szalay MS, Kovács IA, **Korcsmaros T**, Böde C, Csermely P (2007) Stress-induced rearrangements of cellular networks: consequences for protection and drug design. *FEBS Lett.* **581**(19):3675-80
IF: 3,4 **Független idézetek száma: 15**
10. Böde C, Kovacs IA, Szalay MS, Palotai R, **Korcsmaros T**, Csermely P (2007) Network analysis of protein dynamics. *FEBS Lett.* **581**(15):2776-82
IF: 3,4 **Független idézetek száma: 21**
11. Csermely P, **Korcsmaros T**, Sulyok K (eds., 2007) Stress Responses in Biology and Medicine: Stress of Life in Molecules, Cells, Organisms, and Psychosocial Communities. *Annals of the New York Academy of Sciences*, **1113**, pp. 366
Független idézetek száma: 1
12. **Korcsmaros T**, Szalay MS, Böde C, Kovács IA, Csermely P (2007) How to design multi-target drugs: Target-search options in cellular networks. *Exp. Op. Drug Discovery* **2** (6): 799-808
Független idézetek száma: 22
13. Kovacs I, Csermely P, Korcsmaros T, Szalay MS (2007) *WO patent application* WO 2007093960
Független idézetek száma: 1
14. Csermely P, **Korcsmaros T**, Kovács IA, Szalay MS, Söti C (2008) Systems biology of molecular chaperone networks. In: The biology of extracellular molecular chaperones. *Novartis Foundation Symposium Series* **291**, Wiley, pp. 45-58
Független idézetek száma: 5
15. I. J. Farkas, **T. Korcsmáros**, I. A. Kovács, Á. Mihalik, R. Palotai, G. I. Simkó, K. Z. Szalay, M. Szalay-Bekő, T. Vellai, S. Wang, P. Csermely. Network-based tools in the identification of novel drug-targets. *Sci. Signal.* **4**, pt3 (2011).

Összesítés

Tudományos publikációk: **11 cikk, 4 könyvfejezet, 1 könyvszerkesztés, 1 szabadalmi bejelentés**

Impakt faktorok száma: **28,7** Független idézetek száma az ISI és a Scopus adatbázis, valamint a Google Scholar alapján: **132 (2011. áprilisi állapot)**

Köszönetnyilvánítás

Ezúton köszönöm témavezetőimnek, **Prof. Csermely Péternek**, **Dr. Vellai Tibornak** és **Dr. Papp Balázsnak** a munka célkitűzéseinek megfogalmazásában, megvalósításában és a publikációk elkészítésében nyújtott elméleti és gyakorlati segítségeiket.

Az itt bemutatott vizsgálatok egy részét **Dr. Farkas Illés** hálózatkutató biofizikussal, illetve **Szalay-Bekő Máté** informatikussal együttműködve végeztem, valamint TDK hallgatómmal, **Fazekas Dávid** biológus hallgatóval. Hármuk szakértelmének köszönhetően a speciális vizsgálatok is a megfelelő szakmai szintvonalon történtek. **Dr. Farkas Illésnek** külön hálával tartozom a kéziratok elkészítésében végzett munkájáért.

Szintén köszönettel tartozom korábbi TDK hallgatóimnak, **Rovó Petrának**, **Spiró Zoltánnak** és **Zsákai Liliannak**, akikkel a SignaLink adatbázis készítése során dolgoztam együtt 2006 és 2008 között. A statisztikai vizsgálatok elvégzésében és ellenőrzésében folyamatos segítséget kaptam **Dr. Böde Csaba** szakértő statisztikustól. A hálózatok ábrázolását **Palotai Robin** és **Szuromi Gábor** informatikusok segítették.

Külön kiemelném **Dr. Lenti Katalint**, aki kutató tanárként középiskolás koromban bevezetett a kutatás világába, javasolta a Semmelweis Egyetemen a KutDiák-témám elkezdését, majd az elmúlt években, mint a Semmelweis Egyetem ETK Morfológiai és Fiziológiai Tanszékének adjunktusa, közvetlenül is részt vett a SignaLink adatbázis elemzésében. A doktori disszertáció lektorálásában végzett alapos és minden bizonnyal fárasztó munkáját is köszönöm!

Jelenlegi kollégáimra az ELTE TTK Genetikai tanszékén az elmúlt években mindig számíthattam. Ezúton köszönöm a **Vellai-labor** és a **Hálózatbiológiai-csoport összes tagjának** javaslatait és segítségét. Külön kiemelném **Ari Eszter** filogenetikust, **Billes Viktor** PhD. hallgatót és **Földvári-Nagy László** biológust, akik közvetlenül is hozzájárultak a kutatásaink sikeréhez.

Köszönöm **Dr. Jordán Ferencnek** a SignaLink hálózat építésével és elemzésével kapcsolatban megfogalmazott tanácsait, valamint a kézirataim alapos átnézését. Szintén köszönöm a projektben együttműködőként részt vevő **Dr. Vicsek Tamás** akadémikusnak (MTA-ELTE Statisztikus és Biofizikai kutatócsoport) a segítségét, tanácsait és anyagi támogatását. **Dr. Vattay Gábornak** és **Dr. Csabai Istvánnak** az ELTE Komplex rendszerek fizikája tanszék vezetőjének és munkatársának is köszönettel tartozom a SignaLink adatbázis elemzésével kapcsolatos tanácsaikért és támogatásukért. Hálával tartozom a **Semmelweis Egyetem Központi Könyvtárának**, hogy az elektronikus könyvtár-rendszeren keresztül éjjel-nappal kutathattam.

Végül köszönöm **családomnak és egyéb állatfajtáinknak**, hogy az elmúlt években a SignaLink hálózat tervezése, építése és elemzése alatt megértéssel viseltettek irántam. Külön köszönöm feleségemnek, **Papp Diának**, hogy folyamatos támogatásával biztosította a doktori munkám elvégzésének lehetőségét, és nem zavarta, hogy igen gyakran otthon dolgoztam. Nélküle ez a munka nem jött volna létre!