

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

Doktori tézisek

**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

## Bevezetés

A daganatos betegségek világszerte a vezető halálokok közé tartoznak, egyes országokban átvéve a korábban egyeduralkodó, a kardiovaszkuláris betegségekkel összefüggő halálozás elsődleges helyét. A rendszerszintű megközelítések (így például hálózatelemzés) az eddiginél sokkal több adatot felhasználó, sokkal kiterjedtebb elemzésekre adnak lehetőséget, így például az *in vitro* sejtvonalas és *in vivo* emberben megtalálható daganatos molekuláris minták farmakogenomikai elemzésére, mely segítheti a hatékonyabb daganat ellenes küzdelmet.

A magas áteresztőképességű kísérletes módszerek nagy mennyiségű fehérje kölcsönhatásra vonatkozó adatot szolgáltatnak. A fehérje-fehérje interakciós adatok mennyiségének növelésén túl azonban esszenciális az adatok megbízhatóságának javítása is. Minél megbízhatóbbak az adatok, annál precízebb biológiai hipotézisek felállítására van lehetőség az elemzésükkel. A minőség javítására számos megközelítés létezik. Kiemelkedik az adatok integrációjára vonatkozó törekvés, amely segítségével a több forrásból, eltérő kísérletekből származó interakciók létezésének bizonyossága magasabb, mint az egyedi adatbázisokban megtalálható, egy-egy mérésen alapuló kölcsönhatások biológiai valószínűsége.

A fehérjék térbeli elhelyezkedése biztosítja, hogy időben és térben eltérő feladatokat tudjanak ellátni, amely fontos szervező mechanizmusa a sejtekben zajló biokémiai folyamatoknak. A jelátviteli útvonalak kompartment szintű regulációjának szemléletes példája a transzkripció faktorok sejtmagi transzlokáció által mediált aktiválódása. A fehérje transzlokáció fogalmát olyan rendszerbiológiai jelenségként határozzuk meg, ahol a fehérjék poszttranszlációs formában szabályozott módon helyeződnek át a sejt-kompartmentek között. Transzlokáció során a fehérje kölcsönható partnerei és betöltött funkciója változik.

A 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. Ennek egyik példája a ERK2 mitogén-aktivált protein kináz, amely egy tipikus multifunkcionális fehérje. Az

ERK2 eltérő sejtplazmai és sejtmagi funkcióval és biokémiai aktivitással rendelkezik, amely a fehérje foszforiláltságával finoman szabályozott. Megfigyelhető, hogy ez a szabályozás szorosan összefügg a daganatos sejtek malignitásának mértékével is. Így az ERK2 jó példa arra, hogy a szubcelluláris lokalizáció milyen fontos szerepet játszik a fehérje funkcióban és ezzel a jelátvitelben, illetve arra is, hogy a szubcelluláris lokalizáció milyen fontos lehet a daganatos progresszió súlyosságának a megítélésében.

A proliferációval és invazivitással összefüggő, eltérő szabályozások egyensúlya kulcsfontosságú a daganatok malignitásának értékelésében. A daganatokban lassan proliferáló, ellenben igen malignus sejtek jelenléte az egyik oka lehet a konvencionális antiproliferatív terápiákkal szembeni limitált érzékenységnek vagy a kialakuló rezisztenciának, amelyre a metasztázis képzéssel kapcsolatos és a konvencionális onkogén útvonalakat célzó két lépcsős terápia lehet a megoldás.

## 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.
- 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.
- 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 epidemiológiai adatok értelmezésében.

## Módszerek

**Modell organizmusok kiválasztása** - A ComPPI adatbázis a következő négy fajra tartalmaz adatokat: *Saccharomyces cerevisiae* – élesztőgomba, *Caenorhabditis elegans* – fonálféreg, *Drosophila melanogaster* – ecetmuslica, *Homo sapiens* – ember.

**A fehérje lokalizációs és interakciós adatok forrása** - A ComPPI adatbázis 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 szubcelluláris lokalizációs adatok származhatnak kísérletes, prediktált vagy ismeretlen forrásból. 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.

A Translocatome adatbázis fehérje-fehérje interakciós és szubcelluláris lokalizációs adatai a ComPPI adatbázisból származnak. A transzlokálódó fehérjék jelenleg is folyamatban lévő kézi gyűjtése során a PubMed és Google Scholar szervereket használjuk a fehérjéket leíró cikkek keresésére. Az egyes fehérjék azonosításához, és további funkciók kereséséhez döntően a UniProt, NCBI Gene és GeneCards oldalakat alkalmazzuk.

**A fehérjék funkcionális elemzésére használt módszerek** - A fehérjék annotációjára a Gene Ontology adatkészletét használtuk, melyet az AmiGO webserveren keresztül böngészünk. A fehérjék funkcionális elemzésére az interaktómok megjelenítésére alkalmas Cytoscape hálózat megjelenítő és elemző program beépülő alkalmazását, a BiNGO-t választottuk (3.0.2-es verzió), mely alkalmas a megjelenített hálózat elemeinek közvetlen, interaktív vizsgálatára. A Translocatome adatbázis építése során a lokalizációs adatokat a ComPPI nevezéktanának megfelelően rögzítjük, a jelátviteli útvonalakat és a betegségneveket a KEGG adatforrás alapján rendszerezzük.

**A fehérjék hálózatos megjelenítéséhez és elemzéséhez használt eszközök** - A hálózatok megjelenítésére két eszközt használtunk, melyek közül a Gephi 0.8.2beta verzióját alkalmaztuk a nagyobb méretű hálózatok gyors és áttekinthető megjelenítéséhez. A Cytoscape alkalmazást használtuk az egyes fehérjék szubcelluláris lokalizáció specifikus kölcsönhatási hálózatának megjelenítésére (többek között a kutatócsoportunk

által fejlesztett EntOpt hálózat ábrázoló modult), illetve az adatok elemzésére. A hálózatos mérőszámok kiszámolásához az alkalmazásba beépített NetworkAnalyzer eszközt, illetve a munkacsoportunkban fejlesztett ModuLand hálózatos modularizáló algoritmust alkalmaztuk.

**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 alkalmaztuk (3.1.0 verzió). A könnyebb kezelhetőség érdekében az R kódot az RStudio nevű felhasználóbarát alkalmazás segítségével írtuk és futtattuk. Az alapvető táblázatos műveleteket és statisztikai elemzéseket a Microsoft Excel 2013-as verziójával végeztük.

**Adatbázis és webes felület kialakításához használt módszerek** - A ComPPI forrás adatok illesztőfelületek segítségével csatlakoznak az adatbázishoz. Az adatbázis maga MySQL 5 Community Edition alapú, saját nevezéktannal. A webes felület PHP5 nyelven íródott és nginx HTTP szerveren alapul. A Translocatome adatbázis MongoDB, a kézi adatgyűjtést segítő webes felület Ruby on Rails alapú. A gépi tanuláshoz a Python programozási nyelven alapuló scikit-learn csomagot használjuk.

## Eredmények

**A ComPPI adatbázis felépítési folyamatának sematikus bemutatása** - A ComPPI adatbázis igen fontos alkalmazása a biológiailag nem valószínű kapcsolatok kiszűrésének lehetősége a szubcelluláris lokalizáció alapján. Az adatbázis 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 állt össze, mely folyamatot számos helyen kézzel ellenőriztünk és/vagy segítettünk.

Az integrációra való igényt az első, exploratív adatintegrációs lépések során a különböző adatkészletek között észlelt alacsony átfedés alapozta meg. Az elérhető adatforrások keresésének és ellenőrzésének eredményeképpen 9 fehérje-fehérje interakciós adatbázist és 8 fehérje szubcelluláris lokalizációs adatbázist összegeztünk.

A szubcelluláris lokalizációs adatbázisok eltérő szintű adatokat tartalmaznak, így például a predikciós algoritmusok csak a nagy organelumok szerint osztják el a fehérjéket, míg a kísérletes adatok a fehérjék egészen konkrét sejten belüli elhelyezkedéséről is biztosítanak információt. A lokalizációs adatok felbontásán túl az egyes kompartmentek elnevezése is különbözik. Emiatt munkánk első lépésként a kompartmentek eltérő elnevezéseit egységes nevezéktanra emeltük. A folyamat második lépésként kézi gyűjtéssel létrehoztunk egy hierarchikus, redundancia mentes lokalizációs fát, amely a lokalizációkat a Gene Ontology nevezéktanának megfelelően egyértelműen besorolja.

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 névkonvenciókat a végső, legmegbízhatóbb UniProt nevekre. Az adatok sikeres integrációját követően kiszámítottuk a megbízhatósági értéket mind a lokalizációs, mind pedig az interakciós adatokra.

Az adatbázist kiszolgáló felületet a nyilvános használat előtt sokoldalúan teszteltük, ellenőrizve a keresési eredményeket, a letölthető adatokat és a webes funkciókat.

**A megbízhatósági értékek számítása** - A ComPPI lokalizációs megbízhatósági érték egy általunk definiált, új mérőszám, melynek segítségével az egyes fehérjéknek az

adott, nagyobb lokalizáción belüli elhelyezkedésének valószínűségét értékeljük. Ezen információ a lokalizáció-specifikus interakciós mérőszámok eloszlásának segítségével felhasználható magas megbízhatóságú interaktómok építésére. Az érték függ a lokalizációs adat kísérletes, prediktált vagy ismeretlen eredetétől, valamint az adatforrások számától.

A lokalizációs megbízhatósági érték számítása során a különböző lokalizációs evidenciákat vettük figyelembe, kiegészítve azt a ComPPI-ban elérhető forrás lokalizációs adatok számával az egyes fehérjék adott lokalizációjában. 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 a hat nagy lokalizációs kategóriára vonatkozó kompartment-specifikus lokalizációs megbízhatósági értékből számítható.

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.

**A ComPPI adatbázis statisztikája** - A ComPPI 1.1-es verziójának komprehenzív, integrált adatkészlete 127.757 fehérjét és azok 791.059 kapcsolatát, valamint 195.815 nagy lokalizáció szerinti elhelyezkedési adatot foglal magába. A proteóm-szintű adatkészlet az 5 nagy sejten belüli organelumra (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 Gene Ontology sejt-komponenst lefedve.

**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 indíthatóak. A szűrési lehetőségeket mind a keresett fehérjére, mind annak kölcsönható partnereire is érvényesíteni lehet. A ComPPI letöltő oldalán lehetőség nyílik előre meghatározott adatkészletek letöltésére is.

**A ComPPI adatainak felhasználása az egyes fehérjék szintjén** - 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, majd összehasonlítottuk a fokszám eloszlását a teljes interaktómban és a magas megbízhatóságú interaktómban is. Ezek után a megbízható UniProt SwissProt nevezéktanra szűkített találati listából kézzel átnéztük a fokszámra és köztiség mérőszámokra nézve legnagyobb eltérést mutató első 20 fehérjét. A 20 fehérje közül az enoil-CoA hidratáz, vagy más néven krotonáz rendelkezett a legnagyobb abszolút fokszám változással, így ezt választottuk ki további elemzésre.

A krotonáz a zsírsavak béta-oxidációjának második lépését katalizáló enzim, mely a krotonáz fehérje szuperfamilia egyik fő tagja. A zsírsavak béta-oxidációja elsősorban a mitokondriumban zajlik, mely megegyezik a krotonáz kísérletes mitokondriális szubcelluláris lokalizációs adatával. 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 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 irodalomkutatót végeztünk. A krotonáz fokozott kifejeződése volt megfigyelhető májtumoros sejtek sejtplazmájában, ahol az közrejátszott a nyirok áttétek kialakulásában. Ezt a vonalat továbbfejtvé az interakciós partnerek Gene Ontology adatbázisban található biológiai folyamatainak dúsulását vizsgáltuk a BiNGO segítségével. A mitokondriális interaktorok elemzése során a katabolizmussal és

apoptózis negatív szabályozásával összefüggő, valamint ezekkel kapcsolatos folyamatok dúsultak szignifikáns mértékben. Ezzel egyetértésben kimutatták, hogy a krotonáz számos daganattípusban fokozott mértékben fejeződik ki és kiütése esetén májtumorokban csökken a sejtek túlélőképessége, egyúttal növekszik a ciszplatin kezelésre adott apoptotikus válasz. Hasonló szerepét mutatták ki emlődaganatos sejtvonalakban is, ahol a csökkent kifejeződése a PP2-által indukált apoptózist segítette elő.

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 intervenciók lehetőségeire, mint például a krotonáz sejtplazmatikus funkcióinak gátlása máj vagy emlő daganatokban.

**A ComPPI adataira épülő rendszerszintű fehérje transzlokációs adatbázis bemutatása** - A transzlokálódó fehérjék fontos elemei a kompartment szintű interakciónak, 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.

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.

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 funkciókat automatikusan rendeljük hozzá a fehérjékhez a Gene Ontology adatainak felhasználásával. A gépi tanulás során több algoritmust is kipróbáltunk, melyek közül eddigi eredményeink alapján legjobban (80% feletti átlagos pontosság) a neurális hálózat klasszifikálja a transzlokálódó fehérjékre. 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.

A kézzel gyűjtött fehérjék és azok tulajdonságai, a ComPPI adatbázisból származó fehérjelista és a prediktált transzlokációs valószínűségi érték, a fehérjék közti kölcsönhatások, 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/>). 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, mely segíti az adatok kézi gyűjtését, ellenőrzését, és a felhasználói jogosultságok kezelésével lehetőséget ad az adatok szerkesztésére egyszerre számos felhasználó számára.

**A fehérjék térbeli elhelyezkedésének szerepe a daganatos malignitás meghatározásában** - A kutatócsoportunk által készített, saját eredményeinket és az elérhető irodalmi evidenciát összegző hipotézis szerint a daganatok rosszindulatú átalakulása felfogható egy kétlépéses folyamatként, ahol 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. A felállított hipotézisnek megfelelően a daganatok adaptációja során a plasztikusabb molekuláris hálózattal rendelkező fenotípus megfeleltethető a gyorsan osztódó, kevésbé invazív daganatos tulajdonságnak, míg a

rigidebb hálózat a kevésbé osztódó, nyugvó sejtekre jellemző, melyek azonban kifejezetten invazívak is lehetnek.

A daganatok kétlépcsős fejlődéséről szóló hipotézisünket számos olyan molekuláris megfigyelés is alátámasztja, amelyek a fehérjék szubcelluláris lokalizációjával is összefüggenek. A dolgozatban bemutatott példák olyan fehérjék, 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 e hatásaik mértéke szubcelluláris lokalizációjuk függvénye lehet. E molekuláris jelenségeket amerikai, norvég és svéd kollégák vastagbél tumorról kapcsolatos új epidemiológiai adatai értékelésének segítésére is felhasználtuk.

## **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 (<http://compbi.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. 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. A transzlokációval kapcsolatos molekuláris adatok és hipotézisünk segítette az amerikai, norvég és svéd kutatók vastagbél tumorról kapcsolatos új epidemiológiai adatainak az értelmezését, amelyet egy kollaborációs közleményben foglaltuk össze. Munkánk szerint a megfigyelésnek klinikai jelentősége lehet a daganat megelőző szűrések, illetve a terápia tervezéskor, valamint az utánkövetés során.

## Saját publikációk jegyzéke

### A disszertáció témájához kapcsolódó közlemények

1. Adami HO, Csermely P, **Veres DV**, et al. Are rapidly growing cancers more lethal? *Eur. J. Cancer*. 2017;72:210-214. (IF: 6,163, Google Scholar idézettség: 0)
2. Csermely P, Hódsági J, Korcsmáros T, (**Veres DV**) et al. 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.* 2015;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. 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.* 2013;23(4):262-269. (IF: 9,143, Google Scholar idézettség: 16, Web of Science idézettség: 12)
4. **Veres DV**, Gyurkó DM, Thaler B, et al. ComPPI: A cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.* 2015;43(Database issue):D485-D493. (IF: 9,202, Google Scholar idézettség: 21, Web of Science idézettség: 14)

### A disszertáció témájához nem kapcsolódó közlemények

1. Csermely P, Sandhu KS, Hazai E, (**Veres DV**) et al. Disordered Proteins and Network Disorder in Network Descriptions of Protein Structure, Dynamics and Function: Hypotheses and a Comprehensive Review. *Curr. Protein Pept. Sci.* 2012;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. Network strategies to understand the aging process and help age-related drug design. *Genome Med.* 2009;1(9):90. (IF: 0,0, Google Scholar idézettség: 32, Web of Science idézettség: 22)
3. Pákó J, **Veres D**, Tisza J, Horváth I. A COPD a multimorbiditás tükrében. *Orvosi Továbbképző Szemle*. XXII. évf. 11. szám 2015.