DOI:10.14753/SE.2018.2134

A sejtkompartmentek rendszerbiológiájának elemzése, mint a rák kialakulásának egyik új vizsgálati módszere

Doktori értekezés

Dr. Veres Dániel

Semmelweis Egyetem

Molekuláris orvostudományok Doktori Iskola

Témavezető: Dr. Csermely Péter, az MTA lev. tagja, egyetemi tanár

Hivatalos bírálók: Dr. Bödör Csaba, PhD., tudományos főmunkatárs

Dr. Horváth Zsolt, PhD., főorvos

Szigorlati bizottság elnöke:

Dr. Buzás Edit, DSc., egyetemi tanár

Szigorlati bizottság tagjai:

Dr. Cserző Miklós, PhD., tudományos főmunkatárs

Dr. Reményi Attila, PhD., tudományos főmunkatárs

Budapest

2017

DOI:10.14753/SE.2018.2134

Tartalomjegyzék

DOI:10.14753/SE.2018.2134

DOI:10.14753/SE.2018.2134

DOI:10.14753/SE.2018.2134

1. Rövidítések jegyzéke

AD	Activation Domain (aktivációs domén)
AKT	RAC-alpha serine/threonine-protein kinase (RAC-alfa szerin/treonin-protein kináz)
APC/C	Anaphase-Promoting Complex/Cyclosome (anafázis-promótáló komplex/cikloszóma)
ARNT	Aryl Hydrocarbon Receptor Nuclear Translocator (aril hidrokarbon receptor sejtmagi transzlokátor)
ATP	Adenosine TriPhosphate (adenozin-trifoszfát)
BiP	Binding immunoglobulin Protein (kötő immunoglobulin fehérje)
BMI1	Polycomb complex protein BMI-1 (polikomb komplex fehérje BMI-1)
BRET	Bioluminescence Resonance Energy Transfer (biolumineszcencia rezonancia energiatranszfer)
CDKN2A	Cyclin-Dependent Kinase inhibitor 2A (ciklin-dependens kináz inhibitor 2A)
CSV	Comma Separated Values (vesszővel szeparált értékek)
DBD	DNA Binding Domain (DNS-kötő domén)
EGFR	Epidermal Growth Factor Receptor (epidermális növekedési faktor receptor)
EMT	Epithelial–Mesenchymal Transition (epitheliális-mezenchimális átmenet)
ER	Endoplasmic Reticulum (endoplazmatikus retikulum)
ERK	Extracellular signal-Regulated Kinase (extracelluláris szignál-regulált kináz)
FAK	Focal Adhesion Kinase (fokális adhéziós kináz)
FRA1	Fos-Related Antigen 1 (Fos transzkcripciós faktorhoz kötött antigén 1)
FLIM	Fluorescence Lifetime Imaging (fluoreszcencia-élettartam mérésen alapuló mikroszkópia)
FRET	Fluorescence Resonance Energy Transfer (fluoreszcencia rezonancia energia transzfer)
GATA4	Transcription factor GATA-4 (GATA-4 transzkripciós faktor)
GO	Gene Ontology (gén ontológia)

GWAS	Genome-Wide Association Study (teljes genomra kiterjedő asszociációs vizsgálat)
GTP	Guanosine TriPhosphate (guanozin-trifoszfát)
HCM	Hidden Correlation Modeling (rejtett korrelációs modellezés)
HIF	Hypoxia Induced Factor (hipoxia indukált faktor)
HI-FI	High-throughput Interactions by Fluorescence Intensity (fluoreszcencia intenzitás alapú nagy áteresztőképességű kölcsönhatás meghatározás)
hTERT	Human Telomerase Reverse Transcriptase (katalitikus humán telomeráz)
IGFBP-2	Insulin-like Growth Factor-Binding Protein 2 (inzulin-szerű növekedési faktort kötő fehérje)
IL-11	InterLeukin 11 (interleukin 11)
KLF4	Krueppel-Like Factor 4 (Krüppel faktor szerű fehérje 4)
MAPK	Mitogen-Activated Protein Kinase (mitogén aktivált protein kináz)
MEK1	MAPK/ERK Kinase 1 (MAPK/ERK kináz 1)
MITF	Microphthalmia-associated Transcription Factor (microphthalmia-assszociált transzkripciós faktor)
miR-200	miR-200 microRNA family (miR-200 mikroRNS család)
MPS1	MonoPolar Spindle 1 kinase (‘monopolar spindle 1’ kináz)
MYC	Myc proto-oncogene protein (myc proto-onkogén fehérje)
NANOG	Homeobox transcription factor Nanog (homeobox transzkripciós faktor Nanog)
NF-kB	Nuclear Factor kappa B (nukleáris faktor kappa B)
NGS	Next Generation Sequencing (új generációs szekvenálás)
NLS	Nuclear Localization Signal (sejtmagi lokalizációs szignál)
NMR	Nuclear Magnetic Resonance (mágneses magrezonancia spektroszkópia)
NPC	Nuclear Pore Complex (sejtmagi pórus komplex)
NTS	Nuclear Translocation Signal (sejtmagba irányító szignál)
OCT4	Octamer-binding protein 4 (oktamer-kötő fehérje 4)
PHP	Hypertext Preprocessor (‘PHP’ szerveroldali szkriptnyelv)
PP2	4-amino-5-(4-chlorophenyl)-7-(t-butyl)pyrazolo[3,4-d]pyrimidine (4-amino-5-(4-klorofenil)-7-(t-butil)pirazolo [3,4-d]pirimidin)

PTEN	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN (foszfatidil-inozitol 3,4,5-trifoszfát 3-foszfatáz és kettős-specificitású fehérje foszfatáz, PTEN)
P53	Cellular tumor antigen p53 (celluláris tumor antigén p53)
SAC	Spindle Assembly Checkpoint (osztódási orsót ellenőrző pont)
SNAIL2	Zinc finger protein SNAI2 (SNAI2 cink-ujj fehérje)
SOX2	Transcription factor SOX-2 (SOX-2 transzkripciós faktor)
SQL	Structured Query Language (struktúrált lekérdezőnyelv)
SVM	Support Vector Machine (tartóvektor gép)
TF	Transcription Factor (transzkripciós faktor)
TGF-béta	Transforming Growth Factor Beta-1 (transzformáló növekedési faktor béta-1)
trans-OWAS	trans-Ome-Wide Association Study (transz-omika-szintű asszociációs vizsgálat)
TWIST1	Twist-related protein 1 (twist-kapcsolt fehérje 1)
UAS	Upstream Activator Sequence (‘upstream’ aktivátor szekvencia)
VCP	Valosin-Containing Protein (valozin-tartalmú fehérje)
VDAC1	Voltage-Dependent Anion-selective Channel protein 1 (feszültségfüggő anion-szelektív csatorna fehérje 1)
VEGF	Vascular Endothelial Growth Factor (vaszkuláris endoteliális növekedési faktor)
WNT	int/Wingless family (int/Wingless géncsalád)
ZEB	Zinc finger E-box-Binding homeobox (ZEB transzkripciós faktor)
ZEB1	Zinc finger E-box-Binding homeobox 1 (ZEB1 transzkripciós faktor)
ZEB2	Zinc finger E-box-Binding homeobox 2 (ZEB2 transzkripciós faktor)

DOI:10.14753/SE.2018.2134

2. Ábrák és táblázatok jegyzéke

2.1. Ábrák

DOI:10.14753/SE.2018.2134

2.2. Táblázatok

DOI:10.14753/SE.2018.2134

3. Bevezetés

3.1. Általános áttekintés

A daganatos betegségek világszerte a vezető halálokok közé tartoznak[1], egyes országokban átvéve a korábban egyeduralkodó kardiovaszkuláris betegségekkel összefüggő halálozás elsődleges helyét [Siegel és mtsai 2016]. A jelenség okai köré sorolható az egyre öregedő népesség és az öregedéssel járó fokozott előfordulás, de nem elhanyagolható az a tény sem, hogy az egyre hatékonyabb és célzottabb terápiák ellenére a daganatok túlnyomó többsége továbbra sem gyógyítható, csak kezelhető.

A daganat ellenes terápiák fejlesztési folyamata során az első emberen végzett vizsgálatoktól (Fázis 1) a kifejlesztett molekulák 6,7%-a jut el addig, hogy bejegyzett gyógyszer legyen [Hay és mtsai 2014]. Ez az arány sokkal rosszabb, mint más fő terápiás területeken, például a fertőző betegségek indikációban 16,7% a megfelelő mutató [Hay és mtsai 2014]. A kutatás és fejlesztés hatékonyságának növelésére számos megközelítés létezik. Ezek közül kiemelkednek azon megoldások, melyek a sejtvonalakon és állatokon végzett kísérletek emberre történő lefordításának hatékonyságát próbálják növelni.

A tradicionális daganatos gyógyszerfejlesztés során a sejtvonalakon végzett *in vitro* kísérletek tanulságai sokszor nehezen fordíthatók le az állatokon végzett *in vivo* kísérletek eredményeire, és még nehezebb ezek alapján következtetéseket levonni az embereken várható hatékonyságra és toxicitásra. Az *in vitro* és *in vivo* modellek fejlesztése és ezzel prediktív erejük növelése mellett egyre sikeresebb törekvések folynak a folyamat támogatására számítógépes módszerekkel (*in silico*). A számítógépes módszerek képesek hasznosítani az *in vitro* és *in vivo* kísérletek eredményeit, és ezek alapján pontosabb kísérletek tervezésére adnak lehetőséget, mely kísérletek eredményei szintén visszatáplálhatóak a rendszerbe, ezzel egy folyamatos tanulási kört fenntartva [Marcu és Harriss-Phillips 2012].

A számítógépes biológia alapja, hogy a kísérletes módszerek által generált adattömeget feldolgozza, rendszerezze és elemezze. Nem elegendő azonban pusztán az egyes adatpontokat szemlélni, elengedhetetlen az adaton belüli összefüggések keresése. Erre

[1] http://www.who.int/mediacentre/factsheets/fs297/en/

egy lehetőség az egyes adatpontokat egy megadott rendezési elv alapján hálózatokba kapcsolni. Az orvosbiológiai adatokat hálózatként tekintve és elemezve eljutunk a hálózatos orvoslás fogalmához, mely egy fiatal, de egyre elterjedtebb tudományterület [Barabási és mtsai 2011].

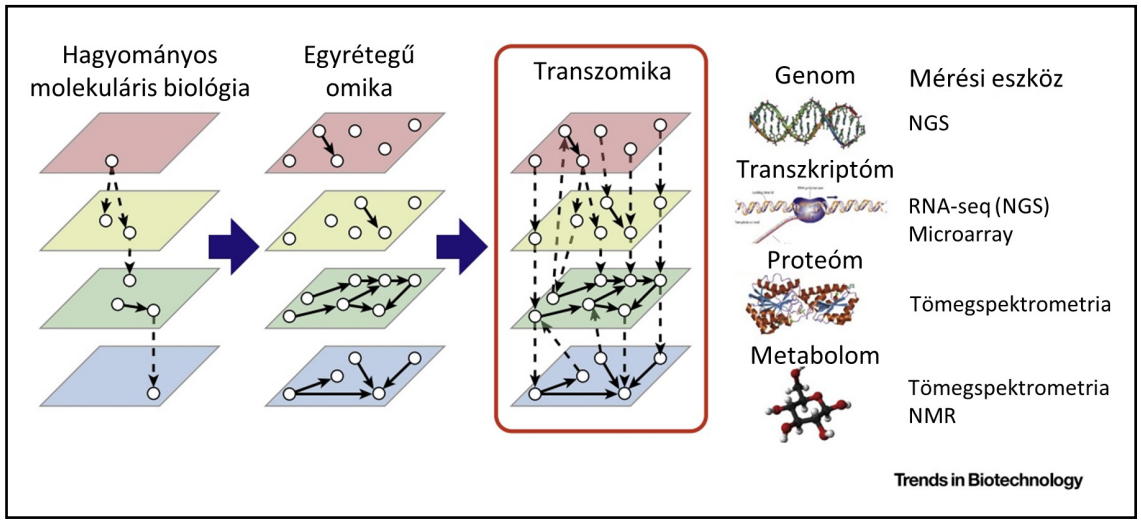
A posztgenomikai korszak magával hozta a különböző szintű biológiai adatok szélesebb körű kinyerésének, és ezek közötti rendszerszintű összefüggések keresésének lehetőségét. Így született meg az epigenomika, transzkriptomika, lipidomika, metabolomika vagy a proteomika tudományága. Ezen rendszerszintű adatokon alapuló, általában hálózatosan szerveződő adatok önállóan is az eddiginél sokkal kiterjedtebb elemzésekre adnak lehetőséget, így például az *in vitro* sejtvonalas és *in vivo* emberben megtalálható daganatos molekuláris minták farmakogenomikai elemzésére [Iorio és mtsai 2016].

Az újabb és fejlettebb nagy áteresztőképességű kísérletes módszereknek köszönhetően az elérhető adatok mennyisége rohamosan növekszik, ami kihívások elé állítja az orvosbiológiai adatokkal foglalkozó kutatókat [Marx 2013]. A kihívások közé tartozik az egyes elszórt adatbázisok és webes felületek fenntartása és folyamatos támogatása, az adatok annotálásának fejlesztése kézi adatgyűjtéssel, vagy éppen az egységes nevezéktanok bevezetése és következetes használata. Mindezek a problémák legfőképp az adatok integrációját nehezítik [Gomez-Cabrero és mtsai 2014], mely probléma megoldására több törekvés is elindult az elmúlt években (pl. az Európai Élettudományi Biológiai Információ Infrastruktúra Program, amelybe hazánk 2017-ben csatlakozott; https://www.elixir-europe.org/).

A technikai nehézségek leküzdésével azonban számtalan lehetőség nyílik az adatok elemzésére, az egyes adatrétegek összekötésére, és ezzel sokkal megbízhatóbb következtetések levonásra. Ennek az a magyarázata, hogy az adatokban található inkonzisztencia kiszűrhető az adatrétegek összekötésével. Erre jó példa a genomika-transzkriptomika-proteomika tengely, ahol az adatok összekötésével például kiszűrhetőek a valójában nem kifejeződő génszakaszok (amennyiben az adott génszakasz törlődött a genomból), vagy egy adott génről átíródó magas hírvivő RNS koncentráció alacsony célfehérje koncentrációval párosulva további szabályozási körök befolyására, például mikro-RNS aktivitásra utalhat.

DOI:10.14753/SE.2018.2134

Ezt a megközelítést nevezik transzomikai elemzésnek (**1. ábra**), ahol a fenotipikus viselkedés megértését az egyes rendszerszintű adatszintek közötti összefüggések feltárása segíti [Yugi és mtsai 2016]. A megközelítésben rejlő lehetőségek miatt az önmagában csak a genomikai szintet felhasználó GWAS tanulmányokat hamarosan felválthatja a mélyebb és átfogóbb trans-OWAS módszer.



***1. ábra: Többszörös adatrétegeket összekötő transzomikai hálózat**. A hagyományos molekuláris biológiai kutatás során egy-egy folyamat megadott elemeit, így például egy adott fehérje kapcsolatait és szabályozó szerepét vizsgáljuk átfogóan, több különböző adatréteg oldaláról egyaránt. Az egyrétegű omikai megközelítés esetében egy vizsgált adatrétegen belüli összefüggéseket keresünk, így például a transzkriptomikai adatrétegben keresünk eltérést a gének kifejeződésében. Ezzel szemben a transzomikai megközelítés ötvözve a hagyományos molekuláris biológiai módszereket az egyrétegű omikai szemlélettel, egyszerre vizsgálja az egyes adatrétegeken belüli, és azok közötti összefüggéseket, ezzel sokkal komplexebb molekuláris mintákat kapcsolva a megfigyelt tulajdonságok mögé. Az egyes adatrétegeket, és az elemzésükre használt módszereket az ábra jobb oldala foglalja össze. Forrás: [Yugi és mtsai 2016]*

Doktori munkám során tudományos diákkörös tevékenységem [Veres 2013] folytatásaként azt vizsgáltam, hogy a proteomikai réteg fehérjék közötti kapcsolataiban

található ellentmondásos, vagy biológiailag nem valószínű kapcsolatok hogyan szűrhetők ki egy újabb adatréteg, a fehérjék sejten belüli lokalizációs adatainak bevonásával.

3.2. Fehérje-fehérje interakciós adatok jellemzése és forrásai

3.2.1. Az interaktómok általános jellemzése

A fehérjék funkciójára sokszor az interakciós partnerek elemzése alapján derül fény. A fehérje-fehérje interakciós adat az egyik legértékesebb forrása a proteóm-szintű elemzéseknek [Koh és mtsai 2012]. Kifejezetten igaz ez akkor, ha a proteomikai adatokat a betegségek rendszerszintű megértése [Vidal és mtsai 2011], vagy hálózatos gyógyszer tervezés céljából szeretnénk hasznosítani [Ivanov és mtsai 2013].

A fehérje-fehérje interakciós hálózatok, vagy más néven interaktómok elemzéséhez és megértéséhez először meg kell határoznunk a hálózat elemeit és azok kapcsolatainak tulajdonságait, illetve meg kell ítélni azok biológiai relevanciáját az adatok forrása alapján. Az interaktómok klasszikus leírása alapján a kölcsönhatást létrehozó illeszkedés nem véletlenszerű, hanem egy szándékos biomolekuláris esemény vagy erő eredménye. Szintén feltétele a kölcsönhatásnak, hogy az illeszkedő felületek a kapcsolatra specifikusak legyenek, ne általános céllal jöjjenek létre [De Las Rivas és Fontanillo 2010].

A klasszikus leírás feltételei azonban a ma ismert interakciók jelentős részénél nem teljesülnek. Az adatok eredete alapján két típust, a genetikai és fizikai kölcsönhatást különítjük el. Előbbiek forrása nem közvetlen fizikai kölcsönhatás megfigyelése, hanem genetikai adatokon alapuló predikció eredménye [Chatr-Aryamontri és mtsai 2017]. A fizikai kölcsönhatások csoportja két alcsoportra különül el: a direkt és indirekt kapcsolatokra.

A fehérje-fehérje interakció detektálási módszerek gyakran fehérje komplexeket vizsgálnak, ahol nehéz elkülöníteni, hogy az egyes fehérjék egymással közvetlen (direkt) bináris kapcsolatban vannak-e, vagy egy másik fehérjén keresztül közvetetten (indirekt) hatnak kölcsön egymásra. Azonban a bináris, direkt interakciók meghatározása is gyakran hamis pozitív eredményhez vezet. Becslések szerint az ismert fehérje-fehérje

kölcsönhatások kb. 20%-a valós, a fennmaradó rész nem jön létre élő sejtekben. Ez is mutatja az adatok mennyiségén túl a minőség emelésének szükségességét.

A fehérje-fehérje interakciós adatok rendszerezése nehéz, ennek ellenére elengedhetetlen. Az elérhető adatbázisok átfedése alacsony [De Las Rivas és Fontanillo 2010], bár az elmúlt években több kezdeményezés is történt az adatok integrációjára [Lehne és Schlitt 2009, Kamburov és mtsai 2013], melyek hatására jelentősen növekedett a több adatforrás által is megerősített kapcsolatok száma. Az integrációt nehezíti a különböző fehérje nevezéktanok használata, azonban az elmúlt évek törekvései egy egységes nevezéktan bevezetésére (kézi adatgyűjtéssel és ellenőrzéssel kiegészítve) [Orchard és mtsai 2007, The UniProt Consortium 2017] jelentősen javították az adatok minőségét és felhasználhatóságát.

3.2.2. Fehérje-fehérje interakciókat meghatározó módszerek

A fehérjék között fellépő kölcsönhatásokat kísérletes módszerrel vagy bioinformatikai eszközökkel határozhatjuk meg. Utóbbi módszert fel lehet használni a kísérletes könyvtárakban még nem szereplő interakciók jóslására, valamint a már létező kölcsönhatások megerősítésére. Elterjedt a Bayes megközelítés használata, mely során génexpressziós adatok, ortológia, domén szerkezet vagy poszttranszlációs módosulások hasonlósága alapján valószínűsítenek kapcsolatot a fehérjék között, melynek mértéke adja a kapcsolat megbízhatósági értékét [McDowall és mtsai 2009].

Alapvetően kétféle kísérletes fehérje-fehérje interakció meghatározó módszert különítünk el. A bináris megközelítés a közvetlen fizikai kölcsönhatást vizsgálja két fehérje között, míg a ko-komplex módszer több fehérje kapcsolódását írja le. A kísérletek áteresztőképessége alapján elkülönítünk alacsony és magas áteresztőképességű technológiákat (*low* és *high throughput*). Előbbiek minősége sokkal megbízhatóbb, mint az utóbbiaké, azonban a proteóm szintű elemzésekhez elengedhetetlen a magas áteresztőképességű technológiák által nyújtott nagy mennyiségű adat használata is.

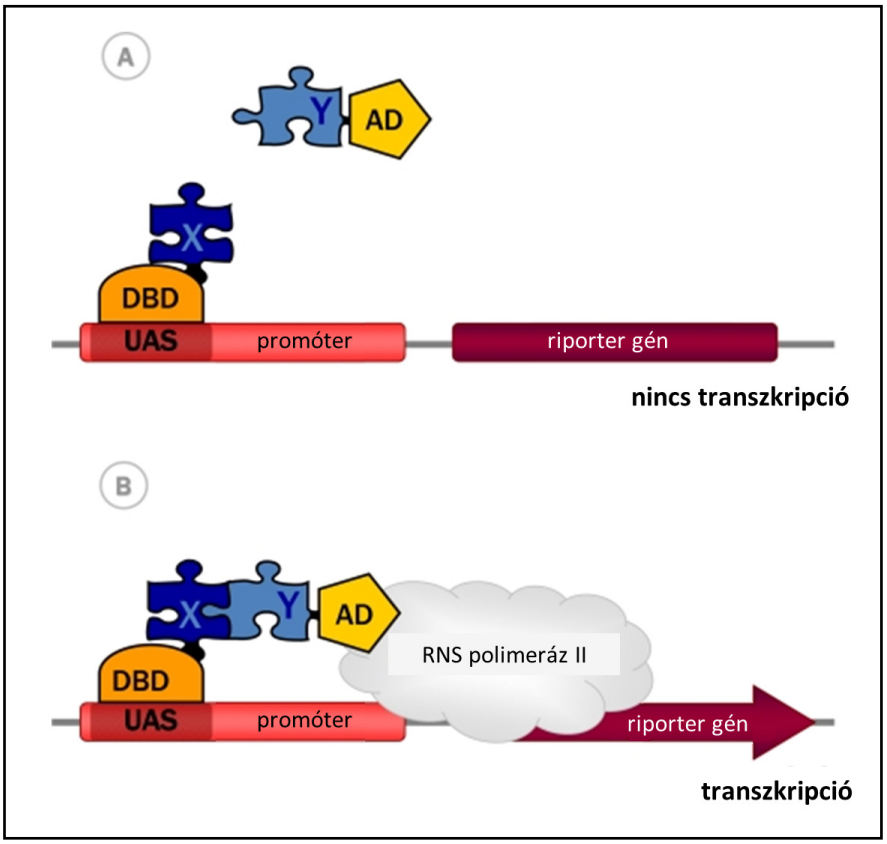
A leggyakrabban használt meghatározási eszközök[2] közül kiemelném az adatok legnagyobb mennyiségét adó nagy áteresztőképességű élesztő kettős hibrid és ko-immunprecipitációs módszert tömegspektrometriával kombinálva, valamint a

[2] https://wiki.thebiogrid.org/doku.php/experimental_systems/

DOI:10.14753/SE.2018.2134

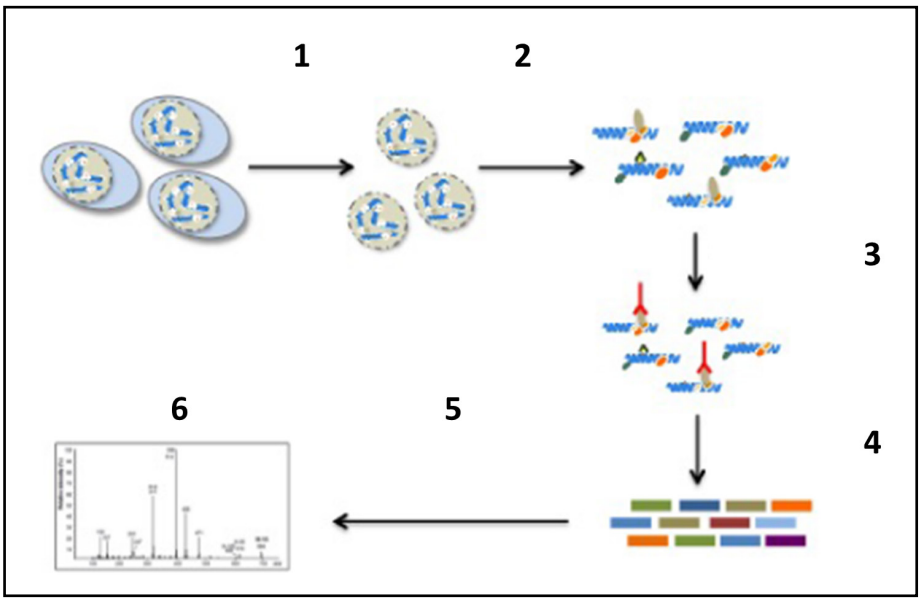
mennyiségi meghatározásra is alkalmas alacsonyabb áteresztőképességű FRET módszereket.

A klasszikus élesztő kettős hibrid módszert mutatja be a **2. ábra** [Fields és Song 1989, Brückner és mtsai 2009].



***2. ábra A klasszikus élesztő kettős hibrid módszer**. A vizsgált X fehérje fuzionál a DNS kötő doménnel (DBD), ezt hívjuk csalinak. A potenciálisan kölcsönható partner Y fehérje az aktivációs doménnel (AD) fuzionál, ezt nevezzük prédának. A csali beköt a riporter gén (LacZ) felső aktivációs szekvenciájához (UAS). Amennyiben X és Y fehérjék kölcsönhatnak, úgy az aktivációs domén jelenlétében létrejön a funkcionális transzkripciós faktor, és megkezdődik a riporter gén átírása az RNS polimeráz II enzim segítségével. A riporter gén egy kifejeződés esetén egyszerűen leolvasható fehérjét kell, hogy kódoljon. Mivel az élesztő kettős hibrid technika gyakran eredményez aspecifikus kötődést és ezzel hamis pozitív kölcsönhatásokat, így általában több, mint egy riporter gént alkalmaznak egyszerre, melynek eredménye egy erőteljesebb transzkripcionális aktivitás. Forrás: [Brückner és mtsai 2009]*

A ko-komplex módszerek közül a leggyakrabban használt az úgynevezett ko-immunoprecipitáció tömegspektrometriával kombinálva [Elion 2006]. A technika lényege (**3. ábra**), hogy a sejt lizátumban úszó vizsgálni kívánt csali fehérjét egy ellene termelt ellenanyaggal immunprecipitáljuk, melyhez később a komplex többi tagja is kötődni fog. A komplexet ezután tömegspektrometriával analizájuk, ezzel meghatározva, hogy mely fehérjék vesznek részt a komplex kialakításában.

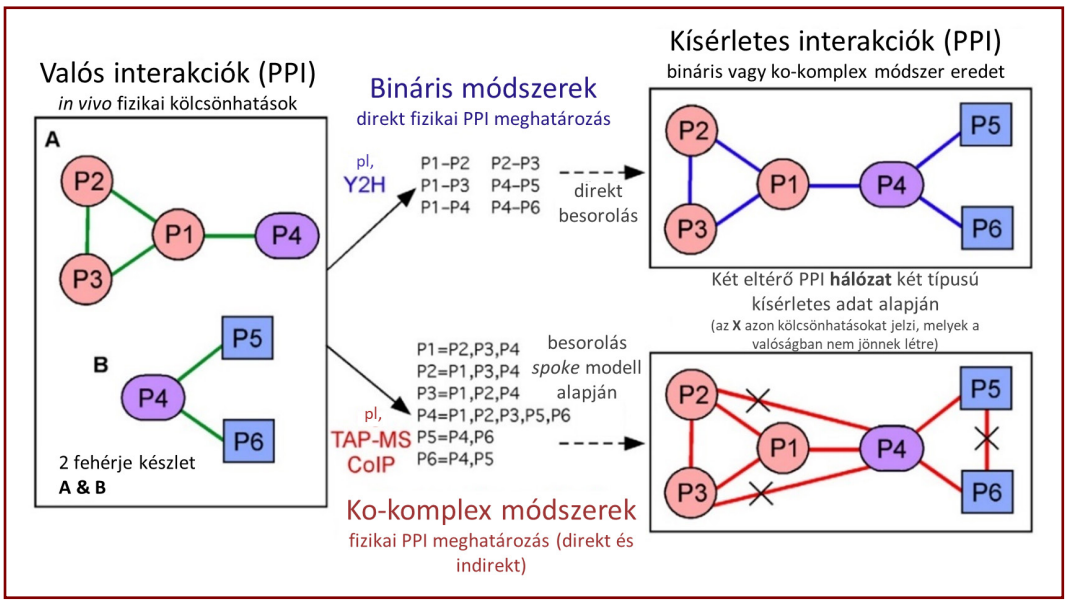


***3. ábra**: **A ko-immunoprecipitációs technika**. A folyamat lépései egy sejtmagi fehérje komplex kimutatás példáján: 1) a két vizsgálni kívánt fehérjét tartalmazó sejtmag meghatározása és a sejtmag izolálása, 2-3) a vizsgált fehérjéket célzó antitest és ellenanyag kötő gyöngyök hozzáadása, 4) kívánt fehérjék immunoprecipitációja, 5) immunoprecipitált fehérjék kimosása és gyűjtése, 6) az immunoprecipitált fehérjék vizsgálata tömegspektrometria segítségével.*
Forrás: https://www.activemotif.com/images/products/RIME-method-2.png/

Mint láthatjuk a két gyakran használt nagy áteresztőképességű technika számos hamis kapcsolatot generálhat. Az élesztő kettős hibrid módszer esetében aspecifikus kölcsönhatások jöhetnek létre a fehérjék túl magas koncentrációja, a fúziós proteinek gátló hatása, vagy az élesztő-idegen fehérjék nem megfelelő konformációja miatt. A technika további hátránya, hogy az élesztő sejtmagjához kötött, így olyan fehérjék is

kapcsolatba tudnak lépni egymással melyek az emberi sejtekben valójában külön sejtkompartmentekben találhatóak, mutatva a kölcsönhatások sejten belüli elhelyezkedése alapján történő szűrésének fontosságát.

A ko-immunoprecipitációs módszer komplexeket vizsgál, így az élesztő kettős hibrid direkt bináris interakcióival szemben itt indirekt kapcsolatokat is kaphatunk, ahol pusztán a kísérlet alapján nem elkülöníthető, hogy mely két fehérje van egymással valós direkt fizikai kölcsönhatásban. A kölcsönhatások valósághoz való közelítését több módszerrel is el lehet érni, melyek közül legelterjedtebb a spoke modell alkalmazása (**4. ábra**).



***4. ábra**: **A ko-komplex módszerrel meghatározott kölcsönhatások szűrése.** Az ábra két példa hálózat (A és B) segítségével mutatja be, hogy a valós fizikai kölcsönhatások milyen formában jelennek meg a bináris, illetve ko-komplex módszereken alapuló meghatározás után. Bináris módszer esetén nem jön létre hamis pozitív kapcsolat, a fizikai kölcsönhatások egyenként jelennek meg a derivált hálózatban. Ezzel ellentétben ko-komplex módszer alkalmazása esetén számos hamis pozitív kapcsolat keletkezik azon fehérjék között is, melyek valójában nem állnak egymással fizikai kölcsönhatásban. Ezek szűrésére alkalmas a spoke modell, melynek segítségével az X-szel jelölt kölcsönhatások kiszűrhetők. Forrás: [De Las Rivas és Fontanillo 2010]*

Összességében a magas áteresztőképességű módszerek nagy mennyiségű, de számos hamis pozitív interakciót tartalmazó adatot szolgáltatnak. Az adatok megbízhatóságának növeléséhez szükség van az alacsonyabb áteresztőképességű technikák alkalmazására, illetve a kapott kölcsönhatások bioinformatikai módszerekkel történő további feldolgozására.

Az alacsony áteresztőképességű technikák közé tartoznak az egy-egy fehérje kapcsolatát elemző, szerkezeti adatokon alapuló ko-kristály módszerek, illetve a kapcsolatok vizuális elemzésére alkalmas fluoreszcens technika. Utóbbi csoportba sorolható az elterjedten alkalmazott fluoreszcens rezonancia energia transzfer (FRET), mely sejten belüli bi-molekuláris kölcsönhatások elemzésére szolgál [Shrestha és mtsai 2015], illetve ennek másik változata a biolumineszcens rezonancia energia transzfer (BRET) [Xie és mtsai 2011]. A technika lényege, hogy a vizsgálni kívánt fehérjéket egy fluoreszcens donorral és egy akceptorral jelöljük meg, melyek között egymás közelébe (<~10 nanométer) kerülve sugárzás mentes energia átadás történik, melyet fluoreszcens mikroszkóp alatt vizsgálható fényjelenség kísér.

A tradicionális technika több lyukú plate-ek alkalmazása esetén is csak limitáltan alkalmas nagy mennyiségű kapcsolat kimutatására, így az automatizációra számos megoldást fejlesztettek ki az elmúlt években. Megjelentek technológiák, melyek speciális labor eszközök nélkül is lehetőséget adnak nagyobb mennyiségű interakció kvantitatív kimutatására, mint például a HI-FI módszer, mely a FRET és a fluoreszcencia kioltást kombinálja [Hieb és mtsai 2012]. Speciálisabb eszközöket igényel, ellenben még magasabb áteresztőképességű a FRET kombinálása fluoreszcencia-élettartam képalkotással (FLIM) [Margineanu és mtsai 2016], ami lehetőséget ad nagy mennyiségű és kvantitatív fehérje-fehérje interakció meghatározására.

3.2.3. Az interakciós adatok minőségének növelésére irányuló törekvések

A fehérje-fehérje interakciós adatok mennyiségének növelésén túl esszenciális az adatok megbízhatóságának javítása is. Minél megbízhatóbbak az adatok, azok elemzésével annál precízebb biológiai hipotézisek felállítására van lehetőség. A minőség javítására számos megközelítés létezik. Kiemelkedik az adatok integrációjára vonatkozó törekvés, melynek segítségével a több forrásban megtalálható, eltérő kísérletekből származó interakciók létezésének bizonyossága magasabb, mint az egyedi adatbázisokban megtalálható egy-

egy mérésen alapuló kölcsönhatások biológiai valószínűsége. Ilyen törekvés az IMEx konzorcium[3] [Orchard és mtsai 2012], mely magas kritériumokat támaszt az adatok minőségével és ellenőrzésével kapcsolatban.

Az adatok minősége javítható a kölcsönhatások meghatározott szempontok alapján történő szelekciójával is. Erre példa az Interactome3D adatbázis és alkalmazás [Mosca és mtsai 2012], mely a fehérjék és fehérje komplexek kristályszerkezetének vizsgálata alapján határoz meg magas megbízhatóságú kölcsönhatásokat. Hasonlóan jó minőségű de eltérő megközelítésű adatokat tartalmaz a MatrixDB [Launay és mtsai 2015], mely magas megbízhatóságú kísérletes adatokon alapul és kizárólag az extracelluláris mátrix kapcsolatait tartalmazza.

Számos számítógépes módszer létezik arra, hogy fehérje-fehérje interakciókat jósoljunk. E jóslások új interakciók elemzésére, vagy a régiek megerősítésére használhatók [Zahiri és mtsai 2013]. A hamis kapcsolatok kiszűrésére leggyakrabban a Bayes-módszert alkalmazzák, melynek segítségével különböző paraméterek mentén pontozható egy-egy kapcsolat megbízhatósága. Ilyen eszköz és adatbázis a HitPredict [López és mtsai 2015], mely a szerkezetből (Pfam domének) adódó sajátosságok mellett a szekvencia homológiát és Gene Ontology (GO) [The Gene Ontology Consortium 2013] funkciókat vesz figyelembe a több forrásból származó kölcsönhatások pontozása során, melynek segítségével magas megbízhatóságú interaktómok letöltésére ad lehetőséget. Ez a módszer tartalmazza a GO [The Gene Ontology Consortium 2013] sejten belüli lokalizációra (*cellular component term*) vonatkozó szűrését is, mely limitáltan, de figyelembe veszi a fehérjék térbeli elrendeződését is.

3.2.4. Fehérje-fehérje interakciós adatbázisok bemutatása

A jelenleg elérhető számos fehérje-fehérje interakciós adatbázis közül dolgozatomban a leggyakrabban használt adatforrásokra fókuszálok.

Az IMEx konzorcium [Orchard és mtsai 2012] tagja a DIP (Database of Interacting Proteins) [Salwinski és mtsai 2004], mely 834 organizmusra tartalmaz folyamatosan frissülő jó minőségű kísérletes interakciós adatokat. Szintén a konzorcium tagja a MINT

[3] http://www.imexconsortium.org/

és IntAct adatbázisok összevonásával keletkezett MIntAct adatbázis [Orchard és mtsai 2014], mely kiterjedt kézi adatgyűjtéssel és ellenőrzéssel biztosítja a folyamatosan magas minőséget.

Elterjedten használt eszköz a BioGrid (Biological General Repository for Interaction Datasets) adatbázis [Chatr-Aryamontri és mtsai 2015], melynek részletes webes felülete [Winter és mtsai 2011] nagyban segíti a felhasználót az adatok böngészésében. A BioGrid fizikai kölcsönhatások mellett genetikai kapcsolatokat is tartalmaz, valamint listázza az interakció meghatározására alkalmazott kísérletes módszert is, így a kölcsönhatások ez alapján is szűrhetők. Közvetlen fizikális kapcsolatokat határoz meg a CCSB (Center for Cancer Systems Biology) adatforrás is, melynek emberi adatkészlete kiemelten minőségi adatokat tartalmaz [Rolland és mtsai 2014]. A hálózat forrása és homogenitása miatt kiválóan alkalmas az interaktóm topológiai elemzésére, azaz az egyes fehérjék közti kapcsolatok térképének elemzésére a proteóm szintjén.

A legnagyobb elérhető interakciós adatforrás a STRING [Szklarczyk és mtsai 2015], mely jelenleg 2031 modell organizmus 9,6 millió fehérjéjére és azok 184 millió kapcsolatára tartalmaz adatokat. Az adatbázis számos forrást integrál és direkt kapcsolatok mellett a funkcionális indirekt kapcsolatokat is tartalmazza. Hasonlóan összesíti az adatokat három organizmusra a ConsensusPathDB [Kamburov és mtsai 2013]. A többi említett adatbázissal ellentétben ez a forrás a direkt és indirekt fehérje-fehérje interakciók mellett genetikai, metabolikus, jelátviteli, gén szabályozási és gyógyszer-célpont elemeket is tartalmaz.

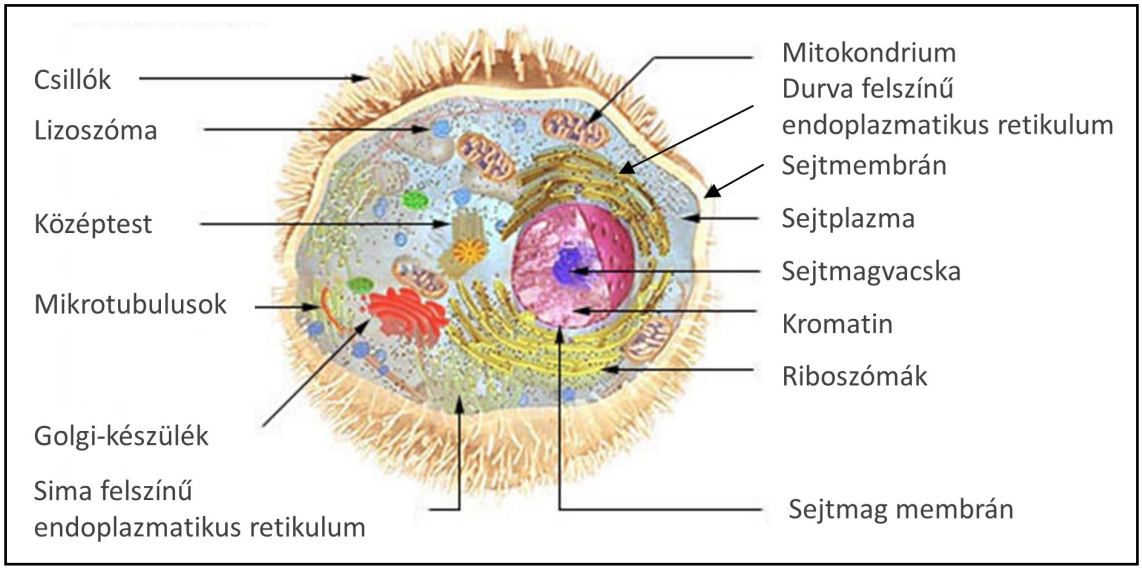
DOI:10.14753/SE.2018.2134

3.3. Szubcelluláris lokalizációs adatok jellemzése és forrásai

3.3.1. A szubcelluláris lokalizációs adatok általános jellemzése

A fehérjék funkciójának megítélésben kapcsolataik mellett kiemelt fontosságú a lokalizációjuk is, mely lehet a sejten belül vagy a sejten kívüli térben. Az eukarióta sejtek molekuláris mechanizmusait (kiemelkedően a jelátvitelt) alapvetően meghatározza a fehérjék és egyéb molekulák térbeli elkülönülése, ezzel biztosítva, hogy a szabályozás időben függetlenül mehessen végbe. Ez a szabályozás biztosítja a szélesebb körű válaszadási képességet, így például a stresszhelyzetre való válaszok kiterjedt varianciáját [García-Yagüe és mtsai 2013].

A fehérjék elhelyezkedését didaktikusan két fő részre bonthatjuk fel, a sejten belüli és sejten kívüli lokalizációra. A plazmamembránnal határolt sejten belüli térben membránnal elhatárolt funkcionális egységek, úgynevezett sejtszervecskék találhatóak (**5. ábra**).



***5. ábra**: **A sejt szerkezete**. Az ábra a sejt főbb szerkezeti és funkcionális elemeit mutatja be. Forrás: https://en.wikipedia.org/wiki/Cell_biology/*

Ide soroljuk például a sejtmagot, a mitokondriumokat, az endoplazmás retikulumot és a Golgi-készüléket. A sejten belül azonban a membránnal határolt sejtszervecskék, vagy organellumok mellett számos, nem membránnal határolt funkcionális egység is található,

mint például a riboszómák, a sejtváz vagy a proteaszóma. A sejten kívül található az extracelluláris mátrix állomány.

Ezen sejtes lokalizációk folyamatos dinamikus anyag és energia transzportot tartanak fenn, melynek kiemelkedően fontos szabályozó eleme a fehérjék transzportja [Elbaz és Schuldiner 2011]. A sejten belüli szerveződés organellumokon belül és azokon kívül is fontos szabályozási mechanizmusokra ad lehetőséget, mint például a makromolekulák besűrűsödésének jelensége (úgynevezett *macromolecular crowding*), mely jól körülhatárolt funkcióval rendelkező fehérjéknek adhat új funkciót pusztán a mikrokörnyezet megváltozásán keresztül [Lewitzky és mtsai 2012]. A fehérje-fehérje interakciós hálózat megértéséhez elengedhetetlen a fehérjék elhelyezkedésének figyelembevétele [Kumar és Ranganathan 2010, Inder és mtsai 2013].

3.3.2. A szubcelluláris lokalizációs adatok forrása és minősége

A fehérjék lokalizációját az aminosav szekvencia úgynevezett lokalizációs szignál régiói kódolják, melyek közül a legismertebb a sejtmagi lokalizációs szignál (NLS). A fehérje ezen szignáloknak mentén jut el a cél kompartmentbe, ahol betölti végső funkcióját. Ezt az információt használják ki a lokalizáció predikcióra alkalmas eszközök, melyek különböző módszerekkel (pl. neurális hálók, SVM) keresnek lokalizációs szignált a fehérjék aminosav szekvenciájának elemzésével [Yu és mtsai 2010]. Az így kapott prediktált lokalizációs adat mennyisége magas, így alkalmas proteóm szintű elemzésekre, azonban a megbízhatósága alacsony és a felbontása is limitált [Hu és mtsai 2009b]. Mivel a lokalizációs szignál a főbb sejtszervecskékbe irányítja a fehérjéket, így ezen algoritmusok a fehérjéket csak ezen főbb lokalizációk szerint tudják rendezni, részletesebb felbontásra nem alkalmasak.

A lokalizációs adatok minőségének növelése érdekében szükség van a kísérletes eredményekre. A legelterjedtebb és legmegbízhatóbb kísérletes lokalizáció meghatározó módszer a fehérjék immunfluoreszcenciával történő jelölésén alapszik. Az immunhisztokémia és annak elektronmikroszkópos változata, az immuncitokémia segítségével a fehérjék lokalizációját sejtekre és sejten belüli organellumokra specifikusan lehet vizsgálni [Al-Shibli és mtsai 2017]. A szubcelluláris lokalizáció sejten belüli rétegenkénti vizsgálatára ad lehetőséget a konfokális mikroszkópia, mely a fluorofór térbeli elhelyezkedését optikai szeleteléssel teszi megismerhetővé.

Ezen módszerek előnye, hogy a megbízható sejten belüli lokalizációs adat mellett a fehérjék mennyiségének becslésére is alkalmasak, így kompartmentre specifikus szubcelluláris lokalizációs adat nyerhető, mely az elemzés és modellezés újabb kapuit nyitja meg. Hátrányuk azonban, hogy ezen adatok túlnyomó részben mikroszkópos, kézi gyűjtésből származnak, melyek áteresztőképessége alacsony, így nem alkalmasak proteóm szintű elemzésre.

A proteóm szintű kísérletes lokalizációs adatokhoz tehát szükség van a folyamat automatizálására, melyre több módszer is létezik. A legtöbb kísérletes sejtes lokalizációt tartalmazó Human Protein Atlas [Pontén és mtsai 2011] immunhisztokémia és immunfluoreszcens képeit kézi kiértékelés mellett képfeldolgozó algoritmusokkal is támogatják [Li és mtsai 2012]. Ezen algoritmusok alkalmasak a kézi annotáció értékelésére, így jelezve azon eredményeket, melyeket érdemesnek tart a felülbírálásra. A már meglévő annotációk kiértékelésén túl a számítógépes feldolgozás alkalmas lehet új annotációk keresésére is, illetve különböző szövet, sejt és sejten belüli lokalizációs minták keresésére [Cornish és mtsai 2015].

Összességében a sejtes lokalizáció pusztán számítógépes predikcióval megvalósuló meghatározása, illetve a kísérletes eredmények algoritmusokkal történő értékelése együttesen alkalmas a fehérjék proteóm szintű térbeli annotálására.

3.3.3. A szubcelluláris lokalizációs adatbázisok bemutatása

A fehérje lokalizációs adatokhoz predikciós és kísérletes eljárások segítségével juthatunk, ennek megfelelően a lokalizációs adatbázisokat is feloszthatjuk prediktált, kísérletes vagy integrált adatokat tartalmazó forrásokra.

A szubcelluláris lokalizáció predikciójára számos eljárás létezik, ezek közül két eszközt emelnék ki, melyek használata elterjedt, vagy éppen a legújabb megoldásokat tükrözik. A Hum-mPLoc 3.0 a több mint egy évtizede aktív mPLoc legújabb verziója, emberi fehérjék lokalizációjának predikciójára alkalmas [Zhou és mtsai 2017]. A predikció egy Hidden Correlation Modeling (HCM) nevű eljárással történik, mely az aminosav szekvencia elemzésén túl többek között fehérjék GO [The Gene Ontology Consortium 2013] annotációját is figyelembe veszi. Az algoritmus 12 szubcelluláris lokalizáció szerint (középtest, sejtplazma, sejtváz, endoplazmatikus retikulum, endoszóma,

extracelluláris mátrix, Golgi-készülék, lizoszóma, mitokondrium, sejtmag, peroxiszóma, és plazma membrán) tudja szelektálni a FASTQ formátumban megadott fehérjéket. Ezen algoritmus is problémába ütközik azon fehérjék esetében, amelyeknek több lokalizációja is lehetséges. Ilyen multikompartment fehérjék predikciójára alkalmas a FUEL-mLOC szerver [Wan és mtsai 2017], mely rendszerezett és egységesített nevezéktannal nyújt megbízható predikciókat számos organizmusra nézve.

A jelenlegi legnagyobb elérhető kísérletes emberi fehérje lokalizációs adatbázis a Human Protein Atlas [Pontén és mtsai 2011], mely az immunfluoreszcens technikát konfokális mikroszkópiával kombinálva 200 nanométeres felbontású képeket alkot a sejt szerveződéséről. A képek elemzése révén az adatbázisban megközelítőleg 12000 fehérje lokalizációja érhető el metalokalizációkra[4] (sejtmag, sejtplazma és szekréciós útvonal) bontva. A fehérjék legnagyobb része a sejtmagban található, ezt követi a sejtplazma és a vezikulák, utóbbi magában foglalja a transzport vezikulákat és más kis membránnal határolt organellumokat, mint például az endoszómák vagy peroxiszómák. A fehérjék 51%-a több, mint egy lokalizációval rendelkezik, és 15% esetében figyeltek meg varianciát attól függően, hogy melyik sejttípusban vizsgálták őket.

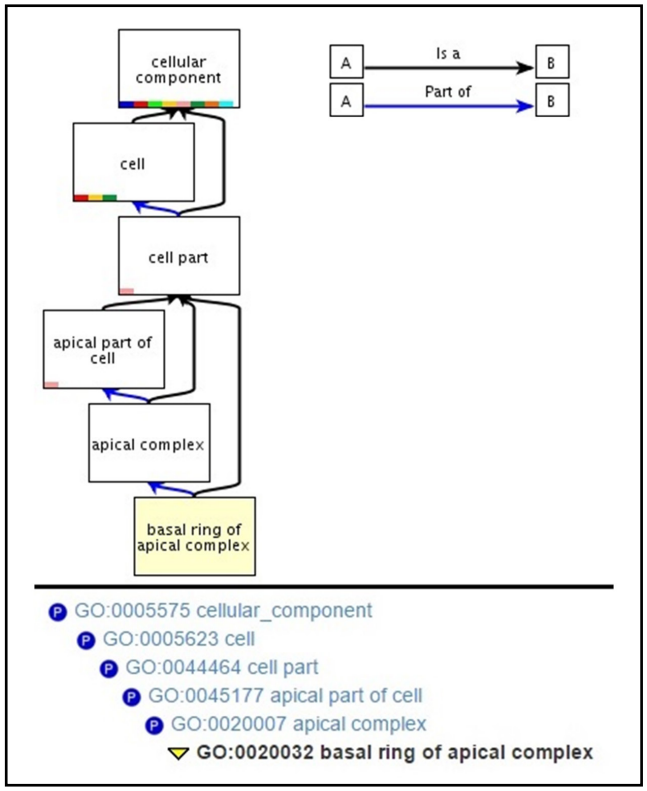
A Human Protein Atlas [Pontén és mtsai 2011] egyik hátránya, hogy a szecernált fehérjéken kívül nem ad információt a sejten kívüli lokalizációkról. A MatrixDB egy olyan [Launay és mtsai 2015] integrált fehérje-fehérje interakciós adatbázis, mely a fehérjék sejten kívüli kapcsolatainak elemzésére szolgál. Az ehhez szükséges lokalizációs adatokat a GO [The Gene Ontology Consortium 2013] és a UniProt [The UniProt Consortium 2017] források mellett a Matrisome Project [Naba és mtsai 2012] minőségi prediktált adataiból nyeri. Az oldalról letölthetők az extracelluláris, szecernált és membrán fehérje lokalizációra vonatkozó adatok.

Az átfogó kísérletes Human Protein Atlas [Pontén és mtsai 2011] adatbázis mellett az irodalom számtalan sporadikus kísérletben írja le egyes fehérjék szubcelluláris lokalizációját. Ezen adatok rendszerezése céljából két nagy adatbázis is létrejött, melyek egységes nevezéktant használnak. A UniProt [The UniProt Consortium 2017] egy a fehérjék rendszerzésére és annotálására létrejött tudástár. Számtalan adatot tartalmaz a

[4] http://www.proteinatlas.org/humancell/

fehérjékről, többek között kézzel gyűjtött kísérletes hivatkozásokon alapulva a fehérjék szubcelluláris lokalizációjáról[5] is, mely adatok könnyen kereshetők és letölthetők. Azonban csak részben tartalmazza a Human Protein Atlas [Pontén és mtsai 2011] adatait, miközben átfedése a GO [The Gene Ontology Consortium 2013] adatokkal magas.

Az eltérő adatbázisok a fehérjék, illetve a szubcelluláris lokalizációk tekintetében sokszor eltérő nevezéktant használnak. Ez utóbbi egységesítése és egy összefüggő hierarchiába történő rendezése azonban elengedhetetlen, amennyiben az igény a predikciós algoritmusok alacsony felbontásának és a kísérletek sokszor nagyon pontos lokalizációs megnevezéseinek összefésülése. Ennek megoldását képezi a GO [The Gene Ontology Consortium 2013] nevezéktana, mely kifejezetten a sejtes elemek nevezéktanát[6] gyűjti össze és rendszerezi (**6. ábra**).



6. ábra: A Gene Ontology lokalizációs fa felépítése az „apikális komplex bazális gyűrűje" lokalizáció példáján bemutatva. *(folytatás a következő oldalon ->)*

[5] http://www.uniprot.org/help/subcellular_location/
[6] http://www.geneontology.org/page/cellular-component-ontology-guidelines/

(-> folytatás az előző oldalról) A GO [The Gene Ontology Consortium 2013] annotáció hierarchikusan csoportosítja az egyes lokalizációs adatokat, így egy adott részletes felbontású lokalizációs információt a fa kibontásával, több lépésen keresztül érhetünk el. Az ábrán bemutatott példa az apikális komplex bazális gyűrűjének elhelyezkedését mutatja a fában, mely alulról felfele az apikális komplex – a sejt apikális része – sejt rész – sejt útvonalon érhető el.

Forrás: http://amigo.geneontology.org/amigo/term/GO:0020032/

Amellett, hogy az egyes lokalizációk helyet kapnak ebben a rendszerben, az adatok forrásának rendezése is megtörténik. Ennek segítségével eldönthető, hogy az egyes lokalizációs információk mennyire megbízhatóak, hány és milyen forrásból származnak (**1. táblázat**).

1. táblázat: A Gene Ontology lokalizációs adatainak eredete

Evidencia kód csoportja	Evidencia kódok
Kísérletes evidencia kódok	• Kísérleten alapuló (*Inferred from Experiment (EXP)*) • Közvetlen meghatározáson alapuló (*Inferred from Direct Assay (IDA)*) • Fizikai kölcsönhatáson alapuló (*Inferred from Physical Interaction (IPI)*) • Mutáns fenotípuson alapuló (*Inferred from Mutant Phenotype (IMP))* • Genetikai kölcsönhatáson alapuló (*Inferred from Genetic Interaction (IGI))* • Expressziós mintázaton alapuló (*Inferred from Expression Pattern (IEP)*)

1. táblázat: A Gene Ontology lokalizációs adatainak eredete (folytatás)

Evidencia kód csoportja	Evidencia kódok
Számítógépes elemzésen alapuló evidencia kódok	• Szekvencia vagy szerkezeti hasonlóságon alapuló (*Inferred from Sequence or structural Similarity (ISS)*) • Szekvencia ortológián alapuló (*Inferred from Sequence Orthology (ISO)*)
Számítógépes elemzésen alapuló evidencia kódok	• Szekvencia illesztésen alapuló (*Inferred from Sequence Alignment (ISA)*) • Szekvencia modellen alapuló (*Inferred from Sequence Model (ISM)*) • Genomikai összefüggésen alapuló (*Inferred from Genomic Context (IGC)*) • Az ősök biológiai vonatkozásain alapuló (*Inferred from Biological aspect of Ancestor (IBA)*) • A leszármazottak biológiai vonatkozásain alapuló (*Inferred from Biological aspect of Descendant (IBD)*) • Kulcs szekvencia elemeken alapuló (*Inferred from Key Residues (IKR)*) • Gyors divergencián alapuló (*Inferred from Rapid Divergence(IRD)*) • Ellenőrzött számítógépes analízisen alapuló (*Inferred from Reviewed Computational Analysis (RCA)*)
Szerzői közlésen alapuló evidencia kódok	• Nyomon követhető szerzői közlésen alapuló (*Traceable Author Statement (TAS)*) • Nem nyomon követhető szerzői közlésen alapuló (*Non-traceable Author Statement (NAS)*)

1. táblázat: A Gene Ontology lokalizációs adatainak eredete (folytatás)

Kurátori következtetésen alapuló evidencia kódok	• Kurátori következtetésen alapuló (*Inferred by Curator (IC)*) • Nincs elérhető biológiai adat (*No biological Data available (ND)*)
Automatikusan hozzárendelt evidencia kódok	• Elektronikus annotáción alapuló (*Inferred from Electronic Annotation (IEA)*)

A GO [The Gene Ontology Consortium 2013] nevezéktanához tartozó egyes bejegyzések eredete eltérő lehet, melyet evidencia kódok segítségével rendszerez az adatkészlet. Az egyes evidencia csoportok eltérő bizonyosságot szolgáltatnak az adott információ hitelességéről. Segítségével lehetőség nyílik az adatok mennyiségén túl azok minőségének alaposabb kiértékelésére is.
Forrás: http://www.geneontology.org/page/guide-go-evidence-codes/

Az adatok összesítésére vonatkozó törekvés magában foglalja az elérhető adatforrások minél szélesebb körű integrációját. A COMPARTMENTS [Binder és mtsai 2014] automatikus szövegbányászat segítségével összesíti az adatforrásokat, azokat folyamatosan frissíti, és egységes GO [The Gene Ontology Consortium 2013] nevezéktanra fordítja. Az ehhez hasonló integratív kezdeményezések szükségesek az átfogó és jó minőségű adatok minél egyszerűbb és folyamatos eléréséhez.

3.4. A sejten belüli szerveződés szerepe az egészséges sejtes jelátvitelben

3.4.1. A sejtkompartmentek szerepe, transzport folyamatok bemutatása

A fehérjék térbeli elhelyezkedése biztosítja, hogy időben és térben eltérő feladatokat tudjanak ellátni, mely fontos eleme a sejtekben zajló biokémiai folyamatoknak. Az eukarióta sejteket alapvetően az különíti el a prokarióta sejtektől, hogy belső membránrendszerrel rendelkeznek, melynek segítségével funkcionális egységeket, sejtkompartmenteket tudnak létrehozni. Ezen kompartmenteken belül lehetőség van a mikrokörnyezet változtatására, így például a mitokondrium vagy a peroxiszóma belső

összetétele jelentősen különbözik. Ennek köszönhetően hasonló metabolikus folyamatok eltérő utakon valósulhatnak meg a két organellumban [Demarquoy és Le Borgne 2015].

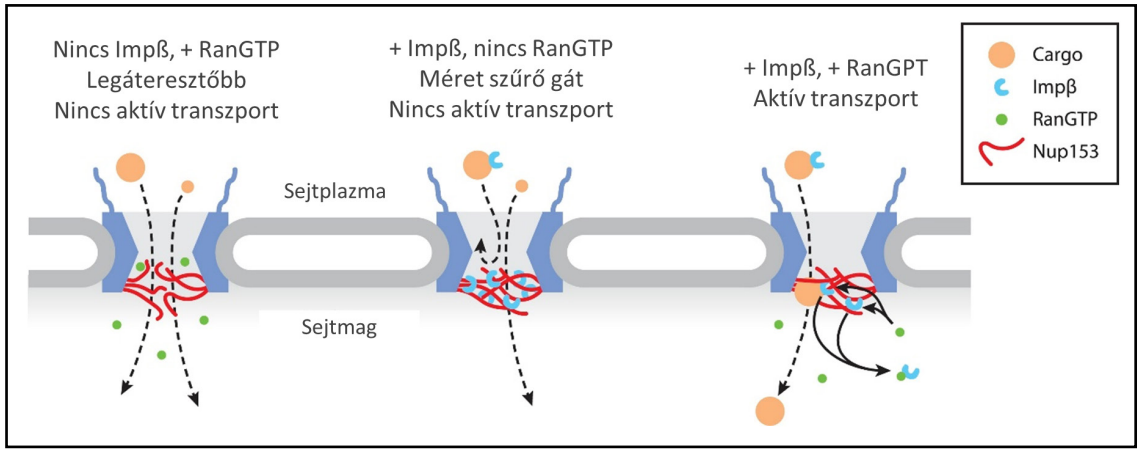
Az egyes kompartmentek között passzív, illetve aktív transzport folyamatok tartják fenn az anyag és energia transzportot. A passzív transzport történhet a koncentrációs grádiensnek megfelelően sima diffúzióval, mely folyamat alapvetően jellemző a kis méretű hidrofób, nem apoláros molekulákra. Nagyobb méretű, töltéssel rendelkező, hidrofil és poláros molekulák transzportja történhet facilitált diffúzióval, integráns membránfehérjék segítségével. Ez utóbbi transzport folyamat kis méretű peptidek, illetve fehérjék esetén is lehetséges. Korábban feltételezték, hogy a ~60kDa alatti molekulák sejtmagi pórus komplexen (NPC) keresztül passzívan is be tudnak jutni a sejtmagba [Rabut és mtsai 2004], azonban egyre több evidencia mutatja, hogy a kisméretű molekulák transzportja is lehet aktív, amennyiben fontos szabályozó feladatot látnak el.

A sejten belüli aktív transzport során a molekulák energia felhasználásával kerülnek át az egyik organellumból a másikba, mely sokszor a koncentráció grádiens ellenében történik. Ilyen aktív transzport történhet a sejtmag és sejtplazma között a sejtmagi pórus komplexen keresztül. Ez az egyik legnagyobb dinamikus fehérje komplex az emberi sejtekben, 25 nanométer széles és 30-50 különböző fehérjét tartalmaz, melyeket nukleoforinoknak hívunk. A legtöbb kis molekula, mint az egyes ionok vagy akár kisebb fehérjék, képesek passzívan átdiffundálni a pórusokon, míg a nagyobb molekulák csak aktív transzport révén juthatnak át a membrán egyik oldaláról a másikra. Ebben hordozó fehérjék segítenek (karioferinek vagy transzportinek), melyek specializálódtak a póruson keresztüli molekula transzportra [Macara 2001].

A karioferin hordozó fehérjéknek két típusát különböztetjük meg: az importin és exportin fehérjéket. Előbbiek a molekulák sejtmag irányú transzportját biztosítják, míg utóbbiak a sejtplazma irányába szállítják a molekulákat. A legtöbb importin heterodimer receptor, mely egy importin-alfa és egy importin-béta egységből áll. Előbbi ismeri fel a sejtmagi lokalizációs jelet (NLS), egyúttal adapter molekulája az importin-bétának, mely a sejtmagi pórus komplexszel való interakciót mediálja, miközben a transzporthoz

DOI:10.14753/SE.2018.2134

RanGTP-ből származó energiát használ fel [Lowe és mtsai 2015]. A komplex működését a **7. ábra** mutatja be.



***7. ábra**: **A sejtmagi pórus komplex működésének szabályozása.** A nukleáris pórus komplex működését az importin-béta és a Ran GTP-kötő fehérje jelenléte szabályozza. Kizárólag Ran jelenlétében kisebb és nagyobb méretű molekulák is képesek passzívan átjutni a membránon. Ezzel szemben egyedül az importin-béta jelenlétében a pórus átmérője szűkül, így csak a kisebb molekulák passzív transzportja lehetséges. Csak abban az esetben jöhet létre aktív transzport, amennyiben mindkét molekula jelen van. Forrás: [Lowe és mtsai 2015]*

Érdekes megfigyelés, hogy a sejtmagi pórus komplex a kor előrehaladtával szerkezetében megváltozik, és azon fehérjéket is átengedi, melyeket korábban nem, így a sejtmag és sejtplazma fehérje állománya a fiatal sejtekben megfigyelthez képest nagyobb mértékben keveredhet [D'Angelo és mtsai 2009]. A sejtmagi pórus komplexitását mutatja működésének allosztérikus szabályozása is, mely szintén alkalmas a pórus, és ezzel az áthaladó molekulák méretének szabályozására [Koh és Blobel 2015].

A sejtmag és sejtplazma közti fehérje áthelyeződés mellett a fehérjék transzportjának kiemelt iránya az endoplazmás retikulum (ER), mint a szekréciós útvonal első eleme. Az ER-be történő transzport a 'translocon' nevű fehérjeszerkezeten keresztül zajlik. A translocon egy fehérje komplex, melyet az ER membránjának tagjai alkotnak. A

riboszómákhoz történő dinamikus kötődés, illetve a további fehérjékhez való kapcsolódás (például a BiP) lehetővé teszi az ER membrán permeabilitásának szabályozását [Johnson és van Waes 1999]. A translocon-on keresztül történő áthelyeződést dajkafehérjék segítik mind a sejtplazmai, mind az ER oldalon, biztosítva a megfelelő konformációt az áthelyeződés során, illetve a végleges konformáció elnyerését az ER lumenben.

3.4.2. A fehérje transzlokáció rendszerszintű definíciója

Az említett fehérjékre jellemző két példa folyamatot, a sejtplazma- sejtmag és sejtplazma-ER irányú kompartmentek közti áthelyeződést transzlokációnak nevezzük. A transzlokáció lehet ko-transzlációs vagy poszttranszlációs. Előbbi esetre példa a sejtplazma-ER irány, ahol a riboszómákon képződő fehérjék az átolvasás során kerülnek az ER lumenébe [Nyathi és mtsai 2013], míg utóbbira tipikus példa a sejtplazma-sejtmag irányú áthelyeződés, amely létrejöhet a sejtplazma-ER viszonylatban is [Johnson és mtsai 2013].

A fehérje transzlokáció definíciója azonban nem egységes az irodalomban, és sokszor eltérő feltételek mentén határozzák meg a fogalmat. A dolgozatban tárgyalt transzlokáció fogalmát rendszerbiológiai szempontból közelítjük meg.

A fehérje transzlokáció fogalmát olyan rendszerbiológiai jelenségként határozzuk meg, ahol a fehérjék poszttranszlációs formában szabályozott módon helyeződnek át a sejtkompartmentek között. Transzlokáció során a fehérje kölcsönható partnerei és betöltött funkciója változik.

Az alap definíciót tudományos diákköröseimmel, Dobronyi Leventével és Mendik Péterrel a következők szerint pontosítottuk [Dobronyi és mtsai 2016, Mendik és mtsai 2017]:

- A sejtciklus ’M’ fázisában a szubcelluláris membránok részben dezintegrálódhatnak és újraszerveződhetnek, az ilyenkor fellépő fehérje áthelyeződés csak limitáltan feleltethető meg transzlokációnak a fenti definíció szerint.
- Jelző peptidek vagy jelző szekvencia szakaszok irányítják a fehérjéket a megfelelő lokalizációjukba. Mivel a fehérjék szintézise során történő ko-transzlokáció a

fehérjék funkcióját és kölcsönható partnereit nem változtatja meg, így ezt a folyamatot kizárjuk a rendszerbiológiai meghatározásból.

- A transzlációs folyamat után a fehérjék eloszlanak a sejtes kompartmentekben. A poszttranszlációs szállítási folyamatot, mely a fehérjék funkciójának betöltését segíti végső lokalizációjukban, szintén nem tekintjük transzlokációnak.
- Egyes sejten belüli folyamatok szintén megváltoztathatják a fehérjék sejten belüli lokalizációját, azonban ezeket sem tekintjük rendszerbiológiai szempontból transzlokációnak. Ezen folyamatok a következők: fehérje degradáció, fehérje kifejeződés csökkenés és a fehérjék passzív diffúziója.

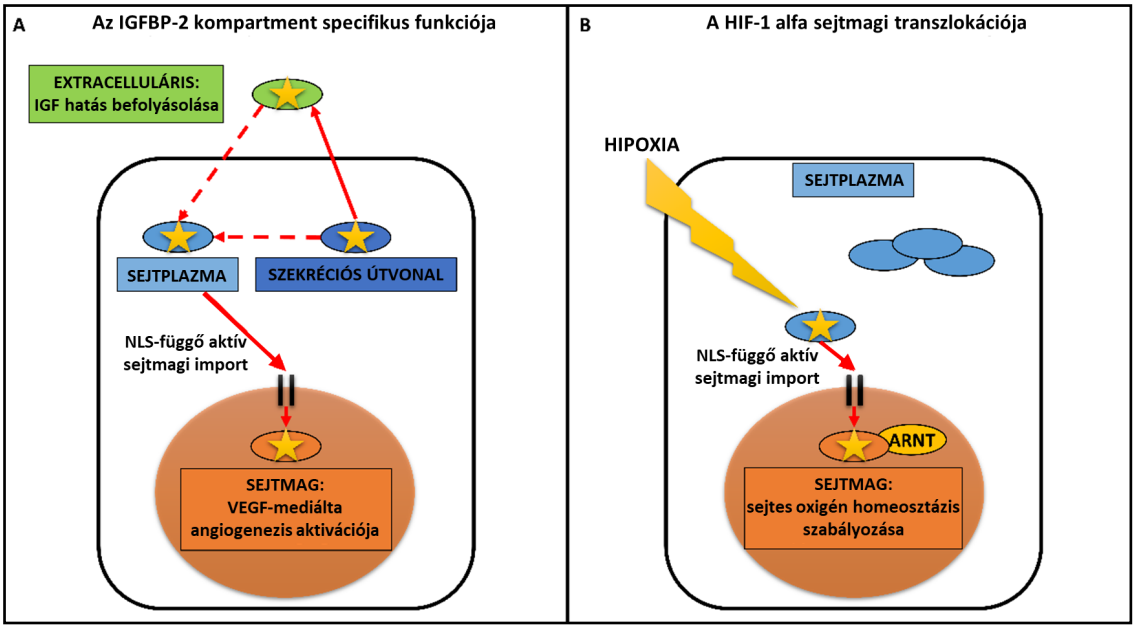
A fehérje transzlokáció rendszerbiológiai definíciója tehát limitálja, hogy mely fehérjéket vizsgáljuk, ezzel segítve elő a konkrét kérdésfeltevést és vizsgálati lehetőséget.

3.4.3. A fehérjék térbeli elhelyezkedésének funkcionális szerepe

Számos fehérje rendelkezik több mint egy szubcelluláris lokalizációval, mely térbeli elkülönülés segíti a biológiai folyamatok pontos szabályozását. A jelátviteli útvonalak kompartment szintű regulációjának szemléletes példája a transzkripciós faktorok sejtmagi transzlokáció által mediált aktiválódása [Hao és O'Shea 2012].

A **8. ábra** két transzlokálódó fehérjét mutat be, melyek funkciója az egészséges sejtekben lokalizációtól függő, és az egyensúly felborulása patológiás következményekkel járhat.

DOI:10.14753/SE.2018.2134



8. ábra: A transzlokáció funkcionális hatásának bemutatása az IGFBP2 és a HIF1A példáján. *Az inzulin-szerű növekedési faktor-kötő fehérje 2 (IGFBP-2) (**A panel**) domináns lokalizációja az extracelluláris tér, ahol fontos szabályozó elemét képezi az inzulin növekedési faktor (IGF) jelátvitelének, növelve az IGF féléletidejét [Firth és Baxter 2002]. A fehérje azonban transzkripciós faktorként is tud viselkedni, NLS-függő importin mediált sejtmagi transzport segítségével a sejtmagba kerülve a vaszkuláris endotheliális növekedési faktor (VEGF) kifejeződését segítve az érújdonképződést aktiválja [Azar és mtsai 2014]. A folyamat segíti a sejtek tápanyaghoz jutását akkor is, amikor kevesebb növekedést serkentő faktorhoz jutnak hozzá. Egy másik fontos példa a hipoxia indukálta faktor (HIF) alfa (**B panel**), mely a sejtplazmából a sejtmagba áthelyeződve a sejtmagban szintén transzkripciós faktorként viselkedik, ahol a sejt oxigén homeosztázisát szabályozza [Semenza 2009]. Az áthelyeződés után heterodimert alkot az aril hidrokarbon receptor sejtmagi transzlokátor (ARNT) fehérjével, mely heterodimerben ko-faktorként stabilizálja a sejtmagi aktivitást, azonban nem elengedhetetlen a transzlokációs folyamathoz [Kietzmann és mtsai 2016]. Forrás: [Veres és mtsai 2015]*

3.5. Jelátviteli sajátosságok a daganatos sejtekben a sejt organellumok szintjén

3.5.1. A szubcelluláris lokalizáció funkcionális szerepe a daganatos jelátvitelben

Számos fehérje rendelkezik több mint egy jól elkülönülő funkcióval. Azon fehérjéket, melyek több autonóm módon elkülönülő, gyakran össze nem függő funkcióval rendelkeznek, mely funkciók nem gén fúzióhoz, az RNS alternatív vágódásához vagy proteolítikus hasításhoz köthetők, multifunkcionális (*moonlighting*) fehérjéknek nevezzük [Tompa és mtsai 2005, Huberts és van der Klei 2010]. Ezen fehérjék a különböző feladatok betöltése céljából eltérő biokémiai funkciókkal rendelkeznek. A multifunkcionális fehérjék rendszerszintű feltérképezése és annotálása még ugyan várat magára, de a MoonProt Database [Mani és mtsai 2015] már most is ~300 ilyen fehérjét tartalmaz, melyek közül ~50 humán.

Ezen multifunkcionális fehérjék sokszor transzlokáció útján változtatnak lokalizációt, így képesek betölteni eltérő funkciójukat. A fehérjék sejten belüli elhelyezkedésének megzavarása, a transzlokációs folyamatok egyensúlyának felborulása az egészséges sejtviselkedést patológiás irányba mozdíthatja el, ezzel potenciálisan hozzájárulva egyéb patológiás események felerősítéséhez vagy fenntartásához.

A RAS-MAPK-ERK pálya az egyik legfontosabb növekedéssel, sejt túléléssel, sejt differenciációval összefüggő központi jelátviteli útvonal. Az útvonal valamilyen szintű normálistól eltérő működése minden daganatos betegségben megfigyelhető, így például az útvonal bemeneteként szolgáló receptor tirozin kinázok fokozott aktivációja, mint például az epidermális növekedési faktor receptor (EGFR) [Scaltriti és Baselga 2006], vagy az útvonal fontos belépő szabályozó elemének, a Ras GTP-áz fehérjének a gyakori mutációja [Prior és mtsai 2012]. A kináz kaszkád utolsó sejtplazmai lépése az extracelluláris szignál-regulált kináz (ERK1/2) foszforilációja, és önálló vagy hetero-, illetve homodimer formában sejtmagba történő transzlokációja [Lidke és mtsai 2010]. A RAS-MAPK-ERK kaszkád komponenseinek, kiemelten az ERK2-nek a szubcelluláris lokalizáció szintű szabályozása meghatározó fontosságú a daganatos jelátvitelben [Plotnikov és mtsai 2011].

3.5.2. Az ERK fehérjék lokalizáció-specifikus funkciója

Az ERK2 tipikus multifunkcionális fehérje, eltérő sejtplazmai és sejtmagi funkcióval és biokémiai aktivitással. Sejtplazmai lokalizációban szerin/treonin kinázként funkcionál, és a sejtplazmai kölcsönható partnereket, például más transzkripciós faktorokat, apoptózis szabályozó fehérjéket, ion csatornákat és receptorokat foszforilál. Fontos sejtplazmai negatív szabályozó kör, ahogy az ERK2 foszforilálja az őt közvetve vagy közvetetten aktiváló Raf és MEK1 fehérjéket, melyek ezáltal gátlódnak, ezzel visszahatva saját sejtmagi transzkripcionális aktivitásának csökkentésére [Dhillon és mtsai 2007]. Az ERK2 dimerizációja elengedhetetlen ahhoz, hogy betölthesse fontos sejtplazmai funkcióit, mivel az ERK2 dimer és az állvány fehérjék kapcsolódása kell ahhoz, hogy sejtplazmai partnereivel kapcsolatba tudjon lépni [Casar és mtsai 2009]. Ennek a dimerizációnak a megakadályozása önmagában is elégséges a sejtosztódás, transzformáció és tumor kialakulás enyhítéséhez.

Mint látjuk, a sejtplazmában az ERK2 fehérje foszforilációs partnerein keresztül negatív és pozitív szabályozó körök résztvevője, funkciója pedig szabályozható a dimer képződésen keresztül. Az ERK2 fehérje sejtmagi áthelyeződése nem szokásos NLS mediált folyamat, hanem sejtmagi transzlokációs szekvencián (NTS) alapul, mely a kináz inzert domén része. A domén foszforilációja segíti elő, hogy a fehérje kapcsolódni tudjon az importin molekulához, és bekerülhessen a sejtmagba [Zehorai és mtsai 2010]. A sejtmagba történő áthelyeződés után transzkripcionális represszor funkciót tölt be a [GS]AAA[GC] konszenzus szekvenciához kötődve, mely funkció teljesen független a sejtplazmai kináz aktivitástól. A sejtmagban számos gén promóteréhez kötődik, többek között az interferon gamma-indukálta gének kifejeződését csökkenti [Hu és mtsai 2009a], mely az ERK2 fokozott aktivációja esetén az interferon által mediált tumor ellenes aktivitást csökkenti [Parker és mtsai 2016].

3.5.3. Az ERK szerepe a daganatok progressziójában

A tradicionális klonális szelekció alapú daganatos iniciáció és progresszió elméletben az agresszív klónok kiválogatódása, és ezzel az invazív és áttétet képző daganatos stádium elérése egy lineáris folyamat, az idő előrehaladtával esélye növekszik. Számos próbálkozás történt már arra vonatkozóan, hogy meghatározzák, mely gének

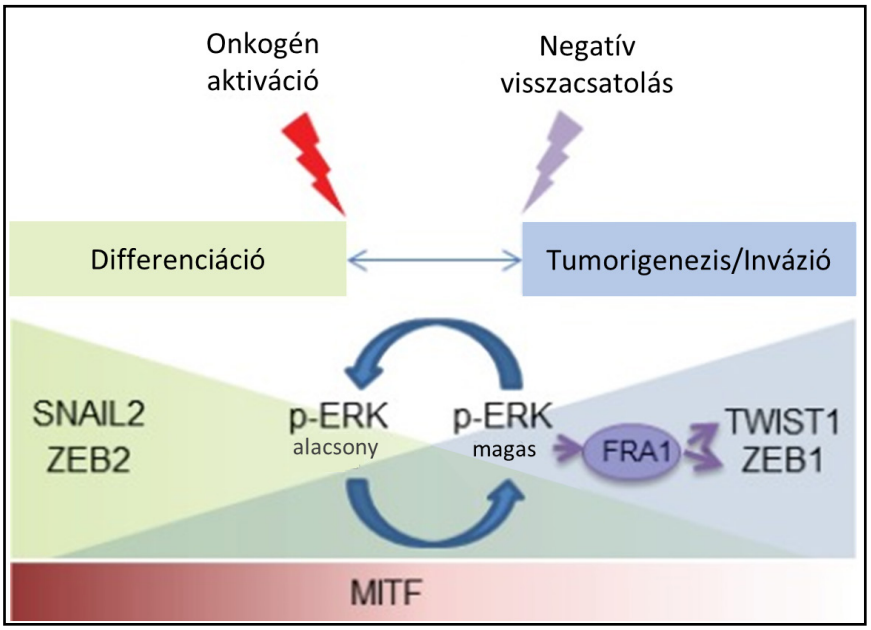
aktiválódnak és okoznak specifikusan áttétképzési hajlamot a késői stádiumú tumorokban, azonban ezidáig nem azonosítottak teljes bizonyossággal kizárólag a metasztatikus folyamat inicializálásával összefüggő géneket. Az új megközelítés szerint a tumorok progressziója és a metasztatikus aktivitás kéz a kézben jár. Azonos útvonalak szabályozzák, mely útvonalak finom szabályozása okozhat invazívabb, vagy éppen benignusabb tumor fenotípusokat [Klein 2010].

Az epitheliális-mezenchimális átmenet (EMT) során visszafordítható módon aktiválódik az embrionális genetikus program. Számos transzkripciós faktor (TF) irányítja a folyamatot, mint például a SNAL2, TWIST1, ZEB1 vagy ZEB2. Az EMT folyamata eredetileg a sejtek gyorsabb mozgását és ezzel az invazív képesség növekedését jelentette, azonban időközben kiderült, hogy az EMT-TF program a sejtek osztódását, túlélését, gyógyszerekkel szemben mutatott rezisztenciáját, tumorgenitását és őssejt-szerű viselkedését is szabályozza [Thiery és mtsai 2009].

Az EMT-TF program komplex szabályozása tehát rámutat az EMT és a fő onkogenikus útvonalak összefüggéseire, melynek szemléletes példája az EMT-TF hálózat és az egyik fő tumor-iniciáló útvonal, a RAS-MAPK-ERK modul kölcsönhatása melanómában [Caramel és mtsai 2013]. A neurális eredetű melanocita sejtekben az embrionális sejtpopulációt az EMT útvonal szabályozza, normál sejtekben SNAIL2 és ZEB2 pozitivitással, ZEB1 és TWIST1 negativitással. A RAS-MAPK-ERK jelátvitel aktivációja megváltoztatja a fehérjék aktivációját, mely a ZEB1 és TWIST1 kifejeződésének növekedéshez vezet. Ez a két fehérje pedig szerepet játszik többek között a BRAF mutált melanoma sejtek invazivitásának növekedésében [Tulchinsky és mtsai 2014], illetve rezisztencia kialakulásában tüdő daganatos betegekben [Chiu és mtsai 2017].

DOI:10.14753/SE.2018.2134

Érdekes, hogy míg a melanómás sejtek szinte kivétel nélkül tartalmaznak RAS-MAPK-ERK útvonalat aktiváló mutáció(ka)t, addig a korábban bemutatott multifunkcionális ERK fehérje foszforilációjának mértéke nem egységes. Amennyiben az ERK jobban foszforilált, rosszabb túlélést figyeltek meg, míg az alacsonyan foszforilált ERK jobb túléléssel párosult. A **9. ábra** az ERK foszforilációjának mértékét, és annak az eltérő tumoros viselkedéssel való összefüggését mutatja. Az ERK foszforilációjának mértéke egy kényes egyensúlyt állít be a sejtek osztódása és differenciálódása, valamint invazivitása és áttétképződési hajlama között, melyben többek között szerepet játszik a MITF fehérje [Hoek és Goding 2010], és a FRA1 is [Tam és mtsai 2013].



***9. ábra: Az ERK szerepe az EMT transzkripcionális szabályozásában**. Az ERK alacsony foszforilációja együtt jár a SNAIL2 és ZEB2 pozitivitással, mely alapvetően az embrionális program proliferációs és differenciációs folyamataira jellemző. Onkogén stimulus hatására az ERK foszforilálódik, és a magasan foszforilált ERK a FRA1-en keresztül a TWIST1 és ZEB1 kifejeződését fokozza, mely invazív fenotípushoz vezet. A proliferatív és invazív viselkedést promótáló ERK foszforilációjának mértéke egy kényes egyensúly eredménye, melyben kulcs szerepe van a MITF fehérjének. Forrás: [Tulchinsky és mtsai 2014]*

DOI:10.14753/SE.2018.2134

A proliferációval és invazivitással összefüggő, eltérő szabályozások egyensúlya kulcsfontosságú tehát a daganatok malignitásának megítélésében. A daganatokban lassan proliferáló, ellenben igen malignus sejtek jelenléte az egyik oka lehet a konvencionális antiproliferatív terápiákkal szembeni limitált érzékenységnek vagy a kialakuló rezisztenciának, melyre az EMT-TF és a konvencionális onkogén útvonalakat (mint a RAS-MAPK-ERK) célzó két lépcsős terápia lehet a megoldás [Sáez-Ayala és mtsai 2013].

Összefoglalva elmondható, hogy az ERK fehérje jelátvitelben betöltött szerepe alapvetően szubcelluláris lokalizáció függő, és finoman szabályozott a fehérje foszforiláltságának a segítségével. Megfigyelhető, hogy ez a szabályozás szorosan összefügg a daganatos sejtek malignitásának mértékével is, ezzel bemutatva a szubcelluláris lokalizáció fehérje funkcióban és ezzel jelátvitelben betöltött általános és esszenciális szerepét, illetve jelentőségét a daganatos progresszió megítélésében.

4. Célkitűzések

Doktori munkám célkitűzései az alábbiak voltak.

1. **Kompartment-specifikus fehérje-fehérje kölcsönhatási adatbázis létrehozása:** A dolgozatban bemutatott ComPPI egy olyan adatbázis és webes felület, mely számos forrásból integrált, jó minőségű kompartment specifikus fehérje-fehérje interakciós adatokat tartalmaz 4 speciesz (élesztő, fonálféreg, gyümölcslégy és ember) 125.757 fehérjéjének 791.059 kölcsönhatására. Az adatbázis mind a szubcelluláris lokalizációk, mind pedig az azonos kompartmentumban létrejövő kölcsönhatások valószínűségét külön értékkel jósolja [Veres és mtsai 2015].

2. **Kézi és számítógépes adatgyűjtésen alapuló adatbázis létrehozása transzlokálódó fehérjék gyűjtésére és predikciójára:** A Translocatome egy fejlesztés alatt álló adatbázis és interaktív webes felület, mely a ComPPI szűrésén és kézi adatgyűjtésen alapulva számos, eddig máshol nem összegzett transzlokálódó fehérjéről tartalmaz részletes adatokat a szubcelluláris lokalizációval kapcsolatos funkcióról, mely adatok alkalmasak egy gépi algoritmus tanítására új transzlokálódó fehérjék jóslása céljából [Dobronyi és mtsai 2016, Mendik és mtsai 2017].

3. **A daganatos iniciáció és progresszió jobb megértése hálózatbiológiai megfigyelések segítségével:** A szubcelluláris lokalizációs viszonyok megváltozása szerepet játszik a daganatok kialakulásában és progressziójában. A doktori munkám során kifejlesztett adatbázisok egyes fehérje és interaktóm szintű adatai alapján lehetőség nyílt a daganatos jelátvitellel összefüggő fehérjék rendszerszintű vizsgálatára, amely segített a rákos folyamatok kialakulásával kapcsolatos hipotézisek kidolgozásában és régebbi valamint újonnan közölt epidemiológiai adatok értelmezésében [Gyurkó és mtsai 2013, Csermely és mtsai 2015, Adami és mtsai 2017].

DOI:10.14753/SE.2018.2134

5. Módszerek

5.1. A ComPPI adatbázis létrehozása során használt források és módszerek

5.1.1. Modell organizmusok kiválasztása

Az adatbázis megalkotásának első lépése során kiválasztottuk azon modell organizmusokat, melyek bioinformatikai elemzése hasznos, új biológiai hipotézisek felállítására adhat lehetőséget. Kiemelt szempont volt, hogy a különböző rendszertani szintről származó modelleken keresztül az evolúciós összehasonlító vizsgálatokra is lehetőség nyíljon, mint például az interakciós ortológok mintájára (úgynevezett interolog [Huang és mtsai 2007]) lokalizációs ortológok (úgynevezett localog [Veres és mtsai 2015]) keresésére. Technikailag elvárás volt, hogy nagy mennyiségű és jó minőségű adat álljon rendelkezésre a kiválasztott modellekre vonatkozóan.

A ComPPI adatbázis ezen megfontolások alapján a következő négy fajra tartalmaz adatokat:

- *Saccharomyces cerevisiae* - élesztőgomba
- *Caenorhabditis elegans* - fonálféreg
- *Drosophila melanogaster* - ecetmuslica
- *Homo sapiens* - ember

5.1.2. A fehérje lokalizációs adatok forrása

A ComPPI adatbázis (http://comppi.linkgroup.hu/, [Veres és mtsai 2015]) 9 fehérje-fehérje kölcsönhatási adatbázist és 8 szubcelluláris lokalizációs adatbázist összegez. Az adatforrások integrációja segíti növelni az adatok mennyiségét és javítani azok minőségét, mivel csökkenti az egyes adatbázisok alacsony átfedéséből fakadó adatvesztés lehetőségét. A felhasznált adatforrások szabadon elérhetőek az akadémiai kutatások számára, és jellemző rájuk a proteóm szintű adat letöltési lehetőség.

A szubcelluláris lokalizációs adatok származhatnak kísérletes, prediktált vagy ismeretlen forrásból. A szubcelluláris lokalizációs adatok rendszerszintű átfogó összehasonlítása hiányzik az irodalomból, annak ellenére, hogy az adatok mennyisége és minősége

jelentősen eltér. Erre példa a **2. táblázat**ban bemutatott forrásadatbázisok integrációja során megfigyelhető, több mint kétszeres adat nyerés.

2. táblázat: Az emberi prediktált szubcelluláris lokalizációs forrás adatbázisok alacsony átfedése

Forrás adatbázis neve	**Lokalizációk száma**
BaCelLo [Pierleoni és mtsai 2006]	33 551
eSLDB [Pierleoni és mtsai 2007]	37 348
PA-GOSUB [Lu és mtsai 2005]	30 733
pTARGET [Guda 2006]	42 480
A források átlaga:	**36 028**
Összesítés után:	**84 635**

Az adatkészletek keresése során megvizsgált négy gyakran használt prediktált szubcelluláris lokalizációkat tartalmazó adatbázis átfedése alacsony, összesítésük után a legnagyobb egyedi adatkészletnél kétszer nagyobb adatbázishoz jutunk.

A forrásadatok alacsony átfedése indokolta a különböző források összevonását, ami megkülönbözteti a ComPPI adatbázist számos más szubcelluláris adatot használó adatbázistól, melyek csak a Gene Ontology [The Gene Ontology Consortium 2013] adatait használják fel. Az adatforrások kiválasztása és válogatása után a **3. táblázat**-ban felsorolt adatbázisokat használtuk fel.

3. táblázat: A ComPPI szubcelluláris lokalizációs forrás adatbázisai

Adatbázis neve	**Verzió / letöltés időpontja**	**Hivatkozás**
Saccharomyces cerevisiae		
eSLDB	dátum: 2008 április	[Pierleoni és mtsai 2007]
Gene Ontology	letöltés időpontja: 2014 június	[The Gene Ontology Consortium 2013]
Organelle DB	dátum: 2007	[Wiwatwattana és mtsai 2007]

3. táblázat: A ComPPI szubcelluláris lokalizációs forrás adatbázisai (folytatás)

Adatbázis neve	Verzió / letöltés időpontja	Hivatkozás
Saccharomyces cerevisiae		
PA-GOSUB	verzió: 2.5	[Lu és mtsai 2005]
Caenorhabditis elegans		
eSLDB	dátum: 2008 április	[Pierleoni és mtsai 2007]
Gene Ontology	letöltés időpontja: 2014 június	[The Gene Ontology Consortium 2013]
Organelle DB	dátum: 2007	[Wiwatwattana és mtsai 2007]
PA-GOSUB	verzió: 2.5	[Lu és mtsai 2005]
Drosophila melanogaster		
eSLDB	dátum: 2008 április	[Pierleoni és mtsai 2007]
Gene Ontology	letöltés időpontja: 2014 június	[The Gene Ontology Consortium 2013]
Organelle DB	dátum: 2007	[Wiwatwattana és mtsai 2007]
PA-GOSUB	verzió: 2.5	[Lu és mtsai 2005]
Homo sapiens		
eSLDB	dátum: 2008 április	[Pierleoni és mtsai 2007]
Gene Ontology	letöltés időpontja: 2014 június	[The Gene Ontology Consortium 2013]
Human Proteinpedia	verzió: 2.0	[Kandasamy és mtsai 2009]
LOCATE	dátum: 2008. november 21.	[Sprenger és mtsai 2008]
MatrixDB	dátum: 2014. március 20.	[Chautard és mtsai 2011]
Organelle DB	dátum: 2007	[Wiwatwattana és mtsai 2007]
PA-GOSUB	verzió: 2.5	[Lu és mtsai 2005]
Human Protein Atlas	verzió: 12 (2013. december 05.)	[Pontén és mtsai 2011]

A táblázat tartalmazza az egyes fajokhoz felhasznált adatbázisok neveit, az integrált verzió azonosítóját, illetve a hivatkozásokat. Az adatbázisok számát tekintve kiemelten több magas minőségű emberi adatokat tartalmazó adatkészlet érhető el. A ComPPI frissítése folyamatban van, hogy az elérhető legfrissebb adatokat tartalmazza.

5.1.3. A fehérje-fehérje kölcsönhatási adatok forrása

A fehérje-fehérje kölcsönhatási adatok forrása lehet kísérletes vagy bioinformatikai predikció, típusát tekintve pedig fizikai vagy genetikai kölcsönhatás. A ComPPI adatbázis kizárólag kísérletes forrásból származó fizikai kölcsönhatásokat tartalmaz, mely biztosítja a kapcsolatok funkcionális elemzésének lehetőségét. Az interakciók kísérletes forrása egyaránt lehet alacsony vagy magas áteresztőképességű kísérletes technika.

Egy átfogó tanulmány bemutatta, hogy a leggyakrabban használt fehérje-fehérje kölcsönhatási adatbázisok (BioGRID [Chatr-Aryamontri és mtsai 2015], BIND [Bader és mtsai 2003], DIP [Salwinski és mtsai 2004], IntAct [Kerrien és mtsai 2012], HPRD [Keshava és mtsai 2009], MINT [Licata és mtsai 2012]) átfedése meglepően alacsony, a hat vizsgált adatbázis mindegyikében szereplő kapcsolatok száma három [De Las Rivas és Fontanillo 2010]. Ez az alacsony átfedés indokolja a fehérje-fehérje kölcsönhatási adatok esetében is a források összesítésére vonatkozó tendenciát, és ezzel új metaadatbázisok létrehozását [Kamburov és mtsai 2013]. A **4. táblázat** bemutatja a ComPPI adatbázis felépítéséhez használt fehérje-fehérje interakciós adatbázisokat.

4. táblázat: A ComPPI fehérje-fehérje kölcsönhatási forrás adatbázisai

Adatbázis neve	Verzió / letöltés időpontja	Hivatkozás
Saccharomyces cerevisiae		
BioGRID	verzió: 3.2.111	[Chatr-Aryamontri és mtsai 2015]
CCSB	dátum: 2011. február	[Rolland és mtsai 2014]
DIP	dátum: 2014. január 17.	[Salwinski és mtsai 2004]
IntAct	dátum: 2012. június 11.	[Kerrien és mtsai 2012]
MINT	dátum: 2013. március 26.	[Licata és mtsai 2012]
Caenorhabditis elegans		
BioGRID	verzió: 3.2.111	[Chatr-Aryamontri és mtsai 2015]
CCSB	dátum: 2011. február	[Rolland és mtsai 2014]

4. táblázat: A ComPPI fehérje-fehérje kölcsönhatási forrás adatbázisai (folytatás)

Adatbázis neve	Verzió / letöltés időpontja	Hivatkozás
Caenorhabditis elegans		
DIP	dátum: 2014. január 17.	[Salwinski és mtsai 2004]
IntAct	dátum: 2012. június 11.	[Kerrien és mtsai 2012]
MINT	dátum: 2013. március 26.	[Licata és mtsai 2012]
Drosophila melanogaster		
BioGRID	verzió: 3.2.111	[Chatr-Aryamontri és mtsai 2015]
DIP	dátum: 2014. január 17.	[Salwinski és mtsai 2004]
DroID	dátum: 2014. január	[Murali és mtsai 2011]
IntAct	dátum: 2012. június 11.	[Kerrien és mtsai 2012]
MINT	dátum: 2013. március 26.	[Licata és mtsai 2012]
Homo sapiens		
BioGRID	verzió: 3.2.111	[Chatr-Aryamontri és mtsai 2015]
CCSB	dátum: 2011. február	[Rolland és mtsai 2014]
DIP	dátum: 2014. január 17.	[Salwinski és mtsai 2004]
HPRD	dátum: 2010. április 13.	[Keshava és mtsai 2009]
IntAct	dátum: 2012. június 11.	[Kerrien és mtsai 2012]
MatrixDB	dátum: 2012. augusztus 01.	[Launay és mtsai 2015]
MINT	dátum: 2013. március 26.	[Licata és mtsai 2012]
MIPS	dátum: 2004	[Pagel és mtsai 2005]

A táblázat tartalmazza az egyes fajokhoz felhasznált adatbázisok neveit, az integrált verzió azonosítóját, illetve a hivatkozásokat. Az adatbázisok számát tekintve kiemelten több magas minőségű emberi adatokat tartalmazó adatkészlet érhető el. A ComPPI frissítése folyamatban van, hogy az elérhető legfrissebb adatokat tartalmazza.

5.1.4. A fehérjék funkcionális elemzésére használt módszerek

A ComPPI adatbázis összesített és feldolgozott adatainak elemzése során az egyes fehérjék és kölcsönható partnereik biológiai funkcióit szubcelluláris lokalizációra specifikusan vizsgáltuk. Ehhez szükség volt a fehérjék biológiai folyamatokban betöltött szerepére vonatkozó adatok felkutatására, melyhez a Gene Ontology [The Gene Ontology

Consortium 2013] adatkészletét használtuk, melyet az AmiGO[7] webszerver alkalmazásával böngésztünk [Carbon és mtsai 2009].

A vizsgált fehérjékhez automatikusan rendeltük hozzá az egyes biológiai folyamatokra vonatkozó információkat, majd megvizsgáltuk, hogy a szubcelluláris lokalizációra specifikus kölcsönhatási hálózatban hogyan dúsul az információ, azaz mely biológiai folyamatok jelenléte a legvalószínűbb a szomszédsági hálózatban. Erre a feladatra az interaktómok megjelenítésére alkalmas Cytoscape hálózat megjelenítő és elemző program [Shannon és mtsai 2003] beépülő alkalmazását, a BiNGO[8]-t választottuk (3.0.2-es verzió) [Maere és mtsai 2005], mely alkalmas a megjelenített hálózat elemeinek közvetlen, interaktív vizsgálatára. A statisztikai elemzés eredménye megjeleníthető, valamint letölthető, ami elősegíti a további feldolgozást.

5.1.5. A fehérjék hálózatos megjelenítéséhez és elemzéséhez használt eszközök

Az interaktóm vizsgálata során elengedhetetlen a hálózat megjelenítése, hiszen a hálózatos megközelítés egyik előnye a vizuális leképezés, mely alapja lehet későbbi biológiai hipotézisek felállításának. A hálózatok megjelentésére két eszközt használtunk, melyek közül a Gephi[9] 0.8.2beta verzióját [Bastian és mtsai 2009] alkalmaztuk a nagyobb méretű hálózatok gyors és áttekintő megjelenítéséhez. A Cytoscape[10] [Shannon és mtsai 2003] a leggyakrabban használt eszköz biológiai hálózatok megjelenítésére és elemzésére, mely számos közösségi alapon fejlesztett beépülő modullal segíti a hálózatok széles körű elemzését. Az alkalmazás 3.1.0 verzióját használtuk az egyes fehérjék szubcelluláris lokalizáció specifikus kölcsönhatási hálózatának megjelenítésére, illetve az adatok elemzésére. A fokszám (mely megadja a hálózat egyes elemeivel kapcsolódó szomszédok számát) és a köztiség mérőszám (mely az egy hálózatos ponton áthaladó legrövidebb utak számát határozza meg) kiszámolásához az alkalmazásba beépített NetworkAnalyzer[11] eszközt használtuk.

[7] http://amigo.geneontology.org/amigo/
[8] http://apps.cytoscape.org/apps/bingo/
[9] https://gephi.org/
[10] http://cytoscape.org/index.html/
[11] http://apps.cytoscape.org/apps/networkanalyzer/

5.1.6. Statisztikai elemzéshez és adat vizualizációhoz használt módszerek

A ComPPI lokalizációs és interakciós megbízhatósági érték számítása és optimalizálása, valamint az adatok általános statisztikai elemzése során, illetve grafikonok generálásához az R programcsomagot[12] alkalmaztuk (3.1.0 verzió) [Ihaka és Gentleman 1996]. A könnyebb kezelhetőség érdekében az R kódot az RStudio[13] nevű felhasználóbarát alkalmazás segítségével írtuk és futtattuk. Az alapvető táblázatos műveleteket és statisztikai elemzéseket a Microsoft Excel 2013-as verziójával végeztük.

5.1.7. A ComPPI adatbázis és webes felület kialakításához használt módszerek

A ComPPI adatbázis felépítése hierarchikus, ahol a forrás adatbázisok egy univerzális bemeneten keresztül, Symfony[14] 2 PHP keretrendszerben megírt illesztőfelületek segítségével csatlakoznak az adatbázishoz. Az adatbázis maga MySQL[15] 5 Community Edition alapú, saját nevezéktannal. Ez a struktúra biztosítja a bemenő adatbázisok gyors frissítésének, illetve régi adatbázisok lecsatolásának vagy újak becsatolásának lehetőségét.

A webes felület szintén Symfony keretrendszerben készült, PHP5 nyelven íródott és nginx[16] HTTP szerveren alapul. A felület és az adatréteg egymással kétirányú kapcsolatban van. A lekért adatok a felületen megjeleníthetőek, illetve CSV szöveges vagy SQL adatbázis formátumban menthetőek.

További felhasznált eszközök az Ubuntu Linux 14.04[17] operációs rendszer, a git[18] verzió kontroll rendszer, a jQuery[19] JavaScript keretrendszer és a D3.js[20] JavaScript könyvtár (hálózatos vizualizációra), valamint a Python 3[21] programozási nyelv.

A ComPPI adatbázis és webes felület a felhasznált eszközökkel összhangban nyílt forráskódú, így a hozzáértők számára ellenőrizhető, átlátható és továbbfejleszthető. A

[12] https://cran.r-project.org/
[13] https://www.rstudio.com/
[14] http://symfony.com/
[15] http://www.mysql.com/
[16] http://nginx.org/
[17] http://ubuntu.com/
[18] http://git-scm.com/
[19] http://jquery.com/
[20] http://d3js.org/
[21] https://python.org/

webes felület kutató- és szerzőtársam, Gyurkó M. Dávid munkája, melynek részletes leírása doktori értekezésében olvasható [Gyurkó 2015].

5.2. A Translocatome adatbázis létrehozása során használt források és módszerek

5.2.1. A fehérje-fehérje kölcsönhatási, szubcelluláris lokalizációs adatok és a kézzel gyűjtött fehérjék forrásai

A Translocatome adatbázis fehérje-fehérje interakciós és szubcelluláris lokalizációs adatai a ComPPI[22] adatbázisból [Veres és mtsai 2015] származnak. Az adatkészlet SQL formátumban történő letöltése után csak azon kölcsönhatások kerülnek be a Translocatome adatbázisba, melyek kölcsönható partnereinek legalább egy kísérletes eredetű szubcelluláris lokalizációja van.

A transzlokálódó fehérjék kézi gyűjtése két módon történik. Egyik esetben a ComPPI adatkészletének alapján szűrünk olyan fehérjéket, melyek valószínűsíthetően transzlokálódnak, majd ezen példákra keresünk evidenciát az irodalomban. Második esetben közvetlenül az irodalomból, kézi gyűjtéssel illesztjük be a fehérjéket az adatkészletbe abban az esetben, ha az még nem tartalmazza őket.

Mindkét esetben a PubMed[23] és Google Scholar[24] szervereket használjuk a fehérjéket leíró cikkek keresésére. Az egyes fehérjék azonosításához, és további funkciók kereséséhez döntően a UniProt[25] [The UniProt Consortium 2017], NCBI Gene[26] és GeneCards[27] [Safran és mtsai 2010] oldalakat alkalmazzuk.

A Translocatome adatbázis két tudományos diákkörös hallgatómmal, Dobronyi Leventével és Mendik Péterrel végzett közös munka eredménye [Dobronyi és mtsai 2016, Mendik és mtsai 2017].

[22] http://comppi.linkgroup.hu/downloads/
[23] https://www.ncbi.nlm.nih.gov/pubmed/
[24] https://scholar.google.hu/
[25] http://www.uniprot.org/
[26] https://www.ncbi.nlm.nih.gov/gene/
[27] http://www.genecards.org/

5.2.2. A fehérjék biológiai folyamatokban betöltött szerepére vonatkozó adatok forrása és feldolgozása

A transzlokálódó fehérjékre vonatkozó kézi adatgyűjtés során az adatok egységesítése elengedhetetlen, így a fehérje funkciókat a legmegfelelőbb Gene Ontology funkcióra [The Gene Ontology Consortium 2013], míg az interakciós partnereket UniProt SwissProt nevezéktanra fordítjuk le [The UniProt Consortium 2017]. A lokalizációs adatokat a ComPPI [Veres és mtsai 2015] nevezéktanának megfelelően rögzítjük, a jelátviteli útvonalakat és a betegségneveket a KEGG adatforrás [Kanehisa és mtsai 2017] alapján rendszerezzük.

Az adatbázis fehérjéinek biológiai folyamatokban betöltött szerepének rendszerszintű megfeleltetését a ComPPI-hoz hasonlóan a Translocatome esetében is a Gene Ontology [The Gene Ontology Consortium 2013] nevezéktanát használva, az AmiGO[28] webszerver [Carbon és mtsai 2009] felhasználásával végezzük, mely adatok automatikusan frissíthetők új fehérjék beemelésének esetén.

5.2.3. A fehérjék hálózatos paramétereinek meghatározása, a hálózat ábrázolása

A transzlokálódó fehérjék predikciójára Hári Ferenc és Kerepesi Csaba közreműködésével egy gépi tanuló algoritmust fejlesztünk, melynek tanulási paraméterei közé tartoznak a fehérjék biológiai funkciói mellett azok hálózatos tulajdonságai is. Legfontosabb felhasznált jellemzők a fokszám, a köztiségi, illetve a hídság mérőszám. A fokszám az egyes hálózatos pontok szomszédjainak számát adja meg. A köztiségi mérőszám egy központiságot kifejező paraméter, mely az áthaladó bármely két pont közti legrövidebb utak számát reprezentálja. A hídság szintén egy központisági mérőszám, mely jellemzi, hogy egy adott pont milyen mértékben tartozik hálózatos modulok átfedő régiójába, összehasonlítva a hálózat többi elemével.

A fokszám és köztiségi mérőszám hálózatos paraméterek számolásához a Translocatome esetében egy automatikus módszert fejlesztettünk, mely a bekerülő fehérjékhez kiszámítja és hozzárendeli ezen paramétereket a nyílt forráskódú NetworkX[29] Python[30] csomag (1.11 verzió) segítségével. A hídsági központiság mérőszámát a

[28] http://amigo.geneontology.org/amigo/
[29] https://networkx.github.io/
[30] https://python.org/

munkacsoportunkban fejlesztett ModuLand [Szalay-Bekő és mtsai 2012] hálózatos modularizáló algoritmus segítségével számoltuk ki.

A Translocatome esetében is használjuk a Cytoscape [Shannon és mtsai 2003] hálózat megjelenítő programot, a NetworkAnalyzer[31] beépített eszközt kiemelten a kisebb elemzésekre, illetve a kutatócsoportunk által fejlesztett EntOpt[32] hálózat ábrázoló modult [Kovács és mtsai 2015] a hálózatok vizualizációjára.

5.2.4. A Translocatome adatbázis és webes felület kialakításához használt módszerek

A Translocatome adatbázis és webes felülete nyílt forráskódú (http://translocatome.linkgroup.hu/). Fejlesztése a dolgozat elkészítésekor folyamatban van, így a felsorolt nyílt forráskódú eszközök aktuálisan elérhető legfrissebb verzióján alapul. Az adatbázis MongoDB[33], a kézi adatgyűjtést segítő webes felület Ruby on Rails[34] alapú. A felhasználókat kiszolgáló adat böngészésre és letöltésre szolgáló felület Express[35] keretrendszerben készül, mely kiszolgálja a Node.js[36] és React[37] alapú fejlesztést. A gépi tanuláshoz a Python[38] programozási nyelven alapuló scikit-learn[39] csomagot használjuk.

[31] http://apps.cytoscape.org/apps/networkanalyzer/
[32] http://apps.cytoscape.org/apps/entoptlayout/
[33] https://www.mongodb.com/
[34] http://rubyonrails.org/
[35] https://expressjs.com/
[36] https://nodejs.org/en/
[37] https://facebook.github.io/react/
[38] https://python.org/
[39] http://scikit-learn.org/stable/

DOI:10.14753/SE.2018.2134

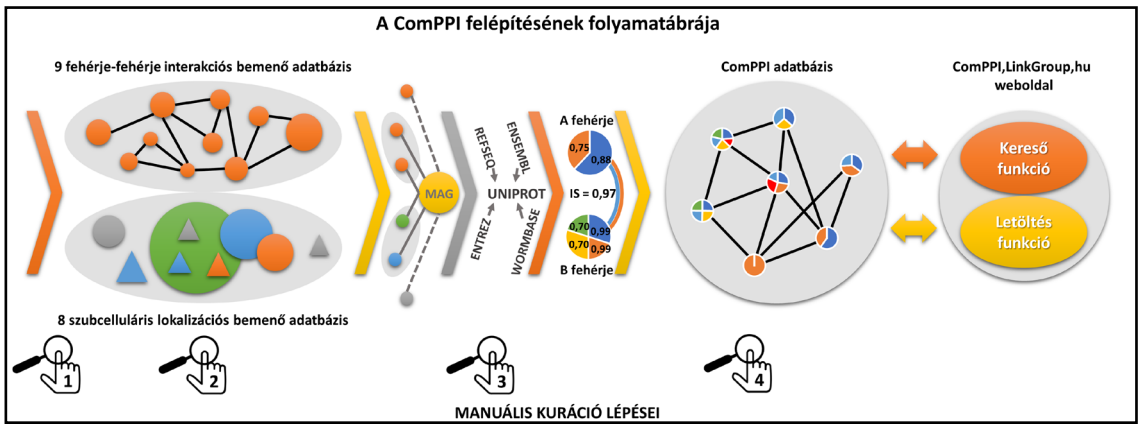
6. Eredmények

6.1. A ComPPI adatbázis általános bemutatása

6.1.1. Az adatbázis felépítési folyamatának sematikus bemutatása

A ComPPI adatbázis (http://comppi.linkgroup.hu/, [Veres és mtsai 2015]) létrehozásának célja az volt, hogy a fehérje-fehérje interakciós adatok kompartment szintű szűrésével egy olyan megbízható adatforrást hozzunk létre, mely alkalmas a biológiai folyamatok sejtszervecske szintű elemzésére. A ComPPI fő tulajdonsága a biológiailag nem valószínű kapcsolatok kiszűrésének lehetősége a szubcelluláris lokalizáció alapján, illetve lehetőséget ad a fehérjék új lokalizáció alapú funkcióinak jóslására is.

A minél szélesebb körű és megbízhatóbb adatok elérése érdekében a végső adatbázis számos bemenő adatbázis integrációjából áll össze, mely folyamatot számos helyen kézileg ellenőriztünk vagy segítettünk. A **10. ábra** ezt a folyamatot mutatja be sematikusan.



***10. ábra: A ComPPI felépítésének folyamatábrája**. A bemenő adatbázisok kiválasztása (1) és ellenőrzése (2) után azok integrációja következik, melynek alapja az eltérő fehérje nevek egyeztetése, és egységes nevezéktanra fordítása (3). Az összesített adatkészleten alapulva kiszámítjuk a lokalizációs és interakciós megbízhatósági értékeket, melyek bekerülnek a végső adatkészletbe. Az illusztratív példa a HSP 90-alfa A2 fehérje (A) és a Survivin (B) lokalizációs viszonyain alapuló megbízhatósági érték számítást mutatja. A ComPPI adatbázis kézi ellenőrzés (4) után böngészhető a webes felületen, illetve letölthető. Forrás: [Veres és mtsai 2015]*

A ComPPI megalkotásának folyamata során először felmértük, hogy milyen adatbázisok állnak rendelkezésre a fehérje-fehérje interakciós és szubcelluláris adatok minél szélesebb körű lefedéséhez. Az integrációra való igényt az első exploratív adat integrációs lépések során mutatkozó alacsony átfedés alapozta meg. Az elérhető adatforrások áttekintése során 24 olyan adatbázist találtunk, melyek potenciálisan integrálhatóak a végső adatkészletbe. Ezen adatbázisok átnézése volt az első kézi lépés a ComPPI felépítésében, a kompatibilis adatszerkezet mellett minőségi ellenőrzéssel, hibás bejegyzések és adat inkonzisztencia keresésével. Ezen lépésekre a minél megbízhatóbb végső adatkészlet céljából volt szükség. Egyes adatbázisok felhasználásával kapcsolatos engedélyek is kizártak forrásokat, mint például a STRING interaktómot [Szklarczyk és mtsai 2015], melynek legfrissebb verziója nem integrálható más adatbázisokba. A forráskeresés és -ellenőrzés eredményeképpen 9 fehérje-fehérje interakciós adatbázist és 8 fehérje szubcelluláris lokalizációs adatbázist összegeztünk, mely adatbázisok 3 modell organizmusra (*Caenorhabditis elegans, Drosophila melanogaster, Saccharomyces cerevisiae*) és az emberre tartalmaznak adatokat.

A szubcelluláris lokalizációs adatbázisok eltérő szintű adatokat tartalmaznak, így például a predikciós algoritmusok csak a nagy organellumok szerint osztják el a fehérjéket, míg a kísérletes adatok a fehérjék egészen konkrét elhelyezkedéséről is biztosítanak információt. A lokalizációs adatok felbontásán túl az egyes kompartmentek elnevezése is különbözik, melyet egységes nevezéktanra emeltünk. Ennek érdekében a folyamat második lépéseként kézi gyűjtéssel létrehoztunk egy hierarchikus, redundancia mentes lokalizációs fát, mely a lokalizációkat a GO [The Gene Ontology Consortium 2013] nevezéktanának megfelelően egyértelműen besorolja.

Az egyedi adatbázisokhoz illesztőfelületeket állítottunk össze, melyek segítségével megadható, hogy az adatszerkezeti eltéréseket, eltérő nevezéktanokat vagy kivételes egyedi adattal összefüggő szituációkat a források betöltése során hogyan kezelje a rendszer. Az illesztőfelület alapú megoldás biztosítja, hogy a régi adatbázisok megközelítőleg automatikusan legyenek frissíthetők, illetve új adatbázisok beemelésére legyen lehetőség.

Az eltérő bemenő adatforrások sokszor eltérő fehérje nevezéktant használnak, melyet egységes formátumba érdemes hozni. Erre a célra kifejlesztettünk egy algoritmust, mely kézzel ellenőrzött fordítótáblák segítségével egységesíti az eltérő név konvenciókat a

végső, legmegbízhatóbb UniProt [The UniProt Consortium 2017] nevekre. Az adatok sikeres integrációját követően kiszámítottuk a megbízhatósági értéket a lokalizációs és interakciós adatokra nézve.

Az integrált ComPPI adatkészletet hat független kutató ellenőrizte, akik szakértői a kapcsolódó kutatási területeknek. Ezen kézi ellenőrzés során a 200-200 random fehérje bemenő és integrált adatai közti egyezést vizsgálták annak érdekében, hogy biztosítsák az adatok magas szintű megbízhatóságát. Az ellenőrzési folyamat során a forrás adatbázisok kiválogatásához hasonló módszereket használva kerestek hamis bejegyzéseket, fehérje név fordítási hibákat, valamint adat inkonzisztenciát, majd ezt követően javítottuk a felmerülő eltéréseket.

Az adatbázist kiszolgáló felületet nyilvános használat előtt teszteltük, ellenőrizve a keresési eredményeket, a letölthető adatokat és a webes funkciókat. A teljes ellenőrzést követően a bemenő adatbázisokat és a lokalizációs fát frissítettük annak érdekében, hogy a legújabb adatokkal publikálhassuk az adatbázist.

A ComPPI a http://ComPPI.LinkGroup.hu/ oldalon érhető el, melynek webes felületén keresztül lehetőség van az adatok felhasználóbarát böngészésére, illetve a szelektált adatkészletek letöltésére.

6.1.2. Az interakciós és lokalizációs adatok integrálásának lépései

6.1.2.1. A forrásadatbázisok kiválasztásának elve

A fehérje-fehérje interakciós és szubcelluláris lokalizációs adatforrások átfedése kicsi [Veres és mtsai 2015], így a teljesebb fehérje lefedettség és jobb minőség érdekében több adatforrást integráltunk. Ehhez nyilvánosan elérhető, kutatási célból letölthető és felhasználható adatbázisokat választottunk, lehetőség szerint a modell organizmusokra és az emberre vonatkozó proteóm szintű adatkészlettel.

A fehérje-fehérje interakciós adatforrások esetében csak a kísérletes evidenciával alátámasztott fizikai kölcsönhatásokat alkalmaztuk. Az adatok kísérleti forrása lehet alacsony, illetve magas áteresztőképességű egyaránt. Szempont volt az adatok megbízhatósága, a frissítési gyakoriság, és a legfrissebb verzió szabad felhasználásának lehetősége kutatók számára. Ezek alapján 9 fehérje-fehérje interakciós forrást összegeztünk, ezek között voltak fajra specifikus adatforrások (DroID [Murali és mtsai 2011], HPRD [Keshava és mtsai 2009], MatrixDB [Launay és mtsai 2015] és MIPS

[Pagel és mtsai 2005]), illetve több fajra kiterjedő általános adatbázisok (BioGRID [Chatr-Aryamontri és mtsai 2015], CCSB [Rolland és mtsai 2014], DIP [Salwinski és mtsai 2004], IntAct [Kerrien és mtsai 2012] és MINT [Licata és mtsai 2012]).

A szubcelluláris lokalizációs adatok forrása lehet kísérletes vagy prediktált. A felhasznált adatbázisok egy része csak kísérletes (Human Proteinpedia [Kandasamy és mtsai 2009], Human Portein Atlas [Pontén és mtsai 2011]), míg mások kizárólag prediktált (PA-GOSUB [Lu és mtsai 2005]) adatokat tartalmaznak. A kiválasztott forrás adatbázisok nagyobb része azonban integrált adatokat tartalmaz mind kísérletes, mind prediktált forrásból (eSLDB [Pierleoni és mtsai 2007], Gene Ontology [The Gene Ontology Consortium 2013], LOCATE [Sprenger és mtsai 2008], MatrixDB [Launay és mtsai 2015], OrganelleDB [Wiwatwattana és mtsai 2007]). A proteóm-szintű adatokat tartalmazó szubcelluláris lokalizációs adatforrások kiválasztása során a jó minőségű kísérletes adatok maximalizálása mellett törekedtünk a megbízható predikciós eredmények integrálására, így olyan predikciós algoritmusok eredményeit használtuk csak fel, melyek több módszert alkalmaznak robosztus, magas megbízhatóságú tanító adatokon validált gépi tanulási eszközökkel kombinálva.

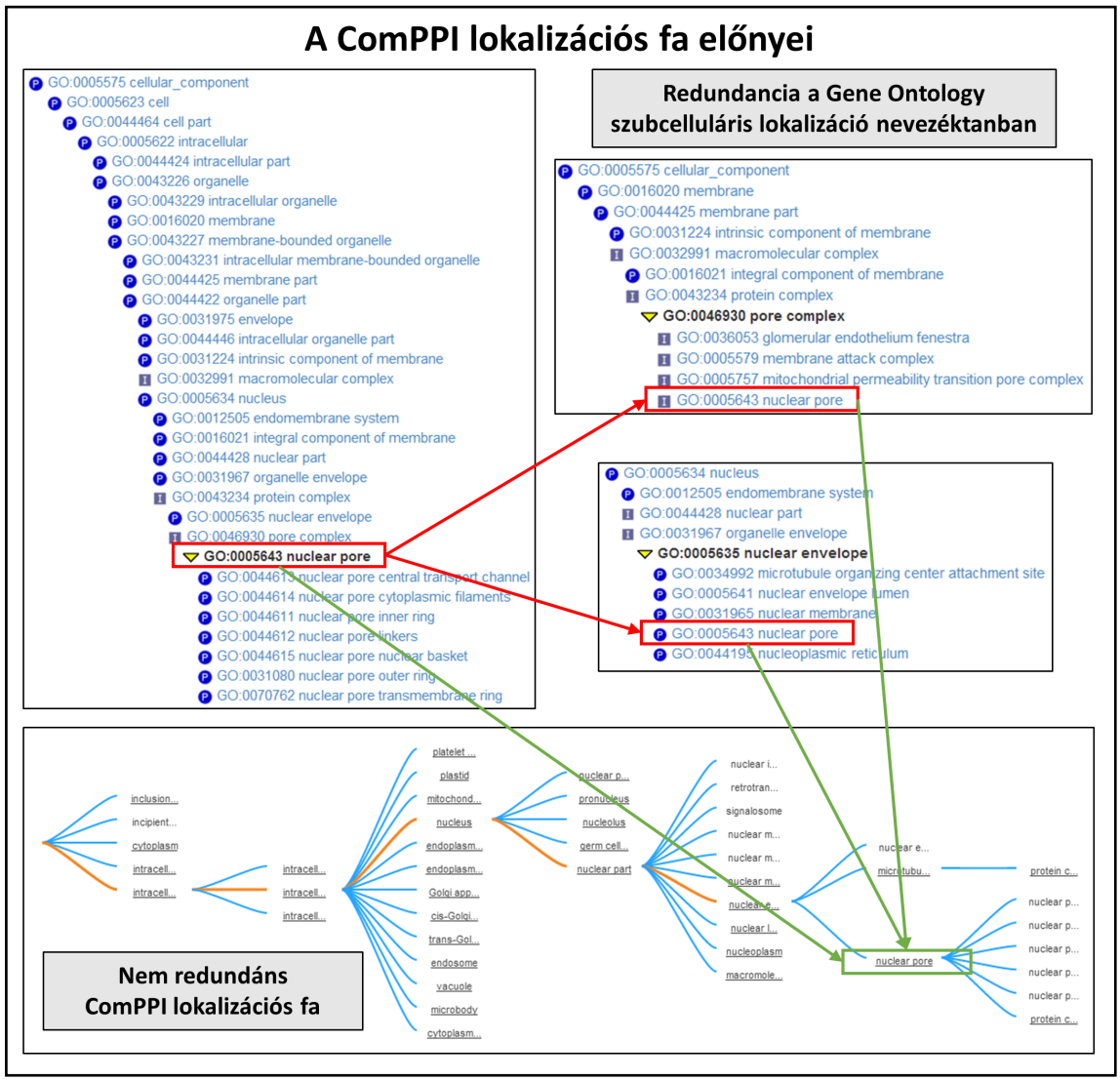
6.1.2.2. A szubcelluláris lokalizációs adatok szerkezete

A szubcelluláris lokalizációs adatok különböző forrásokból származnak, melyek tartalmazhatnak kísérletes, prediktált, vagy ismeretlen eredetű adatokat. A kísérletes adatok felbontása általában magas, a fehérjék pontos lokalizációjáról hordoznak információt, mint például a Nup107-es fehérje sejtmagi pórus komplex lokalizációja [Hoelz és mtsai 2011]. A prediktált adatok felbontása általában limitált, így például a sejtmagi lokalizációs szignál jelenlétében, de kísérletes validáció hiányában pusztán a sejtmagi lokalizációt képes feltételezni [Brameier és mtsai 2007].

Az adatok felbontásában és a lokalizációk nevezéktanában lévő különbségek miatt a szubcelluláris lokalizációs adatokat a GO [The Gene Ontology Consortium 2013] vonatkozó nevezéktana[40] alapján összesítettük. Az itt található egységes nevezéktan hátránya, hogy egy adott lokalizáció több úton is elérhető az irányított aciklikus gráf struktúra miatt (**11. ábra**), mely megnehezíti a magasabb felbontású lokalizációs adatok nagyobb kategóriáknak történő egyértelmű megfeleltetését.

[40] http://geneontology.org/page/cellular-component-ontology-guidelines/

A probléma megoldása céljából kézi gyűjtéssel létrehoztunk egy hierarchikus szubcelluláris lokalizációs fát[41], mely redundancia nélkül rendezi be az egységes nevezéktanra fordított lokalizációs adatokat. A lokalizációs fa segítségével a több mint 1600 forrás adatban előforduló GO [The Gene Ontology Consortium 2013] lokalizációs nevet egyértelműen be tudtuk osztani hat nagy lokalizációs kategóriába. Ezek a következők: sejtplazma, sejtmag, mitokondrium, szekréciós-út, membrán, extracelluláris tér. Ez az új módszer biztosítja, hogy a ComPPI különböző forrásokból származó lokalizációs adatai egységesen és következetesen megfeleltethetők legyenek a hat nagy lokalizációnak (**11. ábra**) [Veres és mtsai 2015].

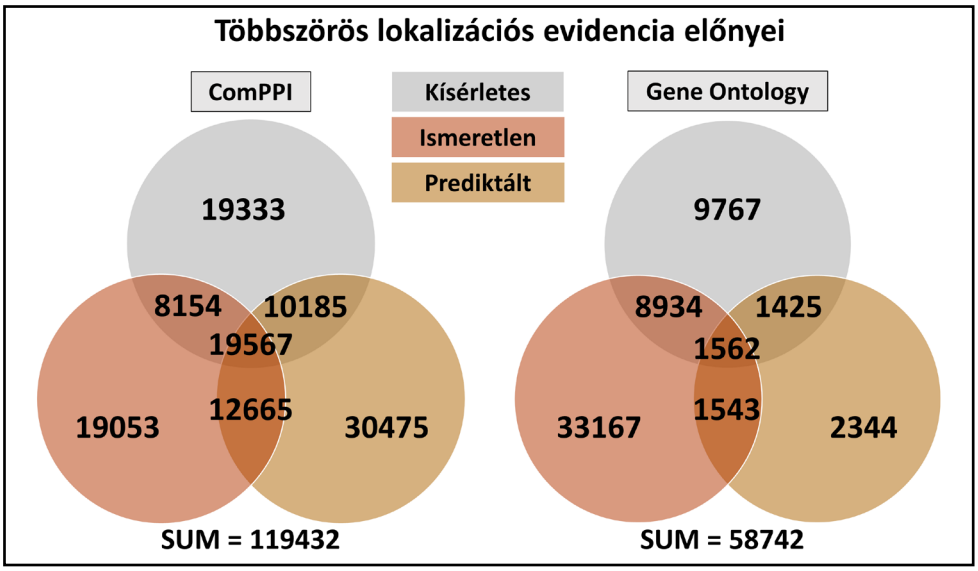


***11. ábra**: **A ComPPI lokalizációs fa előnyei**. (folytatás a következő oldalon ->)*

[41] http://www.linkgroup.hu/pic/loctree.svg/

(-> folytatás az előző oldalról) Az ábra felső részén egy példát mutatunk be a Gene Ontology [The Gene Ontology Consortium 2013] szubcelluláris lokalizációs nevezéktan redundanciájára. A sejtmag pórust (nuclear pore) több útvonalon is elérhetjük, így a sejtmag (nucleus) – sejtmag membrán (nuclear envelope) – sejtmag pórus (nuclear pore) sorrendben, vagy a membrán (membrane) – membrán rész (membrane part) – membrán belső eleme (intrinsic component of the membrane) – membrán integráns eleme (integral component of the membrane) – pórus komplex (pore complex) – sejtmag pórus (nucelar pore) útvonalon. Annak érdekében, hogy a magasabb felbontású lokalizációs adatokat egyértelműen tudjuk nagyobb egységes kategóriákba sorolni, szükséges volt egy nem redundáns lokalizációs fa létrehozása, melynek egy elemét mutatja az ábra alsó része, bemutatva a sejtmag pórus (nuclear pore) egyértelmű követhetőségét. Forrás: [Veres és mtsai 2015]

A szubcelluláris lokalizációs adatok integrációjának eredménye, hogy nemcsak az adatok mennyisége, hanem megbízhatósága is növekedett. Ezt mutatja a **12. ábra** Venn-diagramja, mely a kísérletes, prediktált, és ismeretlen forrásból származó adatok eloszlását szemlélteti a teljes ComPPI adatkészletben, összehasonlítva a ComPPI azon adataival, melyekhez tartozik GO [The Gene Ontology Consortium 2013] forrás.

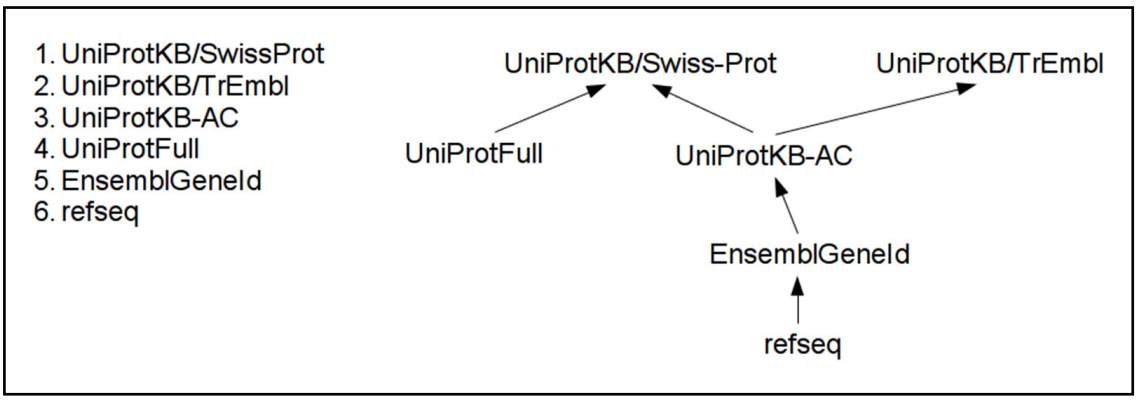


***12. ábra: A lokalizációs adatok integrációjának előnyei**. A szubcelluláris lokalizációs adatok forrása lehet kísérletes módszer (szürke), predikció eredménye (barna), vagy ismeretlen eredetű információ (mogyoró). (folytatás a következő oldalon ->)*

(-> folytatás az előző oldalról) A ComPPI teljes adatkészletében (balra) a csak ismeretlen eredetű lokalizációval rendelkező bejegyzések száma jóval alacsonyabb, mint a GO [The Gene Ontology Consortium 2013] adatokra szűkített (jobbra) esetben (57%), miközben a csak kísérletes (264%) vagy vegyes (376%) lokalizációval rendelkező adatok száma jóval nagyobb a teljes adatkészletben, mint az várható a fehérjék számában lévő különbségből kiindulva (203%). Forrás: [Veres és mtsai 2015]

6.1.2.3. A fehérjenevek fordítása

Az egyes forrás adatbázisok sokszor eltérő fehérje nevezéktanokat alkalmaznak. Az adatok integrációja során ezért szükség van egy egységes nevezéktan használatára. Annak érdekében, hogy biztosítsuk az adatok magas szintű kapcsoltságát, fordító táblák segítségével minden egyedi fehérje nevét lefordítottuk a legmegfelelőbb fehérje nevezéktanra. A választott elérendő egységes nevezéktan a UniProt [The UniProt Consortium 2017] által szolgáltatott kézzel ellenőrzött SwisProt, illetve az automatikusan annotált TrEMBL nevezéktan[42]. Az adatok maximális megbízhatóságát az elsődleges SwissProt nevezéktannal tudtuk elérni, így ezt választottuk a legerősebb név konvenciónak (**13. ábra**). A végső fehérje halmaz 30%-a UniProt SwissProt, míg 70%-a UniProt TrEMBL nevezéktannal szerepel.



13. ábra: Fehérje nevezéktan fordítás a ComPPI-ban. *Amennyiben egy új fehérje név kerül be a rendszerbe, a cél annak lefordítása az elérhető legerősebb nevezéktanra. (folytatás a következő oldalon ->)*

[42] http://www.uniprot.org/help/uniprotkb_sections/

(-> folytatás az előző oldalról) Példaként vehetjük az NP_005205 RefSeq névvel rendelkező gén terméket, melynek következő elérhető fordítási lépése az ENSG00000163599 EnsemblGeneID. A legerősebb UniProtKB/SwissProt név megtalálása zárja a sort, ez esetben a P16410. A fordítás lehetséges teljes elnevezések alapján is (UniProtFull), így a `Cytotoxic T-lymphocyte protein 4 precursor' név az előbbi P16410 SwissProt nevezéktanra fog lefordulni.

A megoldás gyors teljesítménye mellett lehetőséget ad szinonimák keresésére, melyet felhasználtunk a fehérje név szinonimák kiszolgálása során. A fordításhoz publikusan elérhető fordító táblákat (UniProt [The UniProt Consortium 2017], illetve a Human Protein Reference Database [Keshava és mtsai 2009]), valamint a Synergizer (http://llama.mshri.on.ca/synergizer/translate/) és Protein Identifier Cross-Reference (PICR) [Wein és mtsai 2012] web eszközökkel végzett kézi fordítást alkalmaztuk.

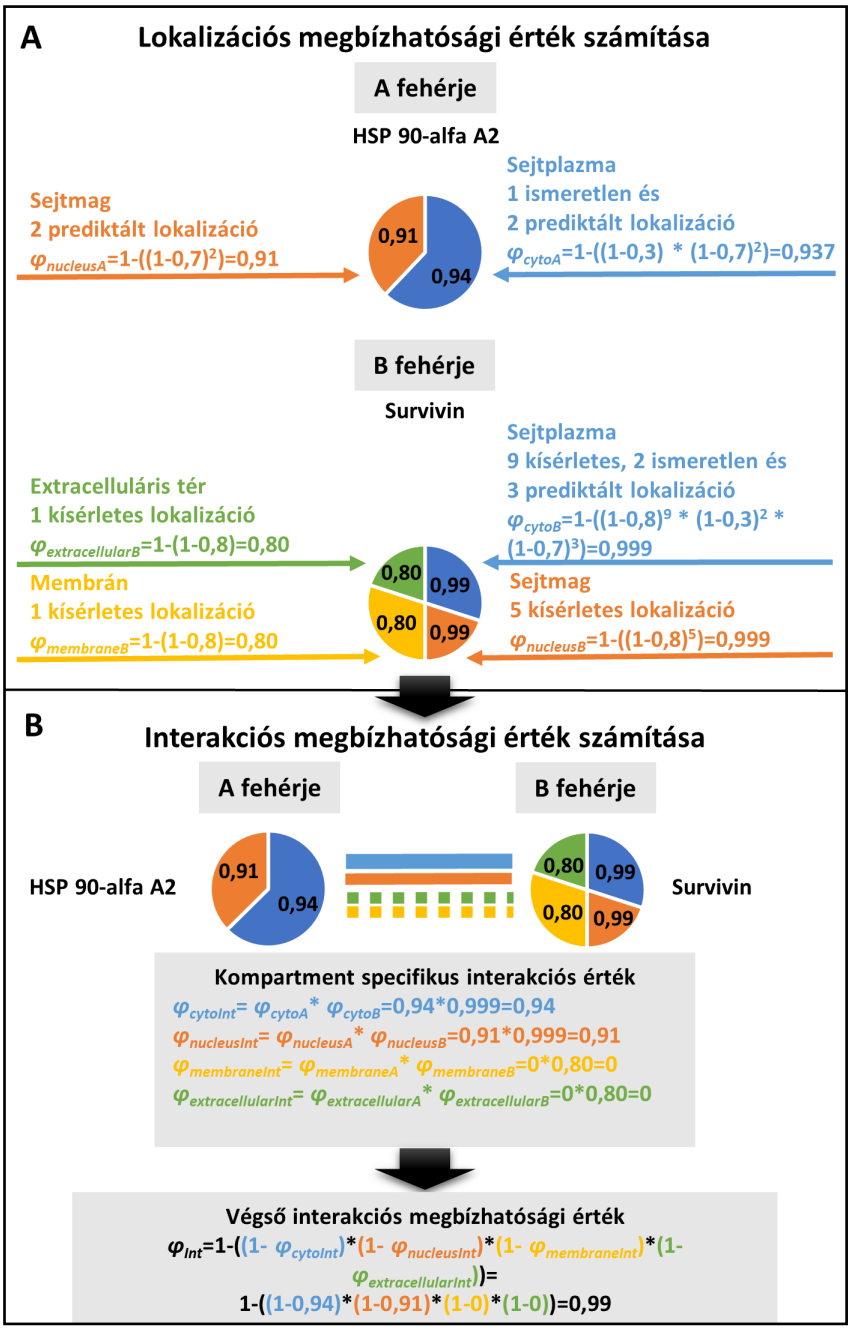
Az adatok átnézése során nem bizonyult magasnak azon esetek száma, ahol a fehérjék fordítása nem volt megfelelő. Ezzel együtt az adatbázis használata során annak limitációit is figyelembe kell venni. A fordítás robosztussága ellenére előfordulhatnak redundáns bejegyzések, főleg az automatikusan annotált UniProt TrEMBL nevekkel rendelkező fehérjék között. Szintén előfordulhatnak fehérje fragmentek a UniProt SwissProt nevek között is, melyekre nem szűr az adatkészlet. Gén nevekről történő fordítás esetén egynél több peptid termék nevére is fordítódhat az adott név (például alternatív splicing eredményeképpen), ezzel duplikált adatokat eredményezve.

DOI:10.14753/SE.2018.2134

6.1.3. A lokalizációs és interakciós megbízhatósági érték számításának módja

6.1.3.1. A megbízhatósági értékek számítása

A ComPPI lokalizációs megbízhatósági érték egy általunk definiált, új mérőszám [Veres és mtsai 2015], melynek segítségével az egyes fehérjék adott nagy lokalizáción belüli elhelyezkedésének valószínűségét értékeljük. Az érték függ a lokalizációs adat kísérletes, prediktált vagy ismeretlen eredetétől, valamint a források számától (**14. ábra**).



(Az ábra magyarázó szövegét lásd a következő oldalon ->)

14. ábra: A ComPPI megbízhatósági érték számításának bemutatása*. A lokalizációs megbízhatósági érték számításának lépéseit a sejtmagi és sejtplazmai lokalizációval rendelkező HSP 90-alfa A2 és a négy lokalizációba (sejtplazma, sejtmag, membrán, extracelluláris tér) is tartozó Survivin fehérjék példáján illusztráljuk. (**A**) A lokalizációs megbízhatósági értéket minden nagy lokalizációnak megfelelően kiszámoljuk, figyelembe véve az elérhető evidenciák számát és azok típusát (felhasznált evidencia típus súlyok: kísérletes = 0,8, prediktált = 0,7, ismeretlen = 0,3, lásd később). (**B**) Az interakciókra vonatkozó megbízhatósági érték számítása az egyes fehérjék lokalizáció specifikus megbízhatósági értékén alapul. Először kiszámolunk a nagy lokalizációknak megfelelően egy konszenzusos kompartmentekre specifikus megbízhatósági értéket, majd ezek alapján számítjuk ki a végső interakciós megbízhatósági értéket. Forrás: [Veres és mtsai 2015]*

A lokalizációs megbízhatósági érték (φLocX) számítása (**1. képlet**) során probabilisztikus diszjunkciót (*V* operátoral jelölve) használtunk, alkalmazva a különböző lokalizációs evidenciákat (pLocX), kiegészítve azt a ComPPI-ban elérhető forrás lokalizációs adatok számával (res) az *X* fehérje *Loc* lokalizációjában.

1. képlet: $$\varphi_{LocX} = V_{res} p_{LocX}$$

Az interakciókra vonatkozó megbízhatósági érték az adott kölcsönhatás szubcelluláris lokalizáció specifikus valószínűségét fejezi ki, melynek alapja a kölcsönható partnerekre vonatkozó kompartment specifikus lokalizációs megbízhatósági értékek konszenzusa. A kölcsönható partnerek hat nagy lokalizációs kategóriára vonatkozó kompartment specifikus lokalizációs megbízhatósági értékét külön-külön számítjuk ki a vonatkozó lokalizációs megbízhatósági értékek szorzataként. A végső interakciós megbízhatósági érték (φ_{Int}) a hat nagy lokalizációs kategóriára vonatkozó kompartment specifikus lokalizációs megbízhatósági értékek probabilisztikus diszjunkciójának (*V* operátorral jelölve) eredménye (**2. képlet**), ahol φ_{LocA} és φ_{LocB} az *A* és *B* kölcsönható partnerek kompartment specifikus lokalizációs megbízhatósági értékét jelöli.

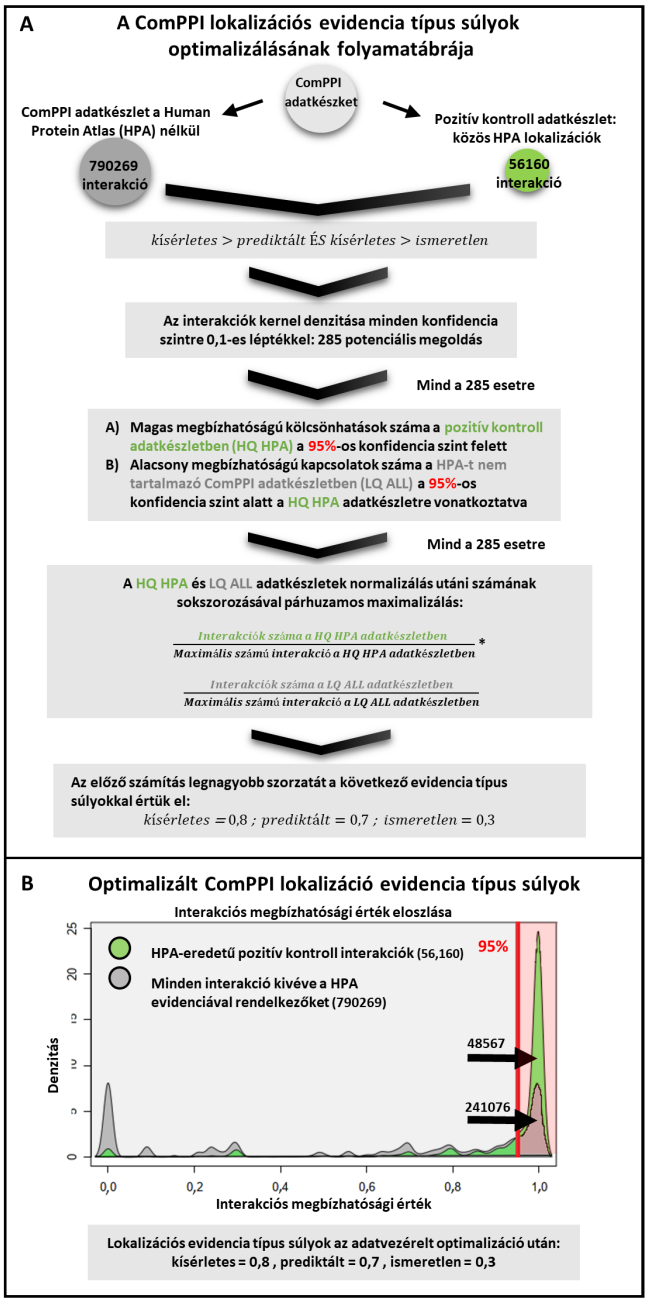
2. képlet: $$\varphi_{Int} = V^{6}_{i=1} \varphi_{LocA} * \varphi_{LocB}$$

A ComPPI megbízhatósági értékek segítségével lehetőségünk van felmérni annak a valószínűségét, hogy az adott fehérje előfordul-e a meghatározott kompartmentben. Ezen

információ a lokalizáció specifikus interakciós mérőszámok eloszlásának segítségével felhasználható magas megbízhatóságú interaktómok építésére [Veres és mtsai 2015].

6.1.3.2. A megbízhatósági értékek optimalizálása

A lokalizációs megbízhatósági érték számítása során a forrás adatok eredete, így a kísérletes, prediktált vagy ismeretlen evidencia szabad paraméterek azok, melyeket súlyoznunk kell annak érdekében, hogy egy minél megbízhatóbb egységes pontozási rendszert állítsunk fel a diverz adatkészletre. Az evidencia típusok súlyozására adatvezérelt optimalizációt végeztünk (**15. ábra**).



(Az ábra magyarázó szövegét lásd a következő oldalon -->)

***15. ábra: A megbízhatósági érték paramétereinek optimalizálása**. Az **A** panel folyamatábrája a ComPPI lokalizációs evidencia típus súlyok optimalizálásának lépéseit mutatja be. A folyamat célja az evidencia típusok egymáshoz képesti súlyozása volt úgy, hogy megfelelő biztonsággal el tudjunk különíteni egy magas és egy alacsony megbízhatóságú adatkészletet. A súlyok meghatározásához meghatároztunk egy pozitív kontroll adatkészletet, mely csak olyan interakciókat tartalmaz, ahol mindkét partner rendelkezik legalább egy Human Protein Atlas-ból (HPA) [Pontén és mtsai 2011] származó kísérletes lokalizáció evidenciával. A zölddel jelölt 56160 pozitív kontroll interakciót összehasonlítottuk a szürkével jelölt 790269 interakcióval, mely az összes olyan ComPPI kölcsönhatást tartalmazza, ahol nem található HPA evidencia. A kiindulási hipotézisünk az volt, hogy a kísérletes adatok megbízhatósága a legmagasabb, így magasabb súlyt várunk rájuk, mint egy prediktált vagy ismeretlen evidencia esetében. Ennek a feltételnek megfelelő összes kísérletes, prediktált és ismeretlen lokalizációs evidencia típus súly kombinációt kiszámoltuk 0 és 1 közötti érték tartományban 0,1-es léptékkel, majd az adatok kernel denzitása alapján vizsgáltuk az egyes konfidenciaszinteket a teljes eloszláshoz viszonyítva. A cél az volt, hogy maximalizáljuk a magas megbízhatóságú interakciók (HQ) számát a pozitív kontroll adatkészletben (HPA), miközben szintén maximalizáljuk az alacsony megbízhatóságú (LQ) kölcsönhatások számát a HPA nélküli teljes ComPPI adatkészletben. Mivel a HQ interakciók száma jóval kisebb, mint az LQ interakcióké, így normalizáltuk az értékeket a statisztikai hibák minimalizálása érdekében. A normalizált értékek szorzásával jutottunk egyedi evidencia típus súlyokhoz, mely a legjobb HQ és LQ szétválasztást a következő esetekben adta: kísérletes = 0,8, prediktált = 0,7 és ismeretlen = 0,3. A **B** panel az interakciós megbízhatósági értek eloszlását mutatja 15 optimalizáló kör után. Az X tengelyen az interakciós megbízhatósági értek látható, míg az Y tengely reprezentálja az interakciós értek eloszlás kernel denzitását. A 95% feletti konfidencia intervallumban található magas megbízhatóságú interakciók száma a pozitív kontroll adatkészletben 48567 a teljes készlet 56160 eleméből. Ezzel párhuzamosan a magas megbízhatóságú csoportba az összes ComPPI interakció 30%-a, 241076 interakció került be a 790269 kölcsönhatásból. Az optimális eloszláshoz tartozó kísérletes evidencia típus súly azt mutatja, hogy egy evidencia önmagában nem elég a megbízható lokalizáció megítéléséhez, és legalább két kísérletes evidencia szükséges ahhoz, hogy egy lokalizáció a magas megbízhatóságú csoportba kerüljön.* *(folytatás a következő oldalon ->)*

(-> folytatás az előző oldalról) A prediktált evidenciákhoz tartozó aránylag magas súly alátámasztja, hogy a szubcelluláris lokalizáció predikciós algoritmusok megbízhatósága magas, míg az ismeretlen eredethez tartozó alacsony súly megerősíti az adatforrások validációjának igényét. Forrás: [Veres és mtsai 2015]

Az interakciós megbízhatósági érték optimalizációja során a Human Protein Atlas (HPA) [Pontén és mtsai 2011] adatbázis megbízható kísérletes adatait használtuk pozitív kontroll adatkészletnek, ahol a kölcsönható partnerek legalább egy közös lokalizációval rendelkeznek a HPA adatbázisban. Célunk volt egy a kísérletes, prediktált és ismeretlen eredetű evidenciákhoz tartozó súlyok arányának meghatározása, melynek segítségével a magas megbízhatósági kölcsönhatások számát maximalizálhatjuk a pozitív adatkészletben. Mindeközben az alacsony megbízhatóságú kapcsolatok számát maximalizáljuk azon kölcsönhatásokra, melyek hiányoznak a pozitív adatkészletből. Ezen feltételek biztosítják, hogy a magas megbízhatóságúnak jelölt interakciók minél inkább megegyezzenek a kísérletesen validált adatokkal.

6.1.4. A ComPPI adatbázis statisztikája

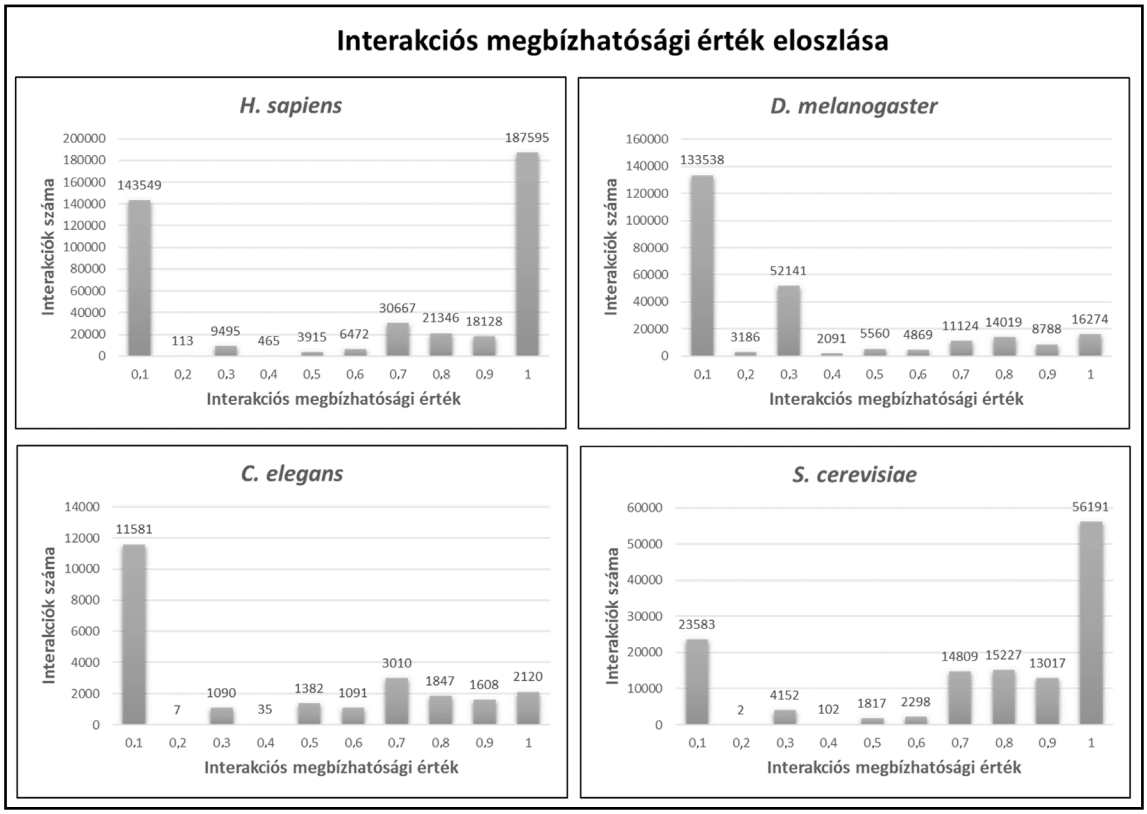
A ComPPI adatbázis 3 modell organizmusra (*Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*) és az emberre (*Homo sapiens*) tartalmaz komprehenzív, integrált adatkészletet. Az 1.1-es verzió 127757 fehérjét és azok 791059 kapcsolatát, valamint 195815 nagy lokalizáció szerinti elhelyezkedési adatot foglal magába (**5. táblázat**). A proteóm-szintű adatkészlet az 5 nagy sejten belüli organellumra (membrán, mitokondrium, sejtmag, sejtplazma, szekréciós-út) és az extracelluláris kompartmentre kategorizáltan tartalmaz lokalizációs adatokat. A lokalizációs bejegyzések több mint 60%-a magas felbontású, 1600 feletti GO [The Gene Ontology Consortium 2013] sejtkomponenst lefedve. Az adatok faj és adat típus szerinti részletes bemutatása a következő oldalon érhető el: http://comppi.linkgroup.hu/help/statistics/

DOI:10.14753/SE.2018.2134

5. táblázat: A ComPPI adatkészlet összegző statisztikája

Adat típus	Kompartmentalizált adatkészlet	Interakciós adatkészlet	Lokalizációs adatkészlet
Fehérjék száma	42 829	53 168	119 432
Lokalizációk száma	86 874	-	195 815
Átlagos lokalizációs megbízhatósági érték	0,76	-	0,73
Interakciók száma	517 461	791 059	-
Átlagos interakciós megbízhatósági érték	0,76	0,49	-

Az interakciók megbízhatósági értek szerinti eloszlását mutatja a **16. ábra**. Látható, hogy az emberben és élesztőben (*Saccharomyces cerevisiae*) jelen lévő magas megbízhatóságú fehérje-fehérje kölcsönhatások száma kiemelkedő. Szintén jelentős azon interakciók száma, melyekben a kölcsönható partnerekre nézve nem elérhető szubcelluláris lokalizációra vonatkozó információ, vagy azok nem azonos lokalizációban találhatóak. Érdekes megfigyelés a *Caenorhabditis elegans* esetében látható alacsony számú jó megbízhatóságú adat, mely a szubcelluláris adatok hiányosságával függhet össze. Az ecetmuslica (*Drospohila melanogaster*) estében kiemelkedően magas az alacsony megbízhatóságú kapcsolatok száma, mely a fehérje-fehérje interakciók detektálási problémáira világíthat rá a modell organizmusban.



16. ábra: Az interakciós megbízhatósági érték eloszlása. *Az ábrán az integrált fehérje-fehérje kölcsönhatási adatok interakciós megbízhatósági érték eloszlása látható a négy faj szerinti bontásban. Az X tengely mutatja az egyes konfidencia intervallumokat, míg az Y tengely a megfelelő kölcsönhatások számát. Forrás: [Veres és mtsai 2015]*

6.1.5. Az ComPPI felhasználói felületének bemutatása

A ComPPI webes felhasználói felületének (http://comppi.linkgroup.hu/) segítségével a bioinformatikai jártassággal nem rendelkező kutatók is felhasználóbarát módon böngészhetik az egyes fehérjék lokalizációs és interakciós adatait. A keresési funkciót segíti a beírt fehérje nevének automatikus kiegészítése, így csökkentve az elgépelésből adódó keresési hibák számát. Az egyszerű keresésen túl lehetőség van kibővített keresésre is, melynek segítségével faj, lokalizáció és lokalizációs megbízhatósági értékre specifikus keresések indíthatók. A szűrési lehetőségeket mind a keresett fehérjére, mind annak kölcsönható partnereire is lehet érvényesíteni.

A keresett fehérje kölcsönható partnerei szubcelluláris lokalizáció, lokalizációs és interakciós megbízhatósági érték alapján szűrhetők a találati oldalon[43]. Az így generált igény szerinti találati lista a felületen keresztül letölthető. A fehérjék nevének szinonimái, a részletes lokalizációs adat, illetve az interakciós adatok forrása mellett letisztult hálózatos megjelenítés segíti a keresett fehérjék első szomszédjainak böngészését. Közvetlen keresésre is lehetőség van a fehérjék UniProt alapú specifikus aloldalainak[44] betöltésével, mely segíti a ComPPI összekapcsolását más adatforrásokkal, illetve a többszörös keresésre is lehetőséget ad adatbányászat céljából.

A ComPPI letöltő oldalán[45] lehetőség nyílik előre meghatározott adatkészletek letöltésére. A kompartmentalizált fehérje-fehérje interakciós adatkészlet csak olyan kölcsönhatásokat tartalmaz, ahol a kapcsolódó fehérjéknek legalább egy lokalizációjuk közös. Ezen adatkészletek nagy lokalizációra specifikus letöltésére is lehetőség van.

Az integrált fehérje-fehérje interakciós adatkészlet az adott fajra vonatkozó összes kölcsönhatási adatot tartalmazza, a szubcelluláris lokalizációtól függetlenül. Az összesített szubcelluláris lokalizációs adatkészlet a hat nagy kategóriának megfelelően biztosítja a fehérjék integrált lokalizációs adatainak letöltését. Ezen kívül lehetőség van az aktuális, és korábbi verziók adatbázis formátumban való letöltésére is.

A ComPPI által kiszolgált adatok tartalmazzák a fehérje-fehérje kölcsönhatások listáját, az egyes fehérjék lokalizációs adatait, illetve az interakciós és lokalizációs megbízhatósági értéket. Mindemellett az interakciós és szubcelluláris lokalizációs adatokhoz is elérhető a forrás adatbázis(ok) neve és a hozzá(juk) tartozó referencia, mely segíti az adatok eredetének visszakövetését.

Az oldal fejlesztése során fontosnak tartottuk a minél egyszerűbb és felhasználóbarátabb használatot, így az oldalon belül hivatkozunk a megfelelő magyarázó oldalakra, melyeket a részletes dokumentáció[46] tartalmaz. A felhasználást segíti a ComPPI használati lehetőségeit szemléltető képes bemutató[47] is.

[43] http://comppi.linkgroup.hu/protein_search/
[44] http://comppi.linkgroup.hu/protein_search/interactors/P04637/
[45] http://comppi.linkgroup.hu/downloads/
[46] http://comppi.linkgroup.hu/help/
[47] http://comppi.linkgroup.hu/help/tutorial/

DOI:10.14753/SE.2018.2134

6.2. A ComPPI adatainak felhasználása az egyes fehérjék szintjén

A szubcelluláris adatok integrációja a fehérje-fehérje kölcsönhatási adatokkal számos felhasználásra ad lehetőséget. A biológiailag nem valószínű kapcsolatok a lokalizációs viszonyok alapján szűrhetők, illetve új valószínű lokalizációk vagy lokalizáció-specifikus biológiai funkciók prediktálhatóak.

6.2.1. A lokalizációs viszonyok jelentősége a krotonáz példáján

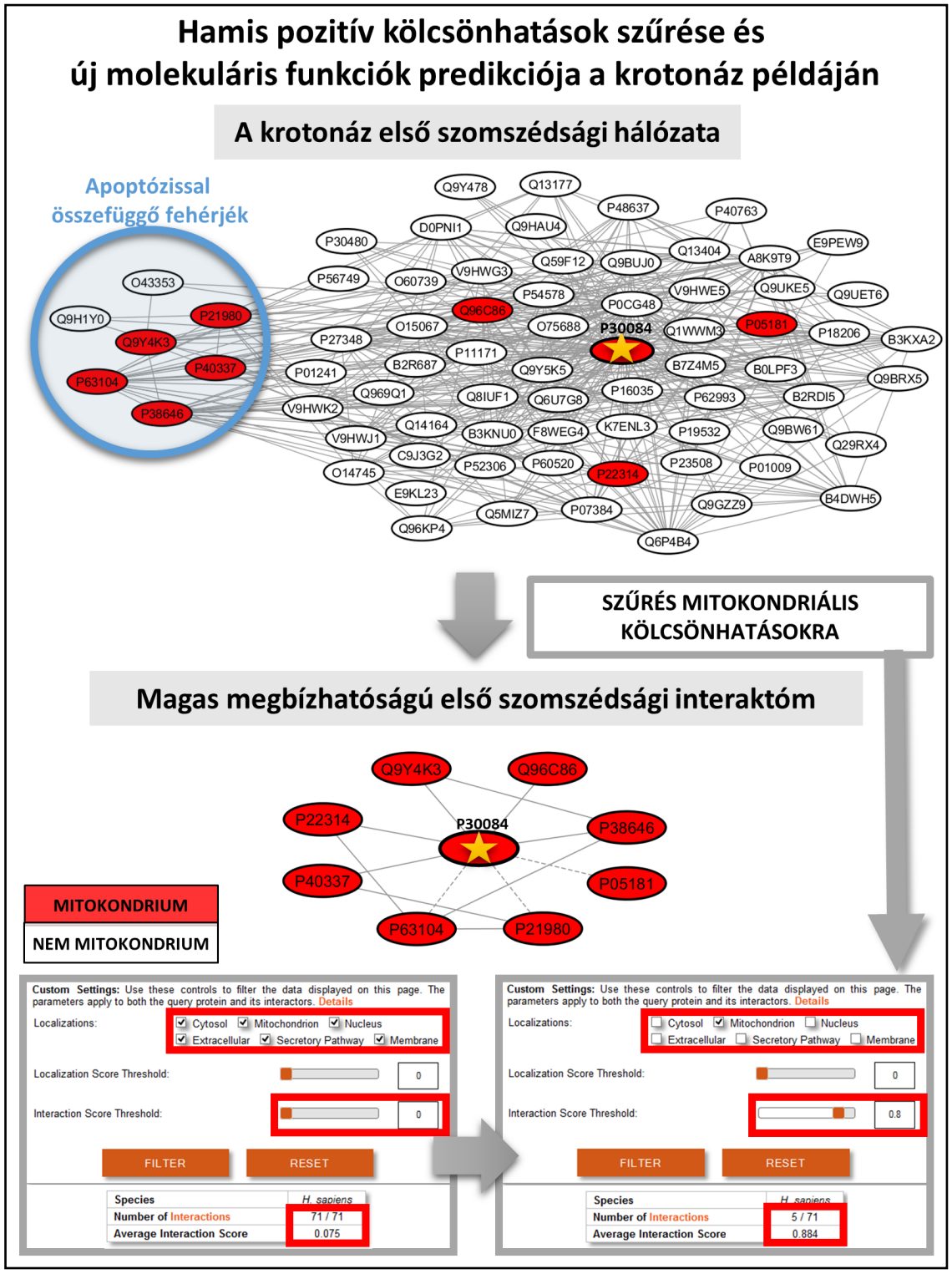
Ahhoz, hogy bemutassuk a lokalizáció-specifikus biológiailag nem valószínű fehérje-fehérje kölcsönhatások szűrésének jelentőségét, rendszerszintű keresést végeztünk az emberi adatkészletben olyan fehérjék után, melyek interaktómja a szűrés végére a legnagyobb változást mutatta. Ehhez kiszámítottuk a fehérjék kölcsönható partnereinek számát (fokszám), majd összehasonlítottuk a fokszám eloszlását a teljes interaktómban és a magas megbízhatóságú interaktómban is. Utóbbi esetben a biológiailag nem valószínű kapcsolatokat közös szubcelluláris lokalizáció hiányában nem vettük figyelembe. A teljes emberi adatkészlet 23.265 fehérjét és 385.481 interakciót, míg a szűrt interaktóm 19.386 fehérjét és azok 260.829 kölcsönhatását tartalmazta. A fokszám eloszlásán túl a köztiség hálózatos mérőszámot is vizsgáltuk, mely az adott fehérjén átmenő legrövidebb utak számát megadva karakterisztikus jellemzője egy hálózatos csúcs fontosságának.

Ezek után a megbízható UniProt SwissProt nevezéktanra szűkített találati listából (15.258 fehérje a 19.386-ból) kézzel átnéztük a fokszámra és köztiség mérőszámokra nézve legnagyobb eltérést mutató első 20 fehérjét [Veres és mtsai 2015]. A 20 fehérje közül az enoil-CoA hidratáz, vagy más néven krotonáz rendelkezett a legnagyobb abszolút fokszám változással, így ezt választottuk ki további elemzésre (**17. ábra**).

A krotonáz a zsírsavak béta-oxidációjának második lépését katalizáló enzim [Waterson és Hill 1972], mely a krotonáz fehérje szupercsalád egyik fő tagja [Hamed és mtsai 2008]. A zsírsavak béta-oxidációja elsősorban a mitokondriumban zajlik [Turteltaub és Murphy 1987], mely megegyezik a krotonáz kísérletes mitokondriális szubcelluláris lokalizációs adatával[48].

[48] http://comppi.linkgroup.hu/protein_search/interactors/P30084/

DOI:10.14753/SE.2018.2134



17. ábra: A ComPPI felhasználása a krotonáz példáján. *Az ábra a krotonáz (enoyl-CoA hidratáz, UniProtAC: P30084) kísérletes evidenciával rendelkező kölcsönhatásait mutatja a mitokondriális lokalizációra történő szűrés előtt és után. A 0,8 alatti megbízhatósági értékkel rendelkező kölcsönhatások szaggatott vonallal vannak elkülönítve.* *(folytatás a következő oldalon ->)*

(-> folytatás az előző oldalról) Amennyiben a krotonáz eredeti 71 interaktorát leszűrjük a mitokondriális lokalizációnak megfelelően, a kapcsolatok száma 8-ra csökken, miközben az átlagos megbízhatósági érték jelentősen növekszik, ezzel mutatva a kompartment specifikus szűrés jelentőségét a hamis pozitív kapcsolatok detektálásában. A kék körrel jelölt kölcsönható partnerek a fehérje apoptotikus szabályozással összefüggő sejtplazmai partnerei, mely a hamis pozitív kapcsolatok szűrése mellett felveti új biológiai funkciók lokalizáció specifikus predikciójának lehetőségét is. Forrás: [Veres és mtsai 2015]

A krotonáz az összesített adatok alapján 71 partnerrel hat kölcsön, melyek közül csak 8 rendelkezik mitokondriális lokalizációval, 0,8 feletti interakciós megbízhatósági értékkel pedig mindösszesen 5. A szomszédok kézi átnézése után kiderült, hogy csak egy mitokondriális partnernek van kísérletes lokalizációja, ez a mitokondriális Hsp70 hősokkfehérje [Bhattacharyya és mtsai 1995]. A fennmaradó 7 mitokondriális interaktor lokalizációs adata nem kísérletes evidencián nyugszik, miközben a maradék 63 interakciós partnernek egyáltalán nincs mitokondriális lokalizációra utaló adata.

A krotonáz 71 első szomszédja között 428 fizikai fehérje-fehérje kölcsönhatás található (**17. ábra**). A mitokondriális részhálózatban csak 13 él marad, melyek közül csak 10 magas megbízhatóságú. A krotonáz második szomszédjaival az interaktóm fehérje tagjainak 81%-át lefedjük, mely összesen 14.803 fehérjét és 319.305 interakciót jelent. A mitokondriumra történő szűrést követően a második szomszédok hálózata jóval kisebbnek bizonyul: 2017 fehérje és azok 8381 kapcsolata.

A 71 interaktor közül 52 rendelkezik sejtplazmai lokalizációval. A 8 mitokondriális partner közül 7-nek sejtplazmai lokalizációja is bizonyított, mely lokalizációk megbízhatósági értéke 0,8 felett van. Ezek alapján felmerül, hogy a krotonáz sejtplazmai lokalizációval is rendelkezhet, ez magyarázná számos sejtplazmai kölcsönhatását.

A feltételezés igazolására irodalomkutatást végeztünk. A krotonáz fokozott kifejeződése volt megfigyelhető májtumoros sejtek sejtplazmájában, ahol az közrejátszott a nyirok áttétek kialakulásában [Zhang és mtsai 2013]. Ezt a vonalat továbbfejtve az interakciós partnerek Gene Ontology [The Gene Ontology Consortium 2013] adatbázisban található

biológiai folyamatainak (*biological process*[49]) dúsulását vizsgáltuk a BiNGO [Maere és mtsai 2005] segítségével. A mitokondriális interaktorok elemzése során a katabolizmussal (GO:0009056) és apoptózis negatív szabályozásával (GO:0043066) összefüggő, valamint ezekkel kapcsolatos folyamatok dúsultak szignifikáns mértékben [Veres és mtsai 2015]. Ezzel egyetértésben kimutatták, hogy a krotonáz számos daganattípusban fokozott mértékben fejeződik ki [Yeh és mtsai 2006], és a kiütése esetén májtumorokban csökken a sejtek túlélőképessége és növekszik a ciszplatin kezelésre adott apoptotikus válasz [Zhu és mtsai 2013]. Hasonló szerepét mutatták ki emlő daganatos sejtvonalakban is, ahol a csökkent kifejeződése a PP2-által indukált apoptózist segítette elő [Liu és mtsai 2010].

Ezen megállapításokból feltételezhető, hogy a lokalizációs alapon biológiailag nem valószínű kapcsolatok magas száma a krotonáz tranziens és dinamikus sejtplazmatikus szubcelluláris lokalizációjának lehet a következménye, mely lokalizációban az eddigiekben csak feltételezetten krotonázzal összefüggő biológiai folyamatokban vesz részt, mint például az apoptózis gátlása. A szubcelluláris lokalizáció alapú funkciók felhívhatják a figyelmet olyan gyógyszeres intervenciós lehetőségekre, mint például a krotonáz sejtplazmatikus funkcióinak gátlása máj vagy emlő daganatokban.

6.2.2. Az MPS1 kináz lokalizáció specifikus interaktómának elemzése

Az MPS1 egy kettős specificitású fehérje, melynek szerin/treonin és tirozin kináz funkciója elsődlegesen a mitotikus ellenőrzőpont szabályozásával függ össze. Ezen funkciójának gátlása fontos terápiás célpont daganatokban [Jemaá és mtsai 2013], azonban a szerzett rezisztencia magas előfordulása miatt egyelőre limitált sikeraránnyal alkalmazható [Gurden és mtsai 2015].

Doktori tanulmányaim részeként három hónapot tölthettem ösztöndíjasként Londonban az Institute of Cancer Research daganatkutató intézet és a Cambridge-i Egyetem közös vendégkutatójaként, ahol fő feladatom a klinikailag releváns kináz mutációk predikciójára szolgáló molekuláris modellezésen alapuló algoritmus megalkotása volt. Példa fehérjeként az MPS1 kináz szerkezetét és mutációinak hatását vizsgáltam a kináz ATP kötő képességére nézve, mely vizsgálatok során merült fel a kináz lokalizáció specifikus funkciójának fontossága.

[49] http://www.geneontology.org/page/biological-process-ontology-guidelines/

A ComPPI adatainak böngészése során felmerült annak a lehetősége, hogy az MPS1 gátlásának limitált klinikai sikeressége, és a potenciális rezisztencia kialakulásának valószínűsége összefügghet a fehérje szubcelluláris lokalizációjával.

Az MPS1 domináns lokalizációja a középtest-asszociált sejtplazmatikus szubcelluláris lokalizáció [Fisk és mtsai 2003], de ismert sejtmagi elhelyezkedése is. A sejtmagban betöltött funkciója és terápiás válaszban betöltött szerepe azonban egyelőre tisztázatlan. A lokalizáció specifikus interaktóm vizsgálat során a fehérje számos, csak sporadikusan leírt funkcióját határoztuk meg.

Az MPS1 a mitotikus orsó összeszerelés ellenőrzőpontjának tagja (*spindle assembly checkpoint* (SAC)) [Musacchio és Salmon 2007], mely folyamat gátlása a mitotikus feltartóztatás (*mitotic arrest*) gátlásához és apoptózishoz vezető aneuplodiát eredményez [Kwiatkowski és mtsai 2010]. Az MPS1 szintén asszociálódhat a kinetokorral, és mint más kinetokor asszociált fehérjék a sejtmagi pórus komplexhez kötődhet [Liu és mtsai 2003]. Kimutatták, hogy a G2/M átmenet során az MPS1 a sejtplazmatikus oldalról a sejtmagi oldalra transzlokálódhat a sejtmaghártya lebontását megelőzően [Zhang és mtsai 2011]. A kondenzin-2 fehérje foszforilációján keresztül a kromatin szerveződésre kifejtett hatását szintén leírták [Kagami és mtsai 2014]. Sejtmagi lokalizációja esetén azonban kináz funkciója nélkülözhető, valamint a specifikus gátlószerek sem befolyásolják a sejtmagi lokalizációt [Zhang és mtsai 2011]. Ezen megfigyelések alapján felmerült, hogy az MPS1-nek a terápia után is fennmaradó funkciója lehet a sejtmagban.

A ComPPI 40 interakciós partnert listáz az MPS1 fehérjére keresés esetén[50]. A **18. ábra** az MPS1 és első szomszédjainak kapcsolatait mutatja a megbízható sejtmagi interaktorokra történő szűrés előtt és után (interakciós megbízhatósági érték > 0,9). A 40 kölcsönható partner 70%-a, összesen 28 fehérje rendelkezik sejtmagi lokalizációval, mely eredmény megerősíti az eddig ismertnél fontosabb sejtmagi funkció valószínűségét.

A sejtmagi funkció vizsgálatához megalkottuk az MPS1 első és második szomszédjainak interaktómát, majd a sejtmagi kölcsönhatásokra nézve szűrést végeztünk. A Gene Ontology [The Gene Ontology Consortium 2013] biológiai folyamatok dúsulásos elemzése ebben az esetben is a BiNGO [Maere és mtsai 2005] eszközzel történt, mely az ismert MPS1-el összefüggő funkciók, mint a sejtciklus szabályozása (GO:0022402),

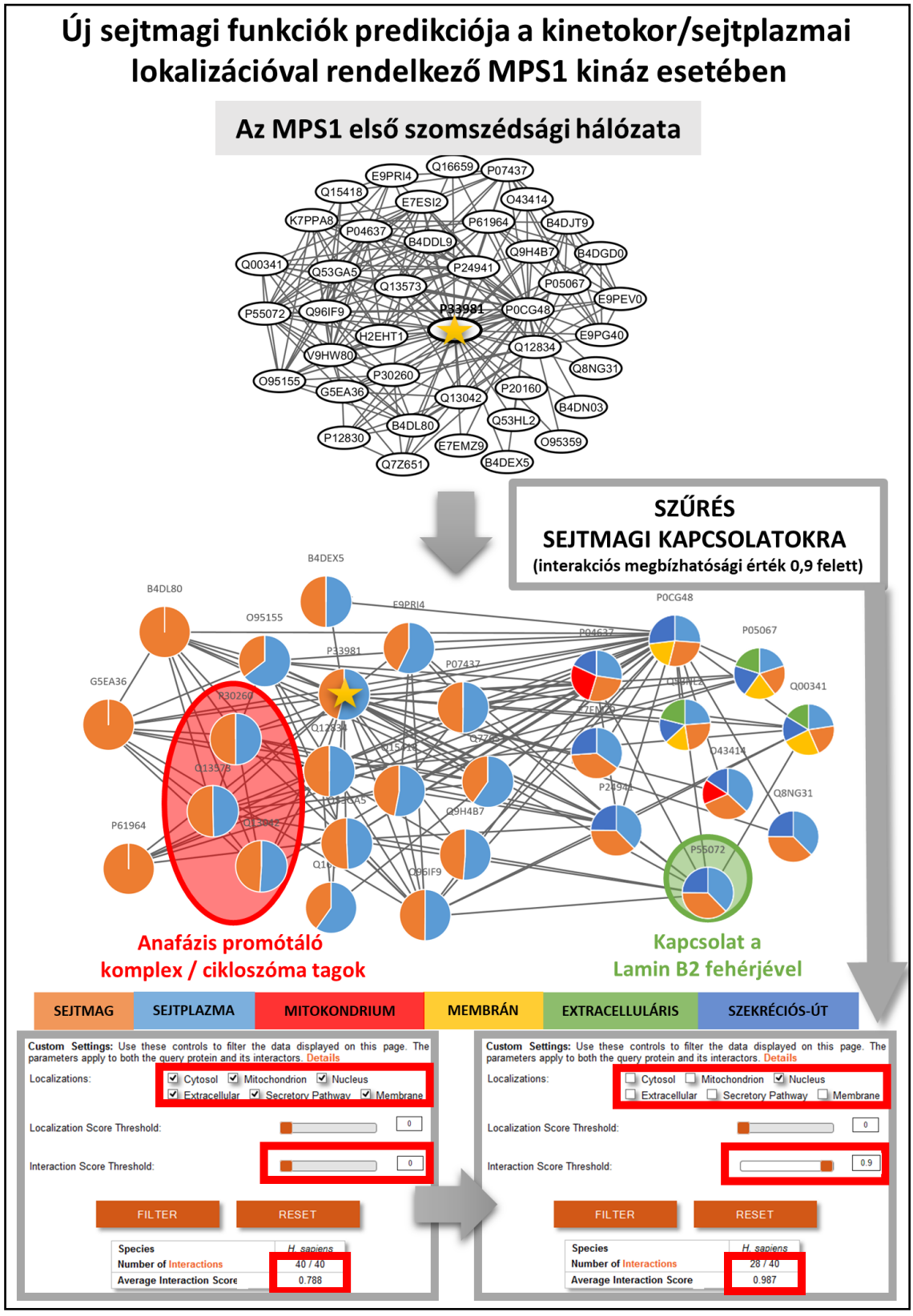
[50] http://comppi.linkgroup.hu/protein_search/interactors/P33981/

mellett új funkciók előfordulását is szignifikánsnak jelezte, mint a sejtes komponensek rendeződésében (GO:0016043) és a sejtorganellum elrendeződésében (GO:0006996) betöltött potenciális szerep.

Ezen tulajdonságok egybevetve az MPS1 korábban leírt funkcióival - így a sejtmagi pórus komplexhez asszociált lokalizációval [Liu és mtsai 2003], a sejtmaghártya lebontását megelőző sejtmagi lokalizációval [Zhang és mtsai 2011], illetve a kromatin szerveződésre kifejtett hatásával [Kagami és mtsai 2014] - felvetik az MPS1 szerepének lehetőségét a mitózis során bekövetkező sejtkomponens szintű újraszerveződésben, így például a sejtmag összeszerelésében. Ez utóbbi folyamatban ismert a nukleáris pórus komplex jelentősége, mely tovább erősíti a hipotézist [Kabachinski és Schwartz 2015].

Az MPS1 kapcsolatainak vizsgálata során három olyan második szomszédot találtunk, melyek kulcsszerepet játszanak a sejtmag összeszerelésében (Lamin-B2 [Dechat és mtsai 2008], Erbin [Gant és mtsai 1999] és Emerin [Haraguchi és mtsai 2001]). Három másik második szomszéd pedig a kromatin kondenzáció és dekondenzáció epigenetikai szabályozásában vesz részt (Aurora kináz B [Sabbattini és mtsai 2007] és C [Balboula és Schindler 2014], valamint a hiszton aciltranszferáz p300 [Wu 1997]).

DOI:10.14753/SE.2018.2134



***18. ábra**: **Az MPS1 lokalizáció specifikus funkciói**. Az MPS1 kináz vizsgálata során először megvizsgáltuk, hogy hány interakciós partnere található a sejtmagban. Csak azokat a kapcsolatokat vettük figyelembe, ahol az* *(folytatás a következő oldalon ->)*

(-> folytatás az előző oldalról) interakciós megbízhatósági érték 0.9 felett volt, így a szűrés után a kezdeti 40 kölcsönható partner száma 28-ra csökkent. Megvizsgáltuk a mitokondriális lokalizációval rendelkező partnerek lokalizációs viszonyait, melyet a színkódok mutatnak. A részletes elemzés során a piros körrel jelölt anafázis promótáló komplex tagjai, illetve a zölddel jelölt VCP fehérje emelkedett ki, mely közvetett kapcsolatot biztosít a Lamin B2 felé. Forrás: [Veres és mtsai 2015]

A Lamin B2 két első szomszédon keresztül is kapcsolódik az MPS1-hez, melyek közül az egyik a VCP fehérje. A VCP ismert szabályozója a sejtmaghártya újrarendeződésének [Güttinger és mtsai 2009], valamint szerepe van a DNS károsodásra adott válaszban is [Meerang és mtsai 2011], mely funkcióval az MPS1-t is összefüggésbe hozták [Maachani és mtsai 2015].

Mindezek mellett az MPS1 három első szomszédja is tagja az anafázis promótáló komplexnek (APC/C), mely komplex a Cdc20 fehérje ubiquitinációján keresztül kulcsfontosságú a mitotikus orsó összeszerelés ellenőrzőpontjának szabályozásában [Nilsson és mtsai 2008]. Az APC/C feladata az MPS1 degradációjának a mediálása is, mely segíti a sejtciklusba való visszatérés koordinációját környezeti stressz esetén [Ostapenko és mtsai 2012]. Ez fontos menekülési útvonal lehet a daganatos sejtek számára is.

Emellett érdekes megfigyelés, hogy az MPS1 első szomszédjai közül három rendelkezik mitokondriális lokalizációval, mely felveti az MPS1 mitokondriális lokalizációjának lehetőségét. Ezt megerősítő evidencia az MPS1 kapcsolódása a VDAC1 ion-csatornához, ezáltal bekerülése a mitokondriumba, mely lokalizációban a citokróm C mediálta apoptózist szabályozza [Zhang és mtsai 2016].

Mindezek alapján az MPS1 szerepe a sejtmag összeszerelésében további vizsgálatot és kísérletes megerősítést igényel, azonban a felsorolt példák alapján jól látható az MPS1 szubcelluláris lokalizáció szintű szabályozása, és ennek a kényes egyensúlynak a felborulása daganatos sejtekben, melynek szerepe lehet a terápiára adott válaszkészségben is.

DOI:10.14753/SE.2018.2134

6.3. A ComPPI adataira épülő rendszerszintű fehérje transzlokációs adatbázis bemutatása

6.3.1. A Translocatome adatbázis általános bemutatása

A fehérjék szubcelluláris kompartmenteken belüli megoszlása dinamikusan változik, mely megoszlást alapvetően befolyásolja a sejten belüli jelátvitel. A fehérjék szabályozott, jelátviteli folyamatok hatására bekövetkező helyváltoztatását nevezzük transzlokációnak, melynek a korábban bemutatott rendszerbiológiai definíció értelmében funkcionális jelentőséget is tulajdonítunk.

A ComPPI adatkészlete egyszerre tartalmazza a fehérjék részletes szubcelluláris lokalizációs adatait, illetve az egyes fehérjék közti fizikális kölcsönhatásokat. Ez alapján lehetőség van a fehérjék interaktómát nemcsak a teljes sejt szintjén, hanem az egyes kompartmentekre specifikusan is vizsgálni, mint ahogy az egyes fehérjék példáin ezt bemutattam. A transzlokálódó fehérjék fontos elemei a kompartment szintű interaktómnak, elengedhetetlen szabályozó funkcióval, gyakran patológiás szereppel. A ComPPI adatainak felhasználásával fejlesztés alatt áll új adatbázisunk, mely a transzlokálódó fehérjéket és azok kölcsönhatásait tartalmazza. Ezt az adatbázist Translocatome-nak neveztük el [Dobronyi és mtsai 2016, Mendik és mtsai 2017].

Mivel nem áll rendelkezésre átfogó adatkészlet arra vonatkozóan, hogy mely fehérjék transzlokálódnak, és melyek nem, így elsődleges célunk ezen fehérjék összegyűjtése, illetve újak prediktálása volt. A transzlokáció valószínűségét a rendszerbiológiai definíció értelmében a hálózatban betöltött szerep és biológiai funkciók alapján határozzuk meg. Ehhez elengedhetetlen a fehérjék kölcsönhatási hálózatának ismerete, illetve az egyes fehérjékhez tartozó biológiai funkciók hozzárendelése.

A fehérjék kölcsönhatásai a ComPPI adatbázisból kerülnek importálásra, beiktatva egy szűrést kizárólag azon emberi UniProt SwissProt [The UniProt Consortium 2017] fehérjék kapcsolataira, melyeknek legalább egy kísérletesen is megerősített szubcelluláris lokalizációja van, az azonban nem feltétel, hogy a kölcsönható partnereknek legyen közös lokalizációjuk. Az így kapott hálózat a ComPPI v1.1 adataira alapulva 12.806 fehérjét, és azok 145.486 fizikai kapcsolatát tartalmazza. (A ComPPI frissítése a dolgozat elkészítésének idejében még folyamatban van, így a Translocatome publikálása során a legfrissebb elérhető adatokat fogja tartalmazni.)

Ahhoz, hogy a fehérjék transzlokációs valószínűségét szisztematikusan meg tudjuk határozni, szükségünk van egy megbízható kézzel gyűjtött adatkészletre, mely tartalmaz biztosan transzlokálódó (pozitív adatkészlet) és biztosan nem transzlokálódó (negatív adatkészlet) fehérjéket, segítve a hasonló tulajdonságú fehérjék elkülönítését. Ehhez részletesen rögzítjük az irodalmi gyűjtésből származó egyes fehérje tulajdonságokat, melyek később felhasználhatóak az eredmények ellenőrzéséhez, és az egyes transzlokálódó fehérjék szerepének mélyrehatóbb vizsgálatához.

A negatív és pozitív adatkészlettel lehetőségünk van gépi tanulás segítségével egy transzlokációs valószínűséget rendelni a ComPPI-ból származó fehérjékhez, melyhez az interaktóm hálózatos paraméterei mellett a fehérjék biológiai funkcióira is szükségünk van. Ezen funckiókat automatikusan rendeljük hozzá a fehérjékhez a Gene Ontology adatainak felhasználásával [The Gene Ontology Consortium 2013].

A ComPPI adatbázisból származó fehérjelista, a fehérjék közti kölcsönhatások, a kézzel gyűjtött fehérjék és azok tulajdonságai, valamint az automatikusan hozzárendelt fehérje funkciók adják a Translocatome adatbázis adatait, melyet a fejlesztés alatt álló webes felületen keresztül lehet böngészni, illetve letölteni (http://translocatome.linkgroup.hu/).

6.3.2. A transzlokálódó fehérjék kézi adatgyűjtése

A Translocatome adatbázis alapja a kézzel gyűjtött fehérjekészlet, mely részletes információkat tartalmaz a biztosan transzlokálódó fehérjékről. Ezek lehetnek egészséges sejtben előforduló áthelyeződések, azonban előfordulnak kizárólag patológiás körülmények között leírt transzlokációk is, melyeket nem használunk fel a pozitív adatkészletben [Dobronyi és mtsai 2016, Mendik és mtsai 2017].

A kézzel gyűjtött készlet folyamatosan növekszik, ezzel is segítve a minél jobb predikciós lehetőségeket. A dolgozat készítésének időpontjában az adatbázis több mint 200 transzlokálódó fehérjére tartalmazott adatokat, mely tulajdonságokat a **6. táblázat** mutatja be egy példafehérjén keresztül.

6. táblázat: A kézzel gyűjtött transzlokálódó fehérjék rögzített tulajdonságai

Adatmező	Leírás	Példa
UniProtAC	A rögzített fehérje UniProt neve.	Q05397
Gén név	A rögzített fehérje gén neve.	FAK | FAK1 | PTK2
UniProt teljes név	A rögzített fehérje teljes neve.	Fokális adhéziós kináz 1
Referencia	A forrás cikk PubMed azonosítója.	26056081
A Lokalizáció	A rögzített fehérje A szubcelluláris lokalizációja nagy lokalizációs csoportokba osztva. Gene Ontology *cellular component term* azonosító használata ajánlott.	Nagy lokalizációs csoport: sejtplazma (GO:0005737) Részletes lokalizáció: fokális adhézió (GO:0005925)
B Lokalizáció	A rögzített fehérje B szubcelluláris lokalizációja nagy lokalizációs csoportokba osztva. Gene Ontology *cellular component term* azonosító használata ajánlott.	Nagy lokalizációs csoport: sejtmag (GO:0005634)
C Lokalizáció	A rögzített fehérje C szubcelluláris lokalizációja nagy lokalizációs csoportokba osztva. Gene Ontology *cellular component term* azonosító használata ajánlott.	Nincs rá adat.
Detekciós módszer	A lokalizáció meghatározására használt kísérletes módszer.	Immunhisztokémia

6. táblázat: A kézzel gyűjtött transzlokálódó fehérjék rögzített tulajdonságai (folytatás)

Adatmező	Leírás	Példa
A Transzlokációs mechanizmus	A transzlokáció mechanizmusa adott két nagy lokalizáció között.	Nincs rá adat.
B Transzlokációs mechanizmus	A transzlokáció mechanizmusa adott két nagy lokalizáció között.	Nincs rá adat.
Szerkezeti információ	Szerkezeti információ a transzlokáció mechanizmusával összefüggésben, így például lokalizációs szignálok jelenléte és elhelyezkedése.	Nukleáris lokalizációs szignál (NLS) a FERM doménben
Biológiai folyamatok az A lokalizációban	Biológiai folyamatokban betöltött szerep az A lokalizációban. Gene Ontology *biological process* azonosító használata ajánlott.	sejtmigráció pozitív szabályozása (GO:0030335) | sejtosztódás szabályozása (GO:0042127) | sejt alak szabályozása (GO:0008360) | fokális adhézió kialakulásának szabályozása (GO:0051893)
Biológiai folyamatok a B lokalizációban	Biológiai folyamatokban betöltött szerep a B lokalizációban. Gene Ontology *biological process* azonosító használata ajánlott.	transzformáló növekedési faktor béta2 termelés serkentése (GO:0032915) | kemokin (C-C motívum) ligand 5 termelés serkentése (GO:0071651)

6. táblázat: A kézzel gyűjtött transzlokálódó fehérjék rögzített tulajdonságai (folytatás)

Adatmező	Leírás	Példa
Biológiai folyamatok a C lokalizációban	Biológiai folyamatokban betöltött szerep a C lokalizációban. Gene Ontology *biological process* azonosító használata ajánlott.	Nincs rá adat.
Interakciók az A lokalizációban	Interakciós partnerek listája az A lokalizációban. UniProt SwissProt nevezéktan használata ajánlott.	Nincs rá adat.
Interakciók a B lokalizációban	Interakciós partnerek listája a B lokalizációban. UniProt SwissProt nevezéktan használata ajánlott.	P20226 (TFIID) | P05412 (JUN) | P19838 (NFKB1) | P37231 (PPARG) | P14859 (OKT1) | P10914 (IRF1) | A6NHT5 (NKX5-1)
Interakciók a C lokalizációban	Interakciós partnerek listája a C lokalizációban. UniProt SwissProt nevezéktan használata ajánlott.	Nincs rá adat.
Jelátviteli útvonal	Jelátviteli útvonalak listája, melyekben az adott fehérje szerepet játszik.	Nincs rá adat.
Patológiás szerep	Az adott fehérje patológiás szerepe, amennyiben nem kizárólag fiziológiás transzlokációról van szó.	T-reg sejtek akkumulációján keresztül a daganatos sejtek immuntoleranciájának kialakulásában van szerepe.

6. táblázat: A kézzel gyűjtött transzlokálódó fehérjék rögzített tulajdonságai (folytatás)

Adatmező	Leírás	Példa
Betegség csoport	Az adott fehérje patológiás szerep esetén milyen betegség csoportban játszik szerepet.	Daganatos betegségek
Konkrét betegség	Az adott fehérje patológiás szerep esetén milyen konkrét betegségben játszik szerepet.	Laphám karcinóma, kolorektális daganat

A táblázat a transzlokálódó fehérjék gyűjtése során kézzel rögzített adatokat listázza, és mutatja be a Fokális Adhéziós Kináz (FAK) példáján. Ezen adatok csak a kézi gyűjtés eredményét tartalmazzák, az automatikusan annotált információkat nem, mely tovább bővítheti az elérhető adatmennyiséget.

A fehérjék adatait egységesen rögzítjük, így segítve elő a folyamatos bővítést, illetve a kiegészítést más forrásokból származó adatokkal. Ez a megoldás egyúttal segíti a Translocatome kézzel gyűjtött adatainak integrálását más, külső adatbázisokba is.

6.3.3. A fehérjék transzlokációjának predikciója tanuló algoritmussal

A transzlokálódó fehérjék kézi gyűjtése magas minőségű adatkészletet eredményez. Alkalmazható egy gépi tanuló algoritmus pozitív adatkészleteként, mely egy adott fehérjehalmazból a kiválasztott paraméterek hasonlósága alapján meg tudja állapítani, milyen valószínűséggel tartozik bele egy fehérje a keresett csoportba. Ahhoz, hogy ezt az elemzést el tudjuk végezni szükség van a pozitív mellett egy negatív adatkészletre is, illetve meg kell határozni azon paramétereket, melyeket az algoritmus a klasszifikáció során felhasznál [Dobronyi és mtsai 2016, Mendik és mtsai 2017].

A pozitív adatkészlet a bizonyosan transzlokálódó fehérjék közül azokat tartalmazza, melyek egészséges körülmények között is előfordulnak a sejten belül, hiszen ilyen transzlokációra jellemző fehérjéket keresünk. A negatív adatkészlet szintén egészséges

sejtekben a bizonyosan nem transzlokálódó fehérjéket tartalmazza, melyek egyaránt lehetnek csak egy lokalizációval rendelkezők, vagy éppen olyan multikompartment fehérjék, melyek nem rendelkeznek különböző funkcióval az egyes kompartmentekben. Az adatkészlet a dolgozat elkészítésekor több mint 200 fehérjét tartalmazott a pozitív, és több mint 140 fehérjét a negatív adatkészletben.

A gépi tanulás során azt elemezzük, hogy egy meghatározott paraméter halmazt vizsgálva mely tulajdonságok azok, melyek mentén a legnagyobb bizonyossággal el tudjuk választani egymástól a negatív és pozitív adatkészletben szereplő fehérjéket. A kapott paraméterek mentén a többi fehérjét is be tudjuk sorolni a transzlokációs valószínűségi érték *(*úgynevezett *evidence score)* meghatározásának segítségével, mely a gépi tanuló algoritmus által számolt 0 és 1 közötti valószínűség, ahol 0 a biztosan nem transzlokálódó, és 1 a biztosan transzlokálódó fehérjékhez rendelt érték. Ehhez a pozitív és negatív adatkészletet két részre osztjuk, egy tanuló és egy validáló halmazra, és az algoritmus minőségét azzal mérjük vissza, hogy a tanuló fehérje halmazok alapján megállapított paraméterek segítségével milyen bizonyossággal kapjuk vissza a validáló halmaz elvárt tulajdonságait, azaz a pozitív teszt adatok valóban pozitívként, míg a negatívak negatívként jelennek meg.

A paramétereket a korábban bemutatott rendszerbiológiai szemléletű transzlokálódó fehérje definíció mentén határoztuk meg, így egyrészt hálózatos mérőszámokat (a fehérjék fokszáma, illetve a hídsági és köztiségi központiság mérőszámok), másrészt biológiai tulajdonságokat (Gene Ontology biológiai folyamat[51], molekuláris funkció[52], sejten belüli komponens[53] [The Gene Ontology Consortium 2013]) vettünk figyelembe. A hálózatos mérőszámok esetében feltételezésünk szerint a transzlokálódó fehérjék általában magasabb fokszámmal, azaz több szomszéddal szerepelnek a hálózatban, illetve számos útvonal halad rajtuk keresztül fontos szabályozó funkciójuk miatt. A biológiai folyamatok terén a nagyobb számú, és egymástól eltérő folyamatokban való részvételt feltételezzük.

A gépi tanulás során több algoritmust is kipróbáltunk, melyek közül az SVM és a neurális hálózat (*neural network*) hozta a legjobb eredményeket az eddigi tesztek alapján. A tesztelés során a tanuló és validáló adatkészletet véletlenszerűen választva a negatív és

[51] http://geneontology.org/page/biological-process-ontology-guidelines/
[52] http://geneontology.org/page/molecular-function-ontology-guidelines/
[53] http://geneontology.org/page/cellular-component-ontology-guidelines/

pozitív adatkészletből 1000 futtatás után nézzük meg a tanulás pontosságát. Eddigi eredményeink alapján legjobban (80% feletti átlagos pontosság) a neurális hálózat klasszifikálja a transzlokálódó fehérjéket (**7. táblázat**). Az algoritmus további pontosítására is lehetőség nyílik a biológiai hűség növelésén keresztül, így például a tanuló és validáló adatkészletek növelésével, illetve a neurális hálózat által kiválasztott jellemző paraméterek biológiai értelmének vizsgálatával [Dobronyi és mtsai 2016, Mendik és mtsai 2017].

7. táblázat: A gépi tanuló algoritmus transzlokáció predikciójának előzetes eredményessége

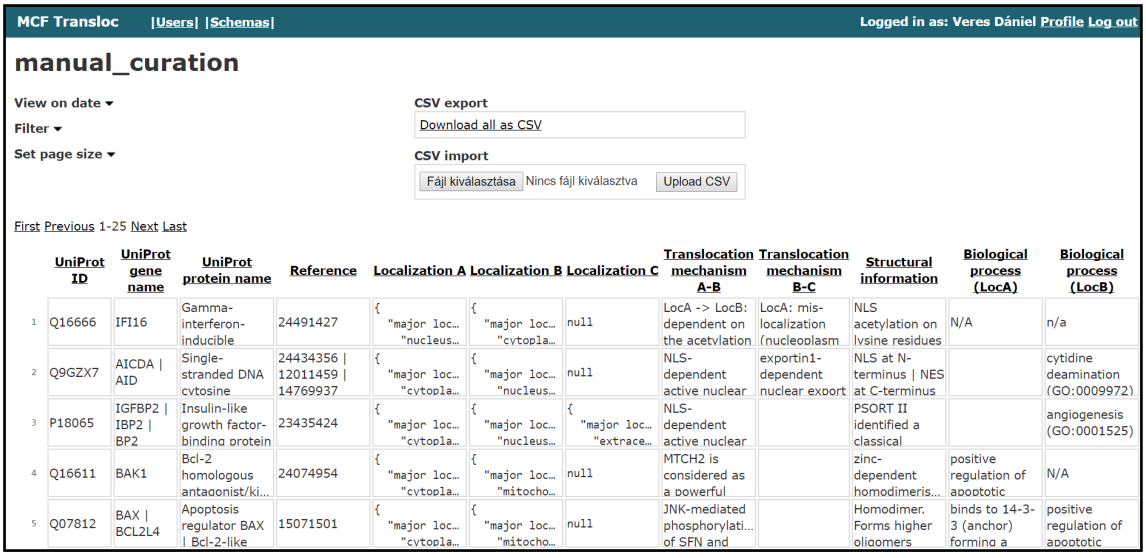
UniProt	Fehérje név	Adatkészlet	Neurális hálózat	SVM	Döntési fa	Legközelebbi szomszédok
P62805	H4 hiszton	negatív	1	-1	-1	1
P00747	Plazminogén	negatív	-1	-1	1	1
P25067	Kollagén alfa-2(VIII) lánc	negatív	-1	-1	-1	-1
P02545	Prelamin-A/C	negatív	1	1	1	1
P02768	Szérum albumin	negatív	-1	-1	-1	1
P42684	Abelson tirozin-fehérje kináz 2	pozitív	1	1	1	1
Q05397	Fokális adhéziós kináz 1	pozitív	1	1	1	1
P56945	Emlő tumor antiösztrogén rezisztencia fehérje 1	pozitív	1	-1	1	1
Q16665	Hipoxia indukáló faktor 1-alfa	pozitív	1	1	1	1
P04150	Glükokortikoid receptor	pozitív	1	1	1	1
Átlagos predikciós hatékonyság:			**85%**	**79%**	**77%**	**70%**

A táblázat 10 példa fehérjén mutatja be a négy gépi tanuló algoritmus klasszifikációjának eredményét. A valós találatok zölddel, míg a hamisak pirossal vannak kiemelve. Jól látható, hogy a predikciók hatékonysága sokkal jobb a pozitív adatkészlet esetében, *(folytatás a következő oldalon ->)*

(-> folytatás az előző oldalról) azaz a tanuló algoritmus jelen stádiumban több hamis pozitív transzlokálódó fehérjét prediktál, mint amennyi hamis negatív eredményt ad. Az átlagos eredmény mutatja, hogy a neurális hálózat túlteljesíti a többi algoritmust. Az adatkészletek növelésével és a neurális hálózat fejlesztésével az előzetes eredményeket meghaladó teljesítmény várható a klasszifikációs algoritmustól [Dobronyi és mtsai 2016, Mendik és mtsai 2017].

6.3.4. A Translocatome közösségi adatfejlesztésre is alkalmas webes felülete

A Translocatome adatbázis webes felülete a felhasználóbarát böngészési és adatletöltési lehetőségek mellett, a ComPPI-tól eltérően az adatok közösségi fejlesztésére is lehetőséget fog adni. Ehhez kifejlesztettünk egy webes eszközt (**19. ábra**), mely segíti az adatok kézi gyűjtését, ellenőrzését, és a felhasználói jogosultságok kezelésével lehetőséget ad az adatok szerkesztésére egyszerre számos felhasználó számára [Dobronyi és mtsai 2016, Mendik és mtsai 2017].



	UniProt ID	UniProt gene name	UniProt protein name	Reference	Localization A	Localization B	Localization C	Translocation mechanism A-B	Translocation mechanism B-C	Structural information	Biological process (LocA)	Biological process (LocB)
1	Q16666	IFI16	Gamma-interferon-inducible	24491427	{ "major loc… "nucleus…	{ "major loc… "cytopla…	null	LocA -> LocB: dependent on the acetylation	LocA: mis-localization (nucleoplasm	NLS acetylation on lysine residues	N/A	n/a
2	Q9GZX7	AICDA | AID	Single-stranded DNA cytosine	24434356 | 12011459 | 14769937	{ "major loc… "cytopla…	{ "major loc… "nucleus…	null	NLS-dependent active nuclear	exportin1-dependent nuclear export	NLS at N-terminus | NES at C-terminus		cytidine deamination (GO:0009972)
3	P18065	IGFBP2 | IBP2 | BP2	Insulin-like growth factor-binding protein	23435424	{ "major loc… "cytopla…	{ "major loc… "nucleus…	{ "major loc… "extrace…	NLS-dependent active nuclear		PSORT II identified a classical		angiogenesis (GO:0001525)
4	Q16611	BAK1	Bcl-2 homologous antagonist/ki…	24074954	{ "major loc… "cytopla…	{ "major loc… "mitocho…	null	MTCH2 is considered as a powerful		zinc-dependent homodimeris…	positive regulation of apoptotic	N/A
5	Q07812	BAX | BCL2L4	Apoptosis regulator BAX | Bcl-2-like	15071501	{ "major loc… "cytopla…	{ "major loc… "mitocho…	null	JNK-mediated phosphorylati… of SFN and		Homodimer. Forms higher oligomers	binds to 14-3-3 (anchor) forming a	positive regulation of apoptotic

19. ábra: A Translocatome kézi adatgyűjtő felületének képe. *A webes felület segítségével rendszerezve és verzió követve lehet majd kézzel adatokat rögzíteni, mely elősegíti a fehérjék tulajdonságainak gyűjtését számos felhasználó számára párhuzamosan. Az ábra öt példa fehérje esetében mutatja be az adatok egy részének megjelenítését a beviteli felületen.*

DOI:10.14753/SE.2018.2134

A felület közvetlenül az adatbázis tartalmát jeleníti meg különböző nézetekben, mely adatbázisban bekövetkező változások másodperces pontossággal visszakereshetők, így adva lehetőséget az adatok korábbi időpontban való megtekintésére vagy visszaállítására, illetve a bevitt változtatások listázására. Az adatok felvitelére, illetve letöltésére egyaránt lehetőség van szöveges dokumentumként (CSV) is.

A Translocatome adatait meghívásos, illetve jelentkezéses alapon lehet szerkeszteni, ezzel ellenőrizve, hogy csak kompetens felhasználók vigyenek be adatokat vagy ellenőrizzék a mások által rögzített információkat. Az újonnan rögzített adatokat a tervek szerint azok mennyiségétől függően, ellenőrzést követően fél évente fogjuk az éles, publikus felhasználói felületen is megjeleníteni.

DOI:10.14753/SE.2018.2134

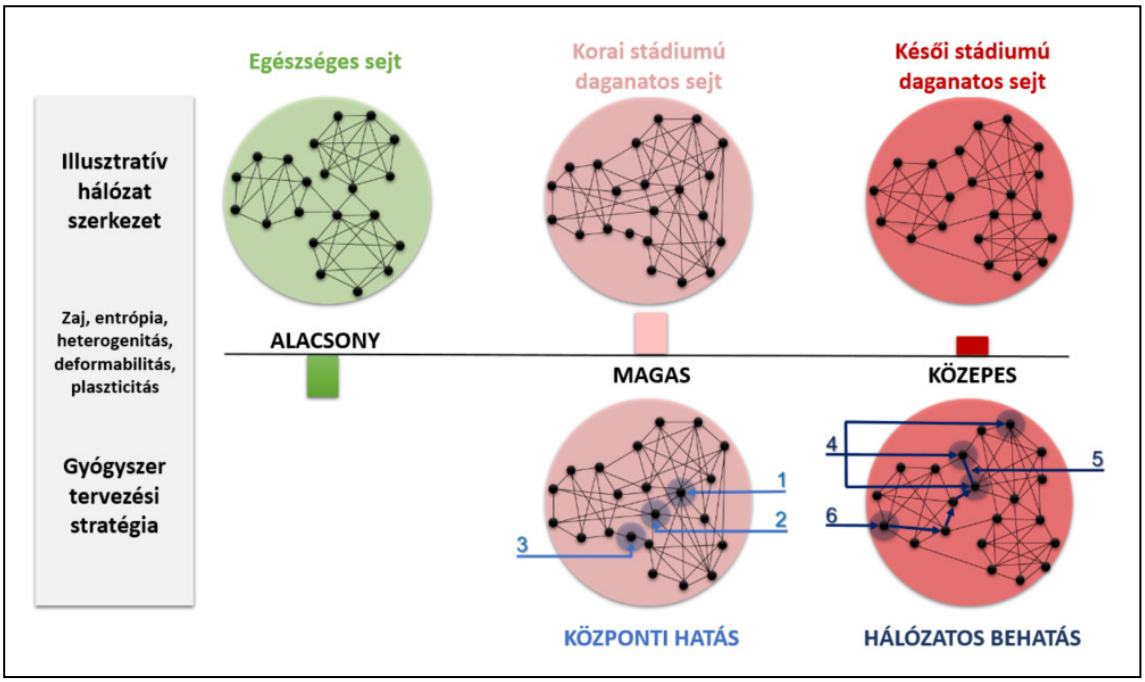
6.4. A fehérjék térbeli elhelyezkedésének szerepe a daganatos malignitás meghatározásában

6.4.1. A malignus transzformáció kétlépcsős hipotézise

A daganatok kialakulását egyre inkább rendszerszintű eseménynek tekintjük, melyet a molekuláris hálózatok kényes egyensúlyának eltolódása eredményez [Hornberg és mtsai 2006]. A kutatócsoportunk által készített, az elérhető irodalmi evidenciát összegző hipotézis szerint a daganatok rosszindulatú átalakulása felfogható egy kétlépéses folyamatként. Ebben a molekuláris hálózat a kezdeti állapotból először egy flexibilisebb, plasztikusabb állapotba megy át a korai stádiumú daganatokban, majd onnan visszakerül egy új, rigidebb szerkezetbe, mely stabilizálja a kései stádiumú daganatos fenotípust [Gyurkó és mtsai 2013, Csermely és Korcsmáros 2013, Csermely és mtsai 2013, Csermely és mtsai 2015]. Ez a végső daganatos fenotípus még mindig plasztikusabb, mint a kiindulási állapot, azonban sokkal rigidebb, mint a karcinogenezis köztes állapotainak molekuláris hálózata. Ennek feltétele az, hogy a daganatos sejt először kiszakadjon a megszokott mikrokörnyezetből, és ehhez a plasztikus hálózatos beállítás maximalizálja a lehetőségeket. Az útkeresés időszaka után azonban a független daganatos fenotípust konzerválva már stabilabb állapotba kerül, amennyiben nem szükséges fenntartania a sérülékenyebb, ugyanakkor flexibilisebb plasztikus állapotot.

A daganatok korai stádiumában látható fokozott hálózatos plaszticitás megfeleltethető a klonális expanzió jelenségének, mely során megjelennek a daganatos átalakulást inicializáló sejtek. A kései stádiumot jellemző sejtek lehetnek elsődleges daganatos sejtek, vagy olyan áttétet képző sejtek, melyek már megtelepedtek az új szövetben és hozzászoktak annak mikrokörnyezetéhez. Ezen megfigyelések segíthetik a molekuláris hálózatok viselkedésének megfelelő, kiemelten a plasztikus-rigid átmenetet célzó daganatos biomarkerek keresését [Chen és mtsai 2012], illetve a megfelelő daganat ellenes terápiák tervezését [Gyurkó és mtsai 2013] (**20. ábra**).

DOI:10.14753/SE.2018.2134



20. ábra: A daganatok kétlépcsős fejlődésének hálózatos illusztrációja, bemutatva az eltérő terápiás megközelítés lehetőségét. *A daganatok molekuláris hálózatának tulajdonságai eltérnek a korai és késői stádiumú daganatos sejtek esetében, mely eltérések egyben különböző terápiás beavatkozási pontokat is jelentenek. A korai fázist jellemző magas plaszticitással rendelkező hálózatban a csomópontok (1-es jelölés), az egyes hálózatos csoportok közti hidak (2-es jelölés), és ezen csoportok közti információ terjedésben fontos, szűk keresztemetszetet biztosító elemek (3-as jelölés) lehetnek az optimális célpontok. Ezzel szemben a rigidebb késői stádiumban a több molekulát célzó beavatkozások (4-es jelölés), az egyes fontos kölcsönhatásokat módosító (5-ös jelölés), vagy a hálózatos hatást kihasználva indirekt módon ható gyógyszerek biztosíthatják a leghatékonyabb beavatkozást. A heterogén daganatos populáció egyszerre tartalmazhat korai és késői stádiumnak megfelelő daganatos sejteket, így az a kombinált gyógyszeres beavatkozás vezethet a legeredményesebb kezeléshez, mely egyszerre vagy szekvenciálisan alkalmazza mind a központi hatáson, mind a hálózatos behatáson alapuló gyógyszereket. Forrás: [Gyurkó és mtsai 2013]*

DOI:10.14753/SE.2018.2134

6.4.2. A daganatok kétlépcsős fejlődési modelljét támogató molekuláris megfigyelések

A daganatok kétlépcsős fejlődéséről szóló hipotézisünket számos olyan molekuláris megfigyelés is alátámasztja, amelyek a fehérjék szubcelluláris lokalizációjával is összefüggenek. A bemutatott példák olyan fehérjék (**8. táblázat**), melyek a ComPPI és Translocatome adatbázisok építése és kézi adatgyűjtése során kerültek elemzésre, és az irodalmi előzmények olyan szubcelluláris lokalizáció specifikus-funkciót tulajdonítanak nekik, illetve olyan lokalizáció-specifikus funkciót prediktáltunk számukra a munkánk során, amely fontos lehet e fehérjéknek a daganatos malignitás kialakulásában játszott szerepével kapcsolatban.

Az MPS1 és ERK2 kinázok a malignus proliferációt a szubcelluláris lokalizációjuktól függő, különböző funkciókkal segítik. A FAK1 kináz, a hTERT telomeráz, a NANOG transzkripciós faktor, a P53 tumor szuppresszor és a ZEB1 transzkripciós faktor kontextus-függően segítheti a rákos sejtek proliferációját és/vagy metasztázisát. A bemutatott fehérjék esetében fontos irodalmi adatok utalnak arra, hogy hatásuk mértéke szubcelluláris lokalizációjuk függvénye lehet. E molekuláris jelenségeket amerikai, norvég és svéd kollégák vastagbél tumorral kapcsolatos új epidemiológiai adatai értékelésének segítésére is felhasználtuk [Adami és mtsai 2017].

8. táblázat: Példák a daganatos malignitás kialakulásával összefüggő, lokalizáció-specifikus funkcióval rendelkező fehérjékre

Gén név	UniProtAC	ComPPI / Translocatome	Lokalizáció
A daganatos proliferációt lokalizáció-specifikus funkcióval elősegítő fehérjék 1.			
MPS1	P33981	ComPPI	sejtplazma, sejtmag
Az MPS1 kináz esetében kutatásaink során felmerült [Veres és mtsai 2015], hogy kompartment alapú biológiai funkcionalitása lehet. A kölcsönhatásokat megvizsgálva kiderült, hogy a dominánsan sejtplazmai MPS1 sejtmagi lokalizációjában kináz funkciójától független, eddig nem tisztázott szereppel rendelkezik [Zhang és mtsai 2016]. A kölcsönható partnerek biológiai funkciójának vizsgálata felvetette annak lehetőségét, hogy az MPS1 részt vesz a sejtes organellumok szerveződésének kialakulásában, kiemelten pedig a sejtmagmembrán organizációjában, *(folytatás a következő oldalon ->)*			

8. táblázat: Példák a daganatos malignitás kialakulásával összefüggő, lokalizáció-specifikus funkcióval rendelkező fehérjékre (folytatás)

Gén név	UniProtAC	ComPPI / Translocatome	Lokalizáció
A daganatos proliferációt lokalizáció-specifikus funkcióval elősegítő fehérjék 1.			
MPS1	P33981	ComPPI	sejtplazma, sejtmag
(-> folytatás az előző oldalról) melyet a sejtmagi pórus komplexhez kötődését leíró evidencia is megerősít [Liu és mtsai 2003]. Mindezen funkciók hasznos betekintést adhatnak az MPS1 szubcelluláris lokalizáció specifikus biológiai funkcióiba, mely az MPS1 gátlók daganatos betegségekben történő felhasználásnak jobb megértését is segíthetik.			
Gén név	**UniProtAC**	**ComPPI / Translocatome**	**Lokalizáció**
A daganatos proliferációt lokalizáció-specifikus funkcióval elősegítő fehérjék 2.			
ERK2	P28482	ComPPI és Translocatome	sejtplazma, sejtmag
Az ERK2 fehérje a sejtplazmában foszforilációs partnerein keresztül negatív és pozitív szabályozó körök résztvevője, funkciója pedig szabályozható a dimer képződésen keresztül. A sejtmagba történő áthelyeződés után transzkripcionális represszor funkciót tölt be a [GS]AAA[GC] konszenzus szekvenciához kötődve, mely funkció teljesen független a sejtplazmai kináz aktivitástól. A sejtmagban számos gén promóteréhez kötődik, többek között az interferon gamma-indukálta gének kifejeződését csökkenti [Hu és mtsai 2009a], mely az ERK2 fokozott aktivációja esetén az interferon által mediált tumor ellenes aktivitást csökkenti [Parker és mtsai 2016]. Az ERK fehérje jelátvitelben betöltött szerepe alapvetően szubcelluláris lokalizáció függő, és finoman szabályozott a fehérje foszforiláltságának a segítségével. Megfigyelhető, hogy ez a szabályozás szorosan összefügg a daganatos sejtek malignitásának mértékével is, bemutatva a szubcelluláris lokalizáció fehérje funkcióban és jelátvitelben betöltött általános és esszenciális szerepét, illetve jelentőségét a daganatos progresszió megítélésében.			

DOI:10.14753/SE.2018.2134

8. táblázat: Példák a daganatos malignitás kialakulásával összefüggő, lokalizáció-specifikus funkcióval rendelkező fehérjékre (folytatás)

Gén név	UniProtAC	ComPPI / Translocatome	Lokalizáció
A proliferációt és a metasztázist valószínűleg lokalizáció-specifikusan elősegítő fehérjék 1.			
FAK1	Q05397	ComPPI és Translocatome	sejtplazma, sejtmag
A fokális adhéziós kináz (FAK) egy tirozin kináz fontos sejtmigráció és proliferáció szabályozó szereppel. A fehérje hagyományos funkciója sejtplazmai, részt vesz az integrin és növekedési faktor receptoriális jelátvitelben. Sejtmagi szerepének ismerete az elmúlt években lett egyre kiemelkedőbb, mely lokalizációban fokozza a P53 és GATA4 degradációját ubiquitináción keresztül, mely fokozott sejtosztódáshoz és csökkent gyulladásos válaszhoz vezet. A FAK szintén működhet ko-transzkripcionális szabályozóként, így szerepe lehet a transzkripcionális szabályozásban is. A FAK jelátvitel a fokális adhéziótól a sejtmagig tehát fontos szabályozó tengelye a sejteknek, melynek befolyásolása terápiás célpont is lehet. Megfigyelték, hogy a FAK kináz funkciójának gátlása a sejtmagi funkciót nem érinti, sőt fokozott sejtmagi felhalmozódást eredményez. Mindez azt mutatja, hogy a FAK lokalizáció specifikus gátlása elengedhetetlen a fehérjét gátló daganatellenes terápiák tervezésekor [Lim 2013].			
Gén név	**UniProtAC**	**ComPPI / Translocatome**	**Lokalizáció**
A proliferációt és a metasztázist valószínűleg lokalizáció-specifikusan elősegítő fehérjék 2.			
hTERT	O14746	ComPPI	sejtplazma, sejtmag
A transzlokáció jelentőségére példa a telomeráz reverz transzkriptáz fehérje (hTERT), mely normál testi sejtekben inaktív, azonban daganatos osztódó sejtekben fokozott az aktivitása. Lokalizációját tekintve a sejtplazma és a sejtmag között mozog, mely egyensúly a sejtciklus fázisától, a fehérje foszforilációs státuszától, illetve a DNS károsodás mértékétől is függ. A hTERT sejtmagi lokalizációja szükséges funkciójának betöltéséhez, a telomer szekvencia meghosszabbításához. A sejtmagba transzlokálódását az AKT általi foszforiláció aktiválja, azonban emellett az NF-kB fehérjére is szükség van, mely modulálja a transzlokációs folyamatot, és gátlása potenciális daganatellenes terápia lehet [Akiyama és mtsai 2003]. *(folytatás a következő oldalon ->)*			

8. táblázat: Példák a daganatos malignitás kialakulásával összefüggő, lokalizáció-specifikus funkcióval rendelkező fehérjékre (folytatás)

Gén név	UniProtAC	ComPPI / Translocatome	Lokalizáció
A proliferációt és a metasztázist valószínűleg lokalizáció-specifikusan elősegítő fehérjék 2.			
hTERT	O14746	ComPPI	sejtplazma, sejtmag
(-> folytatás az előző oldalról) A hTERT szerepét számos kontextusban leírták a daganatos iniciációban és progresszióban egyaránt [Jafri és mtsai 2016], melyek közül kiemelkedik a hTERT metasztázis és daganatos őssejt-szerű viselkedés promótáló hatása [Liu és mtsai 2013]. Ez a kettősség felveti a lehetőségét, hogy a hTERT bizonyos őssejt-szerű, invazív vagy éppen proliferatív viselkedést mediáló funkciója lokalizáció alapon is szabályozott.			
Gén név	**UniProtAC**	**ComPPI / Translocatome**	**Lokalizáció**
A proliferációt és a metasztázist valószínűleg lokalizáció-specifikusan elősegítő fehérjék 3.			
NANOG	Q9H9S0	ComPPI	sejtplazma, sejtmag
A NANOG egy embrionális őssejtekben ismert transzkripciós faktor, melynek fontos szerepét a daganatos sejtek differenciációjában is leírták. A NANOG sejtmagi/perinukleáris, illetve mérsékelt sejtplazmai lokalizációval rendelkezik, mely utóbbi szerepe nem tisztázott. A NANOG szerepét többek között loss-of-function kísérletekben vizsgálták, ahol kimutatták szerepét a tumorok fejlődésében [Jeter és mtsai 2009]. Hatásuk van a proliferáció ütemére, a sejtes differenciációra, illetve újabb vizsgálatok metasztázis elősegítő hatását is leírták [Watanabe és mtsai 2014]. Összefüggés a lokalizációs és funkcionális diverzitás között nem tisztázott, azonban egy friss kutatás összefüggést talált daganatos altípusok metasztázis képessége és a NANOG sejtplazmai lokalizációja között [Tamaki és mtsai 2017], mely megerősíti ezen vizsgálatok fontosságát.			

8. táblázat: Példák a daganatos malignitás kialakulásával összefüggő, lokalizáció-specifikus funkcióval rendelkező fehérjékre (folytatás)

Gén név	UniProtAC	ComPPI / Translocatome	Lokalizáció
A proliferációt és a metasztázist valószínűleg lokalizáció-specifikusan elősegítő fehérjék 4.			
P53	P04637	ComPPI és Translocatome	sejtplazma, sejtmag, mitokondrium
A P53 az egyik legközpontibb tumor szupresszorként ismert daganatos fehérje, mely számos patológiás funkcióval rendelkezik. Ezek közül kiemelkedik loss-of-function mutáció esetén funkciójának kiesése, melynek eredménye többek között az apoptózis indukálásnak elkerülése. A P53 transzkripcionális aktivitását a sejtmagban fejti ki, míg sejtplazmában az apoptózis és autofágia indukciójáért felelős [O'Brate és Giannakakou 2003]. A loss-of-function hatáshoz hasonló eredmény érhető el akkor is, ha a P53 szubcelluláris lokalizációs dinamikáját változtatjuk. Ennek vizsgálatára az irodalomban több kísérletes eredményt is találhatunk [Tian és mtsai 2010]. A P53 gain-of-function mutációi ezzel szemben a metasztatikus viselkedést tudják elősegíteni, többek között a sejtmagi transzkripcionális aktivitás mediálásán keresztül [Kastenhuber és Lowe 2017]. Ezek az eredmények bemutatják a P53 lokalizáció specifikus funkcióit, illetve annak dinamikájának fontosságát fiziológiás és patológiás esetben.			
Gén név	**UniProtAC**	**ComPPI / Translocatome**	**Lokalizáció**
A proliferációt és a metasztázist valószínűleg lokalizáció-specifikusan elősegítő fehérjék 5.			
ZEB1	P37275	ComPPI	sejtplazma, sejtmag
A ZEB1 transzkripciós faktor, mely szolid daganatokban kulcsfontosságú eleme és egyben markere az EMT aktivációjának, míg például köpenysejtes limfómában a WNT jelpályán keresztül a daganat növekedését segíti elő [Sánchez-Tilló és mtsai 2014]. Ez az útvonal kiemelten fontos szerepet lát el az epitheliális differenciációban. Ez a kettős fenotipikus viselkedés összefüggésben lehet a ZEB1 eltérő szubcellluláris lokalizációjával, mely többek között fontos a YAP fehérje szabályozásában, mely a Hippo jelátviteli út tagja. A ZEB1 és YAP fehérjék több jelátviteli utat összekötve, mint például a WNT-t és a Hippo-t, daganat promótáló hatással rendelkeznek, növekedett metasztázis eséllyel és ebből eredően csökkent prognózissal [Lehmann és mtsai 2016].			

DOI:10.14753/SE.2018.2134

A táblázat a ComPPI és Translocatome adatbázis adatainak kézi gyűjtése, ellenőrzése, és feldolgozása során gyűjtött fehérjéket tartalmazza, melyek a daganatok malignitásának meghatározásában lokalizáció-specifikus funkcióval rendelkeznek. Az adott fehérje gén neve, UniProt azonosítója és szubcelluláris lokalizációja mellett feltüntetésre került, hogy az adott fehérje szerepel-e a ComPPI vagy Translocatome adatbázisokban.

7. Megbeszélés

A biológiai folyamatok térbeli és időbeli elkülönítése az egyik alapvető szabályozási lehetőség az eukarióta sejtekben, sejtes szerveződésekben. A sejtorganellumok fizikai jelenléte, azok differenciálása és feladataik megértése a biokémia egyik nagy erőkkel kutatott és jól ismert területe. Ezzel együtt feltételezhetően számos speciális sejten belüli funkcionális egység vár még felfedezésre, melyek a sejten belüli szerveződés dinamikus és plasztikus tulajdonságainak vizsgálatán alapulhatnak. A jelátvitel sejten belüli szintjének jobb megértése elvezethet egyes betegségek patofiziológiájának még jobb megértéséhez, adott esetben hiányzó molekuláris biológiai láncszemek felfedezéséhez [Gough 2016].

A sejtes komplexitás megértéséhez a fizikai szerveződésen túl hozzájárul, ha megértjük a molekulák kapcsolatainak rendszerét, melyben a rendszerbiológia eszköztára lehet segítségünkre. A fehérjék kölcsönhatási hálózata, vagy interaktómja számos olyan szabályozási vagy funkcionális tulajdonságba enged betekintést, melyet az egyes fehérjék vagy jelátviteli útvonalak vizsgálata nem tud feltárni. A magas megbízhatóságú bináris fehérje-fehérje kölcsönhatási hálózatok felhasználásával lehetőségünk van egy referencia interaktóm felépítésére, melynek eltéréseit a referencia genomhoz hasonló módon lehet majd vizsgálni a fiziológiás sejtek perturbációja során, vagy éppen patológiás esetben [Luck és mtsai 2017].

A fizikai sejten belüli szerveződés feltárása, illetve a sejtek molekuláris hálózatainak még teljesebb ismerete azonban csak kéz a kézben vezethet még teljesebb megértéshez. Az adatrétegek összekapcsolásával megszületett a transzomikai megközelítés [Yugi és mtsai 2016], mely új hipotézisek felállítására ad lehetőséget az egyes adatrétegek összefüggéseinek feltárásával.

Az adatok magas minőségű integrálása és felhasználása minden molekuláris szinten aktív tendencia, ahol kiemelkedik a jelátviteli adatok összesítése. Az egyes útvonalak elnevezésükben, de a hozzájuk sorolt fehérjékben is különbözhetnek az egyes adatbázisok között, így azok rendszerezett összesítése alapul szolgál a még jobb jelátviteli megértésnek nemcsak az útvonalak, hanem a teljes jelátviteli hálózat szintjén is [Rahmati és mtsai 2017]. Ezen jelátviteli útvonalak annotálhatók különböző molekuláris adatokkal, vagy éppen az adatokból kiindulva rekonstruálhatók az adott sejt legfontosabb jelátviteli mintázatai [Rudolph és mtsai 2016].

A dolgozatban bemutatott ComPPI és Translocatome adatbázisok célja az adatok integrálása, annotálása magas minőségű kézi adatgyűjtéssel, és ezen keresztül az egyes eltérő biológiai adatok összekötésével új biológiai hipotézisek feltárása, melyek többek között a daganatos patobiokémia jobb megértését teszik lehetővé. A ComPPI az első átfogó fehérje-fehérje kölcsönhatási adatbázis, mely a kapcsolatokhoz hozzárendeli a partner fehérjék szubcelluláris lokalizációját, ezzel lehetőséget adva a biológiailag nem valószínű kapcsolatok szűrésére, illetve új lokalizáció specifikus biológiai funkciók predikciójára.

A ComPPI interaktóm lefedettsége a teljes proteóm szintjén kiemelkedően magas más adatbázisokkal összehasonlítva [Veres és mtsai 2015]. A forrás adatbázisok alacsony átfedését az adatkészletek összevonásával oldottuk fel, így a 9 fehérje-fehérje interakciós adatbázis és 8 szubcelluláris lokalizációs adatkészlet megfelelően nagy mennyiségű adatot szolgáltat. Az adatbázis összeállítása során több manuális kurációs lépés segítette a minőség biztosítását, egyúttal a még teljesebb adatintegrációt. Ebben az integrációs feladatban kiemelkedő jelentősége van a különböző forrásokból származó eltérő felbontású szubcelluláris adatok összesítésének a kézzel felállított lokalizációs fa segítségével, mely egyértelmű megfeleltetést biztosít a magas felbontású lokalizációs adatok számára az egyes sejtszervecskék szintjén.

A lokalizációs és az abból származó interakciós megbízhatósági érték segítségével lehetőség van magas és alacsony megbízhatóságú interaktómok létrehozására a szubcelluláris lokalizáció egyezése alapján. Ennek használatával olyan kölcsönhatások törölhetők, melyek partnerei nem osztoznak közös lokalizációban, mivel a térbeli elválasztás miatt biológiailag nem valószínű, hogy az adott kölcsönhatás létrejön az élő sejtben.

A ComPPI webes felületén keresztül egyes fehérjékre tudunk keresni azok különböző elnevezéseit használva, a bővített keresés segítségével faj- és lokalizáció specifikusan. Az egyes fehérjék találati listája tartalmazza a kapcsolatokat, azok szubcelluláris lokalizációját és a kölcsönhatás forrását. A találati oldal is szűrhető, így kompartmentre, lokalizációs vagy interakciós megbízhatósági értékre specifikus kölcsönhatási adatokat böngészhetünk vagy tölthetünk le. A letöltő felületen keresztül interaktóm szintű adatkészletekhez jutunk. A kompartmentalizált adatkészlet fajra szűrhetően tartalmazza azon kapcsolatokat, melyek legalább egy közös lokalizációval rendelkeznek. Lokalizáció

specifikus letöltés esetén az adott kompartment kapcsolati hálóját elemezhetjük. Az integrált fehérje-fehérje interakciós adatkészlet fajra specifikusan magában foglalja az összes elérhető fizikai kölcsönhatást a fehérjék között, lokalizáció specifikus szűrés nélkül. Az összesített lokalizációs adatkészlet az egyik legnagyobb elérhető fehérje szubcelluláris lokalizáció adatforrás, mely a négy fajra kompartment specifikusan letölthető fehérje listát szolgáltat.

A ComPPI webes felülete a http://ComPPI.LinkGroup.hu/oldalon érhető el, mely nyílt forráskódú, így további fejlesztése nyitott a tudományos közösség számára, melynek köszönhetően potenciálisan ComPPI alapú adatbázisok létrehozása is lehetséges. Az adatkészlet limitációi közé tartozik a kísérletes lokalizációk viszonylag alacsony aránya (29% a teljes adatkészletben), melynek feloldására több lokalizációs kísérletes adat integrálását tervezzük.

A ComPPI adatait a megjelenése óta eltelt rövid időben is már számos kutatócsoport használta már fel más adatbázisok bemeneteként, rendszerszintű elemzésekre, vagy az egyes fehérjék vizsgálatára. A SignaLink[54] [Fazekas és mtsai 2013] és OmniPath[55] [Türei és mtsai 2016] adatbázisok a ComPPI szubcelluláris lokalizációs adatait használják, hogy annotálják az egyes jelátviteli fehérjéket. A kiterjedt ComPPI adatkészlet nemcsak más adatbázisok kiegészítő eleme lehet, hanem önmagában is hasznos forrás az interaktóm szubcelluláris lokalizáció specifikus elemzésére [Gibson és mtsai 2015]. Ota és munkatársai a ComPPI adatkészletén azt vizsgálták, hogy az interaktómban csomópont szerepet betöltő fehérjék milyen lokalizációval rendelkeznek, kiemelt figyelemmel a többszörös lokalizációval rendelkező fehérjékre, melynek segítségével lokalizáció specifikus funkciókat állapítottak meg több fehérje esetében [Ota és mtsai 2016].

Saját vizsgálataink a teljes adatkészleten a krotonáz jelentős lokalizáció dependens szerepét mutatták. A zsírsavak béta-oxidációjában fontos fehérje ismert lokalizációja a mitokondrium, azonban 71 kölcsönható partnere közül mindössze 8 rendelkezik mitokondriális kapcsolattal, melyek közül 5 magas megbízhatóságú (interakció megbízhatósági érték 0,8 felett), és csak egynek van kísérletes mitokondriális lokalizációja. Ezzel szemben a 71 interaktor közül 52 rendelkezik sejtplazmai lokalizációval, miközben a 8 mitokondriális partner közül 7 esetében is megtalálható ez

[54] http://www.signalink.org/
[55] http://www.omnipathdb.org/

a lokalizáció. A megfigyelés felveti a krotonáz sejtplazmai lokalizációjának lehetőségét, és ott betöltött biológiai funkcióját. A sejtplazmai kapcsolatok biológiai funkcióinak elemzése rámutatott, hogy a krotonáz potenciálisan antiapoptotikus szereppel bír. Az eredmények validálására irodalomkutatást végeztünk, mely mind a sejtplazmai lokalizációt [Zhang és mtsai 2013], mind az antiapoptotikus szerepet [Yeh és mtsai 2006, Liu és mtsai 2010, Zhu és mtsai 2013] megerősítette. Ezek alapján a krotonáz példája egyszerre mutatja be a ComPPI szerepét a biológiailag nem valószínű kapcsolatok szűrésében, illetve korábban nem ismert, nem konvencionális lokalizáció specifikus biológiai funkciók predikciójában.

A rendszerszintű megközelítés mellett az egyes fehérjék lokalizáció specifikus interaktómjának elemzése is segítheti új biológiai hipotézisek validálását vagy generálást, mely hipotézisek egyaránt érkezhetnek kísérletes vagy bioinformatikai elemzésekből. Az MPS1 kináz esetében kutatásaink során felmerült, hogy kompartment alapú biológiai funkcionalitása lehet. A kölcsönhatásokat megvizsgálva kiderült, hogy a dominánsan sejtplazmai MPS1 sejtmagi lokalizációjában kináz funkciójától független, eddig nem tisztázott szereppel rendelkezik [Zhang és mtsai 2011]. A kölcsönható partnerek biológiai funkciónak vizsgálata felvetette annak lehetőségét, hogy az MPS1 részt vesz a sejtes organellumok szerveződésének kialakulásában, kiemelten pedig a sejtmagmembrán organizációjában, melyet a sejtmagi pórus komplexhez kötődését leíró evidencia is megerősít [Liu és mtsai 2003].

Az interaktómok vizsgálata proteóm szinten, a szubcelluláris lokalizációt figyelembe véve nemcsak bioinformatikai módszerekkel lehetséges, hanem kísérletes úton is [Mardakheh és mtsai 2017]. A fehérjék ko-lokalizációjának kísérletes vizsgálata még több evidenciát adhat arra vonatkozóan, hogy az egyes fehérje-fehérje interakciók partnerei valójában tartózkodnak-e közös szubcelluláris térben. A fiziológia mellett fontos a patológia vizsgálata is, így például a daganatos sejtek egyedi fehérje kölcsönhatási hálózata. Az OncoPPi [Li és mtsai 2017] egy daganatokra specifikus fehérje kölcsönhatási adatkészlet, mely kapcsolatok nem találhatóak meg más, fiziológiás adatokat tartalmazó adatbázisokban. Ezen daganatos kölcsönhatások elemzése új biológiai szabályozó funkciók feltárására ad lehetőséget, melyek később beépíthetők átfogóbb modellezési rendszerekbe. Ilyen modell például az EMT folyamatot leíró dinamikus jelterjedést vizsgáló hálózat [Steinway és mtsai 2015], mely az EMT

folyamatának kulcsfontosságú szabályozó molekuláit határozza meg a kölcsönhatási hálózat működésének szimulációjával.

Számos elemzés és azokra épülő adatkészlet foglalkozik a fehérjék kompartmenteken belüli ko-lokalizációjával, azonban ez a fajta vizsgálat a statikus állapotot tükrözi. A fehérjék az egyes kompartmentek között helyet tudnak változtatni, melyet transzlokációnak nevezünk. Ennek elemzése azonban dinamikus, az egyes lokalizációkon belüli fehérje koncentráció változásokat figyelembe véve lehet a teljes képet megalkotni a funkcionális aktivitásról. A ComPPI adataira épülve célunk, hogy megalkossuk az első rendszerszintű fehérje transzlokációt is modellező jelátviteli dinamikus modellt, melyhez azonban információra van szükségünk az egyes transzlokálódó fehérjékről.

A Translocatome [Dobronyi és mtsai 2016, Mendik és mtsai 2017] az első adatbázis, mely kézi adatgyűjtésen és gépi tanuló algoritmuson alapuló predikció segítségével listázza azon fehérjéket, melyekről ismert vagy feltételezhető, hogy transzlokálódnak a sejten belül. Az adatkészlet jelenleg több mint 250 fehérjére tartalmaz részletes adatokat a transzlokáció irányáról, ha elérhető annak mechanizmusáról, kapcsolódó fehérjeszerkezeti aspektusokról. Emellett gyűjti, hogy az egyes fehérjék lokalizáció specifikusan milyen funkcióval rendelkeznek, és milyen speciális interakciókat alakítanak ki egyik, vagy másik lokalizációban. Sok transzlokálódó fehérje szerepet játszik patológiás állapotokban is, mely elérhető adat esetén szintén szerepel az adatbázisban.

A kézzel gyűjtött adatkészlet nemcsak biztosan transzlokálódó fehérjéket tartalmaz, hanem biztosan nem transzlokálódókat is, melyek lehetnek csak egy megadott lokalizációban, vagy bárhol a sejten belül. A két adatkészlet használható tanító adatként egy gépi tanuló algoritmus számára, mely a ComPPI-ból átemelt fehérjékre egy transzlokációs valószínűséget prediktál hálózatos és funkcionális paraméterek alapján. Ennek segítségével a kézzel ellenőrzött adatkészlet is folyamatosan bővül, hiszen a transzlokálódónak prediktált fehérjéket az irodalomban található evidenciák alapján be tudjuk emelni a valóban transzlokálódó fehérjék halmazába. A gépi tanuló algoritmus nyílt forráskódú, és a paraméterek is változtathatók, így lehetőség van nemcsak a ComPPI interaktómon, hanem más hálózatokon is tesztelni a rendszert, ami további lehetőségeket nyit a predikciók visszaellenőrzésére, valamint új fehérjék gyűjtésére.

A Translocatome fejlesztés alatt álló webes felülete (http://translocatome.linkgroup.hu/) két fő elemből áll, melyek a kézi adatgyűjtésre szolgáló adatbevitelt támogató webes eszköz, illetve a később telepítésre kerülő, felhasználókat kiszolgáló adatmegjelenítő oldal. Az előbbi eszköz alkalmas arra, hogy több felhasználó egyszerre hozzáférjen az adatkészlethez, ellenőrizze a bevitt adatokat, vagy új fehérjékre vonatkozó információkat vigyen be a rendszerbe. Az adatbevitel meghívásos alapon történik, azaz megfelelő kompetencia esetén ellenőrzés mellett lehetőség van az adatkészlet bővítésére, melyet terveink szerint ellenőrzés után fél évente fogunk frissíteni az éles adatbázisban. Jelenleg négy kurátor dolgozik azon, hogy az adatkészlet terjedelme és minősége egyre nagyobb legyen, ezzel biztosítva a még jobb minőségű predikciók létrehozását.

A felhasználókat kiszolgáló webes eszköz fő célja a transzlokálódó fehérjék listázása és az egyes fehérjékhez tartozó információk rendszerezett megjelenítése. A böngészési felületen listázódnak az aktuális verzióban kézzel gyűjtött transzlokálódó fehérjék, illetve a legmagasabb predikciós értékkel rendelkezők. A letöltési felületen lehetőség van a tejes adatkészlet letöltésére, mely egyszerűsíti a bioinformatikai elemzéshez való felhasználást.

Számos transzlokálódó fehérje ismert, melyek szerepe patológiás állapotokban kiemelkedő. A legtöbb ismert patológiás transzlokáció daganatos betegségekkel függ össze, ahol a transzkripciós faktorok sejtplazma – sejtmag lokalizációs egyensúlya felborul. A transzkripciós faktorok a sejtplazmában transzkripcionális aktivitásuktól független feladatot töltenek be (mint láttuk az ERK2 esetében), vagy nyugalmi helyzetben vannak mielőtt a sejtmagba kerülve kifejtik hatásukat a génkifejeződésre. Ez a transzlokáció egy kényes szabályozási lépés, hiszen a fokozott áthelyeződés emelkedett transzkripcionális aktivitást eredményez, mely végső soron daganat iniciációhoz vagy progresszióhoz vezethet. Ugyanígy a másik irányban a sejtmagi transzlokáció gátlása terápiás célpont lehet egyes daganatos betegségekben [Hill és mtsai 2014].

A transzlokáció jelentőségére példa a telomeráz reverz transzkriptáz fehérje (hTERT), mely normál testi sejtekben inaktív, azonban daganatos osztódó sejtekben fokozott az aktivitása. Lokalizációját tekintve a sejtplazma és a sejtmag között mozog, mely egyensúly a sejtciklus fázisától, a fehérje foszforilációs státuszától, illetve a DNS károsodás mértékétől is függ. A hTERT sejtmagi lokalizációja szükséges funkciójának betöltéséhez, a telomer szekvencia meghosszabbításához. A sejtmagba transzlokálódását

az AKT általi foszforiláció aktiválja[56], azonban emellett az NF-kB fehérjére is szükség van, mely modulálja a transzlokációs folyamatot, és gátlása potenciális daganatellenes terápia lehet [Akiyama és mtsai 2003].

Egy másik példa a fokális adhéziós kináz (FAK), mely egy tirozin kináz fontos sejtmigráció és proliferáció szabályozó szereppel. A fehérje hagyományos funkciója sejtplazmai, részt vesz az integrin és növekedési faktor receptoriális jelátvitelben. Sejtmagi szerepének ismerete az elmúlt években lett egyre kiemelkedőbb, ahol fokozza a P53 és GATA4 degradációját ubiquitináción keresztül, mely fokozott sejtosztódáshoz és csökkent gyulladásos válaszhoz vezet. A FAK szintén működhet ko-transzkripcionális szabályozóként, így szerepe lehet a transzkripcionális szabályozásban is. A FAK jelátvitel a fokális adhéziótól a sejtmagig tehát fontos szabályozó tengelye a sejteknek, melynek befolyásolása terápiás célpont is lehet. Megfigyelték, hogy a FAK kináz funkciójának gátlása a sejtmagi funkciót nem érinti, sőt fokozott sejtmagi felhalmozódást eredményez. Mindez azt mutatja, hogy a FAK lokalizáció specifikus gátlása elengedhetetlen a fehérjét gátló daganatellenes terápiák tervezésekor [Lim 2013].

Ez a két példa is alátámasztja, hogy az egyes kulcsfontosságú transzlokálódó fehérjék komplex szabályozás alatt állnak, melynek rendszerszintű vizsgálata elengedhetetlen a biológiai folyamatok jobb megértésének érdekében. Léteznek modellek egyes fehérje transzlokációk hatásának jelátviteli elemzésére, mint például a P53 esetében [Eliaš és mtsai 2014], azonban a transzlokációs folyamatok átfogó modellezése még várat magára. A munka további tervezett része a Translocatome és a ComPPI adatainak felhasználása egy dinamikus jelátviteli hálózati modell építésére, melynek segítségével a transzlokáció fiziológiája és patológiás hatásai új megközelítésből vizsgálhatók.

A daganatok iniciációja és progressziója számos ponton szabályozott, bonyolult jelátvitel eredménye, mely az egyes sejtek szintjén túl a sejtek között is mediálja a daganatos fenotípust jellemző tulajdonságok kialakulását. A korábbiakban bemutatott sejten belüli elhelyezkedésen alapuló szabályozás, például az ERK2 esetében [Zehorai és mtsai 2010] alátámasztja a különböző molekuláris szabályozási körök által befolyásolt transzlokáció fontos szerepét a daganatok invazivitásának meghatározásában, mely lehet dominánsan proliferáció vagy invazivitás fokozó hatású [Tulchinsky és mtsai 2014]. Ez a két viselkedés alapvetően elkülöníthető egy gyorsan osztódó fenotípusra rapid tumor

[56] http://www.uniprot.org/uniprot/O14746/

növekedéssel, de sokszor limitált invazivitással, valamint egy alapvetően metasztatikus hajlammal járó, lassabb osztódási ütemet mutató malignusabb tumor típusra.

Ezt a kettős viselkedést jellemezhetjük hálózatbiológiai szempontból, melyre kísérletet tettünk kutatócsoportunk számos közleményében, összefoglalva az elérhető irodalmi evidenciákat [Gyurkó és mtsai 2013, Csermely és Korcsmáros 2013, Csermely és mtsai 2013, Csermely és mtsai 2015]. A rosszindulatú daganatok kialakulásának és progressziójának rendszerszintű kétlépcsős modellje szerint elkülöníthető egy kezdeti flexibilis, plasztikus állapot, illetve egy későbbi rigidebb, stabilabb molekuláris hálózat. Az első stádium felfogható egyfajta útkeresésként, ahol a transzformálódó sejtek igyekeznek megtalálni a számukra optimális szabadságot nyújtó molekuláris hálózatos állapotot, mely speciálisan megfelel az adott genotípusnak és mikrokörnyezetnek. Ez a flexibilis állapot kiemelkedő adaptációs képességet jelent, hiszen a változó környezetre is képes gyorsan reagálni, például génexpressziós szabályozás változtatásával. A plaszticitás azonban egyúttal sérülékenységet is jelent, szélsőséges környezeti változások esetén nem képes reagálni a behatásokra, éppen környezeti függősége miatt. A rigidebb kései stádium elérése azt segíti, hogy a daganatos sejt függetlenedjen környezetétől, és képes legyen önálló viselkedésre.

A flexibilisebb állapot tehát egy gyorsabb osztódási ütemmel jellemzett, alapvetően adaptív, de viszonylag sérülékeny állapot, ahol a daganat a mikrokörnyezet és saját molekuláris tulajdonságinak megfelelően keresi az önállósodásra optimális fenotípust. A rigid állapot kevésbé adaptív, azonban az adott mikrokörnyezethez alkalmazkodva sokkal invazívabb fenotípust vehet fel, ezzel elősegítve a lokális és távoli metasztázisok képződését. Ez a két állapot azonban nem lineárisan követi egymást, hanem egy dinamikus egyensúly eredménye mind az egyes sejtek, mind a daganatos sejtek populációjának szintjén, mely állapotok között számos külső tényező hatására válthat a daganatos molekuláris hálózat [Csermely és mtsai 2015]. A két fenotípus közötti átmenet képezi a rendszer legsérülékenyebb állapotát, melynek biomarkerekkel történő felismerése elősegítheti a daganat ellenes terápiák még hatékonyabb célzását [Chen és mtsai 2012].

A kétlépcsős hipotézist számos kísérletes és klinikai evidencia támogatja [Gyurkó és mtsai 2013, Csermely és Korcsmáros 2013, Csermely és mtsai 2013, Csermely és mtsai 2015], azonban a stádium-specifikus adatok mennyisége limitált, mely segítené a

molekuláris hálózat karakterizálását valós betegadatokkal. Rendszerszintű molekuláris adat általában a két végállapotban, az egészséges szövetre és a már kialakult kései stádiumú tumorra érhető el. Az adatok hiánya nehezíti a molekuláris hálózat szerkezeti és dinamikus változásának időbeli lekövetését, és így a malignitást és metasztázist elősegítő molekuláris eltérések meghatározását. A daganatos őssejtek, illetve dormant állapotban lévő sejtek elkülönítése, és ezek elemzése lehetőséget adhat a hiányzó plaszticitással és evolvabilitással összefüggő rendszerszintű hálózatos megfigyelésekre.

A klinikai evidencia és a daganatos molekuláris hálózat álalakulásának kétlépcsős hipotézise együttesen megalapozza a megfigyelést, miszerint a gyorsan növekvő daganatok általánosságban kevésbé hajlamosak metasztázis kialakítására, és megnövekedett szenzitivitásuk miatt jobb terápiás választ mutatnak a klinikumban. Ezzel szemben a lassabb osztódási ütemet mutató daganatok gyakrabban invazívak és adnak távoli metasztázisokat, mely csökkenti terápiás válaszkészségüket [Adami és mtsai 2017]. Ez a kettős viselkedés széleskörben megfigyelhető, azonban kiemelt jelentősége van az egyes daganatos sejtpopulációk egyedi elemzésének, hiszen a különböző fenotípussal rendelkező sejtek egyensúlyának eltérései miatt az agresszíven növekedő és osztódó tumorok is adhatnak metasztázist, és a lassan növekedő tumorok is lehetnek non-invazívak, mint ahogy azt számos aszimptomatikus indolens tumor diagnózisa is mutatja [Adami és mtsai 2017].

A daganatok közti heterogenitás [Burrell és mtsai 2013] miatt kiemelkedő fontosságú a betegekre specifikus omikai adatokon alapuló prediktív terápia tervezés, melynek alapja lehet a mind proliferatív, mind metasztatikus viselkedésre specifikus biomarkereket tartalmazó diagnosztikai eszköztár. A daganatos sejtek plasztikus és rigid állapotát egyaránt gátló kezelés bírhat a legnagyobb hatékonysággal a terápiás válaszkészség maximalizálásában, illetve a későbbi recidíva elkerülésében, a rezisztencia kialakulásának minimalizálásban. A molekuláris hálózatok célzása lehetséges központi molekulák targetálásával (úgynevezett *central-hit* terápia), vagy a hálózatos szomszédok mediálásán keresztül (úgynevezett *network influence* stratégia) [Csermely és mtsai 2013]. A precíziós onkológia fő kihívása [Prasad és mtsai 2016] a különböző daganatos sejtállapotok, és ezen keresztül a diverz daganatos sejtpopuláció minél hatékonyabb célzása jól megtervezett személyreszabott terápiával.

DOI:10.14753/SE.2018.2134

A klinikumban megfigyelhető, hogy a magas osztódási rátával rendelkező daganatok hamarabb mutatnak klinikai tüneteket, ami korai fázisban segíti a diagnosztikát, melyet általában agresszív terápia követ. Ezzel szemben a lassan növekvő tumorok csak késői stádiumban kerülnek felfedezésre, gyakran csak másodlagos tumorok jelenléte okoz klinikai tüneteket, mely stádiumban a kezelések hatékonysága sokkal csekélyebb. A molekuláris hálózat tulajdonságai alapján logikus lenne, hogy a gyorsan osztódó, korai stádiumban felismert tumorok ellen kevésbé drasztikus terápiákat alkalmazunk, míg a késői stádiumban felismert tumorok korai diagnosztikáját sürgessük. Nem ismert azonban a gyorsan növekvő daganatok válaszkészsége enyhébb kezelés hatására, különösen az invazív fenotípusra váltás valószínűsége növekedhet. A lassan növekvő tumorok esetében pedig a korai kezelés potenciális hatásaival kapcsolatban limitált a klinikai evidencia. Mindezek alapján egyértelmű, hogy a minél pontosabb molekuláris diagnosztika elengedhetetlen a valós személyreszabott terápia korának eléréséhez, azonban további kísérletes és klinikai evidencia szükséges ezek megalapozásához.

DOI:10.14753/SE.2018.2134

8. Következtetések

Doktori munkám során a fehérjék térbeli elhelyezkedésének szerepét vizsgáltam a fehérje-fehérje kölcsönhatási hálózatban, mely rendezettség elengedhetetlen feltétele a megfelelő egészséges sejtes jelátvitel fenntartásának. Daganatos sejtekben ez a rendszer gyakran felborul, melyre dolgozatomban számos példát mutatok be. A fehérje-fehérje kölcsönhatások sejtkompartment szintű elemzése lehetőséget ad a fontos jelátviteli folyamatok térbeli vizsgálatára, így például a fehérje transzlokáció rendszerszintű predikciójára és hatásának elemzésére.

A dolgozatban bemutatott legfontosabb új eredmények:

1. Megalkottuk az első kompartmentalizált fehérje-fehérje kölcsönhatási adatbázist, a ComPPI-t [Veres és mtsai 2015, http://comppi.linkgroup.hu/], mely az egyik legnagyobb felhasználó-barát interakciós és lokalizációs adatforrás.

2. Megalkottuk a Translocatome adatbázist (http://translocatome.linkgroup.hu/), amely az emberi sejtekben a transzlokáció útján a jelátviteli szabályozásban szerepet játszó fehérjék első szisztematikusan gyűjtött és annotált adatkészlete [Dobronyi és mtsai 2016, Mendik és mtsai 2017]. Példákat gyűjtöttünk a kompartmentalizáció és a transzlokáció szerepére a malignus elváltozásokban.

3. Hálózatos munkáink tapasztalatai alapján megalkottuk a malignus transzformáció kétlépcsős hipotézisét. Az egészséges sejtből először egy plasztikus, a környezeti hatásokra gyorsan reagálni képes pre-malignus hálózat jön létre. Ezt követően a molekuláris hálózat rigidebbé válik, és ezzel stabilizálja az új, sok esetben metasztatikus fenotípust [Gyurkó és mtsai 2013, Csermely és mtsai 2015]. A transzlokációval kapcsolatos molekuláris adatok és hipotézisünk segítette az amerikai, norvég és svéd kutatók vastagbél tumorral kapcsolatos, új epidemiológiai adatainak értelmezését, amelyet egy kollaborációs közleményben foglaltuk össze [Adami és mtsai 2017]. Munkánk szerint a megfigyelésnek klinikai jelentősége lehet a daganat megelőző szűrések, illetve a terápia tervezéskor, valamint az utánkövetés során.

A legfontosabb új eredmények összefoglalását az alábbi fejezetek tartalmazzák.

DOI:10.14753/SE.2018.2134

8.1. A ComPPI adatbázis és webes felület fehérjék kompartment specifikus funkcionális elemzésére

Megalkottuk az első kompartmentalizált fehérje-fehérje kölcsönhatási adatbázist, a ComPPI-t [Veres és mtsai 2015], mely 9 fehérje-fehérje interakciós és 8 szubcelluláris lokalizációs adatforrást összegezve az egyik legnagyobb elérhető interakciós és lokalizációs adatforrás. A kialakítás során számos kézi ellenőrző és adatfeldolgozó lépés segítette az adatok integrációját, és az elkészült adatkészlet minőségének biztosítását. A felhasználóbarát webes felület segítségével lehetőség nyílik egyes fehérjék szubcelluláris lokalizáció specifikus interaktómának elemzésére, valamint kompartment specifikus interaktómok letöltésére további elemzés céljából.

A ComPPI adatain alapulva rendszerszintű elemzéseket végeztünk, melynek eredményeképpen két fehérje, a krotonáz és az MPS1 esetében azonosítottunk új biológiai funkciókat, melyek egy részére megerősítést találtunk az irodalomban. A krotonáz esetében a szomszédsági hálózat alapján prediktált sejtplazmai lokalizációra találtunk evidenciát, mely konszenzusban volt a már leírt eredményekkel daganatos sejtekben [Zhang és mtsai 2013]. A kölcsönható partnerek elemzése kimutatott apoptózis gátlással összefüggő biológiai funkciókat is, mely megerősíti a krotonáz feltételezett funkcióját rákos megbetegedésekben [Zhu és mtsai 2013].

Az MPS1 kináz interaktóm vizsgálata a sejtmagi lokalizáció fontosságát mutatta, melyre sporadikus irodalmi adatok is elérhetőek [Zhang és mtsai 2011]. Az interaktóm alapján az MPS1 és szomszédjainak biológiai funkcionális elemzése felvetette a lehetőségét, hogy az MPS1 fontos szereppel bír a sejtmag membrán újra szerveződésében osztódás után, melyet számos, eddig nem összegzett evidencia erősít meg, például az MPS1 kapcsolata a sejtmag pórus komplexszel [Liu és mtsai 2003].

Megfigyeléseink bioinformatikai módszereken és az elérhető adatokon alapulnak, így a kísérletes validáció elengedhetetlen a prediktált funkciók bizonyítására. Azonban bizonyítás nélkül is látható, hogy a fehérjék szubcelluláris lokalizációja hogyan tudja befolyásolni azok kapcsolati hálóját, és ezen keresztül biológiai folyamatokban betöltött szerepüket, melynek kiemelt jelentősége van a daganatok kialakulásában és progressziójában.

8.2. A fehérjék transzlokációjának proteóm szintű adatbázisa és vizsgálata

A Translocatome adatbázis az első adatkészlet, mely szisztematikusan gyűjti és annotálja az emberi sejtekben transzlokáció útján jelátviteli szabályozásban szerepet játszó fehérjéket [Dobronyi és mtsai 2016, Mendik és mtsai 2017]. Az adatbázis magja egy kézzel gyűjtött adatkészlet, mely részletesen bemutatja a transzlokálódó fehérjék tulajdonságait, és a kézzel gyűjtött biztosan nem transzlokálódó fehérjékkel kiegészítve alkalmas tanulókészlet további fehérjék transzlokációs valószínűségének predikciójához.

Munkánk során megalkottuk a transzlokálódó fehérjék rendszerbiológiai definícióját, melyhez felhasználtuk a dolgozat elkészítésének időpontjáig összegyűjtött több mint 200 transzlokálódó fehérje feldolgozás közben megismert tulajdonságait. A definíció alapján összegyűjtöttük azokat a fehérjéket jellemző paramétereket, melyeket felhasználhatunk a gépi tanulás során a transzlokáció valószínűségének meghatározására, mely algoritmus a kezdeti elemzések során 80% feletti pontossággal prediktál transzlokálódó fehérjéket.

A Translocatome-mal kapcsolatos munka még folyamatban van. A hátralévő munkafázisok célja az adatbázis teljes megalkotása mellett a felhasználóbarát webes felület véglegesítése, kiegészítve egy bejelentkezés után elérhető felülettel, ahol lehetőség van az adatok áttekintésére, javítására, és új fehérjék vagy hozzájuk tartozó információk bevitelére. Ez a kézi adatgyűjtő felület segíti majd a Translocatome közösségi fejlesztését, így a minél kiterjedtebb és megbízhatóbb adatok rögzítését.

A transzlokálódó fehérjék jelentősége daganatos betegségekben kiemelt fontosságú, számos esetben az egészséges transzlokációs egyensúly felborulása is közrejátszik a daganatok kialakulásában és progressziójában, illetve előfordulnak csak patológiás esetben megfigyelhető áthelyeződések is. Az adatok összegyűjtése után a teljes sejtes jelátvitel kontextusába helyezve rendszerszinten tervezzük modellezni a transzlokáció hatásait, ezzel új terápiás beavatkozási pontokat, vagy érzékenységet jellemző biomarkereket azonosítva.

8.3. A malignus transzformáció kétlépcsős hipotézise és ennek szerepe a daganatos progresszió megítélésében

Az irodalmi evidencia és eddigi kutatási eredményeink összegzése alapján a daganatok kialakulását és progresszióját egy kétlépcsős folyamatként írtuk le, melyre jellemző a molekuláris hálózat viselkedésének változása. A kezdeti nyugalmi állapotból először egy

flexibilis, a környezeti hatásokra plasztikusan reagálni képes hálózat jön létre, mely jellemző a daganatok korai fázisában észlelhető adaptációs folyamatokra. A korai stádiumú daganatok progressziója során a molekuláris hálózat megszilárdul, rigidebbé válik, ezzel beágyazódva és stabilizálva az új daganatos fenotípust [Gyurkó és mtsai 2013, Csermely és Korcsmáros 2013, Csermely és mtsai 2013, Csermely és mstai 2015].

A két állapot azonban nem elengedhetetlenül lineárisan követi egymást, hanem dinamikusan váltakozik, mely segíti a daganatok új környezeti hatásokhoz történő alkalmazkodását, mint amilyenek a daganat ellenes terápiák is. Az alkalmazkodó, plasztikus állapot és a beágyazott, rigidebb állapot nemcsak időben, hanem térben is differenciálódik. A daganatos sejtes populáció egy része őssejt szerű jegyeket hordoz, és adaptációs inger hiányában dormant állapotot vesznek fel [Kleffel és Schatton 2012, Csermely és mtsai 2015].

Megfelelő körülmények esetén a stabilizált, rigidebb molekuláris hálózattal rendelkező daganatos sejtek osztódása jelentős mértékű lehet, így a tumor mérete gyorsan növekszik. Kezelés következtében a külső körülmények negatív irányba mozdulásának hatása visszafoghatja az osztódás ütemét, és a dormant állapotú daganatos őssejtek aktivációjával egy plasztikusabb molekuláris válasz alakul ki, ahol az új környezet keresése, így az áttétképzési hajlam kerül előtérbe.

A két folyamat egyensúlyának szabályozása a teljes jelátviteli hálózat szabályozása alatt ál [Huang és Ingber 1999, Coghlin és Murray 2010], mely nemcsak az egyes sejtek szintjén, hanem a sejtek között is mediálja a kommunikációt. A daganatos sejtek és a mikrokörnyezet kölcsönható sejtjeinek hálózatos szerkezete és dinamikája együtt szabályozza ezt az egyensúlyt, és választja el egymástól a növekedésre és áttétképzésre való hajlamot, ezzel segítve a daganatok minél effektívebb alkalmazkodását és túlélését. Saját munkánkat és újonnan közölt epidemiológiai adatokat is tartalmazó kollaborációs közleményben foglaltuk össze, hogy a megfigyelésnek klinikai jelentősége lehet a daganat megelőző szűrések, illetve a terápia tervezéskor, valamint utánkövetés során [Adami és mtsai 2017].

8.4. A számítógép-vezérelt személyre szabott onkológia, mint a rendszerbiológia új felhasználási módja

A bemutatott adatkészletek, webes alkalmazások és biológiai hipotézisek a számítógépes biológia eszköztárának egy kis szeletét mutatják be, mely eszköztárat az egyre növekvő

DOI:10.14753/SE.2018.2134

kísérletes adattömeg rendszerezésére, újragondolására, és következtetések levonására hasznosíthatjuk. A személyre szabott orvoslás és „big data” korszaka beköszöntött az onkológiába, ahol a felhasználás eredményességének kulcsa az adatok mennyiségén túl azok megbízhatósága, és az adatokon alapuló prediktív modellek megalkotása.

Az adatok integrációja az egyes adatrétegek megfelelő tulajdonságainak párosításával segíti olyan számítógépes modellek megalkotását, ahol a jelátviteli folyamatokat az adott daganatos típusra szabva tudjuk vizsgálni. A rendszerszintű modellek megalkotása után szükséges a hatékony és prediktív algoritmusok felépítése, melyek elemzik és értékelik az adatokat [Fekete és mtsai 2016]. A számítógépes szimulációk párosítva a mesterséges intelligencia előnyeivel számos olyan lehetőséget adnak a kutatók és klinikusok kezébe, mely korábban nem állt rendelkezésre, segítve mind a daganatos megbetegedések megelőzését azok jobb megértésén keresztül, mind a hatékonyabb gyógyszer fejlesztési folyamatot és személyre szabott terápia választást.

9. Összefoglalás

A fehérjék sejten belüli elhelyezkedése alapvetően befolyásolja a különböző biokémiai folyamatok, így a sejtes jelátvitel térbeli és időbeli szabályozását. A fehérjék sejtorganellumok közötti áthelyeződése fontos funkcionális következménnyel bírhat az egyes fehérjék, illetve a teljes fehérje-fehérje hálózat szintjén. Ez a szabályozás szerepet játszhat a daganatos iniciációban és progresszióban, ezen keresztül fontos diagnosztikai vagy terápiás célpont lehet.

Doktori munkám során megalkottam az első kompertmentalizált fehérje-fehérje interakciós adatbázist, mely figyelembe veszi a fehérjék szubcelluláris lokalizációját is, ezzel szűrve a biológiailag nem valószínű kapcsolatokat és lehetőséget teremtve a lokalizáció alapú új biológiai funkciók jóslására. Az adatokat felhasználva, és kézzel gyűjtött fehérjékkel kiegészítve létrehoztuk az első, kifejezetten transzlokálódó fehérjéket gyűjtő és bemutató adatbázist, melyet Translocatome-nak neveztem el. Az összegyűjtött adatok, és az ezeken alapuló elemzések segítségével megalkottunk és példákon keresztül bemutattunk egy, a daganatok proliferatív és metasztatikus viselkedését leíró molekuláris hálózatos modellt.

A munka során létrejött ComPPI (http://comppi.linkgroup.hu/) adatbázis és webes felület egy átfogó fehérje-fehérje kölcsönhatási és szubcelluláris lokalizációs adatbázis, mely 3 modell organizmusra és az emberre tartalmaz magas minőségű adatokat. Az adatbázist már eddig is számos más adatbázis és tanulmány használta fel.

A fejlesztés alatt álló Translocatome (http://translocatome.linkgroup.hu/) adatbázis és webes felület több mint 200 transzlokálódó fehérjére vonatkozóan tartalmaz magas megbízhatóságú, kézzel gyűjtött adatot. Gépi tanuló algoritmus segítségével további fehérjéket sorol be a potenciálisan transzlokálódó vagy nem transzlokálódó csoportba, ezzel alapozva meg egy proteóm szintű átfogó adatkészletet a fehérjék áthelyeződéséről.

A ComPPI és Translocatome fejlesztése során előkerülő példák, mint például a FAK1, hTERT, NANOG, P53 vagy ZEB1 fehérjék fontos szereppel bírnak a daganatok kialakulásában. Mindemellett szubcelluláris funkcióváltással is bemutatják a daganatok eltérő viselkedését a korai, inkább proliferatív, és késői, inkább invazív fenotípus kialakulása során. E kétlépcsős hipotézis alapja lehet a növekedési ütemen és a daganat méreten túlmutató, a molekuláris hálózat egészét elemző precíziós diagnosztikus biomarkerek és terápiás beavatkozások tervezésének.

10. Summary

Subcellular localization of proteins plays a major role in different biochemical processes, especially in the spatial and temporal regulation of cellular signalling. Protein translocation between cellular organelles often has essential consequences both on the function of individual proteins and the whole protein-protein interaction network. Translocation-mediated regulation plays a key role in cancer initiation and progression, leading to potential diagnostic or therapeutic targets.

During my doctoral work my goal was to create the first compartmentalised protein-protein interaction database, which also takes into account the subcellular localisation of the proteins, giving the opportunity to filter out biologically unlikely interactions and to predict subcellular localisation-specific, novel biological functions. Based on this dataset extended with manually curated proteins, my goal was to establish the first database collecting translocating proteins with its connected web interface, the so-called 'Translocatome'. By the analysis based on the collected data my aim was to elaborate the molecular network model of cancer ploriferation and metastatic behaviour through relevant protein examples.

The novel ComPPI database and its web interface (http://comppi.linkgroup.hu/) is a highly comprehensive protein-protein interaction and subcellular localisation database, containing high-quality data for three model organisms and for humans, used by several external databases and research studies.

The Translocatome database and web interface (http://translocatome.linkgroup.hu/) is currently under development, containing manually curated high-confidence data for more than 200 translocating proteins. With the help of a machine-learning algorithm other proteins were also categorized into potentially translocating or non-translocating groups, establishing a proteome-wide comprehensive dataset for protein translocation.

Protein examples identified during the development of ComPPI and Translocatome, such as FAK1, hTERT, NANOG, P53 and ZEB1, have a key role in cancer initiation including their regulation at the subcellular level. Besides, they also show a divergent behaviour in early, more proliferative and late stage, more invasive cancer phenotypes. The emerging two-stage hypothesis of cancer progression can help the development of whole network-based precision diagnostic biomarkers and therapeutic interventions going beyond the growth-rate and tumor size-based classification.

11. Irodalomjegyzék

1. Adami HO, Csermely P, Veres DV, Emilsson L, Løberg M, Bretthauer M, Kalager M. (2017) Are rapidly growing cancers more lethal? *Eur J Cancer,* 72:210-214.
2. Akiyama M, Hideshima T, Hayashi T, Tai Y-T, Mitsiades CS, Mitsiades N, Chauhan D, Richardson P, Munshi NC, Anderson KC. (2003) Nuclear factor-kappaB p65 mediates tumor necrosis factor alpha-induced nuclear translocation of telomerase reverse transcriptase protein. *Cancer Res*, 63(1):18-21.
3. Al-Shibli SM, Amjad NM, Al-Kubaisi MK, Mizan S. (2017) Subcellular localization of leptin and leptin receptor in breast cancer detected in an electron microscopic study. *Biochem Biophys Res Commun*, 482(4):1102-1106.
4. Azar WJ, Zivkovic S, Werther GA, Russo VC. (2014) IGFBP-2 nuclear translocation is mediated by a functional NLS sequence and is essential for its pro-tumorigenic actions in cancer cells. *Oncogene*, 33(5):578-588.
5. Bader GD, Betel D, Hogue CWV. (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res,* 31(1):248-250.
6. Balboula AZ, Schindler K. (2014) Selective disruption of aurora C kinase reveals distinct functions from aurora B kinase during meiosis in mouse oocytes. *PLoS Genet,* 10(2):e1004194.
7. Barabási A-L, Gulbahce N, Loscalzo J. (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet,* 12(1):56-68.
8. Bastian M, Heymann S, Jacomy M. (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third Int AAAI Conf Weblogs Soc Media,* 361-362.
9. Bhattacharyya T, Karnezis AN, Murphys SP, Hoang T, Freeman BC, Phillips B, Morimoto RI. (1995) Cloning and subcellular localization of human mitochondrial HSP70. *J Biol Chem,* 270(4):1705-1710.
10. Binder JX, Pletscher-Frankild S, Tsafou K, Stolte C, O'Donoghue SI, Schneider R, Jensen LJ. (2014) COMPARTMENTS: Unification and visualization of protein subcellular localization evidence. *Database*, 2014:bau012.
11. Brameier M, Krings A, MacCallum RM. (2007) NucPred--predicting nuclear localization of proteins. *Bioinformatics*, 23(9):1159-1160.

12. Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U. (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci*, 10(6):2763-2788.
13. Burrell RA, McGranahan N, Bartek J, Swanton C. (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501(7467):338-345.
14. Caramel J, Papadogeorgakis E, Hill L, Browne GJ, Richard G, Wierinckx A, Saldanha G, Osborne J, Hutchinson P, Tse G, Lachuer J, Puisieux A, Pringle JH, Ansieau S, Tulchinsky E. (2013) A Switch in the Expression of Embryonic EMT-Inducers Drives the Development of Malignant Melanoma. *Cancer Cell,* 24(4):466-480.
15. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics*, 25(2):288-289.
16. Casar B, Pinto A, Crespo P. (2009) ERK dimers and scaffold proteins: Unexpected partners for a forgotten (cytoplasmic) task. *Cell Cycle*, 8(7):1007-1013.
17. Chatr-aryamontri A, Breitkreutz B-J, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, Reguly T, Nixon J, Ramage L, Winter A, Sellam A, Chang C, Hirschman J, Theesfeld C, Rust J, Livstone MS, Dolinski K, Tyers M. (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res*, 43(Database issue):D470-D478.
18. Chatr-aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, O'Donnell L, Oster S, Theesfeld C, Sellam A, Stark C, Breitkreutz B-J, Dolinski K, Tyers M. (2017) The BioGRID interaction database: 2017 update. *Nucleic Acids Res*, 45(Database issue):D369-D379.
19. Chautard E, Fatoux-Ardore M, Ballut L, Thierry-Mieg N, Ricard-Blum S. (2011) MatrixDB, the extracellular matrix interaction database. *Nucleic Acids Res*, 39(Database issue):D235-40.
20. Chen L, Liu R, Liu Z-P, Li M, Aihara K. (2012) Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci Rep*, 2:342.
21. Chiu L-Y, Hsin I-L, Yang T-Y, Sung W-W, Chi J-Y, Chang JT, Ko J-L, Sheu, G-T. (2017) The ERK-ZEB1 pathway mediates epithelial-mesenchymal transition in pemetrexed resistant lung cancer cells with suppression by vinca alkaloids. *Oncogene*, (36):242-253.

22. Coghlin A, Murray GI. (2010) Current and emerging concepts tumor metastasis. *J Pathol,* 222:1-15.

23. Cornish TC, Chakravarti A, Kapoor A, Halushka MK. (2015) HPASubC: A suite of tools for user subclassification of human protein atlas tissue images. *J Pathol Inf*, 6:36.

24. Csermely P, Korcsmáros T. (2013) Cancer-related networks: A help to understand, predict and change malignant transformation. *Semin Cancer Biol*, 23(4):209-212.

25. Csermely P, Hódsági J, Korcsmáros T, Módos D, Perez-Lopez ÁR, Szalay K, Veres DV, Lenti K, Wu LY, Zhang XS. (2015) Cancer stem cells display extremely large evolvability: Alternating plastic and rigid networks as a potential mechanism. Network models, novel therapeutic target strategies, and the contributions of hypoxia, inflammation and cellular senescence. *Semin Cancer Biol*, 30:42-51.

26. Csermely P, Korcsmáros T, Kiss HJ, London G, Nussinov R. (2013) Structure and dynamics of molecular networks: A novel paradigm of drug discovery. *Pharmacol Ther,* 138:333-408.

27. D'Angelo MA, Raices M, Panowski SH, Hetzer MW. (2009) Age-Dependent Deterioration of Nuclear Pore Complexes Causes a Loss of Nuclear Integrity in Postmitotic Cells. *Cell*, 136(2):284-295.

28. De Las Rivas J, Fontanillo C. (2010) Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. Lewitter F, ed. *PLoS Comput Biol*, 6(6):e1000807.

29. Dechat T, Pfleghaar K, Sengupta K, Shimi T, Shumaker DK, Solimando L, Goldman RD. (2008) Nuclear Lamins, Major Factors in the Structual Organization and Function of the Nucleus and Chromain. *Genes Dev*, 22(7):832-853.

30. Demarquoy J, Le Borgne F. (2015) Crosstalk between mitochondria and peroxisomes. *World J Biol Chem*, 6(4):301-309.

31. Dhillon AS, Hagan S, Rath O, Kolch W. (2007) MAP kinase signalling pathways in cancer. *Oncogene*, 26(22):3279-3290.

32. Dobronyi L, Mendik P, Csermely P, Veres DV. (2016) Translocatome: a novel tool for drug target discovery, based on the systematic analysis of protein translocation. *12th International Congress of Cell Biology*, Prague. (poster P173)

33. Elbaz Y, Schuldiner M. (2011) Staying in touch: the molecular era of organelle contact sites. *Trends Biochem Sci*, 36(11):616-623.

34. Eliaš J, Dimitrio L, Clairambault J, Natalini R. (2014) The p53 protein and its molecular network: modelling a missing link between DNA damage and cell fate. *Biochim Biophys Acta*, 1844:232-247.

35. Elion EA. (2006) Detection of Protein-Protein Interactions by Coprecipitation. *Curr Protoc Protein Sci,* Chapter 5:Unit 5.25.

36. Fazekas D, Koltai M, Türei D, Módos D, Pálfy M, Dúl Z, Zsákai L, Szalay-Bekő M, Lenti K, Farkas IJ, Vellai T, Csermely P, Korcsmáros T. (2013) SignaLink 2 - a signaling pathway resource with multi-layered regulatory networks. *BMC Syst Biol*, 7:7.

37. Fekete I, Szalay KZ, Veres DV, Csermely P. (2016) Robust computational prediction of personalized drug combinations. *Applied Bioinformatics in Life Sciences,* Leuven. (poster)

38. Fields S, Song O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245-246.

39. Firth SM, Baxter RC. (2002) Cellular actions of the insulin-like growth factor binding proteins. *Endocr Rev*, 23(6):824-854.

40. Fisk HA, Mattison CP, Winey M. (2003) Human Mps1 protein kinase is required for centrosome duplication and normal mitotic progression. *Proc Natl Acad Sci U S A*, 100(25):14875-14880.

41. Gant TM, Harris CA, Wilson KL. (1999) Roles of LAP2 Proteins in Nuclear Assembly and DNA Replication: Truncated LAP2ß Proteins Alter Lamina Assembly, Envelope Formation, Nuclear Size, and DNA Replication Efficiency in Xenopus laevis Extracts. *J Cell Biol,* 144(6):1083-1096.

42. García-Yagüe ÁJ, Rada P, Rojo AI, Lastres-Becker I, Cuadrado A. (2013) Nuclear import and export signals control the subcellular localization of Nurr1 protein in response to oxidative stress. *J Biol Chem*, 288(8):5506-5517.

43. Gibson TJ, Dinkel H, Van Roey K, Diella F. (2015) Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Commun Signal*, 13(1):42.

44. Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merkenschlager M, Gisel A, Ballestar E, Bongcam-Rudloff E, Conesa A, Tegnér J. (2014) Data

integration in the era of omics: current and future challenges. *BMC Syst Biol*, 8 Suppl 2(2):I1.

45. Gough NR. (2016) Emerging roles for organelles in cellular regulation. *Sci Signal*, 9:eg11.
46. Guda C. (2006) pTARGET: a web server for predicting protein subcellular localization. *Nucleic Acids Res*, 34(Web Server issue):W210-3.
47. Gurden MD, Westwood IM, Faisal A, Naud S, Cheung KM, McAndrew C, Wood A, Schmitt J, Boxall K, Mak G, Workman P, Burke R, Hoelder S, Blagg J, Van Montfort RL, Linardopoulos S. (2015) Naturally occurring mutations in the MPS1 gene predispose cells to kinase inhibitor drug resistance. *Cancer Res*, 75(16):3340-3354.
48. Güttinger S, Laurell E, Kutay U. (2009) Orchestrating nuclear envelope disassembly and reassembly during mitosis. *Nat Rev Mol Cell Biol*, 10(3):178-191.
49. Gyurkó DM, Veres DV, Módos D, Lenti K, Korcsmáros T, Csermely P. (2013) Adaptation and learning of molecular networks as a description of cancer development at the systems-level: Potential use in anti-cancer therapies. *Semin. Cancer Biol*, 23(4):262-269.
50. Gyurkó MD. (2015) A gap gének vizsgálata kísérletes és hálózatos módszerekkel (phd.semmelweis.hu/mwp/phd_live/vedes/export/gyurkomartondavid.m.pdf). *Doktori értekezés.*
51. Hamed RB, Batchelar ET, Clifton IJ, Schofield CJ. (2008) Mechanisms and structures of crotonase superfamily enzymes--how nature controls enolate and oxyanion reactivity. *Cell Mol Life Sci*, 65(16):2507-2527.
52. Hao N, O'Shea EK. (2012) Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat Struct Mol Biol*, 19(1):31-39.
53. Haraguchi T, Koujin T, Segura-Totten M, Lee KK, Matsuoka Y, Yoneda Y, Wilson KL, Hiraoka Y. (2001) BAF is required for emerin assembly into the reforming nuclear envelope. *J Cell Sci*, 114(24):4575-4585.
54. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. (2014) Clinical development success rates for investigational drugs. *Nat Biotech*, 32(1):40-51.
55. Hieb AR, D'Arcy S, Kramer MA, White AE, Luger K. (2012) Fluorescence strategies for high-throughput quantification of protein interactions. *Nucleic Acids Res*, 40(5):1-13.

56. Hill R, Cautain B, de Pedro N, Link W. (2014) Targeting nucleocytoplasmic transport in cancer therapy. *Oncotarget*, 5(1):11-28.

57. Hoek KS, Goding CR. (2010) Cancer stem cells versus phenotype-switching in melanoma. *Pigment Cell Melanoma Res*, 23(6):746-759.

58. Hoelz A, Debler EW, Blobel G. (2011) The structure of the nuclear pore complex. *Annu Rev Biochem*, 80:613-643.

59. Hornberg JJ, Bruggeman FJ, Westerhoff H V, Lankelma J. (2006) Cancer: a Systems Biology disease. *Biosystems*, 83(2-3):81-90.

60. Hu S, Xie Z, Onishi A, Yu X, Jiang L, Lin J, Rho HS, Woodard C, Wang H, Jeong JS, Long S, He X, Wade H, Blackshaw S, Qian J, Zhu H. (2009) Profiling the Human Protein-DNA Interactome Reveals ERK2 as a Transcriptional Repressor of Interferon Signaling. *Cell*, 139(3):610-622.

61. Hu Y, Lehrach H, Janitz M. (2009) Comparative analysis of an experimental subcellular protein localization assay and in silico prediction methods. *J Mol Histol*, 40(5-6):343-352.

62. Huang S, Ingber DE. (1999) The structural and mechanical complexity of cell-growth control. *Nat Cell Biol*, 1(5):E131-8.

63. Huang T-W, Lin C-Y, Kao C-Y. (2007) Reconstruction of human protein interolog network using evolutionary conserved network. *BMC Bioinformatics*, 8:152.

64. Huberts DHEW, van der Klei IJ. (2010) Moonlighting proteins: An intriguing mode of multitasking. *Biochim Biophys Acta – Mol Cell Res*, 1803(4):520-525.

65. Ihaka R, Gentleman R. (1996) R: A Language for Data Analysis and Graphics. *J Comput Graph Stat*, 5(3):299-314.

66. Inder KL, Davis M, Hill MM. (2013) Ripples in the pond-using a systems approach to decipher the cellular functions of membrane microdomains. *Mol Biosyst*, 9(3):330-338.

67. Iorio F, Knijnenburg TA, Bignell GR, Menden MP, Schubert M, Aben N, Gonçalves E, Barthorpe S, Lightfoot H, Cokelaer T, Greninger P, van Dyk E, Chang H, de Silva H, Heyn H, Deng X, Egan RK, Liu Q, Mironenko T, Mitropoulos X, Richardson L, Wang J, Zhang T, Moran S, Sayols S, Soleimani M, Tamborero D, Lopez-Bigas N, Ross-Macdonald P, Esteller M, Gray NS, Haber DA, Stratton MR, Benes CH, Wessels LFA, Saez-Rodriguez

J, McDermott U, Garnett MJ. (2016) A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*, 166(3):740-754.

68. Ivanov AA, Khuri FR, Fu H. (2013) Targeting protein-protein interactions as an anticancer strategy. *Trends Pharmacol Sci*, 34(7):393-400.
69. Jafri MA, Ansari SA, Algahtani MH, Shay JW. (2016) Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med*, 8:69.
70. Jemaá M, Galluzzi L, Kepp O, Senovilla L, Brands M, Boemer U, Koppitz M, Lienau P, Prechtl S, Schulze V, Siemeister G, Wengner AM, Mumberg D, Ziegelbauer K, Abrieu A, Castedo M, Vitale I, Kroemer G. (2013) Characterization of novel MPS1 inhibitors with preclinical anticancer activity. *Cell Death Differ,* 6:1-14.
71. Jeter CR, Badeaux M, Choy G, Chandra D, Patrawala L, Liu C, Calhoun-Davis T, Zaehres H, Daley GQ, Tang DG. (2009) Functional Evidence that the Self-Renewal Gene NANOG Regulates Human Tumor Development. *Stem Cells*, 27(5):993-1005.
72. Johnson AE, van Waes MA. (1999) The Translocon: A Dynamic Gateway at the ER Membrane. *Annu Rev Cell Dev Biol*, 15(1):799-842.
73. Johnson N, Powis K, High S. (2013) Post-translational translocation into the endoplasmic reticulum. *Biochim Biophys Acta – Mol Cell Res*, 1833(11):2403-2409.
74. Kabachinski G, Schwartz TU. (2015) The nuclear pore complex - structure and function at a glance. *J Cell Sci*, 128(3):423-429.
75. Kagami Y, Nihira K, Wada S, Ono M, Honda M, Yoshida K. (2014) Mps1 phosphorylation of condensin II controls chromosome condensation at the onset of mitosis. *J Cell Biol*, 205(6):781-790.
76. Kamburov A, Stelzl U, Lehrach H, Herwig R. (2013) The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res*, 41(Database issue):D793-800.
77. Kandasamy K, Keerthikumar S, Goel R, Mathivanan S, Patankar N, Shafreen B, Renuse S, Pawar H, Ramachandra YL, Acharya PK, Ranganathan P, Chaerkady R, Prasad TSK, Pandey A. (2009) Human Proteinpedia: a unified discovery resource for proteomics research. *Nucleic Acids Res*, 37(Database issue):D773-81.

78. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. (2017) KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*, 45(D1):D353-D361.

79. Kastenhuber ER and Lowe SW. (2017) Putting p53 in Context. *Cell*, 170(6):1062-1078.

80. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeiffenberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H. (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res*, 40(Database issue):D841-6.

81. Keshava PTS, 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, Ramabadran S, Chaerkady R, Pandey A. (2009) Human Protein Reference Database--2009 update. *Nucleic Acids Res*, 37(Database issue):D767-72.

82. Kietzmann T, Mennerich D, Dimova EY. (2016) Hypoxia-Inducible Factors (HIFs) and Phosphorylation: Impact on Stability, Localization, and Transactivity. *Front Cell Dev Biol*, 4(February):11.

83. Kleffel S, Schatton T. (2013) Tumor Dormancy and Cancer Stem Cells: Two Sides of the Same Coin?. In Enderling H, Almog N, Hlatky L (Eds.), *Systems Biology of Tumor Dormancy* (pp. 145-179).

84. Klein CA. (2010) Parallel progression of tumour and metastases. *Nat Rev Cancer*, 10(2):156.

85. Koh GCKW, Porras P, Aranda B, Hermjakob H, Orchard SE. (2012) Analyzing protein-protein interaction networks. *J Proteome Res*, 11(4):2014-2031.

86. Koh J, Blobel G. (2015) Allosteric regulation in gating the central channel of the nuclear pore complex. *Cell*, 161(6):1361-1373.

87. Kovács IA, Mizsei R, Csermely P. (2015) A unified data representation theory for network visualization, ordering and coarse-graining. *Sci Rep*, 5:13786.

88. Kumar G, Ranganathan S. (2010) Network analysis of human protein location. *BMC Bioinformatics*, 11 Suppl 7:S9.

89. Kwiatkowski N, Jelluma N, Filippakopoulos P, Soundararajan M, Manak MS, Kwon M, Choi HG, Sim T, Deveraux QL, Rottmann S, Pellman D, Shah JV, Kops GJ, Knapp S, Gray NS. (2010) Small-molecule kinase inhibitors provide insight into Mps1 cell cycle function. *Nat Chem Biol*, 6(5):359-368.

90. Launay G, Salza R, Multedo D, Thierry-Mieg N, Ricard-Blum S. (2015) MatrixDB, the extracellular matrix interaction database: Updated content, a new navigator and expanded functionalities. *Nucleic Acids Res*, 43(Database issue):D321-D327.

91. Lee K, Sung M-K, Kim J, Kim K, Byun J, Paik H, Kim B, Huh WK, Ideker T. (2014) Proteome-wide remodeling of protein location and function by stress. *Proc Natl Acad Sci*, 111(30):E3157-E3166.

92. Lehmann W, Mossmann D, Kleemann J, Mock K, Meisinger C, Brummer T, Herr R, Brabletz S, Stemmler MP, Brabletz T. (2016) ZEB1 turns into a transcriptional activator by interacting with YAP1 in aggressive cancer types. *Nat Commun*, 7:10498.

93. Lehne B, Schlitt T. (2009) Protein-protein interaction databases: keeping up with growing interactomes. *Hum Genomics*, 3(3):291-297.

94. Levy ED, Landry CR, Michnick SW. (2009) How perfect can protein interactomes be? *Sci Signal*, 2(60):pe11.

95. Lewitzky M, Simister PC, Feller SM. (2012) Beyond “furballs” and “dumpling soups” - towards a molecular architecture of signaling complexes and networks. *FEBS Lett*, 586(17):2740-2750.

96. Li J, Newberg JY, Uhlén M, Lundberg E, Murphy RF. (2012) Automated Analysis and Reannotation of Subcellular Locations in Confocal Images from the Human Protein Atlas. *PLoS One*, 7(11):e50514.

97. Li Z, Ivanov AA, Su R, Gonzalez-Pecchi V, Qi Q, Liu S, Webber P, McMillan E, Rusnak L, Pham C, Chen X, Mo X, Revennaugh B, Zhou W, Marcus A, Harati S, Chen X, Johns MA, White MA, Moreno C, Cooper LA, Du Y, Khuri FR, Fu H. (2017) The OncoPPi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies. *Nat Commun*, 8:14356.

98. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E, Castagnoli L, Cesareni G. (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res*, 40(Database issue):D857-61.

99. Lidke DS, Huang F, Post JN, Rieger B, Wilsbacher J, Thomas JL, Pouysségur J, Jovin TM, Lenormand P. (2010) ERK nuclear translocation is dimerization-independent but controlled by the rate of phosphorylation. *J Biol Chem*, 285(5):3092-3102.

100. Lim STS. (2013) Nuclear FAK: A new mode of gene regulation from cellular adhesions. *Mol Cells*, 36(1):1-6.

101. Liu S, Chan GKT, Hittle JC, Fujii G, Lees E, Yen TJ. (2003) Human MPS1 Kinase Is Required for Mitotic Arrest Induced by the Loss of CENP-E from Kinetochores. *Mol Biol Cell,* 14(4):1638–51.

102. Liu X, Feng R, Du L. (2010) The role of enoyl-CoA hydratase short chain 1 and peroxiredoxin 3 in PP2-induced apoptosis in human breast cancer MCF-7 cells. *FEBS Lett*, 584(14):3185-3192.

103. Liu Z, Li Q, Li K, Chen L, Li W, Hou M, Liu T, Yang J, Lindvall C, Björkholm M, Jia J, Xu D. (2013) Telomerase reverse transcriptase promotes epithelial-mesenchymal transition and stem cell-like traits in cancer cells. *Oncogene*, 32(36):4203-13.

104. López Y, Nakai K, Patil A. (2015) HitPredict version 4: Comprehensive reliability scoring of physical protein-protein interactions from more than 100 species. *Database*, 2015(1):1-10.

105. Lowe AR, Tang JH, Yassif J, Graf M, Huang WY, Groves JT, Weis K, Liphardt JT. (2015) Importin-ß modulates the permeability of the nuclear pore complex in a Ran-dependent manner. *Elife*, 2015(4):1-24.

106. Lu P, Szafron D, Greiner R, Wishart DS, Fyshe A, Pearcy B, Poulin B, Eisner R, Ngo D, Lamb N. (2005) PA-GOSUB: a searchable database of model organism protein sequences with their predicted Gene Ontology molecular function and subcellular localization. *Nucleic Acids Res*, 33(Database issue):D147-53.

107. Luck K, Sheynkman GM, Zhang I, Vidal M. (2017) Proteome-Scale Human Interactomics. *Trends Biochem Sci*, 42(5):342-354.

108. Maachani UB, Kramp T, Hanson R, Zhao S, Celiku O, Shankavaram U, Colombo R, Caplen NJ, Camphausen K, Tandle A. (2015) Targeting MPS1 Enhances Radiosensitization of Human Glioblastoma by Modulating DNA Repair Proteins. *Mol Cancer Res*, 13(5):852-862.

109. Macara IG. (2001) Transport into and out of the nucleus. *Microbiol Mol Biol Rev*, 65(4):570-94.

110. Maere S, Heymans K, Kuiper M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, 21(16):3448-3449.

111. Mani M, Chen C, Amblee V, Liu H, Mathur T, Zwicke G, Zabad S, Patel B, Thakkar J, Jeffery CJ. (2015) MoonProt: A database for proteins that are known to moonlight. *Nucleic Acids Res*, 43(D1):D277-D282.

112. Marcu LG, Harriss-Phillips WM. (2012) In silico modelling of treatment-induced tumour cell kill: Developments and advances. *Comput Math Methods Med,* Article ID 960256.

113. Mardakheh FK, Sailem HZ, Kümper S, Tape CJ, McCully RR, Paul A, Anjomani-Virmouni S, Jørgensen C, Poulogiannis G, Marshall CJ, Bakal C. (2017) Proteomics profiling of interactome dynamics by colocalisation analysis (COLA). *Mol BioSyst*, 13(1):92-105.

114. Margineanu A, Chan JJ, Kelly DJ, Warren SC, Flatters D, Kumar S, Katan M, Dunsby CW, French PM. (2016) Screening for protein-protein interactions using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM). *Sci Rep*, 6(1):28186.

115. Marx V. (2013) Biology: The big challenges of big data. *Nature*, 498(7453):255-260.

116. McDowall MD, Scott MS, Barton GJ. (2009) PIPs: Human protein-protein interaction prediction database. *Nucleic Acids Res*, 37(Database issue):651-656.

117. Meerang M, Ritz D, Paliwal S, Garajova Z, Bosshard M, Mailand N, Janscak P, Hübscher U, Meyer H, Ramadan K. (2011) The ubiquitin-selective segregase VCP/p97 orchestrates the response to DNA double-strand breaks. *Nat Cell Biol*, 13(11):1376-1382.

118. Mendik P, Dobronyi L, Hári F, Karapinar TO, Moco L, Csermely P, Veres DV. (2017) Translocatome: a novel tool for the functional analysis of protein translocation between cellular organelles. *Hungarian Molecular Life Sciences Conference*, Eger. (poster)

119. Mosca R, Céol A, Aloy P. (2012) Interactome3D: adding structural details to protein networks. *Nat Methods*, 10(1):47-53.

120. Murali T, Pacifico S, Yu J, Guest S, Roberts GG, Finley RL. (2011) DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. *Nucleic Acids Res*, 39(Database issue):D736-43.

121. Musacchio A, Salmon ED. (2007) The spindle-assembly checkpoint in space and time. *Nat Rev Mol Cell Biol*, 8(5):379-393.

122. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. (2012) The Matrisome: In Silico Definition and In Vivo Characterization by Proteomics of Normal and Tumor Extracellular Matrices. *Mol Cell Proteomics*, 11(4):M111.014647-M111.014647.

123. Nilsson J, Yekezare M, Minshull J, Pines J. (2008) The APC/C maintains the spindle assembly checkpoint by targeting Cdc20 for destruction. *Nat Cell Biol*, 10(12):1411-1420.

124. Nyathi Y, Wilkinson BM, Pool MR. (2013) Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Biophys Acta – Mol Cell Res*, 1833(11):2392-2402.

125. O'Brate A and Giannakakou P. (2003) The importance of p53 location: nuclear or cytoplasmic zip code? *Drug Resist Updat,* 6(6):313-22.

126. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L, Lovering RC, Meldal B, Melidoni AN, Milagros M, Peluso D, Perfetto L, Porras P, Raghunath A, Ricard-Blum S, Roechert B, Stutz A, Tognolli M, van Roey K, Cesareni G, Hermjakob H. (2014) The MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*, 42(Database issue):358-363.

127. Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, Bridge A, Briganti L, Brinkman FS, Cesareni G, Chatr-aryamontri A, Chautard E, Chen C, Dumousseau M, Goll J, Hancock RE, Hannick LI, Jurisica I, Khadake J, Lynn DJ, Mahadevan U, Perfetto L, Raghunath A, Ricard-Blum S, Roechert B, Salwinski L, Stümpflen V, Tyers M, Uetz P, Xenarios I, Hermjakob H. (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods*, 9(4):345-350.

128. Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stümpflen 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 Rivas J, 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(8):894-898.

129. Ostapenko D, Burton JL, Solomon MJ. (2012) Identification of anaphase promoting complex substrates in S. cerevisiae. *PLoS One*, 7(9):e45895.
130. Ota M, Gonja H, Koike R, Fukuchi S. (2016) Multiple-Localization and Hub Proteins. *PLoS One*, 11(6):1-18.
131. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes HW, Ruepp A, Frishman D. (2005) The MIPS mammalian protein-protein interaction database. *Bioinformatics*, 21(6):832-834.
132. Parker BS, Rautela J, Hertzog PJ. (2016) Antitumour actions of interferons: implications for cancer therapy. *Nat Rev Cancer*, 16(3):131-144.
133. Pierleoni A, Martelli PL, Fariselli P, Casadio R. (2006) BaCelLo: a balanced subcellular localization predictor. *Bioinformatics*, 22(14):e408-16.
134. Pierleoni A, Martelli PL, Fariselli P, Casadio R. (2007) eSLDB: eukaryotic subcellular localization database. *Nucleic Acids Res,* 35(Database issue):D208-12.
135. Plotnikov A, Zehorai E, Procaccia S, Seger R. (2011) The MAPK cascades: Signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochim Biophys Acta – Mol Cell Res*, 1813(9):1619-1633.
136. Pontén FK, Schwenk JM, Asplund A, Edqvist PHD. (2011) The Human Protein Atlas as a proteomic resource for biomarker discovery. *J Intern Med*, 270(5):428-446.
137. Prasad V, Fojo T, Brada M. (2016) Precision oncology: Origins, optimism, and potential. *Lancet Oncol*, 17(2):e81-e86.
138. Prior IA, Lewis PD, Mattos C. (2012) A comprehensive survey of Ras mutations in cancer. *Cancer Res*, 72(10):2457-2467.
139. Pu T, Zhang X-P, Liu F, Wang W. (2010) Coordination of the Nuclear and Cytoplasmic Activities of p53 in Response to DNA Damage. *Biophys J*, 99(6):1696-1705.
140. Rabut G, Doye V, Ellenberg J. (2004) Mapping the dynamic organization of the nuclear pore complex inside single living cells. *Nat Cell Biol*, 6(11):1114-1121.

141. Rahmati S, Abovsky M, Pastrello C, Jurisica I. (2017) PathDIP: An annotated resource for known and predicted human gene-pathway associations and pathway enrichment analysis. *Nucleic Acids Res*, 45(Database issue):D419-D426.

142. Rolland T, Taşan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly MP, Carvunis AR, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan C, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang S, Kiros R, Lin GN, Luck K, MacWilliams A, Menche J, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruyssinck E, Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejeda AO, Trigg SA, Twizere JC, Vega K, Walsh J, Cusick ME, Xia Y, Barabási AL, Iakoucheva LM, Aloy P, De Las Rivas J, Tavernier J, Calderwood MA, Hill DE, Hao T, Roth FP, Vidal M. (2014) A Proteome-Scale Map of the Human Interactome Network. *Cell*, 159:1212-1226.

143. Rudolph JD, de Graauw M, van de Water B, Geiger T, Sharan R. (2016) Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Networks. *Cell Syst*, 3(6):585-593.e3.

144. Sabbattini P, Canzonetta C, Sjoberg M, Nikic S, Georgiou A, Kemball-Cook G, Auner HW, Dillon N. (2007) A novel role for the Aurora B kinase in epigenetic marking of silent chromatin in differentiated postmitotic cells. *EMBO J*, 26(22):4657-4669.

145. Sáez-Ayala M, Montenegro MF, Sánchez-del-Campo L, Fernández-Pérez MP, Chazarra S, Freter R, Middleton M, Piñero-Madrona A, Cabezas-Herrera J, Goding CR, Rodríguez-López JN. (2013) Directed Phenotype Switching as an Effective Antimelanoma Strategy. *Cancer Cell*, 24(1):105-119.

146. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H, Sirota-Madi A, Olender T, Golan Y, Stelzer G, Harel A, Lancet D. (2010) GeneCards Version 3: the human gene integrator. *Database (Oxford)*, 2010:baq020.

147. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res*, 32(Database issue):D449-51.

148. Sánchez-Tilló E, Fanlo L, Siles L, S Montes-Moreno, A Moros, G Chiva-Blanch, R Estruch, A Martinez, D Colomer, B Győrffy, G Roué, and A Postigo. (2014) The EMT activator ZEB1 promotes tumor growth and determines differential response to chemotherapy in mantle cell lymphoma. *Cell Death Differ*, 21(2):247-257.

149. Scaltriti M, Baselga J. (2006) The Epidermal Growth Factor Receptor Pathway: A Model for Targeted Therapy. *Clin Cancer Res*, 12(18):5268-5272.

150. Semenza GL. (2009) Regulation of Oxygen Homeostasis by Hypoxia-Inducible Factor 1. *Physiology (Bethesda),* 24(2):97-106.

151. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, (13):2498-2504.

152. Shrestha D, Jenei A, Nagy P, Vereb G, Szöllősi J. (2015) Understanding FRET as a Research Tool for Cellular Studies. *Int J Mol Sci,* 16(4):6718-6756.

153. Siegel RL, Miller KD, Jemal A. (2016) Cancer statistics. *CA Cancer J Clin*, 66(1):7-30.

154. Sprenger J, Fink JL, Karunaratne S, Hanson K, Hamilton NA, Teasdale RD. (2008) LOCATE: a mammalian protein subcellular localization database. *Nucleic Acids Res*, 36(Database issue):D230-3.

155. Steinway SN, Zañudo JGT, Michel PJ, Feith DJ, Loughran TP, Albert R. (2015) Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *npj Syst Biol Appl*, 1(August):15014.

156. Szalay-Bekő M, Palotai R, Szappanos B, Kovács IA, Papp B, Csermely P. (2012) ModuLand plug-in for Cytoscape: Determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics*, 28(16):2202-2204.

157. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. (2015) STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*, 43(Database issue):D447-D452.

158. Tam WL, Lu H, Buikhuisen J, Soh BS, Lim E, Reinhardt F, Wu ZJ, Krall JA, Bierie B, Guo W, Chen X, Liu XS, Brown M, Lim B, Weinberg RA. (2013)

Protein kinase C α is a central signaling node and therapeutic target for breast cancer stem cells. *Cancer Cell*, 24(3):347-364.

159. Tamaki T, Shimizu T, Niki M, Shimizu M, Nishizawa T, Nomura S. (2017) Immunohistochemical analysis of NANOG expression and epithelial-mesenchymal transition in pulmonary sarcomatoid carcinoma. *Oncol Lett,* 13(5):3695-3702.
160. The Gene Ontology Consortium. (2013) Gene ontology annotations and resources. *Nucleic Acids Res,* 41(D1):530-535.
161. The UniProt Consortium. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res*, 45(Database issue): D158-D169.
162. Thiery JP, Acloque H, Huang RYJ, Nieto MA. (2009) Epithelial-Mesenchymal Transitions in Development and Disease. *Cell*, 139(5):871-890.
163. Tompa P, Szász C, Buday L. (2005) Structural disorder throws new light on moonlighting. *Trends Biochem Sci*, 30(9):484-489.
164. Tulchinsky E, Pringle JH, Caramel J, Ansieau S. (2014) Plasticity of melanoma and EMT-TF reprogramming. *Oncotarget*, 5(1):1-2.
165. Turteltaub W, Murphy A. (1987) Subcellular Localization and Capacity of ß-Oxidation Aldehyde Dehydrogenase in Porcine Liver. *Arch Biochem Biophys*, 255(1):120-6.
166. Türei D, Korcsmáros T, Saez-Rodriguez J. (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods*, 13(12):966-967.
167. Veres D. (2013) ComPPI kompartmentalizált fehérje-fehérje interakciós hálózatok készítése és felhasználása. *Szakdolgozat*.
168. Veres DV, Gyurkó DM, Thaler B, Szalay KZ, Fazekas D, Korcsmáros T, Csermely P. (2015) ComPPI: A cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res*, 43(Database issue):D485-D493.
169. Vidal M, Cusick ME, Barabási A-L. (2011) Interactome networks and human disease. *Cell*, 144(6):986-998.
170. Wan S, Mak M-W, Kung S-Y. (2017) FUEL-mLoc: feature-unified prediction and explanation of multi-localization of cellular proteins in multiple organisms. *Bioinformatics*, 33(5):749-750.

171. Watanabe M, Ohnishi Y, Inoue H, Wato M, Tanaka A, Kakudo K, Nozaki M. (2014) NANOG expression correlates with differentiation, metastasis and resistance to preoperative adjuvant therapy in oral squamous cell carcinoma. *Oncol Lett,* 7(1):35-40.

172. Waterson RM, Hill RL. (1972) Enoyl Coenzyme A Hydratase (Crotonase). *J Biol Chem*, 247(16):5258-5205.

173. Wein SP, Côté RG, Dumousseau M, Reisinger F, Hermjakob H, Vizcaíno JA. (2012) Improvements in the protein identifier cross-reference service. *Nucleic Acids Res*, 40(Web Server issue):276-280.

174. Winter AG, Wildenhain J, Tyers M. (2011) BioGRID REST service, BiogridPlugin2 and BioGRID WebGraph: New tools for access to interaction data at BioGRID. *Bioinformatics*, 27(7):1043-1044.

175. Wiwatwattana N, Kumar A. (2005) Organelle DB: a cross-species database of protein localization and function. *Nucleic Acids Res*, 33(Database issue):D598-604.

176. Wiwatwattana N, Landau CM, Cope GJ, Harp GA, Kumar A. (2007) Organelle DB: an updated resource of eukaryotic protein localization and function. *Nucleic Acids Res*, 35(Database issue):D810-4.

177. Wu C. (1997) Chromatin Remodeling and the Control of Gene Expression. *J Biol Chem*, 272(45):28171-28174.

178. Xie Q, Soutto M, Xu X, Zhang Y, Johnson CH. (2011) Bioluminescence Resonance Energy Transfer (BRET) Imaging in Mammalian Cells. *Methods Mol Biol*, 680:29-43.

179. Yeh CS, Wang JY, Cheng TL, Juan CH, Wu CH, Lin SR. (2006) Fatty acid metabolism pathway play an important role in carcinogenesis of human colorectal cancers by microarray-bioinformatics analysis. *Cancer Lett*, 233(2):297–308.

180. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. (2010) PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 26(13):1608-1615.

181. Yugi K, Kubota H, Hatano A, Kuroda S. (2016) Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple “Omic” Layers. *Trends Biotechnol*, 34(4):276-290.

182. Zahiri J, Bozorgmehr JH, Masoudi-Nejad A. (2013) Computational Prediction of Protein–Protein Interaction Networks: Algorithms and Resources. *Curr Genomics*, 14(6):397-414.

183. Zehorai E, Yao Z, Plotnikov A, Seger R. (2010) The subcellular localization of MEK and ERK-A novel nuclear translocation signal (NTS) paves a way to the nucleus. *Mol Cell Endocrinol*, 314(2):213-220.

184. Zhang J, Sun M, Li R, Liu S, Mao J, Huang Y, Wang B, Hou L, Ibrahim MM, Tang J. (2013) Ech1 is a potent suppressor of lymphatic metastasis in hepatocarcinoma. *Biomed Pharmacother*, 67(7):557-560.

185. Zhang X, Ling Y, Guo Y, Bai Y, Shi X, Gong F, Tan P, Zhang Y, Wei C, He X, Ramirez A, Liu X, Cao C, Zhong H, Xu Q, Ma RZ. (2016) Mps1 kinase regulates tumor cell viability via its novel role in mitochondria. *Cell Death Dis*, 7:e2292.

186. Zhang X, Yin Q, Ling Y, Zhang Y, Ma R, Ma Q, Cao C, Zhong H, Liu X, Xu Q. (2011) Two LXXLL motifs in the N terminus of mps1 Are required for mps1 nuclear import during G2/M transition and sustained spindle checkpoint responses. *Cell Cycle*, 10(16):2742-2750.

187. Zhou H, Yang Y, Shen H-B. (2017) Hum-mPLoc 3.0: Prediction enhancement of human protein subcellular localization through modeling the hidden correlations of gene ontology and functional domain features. *Bioinformatics*, 33(6):843-853.

188. Zhu XS, Dai YC, Chen ZX, Xie JP, Zeng W, Lin YY, Tan QH. (2016) Knockdown of ECHS1 protein expression inhibits hepatocellular carcinoma cell proliferation via suppression of Akt activity. *Crit Rev Eukaryot Gene Exp,* 23(3):275–282.

12. Saját publikációk jegyzéke

12.1. A disszertáció témájához kapcsolódó közlemények

1. Adami HO, Csermely P, **Veres DV**, Emilsson L, Løberg M, Bretthauer M, Kalager M. (2017) Are rapidly growing cancers more lethal? *Eur J Cancer,* 72:210-214. (IF: 6,163, Google Scholar idézettség: 0)
2. Csermely P, Hódsági J, Korcsmáros T, Módos D, Perez-Lopez ÁR, Szalay K, **Veres DV**, Lenti K, Wu LY, Zhang XS. (2015) Cancer stem cells display extremely large evolvability: Alternating plastic and rigid networks as a potential mechanism. Network models, novel therapeutic target strategies, and the contributions of hypoxia, inflammation and cellular senescence. *Semin Cancer Biol*, 30:42-51. (IF: 9,955, Google Scholar idézettség: 39, Web of Science idézettség: 20)
3. Gyurkó DM, **Veres DV**, Módos D, Lenti K, Korcsmáros T, Csermely P. (2013) Adaptation and learning of molecular networks as a description of cancer development at the systems-level: Potential use in anti-cancer therapies. *Semin Cancer Biol*, 23(4):262-269. (IF: 9,143, Google Scholar idézettség: 16, Web of Science idézettség: 12)
4. **Veres DV**, Gyurkó DM, Thaler B, Szalay KZ, Fazekas D, Korcsmáros T, Csermely P. (2015) ComPPI: A cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res*, 43(Database issue):D485-D493. (IF: 9,202, Google Scholar idézettség: 21, Web of Science idézettség: 14)

12.2. A disszertáció témájához nem kapcsolódó közlemények

1. Csermely P, Sandhu KS, Hazai E, Hoksza Z, Kiss HJ, Miozzo F, **Veres DV**, Piazza F, Nussinov R. (2012) Disordered Proteins and Network Disorder in Network Descriptions of Protein Structure, Dynamics and Function: Hypotheses and a Comprehensive Review. *Curr Protein Pept Sci*, 13(1):19-33. (IF: 2,326, Google Scholar idézettség: 55, Web of Science idézettség: 36)
2. Simkó GI, Gyurkó D, **Veres DV**, Nánási T, Csermely P. (2009) Network strategies to understand the aging process and help age-related drug design. *Genome Med*, 1(9):90. (IF: 0,0, Google Scholar idézettség: 32, Web of Science idézettség: 22)
3. Pákó J, **Veres D**, Tisza J, Horváth I. (2015) A COPD a multimorbiditás tükrében. *Orvosi Továbbképző Szemle.* XXII. évf. 11. szám.

DOI:10.14753/SE.2018.2134

13. Köszönetnyilvánítás

Kiemelkedő köszönettel tartozom témavezetőmnek és mentoromnak, Csermely Péter professzor úrnak 2009 óta tartó támogatásáért, mely mind tudományos diákkörös, mind doktori hallgatói időszakom alatt végig kísérte és segítette kutatómunkámat.

Köszönöm a lehetőséget Mandl József professzor úrnak, a Semmelweis Egyetem Molekuláris Orvostudományok Doktori Iskola korábbi vezetőjének, valamint az Orvosi Vegytani, Molekuláris Biológiai és Patobiokémiai Intézet korábbi igazgatójának, illetve Ligeti Erzsébet professzor asszonynak az Iskola jelenlegi vezetőjének és Bánhegyi Gábor professzor úrnak, az intézet jelenlegi igazgatójának, akik jóvoltából a doktori iskolában és az intézetben végezhettem kutatómunkám.

Kutatócsoportunk, a LINK csoport (http://linkgroup.hu/) tagjai számos hasznos kérdéssel és tanáccsal segítették kutatásaim. Kiemelném a ComPPI adatbázis fejlesztésében résztvevő csapatot, Gyurkó M. Dávidot, Szalay Kristóf Zsoltot és Thaler Benedeket a kitartó közös munkáért, illetve a NetBiol (http://netbiol.elte.hu/) hálózatkutatócsoport tagjait: Fazekas Dávidot és Korcsmáros Tamást, akik tanácsaikkal segítették a munkát. A Translocatome projekt csapatnak is külön köszönetem az eredményes közös munkáért, így Dobronyi Leventének, Hári Ferencnek, Kerepesi Csabának, Leonardo Moconak és Mendik Péternek.

Kutatómunkám mellett a Turbine Kft. (http://turbine.ai) alapítójaként köszönettel tartozom az egész csapatnak támogatásukért, kiemelten Mérő Lászlónak a Translocatome projektben is használt kézi adatgyűjtő felület kifejlesztéséért.

Külföldi tanulmányutam során témavezetőm volt Andreas Bender a Cambridge-i Egyetemen, illetve Nathan Brown az Institute of Cancer Research-ben, akiknek külön köszönöm segítő munkájukat a kint töltött idő során.

Köszönöm az anyagi támogatást az Országos Tudományos Kutatási Alapprogramtól, az Új Nemzeti Kiválósági Programtól, illetve a Campus Hungary ösztöndíjat, mely külföldi tanulmányutamat támogatta.

Végül köszönöm családomnak, kiemelten édesapámnak, Veres Gábornak, aki elindított a tudomány megismerésének rögös útján inspiráló beszélgetéseinkkel, valamint feleségemnek, Veres-Antalfi Rékának, aki odaadó támogatásával megteremtette a stabil személyes hátteret doktori munkám során.