

A jelátviteli hálózatok regulációja és tumorokban való vizsgálata

Doktori értekezés

Dr. Módos Dezső

Semmelweis Egyetem
Patológiai tudományok Doktori Iskola



- Témavezető: Dr. Földvári-Nagy Lászlóné Dr. Lenti Katalin, PhD, főiskolai tanár
- Konzulens: Dr. Korcsmáros Tamás, PhD, tudományos főmunkatárs
- Hivatalos bírálók: Dr. Kapuy Orsolya, PhD, egyetemi adjunktus
Dr. Egyed Balázs, PhD, egyetemi adjunktus
- Szigorlati bizottság elnöke: Dr. Tóth Sára, PhD, habilitált egyetemi docens
- Szigorlati bizottság tagjai: Dr. Sipeki Szabolcs, PhD, egyetemi adjunktus
Dr. Hegedűs Tamás, PhD, tudományos főmunkatárs

Budapest
2017

„hogy kaphatnék kapcsolatot, összefüggést az élet elem kerülő apró-cseprő dolgai közt –
hogyan akasszam össze egyik jelenséget a másikkal – a viszonylagost az elmúltot hogy
hozzam vonatkozásba a nem viszonylagossal és– a részt, hogy kössem össze az
egésszel?”

Karinthy Frigyes: Láncszemek
1929

Csillának

Tartalomjegyzék

I.	Rövidítések jegyzéke	5
II.	Bevezetés	8
1.	<i>A sejten belüli jelátviteli útvonalak és daganatokon belüli szerepük</i>	8
1.1.	A MAPK jelátviteli útvonal	9
1.2.	A Notch jelátviteli útvonal	11
1.3.	A WNT/wingless jelátviteli útvonal.....	12
1.4.	A Hedgehog jelátviteli útvonal	13
1.5.	A TGF- β jelátviteli útvonal	14
1.6.	A JAK/STAT jelátviteli útvonal	14
1.7.	A nukleáris hormon receptor jelátviteli útvonal.....	15
2.	<i>A makroautofágia molekuláris mechanizmusa</i>	15
3.	<i>Az oxidációs stressz válasz és az NRF2 transzkripció faktor</i>	19
4.	<i>A jelátviteli útvonalak különböző definíciója az egyes adatbázisokban</i>	19
5.	<i>Fehérje-fehérje interakciós adatbázisok</i>	22
6.	<i>Az autofágiát vizsgáló adatbázisok</i>	26
7.	<i>Regulációs adatbázisok</i>	26
7.1.	Transzkripció faktor célgén adatok meghatározása és adatbázisai.....	26
7.2.	A miRNS reguláció és adatbázisai	30
7.3.	Transzkripció faktor-miRNS adatbázisok	35
8.	<i>Az adatbázisokból a rendszerbiológiai modell felé: hálózatok</i>	36
8.1.	<i>Az esszencialitás hálózat alapú meghatározása</i>	39
9.	<i>A hálózatok szerepe a daganatok vizsgálatában</i>	40
III.	Célkitűzések.....	43
IV.	Módszerek.....	44
1.	<i>Transzkripció faktor pozíció súlyozási mátrix alapján történő DNS kötés jóslása</i> 44	
2.	<i>Használt adatbázisok</i>	46
2.1.	miRNS célpont adatbázisok:	46
2.2.	Transzkripció faktor célgén adatbázisok	47
2.3.	Transzkripció faktor miRNS adatbázisok.....	47
2.4.	Betegség specifikus adatbázisok	47
2.5.	Gyógyszerészeti fontosság és molekuláris célpontok meghatározására használt adatbázisok	48
2.6.	Interakciós és jelátviteli adatbázisok.....	49
2.7.	Egyéb adatbázisok.....	49
3.	<i>Paralóg csoportok meghatározása</i>	53
4.	<i>Szövetspecifitás a kritikus paralóg csoportokban</i>	54
5.	<i>Paralóg csoportok tulajdonságainak specifikitási vizsgálata</i>	56
6.	<i>Microarray elemzés és a differenciáltan expresszálódó gének meghatározása az első szomszédok vizsgálatakor</i>	56
7.	<i>Használt programok</i>	57
8.	<i>Statisztikai módszerek</i>	58
V.	Eredmények és az eredmények értékelése.....	59
1.	<i>Az Autofágia Regulációs Adatbázis az NRF-ome és a SignaLink jelátviteli adatbázis regulációs rétegei</i>	59

1.1.	A transzkripció faktor és célgén adatok integrálása:	59
1.2.	A miRNS adatok integrálása	59
2.	<i>A kritikus paralógok meghatározása a jelátvitelben</i>	61
2.1.	A kritikus paralógok centrálisak az emberi jelátvitelben	64
2.2.	Specifikus reguláció és diverz funkció a kritikus fehérjéken belül.....	67
2.3.	A kritikus paralóg csoportok betegséget okozó hatása és gyógyszerészeti relevanciája.....	68
3.	<i>Az első szomszédok szerepe a daganatok jelátviteli hálózatában</i>	72
3.1.	Az első szomszédok központi szerepet töltenek be a daganatok interakciós hálózatában	72
3.2.	A daganatban mutálódó fehérjék közvetlenül, a daganatban differenciáltan expresszálódó fehérjék közvetetten huzalozzák át a jelátviteli hálózatot	88
3.3.	Lehetséges új utak a gyógyszerfejlesztésben az első szomszédok bevonásával.	98
VI.	Megbeszélés.....	102
1.	<i>A SignaLink, az ARN és az NRF2ome adatbázis lehetséges alkalmazásai</i>	102
2.	<i>Példák a kritikus paralóg csoportokra, szerepük a jelátvitelben</i>	104
1.1.	A NOTCH fehérjék:	105
1.2.	A SMAD 2 és 3 fehérje:.....	107
1.3.	A MAPK fehérjék szövet specifikus expressziója:	108
3.	<i>A kritikus paralóg csoportok jelentősége</i>	110
4.	<i>A kritikus paralóg csoportok meghatározásának limitációi</i>	113
5.	<i>Példák az első szomszédok gyógyszerészeti alkalmazására.</i>	113
6.	<i>A daganatban érintett fehérjék első szomszédjainak a jelentősége</i>	122
7.	<i>A daganatban érintett fehérjék első szomszédjainak meghatározásának limitációi</i> 124	
VII.	Következtetések	125
VIII.	Összefoglalás	127
IX.	Summary	128
X.	Irodalomjegyzék	129
XI.	Saját publikációk jegyzéke	180
1.	<i>A dolgozathoz kapcsolódó publikációk</i>	180
2.	<i>Nem kapcsolódó publikációk</i>	181
XII.	Köszönetnyilvánítás.....	182
XIII.	Ábrák táblázatok és egyenletek jegyzéke	183
1.	<i>Ábrajegyzék</i>	183
2.	<i>Táblázatjegyzék</i>	186
3.	<i>Egyenletjegyzék</i>	187

I. Rövidítések jegyzéke

ADB	<i>Autophagy Database</i>	cMyc	Myelocytomatosis virális onkogén homológja
Akt	vAkt homológ	CNC	<i>Cap and collar</i> motívum
AMP	Adenozin-monofoszfát	CREBP	cAMP-reszponzív elem kötő fehérje
AMPK	AMP regulált kináz	DAG	Diacilglicerol
APC	<i>Adenomatous Polyposis Coli</i>	DE	Daganatban differenciáltan exprezálódó (gén /fehérje)
ATG5	<i>Autophagy-related gene 5</i>	DÉ	Daganatban érintett fehérje
ATG7	<i>Autophagy-related gene 7</i>	DIP	<i>Database of Interacting Proteins</i>
ATG10	<i>Autophagy-related gene 10</i>	DNS	Dezoxiribonukleinsav
ATG12	<i>Autophagy-related gene 12</i>	Drosha	<i>drosha ribonuclease III</i>
ATG13	<i>Autophagy-related gene 13</i>	DUSP	<i>Dual specificity phosphatase</i>
ATG16	<i>Autophagy-related gene 16</i>	DSH	<i>Dishevelled</i>
BioGRID	<i>BIOlogical General Repository for Interaction Datasets</i>	EGF	Epidermális növekedési faktor
BioPlex	<i>biophysical interactions of ORFeome-based complexes</i>	EGFR	Epidermális növekedési faktor receptor
BCL2	B-cell lymphoma 2	ERK	<i>Extracellular Signal-Regulated Kinase</i>
BH3	BCL2 homológ domén tartalmazó fehérje	ENCODE	<i>Encyclopedia of DNA Elements</i>
BIF1	<i>Endophilin B1</i>	FDR	<i>False discovery rate</i> hamis találati arány
cAMP	ciklikus adenozin-monofoszfát	FGF	Fibroblaszt serkentő növekedési faktor
CBP	<i>CREB binding protein</i>	FGFR	Fibroblaszt serkentő növekedési faktor receptor
cFos	Fos protoonkogén	FIP200	200 kiloDalton tömegű fokális adhézis kináz interakciós fehérje
cJun	Jun protoonkogén		
CGC	<i>Cancer Gene Census</i>		
ChIP-Seq	Kromatin immunprecipitációs szekvenálás		
ChiP-chip	DNS microarray elemzés kromatin immunprecipitáció után		

FRAT1	<i>Frequently Rearranged in Advanced T-cell lymphoma 1</i>	LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
FZD	<i>Frizzled</i>	LC3	<i>Microtubule-associated protein 1A/1B-light chain 3</i>
GDP	Guanozil difoszfát	LEF1	<i>Lymphoid enhancer-binding factor 1</i>
GEF	<i>Guanozil exchange factor-</i>	LFNG	<i>Lunatic fringe</i>
Gli	Glioma asszociált fehérje	MAF	Musculoaponeurotic fibrosarcoma protoonkogén
GTP	Guanozil trifoszfát	MAP2K	Mitogén aktivált protein kináz-kináz
GSK3 β	<i>Glycogen Synthase Kinase-3 beta</i>	MAP3K	Mitogén aktivált protin kináz-kináz-kináz
HADb	<i>Human Autophagy Database</i>	MAPK	Mitogén aktivált protein kináz
HCC	Hepatocelluláris carcinoma	MDK	<i>Mideikine</i>
HER2	<i>Human Epidermal growth factor Receptor 2</i>	MEK	<i>Meiotic Chromosome-axis-Associated Kinase</i>
Hh	<i>Hedgehog</i> jelátviteli útvonal vagy fehérje	MFNG	<i>Maniac fringe</i>
HOCTAR	<i>Hostgene Oppositely Correlated TARgets</i>	miRNS	mikro RNS
IL7R	Interleukin 7 receptor	MMTV	<i>Mouse Mammary Tumor Virus</i>
IFN γ R	Interferon gamma receptor	mRNS	<i>messenger RNS</i>
IntAct	<i>Molecular Interaction Database</i>	mTOR	<i>Mammalian Target of Rapamycin</i>
JAK	<i>Janus Activated Kinase</i>	NÉ	Daganatban nem érintett gén/fehérje
KEAP1	<i>Kelch-Like epichlorohydrin-associated protein 1</i>	NHR	Nucleáris hormon receptor
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>	NOTCH	<i>Neurogenic locus notch homolog protein</i>
KFERQ	lizin-fenilalanin-glutaminsav-arginin-glutamin aminosav motívum	NRF2	<i>Nuclear Factor Erythroid 2-Related Factor 2</i>
KP	Kritikus paralóg	NGS	új generációs szekvenálás
KPCS	Kritikus paralóg csoport	NSCLC	nem-kissejtes tüdő rák
LAMP2	Lizoszóma asszociált fehérje 2		

ORF	<i>Open Reading Frame</i>	SMO	<i>Smohend</i>
PDGF	<i>Platelet-Derived Growth Factor</i>	SOCS3	<i>Suppressor Of Cytokine Signalling 3</i>
PDGFR	<i>Platelet-Derived Growth Factor Receptor</i>	SOS	<i>Son of Sevenless</i>
PF	Paralóg fehérje	STAT	<i>Signal Transducer and Activator of Transcription</i>
PicTar	<i>Probabilistic Identification of Combinations of Target Sites</i>	STRING	<i>Search Tool for the Retrieval of Interacting Genes/Proteins</i>
PPI	Fehérje-fehérje interakció	SOCS	<i>Suppressor of Cytokine Signaling</i>
PTEN	<i>Phosphatase and Tensin Homolog</i>	TaLasso	<i>miRNA-Target LASSO</i>
RAF	<i>Rapidly Accelerated Fibrosarcoma</i>	TF	Transzkripció faktor
RAB7	<i>Ras-related protein 7</i>	TCGA	<i>The Cancer Genome Atlas Consortium</i>
Ras	<i>Rat Sarcoma</i>	ULK	<i>Unc-51 like autophagy activating kinase</i>
RMA	<i>Robust Multiarray Average</i>	UV	Ultra viola
RNS	Ribonukleinsav	UVRAG	<i>UV radiation resistance-associated gene protein</i>
ROC	<i>Receiever Opereating Characteristic</i>	VPS	<i>Vacuolar protein sorting</i>
SELEX	<i>Systematic Evolution of Ligands by Exponential Enrichment</i>	WNT	<i>Wingless-type MMTV integration site family member</i>
SH2	<i>Src Homology 2</i>	XPO5	Exportin 5
SMAD	Sma (<i>small body size</i>) és Mad (<i>mothers against decapentaplegic</i>) összevonása		

II. Bevezetés

Az emberi szervezet bonyolult működéséhez rengeteg folyamatnak kell megfelelő szabályozottsággal végbemennie. Egymással kommunikáló rendszerként fonódnak össze a szervek, szövetek, és az élővilág legkisebb önállóan működőképes egysége, a sejt is.

Egy sejt sorsa számtalan bejövő jeltől, ingertől függ, és millióféleképpen válaszolhat pillanatnyi állapotától függően a bejövő jelekre. Ezeknek a megfelelő jeleknek a feldolgozását végzi a jelátviteli rendszer. Egy ilyen bonyolult rendszer megértéséhez szükségünk van arra, hogy ne csak az egyes biokémiai reakciókat figyeljük meg, hanem a rendszer összességét, a bejövő jeleket feldolgozó rendszert is. Ennek a megismeréséhez először is olyan gyűjteményekre, adatbázisokra van szükség, melyek leírják a jelátvitelt, és azok transzkripcionális regulációs visszacsatolásait is tartalmazzák.

Az így felállított adatbázisokból elkészíthető a jelátvitel interakciós modellje. Ennek egyik legegyszerűbb reprezentációja az egyes fehérjéket és a köztük található interakciókat tartalmazó fehérje interakciós hálózat.

A jelátvitel és ezzel együtt a sejt, mint rendszer megértéséhez és beavatkozások tervezéséhez meg kell találnunk a rendszer kulcsfontosságú, kritikus szabályozási pontjait. Ezek lehetnek azon helyek, melyek leginkább érintettek a betegségek, különösen a daganatos betegségek kialakulásakor. Ha meghatároztuk a jelátvitelben történő változást a különböző kórfolyamatokban, akkor elemezhetjük a változások hatásait, megtudva, hogy ezek a jelátviteli hálózat mely területén mennek végbe. Ennek ismeretében képesek lehetünk a betegségek esetén a betegségekhez hasonló módon felülírni a jelátvitelt, de ellenkező előjellel, ezzel gyógyítva a rendszerszintű betegségeket, mint amilyen a rák.

1. A sejten belüli jelátviteli útvonalak és daganatokon belüli szerepük

A sejt információtovábbító és feldolgozó rendszerét jelátvitelnek nevezzük. Ez a rendszer felelős a sejtek döntési mechanizmusaiért. A sejt a rá ható jeleket analizálva képes eldönteni, hogy osztódjon, differenciálódjon, vagy esetleg elpusztuljon valamely sejthalál útvonal aktiválása után (Lim és mtsai, 2015).

A jelátvitelben a sejtet ért hatásokat különböző receptor fehérjék érzékelik. A receptorhoz köthet fehérje természetű közvetítő molekula (ligand fehérje), de

kötődhetnek különböző, nem fehérje természetű kismolekulák is, mint például egyes neurotranszmitterek (acetilkolin, glutamin, adrenalin stb.).

A receptorok az extracelluláris ligand (ezáltal a jel) érzékelése után konformációs átalakuláson mennek keresztül. A konformáció változás után a receptorok intracelluláris irányba továbbítják a jelet. A jel G fehérjéken keresztül haladhat tovább különböző intracelluláris mediátorokra (Audet és Bouvier, 2012), mint a diacilglicerol (DAG) vagy a ciklikus adenzin-monofoszfát (cAMP). Egy másik lehetőség, hogy a receptorok kinázként foszforilálnak különböző fehérjéket (Lemmon és Schlessinger, 2010).

A bejövő jel ezután különböző mediátorokra kerül, ahol felerősödhet. Ilyen mediátorok lehetnek az említett kismolekulák, vagy a mitogén aktivált protein kináz (MAPK) kaszkád esetén az útvonalakon belüli kinázok (Raman és mtsai, 2007; Seger és Krebs, 1995; ZHANG és LIU, 2002).

A bemeneti jel pontos sorsát a jelátviteli kofaktorai szabályozzák. Ezek a fehérjék a jelek terjedését befolyásolják, de a jel terjedéséhez nem feltétlenül szükségesek (Korcsmáros és mtsai, 2010). Közéjük tartoznak például az állvány fehérjék vagy a fehérje lebontásért felelős E3 típusú ubiquitin ligázok¹.

A transzkripciós faktorokat tekintjük a jelátviteli útvonalak végső elemeinek. A transzkripciós faktorok biztosítják a jelek mRNS, majd a transzláció utána a fehérje szintű megjelenését, és ezzel a fenotípusban bekövetkező változást (Zhou, 2012).

A jelátviteli rendszer alapvető egysége a jelátviteli út. Ez a jelet (ligand) érzékelő receptortól a transzkripciós faktorig tartó kaszkád. Ezen utak alkotnak együttesen egy-egy jelátviteli útvonalat (Lim és mtsai, 2015).

Az alábbiakban a dagantokban és a differenciációban legfontosabb útvonalak áttekintése következik (Pires-daSilva és Sommer, 2003).

1.1. A MAPK jelátviteli útvonal

A MAPK jelátviteli útvonala az egyik legjobban felderített jelátviteli útvonal. Kiindulási pontjai a sejtmembránt egyszer átszelő receptor tirozin kinázok, melyekhez különböző ligandok kapcsolódhatnak. Az útvonalat növekedési faktor útvonalnak is nevezik, mivel leggyakrabban a különböző növekedési faktorok a receptor tirozin kinázok ligandjai (epidermális növekedési faktor (EGF) és receptora (EGFR), fibroblaszt serkentő

¹ Az ubiquitin ligázok olyan fehérjék, amelyek ubiquitinnel jelölik meg az egyes fehérjéket és ezzel a proteaszómába irányítják őket. Az ubiquitiniláció folyamatában három enzim vesz részt: E1, E2, E3. Közülük az E1 és az E2 a végrehajtó enzim az E3 pedig a specifikus dokkolást biztosítja.

növekedési faktor (FGF) és receptora (FGFR), platelet-derived növekedési faktor (PDGF) és receptora (PDGFR) stb.). Az útvonal végső fenotipikus eredménye sokszor a sejtosztódás. (Jura és mtsai, 2011; Lemmon és Schlessinger, 2010; Yarden és Ullrich, 1988).

A receptorok dimerizációt követően autofoszforilálják egymást, és képessé válnak különböző SH2 (*Src Homology 2*) doménnel² rendelkező fehérjéket megkötni. Ezek állványfehérjeként szolgálhatnak GTP (guanozil-trifoszfát) kicserélő fehérjéknek (GEF), mint amilyen a SOS (*Son of Sevenless*). Ezek a GEF-ek kicserélhetik a kis GTP kötő fehérjék GDP-jét (guanozil-difoszfát) GTP-re ezzel továbbadva a bejövő jelet. Ilyen GTP fehérjék például a különböző Ras fehérjék, melyek képesek aktiválni a MAP kináz kaszkádot. Itt egymást foszforiláló, egymáshoz hasonló szerkezetű kinázok helyezkednek el. Ez a rendszer amplifikációs és zaj szűrő szereppel bír (Alon, 2007; Ladbury és Arold, 2012). A MAPK jelátvitelt paralóg³ specifikus foszfatázok, állvány fehérjék, feedback körök és szövet specifikus expresszió regulálja (Kolch, 2005; Kondoh és Nishida, 2007; Raman és mtsai, 2007; Seger és Krebs, 1995) Ez a rendszer amplifikációs és zaj szűrő szereppel bír (Alon, 2007; Ladbury és Arold, 2012). A MAPK jelátvitelt paralóg⁴ specifikus foszfatázok, állvány fehérjék, feedback körök és szövet specifikus expresszió regulálja (Kolch, 2005; Kondoh és Nishida, 2007; Raman és mtsai, 2007; Seger és Krebs, 1995). A kinázok egymás után három rétegben helyezkednek el (Qi és Elion, 2005). A Ras fehérjék után a MAP3K típusú kinázok következnek, mint a RAF (*Rapidly Accelerated Fibrosarcoma*). A MAP3K a mitogén aktivált protein kináz-kinázokat aktiválja (MAP2K). Ezek közé tartoznak például a különböző MEK (*meiotic chromosome-axis-associated kinase*) fehérjék. Végül az ERK (extracelluláris szignál által regulált kináz) típusú kinázok következnek, mint például a p38 fehérjék (Raingeaud és mtsai, 1996), amik már maguk a MAP kinázok. A tényleges fenotipikus változást a MAP kinázok célpontjai folytatják, mint a cFos (Fos protoonkogén), cJun (Jun protoonkogén), és a cMyc (myelocytomatosis virális onkogén homológja) transzkripció faktorok.

² Az SH2 (*Src Homology 2*) domén egy konzervált fehérje domén, amely lehetővé teszi a domént tartalmazó fehérjéknek más fehérjék foszforilált tirozin oldalláncához való kapcsolódását.

³ Azonos faj genomjában található egymáshoz teljes szekvenciájában nagyban hasonló, közös evolúciós eredetű gének

⁴ Azonos faj genomjában található egymáshoz teljes szekvenciájában nagyban hasonló, közös evolúciós eredetű gének

A különböző rétegekben a jelátvitelbe lehetősége van más útvonalaknak is beleszólni, keresztbeszélgetést, cross-talkot folytatni. Például a MEK6 cross-talkot folytat a differenciációban jelentős TGF- β (*transforming growth factor*) útvonallal, ahonnan stressz szingálok érik el a MAPK jelátvitelt (Kim és mtsai, 2004). A bejövő jelek egymással részben ellentétes sejtes folyamatra fejtik ki hatásukat, például az ERK kinázok folyamatos aktiválódása vezethet sejthalálhoz, lökés szerű, rövid ideig tartó aktiválódása pedig osztódáshoz (Lu és Xu, 2006). A megfelelő egyensúly biztosítása érdekében a kinázok gátlását foszfatázok végzik, mint amilyen a dual-specific foszfatázok (DUSP) (Cerignoli és mtsai, 2006; Tambe és mtsai, 2016).

A sejtsztódásban betöltött szerepük miatt a daganatokban gyakoriak a mutációk az útvonalon belül, vagy a receptorok amplifikációja. Például tüdőrákban az EGFR 2. és 7. exonja közötti szakasz deléciója történik, ami a sejtmembránon kívüli domén deléciójához, és ezzel a receptor ligand nélküli aktiválódásához vezet (Hanahan és Weinberg, 2000). A mell adenocarcinómák 30%-ban a HER2 (*human epidermal growth factor receptor 2*) amplifikálódik. A HER2 amplifikáció a meghatározása a luminalis B típusú mell adenocarcinómáknak (Zhao és mtsai, 2015). Az összes daganat 25%-ban, mint például egyes tüdőrákokban a RAS fehérje mutálódik, és ezzel folyamatos aktivációs szignált szolgáltat (Medema és Bos, 1993).

1.2. A Notch jelátviteli útvonal

A Notch útvonalnak szerepe van a differenciáció, a szervfejlődés és az apoptózis szabályozásában egyaránt. Ez a jelátviteli útvonal azért különleges, mert a NOTCH fehérjék töltik be egyszerre a receptor, a mediátor és a transzkripciós faktor szerepet is. A delta és a jagged fehérjék által aktivált NOTCH receptor a hasítását követően a sejtmembránról levágódva a sejtmagba jut, és ott transzkripciós faktorként képes működni megfelelő kofaktorokhoz kapcsolódva (Bray, 2006; Guruharsha és mtsai, 2012). A Notch jelátvitel megfelelő működését a különböző jelet fogadó sejten belüli és sejten kívüli szabályozó faktorok befolyásolják NOTCH receptor specifikusan, ahol az egyes NOTCH receptorokra való hatás akár ellentétes is lehet. Ilyen sejten kívüli szabályozó faktor például az LFNG (lunatic fringe) és az MFNG (maniac fringe) fehérje. Az LFNG NOTCH1 specifikus gátlást, míg az MFNG inkább NOTCH2 specifikus gátlást folytat, ezzel serkenti a másik NOTCH receptort (Bray, 2006; Guruharsha és mtsai, 2012). A Notch útvonalnak szerepe van a differenciáció, a szervfejlődés és az apoptózis

szabályozásában egyaránt. Ez a jelátviteli útvonal azért különleges, mert a NOTCH fehérjék töltik be egyszerre a receptor, a mediátor és a transzkripciós faktor szerepet is. A delta és a jagged fehérjék által aktivált NOTCH receptor a hasítását követően a sejtmembránról levágódva a sejtmagba jut, és ott transzkripciós faktorként képes működni megfelelő kofaktorokhoz kapcsolódva (Bray, 2006; Guruharsha és mtsai, 2012). A Notch jelátvitel megfelelő működését a különböző jelet fogadó sejten belüli és sejten kívüli szabályozó faktorok befolyásolják NOTCH receptor specifikusan, ahol az egyes NOTCH receptorokra való hatás akár ellentétes is lehet. Ilyen sejten kívüli szabályozó faktor például az LFNG (lunatic fringe) és az MFNG (maniac fringe) fehérje. Az LFNG NOTCH1 specifikus gátlást, míg az MFNG inkább NOTCH2 specifikus gátlást folytat, ezzel serkenti a másik NOTCH receptort (Bray, 2006; Guruharsha és mtsai, 2012).

A NOTCH fehérjék közül a NOTCH1 folytat cross-talkot más útvonalakkal. A Notch útvonalon kívül a szintén fejlődésbiológiai szerepű WNT (*Wingless-type MMTV integration site family member*) és a mezenchimális TGF- β útvonallal is kapcsolatban áll. (Bray, 2006; Guruharsha és mtsai, 2012). A Notch jelátvitelnek elsősorban fejlődésbiológiai szerepe van, de daganatoknál is leírták szerepét az epithelialis-mesenchimalis átalakulásban (Groth és Fortini, 2012). A NOTCH fehérjék közül a NOTCH1 folytat cross-talkot más útvonalakkal. A Notch útvonalon kívül a szintén fejlődésbiológiai szerepű WNT (*Wingless-type MMTV integration site family member*) és a mezenchimális TGF- β útvonallal is kapcsolatban áll. (Bray, 2006; Guruharsha és mtsai, 2012). A Notch jelátvitelnek elsősorban fejlődésbiológiai szerepe van, de daganatoknál is leírták szerepét az epithelialis-mesenchimalis átalakulásban (Groth és Fortini, 2012).

1.3. A WNT/wingless jelátviteli útvonal

A WNT jelátvitel a differenciáció egyik fontos útvonala. *Drosophilában* írták le, ahol a WNT fehérje homológjának kiesése a szárny nélküli fenotípushoz vezetett. A WNT jelátviteli útvonal működésének központi fehérjéje a β -katenin. A β -katenin sejtmagba jutása jelenti a jelátviteli útvonal aktiválódását. Ligand nélkül a β -katenin folyamatosan keletkezik és lebomlik a proteaszóma rendszeren keresztül. Az érkező jel ezt a lebomlást gátolja meg. A bejövő jel a különböző WNT ligandokon keresztül érkező *Frizzled* (FZD) receptorokra kerül. Ekkor a *Dishevelled* (DSH) fehérje magánál tartja az axin állvány fehérjét, és ezzel nem engedi a GSK3 β -t (glikogén-szintetáz-kináz-3-béta), hogy foszforilálja a β -katenint. Ezáltal a β -katenin a

citoplazmában marad, majd utána a sejtmagba kerülve transzkripciós aktivitást indít be azzal, hogy a TCF 1, 2 vagy 3 (*Transcription factor 1, 2, 3*) és LEF1 (*Lymphoid enhancer-binding factor 1*) transzkripciós faktorhoz kötődik. A jelátvitel működéséhez szükség van az APC (*Adenomatous Polyposis Coli*) fehérjére is, amely állványfehérjeként engedi a GSK3 β β -katenin foszforilációját az axinhoz hasonlóan, de a DSH-től függetlenül. Ezt az utat nevezik a kanonikus WNT jelátviteli útnak (Logan és mtsai, 2004)(Macdonald és mtsai, 2007). A WNT jelátviteli útvonal a LEF1-n keresztül cross-talkot folytat a Notch jelátviteli útvonallal. A β -katenin emellett részt vesz az aktin citoskeleton irányításában és a sejtdhézióban is (Logan és mtsai, 2004).

A jelátviteli útvonal másik ága az úgynevezett nem kanonikus β -katenin - független jelátviteli útvonal. Itt a FZD receptorokon keresztül egyéb útvonalakkal történik cross-talk. Ilyen például a WNT jelátviteli útvonal sejtdhézióban vagy a sejt polaritásban való részvétele (Niehrs, 2012; Semenov és mtsai, 2007).

A WNT jelátvitel megváltozása a daganatok közül leggyakrabban colorectalis carcinomákban fordul elő. Itt vagy az APC fehérje működése romlik el, ezzel gátolva a β -katenin lebomlását, vagy maga a β -katenin mutálódik úgy, hogy nem képes lebomlani. Mindegyik esetben az eredmény az útvonal túlaktiválódása, aminek következtében egy dedifferenciált fenotípus alakul ki (Gregorieff és Clevers, 2005).

1.4. A Hedgehog jelátviteli útvonal

A Hedgehog útvonalnak fejlődésbiológiai szerepe van: részt vesz a polaritás és a sejtsors meghatározásában, de részt vesz a sejtosztódás szabályozásában is. A Hedgehog jelátviteli útvonal működésekor a bejövő jel a Hedgehog (Hh) fehérje. Ez a fehérje kötődik a *Dispatched* receptorhoz. A receptorhoz való kötődés meggátolja a *Dispatch* fehérjét abban, hogy a SMO (*Smoothened*) fehérjét az endoszómában tartsa, így az kikerül a sejtfelszínre. Ekkor a SMO meggátolhatja a Gli (Glioma asszociált fehérje) transzkripciós faktor lebontását, így az bekerülhet a sejtmagba (Briscoe és Thérond, 2013; Riobo és mtsai, 2006). Az útvonal komplexitását a különféle Hedgehog ligandok mint a Sonic, Indian vagy Desert Hh fehérjék biztosítják. A Gli fehérjék lebontásához szükség van a GSK-3 β általi foszforilációra is (Mimeault és Batra, 2010). A Hedgehog jelátviteli útvonal mutációját a bazális-sejtes karcinoma kialakulásában írták le (Briscoe és Thérond, 2013).

1.5. A TGF- β jelátviteli útvonal

TGF- β útvonal a kötőszöveti differenciációban vesz részt és a középső csíralemez szerveinek és sejtjeinek a fejlődését befolyásolja. A TGF- β útvonal működése során a TGF- β receptorról a különböző SMAD (Sma (*small body size*) és Mad (*Mothers against decapentaplegic*) összevonása) mediátorokra kerül a jel. A SMAD-okat funkciójuk szerint csoportosítjuk. Az első csoport az R vagy receptor SMAD. Ezek az egymáshoz szekvenciájukban hasonló paralóg SMAD2 és 3, illetve a SMAD1, 5 és 9. Az R-SMAD-ok a jelet továbbítják a TGF- β receptorok felől a sejtmagba, ahol transzkripció faktor aktivitásuk van. A másik csoportba az I-SMAD-ok tartoznak. Ezeknek gátló funkciója van, ilyen a SMAD6 és 7. A harmadik csoportba a co-SMAD, SMAD4 tartozik. Ez az R-SMAD-okkal működik együtt kofaktorként (Lutz és Knaus, 2002).

Ahhoz, hogy a TGF- β útvonalon a receptorról a jel útvonal-specifikus expresszió-változásokat indukáljon, valamelyik R-SMAD csoport két tagjának (SMAD2 – SMAD3 vagy SMAD1 - SMAD5 - SMAD9) jelen kell lennie (Lutz és Knaus, 2002). Ha csak az egyik SMAD van jelen, akkor a bejövő TGF- β jel egy másik jelátviteli útvonalat befolyásol a domináns paralógon keresztül. Például a TGF- β aktiválhatja a WNT útvonalat a SMAD3-on keresztül, a SMAD3 leállíthatja az axin expressziót, amely a WNT jelátvitel egyik negatív regulátora, ahogy fentebb láttuk (Dao és mtsai, 2007). Ez a cross-talk fontos a chondrociták differenciációjában (Dao és mtsai, 2007). Egy másik cross-talk a SMAD3 és NOTCH1 közös transzkripcionális szabályozása a HES1 fehérjének, mely a neuronális érésben vesz részt (Blokzijl és mtsai, 2003). A TGF- β jelátvitel hiánya a tumorok malignitását fokozza, de részt vesz az epithelialis-mesenchimalis átalakulásban is (Samanta és Datta, 2012).

1.6. A JAK/STAT jelátviteli útvonal

A JAK (Janus kináz) STAT (*Signal Transducer and Activator of Transcription*) jelátvitel elsősorban az immunfolyamatokban és a citokin jelátvitelben vesz részt. Alapját képezik a különböző citokin receptorok, melyek állvány fehérjeként magukhoz gyűjtik a JAK kinázokat. A receptor aktiválódást követően a JAK kinázok foszforilálják a receptorokat, és ezt követően a foszforilált receptorokhoz SH2 doménen keresztül kötődnek a STAT fehérjék. A STAT fehérjék a sejtmagba jutnak, így befolyásolják a génextpressziót, és ezen keresztül a különféle sejten belüli folyamatokat (Koppikar és mtsai, 2012; Pencik és mtsai, 2016).

A jelátviteli útvonal komplexitását az adja, hogy a citokin receptorok száma nagy, és a választ vivő 8 különböző STAT fehérje egymással kombinálódva különféle választ generálnak. A jelátvitelt a SOCS (*Suppressor of cytokine signaling*) fehérjék képesek gátolni, amelyek a JAK fehérjékhez kapcsolódnak, ezzel megakadályozzák a STAT fehérjék kötődését, és ezzel a jelátvitel elindulását (Dutta és Li, 2013; Rawlings és mtsai, 2004). A jelátvitelt a SOCS (*Suppressor of cytokine signaling*) fehérjék képesek gátolni, amelyek a JAK fehérjékhez kapcsolódnak, ezzel megakadályozzák a STAT fehérjék kötődését, és ezzel a jelátvitel elindulását (Dutta és Li, 2013; Rawlings és mtsai, 2004). A JAK/STAT jelátvitel a tumor ellenes immunitásban vesz részt, például az IFN- γ jelet továbbítja (Zaidi és Merlino, 2011).

1.7. A nukleáris hormon receptor jelátviteli útvonal

A nukleáris hormon receptor (NHR) útvonal az egyik legkevésbé komplex sejten belüli útvonal. Ide tartoznak a különböző szteroid receptorok, mint az ösztrogén receptor, vagy a glükokortikoid receptorok, illetve egyéb, a sejtmembránon áthaladni képes molekulák receptorai, mint a tiroxin receptor és a retinsav receptor. Az útvonal működése során a transzkripció faktoraként is szolgáló citoplazmatikus receptorhoz kötődik a nem fehérje természetű, sejtmembránon átjutni képes ligand. Ezt követően a keletkezett komplex bejut a sejtmagba, és ott transzkripcionális aktivitást vált ki (Huang és mtsai, 2010). Az útvonal szabályozásában szerepet játszanak a bejutó ligandot elbontó enzimek, mint például a 11- β hidroxiszteroid-dehidrogenáz a vesében (Tomlinson és mtsai, 2004).

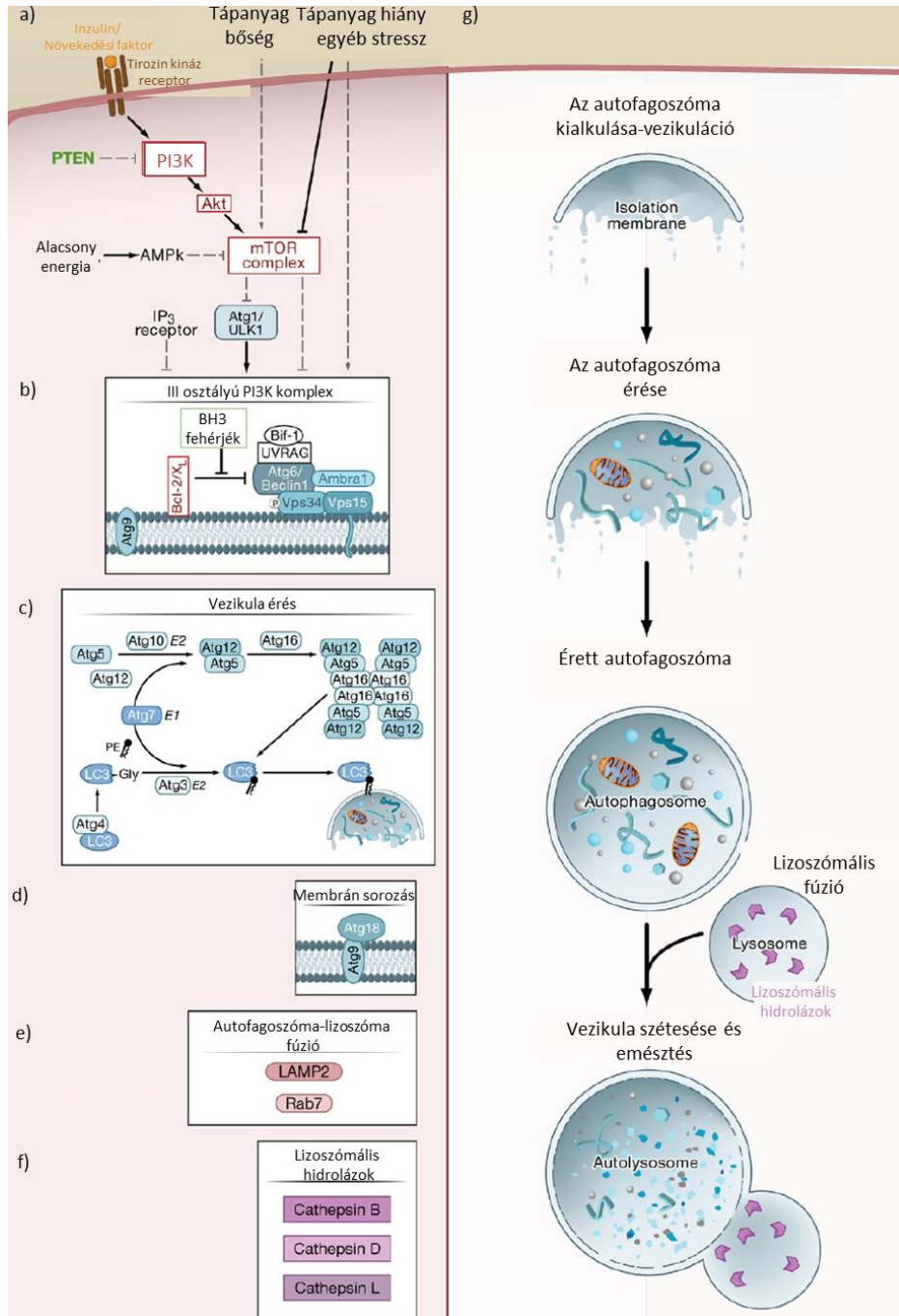
Az NHR útvonal képviselői a nemre specifikus daganatokban fontos, terápiát befolyásoló biomarkerek. Például az ER+ mell daganat a luminalis A típusa a mell daganatoknak, és ez a típus reszponzív a hormon terápiára (Koboldt és mtsai, 2012). Az androgén receptor pozitívitas a prosztata daganatok típusainak elkülönítésére szolgál (Wen és mtsai, 2014).

2. A makroautofágia molekuláris mechanizmusa

Az autofágia a sejtek önmérsztési folyamata, szerepe a károsodott vagy süükségtelen organellumok és fehérjék lebontása. Az autofágia jelentős végkimenetele és szabályozója a jelátvitelnek.

Mechanizmus alapján megkülönböztetünk mikro- makro- és chaperon mediált autofágiát. A mikroautofágia során kis citoplazma részek jutnak közvetlenül a

lizoszómákba membrán-betűródések segítségével. A makroautofágia során citoplazma részek vagy organelumok kerülnek kettős membránnal határolt vezikulákba. (Glick és mtsai, 2010). A chaperon mediált autofágiában chaperon fehérjék segítségével kerülnek be a lizoszómákba a KFERQ (lizin-fenilalanin-glutaminsav-arginin-glutamin) aminosav-szekvenciát tartalmazó, nem megfelelően feltekeredett fehérjék.



1. ábra Az autofágia folyamata emlősökben az egyes résztvevő fehérjékkel.

a) Az autofágia iniciáció szabályozásának áttekintése, b) Az iniciáció folyamata c) Az autofagoszómák érése d) Az autofagoszómák membrán növekedésében részt vevő fehérjék e) Az autofagoszóma lizoszóma fúzióban részt vevő fehérjék, f) A lizoszómális hidrolázok g) A sejten belül az autofág vezikulák kialakulásának a folyamata. (Levine és Kroemer, 2008a) Részleteket lásd a szövegben. Rövidítések: Akt: Akt fehérje (virális Akt fehérje homológ), ATG: *Autophagy related protein*, BCL2: B-sejtes limfóma, BH3: BCL2 homológ domén tartalmazó fehérje, BIF1: *Endophilin B1*, LAMP2: Lizoszóma asszociált fehérje 2, LC3: *Microtubule-associated protein 1A/1B-light chain 3*, mTOR: *mammalian target of rapamycin*, PI3K Foszfinozitol 3-kináz, PTEN *Phosphatase and tensin homolog*, ULK *Unc-51 like autophagy activating kinase*, RAB7 *Ras-related protein 7*, UVRAG UV radiation resistance-associated gene protein, VPS vacuolar protein sorting

A makroautofágia működését tekintem át részletesen. A folyamatot különböző jelátviteli utak szabályozzák, mint a receptor tirozin kináz jelátviteli útvonalba tartozó inzulin jelátvitel. A bejövő szabályozó jelek az mTOR-kompleyre (*mammalian target of rapamycin*) és az AMPK (AMP regulált kináz) kinázra jutnak (1. ábra/a). Az AMPK és az mTOR a sejtek tápanyagellátottságát és energiaszintjét érzékeli. Az AMPK az AMP szint megemelkedésekor lép működésbe, és gátolja az mTOR-t. Az mTOR alapvetően aktív, és az ULK 1/2 (*unc-51 like autophagy activating kinase 1/2*) fehérjék foszforilálásával gátolja az autofágiát. Ha az aminosavsztint leesik vagy extracelluláris szignál érkezik az mTOR irányába, akkor az, gátlás alá kerül, és ilyenkor az ULK 1/2 foszforilációja megszűnik. A defoszforilált ULK 1/2 aktív és kinázként foszforilálja az ULK 1/2 autofágia indukciós-komplex tagjait (ATG13 (*Autophagy-related protein 13*), ATG101 (*Autophagy-related protein 101*) és a FIP200 (200 kiloDalton tömegű fokális adhéziós kináz interakciós fehérje) (Aladzity és mtsai, 2007). Az ULK 1/2-komplex végül a Beclin-1 foszforilációjával serkenti az autofágiát. A Beclin-1 komplexet képez a VPS34, a VPS15, az UVARG és a Bif-1 fehérjével (Levine és Kroemer, 2008a, 2008b; Takahashi és mtsai, 2007). A komplex foszfatidil-inozitol 3 foszfátzként működik. Az autofagoszóma képződésének helyén levő membránokban a foszfatidil-inozitol 3 foszfát fontos jel lesz az autofág vezikula kialakulásában (1. ábra/b).

Az autofágia iniciációs komplex kialakulása után megkezdődik a sejt különböző részeiből a membrán darabok odaszállítása. Ezt két ubiquitinszerű konjugációs rendszer segíti (1. ábra/c). Az első esetben az ATG7 (*Autophagy-related protein 7*) E1 típusú ubiquitin aktiváló enzim aktiválja az ATG12 (*Autophagy-related protein 12*) fehérjét. Ezután az ATG12 az ATG10 (*Autophagy-related protein 10*) fehérje, mint E2 típusú ubiquitin konjugáló enzim közreműködésével az ATG5 (*Autophagy-related protein 5*) fehérjéhez kapcsolódik kovalens kötéssel. Ehhez a komplexhez kapcsolódik nem kovalens kötéssel az ATG16 (*Autophagy-related protein 16*) is (Levine és Kroemer, 2008b; Mizushima és Klionsky, 2007).

A másik ubiquitinszerű konjugációs komplex az LC3 (*Microtubule-associated protein 1A/1B-light chain 3*) fehérjét érinti (1. ábra/c). Itt első lépésben az LC3 fehérje proteolízise történik. A következőkben az ATG7, az ATG3 (*Autophagy-related protein 3*), mint E2 típusú ubiquitin konjugáló enzim, valamint az ATG12-ATG5-ATG16 komplex, mint E3 típusú ubiquitin ligáz enzim segítségével az LC3 utolsó glicin aminosavára egy foszfatidiletanolamin kerül. Ennek a célja az, hogy az LC3 a formálódó autofagoszóma membránhoz kapcsolódjon és ezzel segítse az autofág vezikula kialakulását (Mizushima és Klionsky, 2007; Tanida és mtsai, 2008).

A keletkezett komplexhez az ATG9 és ATG16 fehérjék segítségével szállítódnak a membrán darabok a sejt különböző területeiről az emlősökben (Levine és Kroemer, 2008a). Ez a vezikula növekedés folyamata (*membrane recruitment*) (1. ábra/d). A bezáródott autofág vezikula ezután lizoszómával fúzionál. Ebben a LAMP2 (Lizoszóma asszociált fehérje 2) és a RAB7 (*Ras-related protein 7*) fehérje vesz rész (1. ábra/e). Az autofagoszóma és a lizoszóma fúziója után kialakuló vezikulát autolizoszómának nevezzük. Ebben megtörténik a lizoszómális hidrolázok, pl. a különböző katepszinek, által az autolizoszóma beltartalmának a lebontása (Levine és Kroemer, 2008b; Mizushima és Klionsky, 2007) (1. ábra/f).

Az autofágia szabályozásáról még keveset tudunk. Központi eleme az mTOR komplex és működésének szabályozása, melynek egyik fontos része a sejt energia állapotának érzékelése az AMPK és az inzulin jelátviteli út által (Sarbasov és mtsai, 2005).

3. Az oxidációs stressz válasz és az NRF2 transzkripció faktor

Az NRF2 (*nuclear factor erythroid 2-related factor 2*) transzkripció faktor az oxidatív és xenobiotikum stressz érzékelésében és továbbításában játszik szerepet (Moi és mtsai, 1994a). Maga az NRF2 egy CNC (*cap and collar* - sapka és gallér) motívum alapú leucin cipzár transzkripció faktor (Moi és mtsai, 1994b). Heterodimereket képezve kötődik a DNS-hez a MAF (Musculoaponeurotic Fibrosarcoma Protoonkogén) fehérjékkel. Hat különböző doménje lehetővé teszi, hogy sok, eltérő fehérjével alkosson kapcsolatot. Ezek közül a legfontosabbak az NRF2 negatív regulátora, a KEAP1 (*Kelch-like epichlorohydrin-associated protein 1*), a helikáz és kromodomén funkcióval bíró CHD6 (*Chromodomain Helicase DNA Binding Protein 6*) és a CBP (*CREB(cAMP-reszponzív elem kötő fehérje)-binding protein*) (Itoh és mtsai, 1999; Motohashi és Yamamoto, 2004; Zhang, 2006). A sejtben normál esetben az NRF2 gátolva van, mert kapcsolódik a KEAP1-hez. A beérkező oxidációs stressz esetén az NRF2 felszabadul a KEAP1 gátlása alól, és bekerül a sejtmagba, ahol aktiválja az antioxidáns reszponzív elemeket tartalmazó géneket (Motohashi és Yamamoto, 2004). Több daganatban és egyéb betegségben fontos jelátviteli útvonallal kapcsolatban áll. Például a p53 által regulált p21 képes fokozni az NRF2 antioxidáns jelet, és ezzel a daganatok túlélési esélyét növelni (Taguchi és mtsai, 2011). Ugyanígy a tirozin kináz jelátviteli útvonalba tartozó MAPK és PI3K is képes aktiválni az NRF2 stressz választ (Surh és mtsai, 2008). A KEAP1 mutációját leírták többek között tüdő adenocarcinomákban és húgyhólyagrákban (Padmanabhan és mtsai, 2006; Shibata és mtsai, 2008a; Takahashi és mtsai, 2010). A megfigyelt mutációk közös tulajdonsága, hogy a KEAP1 NRF2 kötésének gátlásához vezetnek. Ez a folyamat pedig az NRF2 aktiválásához vezet, melynek végeredménye az antioxidációs válasz szintjének emelkedése. Magának az NRF2-nek a mutációját is leírták tüdő daganatokban és fej nyaki rákokban a KEAP1 kötő doménben (Shibata és mtsai, 2008b).

4. A jelátviteli útvonalak különböző definíciója az egyes adatbázisokban

A jelátviteli útvonalakat sokféleképpen definiálhatjuk. Az egyes adatforrásokban nagymértékben eltérő jelátviteli útvonal definíciókkal találkozhatunk. A legelterjedtebb definíció egy-egy jelátviteli útvonalra a kiindulási receptorról, vagy jellegzetes fehérjéről való elnevezés. Például EGFR útvonal, ERK útvonal. Ezt használja például NetPath, vagy a Reactome adatbázis (Croft és mtsai, 2014; Hwang és mtsai, 2013). Ezekben az

adatbázisokban ezért megtalálhatóak például a különböző interleukinoknak megfelelő jelátviteli útvonalak. Ez a felsorolás félrevezető, mivel önkényesen meghatározott szempontokat jelöl ki, és szinte korlátlan jelátviteli útvonal létrehozásához vezet. A jelátviteli útvonalakat sokféleképpen definiálhatjuk. Az egyes adatforrásokban nagymértékben eltérő jelátviteli útvonal definíciókkal találkozhatunk. A legelterjedtebb definíció egy-egy jelátviteli útvonalra a kiindulási receptorról, vagy jellegzetes fehérjéről való elnevezés. Például EGFR útvonal, ERK útvonal. Ezt használja például NetPath, vagy a Reactome adatbázis (Croft és mtsai, 2014; Hwang és mtsai, 2013). Ezekben az adatbázisokban ezért megtalálhatóak például a különböző interleukinoknak megfelelő jelátviteli útvonalak. Ez a felsorolás félrevezető, mivel önkényesen meghatározott szempontokat jelöl ki, és szinte korlátlan jelátviteli útvonal létrehozásához vezet.

A másik jellegzetes jelátviteli útvonal definíció, hogy valamely betegségről, vagy folyamatban való részvételtől nevezik el. Erre példák a KEGG (*Kyoto Encyclopedia of Genes and Genomes*) adatbázis daganatokkal kapcsolatos útvonalai (Kanehisa és mtsai, 2014). Ez a fajta meghatározás szintén önkényes, és a betegséghez rendeli az adott jelátviteli funkcióját, az eredeti, egészséges sejtben betöltött funkciójával szemben (Beyer és mtsai, 2007; Vinayagam és mtsai, 2014).

Ha az interakciós hálózatból indulunk ki, lehetőségünk van matematikailag definiálni a hálózaton belüli csoportokat. A hálózat egy adott területén a jobban kapcsolódó pontokat nevezzük moduloknak, vagy csoportoknak (*community*) (Barabási, 2016; Palla és mtsai, 2005). A teljes hálózaton belül a modulok alkothatnak egy útvonalat. A módszerre különböző hálózaton belüli modularizálási (csoport keresési) algoritmusokat fejlesztettek ki. A csoportkeresési algoritmusok eredményei ritkán találhatók meg adatbázisokban, inkább az interakciós hálózatból *de novo* képesek útvonalakat, biológiai funkciós egységeket meghatározni (Barabási, 2016; Newman, 2006). A legelterjedtebb ilyen algoritmus a Girvan–Newman algoritmus (Girvan és Newman, 2002). Ennek lényege, hogy a legmagasabb köztiségű⁵ éleket veszi ki a hálózatból addig, amíg a hálózat nem esik több darabra (komponensre). Egy másik csoportkereső algoritmus a klikk-perkolációs algoritmus (Adamcsek és mtsai, 2006). Ennek lényege, hogy először a hálózatban adott méretű (k), egymással összekötött pontot határoz meg. Ezt nevezzük klikknek. Ezután ezeket a klikkeket megpróbálja a hálózatban

⁵ Köztiség: Egy gráf adott pontján vagy élén átmenő legrövidebb utak száma.

elforgatni: ha egy másik pont csatlakozik a klikkhez legalább két élen keresztül, akkor az eredeti pontok és az új pont is beletartozik a csoportba. Az eredetileg random meghatározott klikkből kiindulva az algoritmus képes meghatározni a modulokat. Ezt a módszert nemrég sikerrel alkalmazták a BioPlex hálózaton a sejten belüli interakciós modulok megtalálására (Huttlin és mtsai, 2015). Egy másik modul-azonosítási módszer Szalay-Bekő Máté és Kovács István által kifejlesztett ModuLand algoritmus (Kovács és mtsai, 2010; Szalay-Beko és mtsai, 2012). Ez az algoritmus az információterjedés szempontjából vizsgálja a hálózatot, és annak lokális maximumait határozza meg. A lokális maximumtól az információ mennyiségének monoton csökkenéséig tartoznak egy modulba a pontok egy hálózaton belül (Kovács és mtsai, 2010; Szalay-Beko és mtsai, 2012). A modularizáción alapuló módszerek a biológiai relevanciától tekintenek el, és csak az interakciós hálózat szerkezetéből jósolják meg a biológiai csoportokat, azaz a jelátviteli útvonalakat.

A legjobb csoportosítás, ha kommunikációs nyelv szerint próbáljuk meg csoportokba foglalni az egyes jelátviteli útvonalakat (Korcsmáros és mtsai, 2010; Türei és mtsai, 2016; Vinayagam és mtsai, 2014). A molekuláris nyelv az egyes jelátviteli útvonalakban eltérő lehet. Például a MAPK jelátviteli útvonalban a tirozin oldalláncok foszforilálása, vagy a WNT kanonikus jelátviteli útvonalban a β -katenin sejtmagba kerülése, ahogy a jelátviteli útvonalakat ismertető fejezetben láttuk. Ez a fajta csoportosítás megőrzi a biológiai jelentőségét az egyes útvonalaknak, és kevésbé önkényes, evolúciósan is konzervált, fajokon átívelő jelátviteli útvonalakat eredményez (Pires-daSilva és Sommer, 2003). Hátránya lehet a jelátviteli beszélt nyelvek megválasztásának a kérdése. Ennek elkerülésére az útvonalak definíciójakor a Signalink adatbázis építésekor több áttekintő (*review*) cikk került átnézésre, ami alapján egy protein az adott jelátviteli útvonalhoz sorolódott (Korcsmáros és mtsai, 2010).

A felsorolt négy módszer közül az első kettő tetszőleges és akár különböző méretű jelátviteli útvonalakat is eredményezhet nagymértékű átfedésekkel. A harmadik módszer esetén a biológiai relevancia kérdéses, ezért a leghatékonyabb a negyedik módszernek megfelelő jelátviteli adatbázissal dolgozni (Korcsmáros és mtsai, 2010).

5. Fehérje-fehérje interakciós adatbázisok

Jelenleg igen sokféle jelátviteli útvonal és fehérje interakciós adatbázis áll a kutatók rendelkezésére (Chowdhury és Sarkar, 2015; Türei és mtsai, 2016; Vinayagam és mtsai, 2014). Ez egyrészt annak köszönhető, hogy a jelátviteli útvonalakat több évtizede kutatják, és nagy mennyiségű adat áll rendelkezésre a jelátviteli fehérjék felépítéséről és működéséről. Másrészt ezeket az adatbázisokat különböző célok, módszerek és források alapján készítették. Éppen ezért a forrásokat sokszor részletekbe menően és külön-külön kell elemezni. Rendszerszintű elemzésekre inkább azok az adatbázisok alkalmasak, amelyeket általános céllal, lehetőleg egységes módszereket alkalmazva hoztak létre (Csermely és mtsai, 2013). Utóbbiak mennyiség szempontjából jelentősebb részét az elmúlt évtizedben elterjedt, ún. nagy áteresztőképességű (*high-throughput*) módszerek adják. Ezek kapcsolatok ezreit képesek kimutatni, ezért igen gyakran használják a hálózatkutatások során. Ugyanakkor, ezen források nagy része a jelátvitel szempontjából jelentős (extracelluláris, membránkötött, illetve sejtmagi) fehérjéket nem tartalmazza. Az adatbázisok gyűjtése során ugyanis legtöbbször élesztő kettős hibrid eljárást alkalmaznak, amely nem képes kimutatni a sejtmagi kapcsolatokat (mivel azok aktiválhatják a riporter transzkripciós faktort), és a membránhoz kötött fehérjék közötti kapcsolatokat (mivel azok nem képesek bejutni a sejtmagba, és ezért nem képesek aktiválni a riporter transzkripciós faktort). További módszertani probléma, hogy az ilyen adatbázisok irányítatlan kapcsolatokat tartalmaznak, és sajnos gyakori a fals pozitívitás, azaz alacsony a specificitásuk (Brückner és mtsai, 2009).

Ilyen nagy áteresztőképességű gyűjtésű adatbázis például az **IntAct** (Orchard és mtsai, 2014) (*Molecular Interaction Database*), a **BioGRID** (Chatr-aryamontri és mtsai, 2015) (*BIOlogical General Repository for Interaction Datasets*), és **DIP** (*Database of Interacting Proteins*)(Salwinski, 2004).

Az európai molekuláris biológiai kutató központ által fenntartott **IntAct** (Orchard és mtsai, 2014) adatbázis jelenleg (2016. szeptember) 653104 interakciót tartalmaz, melyek összesen 14346 publikációból származnak. Ezek között mind nagy áteresztőképességű, mind pedig kis léptékű publikációk megtalálhatók. Az IntAct adatbázist nemrég összevonták a MINT (Licata és mtsai, 2012) adatbázissal, ezzel is növelve a tárolt kapcsolatok számát, a párhuzamosságok megszüntetésével egyetemben.

A másik nagy adatbázis, mely gyűjti a nagy áteresztőképességű adatokat a **BioGRID** adatbázis (Chatr-aryamontri és mtsai, 2015). Az adatbázisban jelenleg (2016. szeptember) 1 072 173 genetikus és fehérje interakciót tartalmaz különböző fajokban. Ezekből fizikai interakció az emberben 365 547, összesen 25 383 publikációból. Ezen kapcsolatok között élesztő kettős hibrid, affinitás tisztítás kromatográfia, ko-lokalizációs adatok és különféle fehérje-, peptid- és RNS interakciós adatok is megtalálhatóak (BioGRID, 2016).

Interakciós adatbázisok közé tartozik még a **STRING** is (*Search Tool for the Retrieval of Interacting Genes/Proteins*) (Szklarczyk és mtsai, 2015). Ennek az adatbázisnak egyik legnagyobb előnye, hogy szövegbányászati algoritmusokkal keresi meg az egyes fehérjék közötti interakciókat. Emiatt megbízhatósága nem olyan magas, mint az alább részletezett, kézi gyűjtésű útvonal adatbázisok, de nagyobb mennyiségben és gyorsabban képes feldolgozni az egyes molekuláris biológiai témájú tanulmányokat. Az adatbázis további előnye, hogy emellett nagy-áteresztőképességű és részletekbe menő kísérleteket is tartalmaz, valamint a kézi gyűjtésű adatbázisokból is tartalmaz információt. Összesen 184 millió kapcsolatot 2031 különböző organizmusban 9,6 millió különböző fehérje között (2016. szeptember).

Az interakciós adatbázisok közül viszonylag új a teljes humán genomot tandem tömeg spektroszkópos affinitás purifikációs technikával vizsgáló adatbázis a **BioPlex** (*biophysical interactions of ORFeome-based complexes*) (Huttlin és mtsai, 2015). Az adatbázis előnye, hogy egységes metodika szerint készült, és nagy mennyiségű adat áll rendelkezésre benne, hátránya, hogy az affinitás purifikációs vizsgálat miatt a komplexek pontos szerkezete és kapcsolatok iránya nem jósolható.

A fent említett tulajdonságok teszik nélkülözhetetlenné az adatbázisok másik típusát, a kézi gyűjtéssel készült adatbázisokat. Ezek általában kevesebb információt tartalmaznak, de részletesebbek és megbízhatóbbak (Vinayagam és mtsai, 2014). Ugyanakkor a legtöbb ilyen (jelátviteli) adatbázis pontos útvonal-definíció és szabványosított gyűjtési feltételek nélkül készült. Ennek az a következménye, hogy még egy-egy adatbázison belül sem lehet (vagy szabad) az ott lévő útvonalakat összehasonlítani, vagy a közöttük lévő kapcsolatokat vizsgálni. A kézzel gyűjtött adatbázisok másik nagy hibája, hogy torzítanak, mivel csak az ismert, publikált kapcsolatokat gyűjtik ki. Ez azt jelenti, hogy a nagyobb mértékben kutatott fehérjékről

több kapcsolat áll rendelkezésre. Mindez tehát nagyobb megbízhatóságot jelent, de a szenzitivitásuk sajnos alacsony, mivel a kutatási torzítás miatt nagyon sok lehetséges kapcsolat elvész (Vinayagam és mtsai, 2014).

A kézi gyűjtésű adatbázisok fenntartására profitorientált cégek is létrejöttek. Ilyen az **Ingenuity Pathway Analysis** (www.ingenuity.com). Ezek legnagyobb problémája az adatok nem teljes elérhetősége és tudományos felhasználhatósága. Az Ingenuity Pathway Analysis használta során a teljes adatbázis nem tölthető le, mivel az a másolás veszélyét rejti magában, de jól használható mRNS microarray adatok útvonalanalíziséhez (Roychowdhury és mtsai, 2017; Zhou és mtsai, 2016).

Az ingyenes kézzel gyűjtött jelátviteli adatbázisokból több is létezik. Ilyen a **KEGG**, a **NetPath** vagy a **Reactome** (Croft és mtsai, 2014; Keshava Prasad és mtsai, 2009; Ogata és mtsai, 1999). Ezek az adatbázisok útvonal annotációja sokszor nem egységes és nem meghatározott. Az ingyenes kézzel gyűjtött jelátviteli adatbázisokból több is létezik. Ilyen a **KEGG**, a **NetPath** vagy a **Reactome** (Croft és mtsai, 2014; Keshava Prasad és mtsai, 2009; Ogata és mtsai, 1999). Ezek az adatbázisok útvonal annotációja sokszor nem egységes és nem meghatározott.

A három adatbázis közül a **Reactome** tartalmazza a legtöbb kapcsolatot és útvonalat (Croft és mtsai, 2014). Jelenleg a Gene Ontology-hoz (Consortium, 2015) hasonlóan az egyes útvonalak egymásba ágyazottak. Ezzel lehetőség van a Gene Ontology-hoz használt módszerekkel a Reactome-mal is gén expressziós dúsulásos (*enrichment*) vizsgálatokat végezni (Creixell és mtsai, 2015; Croft és mtsai, 2014). A reakciók gyűjtése miatt a Reactome nagy mennyiségű komplexet tartalmaz, ezzel emeli a komplexben részt vevő kapcsolatainak számát és a kluszterezettségét, mely hálózatbiológiai elemzéskor eltéréshez vezethet. A hátránya, hogy az útvonalak közötti cross-talkok vizsgálatára nem alkalmas, mivel az adatok újra csoportosítása szükséges ahhoz, hogy egységes, nagyméretű, az evolúción és a molekuláris nyelven alapuló útvonalakat kapjunk. A Reactome a faszzerű annotációs struktúrája miatt kis lépésekben építi fel az útvonalait, ezáltal például a MAPK útvonalat RAF, MAP kináz kaszkádnak nevezi és felosztja több kisebb részre, mint a PDGFR jelátvitel vagy PKC és MAPK jelátvitel.

A **KEGG** a legrégebbi jelátviteli adatbázis (Ogata és mtsai, 1999). Leghasznosabb dúsulásos vizsgálatok végzésére, azon belül is betegség specifikus, vagy metabolikus

Emellett nagyobb biológiai funkciók is megtalálhatók az adatbázisban, mint „apoptosis” vagy „osteogenesis”.

A **XtalkDB** ténylegesen a különböző útvonalakba sorolt ligandok és a transzkripciós faktorok közötti utakat tekinti cross-talknak (Sam és mtsai, 2017; Tegge és mtsai, 2015), de az útvonal annotációját a KEGG-ből nyeri, ami nem különíti el az egyes paralógok közötti kapcsolatokat (Kanehisa és mtsai, 2014).

6. Az autofágiát vizsgáló adatbázisok

Korábban az autofágia regulációjának felderítésében két kézi gyűjtésű adatbázisra támaszkodhattak a kutatók. Az egyik a **HADb** (*Human Autophagy Database*), mely az emberi autofágiában részt vevő fehérjéket gyűjti össze (Moussay és mtsai, 2011). A másik az **ADB** (*Autophagy Database*), mely nem csupán az emberi, hanem több más faj autofágiában részt vevő génjeit tartalmazza (Homma és mtsai, 2011). Az utóbbiban néhány fehérje esetében kapcsolatlista is található az egyes fehérjékre vonatkoztatva. Hiányoznak azonban az irodalmi hivatkozások és nem tartalmazza az autofágia transzkripcionális és poszt-transzkripcionális regulációját sem.

Az autofágia daganatokban is jelentős szereppel bír, két élű kardként viselkedik (Kubisch és mtsai, 2013). A mell, petefészek és prosztata daganatokban leírták a BECLIN1 fehérje mutációját, és ezért az autofágiának valószínű tumorszupresszor hatást tulajdonítanak. Ugyanakkor például hepatocelluláris carcinomában az emelkedett autofágia a daganat malignitásának a fokozódásához vezet EMT-n keresztül (Li és mtsai, 2013a). Ilyen eredmények tették lehetővé, hogy az autofágiát gyógyszeresen célozhassuk, például az mTOR fehérjéken keresztül. Sajnos az ilyen jellegű kísérletek közepes sikert értek az mTOR fehérjék központi szerepe miatt (Rubinsztein és mtsai, 2007).

7. Regulációs adatbázisok

A fentiekben tárgyalt jelátviteli útvonal és fehérje-fehérje interakciós adatbázisok nem tartalmazzák a jelátvitel eredményét, a transzkripcionális interakciókat és a jelátvitel regulációját.

7.1. Transzkripciós faktor célgén adatok meghatározása és adatbázisai

A transzkripciós faktor célgén kapcsolatok lehetővé teszik a jelátviteli visszacsatolási hurkok modellezését.

A transzkripciós faktorok a DNS-hez kötődnek különféle DNS kötő doménekkel, mint amilyen a leucin cipzár, cink ujj vagy a homo domén (Mitchell és Tjian, 1989). A DNS-en felismernek egy speciális 8-15 nukleotid hosszú szekvenciát. Ezt nevezzük transzkripciós faktor kötőhelynek.

7.1.1. A transzkripciós faktor kötőhelyek meghatározásának módszerei

A transzkripciós faktor kötőhelyek meghatározásának leggyakrabban **ChIP-Seq** (kromatin immunprecipitációs szekvenálás) adatokat használnak. Ennek lényege, hogy a sejtek emésztése után a transzkripciós faktorokat a hozzájuk kötődött DNS szakasszal együtt kinyerik (kromatin immunprecipitáció), majd a kötődött nukleotid szekvenciákat megszekvenálják (Johnson és mtsai, 2007). Sajnos az összes transzkripciós faktor pontos genomi kötőhelyének ismerete nem lehetséges, mivel a lehető legtöbb sejtállapotban kellene megvizsgálni, hogy a transzkripciós faktorok mely nukleotidokhoz kötődnek (Johnson és mtsai, 2007).

Régebbi módszer a **ChIP on chip** (DNS microarray elemzés kromatin immunprecipitáció után) módszer. Ilyenkor a kötődött nukleotidokat nem szekvenálják, hanem fluoreszcens jelzéssel ellátva DNS microarray platformon vizsgálják meg. A mért fluoreszcencia a fehérjéhez kötődő nukleotidoktól származik (Aparicio és mtsai, 2004).

A Chip-seq módszer mellett a transzkripciós faktorok célszekvenciájának meghatározására a másik lehetőség a **SELEX** (*Systematic evolution of ligands by exponential enrichment*) módszer. Ennek lényege, hogy véletlen, rövid oligo-nukleotid szakaszokat adnak a transzkripciós faktorokhoz, majd szelekciós-amplifikációs ciklusokat alkalmazva változó erősségű elúcióval megállapítják az egyes transzkripciós faktorok kötődési szekvenciáit (Jolma és mtsai, 2013).

7.1.2. A kapott szekvencia adatokból transzkripciós faktor kötőhely készítése: A pozíció súlyozott mátrix és információ tartalma

A szekvenciákból lehetőség van pozíció súlyozott mátrix készítésére, mely tartalmazza a transzkripciós faktorhoz kapcsolódott nukleotidokat. Ehhez meg kell határozni, hogy hányszor fordul elő az adott nukleotid a transzkripciós faktorhoz kötött hely specifikusan. Ebből lehet egy gyakorisági értéket számolni. Ezek az információk találhatóak meg a JASPAR adatbázisban (Mathelier és mtsai, 2015).

Az egyes transzkripció faktorok kötőhelye szekvencia logóval mutatható be, ami jellemzi, hogy az egyes nukleotidok milyen gyakorisággal szerepelnek az adott pozícióban. A szekvencia logó y tengelye az adott nukleotid információ tartalma. Az információ tartalom azt mutatja meg, hogy hány darab bináris döntést helyettesítve, megspórolva lehet eljutni a kért információhoz. A DNS-ben található négy nukleotid összesen két döntésnek felel meg: például 1. döntés adenin és guanin vagy timin és citozin, 2. döntés adenin vagy guanin, illetve timin vagy citozin. Ezért, ha az adott logóban az adott helyen csak egy nukleotid található, annak az információ tartalma 2, hiszen így két döntést spórolhatunk meg. Ha a két nukleotid egyenlő eséllyel szerepel, akkor az információ tartalma 1, hiszen így 2 különböző nukleotid közül választhatunk, tehát egy döntést spóroltunk meg a kettőből (Schneider és Stephens, 1990) (1. Egyenlet Schneider szerinti számítással).

$$H_s(j) = - \sum_{i=1}^A f_{ij} \log_2(f_{ij})$$

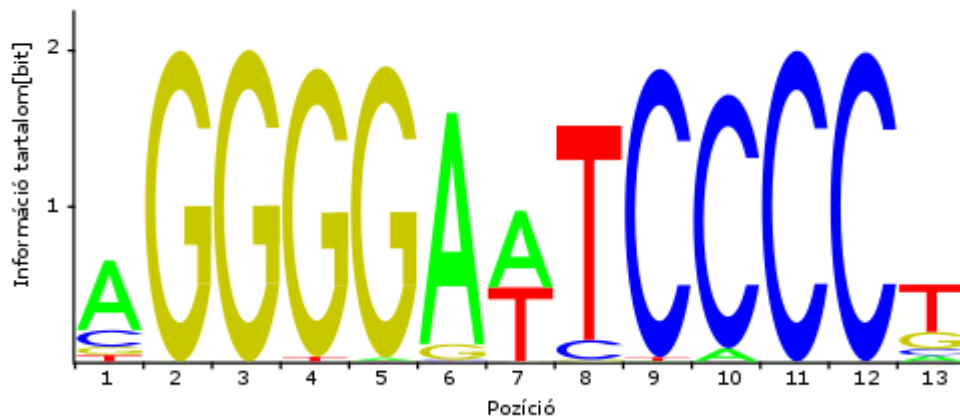
1. Egyenlet

ahol $H_s(j)$ a j-edik helyen mért bizonytalanság. f_{ij} az i-edik betű (nukleotid) előfordulása a j-edik helyen. A bizonytalanságból megállapíthatjuk az információ tartalmat, ahol az $e(n)$ a kevés mérésből származó hiba tényező (2. Egyenlet).

$$R_{seq}(j) = 2 - H_s(j) + e(n)$$

2. Egyenlet

A szekvencia logókból és a pozíció súlyozott mátrixokból meg lehet becsülni az egyes transzkripció faktorok DNS kötő helyét. Sajnos ebből az információból ritkán lehet pontos értéket adni arra vonatkozóan, hogy az adott transzkripció faktor a gén expresszióját serkenti, vagy gátolja, mivel a transzkripció faktorok sokszor egymáshoz is kötődnek, és ezzel új funkciót kaphatnak, illetve kötőhelyüktől akár több ezer nukleotidra is kifejthetik hatásukat.



3. ábra Az NFκB transzkripciós faktor pozíció súlyozott szekvencia logója a JASPAR adatbázisból

7.1.3. Transzkripciós faktor célgén adatbázisok

A transzkripciós faktorokat több adatbázis is tartalmazza. Ilyen például a **JASPAR** adatbázis (Mathelier és mtsai, 2015). Ebben különböző transzkripciós faktorok pozíció súlyozott mátrixai találhatóak ChIP-Seq, ChIP-Chip és SELEX adatokból.

Az egyik legnagyobb, nemzeteken átívelő projekt, mely a genom szabályozásának megértését tűzte ki célul az **ENCODE** (Auerbach és mtsai, 2013; Diehl és Boyle, 2016; Dunham és mtsai, 2012; Qu és Fang, 2013) projekt, mely nemcsak transzkripciós faktor és mRNS, hanem transzkripciós faktor és miRNS kapcsolatokat is tartalmaz.

A **HTRIdb** (Bovolenta és mtsai, 2012) irodalmi gyűjtésű adatbázis, mely nagy áteresztőképességű ChIP-Seq és ChIP-Chip adatokat tartalmaz. Az egyes kapcsolatokhoz súlyokat rendel aszerint, hogy az adott kapcsolat hányszor fordult elő különböző forrásokban.

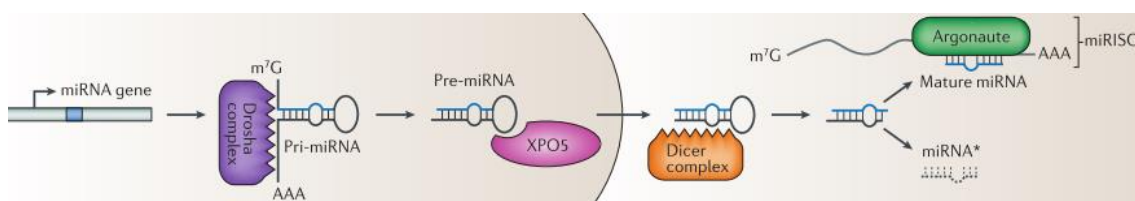
Az **ORegAnno** irodalmi gyűjtésen keresztül működő transzkripciós faktor célgén adatbázis. Az irodalmi gyűjtést számítógépes algoritmusok végzik, de a kutatók hozzá is adhatnak az adatbázishoz (Montgomery és mtsai, 2006). Az ORegAnno adatait, és több más adatbázist is tartalmaz a **PAZAR** (Portales-Casamar és mtsai, 2009) adatbázis, melyről néhány specifikus gyűjtés és az ENCODE projekt adatai is letölthetőek.

Az egyik legnagyobb adatbázis, mellyel lehetőség van kutatásokat végezni a **TRANSFAC**, de sajnos az adatbázis csak licenc vásárlással érhető el (Knüppel és mtsai, 1994).

Gépi algoritmusos keresést és kézzel gyűjtött adatokat egyaránt tartalmaz a **TRRUST** adatbázis (Han és mtsai, 2015).

7.2. A miRNS reguláció és adatbázisai

A miRNS-ek 21-24 nukleotid hosszúságú RNS molekulák, melyek egy 8 nukleotid hosszú felismerő hellyel rendelkeznek, amivel képesek a különböző transzkriptálódott mRNS-eket felismerni (Pasquinelli, 2012). Az eredeti leírásuk *Caenorhabditis elegans*-ből származik (Lee és mtsai, 1993; Wightman és mtsai, 1993). Az emberi szervezetben az alábbi komplex érési folyamaton mennek keresztül. A miRNS kódoló génekről az RNS polimeráz II írja át a pri-miRNS-t (Chekulaeva és Filipowicz, 2009; Kim és mtsai, 2009)(Lee és mtsai, 1993; Wightman és mtsai, 1993). Az emberi szervezetben az alábbi komplex érési folyamaton mennek keresztül. A miRNS kódoló génekről az RNS polimeráz II írja át a pri-miRNS-t (Chekulaeva és Filipowicz, 2009; Kim és mtsai, 2009). Ez a hajtúkanyarulattal összekötött duplaszálú pri-miRNS a sejtmagban a Drosha (drosha ribonuclease III) komplexhez kötődik (Krol és mtsai, 2010). Itt képződik a pre-miRNS struktúrája, amit az exportin 5-ös visz ki a sejtmagból (XPO5). Ezt a Dicer komplex ismeri föl, ami eltávolítja a pre-miRNS hajtúkanyarulatát. Az így képződött két szálból az egyik az Argonaute fehérjekomplexhez kötődik, és mint miRISC komplex részt vesz az mRNS degradációs folyamatokban,. A másik szál a citoplazmában a nukleázok által degradálódik (Pasquinelli, 2012). A folyamatot a 4. ábra szemlélteti (Pasquinelli, 2012).



4. ábra A miRNS-ek érési folyamata (Pasquinelli, 2012)

A képződött RISC komplex emlősökben elsősorban az mRNS-ek 3' UTR régióján kapcsolódik. A nukleotid párosodás a miRNS 8 nukleotidos felismerő (un. *seed*) régiójában a legtrikább esetben pontos. Általában van egy lötyögés a miRNS-ek és a célpont mRNS-ek között. Éppen emiatt roppant módon nehéz a célhelyek pontos jóslása, valamint ez okozza az egyes miRNS-ek nagyon széles célpont-specifikusságát. Egy

miRNS a genomnak akár az egy százalékát is regulálhatja (Pasquinelli, 2012). A különböző adatbázisokban éppen ezeket a problémákat igyekeztek elhárítani.

A miRNS adatbázisok közül két nagy csoportot különíthetünk el: egyrészt magukat a miRNS szekvenciákat gyűjtő adatbázisokat, másrészt a miRNS és azok célszekvenciáit gyűjtő adatbázisokat. A legfontosabb az összegző adatbázisok közül a miRBase, amely az összes eddig ismert miRNS szekvenciát tartalmazza (www.mirbase.org) (Kozomara és Griffiths-Jones, 2011). Mindegyik miRNS-nek külön miRBase azonosítója van, amivel lehetőség nyílik az egyedi azonosításukra.

Az utóbbi csoportba tartozó adatbázisokban több eltérő módszer szerint történik a kapcsolatok azonosítása és definiálása, ezért érdemes lehet ezeket megvizsgálni.

1. Kísérletes alapon igazolt kézzel gyűjtött adatbázisok.
2. Cél régió konzerváltság és seed régió ismeretén alapuló adatbázisok.
3. miRNS-ek és célgénjeik közötti antikorrreláción alapuló adatbázisok és módszerek.
4. A fentieket összegyűjtő adatbázisok, például a miRecords (Xiao és mtsai, 2009).

7.2.1. Kísérleti alapon igazolt miRNS adatbázisok

Ezek az adatbázisok több száz cikk eredményeit tartalmazzák. Ennek megfelelően az ilyen adatbázisok általában kicsik, néhány száz miRNS-t és célgénjeiket tartalmazzák. Megbízhatóságuk viszont nagy, és általában szerepelnek az adott kapcsolatot leíró hivatkozások is, így lehetőség nyílik minden egyes kapcsolatnak utánanézni. Hátrányuk viszont, hogy korlátozott a méretük, és sajnos nem túl gyakori a frissítésük.

Ilyen adatbázisok közé tartozik a **TarBase** (Papadopoulos és mtsai, 2009; Vergoulis és mtsai, 2012), a **miRECORDS** validált célpontokat tartalmazó része (Xiao és mtsai, 2009), és a **miRTarBase** (Hsu és mtsai, 2011). Egy külön kategóriát alkot az ilyen adatbázisokon belül a nagy áteresztőképességű adatokat gyűjtő **StarBase** (Li és mtsai, 2014). Egyes esetekben az is elképzelhető, hogy kutatók az általuk tanulmányozott rendszer mikroRNS szabályozását veszik górcső alá és gyűjtik össze az ez irányú publikációkat. Ilyen adatbázis például az apoptózist és autofágiát szabályozó mikroRNSeket összegyűjtő **mirDeathDB** (Xu és Li, 2012), vagy a betegségekben érintett miRNS-eket tartalmazó **miR2Disease** (Jiang és mtsai, 2009).

Ha megnézzük a fenti adatbázisokat, feltűnik, hogy a TarBase 2012-es (6.0) és a miRTarBase 2013-as (4.5) verziója mind nagyon nagy, méretű. Ennek az oka az, hogy

mindkét esetben az adatok gyűjtésekor szövegbányászati módszereket is alkalmaztak. A miRTarBase a PubMed összes kivonatában kereste meg a miRNS témájú cikkeket, majd ezek teljes szövegében nézték meg, hogy milyen módszereket használtak, és milyen célpontokat találtak.

A TarBase esetében a kivonat gyűjtemény a MedLine volt. Az innen származó találatok közül egy mérőszám segítségével megállapították, melyek a legfontosabbak. Ezt követően a cikkek átolvasás után bekerültek az adatbázisba. Az adatbányászat és az irodalom exponenciális növekedése egy nagyságrenddel megemelte a miRNS célpontok számát. A Tarbase 5.0-ás (Sethupathy és mtsai, 2006) verziójában még mindössze 5000 interakció szerepelt, a 6.0-ban (Vergoulis és mtsai, 2012) pedig több mint 65000, a 7.0 verzióban pedig már félmillió (Vlachos és mtsai, 2015).

Fontos megjegyezni, hogy ezen adatbázisok mindegyike tartalmaz egy kézi gyűjtésű lépést, amely következtében a megbízhatóságuk magas szintű, szemben a csak gépi kereső algoritmusokkal dolgozó adatbázisokkal. Csak gépi algoritmusokkal dolgozó adatbázis például a **miRWalk** validált célpontokat tartalmazó része (Dweep és mtsai, 2011).

7.2.2. Konzerváltsági és kötési energián alapuló miRNS adatbázisok

A miRNS-ek célpontjai az mRNS-ek 3' nem transzlálódó régiójában találhatóak. Ha van komplementerség bizonyos lötyögéssel a miRNS-ek seed régiójával, és a fenti régió több faj között is konzervált területet tartalmaz, akkor az, feltehetően funkcionális, tehát a kapcsolat valós. Erre a hipotézisre építenek az evolúciós konzerváltság alapú adatbázisok, mint a **PicTar** (Krek és mtsai, 2005) és a **TargetScan** (Volinia és mtsai, 2010). Ezen adatbázisok mérete általában egy nagyságrenddel nagyobb, mint a kézzel gyűjtött adatbázisoké.

A **TargetScan** először egy evolúciós fát alkot, ami segítségével megállapítja az egyes régiók konzerváltságát. Maga az algoritmus figyelembe veszi, hogy a miRNS seed régiói körül milyen nukleotidok fordulnak elő. Az első nukleotid a miRNS-ek többségénél uracil, ami számos esetben szintén illeszkedik. Azoknál a miRNS-eknél, ahol az első nukleotid nem uracil, kétszer gyakoribb a mag régiót közvetlenül megelőző pozícióban az adenin konzervált jelenléte, mint más nukleotidoké. Ezen feltételek beépítése az algoritmusba javítja a predikciót.

A **TargetScanS** (Lewis és mtsai, 2005) algoritmus a 6 nukleotidból álló mag tökéletes illeszkedését és/vagy a 8. nukleotid illeszkedését, vagy az 1. pozícióban az mRNS-ben adenin jelenlétét követeli meg. A mag régió kívül alacsony szintű a szekvenciák konzerváltsága. Ahol a 8. nukleotid illeszkedik, a 9-es pozícióban magasabb a konzerváltság mértéke. A 9. pozícióban szintén gyakoribb az adenin az mRNS-ben. Ezen feltétel beépítése tovább javítja az algoritmust. A 9. nukleotidtól upstream irányba eső szakaszok, melyek szerepet játszhatnak a miRNS 3' végének illeszkedésében, nem mutatnak magasabb konzerváltságot, mint az 1. nukleotidtól downstream, az illeszkedő miRNS-sel már nem átfedő szakaszok. Ez alapján kizárva azon illeszkedéseket, melyek hosszabb konzervált szekvenciákba esnek (azaz a miRNS kötőhelyeket rövid konzervált szigetként definiálva), tovább növelhető az algoritmus prediktív ereje. A legtöbb kísérletesen igazolt miRNS-mRNS szabályozási kapcsolat esetében több kötőhely is megfigyelhető egy mRNS 3' UTR régiójában. Számos példa akad, ahol csak egyetlen kötőhely azonosítható a miRNS és az mRNS között, mégis miRNS transzfekcióval⁶ befolyásolható az mRNS-ek szintje. Az illeszkedő oktamerek 43%-a ilyen, míg a heptamereknél 19-25% ez az arány. A több kötőhely hatása összeadódik, kis távolság (8-40 nukleotid) esetén pedig erősítik egymás hatását (Grimson és mtsai, 2007). A TargetScan eredetileg csak a több kötőhellyel rendelkező kapcsolatokat fogadta el, a TargetScanS (Lewis és mtsai, 2005) azonban nincs tekintettel a kötőhelyek számára.

Más megközelítést alkalmaz a **miRanda** algoritmus (Enright és mtsai, 2003). Itt nem konzerváltsági alapon próbálták megkeresni a célpontokat, hanem a szabadenergia szint változás alapján próbálták megtalálni a legerősebb kötéseket. A gond a módszerrel az, hogy a miRNS-ek esetén nem gyakoribbak a legerősebb, legnagyobb szabadenergia változással járó guanin-citozin közötti kötések. A miRNS-ek feltekeredéséből adódó szabadenergia változások nagyobb mértékűek a kötési energiából közvetlenül származó szabadenergia változásoknál. A miRanda algoritmus fejlettebb változata a mirSVR score (Betel és mtsai, 2010), amiben már a konzerváltságot is figyelembe veszik. Az eredményeket a microRNA.org weboldalon tették közzé (Betel és mtsai, 2008).

⁶ Transzfekciónak nevezzük mikor az eukarióta sejt idegen nukleotidot vesz fel, ebben az esetben miRNS-t.

7.2.3. Antikorreláció alapon működő mikroRNS adatbázisok

A miRNS-ek célpontjukra való hatását leírhatjuk az alábbi egyenlettel (Muniategui és mtsai, 2012) (3. Egyenlet).

$$x_j = \sum_{k=1}^K \beta_{jk} \cdot z_k \cdot c_{jk} + x_j^0 + \varepsilon_j$$

3. Egyenlet

Az x_j az mRNS-ek, a z_k a miRNS-ek szintjét jelzi, a β_{jk} a k-dik miRNS hatása a j-ik mRNS-re, az x_j^0 a j-edik mRNS szintjét jelzi miRNS hatás nélkül, az ε_j pedig a hiba tag. A c_{jk} azt mondja meg, hogy van-e szekvencia homológia a cél mRNS és miRNS között, értéke 1 ha van 0 ha nincs. Ennek vizsgálatára az előző két fejezetben használt adatbázisok alkalmasak. Mind a miRNS mind pedig az mRNS szintet a kettesalapú logaritmussal mérik, egyrészt azért, mert így a fenti képletben az összeadások szorzásnak, a szorzások pedig hatványkitevőnek minősülnek a tömeghatásos egyenletek logikáját követve, másrészt az expresszió szintjét a microarray chipeknél ilyen formában adják meg leggyakrabban.

Mint látható az alábbi adatbázisok, algoritmusok nem képesek a legújabb miRNS – mRNS kapcsolatokat feltárni, viszont alkalmasak lehetnek arra, hogy a már létező szekvencia homológia és konzerváltság alapján meglévő adatokból kapott képet finomítsák. Valamennyi módszernek a hátránya, hogy csak a legerősebb kapcsolatokat képes kimutatni. Az alábbiakban néhány a fentieket használó algoritmust szeretnék bemutatni.

Az első ilyen algoritmus a **TaLasso** (Muniategui és mtsai, 2012), mely azt a gyakran használt módszert alkalmazza, hogy a szekvencia alapú, jósló mechanizmusokat összeköti az antikorrelációval működő módszerekkel. A 3. Egyenlet c_{jk} konstansa származik a szekvencia alapú jóslóból. A megmaradó lineáris egyenlet rendszer megoldását a legkisebb négyzetek módszerével végzi a program a LASSO algoritmust alkalmazva (Kim, 2007). Ez alapján nyújt becslést az egyes miRNS-ek egyes mRNS-ekre való hatására.

A *Nature Methods* hasábjain megjelent **Genmir++** algoritmushoz (Huang és mtsai, 2007) a kapcsolatok kialakításához ugyanazon mintákból származó miRNS és mRNS expressziós adatokat gyűjtöttek. A kapcsolati erősséget (a 3. Egyenletben a β_{jk} -t) Bayess

háló⁷ alapján állapították meg, ahol az egyes paramétereket a háló egyes pontjai adták. A c_{jk} (a konstans, ami megadja, hogy az adott miRNS célpontja-e az adott mRNS) vételéhez több adatbázist vizsgáltak meg, hogy melyik lenne a legalkalmasabb a szekvencia homológia keresésre. Végül a TARGETSCAN mellett döntöttek, mivel annak a célpontjai mutattak a legjobb átfedést az egyes biológiai funkciókban a Gene Ontology adatbázis alapján.

A **HOCTAR** a “Hostgene Oppositely Correlated TARgets” rövidítése. A módszer (Gennarino és mtsai, 2009) és adatbázis (Gennarino és mtsai, 2011) a miRNS-eket rávetíti azon mRNSekre, amikben a miRNS saját szekvenciája előfordul. Ezt követően különböző szekvencia alapú miRNS-mRNS célpont adatbázisok alapján (miRanda, TargetScan, PicTar) megvizsgálja, hogy mely miRNS-ekhez mely mRNS-ekhez kapcsolhatók. Végezetül egy sorba rendező algoritmus segítségével összehasonlítja, hogy vajon mik azok az mRNS-ek, melyek legjobban antikorrelálnak az egyes miRNS-ek „gazda” (host)-gének expressziójával. A módszer előnye az, hogy nincs szükség miRNS microarray adatok vizsgálatára, csupán mRNS szint méréséből is képes megmutatni, mely miRNS-ek lehetnek aktívak. Hátránya, hogy a rengeteg egyéb hatás elmoshatja a miRNS-ek hatását, elég csupán a transzkripciós faktorokra, vagy a DNS metilációra gondolni (Gennarino és mtsai, 2009).

7.3. Transzkripciós faktor-miRNS adatbázisok

A **PUTMIR** egyike a transzkripciós faktor- miRNS kötéseket gyűjtő adatbázisoknak (Bandyopadhyay és Bhattacharyya, 2010). A vizsgálatához a Transfac adatbázisból gyűjtötték ki a transzkripciós faktorokat és a hozzájuk tartozó pozíció súlyozott mátrixokat. A miRNS génektől 3' és 5' irányba 10000 nukleotidban kerestek kötő szekvenciákat. A pozíció súlyozott mátrix kötési erősségét meghatározták, normalizálva a transzkripciós faktor kötés hosszával. A kapott eloszlás átlagától 1,64 szórással magasabb értékeket vettek szignifikánsnak.

A **TransmiR** adatbázis kézi gyűjtésű transzkripciós faktor - miRNS kapcsolatokat tartalmaz (Wang és mtsai, 2010). Emberi vonatkozásban összesen 86

⁷ Irányított gráffal való modellezése egymástól független és függő eseményeknek, melyekben az élek a feltételes valószínűségek. Például, ha az mRNS szint leesik, miközben a miRNS szint magas, akkor az mRNS-miRNS kötődés valószínűsíthető. Ez felírható minden egyes mRNS-miRNS közötti kapcsolat esetén és a legvalószínűbbek kiválaszthatók, vagy ez a kapcsolat közvetlenül is használható a kapcsolat jóslására, ahogy a Genmir++ algoritmusban történik.

cikkből 82 gén 243 kapcsolatát gyűjtötte 100 miRNS-re. Sajnos az egyes gének közé nem transzkripciós faktorok is belekeveredtek az adatgyűjtés során.

A már korábban említett **ENCODE** adatbázis szintén tartalmaz ilyen jellegű kapcsolatokat (Qu és Fang, 2013).

8. *Az adatbázisokból a rendszerbiológiai modell felé: hálózatok*

Ha ismertek a biológiai entitások közötti interakciók, akkor azokat modellezhetjük egy gráfként (De Las Rivas és Fontanillo, 2010). Egy gráf pontok és azokat összekötő élek halmaza (Euler, 1736). Fehérje interakciós hálózatok esetén a pontok egyes fehérjék, míg az élek interakciókat jelentenek. A regulációs hálózatokban a pontok lehetnek transzkripciós faktorok és azok célgénjei, vagy miRNS-ek és azok az mRNS-ek, amiket miRNS-ek regulálnak. Egy gráf lehet irányított vagy irányítatlan. Irányított, ha ismert, hogy mely pontból mely pontba mutatnak az egyes élek, míg irányítatlan, ha ismeretlen. Például egy biokémiai reakció irányított, ha az egyik enzim foszforilálja a másikat. Irányítatlan a kapcsolat, ha például két fehérje egyszerűen érintkezik egymással. A gráf lehet súlyozott vagy súlyozatlan. Súlyozott, ha az egyes élekhez hozzá van rendelve egy súly érték. Ez lehet az előbbi példánál a reakciókinetikai állandó, vagy az interakció erőssége. Ez sokszor ismeretlen, ezért a gráf súlyozatlan marad (Csermely és mtsai, 2013). A jelátvitelt általában irányított, súlyozatlan gráfként modellezhetjük toplógiai hálózatokban (Samaga és Klamt, 2013). Több lehetőség is lehet súly hozzárendelésére, például gén expressziós értékek (Bansal és mtsai, 2006; Chen és Zhang, 2016), vagy az említett reakciókinetikai állandók hozzáadásával (leginkább metabolikus hálózatokban (Nikoloski és mtsai, 2015)).

A hálózatnak azt a részét, amelyikben minden további pont elérhető, óriás komponensnek nevezzük (Barabási, 2016). Gyakori kérdés a biológiai hálózatokban, hogy melyek a központi, legfontosabb pontok a hálózaton belül. A természetes rendszerek a nem célzott támadásokkal szemben robusztusak, és csak a központi elemek kiejtésekor válnak sebezhetővé (Barabási, 2016; Jeong és mtsai, 2001)(Barabási, 2016; Jeong és mtsai, 2001). Ilyenkor betegségek alakulhatnak ki. Éppen ezért használják a kritikusság mérésére az egyes hálózati centralitási értékeket, mellyel meghatározható, hogy mely pontok kiesése jár a hálózat és ezzel az organizmus számára végzetes következményekkel. Az egyes centralitási értékek részben korrelálnak egymással, de ez nagyban hálózatfüggő (Valente és mtsai, 2008).

A legegyszerűbb centralitási mérőszám a **fokszám** (*degree*). A fokszám a hálózatban adott pont (v) szomszédjainak (éleinek (e_{vu})) a száma: **4. Egyenlet**.

$$D(v) = \sum_u e_{vu}$$

4. Egyenlet

A fokszám a lokális fontosságát méri a pontoknak. A lokális fontosság mellett egy pont globális fontosságának a megállapításhoz az egész hálózatot figyelembe vevő mérőszámra is szükség van. Ezek közül egy könnyen számolható mérőszám a **köztiségi centralitás** (*betweenness centrality*, **5. Egyenlet**) (Anthonisse, Jac, 1971):

$$BC(v) = \sum_s \sum_t \frac{\rho(s, v, t)}{\rho(s, t)} \quad s \neq t \neq v$$

5. Egyenlet

Ez az adott ponton (v) átmenő legrövidebb utak számát méri ($\rho(s, v, t)$), és osztja vissza az összes legrövidebb úttal a hálózatban ($\rho(s, t)$), ahol s és t pontok a hálózatban.

A **kluszterezettség együttható** (*clustering coefficient*) azt méri, hogy egy adott pont szomszédjai közül a szomszédok hány százaléka áll kapcsolatban egymással is. Más szóval, hány darab háromszög ($\lambda(v)$) írható fel az adott pont (v) körüli összes, három részes motívumból ($\tau(v)$) (**6. Egyenlet**) (Watts és Strogatz, 1998).

$$CC_v = \frac{\lambda(v)}{\tau(v)}$$

6. Egyenlet

Ezzel, a hálózaton belül azt mondja meg, hogy egy adott pont mennyire része egy egymással is összefüggő csoportnak (Sabidussi, 1966; Watts és Strogatz, 1998), tehát a lokális centralitások közé tartozik.

Centrálisnak mondhatunk egy pontot, ha a hálózat szélétől messze található. Ennek mérésére alkalmas a **közelség** (*closeness*) (Jeong és mtsai, 2001), amely megméri az adott ponttól (v) a hálózat összes pontjának (N) a távolságát ($d(u, v)$), és ennek a reciprokát veszi (**7. Egyenlet**).

$$CC = \frac{N - 1}{\sum_v d(u, v)}$$

7. Egyenlet

Mivel itt is a teljes hálózathoz mérjük a pontunk távolságát, ezért a pont egész hálózathoz viszonyított, globális centralitását közelíthetjük meg így.

A jelátviteli hálózatokban specifikus hálózati mérőszámra is szükség lehet. Ez a „**csokornyakkendőség**” (*bowtiness*) (Supper és mtsai, 2009) (8. Egyenlet).

$$BT(v) = \sum_L \sum_T \frac{\rho(L, v, T)}{\rho(L, T)}$$

8. Egyenlet

A jelátviteli utak a sejtben egymással kapcsolatban állnak és a különböző ligandokból kiinduló jelek képesek keresztbeszélgetést folytatni egymással (Fazekas és mtsai, 2013; Korcsmáros és mtsai, 2010). A jelátviteli utak a sejtben egymással kapcsolatban állnak és a különböző ligandokból kiinduló jelek képesek keresztbeszélgetést folytatni egymással (Fazekas és mtsai, 2013; Korcsmáros és mtsai, 2010). Jelátviteli útnak a ligandtól (L) a transzkripciós faktorokig (T) terjedő kaszkádot tekintjük. A sok transzkripciós faktort az egyik, míg a ligandokat a másik oldalra téve egy csokornyakkendő két szélét alkotják, a köztük a jelet vivő, egymással keresztbeszélgető fehérjék pedig a nyakkendő csomóját. A csokornyakkendő az azt méri, hogy a ligandoktól a transzkripciós faktorokig egy adott fehérjén (v) hány darab jelátviteli kaszkád (út) ($\rho(L, v, T)$) megy át, elosztva az összes jelátviteli úttal ($\rho(L, T)$) (Supper és mtsai, 2009) (8. Egyenlet). Látható, hogy míg a csokornyakkendő csak a ligandok és a transzkripciós faktorok között méri a jelátviteli utakat, addig a köztiség minden egyes pont között méri a legrövidebb utakat, és azt, hogy hány út megy át az adott ponton.

A magas fokszámú, lokálisan centrális pontokat nevezzük **hub**oknak. A huboknak két típusát lehet elkülöníteni. Az egyik típus, melynek nagy mennyiségű kapcsolata állandó. Az ilyen csomópontokat partizó csomópontoknak (*party hub*) hívjuk (Han és mtsai, 2004). Azokat a csomópontokat, melyek sok kapcsolattal rendelkeznek, de ezek közül egyszerre kevésben vesznek részt, randizó csomópontoknak (*date hub*) hívjuk. A jelátvitel szempontjából a partizó csomópontok lehetnek például egyes állványfehérjék, melyek mindig azonos, de relatíve nagy mennyiségű fehérjével vannak kapcsolatban (pl axin) (Good és mtsai, 2011). Randizó csomópontok lehetnek például az egyes stresszfehérjék (Palotai és mtsai, 2008). A randizó és partizó csomópontok elkülönítésére időbeli interakciós hálózatok feltérképezésekor van lehetőség, de ilyen hálózatok építése komoly problémákba ütköznek (Agarwal és mtsai, 2010).

A hub fehérjéken sokszor nagy mennyiségű út megy keresztül. Ezen fehérjéknek minden centralitás értéke magas. A magas globális centralitás viszont nem mindig jár együtt magas lokális centralitással. Azon fehérjéket, melyek fokszáma alacsonyabb, míg globális centralitása magas **palacknyak** (*bottleneck*), vagy darázsderék fehérjéknek nevezzük (Yu és mtsai, 2007).

8.1. Az esszencialitás hálózat alapú meghatározása

Az esszencialitás, mint fogalom többféleképpen értelmezhető. Genetikai értelmében esszenciálisnak azokat a géneket tekintjük, melyek nélkül az organizmus vagy nem képes kifejlődni vagy pedig nagyon hamar elpusztul (Jeong és mtsai, 2001). Ahhoz, hogy egy gén esszencialitását meghatározzuk, költséges kísérletekre van szükség. Ezért szükség volt az esszencialitás jóslására.

A gráfelméletben a kritikusságnak más jellegű definíciója van. Itt, azt a kapcsolatot tekintjük kritikusságnak, melynek elvesztése után a gráf darabjaira esik szét (Barabási, 2016). Ez elméletben egyet jelent az organizmus halálával is.

Egyes algoritmusok csupán az interakciós hálózatot vették alapul. A fentiekben ismertetett centralitások korrelálnak a hálózatban a gének biológiai esszencialitásával (Estrada, 2006; Mason és Verwoerd, 2007). Az él kluszterezettségi algoritmust bevezetve a hálózaton belül az élek vizsgálatával lehet kimutatni az esszenciális pontokat (Wang és mtsai, 2012).

A másik lehetőség, hogy a szigorúan gráf elméleti centralitások mellé valamilyen biológiai funkciót is vizsgálnak. Ilyen feltételezés volt, hogy az esszenciális fehérjék magasabb eséllyel elemei komplexeknek (Li és mtsai, 2017; Luo és Qi, 2015), több funkciójuk van (Tew és mtsai, 2007), és megváltozik az expressziójuk a különféle sejtes állapotok között (Li és mtsai, 2012).

Egy további lehetőség a hálózat forrás és nyelő pontjainak⁸ összekötése (Di Summa és mtsai, 2011). Az jelátvitelben a forrás pontoknak tekinthetők a bejövő jelet hordozó ligandok, míg nyelő pontoknak a jelátviteli utak vége, a transzkripciós faktorok tekinthetők. Ezek összekötésével meghatározhatóak az adott jelátviteli hálózatban az esszenciális utak.

⁸ Forrás és nyelő avagy source és sink pontok. A forrás pontok csak kimenő kapcsolattal, a nyelő pontok csak bemenő kapcsolattal rendelkeznek. Vegyük észre, hogy ezek a ligandok és a transzkripciós faktorok a jelátvitelben.

9. A hálózatok szerepe a daganatok vizsgálatában

A daganatokat rendszerbiológiai betegségnek tekintjük (Hornberg és mtsai, 2006). A korábban alkalmazott kemoterápiák a gyorsan osztódó daganatokkal szemben (mint az akut myeloid és limfoid leukémiák) érték el a legnagyobb sikereket, de csak mérsékelt eredményt hoztak a szolid daganatokkal (mint a nem-kissejtes tüdőrák vagy a colorectalis carcinoma) szembeni harcban (John E. Niderhuber és mtsai, 2014). A legújabb, célzott kemoterápiás szerek képesek ugyan javítani a túlélést (Siegel és mtsai, 2012), ám a daganatok gyakran felülkerekednek a gyógyszerekkel folytatott evolúciós fegyverkezési versenyben. Ahhoz, hogy a daganatokat kezelni tudjuk, szükség van új terápiás célpontokra.

A daganatban mutálódó génekből képződő fehérjék (továbbiakban mutált fehérjék) száma kevés. Vogelstein és munkatársai mindössze 138 ilyen driver gént írtak le (Vogelstein és mtsai, 2013). A Cancer Gene Census (CGC) adatbázisban, mely a driver (igazoltan daganatot okozó) mutációk gyűjtőhelye, mindössze 547 (Higgins és mtsai, 2007) ilyen mutáció szerepelt 2016 szeptemberében. A daganatokat okozó driver gének mellett rengeteg passenger (a daganat képződésében nem részt vevő) mutáció is előfordul a daganat genomjában (Stratton és mtsai, 2009; Vogelstein és mtsai, 2013). A driver gének közül nem szükséges mindegyik mutációja az egyes daganatokban a jelátvitel áthuzalozásához⁹, mivel a jelátvitel legközpontibb útvonalaiban foglalnak helyet, úgymint a MAPK, TGF, JAK/STAT, Notch, Hedgehog, WNT, illetve a p53 (Pires-daSilva és Sommer, 2003; Vogelstein és mtsai, 2013). Az áthuzalozás jelenségének megértéséhez nem elég magukat a mutált fehérjéket vizsgálni, elemezni kell a közvetlen környezetüket is. Erre, a leggyakrabban használt metodika az útvonalak elemzése (Chen és mtsai, 2013; Reimand és Bader, 2014; Srihari és Ragan, 2013), mellyel lehetőség van új driver gének jóslására.

Az útvonalak azonban kommunikálnak egymással, cross-talkot folytatnak. Az útvonalak közötti cross-talkok megváltozása képes az egyik útvonalból bejövő jelet átalakítani: például halál jelet túlélési jellé alakítani (Hanahan és Weinberg, 2011, 2000; Mimeault és Batra, 2010). A cross-talkok áthuzalozását a javunkra is fordíthatjuk.

⁹ Áthuzalozás (rewiring): a bejövő szignalizációs mechanizmusok más jelátviteli utakra való terelése például daganatokban, vagy gyógyszeres terápiában.

Megtalálva a cross-talk forrását, meggátolhatjuk, vagy serkenthetjük, így a túlélési jelből halál jelet alakítva (Farkas és mtsai, 2011).

A hálózatbiológiai vizsgálatok képesek meghatározni a mutált fehérjék rendszer szintű hatását (Cui és mtsai, 2007a; Li és mtsai, 2009, 2013c; Xia és mtsai, 2011). A hálózatbiológiai vizsgálatok képesek meghatározni a mutált fehérjék rendszer szintű hatását (Cui és mtsai, 2007a; Li és mtsai, 2009, 2013c; Xia és mtsai, 2011). A daganatban részt vevő fehérjék a hálózat centrumában helyezkednek el. Több fehérje hat rájuk közvetlenül, és több fehérjére vannak ők hatással, azaz irányított hálózatban mind magasabb be- és kimenő fokszámmal rendelkeznek (Cui és mtsai, 2007a). Tehát lokálisan központibb szerepet töltenek be. Összekötik a sejthálózat egyes részeit, mert több legrövidebb út megy át rajtuk, magasabb köztiséggel rendelkeznek (Li és mtsai, 2013c; Xia és mtsai, 2011). A mutált fehérjék magasabb kluszterezettségű együtthatóval rendelkeznek, azaz szomszédjainak jelentős része egymással is szomszédos (Cui és mtsai, 2007a; Li és mtsai, 2009). Ez azt jelenti, hogy a jelátvitelnek egy egymással összekötött területét kell módosítani a tumorok progressziójához.

A daganatban való változások másik csoportját a differenciáltan expresszáldó gének alkotják. A microarray és új generációs szekvenálási (NGS) adatok elterjedésével egyre nagyobb szükség volt megállapítani a legrelevánsabb változásokat. Ehhez alkalmazták a differenciáltan expresszáldó génekből képzett hálózatokat, és a legközpontibb differenciáltan expresszáldó géneket próbálták megtalálni például mell vagy prosztatata daganat esetén (Ummanni és mtsai, 2011; Vellaichamy és mtsai, 2010; Wang és mtsai, 2011). Az ilyen jellegű munkák nem vették figyelembe differenciáltan expresszáldó fehérjék interakciós környezetét. A másik lehetőség itt is az útvonalak és a jelek lefutásának komplexebb modellezése volt a differenciáltan expresszáldó gének útvonalakban való dúsulásának vizsgálatával (Ergün és mtsai, 2007). A közvetlen interakciós környezet felmérése a differenciáltan expresszáldó gének esetén sem történt meg.

A differenciált expresszió és a mutáció együttes vizsgálatára elsősorban a The Cancer Genome Atlas Consortium (TCGA) munkája során került sor. Különböző daganatokat a mutációs és differenciáltan expresszáldó gének segítségével különböző csoportokba sorolták (Bass és mtsai, 2014; Cancer és mtsai, 2014; Cancer Genome Atlas Network és mtsai, 2012; Collisson és mtsai, 2014; Koboldt és mtsai, 2012). Ezzel a

daganatkezelés személyre szóló terápia felé való elmozdulásának adtak lehetőséget. A differenciált expresszió és a mutáció együttes vizsgálatára elsősorban a The Cancer Genome Atlas Consortium (TCGA) munkája során került sor. Különböző daganatokat a mutációs és differenciáltan expresszáldó gének segítségével különböző csoportokba sorolták (Bass és mtsai, 2014; Cancer és mtsai, 2014; Cancer Genome Atlas Network és mtsai, 2012; Collisson és mtsai, 2014; Koboldt és mtsai, 2012). Ezzel a daganatkezelés személyre szóló terápia felé való elmozdulásának adtak lehetőséget.

A TCGA adatsoraiból való munka másik megközelítése az volt, hogy az összes adatból próbáltak általános megállapításra jutni a daganatok alapvető természetére vonatkozóan (Chang és mtsai, 2013; Leiserson és mtsai, 2015). Colorectalis carcinoma (Cancer Genome Atlas Network és mtsai, 2012), tüdő adenocarcinoma (Collisson és mtsai, 2014) és mellldaganat (Koboldt és mtsai, 2012) esetén mind a mutációt, mind a normál szövethez viszonyított expressziót, valamint az egyes gének metiláltságát is figyelembe vették. Az adatokból megállapítható volt, hogy eltérő az egyes daganatok útvonal-preferenciája, például a colorectalis carcinoma esetén a WNT, foszfatidil inozitol 3 kináz (PI3K) és a CGC-ben nem szereplő p53 útvonal fordul elő (Cancer Genome Atlas Network és mtsai, 2012). A megváltozó útvonalak tüdő adenocarcinomák esetén az EGFR, PI3K, nuclear factor erythroid 2-related factor 2 (NRF2), p53 és a sejt ciklus szabályozás (Collisson és mtsai, 2014), míg mellldaganatokban a PI3K és a p53 mellé a retinoblasztoma (RB) útvonal társul.

Az összes daganat vizsgálatánál is a legfontosabb cél az útvonalak analízise volt (Leiserson és mtsai, 2015). Sikerült megállapítani, hogy a leggyakrabban daganatot okozó géneket már ismerjük és egyre inkább a specifikus, néhány beteget érintő mutációk kutatására kell koncentrálni. Egyre jobban ismerjük a ritka mutációk között a driver mutációkat (Lawrence és mtsai, 2014), viszont egyik esetben sem vizsgálták a daganatban érintett fehérjék első szomszédját és az interakciós környezetet közvetlenül, ami számtalan új kutatási irányt nyithat a rákkal való harcban, és a személyre szabott terápiai beavatkozások fejlesztésében.

III. Célkitűzések

Mivel mind a mai napig az útvonalak csoportosítása önkényes, és nem a biológiai és evolúciós utakat követi, ezért kutatócsoportunk létrehozta a Signalink jelátviteli adatbázist. Ez a jelátviteli adatbázis, azonban nem tartalmazott regulációs kapcsolatokat, amikkel a jelátvitel visszacsatolásait és a fenotípust irányító transzkripcionális változásokat bemutathattuk volna. Ezért a Signalink adatbázist továbbfejlesztettük, és több más regulációs adatréteget adtunk hozzá. Ezen rétegek közül a transzkripcionális szabályozás elkészítése volt az én feladatomban.

Az elkészült adatbázissal lehetőség volt két cél megvalósításra:

- A jelátvitelben a paralóg csoportok közül a kritikus paralógok azonosítására és jellemzésére.
- A daganatokban előforduló egyes jelátviteli változások vizsgálatára és azok jelátviteli helyzetének és környezetének bemutatására.

PhD munkám során tehát célul tűztem ki, hogy:

1. elkészítem a Signalink 2 jelátviteli adatbázis transzkripcionális faktor -célgén kapcsolatokat és miRNS- célgén kapcsolatokat tartalmazó rétegeit,
2. elkészítsek egy algoritmust a transzkripcionális faktor célgén kapcsolatok jóslásához és ezt alkalmazom két másik fontos sejtbioológiai folyamat – az autofágia és az NRF2 stressz válasz- regulációs vizsgálatára,
3. meghatározom a jelátvitelben a kritikus paralógokat, és elemzem azok hálózati és biológiai jelentőségét,
4. azonosítom négy nagy prevalenciájú daganatban (colorectalis carcinoma, mell daganat, nem-kissejtes tüdő rák és hepatocellularis carcinoma) érintett fehérjéket a jelátvitelben és bemutatom környezetüket első szomszédjaikon keresztül, ezzel illusztrálva a daganatok jelátvitelt áthuzalozó stratégiáját,
5. megkeresem és bemutatom mind a kritikus paralógokat, mind pedig az első szomszédok gyógyszertervezési és alkalmazási lehetőségeit.

IV. Módszerek

1. *Transzkripció faktor pozíció súlyozási mátrix alapján történő DNS kötés jóslása*

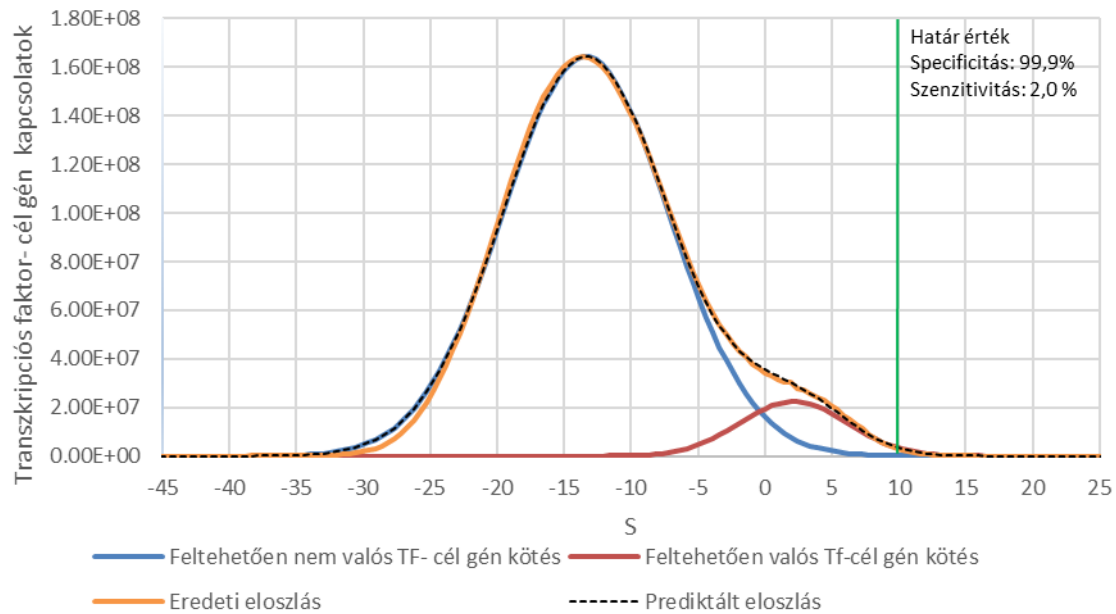
A transzkripció faktor célgénnek vizsgálatához kidolgoztam egy eljárást, amely az összes lehetséges nukleotid kötődésből számolja ki, hogy melyik a lehetséges transzkripció faktor kötőhely. A vizsgálathoz minden egyes transzkripció faktor pozíció súlyozott mátrixát vettem a JASPAR adatbázisból (Mathelier és mtsai, 2015). Ezután minden lehetséges DNS kombinációt elkészítettem az adott transzkripció faktor DNS kötő helyének megfelelően. Kiszámoltam, hogy az adott nukleotid szekvencia mekkora eséllyel tud transzkripció faktort kötni a JASPAR transzkripció faktor kötő mátrixok alapján. Az összehasonlíthatóság érdekében normalizáltam a kapott eredményt ($0,25^n$ a 9. Egyenletben). Elosztottam az átlagos nukleotid frekvencia gyakorisággal. Végül a kapott érték tízes alapú logaritmusát használtam (9. Egyenlet).

$$S = \log \frac{\prod_1^n p_{nukleotid}}{0,25^n}$$

9. Egyenlet

Ahol a transzkripció faktor DNS kötő helyének hossza n , a nukleotid szekvencia és a transzkripció faktor kötésének esélye $\prod_1^n p_{nukleotid}$, $0,25^n$ pedig a normalizálási faktor. A módszer nem vette figyelembe, hogy a genom egyes területein a nukleotidok különböző arányban fordulnak elő. A programozásban Szuromi Gábor, mérnök-informatikus kolléga volt segítségemre.

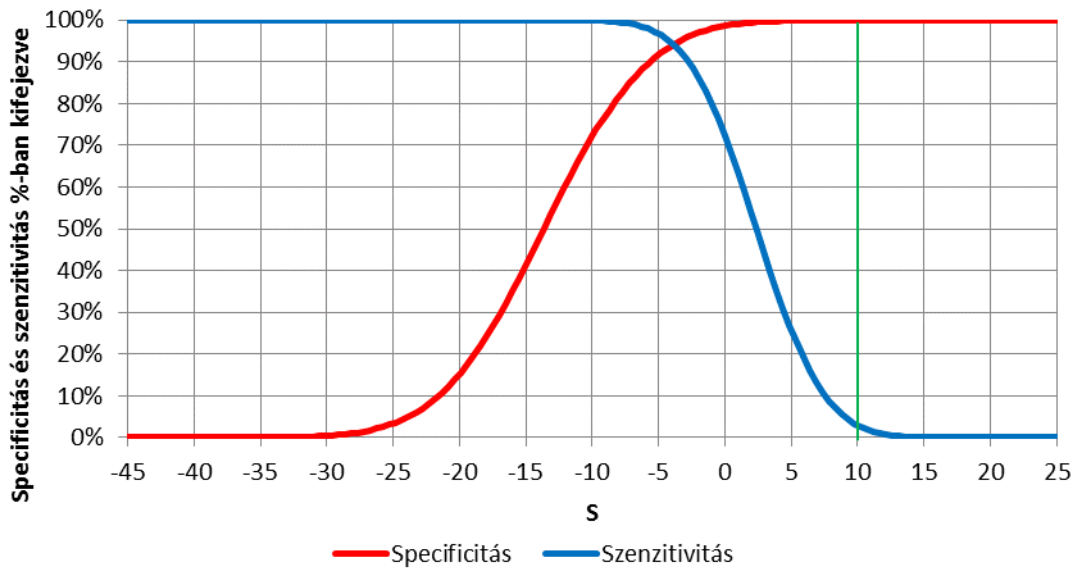
Ezzel az S értékkel dolgoztam tovább. Hisztogramon ábrázolva az eloszlását látható, hogy két Gauss eloszlás található egymáson. A magasabb érték jelzi a valódi kötődést. Ahhoz, hogy meghatározzam a két Gauss görbe várható értékét és szórását, egy Matlab algoritmust írtam, mely heurisztikusan a legkisebb eltérést számolva kereste meg két normál eloszlású görbe legjobb illesztését az általam mért értékre. A programhoz először az eredeti hisztogram első deriváltjait (az inflexiós pontokat) kerestem meg. Ezek a pontok adták a kiindulási középpértékeket. Innen kiindulva a program változtatta a középpérték, illetve a szórás értékét, míg a legjobb fedést meg nem találta. (Az 5. ábra szaggatott fekete és narancssárga görbéje közti különbség minimálisra csökkent.) A heurisztikus algoritmus keresése után az első görbe (5. ábra kék) várható értéke $-13,39$ szórása $2,15$ -nek adódott. A másodiké rendre (5. ábra piros) $6,2$ és $4,01$.



5. ábra A transzkripciós faktor célgén kapcsolatok megoszlása az S érték függvényében. A narancssárga görbe mutatja az eredeti eloszlást. Erre illeszkedik a két normál eloszlású valószínűségi változóból (kék és vörös görbe) készített sűrűség függvény. A zöld vonal jelzi a határértékeket a kapcsolatok esetén.

Azzal a hipotézissel éltem, hogy a magasabb S értékkel rendelkező, azaz a véletlennél nagyobb eséllyel kötődő transzkripciós faktor célgén kapcsolatok találhatóak a kisebb magasságú, második Gauss görbében (Schneider és Stephens, 1990). Sajnos, ahogy látható (5. ábra), nagyon nehéz elkülöníteni a fals pozitívtól a valódi transzkripciós faktor kötések. ROC (*Receiver Operating Characteristic*) analízist használva kerestem meg egy megfelelő határértéket a görbén. A specifitás (valódi negatív/összes negatív, esetünkben határértéktől balra eső terület a kék görbén/összes terület a kék görbén) érték nagyságát szerettem volna magasan tartani, hogy minél kevesebb hamis pozitív transzkripciós faktor célgén kapcsolat kerüljön az adatbázisokba, ezzel természetesen a szenzitivitás (valódi pozitív/összes pozitív esetünkben határértéktől jobbra lévő terület a piros görbén /összes terület a piros görbén) csökkent. Végezetül $S = 10$ határértéket választottam. Ennek a specifitása 99,9%, de a szenzitivitása igen alacsony, mindössze 2.0%. Ez roppant módon pazarlónak tűnhet, de ebben az esetben is 22 026 457 a hipotézis szerint létező transzkripciós faktor célgén kapcsolatról van szó. Az analízishez használt

határérték keresését az 6. ábra mutatja. Végezetül azok a gén szekvenciák, melyek megfelelő határértékkel rendelkeztek és egy adott géntől legfeljebb 2000 nukleotid távolságban helyezkedtek el, transzkripciós faktor kötőhelyként kerültek felismerésre. Mind az introniális és exoniális transzkripciós faktor kötőhelyeket felismertnek tekintettük. A kapott transzkripciós faktor-célgén kötéseket az egyes adatbázisokra lebontva az Eredmények fejezetben közlöm.



6. ábra A specifitás és szenzitivitás meghatározására végzett ROC elemzés

2. Használt adatbázisok

A munkám során nagyon sok különböző adatbázist használtam mind a saját adatbázisaink regulációs rétegeinek felépítéséhez, mind pedig a különböző vizsgálatokhoz. Az egyes adatbázisokat a 2. táblázat tartalmazza, és az alábbiakban részletezem munkafolyamathoz kapcsolatan.

2.1. miRNS célpont adatbázisok:

A miRNS adatbázisoknál a miRNS neveket miRBase azonosítóra fordítottam (Kozomara és Griffiths-Jones, 2011), a célgének különböző azonosítóit pedig UniProt azonosítóra. Az előbbit Python program segítségével végeztem, az utóbbihoz az UniProt ID fordító szolgáltatását és az Európai Bioinformatikai Intézet PICR (Wein és mtsai, 2012) platformját használtam.

A predikciós miRNS adatbázisok közül a PicTar (Krek és mtsai, 2005), TargetSCAN (Volinia és mtsai, 2010) és a miRanda (Betel és mtsai, 2010; Enright és

mtsai, 2003) algoritmust használó microRNA.org (Betel és mtsai, 2008) került integrálásra a Signalink 2 esetén. A kézi gyűjtésű adatbázisok közül a Tarbase (Vergoulis és mtsai, 2012; Vlachos és mtsai, 2015) és a miRecords (Xiao és mtsai, 2009) kézi gyűjtésű része került be mindegyik adatbázisba. Az ARN-be bekerült még a miR2Disease (Jiang és mtsai, 2009), miRDeathDB (Xu és Li, 2012) és a miRTarBase (Hsu és mtsai, 2014). Az adatok integrálását valamennyi adatbázisba Türei Dénessel, biológus PhD hallgató társammal közösen végeztük.

A kritikus paralógok vizsgálatára a miRTarBase adatbázist használtam (Hsu és mtsai, 2014), mivel a Tarbase a vizsgálat ideje alatt nem volt elérhető, és a miRTarBase volt akkor a legnagyobb, elérhető, kézi gyűjtésű miRNS célgén adatbázis.

2.2. Transzkripció faktor célgén adatbázisok

A transzkripció faktor célgén kapcsolatok feltöltésére a különböző adatbázisokban az emberi jelátvitelhez az ABS (Blanco és mtsai, 2006), az ENCODE (Diehl és Boyle, 2016), az ORegAnno (Montgomery és mtsai, 2006) a PAZAR (Portales-Casamar és mtsai, 2009) és a HTRI-t (Bovolenta és mtsai, 2012) használtuk. Ehhez az azonosítókat UniProt azonosítókra fordítottuk a PICR (Wein és mtsai, 2012) szoftver segítségével.

A kritikus paralóg csoportok vizsgálatok az ORegAnno (Montgomery és mtsai, 2006) adatbázist töltöttem le a PAZAR (Portales-Casamar és mtsai, 2009) adatbázisból és a HTRI-t (Bovolenta és mtsai, 2012).

2.3. Transzkripció faktor miRNS adatbázisok

Az adatbázisok közül az ENCODE (Diehl és Boyle, 2016), a PuTmiR (Bandyopadhyay és Bhattacharyya, 2010) és a TransmiR (Wang és mtsai, 2010) tartalmazott transzkripció faktor-célgén kapcsolatokat. A fehérje azonosítókat UniProt azonosítókra fordítottuk, a miRNS azonosítókat miRBase azonosítókra, és így kerültek beépítésre az adatbázisokba.

2.4. Betegség specifikus adatbázisok

A daganatokban érintett fehérjék meghatározáshoz szükség volt egy adatbázisra, mely tartalmazza azon géneket, melyek mutációja daganatot okoz. Ez az adatbázis a Cancer Gene Census (CGC) (Futreal és mtsai, 2004) volt. Innen az első szomszédokhoz daganat specifikusan szűrtük ki a daganatot okozó géneket. Ez a fajta szűrés lehetőséget

teremtett a daganat-specifikusság növelésére. Az egyes kulcsszavakat a 1. táblázat tartalmazza.

1. táblázat A daganatban érintett fehérjék meghatározására használt kereső kifejezések a Cancer Gene Census adatbázisban

Daganat típusa	Keresett kifejezés a CGC-ben
Colorectalis carcinoma	adenomatous polyposis coli, colorectal, adenomatous polyposis coli; Turcot syndrome
Mell daganat	familial breast cancer, lobular breast, breast
Hepatocellular carcinoma	hepatocellular carcinoma, hepatocellular
NSCL	lung, NSCLC, de nem tartalmazta „small cell lung carcinoma”

A kritikus paralóg csoportok daganatban való érintettségének vizsgálatokor a teljes CGC adatbázist használtam.

Az örökletes betegség specifikus adatok kinyerésére az OMIM (Online Mendelian Inheritance in Man) adatbázist használtam (Amberger és mtsai, 2015).

2.5. Gyógyszerészeti fontosság és molekuláris célpontok meghatározására használt adatbázisok

A gyógyszer és kismolekula adatok vizsgálatához a ChEMBL adatbázist (Bento és mtsai, 2014; Gaulton és mtsai, 2012) használtam fel mind az első szomszéd fehérjék vizsgálatához, mind pedig a kritikus paralóg csoportok vizsgálatához. Akkor tekintettem egy molekulát gyógyszernek, ha már eljutott legalább klinikai 4. fázisba (nagy mennyiségű betegre való mértékű tesztelés). Viszonylag erős határértéket alkalmaztam a ChEMBL adatbázis szűrésére a célpontok meghatározásakor: 500 nM IC₅₀¹⁰, K_i¹¹ és K_d¹² értéket tekintettünk határértéknek. A fenti határérték már alkalmas lehet struktúra alapú hasonlósági keresésekre, ha a compoundok között új vezető (*lead*) molekulákat szeretnénk találni a daganat ellenes gyógyszerek kutatására is. A ChEMBL adatbázis letöltésében és szűrésében Krishna Bulusu a Cambridgei Egyetem Kémia Tanszékének

¹⁰ Az IC₅₀ az inhibitory concentration 50 rövidítése. A molekula azon koncentrációja, ahol az adott biokémiai vagy sejtes funkciót 50%-ban képes gátolni. Például a sejtek fele meghal ekkora koncentráció alkalmazásakor.

¹¹ K_i inhibitorikus egyensúlyi állandó.

¹² K_d disszociációs konstans

munkatársa volt a segítségemre. A gyógyszerek indikációihoz az ATC klasszifikációt használtam (Berg és mtsai, 2014). Ez a klasszifikáció a gyógyszerek terápiás indikációit veszi számba, és nem veszi figyelembe a hatásmechanizmust, vagy a molekulák hasonlóságát.

2.6. Interakciós és jelátviteli adatbázisok

A kritikus paralóg csoportok vizsgálatához az általam és a kutató csoportunk által fejlesztett Signalink 2 adatbázist használtam (Fazekas és mtsai, 2013), mivel csak ez az adatbázis tartalmazott megfelelő jelátviteli útvonal besorolást, ami ehhez a vizsgálatához nélkülözhetetlen volt (Vinayagam és mtsai, 2014). A Signalink 2-nek csak a legmegbízhatóbb, a kézi gyűjtésű kapcsolatokat tartalmazó jelátviteli rétegét használtam fel mind az első szomszédok, mind pedig a kritikus paralógok vizsgálata során.

A daganatban érintett fehérjék és első szomszédjaik vizsgálatához, több jelátviteli és interakciós adatbázist is felhasználtam, hogy igazolni lehessen az eredmények adatforrástól független voltát. Jelátviteli adatbázisként a Signalink 2 mellett a Reactome-t (Croft és mtsai, 2014), mint legnagyobb, ilyen jellegű adatbázis és Cui és munkatársai (Cui és mtsai, 2007a) által épített a daganatok jelátvitelét vizsgáló adatbázist használtam.

Interakciós adatforrásként a kézi gyűjtésű HPRD-t (Keshava Prasad és mtsai, 2009) használtam fel. Nagy áteresztőképességű adatok vizsgálatához összevontam a BioGRID (Chatr-aryamontri és mtsai, 2015), IntAct (Orchard és mtsai, 2014) és DIP (Salwinski, 2004) adatbázisokat, hogy egy egységes képet kapjak. Valamennyi adatbázist UniProt (The UniProt Consortium, 2014) azonosítóra fordítottam, és csak a SwissProt (ellenőrzött) emberi azonosítókat tartottam meg. A kapott interakciókból, egy általam írt Python script, illetve az Igraph modul segítségével kikerestem az önmagukra ható kapcsolatokat és a dupla interakciókat.

2.7. Egyéb adatbázisok

A Gene Ontology (Consortium, 2015) molekuláris funkcióra (*Molecular Function*) és biológiai funkcióra (*Biological Process*) vonatkozó adatokat az UniProt (The UniProt Consortium, 2014), a gén expressziós adatokat a Gene Expression Omnibus adatbázisból (Barrett és mtsai, 2013; Edgar és mtsai, 2002) töltöttem le. A

szövetspecifikus expressziók meghatározásához az Ensembl 74 (Flicek és mtsai, 2014) verziójának “egentics” adatait használtam.

2. táblázat Az alkalmazott adatbázisok a felhasznált adatok megnevezésével.

Adatbázis	Tulajdonságok	Referencia	Elérhetőség
SignaLink	Jelátviteli fehérjék kiindulási listája, fehérje-fehérje kapcsolatok, fehérje tulajdonságok (útvonal tagság, cross-talk és jelátviteli szerep)	(Korcsmáros és mtsai, 2010)	http://www.signalink.org/
Reactome	Jelátviteli útvonal adatbázis	(Croft és mtsai, 2014)	http://www.reactome.org/
HPRD	Emberi fehérje-fehérje interakciós adatbázis	(Keshava Prasad és mtsai, 2009)	http://www.hprd.org/
Database of Interacting Proteins	Interakciós adatbázis	(Salwinski, 2004)	http://dip.doe-mpi.ucla.edu/dip/Main.cgi
IntAct	Interakciós adatbázis	(Orchard és mtsai, 2014)	http://www.ebi.ac.uk/intact/
BioGrid	Interakciós adatbázis	(Stark, 2006)	http://thebiogrid.org/
Cui és munkatársai	Daganatok jelátviteli adatbázisa	(Cui és mtsai, 2007a)	http://msb.embopress.org/content/3/1/152.long#DC10
Uniprot	Gén ontológia, fordítás	(The UniProt Consortium, 2014)	http://www.uniprot.org/
Ensembl	Szövettani lokalizáció	(Flicek és mtsai, 2014)	http://www.ensembl.org/
Oreganno	Transzkripciós faktorok és célgénjei	(Griffith és mtsai, 2007)	http://www.oreganno.org/
JASPAR	Transzkripciós faktor pozíciós súlyozott mátrix	(Mathelier és mtsai, 2015)	http://jaspar.genereg.net/
PAZAR	Transzkripciós faktorok és célgénjei összesített adatbázis	(Portales-Casamar és mtsai, 2009)	http://www.pazar.info/
HTRI	Humán transzkripciós faktor -célgén adatbázis	(Bovolenta és mtsai, 2012)	http://www.lbbc.ibb.unesp.br/htri/
OrthoDB	Orthológiai csoportosítás,	(Kriventseva és mtsai, 2015)	http://orthodb.org/
Inparanoid	Izoforma csoportosítás	(Sonnhammer és Östlund, 2015)	http://inparanoid.sbc.su.se/cgi-bin/index.cgi
Mirbase	Általános miRNS információ	(Kozomara és Griffiths-Jones, 2011)	http://www.mirbase.org/
miRTarBase	Kísérletes miRNS célpont információk		http://mirtarbase.mbc.nctu.edu.tw/

Adatbázis	Tulajdonságok	Referencia	Elérhetőség
TarBase	Kísérletes miRNS célpont információk	(Vergoulis és mtsai, 2012)	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index
miRDeathDB	Sejthalállal kapcsolatos kézzel gyűjtött miRNS célpontok	(Xu és Li, 2012)	http://www.rna-world.org/mirdeathdb/
Targetscan	Prediktált miRNS adatbázis	(Friedman és mtsai, 2008)(Friedman és mtsai, 2008)	http://www.targetscan.org/
PicTar	Prediktált miRNS adatbázis	(Krek és mtsai, 2005)	http://pictar.mdc-berlin.de/
miRecords	Kézzel gyűjtött és további miRNS kötőhely adatbázisokat integrál	(Xiao és mtsai, 2009)	http://c1.accurascience.com/miRecords/
Putmir 1.1 és 2.0	miRNS-ek transzkripcionális szabályozását tartalmazó adatbázis	(Bandyopadhyay és Bhattacharyya, 2010)	http://www.isical.ac.in/~bioinfo_miu/TF-miRNA/TF-miRNA2.php
TransmiR	Transzkripció faktor célgén adatbázis	(Wang és mtsai, 2010)	http://www.cuilab.cn/transmir
ENCODE	Transzkripció faktor kötőhelyek és egyéb regulációs nemkódoló régiók gyűjteménye	(Dunham és mtsai, 2012)	https://www.encodeproject.org/
Gene Expression Omnibus	Gén expressziókat tartalmazó adatbázis	(Edgar és mtsai, 2002)	http://www.ncbi.nlm.nih.gov/geo/
Cancer Gene Census	Daganatos mutációk adatbázisa	(Futreal és mtsai, 2004)	http://cancer.sanger.ac.uk/census/
ChEMBL	Különféle molekulák és célpontok	(Gaulton és mtsai, 2012)	https://www.ebi.ac.uk/chembl/
Anatomical Therapeutic Classification	Gyógyszerek indikációja	(Berg és mtsai, 2014)	http://www.whocc.no/atc_ddd_index/
Online Mendelian Inheritance in Man	Örökletes betegségek és tulajdonságok adatbázisa	(Amberger és mtsai, 2015)	http://www.omim.org/

3. Paralóg csoportok meghatározása

A paralóg csoportok meghatározásakor a lehető legnagyobb átfedést próbáltuk biztosítani a Signalink jelátviteli adatbázissal. Ehhez két, egymástól független adatbázist töltöttem le, az OrthoDB-t és az InParanoidot (Kriventseva és mtsai, 2015; Sonnhammer és Östlund, 2015). A két adatbázis egymástól részben eltérő módszerrel határozza meg a paralógokat. Mindegyik esetben szükség volt különböző fajokban ortológ keresésére. Ehhez az embertől egy konszenzus evolúciós távolságot határoztam meg. Erre az ember és a többi emlős közötti evolúciós távolságot választottam, mert így olyan paralóg csoportokat kaptam, amik kellően kicsik voltak, és az emberi jelátvitelt tartalmazó Signalink2 adatbázist jól lefedték. Az OrthoDB a legjobb találati arány esetén mondja ki két génről, hogy egymás ortológjai különböző fajokban, és ezután keres paralógokat az adott genomban (Kriventseva és mtsai, 2015). Az OrthoDB-ről letöltött „ODB8_EukOGs_genes_Mammalia-40674.txt” fájlt használtam.

Az InParanoiddal páronkénti BLAST keresést végezhetünk két faj között (Sonnhammer és Östlund, 2015). Vizsgálatomhoz az InParanoidban található emlős fajokat választottam ki. A fehérjék ortológ kapcsolataiból hálózatot építettünk. Ennek a hálózatnak az óriás komponensei alkotják az ortológ csoportokat. A hálózat építő program megírását Fazekas Dávid biológus kollégám végezte.

Ezek után a két, egymást kiegészítő forrást összeadtam. Olyan paralóg csoportokat akartunk létrehozni, melyeknek nagy az átfedése a Signalink adatbázissal. Ehhez meghatároztam egy távolság-értéket, mely azt méri, hogy mennyi az ortológ csoportnak a mérete, és mekkora abból a Signalink adatbázisbeli fehérjék aránya (10. Egyenlet).

$$D = \sqrt{n^2 + \left(\frac{n}{m}\right)^2}$$

10. Egyenlet

D az n és az n/m változó által meghatározott síkon az origótól mért távolság, n a Signalink fehérjék mennyisége az adott paralóg csoportban, m pedig a paralóg csoport mérete. Egy adott fehérjét a legnagyobb D értékű csoportokba soroltuk be. Ha egy fehérje két azonos D értékű csoportba is bekerült, akkor abba soroltuk be, amelyik a legnagyobb mennyiségű Signalink fehérjét tartalmazta (n).

4. Szövetspecifitás a kritikus paralóg csoportokban

A kritikus paralógok vizsgálatakor minden egyes fehérjéhez hozzárendeltük az expresszálandó szöveteket. Ehhez az Ensembl adatbázisból töltöttem le az egyes fehérjék szöveti expressziós tulajdonságait (Flicek és mtsai, 2014). Az Ensembl azonosítókat itt is UniProt azonosítókra fordítottam, azonban az itt talált szöveti csoportosítás túl részletesnek bizonyult, ezért egy általam készített, új osztályozási rendszer alapján összevontam a szövettani csoportokat. Az osztályozási rendszeremet a 3. táblázat tartalmazza. Minden egyes szövetre készítettem egy szövetspecifikus gráfot, ha az adott szövet több, mint a SignaLink fehérjék felét tartalmazta. A 733 SignaLink2 emberi jelátviteli fehérje közül 165 fehérje nem tartozott egyetlen szövethez sem, ezeket kizártam az analízisből. Ennek a hatásnak a kontrolálására minden egyes fehérjét hozzá is adtam a szövetspecifikus gráfokhoz is. Minden egyes szövetspecifikus hálózatnál meghatároztam a ligandoktól a transzkripciós-faktorokig tartó utakat. Ezután, *in silico* kiütöttem egy-egy fehérjét, és néztem, hogy ugyanannyi vagy kevesebb lesz-e a ligandoktól a transzkripciós faktorokig menő utak száma. Ha csökkent az utak száma, akkor az adott fehérje az adott szövetben részt vesz egy esszenciális útban. Az egyetlen szövetbe sem sorolt fehérjékkel és azok nélkül is elkészítettem a gráfokat.

3. táblázat A módosított szövettani/anatómiai csoportosítás.

Az eredeti forrás és az új csoportosítás is angol nyelven készült, és az egyértelműség miatt ezt itt megtartom. Az utolsó sorban található a nem kategorizált szövetek csoportja, amely azon szövettani csoportokat tartalmazza, amelyeket nem vettünk bele az elemzésbe.

Régi szövettani csoportok	Új szövettani csoport
cerebellum cortex, brain, cerebral cortex, subthalamic nucleus, frontal lobe, corpus striatum, globus pallidus, corpus callosum, cerebrum, pons, substantia nigra, cerebellum, amygdala, parietal lobe, temporal lobe, hippocampus, medulla oblongata, hypothalamus	CNS (central nervous system)
artery, vein, aorta, arterial adventitia, heart, myocardium, atrium	Cardiovascular
nasopharynx, bronchus, larynx, alveolus, trachea, lung, nose	Respiratory
gall bladder, liver, bile duct	Liver and gall
thyroid, islets of Langerhans, adrenal gland, adrenal medulla, adrenal cortex, pituitary gland, parathyroid	Endocrine
breast, mammary gland	Breast
gum, parotid gland, salivary gland, oral cavity, submandibular gland, tongue, hypopharynx	Oral
synovium, muscle, skeletal muscle, bone, cartilage	Musculoskeletal
kidney, ureter, bladder, urinary	Urine
bone marrow, blood, lymph, lymph node, thymus, spleen, lymphoreticular, tonsil	Lymphatic and hematologic
ovary, uterus, myometrium, endometrium, cervix	Female genitalia
cochlea, vestibule, internal ear, middle ear	Ear
rectum, colorectal, pharynx, small intestine, duodenum, ileum, colon, stomach, pancreas, oesophagus, intestine	Alimentary
visual apparatus, cornea, lacrimal gland, lens, retina, macula lutea, iris, optic nerve, fovea centralis, ciliary body, trabecular meshwork	Eye
testis, epididymis, prostate, penis, seminal vesicle, foreskin	Male genitalia
skin, epidermis,	Skin
nervous, peripheral nerve, spinal ganglion, sympathetic chain, ganglion	Peripheral nervous system
placenta, developmental, amnion, choroid, amniotic fluid, trophoblast, umbilical cord	Developmental
smooth muscle, greater omentum, unclassifiable (Anatomical System), meninges, head and neck, Anatomical System (egenetics), motor, peritoneum, germinal center, whole body, pia mater, dura mater, mesenchyma	<i>Nem kategorizált</i>

5. *Paralóg csoportok tulajdonságainak specificitási vizsgálata*

A kritikus paralógok vizsgálatakor szükséges volt meghatározni, hogy mennyire hasonlóak egymáshoz a paralóg csoportokon belül az egyes fehérjék. Ehhez a biológiai funkciót, a regulációt (miRNS és transzkripciós faktor) és az örökletes betegségeket vettem alapul. Első lépésben meghatároztam, hogy melyek azok a tulajdonságok, amelyek nem találhatóak meg minden egyes fehérjében az egyes csoportokon belül. Ezeket részben specifikus tulajdonságnak neveztem el. Megszámoltam, hogy hány ilyen tulajdonság van egy csoportban, és visszaosztottam a csoport nagyságával, hogy ezzel normáljam a nagyobb méretű fehérjecsoportokat. Sajnos, az egyes fehérjék kutatottsága még így is befolyásolhatta a vizsgálatokat, különösen a Gene Ontology elemzéseknél (Geeleher és mtsai, 2013; Huntley és mtsai, 2015).

6. *Microarray elemzés és a differenciáltan expresszálódó gének meghatározása az első szomszédok vizsgálatakor*

Vizsgálatunkhoz a Gene Expression Omnibus (Barrett és mtsai, 2013) adatbázison a 2014 augusztusáig közzé tett valamennyi Affymetrix HGU133 plus 2 típusú colorectalis carcinoma, hepatocellularis carcinoma, mell daganat illetve nem kissejtes tüdő carcinomát tartalmazó chipet összegyűjtöttük Kubisch János mesterszakos biológus hallgatóval közösen. A chipek GEO azonosítóját az 1. mellékletben közlöm. Kereső szavaink „*colon cancer*”, „*liver cancer*”, „*breast cancer*” és „*lung cancer*” voltak. Ezekből válogattuk ki később a különféle daganatokat. Mellé egészséges páciensekből származó kontroll chipeket is gyűjtöttünk. Összesen 1557 microarray chip adatait dolgoztuk fel. Ezek megoszlását mutatja a 4. táblázat.

4. táblázat A vizsgált microarrayek megoszlása az egészséges és daganatos betegek között

	Egészséges	Daganatos
Colorectalis carcinoma	44	486
Mell daganat	162	143
Hepatocellularis carcinoma	30	143
Nem kissejtes tüdőrák	45	504

A vizsgálathoz valamennyi chipet egymáshoz normáltunk az RMA¹³ (*Robust Multi-array Average*) (Irizarry, 2003) módszer segítségével. A megfelelő microarray normalizáló technika kiválasztásában és az chippek feldolgozásában dr. Marczell István az I. Belgyógyászati Klinika munkatársa volt segítségemre.

A kapott expressziós értékeket probesettenkénti¹⁴ maximummal kötöttem UniProt azonosítóhoz. Ezután az egyes hálózatokra leszűrtem a géneket. A szűrt, normált adatokból állapotspecifikus átlagokat (normál, daganatos) készítettem és ezen átlagok alapján az átlagtól több mint egy szórásnyi értékkel alacsonyabb expressziós értéket felvevő géneket „nem expresszálnak” tekintettem.

Ha egy gén expresszáldott az adott szövetben és az adott hálózatban az egyik állapotban például a daganatban de nem expresszáldott a másokban pl. a normál szövetben akkor differenciáltan expresszáldónak tekintettem, függetlenül attól, hogy mennyire volt a meghatározott 1 szórásnyi határértéktől. Ez a módszer lehetővé tette a diszkrét elemzését az expresszáldott géneknek.

7. Használt programok

A hálózat ábrákat a Cytoscape (Shannon és mtsai, 2003) segítségével készítettem. A violinplotokhoz R „vioplot” (Adler, 2015) a további ábrákhoz a Python „matplotlib” paketet használtam (Hunter, 2007). A számításokhoz a Python NumPy (van der Walt és mtsai, 2011), és ScipyPy paketet használtam (Jones és mtsai, 2001). Mindig a gráf óriás komponensét vizsgáltam. A centralitási értékeket a Python Igraph (Csardi és Nepusz, 2006) pluginje segítségével határoztam meg.

¹³ Az RMA módszer során az egyes microarrayeket egymáshoz normáljuk. Az intenzitás minden esetben logaritmikus. Az általam elemzett Affymetrix microarray chipen egy probeset 24 probe (minta) tartozik, melyek közül 12 valódi kötődéssel rendelkezik 12 pedig mismatch probe, azaz a középső nukleotid el van rontva. A microarrayek normálása során, egy chipen belüli egy probesetre számítják ki az expresszió értékeit. Az RMA esetén úgy, hogy összehasonlíthatóak legyenek egymással is az értékek a teljes vizsgálat során. A normálás lépései: 1. háttérkorrekció azzal a feltételezéssel, hogy a jel exponenciális eloszlású, míg a zaj Gauss eloszlást mutat a chipen. 2. kvantilis normalizáció az arrayek között: Minden chipen a probokat sorba rendezik az intenzitás szerint. A chippek között kiszámolják az azonos rangú probok intenzitásának átlagát. Ez az átlag lesz az egymáshoz normált intenzitás. 3. Median polish alapú probeset szummálás: Ennek lényege, hogy egy adott probesetnek az adott chipen található probok mátrixának vesszük a sor és oszlop mediánjait. Ezeket kivonjuk az eredeti mátrixból. Ezt addig ismételjük, amíg a sor és oszlop mediánok mindegyike 0 lesz, vagy maximum 5-ször. Az így kapott reziduális mátrixot kivonjuk az eredeti értékekből és végül az átlagát vesszük a probesetnek tartozó 12 probe intenzitásnak. Ez lesz az adott probeset expressziója.

A módszer hátránya hogy meglehetősen memória igényes mivel az összes chip adatát egyszerre kell tárolni a számítógép memóriájában.

¹⁴ Egy génhez viszont több probeset is tartozhat. Én ezek közül a maximálisat vettem alapul.

8. *Statisztikai módszerek*

A statisztikai vizsgálatokhoz Kolmogorov-Szmirnov és Wilcoxon féle rang próbát használtam, mivel a hálózati paraméterek skála független eloszlást mutatnak (Barabasi és Albert, 1999). A Wilcoxon-rang próba rang alapján méri egymáshoz a vizsgált két minta mediánját, és azzal a null hipotézissel él, hogy a két minta mediánja megegyezik (Fay és Proschan, 2010; Mann és Whitney, 1947). A Kolmogorov-Szmirnov teszt a minta eloszlásának (eloszlás függvényének) az alakjára érzékeny, azaz azt vizsgálja, hogy a két minta ugyanazon típusú eloszláshoz tartozik-e (Kolmogorov, 1933; Smirnov, 1948). A legtöbb vizsgálat esetén mindegyik tesztet elvégeztem. A kontroll minden esetben (ahol máshogy nem említtem a dolgozat során) a teljes hálózat volt, ehhez viszonyítottuk az adott csoportot.

A dúsulási vizsgálatokhoz (*enrichment analysis*) hipergeometrikus tesztet végeztem. A disszertációban csupán a Benjamini-Hochberg által végzett (Hochberg, 1995) hamis találati arány (FDR, *false discovery rate*) általi korrekcióval készült értékeket mutatom be. A hipergeometrikus teszthez használható programot Ari Esztertől az Eötvös Loránd Tudományegyetem Genetikai Tanszékének munkatársától kaptam.

A gráfokban mért óriás komponensek nagyságát statisztikai mintavétellel történő Z score analízissel határoztam meg. Ennek lényege, hogy az adott gráfból kiválasztottam n darab pontot (n a vizsgált típusú pont mennyisége volt), és megvizsgáltam az n pont összekötöttségét. Ezt megismételtem 1000-szer, és az ebből számolt eloszlás átlagához és szórásához viszonyítottam a minta értékét. A minta átlagtól való két szórásnál nagyobb eltérést tekintettem szignifikánsnak. A kapott érték egyfajta Z score (11. Egyenlet).

$$Z = \frac{k_{GC} - \bar{k}_{GC}}{SD_{k_{GC}}}$$

11. Egyenlet

Ahol adott G gráfon k_{GC} a G gráf k elemű ponthalmaznak óriáskomponensbeli pontjainak a száma, \bar{k}_{GC} a G gráfon k darab véletlenszerűen kiválasztott pont óriás komponensen belüli pontjainak számának az átlaga, $SD_{k_{GC}}$ a k véletlenszerűen kiválasztott pont óriás komponensbeli pontjainak számának szórása. A mérőszám függ a gráftól és a k kiválasztott pontok számától.

V. Eredmények és az eredmények értékelése

1. Az Autofágia Regulációs Adatbázis az NRF-ome és a SignaLink jelátviteli adatbázis regulációs rétegei

Mindkét adatbázisban a regulációs adatok integrálását végeztem. A két adatbázisban ez csak a cél fehérjék szintjén tér el, ezért egyben mutatom be a végzett munkát.

1.1. A transzkripció faktor és célgén adatok integrálása:

A transzkripcionális elemek integrálásához kidolgoztam egy spektrumanalízishez hasonló eljárást. A transzkripció faktorokat kigyűjtöttem a JASPAR adatbázisból, és a pozíció súlyozott mátrixokból számított érték segítségével meghatároztam, mely szekvenciákhoz kötődhetnek az egyes gének. A hamis pozitív értékeket lehetőség szerint minimalizáltam (ld. módszertani fejezet).

Emellett munkatársaimmal integráltuk a publikáláskor elérhető nyilvános transzkripció faktor célgén adatbázisokat is (5. táblázat). A kapcsolatokat minden esetben szűrtem az adott adatbázis célfehérjéire, tehát a SignaLink esetében a jelátviteli fehérjékre, az ARN esetében pedig az autofágiában részt vevő fehérjékre. Az integrálás eredményét az 5. táblázat tartalmazza.

1.2. A miRNS adatok integrálása

A miRNS adatokat egyéb adatbázisokból integráltuk a SignaLink adatbázisba. Nem végeztünk önálló predikciókat. A kapcsolatok számát az 5. táblázat tartalmazza. Az elkészült adatbázisok részletes elemzését Türei Dénes kollégám doktori értekezése tartalmazza.

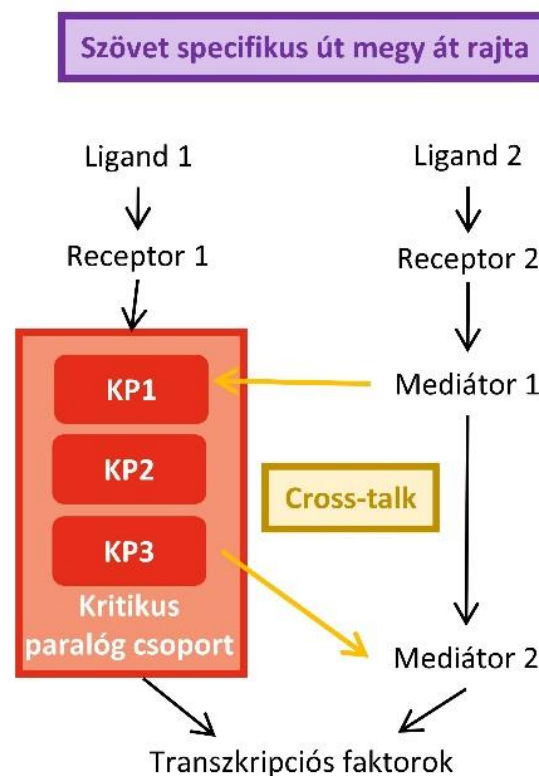
5. táblázat Az integrált transzkripciós faktor célgén adatbázisok miRNS célgén és transzkripciós faktor célgén adatbázisok a SignaLink 2 NRF2ome és az ARN adatbázisban.

Érdeemes megnézni, hogy az általam készített jósló algoritmus egy nagyságrenddel több TF-célgén kapcsolatot prediktál, mint a kézi gyűjtésű adatbázisok – JASPAR sor. A miRNS célgének esetén az eltérés még határozottabban igaz a predikciós algoritmusok irányába (doRiNA, DianaMicroT, miRanda, PicTar, TargetScan).

Forrás adatbázis	Integrált pontok száma			Kapcsolatok darabszáma		
	SignaLink	NRF2ome	ARN	SignaLink	NRF2ome	ARN
Transzkripciós faktor célgén adatbázisok						
ABS	21	2	0	21	1	0
edgeDB	523			259		
ENCODE	636	115	350	2475	186	1064
HTRI	1911	119	141	6.817	293	351
JASPAR	2567	6329	316	23098	8217	2911
ORegAnno	297	41	27	301	35	24
PAZAR	665	94	106	1032	100	111
RedFly	136			179		
miRNS célgén adatbázisok						
miR2Disease	226		84	199		75
doRiNA	13412		1073	277792		15.688
DianaMicroT	8147		885	144982		24446
miRDeathDB	204		54	458		47
miRanda	22068	258	1102	4297167	2206	173446
PicTar	6789		332	32396		1520
TargetScan	12344	538	1354	2627654	5947	126068
TarBase	857	8	75	905	4	62
TF-mikro-RNS adatbázisok						
ENCODE	218	190	253	770	665	806
PuTmiR 1.1	521	460	328	4211	4067	969
PuTmiR 2.0			123			179
TransmiR	183	163	165	246	250	238

2. A kritikus paralógok meghatározása a jelátvitelben

Munkám során a szövetspecifikus interakciókból indultam ki. Azt feltételeztem, hogy azok a fehérjék lehetnek nagyobb eséllyel esszenciálisak vagy kritikusak, amelyek elvesztése egy adott szövetben egy jelátviteli út (egy ligandtól a transzkripciós faktorig tartó kaszkád) elvesztésével jár. A fenti definíciót kiegészítettem azzal, hogy az evolúció során feltehetően az ilyen fehérjék duplikálódhattak, hiszen evolúciós előnnyel járhat, ha egy kritikus funkciót több hasonló vagy ugyanolyan fehérje lát el. Ez alapján alkottam meg a kritikus paralóg csoportok (KPCS) definícióját. (7. ábra)



7. ábra A kritikus paralóg csoportok definíciója

KP kritikus paralóg (Módos és mtsai, 2016)

Egy kritikus paralóg csoport olyan fehérjecsoport, mely

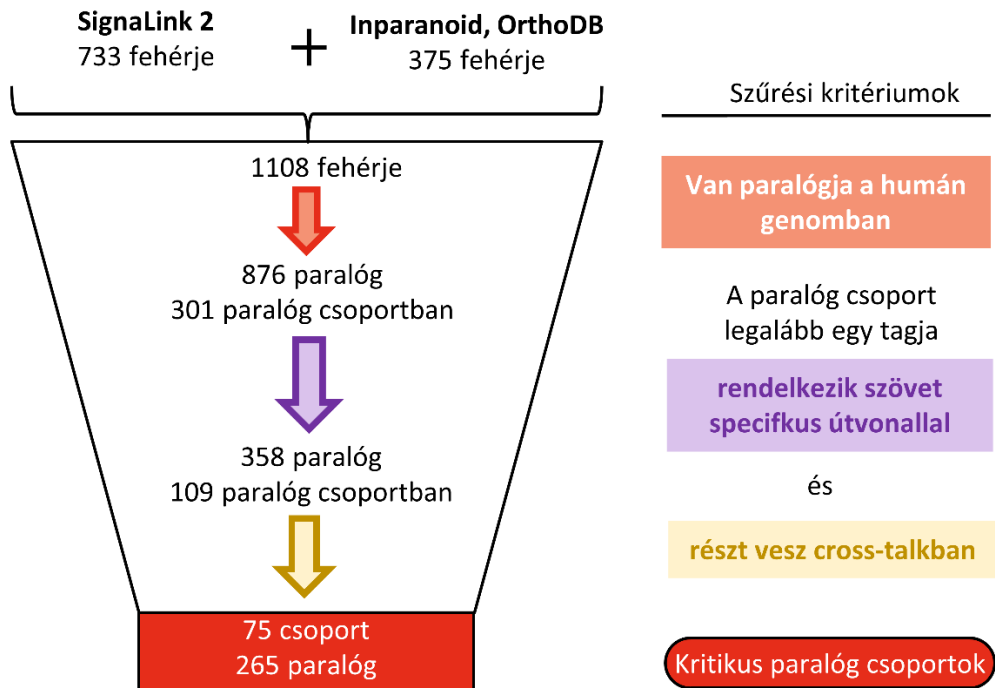
1. rendelkezik paralógokkal a jelátvitelben (evolúciós kritérium)
2. legalább egy tagján átmegy egy szövetspecifikus jelátviteli út (szövetspecifikus gráf elméleti kritérium)
3. legalább egy tagja cross-talkban vesz részt (jelátviteli-biológiai kritérium)

Ezen három kritérium elegendő volt arra, hogy azonosíthassam hét jelátviteli útvonal kritikus paralógjait, és kiterjesszem Kahn és munkatársai az inzulin jelátvitelben végzett kritikus pontokra vonatkozó kutatását (Taniguchi és mtsai, 2006), ami a vizsgálataink kiindulópontját képezte.

A definiálás után a kritikus paralóg csoportok meghatározására egy olyan munkafolyamatot hoztam létre, mely tartalmaz mind jelátviteli, mind evolúciós, mind pedig gráfelméleti paramétereket (7. ábra). A munkámhoz Kahn és munkatársai anyagát vettem alapul és kiterjesztettem további jelátviteli útvonalakra (Taniguchi és mtsai, 2006).

A munkafolyamatot a 8. ábra ismerteti.

1. **Paralóg csoportok a jelátvitelben – evolúciós kritérium:** A Signalink adatbázisban található jelátviteli útvonalakhoz annotált fehérjékből paralóg csoportokat alkottam, felhasználva az Inparanoid és az OrthoDB ortológ csoportosításait, a Módszerek fejezetben leírtaknak megfelelően. Ez 301 darab paralóg csoportot és 876 paralóg fehérjét eredményezett.
2. **Kritikusság a jelátvitelben – szövetspecifikus gráfelméleti kritérium:** A szövettani csoportosítást felhasználva szövetileg specifikus hálózatokat hoztam létre. Ezekben a hálózatokban megvizsgáltam, mely fehérjék azok, amelyek nélkül egy ligandtól a transzkripciós faktorig terjedő út elvész. Azok a paralóg csoportok mentek át ezen a szűrési kritériumon, melyeknek legalább egy ilyen eleme volt. Így összesen 109 paralóg csoportot kaptam 358 jelátviteli fehérjével.
3. **Cross-talk – jelátvitel specifikus kritérium:** A Signalink jelátviteli adatbázis alkalmas arra, hogy az útvonalak közti cross-talkokat felderíthessük. A szűrésen azok a csoportok mentek át, melyek legalább egy tagja összekapcsolt két különböző jelátviteli útvonalat.



8. ábra a kritikus paralógok azonosítása (Módos és mtsai, 2016)

A három szűrési lépés alapján végezetül 75 kritikus paralóg csoportot kaptam az emberi jelátviteli hálózatban, melyben 265 kritikus paralóg (KP) foglalt helyet. Ezek közül 168 származott a Signalink adatbázisból és 97 fehérje volt egy Signalink fehérje paralógja. A dolgozat 2. mellékletében megtalálható táblázat közli az összes kritikus paralóg csoportot és a kritikus paralógokat. A további összehasonlításokhoz a nem kritikus paralógokat tovább bontottam paralóg fehérjékre – olyan jelátviteli fehérjékre melyeknek volt paralógja a humán genomban – és egyéb jelátviteli fehérjékre. A paralóg fehérjék (PF) a dolgozat hátralevő részében a nem kritikus paralóg csoportokat (nKPCS) alkotják. Összesen 226 nKPCS volt 661 PF-vel. Egyéb Signalink fehérje még 232 volt a vizsgálatunkban.

Mivel nem minden fehérje rendelkezett szövetspecifikus annotációval, ezért megismételtem a vizsgálataimat úgy is, hogy a szöveti annotációval rendelkező fehérjéket nem adtam hozzá egyetlen szövet specifikus hálózathoz sem, illetve úgy is, hogy minden szövet specifikus hálózathoz hozzáadtam őket. Ugyanarra az eredményre jutottam és ugyanazokat a kritikus fehérjéket állapítottam meg. A szövetspecifitás megállapítására bármilyen más módszer is alkalmas lehet (például a microarray módszerek). Az általam felállított munkamenet egyszerűen újra futtatható, ha bármelyik adatbázis frissül.

2.1. A kritikus paralógok centrálisak az emberi jelátvitelben

Négy különböző hálózati paramétert vizsgáltam meg a SignLinkben megtalálható fehérjék esetén. Ezek a fokszám (*degree*), a köztiség (*betwenness*), a csokornyakkendőség (*bowtiness*) és a közelség (*closeness*) voltak. Összehasonlítottam egymással a kritikus paralógokat a paralóg fehérjékkel és az egyéb fehérjékkel. A kritikus paralógok a hálózatban lokálisan fontos szerepet játszanak a magas fokszámuknak köszönhetően (medián KP: 6, PF: 2 Egyéb:2 Kolmogorov-Szmirnov teszt, Wilcoxon rang próba $p < 0,001$). Ezek magasabb köztiséggel rendelkeztek, mint az egyéb fehérjék és a paralóg fehérjék, azaz jobban összekötik a hálózat egymástól távol eső részeit (mediánok rendre 264, 0, 0 Kolmogorov-Szmirnov teszt, Wilcoxon rang próba $p < 0,001$). Mindez nem csak az általánosan használt globális hálózati paraméterrel, a köztiséggel mérve igaz, hanem a jelátvitelre specifikus csokornyakkendőség esetén is (medián rendre 0,01, 0,0 0,0 Kolmogorov-Szmirnov teszt, Wilcoxon rang próba $p < 0,001$). Magasabb közelséggel rendelkeztek, tehát inkább a jelátviteli hálózat belsőbb régiójában foglaltak helyet (medián rendre: 0,0643, 0,0636, 0,06328 Kolmogorov-Szmirnov teszt, Wilcoxon rang próba $p < 0,001$). Összességében a kritikus paralógok mind a négy mért centralitásban magasabb értéket vesznek fel, mint az egyéb fehérjék vagy a paralóg fehérjék (9. ábra Wilcoxon rang próba $p < 0,001$, Kolmogorov-Szmirnov teszt $p < 0,001$).

Az eltérő értékek mellett a centralitás mérőszámok görbéjének alakja is eltért (Kolmogorov-Szmirnov teszt $p < 0,001$), tehát különböző típusú fehérjék tartoznak az egyes csoportokba. A magasabb centralitás, több funkciót jelenthet, és nagyobb eséllyel tartozik letális fehérjékhez (Jeong és mtsai, 2001).

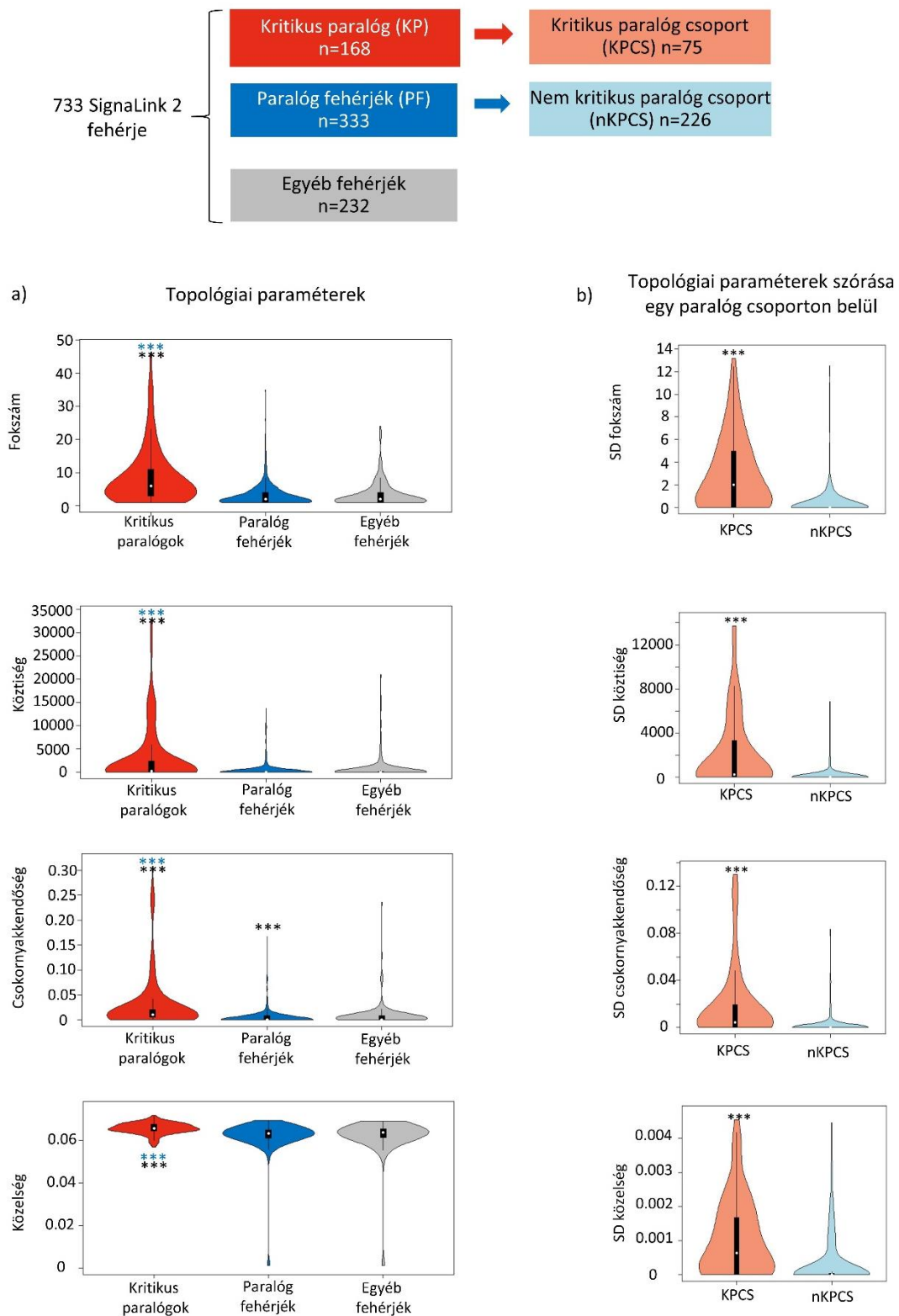
Nemcsak a paraméterek értékére voltunk kíváncsiak, hanem arra is, hogy ezek mennyire változékonyak az egyes csoportokon belül. Ezért kiszámoltuk az egyes több elemű paralóg csoportokon belül a paraméterek szórását. Azt kaptuk, hogy a kritikus paralóg csoportoknak magasabb volt a szórása az egyes paraméterekben, mint a nem kritikus paralóg csoportoknak (Wilcoxon rang próba $p < 0,001$, Komogorov- Szmirnov teszt $p < 0,001$ 9. ábra, 6. táblázat).

6. táblázat A Signalink jelátviteli hálózaton mért centralitási értékek szórása az egyes kritikus paralóg csoportokon (KPCS) illetve nem kritikus paralóg csoportokon (nKPCS) belül mérve.

A medián 25 és 75 percentilis értékek.

Szórás egy paralóg csoporton belül	nKPCS			KPCS		
	25%	Medián	75%	25%	Medián	75%
Fokszám	0	0	0	0	2	4,97
Köztiség	0	0	0	0	223	3335
Közelség	0	0	0,000034	0	0,000644	0,00170
Csokornyakkendőség	0	0	0	0	0,00444	0,0193

Tehát egy kritikus paralóg csoporton belül eltérő központiságú fehérjék találhatóak. **Az, hogy a kritikus paralógok magasabb és egy csoporton belül eltérő centralitással rendelkeznek, azt implicálja, hogy a hálózatban a kritikus paralógok nem felcserélhetők egymással, és felhetően a funkciójuk különbözik. Ezzel szemben a nem kritikus paralóg csoportokon belüli hasonló és alacsony hálózati centralitások azt mutatják, hogy a paralóg fehérjék funkciója felcserélhető.** Ennek igazolására funkcionális annotációs vizsgálatokat végeztem.



9. ábra A kritikus fehérjék hálózatos topológiai tulajdonságai

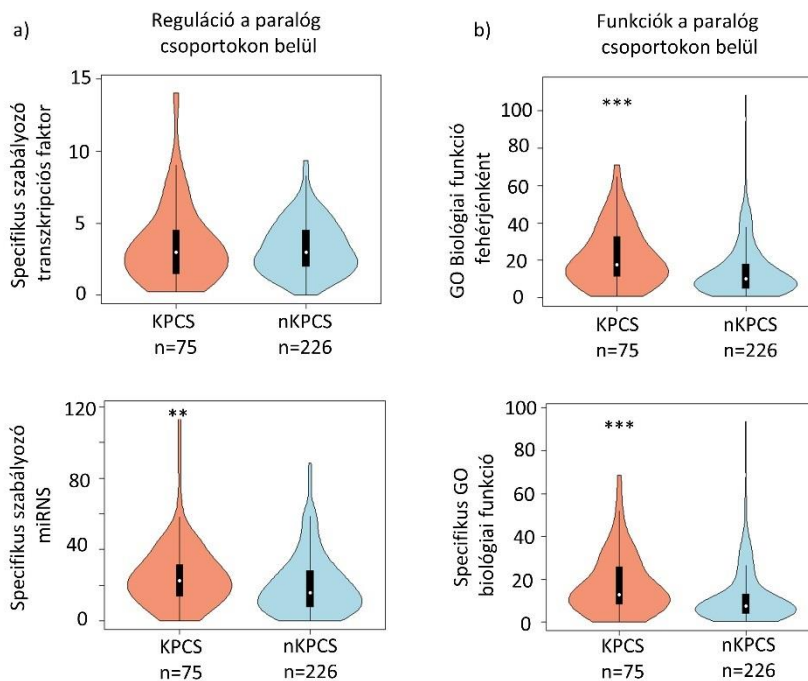
a) Az egyes topológiai paraméterek különböző típusokban fehérjénkénti eloszlása b) A különböző topológiai paraméterek szórása ***: $p < 0,001$ Wilcoxon rang próba. Fekete csillagok az egyéb fehérjékhez képest mutatják a statisztikailag szignifikáns eltérést, a

kék csillagok a paralóg fehérjék és a kritikus paralógok közötti eltérést jelzik. (Módos és mtsai, 2016)

2.2. Specifikus reguláció és diverz funkció a kritikus fehérjéken belül

Az egymáshoz hasonló fehérjéknek szüksége van valamilyen regulációs különbségre. Ehhez megnéztem, hogy az egyes fehérjékhez hány darab miRNS illetve hány darab transzkripciós faktor tartozik, és azok közül hány darab legalább részben specifikus egy paralóg csoport egy tagjához viszonyítva. A vizsgálatom alapján a 301 KPCS és nKPCS közül 294-ben volt valamilyen, legalább részben specifikus reguláció. Kíváncsi voltam, hogy van-e eltérés a kritikus paralóg csoportok és a nem kritikus paralóg csoportok regulációjában. Transzkripcionális szinten mind a két csoport hasonlóan specifikusan regulálódik: a specifikus transzkripciós faktorok számában nincs különbség (medián KPCS 8 és nKPCS 7 10. ábra/a Kolmogorov-Szmirnov teszt, Wilcoxon rang próba $p > 0,05$). Poszt-transzkripcionálisan viszont a reguláció eltér. A legalább részben specifikus miRNS-ek száma a kritikus paralóg csoportokban szignifikánsan magasabb a nem kritikus paralóg csoportokhoz képest (medián rendre 65, 42 10. ábra/a Kolmogorov-Szmirnov teszt, Wilcoxon rang próba $p < 0,01$).

Az előző fejezetben említett hálózati központosság nem feltétlenül jelent több funkciót is. Ennek ellenőrzésére a Gene Ontology adatbázis biológia funkció (*biological process*) doménjét használtam fel. Megvizsgáltam, hogy fehérjénként hány biológiai funkció tartozik egy kritikus fehérjéhez. A kritikus paralógok esetén ez az érték szignifikánsan magasabb, összehasonlítva a paralóg fehérjékkel, illetve az egyéb fehérjékkel (medián rendre 15, 13, 12 Wilcoxon rang próba, Komogorov-Szmirnov teszt $p < 0,001$). Arra is kíváncsi voltam, hogy ezek a funkciók specifikusak-e egyes fehérjékre, vagy ugyanazokat a funkciókat tartalmazzák-e az egyes fehérjék egy paralóg csoporton belül. Ehhez meghatároztam, hogy egy paralóg csoportban hány darab, részben specifikus funkció van, ezt normáltam a csoport nagyságával. A kritikus paralóg csoportokban ez a szám szignifikánsan magasabb, mint a nem kritikus paralóg csoportokban lévő fehérjéknél (rendre 39 illetve 19 Wilcoxon rang próba $p < 0,001$ 10. ábra/b).

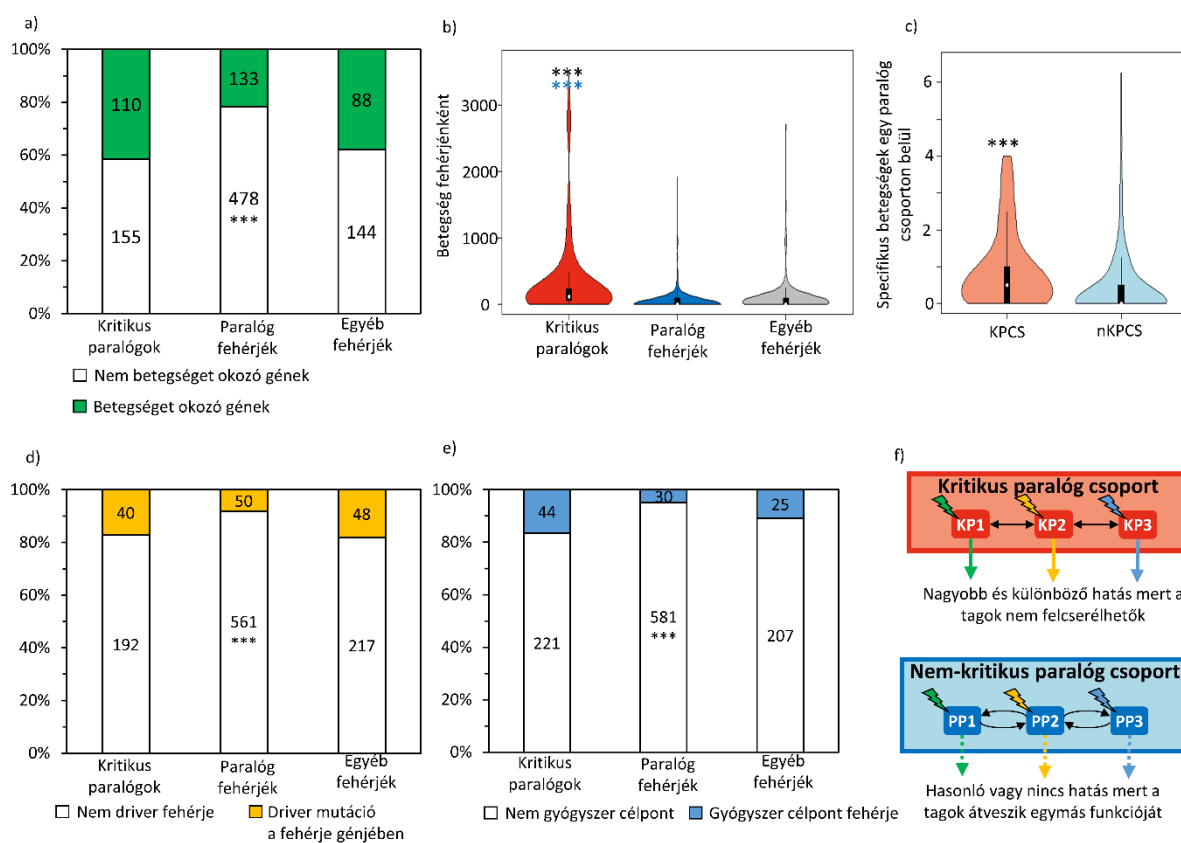


10. ábra A kritikus paralóg csoportok (KPCS), a nem kritikus paralóg csoportok (nKPCS) regulációja a) és biológiai funkciói a Gene Ontology adatbázis alapján b)(Módos és mtsai, 2016)

2.3. A kritikus paralóg csoportok betegséget okozó hatása és gyógyszerészeti relevanciája

Megvizsgáltam a kritikus paralóg csoportok betegségekkel való összefüggését. Megnéztem, hogy vajon a kritikus paralógok génjei között több-e az öröklődő betegségben mutálódó gén, illetve a daganatot okozó mutáció a Cancer Gene Census (Futreal és mtsai, 2004) adatbázis szerint. Azt kaptam, hogy a kritikus paralógok között a betegséget okozó gének ugyanolyan mértékben fordulnak elő, mint a nem érintett fehérjék között. (Rendre 41% és 37 % Khi négyzet teszt $p > 0,05$ 11. ábra/a). Viszont azt találtam, hogy mind a két csoportban gyakoribbak a betegséget okozó gének, mint a paralóg fehérjék között (21% a 41% vagy 37% ellen Khi négyzet teszt $p < 0,001$) (11. ábra/a). A kritikus paralógokra több különböző öröklődő betegség jutott fehérjénként, mint az egyéb fehérjékre vagy a paralóg fehérjékre (mediánok KP: 117, PF: 0, Egyéb:0 Wilcoxon rang próba, Kolmogorov-Szmirnov teszt $p < 0,001$ 11. ábra/b). Ahhoz, hogy megállapítsam, hogy ezek a betegségek mennyire specifikusak az egyes kritikus fehérjékre, elvégeztem a betegségekre is ugyanazt a specifikitási analízist, mint a

regulációra, illetve a biológiai funkciókra tettem. A betegségek specifikusabbak voltak az egyes kritikus paralógokra egy kritikus paralóg csoporton belül, mint a paralóg fehérjékre egy paralóg csoporton belül (medián 0 és 0,5 Wilcoxon rang próba, Kolmogorov-Szmirnov teszt $p < 0.001$ 11. ábra/c). A megtalált hasonló betegséget okozó gén arány a kritikus paralógok és az egyéb jelátviteli fehérjék között azt jelzi, hogy ezek a fehérjék specifikus, nem redundáns jelátviteli szerepet töltenek be. A kritikus fehérjék között megjelenő betegségek magasabb száma mögött a cross-talkban betöltött szerepük állhat. A paralóg fehérjék esetén a mutáció hatását legalább részben tudja ellensúlyozni a másik paralóg fehérje ugyanabból a paralóg csoportból (11. ábra/f). Ezért kevesebb ott a betegséget okozó gén, mint a kritikus paralógok és az egyéb fehérjék között.



11. ábra A kritikus paralóg csoportok tagjai nem felcserélhetők a betegségekben, míg a paralóg csoportok tagjai igen.

a) Az egyes fehérje csoportok öröklődő betegségekben való részvétele az OMIM adatbázis alapján, b) Az egy fehérjére jutó betegségek száma, c) Az egy csoporton belüli betegségek száma d) Az egyes fehérjék valamilyen daganatban driver mutációt tartalmazó génjeinek megoszlása a CGC adatbázis alapján, e) A jelenlegi

gyógyszercélpontok megoszlása a ChEMBL adatbázis alapján, f) A betegségek, mutációk és gyógyszerek hatása a kritikus paralóg csoportokra (villám jelek) és a nem kritikus paralóg csoportokra. A kritikus paralóg csoportokban a kritikus paralógok (KP) hatása nem felcserélhető (egymással szembe fordított nyíl), míg a nem kritikus paralóg csoportokban az egyes tagok célzás hatásra átveszik egymás hatását (nyilak). A kategorikus változók között Khi négyzet tesztet használtunk (a,d,e), a folyamatos változók között Wilcoxon rang próbát (b, c). A szignifikanciát az alábbiak szerint jeleztem: *** $p < 0,001$. A kék csillagok a paralóg fehérjékhez képesti eltérést jelzik, a feketék a nem érintettekhez és a kritikus fehérjékhez képest. (Módos és mtsai, 2016)

Megvizsgáltam azt is, hogy a daganatot okozó, *driver* mutációt elszenvedő gének között a kritikus paralógok aránya hogyan alakul. A *driver* gének alapvetően magasabb centralitással rendelkeznek a fehérje-fehérje interakciós hálózatokban (Xiong és mtsai, 2014), ezért azt vártam, hogy több *driver* gént fogok kapni a kritikus paralógok között, hiszen azok is magas centralitással rendelkeznek (10. ábra). Nem várt eredményként a betegségek megoszlásához hasonló eredményt kaptam. Nem volt különbség a kritikus paralógok és az egyéb fehérjék között (rendre 17,2% és 18,1% Khi négyzet teszt $p > 0,05$ 11. ábra/d). Viszont a paralóg fehérjék között a *driver* gének ritkábban fordultak elő, mint a másik két csoportban (8,2% a 17,2%-al vagy a 18,1%-al szemben Khi négyzet teszt $p < 0,001$ 11. ábra/d). Az oka ennek az eltérésnek feltehetően a Signalink 2.0 adatbázis daganatokra specifikus jelátviteli gyűjtése (Fazekas és mtsai, 2013) és az, hogy a paralóg fehérjék képesek egymás helyét átvenni a jelátvitelben, ezzel a daganatos mutációt legalább részben semlegesíteni (11. ábra/f).

A kritikus paralógok nagy jelátviteli centralitása, nagy mennyiségű különböző biológiai funkciója, a daganatokban való *driver* génként való részvétele potenciális gyógyszercélponttá teszi, hiszen kiterjedt hatást képesek létrehozni a sejten belül (Csermely és mtsai, 2013). Ezért megkerestem a ChEMBL adatbázisban (Bento és mtsai, 2014), hogy hány darab gyógyszercélpont található az egyes fehérjék között. Azt kaptam, hogy a kritikus fehérjék gyakrabban lesznek gyógyszercélpontok, mint a paralóg fehérjék (16% a 4%-kal szemben Khi négyzet teszt $p < 0,001$ 11. ábra/f). Az előzőekhez hasonlóan ennek oka az lehet, hogy a kritikus paralógok nem helyettesíthetők egymással, míg a paralóg fehérjék igen. Ezért, ha egy paralóg fehérjét célzunk, nagyobb dózis szükséges,

hiszen több fehérjét kell elérni, hogy ugyanaz a hatás létre jöjjön. Ezzel szemben a kritikus paralógok célzása sokkal hatékonyabb, mivel azok nem helyettesíthetik egymást. A kritikus paralógok célzásához kritikus paralóg specifikus gyógyszerek és megfelelő körültekintés szükséges, mivel a kritikus fehérjék központiak, és ezért sok mellékhatással járhat célzásuk (Csermely és mtsai, 2013; Perez-Lopez és mtsai, 2015).

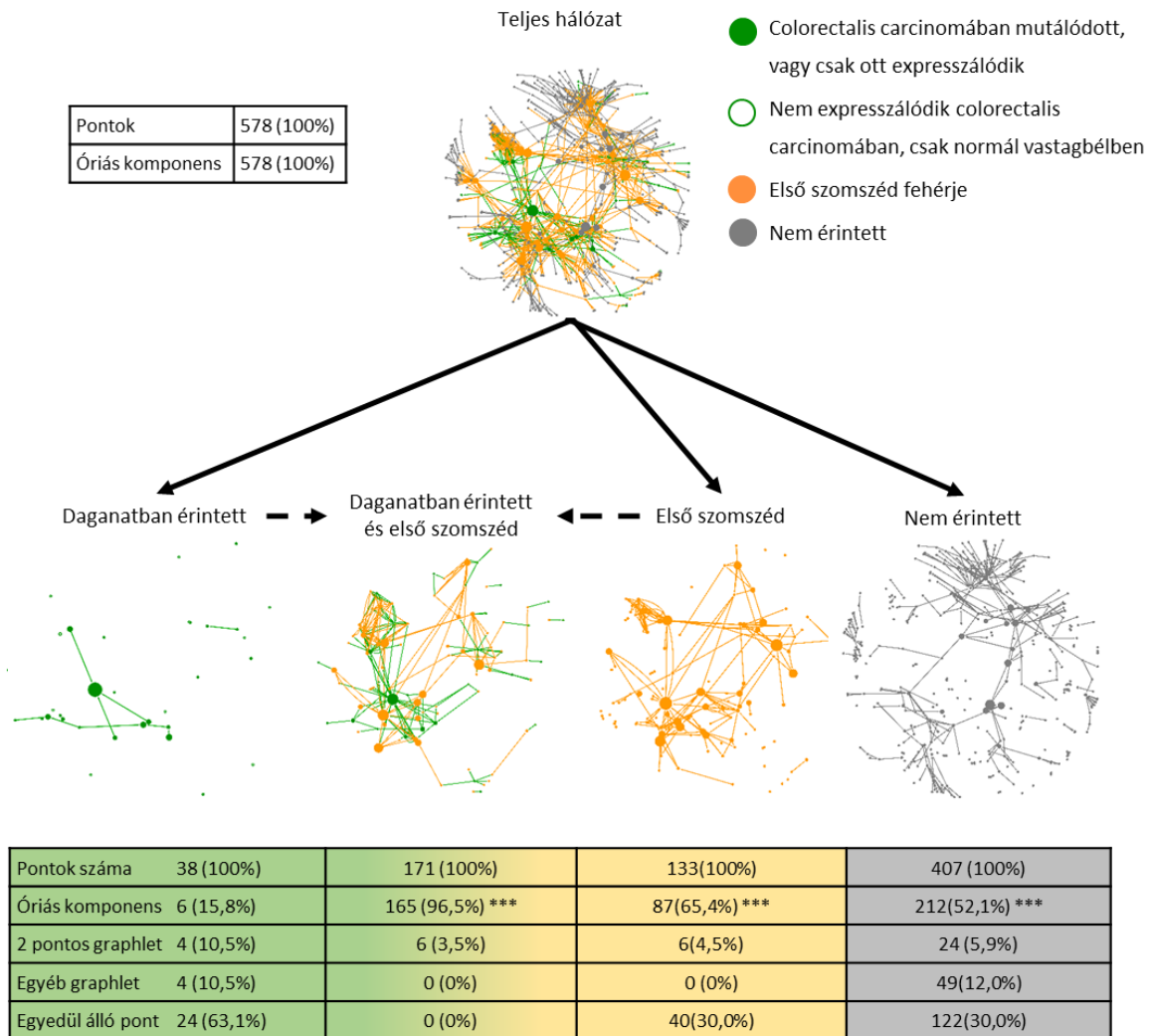
A kritikus paralógok elkülöníthetők a paralóg fehérjéktől mivel centrálisabbak, egy csoporton belül eltérőbb poszt-transzkripcionális regulációval és eltérőbb funkcióval rendelkeznek, nagyobb arányban vesznek részt betegségekben beleértve a daganatos betegségeket külön is, és jelenleg nagyobb arányban gyógyszer-célpontok, mint a nem paralóg fehérjék. **Tehát a kritikus paralógok egymással nem felcserélhetők, míg a paralóg fehérjék egy csoporton belül hasonlóak és felcserélhetőek egymással.**

3. Az első szomszédok szerepe a daganatok jelátviteli hálózatában

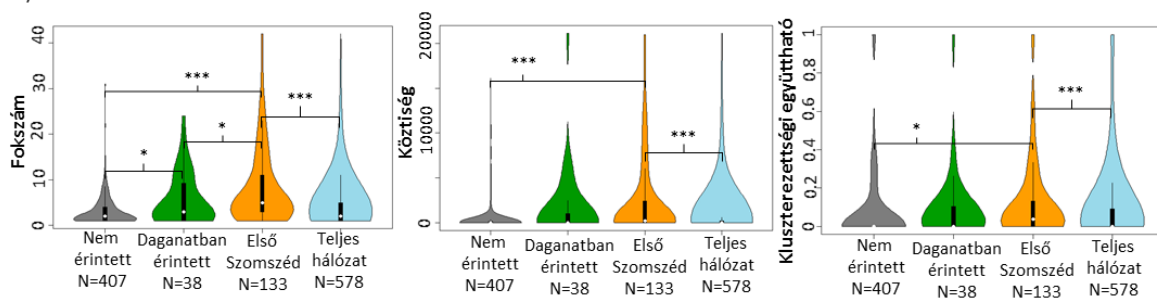
3.1. Az első szomszédok központi szerepet töltenek be a daganatok interakciós hálózatában

A daganatban érintett fehérjék mindössze a teljes jelátviteli hálózat 3-10 százalékát alkotják. (A SignaLink jelátviteli hálózatban vastagbél: 38/578 6,6%, mell: 52/572 9,1%, máj: 21/545 3,8%, tüdő: 31/581 5,3 %). Ez a relatíve kevés fehérje is képes áthuzalozni és átalakítani a jelátvitelt (Pires-daSilva és Sommer, 2003; Vogelstein és mtsai, 2013). Vizsgálatom során a korábbi irodalmi adatokkal ellentétben azt találtam, hogy ezek a fehérjék nincsenek kapcsolatban egymással, nem alkotnak egy óriáskomponenst, ahogy a colorectalis carcinoma jelátviteli hálózatán is látszik (12. ábra). A 38 daganatban érintett fehérje közül mindössze 6 fehérje kapcsolódik egymással direkt módon. A daganatban érintett fehérjéknek szükségük van az első szomszédjaikra, melyek egymással kapcsolatban vannak. A 133 első szomszéd közül 87 kapcsolódik egymáshoz (12. ábra/a). A daganatban nem érintett fehérjék a jelátviteli hálózat szélén találhatóak (12. ábra/a). Ugyan a teljes hálózat több mint felét alkotják, de csupán a fehérjék fele kapcsolódik egymáshoz. Szignifikánsan alacsonyabb az óriás komponensben való részvételük, mint az első szomszédoknak (Z teszt $p < 0,001$). Természetesen ezt nemcsak a colorectalis carcinoma, hanem a további három (mell, HCC, NSCLC) daganatban és valamennyi adatbázisban is megvizsgáltam. Hasonló eredményeket kaptam mindegyik daganatban, az első szomszédok a várható értéknél több, mint 3 szórással eltérő óriás komponenst hoztak létre (7. táblázat). A colorectalis carcinomában érintett fehérjék a várható értéknél több, mint 2 szórással nagyobb összefüggő hálózatot alkotnak a Cui és munkatársai féle jelátviteli hálózatban és a HPRD interakciós hálózatban. Az összes többi esetben a daganatban érintett fehérjék nem térnek el a várható értéktől, sőt a hepatocellularis carcinomában olyan kevés daganatban érintett fehérje fordult elő a gyűjtésem alapján, hogy a SignaLink adatbázisból készített hálózatban egyelemű, egymással nem összefüggő pontokat kaptam. A nem érintett fehérjék minden esetben a hálózatok perifériáját alkották, és a nagy számuk ellenére minden daganatban és minden hálózatban több, mint 4 szórással eltértek negatív irányba a várható értéktől. Ez legerősebben a két irányítatlan hálózatban a Biogrid- IntAct-DIP összevont hálózatban és a HPRD adatbázisból kapott hálózatban jelentkezett (7. táblázat).

a) Colorectalis carcinoma jelátviteli hálózata



b)



12. ábra a) Vastagbél daganat és normál vastagbél SignaLink jelátviteli hálózatra vetítve.

A zöld pontok a daganatban érintett fehérjék, a narancssárgák az első szomszédjaik, a szürkék a nem érintett fehérjék. A fehérjék mérete a köztiséggel egyenesen arányos. Az daganatban érintett fehérjék kapcsolatai zöldek, az első szomszédoké narancssárgák, a

nem érintetteké szürkék. Az ábra alatt az egyes csoportok szétszedett hálózatai láthatók. Az egyes gráfok alatt a gráfok tulajdonságai látszanak. Az óriás komponensek aránya eltér az egyes tulajdonságú pontok között (Z teszt *** $p < 0,001$) **b) Különböző topológiai mérőszámok alakulása a colorectalis carcinomában a daganatban érintett, nem érintett és első szomszéd fehérjék között** hegedű ábra, * $< 0,05$ *** $< 0,001$ Wilcoxon rang próba (Módos és mtsai, 2017)

Az első szomszéd fehérjék fontosságát nemcsak a gráf óriáskomponensében ragasztóként való összekötő szerepe jelzi, hanem a különböző, általunk megvizsgált centralitási értékekben is legalább annyira magas értéket vettek fel, mint a daganatban érintett fehérjék. A colorectalis carcinoma hálózatban az első szomszéd fehérjék magasabb fokszámmal rendelkeznek, mint a daganatban érintett fehérjék (12. ábra medián rendre fokszám esetén 7,5; 5 illetve kluszterezettségi együttható esetén 0 és 0,048 $p = 0,031$, Wilcoxon rang próba), és ugyanolyan köztiséggel (medián rendre 185,52 és 849,441 $p > 0,05$, Wilcoxon rang próba). Viszont mind a három értékük szignifikánsan magasabb a teljes hálózathoz képest ($p < 0,001$ Wilcoxon rang próba).

7. táblázat Az első szomszédok a várhatónál nagyobb óriás komponenszt alkotnak, míg a nem érintett fehérjék a várhatónál kisebbet.

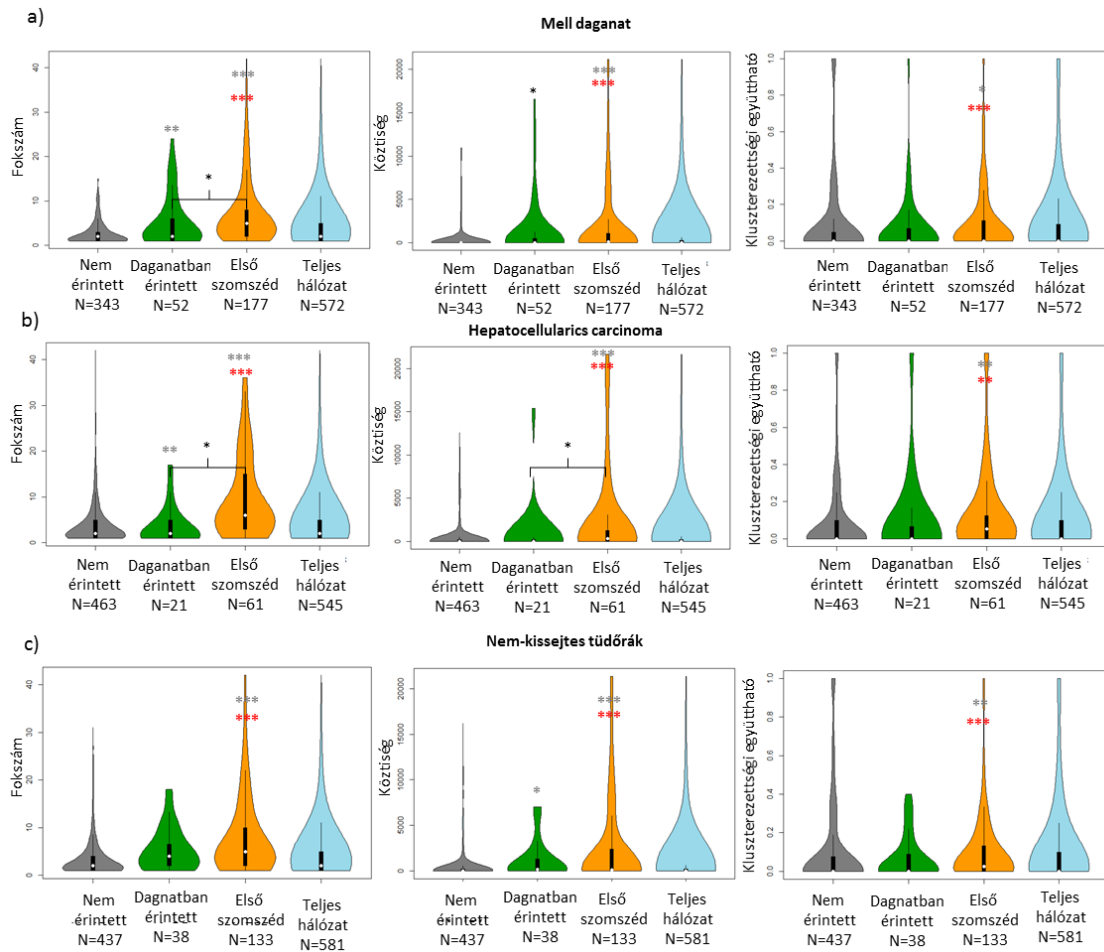
Hálózat	Signalink																							
	Colorectalis carcinoma						Mell						Hepatocellularis carcinoma						Nem kissejtes tüdő rák					
Típus	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	P	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	P	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	P	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	P
Daganatban érintett	38	6,6%	6	15,8%	1,79	0,0727	52	9,1%	3	5,8%	-0,82	0,413	21	1,8%	Egy elemű óriás komponens			31	5,3%	4	12,9%	0,97	0,3304	
Első szomszéd	133	23,0%	87	65,4%	4,29	<0,0001	177	30,9%	128	72,3%	3,04	0,0023	61	5,2%	32	52,5%	13,00	<0,0001	113	19,4%	69	61,1%	4,97	<0,0001
Nem érintett	407	70,4%	212	52,1%	-8,63	<0,0001	343	60,0%	53	15,5%	-12,47	<0,0001	463	39,6%	362	78,2%	-11,33	<0,0001	437	75,2%	306	70,0%	-5,17	<0,0001
Daganatban érintett +Első szomszéd	171	29,6%	165	96,5%	5,22	<0,0001	229	40,0%	229	100%	4,23	<0,0001	82	7,0%	74	90,2%	21,67	<0,0001	144	24,8%	144	100%	7,34	<0,0001
Teljes hálózat	578	100%	578	100%			572	100%	572	100%			545	46,6%	545	100%			581	100%	581	100%		

Hálózat	Cui és munkatársai																							
Daganat	Colorectalis carcinoma						Mell						Hepatocellularis carcinoma						Nem kissejtes tüdő rák					
Típus	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p
Daganatban érintett	81	6,7%	18	22,2%	3.15	0,0016	125	10,4%	14	11,2%	-0,10	0,9216	40	3,4%	3	7,5%	0,11	0,9122	70	5,8%	8	11,4%	0,90	0,3655
Első szomszéd	376	31,2%	291	77,4%	4.23	<0,0001	380	31,6%	320	84,2%	5,38	<0,0001	146	12,5%	119	81,5%	8,31	<0,0001	265	21,8%	220	83,0%	5,1	<0,0001
Nem érintett	749	62,1%	420	56,1%	-10.02	<0,0001	696	58,0%	261	37,5%	-15,44	<0,0001	984	84,1%	779	79,2%	-10,2	<0,0001	880	72,4%	572	65,0%	-11,1	<0,0001
Daganatban érintett +Első szomszéd	457	37,9%	449	98,2%	7.57	<0,0001	505	42,0%	497	98,4%	7,85	<0,0001	186	15,9%	167	89,8%	6,97	<0,0001	335	27,6%	324	96,7%	7,5	<0,0001
Teljes hálózat	1206	100%	1206	100%			1201	100%	1201	100%			1170	100%	1170	100%			1215	100%	1215	100%		

Hálózat	Reactome																							
	Daganat	Colorectalis carcinoma					Mell					Hepatocellularis carcinoma					Nem kissejtes tüdő rák							
Típus	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p
Daganatban érintett	237	6.1%	44	18,6%	-0,14	0,8910	422	11,1%	146	34,6%	-0,83	0,4039	122	3,2%	22	18,0%	1,16	0,2447	247	6,3%	78	31,6%	1,13	0,2598
Első szomszéd	2078	53.3%	1992	95,9%	4,96	<0,0001	2267	59,7%	2186	96,4%	4,31	<0,0001	1473	38,6%	1397	94,8%	6,03	<0,0001	2100	54,0%	2025	96,4%	4,99	<0,0001
Nem érintett	1582	40.6%	246	15,5%	-26,09	<0,0001	1109	29,2%	235	21,2%	-15,93	<0,0001	2223	58,2%	1551	69,8%	-12,41	<0,0001	1545	39,7%	631	40,8%	-16,49	<0,0001
Daganatban érintett +Első szomszéd	2315	59.4%	2271	98,1%	5,78	<0,0001	2689	70,8%	2674	99,4%	5,27	<0,0001	1595	41,8%	1563	98,0%	7,23	<0,0001	2347	60,3%	2312	98,5%	5,60	<0,0001
Teljes hálózat	3897	100%	3897	100%			3798	100%	3798	100%			3818	100%	3818	100%			3892	100%	3892	100%		

Hálózat	Biogrid DIP IntAct																							
Daganat	Colorectalis carcinoma						Mell						Hepatocellularis carcinoma						Nem kissejtes tüdő rák					
Típus	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	p
Daganatban érintett	628	6,2%	244	38,9%	-0,12	0,905E	1057	10,6%	547	51,8%	-0,26	0,796	307	3,1%	68	22,1%	0,08	0,9367	624	6,1%	264	42,3%	0,51	6.0886E-01
Első szomszéd	4724	46,3%	4594	97,2%	5,39	<0,0001	5159	51,6%	4974	96,4%	4,42	<0,0001	3002	30,0%	2960	98,6%	8,36	<0,0001	4634	45,3%	4510	97,3%	5,85	<0,0001
Nem érintett	4850	47,5%	1016	20,9%	-36,24	<0,0001	3782	37,8%	150	4,0%	-37,03	<0,0001	6698	66,9%	3823	57,1%	-27,99	<0,0001	4963	48,6%	1292	26,0%	-33,54	<0,0001
Daganatban érintett +Első szomszéd	5352	52,5%	5252	98,1%	6,54	<0,0001	6216	62,2%	6216	100,0%	5,51	<0,0001	3309	33,1%	3309	100%	8,68	<0,0001	5258	51,4%	5258	100%	6,59	<0,0001
Teljes hálózat	10202	100%	10202	100%			9998	100,0%	9998	100,0%			10007	100,0%	10007	100,0%			10221	100%	10221	100%		

Hálózat	HPRD																							
Daganat	Colorectalis carcinoma						Mell						Hepatocellularis carcinoma						Nem kissejtes tüdő rák					
Típus	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	P	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	P	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	P	Összes fehérje	% teljes hálónak	Pontok óriás komponensben	% óriás komponensben	Z	P
Daganatban érintett	405	6,0%	87	21,5%	2,01	0,0445	669	10,1%	150	22,4%	-0,36	0,7166	213	3,2%	16	7,5%	0,94	0,3474	404	5,9%	61	15,1%	0,90	0,3697
Első szomszéd	1836	27,1%	1668	90,8%	13,90	<0,0001	2221	33,4%	1990	89,6%	12,70	<0,0001	1231	18,6%	1111	90,3%	13,96	<0,0001	1740	25,6%	1573	90,4%	14,06	<0,0001
Nem érintett	4540	67,0%	2600	57,3%	-38,68	<0,0001	3763	56,6%	1438	38,2%	-41,86	<0,0001	5186	78,2%	3900	75,2%	-27,61	<0,0001	4567	67,2%	2884	63,1%	-33,28	<0,0001
Daganatban érintett +Első szomszéd	2241	33,0%	2235	99,7%	17,84	<0,0001	2890	43,4%	2879	99,6%	16,82	<0,0001	1444	21,8%	1428	98,9%	17,63	<0,0001	2144	31,5%	2142	99,9%	17,55	<0,0001
Teljes hálózat	6781	100%	6781	100%			6653	100%	6653	100%			6630	100%	6630	100%			6801	100%	6801	100%		

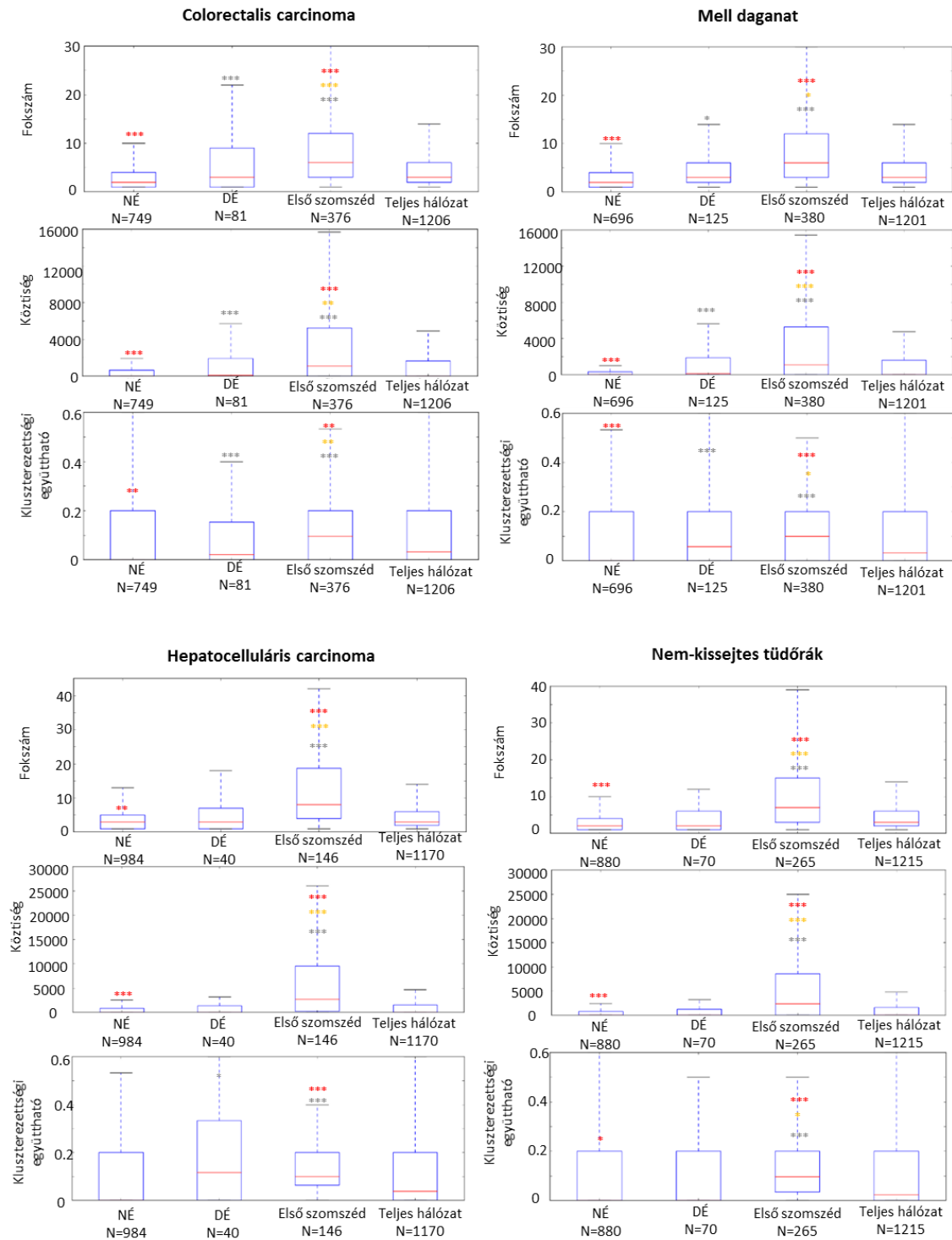


13. ábra A Signalink jelátviteli hálózaton a többi három daganat hálózatos centralitásai.

a) Mell daganat b) Hepatocellularis carcinoma, c) Nem-kissejtes tüdőrák A mutált fehérjék daganat-specifikusan a Cancer Gene Census adatbázisból származnak, a differenciáltan expresszálódó fehérjék azok, amelyek mRNS-e csak daganatban vagy csak a normál állapotban jelent meg a Signalink kézi gyűjtésű jelátviteli hálózaton. * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$. Wilcoxon rang próba a teljes szövetspecifikus hálózatra nézve vörös csillaggal, a nem érintett fehérjékhez képest szürke csillaggal. (Módos és mtsai, 2017)

A többi daganatban (mell, máj, tüdő) az első szomszédok szignifikánsan magasabb köztséggel, fokszámmal és kluszterezettségi együtthatóval rendelkeztek, mint a nem érintett fehérjék, és mint a teljes hálózat fehérjei. ($p < 0,001$ Wilcoxon rang próba) a Signalink jelátviteli hálózaton (13. ábra).

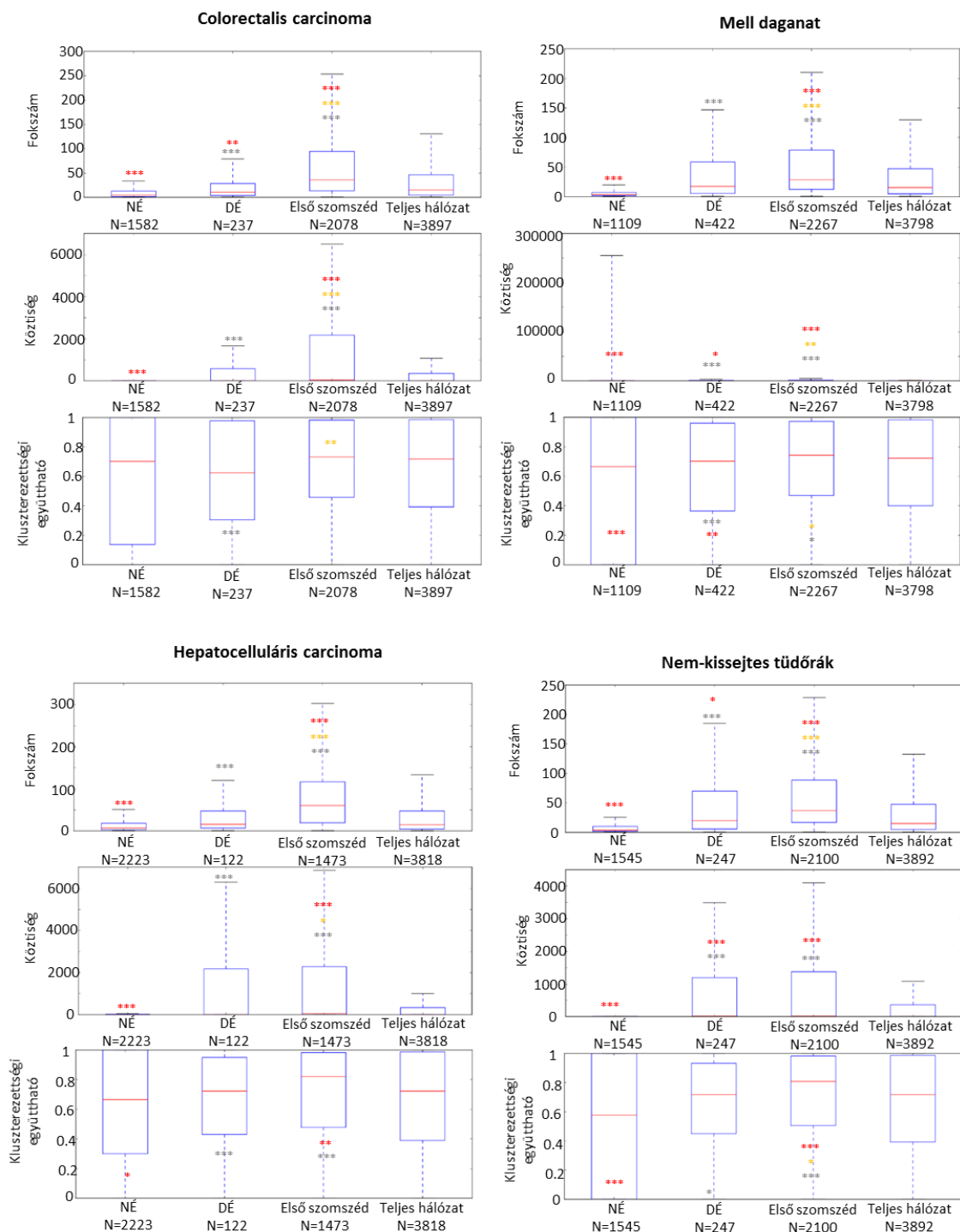
A négy további kontroll hálón (Cui és munkatársai 14. ábra, Reactome 15. ábra, Biogrid- IntAct-DIP összevont háló 16. ábra, HPRD 17. ábra) végzett mérésen hasonló eredmény jött ki a teljes háléhoz és a nem érintett fehérjékhez képest, a daganatban érintett fehérjék első szomszédjai magasabb topológiai centralitásokkal rendelkeztek ($p < 0,05$ Wilcoxon rang próba). Egyetlen kivétel a Reactome hálózatban mért kluszterezettség együttható volt (15. ábra), mivel a komplexeket – ezáltal egymással is szomszédos fehérjéket – nagy számban tartalmazó Reactome esetében a kluszterezettség alapvetően magasabb és nem tér el a fehérjék daganattal kapcsolatos besorolása alapján.



14. ábra Az első szomszédok fontossága a különböző daganatokban a Cui és munkatársai jelátviteli hálózatban.

A mutált fehérjék daganat-specifikusan a Cancer Gene Census adatbázisból származnak, a differenciáltan expresszálódó fehérjék alatt azt értettük, amelyek mRNS-e csak daganatban vagy csak a normál állapotban jelent meg a Cui és munkatársai hálózatban.

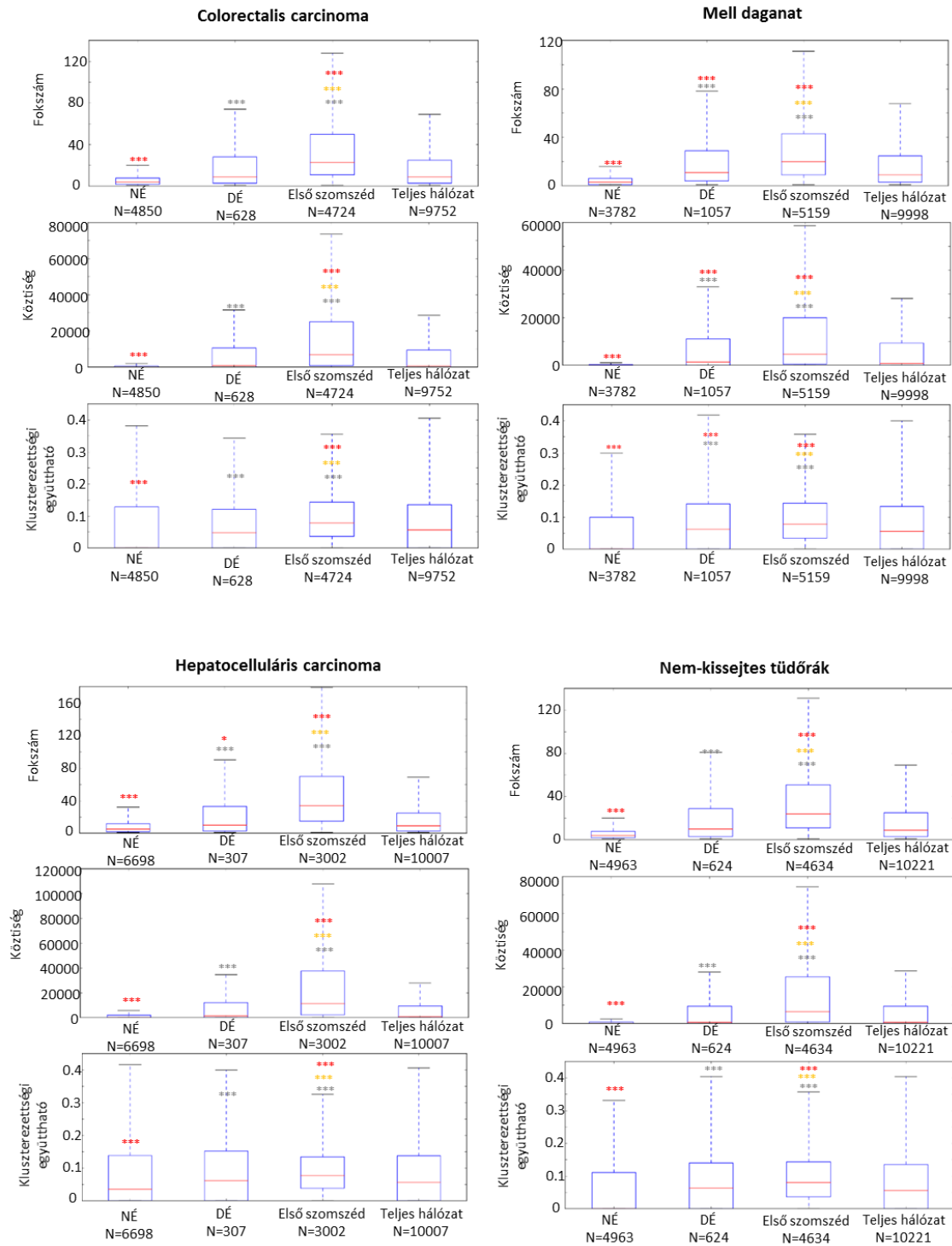
NÉ : nem érintett, DÉ az adott daganatban érintett * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$.
Wilcoxon rang próba a teljes szövetspecifikus hálózatra nézve vörös csillaggal, a daganatban érintett fehérjékhez képest narancssárga csillaggal, a nem érintett fehérjékhez képest szürke csillaggal. (Módos és mtsai, 2017)



15. ábra Az első szomszédok fontossága a különböző daganatokban a Reactome jelátviteli hálózatban.

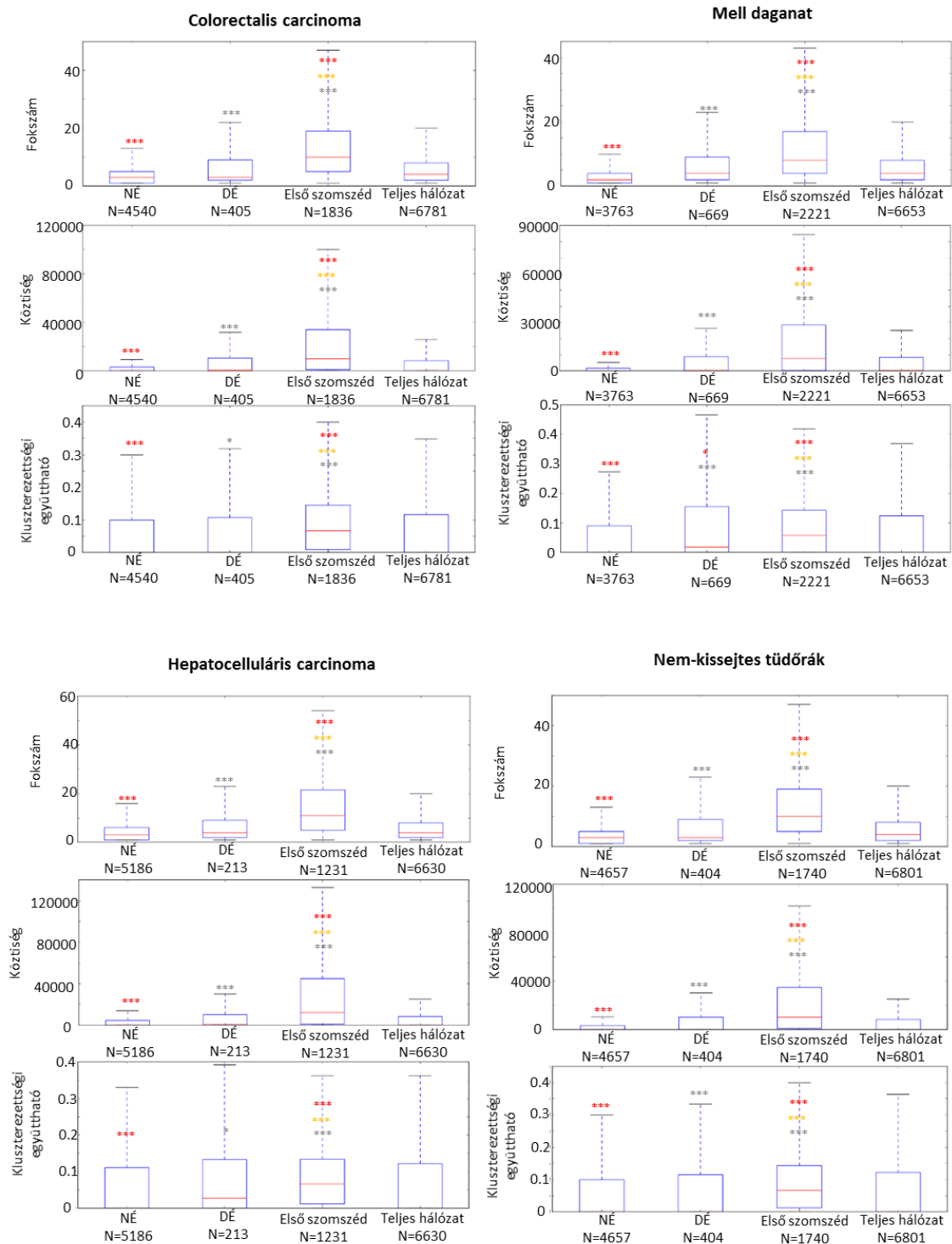
A mutált fehérjék daganat-specifikusan a Cancer Gene Census adatbázisból származnak, a differenciáltan expresszálódó fehérjék alatt azt értettük, amelyek mRNS-e csak daganatban vagy csak a normál állapotban jelent meg a Reactome hálózatban. NÉ : nem érintett, DÉ az adott daganatban érintett * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$.

Wilcoxon rang próba a teljes szövetspecifikus hálózatra nézve vörös csillaggal, a daganatban érintett fehérjékhez képest narancssárga csillaggal, a nem érintett fehérjékhez képest szürke csillaggal. (Módos és mtsai, 2017)



16. ábra Az első szomszédok fontossága a különböző dagantokban a Biogrid, DIP és IntAct összevont fehérje-fehérje interakciós hálózatban.

A mutált fehérjék daganat-specifikusan a Cancer Gene Census adatbázisból származnak, a differenciáltan expresszálódó fehérjék alatt azt értettük, amelyek mRNS-e csak daganatban vagy csak a normál állapotban jelent meg a Biogrid, DIP és IntAct összevont hálózatban. NÉ : nem érintett, DÉ az adott daganatban érintett * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$. Wilcoxon rang próba a teljes szövetspecifikus hálózatra nézve vörös csillaggal, a daganatban érintett fehérjékhez képest narancssárga csillaggal, a nem érintett fehérjékhez képest szürke csillaggal. (Módos és mtsai, 2017)



17. ábra Az első szomszédok fontossága a különböző daganatokban a HPRD kézi gyűjtésű fehérje-fehérje interakciós hálózatban.

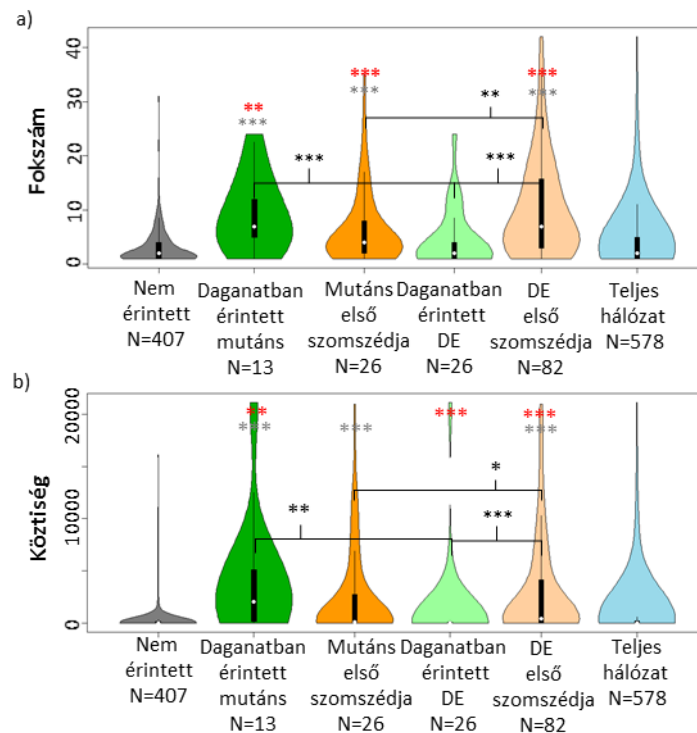
A mutált fehérjék daganat-specifikusan a Cancer Gene Census adatbázisból származnak, a differenciáltan expresszálódó fehérjék alatt azt értettük, amelyek mRNS-e csak daganatban vagy csak a normál állapotban jelent meg a HPRD hálózatban. NÉ: nem

érintett, DE az adott daganatban érintett * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$. Wilcoxon rang próba a teljes szövetspecifikus hálózatra nézve vörös csillaggal, a daganatban érintett fehérjékhez képest narancssárga csillaggal, a nem érintett fehérjékhez képest szürke csillaggal. (Módos és mtsai, 2017)

Összességében tehát elmondható, hogy az első szomszédok úgy viselkednek, mint a jelátviteli hálózat ragasztói: a tumorigenesis során megváltozó fehérjékből az információt tovább juttatják a jelátviteli hálózat egészéig. Így lesznek képesek a daganatban érintett fehérjék első szomszédjaikon keresztül áthuzalozni a teljes jelátviteli hálózatot.

3.2. A daganatban mutálódó fehérjék közvetlenül, a daganatban differenciáltan expresszáldó fehérjék közvetetten huzalozzák át a jelátviteli hálózatot

Az első szomszéd fehérjék tehát központi szerepet töltenek be a jelátviteli és a fehérje interakciós hálózatokban, ezért megnéztem, hogy van-e különbség a daganatban mutálódó fehérjék és első szomszédjaik, valamint a daganatban differenciáltan expresszáldó (DE) fehérjék és első szomszédjaik között.



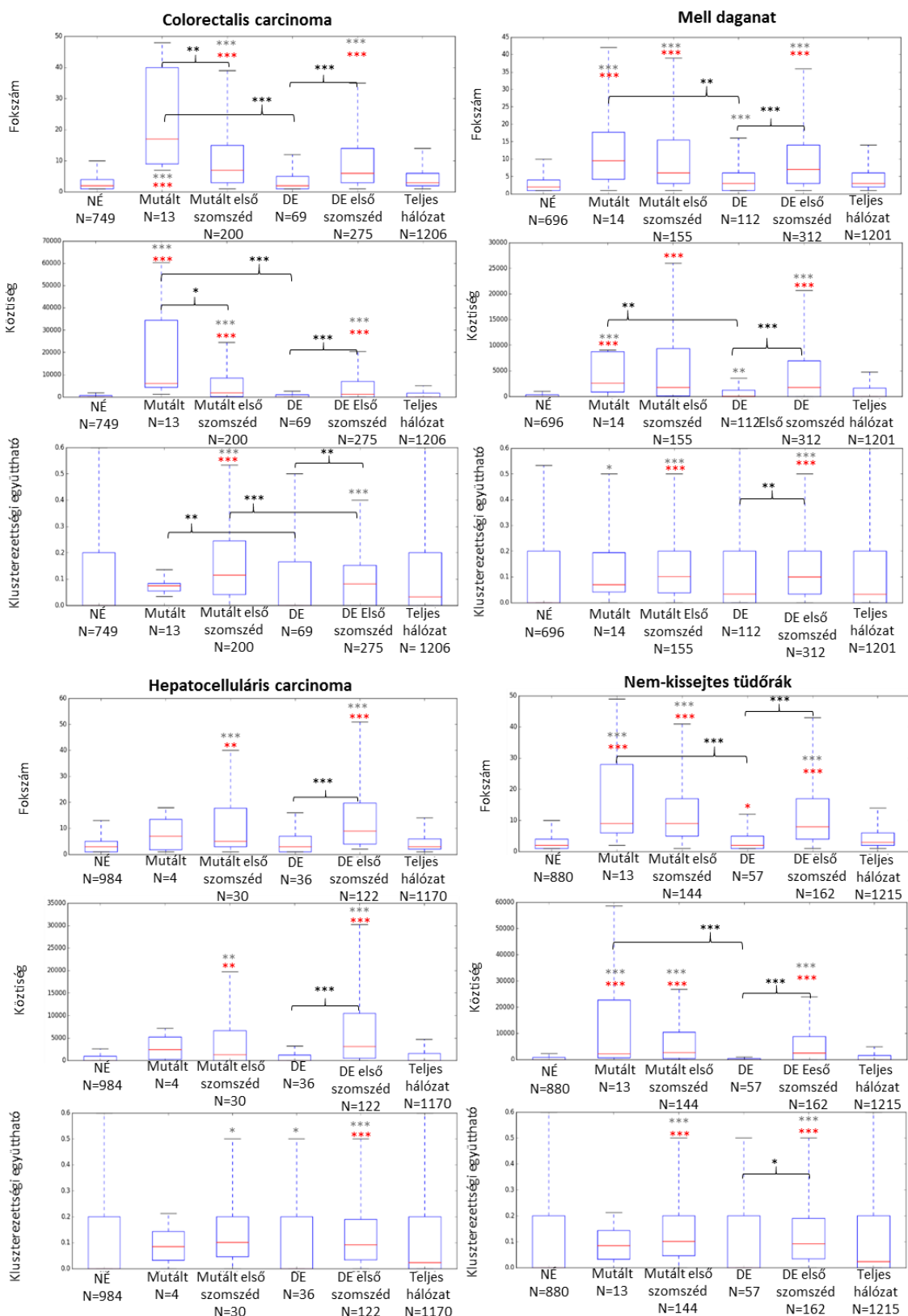
18. ábra A mutálódott és differenciáltan expresszálódó daganatos fehérjék és első szomszédjaik hálózati centralitásai vastagbél daganatban a SignaLink adatbázist használva.

a) fokszám, b) köztség. A hegedűplottokról leolvasható a mért értékek teljes eloszlása. A fehér pontok a mediánokat jelzik, a dobozok 25-75 percentilisig terjednek, a szakállak a dobozok kétszeresei. A csillagok a Wilcoxon rang próba szerinti szignifikanciának felelnek meg: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$. A szürke csillagok a nem érintett pontokhoz képest mérik a szignifikanciát, a pirosok a teljes hálózathoz. (Módos és mtsai, 2017)

A vastagbél daganatban a SignaLink jelátviteli hálózatban a daganatban érintett mutáció során megváltozó fehérjék ugyanannyira központiak, mint az első szomszédjaik ($p > 0,05$ Wilcoxon rang próba), míg a DE fehérjék épp ellenkezőleg, kevésbé központiak, mint első szomszédjaik ($p < 0,05$ Wilcoxon rang próba) (18. ábra). Ez igaz valamennyi további daganatra is (19. ábra, 20. ábra, 21. ábra, 22. ábra). Az elkülönülés vastagbél daganat esetében mind a jelátviteli hálózatokban (Reactome; Cui és munkatársai., 19. ábra, 20. ábra), mind pedig a fehérje-fehérje interakciós hálózaton (DIP, IntAct, BioGrid

egyesített hálózat; HPRD 21. ábra, 22. ábra) megfigyelhető volt. A Cui és munkatársai és az IntAct, DIP, BIoGrid hálózatban a daganatban érintett mutált fehérjék magasabb fokszámmal és köztiséggel rendelkeztek, mint első szomszédjaik. Ezzel szemben a daganatban DE fehérjék első szomszédjainak magasabb volt mind a fokszáma, mind a köztiség értéke, mint a DE daganatban érintett fehérjéknek. A Reactome hálózatban a sok komplex miatt a mutált fehérjék első szomszédjai is magasabb fokszámot vettek fel, mint a daganatban érintett mutált fehérjék (20. ábra). **A jelenség oka, hogy a daganatban mutált fehérjék hatása a jelátvitelre közvetlen. Önmagukban képesek a jelátvitelt áthuzalozni nagyobb központiségük miatt. Ezzel szemben a differenciáltan expresszáldó fehérjéknek inkább indirekt a hatása, amit a központi első szomszédok közvetítenek. A daganatban érintett differenciáltan expresszáldó fehérjék közvetett hatást fejtenek ki a magasabb centralitású első szomszédjukra, és ezáltal képesek befolyásolni a jelátvitelt.**

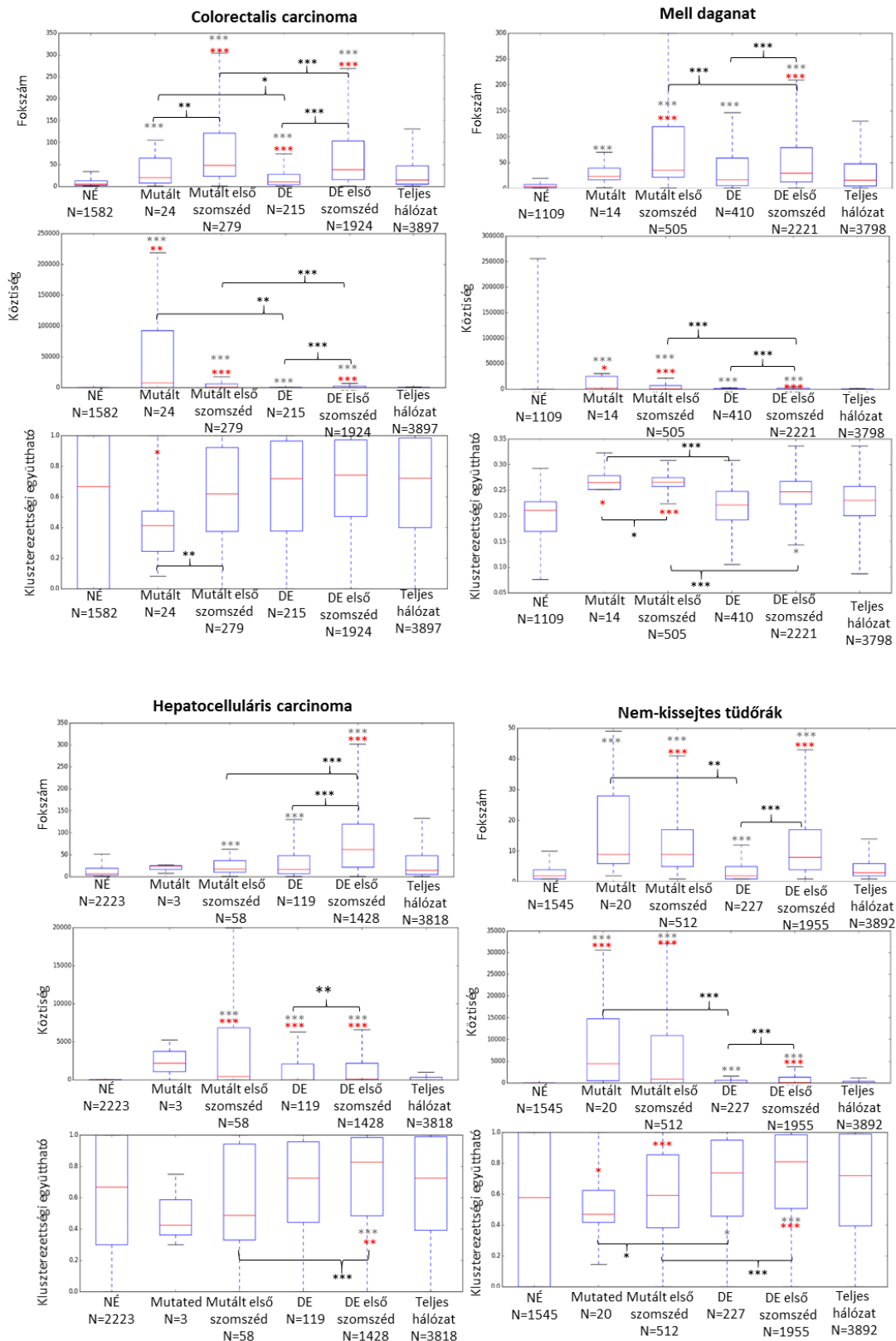
Ez vizsgálatunk során mind a négy daganatban és mind az öt hálózatban megjelent (19. ábra, 20. ábra, 21. ábra, 22. ábra). A hálózatok vizsgálata során a vizsgált fehérjék mennyiségében nagy eltéréseket tapasztaltam. A nagy áteresztő képességű módszerekkel gyűjtött hálózat (DIP, BIOGRID, IntAct) elemszáma az adott szövetben megközelítette a 10 000 fehérjét, míg a HPRD-ben ennek a kétharmadát. Összehasonlításképpen a SignaLinkben 578 fehérje szerepelt a vizsgálatban.



19. ábra A colorectalis carcinoma, a mell daganat, a hepatocelluláris carcinoma és nem kissejtes tüdőrák daganatban mutáns és differenciáltan expresszálódó

fehérjéinek és első szomszédjaiknak hálózati centralitásai a Cui és munkatársai (2007) manuálisan gyűjtött daganatos jelátviteli hálózaton.

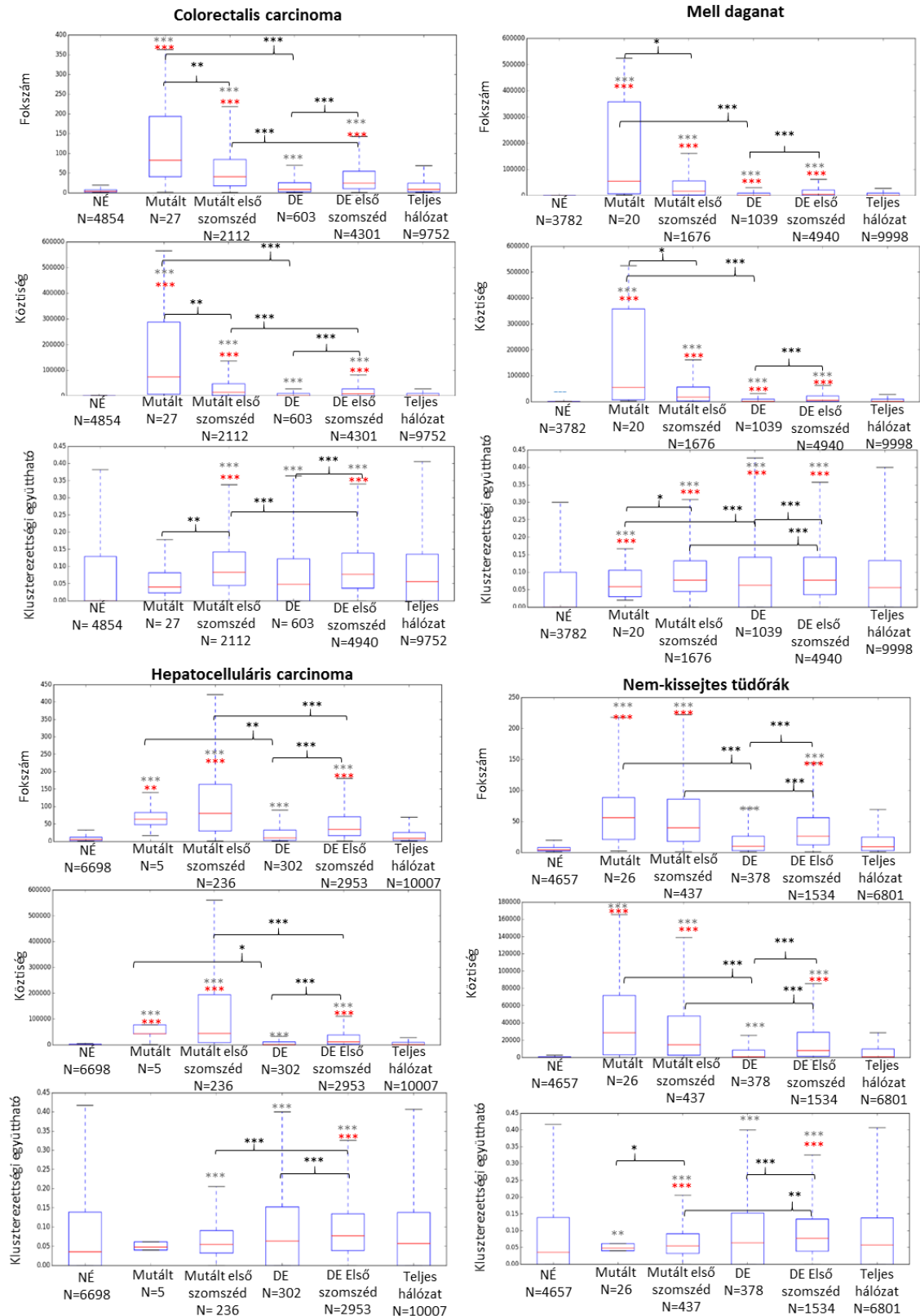
A mutált fehérjék daganat-specifikusan a Cancer Gene Census adatbázis szerint daganatot okozó mutációt szenvedő fehérjék, a differenciáltan expresszálódó fehérjék alatt azt értettük, amelyek mRNS-e csak daganatban vagy csak a normál állapotban jelent meg a Cui és munkatársai hálózatban. DE: az adott daganatban differenciáltan expresszálódó fehérje NÉ: nem érintett fehérje. Wilcoxon rang próba: * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$. A teljes hálózathoz képesti eltérést a piros csillagok, a nem érintett fehérjékhez képesti eltérést a szürke csillagok jelzik. (Módos és mtsai, 2017)



20. ábra A colorectalis carcinoma, a mell daganat, a hepatocelluláris carcinoma és nem kissejtes tüdőrák daganatban mutáns és differenciáltan expresszálódó

fehérjéinek és első szomszédjaiknak hálózati centralitásai a Reactome manuálisan gyűjtött jelátviteli hálózaton.

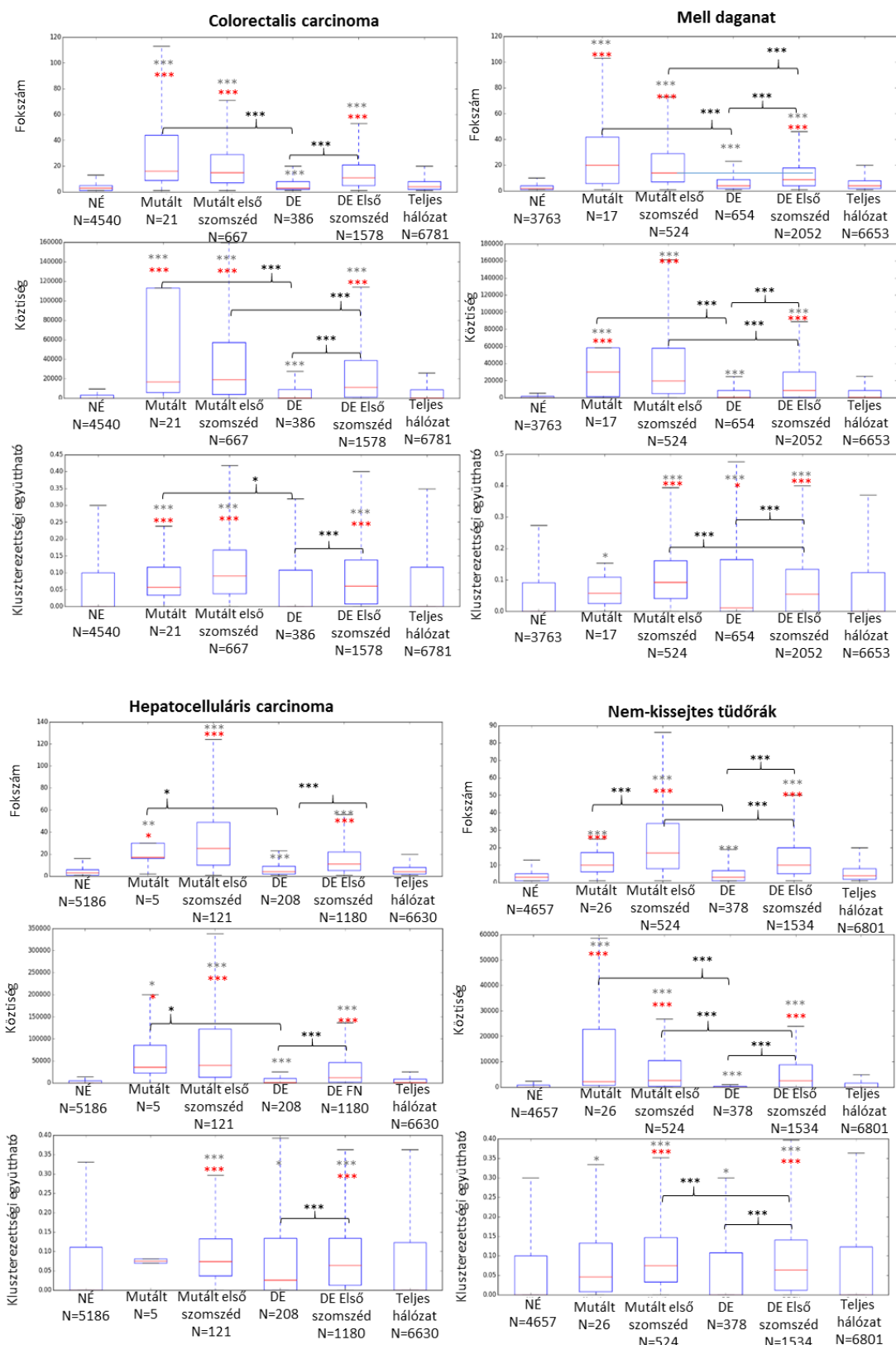
A mutált fehérjék daganat-specifikusan a Cancer Gene Census adatbázis szerint daganatot okozó mutációt szenvedő fehérjék, a differenciáltan expresszáldó fehérjék alatt azt értettük, amelyek mRNS-e csak daganatban vagy csak a normál állapotban jelent meg a Reactome hálózatban. NÉ: nem érintett, DE: differenciáltan expresszáldó. Wilcoxon rang próba * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$. A teljes hálózathoz képesti eltérést a piros csillagok, a nem érintett fehérjékhez képest a szürke csillagok jelzik. (Módos és mtsai, 2017)



21. ábra A colorectalis carcinoma, a mell daganat, a hepatocelluláris carcinoma és nem kissejtes tüdőrák daganatban mutáns és differenciáltan expresszálódó

fehérjéinek és első szomszédjaiknak hálózati centralitásai a Biogrid, DIP és IntAct fehérje interakciós hálózaton.

A mutált fehérjék daganat-specifikusan a Cancer Gene Census adatbázis szerint daganatot okozó mutációt szenvedő fehérjék, a differenciáltan expresszáldó fehérjék alatt azt értettük, a differenciáltan expresszáldó fehérjék alatt azt értettük, amelyek mRNS-e csak daganatban vagy csak a normál állapotban jelent meg a Biogrid, DIP, IntAct hálózatban. NÉ: nem érintett, DE: differenciáltan expresszáldó. Wilcoxon rang próba: * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$. A teljes hálózathoz képesti eltérést a piros csillagok, a nem érintett fehérjékhez képest a szürke csillagok jelzik. (Módos és mtsai, 2017)



22. ábra A colorectalis carcinoma, a mell daganat, a hepatocelluláris carcinoma és nem kissejtes tüdőrák daganatban mutáns és differenciáltan expresszálódó

fehérjéinek és első szomszédjaiknak hálózati centralitásai a kézi gyűjtésű HPRD fehérje interakciós hálózaton.

A mutált fehérjék daganat-specifikusan a Cancer Gene Census adatbázis szerint daganatot okozó mutációt szenvedő fehérjék, a differenciáltan expresszálódó fehérjék alatt azt értettük, amelyek mRNS-e csak daganatban vagy csak a normál állapotban jelent meg a HPRD hálózatban. NÉ: nem érintett, DE: differenciáltan expresszálódó. Wilcoxon rang próba * - $p < 0,05$; ** - $p < 0,01$; *** - $p < 0,001$. A teljes hálózathoz képesti eltérést a piros csillagok, a nem érintett fehérjékhez képest a szürke csillagok jelzik. (Módos és mtsai, 2017)

3.3. Lehetséges új utak a gyógyszerfejlesztésben az első szomszédok bevonásával.

Jelenleg a daganatban érintett fehérjéket tekintjük a gyógyszerfejlesztés elsődleges célpontjainak. Mint ahogy láthattuk, az első szomszédok legalább annyira központi helyet foglalnak el a daganatok Jelenleg a daganatban érintett fehérjéket tekintjük a gyógyszerfejlesztés elsődleges célpontjainak. Mint ahogy láthattuk, az első szomszédok legalább annyira központi helyet foglalnak el a daganatok jelátviteli hálózatában, mint a daganatban érintett fehérjék (12. ábra/b), ezért érdemes lehetséges gyógyszer-célpontként tekinteni rájuk.

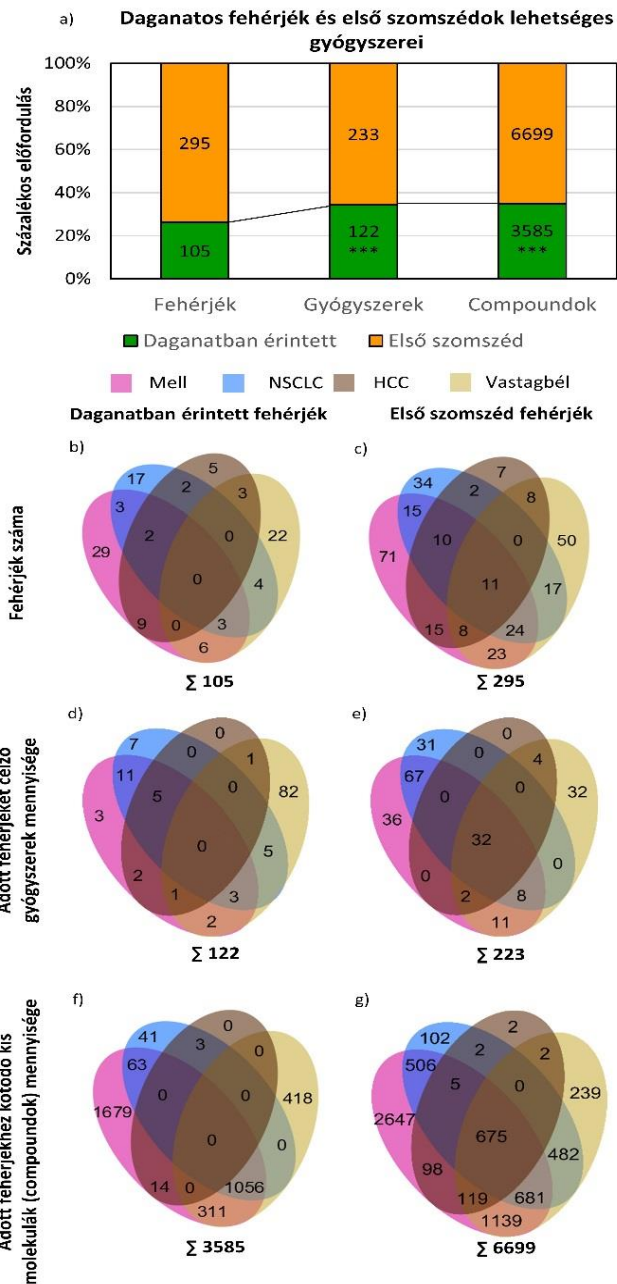
Ezért megvizsgáltuk, hogy a ChEMBL adatbázisban hány jelenlegi gyógyszer, és a ChEMBL adatbázisból hány kémiai molekula (*compound*) hat a daganatban érintett fehérjékre, és hány azok első szomszédjaira (23. ábra). Látható, hogy mind gyógyszerekből, mind pedig compoundokból több jut egy daganatban érintett fehérjére, mint első szomszédjára (rendre 1,14 a 0,76 szemben; és 34,14 a 23,11 szemben $p < 0,001$ Bernoulli teszt a fehérjék arányhoz képest). A daganatban érintett fehérjék felé eltolódó arány alapján elmondható, hogy az első szomszédokat nem vizsgálták olyan kimerítően gyógyszer-célpontként, mint a daganatban érintett fehérjéket (23. ábra).

Megvizsgáltam, hogy a daganatban érintett fehérjékre az első szomszédokhoz képest magasabb gyógyszer, illetve compound szám hogyan mutatkozik az általam vizsgált négy daganatban. A legtöbb daganatban érintett fehérje a mell daganatban fordul elő (52), ezért is meglepő, hogy a legtöbb gyógyszer a colorectalis carcinomában érintett fehérjére van (94 colorectalis carcinoma, 31 nem-kissejtes tüdőrák, 27 mell daganat, 9 hepatocellularis carcinoma) (23. ábra). A legtöbb compound a mell daganatban érintett fehérjék ellen készült (3123 mell daganat, 1785 colorectalis carcinoma, 1163 nem-

kissejtes tüdő rák, 17 hepatocellularis carcinoma) (23. ábra). A hepatocellularis carcinomával összefüggő fehérjék ellen létező compoundok száma (17) aggasztóan alacsony. Egyszerre mind a négy daganat ellen nem találtam sem gyógyszert, sem compoundot, mivel jelen vizsgálatban nem találtam mindegyik daganatban egységesen megváltozó fehérjét a CGC szigorú daganat specifikus szűrése miatt (23. ábra). A legtöbb compoundot a mediator rétegben elhelyezkedő Akt kinázra fejlesztettek (953-at), amely a négy tumorból 3 esetén daganatban érintett (mell, tüdő, vastagbél). jelátviteli hálózatában, mint a daganatban érintett fehérjék (12. ábra/b), ezért érdemes lehetséges gyógyszerként tekinteni rájuk.

Ezért megvizsgáltuk, hogy a ChEMBL adatbázisban hány jelenlegi gyógyszer, és a ChEMBL adatbázisból hány kémiai molekula (compound) hat a daganatban érintett fehérjékre, és hány azok első szomszédjaira (23. ábra). Látható, hogy mind gyógyszerekből, mind pedig compoundokból több jut egy daganatban érintett fehérjére, mint első szomszédjára (rendre 1,14 a 0,76 szemben; és 34,14 a 23,11 szemben $p < 0,001$ Bernoulli teszt a fehérjék arányhoz képest). A daganatban érintett fehérjék felé eltolódó arány alapján elmondható, hogy az első szomszédokat nem vizsgálták olyan kimerítően gyógyszerként, mint a daganatban érintett fehérjéket (23. ábra).

Megvizsgáltam, hogy a daganatban érintett fehérjékre az első szomszédokhoz képest magasabb gyógyszer, illetve compound szám hogyan mutatkozik az általam vizsgált négy daganatban. A legtöbb daganatban érintett fehérje a mell daganatban fordul elő (52), ezért is meglepő, hogy a legtöbb gyógyszer a colorectalis carcinomában érintett fehérjére van (94 colorectalis carcinoma, 31 nem-kissejtes tüdőrák, 27 mell daganat, 9 hepatocellularis carcinoma) (23. ábra). A legtöbb compound a mell daganatban érintett fehérjék ellen készült (3123 mell daganat, 1785 colorectalis carcinoma, 1163 nem-kissejtes tüdő rák, 17 hepatocellularis carcinoma) (23. ábra). A hepatocellularis carcinomával összefüggő fehérjék ellen létező compoundok száma (17) aggasztóan alacsony. Egyszerre mind a négy daganat ellen nem találtam sem gyógyszert, sem compoundot, mivel jelen vizsgálatban nem találtam mindegyik daganatban egységesen megváltozó fehérjét a CGC szigorú daganat specifikus szűrése miatt (23. ábra). A legtöbb compoundot a mediator rétegben elhelyezkedő Akt kinázra fejlesztettek (953-at), amely a négy tumorból 3 esetén daganatban érintett (mell, tüdő, vastagbél).



23. ábra Gyógyszerek és kismolekulák megoszlása a daganatban érintett fehérjék és az első szomszédok között

- a) Az összes daganatban a daganatban érintett és az első szomszédokra vonatkozó gyógyszercélpontok (Bernoulli teszt a fehérjék megoszlásához viszonyítva *** $p < 0,001$)
 b) Daganatban érintett fehérjék megoszlása az általunk vizsgált daganatokban, c) Első szomszéd fehérjék megoszlása az általunk vizsgált daganatokban d) Daganatokban érintett fehérjéket célzó gyógyszerek e) Első szomszédokat célzó gyógyszerek, f) Daganatokban érintett fehérjéket célzó molekulák a ChEMBL adatbázis szerint, g) Első

szomszéd fehérjéket célzó molekulák a ChEMBL adatbázis szerint (A célfehérjéhez való kötődés határértéke <50 nM volt) (Módos és mtsai, 2017)

Az egyes első szomszéd fehérjékre a létező gyógyszerek száma jóval kisebb. Viszont ezek terápiás relevanciájának feltárásával új lehetőségekre mutathatunk rá a gyógyszertervezésben. Mind az első szomszédokra alkalmazott gyógyszerek, mind a daganatban érintett fehérjék ellen alkalmazottak között kevés olyan van, amely rák ellenes szerként használható a gyógyszerek típusát meghatározó Anatomical Therapeutic Chemical (ATC) klasszifikációval rendelkezők közül (Berg és mtsai, 2014) (27, illetve 25 db a 184, illetve 113 közül). A hasonló számú gyógyszer oka, hogy a különböző daganatokban az első szomszédok átfednek a daganatban érintett fehérjékkel.

A kevés daganatban alkalmazott szer miatt mindenképpen gondolkodni kell az egyes gyógyszerek indikáción túli (*off-label*) alkalmazásán. Így lehetőségünk van a rákban alkalmazható gyógyszerek körét kiterjeszteni. **Sőt, ha az első szomszédokat is lehetséges gyógyszer-célpontoknak tekintjük, akkor nemcsak a már alkalmazható készítmények számát növeljük meg, hanem a lehetséges célpontok számát is megsokszorozzuk.**

VI. Megbeszélés

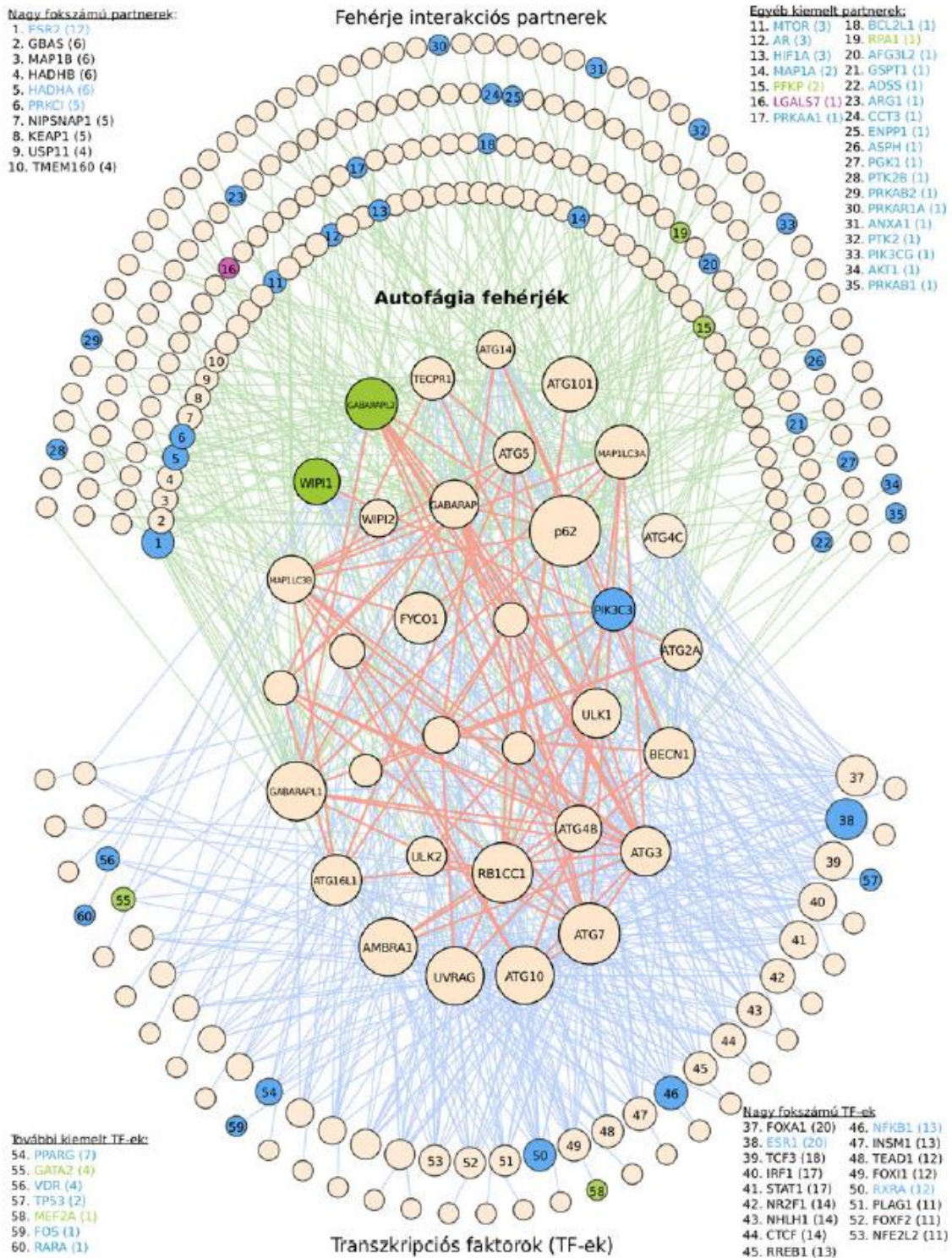
A megbeszélésben szeretném az általam fejlesztett adatbázisok alkalmazásának lehetőségeit, a kritikus paralógok koncepciójának alkalmazását, valamint a rákban érintett fehérjék első szomszédjainak gyógyszerfejlesztési jelentőségét ismertetni.

1. A SignaLink, az ARN és az NRF2ome adatbázis lehetséges alkalmazásai

Mindegyik adatbázis a maga nemében hiánypótló szerepet tölt be. A SignaLink a jelátviteli vizsgálatok, és azon belül is a cross-talk vizsgálatokra nyújt egy lehetséges alternatívát. Előnye, hogy réteges felépítése miatt különböző típusú és megbízhatóságú adatokat tartalmaz. A dolgozat szempontjából csupán a transzkripció-faktor célgén, miRNS-célgén és TF-miRNS réteget emeltem ki, mivel én ezen rétegek elkészítésben vettem részt. Lehetséges alkalmazásai közé tartozik a SignaLink útvonal annotációk használata gyógyszermolekula hatás vizsgálatokban (Bulusu és mtsai, 2016; Melas és mtsai, 2015). Az általam bemutatott két módszer a SignaLink adatbázis lehetséges alkalmazására példa.

Az ARN, mint az autofágia és annak regulációját összesítő adatbázis hiánypótló kiindulópontja az autofági kutatásoknak. Használata lehetőséget ad regulációs körök megtalálására, és ezzel egyes autofágia mechanizmusok tisztázására.

Az autofágiára specifikusan is alkalmazhatóak a daganatban érintett első szomszédok feltérképezésében használt módszerek. Az ARN adatbázissal, pedig lehetőség van a daganatok egyes mutációinak az autofágia hálózatra gyakorolt hatását bemutatni regulációs interakciókon keresztül is (Kubisch és mtsai, 2013) (24. ábra). Az adatbázis használata szerepel az autofágia kutatás jelenlegi ajánlott metodikái között (Klionsky és mtsai, 2016).

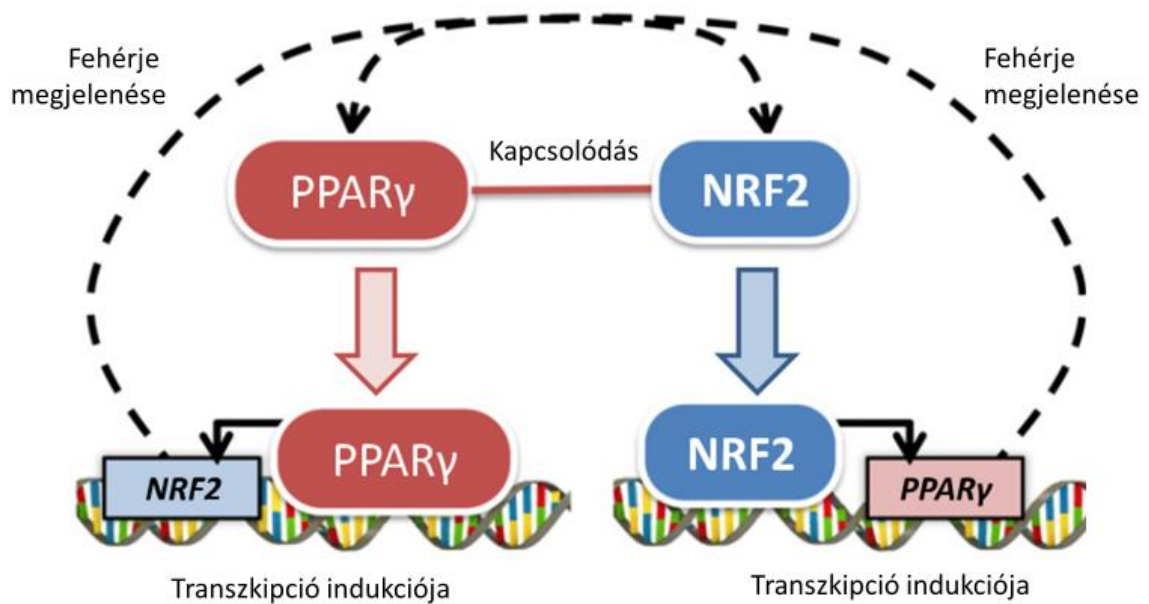


24. ábra Rákban mutációt szenvedő gének és gyógyszercélpontok az autofágia szabályozási hálózatában.

Középen az autofágia központi végrehajtó fehérjéi, felül a poszt-transzlációs szabályozók, alul pedig a transzkripció faktorok láthatók. A pontok nagysága fokszámukkal arányos, azaz megmutatja, hogy hány autofágia fehérjét szabályoznak. Az

ábra szélén a nagy foksámú pontok felsorolása látható. A daganatos megbetegedésekben mutációt szenvedő gének zölddel kerültek kiemelésre, míg a gyógyszercélpontok késsel. A szabályozók közti kapcsolatok nincsenek feltüntetve (Kubisch és mtsai, 2013) alapján)

Az NRF2ome adatbázis az NRF2 és interakciós partnereinek kigyűjtésére szolgál. Lehetőséget nyújt a miRNS és a transzkripciós faktorok regulációs köreinek bemutatására. Egyik legjobb példa erre az NRF2 és a PPAR-gamma közötti regulációs kör (25. ábra) (Papp és mtsai, 2012).



25. ábra A PPAR-gamma és az NRF2 közötti transzkripcionális visszacsatolási kapcsolat.

Az NRF2 serkenti a PPAR-gamma transzkripcióját, míg a PPAR az NRF2 transzkripcióját serkenti. Ha a két fehérje egyszerre jelen van a rendszerben, akkor összekapcsolódnak, ezáltal nem képesek bejutni a sejtmagba, azaz kölcsönösen gátolják egymás transzkripciós hatását. A visszaható (*feed-back*) kapcsolathoz a regulációs adatok nélkülözhetetlenek. Az általam is keresett regulációs adatok kerültek be később az NRF2-ome adatbázisba (Papp és mtsai, 2012).

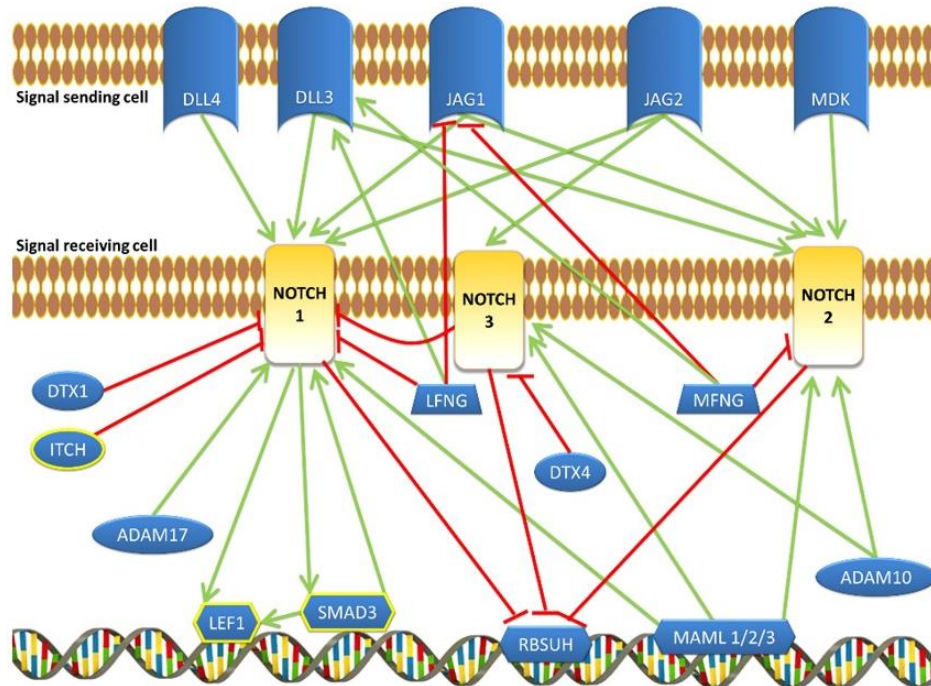
2. Példák a kritikus paralóg csoportokra, szerepük a jelátvitelben

A megbeszélésben szeretném ismertetni néhány kritikus paralóg szerepét és funkcióját az intracelluláris kommunikációban, illusztrálva jelentőségüket a jelátvitelben.

1.1. A NOTCH fehérjék:

A Notch útvonalban a névadó fehérje három izoformája a NOTCH1, a NOTCH2 és a NOTCH3 található egy kritikus paralóg csoportban. (a NOTCH4 evolúciósan nagyobb távolságra található, ezért nem tagja a paralóg csoportnak).

A NOTCH1 fehérje a három NOTCH paralóg közül a legjobban ismert és a legsokoldalúbb. Ez folytat legerősebben cross-talkot, a Notch útvonalon kívül a WNT és a TGF- β útvonallal is kapcsolatban áll. A WNT útvonallal a LEF-1 fehérjén, a TGF- β útvonallal a SMAD3 fehérjén keresztül folytat cross-talkot (Blokzijl és mtsai, 2003; Ross és Kadesch, 2001). Szabályozó fehérjéi közül az ITCH a TGF- β útvonallal is kapcsolatban van (Bai és mtsai, 2004), valamint a TGF- β útvonal egyik központi fehérjéje, a SMAD3 is visszahat a NOTCH1 működésére (Blokzijl és mtsai, 2003) (26. ábra). Mutációk az egyes NOTCH paralógokban eltérő betegségeket okoznak. A NOTCH1 szerepe ismert a T-limfociták érésében (Stanley és Guidos, 2009), és mutációja bikuszpidaális aorta billentyűt okoz (Garg és mtsai, 2005). A NOTCH2 mutációja az Alagile szindrómát okozza, ami a máj epevezetékének az elzáródásával jár (McDaniell és mtsai, 2006). A NOTCH3 génben bekövetkező mutáció vaszkuláris malformációt okoz a központi idegrendszerben (Kalimo és mtsai, 2006). A három receptor a szervezet különböző pontján okoz egymástól eltérő betegségeket.



26. ábra A NOTCH kritikus pont hálózata.

A kritikus pontba három NOTCH izoforma tartozik, közülük a NOTCH1-ről tudjuk, hogy rendelkezik cross-talk kapcsolatokkal (Blokzijl és mtsai, 2003; Ross és Kadesch, 2001). A pontok alakja az útvonal-pozícióban betöltött funkciót, a színek az útvonaltagságot jelzik. A pontok között közvetlen serkentő és gátló kapcsolatok vannak, de az LFN3 és MFNG fehérjéknél szaggatott vonallal közvetett kapcsolatokat is láthatunk. Ezek a ligandok receptor-specifikus serkentését vagy gátlását jelzik. A SMAD3 speciális helyzetű, mivel egyszerre lehet mediátor, kofaktor és transzkripciós faktor is több útvonalban. Az ábrán jól látható, hogy a NOTCH1 receptor a SMAD3 és a LEF1 fehérjékkel cross-talk kapcsolatban van, és ez indokolja, hogy a NOTCH1 erős, pozitív és negatív szabályozás alatt áll. (Módos és mtsai, 2016)

A három izoformának vannak azonos és vannak különböző ligandjai és kofaktorai (26. ábra). Ennek következtében előfordulhat az, hogy akár mindegyik jelenlévő receptor-izoforma aktiválódik, de az is, hogy ligand-függő szelektív módon csak egy adott izoforma típus. A Notch útvonal szabályozása szempontjából a kofaktorok kiemelten fontosak, mivel – szemben például az EGF/MAPK útvonallal – itt nincsen sok központi elem, amely közvetlenül befolyásolná a jel terjedését (Bray, 2006; Korcsmáros és mtsai, 2010)(Bray, 2006; Korcsmáros és mtsai, 2010). A 11. ábrán azt is láthatjuk, hogy vannak általános (MAML1, -2, -3), és vannak receptor-specifikus kofaktorok (DTX1, -4,

ADAM17, -10), amelyek befolyásolhatják, hogy ennek a kritikus pontnak melyik tagja lesz aktív. Fontos azt is kiemelni, hogy az egyes kritikus ponti fehérjék akár egymás hatását is befolyásolhatják. A NOTCH3 képes gátolni a NOTCH1 Notch útvonal-specifikus célgénjeinek átíródását (Beatus és mtsai, 1999). Egy ilyen komplex kritikus ponti hatás növeli a NOTCH1 más útvonalakat aktiváló hatásának specifikusságát.

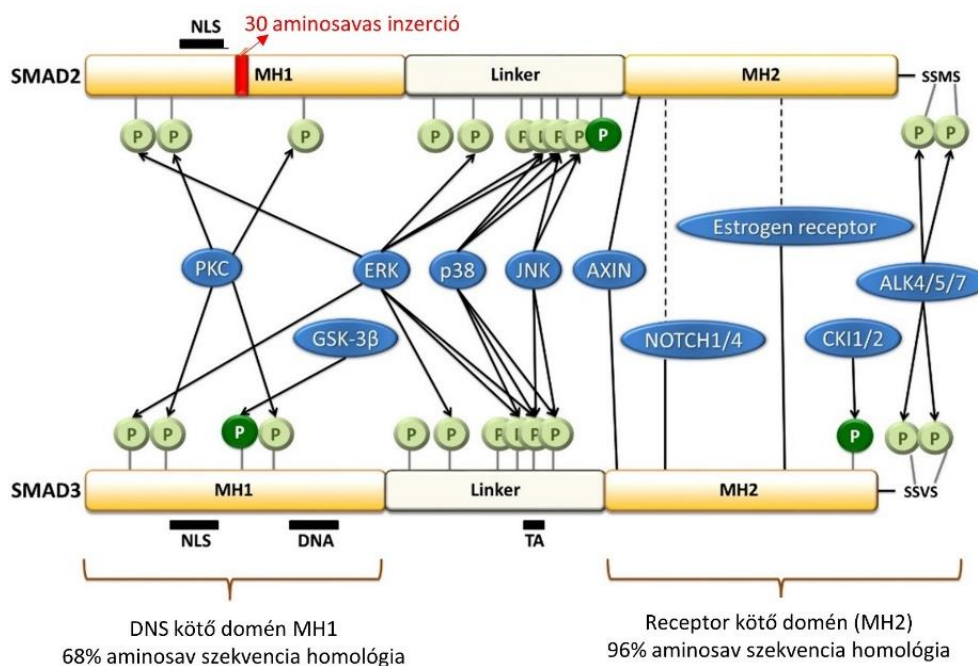
Összességében ez a kritikus pont és szabályozása arra példa, hogy a kritikus ponti fehérjékre ható kofaktorok segítségével lehetőség van a bejövő jelet más útvonalakra terelni. Ez segít abban, hogy a jel jelátviteli útra specifikus maradjon. Elképzelhető, hogy ez a specifikusság okozza a különböző betegségeket az eltérő receptorok mutációiban.

1.2. A SMAD 2 és 3 fehérje:

R-SMADok közé tartozó SMAD2 és SMAD3 fehérje TGF- β jel továbbításában vesz részt. A TGF- β jel továbbításához mindkét izoformának jelen kell lennie, mivel az aktivált receptor jelét egy dupla heterodimer komplex képződésével továbbítják a sejtmag felé (Lutz és Knaus, 2002). Az evolúció során a SMAD2 duplikálódott. A két paralóg szekvenciáinak 90 %-a egyforma, azonban a SMAD2 MH1 doménjén (3. exonon) egy 30 aminosavas inzert található (Dennler és mtsai, 1999). Ennek következtében a SMAD2 – szemben a SMAD3-mal – nem tud a DNS-hez kötődni, és több más fehérjéhez sem tud kapcsolódni (Jayaraman és Massagué, 2000; Shi és mtsai, 1998). Mindkét fehérje részt vesz cross-talkban a WNT és a RTK/MAPK jelátviteli útvonalban (Luo, 2017) (27. ábra.) Mindkét fehérjének fontos szerepe van az embrionális differenciálódás, beleértve az epitéliális-mezenchimális átalakulás szabályozásában. A SMAD2 felelős az embrió dorso-ventralis tengelyének kialakulásáért, míg a SMAD3 negatívan regulálja a sejtciklust és apoptózist indukál (The UniProt Consortium, 2014).

Ha csak az egyik paralóg van jelen, vagy a paralógok aránya eltérő, akkor a bejövő TGF- β jel egy másik jelátviteli útvonalat befolyásol a domináns paralógon keresztül. Például a TGF- β aktiválhatja a WNT útvonalat a SMAD3-on keresztül, a SMAD3 leállíthatja az axin expressziót, amely a WNT jelátvitel egyik negatív regulátora (Dao és mtsai, 2007) (27. ábra). Ez a cross-talk fontos a chondrociták differenciációjában (Dao és mtsai, 2007). A SMAD3 több olyan cross-talk foszforilációs hellyel rendelkezik, amivel a SMAD2 nem (27. ábra). Például a GSK-3 β foszforilációs helye a 66. treonin aminosavon. Ez a foszforiláció a SMAD3 ubiquitinizálódásának a jele (Guo és mtsai, 2008). A foszforilációhoz kapcsolódnia kell az axinhoz, de oda a SMAD2 is

kapcsolódhat, ami nem tartalmazza ezt a treonint, tehát megvédi a SMAD3-t a degradációtól.



27. ábra A SMAD2 és a SMAD3 fehérjék szerkezete.

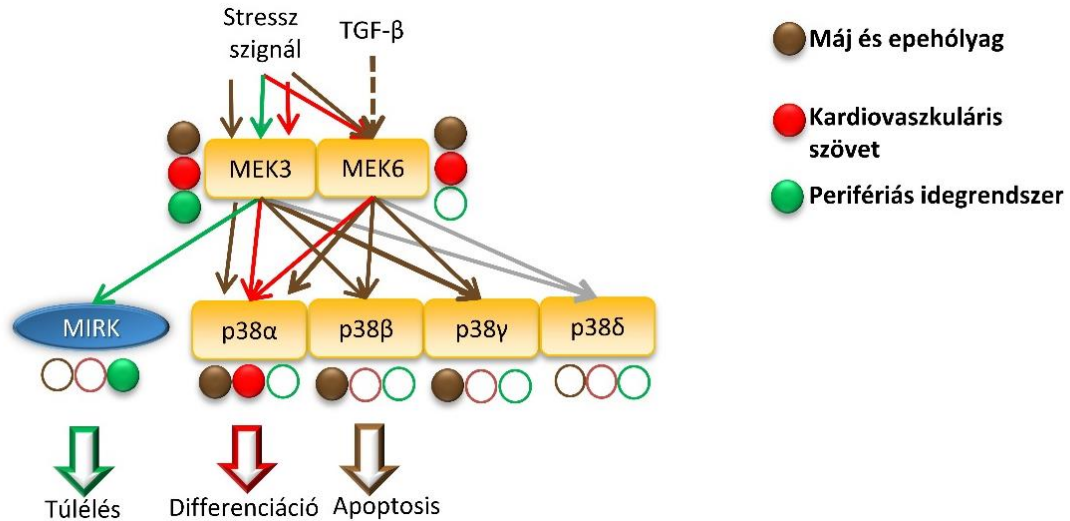
P-vel jelöltük a különböző foszforilációs helyeket. A téglalapok jelzik a különböző doméneket. A SMAD2 DNS kötő doménje rendelkezik egy 30 aminosavból álló inszercióval ami miatt a SMAD2 nem képes a DNS-hez kötődni, csak a SMAD3-mal együtt. (piros téglalap) Mindkét fehérje képes cross-talkra például az axinhoz való kapcsolódás, vagy ERK foszforiláció útján. A SMAD3 kapcsolódni képes a GSK-3 β -val is, ami a WNT jelátvitel egyik központi fehérjéje. (Módos és mtsai, 2016)

Mindebből azt láthatjuk, hogy a kritikus pont két elemének a jelátvitelben eltérő hatása van, bár nagyban homológ a szekvenciájuk. Feltehetően az evolúció során létrejövő duplikáció tette lehetővé a SMAD2-ben a 30 aminosavas inszerciót. Ez az inszerció pedig nagymértékű funkcióváltozást eredményezett. Az új funkciónak fokozatosan újfajta szabályozási lehetőségei alakultak ki. Továbbá láthatjuk, hogy a cross-talkban jobban involvált kritikus ponti fehérje nagyobb mértékben is szabályozott, épp úgy, mint a NOTCH1 fehérje esetén.

1.3. A MAPK fehérjék szövet specifikus expressziója:

Az expressziós reguláció fontosságát a MAPK jelátviteli útvonal kritikus paralóg csoportjain mutatom be. Ismert, hogy a MAPK jelátviteli útvonalat több, egymáshoz

hasonló fehérjéből felépülő, paralóg csoport alkotja (Seger és Krebs, 1995). A példában csak a szövet specifikus expresszióra fókuszálunk, mely befolyásolja a bejövő jel szövetspecifikus hatását (28. ábra).



28. ábra A MAPK jelátvitel kritikus paralóg csoportjai- élet és halál urai

Az egyszerűség és áttekinthetőség kedvéért nem ábrázoltuk a jelátvitel foszfatázait valamint állványfehérjéit. Három kiválasztott szövetben mutatjuk be az expresszióját a két kritikus paralóg csoportnak: A máj és epehólyag területén (barna karika), a kardiovaszkuláris szövetekben (piros karika) és a perifériás idegrendszerben (zöld karika). Ha az adott fehérje expresszálódik az adott szövetben, akkor teli karikával jeleztük. Ha nem expresszálódott, akkor üres karikával. Az egyes fehérjék közötti élek az adott szövet színével láthatók. A három szövetben eltér a jelátvitel útja és a végkifejlete mindössze az expresszió miatt. A részleteket lásd a szövegben. (Módos és mtsai, 2016)

A MEK3 és a MEK6-nak nincs ismert célpont sepcificitása a p38 paralógok között (p38α, p38β, p38γ, p38δ) scaffoldok nélkül. A végső hatás azon múlik, hogy melyik p38 paralóg expresszálódik az adott szövetben. Ha a p38α expresszálódik, akkor a MEK3/6-p38 út indul be, mely myoblastokban a miocitikus differenciálódást jelenti (Choi és mtsai, 2011). A májban a p38β aktiválódása a MEK6-on keresztül ismert proapoptotikus út, melyet a TGF-β útvonallal folytatott cross-talk aktivál (Kim és mtsai, 2004), és hepatocellularis fibrosishoz vezethet (Dooley és ten Dijke, 2012). A perifériális idegrendszerben egyik p38 fehérje sincs jelen. Ilyenkor a MEK szignál a MIRK fehérjék

felé mehet tovább. Ez egy ismert túlélési jel a perifériás idegrendszerben (Mercer és Friedman, 2006).

Összességében tehát elmondható, hogy a sejt sorsa azon is múlik, hogy a bejövő jelet milyen p38 izoformák fogadják, azaz egy kritikus paralóg csoporton belül milyen kritikus fehérjék vannak jelen. A példában bemutatott jelátviteli hálózatban a bejövő jel túlélési jellé (idegrendszer), sejthalált közvetítő jellé (máj), vagy differenciáció felé terelt jellé (szív) alakulhat. Ez a kritikus paralóg csoport kiváló lehetőséget jelent a jelátvitel áthuzalozására: apoptotikus szignálok növekedési vagy túlélési jellé való alakítására. Végeredményben az expresszáldó kritikus paralógok élet és halál urai a sejt számára.

3. A kritikus paralóg csoportok jelentősége

A jelátviteli hálózatok specifikusan képesek nagy mennyiségű különböző bementi jelet feldolgozni és tovább juttatni viszonylag kevés különböző jelátviteli útvonalon. A jelátviteli útvonalak szabályozása kulcs lépés a sejtek életében és döntéseik meghozatalában. Éppen ezért a hálózatbiológia egyik kérdése a jelátvitel kulcsfontosságú elemeinek megtalálása (Jeong és mtsai, 2001). A korábbi munkák, vagy kizárólagosan hálózat alapú (Khuri és Wuchty, 2015; Li és mtsai, 2013b; Wang és mtsai, 2012), vagy csak biológiai módszereket alkalmaztak (Taniguchi és mtsai, 2006), de előfordult ezen módszerek kombinációja is (Li és mtsai, 2017; Luo és Qi, 2015).

Az egyik legnagyobb hatású analízist Kahn és munkatársai végezték az inzulin jelátviteli útvonal vizsgálatakor (Taniguchi és mtsai, 2006). Azonosították az inzulin jelátviteli útvonal kritikus pontjait, ahol egy kritikus pontba olyan fehérjék tartoztak, melyek egymás paralógjai vagy izoformái voltak, részt vettek cross-talkban az inzulin jelátviteli útvonalon kívül más útvonalakkal, és esszenciálisak voltak az inzulin jelátvitelben.

Alternatíva a jelátvitel központi elemeinek megtalálásra a hálózatbeli centralitás alapú keresés. Itt az egymással kapcsolódó jelátviteli útvonalakat gráfként reprezentálják, és a centralitási mérőszámokkal prediktálják az esszenciális kritikus pontokat (Khuri és Wuchty, 2015; Li és mtsai, 2013b; Wang és mtsai, 2012). A gráf reprezentációk hátránya, hogy függenek a hálózatról alkotott *a priori* ismereteinktől, esetünkben a jelátviteli hálózat interakcióitól. Ahhoz, hogy ne essünk bele az ilyen jellegű csapdába, valamilyen biológiai ismeretet hozzá kell tennünk a hálózati mérőszámokhoz. Ilyen lehet, hogy az esszenciális fehérjék nagyobb arányban expresszáldódnak együtt, és alkotnak közös

modulokat (Li és mtsai, 2012), vagy az, hogy az ilyen fehérjék nagyobb arányban tagjai komplexeknek, mint ahogy azt várnánk a közönséges fehérjékkel szemben (Li és mtsai, 2013b, 2017; Luo és Qi, 2015).

Doktori dolgozatom során kombináltam ezeket a tulajdonságokat és kiterjesztettem Kahn és munkatársainak az inzulin jelátvitelre alkalmazott kritikus pont koncepcióját további útvonalakra (Taniguchi és mtsai, 2006). Megőriztem evolúciós kritériumként azt, hogy hasonló gének fontosak lehetnek a jelátvitelben, ezért paralógokat kerestem két egymást kiegészítő forrás segítségével (Kriventseva és mtsai, 2015; Sonhammer és Östlund, 2015). Az esszencialitást szövetspecifikusan és jelátvitelre jellemző módon definiáltam. Azt a fehérjét tekintettem esszenciálisnak, mely nélkül elvész legalább egy szövetben egy ligand és transzkripció faktor közötti jelátviteli út. Ezzel több réteg biológiai ismeretet vittem be a munkába. Egyrészt a ligandtól transzkripció faktorig tartó utak csak a jelátvitelben értelmezhetőek, a gráf elméletben alkalmazott forrás és nyelő pontokkal szemben, másrészt szövetspecifikusságot, még az egyik fenti munka sem alkalmazta, csupán koexpressziót. Ez a kritérium egy szövet specifikus és jelátvitel specifikus azonosítást tette lehetővé a kritikus paralóg csoportoknak. A harmadik kritérium a paralóg csoportok közül legalább egy fehérjének cross-talkban való részvétele volt. Ehhez nélkülözhetetlen volt a megfelelő adatbázis kiválasztása, és ezért választottuk a SignaLink 2 adatbázist a munkához (Fazekas és mtsai, 2013) (7. ábra).

Az általam megalkotott munkafolyamat által sikerült 75 kritikus paralóg csoportot azonosítani, amiben összesen 267 kritikus fehérjét találtam az általam vizsgált hét jelátviteli útvonalban (8. ábra). Az eredmények fejezetben demonstráltam a kritikus paralóg csoportok fontosságát az egyes jelátviteli útvonalakban (26. ábra).

A kritikus paralóg csoportok különböznek a nem kritikus paralóg csoportoktól. Ez nemcsak a hálózati centralitási értékek abszolút értékében mutatkozik meg, amit a kritikus paralóg csoportok szelektálása alapján várnánk, hanem az egy csoporton belüli szórás értékek alapján is. Emellett a kritikus paralóg csoportoknak sokkal több paralóg specifikus funkciója is van. Ez egybevág az irodalom azon téziseivel, miszerint az esszenciális fehérjék magas fokszámmal (Jeong és mtsai, 2001), és különbözőbb biológiai funkcióval (Albert, 2005) rendelkeznek. A paralóg csoportokban található fehérjék definíciójuknak megfelelően szekvenciájuk alapján hasonlóak egymáshoz. Ezek a

hasonló fehérjék az evolúciós adaptáció és innováció forrásai a jelátvitelben, illetve bármilyen funkcióban (Csermely, 2008; Fraser, 2005). A három, általam bemutatott példa közül a SMAD2 és SMAD3 mutatja leginkább azt, hogyan képes az evolúció kihasználni a részben hasonló szerkezetet (27. ábra).

Gén duplikáció után az egyes paralógok regulációja is szelekciós nyomás alá kerülhet. Munkám során azt találtam, hogy a miRNS-eken keresztüli regulációja a kritikus paralóg csoportoknak sokkal diverzebb, a nem kritikus paralóg csoportokkal összehasonlítva. Más kutatások is alátámasztják, hogy a központi fehérjék sokkal erőteljesebb regulációval rendelkeznek (Hsu és mtsai, 2008; Jeong és mtsai, 2001). Érdekes módon a kritikus fehérjék esetén ez poszt-transzkripcionálisan valósul meg, míg a transzkripción keresztüli reguláció hasonló a nem kritikus paralóg csoportok és a kritikus paralóg csoportok között (10. ábra). Ezek alapján egyetértek azzal a feltételezéssel, hogy a miRNS reguláció evolvíbilisabb, mint a transzkripció faktorokon alapuló reguláció (Chen és Rajewsky, 2007). A miRNS-ek a jelátvitel finom hangolásáért felelősek a transzkripciót követően (Ichimura és mtsai, 2011), ezért lehetőségük van az egyes paralógok megfelelő arányának a beállítására. Ezt a MAPK jelátvitelen keresztül mutattam be (28. ábra). Más regulációs lehetőségek is elképzelhetők, mint a kofaktorokon keresztüli reguláció, ahogy a Notch jelátvitelben láttuk (26. ábra), vagy a különböző foszforilációs helyek, amiket a SMAD2 és SMAD3 példáján ismertettem (27. ábra).

A betegséghez és daganatokhoz kapcsolódó mutációk gyakoribbak a kritikus fehérjék között, mint a nem kritikus paralóg fehérjék között (11. ábra). A genetikailag öröklődő betegségek specifikusabbak voltak a kritikus paralóg csoportokon belül, mint a nem kritikus paralóg csoportokon belül (11. ábra). A jelátvitelben a daganatot okozó és betegségekkel kapcsolatos mutációk száma magas (Creixell és mtsai, 2015; Hanahan és Weinberg, 2011). Ezért a kritikus paralógok vizsgálata során is a teljes genomhoz képest relatíve sok mutációra lehetett számítani. Ennek függvényében meglepő, hogy a nem kritikus paralóg fehérjék között a mutációk kevésbé voltak gyakoriak, mint az egyéb fehérjék között. Feltehetően a paralóg fehérjék hasonló hálózati centralitása és biológiai funkciója állhat a jelenség mögött. Ebből arra lehet következtetni, hogy a paralóg fehérjék átvehetik egymás funkcióit. Tehát ahhoz, hogy a jel más útra terelődjön, egy nem kritikus paralóg csoporton belül szükség van minden egyes paralóg fehérjének a megváltoztatására. A más útra terelés lehet a daganatokban, örökletes betegségekben,

vagy éppen terápiásan. Ezért kevesebb a gyógyszer-célpontok száma a paralóg fehérjék között, és emelkedett a kritikus paralógok és az egyéb fehérjék között (11. ábra). A dolgozatban ismertetett módszer segíthet abban, hogy megtaláljuk a betegséget okozó és jó gyógyszer-célpont kritikus paralógokat, és el tudjuk őket különíteni az egyéb paralógoktól. Az új gyógyszer-célpontokat a kritikus paralógok között lehet érdemes keresni.

4. *A kritikus paralóg csoportok meghatározásának limitációi*

Munkámnak természetesen megvannak a limitációi. A legnagyobbat az alkalmazott adatbázis adja. Legjobb tudásom szerint a SignaLink2 adatbázis rendelkezik a legjobb biológiailag releváns cross-talk adatokkal, és precíz, visszakövethető biológiai és evolúciós útvonal besorolással (Fazekas és mtsai, 2013; Türei és mtsai, 2016; Vinayagam és mtsai, 2014). Munkánk tesztelhető egyéb adatbázisokon is, mint a Signor (Perfetto és mtsai, 2016), a Reactome (Croft és mtsai, 2014) vagy akár, ha a paralógok interakcióinak pontos besorolása megtörténik, a KEGG alapokon nyugvó XtalkDB adatbázison is (Sam és mtsai, 2017). Analízisünkben a cross-talk információ a limitáló tényező, de más adatbázis nem rendelkezik megfelelő cross-talk annotációval.

A munka másik kritikus lépése a paralógok csoportosítása. Erre az egymást kiegészítő két adatforrást használtam (OrthoDB és InParanoid (Kriventseva és mtsai, 2015; Sonnhammer és Östlund, 2015)), és ezzel tettem kísérletet az egy adatforrásból való torzítás kikerülésére.

A biológiai funkciók elemzésére használt Gene Ontology (Ashburner és mtsai, 2000) adatbázisban ismert a kutatási torzítás jelensége, mely sajnos elkerülhetetlenül több és jobban elemzett funkcióval látja el a jobban kutatott fehérjéket (Dessimoz és Škunca, 2016; Rhee és mtsai, 2008). A több kapcsolattal rendelkező fehérjék általában jobban is kutatottak, ezért ez a torzítás bármilyen kézi gyűjtésű adatbázisban megjelenik (Dessimoz és Škunca, 2016).

5. *Példák az első szomszédok gyógyszerészeti alkalmazására.*

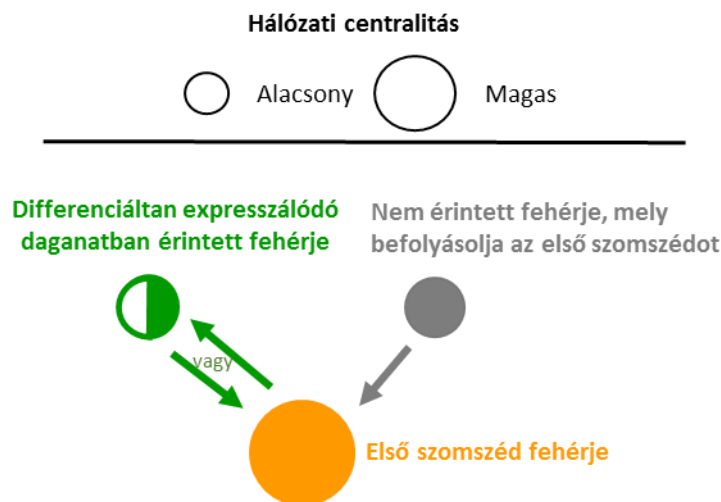
Ahogy a korábbiakban láttuk, a daganatban érintett fehérjék első szomszédjai központi helyet foglalnak el a jelátvitelben. Gyógyszerészeti célzásukra kétféle lehetőség tárulkozik fel. Az első a közvetlen célzása a centrális első szomszédoknak (*central hit*

strategy). Mivel az első szomszédok nagy központiséggel rendelkező fehérjék, célzásuk kétféle következményel járhat:

- (1) az első szomszédok célzása nagyobb mellékhatásokkal járhat, mint a daganatban érintett fehérjék célzása, mivel az első szomszédok által elérhető funkciók száma magasabb, mint a daganatban érintett fehérjék esetén,
- (2) az első szomszédok célzása nagyobb eséllyel vezet sejtthálához, mint a daganatban érintett fehérjék célzása, mivel központibb helyzetűek, és több funkcióval rendelkeznek.

Az első szomszédok célzása a daganatok esetében működhet, de nem minden esetben a legkifizetődőbb a nagy mennyiségű mellékhatás miatt.

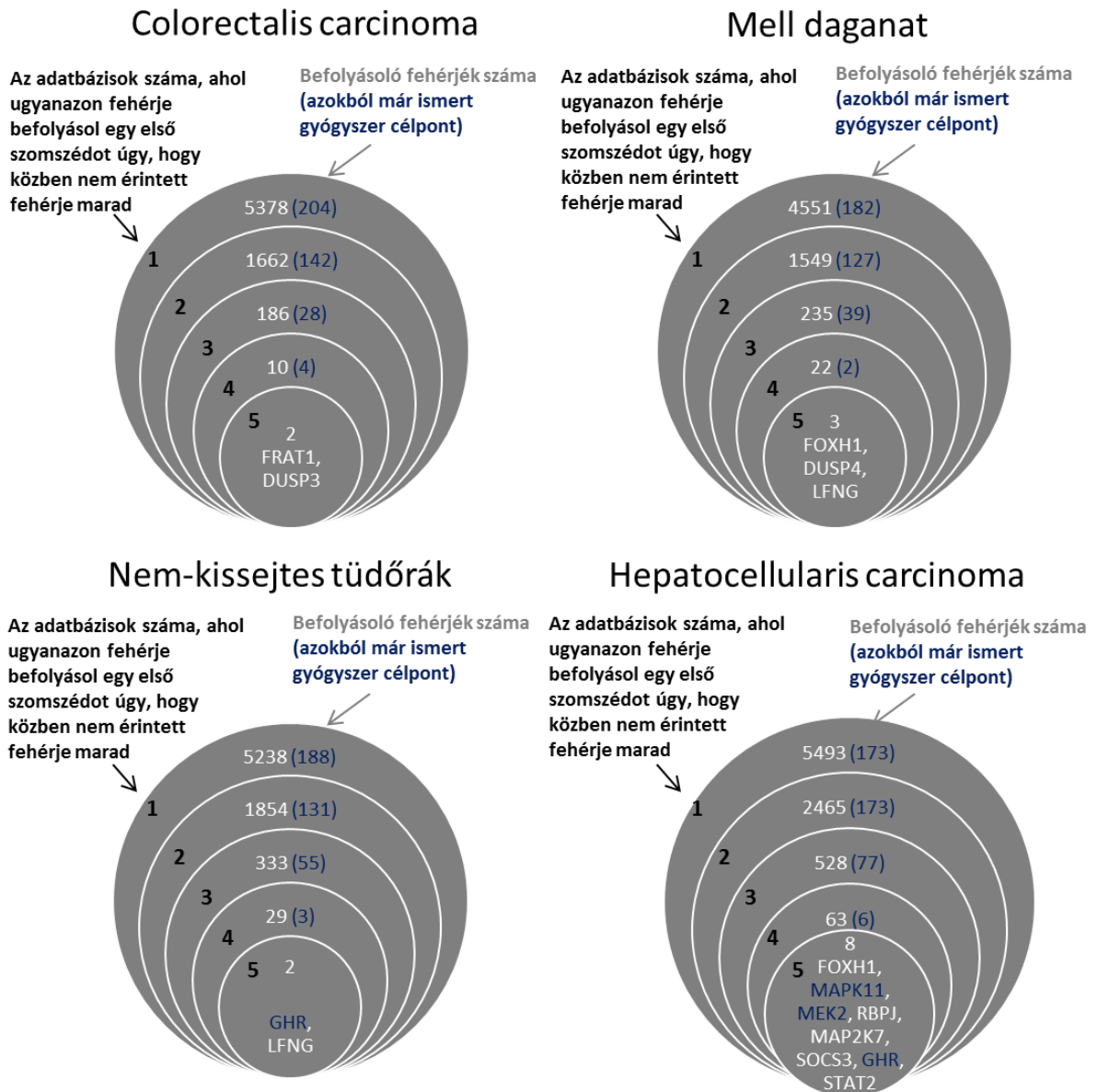
A másik lehetőség az első szomszédok befolyásolása (*influencer strategy*). Ilyenkor nem magát a centrális első szomszédot célozzuk meg, hanem inkább a mellette lévő, azt befolyásoló fehérjét. Ez a stratégia ahhoz hasonlít, mint amikor a daganatban megjelenő vagy eltűnő fehérje hat a centrális első szomszédjára. Ennek a stratégiának az előnye, hogy feltehetően kisebb mellékhatással rendelkezik, hiszen a normál sejtekben a perturbált fehérje kapcsolatainak több eleme hiányzik, amíg daganatokban jelen van. Hátránya viszont, hogy nehezen kimutatható, mivel több esetben sejt típusokon átívelő kísérletekre van szükség az igazolására.



29. ábra A befolyásoló fehérjék a jelátvitelben, mint további gyógyszer-célpont jelöltek. (Módos és mtsai, 2017)

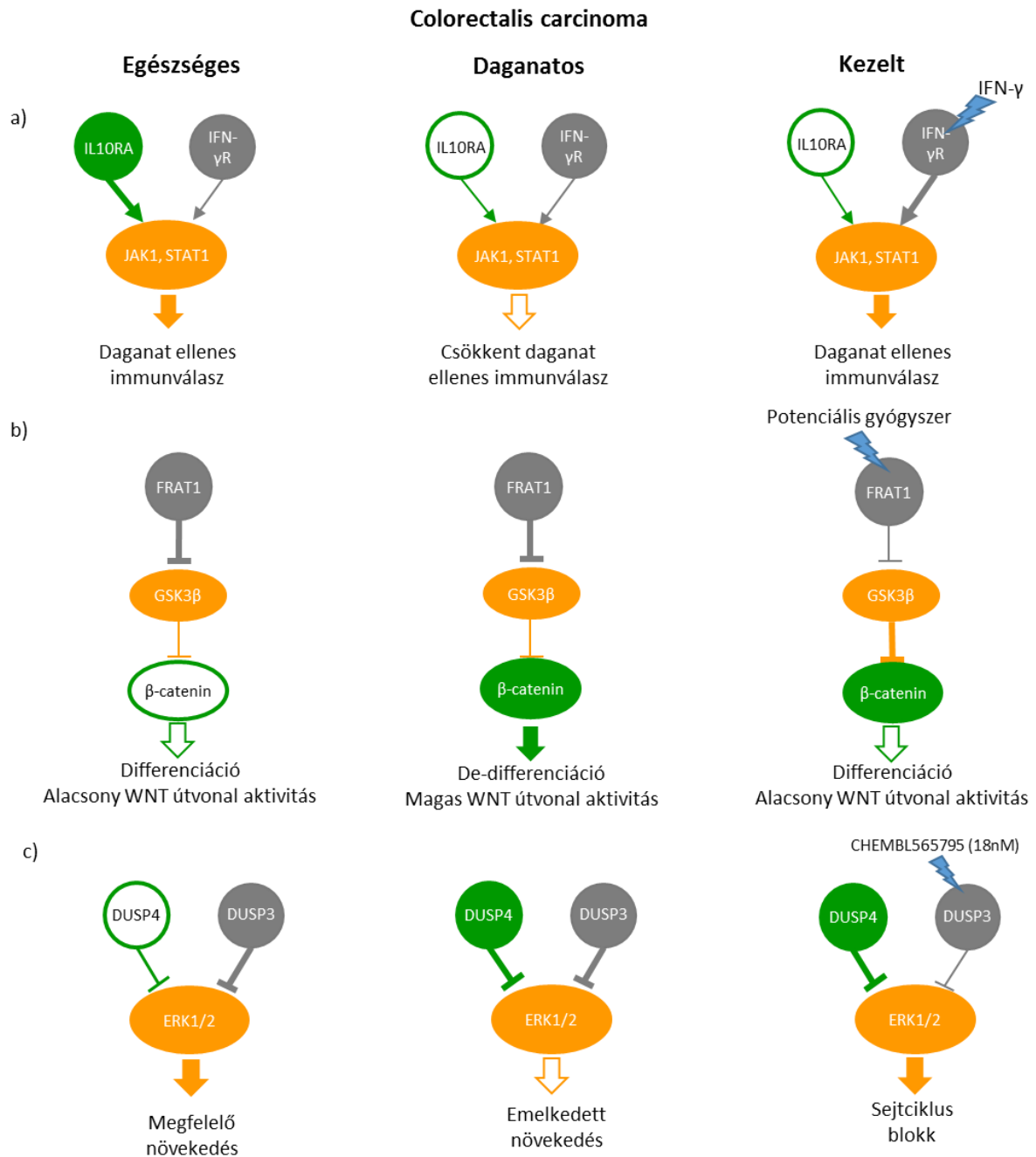
Példáimat aszerint választottam ki, hogy mind az öt, általunk vizsgált adatbázisban szerepeljenek és ott befolyásoló (*influencer*) pozícióban foglaljanak helyet (egy differenciáltan expresszázó daganatban érintett fehérje első szomszédjára hatnak). Az ilyen influencer fehérjék között több a gyógyszer-célpont, mint várnánk az adatbázisok összes fehérjéihez képest ($p < 0,001$ Khi négyzet teszt). A colorectalis carcinomából három példát ismertetek, a többi, általunk vizsgált daganatból egyet-egyet.

A colorectalis carcinomában az interferon- γ receptor (IFN- γ R) volt gyógyszer-célpont, mely legalább négy adatbázisban szerepelt befolyásoló pozícióban (30. ábra, 31. ábra). A DUSP3 (*Dual Specificity protein Phosphatase 3*) és a FRAT1 (*Frequently Rearranged in Advanced T-cell lymphoma*) fehérje bírt valamennyi adatbázisban befolyásoló szereppel, de egyikük sem gyógyszer-célpont. Az IFN- γ R már ismert gyógyszer-célpont, mivel az interferon- γ receptora és magát az interferon- γ -t használják különféle terápiákban (Miller és mtsai, 2009; Zaidi és Merlino, 2011). A három vastagbél daganatban található befolyásoló fehérje eltérő módon éri el az első szomszédot.



30. ábra A befolyásoló fehérjék előfordulása az egyes daganatokban aszerint, hogy hány adatbázisban fordultak elő, mint a daganatban érintett első szomszédokat befolyásoló, de az adott daganatban nem érintett fehérjék.

Késsel kiemelve közülük a már gyógyszer-célpont fehérjéket. (Módos és mtsai, 2017)



31. ábra A colorectalis carinomában három különböző befolyásoló fehérje gyógyszerként való alkalmazása

IFN- γ R a JAK/STAT, a FRAT1 a WNT és a DUSP3 MAPK jelátviteli útvonalakból. A zöld pontok az általunk vizsgált daganatban (mell, NSCLC, HCC) differenciáltan expresszált fehérjét jelzik. A narancssárga pontok az első szomszédokat jelzik, a szürke pontok a nem érintetteket, ebben a helyzetben a befolyásoló fehérjéket. A nyilak serkentést, a kalapács nyilak gátlást fejeznek ki. A bejövő jelet látjuk az egyes daganatokhoz tartozó ábrák felső részén. A villámok az adott fehérjére kifejtett

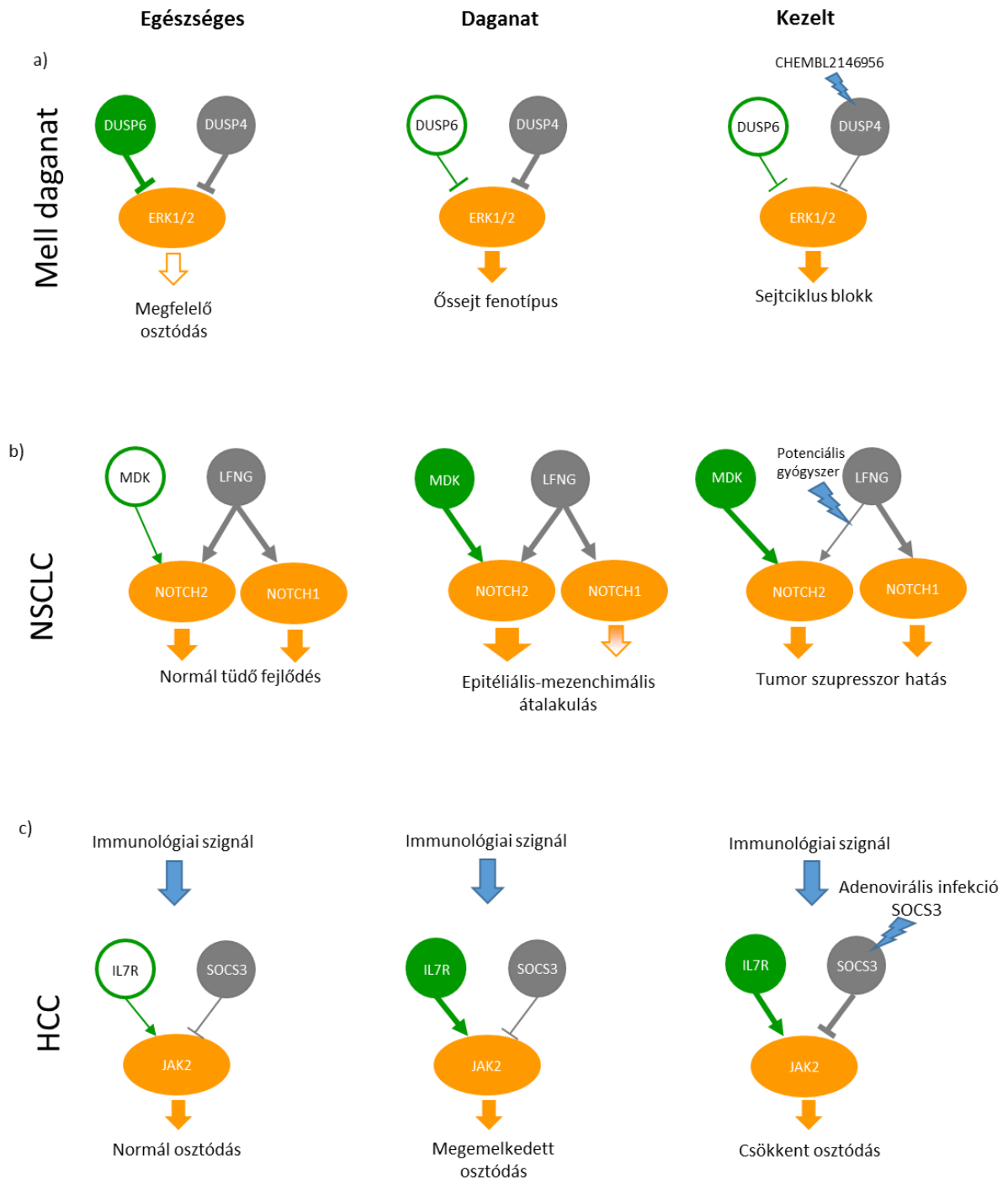
gyógyszerhatást jelölik. Az egyszerűség kedvéért a két INF- γ R alegységet (IFN- γ R1 és IFN- γ R2) egyként ábrázoltam. (Módos és mtsai, 2017)

Az IFN- γ R heterodimer, az IFN- γ R1 és az IFN- γ R2-ből formálódik. Az IFN- γ válasz részt vesz a tumorelles adaptív immunválaszban (Zaidi és Merlino, 2011). Az IFN- γ R ugyanazon JAK-STAT jelátviteli útvonalat, és ugyanúgy a JAK1 és a STAT1 fehérjéket használja, mint az IL-10R (interleukin 10 receptor). Viszont az IL-10 vizsgálatunk szerint eltűnik colorectalis carcinomában. Ezért elképzelhető, hogy az IFN- γ R helyettesítheti az IL-10 receptort a JAK-STAT jelátvitelben, és ezzel tumor ellenes hatást válthat ki. Az IFN- γ -t használják különböző tumorok ellen gyógyszerként (Zaidi és Merlino, 2011). Sajnos az IFN- γ kezelés nem mindig hatásos, és melanomában előfordult, hogy a kezelés inflamatorikus hatással bírt, mely végül a tumor növekedéséhez vezetett (Zaidi és Merlino, 2011). Viszont a kutatások egyetértenek abban, hogy érdemes megvizsgálni, mint lehetséges, új terápiás útvonalat.

A FRAT1, mely ismert protoonkogénként megtalálható mind az öt általunk vizsgált adatbázisban befolyásoló pozícióban (30. ábra, 31. ábra). A FRAT1 a WNT jelátviteli útvonal aktiválódásához járul hozzá a β -catenin GSK3 β általi foszforilálásának gátlásával (Guo és mtsai, 2013; Zhang és mtsai, 2012). A GSK3 β az első szomszédja a β -cateninnek, magas centralitási értékekkel rendelkező, multifunkcionális fehérje, ezért gátlása toxikus mellékhatásokkal járhat (Rayasam és mtsai, 2009). Glioblastomában és nem-kissejtes tüdőrákban korábbi vizsgálatok azt találták, hogy a FRAT1 gátlása emeli a GSK3 β β -catenin foszforiláló aktivitását, és ezzel gátolja a WNT jelátviteli utat (Guo és mtsai, 2013; Zhang és mtsai, 2012). A FRAT1 szerepe colorectalis carcinomákban kevésbé ismert, de más daganatokban betöltött szerepe, és a jelátviteli hálózatban betöltött befolyásoló pozíciója alapján releváns célpont lehet a colorectalis carcinoma ellenes gyógyszerek vizsgálata során.

DUSP3 volt a másik olyan fehérje, amely colorectalis carcinomában mind az öt adatbázisban befolyásoló fehérjeként szerepelt (30. ábra, 31. ábra). A DUSP3 befolyásoló szerepe az ERK1 és ERK2 (*Extracellular signal-regulated kinase*) első szomszédra irányuló gátló hatása adja (Patterson és mtsai, 2009). Az ERK1 és ERK2-re pedig egy másik dual specific foszfatáz, a DUSP4 hat daganatban érintett fehérjeként. Mindkét DUSP gátolja az ERK működéseket. Az ERK fehérjék mind növekedési és osztódási,

mind pedig apoptotikus jeleket közvetítenek a MAPK jelátvitelen keresztül (Lu és Xu, 2006). A daganat kialakulásához tehát megfelelő arányban szükségesek az ERK fehérjéken közvetített jelek. A DUSP4-ről ismert, hogy colorectalis carcinoma sejttenyészetben való expressziója és az ERK jelátvitel blokkolása proliferációhoz vezet (Gröschl és mtsai, 2013). A DUSP3 beépített fékként működik a sejtosztódás során: ERK1/2 általi proliferációs szignál megfelelő időpontban való gátlásában vesz részt. Kismolekulákkal való gátlása során a túlpörgetett sejtosztódási szignálok a sejtciklus gátlásába mennek át (Cerignoli és mtsai, 2006; Tambe és mtsai, 2016). Épp ezért colorectalis carcinomában a DUSP3 potenciális gyógyszer-célpont lehet, épp úgy mint cervicalis carcinomában, ahol DUSP3-gátló molekulák csökkentett osztódáshoz vezettek (Wu és mtsai, 2009). A DUSP3-t blokkoló molekulák nanomoláris tartományban találhatóak a ChEMBL adatbázisban, ami kiváló gyógyszerfejlesztési iránymutató (*lead*) molekulákká teszik őket a további gyógyszerfejlesztéshez (Wu és mtsai, 2009).



32. ábra A mell daganat a nem-kissejtes tüdőrákban és a hepatocellularis carcinómában mind az öt (Signalink, Cui és munkatársai, Reactome, HPRD, IntAct -DIP-BioGrid), általunk vizsgált interakciós hálózat szerint befolyásoló pozícióban lévő fehérjék, mint lehetséges gyógyszer-célpontok.

a) mell daganat b) nem-kissejtes tüdőrák (NSCLC) c) hepatocellularis carcinoma (HCC). A zöld pontok az általunk vizsgált daganatban (mell, NSCLC, HCC) differenciáltnan expresszált fehérjét jeleznek. A narancssárga pontok az első szomszédokat jelzik, a

szürke pontok a nem érintetteket, ebben a helyzetben a befolyásoló fehérjét. A nyilak serkentést, a kalapács nyilak gátlást fejeznek ki. A bejövő jelet látjuk az egyes daganatokhoz tartozó ábrák tetején. A villámok az adott fehérjére való gyógyszer általi hatást mutatják. (Módos és mtsai, 2017)

A mell daganatokban a DUSP4 nem a dagantokban érintett fehérje, hanem befolyásoló pozícióban található, és ugyanúgy az ERK1/2 fehérjén hat a daganatban érintett DUSP6-tal (32. ábra) A mell daganat kialakulása során a DUSP6 expressziója csökken adataink és az irodalom szerint egyaránt (Manzano és mtsai, 2014). Ez azt jelenti, hogy a mell daganat progressziója során az ERK1 és 2 egyre aktívabb lesz, és ebben az esetben ez az ERK onkogén mivoltát erősíti. Azonban a DUSP4 továbbra is elegendő ahhoz, hogy a sejtciklus megfelelő szakaszában gátolhassa az ERK szignált, és ezzel epithelialis-mesenchimalis átalakulást (EMT) idézzen elő a sejtciklus gátlásával ellentétben (Boulding és mtsai, 2016). A DUSP4 gátlása megelőzte az EMT-t, és sejtciklus gátlásához vezetett hasonlóan a DUSP3 vastagbél daganatokban történő gátlásához (Tambe és mtsai, 2016). Ezért a DUSP4 gátlók, mint amilyen a CHeMBL 2146956, lehetséges vezér molekulák lehetnek a daganatok ellen.

A nem-kissejtes tüdőrákban a Notch jelátviteli útvonal egyik tagja, az LFNG (*lunatic fringe*) található befolyásoló pozícióban mindegyik általunk vizsgált adatbázisban (32. ábra). Az LFNG képes mind a NOTCH1-en, mind a NOTCH2-n keresztüli jelet befolyásolni (Hicks és mtsai, 2000). A daganatban érintett fehérje ebben az esetben a mideikine (MDK), amely vizsgálatunk szerint a nem-kissejtes tüdőrákban expresszálódik, de a normál tüdőben nem. Az MDK ismert lehetséges daganat marker nem-kissejtes tüdőrákban (Yuan és mtsai, 2015). Medulloblastomában írták le az MDK NOTCH2 aktiváló hatását (Kishida és mtsai, 2013). Az LFNG a különböző Notch jelátviteli útvonalban található ligandok által közvetített serkentő jelet eltérő módon befolyásolja: mind a NOTCH1, mind a NOTCH2 Delta1 általi aktiválást serkenti. Ezzel ellentétben a Jagged ligand által továbbított jelet csak a NOTCH1 fehérje irányában aktiválja. Az aktív NOTCH1 tumor szupresszor hatással bír (Hicks és mtsai, 2000). A bejövő MDK jel viszont NOTCH2 irányt vesz fel, és a már meglévő LFNG jellel karöltve EMT átalakulását serkenti (Groth és Fortini, 2012). Lehetséges terápiás cél, hogy az LFNG NOTCH2 serkentő hatását megszüntessük, és NOTCH1 irányba tereljük a bejövő

jelet, ezzel ellenpontozva az MDK NOTCH2 irányú serkentését. Az LFNG NOTCH1 irányú tumor szupresszor hatása leírásra került már hasnyálmirigy daganatokban (Zhang és mtsai, 2016). Az ehhez hasonló ú.n. „*edgetic*”, élhatású gyógyszertervezés a dolgozatomban megírásakor az egyik felkapott, lehetséges gyógyszertervezési irány (Csermely és mtsai, 2013; Scott és mtsai, 2016). Egy lehetséges példa lehet az LFNG célzása és csak NOTCH1 irányba való aktiválása.

A hepatocelluláris carcinomában a SOCS3 (*Suppressor Of Cytokine Signalling 3*) befolyásoló pozícióban van mind az öt adatbázisban (32. ábra). A SOCS3 a JAK2 fehérjét gátolja (Subramaniam és mtsai, 2013). A daganatban érintett, differenciáltan expresszáldó fehérje HCC-ben az IL7R (Interleukin 7 receptor), mely nem expresszáldódik a normál májban, de megjelenik hepatocellularis carcinomában. Az IL7R aktiválódásával JAK2-n keresztül serkenti a sejtosztódást (Inagaki-Ohara és mtsai, 2013). A JAK2 gátlásával a SOCS3 képes csökkent proliferációt okozni a máj daganat sejtekben. Ezért a máj daganatokkal szembeni terápiás megoldás lehet a SOCS3 expressziójának a megemelése. Erre történtek már kísérletek: a SOCS3-t adenovirális vektorral juttatták be hepatocellularis carcinoma sejtekbe. Ez a HCC sejtek líziséhez vezetett, míg a normál májsejtekben nem történt változás (Cui és mtsai, 2007b). A SOCS3 tehát kiváló célpontja lehet a daganat elleni terápiának.

6. A daganatban érintett fehérjék első szomszédjainak a jelentősége

A daganatban érintett fehérjék első szomszédjai legalább annyira központiak, mint maguk a daganatban érintett fehérjék. A teljes jelátviteli hálózat megváltozásához az első szomszédok nélkülözhetetlen közvetítő elemek a daganatban érintett fehérjék és a jelátviteli hálózat többi része között. Úgy működnek, mint a ragasztó, összekötik a daganatban érintett fehérjéket a jelátvitel többi részével (12. ábra). Jelenleg jóval kevesebb gyógyszer és *compound* jut egy első szomszéd fehérjére, mint egy daganatban érintett fehérjére, ám ezek száma még így is jelentős (23. ábra). Munkánk során két eltérő stratégiát találtunk, amellyel a daganatban érintett fehérjék képesek áthuzalozni a jelátvitelt. 1. A mutálódó fehérjék önmagukban közvetlenül huzalozzák át a jelátvitelt. Első szomszédjaik egyenrangú partnerek a hálózati centralitás szempontjából. 2. Ezzel szemben a differenciáltan expresszáldó daganatban érintett fehérjék befolyásolják nagyobb központiságú első szomszédjaikat, és első szomszédjaikon keresztül érnek el változást a jelátvitelben. Mindkét esetet kihasználhatjuk gyógyszertervezésre. A mutáció

esetén *downstream* eső a mutált fehérjéhez hasonló központiságú első szomszédra tervezhetünk gyógyszert, ezzel meggátolva/adott irányba terelve a mutáció hatását a jelátvitelben. Erre a legismertebb, jelenleg is inezíven kutatott példa az onkológiában a K-RAS klinika alkalmazása (Asati és mtsai, 2017). A differenciáltan expresszálódó gének esetén ugyanolyan stratégiát alkalmazhatunk a jelátvitel normálissá való huzalozására, mint amit a rák alkalmazott. A centrális első szomszédokra ható befolyásoló fehérjéket célozva az eltűnő fehérje hatását helyettesíthetjük, mint ahogy az INF γ receptor esetén láttuk (29. ábra), vagy a bejövő jelet terelhetjük megfelelő irányába, ahogy a NOTCH fehérjéknél láttuk a nem-kissejtes tüdőrák esetén (32. ábra).

A vizsgálatunk során öt különféle hálózatot vizsgáltunk, 3 jelátviteli és 2 fehérje-fehérje interakciós hálózatot. Mindegyikből vizsgáltam egy-egy nagy áteresztőképességű módszerekkel készültet (Reactome illetve BIOGRID-IntAct -DIP összevont hálózat), és két kézi gyűjtésű jelátvitelit (SingaLink, Cui és munkatársai), illetve egy kézi gyűjtésű fehérje-fehérje interakciós hálózatot (HPRD). Mindegyik esetben az eredmények konzisztensek voltak.

A daganatban érintett fehérjék interakciós hálózatbeli központiságát több tanulmány is igazolta, ezzel saját munkák során is egybehangzó eredményeket kaptam (Cui és mtsai, 2007a; Li és mtsai, 2009, 2013c; Xia és mtsai, 2011) A daganatban érintett fehérjék interakciós hálózatbeli központiságát több tanulmány is igazolta, ezzel saját munkák során is egybehangzó eredményeket kaptam (Cui és mtsai, 2007a; Li és mtsai, 2009, 2013c; Xia és mtsai, 2011). Viszont a saját vizsgálatainkban a daganatban érintett fehérjék nem alkottak egy összefüggő gráfot, mint más tanulmányokban (Chang és mtsai, 2013; Menche és mtsai, 2015). Ennek két oka lehet: egyrészt a hálózatot leszűkítettük az adott daganatra, specifikus hálózatra az expressziós értékeke alapján, másrészt a daganatban érintett fehérjéket nem minden daganatra, hanem az adott daganatra jellemzően választottuk ki. Mindennek tükrében viszont fontos elmondani, hogy az első szomszédok így is legalább akkora centralitási értékekkel (köztiség, foksám) rendelkeztek öt hálózatban, mint a daganatban érintett fehérjék.

A gyógyszerhatások hálózatos elemzése kimutatta, hogy a legnagyobb hatékonyságot és mellékhatást egyaránt a legnagyobb centralitású célpontokat célzó gyógyszerek okozzák (Qabaja és mtsai, 2014; Wang és mtsai, 2013). A mellékhatás a betegségekben érintett géntől való távolsággal egyenesen arányos, viszont az első

szomszédok még kellő távolságban találhatóak, hogy a gyógyszereknek lehetőségük legyen terápiás hatást kifejteni.

Természetesen az első szomszédok gyógyszerként történő alkalmazásának megvannak a korlátai, elég, ha magas centralitásukra gondolunk, amely feltehetően sok mellékhatást vonhat maga után (Csermely és mtsai, 2013). Szükség van pontos kísérletes elemzésre. Erre jelenleg is folytatott vizsgálatokat a Cambridge-i Egyetemen és az indiai Bangalore Egyetemen. A befolyásolási stratégia esetén a daganatban érintett fehérjétől való nagy távolság lehet a limitáló tényező ezért a hatás elmaradhat a kívánt mértéktől. A fentiekben bemutatott két stratégia egymással komplementer lehet. Ha találunk olyan gyógyszereket, melyek jelenleg nem alkalmazottak az adott daganat terápiájában, de egy első szomszédra hatnak, új indikációkat mutathatunk ezen gyógyszerek számára (*off label* alkalmazhatjuk őket). A bemutatott példákhoz hasonló befolyásoló stratégiát alkalmazva pedig a daganatokhoz hasonló stratégiával felülírhatjuk a hibás jelátvitelt.

7. *A daganatban érintett fehérjék első szomszédjainak meghatározásának limitációi*

Nem vettük figyelembe sem a transzkripcionális (transzkripció faktorok), sem a poszt transzkripcionális (miRNA) szabályozást. A hálózatokon csak azt vizsgáltuk, mely gének találhatóak meg a hálózatokban microarray vizsgálatok alapján. A későbbiekben lehetőség lenne RNA-seq, vagy protein-chip adatok alapján újra vizsgálni az eredményeket. A daganatban érintett fehérjék meghatározásakor mindössze a mutálódó és differenciáltan expresszálódó elemeket vettük figyelembe, nem vizsgáltuk a metilációt, kromoszóma átrendeződéseket. A változásokat daganat specifikusan néztük. A CGC szűk listáját használtuk mutált fehérjék azonosításához, és nem vontuk össze a különböző daganatok hálózatát.

A vizsgálatunk során természetesen további adatbázisok bevonására is lehetőség lett volna, de mivel az eredmények konzisztensek voltak a vizsgált 5 különböző hálózatban, melyet több adatbázisból állítottam össze, ezért megbízhatónak tekinthetőek.

VII. Következtetések

PhD munkám során sikeresen felépítettem a SignaLink, az Autofágia Regulációs Hálózat (ARN) és az NRF2ome adatbázis regulációs rétegeit. A fenti adatbázisok közreműködésemmel alkalmassá váltak a regulációs körök vizsgálatára, amik egyes lehetőségeit a dolgozat diszkussziójában ismertettem.

A három adatbázis közül a SignaLink2-t használtam a jelátviteli hálózatok analízisének bemutatására. Sikeresen azonosítottam és analizáltam a kritikus paralóg csoportokat és kritikus paralógokat az ember legfontosabb 7 jelátviteli útvonalában. Ehhez két egymást kiegészítő paralóg kereső adatforrást használtam. Az általam bemutatott metodika képes lehet a jelenleg nem a jelátvitelbe tartozó paralógok jelátviteli útvonalakhoz való rendelésére.

A paralógok megtalálása után összeállítottam egy munkafolyamatot a kritikus paralógok és paralóg csoportok azonosítására. Ehhez jelátviteli tulajdonságként a cross-talkokban betöltött szerepet használtam, és esszenciálisnak vettem azokat a fehérjéket melyek nélkül eltűnik a jelátvitelben egy adott szövetben a ligandtól a transzkripció faktorig tartó jelátviteli út. Mindössze ez a két kritérium elegendőnek bizonyult, hogy megkülönböztethessük a kritikus és a nem kritikus paralóg csoportokat egymástól. A jelátviteli útvonalakhoz való annotáció segíthet abban, hogy megállapítsuk, melyek lesznek a központi elemek az egymáshoz hasonló fehérjék között. Ezzel lehetőség nyílik a gyógyszercélpontok keresésére és azonosítására, hiszen a kritikus paralóg fehérjék között jóval több a betegségben részt vevő fehérje és a gyógyszercélpont, mint a paralóg fehérjék között. Három egymástól eltérő példával illusztráltam hogyan alakulhat át a jelátvitel egy kritikus paralóg fehérje csoporton belül. A fehérje struktúra evolúciós változása új lehetséges utakat tárhat fel. A kofaktorokban való eltérés a cross-talkok finomhangolását teszi lehetővé. A paralógok specifikus aránya a kritikus paralóg csoportokon belül a jeleket a sejt életben maradása vagy halála között tudják mozgatni. A kritikus paralógok és a nem kritikus paralógok figyelembe vétele segítheti a gyógyszercélpontok azonosítását.

A daganatban érintett első szomszédok vizsgálatával rámutattam, hogy az alkalmazott gyógyszercélpontok körét érdemes kiterjeszteni a daganatban érintett fehérjék első szomszédjaival. Az első szomszédok ugyanolyan vagy magasabb centralitás

értékekkel rendelkeznek, mint a tumor fehérjék, és mint ragasztó tartják össze a jelátviteli hálózatot és küldik tovább a daganatban érintett fehérjék változásait.

Az egyes daganatokban a daganatban érintett fehérjék kétféle stratégiával érik el a teljes hálózatot: a drasztikusabb mutáció közvetlenül, jelátvitel centrálisabb elemeire hat. Ezzel szemben a differenciáltan expresszálódó daganatban érintett fehérjék közvetetten befolyásolják a jelátvitelt az első szomszédjaikon keresztül. Az első szomszédok figyelembe vételével egyes, már használt compoundoknak, vagy gyógyszereknek adhatunk új indikációt, és megnagyobbíthatjuk a gyógyszer célzási teret.

PhD munkám során tehát részt vettem több adatbázis létrehozásában, melyekben specifikus intracelluláris folyamatok vizsgálhatók, úgymint az autofágia, a jelátvitel és az oxidációs stressz. Az adatbázisok közül a SignaLink adatbázis elemzése során megállapítottam, melyek lehetnek azok a paralóg csoportok melyek jó gyógyszercélpontok lehetnek és munkámmal felhívtam a figyelmet a paralógok fontosságára a jelátvitelben. Végezetül a jelátviteli és fehérje-fehérje interakciós hálózatok figyelembe vételével bemutattam a daganatos fehérjék első szomszédjainak fontosságát, és azt, hogy ezek a fehérjék hogyan képesek a gyógyszercélpontok körét kitágítani.

VIII. Összefoglalás

A sejteket érő milliárdnyi bejövő információt a jelátviteli rendszer összegzi. Ennek megértéséhez jelátviteli adatbázisokra van szükség, de mind a mai napig hiányoznak a jelátviteli adatbázisokból a regulációs kapcsolatok. Doktori munkám során részt vettem az NRF2Ome, a SignaLink2 és az Autofágia Regulációs Adatbázis regulációs kapcsolatainak létrehozásában. A transzkripció adatok integrálása során transzkripció faktor kötőhely mátrixokat illesztettem az emberi génekhez és egy 2 kilobázisos flank szakaszhoz mindegyik irányban. Összesen 34 226 transzkripció faktor-célgén kapcsolatot határoztam meg.

A jelátvitel központi elemei megváltoznak betegségek kialakulása során. Doktori munkámban megvizsgáltam, hogy van-e kitüntetett szerepe a paralógoknak a jelátvitelben. Azt találtam, hogy a paralógok két csoportra oszthatók, egyrészt a többi jelátviteli fehérjéhez hasonlóan fontos és központi szereppel bíró kritikus paralógokra, valamint az egy paralóg csoporton belül egymást helyettesíteni képes, nem kritikus paralóg csoportokra. Közülük a kritikus paralóg csoportok gyakrabban vesznek részt daganatos és más betegségek kialakításában.

A daganatok vizsgálatakor legtöbbször magukat a jelátvitelben megváltozó fehérjéket elemzik, de figyelmen kívül hagyják interakciós környezetüket. Munkám során megkerestem négy szolid daganatban (colrectalis carcinoma, hepatocellularis carcinoma, nem-kissejtes tüdőrák, mell daganat) a mutálódó és differenciáltan expresszálódó géneket. Ezeket projektáltam három jelátviteli és két fehérje interakciós hálózatra, hogy megkeressem az első szomszédjaikat. A vizsgálat során megállapítottam, hogy a daganatban érintett fehérjék első szomszédjai legalább annyira központiak, mint a daganatban érintett fehérjék. Sőt, elmondható, hogy a daganatok kétféleképpen húzalozzák át a jelátvitelt. Egyrészt direkt mutációval, amikor centrális fehérjékre hatnak, másrészt indirekt módon, a differenciáltan expresszálódó fehérjék első szomszédjain keresztül. Ezt az információt felhasználva kerestem a daganatban érintett fehérjék első szomszédjai és interakciós partnereik között új gyógyszer-célpontokat.

Összefoglalásként PhD munkám során több biológiai folyamatot is érintve (jelátvitel, autofágia, stressz válasz) létrehoztam regulációs adatbázisokat, és azokat alkalmazva megmutattam a jelátvitelben a paralógok és a daganatokban az első szomszédok szerepét.

IX. Summary

Cell signalling systems process the incoming myriad information from outside the cells. To understand it signalling databases are required, but even today these often lack regulatory -transcriptional- interactions. During my PhD, I was involved in the construction of the NRF2Ome, the Signalink2 and the Autophagy Regulatory Network resources regulatory layers. To integrate the transcriptional data I aligned the transcription factor binding matrix is to the gens and a 2000 base pairs long flank sequence upstream and downstream. I predicted 34,226 transcription factor - target gene interactions.

The central part of the signalling network is changed during the pathogenesis of diseases. During my PhD, I researched whether the paralogs have any special importance in the signalling. I have found that the paralogs could be grouped into two groups: the critical paralogs (important and central) and the non-critical paralogs. The critical paralogs are not interchangeable, while the non-critical paralogs can take each other's function. The critical-paralogs are more often involved in cancer and other diseases.

In cancer signalling research, the directly mutated proteins tend to be first and foremost under investigation, but not their interaction partners. In my work, I searched for four cancer types (colorectal carcinoma, hepatocellular carcinoma, non-small cell lung cancer and breast cancer) the mutated genes and the differently expressed genes. I projected these cancer-related proteins, to three signalling networks (Signalink 2.0, Cui et al, Reactome) and two protein interaction networks to identify their first neighbours and to measure the network parameters. I found the first neighbours are at least as important as the cancer-related proteins themselves. Moreover, I found cancer rewire the signalling with different methods; on one hand by direct mutation of central proteins, on the other hand indirectly through the first neighbours of differently expressed proteins. After describing this phenomenon, I started identifying possible novel drug targets in the first neighbours of cancer-related proteins and their interactors.

In conclusion during my PhD work, I created regulatory databases in multiple biological processes (signalling, autophagy, stress response) and used them to show the paralogues' function in signalling and the first neighbours' role in cancer.

X. Irodalomjegyzék

Adamcsek B, Palla G, Farkas IJ, Derényi I és Vicsek T (2006) CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 22, 1021–1023.

Adler D (2015) Violin plot.

Agarwal S, Deane CM, Porter MA és Jones NS (2010) Revisiting Date and Party Hubs: Novel Approaches to Role Assignment in Protein Interaction Networks. *PLoS Comput. Biol.* 6, e1000817.

Aladzsity I, Toth ML, Sigmond T, Szabo E, Bicsak B, Barna J, Regos A, Orosz L, Kovacs AL és Vellai T (2007) Autophagy Genes *unc-51* and *bec-1* Are Required for Normal Cell Size in *Caenorhabditis elegans*. *Genetics* 177, 655–660.

Albert R (2005) Scale-free networks in cell biology. *J. Cell Sci.* 118, 4947–4957.

Alon U (2007) Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461.

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF és Hamosh A (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 43, D789–D798.

Anthonisse, Jac M (1971) The rush in a directed graph. *Sticht. Math. Cent. Math. Besliskd.* BN 9/71, 1–10.

Aparicio O, Geisberg J V és Struhl K (2004) Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. *Curr. Protoc. cell Biol.* Chapter 17, Unit 17.7.

Asati V, Mahapatra DK és Bharti SK (2017) K-Ras and its inhibitors towards personalized cancer treatment: Pharmacological and structural perspectives. *Eur. J. Med. Chem.* 125, 299–314.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S,

Matese JC, Richardson JE, Ringwald M, Rubin GM és Sherlock G (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.

Audet M, és Bouvier M (2012) Restructuring G-Protein- Coupled Receptor Activation. *Cell* 151, 14–23.

Auerbach RK, Chen B és Butte AJ (2013) Relating Genes to Function: Identifying Enriched Transcription Factors using the ENCODE ChIP-Seq Significance Tool. *Bioinformatics* 29, 1922–1924.

Bai Y, Yang C, Hu K, Elly C és Liu Y-C (2004) Itch E3 Ligase-Mediated Regulation of TGF- β Signaling by Modulating Smad2 Phosphorylation. *Mol. Cell* 15, 825–831.

Bandyopadhyay S, és Bhattacharyya M (2010) PuTmiR: a database for extracting neighboring transcription factors of human microRNAs. *BMC Bioinformatics* 11, 190.

Bansal M, Gatta GD és di Bernardo D (2006) Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics* 22, 815–822.

Barabasi, és Albert (1999) Emergence of scaling in random networks. *Science* 286, 509–512.

Barabási A-L (2016) *Network Science* Cambridge University Press.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S és Soboleva A (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41, D991–D995.

Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, Hinoue T, Laird PW, Curtis C, Shen H, Weisenberger DJ, Schultz N, Shen R, Weinhold N, Kelsen DP, Bowlby R, Chu A, Kasaian K, Mungall AJ, Gordon Robertson A, Sipahimalani P, Cherniack A, Getz G, Liu Y, Noble MS, Peadarallu C, Sougnez C, Taylor-Weiner A, Akbani R, Lee J-S, Liu W, Mills GB, Yang D, Zhang W, Pantazi A, Parfenov M, Gulley M, Blanca Piazuelo M, Schneider BG, Kim J, Boussioutas A, Sheth M, Demchok J a.,

Rabkin CS, Willis JE, Ng S, Garman K, Beer DG, Pennathur A, Raphael BJ, Wu H-T, Odze R, Kim HK, Bowen J, Leraas KM, Lichtenberg TM, Weaver S, McLellan M, Wiznerowicz M, Sakai R, Lawrence MS, Cibulskis K, Lichtenstein L, Fisher S, Gabriel SB, Lander ES, Ding L, Niu B, Ally A, Balasundaram M, Birol I, Brooks D, Butterfield YSN, Carlsen R, Chu J, Chuah E, Chun H-JE, Clarke A, Dhalla N, Guin R, Holt R a., Jones SJM, Lee D, Li H a., Lim E, Ma Y, Marra M a., Mayo M, Moore R a., Mungall KL, Ming Nip K, Schein JE, Tam A, Thiessen N, Beroukhim R, Carter SL, Cherniack AD, Cho J, DiCara D, Frazer S, Gehlenborg N, Heiman DI, Jung J, Kim J, Lin P, Meyerson M, Ojesina AI, Sekhar Pedamallu C, Saksena G, Schumacher SE, Stojanov P, Tabak B, Voet D, Rosenberg M, Zack TI, Zhang H, Zou L, Protopopov A, Santoso N, Lee S, Zhang J, Mahadeshwar HS, Tang J, Ren X, Seth S, Yang L, Xu AW, Song X, Xi R, Bristow C a., Hadjipanayis A, Seidman J, Chin L, Park PJ, Kucherlapati R, Ling S, Rao A, Weinstein JN, Kim S-B, Lu Y, Mills G, Bootwalla MS, Lai PH, Triche Jr T, Van Den Berg DJ, Baylin SB, Herman JG, Murray B a., Arman Askoy RB, Ciriello G, Dresdner G, Gao J, Gross B, Jacobsen A, Lee W, Ramirez R, Sander C, Senbabaoglu Y, Sinha R, Onur Sumer S, Sun Y, Thorsson V, Iype L, Kramer RW, Kreisberg R, Rovira H, Tasman N, Haussler D, Stuart JM, Verhaak RGW, Leiserson MDM, Taylor BS, Black AD, Ann Carney J, Gastier-Foster JM, Hessel C, McAllister C, Ramirez NC, Tabler TR, Wise L, Zmuda E, Penny R, Crain D, Gardner J, Lau K, Curely E, Mallery D, Morris S, Paulauskis J, Shelton T, Shelton C, Sherman M, Benz C, Lee J-H, Fedosenko K, Manikhas G, Potapova O, Voronina O, Belyaev S, Dolzhansky O, Kimryn Rathmell W, Brzezinski J, Ibbs M, Korski K, Kycler W, Łażniak R, Leporowska E, Mackiewicz A, Murawa D, Murawa P, Spychała A, Suchorska WM, Tatka H, Teresiak M, Abdel-Misih R, Bennett J, Brown J, Iacocca M, Rabeno B, Kwon S-Y, Kemkes A, Curley E, Alexopoulou I, Engel J, Bartlett J, Albert M, Park D-Y, Dhir R, Luketich J, Landreneau R, Janjigian YY, Cho E, Ladanyi M, Tang L, McCall SJ, Park YS, Cheong J-H, Ajani J, Constanza Camargo M, Alonso S, Ayala B, Jensen M a., Pihl T, Raman R, Walton J, Wan Y, Eley G, Mills Shaw KR, Tarnuzzer R, Wang Z, Yang L, Claude Zenklusen J, Davidsen T, Hutter CM, Sofia HJ, Burton R, Chudamani S és Liu J (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513, 202–209.

Beatus P, Lundkvist J, Oberg C és Lendahl U (1999) The notch 3 intracellular domain represses notch 1-mediated activation through Hairy/Enhancer of split (HES) promoters.

Development 126, 3925–3935.

Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, Krüger FA, Light Y, Mak L, McGlinchey S, Nowotka M, Papadatos G, Santos R és Overington JP (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* 42, D1083–D1090.

Berg C, Salvesen HB, Fenne O, Litleskare I, Rønning M, Sakshaug S, Strøm H, Granum T és Ullereng SG (2014) Guidelines for ATC classification and DDD assignment 2015 WHO Collaborating Centre for Drug Statistics Methodology Norwegian Institute of Public Health, Oslo.

Betel D, Wilson M, Gabow A, Marks DS és Sander C (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.* 36, D149–D153.

Betel D, Koppal A, Agius P, Sander C és Leslie C (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 11, R90.

Beyer A, Bandyopadhyay S és Ideker T (2007) Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet.* 8, 699–710.

BioGRID TC (2016) Physical Experimental Systems in BioGRID.

Blanco E, Farré D, Albà MM, Messeguer X és Guigó R (2006) ABS: a database of Annotated regulatory Binding Sites from orthologous promoters. *Nucleic Acids Res.* 34, D63–D67.

Blokzijl A, Dahlgvist C, Reissmann E, Falk A, Moliner A, Lendahl U és Ibáñez CF (2003) Cross-talk between the Notch and TGF- β signaling pathways mediated by interaction of the Notch intracellular domain with Smad3. *J. Cell Biol.* 163, 723–728.

Boulding T, Wu F, McCuaig R, Dunn J, Sutton CR, Hardy K, Tu W, Bullman A, Yip D, Dahlstrom JE és Rao S (2016) Differential Roles for DUSP Family Members in Epithelial-to-Mesenchymal Transition and Cancer Stem Cell Regulation in Breast Cancer. *PLoS One* 11, e0148065.

Bovolenta LA, Acencio ML és Lemke N (2012) HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics* 13, 405.

Bray SJ (2006) Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* 7, 678–689.

Breitkreutz D, Hlatky L, Rietman E és Tuszynski JA (2012) Molecular signaling network complexity is correlated with cancer patient survivability. *Proc. Natl. Acad. Sci. U. S. A.* 109, 9209–9212.

Briscoe J, és Théron PP (2013) The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* 14, 418–431.

Brückner A, Polge C, Lentze N, Auerbach D és Schlattner U (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* 10, 2763–2788.

Bulusu KC, Guha R, Mason DJ, Lewis RPI, Muratov E, Kalantar Motamedi Y, Cokol M és Bender A (2016) Modelling of compound combination effects and applications to efficacy and toxicity: State-of-the-art, challenges and perspectives. *Drug Discov. Today* 21, 225–238.

Cancer T, Atlas G, Weinstein JN, Akbani R, Broom BM, Wang W, Verhaak RGW, McConkey D, Lerner S, Morgan M, Creighton CJ, Smith C, Kwiatkowski DJ, Cherniack AD, Kim J, Sekhar Pedamallu C, Noble MS, Al-Ahmadie H a., Reuter VE, Rosenberg JE, Bajorin DF, Bochner BH, Solit DB, Koppie T, Robinson B, Gordenin D a., Fargo D, Klimczak LJ, Roberts S a., Au J, Laird PW, Hinoue T, Schultz N, Ramirez R, Hansel D, Hoadley K a., Kim WY, Damrauer JS, Baylin SB, Mungall AJ, Gordon Robertson A, Chu. A, Sougnez C, Cibulskis K, Lichtenstein L, Sivachenko A, Stewart C, Lawrence MS, Getz G, Lander E, Gabriel. SB, Donehower L, Carter SLS, Saksena G, Schumacher SE, Freeman SS, Jung J, Bhatt AS, Pugh T, Beroukhim R, Gabriel SB, Meyerson M, Chu A, Ally A, Balasundaram M, Butterfield YSN, Dhalla N, Hirst C, Holt R a., Jones SJM, Lee D, Li HI, Marra M a., Mayo M, Moore R a., Schein JE, Sipahimalani P, Tam A, Thiessen N, Wong T, Wye N, Bowlby R, Chuah E, Guin R, Shen H, Bootwalla MS, Triche Jr T, Lai PH, Van Den Berg DJ, Weisenberger DJ, Balu S, Bodenheimer T,

Damrauer Alan P, Hoyle JS, Jefferys SR, Meng S, Mose LE, Simons J V., Soloway MGM, Wu J, Parker JS, Neil Hayes D, Roach J, Buda E, Jones CD, Mieczkowski P a., Tan D, Veluvolu U, Waring S, Todd Auman J, Perou CM, Wilkerson MD, Santoso N, Parfenov M, Ren X, Pantazi A, Hadjipanayis A, Seidman J, Kucherlapati R, Lee S, Yang LL, Park PJ, Wei Xu A, Protopopov A, Zhang JJ, Bristow C, Mahadeshwar HS, Seth S, Song X, Tang J, Zeng D, Chin L, Guo C, Casasent TD, Liu W, Ju Z, Motter T, Peng B, Ryan M, Su X, Yang J-Y, Lorenzi PL, Yao H, Zhang N, Mills GB, Bernard B, Kreisberg D, Reynolds S, Rovira H, Shmulevich I, Gao J, Jacobsen A, Arman Aksoy B, Antipin Y, Ciriello G, Dresdner G, Gross B, Lee W, Reva B, Shen R, Sinha R, Onur Sumer S, Weinhold N, Ladanyi M, Sander C, Benz C, Carlin D, Haussler D, Ng S, Paull EO, Stuart J, Zhu J, Liu YY, Zhang W, Taylor BS, Lichtenberg TM, Zmuda E, Barr T, Black AD, George M, Hanf B, Helsel C, McAllister C, Ramirez NC, Tabler TR, Weaver S, Wise L, Bowen J, Gastier-Foster JM, Jian W, Tello S, Ittman M, Castro P, McClenden WD, Gibbs R, Saller C, Tarvin K, DiPiero JM, Owens J, Bollag R, Li Q, Weinberger P, Czerwinski C, Huelsenbeck-Dill L, Iacocca M, Petrelli N, Rabeno B, Swanson P, Shelton T, Curley E, Gardner J, Mallery D, Penny R, Van Bang N, Thi Hanh P, Kohl B, Van Le X, Duc Phu B, Thorp R, Viet Tien N, Quang Vinh L, Sandusky G, Burks E, Christ K, Gee J, Holway A, Moinzadeh A, Sorcini A, Sullivan T, Garcia-Grossman IR, Regazzi AM, Boice L, Kimryn Rathmell W, Thorne L, Bastacky S, Davies B, Dhir R, Gingrich J, Hrebinko R, Maranchie J, Nelson J, Parwani A, Bshara W, Gaudioso C, Morrison C, Alexopoulou V, Bartlett J, Engel J, Kodeeswaran S, Antic T, O'Donnell PH, Smith ND, Steinberg GD, Egea S, Gomez-Fernandez C, Herbert L, Jorda M, Beaver A, Kapur P, Lewis C, Lotan Y, Bondaruk J, Czerniak B, Skinner E, Aldape K, Jensen M a., Kahn AB, Pihl TD, Pot D a., Srinivasan D, Wan Y, Ferguson ML, Claude Zenklusen J, Davidsen T, Demchok J a., Mills Shaw KR, Sheth M, Tarnuzzer R, Wang Z, Hutter C, Ozenberger B a., Sofia HJ, Eley G és Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507, 315–322.

Cancer Genome Atlas Network, Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, Kovar CL, Lewis LR, Morgan MB, Newsham IF, Reid JG, Santibanez J, Shinbrot E, Trevino LR, Wu Y-Q, Wang M, Gunaratne P, Donehower LA, Creighton CJ, Wheeler DA, Gibbs RA, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, Stojanov P, McKenna A, Lander ES, Gabriel S, Getz G, Ding L, Fulton

RS, Koboldt DC, Wylie T, Walker J, Dooling DJ, Fulton L, Delehaunty KD, Fronick CC, Demeter R, Mardis ER, Wilson RK, Chu AAL, Chun H-JE, Mungall AJ, Pleasance E, Gordon Robertson A, Stoll D, Balasundaram M, Birol I, Butterfield YSN, Chuah E, Coope RJN, Dhalla N, Guin R, Hirst C, Hirst M, Holt RA, Lee D, Li HI, Mayo M, Moore RA, Schein JE, Slobodan JR, Tam A, Thiessen N, Varhol R, Zeng T, Zhao Y, Jones SJM, Marra MA, Bass AJ, Ramos AH, Saksena G, Cherniack AD, Schumacher SE, Tabak B, Carter SL, Pho NH, Nguyen H, Onofrio RC, Crenshaw A, Ardlie K, Beroukhir R, Winckler W, Getz G, Meyerson M, Protopopov A, Zhang J, Hadjipanayis A, Lee E, Xi R, Yang LL, Ren X, Zhang H, Sathiamoorthy N, Shukla S, Chen P-C, Haseley P, Xiao Y, Lee S, Seidman J, Chin L, Park PJ, Kucherlapati R, Todd Auman J, Hoadley KA, Du Y, Wilkerson MD, Shi Y, Liquori C, Meng S, Li L, Turman YJ, Topal MD, Tan D, Waring S, Buda E, Walsh J, Jones CD, Mieczkowski PA, Singh D, Wu J, Gulabani A, Dolina P, Bodenheimer T, Hoyle AP, Simons J V., Soloway M, Mose LE, Jefferys SR, Balu S, O'Connor BD, Prins JF, Chiang DY, Neil Hayes D, Perou CM, Hinoue T, Weisenberger DJ, Maglinte DT, Pan F, Berman BP, Van Den Berg DJ, Shen H, Triche Jr T, Baylin SB, Laird PW, Getz G, Noble M, Voet D, Saksena G, Gehlenborg N, DiCara D, Zhang J, Zhang H, Wu C-J, Yingchun Liu S, Shukla S, Lawrence MS, Zhou L, Sivachenko A, Lin P, Stojanov P, Jing R, Park RW, Nazaire M-D, Robinson J, Thorvaldsdottir H, Mesirov J, Park PJ, Chin L, Thorsson V, Reynolds SM, Bernard B, Kreisberg R, Lin J, Iype L, Bressler R, Erkkilä T, Gundapuneni M, Liu Y, Norberg A, Robinson T, Yang D, Zhang W, Shmulevich I, de Ronde JJ, Schultz N, Cerami E, Ciriello G, Goldberg AP, Gross B, Jacobsen A, Gao J, Kaczkowski B, Sinha R, Arman Aksoy B, Antipin Y, Reva B, Shen R, Taylor BS, Chan TA, Ladanyi M, Sander C, Akbani R, Zhang N, Broom BM, Casasent T, Unruh A, Wakefield C, Hamilton SR, Craig Cason R, Baggerly KA, Weinstein JN, Haussler D, Benz CC, Stuart JM, Benz SC, Zachary Sanborn J, Vaske CJ, Zhu J, Szeto C, Scott GK, Yau C, Ng S, Goldstein T, Ellrott K, Collisson E, Cozen AE, Zerbino D, Wilks C, Craft B, Spellman P, Penny R, Shelton T, Hatfield M, Morris S, Yena P, Shelton C, Sherman M, Paulauskis J, Gastier-Foster JM, Bowen J, Ramirez NC, Black A, Pyatt R, Wise L, White P, Bertagnolli M, Brown J, Chan TA, Chu GC, Czerwinski C, Denstman F, Dhir R, Dörner A, Fuchs CS, Guillem JG, Iacocca M, Juhl H, Kaufman A, Kohl III B, Van Le X, Mariano MC, Medina EN, Meyers M, Nash GM, Paty PB, Petrelli N, Rabeno B, Richards WG, Solit D, Swanson P, Temple L, Tepper

JE, Thorp R, Vakiani E, Weiser MR, Willis JE, Witkin G, Zeng Z, Zinner MJ, Zornig C, Jensen MA, Sfeir R, Kahn AB, Chu AAL, Kothiyal P, Wang Z, Snyder EE, Pontius J, Pihl TD, Ayala B, Backus M, Walton J, Whitmore J, Baboud J, Berton DL, Nicholls MC, Srinivasan D, Raman R, Girshik S, Kigonya PA, Alonso S, Sanbhadti RN, Barletta SP, Greene JM, Pot DA, Mills Shaw KR, Dillon LAL, Buetow K, Davidsen T, Demchok JA, Eley G, Ferguson M, Fielding P, Schaefer C, Sheth M, Yang LL, Guyer MS, Ozenberger BA, Palchik JD, Peterson J, Sofia HJ és Thomson. E (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337.

Cerignoli F, Rahmouni S, Ronai Z és Mustelin T (2006) Regulation of MAP Kinases by the VHR Dual-Specific Phosphatase – Implications for Cell Growth and Differentiation. *Cell Cycle* 5, 2210–2215.

Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, Ally A, Balasundaram M, Birol I, Butterfield YSN, Chu A, Chuah E, Chun H-JE, Dhalla N, Guin R, Hirst M, Hirst C, Holt R a, Jones SJM, Lee D, Li HI, Marra M a, Mayo M, Moore R a, Mungall AJ, Robertson a G, Schein JE, Sipahimalani P, Tam A, Thiessen N, Varhol RJ, Beroukhim R, Bhatt AS, Brooks AN, Cherniack AD, Freeman SS, Gabriel SB, Helman E, Jung J, Meyerson M, Ojesina AI, Pedamallu CS, Saksena G, Schumacher SE, Tabak B, Zack T, Lander ES, Bristow C a, Hadjipanayis A, Haseley P, Kucherlapati R, Lee S, Lee E, Luquette LJ, Mahadeshwar HS, Pantazi A, Parfenov M, Park PJ, Protopopov A, Ren X, Santoso N, Seidman J, Seth S, Song X, Tang J, Xi R, Xu AW, Yang LLL, Zeng D, Auman JT, Balu S, Buda E, Fan C, Hoadley K a, Jones CD, Meng S, Mieczkowski P a, Parker JS, Perou CM, Roach J, Shi Y, Silva GO, Tan D, Veluvolu U, Waring S, Wilkerson MD, Wu J, Zhao W, Bodenheimer T, Hayes DN, Hoyle AP, Jeffreys SR, Mose LE, Simons J V, Soloway MG, Baylin SB, Berman BP, Bootwalla MS, Danilova L, Herman JG, Hinoue T, Laird PW, Rhie SK, Shen H, Triche T, Weisenberger DJ, Carter SL, Cibulskis K, Chin L, Zhang JJJ, Getz G, Sougnez C, Wang M, Dinh H, Doddapaneni HV, Gibbs R, Gunaratne P, Han Y, Kalra D, Kovar C, Lewis L, Morgan M, Morton D, Muzny D, Reid J, Xi L, Cho J, DiCara D, Frazer S, Gehlenborg N, Heiman DI, Kim J, Lawrence MS, Lin P, Liu YY, Noble MS, Stojanov P, Voet D, Zhang H, Zou L, Stewart C, Bernard B, Bressler R, Eakin A, Iype L, Knijnenburg T, Kramer R, Kreisberg R, Leinonen K, Lin J, Liu YY, Miller MLMML, Reynolds SM,

Rovira H, Shmulevich I, Thorsson V, Yang D, Zhang W, Amin S, Wu C-JC-CC-J, Wu C-JC-CC-J, Akbani R, Aldape K, Baggerly K a, Broom B, Casasent TD, Cleland J, Creighton CJ, Dodda D, Edgerton M, Han L, Herbrich SM, Ju Z, Kim H, Lerner S, Li J, Liang H, Liu W, Lorenzi PL, Lu Y, Melott J, Mills GB, Nguyen L, Su X, Verhaak R, Wang W, Weinstein JN, Wong A, Yang Y, Yao J, Yao R, Yoshihara K, Yuan Y, Yung AK, Zhang N, Zheng S, Ryan M, Kane DW, Aksoy BA, Ciriello G, Dresdner G, Gao J, Gross B, Jacobsen A, Kahles A, Ladanyi M, Lee W, Lehmann K-V, Miller MLMML, Ramirez R, Rättsch G, Reva B, Sander C, Schultz N, Senbabaoglu Y, Shen R, Sinha R, Sumer SO, Sun Y, Taylor BS, Weinhold N, Fei S, Spellman P, Benz C, Carlin D, Cline M, Craft B, Ellrott K, Goldman M, Haussler D, Ma S, Ng S, Paull E, Radenbaugh A, Salama S, Sokolov A, Stuart JM, Swatloski T, Uzunangelov V, Waltman P, Yau C, Zhu J, Hamilton SR, Abbott S, Abbott R, Dees ND, Delehaunty K, Ding L, Dooling DJ, Eldred JM, Fronick CC, Fulton R, Fulton LL, Kalicki-Veizer J, Kanchi K-L, Kandoth C, Koboldt DC, Larson DE, Ley TJ, Lin L, Lu C, Magrini VJ, Mardis ER, McLellan MD, McMichael JF, Miller C a, O'Laughlin M, Pohl C, Schmidt H, Smith SM, Walker J, Wallis JW, Wendl MC, Wilson RK, Wylie T, Zhang Q, Burton R, Jensen M a, Kahn A, Pihl T, Pot D, Wan Y, Levine D a, Black AD, Bowen J, Frick J, Gastier-Foster JM, Harper H a, Hessel C, Leraas KM, Lichtenberg TM, McAllister C, Ramirez NC, Sharpe S, Wise L, Zmuda E, Chanock SJ, Davidsen T, Demchok J a, Eley G, Felau I, Ozenberger B a, Sheth M, Sofia H, Staudt L, Tarnuzzer R, Wang Z, Yang LLL, Zhang JJJ, Omberg L, Margolin A, Raphael BJ, Vandin F, Wu H-T, Leiserson MDM, Benz SC, Vaske CJ, Noushmehr H, Wolf D, Veer LV, Collisson E a, Anastassiou D, Yang T-HO, Lopez-Bigas N, Gonzalez-Perez A, Tamborero D, Xia Z, Li W, Cho D-Y, Przytycka T, Hamilton M, McGuire S, Nelander S, Johansson P, Jörnsten R, Kling T, Sanchez J, Shaw KRM, Mills GB, Shaw KRM, Ozenberger B a, Ellrott K, Shmulevich I, Sander C, Stuart JM, Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, Ally A, Balasundaram M, Birol I, Butterfield YSN, Chu A, Chuah E, Chun H-JE, Dhalla N, Guin R, Hirst M, Hirst C, Holt R a, Jones SJM, Lee D, Li HI, Marra M a, Mayo M, Moore R a, Mungall AJ, Robertson a G, Schein JE, Sipahimalani P, Tam A, Thiessen N, Varhol RJ, Beroukhim R, Bhatt AS, Brooks AN, Cherniack AD, Freeman SS, Gabriel SB, Helman E, Jung J, Meyerson M, Ojesina AI, Peadamallu CS, Saksena G, Schumacher SE, Tabak B, Zack T, Lander ES, Bristow C a, Hadjipanayis A, Haseley P, Kucherlapati R, Lee S, Lee E, Luquette LJ,

Mahadeshwar HS, Pantazi A, Parfenov M, Park PJ, Protopopov A, Ren X, Santoso N, Seidman J, Seth S, Song X, Tang J, Xi R, Xu AW, Yang LLL, Zeng D, Auman JT, Balu S, Buda E, Fan C, Hoadley K a, Jones CD, Meng S, Mieczkowski P a, Parker JS, Perou CM, Roach J, Shi Y, Silva GO, Tan D, Veluvolu U, Waring S, Wilkerson MD, Wu J, Zhao W, Bodenheimer T, Hayes DN, Hoyle AP, Jeffreys SR, Mose LE, Simons J V, Soloway MG, Baylin SB, Berman BP, Bootwalla MS, Danilova L, Herman JG, Hinoue T, Laird PW, Rhie SK, Shen H, Triche T, Weisenberger DJ, Carter SL, Cibulskis K, Chin L, Zhang JJJ, Getz G, Sougnez C, Wang M, Dinh H, Doddapaneni HV, Gibbs R, Gunaratne P, Han Y, Kalra D, Kovar C, Lewis L, Morgan M, Morton D, Muzny D, Reid J, Xi L, Cho J, DiCara D, Frazer S, Gehlenborg N, Heiman DI, Kim J, Lawrence MS, Lin P, Liu YY, Noble MS, Stojanov P, Voet D, Zhang H, Zou L, Stewart C, Bernard B, Bressler R, Eakin A, Iype L, Knijnenburg T, Kramer R, Kreisberg R, Leinonen K, Lin J, Liu YY, Miller MLMML, Reynolds SM, Rovira H, Shmulevich I, Thorsson V, Yang D, Zhang W, Amin S, Wu C-JC-CC-J, Wu C-JC-CC-J, Akbani R, Aldape K, Baggerly K a, Broom B, Casasent TD, Cleland J, Creighton CJ, Dodda D, Edgerton M, Han L, Herbrich SM, Ju Z, Kim H, Lerner S, Li J, Liang H, Liu W, Lorenzi PL, Lu Y, Melott J, Mills GB, Nguyen L, Su X, Verhaak R, Wang W, Weinstein JN, Wong A, Yang Y, Yao J, Yao R, Yoshihara K, Yuan Y, Yung AK, Zhang N, Zheng S, Ryan M, Kane DW, Aksoy BA, Ciriello G, Dresdner G, Gao J, Gross B, Jacobsen A, Kahles A, Ladanyi M, Lee W, Lehmann K-V, Miller MLMML, Ramirez R, Räscht G, Reva B, Sander C, Schultz N, Senbabaoglu Y, Shen R, Sinha R, Sumer SO, Sun Y, Taylor BS, Weinhold N, Fei S, Spellman P, Benz C, Carlin D, Cline M, Craft B, Ellrott K, Goldman M, Haussler D, Ma S, Ng S, Paull E, Radenbaugh A, Salama S, Sokolov A, Stuart JM, Swatloski T, Uzunangelov V, Waltman P, Yau C, Zhu J, Hamilton SR, Abbott S, Abbott R, Dees ND, Delehaunty K, Ding L, Dooling DJ, Eldred JM, Fronick CC, Fulton R, Fulton LL, Kalicki-Veizer J, Kanchi K-L, Kandoth C, Koboldt DC, Larson DE, Ley TJ, Lin L, Lu C, Magrini VJ, Mardis ER, McLellan MD, McMichael JF, Miller C a, O’Laughlin M, Pohl C, Schmidt H, Smith SM, Walker J, Wallis JW, Wendl MC, Wilson RK, Wylie T, Zhang Q, Burton R, Jensen M a, Kahn A, Pihl T, Pot D, Wan Y, Levine D a, Black AD, Bowen J, Frick J, Gastier-Foster JM, Harper H a, Helsen C, Leraas KM, Lichtenberg TM, McAllister C, Ramirez NC, Sharpe S, Wise L, Zmuda E, Chanock SJ, Davidsen T, Demchok J a, Eley G, Felau I, Ozenberger B a, Sheth M, Sofia H, Staudt L, Tarnuzzer R,

Wang Z, Yang LLL, Zhang JJJ, Omberg L, Margolin A, Raphael BJ, Vandin F, Wu H-T, Leiserson MDM, Benz SC, Vaske CJ, Noushmehr H, Wolf D, Veer LV, Collisson EA, Anastassiou D, Yang T-HO, Lopez-Bigas N, Gonzalez-Perez A, Tamborero D, Xia Z, Li W, Cho D-Y, Przytycka T, Hamilton M, McGuire S, Nelander S, Johansson P, Jörnsten R, Kling T, Sanchez J és Shaw KRM (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120.

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 és Tyers M (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* 43, D470–D478.

Chekulaeva M, és Filipowicz W (2009) Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr. Opin. Cell Biol.* 21, 452–460.

Chen J, és Zhang S (2016) Integrative analysis for identifying joint modular patterns of gene-expression and drug-response data. *Bioinformatics* 32, 1724–1732.

Chen K, és Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* 8, 93–103.

Chen Y, Hao J, Jiang W, He T, Zhang X, Jiang T és Jiang R (2013) Identifying potential cancer driver genes by genomic data integration. *Sci. Rep.* 3, 3538.

Choi TG, Lee J, Ha J és Kim SS (2011) Apoptosis signal-regulating kinase 1 is an intracellular inducer of p38 MAPK-mediated myogenic signalling in cardiac myoblasts. *Biochim. Biophys. Acta* 1813, 1412–1421.

Chowdhury S, és Sarkar RR (2015) Comparison of human cell signaling pathway databases--evolution, drawbacks and challenges. *Database* 2015, bau126.

Collisson EA, Campbell JD, Brooks AN, Berger AH, Lee W, Chmielecki J, Beer DG, Cope L, Creighton CJ, Danilova L, Ding L, Getz G, Hammerman PS, Neil Hayes D, Hernandez B, Herman JG, Heymach J V., Jurisica I, Kucherlapati R, Kwiatkowski D, Ladanyi M, Robertson G, Schultz N, Shen R, Sinha R, Sougnez C, Tsao M-S, Travis WD,

Weinstein JN, Wigle DA, Wilkerson MD, Chu A, Cherniack AD, Hadjipanayis A, Rosenberg M, Weisenberger DJ, Laird PW, Radenbaugh A, Ma S, Stuart JM, Averett Byers L, Baylin SB, Govindan R, Meyerson M, Rosenberg M, Gabriel SB, Cibulskis K, Sougnez C, Kim J, Stewart C, Lichtenstein L, Lander ES, Lawrence MS, Getz G, Kandoth C, Fulton R, Fulton LL, McLellan MD, Wilson RK, Ye K, Fronick CC, Maher CA, Miller CA, Wendl MC, Cabanski C, Ding L, Mardis E, Govindan R, Creighton CJ, Wheeler D, Balasundaram M, Butterfield YSN, Carlsen R, Chu A, Chuah E, Dhalla N, Guin R, Hirst C, Lee D, Li HI, Mayo M, Moore RA, Mungall AJ, Schein JE, Sipahimalani P, Tam A, Varhol R, Gordon Robertson A, Wye N, Thiessen N, Holt RA, Jones SJM, Marra M a., Campbell JD, Brooks AN, Chmielecki J, Imielinski M, Onofrio RC, Hodis E, Zack T, Sougnez C, Helman E, Sekhar Pedomallu C, Mesirov J, Cherniack AD, Saksena G, Schumacher SE, Carter SL, Hernandez B, Garraway L, Beroukhir R, Gabriel SB, Getz G, Meyerson M, Hadjipanayis A, Lee S, Mahadeshwar HS, Pantazi A, Protopopov A, Ren X, Seth S, Song X, Tang J, Yang L, Zhang J, Chen P-C, Parfenov M, Wei Xu A, Santoso N, Chin L, Park PJ, Kucherlapati R, Hoadley K a., Todd Auman J, Meng S, Shi Y, Buda E, Waring S, Veluvolu U, Tan D, Mieczkowski PA, Jones CD, Simons J V., Soloway MG, Bodenheimer T, Jefferys SR, Roach J, Hoyle AP, Wu J, Balu S, Singh D, Prins JF, Marron JS, Parker JS, Neil Hayes D, Perou CM, Liu J, Cope L, Danilova L, Weisenberger DJ, Maglinte DT, Lai PH, Bootwalla MS, Van Den Berg DJ, Triche Jr T, Baylin SB, Laird PW, Rosenberg M, Chin L, Zhang J, Cho J, DiCara D, Heiman D, Lin P, Mallard W, Voet D, Zhang H, Zou L, Noble MS, Lawrence MS, Saksena G, Gehlenborg N, Thorvaldsdottir H, Mesirov J, Nazaire M-D, Robinson J, Getz G, Lee W, Arman Aksoy B, Ciriello G, Taylor BS, Dresdner G, Gao J, Gross B, Seshan VE, Ladanyi M, Reva B, Sinha R, Onur Sumer S, Weinhold N, Schultz N, Shen R, Sander C, Ng S, Ma S, Zhu J, Radenbaugh A, Stuart JM, Benz CC, Yau C, Haussler D, Spellman PT, Wilkerson MD, Parker JS, Hoadley K a., Kimes PK, Neil Hayes D, Perou CM, Broom BM, Wang J, Lu Y, Kwok Shing Ng P, Diao L, Averett Byers L, Liu W, Heymach J V., Amos CI, Weinstein JN, Akbani R, Mills GB, Curley E, Paulauskis J, Lau K, Morris S, Shelton T, Mallery D, Gardner J, Penny R, Saller C, Tarvin K, Richards WG, Cerfolio R, Bryant A, Raymond DP, Pennell N a., Farver C, Czerwinski C, Huelsenbeck-Dill L, Iacocca M, Petrelli N, Rabeno B, Brown J, Bauer T, Dolzhanskiy O, Potapova O, Rotin D, Voronina O, Nemirovich-Danchenko E, Fedosenko K V., Gal A, Behera M,

Ramalingam SS, Sica G, Flieder D, Boyd J, Weaver J, Kohl B, Huy Quoc Thinh D, Sandusky G, Juhl H, Duhig E, Illei P, Gabrielson E, Shin J, Lee B, Rogers K, Trusty D, Brock M V., Williamson C, Burks E, Rieger-Christ K, Holway A, Sullivan T, Wigle DA, Asiedu MK, Kosari F, Travis WD, Rekhtman N, Zakowski M, Rusch VW, Zippile P, Suh J, Pass H, Goparaju C, Owusu-Sarpong Y, Bartlett JMS, Kodeeswaran S, Parfitt J, Sekhon H, Albert M, Eckman J, Myers JB, Cheney R, Morrison C, Gaudioso C, Borgia J a., Bonomi P, Pool M, Liptay MJ, Moiseenko F, Zaytseva I, Dienemann H, Meister M, Schnabel P a., Muley TR, Peifer M, Gomez-Fernandez C, Herbert L, Egea S, Huang M, Thorne LB, Boice L, Hill Salazar A, Funkhouser WK, Kimryn Rathmell W, Dhir R, Yousem S a., Dacic S, Schneider F, Siegfried JM, Hajek R, Watson M a., McDonald S, Meyers B, Clarke B, Yang IA, Fong KM, Hunter L, Windsor M, Bowman R V., Peters S, Letovanec I, Khan KZ, Jensen M a., Snyder EE, Srinivasan D, Kahn AB, Baboud J, Pot DA, Mills Shaw KR, Sheth M, Davidsen T, Demchok JA, Yang L, Wang Z, Tarnuzzer R, Claude Zenklusen J, Ozenberger BA, Sofia HJ, Travis WD, Cheney R, Clarke B, Dacic S, Duhig E, Funkhouser WK, Illei P, Farver C, Rekhtman N, Sica G, Suh J és Tsao M-S (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550.

Consortium TGO (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43, D1049–D1056.

Creixell P, Reimand J, Haider S, Wu G, Shibata T, Vazquez M, Mustonen V, Gonzalez-Perez A, Pearson J, Sander C, Raphael BJ, Marks DS, Ouellette BFF, Valencia A, Bader GD, Boutros PC, Stuart JM, Linding R, Lopez-Bigas N és Stein LD (2015) Pathway and network analysis of cancer genomes. *Nat. Methods* 12, 615–621.

Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L és D'Eustachio P (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res.* 42, D472–D477.

Csardi G, és Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Sy*, 1695.

Csermely P (2008) Creative elements: network-based predictions of active centres in proteins and cellular and social networks. *Trends Biochem. Sci.* 33, 569–576.

Csermely P, Korcsmáros T, Kiss HJM, London G és Nussinov R (2013) Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* 138, 333–408.

Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O’Connor-McCourt M, Purisima EO és Wang E (2007a) A map of human cancer signaling. *Mol. Syst. Biol.* 3.

Cui Q, Jiang W, Wang Y, Lv C, Luo J, Zhang W, Lin F, Yin Y, Cai R, Wei P és Qian C (2007b) Transfer of suppressor of cytokine signaling 3 by an oncolytic adenovirus induces potential antitumor activities in hepatocellular carcinoma. *Hepatology* 47, 105–112.

Dao DY, Yang X, Chen D, Zuscik M és O’Keefe RJ (2007) Axin1 and Axin2 are regulated by TGF- and mediate cross-talk between TGF- and Wnt signaling pathways. *Ann. N. Y. Acad. Sci.* 1116, 82–99.

Dennler S, Huet S és Gauthier J-M (1999) A short amino-acid sequence in MH1 domain is responsible for functional differences between Smad2 and Smad3. *Oncogene* 18, 1643–1648.

Dessimoz C, és Škunca N (2016) *The Gene Ontology Handbook* Humana Press.

Diehl AG, és Boyle AP (2016) Deciphering ENCODE. *Trends Genet.* 32, 238–249.

Dooley S, és ten Dijke P (2012) TGF- β in progression of liver disease. *Cell Tissue Res.* 347, 245–256.

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee B-K, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shores N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S,

Kellis M, Khatun J, Kheradpour P, Kundaje A, Lassmann T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SCJ, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LAL, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elnitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigó R, Hardison RC, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Snyder M, Stamatoyannopoulos JA, Tenenbaum SA, Weng Z, White KP, Wold B, Khatun J, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Giddings MC, Bernstein BE, Epstein CB, Shores N, Ernst J, Kheradpour P, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Ward LD, Altshuler RC, Eaton ML, Kellis M, Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Dutttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena HP, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Risk BA, Robyr D, Ruan X, Sammeth M, Sandhu KS, Schaeffer L, See L-H, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Hayashizaki Y, Harrow J, Gerstein M, Hubbard TJ, Reymond A, Antonarakis SE, Hannon GJ, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR, Rosenbloom KR, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kent WJ, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Furey TS, Song L, Grasfeder LL, Giresi PG, Lee B-K, Battenhouse A, Sheffield NC, Simon JM, Showers KA, Safi A, London D, Bhinge AA, Shestak C, Schaner MR, Ki Kim S, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniell RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Birney E, Iyer VR, Lieb JD, Crawford GE, Li G, Sandhu KS, Zheng M, Wang P, Luo OJ, Shahab A, Fullwood MJ, Ruan X, Ruan Y, Myers RM, Pauli F, Williams BA, Gertz J, Marinov GK, Reddy TE, Vielmetter

J, Partridge E, Trout D, Varley KE, Gasper C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, King B, Muratet MA, Antoshechkin I, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Gunter C, Newberry JS, Levy SE, Absher DM, Mortazavi A, Wong WH, Wold B, Blow MJ, Visel A, Pennachio LA, Elnitski L, Margulies EH, Parker SCJ, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Chrast J, Davidson C, Derrien T, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Howald C, Hunt T, Jungreis I, Kay M, Khurana E, Kokocinski F, Leng J, Lin MF, Loveland J, Lu Z, Manthravadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Sboner A, Searle S, Sisu C, Snow C, Steward C, Tanzer A, Tapanari E, Tress ML, van Baren MJ, Walters N, Washietl S, Wilming L, Zadissa A, Zhang Z, Brent M, Haussler D, Kellis M, Valencia A, Gerstein M, Reymond A, Guigó R, Harrow J, Hubbard TJ, Landt SG, Fietze S, Abyzov A, Addleman N, Alexander RP, Auerbach RK, Balasubramanian S, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Cheng C, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harmanci A, Iyengar S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Lamarre-Vincent N, Leng J, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patacsil D, Pei B, Raha D, Ramirez L, Reed B, Rozowsky J, Sboner A, Shi M, Sisu C, Slifer T, Witt H, Wu L, Xu X, Yan K-K, Yang X, Yip KY, Zhang Z, Struhl K, Weissman SM, Gerstein M, Farnham PJ, Snyder M, Tenenbaum SA, Penalva LO, Doyle F, Karmakar S, Landt SG, Bhanvadia RR, Choudhury A, Domanus M, Ma L, Moran J, Patacsil D, Slifer T, Victorsen A, Yang X, Snyder M, White KP, Auer T, Centanin L, Eichenlaub M, Gruhl F, Heermann S, Hoeckendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Weng Z, Whitfield TW, Wang J, Collins PJ, Aldred SF, Trinklein ND, Partridge EC, Myers RM, Dekker J, Jain G, Lajoie BR, Sanyal A, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Frum T, Garg K, Gist E, Hansen RS, Boatman L, Haugen E, Humbert R, Jain G, Johnson AK, Johnson EM, Kuttyavin T V., Lajoie BR, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri F V., Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sabo P, Sanchez ME, Sandstrom RS, Sanyal

A, Shafer AO, Stergachis AB, Thomas S, Thurman RE, Vernot B, Vierstra J, Vong S, Wang H, Weaver MA, Yan Y, Zhang M, Akey JM, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Kaul R, Dekker J, Stamatoyannopoulos JA, Dunham I, Beal K, Brazma A, Flicek P, Herrero J, Johnson N, Keefe D, Lusk M, Luscombe NM, Sobral D, Vaquerizas JM, Wilder SP, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Kundaje A, Hardison RC, Miller W, Giardine B, Harris RS, Wu W, Bickel PJ, Banfai B, Boley NP, Brown JB, Huang H, Li Q, Li JJ, Noble WS, Bilmes JA, Buske OJ, Hoffman MM, Sahu AD, Kharchenko P V., Park PJ, Baker D, Taylor J, Weng Z, Iyer S, Dong X, Greven M, Lin X, Wang J, Xi HS, Zhuang J, Gerstein M, Alexander RP, Balasubramanian S, Cheng C, Harmanci A, Lochovsky L, Min R, Mu XJ, Rozowsky J, Yan K-K, Yip KY és Birney E (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.

Dutta P, és Li WX (2013) Role of the JAK-STAT Signalling Pathway in Cancer. In eLS, John Wiley & Sons, Ltd, Chichester, UK, o. 1–9.

Dweep H, Sticht C, Pandey P és Gretz N (2011) miRWalk – Database: Prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J. Biomed. Inform.* 44, 839–847.

Edgar R, Domrachev M és Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210.

Enright AJ, John B, Gaul U, Tuschl T, Sander C és Marks DS (2003) MicroRNA targets in *Drosophila*. *Genome Biol.* 5, R1.

Ergün A, Lawrence C a, Kohanski MA, Brennan T a és Collins JJ (2007) A network biology approach to prostate cancer. *Mol. Syst. Biol.* 3, 82.

Estrada E (2006) Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics* 6, 35–40.

Euler L (1736) *Solutio problematis ad geometriam situs pertinentis*. *Coment. Acad. Sci. Petropolitanae* 8, 128–140.

Farkas IJ, Korcsmaros T, Kovacs IA, Mihalik A, Palotai R, Simko GI, Szalay KZ, Szalay-Beko M, Vellai T, Wang S és Csermely P (2011) Network-Based Tools for the Identification of Novel Drug Targets. *Sci. Signal.* 4, pt3.

Fay MP, és Proschan MA (2010) Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Stat. Surv.* 4, 1–39.

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

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt S, Johnson N, Juettemann T, Kähäri AK, Keenan S, Kulesha E, Martin FJ, Maurel T, McLaren WM, Murphy DN, Nag R, Overduin B, Pignatelli M, Pritchard B, Pritchard E, Riat HS, Ruffier M, Sheppard D, Taylor K, Thormann A, Trevanion SJ, Vullo A, Wilder SP, Wilson M, Zadissa A, Aken BL, Birney E, Cunningham F, Harrow J, Herrero J, Hubbard TJP, Kinsella R, Muffato M, Parker A, Spudich G, Yates A, Zerbino DR és Searle SMJ (2014) Ensembl 2014. *Nucleic Acids Res.* 42, D749–D755.

Fraser HB (2005) Modularity and evolutionary constraint on proteins. *Nat. Genet.* 37, 351–352.

Friedman RC, Farh KK-H, Burge CB és Bartel DP (2008) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N és Stratton MR (2004) A census of human cancer genes. *Nat. Rev. Cancer* 4, 177–183.

Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, Grossfeld PD és Srivastava D (2005) Mutations in NOTCH1 cause aortic valve disease. *Nature* 437, 270–274.

Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B és Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107.

Geeleher P, Hartnett L, Egan LJ, Golden a., Raja Ali R a. és Seoighe C (2013) Gene-Set Analysis is Severely Biased When Applied to Genome-wide Methylation Data. *Bioinformatics* 29, 1851–1857.

Gennarino VA, Sardiello M, Avellino R, Meola N, Maselli V, Anand S, Cuttillo L, Ballabio A és Banfi S (2009) MicroRNA target prediction by expression analysis of host genes. *Genome Res.* 19, 481–490.

Gennarino VA, Sardiello M, Mutarelli M, Dharmalingam G, Maselli V, Lago G és Banfi S (2011) HOCTAR database: A unique resource for microRNA target prediction. *Gene* 480, 51–58.

Girvan M, és Newman MEJ (2002) Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* 99, 7821–7826.

Glick D, Barth S és Macleod KF (2010) Autophagy: cellular and molecular mechanisms. *J. Pathol.* 221, 3–12.

Good MC, Zalatan JG és Lim WA (2011) Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science* (80-.). 332, 680–686.

Gregorieff A, és Clevers H (2005) Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev.* 19, 877–890.

Griffith OL, Montgomery SB, Bernier B, Chu B, Kasaian K, Aerts S, Mahony S, Sleumer MC, Bilenky M, Haeussler M, Griffith M, Gallo SM, Giardine B, Hooghe B, Van Loo P, Blanco E, Ticoll A, Lithwick S, Portales-Casamar E, Donaldson II, Robertson G, Wadelius C, De Bleser P, Vlieghe D, Halfon MS, Wasserman W, Hardison R, Bergman CM és Jones SJM (2007) ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res.* 36, D107–D113.

Grimson A, Farh KK-H, Johnston WK, Garrett-Engle P, Lim LP és Bartel DP (2007) MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Mol. Cell* 27, 91–105.

Gröschl B, Bettstetter M, Giedl C, Woenckhaus M, Edmonston T, Hofstädter F és

Dietmaier W (2013) Expression of the MAP kinase phosphatase DUSP4 is associated with microsatellite instability in colorectal cancer (CRC) and causes increased cell proliferation. *Int. J. Cancer* 132, 1537–1546.

Groth C, és Fortini ME (2012) Therapeutic approaches to modulating Notch signaling: Current challenges and future prospects. *Semin. Cell Dev. Biol.* 23, 465–472.

Guo G, Kuai D, Cai S, Xue N, Liu Y, Hao J, Fan Y, Jin J, Mao X, Liu B, Zhong C, Zhang X, Yue Y, Liu X, Ma N és Guo Y (2013) Knockdown of FRAT1 Expression by RNA Interference Inhibits Human Glioblastoma Cell Growth, Migration and Invasion. *PLoS One* 8, e61206.

Guo X, Ramirez A, Waddell DS, Li Z, Liu X és Wang X-F (2008) Axin and GSK3-control Smad3 protein stability and modulate TGF- signaling. *Genes Dev.* 22, 106–120.

Guruharsha KG, Kankel MW és Artavanis-Tsakonas S (2012) The Notch signalling system: recent insights into the complexity of a conserved pathway. *Nat. Rev. Genet.* 13, 654–666.

Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, Kim H, Cho A, Kim E, Lee T, Kim H, Kim K, Yang S, Bae D, Yun A, Kim S, Kim CY, Cho HJ, Kang B, Shin S és Lee I (2015) TRRUST: a reference database of human transcriptional regulatory interactions. *Sci. Rep.* 5, 11432.

Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang L V, Dupuy D, Walhout AJM, Cusick ME, Roth FP és Vidal M (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* 430, 88–93.

Hanahan D, és Weinberg RA (2011) Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674.

Hanahan D, és Weinberg RARA (2000) The hallmarks of cancer. *Cell* 100, 57–70.

Hicks C, Johnston SH, DiSibio G, Collazo A, Vogt TF és Weinmaster G (2000) Fringe differentially modulates Jagged1 and Delta1 signalling through Notch1 and Notch2. *Nat. Cell Biol.* 2, 515–520.

Higgins ME, Claremont M, Major JE, Sander C és Lash AE (2007) CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res.* 35, D721–D726.

Hochberg B (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.

Homma K, Suzuki K és Sugawara H (2011) The Autophagy Database: an all-inclusive information resource on autophagy that provides nourishment for research. *Nucleic Acids Res.* 39, D986–D990.

Hornberg JJ, Bruggeman FJ, Westerhoff H V és Lankelma J (2006) Cancer: A Systems Biology disease. *Biosystems* 83, 81–90.

Hsu C-W, Juan H-F és Huang H-C (2008) Characterization of microRNA-regulated protein-protein interaction network. *Proteomics* 8, 1975–1979.

Hsu S-D, Lin F-M, Wu W-Y, Liang C, Huang W-C, Chan W-L, Tsai W-T, Chen G-Z, Lee C-J, Chiu C-M, Chien C-H, Wu M-C, Huang C-Y, Tsou A-P és Huang H-D (2011) miRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.* 39, D163–D169.

Hsu S-D, Tseng Y-T, Shrestha S, Lin Y-L, Khaleel A, Chou C-H, Chu C-F, Huang H-Y, Lin C-M, Ho S-Y, Jian T-Y, Lin F-M, Chang T-H, Weng S-L, Liao K-W, Liao I-E, Liu C-C és Huang H-D (2014) miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.* 42, D78–D85.

Huang JC, Babak T, Corson TW, Chua G, Khan S, Gallie BL, Hughes TR, Blencowe BJ, Frey BJ és Morris QD (2007) Using expression profiling data to identify human microRNA targets. *Nat. Methods* 4, 1045–1049.

Huang P, Chandra V és Rastinejad F (2010) Structural Overview of the Nuclear Receptor Superfamily: Insights into Physiology and Therapeutics. *Annu. Rev. Physiol.* 72, 247–272.

Hunter JD (2007) Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* 9, 90–95.

Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ és O'Donovan C (2015) The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Res.* 43, D1057–D1063.

Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, W??hr M, Chick J, Zhai B, Kolippakkam D, Mintseris J, Obar RA, Harris T, Artavanis-Tsakonas S, Sowa ME, De Camilli P, Paulo JA, Harper JW, Gygi SP, W??hr M, Chick J, Zhai B, Kolippakkam D, Mintseris J, Obar RA, Harris T, Artavanis-Tsakonas S, Sowa ME, De Camilli P, Paulo JA, Harper JW és Gygi SP (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162, 425–440.

Hwang T, Atluri G, Kuang R, Kumar V, Starr T, Silverstein KA, Haverty PM, Zhang Z és Liu J (2013) Large-scale integrative network-based analysis identifies common pathways disrupted by copy number alterations across cancers. *BMC Genomics* 14, 440.

Ichimura A, Ruike Y, Terasawa K és Tsujimoto G (2011) miRNAs and regulation of cell signaling. *FEBS J.* 278, 1610–1618.

Inagaki-Ohara K, Kondo T, Ito M és Yoshimura A (2013) SOCS, inflammation, and cancer. *JAK-STAT* 2, e24053.

Irizarry R a (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264.

Itoh K, Wakabayashi N, Katoh Y, Ishii T, Igarashi K, Engel JD és Yamamoto M (1999) Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev.* 13, 76–86.

Jayaraman L, és Massagué J (2000) Distinct Oligomeric States of SMAD Proteins in the Transforming Growth Factor- β Pathway. *J. Biol. Chem.* 275, 40710–40717.

Jeong H, Mason SP, Barabási A-L és Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411, 41–42.

Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G és Liu Y (2009)

miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 37, D98–D104.

John E. Niderhuber, Armitage JO, Tepper JE, Doroshow JH és Kastan MB (2014) *Clinical Oncology* Elsevier GmbH., Philadelphia,.

Johnson DS, Mortazavi A, Myers RM és Wold B (2007) Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* (80-.). 316, 1497–1502.

Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin K, Vaquerizas JM, Vincentelli R, Luscombe NM, Hughes TR, Lemaire P, Ukkonen E, Kivioja T és Taipale J (2013) DNA-Binding Specificities of Human Transcription Factors. *Cell* 152, 327–339.

Jones E, Oliphant T, Peterson P és others (2001) {SciPy}: Open source scientific tools for {Python}.

Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T és Kuriyan J (2011) Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms. *Mol. Cell* 42, 9–22.

Kalimo H, Ruchoux M-M, Viitanen M és Kalaria RN (2006) CADASIL: a Common Form of Hereditary Arteriopathy Causing Brain Infarcts and Dementia. *Brain Pathol.* 12, 371–384.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M és Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205.

Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R és Pandey A (2009) Human Protein Reference Database--2009 update. *Nucleic Acids Res.* 37, D767–D772.

Khuri S, és Wuchty S (2015) Essentiality and centrality in protein interaction networks revisited. *BMC Bioinformatics* 16, 109.

Kim S (2007) An Interior-Point Method for Large-Scale Logistic Regression. *J. Mach. Learn. Res.* 8, 1519–1555.

Kim K-Y, Kim B-C, Xu Z és Kim S-J (2004) Mixed Lineage Kinase 3 (MLK3)-activated p38 MAP Kinase Mediates Transforming Growth Factor- β -induced Apoptosis in Hepatoma Cells. *J. Biol. Chem.* 279, 29478–29484.

Kim VN, Han J és Siomi MC (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* 10, 126–139.

Kishida S, Mu P, Miyakawa S, Fujiwara M, Abe T, Sakamoto K, Onishi A, Nakamura Y és Kadomatsu K (2013) Midkine promotes neuroblastoma through notch2 signaling. *Cancer Res.* 73, 1318–1327.

Klionsky D, Abdelmohsen K, Abe A, Abedin M, Abeliovich H, Acevedo, Arozena A, Adachi H, Adams C, Adams P, Adeli K, Adhietty P, Adler S, Agam G, Agarwal R, Aghi M, Agnello M, Agostinis P, Aguilar P, Aguirre-Ghiso J, Airoidi E, Ait-Si-Ali S, Akematsu T, Akporiaye E, Al-Rubeai M, Albaiceta G, Albanese C, Albani D, Albert M, Aldudo J, Algül H, Alirezai M, Alloza I, Almasan A, Almonte-Beceril M, Alnemri E, Alonso C, Altan-Bonnet N, Altieri D, Alvarez S, Alvarez-Erviti L, Alves S, Amadoro G, Amano A, Amantini C, Ambrosio S, Amelio I, Amer A, Amessou M, Amon A, An Z, Anania F, Andersen S, Andley U, Andreadi C, Andrieu-Abadie N, Anel A, Ann D, Anoopkumar-Dukie S, Antonioli M, Aoki H, Apostolova N, Aquila S, Aquilano K, Araki K, Arama E, Aranda A, Araya J, Arcaro A, Arias E, Arimoto H, Ariosa A, Armstrong J, Arnould T, Arsov I, Asanuma K, Askanas V, Asselin E, Atarashi R, Atherton S, Atkin J, Attardi L, Auburger P, Auburger G, Aurelian L, Autelli R, Avagliano L, Avantaggiati M, Avrahami L, Awale S, Azad N, Bachetti T, Backer J, Bae D, Bae J, Bae O, Bae S, Baehrecke E, Baek S, Baghdiguian S, Bagniewska-Zadworna A, Bai H, Bai J, Bai X, Bailly Y, Balaji K, Balduini W, Ballabio A, Balzan R, Banerjee R, Bánhegyi G, Bao H, Barbeau B, Barrachina M, Barreiro E, Bartel B, Bartolomé A, Bassham D, Bassi M, Bast, RC J, Basu A, Batista M, Batoko H, Battino M, Bauckman K, Baumgarner B, Bayer K,

Beale R, Beaulieu J, Beck, GR J, Becker C, Beckham J, Bédard P, Bednarski P, Begley T, Behl C, Behrends C, Behrens G, Behrns K, Bejarano E, Belaid A, Belleudi F, Bénard G, Berchem G, Bergamaschi D, Bergami M, Berkhout B, Berliocchi L, Bernard A, Bernard M, Bernassola F, Bertolotti A, Bess A, Besteiro S, Bettuzzi S, Bhalla S, Bhattacharyya S, Bhutia S, Biagosch C, Bianchi M, Biard-Piechaczyk M, Billes V, Bincoletto C, Bingol B, Bird S, Bitoun M, Bjedov I, Blackstone C, Blanc L, Blanco G, Blomhoff H, Boada-Romero E, Böckler S, Boes M, Boesze-Battaglia K, Boise L, Bolino A, Boman A, Bonaldo P, Bordi M, Bosch J, Botana L, Botti J, Bou G, Bouché M, Bouchecareilh M, Boucher M, Boulton M, Bouret S, Boya P, Boyer-Guittaut M, Bozhkov P, Brady N, Braga V, Brancolini C, Braus G, Bravo-San, Pedro J, Brennan L, Bresnick E, Brest P, Bridges D, Bringer M, Brini M, Brito G, Brodin B, Brookes P, Brown E, Brown K, Broxmeyer H, Bruhat A, Brum P, Brumell J, Brunetti-Pierri N, Bryson-Richardson R, Buch S, Buchan A, Budak H, Bulavin D, Bultman S, Bultynck G, Bumbasirevic V, Burelle Y, Burke R, Burmeister M, Bütikofer P, Caberlotto L, Cadwell K, Cahova M, Cai D, Cai J, Cai Q, Calatayud S, Camougrand N, Campanella M, Campbell G, Campbell M, Campello S, Candau R, Caniggia I, Cantoni L, Cao L, Caplan A, Caraglia M, Cardinali C, Cardoso S, Carew J, Carleton L, Carlin C, Carloni S, Carlsson S, Carmona-Gutierrez D, Carneiro L, Carnevali O, Carra S, Carrier A, Carroll B, Casas C, Casas J, Cassinelli G, Castets P, Castro-Obregon S, Cavallini G, Ceccherini I, Cecconi F, Cederbaum A, Ceña V, Cenci S, Cerella C, Cervia D, Cetrullo S, Chaachouay H, Chae H, Chagin A, Chai C, Chakrabarti G, Chamilos G, Chan E, Chan M, Chandra D, Chandra P, Chang C, Chang R, Chang T, Chatham J, Chatterjee S, Chauhan S, Che Y, Cheetham M, Cheluvappa R, Chen C, Chen G, Chen G, Chen G, Chen H, Chen J, Chen J, Chen M, Chen M, Chen P, Chen Q, Chen Q, Chen S, Chen S, Chen S, Chen W, Chen W, Chen W, Chen W, Chen X, Chen Y, Chen Y, Chen Y, Chen Y, Chen Y, Chen Y, Chen Y, Chen Y, Chen Y, Chen Z, Chen Z, Cheng A, Cheng C, Cheng H, Cheong H, Cherry S, Chesney J, Cheung C, Chevet E, Chi H, Chi S, Chiacchiera F, Chiang H, Chiarelli R, Chiariello M, Chieppa M, Chin L, Chiong M, Chiu G, Cho D, Cho S, Cho W, Cho Y, Cho Y, Choi A, Choi E, Choi E, Choi J, Choi M, Choi S, Chou T, Chouaib S, Choubey D, Choubey V, Chow K, Chowdhury K, Chu C, Chuang T, Chun T, Chung H, Chung T, Chung Y, Chwae Y, Cianfanelli V, Ciarcia R, Ciechomska I, Ciriolo M, Cirone M, Claerhout S, Clague M, Clària J, Clarke P, Clarke R, Clementi E, Cleyrat C, Cnop M, Coccia E, Cocco T,

Codogno P, Coers J, Cohen E, Colecchia D, Coletto L, Coll N, Colucci-Guyon E, Comincini S, Condello M, Cook K, Coombs G, Cooper C, Cooper J, Coppens I, Corasaniti M, Corazzari M, Corbalan R, Corcelle-Termeau E, Cordero M, Corral-Ramos C, Corti O, Cossarizza A, Costelli P, Costes S, Cotman S, Coto-Montes A, Cottet S, Couve E, Covey L, Cowart L, Cox J, Coxon F, Coyne C, Cragg M, Craven R, Crepaldi T, Crespo J, Criollo A, Crippa V, Cruz M, Cuervo A, Cuezva J, Cui T, Cutillas P, Czaja M, Czyzyk-Krzeska M, Dagda R, Dahmen U, Dai C, Dai W, Dai Y, Dalby K, Dalla, Valle L, Dalmaso G, D'Amelio M, Damme M, Darfeuille-Michaud A, Dargemont C, Darley-Usmar V, Dasarathy S, Dasgupta B, Dash S, Dass C, Davey H, Davids L, Dávila D, Davis R, Dawson T, Dawson V, Daza P, de, Belleruche J, de, Figueiredo P, de, Figueiredo R, de, la, Fuente J, De, Martino L, De, Matteis A, De, Meyer G, De, Milito A, De, Santi M, de, Souza W, De, Tata V, De, Zio D, Debnath J, Dechant R, Decuypere J, Deegan S, Dehay B, Del, Bello B, Del, Re D, Delage-Mourroux R, Delbridge L, Deldicque L, Delorme-Axford E, Deng Y, Dengjel J, Denizot M, Dent P, Der C, Deretic V, Derrien B, Deutsch E, Devarenne T, Devenish R, Di, Bartolomeo S, Di, Daniele N, Di, Domenico F, Di, Nardo A, Di, Paola S, Di, Pietro A, Di, Renzo L, DiAntonio A, Díaz-Araya G, Díaz-Laviada I, Diaz-Meco M, Diaz-Nido J, Dickey C, Dickson R, Diederich M, Digard P, Dikic I, Dinesh-Kumar S, Ding C, Ding W, Ding Z, Dini L, Distler J, Diwan A, Djavaheri-Mergny M, Dmytruk K, Dobson R, Doetsch V, Dokladny K, Dokudovskaya S, Donadelli M, Dong X, Dong X, Dong Z, Donohue, TM J, Doran K, D'Orazi G, Dorn, GW 2nd, Dosenko V, Dridi S, Drucker L, Du J, Du L, Du L, du, Toit A, Dua P, Duan L, Duann P, Dubey V, Duchon M, Duchosal M, Duez H, Dugail I, Dumit V, Duncan M, Dunlop E, Dunn, WA J, Dupont N, Dupuis L, Durán R, Durcan T, Duvezin-Caubet S, Duvvuri U, Eapen V, Ebrahimi-Fakhari D, Echard A, Eckhart L, Edelstein C, Edinger A, Eichinger L, Eisenberg T, Eisenberg-Lerner A, Eissa N, El-Deiry W, El-Khoury V, Elazar Z, Eldar-Finkelman H, Elliott C, Emanuele E, Emmenegger U, Engedal N, Engelbrecht A, Engelender S, Enserink J, Erdmann R, Erenpreisa J, Eri R, Eriksen J, Erman A, Escalante R, Eskelinen E, Espert L, Esteban-Martínez L, Evans T, Fabri M, Fabrias G, Fabrizi C, Facchiano A, Færgeman N, Faggioni A, Fairlie W, Fan C, Fan D, Fan J, Fang S, Fanto M, Fanzani A, Farkas T, Faure M, Favier F, Fearnhead H, Federici M, Fei E, Felizardo T, Feng H, Feng Y, Feng Y, Ferguson T, Fernández Á, Fernandez-Barrena M, Fernandez-Checa J, Fernández-López A, Fernandez-Zapico M, Feron O, Ferraro E,

Ferreira-Halder C, Fesus L, Feuer R, Fiesel F, Filippi-Chiela E, Filomeni G, Fimia G, Fingert J, Finkbeiner S, Finkel T, Fiorito F, Fisher P, Flajolet M, Flamigni F, Florey O, Florio S, Floto R, Folini M, Follo C, Fon E, Fornai F, Fortunato F, Fraldi A, Franco R, Francois A, François A, Frankel L, Fraser I, Frey N, Freyssenet D, Frezza C, Friedman S, Frigo D, Fu D, Fuentes J, Fueyo J, Fujitani Y, Fujiwara Y, Fujiya M, Fukuda M, Fulda S, Fusco C, Gabryel B, Gaestel M, Gailly P, Gajewska M, Galadari S, Galili G, Galindo I, Galindo M, Galliciotti G, Galluzzi L, Galluzzi L, Galy V, Gammoh N, Gandy S, Ganesan A, Ganesan S, Ganley I, Gannagé M, Gao F, Gao F, Gao J, García, Nannig L, García, Vécovi E, Garcia-Macía M, Garcia-Ruiz C, Garg A, Garg P, Gargini R, Gassen N, Gatica D, Gatti E, Gavard J, Gavathiotis E, Ge L, Ge P, Ge S, Gean P, Gelmetti V, Genazzani A, Geng J, Genschik P, Gerner L, Gestwicki J, Gewirtz D, Ghavami S, Ghigo E, Ghosh D, Giammarioli A, Giampieri F, Giampietri C, Giatromanolaki A, Gibbings D, Gibellini L, Gibson S, Ginet V, Giordano A, Giorgini F, Giovannetti E, Girardin S, Gispert S, Giuliano S, Gladson C, Glavic A, Gleave M, Godefroy N, Gogal, RM J, Gokulan K, Goldman G, Goletti D, Goligorsky M, Gomes A, Gomes L, Gomez H, Gomez-Manzano C, Gómez-Sánchez R, Gonçalves D, Goncu E, Gong Q, Gongora C, Gonzalez C, Gonzalez-Alegre P, Gonzalez-Cabo P, González-Polo R, Goping I, Gorbea C, Gorbunov N, Goring D, Gorman A, Gorski S, Goruppi S, Goto-Yamada S, Gotor C, Gottlieb R, Gozes I, Gozuacik D, Graba Y, Graef M, Granato G, Grant G, Grant S, Gravina G, Green D, Greenhough A, Greenwood M, Grimaldi B, Gros F, Grose C, Groulx J, Gruber F, Grumati P, Grune T, Guan J, Guan K, Guerra B, Guillen C, Gulshan K, Gunst J, Guo C, Guo L, Guo M, Guo W, Guo X, Gust A, Gustafsson Å, Gutierrez E, Gutierrez M, Gwak H, Haas A, Haber J, Hadano S, Hagedorn M, Hahn D, Halayko A, Hamacher-Brady A, Hamada K, Hamai A, Hamann A, Hamasaki M, Hamer I, Hamid Q, Hammond E, Han F, Han W, Handa J, Hanover J, Hansen M, Harada M, Harhaji-Trajkovic L, Harper J, Harrath A, Harris A, Harris J, Hasler U, Hasselblatt P, Hasui K, Hawley R, Hawley T, He C, He C, He F, He G, He R, He X, He Y, He Y, Heath J, Hébert M, Heinzen R, Helgason G, Hensel M, Henske E, Her C, Herman P, Hernández A, Hernandez C, Hernández-Tiedra S, Hetz C, Hiesinger P, Higaki K, Hilfiker S, Hill B, Hill J, Hill W, Hino K, Hofius D, Hofman P, Höglinger G, Höhfeld J, Holz M, Hong Y, Hood D, Hoozemans J, Hoppe T, Hsu C, Hsu C, Hsu L, Hu D, Hu G, Hu H, Hu H, Hu M, Hu Y, Hu Z, Hua F, Hua Y, Huang C, Huang H, Huang K, Huang K, Huang S, Huang S, Huang

W, Huang Y, Huang Y, Huang Y, Huber T, Huebbe P, Huh W, Hulmi J, Hur G, Hurley J, Husak Z, Hussain S, Hussain S, Hwang J, Hwang S, Hwang T, Ichihara A, Imai Y, Imbriano C, Inomata M, Into T, Iovane V, Iovanna J, Iozzo R, Ip N, Irazoqui J, Iribarren P, Isaka Y, Isakovic A, Ischiropoulos H, Isenberg J, Ishaq M, Ishida H, Ishii I, Ishmael J, Isidoro C, Isobe K, Isono E, Issazadeh-Navikas S, Itahana K, Itakura E, Ivanov A, Iyer A, Izquierdo J, Izumi Y, Izzo V, Jäättelä M, Jaber N, Jackson D, Jackson W, Jacob T, Jacques T, Jagannath C, Jain A, Jana N, Jang B, Jani A, Janji B, Jannig P, Jansson P, Jean S, Jendrach M, Jeon J, Jessen N, Jeung E, Jia K, Jia L, Jiang H, Jiang H, Jiang L, Jiang T, Jiang X, Jiang X, Jiang X, Jiang Y, Jiang Y, Jiménez A, Jin C, Jin H, Jin L, Jin M, Jin S, Jinwal U, Jo E, Johansen T, Johnson D, Johnson G, Johnson J, Jonasch E, Jones C, Joosten L, Jordan J, Joseph A, Joseph B, Joubert A, Ju D, Ju J, Juan H, Juenemann K, Juhász G, Jung H, Jung J, Jung Y, Jungbluth H, Justice M, Jutten B, Kaakoush N, Kaarniranta K, Kaasik A, Kabuta T, Kaeffer B, Kågedal K, Kahana A, Kajimura S, Kakhlon O, Kalia M, Kalvakolanu D, Kamada Y, Kambas K, Kaminsky V, Kampinga H, Kandouz M, Kang C, Kang R, Kang T, Kanki T, Kanneganti T, Kanno H, Kanthasamy A, Kantorow M, Kaparakis-Liaskos M, Kapuy O, Karantza V, Karim M, Karmakar P, Kaser A, Kaushik S, Kawula T, Kaynar A, Ke P, Ke Z, Kehrl J, Keller K, Kemper J, Kenworthy A, Kepp O, Kern A, Kesari S, Kessel D, Ketteler R, Kettelhut, Ido C, Khambu B, Khan M, Khandelwal V, Khare S, Kiang J, Kiger A, Kihara A, Kim A, Kim C, Kim D, Kim D, Kim E, Kim H, Kim H, Kim J, Kim J, Kim J, Kim J, Kim K, Kim M, Kim M, Kim P, Kim S, Kim S, Kim Y, Kim Y, Kimchi A, Kimmelman A, Kimura T, King J, Kirkegaard K, Kirkin V, Kirshenbaum L, Kishi S, Kitajima Y, Kitamoto K, Kitaoka Y, Kitazato K, Kley R, Klimecki W, Klinkenberg M, Klucken J, Knævelsrud H, Knecht E, Knuppertz L, Ko J, Kobayashi S, Koch J, Koechlin-Ramonatxo C, Koenig U, Koh Y, Köhler K, Kohlwein S, Koike M, Komatsu M, Kominami E, Kong D, Kong H, Konstantakou E, Kopp B, Korcsmaros T, Korhonen L, Korolchuk V, Koshkina N, Kou Y, Koukourakis M, Koumenis C, Kovács A, Kovács T, Kovacs W, Koya D, Kraft C, Krainc D, Kramer H, Kravic-Stevovic T, Krek W, Kretz-Remy C, Krick R, Krishnamurthy M, Kriston-Vizi J, Kroemer G, Kruer M, Kruger R, Ktistakis N, Kuchitsu K, Kuhn C, Kumar A, Kumar A, Kumar A, Kumar D, Kumar D, Kumar R, Kumar S, Kundu M, Kung H, Kuno A, Kuo S, Kuret J, Kurz T, Kwok T, Kwon T, Kwon Y, Kyrmizi I, La, Spada A, Lafont F, Lahm T, Lakkaraju A, Lam T, Lamark T, Lancel S, Landowski

T, Lane D, Lane J, Lanzi C, Lapaquette P, Lapierre L, Laporte J, Laukkarinen J, Laurie G, Lavandero S, Lavie L, LaVoie M, Law B, Law H, Law K, Layfield R, Lazo P, Le, Cam L, Le, Roch K, Le, Stunff H, Leardkamolkarn V, Lecuit M, Lee B, Lee C, Lee E, Lee G, Lee H, Lee H, Lee J, Lee J, Lee J, Lee J, Lee M, Lee M, Lee P, Lee S, Lee S, Lee S, Lee S, Lee S, Lee S, Lee S, Lee S, Lee S, Lee Y, Lee Y, Lee Y, Leeuwenburgh C, Lefort S, Legouis R, Lei J, Lei Q, Leib D, Leibowitz G, Lekli I, Lemaire S, Lemasters J, Lemberg M, Lemoine A, Leng S, Lenz G, Lenzi P, Lerman L, Lettieri, Barbato D, Leu J, Leung H, Levine B, Lewis P, Lezoualc'h F, Li C, Li F, Li F, Li J, Li K, Li L, Li M, Li M, Li Q, Li R, Li S, Li W, Li W, Li X, Li Y, Lian J, Liang C, Liang Q, Liao Y, Liberal J, Liberski P, Lie P, Lieberman A, Lim H, Lim K, Lim K, Lima R, Lin C, Lin C, Lin F, Lin F, Lin F, Lin K, Lin K, Lin P, Lin T, Lin W, Lin Y, Lin Y, Linden R, Lindholm D, Lindqvist L, Lingor P, Linkermann A, Liotta L, Lipinski M, Lira V, Lisanti M, Liton P, Liu B, Liu C, Liu C, Liu F, Liu H, Liu J, Liu J, Liu J, Liu K, Liu L, Liu L, Liu Q, Liu R, Liu S, Liu S, Liu W, Liu X, Liu X, Liu X, Liu X, Liu X, Liu X, Liu Y, Liu Y, Liu Z, Liu Z, Liuzzi J, Lizard G, Ljubic M, Lodhi I, Logue S, Lokeshwar B, Long Y, Lonial S, Loos B, López-Otín C, López-Vicario C, Lorente M, Lorenzi P, Lőrincz P, Los M, Lotze M, Lovat P, Lu B, Lu B, Lu J, Lu Q, Lu S, Lu S, Lu Y, Luciano F, Luckhart S, Lucocq J, Ludovico P, Lugea A, Lukacs N, Lum J, Lund A, Luo H, Luo J, Luo S, Luparello C, Lyons T, Ma J, Ma Y, Ma Y, Ma Z, Machado J, Machado-Santelli G, Macian F, MacIntosh G, MacKeigan J, Macleod K, MacMicking J, MacMillan-Crow L, Madeo F, Madesh M, Madrigal-Matute J, Maeda A, Maeda T, Maegawa G, Maellaro E, Maes H, Magariños M, Maiese K, Maiti T, Maiuri L, Maiuri M, Maki C, Malli R, Malorni W, Maloyan A, Mami-Chouaib F, Man N, Mancias J, Mandelkow E, Mandell M, Manfredi A, Manié S, Manzoni C, Mao K, Mao Z, Mao Z, Marambaud P, Marconi A, Marelja Z, Marfe G, Margeta M, Margittai E, Mari M, Mariani F, Marin C, Marinelli S, Mariño G, Markovic I, Marquez R, Martelli A, Martens S, Martin K, Martin S, Martin S, Martin-Acebes M, Martín-Sanz P, Martinand-Mari C, Martinet W, Martinez J, Martinez-Lopez N, Martinez-Outschoorn U, Martínez-Velázquez M, Martinez-Vicente M, Martins W, Mashima H, Mastrianni J, Matarese G, Matarrese P, Mateo R, Matoba S, Matsumoto N, Matsushita T, Matsuura A, Matsuzawa T, Mattson M, Matus S, Maugeri N, Mauvezin C, Mayer A, Maysinger D, Mazzolini G, McBrayer M, McCall K, McCormick C, McInerney G, McIver S, McKenna S, McMahan J, McNeish I, Mechta-Grigoriou F, Medema J, Medina D, Megyeri K,

Mehrpour M, Mehta J, Mei Y, Meier U, Meijer A, Meléndez A, Melino G, Melino S, de Melo E, Mena M, Meneghini M, Menendez J, Menezes R, Meng L, Meng L, Meng S, Menghini R, Menko A, Menna-Barreto R, Menon M, Meraz-Ríos M, Merla G, Merlini L, Merlot A, Meryk A, Meschini S, Meyer J, Mi M, Miao C, Micale L, Michaeli S, Michiels C, Migliaccio A, Mihailidou A, Mijaljica D, Mikoshiba K, Milan E, Miller-Fleming L, Mills G, Mills I, Minakaki G, Minassian B, Ming X, Minibayeva F, Minina E, Mintern J, Minucci S, Miranda-Vizueté A, Mitchell C, Miyamoto S, Miyazawa K, Mizushima N, Mnich K, Mograbi B, Mohseni S, Moita L, Molinari M, Molinari M, Møller A, Mollereau B, Mollinedo F, Mongillo M, Monick M, Montagnaro S, Montell C, Moore D, Moore M, Mora-Rodriguez R, Moreira P, Morel E, Morelli M, Moreno S, Morgan M, Moris A, Moriyasu Y, Morrison J, Morrison L, Morselli E, Moscat J, Moseley P, Mostowy S, Motori E, Mottet D, Mottram J, Moussa C, Mpakou V, Mukhtar H, Mulcahy, Levy J, Muller S, Muñoz-Moreno R, Muñoz-Pinedo C, Münz C, Murphy M, Murray J, Murthy A, Mysorekar I, Nabi I, Nabissi M, Nader G, Nagahara Y, Nagai Y, Nagata K, Nagelkerke A, Nagy P, Naidu S, Nair S, Nakano H, Nakatogawa H, Nanjundan M, Napolitano G, Naqvi N, Nardacci R, Narendra D, Narita M, Nascimbeni A, Natarajan R, Navegantes L, Nawrocki S, Nazarko T, Nazarko V, Neill T, Neri L, Netea M, Netea-Maier R, Neves B, Ney P, Nezis I, Nguyen H, Nguyen H, Nicot A, Nilsen H, Nilsson P, Nishimura M, Nishino I, Niso-Santano M, Niu H, Nixon R, Njar V, Noda T, Noegel A, Nolte E, Norberg E, Norga K, Noureini S, Notomi S, Notterpek L, Nowikovsky K, Nukina N, Nürnberger T, O'Donnell V, O'Donovan T, O'Dwyer P, Oehme I, Oeste C, Ogawa M, Ogretmen B, Ogura Y, Oh Y, Ohmuraya M, Ohshima T, Ojha R, Okamoto K, Okazaki T, Oliver F, Ollinger K, Olsson S, Orban D, Ordonez P, Orhon I, Orosz L, O'Rourke E, Orozco H, Ortega A, Ortona E, Osellame L, Oshima J, Oshima S, Osiewacz H, Otomo T, Otsu K, Ou J, Outeiro T, Ouyang D, Ouyang H, Overholtzer M, Ozbun M, Ozdinler P, Ozpolat B, Pacelli C, Paganetti P, Page G, Pages G, Pagnini U, Pajak B, Pak S, Pakos-Zebrucka K, Pakpour N, Palková Z, Palladino F, Pallauf K, Pallet N, Palmieri M, Paludan S, Palumbo C, Palumbo S, Pampliega O, Pan H, Pan W, Panaretakis T, Pandey A, Pantazopoulou A, Papackova Z, Papademetrio D, Papassideri I, Papini A, Parajuli N, Pardo J, Parekh V, Parenti G, Park J, Park J, Park O, Parker R, Parlato R, Parys J, Parzych K, Pasquet J, Pasquier B, Pasumarthi K, Patschan D, Patterson C, Patingre S, Pattison S, Pause A, Pavenstädt H, Pavone F, Pedrozo Z, Peña F, Peñalva M,

Pende M, Peng J, Penna F, Penninger J, Pensalfini A, Pepe S, Pereira G, Pereira P, Pérez-de, la, Cruz V, Pérez-Pérez M, Pérez-Rodríguez D, Pérez-Sala D, Perier C, Perl A, Perlmutter D, Perrotta I, Pervaiz S, Pesonen M, Pessin J, Peters G, Petersen M, Petrache I, Petrof B, Petrovski G, Phang J, Piacentini M, Pierdominici M, Pierre P, Pierrefite-Carle V, Pietrocola F, Pimentel-Muñños F, Pinar M, Pineda B, Pinkas-Kramarski R, Pinti M, Pinton P, Piperdi B, Piret J, Platanius L, Platta H, Plowey E, Pöggeler S, Poirot M, Polčić P, Poletti A, Poon A, Popelka H, Popova B, Poprawa I, Poulouse S, Poulton J, Powers S, Powers T, Pozuelo-Rubio M, Prak K, Prange R, Prescott M, Priault M, Prince S, Proia R, Proikas-Cezanne T, Prokisch H, Promponas V, Przyklenk K, Puertollano R, Pugazhenth S, Puglielli L, Pujol A, Puyal J, Pyeon D, Qi X, Qian W, Qin Z, Qiu Y, Qu Z, Quadrilatero J, Quinn F, Raben N, Rabinowich H, Radogna F, Ragusa M, Rahmani M, Raina K, Ramanadham S, Ramesh R, Rami A, Randall-Demllo S, Randow F, Rao H, Rao V, Rasmussen B, Rasse T, Ratovitski E, Rautou P, Ray S, Razani B, Reed B, Reggiori F, Rehm M, Reichert A, Rein T, Reiner D, Reits E, Ren J, Ren X, Renna M, Reusch J, Revuelta J, Reyes L, Rezaie A, Richards R, Richardson D, Richetta C, Riehle M, Rihn B, Rikihisa Y, Riley B, Rimbach G, Rippo M, Ritis K, Rizzi F, Rizzo E, Roach P, Robbins J, Roberge M, Roca G, Roccheri M, Rocha S, Rodrigues C, Rodríguez C, de, Cordoba S, Rodriguez-Muela N, Roelofs J, Rogov V, Rohn T, Rohrer B, Romanelli D, Romani L, Romano P, Roncero M, Rosa J, Rosello A, Rosen K, Rosenstiel P, Rost-Roszkowska M, Roth K, Roué G, Rouis M, Rouschop K, Ruan D, Ruano D, Rubinsztein D, Rucker, EB 3rd, Rudich A, Rudolf E, Rudolf R, Ruegg M, Ruiz-Roldan C, Ruparelia A, Rusmini P, Russ D, Russo G, Russo G, Russo R, Rusten T, Ryabovol V, Ryan K, Ryter S, Sabatini D, Sacher M, Sachse C, Sack M, Sadoshima J, Saftig P, Sagi-Eisenberg R, Sahni S, Saikumar P, Saito T, Saitoh T, Sakakura K, Sakoh-Nakatogawa M, Sakuraba Y, Salazar-Roa M, Salomoni P, Saluja A, Salvaterra P, Salvioli R, Samali A, Sanchez A, Sánchez-Alcázar J, Sanchez-Prieto R, Sandri M, Sanjuan M, Santaguida S, Santambrogio L, Santoni G, Dos, Santos C, Saran S, Sardiello M, Sargent G, Sarkar P, Sarkar S, Sarrias M, Sarwal M, Sasakawa C, Sasaki M, Sass M, Sato K, Sato M, Satriano J, Savaraj N, Saveljeva S, Schaefer L, Schaible U, Scharl M, Schatzl H, Schekman R, Scheper W, Schiavi A, Schipper H, Schmeisser H, Schmidt J, Schmitz I, Schneider B, Schneider E, Schneider J, Schon E, Schönenberger M, Schönthal A, Schorderet D, Schröder B, Schuck S, Schulze R, Schwarten M, Schwarz T, Sciarretta S, Scotto K, Scovassi A, Screaton R,

Screen M, Seca H, Sedej S, Segatori L, Segev N, Seglen P, Seguí-Simarro J, Segura-Aguilar J, Seki E, Sell C, Seiliez I, Semenkovich C, Semenza G, Sen U, Serra A, Serrano-Puebla A, Sesaki H, Setoguchi T, Settembre C, Shacka J, Shajahan-Haq A, Shapiro I, Sharma S, She H, Shen C, Shen C, Shen H, Shen S, Shen W, Sheng R, Sheng X, Sheng Z, Shepherd T, Shi J, Shi Q, Shi Q, Shi Y, Shibutani S, Shibuya K, Shidoji Y, Shieh J, Shih C, Shimada Y, Shimizu S, Shin D, Shinohara M, Shintani M, Shintani T, Shioi T, Shirabe K, Shiri-Sverdlov R, Shirihai O, Shore G, Shu C, Shukla D, Sibirny A, Sica V, Sigurdson C, Sigurdsson E, Sijwali P, Sikorska B, Silveira W, Silvente-Poirot S, Silverman G, Simak J, Simmet T, Simon A, Simon H, Simone C, Simons M, Simonsen A, Singh R, Singh S, Singh S, Sinha D, Sinha S, Sinicrope F, Sirko A, Sirohi K, Sishi B, Sittler A, Siu P, Sivridis E, Skwarska A, Slack R, Slaninová I, Slavov N, Smaili S, Smalley K, Smith D, Soenen S, Soleimanpour S, Solhaug A, Somasundaram K, Son J, Sonawane A, Song C, Song F, Song H, Song J, Song W, Soo K, Sood A, Soong T, Soontornniyomkij V, Sorice M, Sotgia F, Soto-Pantoja D, Sotthibundhu A, Sousa M, Spaink H, Span P, Spang A, Sparks J, Speck P, Spector S, Spies C, Springer W, Clair D, Stacchiotti A, Staels B, Stang M, Starczynowski D, Starokadomskyy P, Steegborn C, Steele J, Stefanis L, Steffan J, Stellrecht C, Stenmark H, Stepkowski T, Stern S, Stevens C, Stockwell B, Stoka V, Storchova Z, Stork B, Stratoulas V, Stravopodis D, Strnad P, Strohecker A, Ström A, Stromhaug P, Stulik J, Su Y, Su Z, Subauste C, Subramaniam S, Sue C, Suh S, Sui X, Sukserree S, Sulzer D, Sun F, Sun J, Sun J, Sun S, Sun Y, Sun Y, Sun Y, Sundaramoorthy V, Sung J, Suzuki H, Suzuki K, Suzuki N, Suzuki T, Suzuki Y, Swanson M, Swanton C, Swärd K, Swarup G, Sweeney S, Sylvester P, Szatmari Z, Szegezdi E, Szlosarek P, Taegtmeier H, Tafani M, Taillebourg E, Tait S, Takacs-Vellai K, Takahashi Y, Takáts S, Takemura G, Takigawa N, Talbot N, Tamagno E, Tamburini J, Tan C, Tan L, Tan M, Tan M, Tan Y, Tanaka K, Tanaka M, Tang D, Tang D, Tang G, Tanida I, Tanji K, Tannous B, Tapia J, Tasset-Cuevas I, Tatar M, Tavassoly I, Tavernarakis N, Taylor A, Taylor G, Taylor G, Taylor J, Taylor M, Tchetina E, Tee A, Teixeira-Clerc F, Telang S, Tencomnao T, Teng B, Teng R, Terro F, Tettamanti G, Theiss A, Theron A, Thomas K, Thomé M, Thomes P, Thorburn A, Thorner J, Thum T, Thumm M, Thurston T, Tian L, Till A, Ting J, Titorenko V, Toker L, Toldo S, Tooze S, Topisirovic I, Torgersen M, Torosantucci L, Torriglia A, Torrisi M, Tournier C, Towns R, Trajkovic V, Travassos L, Triola G, Tripathi D, Trisciuglio D, Troncoso R,

Trougakos I, Truttmann A, Tsai K, Tschan M, Tseng Y, Tsukuba T, Tsung A, Tsvetkov A, Tu S, Tuan H, Tucci M, Tumbarello D, Turk B, Turk V, Turner R, Tveita A, Tyagi S, Ubukata M, Uchiyama Y, Udelnow A, Ueno T, Umekawa M, Umemiya-Shirafuji R, Underwood B, Ungermann C, Ureshino R, Ushioda R, Uversky V, Uzcátegui N, Vaccari T, Vaccaro M, Váchová L, Vakifahmetoglu-Norberg H, Valdor R, Valente E, Vallette F, Valverde A, Van, den, Berghe G, Van, Den, Bosch L, van, den, Brink G, van, der, Goot F, van, der, Klei I, van, der, Laan L, van, Doorn W, van, Egmond M, van, Golen K, Van, Kaer L, van, Lookeren, Campagne M, Vandenabeele P, Vandenberghe W, Vanhorebeek I, Varela-Nieto I, Vasconcelos M, Vasko R, Vavvas D, Vega-Naredo I, Velasco G, Velentzas A, Velentzas P, Vellai T, Vellenga E, Vendelbo M, Venkatachalam K, Ventura N, Ventura S, Veras P, Verdier M, Vertessy B, Viale A, Vidal M, Vieira H, Vierstra R, Vigneswaran N, Vij N, Vila M, Villar M, Villar V, Villarroja J, Vindis C, Viola G, Viscomi M, Vitale G, Vogl D, Voitsekhovskaja O, von, Haefen C, von, Schwarzenberg K, Voth D, Vouret-Craviari V, Vuori K, Vyas J, Waeber C, Walker C, Walker M, Walter J, Wan L, Wan X, Wang B, Wang C, Wang C, Wang C, Wang C, Wang C, Wang D, Wang F, Wang F, Wang G, Wang H, Wang H, Wang H, Wang H, Wang H, Wang J, Wang J, Wang M, Wang M, Wang P, Wang P, Wang R, Wang S, Wang T, Wang X, Wang X, Wang X, Wang X, Wang Y, Wang Y, Wang Y, Wang Y, Wang Y, Wang Y, Wang Y, Wang Y, Wang Y, Wang Z, Wappner P, Ward C, Ward D, Warnes G, Watada H, Watanabe Y, Watase K, Weaver T, Weekes C, Wei J, Weide T, Wehl C, Weindl G, Weis S, Wen L, Wen X, Wen Y, Westermann B, Weyand C, White A, White E, Whitton J, Whitworth A, Wiels J, Wild F, Wildenberg M, Wileman T, Wilkinson D, Wilkinson S, Willbold D, Williams C, Williams K, Williamson P, Winklhofer K, Witkin S, Wohlgemuth S, Wollert T, Wolvetang E, Wong E, Wong G, Wong R, Wong V, Woodcock E, Wright K, Wu C, Wu D, Wu G, Wu J, Wu J, Wu M, Wu M, Wu S, Wu W, Wu Y, Wu Z, Xavier C, Xavier R, Xia G, Xia T, Xia W, Xia Y, Xiao H, Xiao J, Xiao S, Xiao W, Xie C, Xie Z, Xie Z, Xilouri M, Xiong Y, Xu C, Xu C, Xu F, Xu H, Xu H, Xu J, Xu J, Xu J, Xu L, Xu X, Xu Y, Xu Y, Xu Z, Xu Z, Xue Y, Yamada T, Yamamoto A, Yamanaka K, Yamashina S, Yamashiro S, Yan B, Yan B, Yan X, Yan Z, Yanagi Y, Yang D, Yang J, Yang L, Yang M, Yang P, Yang P, Yang Q, Yang W, Yang W, Yang X, Yang Y, Yang Y, Yang Z, Yang Z, Yao M, Yao P, Yao X, Yao Z, Yao Z, Yasui L, Ye M, Yedvobnick B, Yeganeh B, Yeh E, Yeyati P, Yi F, Yi L, Yin X, Yip C, Yoo Y, Yoo Y,

Yoon S, Yoshida K, Yoshimori T, Young K, Yu H, Yu J, Yu J, Yu J, Yu L, Yu W, Yu X, Yu Z, Yuan J, Yuan Z, Yue B, Yue J, Yue Z, Zacks D, Zacksenhaus E, Zaffaroni N, Zaglia T, Zakeri Z, Zecchini V, Zeng J, Zeng M, Zeng Q, Zervos A, Zhang D, Zhang F, Zhang G, Zhang G, Zhang H, Zhang H, Zhang H, Zhang H, Zhang J, Zhang J, Zhang J, Zhang J, Zhang J, Zhang L, Zhang L, Zhang L, Zhang L, Zhang M, Zhang X, Zhang X, Zhang Y, Zhang Y, Zhang Y, Zhang Y, Zhang Y, Zhao M, Zhao W, Zhao X, Zhao Y, Zhao Y, Zhao Y, Zhao Y, Zhao Z, Zhao Z, Zheng D, Zheng X, Zheng X, Zhivotovsky B, Zhong Q, Zhou G, Zhou G, Zhou H, Zhou S, Zhou X, Zhu H, Zhu H, Zhu W, Zhu W, Zhu X, Zhu Y, Zhuang S, Zhuang X, Ziparo E, Zois C, Zoladek T, Zong W, Zorzano A és Zughaier S (2016) Guidelines for use and interpretation of assays for monitoring autophagy (3rd edition). *Autophagy* 12, 1–222.

Knüppel R, Dietze P, Lehnberg W, Frech K és Wingender E (1994) TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J. Comput. Biol.* 1, 191–198.

Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L, Mardis ER, Wilson RK, Ally A, Balasundaram M, Butterfield YSN, Carlsen R, Carter C, Chu A, Chuah E, Chun H-JE, Coope RJN, Dhalla N, Guin R, Hirst C, Hirst M, Holt RA, Lee D, Li HI, Mayo M, Moore RA, Mungall AJ, Pleasance E, Gordon Robertson A, Schein JE, Shafiei A, Sipahimalani P, Slobodan JR, Stoll D, Tam A, Thiessen N, Varhol RJ, Wye N, Zeng T, Zhao Y, Birol I, Jones SJM, Marra MA, Cherniack AD, Saksena G, Onofrio RC, Pho NH, Carter SL, Schumacher SE, Tabak B, Hernandez B, Gentry J, Nguyen H, Crenshaw A, Ardlie K, Beroukhi R, Winckler W, Getz G, Gabriel SB, Meyerson M, Chin L, Park PJ, Kucherlapati R, Hoadley KA, Todd Auman J, Fan C, Turman YJ, Shi Y, Li L, Topal MD, He X, Chao H-H, Prat A, Silva GO, Iglesia MD, Zhao W, Usary J, Berg JS, Adams M, Booker J, Wu J, Gulabani A, Bodenheimer T, Hoyle AP, Simons J V., Soloway MG, Mose LE, Jefferys SR, Balu S, Parker JS, Neil Hayes D, Perou CM, Malik S, Mahurkar S, Shen H, Weisenberger DJ, Triche Jr T, Lai PH, Bootwalla MS, Maglinte DT, Berman BP, Van Den Berg DJ, Baylin SB, Laird PW, Creighton CJ, Donehower LA, Getz G, Noble M, Voet D, Saksena G, Gehlenborg N, DiCara D, Zhang J, Zhang H, Wu C-J, Yingchun Liu S, Lawrence MS, Zou L, Sivachenko A, Lin P, Stojanov P, Jing R, Cho J, Sinha R, Park RW, Nazaire M-

D, Robinson J, Thorvaldsdottir H, Mesirov J, Park PJ, Chin L, Reynolds S, Kreisberg RB, Bernard B, Bressler R, Erkkila T, Lin J, Thorsson V, Zhang W, Shmulevich I, Ciriello G, Weinhold N, Schultz N, Gao J, Cerami E, Gross B, Jacobsen A, Sinha R, Arman Aksoy B, Antipin Y, Reva B, Shen R, Taylor BS, Ladanyi M, Sander C, Anur P, Spellman PT, Lu Y, Liu W, Verhaak RRG, Mills GB, Akbani R, Zhang N, Broom BM, Casasent TD, Wakefield C, Unruh AK, Baggerly K, Coombes K, Weinstein JN, Haussler D, Benz CC, Stuart JM, Benz SC, Zhu J, Szeto CC, Scott GK, Yau C, Paull EO, Carlin D, Wong C, Sokolov A, Thusberg J, Mooney S, Ng S, Goldstein TC, Ellrott K, Grifford M, Wilks C, Ma S, Craft B, Yan C, Hu Y, Meerzaman D, Gastier-Foster JM, Bowen J, Ramirez NC, Black AD, Pyatt RE, White P, Zmuda EJ, Frick J, Lichtenberg TM, Brookens R, George MM, Gerken MA, Harper HA, Leraas KM, Wise LJ, Tabler TR, McAllister C, Barr T, Hart-Kothari M, Tarvin K, Saller C, Sandusky G, Mitchell C, Iacocca M V., Brown J, Rabeno B, Czerwinski C, Petrelli N, Dolzhansky O, Abramov M, Voronina O, Potapova O, Marks JR, Suchorska WM, Murawa D, Kycler W, Ibbs M, Korski K, Spychała A, Murawa P, Brzeziński JJ, Perz H, Łażniak R, Teresiak M, Tatka H, Leporowska E, Bogusz-Czerniewicz M, Malicki J, Mackiewicz A, Wiznerowicz M, Van Le X, Kohl B, Viet Tien N, Thorp R, Van Bang N, Sussman H, Duc Phu B, Hajek R, Phi Hung N, Viet The Phuong T, Quyet Thang H, Zaki Khan K, Penny R, Mallery D, Curley E, Shelton C, Yena P, Ingle JN, Couch FJ, Lingle WL, King TA, Maria Gonzalez-Angulo A, Mills GB, Dyer MD, Liu S, Meng X, Patangan M, Waldman F, Stöppler H, Kimryn Rathmell W, Thorne L, Huang M, Boice L, Hill A, Morrison C, Gaudioso C, Bshara W, Daily K, Egea SC, Pegram MD, Gomez-Fernandez C, Dhir R, Bhargava R, Brufsky A, Shriver CD, Hooke JA, Leigh Campbell J, Mural RJ, Hu H, Somiari S, Larson C, Deyarmin B, Kvecher L, Kovatich AJ, Ellis MJ, King TA, Hu H, Couch FJ, Mural RJ, Stricker T, White K, Olopade O, Ingle JN, Luo C, Chen Y, Marks JR, Waldman F, Wiznerowicz M, Bose R, Chang L-W, Beck AH, Maria Gonzalez-Angulo A, Pihl T, Jensen M, Sfeir R, Kahn A, Chu A, Kothiyal P, Wang Z, Snyder E, Pontius J, Ayala B, Backus M, Walton J, Baboud J, Berton D, Nicholls M, Srinivasan D, Raman R, Girshik S, Kigonya P, Alonso S, Sanbhadti R, Barletta S, Pot D, Sheth M, Demchok JA, Mills Shaw KR, Yang L, Eley G, Ferguson ML, Tarnuzzer RW, Zhang J, Dillon LAL, Buetow K, Fielding P, Ozenberger BA, Guyer MS, Sofia HJ és Palchik JD (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70.

Kolch W (2005) Coordinating ERK/MAPK signalling through scaffolds and inhibitors. *Nat. Rev. Mol. Cell Biol.* 6, 827–837.

Kolmogorov AN (1933) Sulla Determinazione Empirica di una Legge di Distribuzione. *G. dell'Istituto Ital. degli Attuari* 4, 83–91.

Kondoh K, és Nishida E (2007) Regulation of MAP kinases by MAP kinase phosphatases. *Biochim. Biophys. Acta* 1773, 1227–1237.

Koppikar P, Bhagwat N, Kilpivaara O, Manshoury T, Adli M, Hricik T, Liu F, Saunders LM, Mullally A, Abdel-Wahab O, Leung L, Weinstein A, Marubayashi S, Goel A, Gönen M, Estrov Z, Ebert BL, Chiosis G, Nimer SD, Bernstein BE, Verstovsek S és Levine RL (2012) Heterodimeric JAK–STAT activation as a mechanism of persistence to JAK2 inhibitor therapy. *Nature* 489, 155–159.

Korcsmáros T, Farkas IJ, Szalay MS, Rovó P, Fazekas D, Spiró Z, Böde C, Lenti K, Vellai T és Csermely P (2010) Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics* 26, 2042–2050.

Kovács IA, Palotai R, Szalay MS és Csermely P (2010) Community Landscapes: An Integrative Approach to Determine Overlapping Network Module Hierarchy, Identify Key Nodes and Predict Network Dynamics. *PLoS One* 5, e12528.

Kozomara A, és Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39, D152–D157.

Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M és Rajewsky N (2005) Combinatorial microRNA target predictions. *Nat. Genet.* 37, 495–500.

Kriventseva E V, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, Ioannidis P és Zdobnov EM (2015) OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* 43, D250–D256.

Krol J, Loedige I és Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* 11, 597–610.

Kubisch J, Túrei D, Földvári-Nagy L, Dunai Z, Zsákai L, Varga M, Vellai T, Csermely P és Korcsmáros T (2013) Complex regulation of autophagy in cancer—Integrated approaches to discover the networks that hold a double-edged sword. *Semin. Cancer Biol.* 23, 252–261.

Ladbury JE, és Arold ST (2012) Noise in cellular signaling pathways: causes and effects. *Trends Biochem. Sci.* 37, 173–178.

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

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway L a., Golub TR, Meyerson M, Gabriel SB, Lander ES és Getz G (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501.

Lee RC, Feinbaum RL és Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854.

Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge J V, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, Lawrence MS, Gonzalez-Perez A, Tamborero D, Cheng Y, Ryslik G a, Lopez-Bigas N, Getz G, Ding L és Raphael BJ (2015) Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* 47, 106–114.

Lemmon MA, és Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. *Cell* 141, 1117–1134.

Levine B, és Kroemer G (2008a) SnapShot: Macroautophagy. *Cell* 132, 2007–2009.

Levine B, és Kroemer G (2008b) Autophagy in the Pathogenesis of Disease. *Cell* 132, 27–42.

Lewis BP, Burge CB és Bartel DP (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* 120, 15–20.

Li J, Yang B, Zhou Q, Wu Y, Shang D, Guo Y, Song Z, Zheng Q és Xiong J (2013a) Autophagy promotes hepatocellular carcinoma cell invasion through activation of epithelial–mesenchymal transition. *Carcinogenesis* 34, 1343–1351.

Li J-H, Liu S, Zhou H, Qu L-H és Yang J-H (2014) starBase v2.0: decoding miRNA–ceRNA, miRNA–ncRNA and protein–RNA interaction networks from large-scale CLIP–Seq data. *Nucleic Acids Res.* 42, D92–D97.

Li L, Zhang K, Lee J, Cordes S, Davis DP és Tang Z (2009) Discovering cancer genes by integrating network and functional properties. *BMC Med. Genomics* 2, 61.

Li M, Zhang H, Wang J és Pan Y (2012) A new essential protein discovery method based on the integration of protein–protein interaction and gene expression data. *BMC Syst. Biol.* 6, 15.

Li M, Wang J-X, Wang H és Pan Y (2013b) Identification of essential proteins from weighted protein–protein interaction networks. *J. Bioinform. Comput. Biol.* 11, 1341002.

Li M, Lu Y, Niu Z és Wu F (2017) United Complex Centrality for Identification of Essential Proteins from PPI Networks. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 14, 370–380.

Li WW, Chen L, Li X, Jia X, Feng C, Zhang L, He W, Lv J, He Y, Li WW, Qu X, Zhou Y és Shi Y (2013c) Cancer-related marketing centrality motifs acting as pivot units in the human signaling network and mediating cross-talk between biological pathways. *Mol. Biosyst.* 9, 3026–3035.

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

Lim W, Mayer B és Pawson T (2015) *Cell signaling : principles and mechanisms* Garland Science, Taylor & Francis Group, New York.

Logan CY, Nusse R, Catriona Y. Logan és Roel Nusse (2004) The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* 20, 781–810.

Lu Z, és Xu S (2006) ERK1/2 MAP kinases in cell survival and apoptosis. *IUBMB Life* 58, 621–631.

Luo K (2017) Signaling Cross Talk between TGF- β /Smad and Other Signaling Pathways. *Cold Spring Harb. Perspect. Biol.* 9, a022137.

Luo J, és Qi Y (2015) Identification of Essential Proteins Based on a New Combination of Local Interaction Density and Protein Complexes. *PLoS One* 10, e0131418.

Lutz M, és Knaus P (2002) Integration of the TGF-beta pathway into the cellular signalling network. *Cell. Signal.* 14, 977–988.

Macdonald BT, Semenov M V. és He X (2007) SnapShot: Wnt/beta-catenin signaling. *Cell* 131, 1204.

Mann HB, és Whitney DR (1947) On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Stat.* 18, 50–60.

Manzano RG, Martinez-Navarro EM, Forteza J és Brugarolas A (2014) Microarray phosphatome profiling of breast cancer patients unveils a complex phosphatase regulatory role of the MAPK and PI3K pathways in estrogen receptor-negative breast cancers. *Int. J. Oncol.* 45, 2250–2266.

Mason O, és Verwoerd M (2007) Graph theory and networks in biology. *IET Syst. Biol.* 1, 89–119.

Mathelier A, Fornes O, Arenillas DJ, Chen C-Y, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A és Wasserman WW (2015) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44, D110-5.

McDaniell R, Warthen DM, Sanchez-Lara PA, Pai A, Krantz ID, Piccoli DA és Spinner NB (2006) NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the Notch signaling pathway. *Am. J. Hum. Genet.* 79, 169–173.

Medema RH, és Bos JL (1993) The role of p21ras in receptor tyrosine kinase signaling.

Crit. Rev. Oncog. 4, 615–661.

Melas IN, Sakellaropoulos T, Iorio F, Alexopoulos LG, Loh W-Y, Lauffenburger D a., Saez-Rodriguez J és Bai JPF (2015) Identification of drug-specific pathways based on gene expression data: application to drug induced lung injury. *Integr. Biol.* 7, 904–920.

Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J és Barabási A-L (2015) Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 347, 1257601.

Mercer SE, és Friedman E (2006) Mirk/Dyrk1B: a multifunctional dual-specificity kinase involved in growth arrest, differentiation, and cell survival. *Cell Biochem. Biophys.* 45, 303–315.

Miller CHT, Maher SG és Young HA (2009) Clinical Use of Interferon-gamma. *Annu. New York Acad. Sci.* 1182, 69–79.

Mimeault M, és Batra SK (2010) Frequent deregulations in the hedgehog signaling network and cross-talks with the epidermal growth factor receptor pathway involved in cancer progression and targeted therapies. *Pharmacol. Rev.* 62, 497–524.

Mitchell PJ, és Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245, 371–378.

Mizushima N, és Klionsky DJ (2007) Protein Turnover Via Autophagy: Implications for Metabolism. *Annu. Rev. Nutr.* 27, 19–40.

Módos D, Brooks J, Fazekas D, Ari E, Vellai T, Csermely P, Korcsmáros T és Lenti K (2016) Identification of critical paralog groups with indispensable roles in the regulation of signaling flow. *Sci. Rep.* 6, 38588.

Módos D, Bulusu KC, Fazekas D, Kubisch J, Brooks J, Marczell I és Szabó PM (2017) Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anticancer therapies. *npj Syst. Biol. Appl.* 3.

Moi P, Chan K, Asunis I, Cao A és Kan YW (1994a) Isolation of NF-E2-related factor 2

(Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the beta-globin locus control region. *Proc. Natl. Acad. Sci. U. S. A.* 91, 9926–9930.

Moi P, Chant K, Asunis I, Cao A, Kant YW és Kan YW (1994b) Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the f-globin locus control region. *Genetics* 91, 9926–9930.

Montgomery SB, Griffith OL, Sleumer MC, Bergman CM, Bilenky M, Pleasance ED, Prychyna Y, Zhang X és Jones SJM (2006) ORegAnno: an open access database and curation system for literature-derived promoters, transcription factor binding sites and regulatory variation. *Bioinformatics* 22, 637–640.

Motohashi H, és Yamamoto M (2004) Nrf2–Keap1 defines a physiologically important stress response mechanism. *Trends Mol. Med.* 10, 549–557.

Moussay E, Kaoma T, Baginska J, Muller A, Van Moer K, Nicot N, Nazarov P V, Vallar L, Chouaib S, Berchem G és Janji B (2011) The acquisition of resistance to TNF α in breast cancer cells is associated with constitutive activation of autophagy as revealed by a transcriptome analysis using a custom microarray. *Autophagy* 7, 760–770.

Muniategui A, Nogales-Cadenas R, Vázquez M, Aranguren XL, Agirre X, Luttun A, Prosper F, Pascual-Montano A, Rubio A, L Aranguren X, Agirre X, Luttun A, Prosper F, Pascual-Montano A, Rubio A és Aranguren XL (2012) Quantification of miRNA-mRNA interactions. *PLoS One* 7, e30766.

Newman MEJ (2006) Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* 103, 8577–8582.

Niehhs C (2012) The complex world of WNT receptor signalling. *Nat. Rev. Mol. Cell Biol.* 13, 767–779.

Nikoloski Z, Perez-Storey R és Sweetlove LJ (2015) Inference and Prediction of Metabolic Network Fluxes. *Plant Physiol.* 169, 1443–1455.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H és Kanehisa M (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27, 29–34.

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 és Hermjakob H (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42, D358–D363.

Padmanabhan B, Tong KI, Ohta T, Nakamura Y, Scharlock M, Ohtsuji M, Kang M-I, Kobayashi A, Yokoyama S és Yamamoto M (2006) Structural Basis for Defects of Keap1 Activity Provoked by Its Point Mutations in Lung Cancer. *Mol. Cell* 21, 689–700.

Palla G, Farkas I, Derényi I és Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 814–818.

Palotai R, Szalay MS és Csermely P (2008) Chaperones as integrators of cellular networks: Changes of cellular integrity in stress and diseases. *IUBMB Life* 60, 10–18.

Papadopoulos GL, Reczko M, Simossis V a, Sethupathy P és Hatzigeorgiou AG (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res.* 37, D155–D158.

Papp D, Lenti K, Módos D, Fazekas D, Dúl Z, Türei D, Földvári-Nagy L, Nussinov R, Csermely P és Korcsmáros T (2012) The NRF2-related interactome and regulome contain multifunctional proteins and fine-tuned autoregulatory loops. *FEBS Lett.* 586, 1795–1802.

Pasquinelli AE (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.* 13, 271–282.

Patterson KI, Brummer T, O'Brien PM és Daly RJ (2009) Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochem. J.* 418, 475–489.

Pencik J, Pham HTT, Schmoellerl J, Javaheri T, Schlederer M, Culig Z, Merkel O, Moriggl R, Grebien F és Kenner L (2016) JAK-STAT signaling in cancer: From cytokines to non-coding genome. *Cytokine* 87, 26–36.

Perez-Lopez ÁR, Szalay KZ, Türei D, Módos D, Lenti K, Korcsmáros T és Csermely P (2015) Targets of drugs are generally, and targets of drugs having side effects are specifically good spreaders of human interactome perturbations. *Sci. Rep.* 5, 10182.

Perfetto L, Briganti L, Calderone A, Cerquone Perpetuini A, Iannuccelli M, Langone F, Licata L, Marinkovic M, Mattioni A, Pavlidou T, Peluso D, Petrilli LL, Pirrò S, Posca D, Santonico E, Silvestri A, Spada F, Castagnoli L és Cesareni G (2016) SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.* 44, D548–D554.

Pires-daSilva A, és Sommer RJ (2003) The evolution of signalling pathways in animal development. *Nat. Rev. Genet.* 4, 39–49.

Portales-Casamar E, Arenillas D, Lim J, Swanson MI, Jiang S, McCallum A, Kirov S és Wasserman WW (2009) The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Res.* 37, D54–D60.

Qabaja A, Alshalalfa M, Alanazi E és Alhajj R (2014) Prediction of novel drug indications using network driven biological data prioritization and integration. *J. Cheminform.* 6, 1.

Qi M, és Elion EA (2005) MAP kinase pathways. *J. Cell Sci.* 118, 3569–3572.

Qu H, és Fang X (2013) A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. *Genomics. Proteomics Bioinformatics* 11, 135–141.

Raingaud J, Whitmarsh AJ, Barrett T, Dérijard B és Davis RJ (1996) MKK3- and MKK6-regulated gene expression is mediated by the p38 mitogen-activated protein kinase signal transduction pathway. *Mol. Cell. Biol.* 16, 1247–1255.

Raman M, Chen W és Cobb MH (2007) Differential regulation and properties of MAPKs.

Oncogene 26, 3100–3112.

Rawlings JS, Rosler KM és Harrison DA (2004) The JAK/STAT signaling pathway. *J. Cell Sci.* 117, 1281–1283.

Rayasam GV, Tulasi VK, Sodhi R, Davis JA és Ray A (2009) Glycogen synthase kinase 3: more than a namesake. *Br. J. Pharmacol.* 156, 885–898.

Reimand J, és Bader GD (2014) Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers. *Mol. Syst. Biol.* 9, 637.

Rhee SY, Wood V, Dolinski K és Draghici S (2008) Use and misuse of the gene ontology annotations. *Nat. Rev. Genet.* 9, 509–515.

Riobo N a., Lu K, Emerson CP, Emerson, Jr CP és Emerson CP (2006) Hedgehog signal transduction: Signal integration and cross talk in development and cancer. *Cell Cycle* 5, 1612–1615.

Ross DA, és Kadesch T (2001) The Notch Intracellular Domain Can Function as a Coactivator for LEF-1. *Mol. Cell. Biol.* 21, 7537–7544.

Roychowdhury A, Samadder S, Das P, Mandloi S, Addya S, Chakraborty C, Basu PS, Mondal R, Roy A, Chakrabarti S, Roychoudhury S és Panda CK (2017) Integrative genomic and network analysis identified novel genes associated with the development of advanced cervical squamous cell carcinoma. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 2899–2911.

Rubinsztein DC, Gestwicki JE, Murphy LO és Klionsky DJ (2007) Potential therapeutic applications of autophagy. *Nat. Rev. Drug Discov.* 6, 304–312.

Sabidussi G (1966) The centrality of a graph. *Psychometrika* 31, 581–603.

Salwinski L (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* 32, D449–D451.

Sam SA, Teel J, Tegge AN, Bharadwaj A és Murali TM (2017) XTalkDB: a database of signaling pathway crosstalk. *Nucleic Acids Res.* 45, D432–D439.

Samaga R, és Klamt S (2013) Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. *Cell Commun. Signal.* 11, 43.

Samanta D, és Datta PK (2012) Alterations in the Smad pathway in human cancers. *Front. Biosci. (Landmark Ed.* 17, 1281–1293.

Sarbassov DD, Ali SM és Sabatini DM (2005) Growing roles for the mTOR pathway. *Curr. Opin. Cell Biol.* 17, 596–603.

Schneider TD, és Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100.

Scott DE, Bayly AR, Abell C és Skidmore J (2016) Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nat. Rev. Drug Discov.* 15, 533–550.

Seeger R, és Krebs EG (1995) The MAPK signaling cascade. *FASEB J.* 9, 726–735.

Semenov M V., Habas R, MacDonald BT és He X (2007) SnapShot: Noncanonical Wnt Signaling Pathways. *Cell* 131, 11–12.

Sethupathy P, Corda B és Hatzigeorgiou AG (2006) TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA* 12, 192–197.

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

Shi Y, Vattam KM, Sood R, An J, Liang J, Stramm L és Wek RC (1998) Identification and characterization of pancreatic eukaryotic initiation factor 2 alpha-subunit kinase, PEK, involved in translational control. *Mol. Cell. Biol.* 18, 7499–7509.

Shibata T, Kokubu A, Gotoh M, Ojima H, Ohta T, Yamamoto M és Hirohashi S (2008a) Genetic Alteration of Keap1 Confers Constitutive Nrf2 Activation and Resistance to Chemotherapy in Gallbladder Cancer. *Gastroenterology* 135, 1358–1368.e4.

Shibata T, Ohta T, Tong KI, Kokubu A, Odogawa R, Tsuta K, Asamura H, Yamamoto

M és Hirohashi S (2008b) Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc. Natl. Acad. Sci.* 105, 13568–13573.

Siegel R, Naishadham D és Jemal A (2012) Cancer statistics, 2012. *CA. Cancer J. Clin.* 62, 10–29.

Smirnov N (1948) Table for Estimating the Goodness of Fit of Empirical Distributions. *Annu. Math. Stat.* 19, 279–281.

Sonnhammer ELL, és Östlund G (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* 43, D234-9.

Srihari S, és Ragan M a (2013) Systematic tracking of dysregulated modules identifies novel genes in cancer. *Bioinformatics* 29, 1553–1561.

Stanley P, és Guidos CJ (2009) Regulation of Notch signaling during T- and B-cell development by O-fucose glycans. *Immunol. Rev.* 230, 201–215.

Stark C (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539.

Stratton MR, Campbell PJ és Futreal PA (2009) The cancer genome. *Nature* 458, 719–724.

Subramaniam A, Shanmugam MK, Perumal E, Li F, Nachiyappan A, Dai X, Swamy SN, Ahn KS, Kumar AP, Tan BKH, Hui KM és Sethi G (2013) Potential role of signal transducer and activator of transcription (STAT)3 signaling pathway in inflammation, survival, proliferation and invasion of hepatocellular carcinoma. *Biochim. Biophys. Acta* 1835, 46–60.

Di Summa M, Grosso A és Locatelli M (2011) Complexity of the critical node problem over trees. *Comput. Oper. Res.* 38, 1766–1774.

Supper J, Spangenberg L, Planatscher H, Dräger A, Schröder A és Zell A (2009) BowTieBuilder: modeling signal transduction pathways. *BMC Syst. Biol.* 3, 67.

Surh Y-J, Kundu J és Na H-K (2008) Nrf2 as a Master Redox Switch in Turning on the Cellular Signaling Involved in the Induction of Cytoprotective Genes by Some Chemopreventive Phytochemicals. *Planta Med.* 74, 1526–1539.

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

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 és von Mering C (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452.

Taguchi K, Motohashi H és Yamamoto M (2011) Molecular mechanisms of the Keap1-Nrf2 pathway in stress response and cancer evolution. *Genes to Cells* 16, 123–140.

Takahashi T, Sonobe M, Menju T, Nakayama E, Mino N, Iwakiri S, Nagai S, Sato K, Miyahara R, Okubo K, Hirata T, Date H és Wada H (2010) Mutations in Keap1 are a potential prognostic factor in resected non-small cell lung cancer. *J. Surg. Oncol.* 101.

Takahashi Y, Coppola D, Matsushita N, Cualing HD, Sun M, Sato Y, Liang C, Jung JU, Cheng JQ, Mul JJ, Pledger WJ és Wang H-G (2007) Bif-1 interacts with Beclin 1 through UVRAG and regulates autophagy and tumorigenesis. *Nat. Cell Biol.* 9, 1142–1151.

Tambe MB, Narvi E és Kallio M (2016) Reduced levels of Dusp3/Vhr phosphatase impair normal spindle bipolarity in an Erk1/2 activity-dependent manner. *FEBS Lett.* 590, 2757–2767.

Tanida I, Ueno T és Kominami E (2008) LC3 and Autophagy. *Methods Mol. Biol.* 445, 77–88.

Taniguchi CM, Emanuelli B és Kahn CR (2006) Critical nodes in signalling pathways: insights into insulin action. *Nat. Rev. Mol. Cell Biol.* 7, 85–96.

Tegge AN, Sharp N és Murali TM (2015) Xtalk: A path-based approach for identifying crosstalk between signaling pathways. *Bioinformatics* 32, 242–251.

Tew KL, Li X-L és Tan S-H (2007) Functional centrality: detecting lethality of proteins in protein interaction networks. In *Genome Informatics. International Conference on Genome Informatics*, o. 166–177.

The UniProt Consortium (2014) UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212.

Tomlinson JW, Walker EA, Bujalska IJ, Draper N, Lavery GG, Cooper MS, Hewison M és Stewart PM (2004) 11 β -Hydroxysteroid Dehydrogenase Type 1: A Tissue-Specific Regulator of Glucocorticoid Response. *Endocr. Rev.* 25, 831–866.

Türei D, Korcsmáros T és Saez-Rodriguez J (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* 13, 966–967.

Ummanni R, Mundt F, Pospisil H, Venz S, Scharf C, Barrett C, Fälth M, Köllermann J, Walther R, Schlomm T, Sauter G, Bokemeyer C, Sültmann H, Schuppert A, Brümmendorf TH és Balabanov S (2011) Identification of clinically relevant protein targets in prostate cancer with 2D-DIGE coupled mass spectrometry and systems biology network platform. *PLoS One* 6, e16833.

Valente TW, Coronges K, Lakon C és Costenbader E (2008) How Correlated Are Network Centrality Measures? *Connections* 28, 16–26.

Vellaichamy A, Dezso Z, JeBailey L, Chinnaiyan AM, Sreekumar A, Nesvizhskii AI, Omenn GS és Bugrim A (2010) „Topological significance” analysis of gene expression and proteomic profiles from prostate cancer cells reveals key mechanisms of androgen response. *PLoS One* 5, e10936.

Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, Gerangelos S, Koziris N, Dalamagas T és Hatzigeorgiou AG (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res.* 40, D222–D229.

Vinayagam A, Zirin J, Roesel C, Hu Y, Yilmazel B, Samsonova A a, Neumüller R a, Mohr SE és Perrimon N (2014) Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nat. Methods* 11, 94–99.

Vlachos IS, Paraskevopoulou MD, Karagkouni D, Georgakilas G, Vergoulis T, Kanellos I, Anastasopoulos I-L, Maniou S, Karathanou K, Kalfakakou D, Fevgas A, Dalamagas T és Hatzigeorgiou AG (2015) DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res.* 43, D153–D159.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz L a és Kinzler KW (2013) Cancer genome landscapes. *Science* 339, 1546–1558.

Volinia S, Visone R, Galasso M, Rossi E és Croce CM (2010) Identification of microRNA activity by Targets' Reverse EXpression. *Bioinformatics* 26, 91–97.

van der Walt S, Colbert SC és Varoquaux G (2011) The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* 13, 22–30.

Wang J, Lu M, Qiu C és Cui Q (2010) TransmiR: a transcription factor–microRNA regulation database. *Nucleic Acids Res.* 38, D119–D122.

Wang J, Chen G, Li M és Pan Y (2011) Integration of breast cancer gene signatures based on graph centrality. *BMC Syst. Biol.* 5, S10.

Wang J, Li M, Wang H és Pan Y (2012) Identification of essential proteins based on edge clustering coefficient. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 9, 1070–1080.

Wang X, Thijssen B és Yu H (2013) Target Essentiality and Centrality Characterize Drug Side Effects. *PLoS Comput. Biol.* 9, e1003119.

Watts DJ, és Strogatz SH (1998) Collective dynamics of „small-world” networks. *Nature* 393, 440–442.

Wein SP, Côté RG, Dumousseau M, Reisinger F, Hermjakob H és Vizcaíno JA (2012) Improvements in the Protein Identifier Cross-Reference service. *Nucleic Acids Res.* 40, W276–W280.

Wen S, Niu Y, Lee SO és Chang C (2014) Androgen receptor (AR) positive vs negative roles in prostate cancer cell deaths including apoptosis, anoikis, entosis, necrosis and

autophagic cell death. *Cancer Treat. Rev.* 40, 31–40.

Wightman B, Ha I és Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862.

Wu S, Vossius S, Rahmouni S, Miletic A V, Vang T, Vazquez-Rodriguez J, Cerignoli F, Arimura Y, Williams S, Hayes T, Moutschen M, Vasile S, Pellecchia M, Mustelin T és Tautz L (2009) Multidentate small-molecule inhibitors of vaccinia H1-related (VHR) phosphatase decrease proliferation of cervix cancer cells. *J. Med. Chem.* 52, 6716–6723.

Xia J, Sun J, Jia P és Zhao Z (2011) Do cancer proteins really interact strongly in the human protein-protein interaction network? *Comput. Biol. Chem.* 35, 121–125.

Xiao F, Zuo Z, Cai G, Kang S, Gao X és Li T (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.* 37, D105–D110.

Xiong W, Xie L, Zhou S, Liu H és Guan J (2014) The centrality of cancer proteins in human protein-protein interaction network: a revisit. *Int. J. Comput. Biol. Drug Des.* 7, 146–156.

Xu J, és Li Y-H (2012) miRDeathDB: a database bridging microRNAs and the programmed cell death. *Cell Death Differ.* 19, 1571.

Yarden Y, és Ullrich A (1988) Growth Factor Receptor Tyrosine Kinases. *Annu. Rev. Biochem.* 57, 443–478.

Yu H, Kim PM, Sprecher E, Trifonov V és Gerstein M (2007) The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. *PLoS Comput. Biol.* 3, e59.

Yuan K, Chen Z, Li W, Gao CE, Li G, Guo G, Yang Y, Ai Y, Wu L és Zhang M (2015) MDK Protein Overexpression Correlates with the Malignant Status and Prognosis of Non-small Cell Lung Cancer. *Arch. Med. Res.* 46, 635–641.

Zaidi MR, és Merlino G (2011) The two faces of interferon- γ in cancer. *Clin. Cancer Res.* 17, 6118–6124.

Zhang DD (2006) Mechanistic Studies of the Nrf2-Keap1 Signaling Pathway. *Drug Metab. Rev.* 38, 769–789.

Zhang S, Chung W-C és Xu K (2016) Lunatic Fringe is a potent tumor suppressor in Kras-initiated pancreatic cancer. *Oncogene* 35, 2485–2495.

Zhang Y, Han Y, Zheng R, Yu J-H, Miao Y, Wang L és Wang E-H (2012) Expression of Frat1 correlates with expression of β -catenin and is associated with a poor clinical outcome in human SCC and AC. *Tumor Biol.* 33, 1437–1444.

ZHANG W, és LIU HT (2002) MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res.* 12, 9–18.

Zhao Q, Shi X, Xie Y, Huang J, Shia B és Ma S (2015) Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA. *Brief. Bioinform.* 16, 291–303.

Zhou T-T (2012) Network systems biology for targeted cancer therapies. *Chin. J. Cancer* 31, 134–141.

Zhou J, Lam B, Neogi SG, Yeo GSH, Azizan EAB és Brown MJ (2016) Transcriptome Pathway Analysis of Pathological and Physiological Aldosterone-Producing Human Tissues Novelty and Significance. *Hypertension* 68, 1424–1431.

XI. Saját publikációk jegyzéke

1. A dolgozathoz kapcsolódó publikációk

Módos D., Bulusu K. C., Fazekas D, Kubisch J., Brooks J., Marczell I, Szabó P. M., Vellai T., Csermely P., Lenti K., Bender A. és Korcsmáros T. (2017) Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anti-cancer therapies npj Systems Biology and Applications,3: (1) 2 13 p.

Modos D., Brooks J., Fazekas D., Ari E., Vellai T., Csermely P., Korcsmaros T. és Lenti K. (2016) Identification of critical paralog groups with indispensable role in the regulation of signaling flow Scientific Reports 6: 38588 13 p. (IF: 4,259)

Türei D, Földvári-Nagy L, Fazekas D, **Módos D**, Kubisch J, Kadlecsik T, Demeter A, Lenti K, Csermely P, Vellai T és Korcsmáros T. (2015) Autophagy Regulatory Network – a systems-level bioinformatics resource for studying autophagy components and their regulation Autophagy, 11(1):155-165. (IF: 9,108)

Fazekas D.*, Koltai M.*, Türei D.*, **Módos D**, Pálffy M, Dúl Z, Zsákai L, Szalay-Bekó M, Lenti K, Farkas I, Vellai T, Csermely P és Korcsmáros T. (2013) Signalink 2 – a signaling pathway resource with multi-layered regulatory networks. BMC Syst Biol, 7:7. (IF: 2,853)

Türei D., Papp D., Fazekas D., Földvári-Nagy L., **Módos D.**, Lenti K., Csermely P. és Korcsmáros T. (2013) NRF2-ome: an integrated web resource to discover protein interaction and regulatory networks of NRF2. Oxid Med Cell Longev, 2013: 737591. (IF: 3,363)

Papp D, Lenti K, **Módos D**, Fazekas D, Dúl Z, Türei D, Földvári-Nagy L, Nussinov R, Csermely P, Korcsmáros T (2012) The NRF2-related interactome and regulome contain multifunctional proteins and fine-tuned autoregulatory loops. FEBS Letters 586, 13, 1795–1802. (IF: 3,582)

2. *Nem kapcsolódó publikációk*

Csályi K, Fazekas D, Kadlecsek T, Türei D, Gul L, Horváth B, **Módos D**, Demeter A, Pápai N, Lenti K, Csermely P, Vellai T, Korcsmáros T és Varga M (2016) SignaFish: A Zebrafish-Specific Signaling Pathway Resource. Zebrafish April 2016 (IF: 2,242)

Cs. Földvári-Nagy, K. Cs. Földvári-Nagy, **D. Módos**, K. Lenti (2016) Comparative study of hygiene habits in three different groups in Hungary New Medicine 20 4 pp. 141-147

Csermely P., Hódsági J., Korcsmáros T., **Módos D.**, Perez-Lopez Á. R., Szalay K., Veres D. V., Lenti K., Wu L.Y., Zhang X. S. (2014) 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. Seminars in Cancer Biology 30: pp 42-51 (IF: 9.955 – összefoglaló cikk)

Földvári-Nagy Cs, **Módos D.**, Feith H. J., Lenti K. (2015) Quantitative study of the generational changes among relationship habits in highly educated Hungarian population New Medicine 19: (3) pp 100-109

Perez-Lopez Á. R. Szalay K. Z., Türei D., **Módos D.**, Lenti K., Korcsmáros T., és Csermely P. (2015) Targets of drugs are generally, and targets of drugs having side effects are specifically good spreaders of human interactome perturbations. Scientific Reports 5 article number 10182. (IF:5.288)

Pálffy M., Földvári-Nagy L., **Módos D.**, Lenti K., Korcsmáros T. (2013) Reconstruction and Comparison of Cellular Signaling Pathway Resources for the Systems-Level Analysis of Cross-Talks. In: Prokop A., Csukás B. (szerk) Systems Biology: Integrative Biology and Simulation Tools. Springer Science+Business Media, Dordrecht pp. 463-477 (ISBN: 978-94-007-6802-4)

Gyurkó D M, Veres D V, **Módos D**, Lenti K, Korcsmáros T és 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. Seminars in Cancer Biology 23: 262–269 (IF:9.143 – összefoglaló cikk)

XII. Köszönetnyilvánítás

A dolgozatban szeretném megköszönni a rengeteg támogatást Lenti Katalinnak, témavezetőmnek, akihez nemcsak szakmai, hanem lelki segítségért is fordulhattam. Korcsmáros Tamásnak, konzulensemnek, akivel (ha kellett) az éjszaka közepén is alkottunk a dolgozat alapját szolgáló TDK előadást és a cikkeket. Csermely Péternek, aki átadta hálózatos szemléletét, és megismertetett Tamással. Dörnyei Gabriellának, aki lehetőséget adott gyakorlatokat és előadásokat tartani PhD hallgatóként az Egészségtudományi Kar Morfológiai és Fiziológiai Tanszékén, ezzel elősegítve előadói képességem fejlődését. Vellai Tibornak, aki lehetőséget adott, hogy az ELTE Genetikai Tanszéken dolgozhassak. Fazekas Dávidnak, aki segített első lépéseimnél a programozás területén, sok hasznos ötletet adott, és közben pár sört is elfogyasztottunk. Türei Dénesnek, aki elkészítette az adatbázisok magját alkotó SQL honlapokat és a dizájnt. Szuromi Gábornak, aki segített nekem a TF-célgének szekvenciáinak a legenerálásában. Marczell Istvánnak és Szabó Péternek, akik segítettek megoldani a microarrayek egymáshoz normálást. Kubish Jánosnak, akivel együtt kerestük ki azt az ezeröttszáz microarrayt, amivel dolgoztunk. Ari Eszternek hasznos megjegyzéseiért és a közös, gyümölcsöző ebédékért. Földvári-Nagy Lászlónak, aki átnézte az autofágia fejezetet, és az ő irodalmi ismereteivel tudtuk elkészíteni az autofágia regulációs adatbázist. Johanne Brooksnak a cikkeim angol lektorálásáért és megjegyzéseiért. Krishna Bulusunak, aki megtanított, hogy működik a ChEMBL adatbázis, és Andreas Bendernek, aki alkalmazott, még PhD nélküli post-dokként.

Hálával tartozom a Morfológiai és Fiziológiai Tanszék valamennyi munkatársának, akik meghallgattak és segítettek. A NetBiol csoport és az ELTE Genetikai Tanszék valamennyi tagjának a rengeteg hasznos tanácsért, ötletért, közös kávéért.

Végül családomnak és barátaimnak, akik elviselték, hogy távol vagyok tőlük, mert a kutatásaimon dolgozom, és a végén ők hajtottak, hogy befejezzem disszertációm.

XIII. Ábrák táblázatok és egyenletek jegyzéke

1. Ábrajegyzék

1. ábra Az autofágia folyamata emlősökben az egyes résztvevő fehérjékkel.	16
2. ábra Miért ne használjuk a KEGG adatbázist jelátviteli hálózatos modellezésre?.....	25
3. ábra Az NF κ B transzkripció faktor pozíció súlyozott szekvencia logója a JASPAR adatbázisból	29
4. ábra A miRNS-ek érési folyamata (Pasquinelli, 2012)	30
5. ábra A transzkripció faktor célgén kapcsolatok megoszlása az S érték függvényében. A narancssárga görbe mutatja az eredeti eloszlást. Erre illeszkedik a két normál eloszlású valószínűségi változóból (kék és vörös görbe) készített sűrűség függvény. A zöld vonal jelzi a határértékeket a kapcsolatok esetén.	45
6. ábra A specificitás és szenzitivitás meghatározására végzett ROC elemzés.....	46
7. ábra A kritikus paralóg csoportok definíciója	61
8. ábra a kritikus paralógok azonosítása (Módos és mtsai., 2016)	63
9. ábra A kritikus fehérjék hálózatos toplógiai tulajdonságai	66
10. ábra A kritikus paralóg csoportok (KPCS), a nem kritikus paralóg csoportok (nKPCS) regulációja a) és biológiai funkciói a Gene Ontology adatbázis alapján b)(Módos és mtsai., 2016)	68
11. ábra A kritikus paralóg csoportok tagjai nem felcserélhetők a betegségekben, míg a paralóg csoportok tagjai igen.....	69
12. ábra a) Vastagbél daganat és normál vastagbél SignaLink jelátviteli hálózatra vetítve.	73
13. ábra A SignaLink jelátviteli hálózaton a többi három daganat hálózatos centralitásai.	80
14. ábra Az első szomszédok fontossága a különböző daganatokban a Cui és munkatársai jelátviteli hálózatban.....	82
15. ábra Az első szomszédok fontossága a különböző dagantokban a Reactome jelátviteli hálózatban.....	84
16. ábra Az első szomszédok fontossága a különböző dagantokban a Biogrid, DIP és IntAct összevont fehérje-fehérje interakciós hálózatban.....	85
17. ábra Az első szomszédok fontossága a különböző dagantokban a HPRD kézi gyűjtésű fehérje-fehérje interakciós hálózatban.....	87

18. ábra A mutálódott és differenciáltan expresszáldó daganatos fehérjék és első szomszédjaik hálózati centralitásai vastagbél daganatban a Signalink adatbázis használva.	89
19. ábra A colorectalis carcinoma, a mell daganat, a hepatocelluláris carcinoma és nem kissejtes tüdőrák daganatban mutáns és differenciáltan expresszáldó fehérjéinek és első szomszédjaiknak hálózati centralitásai a Cui és munkatársai (2007) manuálisan gyűjtött daganatos jelátviteli hálózaton.	91
20. ábra A colorectalis carcinoma, a mell daganat, a hepatocelluláris carcinoma és nem kissejtes tüdőrák daganatban mutáns és differenciáltan expresszáldó fehérjéinek és első szomszédjaiknak hálózati centralitásai a Reactome manuálisan gyűjtött jelátviteli hálózaton.....	93
21. ábra A colorectalis carcinoma, a mell daganat, a hepatocelluláris carcinoma és nem kissejtes tüdőrák daganatban mutáns és differenciáltan expresszáldó fehérjéinek és első szomszédjaiknak hálózati centralitásai a Biogrid, DIP és IntAct fehérje interakciós hálózaton.....	95
22. ábra A colorectalis carcinoma, a mell daganat, a hepatocelluláris carcinoma és nem kissejtes tüdőrák daganatban mutáns és differenciáltan expresszáldó fehérjéinek és első szomszédjaiknak hálózati centralitásai a kézi gyűjtésű HPRD fehérje interakciós hálózaton.....	97
23. ábra Gyógyszerek és kismolekulák megoszlása a daganatban érintett fehérjék és az első szomszédok között	100
24. ábra Rákban mutációt szenvedő gének és gyógyszercélpontok az autofágia szabályozási hálózatában.....	103
25. ábra A PPAR-gamma és az NRF2 közötti transzkripcionális visszacsatolási kapcsolat.	104
26. ábra A NOTCH kritikus pont hálózata.	106
27. ábra A SMAD2 és a SMAD3 fehérjék szerkezete.	108
28. ábra A MAPK jelátvitel kritikus paralóg csoportjai- élet és halál urai	109
29. ábra A befolyásoló fehérjék a jelátvitelben, mint további gyógyszercélpont jelöltek. (Módos és mtsai., 2017)	114

30. ábra A befolyásoló fehérjék előfordulása az egyes daganatokban aszerint, hogy hány adatbázisban fordultak elő, mint a daganatban érintett első szomszédokat befolyásoló, de az adott daganatban nem érintett fehérjék.	116
31. ábra A colorectalis carcinomában három különböző befolyásoló fehérje gyógyszer-célpontként való alkalmazása.....	117
32. ábra A mell daganat a nem-kissejtes tüdőrákban és a hepatocellularis carcinomában mind az öt (SignaLink, Cui és munkatársai, Reactome, HPRD, IntAct -DIP-BioGrid), általunk vizsgált interakciós hálózat szerint befolyásoló pozícióban lévő fehérjék, mint lehetséges gyógyszer-célpontok.	120

2. Táblázatjegyzék

1. táblázat A daganatban érintett fehérjék meghatározására használt kereső kifejezések a Cancer Gene Census adatbázisban	48
2. táblázat Az alkalmazott adatbázisok a felhasznált adatok megnevezésével.....	51
3. táblázat A módosított szövettani/anatómiai csoportosítás.....	55
4. táblázat A vizsgált microarrayek megoszlása az egészséges és daganatos betegek között	56
5. táblázat Az integrált transzkripciós faktor célgén adatbázisok miRNS célgén és transzkripciós faktor célgén adatbázisok a SignaLink 2 NRF2ome és az ARN adatbázisban.....	60
6. táblázat A SignaLink jelátviteli hálózaton mért centralitási értékek szórása az egyes kritikus paralóg csoportokon (KPCS) illetve nem kritikus paralóg csoportokon (nKPCS) belül mérve.	65
7. táblázat Az első szomszédok a várhatónál nagyobb óriás komponenst alkotnak, míg a nem érintett fehérjék a várhatónál kisebbet.....	75

3. *Egyenletjegyzék*

1. Egyenlet.....	28
2. Egyenlet.....	28
3. Egyenlet.....	34
4. Egyenlet.....	37
5. Egyenlet.....	37
6. Egyenlet.....	37
7. Egyenlet.....	37
8. Egyenlet.....	38
9. Egyenlet.....	44
10. Egyenlet.....	53
11. Egyenlet.....	58