

Komplex jelátviteli és szabályozási hálózatok összeállítása és vizsgálata

Doktori értekezés

Türei Dénes

Semmelweis Egyetem
Molekuláris Orvostudományok Doktori Iskola



Témavezető:

Dr. Csermely Péter,
egyetemi tanár, az MTA levelező tagja

Hivatalos bírálók:

Dr. Ari Eszter, egyetemi tanársegéd, PhD
Dr. Cserző Miklós,
tudományos főmunkatárs, PhD

Szigorlati bizottság elnöke:

Dr. Tóth Sára, egyetemi docens, PhD

Szigorlati bizottság tagjai:

Dr. Káldi Krisztina, egyetemi docens, PhD
Dr. Hetényi Csaba,
tudományos főmunkatárs, PhD

Budapest
2014

Tartalomjegyzék

Tartalomjegyzék	2
Ábrák jegyzéke	5
Táblázatok jegyzéke	5
Rövidítések jegyzéke	6
1. Bevezető	12
1.1. Jelátviteli útvonalak	14
1.2. Jelátviteli rendszerek és hálózatok	14
1.3. A poszt-transzlációs kapcsolatok vizsgálata nagy teljesítményű módszerekkel	19
1.4. A transzkripcionális szabályozás és vizsgálati módszerei	21
1.5. Poszt-transzkripcionális szabályozás	22
1.6. Bioinformatikai adatforrások	25
1.6.1. Általános fehérje interakciós adatbázisok	25
1.6.2. A fehérje-fehérje interakciók megbízhatóságának becslése	26
1.6.2.1. PRINCESS.	26
1.6.2.2. A Gene Ontology tulajdonságok szemantikus hasonlósága.	26
1.6.3. Jelátviteli útvonal adatbázisok	27
1.6.4. Transzkripció faktor–gén regulációs adatbázisok	32
1.6.5. mikro-RNS-mRNS interakciós adatbázisok	34
1.6.6. Transzkripció faktor–mikro-RNS gén regulációs adatbázisok	35
1.7. Az NRF2 transzkripció faktor	36
1.8. Az autofágia	39
1.8.1. Az autofágiát szabályozó útvonalak	44
1.8.2. Az autofágia transzkripcionális szabályozása	47
1.8.3. Az autofágia poszt-transzkripcionális szabályozása	48
1.8.4. Az autofágia rendszerszintű megközelítése	48
2. Célkitűzések	51
3. Módszerek	53
3.1. A fejlesztések kiindulópontjai	53
3.2. Adatforrások	53
3.2.1. Kézi gyűjtés	54
3.2.2. Fehérje és mikro-RNS referencia adatbázisok	55
3.2.2.1. UniProt.	55
3.2.2.2. WormBase, FlyBase, Ensembl.	55
3.2.2.3. miRBase.	56
3.2.3. Fehérje-fehérje interakciós adatbázisok	56
3.2.4. A predikciók során felhasznált adatbázisok	57

3.2.5.	Predikció domén-motívum kölcsönhatások alapján	59
3.2.6.	Predikció domén-domén kapcsolatok alapján	60
3.2.6.1.	Kapcsolat predikció.	61
3.2.6.2.	Írány predikció.	61
3.2.7.	Transzkripciós faktor–gén regulációs adatbázisok	62
3.2.8.	mikro-RNS–mRNS interakciós adatbázisok	63
3.2.9.	Transzkripciós faktor–miRNS gén regulációs adatbázisok	63
3.3.	Az interakciók tulajdonságai	64
3.3.1.	Írány, hatás, közvetettség	64
3.3.2.	Megbízhatósági mérőszámok	64
3.3.2.1.	PRINCESS.	64
3.3.2.2.	A Gene Ontology tulajdonságok szemantikus hasonlósága.	64
3.3.2.3.	Predikciók megbízhatósági értékei.	65
3.4.	Implementáció	65
3.4.1.	Adatbázis struktúra	66
3.4.2.	Az adatbázisok építése	67
3.4.2.1.	Adatbázis azonosítók fordítása.	67
3.4.3.	Az adatbázis összeállításának folyamata	69
3.4.3.1.	Signalink 2.	69
3.4.3.2.	NRF2ome és ARN.	71
3.4.4.	Weboldalak	72
3.4.5.	A letöltő modul	72
4.	Eredmények	73
4.1.	Az adatbázisok felépítése	73
4.1.1.	A Signalink 2	73
4.1.1.1.	Jelátviteli komponensek.	73
4.1.1.2.	Többretegű szabályozási hálózatok.	74
4.1.2.	Az NRF2ome	75
4.1.3.	Az Autophagy Regulatory Network (ARN)	77
4.2.	Az adatbázisok mennyiségi tulajdonságai	78
4.2.1.	A Signalink 2	78
4.2.2.	Az NRF2ome	79
4.2.3.	Az Autophagy Regulatory Network (ARN)	80
4.2.4.	Az autofágia kapcsolatai jelátviteli útvonalakkal	80
4.2.5.	A Signalink 2, az NRF2ome és az ARN összehasonlítása	81
4.3.	Az adatbázisok internetes felületei	85
4.3.1.	A weboldalak	85
4.3.1.1.	Intelligens névkeresés.	86
4.3.1.2.	Fehérje és mikro-RNS adatlap.	87
4.3.1.3.	Személyre szabott letöltés.	90
4.3.2.	Letöltő modul	91
4.3.3.	Látogatottság monitorozás	95

4.3.4.	Kompatibilitás	95
4.4.	Az adatbázisok elemzése	95
4.4.1.	A Signalink 2 összehasonlítása más jelátviteli adatbázisokkal	95
4.4.2.	A jelátviteli keresztbeszélgetések különböző rétegei	96
4.4.3.	Szabályozási hurkok	98
5.	Megbeszélés	100
5.1.	A Signalink 2 szerepe a jelátvitel kutatásában	100
5.2.	A Signalink 2 jövőbeli fejlesztési irányai	102
5.3.	Az NRF2ome szerepe az NRF2 kutatásában	104
5.3.1.	Poszt-transzlációs szabályozás	104
5.3.2.	Transzkripcionális szabályozás	106
5.3.3.	Poszt-transzkripcionális szabályozás	108
5.3.3.1.	Az NRF2 jelátviteli útvonalak által történő szabályozása.	109
5.4.	Az ARN jelentősége az autofágia kutatásában	109
5.5.	Elemzési lehetőségek a bemutatott adatbázisok felhasználásával	114
5.5.1.	Differenciál expressziós vizsgálatok	114
5.5.2.	Mutációs vizsgálatok	115
5.5.3.	Hálózati perturbációs szimulációk	117
5.5.4.	Modularizációs vizsgálatok	117
5.5.5.	Kritikus pontok vizsgálata	118
5.5.6.	Hálózati motívumok keresése	119
5.5.7.	Extrém útvonal analízis	119
5.5.8.	A hálózatok szerepe a gyógyszerkutatásban	120
6.	Következtetések	122
7.	Összefoglalás	124
8.	Summary	125
9.	Irodalomjegyzék	126
10.	Saját publikációk jegyzéke	151
11.	Köszönetnyilvánítás	154

Ábrák jegyzéke

1.	Szemléletváltás a jelátvitel kutatásában – I.	15
2.	Szemléletváltás a jelátvitel kutatásában – II.	18
3.	Az NRF2 szabályozásának áttekintése	37
4.	A makroautofágia folyamatának áttekintése	41
5.	Az autofágia fehérjék szerepe a fagofór képződésében	43
6.	Az autofágiát szabályozó fontosabb útvonalak	45
7.	A SignaLink 2 adatbázis réteges felépítése	69
8.	A SignaLink 2 adatbázis sémája	71
9.	Az NRF2ome adatbázis felépítése	76
10.	Az ARN felépítése	77
11.	Szabályozási hurkok	85
12.	A weboldal keresőmezőjének működése	86
13.	Fehérje adatlap – I.	87
14.	Fehérje adatlap – II.	88
15.	Az NRF2 kapcsolatait bemutató oldal	90
16.	Letöltési beállítások a SignaLink 2 weboldalán	92
17.	Részletes szűrés a SignaLink 2 weboldalán	93
18.	Letöltési beállítások az NRF2ome weboldalán	94
19.	Letöltési beállítások az ARN weboldalán	94
20.	A TGF- β és a Notch útvonal a SignaLink 2-ben	98
21.	Szabályozási hurkok az NRF2ome adatbázisban – I.	99
22.	Szabályozási hurkok az NRF2ome adatbázisban – II.	105
23.	Hálózati motívumok az NRF2 transzkripcionális szabályozásában	107
24.	A jelátviteli útvonalak és az autofágia fehérjék kapcsolatai	111
25.	Az autofágia szexuálszteroid receptorok általi szabályozása	114
26.	Rákban mutációt szenvedő gének és gyógyszercélpontok az autofágia szabályozási hálózatában	116

Táblázatok jegyzéke

1.	A SignaLink 2 rétegek és fajok szerinti áttekintése	79
2.	Az NRF2ome rétegek szerinti áttekintése	79
3.	Az ARN rétegek szerinti áttekintése	80
4.	Az autofágia közvetlen kapcsolatai jelátviteli útvonalakkal	81
5.	A SignaLink 2, az NRF2ome és az ARN összehasonlítása	82
6.	Az interakciók tulajdonságok szerinti összehasonlítása	83
7.	A más adatbázisokból átvett adatok áttekintése	84
8.	Az adatbázisok weboldalai	85
9.	A SignaLink 2 összehasonlítása más jelátviteli adatbázisokkal	96

Rövidítések jegyzéke

ABS <i>Annotated Regulatory Binding Sites</i> adatbázis	GO BP <i>Gene Ontology, Biological Process</i> – a GO adatbázis biológiai folyamatokat tartalmazó része
AC <i>Accession number</i>	BRAF <i>Rapidly Accelerated Fibrosarcoma protein B</i>
ADB <i>Autophagy Database</i>	BRCA1 <i>Breast cancer type 1 susceptibility protein</i>
ADDA <i>Automatic Domain Decomposition Algorithm</i>	bZIP Leucin-cipzár transzkripció faktor
ADM <i>Adjacent Dinucleotide Model</i>	CAMKK Kalcium/kalmodulin-függő protein kináz kináz
AGO <i>Argonata protein</i>	CBP CREB-kötő fehérje
AIM ATG8 interakciós motívum	CC <i>Gene Ontology, Cellular Component</i> – a GO adatbázis sejt kompartmentumokat tartalmazó része.
AJAX <i>Asynchronous Javascript and XML</i> – webes technológia	CK Kazein kináz
AMBRA <i>Autophagy/Beclin-1 regulator protein</i>	CLEAR <i>Coordinated Lysosomal Expression and Regulation</i> promóter motívum
AMPK AMP aktiválta protein kináz	COSMIC <i>Catalogue of Somatic Mutations in Cancer</i> adatbázis
AP-MS <i>Affinity Purification and Mass Spectrometry</i> – Affinitás-tisztítást követő tömegspektrometria	COX6C Citokróm-c oxidáz, VIc. alegység
ARE <i>Antioxidant Response Element</i> promóter motívum	CREB1 <i>CAMP responsive element binding protein 1</i>
ARN <i>Autophagy Regulatory Network</i> adatbázis	CUL3 Cullin 3 protein
ATF4 <i>Activating transcription factor 4</i>	DAF-16 <i>Forkhead box protein O</i>
ATG Autofágia gén	DAPK <i>Death-Associated Protein kinase</i>
ATM <i>Ataxia telangiectasia mutated protein</i>	DIMA <i>Domain Interaction Map</i> algoritmus
ATP Adenozin-trifoszfát	DNS Dezoxiribonukleinsav
AXIN1 <i>Axis inhibition protein 1</i>	DOMINE <i>Database of Protein Domain Interactions</i> adatbázis
BACH1 <i>BTB and CNC homology protein 1</i>	DPEA <i>Domain Pair Exclusion Analysis</i> algoritmus
BH3 <i>Bcl-2 homology domain 3</i>	

DPP Dipeptidil-peptidáz 3	GABARAPL <i>Gamma-aminobutyric acid receptor-associated protein like protein</i>
DRAM <i>Damage-regulated autophagy modulator protein</i>	GATA3 <i>Trans-acting T-cell-specific transcription factor GATA-3</i>
DUF <i>Domain of unknown function – ismeretlen funkciójú domén</i>	GEO <i>Gene Expression Omnibus</i> adatbázis
DVL2 <i>Dishevelled segment polarity protein 2</i>	GFP <i>Green fluorescent protein – zöld fluoreszcens fehérje</i>
EDGEdb <i>C. elegans Differential Gene Expression Database</i>	GO <i>Gene Ontology</i> adatbázis
EGF Epidermális növekedési faktor	BioGRID <i>The Biological General Repository for Interaction Datasets</i> adatbázis
ELM <i>Eukaryotic Linear Motifs</i> adatbázis	GTP Guanozin-trifoszfát
ENCODE <i>Encyclopedia of DNA Elements</i> adatbázis	gzip GNU zip veszteségmentes tömörítés
ER Endoplazmatikus retikulum	HADB <i>Human Autophagy Database</i> adatbázis
ERK <i>Extracellular signal-regulated kinases</i>	HGNC <i>HUGO Gene Nomenclature Committee</i> nevezéktan
ESR1 Ösztrogén receptor 1	HITS-CLIP <i>High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation</i>
ESR2 Ösztrogén receptor 2	HMMER Rejtett Markov modell <i>Hidden Markov Model</i> szekvencia illesztő algoritmus
EWSR1 EWS 1 RNS-kötő fehérje	HNF1A <i>Hepatocyte nuclear factor 1 homeobox A</i>
BiFC <i>Bimolecular Fluorescence Complementation</i> módszer	HO-1 Hem-oxigenáz 1
FIP200 <i>RB1-inducible coiled-coil protein 1</i>	HOPS <i>Homotypic fusion and protein sorting complex</i>
FLI1 <i>Friend leukemia integration 1 transcription factor</i>	HPRD <i>Human Protein Reference Database</i> adatbázis
FM Frekvencia Moduláció	HSF-1 <i>Heat shock factor protein 1 – hőshockfehérje 1</i>
FOXO <i>Forkhead box protein O</i>	
FRET <i>Fluorescence Resonance Energy Transfer – Fluoreszcencens Rezonancia Energia Transzfer</i>	
GABARAP <i>Gamma-aminobutyric acid receptor-associated protein</i>	

HTRI <i>Human Transcriptional Regulation Interaction Database</i> adatbázis	MAPK <i>Mitogen-activated protein kinase</i>
ID <i>Identification Number</i> – adatbázis azonosító	MAX <i>Myc-associated factor X</i>
IGF <i>Insulin-like growth factor</i> – inzulin-szerű növekedési faktor	MEF <i>Myelin expression factor</i>
IKK <i>Inhibitor of nuclear factor kappa-B kinase</i>	GO MF <i>Gene Ontology, Molecular Function</i> – a GO adatbázis molekuláris funkciókat tartalmazó része
IL Interleukin	MIAME <i>Minimal Information About Microarray Experiment</i> szabvány
IMID <i>Integrated Molecular Interaction Database</i> adatbázis	MINT <i>Molecular Interaction Database</i> adatbázis
INOH <i>Integrating Network Objects with Hierarchies</i> adatbázis	miRNS Mikro-RNS
JAK Janus kináz	MLE <i>Maleless protein</i>
JNK C-Jun N-terminális kináz	mRNS <i>messenger RNS</i>
JSON <i>Javascript Object Notation</i> adatformátum	MS <i>Mass Spectrometry</i> – tömegspektrometria
JUN AP-1 transzkripció faktor	NBR1 <i>Neighbor of BRCA1 gene 1 protein</i>
KEAP <i>Kelch-like ECH-associated protein 1</i>	NCBI <i>National Center for Biotechnology Information</i> intézet
KEGG <i>Kyoto Encyclopedia of Genes and Genomes</i> adatbázis	NCOR1 <i>Nuclear receptor co-repressor 1</i>
LC3 <i>Microtubule-associated proteins 1A/1B light chain 3</i>	NFκB <i>Nuclear factor kappa-B</i>
LIR LC3 interakciós régió (motívum)	NFAT4 <i>Nuclear factor of activated T-cells 4</i>
LKB1 <i>Liver kinase B1</i>	NFE2 <i>Nuclear Factor, Erythroid-Derived 2</i>
LP Lináris Programozás	NFIL3 <i>Nuclear factor, interleukin 3 regulated</i>
MAF <i>Musculoaponeurotic fibrosarcoma oncogene homolog</i>	NHR <i>Nuclear hormone receptor</i> – nukleáris hormon receptor és útvonal
MAFG MafG transzkripció faktor	NID <i>Non-interacting Domains</i> – nem kölcsönható domének
MAP <i>Microtubule-associated protein</i>	NIP <i>Non-interacting Proteins</i> – nem kölcsönható fehérjék

NIX <i>BCL2/adenovirus E1B 19 kDa protein-interacting protein 3-like protein</i>	PINK1 <i>PTEN-induced putative kinase 1</i>
NLP <i>Natural Language Processing – gépi szövegfeldolgozás</i>	PIP3 <i>Phosphatidylinositol 3-phosphate – foszfatidil-inozitol-3-foszfát</i>
LIR_NRBOX <i>Nuclear receptor box</i>	PKA <i>Protein kináz A</i>
NRF2 <i>Nuclear factor (erythroid-derived 2)-like 2</i>	PKB <i>Protein kináz B</i>
OSX <i>Macintosh Operating System X</i>	PKC <i>Protein kináz C</i>
PAR-CLIP <i>Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation</i>	PPARγ <i>Peroxisome proliferator-activated receptor gamma</i>
PARK2 <i>Parkin RBR E3 ubiquitin protein ligase</i>	PPARG <i>Peroxisome proliferator-activated receptor gamma</i>
PAS <i>Phagophore Assembly Site</i>	PPI <i>Protein-protein Interaction – fehérje-fehérje kölcsönhatás</i>
iPAVS <i>Integrated Pathway Resources, Analysis and Visualization System adatbázis</i>	PRINCESS <i>Protein Interaction Confidence Evaluation System with Multiple Data Source algoritmus</i>
BioPAX <i>Biological Pathways Exchange adatformátum</i>	PSI-MI <i>Proteomics Standards Initiative Molecular Interactions adatformátum</i>
PBM <i>Protein Binding Microarray módszer</i>	PSIMI-TAB <i>Proteomics Standards Initiative Molecular Interactions Tab Delimited adatformátum</i>
PDK <i>Phosphoinositide-dependent kinase-1</i>	PSIMI-XML <i>Proteomics Standards Initiative Molecular Interactions Extensible Markup Language adatformátum</i>
PE <i>Phosphatidylethanolamine – foszfatidil-etanolamin</i>	PTEN <i>Phosphatase and tensin homolog</i>
PERK <i>Eukaryotic translation initiation factor 2-alpha kinase 3</i>	PWM <i>Position Weight Model – pozíció-súly modell</i>
PHP <i>PHP Hypertext Preprocessor programozási nyelv</i>	RCDP <i>Relative Co-evolution of Domain Pairs algoritmus</i>
PID <i>Pathway Interaction Database adatbázis</i>	RDFE <i>Random Decision Forest Framework algoritmus</i>
PIK3C3 <i>Phosphatidylinositol 3-kinase catalytic subunit type 3 – foszfatidil inozitol 3 kináz 3-as típusú katalitikus alegysége</i>	EpRE <i>Electrophile Response Element promóter motívum</i>

RISC <i>RNA induced silencing complex</i>	STAT <i>Signal transducer and activator of transcription protein</i>
RNAi RNS interferencia	STING <i>Stimulator of interferon genes protein</i>
RNS Ribonukleinsav	STRING <i>Search Tool for the Retrieval of Interacting Genes/Proteins adatbázis</i>
ROC <i>Receiver Operating Characteristic statisztikai módszer</i>	SVM <i>Support Vector Machine – tanulóvektor-gép</i>
RORA <i>RAR-Related Orphan Receptor A</i>	SVR <i>Support Vector Regression – tanulóvektor-gép regresszió</i>
ROS <i>Reactive Oxygen Species – reaktív oxigén gyökök</i>	TAK1 <i>TGF-beta activated kinase 1</i>
RTK Receptor tirozin-kináz	TCF4 <i>Transcription factor 4</i>
SBML <i>Systems Biology Markup Language</i>	TF Transzkripció faktor
SELEX <i>Systematic Evolution of Ligands by Exponential Enrichment módszer</i>	TFEB EB transzkripció faktor
SILAC <i>Stable Isotope Labeling by Amino Acids in Cell Culture – stabil izotóp címkézés aminosavakkal, sejtkultúrában</i>	TGF <i>Transforming growth factor</i>
SIRT1 Szirtuin 1	TMYC <i>Protein tag encoding a c-myc epitope</i>
SLiM <i>Short Linear Motif – rövid lineáris motívum</i>	TOR <i>Target of rapamycin</i>
SMURF1 <i>E3 ubiquitin-protein ligase SMURF1</i>	mTORC <i>Mammalian target of rapamycin complex – emlős TOR komplex</i>
SP <i>Specificity protein 1 transzkripció faktor</i>	TSC1 <i>Tuberous sclerosis protein 1</i>
SPIKE <i>Signaling Pathways Integrated Knowledge Engine adatbázis</i>	TSC2 <i>Tuberous sclerosis protein 2</i>
SQL <i>Structured Query Language programozási nyelv</i>	TSS <i>Transcription Start Site – transzkripció kezdőhely</i>
SREBP <i>Sterol regulatory element-binding protein</i>	TWIST1 <i>Twist-related protein 1</i>
SS <i>Semantic Similarity – szemantikus hasonlóság</i>	UCSC <i>University of California, Santa Cruz adatbázis</i>
	UI <i>User Interface – felhasználói felület</i>
	ULK <i>unc-51 like autophagy activating kinase 1</i>

UTR <i>Untranslated Region</i>	WIS <i>Worm Interactome version 8 adatbázis</i>
UV <i>Ultraviola</i>	WIPI <i>WD-repeat protein interacting with phosphoinosides</i>
UVRAG <i>UV radiation resistance-associated gene protein</i>	WNT <i>Wingless and Int related protein útvonal és ligandum</i>
VMP1 <i>Vacuole membrane protein 1</i>	XML <i>Extensible Markup Language adatformátum</i>
VPS <i>Vacuolar protein sorting-associated protein homolog</i>	ZIP <i>PKZip veszteségmentes tömörítés</i>

1. Bevezető

A bioinformatikai eszközök tárháza és a rendelkezésre álló molekuláris biológiai adatok mennyisége gyorsuló ütemben növekszik. Az informatikai eszköztár kiterjedt használatát megköveteli a másképpen kezelhetetlen adatmennyiség, valamint az igény a matematikai módszerekkel történő predikcióra, mely támpontot adhat a további kutatás megtervezéséhez. A nagy számú molekuláris biológiai adatbázis között külön kategóriát képeznek az interakciós adatbázisok, melyek gyakran csak fehérjék, esetenként fehérjék és más molekulák (RNS-ek, lipidek, kis molekulák) közt létrejövő kölcsönhatásokat gyűjtik rendszerbe. Az interakciós adatbázisokat tekinthetjük hálózatoknak, amennyiben minden molekulát egy csomópontnak, és minden kölcsönhatást egy élnek tekintünk egy gráfban. Ez az absztrakció azért is előnyös, mert a gráfelmélet matematikai módszertana segítségével elemezhetjük a biológiai rendszereket.

Dolgozatomban három, általam összeállított interakciós adatbázist mutatok be. Időrendben az elsőként elkészült adatbázis a SignaLink 2, melynek célja hét jelátviteli útvonal kapcsolatainak és szabályozásának feltérképezése. Ezután egy, az oxidatív stresszválaszban és számos betegségben fontos szerepet játszó transzkripciós faktor, az NRF2 szabályozási hálózatáról készítettem egy adatbázist. Harmadikként a rákban és öregedésben, valamint neurodegeneratív betegségekben kulcsszerepet játszó folyamat, az autofágia szabályozási hálózatát gyűjtöttem egy adatbázisba.

A három adatbázist elsősorban kialakításuk alapelvei és módszertana köti össze. Ezen alapelvek közül az egyik legfontosabb a molekuláris szabályozás különböző módjainak – transzkripcionális, poszt-transzkripcionális és poszt-transzlációs szabályozás – egységes rendszerbe foglalása. Ahol mégis eltérés mutatkozik e három adatbázis közt, azt mindenhol külön kiemelem. A másik közös elem, hogy az általam összeállított adatbázisok eredeti irodalmi gyűjtés köré épülnek, mely jó minőségű, megbízható ismereteket jelent a fehérjék közti kölcsönhatásokról. A harmadik összekötő elem a jelátviteli hálózat, mely kapcsolatot teremt a sejt kommunikációs- és információfeldolgozó alrendszerei, valamint a vizsgált folyamatok szabályozása között.

Mindhárom adatbázis hiánypótló szerepet tölt be, mivel nem áll rendelkezésre más adatforrás, ahol ezek az adatok egységes rendszerben elérhetőek lennének – ami előfeltétele az ezeket felhasználó elemző munkának. Hiba lenne azonban ezeket az adatbázisokat más adatbázisok összeömlésztésének tekinteni, az adatintegráló munka csak egy része céljaink megvalósításának. Egyfelől, mindegyik adatbázis saját irodalom feldolgozáson alapul, azaz a szövegekben fellelhető információkat kutatócsoportunk tagjai összegyűjtötték, és mint fehérje-fehérje kölcsönhatásokat, hozzáadták az adatbázishoz. Így tehát több ezer cikkből – mely hasonló mennyiségű kísérletet jelent – nyertünk egy fehérje interakciós hálózatot. Másrészt, a további adatok más adatbázisokból való átvételét is gondosan mérlegeltük, és ezeket az adatokat, a bioinformatikai szabványok közti eltéréseket áthidalva, egységes formátumra hoztuk. Továbbá, adatbázisaink tartalmazzak predikciókat és megbízhatósági mérőszámokat (*confidence score*) is, melyek kiszámítását részben saját magunk végeztük, részben pedig más forrásokból átvettük.

Dolgozatomban csak olyan vizsgálatok és eredmények szerepelnek, melyek elkészítésében részt vettem. Az általam kivitelezett részfeladatoknál egyes szám első személyt, míg a kutatócsoport többi tagjával együttműködésben végzeteknél többes számot használok.

A doktori munkám háttérét alkotó ismeretek bemutatása során elsőként a jelátviteli hálózatok általános jellemzőit tárgyalom. Ezután bemutatom a poszt-transzlációs, a transzkripcionális, majd a poszt-transzkripcionális szabályozás megismerését szolgáló leggyakoribb nagy teljesítményű, ún. *high-throughput* kísérletes módszereket. Ezt követően áttekintem az ezekből a módszerekből származó adatokat összegyűjtő bioinformatikai adatbázisokat, és kitérek az *in silico* predikciós módszerek jellemzőire is. A bevezető rész utolsó két fejezetében az NRF2 transzkripciós faktor és az autofágia szabályozásáról rendelkezésre álló ismereteinket foglalom össze.

1.1. Jelátviteli útvonalak

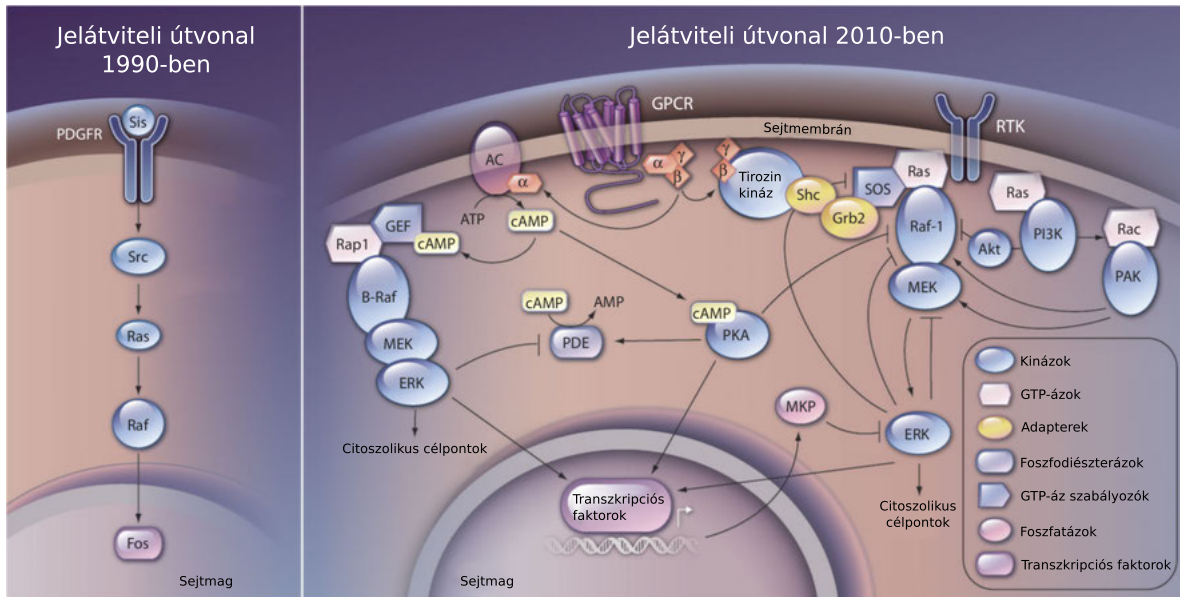
A jelátvitel a sejt információ feldolgozó alrendszere, melynek biológiai funkciója a kívülről és belülről érkező információk sokaságának észlelése, feldolgozása, és specifikus válaszok generálása a sejt működésének befolyásolása által. A hagyományos elképzelés szerint, a jelátvitel útvonalak szerint szerveződik, azaz a jelet egymásnak továbbító fehérjék – leggyakrabban kinázok vagy proteázok – lineáris, kaszkád szerű sorozataiból áll (Kholodenko és mtsai 2012). Ezen kaszkádok kezdőpontja egy többnyire sejtmembránban található receptor, melyek működése a megfelelő ligand kötésétől függ; végpontjuk pedig a transzkripciós faktorok hálózata, melyek aktivitási mintázata a jelátvitel függvénye. Szélsőséges esetet képvisel az NHR (*nukleáris hormon receptor*) útvonal, melyben a receptorok egyben transzkripciós faktorok is (például az ösztrogén receptor).

A legtöbb egyedfejlődési folyamatban evolúciósan konzervált módon, mindössze hét jelátviteli útvonal vesz részt: a Hedgehog, a receptor-tirozinkináz (RTK), a Wingless (WNT), a TGF- β , a Notch, a Janus-kináz (JAK/STAT) és az NHR útvonal (Pires-daSilva és mtsai 2003). A flexibilitás és a robosztusság a jelátviteli útvonalak két sajátossága, mely szükséges a sokféle funkció megvalósításához. A receptorok expressziója sejttípusra jellemző, a ligandumok jelenléte pedig az aktuális fiziológiai állapotot tükrözi. A különböző útvonalakon érkező jelek az útvonalak közti keresztbeszélgetések révén kombinálódnak, a jelátviteli hálózatban információ feldolgozás zajlik. Ebben a folyamatban a sejten belüli kompartmentalizáció, a jelátviteli molekulák lokalizációja is fontos szerepet játszik. Végül a szövetspecifikusan jelen lévő eltérő transzkripciós faktorok és ezek egymás közti interakciói is befolyásolják a sejt külvilág ingereire adott válaszát. Az itt felsorolt mechanizmusok eredményezik, hogy a hét alapvető útvonal képes nagy számú egyedfejlődési és egyéb funkció megvalósítására, flexibilis jelátviteli hálózatot alkotva (Pires-daSilva és mtsai 2003).

1.2. Jelátviteli rendszerek és hálózatok

A lineáris, kaszkád szerű jelátviteli útvonalak izolált vizsgálata nem ad magyarázatot a komplex biológiai folyamatok szabályozásának mikéntjére (1. és 2. ábra). Hogy megértsük a

molekuláris interakciók – külső ingerektől befolyásolt – dinamikája és a sejt fenotípusa közti összefüggést, szükséges a jelátviteli fehérjék kölcsönhatásainak hálózatos reprezentációjának elkészítése, és annak matematikai módszerekkel történő vizsgálata (Bauer-Mehren és mtsai 2009; Kholodenko és mtsai 2012; Taylor és mtsai 2012).



1. ábra. **Szemléletváltás a jelátvitel kutatásában – I.** A lineáris, kaszkád szerű, önálló útvonalak kezdeti modellje helyett, ma bonyolult jelátviteli hálózat képe él bennünk. Az új adatoknak köszönhetően világossá vált, hogy a jelátviteli komponensek igen nagy számú fehérje és mikro-RNS alkotta bonyolult szabályozási hálózatba ágyazódnak. – Az ábra forrása: Kholodenko és mtsai 2012.

Az egy útvonallal kapcsolatban álló komponensek száma növekszik: egy ERK aktivitást RTK stimuláció hatása alatt mérő szűrés 331 új szabályozóját derítette fel ennek az útvonalnak (2. ábra, Friedman és mtsai 2007). Sőt mi több, ezen szabályozók száma nem lehatárolható: RNS interferencia szűrések (*RNAi screening*) adataiból kiderült, hogy az útvonalat szinte nem befolyásoló gének, és jól ismert, legerősebb szabályozói között folytonos az átmenet, utóbbiak nem alkotnak jól elhatárolható csoportot (Friedman és mtsai 2007). Az összetett interakciós hálózatokról a *high-throughput* kísérletekből nyert információ pillanatképet adhat egy jelenségről, azonban elfedi az okokat: egy ilyen bonyolult rendszert akár egy ponton ért hatás számtalan úton terjedhet tovább, és számtalan ponton okozhat változást, azonban a

közreműködő mechanizmusok az egyes pillanatképekből nem láthatók (Kholodenko és mtsai 2012).

Az interakciós hálózat tér- és időbeli dinamikája az expresszió, a poszt-transzlációs módosítások és a dinamikus szubcelluláris lokalizáció hatása révén alakul ki (Taylor és mtsai 2012). A hálózatokon alkalmazott mutációs és expressziós elemzések megmutatják, hogy a hálózat szerkezete, konnektivitása és modularitása, miként változik egyes betegségekben, így rámutathatnak a betegségben kulcsszerepet játszó fehérjékre. A sejtekben, aktuális fiziológiai állapotuktól függően, a rájuk jellemző fehérje interakciós hálózatnak egy része valósul meg. Az élesztő sejtciklusa során a dinamikusan változó mennyiségű fehérjék egymással, és állandóan jelenlévő partnereikkel lépnek interakcióba, így az interakciós hálózat a sejtciklus során folyamatosan módosul (Lichtenberg és mtsai 2005).

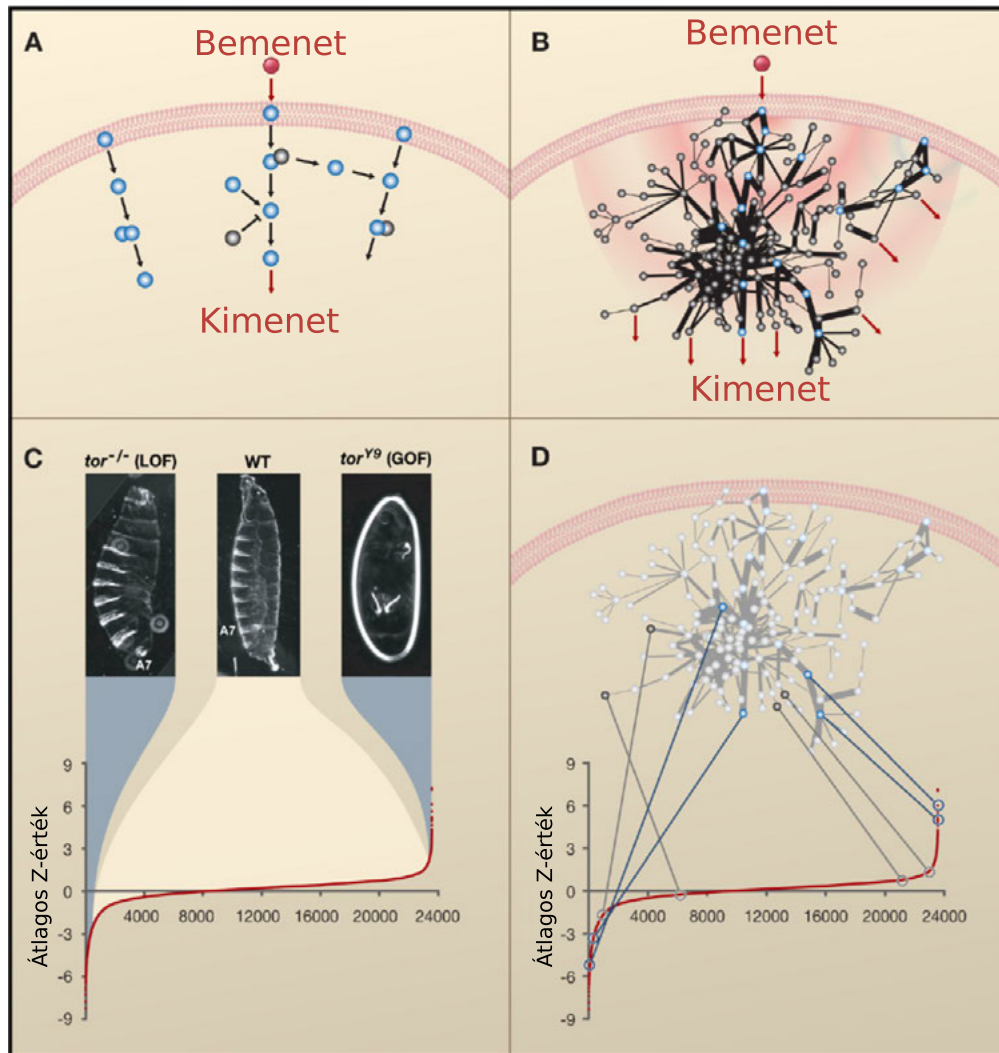
A legtöbb fehérje interakciós hálózat skálafüggetlen, bár ennek precíz megvalósulásáról ismereteink még hiányosak. Az interakciós hálózatokban azonosíthatóak ún. *hub* fehérjék, melyek sok másik fehérjével vannak kapcsolatban – gráfelméleti kifejezéssel, nagy fokszám jellemzi őket. Ezen fehérjék egy része szövetspecifikusan (ezek általában intermoduláris, azaz modulok közti *hubok*) más részük általánosan (ezek általában intramoduláris, azaz modulokon belüli *hubok*) koexpresszálódnak partnereivel. A jelátviteli domének gyakoribbak az intermoduláris *hubok*ban, és a daganatok szempontjából is jelentősebbek ezek a fehérjék. Egyes tanulmányok szerint, a hálózatban nagyobb fokszámú fehérjék rákos sejtekben gyakran nagyobb mennyiségben expresszálódnak (Taylor és mtsai 2009).

A jelátviteli útvonalak robusztusságát a negatív és pozitív visszacsatolások, valamint a redundanciák biztosítják. A pozitív visszacsatolás erősíti az útvonal aktivitását, hozzájárul az elköteleződéshez a differenciáció során. A negatív visszacsatolások korlátozhatják egy útvonal aktivitását, vagy oszcillációkat idézhetnek elő. Ezek a jelentős jelátviteli útvonalak minden többsejtűben megtalálhatók, azonban a paralógok számában különböznek a taxonómiai csoportok. Az evolúció során számos gén duplikálódott, és a homológok új funkciókat nyertek, vagy a korábban egy gén által gyakorolt funkciókat megosztva, komplementer módon vitték tovább. A redundanciák biztosíthatják a létfontosságú funkciók fenntartását a redundáns elemek egy részének kiesése esetén (Pires-daSilva és mtsai 2003).

A szignál integrációban fontos szerepet játszanak a jelátviteli fehérjék, valamint a transzkripciós faktorok aktivitásának időbeli dinamikájában megfigyelhető jellegzetesség, a frekvencia modulált (FM) pulzálás. Levine és mtsai fluoreszcens jelöléssel vizsgálták a p53 és az NF- κ B transzkripciós faktorok aktivitásának időbeli dinamikáját. A rövidebb és hosszabb periódusú pulzálások interakciója segít a megfelelő funkcionális válasz kialakításában. Egyes esetekben AC-DC konverzió is megfigyelhető a jelátvitelben, amikor egy fehérje koncentrációja egy másik pulzálásának gyakoriságát határozza meg. A pulzálás fenntartása negatív és pozitív visszacsatolások együttműködése, vagy késleltetett negatív visszacsatolás révén valósulhat meg: például a p53 periodikusan bekövetkező, ATM-ből kiinduló aktiválása során saját degradációját is kiváltja az Mdm2 transzkripciója révén, másrészt az ATM-et is inaktíválja, a Wip1 segítségével. Amennyiben a DNS károsodás továbbra is fennáll, újabb aktivációs hullám indul meg az ATM irányából (Levine és mtsai 2013).

Az NF- κ B nukleáris lokalizációja periodikusan, mindent-vagy-semmit alapon következik be. Egy gyors negatív visszacsatolás az I κ B transzkripciója révén inaktíválja az NF- κ B-t. Egy pozitív, de lassabb hurok is megfigyelhető, mely az A20 fehérje transzkripciója által, az IKK kináz inaktíválásával serkenti az NF- κ B aktivitását. A pulzáló aktivitások a jelátviteli utakban képesek idő alapú kombinatorikus szabályozás megvalósítására. Például az extracelluláris kalcium szint hatására az NFAT4 transzkripciós faktor percnél rövidebb impulzusok formájában aktiválódik, míg az NFAT1 lassan, nem pulzáló módon. Ily módon más információt közvetít mindkét TF jelenléte, mint az NFAT4 egyedül. Az azonos promóterhez kötődő több transzkripciós faktor aktivitási periódusai közti átfedések, azaz a pulzusok szinkronitásának mértéke a szignál integráció idő alapú megvalósítását jelenti. Továbbá, bizonyos sejtek egymással összeférhetetlen funkciókat időbeni elválasztással valósítanak meg. Ilyen esetekben a szabályozási rendszer bistabil, vagy az egyik, vagy a másik állapot felé köteleződik el (Levine és mtsai 2013).

Az időbeli dinamika a szabályozás különféle módjainak együttes vizsgálatánál is fontos. A jelátvitel kémiai reakciói (pl. foszforiláció) nagyságrendekkel rövidebb idő alatt mennek végbe, mint a transzkripcionális szabályozás, vagy a sejt szintű folyamatok (Papin és mtsai 2004).



2. ábra. **Szemléletváltás a jelátvitel kutatásában – II.** A jelátviteli útvonalak leggyakoribb reprezentációi viszonylag kevés komponensből álló, izolált, kvantitatív információkat nélkülöző kaszkádként mutatják be azokat (A). Egy valósághűbb képet mutat a bonyolult jelátviteli hálózat, melyben az egyes kapcsolatok relatív erőssége eltérő, és amely több irányból kap bemenő jeleket, majd az információ-feldolgozást követően több hatást érvényesít (B). Az egyedfejlődésre ható mutációk relatív hatása (az ábrán ezt fejezi ki a *z-érték*) az összes többi gén meghibásodásának hatásához hasonlítva a skála extrém értékeit képviselik; a legtöbb gén funkcióvesztésének hatását kivédi a rendszer robosztussága, és a vad fenotípus jelenik meg (C). Genom léptékű szűrések mutatják, hogy az egyes gének hatása a jelátviteli hálózat egy adott pontján mért aktivitásra folytonos eloszlás szerint alakul. Ez azt jelenti, hogy nem választhatók szét egy adott útvonalban számottevő szerepet játszó, illetve az azon kívül álló, azt nem befolyásoló gének. Ez a felismerés a sejtek molekuláris szabályozási hálózatának egységességét sugallja (D). – Az ábra forrása: Friedman és mtsai 2007.

1.3. A poszt-transzlációs kapcsolatok vizsgálata nagy teljesítményű módszerekkel

Az utóbbi két évtizedben megsokszorozódott a fehérjék közti interakciókról rendelkezésre álló adatok mennyisége. A molekuláris biológiai módszerek általános fejlődése mellett, ez a nagyságrendi növekedés elsősorban néhány nagy teljesítményű, ún. *high-throughput* módszernek köszönhető. Ezek a módszerek alkalmasak több száz, vagy több ezer fehérje közti interakciók gyors, automatizált feltérképezésére. Ebben a fejezetben bemutatom a legfontosabb *high-throughput* módszereket, melyekből a 1.6.1. fejezetben ismertetett bioinformatikai adatforrások az adataikat nyerik.

Az irodalmi gyűjtéssel állítható elő a legmegbízhatóbb adatsor a fehérjék közti kölcsönhatásokról. A kutatási érdeklődés fókuszában álló fehérjékről és molekuláris folyamatokról aránytalanul sok hivatkozás található, így a kísérleteket bemutató cikkekből gyűjtött adatok torzítanak a leginkább kutatott útvonalak irányába. További problémát jelent, hogy gyakran hiányoznak a kísérletes bizonyítékok az interakciók hiányáról, mivel a negatív eredményeket nehezebb publikálni (Dirnagl és mtsai 2010). Ez utóbbi hiányosság kiküszöbölését a Negatome adatbázis tűzte ki céljául (Smialowski és mtsai 2009). A nagy teljesítményű, ún. *high-throughput* módszerek nagy számú interakcióról szolgáltatnak információt, beleértve a negatív adatokat is, azonban ezen az adatok mögött gyengébb kísérletes bizonyíték áll, és nem adnak információt a kapcsolatok biológiai funkciójáról. A *high-throughput* módszerek egy része – mint az élesztő két hibrid módszer – bináris fehérje-fehérje interakciók feltérképezésére, míg másik részük – például az affinitás-tisztítás és tömegspektrometria (AP-MS) – fehérje komplexek azonosítására is alkalmas (Parrish és mtsai 2006).

Az élesztő két hibrid technikát a 90-es évek elején fejlesztették ki, és már tíz évvel ezelőtt több ezer fehérje-fehérje interakciót tartalmazó hálózatok álltak rendelkezésre. Például egyetlen tanulmány a *Drosophila* 4.679 fehérjéje között 4.780 kapcsolatot tárt fel 2003-ban (Giot és mtsai 2003). 2005-ben hasonló vizsgálat 1.549 emberi fehérje közt 2.754 interakciót (Rual és mtsai 2005), Stelzl és mtsai 2005 pedig 3.186 humán interakciót azonosított.

Az élesztő két hibrid módszer sok hamis negatív eredményt ad, egyrészt a klón könyvtárak hiányossága miatt, másrészt a technika alkalmatlansága miatt bizonyos típusú interakciók kimutatására. A hamis pozitívok azonban még nagyobb problémát okoznak, hiszen a nagy fehérje interakciós adatbázisokban az adatok jelentős hányada csak egy módszerrel, esetenként csak egyetlen élesztő két hibrid kísérlettel lett kimutatva. A megbízhatóságot növelheti, ha megköveteljük, hogy legalább két különböző kísérlet igazolja egy kapcsolat létét; vagy akár bioinformatikai módszerekkel, megbízhatósági mérőszámokat alakíthatunk ki, melyek a kapcsolat valószínűségét az adott fehérjék tulajdonságai alapján becslik (Parrish és mtsai 2006). A HitPredict például az összes nagy protein-protein interakciós adatbázisból átvett kapcsolatokra ad egy komplex megbízhatósági becslést, mely a fehérjék domén összetételén, GO tulajdonságain, és ortológjaiknak más fajokban azonosított interakcióin alapul (Patil és mtsai 2011). Hasonló megközelítést valósít meg a PRINCESS mérőszám, melyet felhasználtam a munkám során, és melyről bővebben a 1.6.2.1. fejezetben írok.

Fehérjekomplexek tömeges feltérképezésére a másik leggyakrabban alkalmazott módszer az affinitás-tisztítást követő tömegspektrometria (*[Tandem] Affinity Purification & Mass Spectrometry – [T]AP-MS*). 2007-ben egy tanulmány tömegspektrometriával 2.235 emberi fehérje 6.463 interakcióját azonosította (Ewing és mtsai 2007).

Ezen módszerek egyik hátránya, hogy membránfehérjék esetében csak módosításokkal működnek; bár az élesztő két hibrid módszer egyes változatai (Snider és mtsai 2010) és az AP-MS (Gavin és mtsai 2006; Wu és mtsai 2003) képesek membránfehérjék vizsgálatára. Emellett az élesztő két hibrid módszer csak olyan fehérjék interakcióit képes kimutatni, melyek működőképesek a többsejtű állatokra jellemző poszt-transzlációs módosítások nélkül is, továbbá nagyobb eséllyel mutatja ki a sejtmagi lokalizációjú fehérjéket (Venkatesan és mtsai 2008).

A tranziens fehérje interakciók azonosítására alkalmas a FRET (*Fluorescence resonance energy transfer*), a BiFC (*Bimolecular fluorescence complementation*), az élesztő két hibrid módszer, illetve a kémiai keresztkapcsolást követő AP-MS (Acuner Ozbabacan és mtsai 2011).

1.4. A transzkripcionális szabályozás és vizsgálati módszerei

A transzkripciós faktorok (TF-ek) a jelátviteli hálózat végpontján álló effektorok, jelenlétük és aktivitásuk mintázata meghatározza a sejt anyagcseréjét, funkcióját és differenciálódását. A WNT, a Notch, a Hedgehog és az NHR útvonalak esetében a transzkripciós faktorok gátló funkciót töltenek be amikor az útvonal inaktív, míg aktív jelátvitel esetén stimulálják a célgének expresszióját. Ezzel szemben az RTK és a TGF- β útvonalak külön gátló és serkentő transzkripciós faktorokat szabályoznak (Pires-daSilva és mtsai 2003).

*C. elegans*ban 940, emberben 1.500-1.800 transzkripciós faktor ismert (Qu és mtsai 2013; Vaquerizas és mtsai 2009; Walhout 2011). A TF-gén szabályozási kapcsolatok feltérképezésére leggyakrabban a ChIP-Seq módszert alkalmazzák. Ez a technika lehetővé teszi egy TF által *in vivo* kötött DNS szekvenciák megállapítását, melyek aztán megkereshetők a genomban. Egyelőre távolról sem vagyunk a teljes transzkripciós faktor kód birtokában, mivel ennek leolvasásához kivitelezhetetlenül sok ChIP-Seq mérés lenne szükséges: minden sejttípus minden fejlődési stádiumában és minden fiziológiai kontextusában az összes transzkripciós faktorra el kellene végezni egy mérést (Jolma és mtsai 2013; Walhout 2011). Ráadásul, a ChIP-Seq adatokból nem derül ki, hogy az adott TF serkentő vagy gátló hatású-e. Egy génre ható TF-ek felderítésére alkalmas az élesztő egy-hibrid technika. Ennek során a vizsgált promóter mögé riporter gént ültetnek, és az egyes transzkripciós faktorokhoz erős transzkripció aktiváló domént csatolnak. Amennyiben a TF kötődik a vizsgált szekvenciához, a riporter gén expresszálódik (Ouwerkerk és mtsai 2001).

Az ENCODE projekt keretében 2012 szeptemberéig 119 humán TF-fel végeztek ChIP-Seq kísérleteket több sejtvonalon, azonosították a DNS-kötő doménjeiket, és számos TF-TF kapcsolatot is felderítettek (Qu és mtsai 2013). A ChIP-Seq alternatívája a SELEX módszer, mely véletlen oligonukleotid könyvtárból indulva, szelekciós-amplifikációs ciklusok és változó erősségű elúció alkalmazásával kvantitatív adatot ad az egyes szekvenciákhoz való kötődés erősségéről. Egy SELEX vizsgálat keretében 830 humán TF kötési profilját mérték meg (Jolma és mtsai 2013). A ChIP-Seq és SELEX adatok kiegészítésére jól használható a DNase-Seq módszer (Boyle és mtsai 2010). További alternatíva a fehérje kötő *microarray* (*Protein Binding*

Microarray – PBM) (Berger és mtsai 2006), melynek hátránya, hogy hosszabb kötőhelyek esetén rossz eredményt ad (Jolma és mtsai 2013).

Csak a TF-ek egy része szekvencia specifikus, azonban a szabályozás szempontjából ezek a legfontosabbak. A legtöbb kötőhely 10-20 bázispár hosszúságú. Általában a szekvenciát vagy a DNS térszerkezetet ismeri fel a TF, és gyakran hidrogénkötést létesít a bázisokkal, vagy egy protein hélix a DNS nagy árkába illeszkedik (Jolma és mtsai 2013). A ChIP-Seq és a SELEX kísérletek eredményeiből létrehozható a TF-re jellemző pozíció-súly mátrix (*Position Weight Matrix – PWM*), mely a kötési hely minden pozíciójában megadja, hogy mely nukleotid milyen arányban fordul elő, és ennek megfelelően, mennyire kedvezően befolyásolja az adott TF kötődését. Egy precízebb modell az ADM (*Adjacent Dinucleotide Model*), mely a nukleotidok szomszédságát is figyelembe veszi. A TF-ek döntő többségénél kizárólag a DNS-kötő domén határozza meg a kötés minőségét, ezek a domének önmagukban is hasonló módon reagálnak a DNS-sel (Jolma és mtsai 2013). Egyes TF-ek azonban nem a DNS-hez kötődnek, hanem más TF-ekhez (Qu és mtsai 2013). A TF-ek működése során nem csak a kötőhely szekvenciája számít, hanem a távolsága a start kodontól, és az iránya is – ez főleg a dimerizálódó TF-eknél fontos. Az ilyen TF-ek célgénjeinek predikciójára alkalmas az irány-távolság modell (Jolma és mtsai 2013).

A fehérje és mikro-RNS kódoló gének transzkripcionális szabályozásáról számos adatbázis gyűjt adatokat: többek között az ABS, a DroiDB, az edgeDB, az ENCODE, a HTRI, az ORegAnno, a PAZAR, a PuTmiR, a RedFly, a TransmiR és a wTF. Ezek többnyire az említett *high-throughput* módszerekből, a JASPAR esetében pedig bioinformatikai predikcióból származnak (Portales-Casamar és mtsai 2010). Az általam felhasznált adatforrások rövid leírása olvasható az 1.6.4. fejezetben, az említett adatbázisok hivatkozásait is ott helyeztem el.

1.5. Poszt-transzkripcionális szabályozás

A mikro-RNS-ek (miRNS-ek) ~22 nukleotid hosszúságú, nem fehérje kódoló RNS-ek, melyek képesek mRNS-ek UTR (*untranslated region*) szakaszához kötődni, és így általában ezek translációját megakadályozni, vagy az mRNS lebontását előidézni (Grimson és mtsai 2007). Előbbi hatásukat állatokban, utóbbit növényekben tartották jellemzőnek, azonban ma már iga-

zolt, hogy állatokban is hozzájárulhatnak az mRNS degradációhoz (Huntzinger és mtsai 2011). A miRNS-ek – az RNS kötő fehérjékkel együtt – a gének kifejeződését poszt-transzkripcionálisan szabályozzák. A miRNS-ek transzkripciója transzkripciós faktorok szabályozása alatt áll, melyek aktivitását a jelátviteli hálózat szabályozza; ezért tekinthetjük úgy, hogy a miRNS-ek is részei az útvonalak közötti keresztbeszélgetések és visszacsatolások rendszerének.

Noha általános nézet, hogy a miRNS-ek gátolják a velük kölcsönhatásba kerülő mRNS-ek translációját, érdemes megemlíteni, hogy néhány esetben translációt aktiváló hatást is kimutattak. Ráadásul a hatás iránya kontextusfüggően változhat (Vasudevan és mtsai 2007). Mivel a kísérletek elenyésző hányadában figyeltek meg aktivációt, nem követünk el nagy hibát, ha – ameddig ismereteink e kérdéskört illetően nem bővülnek – minden miRNS-t gátló hatásúnak tekintünk.

A legtöbb miRNS–mRNS kapcsolatot a vizsgált miRNS túlexpresszállását követően végzett *microarray* és proteomikai mérésekkel derítették fel. Ez a módszer azonban indirekt volta miatt kétségeket ébreszthet (Thomson és mtsai 2011). A miRNS–mRNS kötődés közvetlen detektálására a RISC (*RNA induced silencing complex*) fehérjéi, például az argonauta fehérjék immun-precipitációja, majd az általuk kötött RNS-ek *microarray*-jel vagy szekvenálással történő azonosítása ad lehetőséget. Ezt nevezik AGO-IP technikának. Egy másik módszer, a HITS-CLIP során az RNS-ek között létrehozott kémiai keresztkötést követően izolálják, majd szekvenálják az interakcióban részt vevő RNS párokat. Ennek a technikának egy hatékonyabb változata a PAR-CLIP, mely *in vivo* fotoreaktív nukleotidokat (4-tiouridin) épít be az RNS-ekbe, majd UV fényvel idézi elő a keresztkötések kialakulását (Thomson és mtsai 2011).

A kísérletes adatok alacsony lefedettsége miatt számos *in silico* módszer született a miRNS–mRNS interakciók szekvenciák alapján történő predikciójára. Ezek a módszerek mind a miRNS kötőhelyek kísérletek alapján megismert általános jellemzőit használják fel. Ezeket a szabályokat az alábbiak szerint foglalhatjuk össze:

- a miRNS-ek mag (*seed*) régiója – a 2–7. nukleotidig terjedő szakasz – az általuk szabályozott gének mRNS-einek 3' UTR régióján belüli kötőhelyhez illeszkedik (Kozomara és mtsai 2010);

- ez a régió általában evolúciósan jobban konzervált, mint az mRNS-ek UTR régiójának többi része (Lewis és mtsai 2005);
- gyakori, de nem szükségszerű a 8. nukleotid illeszkedése (Lewis és mtsai 2005);
- az első nukleotid leggyakrabban adenin (Lewis és mtsai 2005);
- gyakori, hogy egy miRNS-nek több kötőhelye van egy 3' UTR régió belül;
- a több kötőhely transláció gátló hatása összeadódik, ha pedig 8–40 nukleotid közelségben találhatók, akkor erősítik egymás hatását (Grimson és mtsai 2007);
- a miRNS-ek ~2%-a a mag régióban nem illeszkedik tökéletesen, mégis gátolja a translációt (Grimson és mtsai 2007);
- ilyen esetekben a miRNS 3' felén, a 13–16. nukleotid jó illeszkedést mutat (Grimson és mtsai 2007);
- a miRNS kötőhelyek környezete 30 nukleotid távolságon belül AU gazdag szekvencia (Grimson és mtsai 2007);
- leginkább a 3' UTR régió két végéhez közel eső kötőhelyek bírnak hatással (Grimson és mtsai 2007);
- a stop kodonhoz 15 bázispárnál közelebb eső kötőhelyek gyengén funkcionálnak (Friedman és mtsai 2009; Grimson és mtsai 2007);
- a hibridizálódott miRNS-mRNS szabadenergiája befolyásolja a kötődés hatékonyságát (Grimson és mtsai 2007).

A mag régió illeszkedésén alapuló kötőhelyeket konvencionálisnak, az egyéb pontokon történő illeszkedést nem konvencionálisnak, a miRNS–mRNS hibridizáció szakaszán kívüli szempontokat pedig kontextus jellemzőknek nevezik. Az egyes predikciós módszerek annyiban különböznek, hogy a fenti szabályok közül melyeket és milyen módon építik be modelljeikbe. Az általam felhasznált predikciók sajátosságait a 1.6.5. fejezetben ismertetem. A kísérletek

alapján egyre diverzebb kép rajzolódik ki a miRNS-ek transzlációt befolyásoló működési mechanizmusait illetően. Például egyes kísérletesen igazolt miRNS–mRNS interakciók esetében a miRNS középső, 11-12. bázispárja kötődik az mRNS-hez (Thomson és mtsai 2011). A predikációs algoritmusok egyike sem képes az összes tényezőt számításba venni, de a tipikusnak tekinthető esetekben jól működnek.

1.6. Bioinformatikai adatforrások

1.6.1. Általános fehérje interakciós adatbázisok

A humán interaktóm becslések szerint 100-250 ezer kapcsolatot tartalmazhat (Tran és mtsai 2013; Venkatesan és mtsai 2008). Az általános fehérje-fehérje interakciós adatbázisok irodalmi gyűjtésből, és *high-throughput* kísérletek eredményeiből gyűjtik össze a fehérjék közti kölcsönhatásokat. Ezen adatbázisok tartalma részben átfedő, számos más adatbázis pedig átveszi, és egyesítve kezeli őket. Az IntAct és a MINT az utóbbi egy évben egyesült, és **MIIntAct** néven 439.013 interakciót tartalmazó adatbázissá vált, a benne szereplő taxonok a teljes filogenetikai fát lefedik. Az adatokat 12.397 publikációban leírt 32.135 kísérletről gyűjtötték össze (Orchard és mtsai 2014). A **BioGRID** (Chatr-aryamontri és mtsai 2013), a **HPRD** (Keshava Prasad és mtsai 2009) és az **InnateDB** (Breuer és mtsai 2013) a MIIntActhoz hasonló adatbázisok, melyeket adatforrásként felhasználtam a általam készített adatbázisokban. Ezekről a 3.2.3. fejezetben részletesebben írok. A STRING adatbázis még ennél is nagyobb méretű, 1.100 faj 5 millió fehérjéjének több mint 200 millió interakcióját tartalmazza. A kísérletes adatok mellett, a STRING automatikus számítógépes szövegfeldolgozási algoritmusokat is használ, melyek 2 millió cikk szövegében, és a MedLine absztrakt gyűjteményben keresnek fehérje interakciókat. A STRING ortológ fehérjék közt meglévő interakciókat más fajokban is meglévőnek tekinti, függetlenül attól, hogy az adott fajban ez kísérletesen bizonyítást nyert-e. Az interakciók különféle forrásai elkülönítve kezelhetők, így a kevésbé megbízható adatokat ki lehet zárni a vizsgálatokból (Franceschini és mtsai 2014).

Ezek az általános interakciós adatbázisok egységesek, nem kategorizálják funkcionális alapon az interakciókat, noha számos ezekre az adatokra építkező útvonal adatbázis ezt pótolja. Az interakciók iránya nem ismert, ezek szimmetrikus viszonyt jelentenek a fehérjék között.

További problémát jelent, hogy módszertani okokból a membránproteinek kapcsolatai ezekben az adatbázisokban kisebb arányban szerepelnek.

1.6.2. A fehérje-fehérje interakciók megbízhatóságának becslése

Amint az 1.3. fejezetben láttuk, a *high-throughput* kísérletek esetében jelentős probléma a hamis pozitív és hamis negatív eredmények számottevő aránya. Ezek kiküszöbölésére az egyik út az interakciók valószínűségének bioinformatikai módszerekkel történő becslése. Az alábbiakban két ilyen megközelítést mutatok be, melyeket adatbázisaink kialakítása során is alkalmaztam.

1.6.2.1. PRINCESS. A **PRINCESS** (*protein interaction confidence evaluation system with multiple data sources*) egy fehérje-fehérje kapcsolat több tulajdonságát felhasználó, komplex mérőszám, mely a nagy teljesítményű kísérleti eljárások alkalmazása kapcsán felmerült igényekre próbál választ adni. A *high-throughput* technikák adatai közt előfordulnak hamis pozitívok, módszertani okokból adódóan – még ha két fehérje valóban kötődik is egymáshoz, nem biztos, hogy az adott interakciónak van biológiai relevanciája, és *in vivo* megvalósul (Parrish és mtsai 2006).

A PRINCESS Bayes-i következtetést alkalmazva, ismert fehérje interakciók hálózatát, domén-domén kapcsolatokat, funkcionális annotációkat, expressziós adatokat, genomikus kontextust, és hálózattopológiai adatokat használ a fehérjék közti interakció valószínűségének becsléséhez. Ezen adatok közül a legtöbbet már mások is használták hasonló célra, a PRINCESS különlegessége azonban, hogy egyszerre alkalmazza mindegyiket. Ennek érdekében egy naiv Bayes-i osztályozó algoritmus segítségével illesztettek modellt, mely egy minden bemeneti tulajdonságot függetlenül kezelő, Bayes-i valószínűségeen alapuló tanuló algoritmus (Li és mtsai 2008).

1.6.2.2. A Gene Ontology tulajdonságok szemantikus hasonlósága. A Gene Ontology (GO) génekhez és géntermékekhez tulajdonságokat rendel hozzá, és ezen tulajdonságokat hierarchikusan kategorizálja. A GO tulajdonságok (*GO terms*) egy fát alkotnak, melyben a

felsőbb szinteken álló tulajdonságokat (szülőket, [Sz]) a leszármazottaikkal [L] összekötő élek jelentése három féle lehet: „az *L* egy *Sz*” (is-a, például „a *mikroszóma* egy *vezikuláris frakció*), az *L* az *Sz* része” (part-of, például „a *replikációs villa* a *kromoszóma* része” illetve „*L* szabályozza *Sz-t*” (regulates, például „a *sejtciklus ellenőrzési pontja* szabályozza a *sejtciklust*) (Smith és mtsai 2005). Az egész GO három ún. doménre osztható, tulajdonképpen három különálló ontológiából áll: sejt összetevő (cellular component – CC), biológiai folyamat (biological process – BP) és molekuláris funkció (molecular function – MF). Egy tulajdonságnak több szülője is lehet, a három domén három fát alkot, melyeknek egy gyökere van (Ashburner és mtsai 2000). A GO-ban szereplő annotációk 98 %-a automatikus módszerekkel készült, csupán 2 % származik manuális gyűjtésből. Minden annotációnál szerepel a forrás publikáció, illetve a módszer, amelyből az adat származik.

A BP és a CC domének esetében logikus a feltevés, miszerint azok a fehérjék, melyek kapcsolatban vannak egymással, hasonló annotációkkal rendelkeznek: hasonló biológiai folyamatokban vesznek részt, és a sejt hasonló komponenseiben fordulnak elő. Így az, hogy két fehérje GO tulajdonságai mennyire hasonlóak, jó támpontot adhat az interakciójuk valószínűségének megítélésében.

Nem elég azonban az azonos és különböző GO tulajdonságok arányát venni, ugyanis két különböző GO tulajdonság esetében is kvantitatív módon kifejezhető, hogy azok jelentése mennyire hasonló. Minél közelebb van közös ősök a GO fán, és minél mélyebb szintjén helyezkednek el a fának, annál inkább hasonlóak. Ezen alapulnak az ún. gráf alapú szemantikus hasonlósági (semantic similarity – SS) mérőszámok. Más SS mérőszámok figyelembe veszik, hogy az adott GO tulajdonságok milyen gyakorisággal vannak hozzárendelve génekhez. Minél ritkább az adott tulajdonság előfordulása, annál nagyobb súllyal esik latba a jelenléte (Alvarez és mtsai 2011).

1.6.3. Jelátviteli útvonal adatbázisok

A **PathGuide** (<http://pathguide.org/>) kb. 547, folyamatosan növekvő számú interakciós és útvonal adatbázis összehasonlító gyűjteménye (Bader és mtsai 2006). Ezek közül több, mint 93 jelátviteli útvonal adatbázis, hasonló számú metabolikus útvonal adatbázis mellett. A

több adatbázist felhasználó elemzésekben, és az adatok integrációja szempontjából fontosak a szabványos adatcserélő formátumok. A PathGuide-ben listázott jelátviteli adatbázisok közül 20 támogatja a BioPax, 16 az SBML, 8 a PSI-MI, 2 pedig a CellML formátumot. Az adatbázisok többsége szabadon felhasználható kutatási célokra, csupán 7 adatbázis érhető el kizárólag licenc megvásárlása esetén. 10 adatbázis jelenleg egyáltalán nem elérhető, melynek oka feltehetően a karbantartás hiánya.

Számos útvonal adatbázisként listázott adatforrás nem tartalmaz valódi útvonal információkat. A továbbiakban áttekintem azokat az adatbázisokat, melyek valamilyen szempontból egybevágnak az én törekvéseimmel, így összehasonlíthatók az általam létrehozni kívánt jelátviteli útvonal adatbázissal.

A **KEGG** (*Kyoto Encyclopedia of Genes and Genomes*) egy nagy lefedettségű, többféle adatot integráló adatbázis (Kanehisa és mtsai 2005). A genomok, gének, ortológiai kapcsolatok, reakciók, gyógyszerhatóanyagok és betegségekkel kapcsolatos információk mellett tartalmaz metabolikus és jelátviteli útvonalakat, valamint fehérje interakciókat is. Az összesen 455 útvonal kialakítása önkényes, így csak az átláthatóságot segítik. Azonban számos tanulmány és más adatbázis átvette ezeket, így általában valamiféle funkcionális ontológia szerepét töltik be. A fehérjék közti kapcsolatok eredete nincs pontosan megjelölve, és gyakran fehérje csoportok között valósul meg interakció, ami lehetlenné teszi a KEGG fehérje interakciós hálózatként valló felhasználását az elemzésekben (Kanehisa és mtsai 2005). Számos adatbázis a KEGG-ből veszi át az útvonalakat, és az előző fejezetben bemutatott nagy fehérje interakciós adatbázisok közül egy vagy több szolgál a kapcsolatok forrásául. Így épül fel az IntegromeDB, a Cell Cycle Database (Baitaluk és mtsai 2012), az Arabidopsis Reactome (Tsesmetzis és mtsai 2008), a FlyMine (Lyne és mtsai 2007), a Genomics of Lipid Associated Disorders Database, a Human Integrated Pathway Database (Yu és mtsai 2013), az Integrated Molecular Interaction Database (Balaji és mtsai 2012), az iPAVS (Sreenivasaiah és mtsai 2012).

A **Coronary Hearth Disease** adatbázis (Wu és mtsai 2013) egy hálózati kutatási platform kialakítását célozza meg, és ilyen tekintetben egy előremutató próbálkozás a rendelkezésre álló adatok rendszerszintű integrálására egy kutatási terület támogatása érdekében. Az adatbázis magját számítógépes szövegfeldolgozással segített irodalmi gyűjtés alkotja, mely a 45

szívkoszorúér betegségben érintett gén közti interakciókat térképezi fel. A fehérjék közti interakciós hálózat a BioGRID-ből és a HPRD-ből, míg a gyógyszerhatóanyagok és fehérjék interakciói a DrugBank adatbázisból származik. Az útvonal annotációkat a KEGG-ből vették át az adatbázis készítői.

A Semmelweis Egyetem LINK csoportjában Prof. Csermely Péter és Dr. Korcsmáros Tamás részvételével fejlesztett SignaLink 1 (Korcsmáros és mtsai 2010) három faj – *C. elegans*, *D. melanogaster* és ember – nyolc jelátviteli útvonala szerepel, melyeket fejlődéstani és patobiokémiai jelentőségük alapján választottak ki (Pires-daSilva és mtsai 2003). Az adatbázis csaknem ezer publikációt feldolgozó irodalmi gyűjtésen alapul, 1.441 fehérje 1.462 kapcsolatát írja le. A SignaLink 1 egységes hálózat, mely magában foglalja az útvonalak közti kapcsolatokat, és az első érintett transzkripció faktorokig fedti le az egyes útvonalakat. A SignaLink 1-ben alkalmazott útvonal definíció az útvonalakat valóban létező biológiai egységeknek tekinti, melyeket biokémiai sajátosságaik és evolúciós hátterük különböztet meg egymástól. Ez a megközelítés lényegileg eltér a többi útvonal adatbázisétól, melyek az útvonalakat inkább önkényesen meghatározott, didaktikai és tájékoztató szempontból hasznos, mesterséges entitásokként kezelik. A SignaLink 1-ben szereplő útvonal besorolások ezért felhasználhatók összehasonlító elemzésekben, valamint értelmezhetők és vizsgálhatók az útvonalak közti keresztbeszélgetések (Korcsmáros és mtsai 2010). A SignaLink 1 szolgált az ebben a dolgozatban bemutatott SignaLink 2 adatbázis fejlesztésének kiindulópontjaként.

Az **INOH** (*Integrating Network Objects with Hierarchies*) (Yamamoto és mtsai 2011) 59 humán, és kisebb számú, más fajokban leírt jelátviteli útvonalat, valamint 29 humán metabolikus útvonalat tartalmaz, melyeket hierarchikus módon alútvonalakra bont. 3.395 fehérje 6.155 interakciója szerepel az adatbázisban. Az útvonalak saját irodalmi gyűjtés eredményei, az útvonaltagságot a fehérjék INOH által létrehozott többféle ontológiába történő besorolásában jelentkező hasonlóság, illetve a gyűjtést végző szakértők ítélete határozza meg, azonban nincs pontos kritériuma. Ez a hierarchikus struktúra miatt ebben az esetben nem is szükséges. Az interakciós hálózat szintén hierarchikus felépítésű, minden pont reprezentálhat egy alhálózatot, például egy protein komplex tagjait. A saját ontológiák, az átgondolt, hierarchikus adatstruktúra és a saját irodalmi gyűjtés által az INOH fejlesztőinek sikerült számos releváns problémára

valóban új és előremutató megoldással szolgálni. Emiatt az INOH kiemelkedik a többi útvonal adatbázis köréből. Útvonal definíciója pedig a legközelebb áll a Signalinkben alkalmazotthoz. A weboldal által nyújtott lehetőségek az adatok böngészésére hiányosak, azonban az adatsorok letölthető szabványos formátumokban.

A **NetPath**-ban (Kandasamy és mtsai 2010) szereplő útvonalak az INOH-hoz hasonlóan saját irodalmi gyűjtéssel kerültek kialakításra, 2.228 cikk feldolgozásával. Összesen 1.053 fehérje 2.448 interakciója található az adatbázisban. A fehérjék több útvonalba is tartozhatnak. Az interakciók lehetnek binárisak, vagy komplexek esetén több fehérje vehet részt egy interakcióban. Négy útvonal (Hedgehog, TGF- β , Notch, WNT)) megegyezik a Signalinkben szereplő útvonalakkal, míg az androgén receptor útvonal az NHR útvonal részének tekinthető. A NetPath összesen 36 humán jelátviteli útvonalat tartalmaz, melyeket az immunitásban és a rákban játszott szerepük alapján választottak ki. Az útvonalak kialakításánál szempont volt az összehasonlíthatóság. A NetPath tartalmazza az útvonalakhoz közvetlenül kapcsolódó transzkripciós faktorokat, és összesen 7.401 TF-gén kapcsolatot .

A **PID** (*Pathway Interaction Database*) (Schaefer és mtsai 2010) egy jelátviteli útvonal adatbázis, melynek célja, hogy átfogó referenciaként szolgáljon a jelátvitel területén, és széles körben elfogadott konvenciót adjon az egyes fehérjék útvonal besorolására nézve. Jelenleg 137 humán jelátviteli útvonalat és 9.248 irodalomból gyűjtött interakciót tartalmaz. Az útvonalak kialakítása irodalmi adatokon és szakértők döntésén alapul. Az adatbázisban fehérjék, komplexek, RNS-ek és kis molekulák közti reakciók, transzkripcionális szabályozás és transzport folyamatok szerepelnek. Elkülönített módon, de a PID adatsémájába importálták a Reactome és a BioCarta adatait is. A weboldal sajnos hiányosságokkal küzd, ráadásul a kényelmetlen grafikai megjelenítés miatt kétséges, hogy teljesíteni tudja-e eredeti célkitűzését.

A **Reactome** (Croft és mtsai 2014; Joshi-Tope 2005) 7.088 fehérje 6.744 interakcióját tartalmazó adatbázis, mely 15.107 publikáció feldolgozásán alapul. Több faj szerepel benne, és támogatja az ortológok keresését. A Reactome adatstruktúrájának alapegysége a reakció, célja az élőlényekben előforduló összes kémiai reakció összegyűjtése, melynek a jelátviteli útvonalak csak kis részét jelentik. A reakciók funkcionális láncolatai az útvonalak, melyeket a gyűjtést végző kutatók alakítanak ki. Az útvonalak egymásba ágyazottak is lehetnek, a

nagyobb útvonal csoportokat pedig folyamatokba kategorizálják. Az útvonalak ilyen módon csak a tájékozódást szolgálják. A Reactome alapvetően különbözik a hálózatos adatbázisoktól, mivel nem egyenrangú entitások közti kapcsolatokat, hanem reakciók halmazát tartalmazza; egy fehérje sok komplexben és sok foszforilációs állapotban is szerepelhet egy-egy reakció csoport ábrázolása során.

A **BioModels** (Chelliah és mtsai 2013) és a **CellML** (Lloyd és mtsai 2008) adatbázis csaknem 500 illetve 330 molekuláris mechanizmus, köztük jelátviteli útvonalak logikai, reakciókinetikai, *rule-based* és egyéb kvantitatív modelljeit tartalmazza, melyek SBML formátumban letölthetők.

Különlegessége miatt említést érdemel a **WikiPathways** (Kelder és mtsai 2012), mely a Wikipedia közösségi alapú fejlesztési modelljét alkalmazza egy útvonal adatbázis fejlesztésére. Ennek megfelelően, az útvonalak definíciója tetszőleges, átfedőek lehetnek. Ugyanakkor a szolgáltatott tartalom megfelel a legtöbb hasonló elvek szerint kialakított adatbázisnak. Az aktív fejlesztők száma jelenleg 660, eddig 1.782 útvonal szerepel benne több tucat fajból.

Egyes adatbázisok a létező nagy számú útvonal adatbázis adatait kívánják egyesíteni, lehetővé téve az adatok együttes felhasználását és elemzését, valamint az összehasonlító megközelítéseket.

A **ConsensusPathDB** (Kamburov és mtsai 2010) funkcionális szempontok alapján rendezett molekuláris interakciós adatbázisok egyesítését tűzte ki célul. Jelenleg 159.131 molekula 413.392 interakcióját tartalmazza. Ezek között szerepelnek fizikai és genetikai interakciók, a transzkripcionális szabályozás, valamint gyógyszerek hatóanyag-célpont kölcsönhatásai. Az adatok 32 különböző adatbázisból származnak, melyek között megtalálható az összes nagy fehérje interakciós adatbázis. A hálózat 4.948 útvonalat tartalmaz, azonban ezeket a forrás adatbázisokból veszi át, így az útvonal definíciók különbözősége miatt nem hasonlíthatók össze.

A **Pathway Commons** (Cerami és mtsai 2011) modern és letisztult felhasználói felületével, CytoScape Web alapú, fejlett hálózat vizualizációs rendszerével tűnik ki az útvonal adatbázisok közül. Kilenc adatbázisból, BioPax formátumban integrálja az adatokat, így 2011-ben már 1.400 útvonal 687.000 interakciója volt elérhető. Az adatforrások a legnagyobb

fehérje interakciós adatbázisok, a BioGRID, a HPRD, az IntAct, a MINT, valamint más útvonal adatbázisok, a HumanCyc, az IMID, a Cancer CellMap, a Nature PID és a Reactome. Az előbbieket nem tartalmazzák a fehérjék útvonal besorolását, az utóbbiak útvonal definíciói pedig eltérők, így nem használhatók fel elemzésekben, csupán a tájékozódáshoz adhatnak támpontot. A hálózatban fehérjék, RNS, DNS és kis molekulák egyaránt szerepelnek. Egy modellezési keretrendszer, a „virtuális sejt” (VCell) segítségével a hálózatból automatikusan modellt készíthet a felhasználó. Hasonló funkciókat valósít meg a Genemania (Zuberi és mtsai 2013), azonban az adatok forrásainak elkülönítése és visszakövetése ebben az adatbázisban egyszerűbb a felhasználó számára. Ezen kívül, a Genemania miRNS-gén interakciók, a transzkripcionális szabályozás, valamint a hatóanyagok és célpontjaik hálózatát is tartalmazza.

1.6.4. Transzkripciós faktor-gén regulációs adatbázisok

Az **ABS** az egyik első transzkripciós faktorok kötőhelyeire és célgénjeire vonatkozó irodalmi gyűjtés (Blanco és mtsai 2006). 68 TF 100 gén promóterén elhelyezkedő, 650 kötőhelyéről tartalmaz adatot. Azok a kötőhelyek szerepelnek benne, amelyek a vizsgált négy gerinces faj közül legalább kettőben megtalálhatóak voltak ortológ géneken.

A **DroID** adatbázis a *Drosophila* fehérje és gén interakciók gyűjtőhelye, mely kísérletekből származó, irodalomból gyűjtött és prediktált interakciókat is tartalmaz (Murali és mtsai 2011). A DroID által kínált adatok közül a Signalink 2-ben a fehérje-fehérje interakciókat, és a prediktált transzkripciós faktor-gén interakciókat használtuk fel.

Az **EDGEDb** (*C. elegans differential gene expression database*) sok egyéb annotáció mellett 605 TF-promóter kapcsolatot tartalmaz, 176 TF és 115 szabályozott gén között (Barrasa és mtsai 2007).

Az **ENCODE** (*Encyclopedia of DNA elements*) 119 humán transzkripciós faktor ChIP-Seq technikával felderített DNS kötési profilját tartalmazza (Gerstein és mtsai 2012). A humán genomban összesen 7.424.765 kötési helyet sikerült gének promóterein azonosítani, ezek 40 %-a proximális, azaz a TSS-től (*transcription start site*) számított 2,5 kilobázison belül található. Az ENCODE az adatok szűrése után 45.328 kapcsolatot tesz közzé transzkripciós

faktorok és fehérje kódoló gének között, illetve 1.648 kapcsolatot TF-ek és általuk szabályozott mikro-RNS-ek között. Ezeket az adatsorokat mindhárom adatbázisba beépítettem.

A **HTRIdb** 284 humán TF és 18.302 célgén között összesen 51.871 kapcsolatot tartalmazó, irodalmi gyűjtésen alapuló adatbázis (Bovolenta és mtsai 2012). A kapcsolatok megbízhatóságát aszerint pontozza, hogy hány cikk és hányféle technikával igazolta a kapcsolat létét. Az interakciók döntő többsége nagy teljesítményű módszerekből származik, mint a ChIP-chip, vagy a ChIP-Seq.

A **JASPAR** adatbázis a transzkripciós faktorok DNS kötési profiljait mátrixok formájában tárolja, melyek megadják, hogy a DNS szekvencia egyes pozícióiban mely bázis jelenléte milyen gyakorisággal fordul elő a kísérletes adatokban (Portales-Casamar és mtsai 2010). Ez az ún. pozíció-súly modell (*Position Weight Model – PWM*), mely az egyes pozíciókban szereplő különböző nukleotidokhoz súlyokat rendel, annak arányában, hogy a nukleotid jelenléte mekkora mértékben járul hozzá az adott fehérje DNS kötéshez. A mátrixok jórészt ChIP-Seq, ChIP-chip és SELEX technikákkal nyert adatokból származnak.

Az **ORegAnno** egy irodalmi gyűjtésre épülő adatbázis, mely 962 tanulmányból származó, 8.011 humán, 1.445 *Drosophila* és 218 *C. elegans* TF kötőhelyről tartalmaz adatot (Griffith és mtsai 2008). Az adatok beküldésében a tudományos közösség is részt vehet. Az irodalmi adatok gyűjtését számítógépes szövegfeldolgozó módszerekkel segítik. Egyes nagyobb adatsorokat más adatbázisokból (pl. RedFly), illetve nagyobb ChIP-Seq kísérletekből is átvettek.

A **PAZAR** adatbázishoz a kutatóknak lehetőségük van hozzáadniuk saját adatsoraikat (Portales-Casamar és mtsai 2009). Ezeket a beküldött adatokon kívül az ENCODE és az ORegAnno adatbázisokból átvett, és JASPAR predikcióból származó adatokat foglal magában. A teljes PAZAR adatbázis 1.284 gén és 708 TF közötti kapcsolatokat tartalmaz.

A **RedFly** 158 TF 2.051 kötőhelyét tartalmazza, melyek 209 gén promóterén helyezkednek el (Gallo és mtsai 2011). Az adatokat 662 cikkből, manuálisan gyűjtötték össze. Az adatbázis *Drosophila* fajokról gyűjt adatokat, és expressziós mintázatokat is tartalmaz.

A **wTF2.0** (*Worm Transcription Factors*) a doméneket és Gene Ontology tulajdonságokat figyelembe véve 934 transzkripciós faktort prediktál *C. elegans* esetébe, és tartalmazza az ezek közti interakciókat is (Reece-Hoyes és mtsai 2005).

1.6.5. mikro-RNS-mRNS interakciós adatbázisok

A **micro-T v4** algoritmus az evolúciós konzerváltságot és a szekvencia kontextuális jellemzőit egyaránt figyelembe veszi (Maragkakis és mtsai 2009, 2011). 7–9 bázispárnyi illeszkedést követel meg, de ha a 3' végen is van illeszkedés, akkor elfogad egy G:U nukleotid párt vagy csak 6 bázispárnyi illeszkedést a mag régióban. A konzervációs profil számításához 27 faj genomját használja fel, majd súlyozva összegzi a predikció két komponensét. Az így jelzett kapcsolatok között megtalálhatóak a nem konzervált, de kontextuális paraméterek alapján nagy valószínűségű kötőhelyek is.

A **miR2Disease** elsődleges célja az adatgyűjtés a miRNS-ek betegségekben játszott szerepéről (Jiang és mtsai 2009). Emellett azonban 299 miRNS célpontjairól tartalmaz, összesen 600 cikkből gyűjtött, megbízható, kísérleteken alapuló információkat.

A **miRDB** a MirTarget2 elnevezésű, SVM tanulóalgoritmuson (SVM – *Support Vector Machine* – Tanulóvektor-gép) alapuló predikciós módszer eredményeit tartalmazza (Wang és mtsai 2008). A modell figyelembe veszi a konzerváltságot, a mag régió illeszkedését, nukleotid összetételét, szabadenergiáját, a kötőhely UTR szakaszon belüli pozícióját és hozzáférhetőségét.

A **miRDeathDB** 92 miRNS és 106 fehérje közötti kapcsolatok irodalmi gyűjtésen alapuló adatsora (Xu és mtsai 2012). Célja az apoptózis és az autofágia miRNS-eken keresztül történő szabályozásának feltárása.

A microRNA.org miRNS expressziós profilokat és a **miRanda** valamint a **mirSVR** algoritmusokkal előállított kötőhely predikciókat tartalmazó adatbázis (Betel és mtsai 2008). A miRanda algoritmus három lépésből áll: először illeszkedést számol a miRNS teljes hosszára, csupán alacsonyabban súlyozza a 3' vég nem fontos bázisait. Tökéletlen illeszkedést, G:U párok esetén büntetőpontokat alkalmaz. Az egyes pozíciók súlyozása kísérletes adatokon alapul. Ezután kiszámolja a hibridizáció szabadenergiáját, végül pedig szűrést végez evolúciós konzerváltság alapján (Enright és mtsai 2003).

A miRanda segítségével talált kötőhelyeken később egy újabb predikciós módszert is alkalmaztak. A mirSVR (SVR – *Support Vector Regression* – SVM Regresszió) gépi tanuláson

alapul, és a predikció által jelzett kötőhelyek kontextus paramétereit (Grimson és mtsai 2007) veti össze a miRNS transzfekciós, argonauta protein immun-precipitációs (AGO-IP), valamint PAR-CLIP (*Photoactivatable Ribonucleoside Enhanced Cross-linking Immunoprecipitation*) eredményeivel (Betel és mtsai 2010). A figyelembe vett paraméterek és az alkalmazott matematikai módszerek alapján a mirSVR mindenképp az egyik legfejlettebb predikciós algoritmus, különös tekintettel a nem kanonikus és nem konzervált kötőhelyek előrejelzésében mutatott teljesítményére (Betel és mtsai 2010).

A **PicTar** a konzerváltságot és a kontextuális jellemzőket egy modell keretein belül veszi figyelembe (Krek és mtsai 2005). A miRNS-ek versengenek a kötőhelyekért, az egymáshoz közel eső kötőhelyek erősítik egymást. A hibridizáció szabad energiáját 7 nukleotid hosszúságú mag szekvenciákra számolja, a pontatlan illeszkedésekhez alacsony súlyt rendel, így ezek általában kiesnek. A felhasználó szabhatja meg, hogy hány szekvenciában kell konzerválnak lennie az adott kötőhelynek. A **doRiNA** a PicTar algoritmusnak a friss adatokkal, három konzerváltsági szintre lefuttatott eredményét tartalmazza (Anders és mtsai 2012).

A **TarBase** a kísérletesen igazolt miRNS-mRNS kapcsolatok legnagyobb gyűjteménye (Vergoulis és mtsai 2012). A MedLine absztrakt gyűjtemény számítógépes szövegbányászatán alapul, de a kapcsolatok manuális gyűjtés útján kerülnek az adatbázisba. Több mint 66.000 kapcsolatot foglal magában. Az adatok nagy része *high-throughput* kísérletekből származik.

A **TargetScanS** algoritmus (Friedman és mtsai 2009; Lewis és mtsai 2005) a 6 nukleotidból álló mag tökéletes illeszkedését, és vagy a 8. nukleotid illeszkedését, vagy az 1. pozícióban az mRNS-ben adenin jelenlétét követeli meg. Az eredeti TargetScan algoritmus konzerváltságon alapult, de később kifejlesztettek egy kontextusfüggő modellt, mely a 13–16. bázispárok illeszkedését és a szekvencia AU arányát is figyelembe veszi. A P_{CT} filogenetikai fán mért relatív konzerváltság alapú pontszám, és a *context score* együttesen határozzák meg a predikciók jóságát.

1.6.6. Transzkripció faktor–mikro-RNS gén regulációs adatbázisok

Az **ENCODE** adatsor tartalmaz TF–miRNS kapcsolatokat is. Az ENCODE adatokról részletesen a 1.6.4. fejezetben írok.

A **PuTmiR** predikción alapuló TF–miRNS szabályozási kapcsolatokat tartalmaz (Bandyopadhyay és mtsai 2010). A miRBase-ből származó miRNS szekvencia adatok, az UCSC genome browserből származó humán genom, illetve konzervált TF kötőhely adatok alapján, a miRNS géneket körülvevő, 10.000 bázispár hosszú *upstream* és *downstream* régióban kerestek 258 féle kötőhely szekvenciát. Az illeszkedések tökéletességét a bázisgyakoriságok szerint korrigálva, minden kapcsolathoz egy mérőszámot is hozzárendelnek.

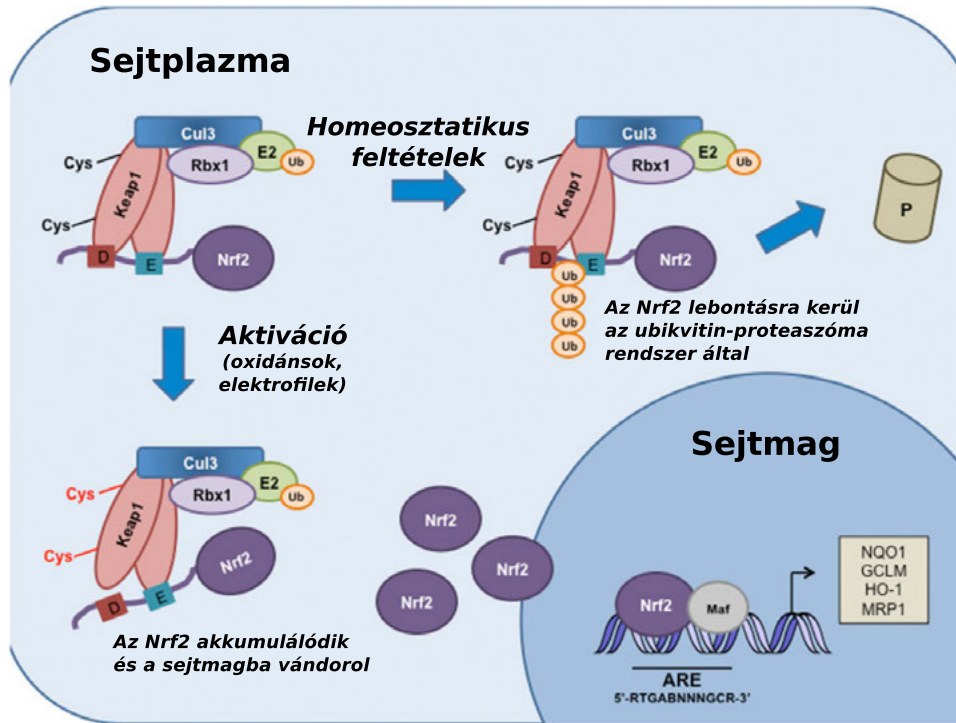
A **TransmiR** irodalmi gyűjtésből származó TF–miRNS kapcsolatok gyűjteménye (Wang és mtsai 2008). 86 cikkből, 82 TF és 100 miRNS közti 243 kapcsolatról tartalmaz adatot.

1.7. Az NRF2 transzkripció faktor

Az eddigiekben bemutatott azokat az adatforrásokat és szabályozási mechanizmusokat, melyek a doktori munkám részét képező mindhárom adatbázis – a Signalink 2, az NRF2ome és az ARN – felépítésében kulcsszerepet játszanak. Áttekintettem a jelátviteli hálózatok jellemzőit, melyek mindhárom adatbázis közös részét képezik. Ebben a fejezetben körvonalazom az NRF2 transzkripció faktor biológiai funkciójáról és szabályozásáról az eddigi kutatások alapján kialakult képet, mely az NRF2ome adatbázis létrehozására ösztönzött.

Az NRF2 az oxidatív és a xenobiotikus stresszválasz fő transzkripcionális szabályozója. bZIP szerkezetének köszönhetően heterodimereket képez a kis MAF fehérjék ZIP doménjével. Hat Neh domént tartalmaz, melyek lehetővé teszik a DNS kötést, valamint számos más fehérjével, köztük transzkripció faktorokkal történő interakciót (Papp és mtsai 2012). Az NRF2 Neh2 doménjén keresztül a KEAP1-hez kötődve gátlás alatt áll, és hamar proteasomális lebontásra kerül (Kensler és mtsai 2007). Oxidatív stressz esetén ez a gátlás feloldódik, és az NRF2 a sejtmagba vándorol, ahol MAF fehérjékkel heterodimert képez (3. ábra).

Az NRF2 az ARE (*antioxidant response element*) vagy EpRE (*electrophile response element*) szegmenst tartalmazó promóterekhez kötődik, így aktiválva a sejt redox homeosztázisának helyreállításához szükséges gének (például, NAD(P)H-kinon-oxidoreduktáz-1, hemoxigenáz-1, glutamát-cisztein ligáz, glutation-S-transzferáz, glutation-peroxidáz, tioredoxin, stb.) transzkripcióját (Surh és mtsai 2008). Az NRF2-t számos jelátviteli fehérje képes fel szabadítani a KEAP1 gátlása alól, mint a MAPK, PI3K/Akt, a PKC, a CK (kazein-kináz),



3. ábra. **Az NRF2 szabályozásának áttekintése.** Nyugalmi állapotban az NRF2 a KEAP1-hez kötött, a citoplazmában található. Ilyen körülmények között féléletideje rövid, hamar ubiquitinálódik, és lebontásra kerül. Oxidatív stressz esetén – melynek hatását többféle mechanizmus közvetítheti, ezekről részletesebben a szövegben – az NRF2 felszabadul a KEAP1 gátlása alól, és sejtmagba vándorol, ahol megindítja a promóterükben ARE szegmenst tartalmazó gének transzkripcióját. – Az ábra forrása: Jaramillo és mtsai 2013.

az ERK1/2, vagy a p38. Cisztein oldalláncainak köszönhetően a KEAP1 maga is funkcionálhat oxidatív ágensek érzékelőjeként (Kensler és mtsai 2007). Egyes fehérjék más módon szabályozzák az NRF2-KEAP1 kapcsolatot: a p62 a KEAP1 ubiquitinációja révén stabilizálja az NRF2-t, a p21 az NRF2-KEAP1 kötődést gátolja, a DPP3 pedig a KEAP1-hez kötődve megakadályozza az NRF2 ubiquitinációját (Jaramillo és mtsai 2013).

Az NRF2 számos fiziológiás és patológiai folyamatban kulcsszerepet tölt be. Az NRF2 *knock-out* egerek sérülékenyebbek karcinogének és mérgeanyagok hatása alatt, valamint gyulladás esetén. Az oxidatív stresszválasz fontos szerepet tölt be szív- és érrendszeri betegségekben, rákban, krónikus gyulladásban és neurodegeneratív betegségekben egyaránt. A

DNS-t károsító oxidatív ágensek a rák kialakulásához is hozzájárulhatnak. Az NRF2 aktiválja a proteaszóma rendszer génjeinek transzkripcióját is, mely segít a károsodott fehérjék lebontásában (Kensler és mtsai 2007). A rákos sejtek gyakran sok NRF2-t expresszálnak, mivel segít kivédeni a kemoterápia, a sugárterápia és az oxidatív stressz hatásait (Jaramillo és mtsai 2013).

A PERK szintén foszforilálhatja az NRF2-t, az endoplazmatikus retikulum stresszválasz részeként. Az NRF2 a saját transzkripcióját is aktiválja, biztosítva a stresszválasz fenntartását. Továbbá, az NRF2 transzkripcióját szabályozza az AhR xenobiotikus válaszáért felelős receptor is. Hasonló módon, az NF- κ B immunválaszban és glutation homeosztázisban betöltött funkciója is összefüggésben van az NRF2 működésével (Kensler és mtsai 2007).

Az NRF2 a már említett biológiai funkciói miatt nagy mértékben a kutatások középpontjában áll, azonban a nagy fehérje interakciós adatbázisok (BioGRID, MINT, STRING, HPRD and InnateDB) csupán pár tucat interakciós partnerét tartalmazzák. A Semmelweis Egyetem LINK hálózatkutató csoportja és az ELTE NetBiol kutatócsoportja által készített gyűjtés némileg javít ezen a helyzeten, az NRF2 és a KEAP1 közvetlen szabályozási hálózatát alkotó 289 interakció irodalomból és adatbázisokból történő összegyűjtésével (Papp és mtsai 2012). Szembetűnő az NRF2 transzkripcionális szabályozása alá eső gének nagy száma: egy ChIP-Seq kísérletben (Malhotra és mtsai 2010) az NRF2 1.054 célgénjét sikerült felderíteni, míg a JASPAR predikciós módszerrel további 6.426 potenciális NRF2 által szabályozott gént. Az összesen 7.469 célgén a szabályozási mechanizmusok, visszacsatolások bioinformatikai eszközök nélkül áttekinthetetlen, óriási halmazát vetíti elénk. Ráadásul, az NRF2 translációját 85 miRNS szabályozza, melyek közül 63 transzkripciójára az NRF2 által szabályozott 18 transzkripció faktor van hatással (Papp és mtsai 2012). Az idézett tanulmány az NRF2 szabályozására vonatkozó legszélesebb körű rendszerszintű elemzést nyújtja, azonban még az ebben publikált adatok is jórészt egy csillagpontos hálózatot alkotnak, melynek közepén az NRF2 áll, és nem tartalmazzák az NRF2-től két, három, vagy több lépésnyi távolságban szereplő molekulákat. További probléma, hogy az adatok táblázatokban vannak tárolva, ahol minden fehérjének csupán egy elnevezése van, ami a keresést nehezzé teszi. Az adatok hálózatos

elemzése vagy vizualizálása pedig bioinformatikai tudást igényel, így sok biológus kutató számára nehézségekbe ütközik.

Az NRF2 szabályozási hálózatának elkészítésében én is részt vettem, a kapcsolatok adatbázisokból történő beszerzése, és az adatok bioinformatikai feldolgozása során.

1.8. Az autofágia

Az autofágia áll a doktori munkám harmadik részét alkotó adatbázis, az ARN (*Autophagy Regulatory Network*) fókuszában. Ebben a fejezetben összefoglalom az autofágia mechanizmusáról és annak szabályozásáról alkotott ismereteinket, rámutatva a kutatási igényekre, melyek az ARN létrehozását indokolják.

Az autofágia a károsodott organellek és fehérjék eltávolítása révén biztosítja ezek folyamatos újratermelődését. Az öregedés során ez a folyamat lelassul. Egerekben az autofágia gátlása gyorsabb öregedést okoz, míg az *Atg5* túlexpressziója megnöveli az élethosszot. A kalóriaszegény táplálkozás serkenti az autofág aktivitást, és számos modell organizmusban az élethossz növekedését is eredményezi. A sejt energiaszintje közvetlenül szabályozza az autofágiát, a SIRT1–IGF útvonalon, valamint az AMPK-n keresztül (Pyo és mtsai 2013).

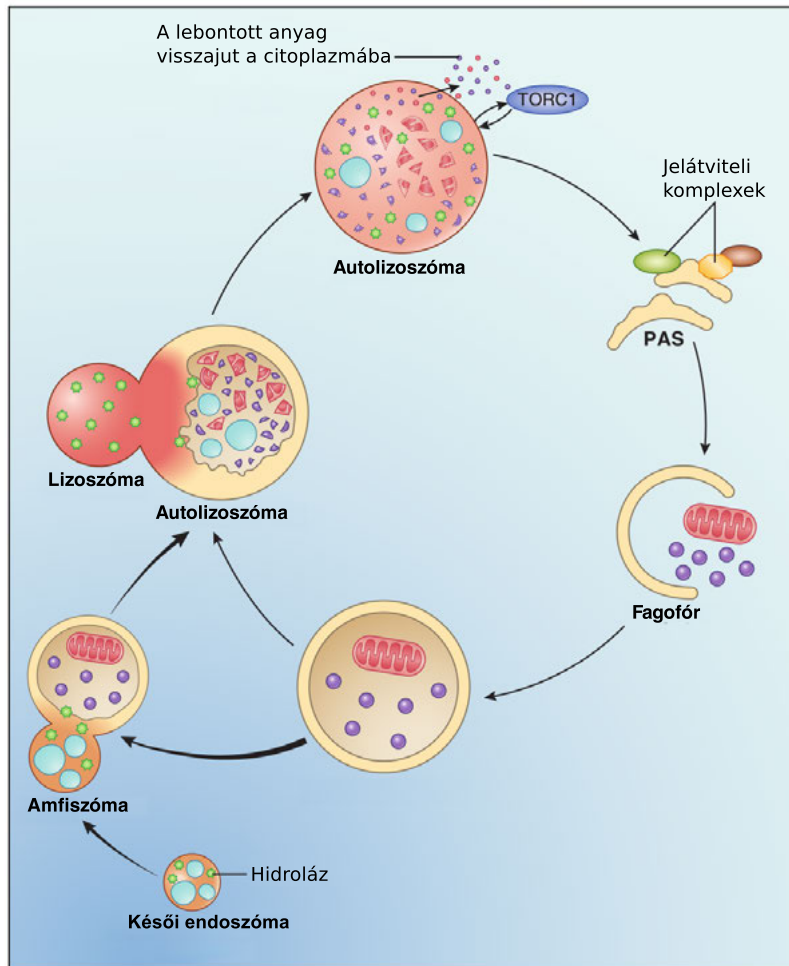
Számos betegség esetén megfigyeltek eltéréseket az autofág aktivitásban, és az autofágia sokszor fontos tényező ezen betegségek patomechanizmusában. A túl magas autofág aktivitás sejthalálhoz vezethet, ugyanakkor az autofágia nélkülözhetetlen a sejtek egészséges állapotának fenntartásához, a stressz elleni védekezéshez, éhezés esetén az anyagcsere fenntartásához, és általában a túlélést szolgálja, az apoptózissal szemben. Az autofágia rákban betöltött szerepe már régóta a kutatások fókuszában van, ugyanakkor neurodegeneratív betegségek, vírusfertőzések, kardiovaszkuláris betegségek és stroke esetében is jelentős.

A felsorolt betegségek mindegyikében az autofágia szerepét a „kétélű kard” analógiájával illetik, mivel bizonyos feltételek esetén hatása pozitív, míg máskor negatív a betegség prognózisa szempontjából (Gurusamy és mtsai 2009; Kenific és mtsai 2010; Lin és mtsai 2010; Martinet és mtsai 2009; Tung és mtsai 2012; Wei és mtsai 2012; White és mtsai 2009). Daganatos megbetegedésekben az autofágia jótékony hatású, amennyiben segít megőrizni a kromoszomális stabilitást, védekezni a stresszhatások, például az oxidatív stressz ellen, gyul-

adás csökkentő hatású, és szenescenciához vezethet. Ugyanakkor az autofágia hozzájárulhat a tumorsejtek túléléséhez, mivel elősegíti a terápiás és metabolikus stressz átvészelését, ezzel hozzájárulva a sugár- és kemoterápiákkal szembeni rezisztenciához, és a metasztázis képzéshez (Kubisch és mtsai 2013). Az ellentmondásos funkció miatt a közvetlen autofágia serkentők és gátlók – melyek többnyire az mTOR komplexen keresztül hatnak az autofágiára – terápiás alkalmazása a legtöbb esetben nem bizonyult sikeresnek (Chen és mtsai 2011; Kimura és mtsai 2013). Az autofágia gyógyszeres szabályozása azonban a különféle daganatok terápiájában továbbra is ígéretes elemnek tűnik.

A neuronok különösen érzékenyek az autofágia zavaraira, számos neurodegeneratív betegségben kulcsszerepet játszik az autofágia (Nixon 2013). Parkinson-kórban a PARK2 és a PINK1 mutációja miatt nem képes részt venni a mitofágia (mitokondriumokat lebontó autofágia) indukciójában, ami végül az idegsejtek károsodásához vezet (Jin és mtsai 2012). Az Alzheimer-kór során a preszenilin-1 mutációja megakadályozza a kellően savas pH kialakulását a lizoszómákban, így a túlterhelt lizoszomális rendszer képtelen lebontani a károsodott fehérje aggregátumokat. Más neurodegeneratív betegségek molekuláris patomechanizmusában is megfigyelhetők a lizoszomális rendszer, illetve az autofágia szabályozásában bekövetkezett hibák (Nixon 2013). Ezenkívül az autofágiának fontos szerepe van az antigén bemutatásban, és az immunválasz kiváltásában (Ma és mtsai 2013).

Valószínűsíthető, hogy az autofágiát célzó terápiáknak kontextusfüggőeknek kell lenniük, hogy megfelelő hatást érjenek el a különféle szövetekben és daganattípusokban (Kubisch és mtsai 2013). A jelek szerint jó pár év kutatómunkája szükséges ahhoz, hogy ezt elérhessük. A rendszer komplexitása elkerülhetlenné teszi a bioinformatika eszköztárának bevetését, melynek célja lehet az autofágia – jelátviteli hálózaton keresztül érvényesülő – szabályozásának, az egyes gyógyszerhatóanyagok hatásának modellezése. A modellezés kiindulópontja azonban mindig az *a priori* tudás összegyűjtése, ami ebben az esetben többek között a molekuláris interakciók hálózatát jelenti. A sikeres modellezés megalapozásához össze kell gyűjteni és egyesíteni kell a különféle rendelkezésre álló módszerekből származó, és a biológiai rendszerekben előforduló szabályozási mechanizmusokat lefedő adatokat.



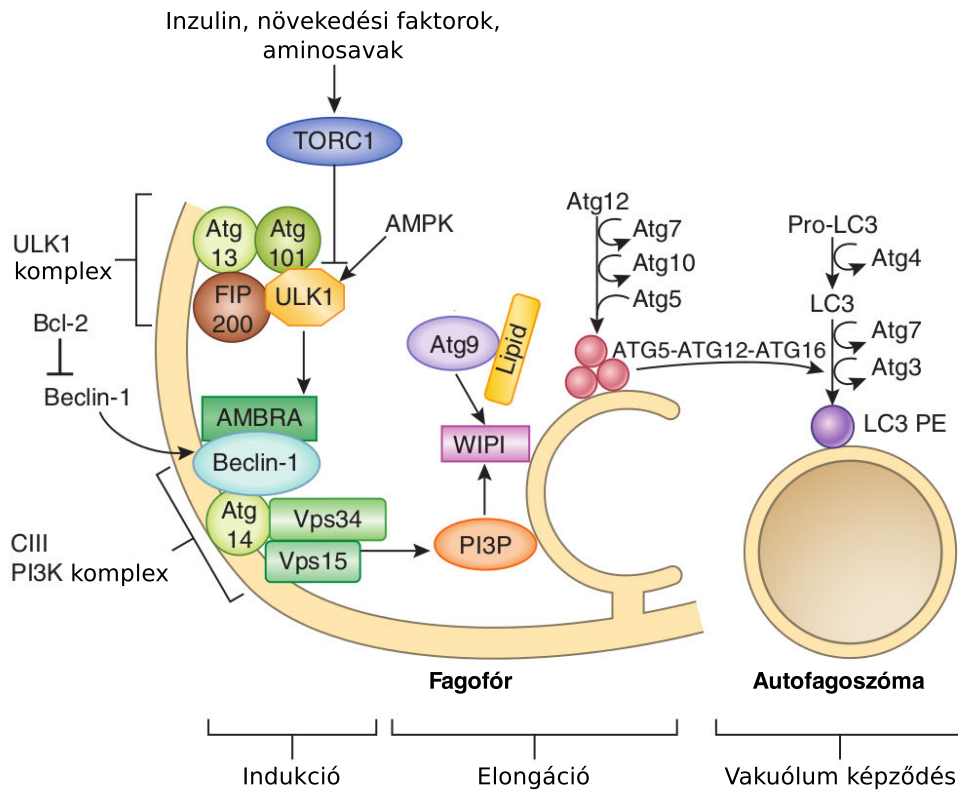
4. ábra. A makroautofágia folyamatának áttekintése. A PAS (*phagophore assembly site*) területén indul meg a fagofór képződése. A fagofórt felépítő membránok helyszínre juttatásában és a képződő fagofórba való beépítésében számos fehérje közreműködik, a folyamatot pedig az autofágia fehérjék bonyolult kölcsönhatása szabályozza. A záródó fagofór az autofágia receptorok segítségével magába zárja a lebontásra ítélt elemeket. A záródott kettős membrán vezikula az autofagoszóma, mely lizoszómával és késői endoszómával egyesülve autofago-lizoszómát képez. A belső membrán és a vezikula tartalma ezután emésztésre kerül, és az így képződött metabolitok a citoplazmába transzportálódnak. – Az ábra forrása: Nixon 2013.

Az autofágia kettős membrán vezikula képződésével kezdődik, mely az endoplazmatikus retikulumból kialakuló pre-autofagoszómális struktúra (PAS, más néven *phagophore assembly site*) területén veszi kezdetét (4. ábra). A formálódó autofág vezikulába Golgi-, endoszóma-

és plazmamembrán eredetű membránok is beépülnek. A kész autofagoszóma lizoszómákkal egyesül, ahol a tartalmát a lizoszomális hidrolázok lebontják (Boya és mtsai 2013).

Az autofágia fehérjék számos más fehérjével való kölcsönhatásuk, és az autofágiától eltérő egyéb funkcióik révén bonyolult szabályozás alatt állnak. Ennek köszönhetően az autofágia szelektív módon képes lebontani mitokondriumokat, peroxiszómákat, endoplazmás retikulumot, endoszómákat, szekréción vezikulákat, lipid cseppeket, glikogént, fehérje aggregátumokat, riboszómákat, valamint patogéneket. Ezt a szelektivitást autofágia receptorok biztosítják, a képződő autofagoszóma belső felületén jelenlévő ATG8 homológokkal történő, LC3 interakciós régiójuk (LIR) vagy ATG8 interakciós motívumaik (AIM) által megvalósuló kölcsönhatásaik révén. Élesztőben például az ATG32 és az ATG36 mitokondrium illetve peroxiszóma specifikus autofágia receptoroknak bizonyultak (Boya és mtsai 2013). Léteznek az ubikvitinált fehérjék autofág lebontására szolgáló autofágia receptorok, mint a p62, az NBR1, az NDP52, az optineurin, a hiszton-deacetiláz 6 és a NIX (Nixon 2013), valamint a patogéneket célzó autofágiában szerepet játszó SMURF1 ligáz és STING (Boya és mtsai 2013).

Az autofágia iniciációja és végrehajtása magában foglalja a membrán lipidek és autofágia fehérjék toborzását a PAS területére, a fagofór membrán növekedését, a lebontandó anyagok specifikus összegyűjtését a fagofór belső membránjához kapcsolódóan, majd a kettős membrán vezikula záródását, mely az autofagoszóma kialakulását jelenti. Ezután az autofagoszóma többnyire lizoszómával egyesül, így a belső membrán és tartalma megemésztődik. Emlősökben az ATG1 homológjai (ULK1, ULK2, ULK3) komplexet alkotnak a FIP200-zal (ATG17 homológ) és az ATG13-mal, valamint az ATG101-gyel (5. ábra). Az mTOR az ULK-hoz kötődik, és foszforilálja az ATG13-at. Az mTOR disszociációja esetén az ULK1/2/3 defoszforilálódik, és foszforilálja az FIP200-at és az ATG13-at. Az autofágia iniciációban közreműködő másik fehérje csoport a PI3K komplex, mely a fagofór membránban a foszfatidil-inozitol-3-foszfát előállításához szükséges. Ezt a komplexet a Vps34 (PI3K) mellett a p150, a Beclin-1 (ATG6 homológ) és az ATG14L vagy az UVRAG (Vps38 homológ) alkotja (Reggiori 2012). Az ULK1 az ATG14L segítségével kölcsönhatásba kerül a Vps34-Beclin-1 komplexszel, és foszforilálja a Beclin-1-et, ezzel elősegítve annak komplexben maradását. Ez az interakció felelős



5. ábra. **Az autofágia fehérjék szerepe a fagofór képződésében.** Az autofágia fehérjék többsége az autofágia iniciációja során a PAS (*phagophore assembly site*) területére lokalizálódik. A folyamat kezdőpontját az ULK1 és a PI3K komplex aktiválódása jelenti, mely több lépésen keresztül a fagofór membrán növekedéséhez, és az ATG8 homológok membránkötött formájának kialakulásához vezet. Ez utóbbi kulcsfontosságú az autofágia receptorok működése, így a lebontásra szánt anyagok autofagoszómába juttatása szempontjából. A további részleteket lásd a szövegben. – Az ábra forrása: Nixon 2013.

az mTOR–ULK1 útvonalon történő, aminosav hiány indukálta autofágia iniciációért (Russell és mtsai 2013). A PI3K komplex elősegíti a WIPI fehérjék (ATG18 homológok, emlősökben négy van belőlük: WIPI1, -2, -3 és -4) toborzását a PAS és a fagofór membrán területére (5. ábra). A WIPI fehérjék és az emlősökben jelenlévő két ATG2 homológ szükségesek az ATG9 visszaszállításához, mely a Golgi és a késői endoszóma membránból jut el a PAS/fagofór membránba (Reggiori 2012). A PI3K komplex közvetlenül segíti az ATG9 cirkulációját, míg az ULK1 komplex hatása lehet közvetett is, az előbbi serkentése által (Russell és mtsai 2013). Az ATG9 a VMP1 mellett az egyetlen transzmembrán autofágia fehérje (Reggiori 2012). Az

ATG9 funkciója feltehetően a membrán darabok szállítása a PAS területére, míg a VMP1 a PI3K komplex tagjainak fagofórhoz toborzását segítheti elő (Yang és mtsai 2010).

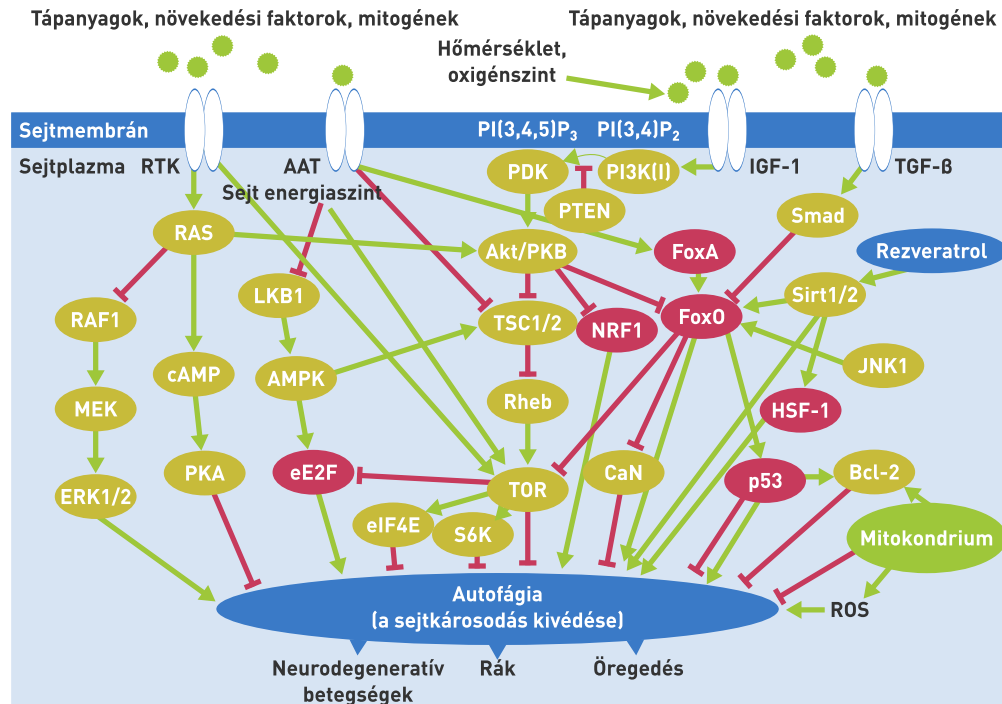
Az ATG12 ubikvitin-szerű fehérje az ATG7 (E1-szerű), és az ATG10 (E2-szerű) fehérjék segítségével az ATG5-tel konjugálódik, majd az ATG16-tal együtt, egy E3-szerű enzimet alkotnak, mely szükséges az ATG8 homológok lipidált, membránkötött formájának kialakításához. Emlősökben hat ATG8 homológ fordul elő (LC3A, LC3B, GABARAP, GABARAPL1, GABARAPL2 és GABARAPL3), melyek C-terminális végét az ATG4 hasítja le, így lehetővé teszi, hogy az ATG8 homológok az ATG7 (E1-szerű), majd az ATG3 (E2-szerű) fehérjékkel konjugálódva, végül a foszfatidil-etanolaminhoz (PE) kapcsolódjanak (5. ábra). E két ubikvitinszerű rendszer szorosan összefügg. LC3-II hiányában nem jön létre az ATG16L komplex, mely E3 szerű enzimeként közreműködhet az ATG8 homológok PE-hez való kapcsolódásában (Yang és mtsai 2010). A fagofór elkészültével szintén az ATG4 hasítja le az ATG8 homológokat a membránról, így az érett autofagoszómán nem találhatók a membránkötött formában. Az ATG8 homológok jelenléte fontos a membránfúziók kivitelezésében, és az autofagoszóma tartalom autofágia receptorok révén történő begyűjtésében (Reggiori 2012).

1.8.1. Az autofágiát szabályozó útvonalak

Az autofágia közvetlen szabályozóiról és az ezekre ható útvonalakról egyre több információra derül fény. Az autofágia központi szerepet tölt be a sejt homeosztázisában: egyrészt önlebontó folyamat, mely szélsőséges esetben képes sejthalált okozni, miközben elengedhetetlen a sejt egészséges működésének fenntartásához, többféle stresszhatás kivédéséhez. Ezeket figyelembe véve, az autofágiának precíz és kontextusfüggő szabályozás alatt kell állnia.

Az autofágia talán legismertebb közvetlen szabályozója az mTOR komplex, melynek emlősökben két változata található meg – mTORC1 és mTORC2. Az mTORC1 gátolja az autofágiát inicializáló ULK1 kinázt. Az mTORC1 több irányból érkező jelet integrál, melyek a sejt energiaszintjét, az aminosav anyagcsere állapotát és növekedési faktorok hatását közvetítik (6. ábra; Boya és mtsai 2013).

Az mTORC1-et szabályozó egyik fontos útvonal az AMPK felől érkezik, mely a sejt energiaszintjének szenzora: energiaszegény körülmények között foszforilálja a TSC1-TSC2



6. ábra. **Az autofágiát szabályozó fontosabb útvonalak áttekintése.** Az autofágiát szabályozó számos útvonalból érkező jel, részben a TSC1/2 közvetítésével, az mTOR komplexen integrálódik. Az AMPK a sejt energia szenzoraként funkcionál. Az IGF receptorból érkező jel a PI3K–PDK–Akt kaszkádon keresztül éri el a TSC1/2-t. A Bcl-2 fontos az autofágia és az apoptózis közti döntés szabályozásában. A FoxO az autofágia fehérjék génjeinek transzkripcióját serkenti. További részletek a szövegben. – Az ábra Vellai és mtsai 2009 nyomán készült, kisebb módosításokkal.

komplexet, ezáltal gátolva az mTOR-t, és elősegítve az autofágiát. Ezzel párhuzamosan az AMPK foszforilálja az mTOR egyik alegységét, a Raptor-t, ezzel is gátolva az mTOR hatását. Az AMPK útvonal stressz, citoplazmikus Ca^{2+} koncentráció növekedés, illetve hipoxia hatását is közvetítheti (Yang és mtsai 2010). A nukleáris p53 transzkripcionális szabályozás útján serkenti, míg a citoplazmikus p53 eddig ismeretlen módon gátolja az autofágiát (Tasdemir és mtsai 2008; Yang és mtsai 2010).

A receptor-tirozinkináz családba tartozó IGF1R ligand kötése esetén autofoszforilálódik, és aktiválja a PI3K-t. Ez utóbbi PIP_3 termelése révén a PDK-t aktiválja, mely a PKB–TSC1/TSC2–Rheb kaszkádon keresztül stimulálja az mTOR-t, következésképp gátolja az autofágiát. A PI3K aktivitását a PTEN gátolja, ezáltal pozitívan serkentve az autofágiát. Az mTORC2

szükséges a PKB aktiválásához, emellett hiányában a FoxO3 transzkripciós faktor aktiválódik, mely sok autofágia gént pozitívan szabályoz (Yang és mtsai 2010).

Az ATG14L szintén fontos szabályozási pont. Táplálékhiányban az ER-en lokalizálódik, míg éhezés hatására az ATG16L komplexszel a fagofór, illetve az LC3-mal együtt az autofagoszóma membrán közelében helyezkedik el. Az ATG14L-lel kompetitív módon kötődik a PI3K komplexhez az UVRAG, mely a Bif-1-gyel kerül kapcsolatba. A Bif-1-nek szerepe lehet a membrán görbülésének kiváltásában. Az UVRAG továbbá a HOPS (C Vps) fehérjékkel való interakción keresztül hozzájárulhat az autofagoszóma és a késői endoszóma, valamint a lizoszóma fúziójához. Az UVRAG–PI3K komplexhez a Rubicon fehérje kötődése esetén az endoszóma és lizoszóma területére helyeződik át, ezáltal késleltetve az autofagoszóma érését.

Az AMBRA1 serkenti az autofágiát, a Vps34-gyel, az ATG14L-lel és a Beclin-1-gyel kerül kapcsolatba. A Beclin-1 a Bcl-2-höz vagy a Bcl-X_L-hez kötődve nem vesz részt a komplexben, így az működésképtelen, ameddig a Bcl-2 foszforilációja nem teszi lehetővé a Beclin-1 csatlakozását a PI3K komplexhez. Az AMBRA1 és a Bif-1 segíti a Beclin-1 asszociációját a komplexhez (Reggiori 2012). A BH3 doménnel rendelkező fehérjék, mint a Bad, a Beclin-1-gyel kompetitív módon kötődnek a Bcl-2/Bcl-X_L-hez (Yang és mtsai 2010). A JNK1 foszforilálja a Bcl-2-t, így szabaddá téve a Beclin-1-et (Reggiori 2012). A DAPK pedig a Beclin-1 foszforilációja révén képes aktiválni azt (Yang és mtsai 2010).

Az autofágia útvonalon belüli negatív visszacsatolásként értelmezhető, hogy az ATG8 homológok az ULK1-hez kapcsolódva, az ULK1-ATG13 komplex autofág lebontását idézik elő (Kraft és mtsai 2012).

Az autofágia szoros kapcsolatban van a szekréciós és endocitotikus folyamatokkal: az ER, a Golgi, a szekréciós vezikulák, az endoszómák membránjai és a plazmamembrán egyaránt részt vesz a fagofór kettős membránjának kialakításában. Ezen membránmozgások szabályozásában kulcsszerepet játszanak egyes monomer GTP-ázok, mint a Rab, Ral, Arf, Arl, Sec családok tagjai (Bento és mtsai 2013).

Az autofágia fehérjék szabályozását számos poszt-transzlációs módosítás (foszforiláció, lipidáció, illetve acetiláció), valamint komplexek képződése biztosítja. Például tápanyaghiány-

ben az ATG5, ATG7, LC3 és ATG12 fehérjék p300 általi acetilációja gátolja, tápanyag hiány esetén a SIRT1 általi deacetilációja stimulálja az autofágiát (Boya és mtsai 2013).

1.8.2. Az autofágia transzkripcionális szabályozása

Az autofágia transzkripcionális szabályozásában érintett transzkripciós faktoroknak csak egy kis részéről van részletes irodalmi adat. DNS károsodás esetén a p53 az AMPK stimulálása, és a DRAM transzkripciós aktivációja révén serkenti az autofágiát (Yang és mtsai 2010). Több autofág gén, az *UVRAG*, a *WIPI*, az *LC3B* és a *vATPáz* promóterén található CLEAR (*Coordinated Lysosomal Expression and Regulation*) szegmenshez kötődik a TFEB, melynek hatását ChIP-Seq kísérletekkel igazolták (Settembre és mtsai 2012). A TFEB túlexpressziója HeLa sejtekben 11 autofágiával kapcsolatos gén, legnagyobb mértékben a *WIPI*, az *LC3B*, a *p62*, a *VPS11*, a *VPS18* és az *ATG9B* expresszióját növelte. Tápanyagbőség esetén a foszforilált TFEB az mTORC1-hez kötődik, és az mTOR gátlása esetén jut a sejtmagba, ahol aktiválja a felsorolt gének transzkripcióját (Settembre és mtsai 2011). A TFEB nukleáris lokalizációját a lizoszomális aminosav szinttől függően a Rag GTP-áz is szabályozza, ez a mechanizmus fontos a lizoszómák autofágia melletti regenerációjában (Peña-Llopis és mtsai 2011). A TFEB az mTOR komplex tagjainak a transzkripcióját is befolyásolja. A tápanyagok szintjének függvényében az ERK1/2 foszforilálja a TFEB-t, megakadályozva sejtmagi lokalizációját. A TFEB siRNS-sel történő gátlása megghiúsította az éhezéskor jellemző expressziós mintázat kialakulását (Settembre és mtsai 2012).

A FoxO transzkripciós faktor család tagjai, elsősorban a FoxO1 és a FoxO3 szintén autofágia serkentők (Hamacher-Brady 2012; Lavallard és mtsai 2012). Jó tápanyag ellátottság esetén, az inzulin útvonal hatása révén az Akt/PKB foszforilálja őket, így inaktív állapotban, a citoplazmában találhatók. Inzulin hatás alatt a Vps34, az ATG12 és a Gabarapl1 expressziója csökken, és ez a hatás FoxO1 túlexpresszállásával megszüntethető (Lavallard és mtsai 2012). A FoxO3-at éhezés esetén az AMPK általi foszforiláció aktiválja (Salih és mtsai 2008). Stressz esetén a FoxO1 acetilálódik, és az ATG7-tel kölcsönhatva serkenti az autofágiát (Lavallard és mtsai 2012). A FoxO3 számos autofágiában érintett gén transzkripcióját segíti, mint a *Beclin-1*, az *LC3*, a *GABARAPL1*, az *ATG4B*, az *ATG12*, a *Vps34* és a *Bnip3* (Hamacher-Brady

2012; Lavallard és mtsai 2012; Mammucari és mtsai 2007; Salih és mtsai 2008; Sengupta és mtsai 2009; Xu és mtsai 2011). A FoxO transzkripciós faktorokat a JNK útvonal gátolja, míg az apoptózist elősegítő Bim-et aktiválja. Ez a mechanizmus része az autofágia és apoptózis közti váltásnak (Xu és mtsai 2011).

Az SREBP-2 a koleszterin szintjének függvényében aktiválja az autofágiát, ahogy a TFEB vagy a FoxO a szénhidrát szinttől függően. Az SREBP-2 serkenti az *LC3B* és az *ATG4* transzkripcióját, bár a szabályozás pontos mechanizmusa nem ismert (Hamacher-Brady 2012; Seo és mtsai 2011). Az autofágia lipofágiának nevezett fajtája szükséges a sejt lipid raktárainak mobilizálásához.

Megfigyelték, hogy a KEAP1 autofágia során bekövetkező lebontása felszabadítja az NRF2-t a gátlás alól, mely a ROS válasz fő transzkripcionális szabályozója. Ezáltal az autofágia tovább fokozódik (Hamacher-Brady 2012). Hasonló módon, a p62 lebontásában is része van az autofágiának, emiatt figyelhető meg magasabb p62 szint a csökkent autofágiával rendelkező tumorsejtekben (Mathew és mtsai 2009). A *Beclin-1* promóterén NF- κ B kötőhelyet is kimutattak, ennek megfelelően a *Beclin-1* expressziója p65-tel stimulálható (Hamacher-Brady 2012).

1.8.3. Az autofágia poszt-transzkripcionális szabályozása

Számos mikro-RNS képes az autofágia fehérjék expresszióját gátolni, például a miR-376b az *ATG4* és a *Beclin-1*, a miR-630 pedig az *ATG12* és az *UVRAG* mRNS-éhez kötődik. A legtöbb ilyen miRNS az autofágia útvonal korai lépéseit gátolja, feltehetően az autofagoszómák kedvezőtlen felhalmozódásának elkerülése végett. Egyes autofágiát gátló miRNS-ek (pl. miR-30a, miR-34a, miR-101) tumor szupresszor hatást mutatnak, így terápiás felhasználásuk is elképzelhető (Frankel és mtsai 2012).

1.8.4. Az autofágia rendszerszintű megközelítése

Az eddigiekben áttekintettem az autofágiát szabályozó főbb útvonalakat, azonban ezen szabályozók mindegyike más jelátviteli útvonalakkal van kapcsolatban, a teljes szabályozási rendszer tehát sokkal bonyolultabb az eddig elmondottaknál. Egy ilyen bonyolult szabályozási

hálózat vizsgálatához elengedhetetlen az adatok együttes, rendszerszintű vizsgálata. Az autofágia szabályozásáról rendelkezésre álló információk összegyűjtésére és a rendszerszintű elemzés lehetőségének megteremtésére az utóbbi években két adatbázis született. A legkorábban publikált HADb (*Human Autophagy Database*) egy 234 autofágiában szerepet játszó gén expressziójának monitorozására alkalmas *microarray* kifejlesztéséhez kapcsolódik. Ez az adatbázis rövid leírást, a nagy gén- és fehérje adatbázisokra mutató hivatkozásokat, szekvencia adatokat tartalmaz a kiválasztott génekről. Meglepő módon interakciókat nem tartalmaz (Moussay és mtsai 2011).

A 2011-ben publikált ADB (*Autophagy Database*) (Homma és mtsai 2011) nem csak ebben a tekintetben lépi túl az HADb-t, hanem az emberen kívül 40 másik fajban azonosít ortológokat, valamint szekvencia alapján is kereshető. Az adatbázis hasznos funkciója az ortológok és homológok részletes, összehasonlító listázása, a szinonimák megfeleltetése, és a hivatkozások a nagy genom adatbázisokra. A kereső szintén ismeri a szinonimákat, így nem a felhasználóra hárul a megfelelő elnevezés kikeresése. Az ADB adatbázis ember esetében 206 fehérjéről tartalmaz adatokat, melyeket autofágiában érintett, vagy autofágiát szabályozó fehérjéknek tekint. Néhány, többnyire az autofágia végrehajtásában közvetlenül résztvevő fehérje esetében az interakciós partnerek listáját is tartalmazza. A kapcsolatok listája azonban a weboldalon csak fehérjénként külön böngészhető – az összesen 641 humán fehérje-fehérje interakció letöltése és egy adatsorrá alakítása informatikai tudást igényel. További probléma, hogy az interakciók forrása, valamint mind a fehérje lista, mind az interakciók gyűjtési köre ismeretlen, így az interakciós adatok felhasználása reprodukálható elemzésekben nehézségekbe ütközik.

Az autofágiában érintett fehérjék közti kapcsolatok legnagyobb gyűjteményét egy ChIP-MS kísérleten alapuló tanulmány kínálja (Behrends és mtsai 2010). Ez azonban kizárólag a ChIP-Seq módszerrel kimutatható 751 interakciót tartalmazza, 409 fehérje között. Az adatsor kiválóan alkalmas további bioinformatikai elemzésre, azonban nem böngészhető, kereshető és vizualizálható informatikai tudás hiányában.

A fentiekből kiderül, hogy nem áll rendelkezésre olyan adatforrás, amely tartalmazná az autofágia gének és szabályozóik közti kölcsönhatások irodalmi gyűjtésből származó listáját. Ugyanakkor az irodalmi adatok és a hatékony *high-throughput* módszerek – mint a ChIP-Seq

– eredményeinek egyesítése biztosítaná a legteljesebb adatsort. A szabályozás mechanizmusainak minél teljesebb körű feltérképezéséhez azonban érdemes felhasználni az általános fehérje interakciós adatbázisban elérhető további kapcsolati adatokat, melyek szintén nagy teljesítményű módszerekből, vagy irodalmi gyűjtésből származnak; a transzkripcionális szabályozásról rendelkezésre álló adatokat; valamint az érintett fehérjék expresszióját szabályozó miRNS-eket. A mindezen adatokat egységes formában tartalmazó adatbázisnak alkalmasnak kell lennie bioinformatikai elemzésekre, és emellett kutató biológusok számára gyorsan és könnyen használható, lexikon szerű adatforrásként kell funkcionálnia.

2. Célkitűzések

A jelátvitel és a molekuláris szabályozási hálózatok tanulmányozásának előbbiekben bemutatott új igényeinek kielégítését, valamint a modern kísérletes módszerek által szolgáltatott adatok lehető leghatékonyabb felhasználását szem előtt tartva, célul tűztem ki a SignaLink 1 jelátviteli adatbázis többirányú kibővítését. Ez a bővítés az alábbi elemeket foglalja magában:

- a SignaLink 1 adatbázis irodalmi gyűjtés alapján felépített jelátviteli hálózatának kiegészítése más adatbázisokból származó poszt-transzlációs szabályozási kapcsolatokkal;
- domén-motívum interakciók alapján további poszt-transzlációs szabályozók prediktálása;
- predikció készítése a *high-throughput* módszerekkel nyert, irányítatlan kapcsolatok irányának megállapítására, domén-domén interakciók alapján;
- az előző pontokban felvázolt fehérje interakciós hálózat transzkripcionális, valamint; poszt-transzkripcionális (mikro-RNS-ek általi) szabályozásának beépítése az adatbázisba, a mikro-RNS-ek transzkripcionális szabályozásával együtt;
- az előző pontokban leírt szabályozási hálózat tárolására alkalmas, hatékony és bővíthető adatbázis struktúra létrehozása;
- az adatok böngészésre lehetőséget biztosító, modern és interaktív weboldal kifejlesztése;
- az adatok hozzáférhetővé tétele bioinformatikai elemzések céljából, több szabványos formátumban, az adatok szűrését megkönnyítő testreszabható letöltő felülettel.

Doktori munkám második részének célja egy nagy jelentőséggel bíró transzkripciós faktor, az NRF2 szabályozási hálózatának felépítése. Ennek megvalósítása érdekében az alábbi célokat tűztem ki:

- egy irodalmi gyűjtésen alapuló, más adatbázisokból és predikciókból származó poszt-transzlációs, transzkripcionális és poszt-transzkripcionális szabályozókat magában foglaló adatbázis létrehozása;

- az így létrejött, az NRF2 szabályozását leíró hálózat összekapcsolása az általam készített Signalink 2-ben szereplő jelátviteli útvonalakkal;
- az adatok böngészését és letöltését biztosító weboldal és letöltő modul elérhetővé tétele.

Munkám harmadik részében, hasonló módszertani alapokra építkezve, egy a kutatások fókuszpontjában álló, jelenleg is számos megválaszolatlan kérdést tartogató, és több betegség patomechanizmusában kulcsszerepet játszó sejttani folyamat, az autofágia szabályozásának feltérképezését, és a jelátviteli hálózattal való összekapcsolását tűztem ki célul. Ennek megvalósítása során célom volt:

- az autofágia végrehajtásában részt vevő fehérjék és szabályozók irodalmi gyűjtéssel összeállított hálózatának kiegészítése más adatbázisokból és predikciókból származó poszt-transzlációs szabályozókkal;
- az így elkészült fehérje interakciós hálózat transzkripcionális és poszt-transzkripcionális szabályozóinak beépítése az adatbázisba;
- az adatok böngészését és letöltését biztosító weboldal és letöltő modul elérhetővé tétele.

3. Módszerek

3.1. A fejlesztések kiindulópontjai

A dolgozatomban bemutatott három adatbázis mindegyike irodalmi gyűjtés köré épül. Ezt az ARN esetében kutatócsoportunk tagja, Földvári-Nagy László végezte, kimondottan az adatbázis létrehozása érdekében.

A másik két adatbázis, a Signalink 2 és az NRF2ome nem előzmények nélkül született, hanem egy-egy korábban kutatócsoportom tagjai és témavezetőm, Prof. Csermely Péter által létrehozott adatforrás továbbfejlesztése.

A Signalink 2 fejlesztésének kiindulópontja az előző publikált verzió, a Signalink 1 (Korcsmáros és mtsai 2010). A Signalink 1 nyolc jelátviteli útvonal 3 fajból származó kapcsolatainak irodalmi gyűjtése, mely az egyes útvonalakat receptoroktól és ligandumoktól az első érintett transzkripciós faktorokig foglalja magában.

Az NRF2ome előzménye az NRF2 szabályozási hálózat (Papp és mtsai 2012). Az NRF2 és a KEAP1 közvetlen szabályozóit tartalmazó hálózat majdnem teljesen csillagpontos, azaz a kapcsolatok többségét az NRF2 partnerei alkotják. Ugyanakkor tartalmaz néhány, szubjektív szempontok alapján fontosnak ítélt további elemet is. Papp és mtsai mellékleteként, hat külön táblázatban érhető el. A hálózat 289 fehérje-fehérje interakcióból, 7.469 NRF2 által transzkripciósan szabályozott génből, és 63, az NRF2 transzlációját gátló mikro-RNS-ből, és 35, a miRNS génekre ható transzkripciós faktorból áll.

A továbbiakban áttekintem az adatforrások legfontosabb jellemzőit, melyek többsége mindhárom adatbázishoz felhasználásra került, majd feldolgozásuk és az új adatbázisok összeállításának módszertanát ismertetem.

3.2. Adatforrások

A Signalink 2, az NRF2ome és az ARN magját az irodalomból történő kézi gyűjtéssel összeállított adatsor adja. Ezen kívül az adatok származhatnak más adatbázisokból vagy predikcióból.

3.2.1. Kézi gyűjtés

A feldolgozott cikkek PubMed azonosítóit tároljuk az adatbázisban, és az egyes interakciókhoz hozzá vannak rendelve a cikkek, melyekben az adott interakciót leírták. Kutatócsoportunk irodalmi gyűjtést végző tagjai a PubMed adatbázisban keresték a fehérjék közötti kapcsolatokat leíró tanulmányokat. A Signalink 1 adatbázis esetében mind a nyolc útvonal, mind a három fajban (*C. elegans*, *D. melanogaster* és az ember) azonos módon, de egymástól függetlenül került feldolgozásra (Korcsmáros és mtsai 2010). A Signalink 2 készítésekor ez az irodalmi gyűjtés az időközben eltelt két évben megjelent cikkekből frissítésre, valamint az útvonalak tagjainak endocitotikus- és állvány-(*scaffold*-)fehérjékkel való kapcsolataival bővítve lett. Továbbá két útvonal, a MAPK és az IGF összevonásra került, és a Signalink 2-ben RTK (receptor-tirozinkináz) néven, egy útvonalként szerepel.

Az NRF2ome magját egyetlen kézi gyűjtés alkotja, mely az NRF2-vel közvetlenül kapcsolatba kerülő fehérjékre terjed ki. Tartalmaz továbbá néhány fontosnak tűnő kapcsolatot ezen fehérjék között, illetve az NRF2 legfontosabb szabályozójának, a KEAP1-nek néhány fontos partnerét. Az így elkészült adatsor összesen 108 fehérje 146 interakcióját tartalmazza (Papp és mtsai 2012).

Az ARN adatbázis szintén fehérje-fehérje interakciók irodalmi gyűjtésén alapul, mely az autofágia végrehajtásában közvetlenül részt vevő 38 fehérjét (*autophagic machinery*) és az ezekkel közvetlenül kölcsönhatásba kerülő további fehérjéket tartalmazza. Az autofágia végrehajtásában közvetlenül részt vevő fehérjékre a továbbiakban mint *autofágia fehérjékre* hivatkozom.

Egyes hivatkozások nem saját kézi gyűjtésből származnak: a felhasznált adatbázisokból minden esetben amikor rendelkezésre álltak, átvettem a hivatkozásokat. A felhasznált adatbázisok publikációinak hivatkozásait szintén hozzárendeltem az adott adatbázishoz, mint adatforráshoz.

3.2.2. Fehérje és mikro-RNS referencia adatbázisok

A referencia adatbázisok a fehérjék és miRNS-ek lehető legteljesebb gyűjteményét kívánják folyamatosan frissített formában közzétenni. A bioinformatikai adatkezelésnél az egységes és jól meghatározott molekula azonosítók miatt a legjobb ezen adatbázisok azonosítóit használni. Munkám során a fehérjék esetében a UniProt, a miRNS-eknél pedig a miRBase azonosítóit használtam.

3.2.2.1. UniProt. Az UniProt (*Universal Protein Resource*) adatbázis (UniProt Consortium 2012) az annotált fehérje szekvenciák legszélesebb körben használt és legteljesebb adatbázisa. Két fő részből áll: a manuálisan ellenőrzött (*reviewed*) adatokat tartalmazó SwissProt, illetve az automatikusan, informatikai módszerekkel bővülő TrEMBL adatbázisokból. Utóbbi számos, szekvenciák alapján feltételezett, de nem biztos hogy létező polipeptidet tartalmaz. Ezeket az adatbázis karbantartói igyekeznek bioinformatikai módszerekkel és manuális ellenőrzéssel kizárni az adatbázisból. Az egyes polipeptid szekvenciákat ún. UniProtKB (*UniProt Knowledge Base*) azonosítóval jelölik, mely hat karakterből (számból és betűből) áll. Az adatok kezelése során minden adatbázis géneket vagy fehérjéket jelölő azonosítóit UniProtKB azonosítóknak feleltettem meg. Amely fehérjéhez nem sikerült UniProt azonosítót rendelni, azt nem tekintettem létező fehérjének, és kizártam a további adatfeldolgozásból.

3.2.2.2. WormBase, FlyBase, Ensembl. A WormBase több *Nematoda* faj között a *C. elegans* genomját és fehérje szekvenciáit tartalmazza, részletes annotációkkal. A géneket WB-Gene kezdetű WormBase ID azonosítók jelölik, melyekre minden *C. elegans* gén esetében található hivatkozás a Signalink 2 fehérje adatlapjain, illetve a Signalink 2 weboldalára a WormBase adatlapjain (Harris és mtsai 2010).

A FlyBase a *Drosophila* fajok elterjedt referencia adatbázisa, melyet szintén kölcsönös hivatkozások kapcsolnak össze a Signalink 2-vel, a *Drosophila* gének FBgn kezdetű azonosítói alapján (Drysdale és mtsai 2008). Az Ensembl adatbázis 70 faj genomját tartalmazza, az emberi fehérjék esetében adatbázisainkban hivatkozunk azok ENSP kezdetű Ensembl Protein azonosítókra (Flicek és mtsai 2013).

3.2.2.3. miRBase. A miRBase a mikro-RNS szekvenciák és annotációik elsődleges referencia adatbázisa. 140 faj több mint 15.000 miRNS génjét, és több mint 17.000 érett miRNS adatait tartalmazza. Ezek egy része manuálisan azonosított, más részük szekvenálásból és *microarray* kísérletek eredményeiből származik. Az MI kezdetű miRBase azonosítókkal, és az itt szereplő standard nevekkel azonosítjuk a miRNS-eket (Kozomara és mtsai 2010).

3.2.3. Fehérje-fehérje interakciós adatbázisok

Adatbázisaink kialakítása során a HPRD-t és a BioGRID-et használtam fel a protein-protein interakciók (PPI) forrásaként, mivel ezek az adatbázisok tekinthetők a különféle módszerekkel felderített interakciók legteljesebb gyűjteményének (Chatr-aryamontri és mtsai 2013; Keshava Prasad és mtsai 2009). Általában a kézi gyűjtésen alapuló hálózatban szereplő fehérjék illetve első szomszédaik közti, valamint a transzkripciós faktorok közti kapcsolatokat válogattam ki ezen forrásokból. A BioGRID-ből és a HPRD-ből származó interakciók irányítatlanok, így ahol csak lehetséges volt, megkíséreltük a bioinformatikai módszerekkel történő irány predikciót.

A **BioGRID** (*Biological General Repository for interaction Datasets*) 209.354 PPI-t és hasonló mennyiségű genetikai interakciót tartalmazó adatbázis (Chatr-aryamontri és mtsai 2013). Az adatok *high-throughput* és hagyományos kísérletekből, összesen 33.858 publikáció feldolgozásából származnak. A BioGRID 50 fajról gyűjt adatokat, melyek között a *C. elegans*, a *D. melanogaster* és az ember is szerepel. A humán PPI-k száma 123.436.

A **HPRD** (*Human Protein Reference Database*) egy faj specifikus fehérje referencia adatbázis, mely a fehérjék sokféle annotációja (pl. foszforilációs motívumok) mellett 38.167 interakciót is tartalmaz (Keshava Prasad és mtsai 2009). Ezek többnyire élesztő két hibrid kísérletekből, *in vivo* illetve *in vitro* vizsgálatokból származnak, jelentős mennyiségű a közösségi együttműködéssel (*Human Proteinpedia*) gyűjtött interakciók száma.

Az **InnateDB** az emlősök veleszületett immunválaszának rendszerszintű elemzését célzó adatbázis. 196.000 kísérletesen igazolt fehérje-fehérje interakciót tartalmaz, emberből, egérből és szarvasmarhából. 18.000 kapcsolat irodalmi gyűjtésből származik (Breuer és mtsai 2013).

Az NRF2ome összeállításakor a BioGRID és a HPRD mellett az InnateDB-t is felhasználtam általános fehérje interakciós adatbázisként.

Az ARN-ben felhasználtam az **ADB** (*Autophagy Database*) és egy autofágiához kapcsolódó fehérjéket vizsgáló CHIP-MS tanulmány adatait (Behrends és mtsai 2010; Homma és mtsai 2011). Ezekről a 1.8.4. fejezetben írok.

3.2.4. A predikciók során felhasznált adatbázisok

A **Pfam** (*Protein Families*) adatbázis az UniProtban megtalálható fehérjék több mint 80 %-ának tartalmazza a domén összetételét (Punta és mtsai 2011). Többszörös szekvencia illesztéseken alapul, melyek a rejtett Markov-lánc modellekkel dolgozó HMMER illesztő algoritmussal készültek. Az adatbázis Pfam-A része több mint 10.000 fehérjéhez tartalmaz manuálisan ellenőrzött domén annotációkat. A Pfam-B szekció egy ADDA nevű, automatikus módszerrel talál doméneket a fehérje szekvenciákban, ezeket akkor érdemes használni, ha az adott szekvenciához Pfam-A domének még nem állnak rendelkezésre. Szakértők határozzák meg, hogy milyen mértékű illeszkedés esetén tartoznak egy családba az egymáshoz hasonló domének, ez családonként eltérő lehet. A családok több mint egy negyede (a domének 6,7 %-a) ismeretlen funkciójú, ún. DUF (*Domain of Unknown Function*).

A domének a fehérjéket felépítő, több fehérjében előforduló, evolúciósan konzervált egységek, a fehérjék harmadlagos szerkezetének elemei. A **DOMINE** 4.036 Pfam domén közti összesen 20.513 domén-domén interakciót tartalmazó adatbázis (Raghavachari és mtsai 2007). Ezek az interakciók összesen nyolc féle predikciós, és két kísérletes forrásból származnak. A kísérletes adatok az iPfam és a 3did térbeli szerkezeteket tartalmazó adatbázisokból valók. A prediktív módszerek sok különböző megközelítést alkalmaznak a domének közti kapcsolatok valószínűségének becslésére.

Az **MLE** (*maximum-likelihood estimation*) módszer Bayes-i statisztika alkalmazásával, ismert PPI-k, szekvenciák és GO annotációk felhasználásával becsüli a domének közti kapcsolatok valószínűségét.

Az **RCDP** (*Relative Co-evolution of Domain Pairs*) elnevezésű módszer minden domén-domén páros szekvenciái esetében megbecsüli a koevolúció mértékét, és ismert PPI-k esetében

feltételezi, hogy azok a domének vesznek részt az interakcióban, melyek leginkább koevolválódnak.

A **P-érték** eljárás szupercsaládonként veszi a doméneket, és az ezen doméneket tartalmazó fehérjék közti interakciók alapján kiszámolja annak a valószínűségét, hogy két szupercsalád tagjai közt lép fel interakció. Ezen valószínűségekből az egyes PPI-k esetében kiszámolja, hogy a fehérje páros mely doménjei vesznek részt legnagyobb valószínűséggel az adott interakcióban.

A **Fusion** nevű módszer azon a megfigyelésen alapul, hogy egyes egymással kapcsolatba kerülő fehérjék homológjai vagy ortológjai néha egy polipeptid lánccá olvadnak össze; így ha a szekvenciákban ilyen összeolvadást tapasztalunk, van rá bizonyos esély, hogy a különálló fehérje pár egymáshoz kapcsolódva funkcionál.

Az **LP** (*Linear Programming*) megközelítésben ismert PPI hálózat és a benne szereplő fehérjék ismert domén összetétele alapján, a parszimónia elvet követve megkeressük a legrövidebb domén-domén interakció listát mellyel magyarázható a hálózatban szereplő interakciók minél nagyobb hányada.

A **DPEA** (*Domain Pair Exclusion Analysis*) minden domén pár esetében megvizsgálja az adott doméneket tartalmazó, egymással kölcsönhatásba kerülő fehérjék arányát, az egyes domének gyakoriságából adódó véletlen előforduláshoz képest. Ezután maximum likelihood módszerrel megnézi, hogy az egyes domén párok kizárása esetén mennyire változik az ismert PPI hálózat domén-domén interakciókkal való magyarázhatósága.

Az **RDF** (*Random Decision Forest Framework*) sok döntési fa modellt generál, és ezek módusát választja ki, mint a legvalószínűbb lehetőséget a tapasztalt mintázat magyarázatára.

A **DIMA** (*Domain interaction Map*) filogenetikai profilok alapján ad becslést az egyes domén-domén interakciók valószínűségére, azaz két doménnek minél inkább hasonló fajokban léteznek ortológjai, annál nagyobb eséllyel van köztük interakció (Raghavachari és mtsai 2007).

A DOMINE adatbázis ezekből a forrásokból általában a predikciók közül csak a legmagasabb pontszámot kapott kapcsolatokat veszi át. Ahogy látni fogjuk, az adatbázisainkban

alkalmazott predikciók nagyban alapoznak az itt felsorolt módszerekkel nyert domén-domén interakciókra, melyeket a DOMINE adatbázisból vettünk át.

A **Negatome** adatbázis egyedülálló módon azokat a fehérje illetve domén párokat gyűjti, melyek kísérletes adatok és térbeli struktúrák alapján nagy valószínűséggel nem kerülnek egymással fizikai interakcióba (Smialowski és mtsai 2009). Összesen 1.892 nem kölcsönható fehérje pár (NIP), és 979 nem kölcsönható domén pár (NID) szerepel a Negatome-ban, melyeket mint negatív referencia adatsort felhasználtunk predikcióinkban.

A fehérjék motívum annotációit az **ELM** adatbázisból szereztem be. A rövid lineáris motívumok (SLiM – *Short Linear Motifs*) a doméneknél és a strukturális motívumoknál rövidebb, 3-11, átlagosan 6 aminosavból álló, evolúciós skálán átlagosan változékonnyabb, egyes esetekben mégis erősen konzervált fehérje szekvencia szakaszok. Sok motívumnak központi szerepe van számos fehérje-fehérje interakcióban (Ren és mtsai 2008). Az ELM (*Eukaryotic Linear Motif*) adatbázis több mint 1.800 motívumot tartalmaz, melyek funkciója valamely fehérje esetében már kísérletes igazolást nyert. Az ELM képes bármely fehérje szekvenciában motívumokat keresni, adatokat tartalmaz ezek funkciójáról és evolúciós konzerváltságáról. Funkciójuk alapján 170 osztályba sorolja be őket, valamint szerkezeti jellemzők alapján megpróbálja kizárni ezen rövid szekvenciák véletlenszerű, de nem funkcionáló előfordulásait (Dinkel és mtsai 2013).

A **Reactome** – melyet a Bevezetőben az útvonal adatbázisok közt említettem – molekulák közti interakciók adatbázisa, mely 7.088 fehérje 6.744 interakcióján kívül tartalmaz számos kis molekulával fellépő reakciót is, részletgazdag annotációkkal (Croft és mtsai 2014; Vastrik és mtsai 2007). A PPI-k cikkekből történő gyűjtése és feldolgozása szakértők munkája. A kölcsönhatások típusa és iránya ily módon többnyire ismert, ezért használtuk a Reactome-ből származó adatsort, mint pozitív standardot az ismeretlen irányú PPI-k irány predikciója során.

3.2.5. Predikció domén-motívum kölcsönhatások alapján

A motívumok más fehérjék aktivitásának célpontjai lehetnek, azonban a motívum szekvencia megléte nem elegendő, hogy feltételezzük ezen interakciók létrejöttének lehetőségét. Ha a motívum a fehérje térszerkezetéből adódóan olyan pozícióba kerül, hogy a vele reakcióba

lépő fehérjék nem férhetnek hozzá, nem tekinthető funkcionálisnak. A már említett ELM adatbázishoz kapcsolódóan készült egy *Structure Filter* nevű algoritmus, mely a szerkezeti jellemzők figyelembe vételével becslést ad a motívum funkcionalitására (Via és mtsai 2009). A *Structure Filter* két értéket ad meg a fehérje térszerkezete alapján: az egyik a hozzáférhetőséget méri (*Accessibility Score*), azaz a motívum kitétségét az oldat, a fehérje környezete felé; a másik érték (*Secondary Structure Score*) a másodlagos szerkezet jellemzőit veszi figyelembe, azaz hogy a motívum hélixben, béta redőben, vagy rendezetlen szakaszban helyezkedik el. Jellemzően a rendezettebb másodlagos szerkezeti elemeken belül a motívumok kisebb valószínűséggel vesznek részt interakciókban, mint a rendezetlen szakaszokban.

Az ELM *Structure Filter* szolgáltatása nyilvánosan elérhető, így az algoritmus saját implementálására nem volt szükség, csupán a fehérje szekvenciák alapján indítottunk lekérdezéseket az ELM szerver felé (Dinkel és mtsai 2013; Via és mtsai 2009).

A domén-motívum interakciók alapján készült fehérje-fehérje kapcsolatokra, a fehérjék Gene Ontology *Biological Process* tulajdonságai alapján szemantikus hasonlósági indexet számoltunk (Alvarez és mtsai 2011; Ashburner és mtsai 2000. – a szemantikus hasonlósági indexről bővebben a 1.6.2.2. fejezetben írok). ROC (*Receiver Operating Characteristic*) elemzéssel – Módos Dezső segítségével – fajonként meghatároztuk a kritikus értéket, mely alatt kizártuk a kapcsolatokat, így csökkentve a hamis pozitívok arányát.

3.2.6. Predikció domén-domén kapcsolatok alapján

A fehérjék közti kölcsönhatásokban gyakran a fehérjéket alkotó domének vesznek részt, az adott doménre jellemző reakciótypus vagy kötődési preferenciák által meghatározott módon. A rendelkezésre álló kísérletes adatok többnyire fehérjék közti interakciókról szólnak, ugyanakkor amennyiben ismerjük, hogy mely domének vesznek részt ezekben az interakciókban, és mely más fehérjékben fordulnak elő ezek a domének, valószínűsíthetjük, hogy ezen domének más fehérjékben is kölcsönhatásba lépnek egymással. Ezen elképzelés szerint egyrészt a fehérjék domén összetétele, másrészt a domének közti interakciók ismeretében következtethetünk fehérjék közti interakciókra. A domének közti kapcsolatokról azonban kevés kísérletes adat áll

rendelkezésre, ezeket többnyire a fehérjék közti interakciók és a fehérjék domén összetétele alapján, *in silico* predikcióval igyekeznek kideríteni (Chen és mtsai 2005).

A domén-domén kapcsolatokon alapuló predikciót adatbázisainkban egyrészt létező, de ismeretlen irányú kapcsolatok irányának becslésére, másrészt új fehérje-fehérje kapcsolatok predikciójára használtuk fel.

3.2.6.1. Kapcsolat predikció. A Pfam (Punta és mtsai 2011) és a DOMINE (Raghavachari és mtsai 2007) adatbázisok adatainak kombinálásával létrehoztunk egy adatsort az egymással kölcsönható doméneket tartalmazó fehérje párokról. A Negatome adatai alapján kizártuk a nem kölcsönható domén-domén párokat. Ezután a domén-motívum alapú predikcióhoz hasonlóan jártunk el: a GO szemantikus hasonlósági index mentén, ROC analízissel csökkentettük a hamis pozitív kapcsolatok arányát.

3.2.6.2. Irány predikció. Az általános PPI adatbázisokból származó interakciók általában nem tartalmazzák az irányt, a két fehérje kölcsönhatását szimmetrikusan írják le. Ugyanakkor az irány – például A foszforilálja B-t, és nem fordítva – fontos információ a jelátviteli hálózatokban, mely nélkül az információ áramlás irányának megállapítása, és a jelátvitel modellezése lehetetlen (Liu és mtsai 2009). Az irány az interakció típusának, vagy a résztvevő domének ismeretében meghatározható lenne, azonban a *high-throughput* technikák erről nem szolgáltatnak adatot. Az ilyen interakciónál a domén-domén kölcsönhatások alapján megkíséreltük az irány predikcióját. A Signalink 2 adatbázis esetében a kapcsolatokat, melyekhez sikerült megbízható módon irányt rendelni, egy külön rétegbe soroltuk, míg a továbbra is ismeretlen irányú kapcsolatokat a másik rétegbe.

Az irány predikcióhoz a DOMINE adatbázisból (Raghavachari és mtsai 2007) beszerzett domén-domén kapcsolati adatokat, illetve a Pfam adatbázisból származó (Punta és mtsai 2011), a fehérjék domén összetételére vonatkozó adatokat, valamint a Reactome-ből letöltött, ismert irányú interakciókat használtuk fel (Croft és mtsai 2014). A predikciót gépi tanuló algoritmussal (*Machine Learning – ML*) végeztük, a következőképpen: az említett három adatbázis adatait kombinálva kiszámoltuk, hogy az egyes domén-domén kapcsolatok különböző

irányú interakciókban milyen arányban fordulnak elő. Az 1. képlet szerint (Liu és mtsai 2009; Rhodes és mtsai 2005) kiszámoltuk az F értékét az előforduló domén-domén interakciókra. Látható, hogy amennyiben az adott kapcsolat nagyobb arányban fordul elő $m \rightarrow n$ irányban $F > 0$, míg ha az $n \rightarrow m$ irány a jellemző, $F < 0$ lesz. Ábrázoltuk az így kapott pontszámok ROC (Receiver Operating Characteristic) görbéjét, és ez alapján határoztuk meg az F értéket, mely felett az irány predikciót elfogadhatónak tekintjük.

$$F(d_{mn}) = \frac{\text{PR}(d_m \rightarrow d_n) - \text{PR}(d_n \rightarrow d_m)}{\text{PR}(d_m) \times \text{PR}(d_n)} \quad (1)$$

3.2.7. Transzkripciós faktor–gén regulációs adatbázisok

A TF–promóter kapcsolatokról kísérletes adatokat tartalmazó adatbázisok közül letöltöttem az ABS, a DroiDB, az edgeDB, az ENCODE, a RedFly, az ORegAnno, a PAZAR és a wTF weboldaláról a megfelelő adatsorokat. A különböző azonosítók elsődleges UniProtKB azonosítóra való lefordítása után kiválasztottam a fehérje-fehérje interakciós rétegekben szereplő fehérjéket szabályozó transzkripciós faktorokat. Ezután minden kapcsolatot forrás megjelöléssel ellátva bevitettem az adatbázisokba.

A JASPAR módszer szerinti predikciót kutatócsoportunk végezte. Az egyes transzkripciós faktorokhoz tartozó, JASPAR-ból származó PWM mátrixok illeszkedését megvizsgáltuk az Ensembl adatbázisból származó minden gén start kodonját megelőző 2.000, illetve stop kodonját követő 2.000 bázispár hosszúságú szekvencián. A mátrix hosszának megfelelő ablakot nukleotidoként csúsztattuk a vizsgált szakaszon, figyelve az illeszkedés mértékét. Az illeszkedést az egyes szakaszok összes nukleotidjának a mátrix adott pozíciójában megfelelő értékek szorzatával mértük, melynek aztán a logaritmusát vettük, és levontuk belőle a szekvencia hosszának (n) hatványára emelt $1/4$ logaritmusát, ami a nukleotidok egyenlő, véletlenszerű előfordulásának valószínűsége.

$$\text{JASPAR} = \log \prod_{i=1}^n p_i - \log(0, 25^n) \quad (2)$$

Ez a mérőszám akár kis számú, az illeszkedést rosszul befolyásoló nukleotid esetén is nagyon kis értéket adott, lehetővé téve a legjobb illeszkedések kiszűrését. A kritikus érték meghatározása kutatócsoportunk tagja, Módos Dezső közreműködésével történt, aki az értékekre két Gauss-görbe összegét illesztette. Feltételeztük, hogy a rossz, véletlenszerű illeszkedések egy alacsonyabb várható érték körül, míg a jó, funkcióképes kötőhelyek egy magasabb várható érték körül log-normális eloszlást alkotnak. A két illesztett görbe alapján ROC analízist végeztünk, és a kritikus értéket 10-nél állapítottuk meg, mely esetben a hamis pozitívok arányának várható értéke 0,01. A TF-promóter szabályozási kapcsolatokkal együtt az illeszkedés jóságát kifejező mérőszám értékét is beépítettem az adatbázisokba.

3.2.8. mikro-RNS–mRNS interakciós adatbázisok

A miRNS–mRNS interakciókat a microT v4, a miRanda, a miRDB, a PicTar és a TargetScan predikciós adatbázisokból és a miR2Disease, a miRDeathDB, valamint a TarBase kísérletes adatokat tartalmazó adatbázisokból vettem át. Az adatbázisokról részletesen írtam a [1.6.5.](#) fejezetben, és a kapcsolódó publikációkra is ott hivatkozom. Az azonosítókat elsődleges UniProtKB azonosítóknak, és érett miRNS-ek miRBase azonosítóinak feleltettem meg. Ezután a fehérje-fehérje interakciós rétegekben szereplő fehérjéket szabályozó miRNS-eket beépítettem az adatbázisba. A miR2Disease és a miRDeathDB esetében a kapcsolatokhoz tartozó publikációk PubMed azonosítóit, a predikciós adatok esetében pedig a predikció jóságát kifejező mérőszámokat is bevitettem a megfelelő táblákba.

3.2.9. Transzkripciós faktor–miRNS gén regulációs adatbázisok

Az ENCODE, a PuTmiR és a TransmiR adatbázisokból származó kapcsolatokat az azonosítók UniProtKB illetve miRBase azonosítókra történő lefordítása után, a miRNS–mRNS interakciók rétegében szereplő miRNS-ek és transzkripciós faktoraik közti kapcsolatokat hozzáadtam az adatbázisokhoz. A PuTmiR esetében a predikcióból származó megbízhatósági mérőszámokat is beépítettem az adatbázisokba. Ezen adatbázisok leírása és hivatkozásai megtalálhatók a [1.6.6.](#) fejezetben.

3.3. Az interakciók tulajdonságai

3.3.1. Irány, hatás, közvetettség

Az irodalomból gyűjtött kapcsolatoknál legtöbbször rendelkezésre áll a kölcsönhatás iránya – pl. A foszforilálja B-t. A más adatbázisokból átvett, nagy teljesítményű módszerekkel felderített kapcsolatoknál ez általában nem ismert. Az irány az adatbázisokban az `interaction.is_directed` mezőjében szerepel, melynek értéke 0 az irányítatlan, 1 az irányított, és 2 amennyiben domén-domén alapú predikció segítségével sikerült az interakció irányát megállapítani. A TF–promóter és a miRNS–mRNS interakciónál az irány eleve adott. Az irányított interakciók lehetnek gátló vagy serkentő hatásúak. Ezt az információt az `interaction.is_stimulation` mező tartalmazza, melynek értéke 1 serkentő, –1 gátló, és 0 az ismeretlen hatású interakciók esetén. A harmadik tulajdonság az indirekt kapcsolatok megkülönböztetésére hivatott. Az `interaction.is_direct` mező értéke 1 közvetlen kapcsolat esetén, és 0 közvetett – például TF–promóter – interakciónál.

Két fehérje között egy rétegben több interakció is lehetséges, például ha ezek fordított irányúak, vagy különböző hatásúak. Ugyanazon két fehérje között több rétegben is lehetnek kapcsolatok. Mivel az adatbázisok kialakítása során szerettem volna megőrizni az adatforrások információ tartalmát, elképzelhető, hogy az egyik rétegben irányított, míg a másokban irányítatlan kapcsolat szerepel ugyanazon két fehérje között. Ez az adatstruktúra – mely a weboldalakon is megjelenítésre kerül – lehetővé teszi a felhasznált források összehasonlító elemzését.

3.3.2. Megbízhatósági mérőszámok

3.3.2.1. PRINCESS. A PRINCESS komplex mérőszám háttéréről a 1.6.2.1. fejezetben részletesen írtam (Li és mtsai 2008). A módszer alkalmazását eddig csak humán adatokra dolgozták ki, így adatbázisainkban csak ember fehérje interakciónál található PRINCESS értékek. A számításokat Dong Li kutatócsoportja végezte, a Pekingi Proteom Kutatóközpontban.

3.3.2.2. A Gene Ontology tulajdonságok szemantikus hasonlósága. A GO tulajdonságok szemantikus hasonlóságának számítása során egy tisztán gráf alapú indexeket választot-

tunk, hiszen bár maga a GO tulajdonságok fája is folyamatosan fejlesztés alatt áll, az egyes génekhez kapcsolt annotációk mindig aktuális biológiai tudásunkat tükrözik, melynek szintje ráadásul fajoként eltérő lehet.

Az Alvarez és mtsai által leírt gráf alapú módszert Fazekas Dávid és én Python nyelven implementáltuk, majd a Gene Ontology weboldaláról letöltött adatok alapján (Ashburner és mtsai 2000) kiszámoltuk a *Biological Process* domén egyes terminusai közti hasonlósági értékeket. Fehérjék esetén mindkét fehérje összes GO tulajdonsága közti hasonlóságok mediánját vettük a két fehérje közti kapcsolat GO szemantikus hasonlóság értékének, ezek a számok szerepelnek az adatbázisokban. Ezenkívül, a szemantikus hasonlóság adatok mentén Módos Dezső által elvégzett ROC (*Receiver Operating Characteristic*) elemzéssel meghatározott kritikus értékek alatti részt kizártuk a domén-domén és domén-motívum kapcsolatok alapján készült predikciók eredményéből, így csökkentve a hamis pozitívok arányát.

3.3.2.3. Predikciók megbízhatósági értékei. A miRNS-ek által történő poszt-transzkripcionális szabályozás, valamint a JASPAR és a PuTmiR transzkripciók kapcsolatainál többféle mérőszám szerepelhet. Ezek minden esetben a predikciós algoritmusok eredményei, melyekről adatforrásonként külön írtam a 1.6.5. fejezetben. A predikciókat tartalmazó adatbázisok rendszerint alkalmaznak valamilyen határértéket, mely a hamis pozitívok kiszűrését szolgálja. Az adatbázisok építése során esetenként ennél szigorúbb határértékeket szabtam meg, és minden mérőszámnál lehetőséget biztosítok a felhasználó számára, hogy a weboldal letöltő felületén még szigorúbb szintre emelje a predikciók szűrését.

3.4. Implementáció

Az adatbázisok kezeléséhez MySQL adatbázis szerveret használok, melyet Linux operációs rendszeren futtatok. A weboldal elkészítéséhez PHP-t, jQuery-t, jQuery UI-t, valamint Cytoscape Webet használtam, az export modul pedig egy különálló, Python nyelven írt program.

3.4.1. Adatbázis struktúra

A MySQL szerveren futó adatbázis tábláit MyISAM tároló motor kezeli. A weboldal és az export modul kérdezi le az adatokat, így a táblák indexelését az itt előforduló lekérdezésekre optimalizáltam. Az indexelés akár többszörösére gyorsíthatja az adatok elérését. Az általam alkalmazott indexekben általában a JOIN ON, GROUP BY és WHERE záradékokban szereplő mezők, illetve a tábla elsődleges kulcsa szerepel.

Az adatbázis tábláinak felépítését, és egymás közti hivatkozásait a 8. ábrán követhetjük. Fehérje-fehérje interakciós adatbázisaink központi táblája az `interaction` tábla, mely az összes kapcsolatot tartalmazza. Az `interaction.source` és `interaction.target` mezők az interakció forrását és célját jelölik, a `protein` és a `mirna` táblákra hivatkozva. További fontos mezők az `interaction.is_direct`, `interaction.is_directed` és az `interaction.is_stimulation`: ezek a kapcsolat alapvető tulajdonságait, közvetlenségét, irányítottságát, illetve serkentő vagy gátló jellegét írják le. Az `interaction.layer` mező a kapcsolatok kategóriáit jelöli, értéke jelenleg 0 és 7 közötti szám. A 4-es érték transzkripciós faktor és gén közti kapcsolatot jelent. Az 5. réteg esetében az `interaction.source` oszlop csak a miRNS-eket tartalmazó `mirna` táblára hivatkozhat, míg a 7. réteg esetében ez a `interaction.target` mezőre igaz. Az összes többi esetben ezek a mezők a fehérjéket tartalmazó `protein` táblára hivatkoznak. Az `interaction.layer` mező értéke azonban egyértelművé teszi, hogy az interakció melyik résztvevője fehérje, és melyik miRNS.

A `protein` és a `mirna` táblák tartalmazzák az interakciós hálózat csomópontjait, a fehérjéket valamint miRNS-eket. Előbbiek UniProtKB azonosítója, utóbbiak neve és miRBase azonosítója szerepel a táblákban. A `pathway` és a `topology` táblában útvonalak (jelenleg hét), és topológiai tulajdonságok (lehetséges értékek: ligand, receptor, mediátor, kofaktor, állványfehérje (*scaffold*), transzkripciós faktor, endocitotikus protein) vannak felsorolva. Az egyes fehérjék útvonal besorolásait és topológiai tulajdonságait a `protein_pathway` és a `protein_topology` táblák tartalmazzák.

Hasonló módon, az `interaction_source`, az `interaction_reference`, valamint az `interaction_weight` táblák a `source`, `reference` és a `weight` táblákra hivatkozva, az interakciók tulajdonságait tartalmazzák.

A `reference` tábla a tábla az általunk, vagy mások által manuálisan gyűjtött kapcsolatok hivatkozásait tartalmazza, PubMed ID-kal jelölve a cikkeket. Az `interaction_reference` tábla kapcsolja össze az egyes interakciókat és hivatkozásokat, az `interaction_source` pedig a forrás adatbázisok hivatkozásait tartalmazza.

A `name` tábla a fehérjék UniProt adatlapján található azonosítóit és neveit tartalmazza. Egy fehérjére több sora is hivatkozik, és szerepel benne a név típusa. A weboldalon működő kereső ezen elnevezések és azonosítók között keres.

3.4.2. Az adatbázisok építése

3.4.2.1. Adatbázis azonosítók fordítása. A más adatbázisokból letöltött adatsorok a fehérjéket illetve géneket különféle azonosítókkal jelölik. Ezek közül leggyakoribbak az Ensembl Protein illetve Gene ID, az Entrez Gene ID, a HGNC Gene Symbol, az NCBI RefSeq ID, valamint az UniProtKB ID (Koh és mtsai 2012). A gének, a róluk átíródó mRNS-ek és az utóbbiak alapján képződő fehérjék az adatbázisban azonos entitásként vannak reprezentálva, azaz a hálózatban a gén, melynek átíródását transzkripciós faktorok szabályozzák, az mRNS, melynek translációját miRNS-ek gátolják, illetve a fehérje, mely más fehérjékkel kerül interakcióba egyazon pontnak felel meg az interakciós hálózatban.

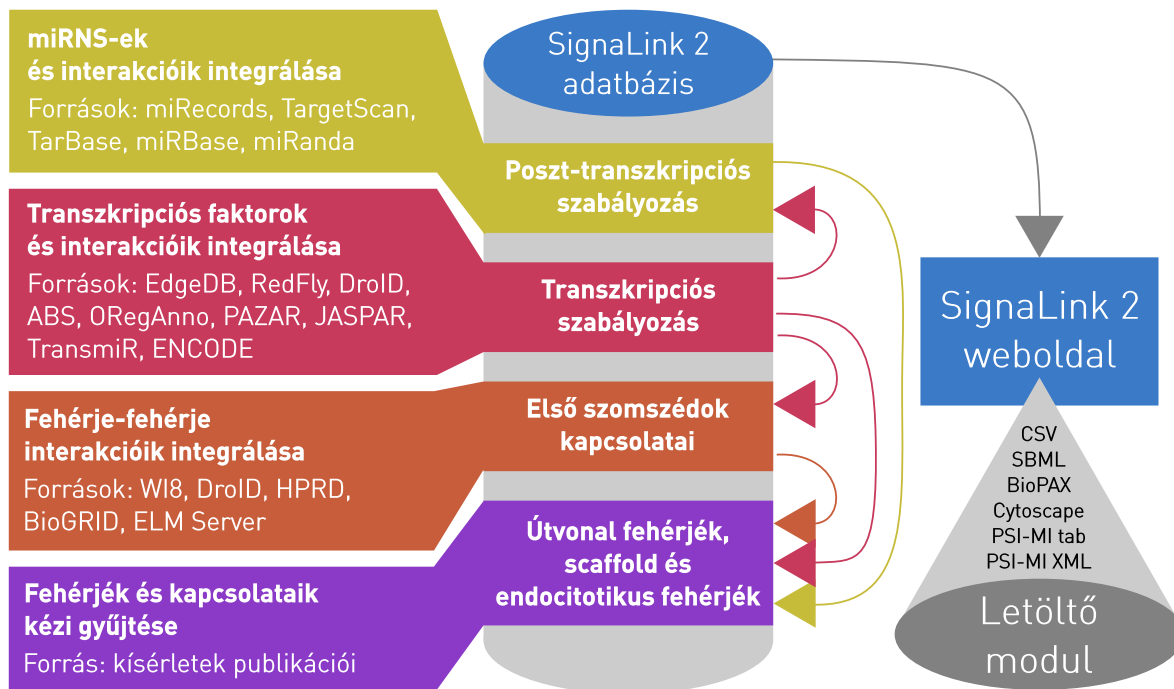
Az UniProt weboldaláról (uniprot.org) letöltöttem az UniProtKB ID-k és más azonosítók közti megfeleltetést tartalmazó táblázatokat. A hatékony adatfeldolgozás érdekében ezekből MySQL adatbázisban megfelelően indexelt táblákat hoztam létre. Az indexelés meggyorsítja az adatok lekérdezését az adatbázisból. Az általam alkalmazott indexek tartalmazzák a fordítandó azonosítót, az UniProt azonosítót, illetve a tábla elsődleges kulcsát. Ezek a táblákat tekinthetjük szótáraknak, én a fordító tábla elnevezést használom. Adatbázis műveletek segítségével, az importálni kívánt adatbázisban szereplő azonosítók és a fordító táblában szereplő azonos típusú azonosítók közötti azonosság alapján megtalálható az adott fehérjéhez tartozó UniProtKB azonosító, melyet az importálandó kapcsolati adatokat tartalmazó tábla

külön oszlopába írtam. Egyes adatbázisok többféle azonosítót tartalmaznak, így mindegyik alapján lehetséges a fordítás. Ezekben az esetekben mindegyik lehetőséget kihasználtam, így nagyobb arányú sikeres megfeleltetést értem el. A legtöbb adatbázisban szerepeltek olyan azonosítók, melyeket nem sikerült UniProtKB azonosítóra lefordítani. Ennek oka lehet az eltérő írásmód vagy régi név használata, például gén nevek esetében; egy másik ok lehet az egyik adat a fehérje egy bizonyos izoformáját vagy translációs variánsát jelöli, míg a másik táblában szereplő azonosító nem ennyire specifikus; de talán a legjellemzőbb ok a törölt vagy elavult adatbázis azonosítók használata. A RefSeq és az UniProtKB azonosítók esetében a megváltozott azonosítókat frissítettem mindegyik adattáblában, és a végleg törölt azonosítókhoz tartozó adatokat töröltem – ezek többnyire nem jelölnek létező fehérjét. A folyamat végén az importálandó adatokat tartalmazó táblában az UniProtKB azonosítókat az elsődleges (*primary*) azonosítóra cseréltem, ahol nem ez szerepelt. Egy fehérjének több UniProtKB azonosítója lehet, de ezek közül csak egy az ún. elsődleges azonosító. Ezzel az eljárással elértem, hogy az összes importálandó adatbázisban a fehérjék egységes, összehasonlítható, egymásnak megfeleltethető elnevezéssel szerepeljenek; így lehetőségessé vált a különböző adatbázisokból származó interakciók egy hálózattá egyesítése.

A fehérjéken kívül mindhárom adatbázisban miRNS-ek is szerepelnek. Ezek az adatkezelés szempontjából ugyanolyan entitások, mint a fehérjék. miRNS-mRNS interakciókat, illetve a miRNS-ek transzkripcióját szabályozó transzkripciós faktorokat integráltunk az adatbázisokba. A más adatbázisokban szereplő miRNS-ek neveit minden esetben a miRBase primer, éretlen, hajtú alakú miRNS (pri-miRNS) AC-knek (*Accession Number*) feleltettem meg. Ezek az azonosítók kilenc karakterből álló, MIXXXXXXX formátumúak, ahol az X helyeken számjegyek szerepelnek. Az érett miRNS-eket jelölő, MIMAT kezdetű azonosítók alapján a megfelelő primer miRNS azonosítóval jelöltem az adattáblákban szereplő miRNS-eket. Egyes adatbázisokban csak nevekkkel jelölik a miRNS-eket. Ezekben az esetekben a nevekben az érett miRNS-re jellemző részt levágtam: a -3p, -5p és * toldalékokat. Azért volt erre szükség, mert egyes adatbázisok az érett, mások a primer miRNS neveket tartalmazzák. Az így egységesített neveket a miRBase adatbázisból származó fordító táblával miRBase AC-re fordítottam, így biztosítva a miRNS-ek egységes jelölését.

3.4.3. Az adatbázis összeállításának folyamata

3.4.3.1. Signalink 2. Az adatbázis összeállításának folyamatát részleteiben a Signalink 2 példáján mutatom be, az ARN és az NRF2ome esetében csak az ettől való eltéréseket említem. Az adatbázis építése a korábbiakban felsorolt adatforrásokat külön táblákban tartalmazó fejlesztői adatbázisban zajlik. Az építést az általam írt SQL szkript automatikusan végzi, a folyamat sikeres végrehajtása esetén a dátumot, és az eggyel megnövelt belső verziószámot a `version` táblába írja. Egy külön SQL szkript a fejlesztői adatbázis tábláit – a forrás táblák kivételével – kiüríti. Egy másik szkript szolgál az építést követően az adatok integritásának ellenőrzésére.



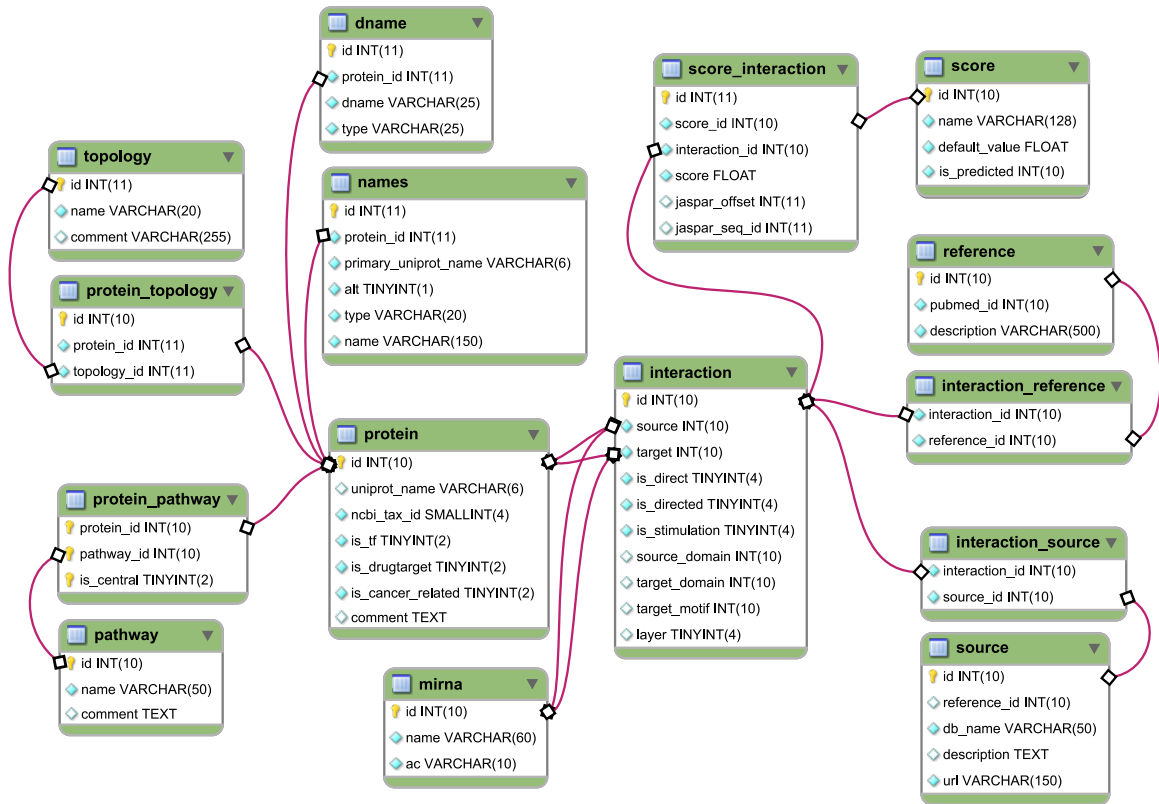
7. ábra. A **Signalink 2** adatbázis réteges felépítése. A szövegből kiderül, hogy az ábra egyszerűsített, a 0-1. réteget, illetve a 2., 3. és 6. réteget, valamint a 4. és a 7. réteget egyként jeleníti meg.

Az adatbázis rétegekből áll, egy kapcsolat egy réteghez tartozik, egy fehérje vagy miRNS azonban több rétegben is szerepelhet. Két fehérje között akár több rétegben is létezhet kapcsolat. Az adatbázis építése a legbelső, 0. rétegtől indul, és rétegenként kifelé haladva történik. Ez

a metódus határozza meg, hogy mely kapcsolatok – és ezáltal mely fehérjék és miRNS-ek – kerülnek be az adatbázisba: minden lépésben csak azok a kapcsolatok kerülhetnek be, amelyek a megelőző rétegekben már szereplő elemekhez kapcsolódnak. A rétegek tartalmát a 7. ábrán követhetjük, ahol a nyilak jelzik az egyes rétegek egymásra épülését.

A Signalink 2 magja a 0. réteg, melynek minden kapcsolata saját irodalmi gyűjtésből származik, és a hét jelátviteli útvonal tagjait tartalmazza. Az 1. réteg szintén saját gyűjtés, ezek a kapcsolatok a jelátviteli fehérjékkel kölcsönhatásba lépő endocitotikus- és állványfehérjéket foglalja magába. A 2. réteg az előző két réteg fehérjéi között, illetve az ezekkel közvetlen szomszédságban álló fehérjékkel való kapcsolatokat tartalmaz, melyek domén-domén illetve domén-motívum interakciókon alapuló predikcióból származnak. A 3. rétegbe általános PPI adatbázisokból – BioGRID, HPRD, DroID, WI8 – átvett kapcsolatok kerültek, szintén a belsőbb rétegekben már szereplő fehérjék és első szomszédai közt. Ezen adatbázisok irány megjelölés nélküli kapcsolatokat tartalmaznak, de a 3. réteghez csak olyan kapcsolatokat adtam hozzá, melyekhez a domén-domén interakciókon alapuló irány predikció segítségével sikerült megbízható módon irányt meghatározni. Azok a kapcsolatok, melyeknél az irány predikció alacsony megbízhatóságú eredményt adott, a 6. rétegbe kerültek, irány megjelölés nélkül. A 4. rétegben transzkripciós faktorok és általuk szabályozott gének kapcsolatai találhatóak, a 1.6.4 fejezetben felsorolt adatbázisokból. A 0-3. és 6. rétegekben szereplő összes fehérje transzkripciós faktorait belefoglaltam ebbe a rétegbe. Ebben a lépésben új fehérjék is bekerültek az adatbázisba, így egy további művelettel az általános PPI adatbázisokból átvett a transzkripciós faktorok közti interakciókat is. Az 5. réteggel az összes eddigi fehérje miRNS szabályozóit csatoltam az adatbázishoz, a 1.6.5 fejezetben felsorolt adatbázisokból. Végül, a 7. réteg az ezen miRNS-eket szabályozó transzkripciós faktorokat tartalmazza, a 1.6.6 fejezetben felsorolt adatbázisok alapján. A több adatbázisban szereplő kapcsolatok csak egyszer szerepelnek az *interaction* táblában, az *interaction_source* táblában több forrásmegjelöléssel. Azonban ha két fehérje közt a feldolgozott cikkek alapján többféle kapcsolat is létrejön: például serkentő és gátló, közvetlen és indirekt egyaránt, akkor ezek külön kapcsolatként szerepelnek az *interaction* táblában. Minden itt felsorolt lépéssel az

egyres kapcsolatokhoz tartozó forrásokat, hivatkozásokat és megbízhatósági mérőszámokat is beleírtam a megfelelő táblákba.



8. ábra. A **Signalink 2 adatbázis sémája**. Ez az adatmodell minimális eltérésektől eltekintve azonos az NRF2ome-nál és ARN-nél alkalmazottal. A zöld mezőkben a táblák nevei, alatta a mezők nevei és típusai szerepelnek. A bíbor színű vonalak a táblák közti relációkat, azaz *foreign key*-ket jelölik.

3.4.3.2. NRF2ome és ARN. Az NRF2ome adatbázisban az 1. réteg az NRF2 kézzel gyűjtött, első szomszédsági hálózatát tartalmazza, 0. és 2. réteg nincsen. Az ARN esetében a 0. réteg az autofágia fehérjék egymás közti interakcióit, az 1. réteg ezen fehérjék első szomszédait tartalmazza. Ezek a kapcsolatok saját irodalmi gyűjtésből, az Autophagy database-ből, illetve

a felhasznált ChIP-MS adatokból (Behrends és mtsai 2010) származnak. A 2. réteg az ARN esetében hiányzik. A 3. réteg az NRF2ome-ban és az ARN-ben a predikciókból származó kapcsolatokat tartalmazza. A 6. rétegben pedig a Signalink 2 0. rétegéből átvett, saját irodalmi gyűjtésből származó kapcsolatok találhatóak, melyek összekötik a jelátviteli útvonalakat az NRF2 illetve az autofágia fehérjék szabályozási hálózatával (10. ábra).

3.4.4. Weboldalak

A webes felületek szerver oldali részét PHP-ban programoztam, Apache webservert szolgálja ki a tartalmakat. A weboldal a gyors és dinamikus működés érdekében aszinkron kérésekkel tölti le a felhasználó által kért információkat, például egy kapcsolat forrásait és hivatkozásait. Kliens oldalon ezt a jQuery-ben elérhető AJAX eljárások teszik lehetővé. A hálózatok vizualizálását a Cytoscape Web segítségével alakítottam ki. A Cytoscape Web egy flash alapú, JavaScriptből vezérelhető hálózat rajzoló program, mely a weboldalba beágyazva működik. A weboldalak funkcióit részletesen az Eredmények részben ismertetem.

3.4.5. A letöltő modul

A letöltő modul egy különálló program, melyet az ELTE NetBiol csoportjának tagja, Fazekas Dávid írt Python nyelven. A weboldalon megadott paramétereket feldolgozva, kinyeri a megfelelő adatokat az adatbázisból, és a kívánt formátummá alakítja azokat.

4. Eredmények

Ebben a részben bemutatom az elkészült adatbázisok jellemzőit, és összehasonlító módon szemléltetem azok mennyiségi tulajdonságait. A felhasznált forrás adatbázisokra a módszertani részben hivatkozom, így itt csak a nevük kerül említésre.

4.1. Az adatbázisok felépítése

4.1.1. A SignaLink 2

4.1.1.1. Jelátviteli komponensek. A SignaLink 2 hét jelátviteli útvonal: a Hedgehog, a WNT/Wingless, a TGF- β , az RTK, a JAK/STAT, a Notch és az NHR útvonalak szabályozását feldolgozó adatbázis. Ugyanezek az útvonalak szerepeltek a SignaLink 1-ben is, mely a SignaLink 2 fejlesztésének kiindulópontját képezte, ugyanakkor az IGF (inzulin) és az EGF/MAPK útvonal összevonásra került, és együtt az RTK (receptor-tirozinkináz) útvonalat alkotják. Kiválasztásukat az egyedfejlődésben és betegségek patomechanizmusában játszott jelentős szerepük indokolta (Pires-daSilva és mtsai 2003). Ezek az útvonalak jól megkülönböztethető biokémiai mechanizmusokat használnak a jel továbbítására, melyek az egyes útvonalakon belül közös evolúciós eredetet mutatnak.

A SignaLink 2 tartalmazza a jelátviteli útvonalakat alkotó fehérjék közti kapcsolatokat, valamint feltérképezi ezen fehérjék poszt-transzlációs, transzkripciós és poszt-transzkripcionális szabályozását. A SignaLink 2-ben három faj szerepel: két fontos modellorganizmusként a *Caenorhabditis elegans* fonálféreg és az ecetmuslinca (*Drosophila melanogaster*), továbbá az ember.

Az útvonalakban szereplő fehérjék tartozhatnak az útvonal központi (*core*), vagy mellék (*non-core*) régiójához. A központi régió nélkülözhetetlen a jel receptorok és transzkripciós faktorok közti továbbításához, míg a mellék régió fehérjéi csak szabályozzák, módosítják a jel továbbadását. Egy jelátviteli fehérje több útvonal tagja is lehet, az ilyen fehérjéket *multi-pathway* fehérjének nevezzük.

Minden jelátviteli fehérje rendelkezik topológiai tulajdonságokkal, melyek a jelátvitelben betöltött funkcióját jelzik. Ez alapján lehet ligandum, receptor, mediátor, kofaktor, vagy

transzkripció faktor. Egy fehérjének legfeljebb két topológiai tulajdonsága lehet (Korcsmáros és mtsai 2010).

A legbelső réteget a jelátviteli útvonalakban résztvevő fehérjék közti kapcsolatok alkotják. Ez a rész ilyen tekintetben megfelel a SignaLink 1-nek, melyben kizárólag a jelátviteli fehérjék szerepeltek, a ligandumoktól és receptoroktól kiindulva az első érintett transzkripció faktorokig. A SignaLink 1. verziója 2008-ban készült el, 8 útvonalat tartalmazott, melyek elemeit 170 review cikk alapján lettek meghatározva. Az útvonal tagok közti további interakciókat összesen 941 cikkből feldolgozásával gyűjtöttük össze. A SignaLink jelenlegi, verziójában ez a szám 2.197-re emelkedett. Az útvonalak és a fajok száma nem bővült, így a cikkek számának növekedése az időközben publikált új kísérletes adatoknak tudható be.

A következő rétegben további saját gyűjtésből származó adatok találhatóak a jelátviteli fehérjéket poszt-transzlációs módosítások révén közvetlenül szabályozó enzimekről (kinázok, foszfatázok, ligázok, proteázok), endocitotikus és állványfehérjékről, valamint az ubikvitin-proteaszóma rendszer fehérjéiről. Ezek a kapcsolatok kutatócsoportunk tagjai által 1.075 cikkből kerültek kigyűjtésre. Az útvonalak és szabályozóik kapcsolatainak irodalmi gyűjtése 2011 áprilisában zárult le.

4.1.1.2. Többrétegű szabályozási hálózatok. A SignaLink 2 a fehérjék és mikro-RNS-ek kapcsolatait kettős szempontrendszer alapján, hagyma-szerűen egymásra épülő rétegekbe sorolja. A rétegek megfelelnek a szabályozás molekuláris biológiai rendszerekben megfigyelhető különböző módjainak. Ezenkívül, a rétegek eltérnek az adatforrás típusában is. Ez az elkülönítés lehetőséget ad, hogy a SignaLink 2-t elemzéseikben felhasználó kutatók dönthessenek a szabályozás különböző szintjeinek, illetve az eltérő megbízhatósági fokú és eltérő mennyiségű adatot szolgáltató adatforrásoknak a felhasználásáról, vagy az elemzésből való kizárásáról.

A harmadik rétegbe az ELM felhasználásával készült predikció alapján, az előző két rétegben szereplő fehérjékben megtalálható motívumokat módosítani képes fehérjék kerültek. A negyedik réteg hasonló módon fehérje-fehérje kapcsolatokat tartalmaz, azonban más adatbázisokból – a BioGRID-ből és a HPRD-ből – átvett, kísérletesen igazolt adatok alapján. Külön rétegbe kerültek azok a kapcsolatok, melyekhez a domén-domén interakciók alapján

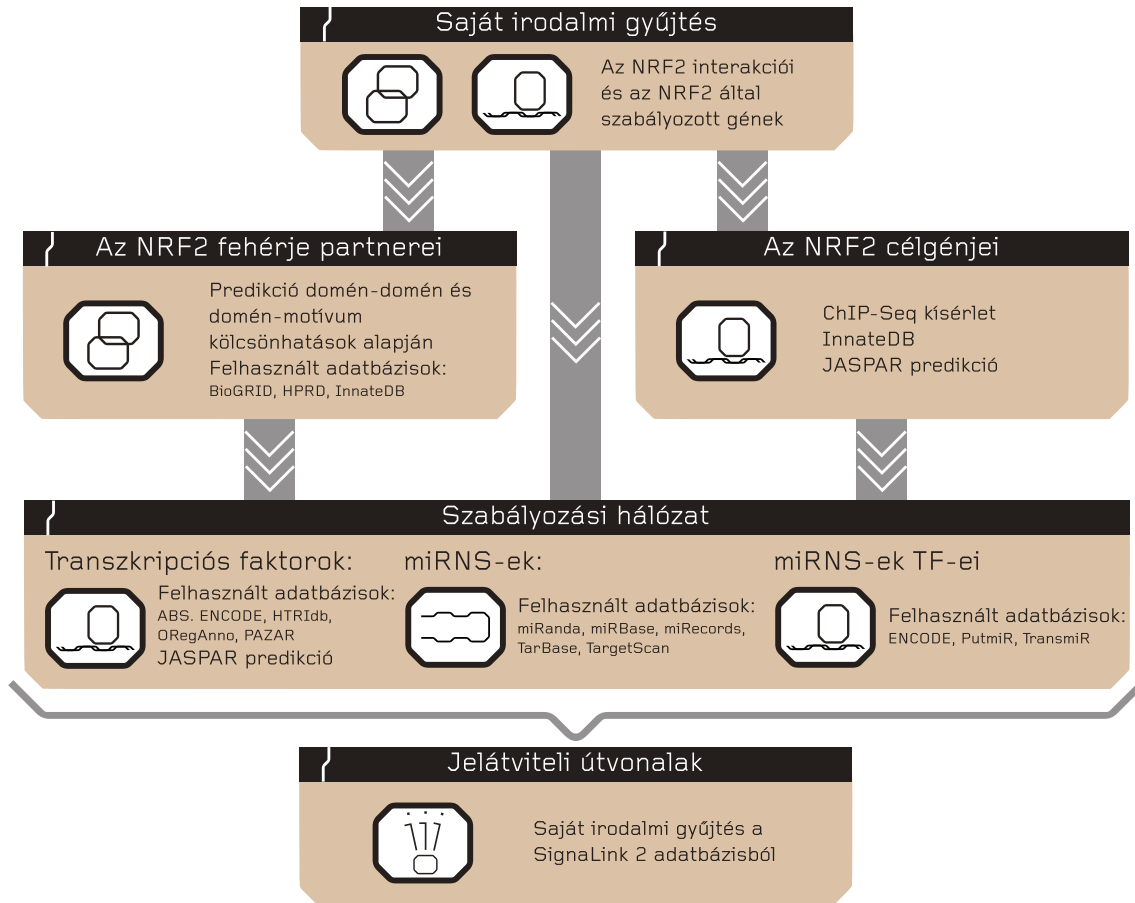
megfelelő megbízhatósággal sikerült irányt prediktálni. Ebben az esetben tehát a kapcsolat léte kísérletesen igazolt, az iránya azonban prediktált. Az irányítatlan kapcsolatok külön réteget alkotnak „további kapcsolatok” (*further interactions*) néven, ez a legkülső réteggént szerepel.

A fehérje-fehérje interakciók rétegein kívül helyezkedik el az ezekben foglalt fehérjék transzkripciós poszt-transzkripcionális szabályozását leíró két réteg. Előbbi több transzkripciós faktor adatbázisból – ABS, DroID, EDGEDb, ENCODE, HTRI, ORegAnno, RedFly, wTF – átvett kapcsolatokat, illetve a JASPAR algoritmussal készült predikciókat tartalmazza. Ezek a kapcsolatok szükségszerűen indirektek és irányítottak, a forrás a transzkripciós faktor, a cél pedig a szabályozott gén. A poszt-transzkripcionális szabályozás rétegében szintén irányított kapcsolatok találhatók. Amennyiben egy miRNS képes kötődni egy fehérjét kódoló mRNS 3' UTR régiójához, a kapcsolat forrásaként a miRNS, a célpontjaként pedig a fehérje szerepel. Ezek a kapcsolatok gátló hatásúként szerepelnek, és több predikciós algoritmusból – microT v4, miRDB, miRanda, PicTar, TargetScan – és három kísérletesen igazolt adatokat tartalmazó adatbázisból – miR2Disease, miRDeathDB és TarBase – származnak. Amennyiben egy fehérje transzkripciós faktorként szabályozza egy miRNS gén átíródását, a kapcsolat forrása fehérje, a célpontja pedig miRNS. Ezek a kapcsolatok a miRNS-e ENCODE, a PuTmiR és a TransmiR adatbázisokból kerültek átvételre, és a transzkripcionális szabályozás kapcsolataival megegyezően indirektek.

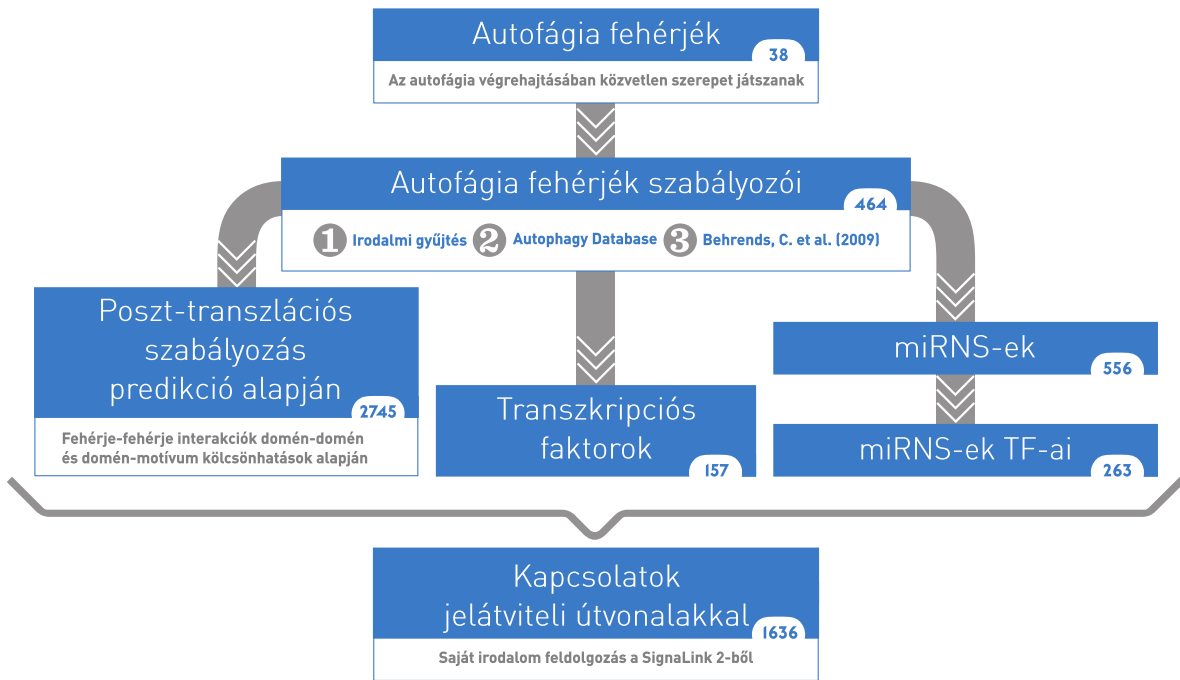
4.1.2. Az NRF2ome

Az NRF2ome-ban átvettem a Signalink 2 egymásra épülő, réteges felépítését, azonban néhány ponton módosítottam azt (9. ábra). A legbelső réteget itt is a kutatócsoportunk által végzett irodalmi gyűjtés alkotja, mely az NRF2 és a KEAP1 közvetlen szabályozóinak hálózatát tartalmazza (Papp és mtsai 2012). Ezt a fehérje interakciós hálózatot egészítik ki a más adatbázisokból (BioGRID, HPRD, InnateDB, IntAct) átvett kapcsolatok, és a domén-domén, valamint domén-motívum interakciók alapján készült predikciók. Az NRF2ome adatbázisában ezek egyetlen fehérje-fehérje interakciós rétegben találhatók. A következő két réteg a transzkripcionális szabályozás és a poszt-transzkripcionális szabályozás kapcsolatait tartalmazza, a Signalink-vel megegyező forrásokból. A legkülső réteget az NRF2ome-ban a jelátviteli

útvonalak alkotják. A jelátviteli fehérjét mind fehérje-fehérje kapcsolatok, mind transzkripció faktorok – és ezeken keresztül miRNS-ek – által összekapcsoltam az NRF2 szabályozási hálózatával.



9. ábra. **Az NRF2ome adatbázis felépítése.** A panelek az egyes rétegeket jelölik. A rétegek egymásra épülő struktúrája, azaz az adatbázis összeállításának menete az ábrát felülről lefelé szemlélve követhető végig (Türei és mtsai [előkészületben](#); Türei és mtsai [2013](#)).



10. ábra. **Az ARN adatbázis felépítése.** A panelek az egyes rétegeket, a számok az adott rétegben kapcsolatokkal rendelkező fehérjék, illetve miRNS-ek számát jelölik. (Türei és mtsai előkészületben; Türei és mtsai 2013).

4.1.3. Az Autophagy Regulatory Network (ARN)

Az ARN felépítése nagyban hasonlít az NRF2ome struktúrájára, azonban ez az adatbázis nem egyetlen, hanem 38 – az autofágia végrehajtásában közvetlenül érintett – fehérje szabályozási hálózatát tartalmazza (10. ábra, alsó rész). Az ARN egy adatforrásban egyesíti az autofágia fehérjék és szabályozói fehérje interakciós hálózatát, melyet saját irodalmi kuráció, egy autofágiára koncentrált ChIP-MS vizsgálat (Behrends és mtsai 2010), az ADB, a BioGRID és a HPRD adatbázisok, valamint domén-domén és domén-motívum alapú predikciók alapján állítottam össze. Maguk az autofágia fehérjék, az autofágia fehérjék kísérletesen igazolt, illetve prediktált szabályozói három külön réteget alkotnak. A transzkripcionális szabályozás és a poszt-transzkripcionális szabályozás rétegeinek felépítése és adatforrásai azonosak a Signalink 2 és NRF2ome esetében leírtakkal. Az NRF2ome-hoz hasonló módon, az ARN-ben is összekapcsoltam az összes szabályozási réteget a Signalink 2 jelátviteli útvonalaival.

4.2. Az adatbázisok mennyiségi tulajdonságai

Az adatbázisokban található entitások – fehérjék és mikro-RNS-ek –, valamint kapcsolatok számát táblázatos formában tekintem át. Ezen mennyiségek ismerete fontos az elemzésekben való felhasználhatóság értékelésénél. Előbb három külön táblázatban rétegek szerint, majd összehasonlító módon, források szerint részletezve mutatom be az adatok mennyiségét.

4.2.1. A Signalink 2

Az 1. táblázatban összefoglalom a Signalink 2 egyes rétegeiben található kapcsolatok és molekulák számát. A Signalink 2 három faj, a *C. elegans*, a *D. melanogaster* és az ember jelátviteli útvonalait és ezek szabályozóit tartalmazza. A legtöbb adat a humán interakciós hálózatról áll rendelkezésre. Ennek oka a kutatottság magasabb szintje mellett, az emlősökben előforduló paralógok nagyobb száma. Az evolúció során egyes gének duplikálódtak, és a homológok között funkcionális különbségek alakultak ki. Ezenkívül egyes domének bizonyos fajokban egy fehérjét alkotnak, míg más fajokban külön fehérjeként vesznek részt a komplexekben (Pires-daSilva és mtsai 2003). A *C. elegans* és a *D. melanogaster* esetében hasonló nagyságrendű fehérjével és kapcsolattal találkozunk. Ezekben a fajokban lényegesen kisebb az ismert miRNS-ek száma is.

Figyelemre méltó a poszt-transzkripcionális szabályozás rétegében szereplő kapcsolatok nagy száma. A többféle perdikciós módszer igen nagy számú adatot eredményez, mely még a szigorú szűrés után is jelentős mennyiségű kapcsolatot jelent. Az egyes módszerek eredményei átfednek, de néha aggasztó mértékű különbség tapasztalható. Amennyiben az elemzésekhez csak a kísérletesen igazolt miRNS–mRNS kapcsolatokat kívánjuk felhasználni, a letöltő modul lehetőséget nyújt erre. A kísérletes adatok lefedettsége azonban alacsonyabb, kevesebb miRNS-ről áll rendelkezésre ilyen adat.

A miRNS-ek transzkripcionális szabályozását külön sorban tüntettem fel.

1. táblázat. A Signalink 2 rétegek és fajok szerinti áttekintése

Organizmus	<i>C. elegans</i>		<i>D. melanogaster</i>		<i>H. sapiens</i>	
	Entitások	Kapcsolatok	Entitások	Kapcsolatok	Entitások	Kapcsolatok
Jelátviteli útvonalak	198	253	210	260	1.150	1.640
Útvonalak szabályozói	0	0	0	0	751	2.122
Poszt-transzlációs szabályozás	916	3.072	1.713	6.896	4.682	82.852
Irányított fehérje-fehérje kapcsolatok	49	47	128	171	951	3.252
Irányítatlan fehérje-fehérje kapcsolatok	100	245	166	496	1.387	6.086
transzkripcionális szabályozás	152	187	998	16.319	2.585	30.736
mikro-RNS-fehérje kapcsolatok	806	9.658	939	5.308	2.844	245.857
mikro-RNS-ek transzkripcionális szabályozása	25	19	0	0	716	5.209

4.2.2. Az NRF2ome

Az NRF2ome csak humán interakciókat tartalmaz, és jóval kisebb méretű szabályozási hálózatot jelent, hiszen a több száz jelátviteli molekula helyett csupán az NRF2 köré épül. A fehérje-fehérje interakciók egy réteget alkotnak, melyben az irodalmi gyűjtés, a más adatbázisokból átvett kapcsolatok és a predikciók szerepelnek. A 2. táblázatban ezeket külön sorban tüntettem fel. A legkülső réteget itt a Signalink 2 irodalmi gyűjtésből származó jelátviteli hálózata alkotja, melyben 508 fehérje 906 kapcsolata szerepel. Ezek azok a jelátviteli komponensek, melyek kapcsolatban vannak az NRF2-t szabályozó fehérjékkel, az ezeket szabályozó transzkripció faktorokkal, és az ezeket szabályozó miRNS-ek transzkripció faktoraiival.

2. táblázat. Az NRF2ome rétegek szerinti áttekintése

Rétegek	Kapcsolatok	Fehérjék	mikro-RNS-ek
Irodalmi gyűjtés	108	85	–
Kísérletesen igazolt fehérje-fehérje kapcsolatok	15.836	2.274	–
Prediktált fehérje-fehérje kapcsolatok	130	131	–
Transzkripcionális szabályozás	7.354	10.263	–
mikro-RNS-fehérje kapcsolatok	7.430	82	541
mikro-RNS-ek transzkripcionális szabályozása	4.963	257	408
Jelátviteli útvonalak	906	508	–

4.2.3. Az Autophagy Regulatory Network (ARN)

Az ARN rétegei az NRF2ome-hoz hasonlóan épülnek egymásra, tartalmukról a 3. táblázatban olvashatók adatok. A legbelső réteget a 38 autofágia fehérje alkotja, azonban nem mind-egyiknek ismertek kapcsolatai más autofágia fehérjékkel. Szélsőséges eset a GABARAPL3, melynek egyetlen kapcsolata sem ismert, egyik rétegben sem. A hat humán Atg8 homológ közül csak erről a fehérjéről nem áll rendelkezésre adat, más adatbázisok (pl. a STRING) a homológok kapcsolatait hozzárendelik a GABARAPL3-hoz is. Az ARN esetében nem végeztem ilyen predikciót, a hamis pozitív kapcsolatok elkerülése érdekében.

3. táblázat. Az ARN rétegek szerinti áttekintése

Rétegek	Kapcsolatok	Fehérjék	mikro-RNS-ek
Autofágia fehérjék	184	31	–
Autofágia fehérjék szabályozói	975	472	–
Prediktált szabályozók	7.082	2.893	–
Transzkripcionális szabályozás	4.962	565	–
mikro-RNS-fehérje kapcsolatok	363.998	769	1.377
mikro-RNS-ek transzkripcionális szabályozása	2.077	394	208
Jelátviteli útvonalak	1.634	741	–

4.2.4. Az autofágia kapcsolatai jelátviteli útvonalakkal

Az ARN alapján a 38 autofágia fehérje poszt-transzlációs szabályozásában 247 fehérje vesz részt. Az ARN tartalmazza az autofágia fehérjék és szabályozóik transzkripció faktorait. A 38 autofágia fehérje közül 33-at, 60 TF összesen 378 TF–promóter kapcsolaton keresztül szabályoz. Ezen kapcsolatok közül 58 kísérletesen igazolt, míg 325 származik a JASPAR szerinti predikcióból. A 60 TF közül 19 jelátviteli útvonalakhoz tartozik, legtöbb az NHR és az RTK útvonalhoz.

A 38 autofágia fehérje közül 37 poszt-transzkripcionális szabályozásáról találunk adatokat az ARN-ben, melyeket 1.319 miRNS 18.246 kapcsolaton keresztül szabályoz. Ezen kapcsolatok közül 9.062 kísérletesen igazolt, a többi pedig különböző predikciós módszerekből

származik. Az említett miRNS-ek közül 393 transzkripcionális szabályozásában 208 TF vesz részt 2.073 TF–promóter kapcsolaton keresztül.

Ha kiterjesztjük a vizsgálódást az autofágia fehérjék szabályozóira is, megállapíthatjuk, hogy 1.238 fehérje vesz részt a 247 autofágiát közvetlenül szabályozó fehérje poszt-transzlációs szabályozásában, 130 TF szabályozza ezek transzkripcióját, míg 1.366 miRNS poszt-transzkripcionálisan szabályozza őket. A 247 közvetlen szabályozó közül öt, míg az 1.238 közvetett szabályozó közül 96 a SignaLink 2-ből átvett jelátviteli útvonalakhoz tartozik. Az öt autofágiával közvetlen kapcsolatban álló útvonal fehérje közül az mTOR az RTK útvonal tagja, a raptor, a DVL2, a TSC1 és a β -katenin a WNT útvonalhoz tartozik, ez utóbbi pedig a TGF- β útvonalban is részt vesz (4. táblázat, 24. ábra).

4. táblázat. Az autofágia közvetlen kapcsolatai jelátviteli útvonalakkal

Jelátviteli fehérje	Útvonalak	Autofágia fehérjék	Adatforrások
raptor	WNT	PIK3C3	saját irodalmi gyűjtés
mTOR	RTK	ULK1, ULK2, ATG13	saját irodalmi gyűjtés
DVL2	WNT	MAP1LC3B, GABARAP	saját irodalmi gyűjtés
TSC1	WNT	Beclin-1	ADB, Behrends és mtsai 2010
β -katenin	TGF- β , WNT	ATG4B, p62, PIK3C3, GABARAP, Beclin-1	domén-motívum alapú predikció

4.2.5. A SignaLink 2, az NRF2ome és az ARN összehasonlítása

Az 5. táblázatban összehasonlítom a három adatbázis átfogó statisztikáit. A legtöbb kapcsolat és a legtöbb fehérje a SignaLink 2-ben található, ezt követi az ARN, a legkisebb méretű pedig az NRF2ome. Ez a rangsor követi az adatbázis magját alkotó irodalmi gyűjtésből származó interakciós hálózat méretét. Az összes többi réteg – kis eltérésektől eltekintve – azonos adatforrásokból építkezik. Az egyes adatforrásokból az adatbázisokba átvett kapcsolatok

5. táblázat. A **Signalink 2**, az **NRF2ome** és az **ARN** összehasonlítása

	Signalink 2	NRF2ome	ARN
Kapcsolatok	363.998	36.139	263.411
Fehérjék	33.105	7.891	4.034
mikro-RNS-ek	872	541	1.380
Források	59	47	59
Hivatkozások	8.446	2.846	2.023

száma így elsősorban annak a hálózatnak a méretétől függ, amely köré a szabályozási rétegeket felépítjük.

A 6. táblázatban az interakciók tulajdonságai: az irányítottság, a közvetlenség és a hatás iránya (serkentés vagy gátlás) szerepelnek. Sok elemzés szempontjából fontos a hálózat irányítottsága. Számos módszer esetében gondot okoz, ha az interakcióknak előjele van, ilyenkor figyelmen kívül lehet hagyni ezt a paramétert. A transzkripcionális szabályozás kapcsolatai minden esetben indirektek. A poszt-transzkripcionális szabályozás kapcsolatai – feltételezvé, hogy minden miRNS gátolja a transzlációt – gátló hatásúak. Mivel a fehérje-fehérje interakcióknál ezek a paraméterek bármilyen értéket felvehetnek, ezen szempontokat a fehérje-fehérje interakciókra szűrve adtam meg a táblázatban.

A 7. táblázatban az egyes adatforrásokat vizsgálom, az egyes adatbázisokba átvett interakciók és interakciós partnerek alapján. Az egyes adatforrások tulajdonságairól részletesen írok az 1.6. fejezetben. A kísérletes adatokat tartalmazó adatbázisok általában kisebbek. Például az ABS-ben, vagy a miRDeathDB-ben igen kis számú kapcsolat szerepel, azonban ezek megbízhatóbb módszerekből származnak, kisebb a hamis pozitívok aránya. Az átvett interakciók számát alapvetően meghatározza egyrészt az adatbázis lefedettsége, másrészt a szabályozott rétegben szereplő fehérjék vagy miRNS-ek száma.

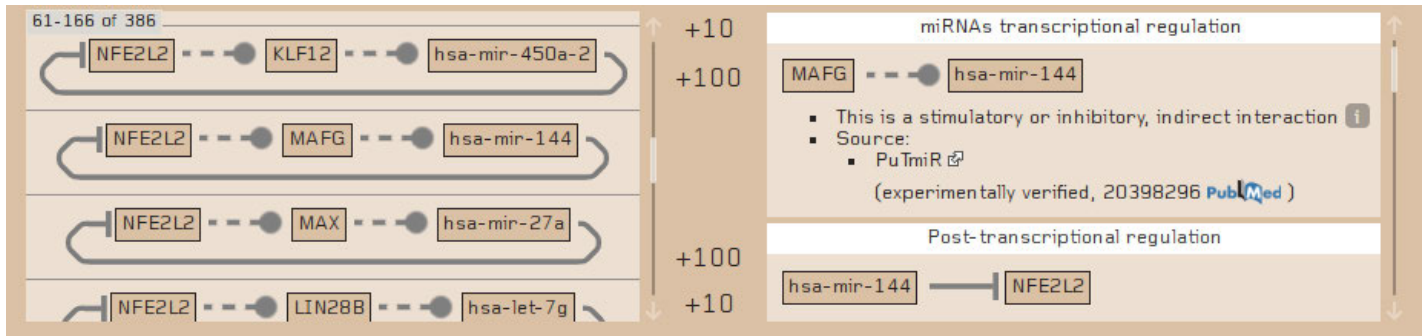
6. táblázat. Az adatbázisokban szereplő fehérjék és miRNS-ek interakcióinak tulajdonságok szerinti összehasonlítása

	Signalink 2	NRF2ome	ARN
<i>Irányítottság – fehérje-fehérje kapcsolatok</i>			
Irányított	99.712	1.755	7.294
Ebből irány predikció	3.470	0	0
Irányítatlan	7.643	12.165	1.312
<i>Hatás – összes kapcsolat</i>			
Serkentő	2.038	1.587	1.658
Gátló	204.126	7.501	248.012
Ismeretlen hatású	157.834	27.051	13.741
<i>Hatás – fehérje-fehérje kapcsolatok</i>			
Serkentő	2.038	1.570	1.568
Gátló	94	71	30
Ismeretlen hatású	105.223	12.279	7.008
<i>Közvetettség – fehérje-fehérje kapcsolatok</i>			
Közvetlen	107.045	13.727	8.467
Indirekt	297	182	128

7. táblázat. A más adatbázisokból átvett adatok áttekintése

Adatbázis neve	Összes kapcsolat	Átvett kapcsolatok – <i>molekulák</i>			Réteg
		Signalink 2	NRF2ome	ARN	
<i>Általános PPI adatbázisok</i>					
BioGRID	110.989 – 12.917	22.016 – 6.432	9.926 – 2.014	384 – 81	3; 6
DroID	158.266 – 12.213	13.894 – 3.238			3; 6
HPRD	39.240 – 9.520	12.277 – 4.409	4.994 – 1.521	119 – 55	3; 6
WI8	3.864 – 2.298	303 – 358			3; 6
<i>Transzkripciós faktor adatbázisok</i>					
ABS	650 – 124	21 – 21	1 – 2	0 – 0	4
DroID	157.914 – 12.248				4
edgeDB	1.406 – 302	523 – 259			4
ENCODE	45.328 – 9.530	2.475 – 636	186 – 115	1.064 – 350	4
HTRI	52.467 – 16.232	6.817 – 1.911	293 – 119	351 – 141	4
JASPAR	183.781 – 26.351	23.098 – 2.567	8.217 – 6.329	2.911 – 316	4
ORegAnno	6.143 – 2.204	301 – 297	35 – 41	24 – 27	4
PAZAR	47.173 – 10.291	1.032 – 665	100 – 94	111 – 106	4
RedFly	2.998 – 276	179 – 136			4
<i>mikro-RNS adatbázisok</i>					
doRiNA	766.682 – 23.548	277.792 – 13.412		15.688 – 1.073	5
microT v4	972.174 – 27.381	144.982 – 8.147		24.446 – 885	5
miR2Disease	813 – 531	199 – 226		75 – 84	5
miRDB	1.884.403 – 56.738	365.645 – 11.439		18.307 – 1.263	5
miRDeathDB	458 – 204			47 – 54	5
miRanda	16.805.191 – 39.513	4.297.167 – 22.068	2.206 – 258	173.446 – 1.102	5
PicTar	104.026 – 13.400	32.396 – 6.789		1.520 – 332	5
TarBase	1.331 – 1.040	905 – 857	4 – 8	62 – 75	5
TargetScan	9.606.509 – 21.406	2.627.654 – 12.344	5.947 – 538	126.068 – 1.354	5
<i>TF–mikro-RNS adatbázisok</i>					
ENCODE	1.648 – 348	770 – 218	665 – 190	806 – 253	7
PuTmiR 1.1	12.119 – 775	4.211 – 521	4.067 – 460	969 – 328	7
PuTmiR 2.0	10.209 – 449			179 – 123	7
TransmiR	833 – 331	246 – 183	250 – 163	238 – 165	7

A táblázatban szereplő adatbázisok hivatkozásai. Fehérje-fehérje interakciós adatbázisok: BioGRID – Chatr-aryamontri és mtsai 2013, HPRD – Keshava Prasad és mtsai 2009, InnateDB – Breuer és mtsai 2013. TF–gén regulációs adatbázisok: ABS – Blanco és mtsai 2006, DroID – Murali és mtsai 2011, edgeDB – Barrasa és mtsai 2007, ENCODE – Gerstein és mtsai 2012, HTRI – Bovolenta és mtsai 2012, JASPAR – Portales-Casamar és mtsai 2010, ORegAnno – Griffith és mtsai 2008, PAZAR – Portales-Casamar és mtsai 2009, RedFly – Gallo és mtsai 2011, wTF – Reece-Hoyes és mtsai 2005. miRNS–mRNS interakciós adatbázisok: doRiNA – Anders és mtsai 2012, micro-T v4 – Maragkakakis és mtsai 2011, miR2Disease – Jiang és mtsai 2009, miRDB – Wang és mtsai 2008, miRDeathDB – Xu és mtsai 2012, miRanda – Betel és mtsai 2010, 2008; Grimson és mtsai 2007, PicTar – Krek és mtsai 2005, TarBase – Vergoulis és mtsai 2012 és TargetScan – Friedman és mtsai 2009; Lewis és mtsai 2005. TF–miRNS gén regulációs adatbázisok: ENCODE – Gerstein és mtsai 2012, PuTmiR – Bandyopadhyay és mtsai 2010, TransmiR – Wang és mtsai 2008.



11. ábra. **Szabályozási hurkok megjelenítése az NRF2ome weboldalon.** A három lépésből álló szabályozási hurkok egy NRF2 transzlációját gátló miRNS-ből, és egy, a miRNS transzkripcióját szabályozó transzkripciós faktorból állnak, mely utóbbinak a transzkripcióját az NRF2 szabályozza. A kerek végű nyilak itt azt jelzik, hogy nincs adatunk a transzkripciós faktorok serkentő vagy gátló hatásáról. Jobb oldalt a szabályozási hurkot felépítő kapcsolatok részletes adatai láthatók. A szabályozási hurkok listázása csak az NRF2ome weboldalon, az NRF2-t érintő visszacsatolások böngészésére áll rendelkezésre, ugyanakkor a három adatbázis bármelyikét letöltve, bármelyik fehérje esetében kigyűjthetők az ilyen visszacsatolások (<http://nrf2.elte.hu/>).

4.3. Az adatbázisok internetes felületei

4.3.1. A weboldalak

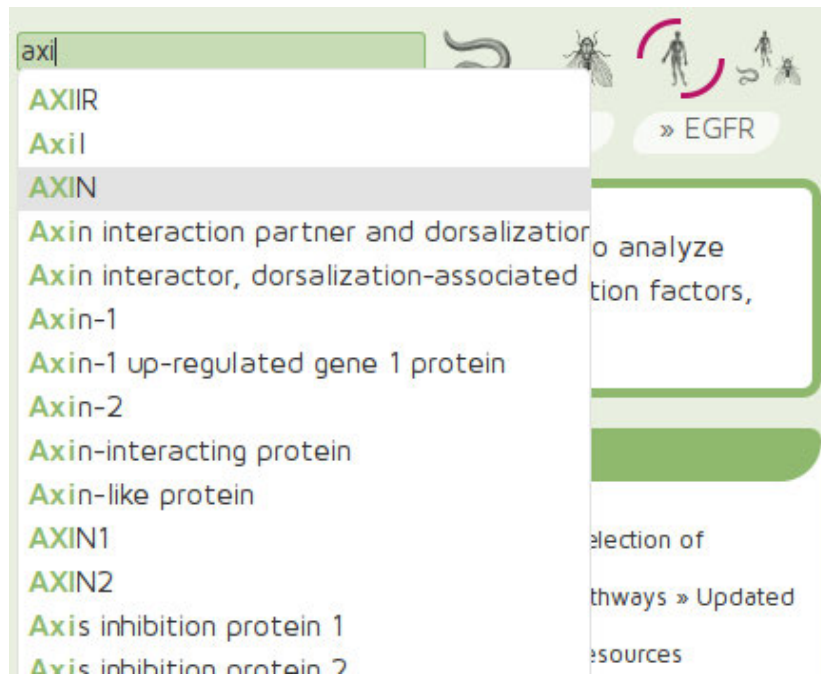
Mindhárom adatbázis rendelkezik saját weboldallal, melyek a 8. táblázatban felsorolt címeken érhetők el.

8. táblázat. **Az adatbázisok weboldalai**

Signalink 2	http://signalink.org/
NRF2ome	http://nrf2.elte.hu/
ARN	http://arn.elte.hu/

A weboldalak lehetőséget adnak az adatbázis teljes tartalmának interaktív böngészésére, és a letöltő modul kezelésére. A felhasználó rákereshet egy-egy fehérjére, lekérdezheti a kapcsolatok listáját, és megismerheti ezek forrásait.

4.3.1.1. Intelligens névkeresés. Mindhárom honlap főoldalán, balra fent egy keresőmező található, ahol különféle elnevezések és azonosítók alapján kereshetünk fehérjéket és miRNS-eket (12. ábra). A keresőmező az elkezdett gépelést automatikusan kiegészíti. Ez a funkció az adatbázisban előforduló összes fehérje és miRNS sokféle szinonim elnevezését tartalmazó lista alapján történik. A keresés a felajánlott lehetőségekre kattintva, vagy a „search” gombra kattintva, illetve az enter leütésével egyaránt elindítható. A Signalink 2 esetében lehetőség van csak egy faj, vagy mindhárom faj adatai közt keresni.



12. ábra. **Keresőmező a Signalink 2 weboldalon.** A felhasználó által elkezdett gépelést automatikusan kiegészíti, például az „axi” szó beírása esetén felajánlja az annexin receptor és az AXIN fehérjék különböző elnevezéseit. Amennyiben az ki van választva egy organizmus, mint ahogy ebben az esetben az ember, csak az adott faj fehérje és mikro-RNS nevei között keres (<http://signalink.org/>).

A kereső algoritmus a keresőszó többféle névvel és azonosítóval való teljes és részleges illeszkedését vizsgálja. Ezt olyan sorrendben teszi, hogy a leggyorsabb (teljes illeszkedés) és leggyakoribb (pl. *gene symbol*) lehetőségeket próbálja ki előbb, így a keresés a lehető leggyorsabban megy végbe. Amennyiben nem egyértelmű az illeszkedés, a felhasználó egy találati listából választhatja ki a kívánt elemet. Ha egyáltalán nincs találat, erről egy üzenet

jelenik meg. Amennyiben a keresés egyetlen találattal zárul, a weboldal az adott fehérje vagy miRNS adatlapjára ugrik (13. ábra).

243 interactions:

Search or browse by interactors and regulators:

A8K0K2

Click on the arrow for details of an interaction

Interactions between pathway members (15)

GSK3B RTK (non-core) RTK (core) Hedgehog (core) TGF (core) WNT/Wingless (core)	→	AXIN1 RTK (non-core) TGF (non-core) TGF (core) WNT/Wingless (core)
PPP2CA WNT/Wingless (non-core)	→	AXIN1 RTK (non-core) TGF (non-core) TGF (core) WNT/Wingless (core)
MACF1 WNT/Wingless (non-core)	---	AXIN1 RTK (non-core) TGF (non-core) TGF (core) WNT/Wingless (core)

Interactions of pathway regulators (8)

Post-translational modifications (136)

Directed protein-protein interactions (9)

Transcriptional regulation (17)

Post-transcriptional regulation (28)

Further interactions (30)

PPP2CA → **AXIN1**

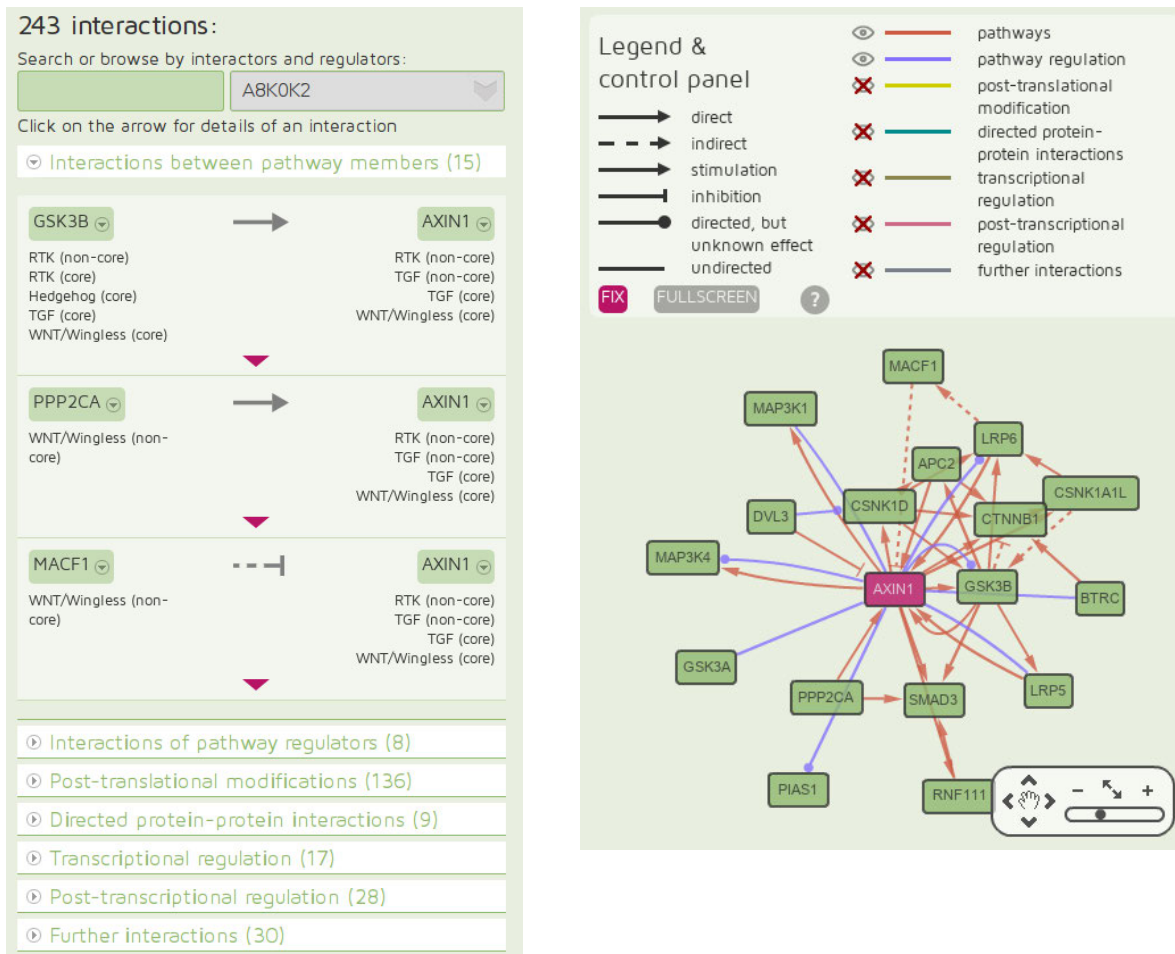
WNT/Wingless (non-core) → RTK (non-core), TGF (non-core), TGF (core), WNT/Wingless (core)

- This is a stimulatory, direct interaction ?
- Sources:
 - HPRD (experimentally verified, 18988627 PubMed)
 - Biogrid (experimentally verified, 21071413 PubMed)
 - Signalink 2.0 (manual curation) (experimentally verified, 20542890 PubMed)
- References:
 - 9920888 PubMed
 - 19131971 PubMed
- Confidence score:
 - GO Semantic Similarity: 0.526758 PubMed ?

13. ábra. **Fehérje adatlap a Signalink 2 weboldalon.** A háromszögekre kattintva az egyes kapcsolatokról további információk érhetők el: a kapcsolat forrása, a hozzá tartozó hivatkozások és a megbízhatósági mérőszámok értékei. A PubMed logókra kattintva a forrás adatbázisokhoz vagy a hivatkozásokhoz tartozó publikációk érhetők el (<http://signalink.org/>).

4.3.1.2. Fehérje és mikro-RNS adatlap. A fehérjék és a miRNS-ek adatlapjain balra fent alapvető információk találhatóak: a legfontosabb azonosítók és elnevezések (Gene Name, UniProt ID, Ensembl Protein ID, FlyBase ID, WormBase ID, miRBase AC); a Signalink 2 esetében a fehérje útvonal besorolásai, és topológiai tulajdonságai (pl. receptor, transzkripció)

faktor, stb.); miRNS-ek esetében a miRNS szerkezetét ábrázoló diagram, mely a miRBase-ből származik (14. és 13. ábra). A fontos referencia adatbázisok (UniProt, miRBase, Ensembl, FlyBase, WormBase) azonosítói egyben linkek is, melyek az adott adatbázisban a megfelelő molekula adatlapjára irányítanak. Az UniProt, a FlyBase és a WormBase esetében ez fordítva is működik, a SignaLink 2-ben szereplő fehérjék adatlapjáról link mutat a SignaLink 2 weboldalára.



14. ábra. **Fehérje adatlap a SignaLink 2 weboldalon.** Bal oldalt az adott fehérje kapcsolatainak listája, rétegek szerinti csoportosításban. A lista elején keresőmező és legördülő menü segíti az interakciós partnerek keresését. Jobb oldalt a fehérje első szomszédsági hálózatát bemutató interaktív ábra. Kezdetben csak az első kettő réteg irodalmi gyűjtésből származó kapcsolatai láthatók, de a menü segítségével a további rétegek is megjeleníthetők (<http://signalink.org>).

Az adatlap bal oldalán egy lista található, mely az adatbázis `interaction.layer` mezőjének megfelelően listázza a kapcsolatokat. Az egyes rétegek címsorára kattintva legördül a kapcsolatok listája, ahol a két fehérje neve közti nyíl alakja jelzi az interakció típusát: közvetlen vagy indirekt, irányított vagy irányítatlan, illetve gátló vagy serkentő. A weboldalakon a kapcsolatok listázásánál vizuális szimbólumok, a részletes nézetben pedig írásos információ jelenik meg a kapcsolatok tulajdonságairól. A háromszög alakú nyíl serkentést, a T-végű nyíl gátlást, a kerek végű nyíl pedig ismeretlen hatást jelent. Az irányítatlan kapcsolatok nyílhegy nélküli vonalként kerülnek megjelenítésre. Az indirekt kapcsolatot szaggatott vonal jelzi (13. és 14. ábra).

A Signalink 2 weboldalán a fehérjék neve alatt szerepel az útvonal besorolás is. A kapcsolatokra kattintva további információ jelenik meg: a források, a hivatkozások, illetve a megbízhatósági mérőszámok. Előbbi kettő linkeket tartalmaz, így a felhasználó eljuthat a forrás adatbázishoz, vagy a hivatkozott publikációhoz. Az ARN és az NRF2ome weboldalán a kapcsolati lista kiegészül a jelátviteli útvonalakkal való egy vagy két lépésben megvalósuló kapcsolatok listájával (*pathway connections*), így gyorsan áttekinthető, hogy milyen irányból érheti szabályozó stimulus az adott fehérjét.

Az adatlap jobb oldalán egy hálózati ábra látható, mely a vizsgált fehérje vagy miRNS első szomszédainak hálózatát mutatja. Alapértelmezetten csak a 0. és 1. réteg, a Signalink 2-nél tehát az útvonalak tagjai, és a hozzájuk kapcsolódó endocitotikus- és állványproteinek láthatók, mivel a legtöbb fehérjénél az összes kapcsolat megjelenítése átláthatatlan ábrát eredményezne. Az ábra fölött található menüből válogatva azonban bármely réteg kapcsolatai ábrázolhatók. Az ábra teljes böngészőablak méretűre nagyítható, az egyes fehérjékre kattintva pedig elérhető azok adatlapja. A hálózati ábra megjelenítését a JavaScriptben íródott Cytoscape Web végzi, mely JSON formátumban tölti le az adatokat.

Az NRF2ome adatbázisban az NRF2-nek kiemelt szerepe van, emiatt a weboldalon egyedi adatlapot hoztam létre: itt az előbb felsorolt információkon kívül böngészhetők a szabályozási hurkok (11. ábra), melyek transzkripció faktorok és miRNS-ek bevonásával visszacsatolásokat alkotnak a hálózatban (15. ábra).

The screenshot displays the NRF2ome web interface. At the top, the logo 'nrf2ome' is on the left, and navigation links 'nrf2 search download' and 'tools people contact' are on the right. Below the logo, there are tabs for 'Upstream' and 'Downstream', and a filter menu with categories: RTK/MAPK, Hedgehog, WNT, NHR, Notch, TGF, and JAK/STAT. A list of upstream proteins is shown, with 'MAPK14' selected. To the right, a detailed view of the interaction between MAPK14 and NFE2L2 is provided. It includes a diagram of an inhibitory interaction and a transcriptional regulation diagram. The transcriptional regulation section notes that NFE2L2 binds to the promoter of MAPK14. At the bottom, a 'Pathway connections (231)' section offers icons for different types of interactions: Protein-protein interactions, Transcriptional regulation, Post-transcriptional regulation, Pathway connections (highlighted in yellow), Regulatory loops, and Network image.

15. ábra. Az NRF2 kapcsolatait bemutató oldal. Az NRF2ome weboldalán egy külön nézet jeleníti meg az NRF2 kapcsolatait. Itt az alul látható menüből választva lehetséges az NRF2 poszt-transzlációs, transzkripcionális és poszt-transzkripcionális szabályozóit, a szabályozási hurkokat és a jelátviteli útvonalakkal fennálló kapcsolatokat böngészni. Ezen a képen az NRF2 RTK útvonalhoz tartozó, *upstream* szabályozóinak listája látható, a kijelölt kapcsolat részletes adatai a jobb oldalon jelennek meg (<http://nrf2.elte.hu/>).

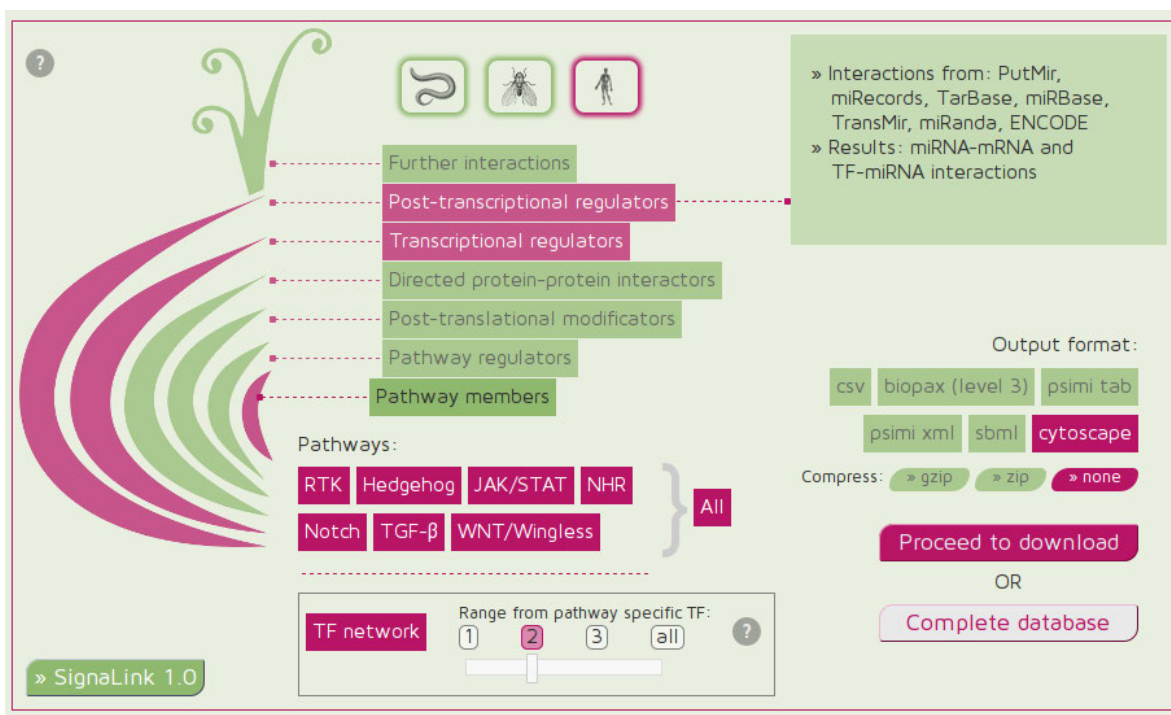
4.3.1.3. Személyre szabott letöltés. A SignaLink 2, az NRF2ome és az ARN rendelkezik letöltő modullal, melynek célja egyrészt többféle szabványos formátumban kinyerni az adatokat, másrészt a felhasználók számára lehetőséget biztosítani, hogy mélyebb informatikai tudás nélkül is precízen szűrjék az adatokat. A letöltés paramétereit beállító felület a weboldalon érhető el (16–19. ábrák). A beállítások két lépésben végezhetők el. Először a legalapvetőbb kérdésekben dönthet a felhasználó: mely rétegeket, a SignaLink 2 esetében mely

útvonalakat és mely fajok adatait tartalmazza a letöltés. Itt állítható be a kívánt formátum, és a tömörítés is (gzip vagy zip). A következő lépésben a korábbi beállítások is megváltoztathatók, de itt lehetőség van az interakciók részletes szűrésére források illetve megbízhatósági értékek alapján (17. ábra). Utóbbiakból tetszőleges limit állítható be, így csökkentve a hamis pozitívok arányát.

4.3.2. Letöltő modul

A letöltő modul a weboldaltól független program, de azonos adatbázisból dolgozik. Feldolgozza a felhasználó által a weboldal letöltő felületén beállított paramétereit; ezek alapján kiválogatja az adatbázisból a kért adatokat; végül ezeket megfelelő formátummá alakítja. Jelenleg hat féle, a molekuláris biológiában leggyakrabban használt, szabványos formátumot képes kiszolgálni: csv, BioPAX (level 3), PSIMI-TAB, PSIMI-XML, SBML és Cytoscape. Ezek bármelyike letölthető tömörítés nélkül, illetve gzip vagy zip által tömörítve. A kérések bonyolultságától függően egyes esetekben a letöltendő fájl elkészülte akár több percig is eltart. Ilyenkor igény szerint e-mail értesítés kérhető, mely tartalmazza a linket, ahol legalább egy hétig elérhető az exportált adat.

A letöltő modultól függetlenül a Signalink 2, az NRF2ome és az ARN letölthető SQL formátumú adatbázis képként. Ezt a fájlt a felhasználók egy MySQL szerverbe betöltve létrehozhatják az adatbázis saját példányát. Ez a lehetőség biztosítja legnagyobb szabadságot az adatok szűrésére, egyéni igények szerinti lekérdezések futtatására, és saját elemzések előkészítésére.



16. ábra. Letöltési beállítások a Signalink 2 weboldalon. A letöltés első lépésében a felhasználó kiválaszthatja a fajokat, a rétegeket és az útvonalakat. Ezután elérhetővé válnak a részletes beállítások (<http://signalink.org/>).

Detailed settings: miRNA regulators

Source	Select	Cut off
TargetScan	✓	0.50 0.5
Include non conserved targets:	✓	
TarBase	✓	No scores available
miRanda	✓	Scores of interactions from Miranda:
Alignment score	✗	
Free energy score	✗	
mirSVR score	✓	-0.50 -1.37 -0.01 -0.5
TransmiR	✓	No scores available
PuTmiR	✓	1.63 1.63 5.48 1.63
ENCODE	✓	No scores available

You are able to filter the interactions of Further interactions by score values or source of information. Description of scores and information sources are available by clicking on the question marks on the right. If You are not sure, the default settings usually give an optimal result.

Default Continue

Hedgehog ✓
JAK/STAT ✓

17. ábra. **Részletes szűrés a Signalink 2 weboldalon.** Ez a kép a poszt-transzkripcionális szabályozás beállításait mutatja, ahol kiválaszthatóak az egyes adatforrások. Ahol rendelkezésre állnak megbízhatósági mérőszámok, egy csúszka segítségével beállítható a kritikus érték, mely alatt a kapcsolatok kizárásra kerülnek (<http://signalink.org/>).

The screenshot shows the NRF2ome web application interface. At the top, there are four icons representing different data types: Protein-protein interactions, Transcriptional regulation, Post-transcriptional regulation, and Pathway connections. The 'Transcriptional regulation' option is selected. Below the icons, there are two checkboxes: 'Experimentally verified' and 'Predicted', both of which are checked. The 'Output format' section includes options for 'csv', 'biopax (level 3)', 'psimi tab', 'psimi xml', 'sbml', and 'cytoscape'. The 'Compress' section includes options for 'gzip', 'zip', and 'none'. At the bottom, there are two buttons: 'Proceed to download' and 'Complete database'.

18. ábra. **Letöltési beállítások az NRF2ome weboldalon.** Ez a felület egyszerű választást tesz lehetővé az adatbázis főbb elemei, a fehérje-fehérje interakciók, a transzkripcionális és poszt-transzkripcionális szabályozás, valamint a jelátviteli útvonalak között. Elkülöníthetők a kísérletesen igazolt kapcsolatok és a predikciók. Innen tovább haladva részletes, adatforrások szerinti szűrésre is lehetőség van (<http://nrf2.elte.hu/>).

The screenshot shows the ARN web application interface. At the top, there are four icons representing different data types: Pathway connections, Predicted post-translational regulators, TFs of miRNAs, and Transcriptional regulators. The 'Pathway connections' option is selected. Below the icons, there are two checkboxes: 'Experimentally verified' and 'Predicted', both of which are checked. The 'Output format' section includes options for 'csv', 'biopax (level 3)', 'psimi tab', 'psimi xml', 'sbml', and 'cytoscape'. The 'Compress' section includes options for 'gzip', 'zip', and 'none'. At the bottom, there are two buttons: 'Proceed to download' and 'Complete database'.

19. ábra. **Letöltési beállítások az ARN weboldalon.** Először a letölteni kívánt rétegek kerülnek kiválasztásra, majd ezeken belül adatforrásonként is lehet szűrni a kapcsolatokat (<http://arn.elte.hu/>).

4.3.3. Látogatottság monitorozás

A weboldalak látogatottságát a Piwik nevű programmal folyamatosan mérjük. A látogatókról minden elérhető adatot naplózunk, például földrajzi helyzetüket és a kliensek technikai paramétereit. A letöltések esetében kötelező megadni az E-mail címet és az intézményt, a név megadása opcionális. A Signalink 2 weboldalát naponta átlagosan 30-an keresik fel, a cikk publikálása óta több mint ezer látogató, a letöltő modulon keresztül pedig több mint 500 letöltés volt.

4.3.4. Kompatibilitás

A weboldalak működését Firefox, Chrome, Safari és Internet Explorer böngészőkben teszteltem, Linux, Mac OSX és Windows platformokon. A modern szabványokat követő böngészőkben az eddigi tesztek alapján a weboldalak minden funkciója működik. A „feedback” felületen keresztül jelentett kisebb hibákat folyamatosan javítani szoktam. Az Internet Explorer esetében, a 9.0-nál régebbi verziók elavultak és nem tartják be a szabványokat, így ezekben a böngészőkben megnyitva az oldalak hibaüzenetet adnak, felkínálva a lehetőséget, hogy a csökkent funkcionalitással együtt a felhasználó megpróbálja használni azokat.

4.4. Az adatbázisok elemzése

4.4.1. A Signalink 2 összehasonlítása más jelátviteli adatbázisokkal

Több adatbázis tartalmaz a Signalink 2-höz hasonló információkat. A 9. táblázatban öt ilyen adatbázist, és a Signalink 1-et hasonlítom össze a Signalink 2-vel. Mindegyik adatforrás számos értékes tulajdonsággal bír, de egyik sem képes a Signalink 2 összes fontos előnyét nyújtani. A Signalink 2 egyedülálló tulajdonságai, hogy 1) irodalmi gyűjtésből származó adatsort tartalmaz, nem csak emberből, hanem két másik modellorganizmusból is; 2) a nagy-teljesítményű kísérletek adatsorait egyesíti az irodalmi gyűjtésből nyert adatokkal; 3) megbízhatósági mérőszámokat tartalmaz az egyes interakciókhoz; 4) testreszabott letöltést biztosít többféle szabványos formátumon keresztül.

A legtöbb adatbázis a különféle adatokat egymással vegyítve tárolja, így nehézséget jelent ezek elkülönítése. A Signalink 2 az egyetlen adatbázis, ahol a szabályozás rétegei könnyen elkülöníthetők, és útvonalak szerint is szelektálhatók. Így lehetséges például csak egy kiválasztott útvonal transzkripciós szabályozását letölteni.

9. táblázat. A Signalink 2 összehasonlítása más jelátviteli adatbázisokkal (Fazekas és mtsai 2013)

	KEGG	Reactome	SPIKE	ConsensusPathDB	TranscriptomeBrowser	Signalink 1	Signalink 2
Modellorganizmusok				✓		✓	✓
Saját irodalmi gyűjtés	✓	✓	✓			✓	✓
Más adatbázisok integrálása				✓		✓	✓
Kapcsolatonkénti hivatkozások		✓	✓			✓	✓
Keresztbeszélgetések és <i>multi-pathway</i> fehérjék		✓				✓	✓
Irányított fehérje-fehérje interakciók	✓	✓		✓	✓		✓
Irányított fehérje-fehérje interakciók					✓		✓
Transzkripcionális szabályozás		✓		✓	✓		✓
Poszt-transzkripcionális szabályozás					✓		✓
mikro-RNS-ek transzkripcionális szabályozása							✓
Megbízhatósági mérőszámok							✓
Testreszabható letöltés					✓	✓	✓
Szabad felhasználhatóság	✓	✓	✓	✓	✓	✓	✓

A táblázatban szereplő adatbázisok hivatkozásai: ConsensusPathDB – Kamburov és mtsai 2010, KEGG – Kanehisa és mtsai 2005, Signalink 1 – Korcsmáros és mtsai 2010, SPIKE – Paz és mtsai 2012, TranscriptomeBrowser – Lopez és mtsai 2008

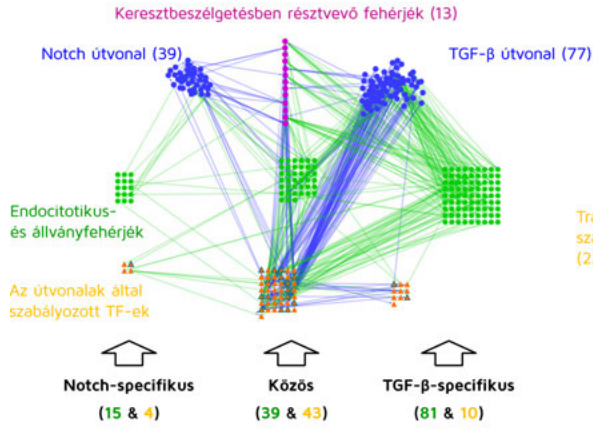
4.4.2. A jelátviteli keresztbeszélgetések különböző rétegei

A Signalink 2 alkalmas a jelátviteli útvonalak között a poszt-transzlációs-, a transzkripcionális és a poszt-transzkripcionális szabályozás szintjén megvalósuló keresztbeszélgetések tanulmányozására. Ennek bemutatására a Notch és a TGF- β útvonalat választottam ki. Ezen

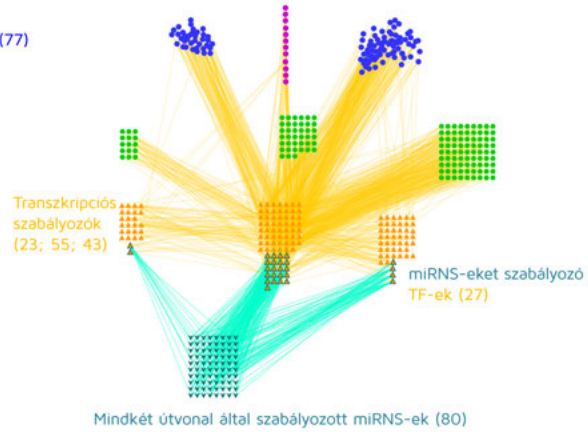
útvonalak eltérő típusú reakciókat tartalmaznak, de kapcsolatban állnak egymással. Funkcióik különböznek, de egyes funkciók megvalósításában együttműködnek (Guo és mtsai 2009; Klüppel és mtsai 2005). A két útvonal kapcsolatának elemzéséhez szükségünk van a poszt-transzlációs, transzkripcionális és poszt-transzkripcionális szabályozók kapcsolataira. Ezeket a hálózatokat láthatjuk a 20. ábrán, melyeket a weboldalon elérhető letöltő modullal nyertünk a SignaLink 2 adatbázisból. A 20.a. ábrán a két útvonalhoz tartozó fehérjék kékkel, poszt-transzlációs szabályozók zölddel, a hozzájuk kapcsolódó transzkripció faktorok narancs színnel vannak feltüntetve. Látható, hogy az útvonalak tagjai közül 13 vesz részt keresztbeszélgetésekben, ezek lila színnel vannak kiemelve. A letöltés során beállított szűréssel csak a kísérletesen igazolt TF–promóter kapcsolatokat válogattuk ki ebbe az adatsorba. A 20.b. ábrán az útvonalakat és poszt-transzlációs szabályozóikat szabályozó transzkripció faktorok, illetve a TF-ek által szabályozott miRNS-ek láthatók. A 20.c. ábrán az útvonal fehérjékre és szabályozókra ható miRNS-ek kapcsolatait mutatom be. Öt miRNS csak a TGF- β útvonal tagjait szabályozza, míg a többi 180 mindkét útvonal elemeire hatással van. A 20.d. ábrán a három hálózat egyesítve szerepel. Ez a hálózat számtalan szabályozási hurkot tartalmaz, melyekben az útvonalak által szabályozott TF-ek miRNS-ek transzkripciójára hatnak, amelyek aztán útvonal fehérjéket szabályoznak, vagy egyes TF-ek szabályozók transzkripcióját befolyásolják. Emellett nagy számú keresztbeszélgetés is megfigyelhető, a poszt-transzlációs, transzkripcionális és poszt-transzkripcionális szabályozás szintjein. Külön csoportosítottuk azt a 80 miRNS-t, amely mindkét útvonal szabályozása alatt áll, és mindkét útvonal fehérjéire hatással van. Ezek képviselik a keresztbeszélgetések poszt-transzkripcionális rétegét.

A miRNS-ek szerepe az útvonalak szabályozásában előtérbe került különböző daganattípusok, például hepatocelluláris karcinóma esetében (Pang és mtsai 2012). A SignaLink 2 segítségével összeállítható az egyes útvonalakat szabályozó, illetve a keresztbeszélgetésekben részt vevő miRNS-ek listája. Noha az itt felvázolt szabályozási hurkok és keresztbeszélgetések szövetspecifikusságára és kontextusfüggésére a SignaLink 2-ben nincsenek adatok, expressziós és proteomikai adatokkal kombinálva a SignaLink 2 hálózatát ez a hiányosság részben kiküszöbölhető. A SignaLink 2-vel végzett elemzések támogatják a kísérletes munkát, mely szükséges a keresztbeszélgetések és szabályozási hurkok funkciójának felderítéséhez.

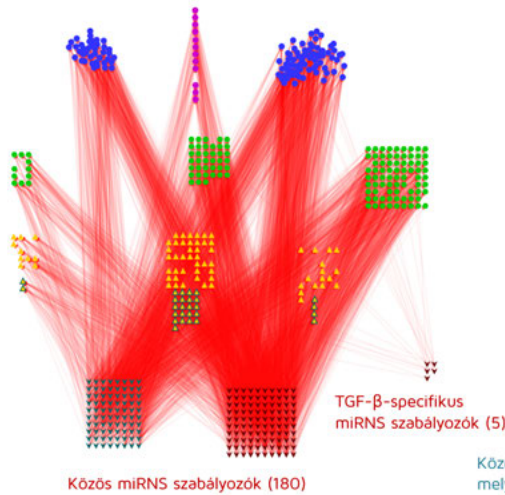
a) Fehérje-fehérje interakciók az útvonalak tagjai, a transzkripciós faktorok, az endocitotikus- és az állványfehérjék között



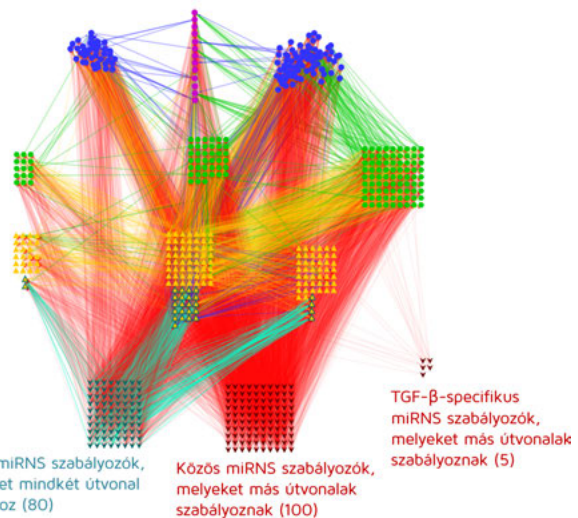
b) Az útvonalak tagjainak, az endocitotikus- és állványfehérjéknek, a transzkripciós faktoroknak és a miRNS-eknek a transzkripcionális szabályozása



c) Az útvonalak tagjainak, a transzkripciós faktoroknak, illetve az endocitotikus- és állványfehérjéknek a poszt-transzkripcionális szabályozása



d) Egyesített hálózati térkép: a Notch és a TGF- β útvonal közös szabályozási hálózata



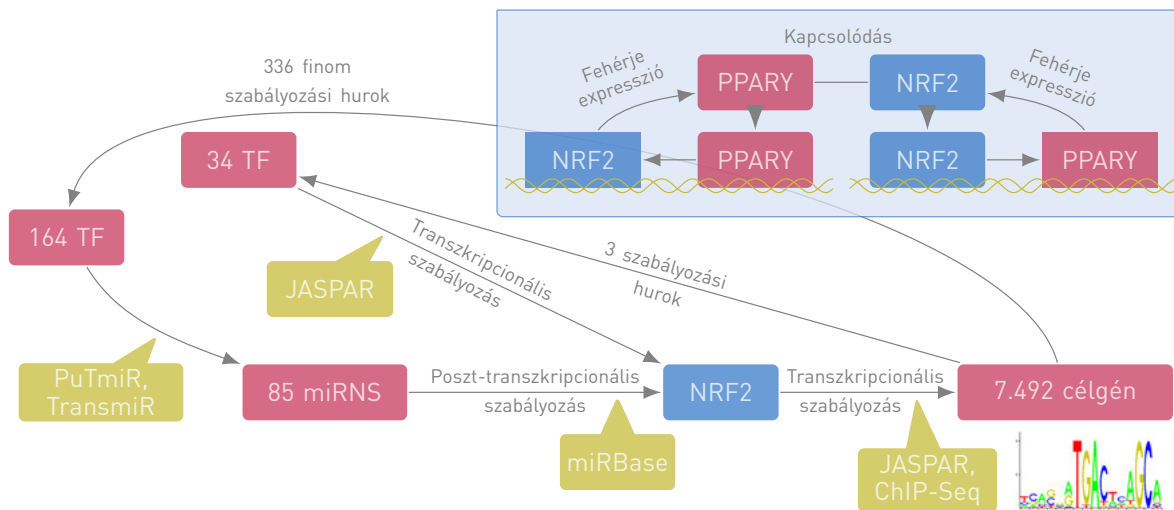
20. ábra. A TGF- β és a Notch útvonal szabályozása a **Signalink 2**-ben. Az első három ábra a szabályozás különböző rétegeit mutatja: (a) poszt-transzlációs szabályozók, (b) transzkripcionális szabályozók, (c) poszt-transzkripcionális szabályozók. A negyedik ábra a három szabályozási réteget egy hálózatként szemlélteti (Fazekas és mtsai 2013).

4.4.3. Szabályozási hurkok

Az előző fejezetben láthattuk a visszacsatolások tömeges jelenlétét az egyes fehérjék szabályozásában. Az NRF2ome adatai alapján összegyűjtöttem egyetlen fehérje, az NRF2 sa-

bályozásában szerepet játszó visszacsatolásokat. A 21. ábrán látható, hogy összesen három szabályozási hurok valósul meg a transzkripcionális szabályozás szintjén, azaz három NRF2 által szabályozott TF befolyásolja az NRF2 transzkripcióját. Ezek közül az egyik a PPAR γ , mely a 21. ábra bal felső sarkában kiemelve látható. A PPAR γ képes kötődni az NRF2-höz, így a transzkripcionális és poszt-transzlációs szabályozás e két fehérje közt egymással párhuzamosan érvényesül.

Nagyságrendekkel nagyobb számú, 336 hurok írható fel a poszt-transzkripcionális szabályozásban. Ezekben a hurkokban az NRF2 által transzkripcionálisan szabályozott TF-ek befolyásolják a miRNS-ek expresszióját, melyek szabályozzák az NRF2 transzlációját (22. ábra). Ezen szabályozási kapcsolatok relatív erősségének, szövet- és kontextusfüggésének értékelése a jövőben kísérletek elvégzésével lehetséges. Az NRF2ome weboldala lehetővé teszi az NRF2-t érintő szabályozás hurkok böngészését (11. ábra).



21. ábra. **Szabályozási hurkok az NRF2ome adatbázisban.** Az NRF2 7.492 célgénje közül 3 megtalálható az NRF2-t szabályozó 34 transzkripciós faktor között. Az NRF2 által szabályozott gének és az NRF2-t poszt-transzkripcionálisan szabályozó 164 miRNS kombinálásával 336, 3 lépésből álló szabályozási hurok állítható össze. Kísérletek szükségesek annak felderítésére, hogy az itt felvázolt lehetséges szabályozási mechanizmusok közül melyek játszanak igazán fontos szerepet az NRF2 precíz, kontextusfüggő szabályozásában. Az ábra jobb felső sarkában az NRF2 és a PPAR γ közti reciprok kapcsolat látható, mely a kölcsönös transzkripcionális szabályozásból, és a két fehérje közti kapcsolódásból tevődik össze.

5. Megbeszélés

5.1. A SignaLink 2 szerepe a jelátvitel kutatásában

A SignaLink 1-et több kutatásban felhasználták, és ezek a tanulmányok megmutatják a SignaLink 2 magját alkotó precízen definiált jelátviteli útvonalak előnyeit. A SignaLink útvonal definíciója lehetővé teszi, hogy az útvonalakat összehasonlító módon vizsgáljuk, és meghatározzuk a köztük zajló keresztbeszélgetéseket. Lázár és mtsai a SignaLink 1 felhasználásával a melanómában megváltozott útvonalak közti keresztbeszélgetéseket elemezték (Lázár és mtsai 2012). Kiderült, hogy a BRAF mutáns és nem BRAF mutáns melanóma sejtek között az EGF és JAK útvonalak közti keresztbeszélgetésekben tapasztalható a legnagyobb különbség. Az elemzés az egyes útvonalakon belüli és az útvonalak közti kapcsolatokban tapasztalt eltéréseket összehasonlítható, kvantitatív módon mutatta ki (Lázár és mtsai 2012).

Az egyes ráktípusokban kulcsszerepet játszó útvonalak megtalálása céljából Gu és mtsai ún. útvonal feldúsulási elemzést végeztek, mely az adott ráktípusban mutációt szenvedő, vagy megváltozott expressziójú gének arányát vizsgálja az egyes útvonalakban. Több ráktípus összehasonlító vizsgálata alapján megállapítható, hogy mely útvonalak érintettsége mutat összefüggést. Egy ilyen elemzés eredményéből következtethetünk az útvonalak közti funkcionális kapcsolatra (Gu és mtsai 2011). Megvizsgálták, hogy a *multi-pathway* fehérjék nagyobb eséllyel érintettek-e daganatos betegségekhez vezető mutációkban.

Több SignaLinket felhasználó tanulmány mutatja, hogy a megbízható, irodalmi gyűjtéssel készült jelátviteli hálózat szintén erős érv a SignaLink 2 használata mellett. Vinayagam és mtsai fehérje-fehérje kapcsolatok hatásának predikciójára használták a SignaLink 2 irodalmi gyűjtésből származó hálózatát, mint pozitív standardot (Vinayagam és mtsai 2013). Kondor és mtsai egy elméleti jelátviteli hálózatok dinamikájára vonatkozó szimuláció tesztelésére használta a SignaLink 1 adatait (Kondor és mtsai 2013). A vizsgálat igazolta, hogy a jelátviteli hálózatok a véletlenszerűen várhatónál kevesebb pozitív, öngerjesztő hurkot tartalmaznak, mely tulajdonságuk hozzájárul a stabil kikapcsolt (jelmentes) állapothoz, és ily módon a külső zaj folytán bekövetkező véletlen aktivitás kiküszöböléséhez. Khurana és mtsai a SignaLink 1 adatait felhasználva hálózati topológia és az emberi populációban tapasztalható genetikai

variabilitás mértékének összefüggését vizsgálták (Khurana és mtsai 2013). Eredményeik szerint az esszenciális gének egybeesnek a jelátviteli hálózatban nagy fokszámú fehérjékkel, ezeknél a géneknél kisebb variabilitás tapasztalható.

A felsorolt vizsgálatok megmutatják, hogy a SignaLink 1 fő újításai: az irodalmi gyűjtésen alapuló, keresztbeszélgetések elemzésére alkalmas hálózat hasznosnak bizonyultak az elemzések során. A SignaLink 2 még túl fiatal ahhoz, hogy publikációk születhettek volna a felhasználásával. Ugyanakkor a SignaLink 2 jelentősen kiterjeszti a lehetséges vizsgálatok körét. Az útvonalak továbbra is valós, funkcionális biológiai entitásokként kezelhetők (Korcsmáros és mtsai 2010). A köztük zajló keresztbeszélgetések azonban több szinten is vizsgálhatók. A poszt-transzlációs kapcsolatok mellett, létezhetnek keresztbeszélgetések a transzkripcionális szabályozás szintjén, amikor egy útvonal befolyásolja egy másik útvonal tagjának a transzkripcióját. Ugyanez lehetséges a poszt-transzkripcionális szabályozásban is, amennyiben egy útvonal olyan miRNS transzkripcióját szabályozza, amely egy másik útvonal tagjára van hatással. A különböző szinteken megnyilvánuló keresztbeszélgetések összehasonlító elemzése értékes többlet információt adhat az útvonalak funkcionális vizsgálatában.

A SignaLink 1 útvonalai esetében is megfigyelhetőek voltak visszacsatolások és szabályozási hurkok. A SignaLink 2 a poszt-transzlációs szabályozók többféle adatforrásból történő összegyűjtésével megnöveli ezen hálózati motívumok számát. Bár az egyes jelátviteli útvonalakra ható gének köre nem határolható le, a SignaLink 2 minden bizonnyal nagyobb részt foglal magában a 2.d. ábrán látható görbe meredek szakaszaiból, azaz az egyes útvonalakra erős hatással bíró komponensekből (Friedman és mtsai 2007). Emiatt a SignaLink 2 fehérje interakciós hálózatára épülő modellek nagyobb arányban fedhetik le a jelátviteli fehérjéket befolyásoló releváns tényezőket, és ennek köszönhetően megalapozottabb biológiai következtetésekre juthatnak.

A szabályozási hurkok vizsgálhatók a transzkripcionális szabályozás és a poszt-transzkripcionális szabályozás szintjén is. Erre az NRF2ome kapcsán, az 5.3. fejezetben mutatok konkrét példát. Ezen szabályozási kapcsolatok időbeli dinamikája nagyságrendekkel tér el a jelátviteli kémiai reakcióinak sebességétől (Papin és mtsai 2004). E szempont figyelembevétele fontos a

vizsgálatok során, emellett jelentős kihívás annak megértése, hogy ezek a visszacsatolások miként szabályozzák, finomhangolják a sejt jelátviteli hálózatának működését.

A jelátviteli hálózatok funkcióinak biztosításához, a robosztusság és a flexibilitás megteremtéséhez a jelátviteli fehérjék mennyiségének és aktivitásának precíz szabályozás alatt kell állniuk. E szabályozás megértése érdekében nem kerülhetjük el a transzkripcionális és poszt-transzkripcionális mechanizmusok vizsgálatát. A SignaLink 2 lényeges újítása, hogy ezeket az elemeket egységes rendszerben kezeli, megteremtve vizsgálatuk alapfeltételeit. Fontos ugyanakkor, hogy ezek a komponensek, a réteges felépítésnek köszönhetően tetszőlegesen elkülöníthetők, így egymástól függetlenül is vizsgálhatók a szabályozás egyes szintjei.

5.2. A SignaLink 2 jövőbeli fejlesztési irányai

A SignaLink következő kiadása, a SignaLink 3 fejlesztése a SignaLink 2 publikálását követően szinte azonnal elkezdődött. Az új verzió számos területen hoz majd előrelépést, azonban a kutatócsoport korlátozott erőforrásai miatt nem valósítható meg az összes kívánatos és hasznos újítás. Itt számba veszem a fejlesztés lehetséges irányait, függetlenül attól, hogy azok ténylegesen megvalósításra kerülnek-e majd a SignaLink 3-as kiadásában.

A főbb lehetőségek közül az egyik az adatbázis extenzív bővítése, azaz újabb fajok, útvonalak és rétegek hozzáadása. A fajokat tekintve, további fontos modell organizmusok, leginkább a zebradánió (*Danio rerio*), illetve a sörélesztő (*Saccharomyces cerevisiae*) hozzáadására jelentkezik a legnagyobb igény. Számos további jelátviteli útvonal, például az NF- κ B kidolgozása is fontos fejlesztési irány. Az új fajok és útvonalak hozzáadása saját irodalmi gyűjtést jelent, a már felállított protokoll szerint. Ez tehát egy igen munkaigényes, de értékes eredménnyel járó folyamat.

Az intenzív fejlesztési irányok egyrészt a már meglévő adatok frissítését, másrészt további annotációk hozzáadását jelenthetik. Számításba kell venni a közösségi alapú fejlesztés lehetőségét is, amit a feldolgozandó adatmennyiség egyre növekvő mérete, és a fejlesztői csapat véges kapacitása indokol.

- **Új adatforrások integrálása.** A rendelkezésre álló bioinformatikai adatok mennyisége egyre gyorsuló ütemben, folyamatosan növekszik. Újabb és hatékonyabb technikák jelentkeznek mind a laboratóriumi technológiák, különösen a *high-throughput* módszerek területén, és a bioinformatikai analízis matematikai eszköztárában egyaránt.
- **Kapcsolatok hatásának predikciója.** A Signalink 2-ben domén-domén interakciók alapján predikció készült a kapcsolatok irányára vonatkozóan. Azonban a másik fontos attribútum, a kapcsolat hatása – gátló vagy stimuláló – a nem irodalmi gyűjtésből származó kapcsolatoknál továbbra is ismeretlen. Egy friss tanulmány RNS-interferencia alapú nagy teljesítményű szűrés (*RNAi high-throughput screening*) adatok felhasználásával vizsgálta egyes gének különböző fenotípusokra gyakorolt hatása közti korrelációt. Ezeket az összefüggéseket az ismeretlen hatású kapcsolatokkal rendelkező fehérje interakciós hálózattal összevetve, lehetővé vált a kapcsolatok hatásának (gátló vagy serkentő) predikciója (Vinayagam és mtsai 2013).
- **Közösségi alapú adatgyűjtés.** A GEO expressziós és metilációs microarray adatokat tartalmazó adatbázis például 800.000 minta 13.000 labor által beküldött adatait tartalmazza (Barrett és mtsai 2013). A PAZAR adatbázis esetében a kutatók regisztrálhatnak, és feltölthetik a kísérleteikből származó adatokat (Portales-Casamar és mtsai 2009). Szintén nagyban alapoz a kutató közösség bevonására a Gene Ontology (Ashburner és mtsai 2000). A Signalink esetében is megvalósítható lenne egy felület, ahol új kapcsolatokat lehet hozzáadni a hálózathoz. Ezek a kapcsolatok egy új réteget alkotnának az adatbázisban, hiszen gyűjtésük más módszerekkel történne, mint a Signalink szabványos gyűjtési protokollja. Valószínűleg szükség lenne a beküldött adatok jóváhagyására, a megbízhatatlan adatok bekerülésének kivédésére. Ez a megoldás saját gyűjtési munkánkban is segítené, hiszen eleve rendelkezésre állna számos interakció és hivatkozás. Ezen kívül információt adna a kutató közösség igényeiről is, mert látható lenne, hogy mely útvonalak és mely fehérjék felé irányul a legnagyobb érdeklődés.
- **Adatfeldolgozási szolgáltatás.** Számos bioinformatikai szolgáltatás lehetőséget ad a kutatóknak saját adatok feltöltésére, elemzés céljából. Ezek a szolgáltatások általában

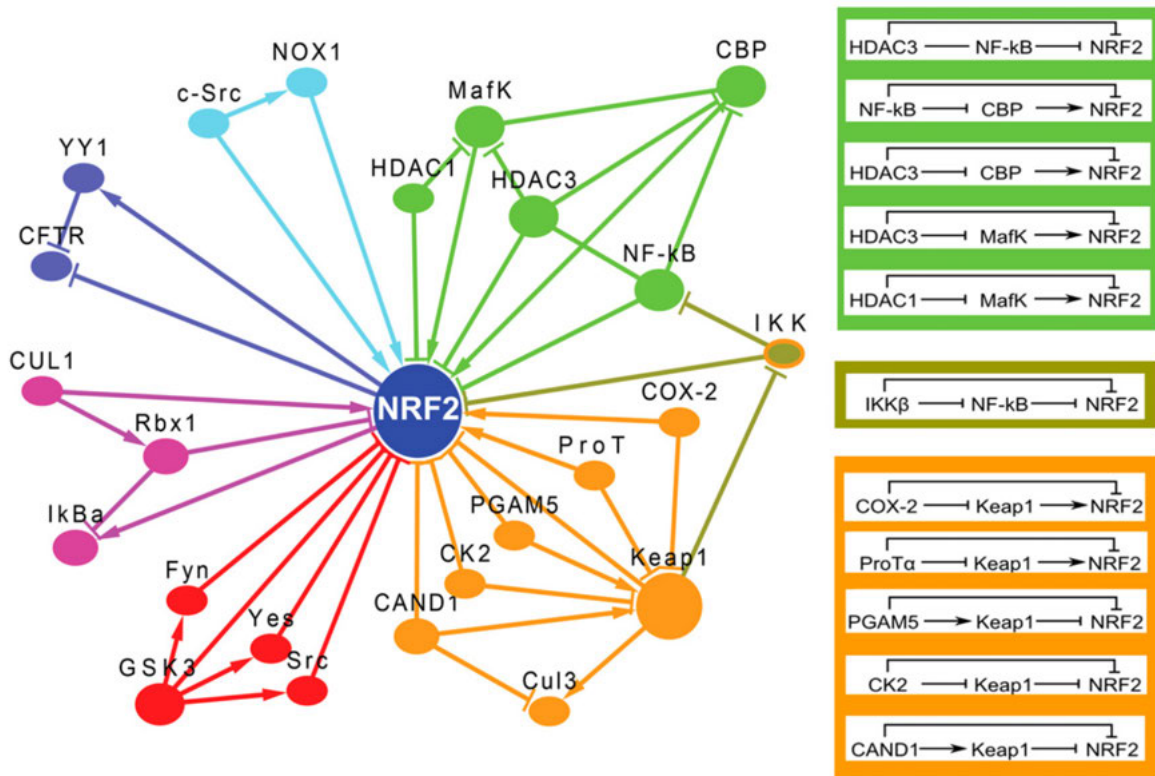
egy előre kidolgozott elemzést (*pipeline, workflow*) hajtanak végre tetszőleges adatokon, például egy fehérje listán. A GEO adatbázishoz kapcsolódó GEO2R webes felülete microarray adatok differenciál expressziós analízisére kínál egyszerű megoldást (Barrett és mtsai 2013). A Reactome weboldalán elérhető elemző modul gén listát vagy expressziós adatokat fogad bemenetként, és útvonal kapcsolatok listáját adja kimenetként (Croft és mtsai 2014). Az NRF2ome és az ARN egy már részben automatizált, tetszőleges fehérje listán alkalmazható adatintegrációs sémát követ. Ezt a módszert webes szolgáltatássá fejlesztve, lehetségessé válna a szabályozási rétegek és a jelátviteli útvonalakkal való kapcsolatok integrálása a felhasználó által megadott fehérje listára építve.

5.3. Az NRF2ome szerepe az NRF2 kutatásában

5.3.1. Poszt-transzlációs szabályozás

Az NRF2ome az első bioinformatikai adatforrás az NRF2 szabályozásával kapcsolatban – sőt, tudomásunk szerint egyetlen más fehérjéről sem készítettek hasonló adatbázist. Létrehozását az NRF2 jelentős patobiokémiai szerepe, és az NRF2 transzkripcionális szabályozása alatt álló gének nagy száma indokolta. Az elkészült adatbázis az NRF2 kísérletesen igazolt poszt-transzlációs szabályozóinak legteljesebb gyűjteménye: az NRF2ome 75 ilyen kapcsolatot tartalmaz, míg a BioGRID-ben 51, a MintActben 9, a HPRD-ben 15, az InnateDB-ben 47, a STRING-ben 10 kapcsolat szerepel. Az utolsó kivételével, ezen adatbázisok adatait az NRF2ome tartalmazza, és irodalmi gyűjtésből, valamint domén-domén és domén-motívum interakciókból származó predikciókkal egészíti ki (Türei és mtsai 2013).

Az NRF2 szabályozási hálózat vizsgálata során sikerült hét reciprok kölcsönhatást, azaz két elemű *feedback* motívumot felderíteni, melyek egyik résztvevője az NRF2, partnerei pedig a BRCA1, a KEAP1, az NFE2, a BACH1, az ATF4, a CUL3 és a NCOR1 (22. ábra). Három elemű *feed-forward* motívumból 23-at sikerült találni, melyben összesen 27 fehérje érintett az NRF2 mellett (Papp és mtsai 2012). Amikor a közvetlen és a közvetett hatás koherens, tehát egyaránt serkentő vagy gátló, ezek a motívumok a jelet felerősítik, és szabályozzák annak időbeli dinamikáját. Amikor a két hatás inkoherens, például közvetlen gátlás mellett közvetett



22. ábra. **Szabályozási hurkok az NRF2ome adatbázisban.** Az NRF2 7.492 célgénje közül 3 megtalálható az NRF2-t szabályozó 34 transzkripciófaktor között. Az NRF2 által szabályozott gének és az NRF2-t poszt-transzkripcionálisan szabályozó 164 miRNS kombinálásával 336, 3 lépésből álló szabályozási hurok állítható össze. Kísérletek szükségesek annak felderítésére, hogy az itt felvázolt lehetséges szabályozási mechanizmusok közül melyek játszanak igazán fontos szerepet az NRF2 precíz, kontextusfüggő szabályozásában. Az ábra jobb felső sarkában az NRF2 és a PPAR γ közti reciprok kapcsolat látható, mely a kölcsönös transzkripcionális szabályozásból, és a két fehérje közti kapcsolódásból tevődik össze (Papp és mtsai 2012).

serkentés valósul meg, hozzájárulnak az útvonal robusztusságához, és megvalósítják a kontextusfüggő szabályozást (Bleris és mtsai 2011). A CK2 például egyaránt gátolja a KEAP1-et és az NRF2-t, azonban a KEAP1 gátlása által pozitív hatást gyakorol az NRF2 aktivitására, míg a közvetlen gátlás az NRF2 felé kontextusfüggő (Pi és mtsai 2007). Három komponensű *feedback* motívumot nem sikerült az NRF2 szabályozási hálózatában felderíteni. Az NRF2ome az adatforrások egyesítésének és a kapcsolatok irány predikciójának köszönhetően, hatékony eszköz a szabályozás hálózati motívumainak felderítésére.

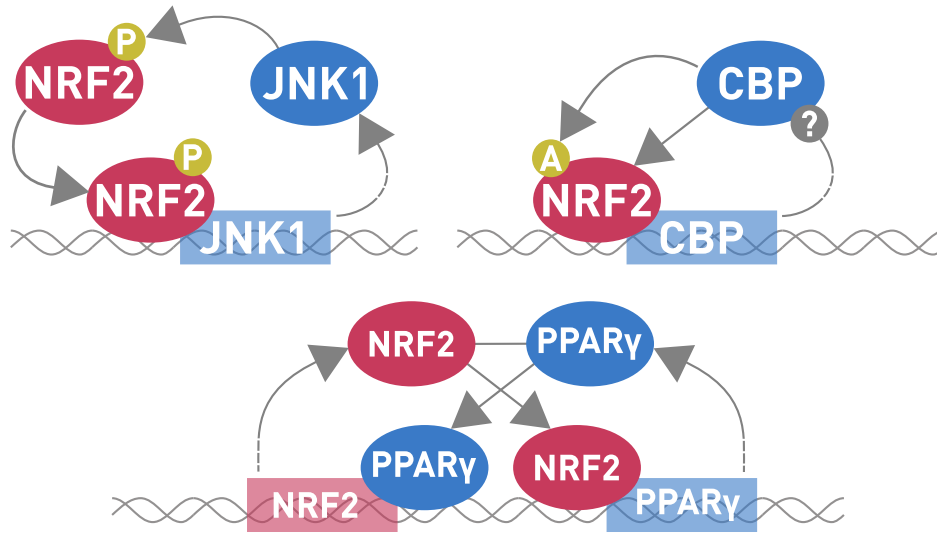
5.3.2. Transzkripcionális szabályozás

Az NRF2 szabályozásáról alkotott kép eddig főleg a KEAP1 általi gátlásra, és az ezt befolyásoló tényezőkre szorítkozott (Zhang 2006). Az NRF2^{ome} igyekszik az elérhető információk teljes körű összegyűjtésére az NRF2 és az interakciós partnerei transzkripció és poszt-transzkripcionális szabályozását illetően. Az NRF2 JASPAR predikcióval nyert transzkripcionális szabályozói közül tíz – MEF2A, ESR1, ESR2, NF- κ B, PPAR γ , SP1, NFE2L2, valamint a TMYC-MAX és EWSR1-FLI1 komplexek – esetében sikerült az irodalomban kísérletes igazolást találni (Papp és mtsai 2012), ami mutatja, hogy a predikció valóban képes hasznos adatokat szolgáltatni. Emiatt a további 24 prediktált transzkripcionális szabályozót érdemes lenne kísérletesen is megvizsgálni (Papp és mtsai 2012).

Az NRF2 által transzkripciósan szabályozott gének esetében figyelemre méltó a kis átfedés az irodalomból gyűjtött 29, az InnateDB-ben szereplő 6, a ChIP-Seq kísérletből származó 1.054, és a JASPAR predikcióból származó 6.426 célgén között. 44 gén szerepel a ChIP-Seq-ből és a JASPAR-ból származó adatsorokban egyaránt, a COX6C szabályozását pedig kísérlet is igazolja. Ez az eltérés valószínűleg a ChIP-Seq és a JASPAR eredményein alkalmazott szűrésnek, illetve a JASPAR esetében a hosszabb vizsgált szekvencia szakaszoknak tudható be. Emellett a ChIP-Seq módszer csak az NRF2 serkentő hatását képes kimutatni.

Az NRF2 interakciós partnerei közül összesen 39 fehérje áll az NRF2 transzkripcionális szabályozása alatt. Például a JNK1 foszforilálja az NRF2-t, melynek hatására az NRF2 a sejtmagba vándorol, és aktiválja a JNK1 transzkripcióját (23. ábra; Owuor és mtsai 2002). Hasonló módon, a p300/CBP promóteréhez kötődik az NRF2, ugyanakkor az NRF2-t acetilálva megakadályozza annak lebontását (23. ábra; Sun és mtsai 2009). Az NRF2 transzkripció célpontjait és interakciós partnereit összevetve, összesen 10 pozitív, 3 negatív, és 26 ismeretlen hatású *feedback* motívumot találtunk. Ezen mechanizmusok biológiai funkciójának megismeréséhez további kísérletes munka szükséges.

A 21. ábrán látható, hogy az NRF2 célgénjei közül három – az NFIL3, a PPAR γ és a RORA – szabályozza az NRF2 transzkripcióját. Az NRF2^{ome} adataiból kiderül, hogy az NRF2 és a PPAR γ kölcsönösen serkentik egymás transzkripcióját (Cho és mtsai 2010; Pi



23. ábra. **Hálózati motívumok az NRF2 transzkripcionális szabályozásában.** Az ábra felső részén két elemű motívumokra láthatunk példát. A JNK1 foszforiláció útján aktiválja az NRF2-t, mely ennek hatására serkenti a JNK1 transzkripcióját. A p300/CBP acetilálja az NRF2-t, és kötődik hozzá, így aktiválva azt. A predikció szerint az NRF2 kötődik a CBP promóteréhez. Ez esetben nem tudjuk azonban, hogy hatása serkentő vagy gátló. Alul látható a PPAR γ és az NRF2 közötti reciprok kapcsolat: a két fehérje kölcsönösen aktiválja egymást, és egymás transzkripcióját serkentik (Papp és mtsai 2012).

és mtsai 2010), miközben a két fehérje kötődik is egymáshoz (Ikeda és mtsai 2000). Ez a kapcsolat négy egymásba ágyazott, pozitív *feedback* motívumból, és két koherens, pozitív *feed-forward* motívumból áll. Ilyen módon komplex szabályozási funkció megvalósítására képes, melynek további kísérletes vizsgálata indokolt. Az NFIL3 és a RORA esetében nincs kísérletes bizonyíték az NRF2-vel való interakcióra. Az NFIL3 az immunválasz szempontjából fontos gének transzkripcióját szabályozza (Motomura és mtsai 2011), az NFIL2 és NFIL3, illetve NRF2 által alkotott visszacsatolási hurok fontos lehet az antioxidáns jelátvitel és az immunválasz összekapcsolásában. A RORA az anyagcsere gyulladással kapcsolatos szabályozásáért felelős (Jetten 2009), így az NRF2-vel való kapcsolatának szintén gyulladás szempontjából lehet funkciója.

5.3.3. Poszt-transzkripcionális szabályozás

Egyes miRNS-ek esetében (pl. miR-28, miR-144) kimutatták az NRF2-t gátló hatását, és ezek meghibásodásának szerepét betegségekben (Sangokoya és mtsai 2010; Yang és mtsai 2011). A többi prediktált miRNS szabályozó funkciójának megállapítása szintén kísérletes igazolásra vár.

Az NRF2-re ható miRNS-ek transzkripcióját 164 TF szabályozza, melyek közül 35 az NRF2 transzkripcionális szabályozása alatt áll, ezek közül 18 esetben serkentő, 17 esetben pedig ismeretlen az NRF2 hatása. A 21. ábrán bemutatott 336 finom szabályozási hurokban 63 miRNS vesz részt, mely 74 %-át alkotja az NRF2-t szabályozó összesen 85 miRNS-nek. Ezek közül négynek (miR-27a, miR-28, miR-93 és miR-144) az NRF2-vel való kapcsolata kísérletesen igazolt. Ezek a miRNS-ek 20 szabályozási hurkot alkotnak, összesen 14 TF-en keresztül. A MAFG transzkripció faktor például két miRNS expresszióját is szabályozza, emellett kísérletesen igazolt, hogy az NRF2 serkenti a MAFG transzkripcióját (Malhotra és mtsai 2010). A MAFG az NRF2-vel heterodimert képezve elősegíti a transzkripciót az ARE elemet tartalmazó promótereken (Zhang 2006). Azonban a MAFG homodimereket, más MAF proteinekkel pedig heterodimereket képez, így kompetitív úton gátolja az NRF2-t (Dhakshinamoorthy és mtsai 2000; Jaiswal 2004). A két említett szabályozási hurok felderítésének köszönhetően valószínűsíthető, hogy a MAFG a miR-93-on és a miR-144-en keresztül gátolja az NRF2 transzlációját.

A 336 szabályozási hurok vizsgálata során egyetlen TF, a TWIST1 esetében találtuk, hogy gátolja egy NRF2 transzlációját gátló miRNS – a miR-200a – transzkripcióját, így a hurok biztosan pozitív visszacsatolást jelent. Az összes többi hurokban a TF serkenti a miRNS-ek transzkripcióját, negatív visszacsatolást alkotva. A szabályozási hurkok nagy száma (336) egy komplex és differenciált szabályozási rendszer képét sejteti, mely képes az NRF2-t szövetspecifikusan, a stressztől függően, illetve más folyamatokkal összhangban szabályozni. Hasonló szabályozási rendszereket leírtak már más mester transzkripció faktorokkal kapcsolatban, mint a DAF-16/FOXO, vagy a HSF-1 (Doshi és mtsai 2008; Libina és mtsai 2003; Pirkkala és mtsai 2001). Az NRF2 szabályozásában résztvevő miRNS-ek és TF-ek expressziójának

időbeli dinamikáját és kontextusfüggését vizsgáló kísérletek fényt deríthetnének e rendszer működésére.

5.3.3.1. Az NRF2 jelátviteli útvonalak által történő szabályozása. Az NRF2ome alkalmas arra, hogy a SignaLink 2-ből származó jelátviteli útvonalak segítségével feltérképezzük az NRF2 választ a jelátvitel által közvetített hatásokra. A JAK/STAT útvonal például jelentős funkcionális átfedést mutat az NRF2-vel. Az JAK/STAT útvonal tagjai közül a STAT1 szabályozza az NRF2 transzkripcióját, míg a STAT3 kötődik az NRF2-höz (Murray 2007). Vírusfertőzés esetén a STAT1 aktiválódik az interferonok hatására (Kimura és mtsai 2006; Kosmider és mtsai 2012). Egerekben kimutatták, hogy NRF2 expressziója növekszik vírusfertőzés hatására, míg Nrf2 (-/-) egerek súlyosabb károsodásokat szenvedtek vírusfertőzésben (Cho és mtsai 2009).

A STAT3 kötődése az NRF2-höz domén-motívum interakciók alapján prediktált, erre nem sikerült kísérletes bizonyítékot találni. Ugyanakkor számos tény szól amellett, hogy ez a kapcsolat funkcióval bír az NRF2 szabályozásában. Alkohol hatása alatt, az NRF2 és a STAT3 egyaránt anti-inflammatorikus és antiapoptotikus szerepet tölt be (Miller és mtsai 2010; Wu és mtsai 2012). Májsejtekben az etanol okozta gyulladást és apoptózist a globuláris adiponektin gátolja, serkentve az 1-es hem-oxigenázt (HO-1), az NRF2-n és az IL10–STAT3 útvonalon keresztül (Mandal és mtsai 2010; Nepal és mtsai 2012). A HO-1 expressziója NRF2-től független módon is előidézhető (Nepal és mtsai 2012). Az IL10–STAT3 útvonal szerepének tisztázása az NRF2-függő HO-1 indukcióban, valamint annak megvizsgálása, hogy az IL10 képes-e aktiválni az NRF2-t, további kísérleteket igényel. Ugyanakkor elképzelhető, hogy az NRF2 és a STAT3 együttműködnek a gyulladási reakció gátlásában alkoholnak kitett májsejtek esetén.

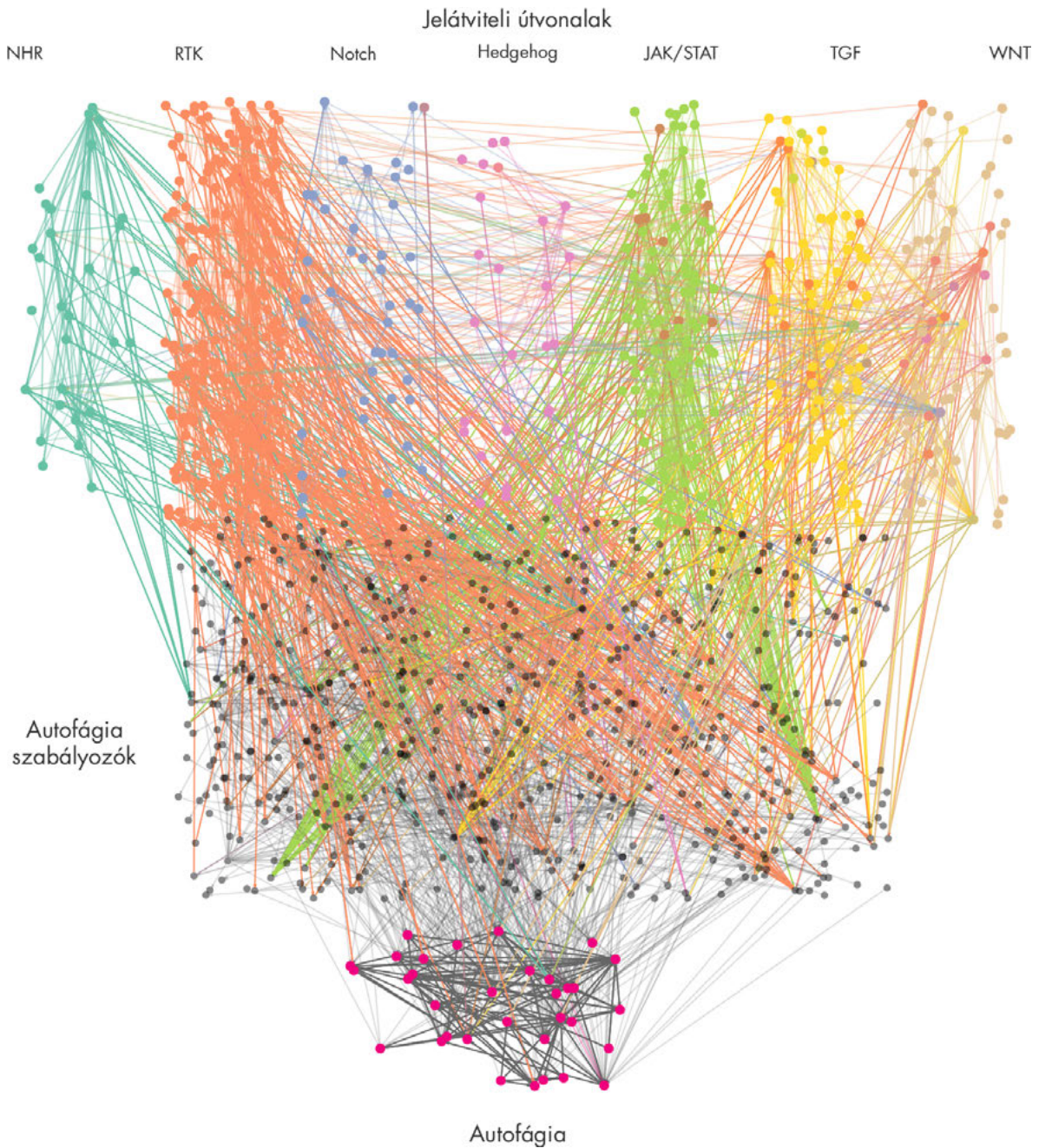
5.4. Az ARN jelentősége az autofágia kutatásában

Az ARN azzal a céllal jött létre, hogy az autofágia szabályozásának megértését előremozdítsa, egyrészt a kísérlettervezés támogatásával, másrészt bioinformatikai elemzések és modellezés alapjául szolgáló interakciós hálózat létrehozásával. Az ARN 38 fehérjét tekint az autofágia

végrehajtásában közvetlenül résztvevő apparátus részének. A nagy fehérje interakciós adatbázisok közül a STRING-ben 37 autofágia fehérje 335 kapcsolata, a BioGRID-ben 36 autofágia fehérje 641 kapcsolata, míg az IntAct-ban 36 autofágia fehérje 2.702 kapcsolata szerepel. Az autofágia-specifikus ADB adatbázisban (<http://tanpaku.org/autophagy/>, Homma és mtsai 2011) azonban csupán 28 autofágia fehérje 179 kapcsolatát találjuk, azonban ennek az adatbázisnak nagy hiányossága, hogy nem ad információt a kapcsolatok forrásáról. Valójában az ADB nem interakciós adatbázisnak készült, a célja inkább az autofágiához köthető gének összegyűjtése, és az ortológok megtalálása, rendszerezése. Csak néhány fehérjéhez kínál interakciókat, és hogy mi alapján választották a fejlesztők ezeket, arról szintén nincs információ. Az egyes fehérjék kapcsolatainak listája kis felugró ablakokban jelenik meg. Az ADB emellett nem kínál adatletöltési lehetőséget, így az adatok kinyerése sem egyszerű. Az elmondottak miatt az ADB nem alkalmas arra, hogy bioinformatikai elemzések alapjául szolgáljon, hiszen ennek előfeltétele, hogy ismerjük az adatok keletkezésének módját, gyűjtésük körét.

Az ARN a 38 autofágia fehérje 710 interakcióját tartalmazza, a transzkripció és a poszt-transzkripcionális szabályozásukkal együtt, emellett jó minőségű jelátviteli hálózatot is kínál (24. ábra). Ezzel jelenleg az elérhető legszélesebb körű autofágiára specializálódott bioinformatikai forrás. Az autofágia számos betegség patomechanizmusában játszott „kétélű kard” szerepe kihívás elé állítja az autofágiát célzó terápiákat. Emiatt szükséges az autofágia szabályozásának minden mechanizmusra kiterjedő, pontos megismerése, és a bioinformatika eszköztárának bevetése. A sikeres modellezés előfeltétele az *a priori* tudás összegyűjtése, az ARN elsősorban ezen a téren kívánja segíteni az autofágia kutatását. A réteges felépítésnek, a pontos forrás annotációknak, és a megbízhatósági mérőszámoknak köszönhetően, az ARN-t felhasználó elemzések készítői kiválaszthatják a számukra legmegfelelőbb adatokat, és összehasonlító jellegű elemzésekre is lehetőség nyílik.

A 4. táblázatban felsorolt, az autofágia fehérjék és jelátviteli útvonalak között kapcsolatot teremtő öt közvetlen poszt-transzlációs szabályozó közül a β -katenin kapcsolatait doménmotívum interakciók alapján prediktáltuk az ARN készítése során. Ugyanakkor ez a jelátviteli fehérje szabályozza közvetlenül a legtöbb autofágia fehérjét: az ATG4B-t, a GABARAP-ot, a p62-t, a Beclin-1-et és a PIK3C3-at. A β -katenin a WNT útvonal része, mely ligand kötődés



24. ábra. **A jelátviteli útvonalak és az autofágia fehérjék kapcsolatai.** Felül eltérő színekkel jelölve a Signalink 2-ből származó hét jelátviteli útvonal látható. A *multi-pathway* fehérjék az útvonalaiuknak megfelelő színátmenettel vannak színezve. Az jelátviteli fehérjékből kiinduló kapcsolatok az útvonalnak megfelelő színű, vastag vonalak. Alul látható az autofágia fehérjék hálózata. Az útvonalakhoz nem tartozó, de az autofágia fehérjéket és az útvonalakat két vagy több lépésben összekötő fehérjék interakciós hálózatát szürke pontok és kapcsolatok mutatják. A transzkripciós és poszt-transzkripcionális szabályozás, valamint a predikciókból származó kapcsolatok nem szerepelnek ezen az ábrán, azonban ezek még több kapcsolatot teremtenek a jelátvitel és az autofágia szabályozása között.

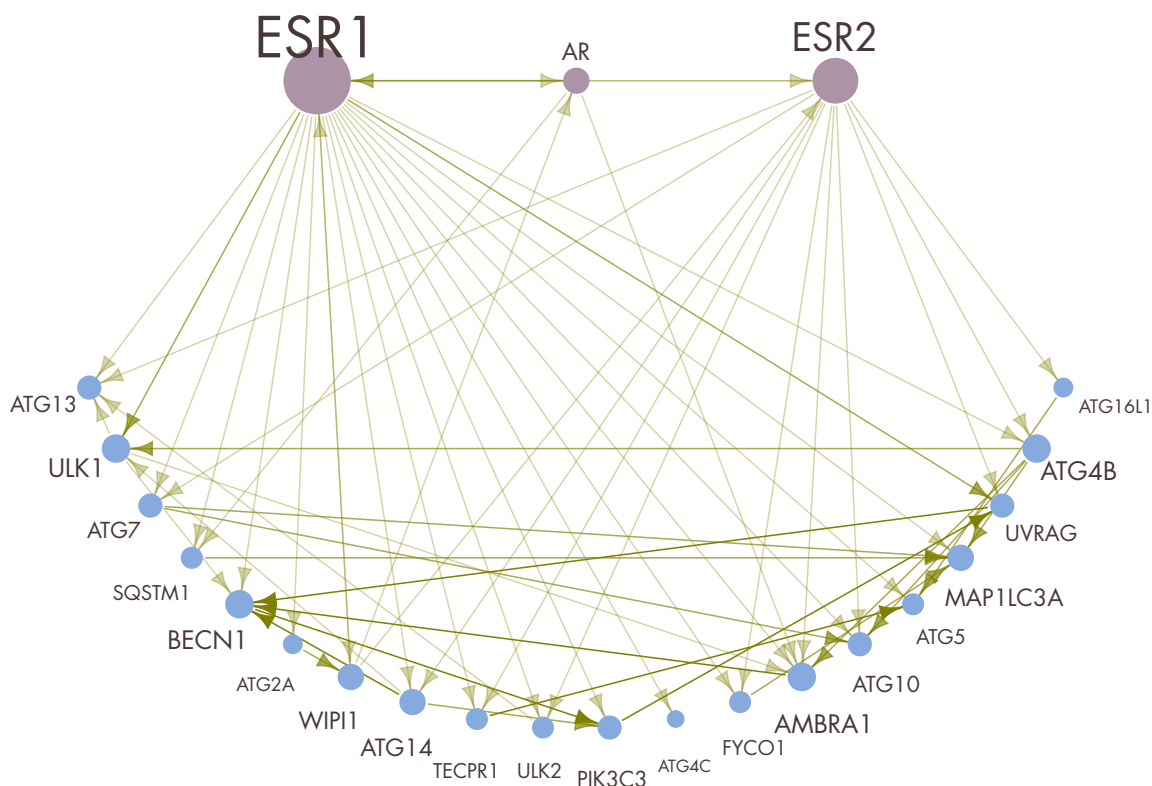
esetén aktiválódik, míg annak hiányában degradálódik. A β -katenin degradációjában fontos szerepet tölt be az autofágia: megtalálható benne egy LC3 interakciós régió (LIR) nevezett motívum, mely képes az LC3-II-höz, és valószínűleg más ATG8 homológokhoz (így a GABARAP-hoz is) kötődni, ebben az esetben a β -katenin autofagoszomális lebontásra kerül (Chang és mtsai 2013). Ezzel párhuzamosan, a WNT útvonalban a β -katenin felett elhelyezkedő, annak lebontását gátló DVL is az autofágia útján kerül lebontásra, hasonló mechanizmus által, összességében egy negatív koherens *feed-forward* motívumot alkotva a szabályozásban. Ez azonban csak az autofágia és a WNT- β -katenin útvonal közti keresztbeszélgetés egyik fele. A β -katenin a TCF4-en keresztül gátolja a p62 transzkripcióját, így gátolva az autofágiát (Petherick és mtsai 2013). Eddig láthattuk, hogy a domén-motívum predikció két olyan kapcsolatot eredményezett, melyek csak 2013-ban megjelent publikációkban kerültek kísérletes igazolásra. A β -katenin emellett az mTOR komplexről *upstream* pontokon is szabályozza az autofágiát. A β -katenin megemelkedett szintje befolyásolja a sejtek sugárterápiára adott választát, és általában rossz prognózist jelez. Laphámsejtes fej-nyak karcinómában a β -katenin RNS csendesítése hatására ezek a sejtek elpusztulnak. Ilyenkor megnő a Bax és a kaszpáz-3 expressziója, és a sejtek nagy arányban apoptózis útjára lépnek. Ugyanakkor a β -katenin csendesítése autofágiát is indukál: kimutatták, hogy megnő az LC3-II típusának aránya. A β -katenin feltehetően a PKA-LKB1-AMPK kaszkád aktiválása útján – mely valószínűleg az ATM, CAMKK vagy TAK1 kinázokon keresztül történik – serkenti az autofágiát. Ezt igazolja, hogy LKB1 csendesítéssel a β -katenin csendesítés hatása kivédhető volt (Chang és mtsai 2013).

A 4. táblázatban látható, hogy az mTOR útvonalban szereplő raptor és TSC1 a WNT útvonalnak is tagja. A WNT útvonal és az autofágia közti keresztbeszélgetés több ponton, a poszt-transzlációs és a transzkripcionális szabályozás szintjén egyaránt megvalósul. Az ARN lehetőséget ad az ezzel párhuzamosan működő poszt-transzkripcionális keresztbeszélgetések feltérképezésére: a WNT útvonal öt TF (CREB1, JUN, HNF1A, GATA3 és TCF4) 12 miRNS-en keresztül 35 autofágia fehérje szabályozására van hatással.

Az autofágia fehérjéket közvetlenül szabályozó transzkripciós faktorok közül a legtöbb az NHR útvonalhoz tartozik. A β -kateninben megtalálható LIR_NRBOX motívum lehetővé teszi,

hogyan az androgén receptorhoz és az ösztrogén receptorhoz kötődve, azok koaktivátoraként szolgáljon (Yang és mtsai 2002). Az ARN adatai szerint az autofágia fehérjék több mint felének, húsznak a promóterén található androgén vagy ösztrogén receptor kötőhely. Ismeretes, hogy több betegség esetén kimutattak nemi különbségeket az autofágia szintjében. Az ezen eltérések mögött álló mechanizmusokról azonban keveset tudunk (Lista és mtsai 2011). Szívizom szövetben és idegszövetben *ischaemia-reperfúziót* követően részben az autofágia eltérő szintje közvetíti az ösztrogén citoprotektív hatását (Bhupathy és mtsai 2010; Bouma és mtsai 2010; Campesi és mtsai 2013; Chen és mtsai 2013; Reichelt és mtsai 2013; Straface és mtsai 2013), ezért hímelekben magasabb az apoptózist szenvedő sejtek aránya (Weis és mtsai 2014; Zhu és mtsai 2006). Neurodegeneratív betegségekben szintén megfigyeltek nemi különbségeket az autofágiában (Barbati és mtsai 2012), ráadásul szinte az összes neurodegeneratív betegség jelentősen nagyobb arányban fordul elő nőknél (Lista és mtsai 2011). Egyes prosztaták sejttípusoknál az androgén jelátvitel jelentős szerepet játszik a túlélést és metasztázist segítő autofágia, és a tumor növekedését hátráltató apoptózis közötti döntésben (Wen és mtsai 2014).

Az ARN segít összekapcsolni a szexuáliszteroid jelátvitelt az autofágiával. A 25. ábrán láthatóak az androgén receptor és a két ösztrogén receptor transzkripcionális szabályozása alatt álló autofágia fehérjék. Az ULK1 és az UVRAG ESR1 általi szabályozását a HTRI adatbázisból átvett kísérletes adatok is alátámasztják, az összes többi TF-promóter kapcsolatot a JASPAR segítségével prediktáltuk. Érdekes módon a progeszteron receptor egyik autofágia fehérje transzkripcionális szabályozásában sem vesz részt. A 20 transzkripciósan szabályozott fehérje közül a WIPI1 LXXLL motívumot tartalmaz, így képes kötődni a szexuáliszteroid receptorokhoz, mely kötődés hormonfüggetlen (Proikas-Cezanne és mtsai 2004). A WIPI1 lokalizációja függ az autofágia aktivitásától, miközben a WIPI1 az ösztrogén és tesztoszteron receptorokhoz kötődve befolyásolja a szexuáliszteroid jelátvitelt, mely számos autofágia fehérje, köztük a WIPI1 transzkripcióját is szabályozza.



25. ábra. **Az autofágia szexuáliszteroid receptorok általi szabályozása.** A szaggatott vonalak a TF–promóter kötődéseket, míg a folytonos vonalak a fehérje-fehérje interakciókat jelölik. A vastagabb szaggatott vonalakkal jelölt kapcsolatokról kísérletes adat található a HTRI adatbázisban, a vékonyabbak a JASPAR algoritmussal készült predikcióból származnak.

5.5. Elemzési lehetőségek a bemutatott adatbázisok felhasználásával

A jelátviteli adatbázisok lehetővé teszik a matematikai modellezés eszköztárának használatát a jelátvitel vizsgálatára. A modellekből laboratóriumban tesztelhető hipotéziseket generálhatunk (Bauer-Mehren és mtsai 2009).

5.5.1. Differenciál expressziós vizsgálatok

A DNS microarray-ek elterjedésének, és a belőlük származó kísérletes adatok publikus hozzáférhetőségének köszönhetően, számos szövetből és sejttípusból, számos daganatból és beteg mintából állnak rendelkezésre expressziós adatok. A dolgozatomban bemutatott adatbázisok

nem szövetspecifikusak, de a bennük szereplő interakciós hálózatot expressziós vizsgálatok adataival kombinálva szövetspecifikus hálózatot nyerhetünk. A microarray-ek leolvasásával nyert adatokat a GEO (Barrett és mtsai 2013), illetve az ArrayExpress (Rustici és mtsai 2012) adatbázisokba töltik fel a laboratóriumok, a MIAME (*Minimal Information About Microarray Experiment*) szabványnak megfelelő információkkal. Az adatok kezeléséhez mindkét adatbázis kínál egy R Bioconductor csomagot: ezek a GEOquery (Davis és mtsai 2007), illetve az ArrayExpress (Kauffmann és mtsai 2009). A nyers adatok normalizálását és háttér korrekcióját követően, az összehasonlítani kívánt csoportok közt többszörös lineáris regresszióval, vagy Bayes statisztikai módszerekkel (Smyth 2005) állapítható meg az egyes gének mRNS-szintjeiben mutatkozó eltérés. Ezután vizsgálható hálózatban – mely lehet a Signalink, bármely más fehérje interakciós adatbázis, vagy annak egy része – szereplő fehérjék eltérő expressziójának mintázata. Összevethető gráfelméleti mérőszámokkal, vagy akár különböző daganat típusok között is végezhető összehasonlítás.

5.5.2. Mutációs vizsgálatok

Hasonlóan az expressziós mintázatok elemzéséhez, az egyes tumorokhoz társítható mutációk által érintett fehérjék hálózati pozíciója is vizsgálható. A legnagyobb tumor mutációs adatbázis, a COSMIC (Catalogue of Somatic Mutations in Cancer) 18.490 gén 542.000 mintában talált 136.000 mutációjáról tartalmaz adatokat (Forbes és mtsai 2010). A mutációs adatok és a szabályozási hálózatok kombinálásával vizsgálható a mutációt szenvedett gének útvonal tagsága, szerepe a keresztbeszélgetésekben, szabályozási folyamatokban. A 26. ábrán az autofágia fehérjék poszt-transzlációs és transzkripcionális szabályozóira illesztett, több daganattípusból származó mutációs adatok láthatók. A zöld és bíbor színű pontok szenvednek mutációt a vizsgált tumorokban.

A mutációt szenvedő fehérjék funkciójának elvesztése gyakran nagy befolyással bír az egész hálózat szerkezetére. A Signalink 2 jelátviteli hálózatának vizsgálata során kiderült, hogy egyes rákban érintett fehérjék intermoduláris pozícióban vannak, így kiesésük megváltoztatja a hálózat nagy moduljai közti összeköttetést.

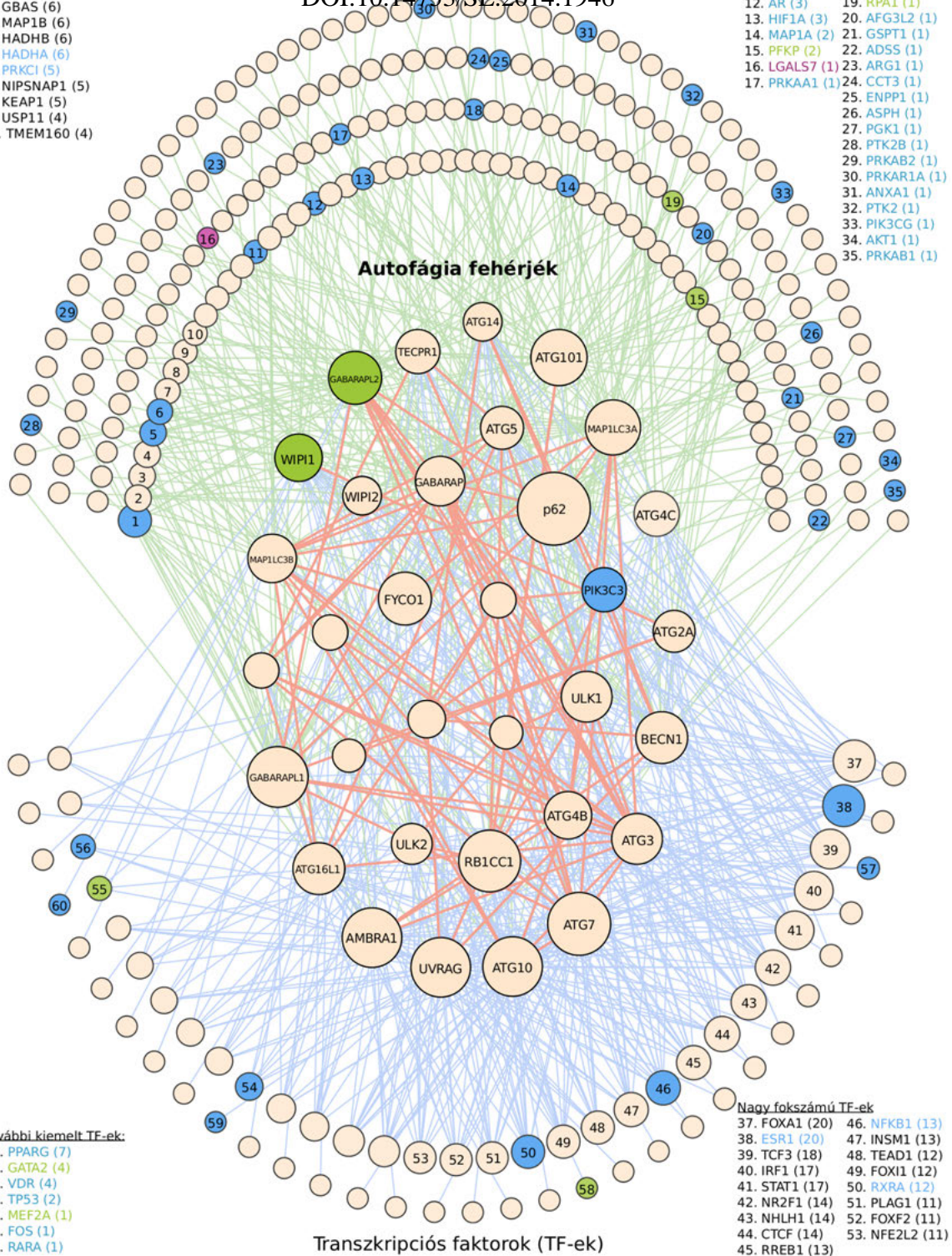
Nagy fokszámú partnerek:

1. ESR2 (12)
2. GBAS (6)
3. MAP1B (6)
4. HADHB (6)
5. HADHA (6)
6. PRKCI (5)
7. NIPSNAP1 (5)
8. KEAP1 (5)
9. USP11 (4)
10. TMEM160 (4)

Fehérje interakciós partnerek
DOI:10.14753/SE.2014.1946

Egyéb kiemelt partnerek:

11. MTOR (3)
12. AR (3)
13. HIF1A (3)
14. MAP1A (2)
15. PFKFB (2)
16. LGALS7 (1)
17. PRKAA1 (1)
18. BCL2L1 (1)
19. RPA1 (1)
20. AFG3L2 (1)
21. GSPT1 (1)
22. ADSS (1)
23. ARG1 (1)
24. CCT3 (1)
25. ENPP1 (1)
26. ASPH (1)
27. PGK1 (1)
28. PTK2B (1)
29. PRKAB2 (1)
30. PRKAR1A (1)
31. ANXA1 (1)
32. PTK2 (1)
33. PIK3CG (1)
34. AKT1 (1)
35. PRKAB1 (1)



26. ábra. **Rákban mutációt szenvedő gének és gyógyszercélpontok az autofágia szabályozási hálózatában.** Középen az autofágiát végrehajtó fehérjék, felül a poszt-transzlációs szabályozók, alul pedig a transzkripciós faktoraik láthatók. A pontok nagysága fokszámukkal arányos, azaz hogy hány autofágia fehérjét szabályoznak. A nagy fokszámú pontok külön listázva vannak. A daganatos megbetegedésekben mutációt szenvedő gének zölddel, míg a gyógyszercélpontok kézzel kiemelve szerepelnek. A szabályozók közti kapcsolatok nincsenek feltüntetve az ábrán (Kubisch és mtsai 2013).

5.5.3. Hálózati perturbációs szimulációk

Hálózatok vizsgálatára alkalmas módszer az egy-egy csomópontból kiinduló perturbációk terjedésének számítógépes szimulációja. A Semmelweis Egyetem LINK Group kutatócsoportjában, Szalay Kristóf által kifejlesztett szoftver, a Turbine alkalmas egy hálózat minden csomópontjából, meghatározott karakterisztikával terjedő perturbációk szimulációjára (Szalay és mtsai 2013). A szimuláció alapján minden csomóponthoz egy elcsendesedési időt (*silencing time*) rendelhetünk, mely megadja, hogy az adott pontból kiinduló perturbáció milyen gyorsan nyelődik el a hálózatban. Ennek reciproka a perturbációs centralitás index, mely nagy fokszámú pontok és hálózati modulok között elhelyezkedő pontok esetében magas értéket ad. A Turbine-t bemutató publikációban az élesztő interaktómján kerül lefuttatásra a perturbációs szimuláció, expressziós mintázatok alapján összehasonlítva a hőstressznek kitett élesztő interaktómját a kontroll populációéval (Szalay és mtsai 2013).

A perturbáció alkalmas gyógszerhatóanyagok hatásának modellezésére. Ennek érdekében az általam kifejlesztett szabályozási hálózatokat hatóanyag-célpont kapcsolatokkal érdemes kombinálni, melyek például a DrugBank adatbázisból nyerhetők (Knox és mtsai 2012).

5.5.4. Modularizációs vizsgálatok

A hálózatok modularizálása során arra törekszünk, hogy a csomópontok olyan csoportjait alakítsuk ki, melyek között a csoporton belül több kapcsolat, szorosabb összeköttetés van, mint a hálózat csoporton kívül eső pontjaival. A modulok lehetnek diszkrét vagy átfedőek. Már igen kis hálózat méret felett a számítási kapacitás korlátaiba ütközik a probléma egzakt megoldása, ezért számos heurisztikus algoritmust fejlesztettek ki a jó megoldás hatékony közelítésére (ezek átfogó összehasonlítása megtalálható a Kovács és mtsai 2010 S1 függelékében). A modularizálás elősegíti a hálózat topológiájának és dinamikájának megértését, illetve a modellezést. Egy diszkrét modulokra osztott hálózat esetében lehetőség van a modulokon belüli alhálózatok külön modellezésére, például reakciókinetikai differenciálegyenletekkel (Li és mtsai 2012). Az idézett tanulmányban a TGF- β útvonalat – nem matematikai modularizáló eljárásokkal, hanem önkényesen – három modulra osztották, és ezek mindegyikét

külön modellezték. Hasonló elemzést alkalmaz Harrington és mtsai 2008 az apoptózis fehérje interakciós hálózatának modellezésére.

A LINK Group kutatócsoportban fejlesztett ModuLand programcsomag (Kovács és mtsai 2010; Szalay-Beko és mtsai 2012) átfedő modulokat hoz létre, melyek hálózata akár hierarchikus módon tovább modularizálható. Minden modulnak egy meghatározott csomópont a középpontja, és minden további pontnak számszerűsíthető az egyes modulokhoz az affinitása.

A Signalink útvonal fehérjék közti kapcsolatokat tartalmazó, 0. réteget modularizálva, a legtöbb modularizáló algoritmus által talált modulok megfelelnek az egyes útvonalaknak. Ez arra utal, hogy a jelátviteli útvonalak nem önkényesen meghatározott, mesterséges kategóriák, hanem valóban létező entitások. A Signalink 2 többrétegű hálózatának vizsgálatával szintén lehetséges lenne funkcionális szempontból releváns csoportokat azonosítani. Érdekes lenne a szabályozás egyes rétegeinek bevonását összehasonlító módon vizsgálni.

5.5.5. Kritikus pontok vizsgálata

Az első jelátviteli útvonalak kísérletes leírásával lineáris kaszkádok képe alakult ki, melyben az útvonal egyes elemei egymásnak továbbítják a receptortól eredő jelet. Idővel láthatóvá vált, hogy az útvonalak között számos kapcsolat, keresztbeszélgetés (*cross-talk*) létezik, így ezek egy összefüggő jelátviteli hálózatot alkotnak. A hálózat elemei továbbítják és módosítják a receptor-ligand interakcióból eredő jelet (Taniguchi és mtsai 2006).

A kritikus pontok (*critical node proteins*) két vagy több fehérjéből álló csoportok, melyek nélkülözhetetlenek a receptortól eredő jel továbbításában (gyakran izoformák); melyek közül legalább kettő eltérő biológiai funkciót valósít meg, így a jel továbbhaladásának irányában döntő szerepet játszanak; melyek erős negatív és pozitív szabályozás alatt állnak; valamint útvonalak közti keresztbeszélgetéseket valósítanak meg. A kritikus pontok koncepciója segít leegyszerűsíteni a jelátviteli hálózatok reprezentációját, melyekben a kombinatorikus úton megvalósuló pályák száma igen nagy értéket ér el (Taniguchi és mtsai 2006).

5.5.6. Hálózati motívumok keresése

Egyes mintázatok, mint a hálózati motívumok, gyakrabban fordulnak elő a jelátviteli és szabályozási hálózatokban, mint a véletlenszerű gráfokban. A motívumok rendszerint két, három vagy négy pontot (fehérjét) magukban foglaló mintázatok, egy többféle élt (pl. serkentés vagy gátlás, fehérje-fehérje kapcsolat és transzkripcionális szabályozás) tartalmazó hálózatban. A négy fehérjéből álló motívumok többsége kisebb motívumokból épül fel. Az egyes motívumok az interakciós hálózat ismétlődő funkcionális egységei, például protein komplexek, közös transzkripcionális szabályozás alatt álló és egymással kölcsönható fehérjék, egy géneket szabályozó transzkripció faktorok, negatív visszacsatolás, pozitív visszacsatolás, ritmus generátor, stb. Az átfedő motívumok kombinálásával a fehérjék funkcionális moduljai alakíthatóak ki (Yeager-Lotem és mtsai 2004).

A hálózati motívumok meghatározzák a jelátviteli hálózat viselkedését periodikus stimulációk alatt. A negatív visszacsatolásokat tartalmazó kaszkádok (negatív *feed-forward* hurok) frekvencia szűrőként viselkedhetnek, míg például a közvetlen negatív visszacsatolás oszcillátorként működhet. A hálózati motívumokat definiálhatjuk meghatározott információ feldolgozási tulajdonságokkal rendelkező hálózati mintázatokként (Cournac és mtsai 2009).

Ahogy a 5.3. fejezetben az NRF2ome példáján bemutattam, az általam kifejlesztett adatbázisok lehetőséget adnak a hálózati motívumok azonosítására, nem csak a fehérjék közti interakciók szintjén, hanem a transzkripcionális és poszt-transzkripcionális szabályozás bevonásával is.

5.5.7. Extrém útvonal analízis

A hálózati motívumok egy másik definíciója korrelált reakciók csoportjainak (*co-sets*) tekinti azokat, így sztöchiometriai alapon, állandó arányú fluxusok szerint azonosítja őket, vagy reakciók olyan csoportjaként, melyek mindig együtt fordulnak elő, azaz olyan reakciókként, melyek feltételezik más reakciók végbemenetelét. Ilyen átfedő motívumok figyelhetők meg a JAK útvonal esetében, ahol a prolaktin receptor a JAK1 kináz útján továbbítja a jelet, míg az interleukin-4 receptor a JAK1 és a JAK3 kináz útján egyaránt. (Papin és mtsai 2004). Ezen

fehérje csoportok aktivitásának koordinált szabályozás alatt kell állnia. Az interakciós hálózat funkcionális moduljainak pontos ismerete távlatokat nyithat ezen modulok kombinálása, és a hálózatok mérnöki tervezése felé (Papin és mtsai 2004).

A jelátviteli útvonalak sztöchiometriai mátrixából lineáris algebrai módszerekkel azonosíthatók az extrém útvonalak, melyek a *steady-state* fluxus eloszlások lineárisan független vektoraiból állnak, és kombinációjukként a hálózat minden útvonala előállítható. Ez a megközelítés a *cross-talk*okat az extrém útvonalak nem negatív lineáris kombinációjaként definiálja. Az extrém útvonalak rendelkezhetnek azonos, átfedő, vagy disztinkt bemenettel, és azonos vagy eltérő kimenettel. Az egymással *cross-talk*okat képező extrém útvonalak kombinációi különböző biológiai funkcióknak feleltethetők meg. Ily módon a jelátviteli hálózat képes ezen eltérő funkciókat gazdaságosan megvalósítani: többnyire azonos reakciók felhasználásával, a bemenetben tapasztalható kis különbségek nyomán más választ produkálni. További rendszerszintű tulajdonság az útvonalakban mutatkozó redundancia, mely az azonos bemenettel és kimenettel rendelkező extrém útvonalakat jelenti, míg a bemeneti redundancia az azonos bemenettel rendelkező extrém útvonalak párját. Az egyes reakciók esetében megállapítható azon extrém útvonalak száma, melyben az adott reakció részt vesz. Azok a reakciók, melyek minden őket érintő extrém útvonalban együtt szerepelnek, korrelált reakciók csoportjait, azaz *co-set*eket alkotnak. A *co-set*ek tagjai nagy valószínűségek esnek azonos transzkripcionális szabályozás alá.

A Signalink 2-ben szereplő topológiai tulajdonságok annotációja az extrém útvonal elemzés során lehetővé teszi a bemenetek (receptorok és ligandumok interakciói), illetve a kimenetek (transzkripciós faktorok) azonosítását. Az adatbázisainkban szereplő transzkripciós szabályozási réteg segítségével vizsgálható a *co-set*ek közös transzkripciós szabályozása.

5.5.8. A hálózatok szerepe a gyógyszerkutatásban

A gyógyszerhatóanyag keresésben egyre inkább előtérbe kerülnek a rendszerbiológiai szemléletű módszerek. A polifarmakológia azon a felismerésen alapul, hogy számos hatóanyag nem egy célpontra hat szelektív módon, hanem több célpont párhuzamos szabályozásával éri el a kívánt hatást. A hálózati farmakológia ezt a megközelítést ötvözi a hálózati biológiával,

figyelembe véve a hatóanyagok hatásának az interakciós és szabályozási hálózatokon történő terjedését, az így érvényesülő közvetett hatásokat. A hálózatok és a rájuk épülő modellek lehetővé teszik az akár több ponton beavatkozó hatóanyagok hatásának szimulációját, így segítve a döntéshozatalt a kísérletes munkában (Csermely és mtsai 2005; Hopkins 2008). Hasonló eszközöket igényel a hatóanyag kombinációs terápiák tervezése, amikor több hatóanyag együttes alkalmazásával több célpontra gyakorlunk hatást az interakciós hálózatban (Korcsmaros és mtsai 2007). Az egy célpontra specifikus hatóanyagok sokszor nem hozzák a várt hatást, mivel a befolyásolni kívánt bonyolult rendszerben általában vannak redundanciák és visszacsatolások, melyek kompenzálják a hatóanyag hatását. A több célpontú és több hatóanyagú terápiák azonban ilyen feltételek mellett is hatásosak lehetnek (Csermely és mtsai 2005). A hálózati biológia által támogatott másik farmakológiai koncepció az *allo-network* célpontok keresése. Ebben az esetben nem az elsődleges célpontokat, hanem az azokkal kölcsönhatásban lévő más fehérjéket, illetve RNS-eket vesszük célba hatóanyagokkal (Nussinov és mtsai 2011).

Az NRF2ome hozzájárul az *allo-network* gyógyszer-célpont kereséshez, ami különösen fontos az NRF2 „kétélű kard” szerepének megfelelő kezelésére a terápiák során. Az NRF2 egészséges sejtekben segít megelőzni egyes daganattípusok kialakulását, míg folyamatos aktivitása hozzájárul a tumorsejtek kemoterápiával szembeni rezisztenciájához.

Az autofágia terápiás célpontként való felhasználása – e folyamat rákban játszott ellentmondásos szerepe miatt – hasonló kihívások elé állítja a kutatókat. Az ARN hasznos adatforrás az *allo-network* és *multi-target* gyógyszer-célpontok azonosításában, és az egyes terápiák hatásának modellezésében is kiindulópontként szolgálhat. Az interakciós hálózat és gyógyszer-célpont adatok kombinálására a 26. ábrán láthatunk példát: itt kék színnel vannak kiemelve az autofágia fehérjék ARN-ből származó szabályozási hálózatában a gyógyszer-célpontok.

A Signalink 2 egységes jelátviteli hálózata, a több szabályozási réteggel együtt alkalmas a hatóanyagok hatásának modellezésére, például perturbációs szimulációk segítségével. Az egyes célpontok hatásának becslésével csökkenthető a nem várt keresztbeszélgetések miatt bekövetkező kedvezőtlen mellékhatások, vagy a rendszer rezilienciája miatt elmaradó várt hatások valószínűsége.

6. Következtetések

A molekuláris szabályozás poszt-transzlációs, transzkripcionális és poszt-transzkripcionális módjairól a molekuláris biológiai kísérletes technikák és a bioinformatikai predikciós módszerek fejlődésének köszönhetően, egyre növekvő mennyiségű adat áll rendelkezésre. Ezeket az adatokat számos adatbázis gyűjti, és elérhetővé teszi tudományos közösség számára. Ezen adatbázisok eltérő, részben átfedő tartalma miatt nehézségekbe ütközik az olyan jellegű, kutatásban felmerülő igények kielégítése, ha például az összes rendelkezésre álló transzkripcionális szabályozási kapcsolatot szeretnénk felhasználni egy elemzésben. A különböző kísérletes módszerek és predikciók eltérő megbízhatósága miatt ugyanakkor bizonyos vizsgálatokban szükséges lehet az adatok szűrése. A legmegbízhatóbb adatokhoz irodalmi gyűjtéssel juthatunk, azonban ez igen munkaigényes folyamat, ráadásul az összehasonlíthatóság érdekében jól meghatározott gyűjtési protokoll szerint kell történnie.

Ahogy a Bevezetőben a bioinformatikai adatforrások áttekintése során láthattuk, ezekre a rendszerszintű elemzések és a molekuláris biológiai folyamatok modellezése során felmerülő alapvető problémákra kevés adatbázis kínál akár csak részleges megoldást. Az egyes biológiai folyamatok szabályozásának feltérképezésére az útvonal adatbázisok hivatottak, azonban ezek gyakran nem tartalmazzák információt a transzkripcionális és poszt-transzkripcionális szabályozásról, ezenkívül az útvonalak definíciója önkényes. A Signalink 2 megalkotásával olyan hiánypótló adatforrás jött létre, mely orvosolja a mindezeket a problémákat, egységes kapcsolati hálózatban integrálja a szabályozás különböző szintjeit, miközben lehetőséget biztosít az adatok több szempont alapján történő szűrésére. A Signalink 2-ben a jelátviteli útvonalak biokémiai és evolúciós szempontból különböző, valódi biológiai egységek. A Signalink 2-vel a jelátviteli útvonalak közötti keresztbeszélgetéseket a poszt-transzlációs, transzkripcionális és poszt-transzkripcionális szabályozás szintjén összehasonlító módon vizsgálhatjuk, amire eddig egyetlen más útvonal adatbázis sem biztosított lehetőséget. Nem állt rendelkezésre korábban olyan szabályozási hálózat sem, melyben a poszt-transzkripcionális szabályozás és a miRNS-ek transzkripcionális szabályozása egyaránt szerepelt volna, noha

az adat integráció ezen lépése elengedhetetlen a miRNS-ek által megvalósuló szabályozási mechanizmusok vizsgálatában.

A dolgozatomban bemutatott másik két új adatforrás, az NRF2ome és az ARN két nagy jelentőséggel bíró molekuláris biológiai mechanizmus szabályozását térképezi fel. Az NRF2 az antioxidáns válasz mester transzkripció faktor, több mint 7.500 gén transzkripcióját szabályozza. Fontos szerepet játszik a mérgező anyagok elleni védekezésben, szív- és érrendszeri betegségekben, gyulladásban és daganatos betegségekben. Az autofágia szintén kulcsszereppel bír számos betegség patomechanizmusában, mint az *ischaemia-reperfúzió* okozta károsodás, a neurodegeneratív betegségek és a rák. Az NRF2 és az autofágia „kétélű kard” hasonlattal leírt, ellentmondásos szerepe jelzi a terápiás megközelítések kidolgozásában rejlő kihívásokat. Ezen folyamatok szabályozásáról eddig nem állt rendelkezésre rendszerszintű adatforrás, noha ez elengedhetetlen az elérhető tudás szintetizálásához, új összefüggések felderítése és a szabályozás kontextusfüggő működésének, valamint a gyógyszerhatóanyagok hatásának modellezése érdekében.

A Signalink 2, az NRF2ome és az ARN többféle elemzési módszer alkalmazásához biztosít megfelelő kiindulási alapot. A különböző modellezési eljárásokkal különböző típusú problémák, különböző méretű rendszerek, eltérő részletességgel vizsgálhatók. A differenciálegyenlet rendszerek vagy a *rule-based* modellek kb. 100 molekulából álló interakciós hálózatok viselkedésének igen pontos leírására alkalmasak, míg a modularizálás, a hálózati topológia elemzése vagy a perturbációs szimulációk segítségével nagy hálózatok tulajdonságairól kaphatunk átfogó képet. Egyes elemzések során az általam létrehozott adatbázisokban található általános interakciós hálózatokból szövet- és állapot-specifikus hálózatok nyerhetők, expressziós és mutációs adatok felhasználásával.

A Signalink 2, az NRF2ome és az ARN számottevő előrelépést jelent az adatok integrálása és a molekuláris szabályozási mechanizmusok feltérképezése terén. Az informatív és felhasználóbarát weboldalak, valamint a többféle szabványos formátumban biztosított letöltési lehetőség lehetővé teszi ezen korszerű bioinformatikai adatforrásoknak a széleskörű alkalmazását a kutatásban.

7. Összefoglalás

A molekuláris biológiai és bioinformatikai módszerek egymást erősítő fejlődésével lehetővé vált a szabályozás poszt-transzlációs, transzkripcionális és poszt-transzkripcionális rétegeinek együttes, rendszerszintű elemzése. Dolgozatomban ismertetem az e három réteget tartalmazó szabályozási hálózatok felépítésének általam kidolgozott alapelveit és módszertanát, valamint bemutatom az ezek segítségével általam megalkotott három új bioinformatikai adatforrást. Ezt követően példákkal szemléltetem, hogy az általam létrehozott adatbázisok a szabályozási folyamatok kis léptékű és rendszerszintű vizsgálatára egyaránt alkalmasak.

Doktori munkám során az alábbi új tudományos eredményeket értem el:

1. Kidolgoztam a fehérje-fehérje kölcsönhatásokról, a transzkripcionális és a poszt-transzkripcionális szabályozásról rendelkezésre álló adatok integrálásának bioinformatikai módszertanát.
2. Létrehoztam a *C. elegans* fonálféreg, a *D. melanogaster* ecetmuslinca és az ember hét jelátviteli útvonalának többretegű szabályozását tartalmazó Signalink 2 adatbázist (<http://signalink.org/>), mely lehetővé teszi a jelátviteli hálózat szabályozásának rendszerszintű elemzését.
3. Az antioxidáns válasz fő transzkripciós faktora, az NRF2 szabályozásának feltérképezésére összeállítottam az NRF2ome elnevezésű bioinformatikai adatforrást (<http://nrf2.elte.hu/>), hogy elősegítsem ezen a területen az ismeretek integrálását és a további kutatások megalapozását.
4. A számos betegség patomechanizmusában kulcsszerepet játszó, és a kutatások által ígéretes terápiás célpontnak tekintett sejttani folyamat, az autofágia kutatásának elősegítésére, létrehoztam az Autofágia Szabályozási Hálózatot (<http://arn.elte.hu/>).
5. Interaktív és felhasználóbarát weboldalak, valamint testreszabható letöltő felület kialakításával, az általam létrehozott új bioinformatikai adatforrásokat a tudományos közösség számára elérhetővé tettem.

8. Summary

Recent developments of molecular biology and bioinformatics enhanced each other mutually, making possible an integrated, systems-level analysis of post-translational, transcriptional and post-transcriptional layers of molecular regulation. In my thesis, I present the principles and methodology I elaborated to create regulatory networks consisting all of these three layers, and introduce three new bioinformatic resources I constructed based on this concept. I show examples to demonstrate the capability of the resources created, to be used in small-scale and systems-level analysis of regulatory processes.

In my PhD work I achieved the following new scientific results:

1. I developed a bioinformatic methodology to integrate protein-protein interaction, transcriptional and post-transcriptional regulation data.
2. I created the Signalink 2 database (<http://signalink.org/>), containing the multi-layered regulation of seven signalling pathways of *C. elegans*, *D. melanogaster* and human, providing the opportunity for a systems-level analysis of the signalling network regulation.
3. As an extensive mapping of the regulation of the master antioxidant transcription factor, NRF2, I assembled the NRF2ome bioinformatic resource (<http://nrf2.elte.hu/>) to facilitate the knowledge integration and support novel research works in this field. I developed the Autophagy Regulatory Network (<http://arn.elte.hu/>), a resource to study the regulation of autophagy, a cellular process having key role in pathomechanisms of several diseases and considered as a promising therapeutical target.
4. I constructed interactive and user-friendly webpages, and customizable download interfaces for the new bioinformatic resources to provide a convenient access for the scientific community.

9. Irodalomjegyzék

1. Acuner Ozbabacan S. E., Engin H. B., Gursoy A. és Keskin O. (2011). Transient protein-protein interactions. *Protein Eng Des Sel*, 24(9): 635–648.
2. Alvarez M. A. és Yan C. (2011). A graph-based semantic similarity measure for the gene ontology. *J Bioinform Comput Biol*, 9(6): 681–695.
3. Anders G., Mackowiak S. D., Jens M., Maaskola J., Kuntzagk A., Rajewsky N., Landthaler M. és Dieterich C. (2012). doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res*, 40(Database issue): D180–D186.
4. Ashburner M., Ball C. A., Blake J. A., Botstein D., Butler H., Cherry J. M., Davis A. P., Dolinski K., Dwight S. S., Eppig J. T., Harris M. A., Hill D. P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J. C., Richardson J. E., Ringwald M., Rubin G. M. és Sherlock G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1): 25–29.
5. Bader G. D., Cary M. P. és Sander C. (2006). Pathguide: a pathway resource list. *Nucleic Acids Res*, 34(Database issue): D504–D506.
6. Baitaluk M., Kozhenkov S., Dubinina Y. és Ponomarenko J. (2012). IntegromeDB: an integrated system and biological search engine. *BMC Genomics*, 13(1): 35.
7. Balaji S., McClendon C., Chowdhary R., Liu J. S. és Zhang J. (2012). IMID: integrated molecular interaction database. *Bioinformatics*, 28(5): 747–749.
8. Bandyopadhyay S. és Bhattacharyya M. (2010). PuTmiR: a database for extracting neighboring transcription factors of human microRNAs. *BMC Bioinformatics*, 11: 190.
9. Barbati C., Pierdominici M., Gambardella L., Malchiodi Albedi F., Karas R. H., Rosano G., Malorni W. és Ortona E. (2012). Cell surface estrogen receptor alpha is upregulated during subchronic metabolic stress and inhibits neuronal cell degeneration. *PLoS ONE*, 7(7): e42339.
10. Barrasa M. I., Vaglio P., Cavasino F., Jacotot L. és Walhout A. J. M. (2007). EDGEDb: a transcription factor-DNA interaction database for the analysis of *C. elegans* differential gene expression. *BMC Genomics*, 8: 21.

11. Barrett T., Wilhite S. E., Ledoux P., Evangelista C., Kim I. F., Tomashevsky M., Marshall K. A., Phillippy K. H., Sherman P. M., Holko M., Yefanov A., Lee H., Zhang N., Robertson C. L., Serova N., Davis S. és Soboleva A. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*, 41(Database issue): D991–D995.
12. Bauer-Mehren A., Furlong L. I. és Sanz F. (2009). Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol*, 5: 290.
13. Behrends C., Sowa M. E., Gygi S. P. és Harper J. W. (2010). Network organization of the human autophagy system. *Nature*, 466(7302): 68–76.
14. Bento C. F., Puri C., Moreau K. és Rubinsztein D. C. (2013). The role of membrane-trafficking small GTPases in the regulation of autophagy. *J Cell Sci*, 126(5): 1059–1069.
15. Berger M. F., Philippakis A. A., Qureshi A. M., He F. S., Estep Preston W 3. és Bulyk M. L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol*, 24(11): 1429–1435.
16. 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(8): R90.
17. Betel D., Wilson M., Gabow A., Marks D. S. és Sander C. (2008). The microRNA.org resource: targets and expression. *Nucleic Acids Res*, 36(Database issue): D149–D153.
18. Bhupathy P., Haines C. D. és Leinwand L. A. (2010). Influence of sex hormones and phytoestrogens on heart disease in men and women. *Womens Health*, 6(1): 77–95.
19. Blanco E., Farré D., Albà M. M., Messeguer X. és Guigó R. (2006). ABS: a database of annotated regulatory binding sites from orthologous promoters. *Nucleic Acids Res*, 34(Database issue): D63–D67.
20. Bleris L., Xie Z., Glass D., Adadey A., Sontag E. és Benenson Y. (2011). Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol Syst Biol*, 7: 519.

21. Bouma W., Noma M., Kanemoto S., Matsubara M., Leshnower B. G., Hinmon R., Gorman J. H. és Gorman R. C. (2010). Sex-related resistance to myocardial ischemia-reperfusion injury is associated with high constitutive ARC expression. *Am J Physiol Heart Circ Physiol*, 298(5): H1510–H1517.
22. Bovolenta L. A., Acencio M. L. és Lemke N. (2012). HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics*, 13: 405.
23. Boya P., Reggiori F. és Codogno P. (2013). Emerging regulation and functions of autophagy. *Nat Cell Biol*, 15(7): 713–720.
24. Boyle A. P., Song L., Lee B.-K., London D., Keefe D., Birney E., Iyer V. R., Crawford G. E. és Furey T. S. (2010). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res*, 21(3): 456–464.
25. Breuer K., Foroushani A. K., Laird M. R., Chen C., Sribnaia A., Lo R., Winsor G. L., Hancock R. E. W., Brinkman F. S. L. és Lynn D. J. (2013). InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res*, 41(Database issue): D1228–D1233.
26. Campesi I., Straface E., Occhioni S., Montella A. és Franconi F. (2013). Protein oxidation seems to be linked to constitutive autophagy: a sex study. *Life Sci*, 93(4): 145–152.
27. Cerami E. G., Gross B. E., Demir E., Rodchenkov I., Babur O., Anwar N., Schultz N., Bader G. D. és Sander C. (2011). Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res*, 39(Database issue): D685–D690.
28. Chang H. W., Lee Y. S., Nam H. Y., Han M. W., Kim H. J., Moon S. Y., Jeon H., Park J. J., Carey T. E., Chang S. E., Kim S. W. és Kim S. Y. (2013). Knockdown of beta-catenin controls both apoptotic and autophagic cell death through LKB1/AMPK signaling in head and neck squamous cell carcinoma cell lines. *Cell Signal*, 25(4): 839–847.

29. Chatr-aryamontri A., Breitkreutz B.-J., Heinicke S., Boucher L., Winter A., Stark C., Nixon J., Ramage L., Kolas N., O'Donnell L., Reguly T., Breitkreutz A., Sellam A., Chen D., Chang C., Rust J., Livstone M., Oughtred R., Dolinski K. és Tyers M. (2013). The BioGRID interaction database: 2013 update. *Nucleic Acids Res*, 41(Database issue): D816–D823.
30. Chelliah V., Laibe C. és Le Novère N. (2013). BioModels database: a repository of mathematical models of biological processes. *Methods Mol Biol*, 1021: 189–199.
31. Chen C., Hu L.-X., Dong T., Wang G.-Q., Wang L.-H., Zhou X.-P., Jiang Y., Murao K., Lu S.-Q., Chen J.-W. és Zhang G.-X. (2013). Apoptosis and autophagy contribute to gender difference in cardiac ischemia–reperfusion induced injury in rats. *Life Sci*, 93(7): 265–270.
32. Chen N. és Karantza V. (2011). Autophagy as a therapeutic target in cancer. *Cancer Biol Ther*, 11(2): 157–168.
33. Chen X.-W. és Liu M. (2005). Prediction of protein-protein interactions using random decision forest framework. *Bioinformatics*, 21(24): 4394–4400.
34. Cho H.-Y., Gladwell W., Wang X., Chorley B., Bell D., Reddy S. P. és Kleeberger S. R. (2010). Nrf2-regulated PPAR-gamma expression is critical to protection against acute lung injury in mice. *Am J Respir Crit Care Med*, 182(2): 170–182.
35. Cho H.-Y., Imani F., Miller-DeGraff L., Walters D., Melendi G. A., Yamamoto M., Polack F. P. és Kleeberger S. R. (2009). Antiviral activity of Nrf2 in a murine model of respiratory syncytial virus disease. *Am J Respir Crit Care Med*, 179(2): 138–150.
36. Cournac A. és Sepulchre J.-A. (2009). Simple molecular networks that respond optimally to time-periodic stimulation. *BMC Syst Biol*, 3: 29.
37. Croft D., Mundo A. F., Haw R., Milacic M., Weiser J., Wu G., Caudy M., Garapati P., Gillespie M., Kamdar M. R., 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(Database issue): D472–D477.

38. Csermely P., Agoston V. és Pongor S. (2005). The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol Sci*, 26(4): 178–182.
39. Davis S. és Meltzer P. (2007). Geoquery: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics*, 23(14): 1846–1847.
40. Dhakshinamoorthy S. és Jaiswal A. K. (2000). Small maf (MafG and MafK) proteins negatively regulate antioxidant response element-mediated expression and antioxidant induction of the NAD(P)H:Quinone oxidoreductase1 gene. *J Biol Chem*, 275(51): 40134–40141.
41. Dinkel H., Michael S., Weatheritt R. J., Davey N. E., Van Roey K., Altenberg B., Toedt G., Uyar B., Seiler M., Budd A., Jödicke L., Dammert M. A., Schroeter C., Hammer M., Schmidt T., Jehl P., McGuigan C., Dymecka M., Chica C., Luck K., Via A., Chatr-Aryamontri A., Haslam N., Grebnev G., Edwards R. J., Steinmetz M. O., Meiselbach H., Diella F. és Gibson T. J. (2013). ELM—the database of eukaryotic linear motifs. *Nucleic Acids Res*, 40(Database issue): D242–D251.
42. Dirnagl U. és Lauritzen M. (2010). Fighting publication bias: introducing the negative results section. *J Cerebr Blood F Met*, 30(7): 1263–1264.
43. Doshi B. M., Perdrizet G. A. és Hightower L. E. (2008). Wound healing from a cellular stress response perspective. *Cell Stress Chaperones*, 13(4): 393–399.
44. Drysdale R és a FlyBase Consortium (2008). FlyBase: a database for the *Drosophila* research community. *Methods Mol Biol*, 420: 45–59.
45. Enright A. J., John B., Gaul U., Tuschl T., Sander C. és Marks D. S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol*, 5(1): R1.
46. Ewing R. M., Chu P., Elisma F., Li H., Taylor P., Climie S., McBroom-Cerajewski L., Robinson M. D., O'Connor L., Li M., Taylor R., Dharsee M., Ho Y., Heilbut A., Moore L., Zhang S., Ornatsky O., Bukhman Y. V., Ethier M., Sheng Y., Vasilescu J., Abu-Farha M., Lambert J.-P., Duewel H. S., Stewart I. I., Kuehl B., Hogue K., Colwill K., Gladwish K., Muskat B., Kinach R., Adams S.-L., Moran M. F., Morin G. B., Topaloglou T. és Figeys D. (2007). Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol Syst Biol*, 3: 89.

47. Flicek P., Ahmed I., Amode M. R., Barrell D., Beal K., Brent S., Carvalho-Silva D., Clapham P., Coates G., Fairley S., Fitzgerald S., Gil L., Garcia-Giron C., Gordon L., Hourlier T., Hunt S., Juettemann T., Kahari A. K., Keenan S., Komorowska M., Kulesha E., Longden I., Maurel T., McLaren W. M., Muffato M., Nag R., Overduin B., Pignatelli M., Pritchard B., Pritchard E., Riat H. S., Ritchie G. R. S., Ruffier M., Schuster M., Sheppard D., Sobral D., Taylor K., Thormann A., Trevanion S., White S., Wilder S. P., Aken B. L., Birney E., Cunningham F., Dunham I., Harrow J., Herrero J., Hubbard T. J. P., Johnson N., Kinsella R., Parker A., Spudich G., Yates A., Zadissa A. és Searle S. M. J. (2013). Ensembl 2013. *Nucleic Acids Res*, 41(Database issue): D48–D55.
48. Forbes S. A., Bindal N., Bamford S., Cole C., Kok C. Y., Beare D., Jia M., Shepherd R., Leung K., Menzies A., Teague J. W., Campbell P. J., Stratton M. R. és Futreal P. A. (2010). COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res*, 39(Database issue): D945–D950.
49. Franceschini A., Szklarczyk D., Frankild S., Kuhn M., Simonovic M., Roth A., Lin J., Minguez P., Bork P., Mering C. és Jensen L. J. (2014). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*, 41(Database issue): D808–D815.
50. Frankel L. B. és Lund A. H. (2012). MicroRNA regulation of autophagy. *Carcinogenesis*, 33(11): 2018–2025.
51. Friedman A. és Perrimon N. (2007). Genetic screening for signal transduction in the era of network biology. *Cell*, 128(2): 225–231.
52. Friedman R. C., Farh K. K.-H., Burge C. B. és Bartel D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*, 19(1): 92–105.
53. Gallo S. M., Gerrard D. T., Miner D., Simich M., Des Soye B., Bergman C. M. és Halfon M. S. (2011). REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in *Drosophila*. *Nucleic Acids Res*, 39(Database issue): D118–D123.

54. Gavin A.-C., Aloy P., Grandi P., Krause R., Boesche M., Marzioch M., Rau C., Jensen L. J., Bastuck S., Dümpelfeld B., Edelmann A., Heurtier M.-A., Hoffman V., Hoefert C., Klein K., Hudak M., Michon A.-M., Schelder M., Schirle M., Remor M., Rudi T., Hooper S., Bauer A., Bouwmeester T., Casari G., Drewes G., Neubauer G., Rick J. M., Kuster B., Bork P., Russell R. B. és Superti-Furga G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440(7084): 631–636.
55. Gerstein M. B., Kundaje A., Hariharan M., Landt S. G., Yan K.-K., Cheng C., Mu X. J., Khurana E., Rozowsky J., Alexander R., Min R., Alves P., Abyzov A., Addleman N., Bhardwaj N., Boyle A. P., Cayting P., Charos A., Chen D. Z., Cheng Y., Clarke D., Eastman C., Euskirchen G., Fietze S., Fu Y., Gertz J., Grubert F., Harmanci A., Jain P., Kasowski M., Lacroute P., Leng J., Lian J., Monahan H., O’Geen H., Ouyang Z., Partridge E. C., Patacsil D., Pauli F., Raha D., Ramirez L., Reddy T. E., Reed B., Shi M., Slifer T., Wang J., Wu L., Yang X., Yip K. Y., Zilberman-Schapira G., Batzoglou S., Sidow A., Farnham P. J., Myers R. M., Weissman S. M. és Snyder M. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature*, 489(7414): 91–100.
56. Giot L., Bader J. S., Brouwer C., Chaudhuri A., Kuang B., Li Y., Hao Y. L., Ooi C. E., Godwin D., Vitols E., Vijayadamodar G., Pochart P., Machineni H., Welsh M., Kong Y., Zerhusen B., Malcolm R., Varrone Z., Collis A., Minto M., Burgess S., McDaniel L., Stimpson E., Spriggs F., Williams J., Neurath K., Ioime N., Agee M., Voss E., Furtak K., Renzulli R., Aanensen N., Carrola S., Bickelhaupt E., Lazovatsky Y., DaSilva A., Zhong J., Stanyon C. A., Finley R. L. J., White K. P., Braverman M., Jarvie T., Gold S., Leach M., Knight J., Shimkets R., McKenna M., Chant J. és Rothberg J. M. (2003). A protein interaction map of *Drosophila melanogaster*. *Science*, 302(5651): 1727–1736.
57. Griffith O. L., Montgomery S. B., Bernier B., Chu B., Kasaian K., Aerts S., Mahony S., Sleumer M. C., Bilenky M., Haeussler M., Griffith M., Gallo S. M., Giardine B., Hooghe B., Van Loo P., Blanco E., Ticoll A., Lithwick S., Portales-Casamar E., Donaldson I. J., Robertson G., Wadelius C., De Bleser P., Vlieghe D., Halfon M. S., Wasserman W., Hardison R., Bergman C. M., Jones S. J. M. és Open Regulatory Annotation Consortium (2008). ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res*, 36(Database issue): D107–D113.

58. Grimson A., Farh K. K.-H., Johnston W. K., Garrett-Engle P., Lim L. P. és Bartel D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*, 27(1): 91–105.
59. Gu Y., Zhao W., Xia J., Zhang Y., Wu R., Wang C. és Guo Z. (2011). Analysis of pathway mutation profiles highlights collaboration between cancer-associated superpathways. *Hum Mutat*, 32(9): 1028–1035.
60. Guo X. és Wang X.-F. (2009). Signaling cross-talk between TGF-beta/BMP and other pathways. *Cell Res*, 19(1): 71–88.
61. Gurusamy N. és Das D. K. (2009). Is autophagy a double-edged sword for the heart? *Acta Physiol Hung*, 96(3): 267–276.
62. Hamacher-Brady A. (2012). Autophagy regulation and integration with cell signaling. *Antioxid Redox Signal*, 17(5): 756–765.
63. Harrington H. A., Ho K. L., Ghosh S. és Tung K. (2008). Construction and analysis of a modular model of caspase activation in apoptosis. *Theor Biol Med Model*, 5(1): 26.
64. Harris T. W., Antoshechkin I., Bieri T., Blasiar D., Chan J., Chen W. J., De La Cruz N., Davis P., Duesbury M., Fang R., Fernandes J., Han M., Kishore R., Lee R., Müller H.-M., Nakamura C., Ozersky P., Petcherski A., Rangarajan A., Rogers A., Schindelman G., Schwarz E. M., Tuli M. A., Van Auken K., Wang D., Wang X., Williams G., Yook K., Durbin R., Stein L. D., Spieth J. és Sternberg P. W. (2010). WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res*, 38(Database issue): D463–D467.
65. 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(Database issue): D986–D990.
66. Hopkins A. L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*, 4(11): 682–690.
67. Huntzinger E. és Izaurralde E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet*, 12(2): 99–110.

68. Ikeda Y., Sugawara A., Taniyama Y., Uruno A., Igarashi K., Arima S., Ito S. és Takeuchi K. (2000). Suppression of rat thromboxane synthase gene transcription by peroxisome proliferator-activated receptor gamma in macrophages via an interaction with Nrf2. *J Biol Chem*, 275(42): 33142–33150.
69. Jaiswal A. K. (2004). Nrf2 signaling in coordinated activation of antioxidant gene expression. *Free Radic Biol Med*, 36(10): 1199–1207.
70. Jaramillo M. C. és Zhang D. D. (2013). The emerging role of the Nrf2-Keap1 signaling pathway in cancer. *Genes Dev*, 27(20): 2179–2191.
71. Jetten (2009). Retinoid-related orphan receptors (RORs): critical roles in development, immunity, circadian rhythm, and cellular metabolism. *Nucl Recept Signal*, 7: e003.
72. 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(Database issue): D98–D104.
73. Jin S. M. és Youle R. J. (2012). PINK1- and parkin-mediated mitophagy at a glance. *J Cell Sci*, 125(4): 795–799.
74. Jolma A., Yan J., Whittington T., Toivonen J., Nitta K. R., Rastas P., Morgunova E., Enge M., Taipale M., Wei G., Palin K., Vaquerizas J. M., Vincentelli R., Luscombe N. M., Hughes T. R., Lemaire P., Ukkonen E., Kivioja T. és Taipale J. (2013). DNA-Binding specificities of human transcription factors. *Cell*, 152(1-2): 327–339.
75. Joshi-Tope G. (2005). Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res*, 33(Database issue): D428–D432.
76. Kamburov A., Wierling C., Lehrach H. és Herwig R. (2010). ConsensusPathDB—a database for integrating human functional interaction networks. *Nucleic Acids Res*, 37(Database issue): D623–D628.

77. Kandasamy K., Mohan S., Raju R., Keerthikumar S., Kumar G. S. S., Venugopal A. K., Telikicherla D., Navarro D. J., Mathivanan S., Pecquet C., Gollapudi S. K., Tattikota S. G., Mohan S., Padhukasahasram H., Subbannayya Y., Goel R., Jacob H. K. C., Zhong J., Sekhar R., Nanjappa V., Balakrishnan L., Subbaiah R., Ramachandra Y. I., Rahiman A., Keshava Prasad T. S., Lin J.-X., Houtman J. C. D., Desiderio S., Renauld J.-C., Constantinescu S., Ohara O., Hirano T., Kubo M., Singh S., Khatri P., Draghici S., Bader G. D., Sander C., Leonard W. J. és Pandey A. (2010). NetPath: a public resource of curated signal transduction pathways. *Genome Biol*, 11(1): R3.
78. Kanehisa M., Goto S., Kawashima S., Okuno Y. és Hattori M. (2005). The KEGG resource for deciphering the genome. *Nucleic Acids Res*, 32(Database issue): D277–D280.
79. Kauffmann A., Rayner T. F., Parkinson H., Kapushesky M., Lukk M., Brazma A. és Huber W. (2009). Importing arrayexpress datasets into r/bioconductor. *Bioinformatics*, 25(16): 2092–2094.
80. Kelder T., Iersel M. P., Hanspers K., Kutmon M., Conklin B. R., Evelo C. T. és Pico A. R. (2012). WikiPathways: building research communities on biological pathways. *Nucleic Acids Res*, 40(Database issue): D1301–D1307.
81. Kenific C. M., Thorburn A. és Debnath J. (2010). Autophagy and metastasis: another double-edged sword. *Curr Opin Cell Biol*, 22(2): 241–245.
82. Kensler T. W., Wakabayashi N. és Biswal S. (2007). Cell survival responses to environmental stresses via the keap1-nrf2-ARE pathway. *Annu Rev Pharmacol Toxicol*, 47: 89–116.
83. Keshava Prasad T. S., 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 D. S., Sebastian A., Rani S., Ray S., Harrys Kishore C. J., Kanth S., Ahmed M., Kashyap M. K., Mohmood R., Ramachandra Y. L., Krishna V., Rahiman B. A., Mohan S., Ranganathan P., Ramabadran S., Chaerkady R. és Pandey A. (2009). Human protein reference database–2009 update. *Nucleic Acids Res*, 37(Database issue): D767–D772.

84. Kholodenko B., Yaffe M. B. és Kolch W. (2012). Computational approaches for analyzing information flow in biological networks. *Sci Signal*, 5(220): re1.
85. Khurana E., Fu Y., Chen J. és Gerstein M. (2013). Interpretation of genomic variants using a unified biological network approach. *PLoS Comput Biol*, 9(3): e1002886.
86. Kimura A., Ishida Y., Hayashi T., Wada T., Yokoyama H., Sugaya T., Mukaida N. és Kondo T. (2006). Interferon-gamma plays protective roles in sodium arsenite-induced renal injury by up-regulating intrarenal multidrug resistance-associated protein 1 expression. *Am J Pathol*, 169(4): 1118–1128.
87. Kimura T., Takabatake Y., Takahashi A. és Isaka Y. (2013). Chloroquine in cancer therapy: a double-edged sword of autophagy. *Cancer Res*, 73(1): 3–7.
88. Klüppel M. és Wrana J. L. (2005). Turning it up a notch: cross-talk between TGF beta and notch signaling. *Bioessays*, 27(2): 115–118.
89. Knox C., Law V., Jewison T., Liu P., Ly S., Frolkis A., Pon A., Banco K., Mak C., Neveu V., Djoumbou Y., Eisner R., Guo A. C. és Wishart D. S. (2012). DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*, 39(Database issue): D1035–D1041.
90. Koh G. C. K. W., Porrás P., Aranda B., Hermjakob H. és Orchard S. E. (2012). Analyzing protein-protein interaction networks. *J Proteome Res*, 11(4): 2014–2031.
91. Kondor D. és Vattay G. (2013). Dynamics and structure in cell signaling networks: off-state stability and dynamically positive cycles. *PLoS ONE*, 8(3): e57653.
92. Korcsmaros T., Szalay M., Böde C., Kovács I. és Csermely P. (2007). How to design multi-target drugs: target-search options in cellular networks. *Exp Op Drug Discovery*, 6: 799–808.
93. Korcsmáros T., Farkas I. J., Szalay M. S., 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(16): 2042–2050.

94. Kosmider B., Messier E. M., Janssen W. J., Nahreini P., Wang J., Hartshorn K. L. és Mason R. J. (2012). Nrf2 protects human alveolar epithelial cells against injury induced by influenza A virus. *Respir Res*, 13(1): 43.
95. Kovács I. A., Palotai R., Szalay M. S. 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(9): e12528.
96. Kozomara A. és Griffiths-Jones S. (2010). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*, 39(Database issue): D152–D157.
97. Kraft C., Kijanska M., Kalie E., Siergiejuk E., Lee S. S., Semplicio G., Stoffel I., Brezovich A., Verma M., Hansmann I., Ammerer G., Hofmann K., Tooze S. és Peter M. (2012). Binding of the Atg1/ULK1 kinase to the ubiquitin-like protein atg8 regulates autophagy: regulation of atg1 kinase function. *EMBO J*, 31(18): 3691–3703.
98. Krek A., Grün D., Poy M. N., Wolf R., Rosenberg L., Epstein E. J., MacMenamin P., Piedade I., Gunsalus K. C., Stoffel M. és Rajewsky N. (2005). Combinatorial microRNA target predictions. *Nat Genet*, 37(5): 495–500.
99. Lavallard V. J., Meijer A. J., Codogno P. és Gual P. (2012). Autophagy, signaling and obesity. *Pharmacol Res*, 66(6): 513–525.
100. Lázár V., Ecsedi S., Vízkeleti L., Rákosy Z., Boross G., Szappanos B., Bégány A., Emri G., Adány R. és Balázs M. (2012). Marked genetic differences between BRAF and NRAS mutated primary melanomas as revealed by array comparative genomic hybridization. *Melanoma Res*, 22(3): 202–214.
101. Levine J. H., Lin Y. és Elowitz M. B. (2013). Functional roles of pulsing in genetic circuits. *Science*, 342(6163): 1193–1200.
102. Lewis B. P., Burge C. B. és Bartel D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1): 15–20.

103. Li D., Liu W., Liu Z., Wang J., Liu Q., Zhu Y. és He F. (2008). PRINCESS, a protein interaction confidence evaluation system with multiple data sources. *Mol Cell Proteomics*, 7(6): 1043–1052.
104. Li Y., Wang M., Carra C. és Cucinotta F. A. (2012). Modularized smad-regulated TGF-beta signaling pathway. *Math Biosci*, 240(2): 187–200.
105. Libina N., Berman J. R. és Kenyon C. (2003). Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. *Cell*, 115(4): 489–502.
106. Lichtenberg U., Jensen L. J., Brunak S. és Bork P. (2005). Dynamic complex formation during the yeast cell cycle. *Science*, 307(5710): 724–727.
107. Lin L.-T., Dawson P. W. H. és Richardson C. D. (2010). Viral interactions with macroautophagy: a double-edged sword. *Virology*, 402(1): 1–10.
108. Lista P., Straface E., Brunelleschi S., Franconi F. és Malorni W. (2011). On the role of autophagy in human diseases: a gender perspective. *J Cell Mol Med*, 15(7): 1443–1457.
109. Liu W., Li D., Wang J., Xie H., Zhu Y. és He F. (2009). Proteome-wide prediction of signal flow direction in protein interaction networks based on interacting domains. *Mol Cell Proteomics*, 8(9): 2063–2070.
110. Lloyd C. M., Lawson J. R., Hunter P. J. és Nielsen P. F. (2008). The CellML model repository. *Bioinformatics*, 24(18): 2122–2123.
111. Lopez F., Textoris J., Bergon A., Didier G., Remy E., Granjeaud S., Imbert J., Nguyen C. és Puthier D. (2008). TranscriptomeBrowser: a powerful and flexible toolbox to explore productively the transcriptional landscape of the gene expression omnibus database. *PLoS ONE*, 3(12): e4001.
112. Lyne R., Smith R., Rutherford K., Wakeling M., Varley A., Guillier F., Janssens H., Ji W., McLaren P., North P., Rana D., Riley T., Sullivan J., Watkins X., Woodbridge M., Lilley K., Russell S., Ashburner M., Mizuguchi K. és Micklem G. (2007). FlyMine: an integrated database for *Drosophila* and anopheles genomics. *Genome Biol*, 8(7): R129.
113. Ma Y., Galluzzi L., Zitvogel L. és Kroemer G. (2013). Autophagy and cellular immune responses. *Immunity*, 39(2): 211–227.

114. Malhotra D., Portales-Casamar E., Singh A., Srivastava S., Arenillas D., Happel C., Shyr C., Wakabayashi N., Kensler T. W., Wasserman W. W. és Biswal S. (2010). Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through ChIP-Seq profiling and network analysis. *Nucleic Acids Res*, 38(17): 5718–5734.
115. Mammucari C., Milan G., Romanello V., Masiero E., Rudolf R., Del Piccolo P., Burden S. J., Di Lisi R., Sandri C., Zhao J., Goldberg A. L., Schiaffino S. és Sandri M. (2007). FoxO3 controls autophagy in skeletal muscle in vivo. *Cell Metab*, 6(6): 458–471.
116. Mandal P., Park P.-H., McMullen M. R., Pratt B. T. és Nagy L. E. (2010). The anti-inflammatory effects of adiponectin are mediated via a heme oxygenase-1-dependent pathway in rat kupffer cells. *Hepatology*, 51(4): 1420–1429.
117. Maragkakis M., Alexiou P., Papadopoulos G. L., Reczko M., Dalamagas T., Giannopoulos G., Goumas G., Koukis E., Kourtis K., Simossis V. A., Sethupathy P., Vergoulis T., Koziris N., Sellis T., Tsanakas P. és Hatzigeorgiou A. G. (2009). Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, 10: 295.
118. Maragkakis M., Vergoulis T., Alexiou P., Reczko M., Plomaritou K., Gousis M., Kourtis K., Koziris N., Dalamagas T. és Hatzigeorgiou A. G. (2011). DIANA-microT web server upgrade supports fly and worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic Acids Res*, 39(Web Server issue): W145–W148.
119. Martinet W., Agostinis P., Vanhooecke B., Dewaele M. és De Meyer G. R. Y. (2009). Autophagy in disease: a double-edged sword with therapeutic potential. *Clin Sci*, 116(9): 697–712.
120. Mathew R., Karp C. M., Beaudoin B., Vuong N., Chen G., Chen H.-Y., Bray K., Reddy A., Bhanot G., Gelinas C., Dipaola R. S., Karantza-Wadsworth V. és White E. (2009). Autophagy suppresses tumorigenesis through elimination of p62. *Cell*, 137(6): 1062–1075.

121. Miller A. M., Wang H., Park O., Horiguchi N., Lafdil F., Mukhopadhyay P., Moh A., Fu X. Y., Kunos G., Pacher P. és Gao B. (2010). Anti-inflammatory and anti-apoptotic roles of endothelial cell STAT3 in alcoholic liver injury. *Alcohol Clin Exp Res*, 34(4): 719–725.
122. Motomura Y., Kitamura H., Hijikata A., Matsunaga Y., Matsumoto K., Inoue H., Atarashi K., Hori S., Watarai H., Zhu J., Taniguchi M. és Kubo M. (2011). The transcription factor E4BP4 regulates the production of IL-10 and IL-13 in CD4+ t cells. *Nat Immunol*, 12(5): 450–459.
123. 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-alpha in breast cancer cells is associated with constitutive activation of autophagy as revealed by a transcriptome analysis using a custom microarray. *Autophagy*, 7(7): 760–770.
124. Murali T., Pacifico S., Yu J., Guest S., Roberts George G 3. és Finley Russell L J. (2011). DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic Acids Res*, 39(Database issue): D736–D743.
125. Murray P. J. (2007). The JAK-STAT signaling pathway: input and output integration. *J Immunol*, 178(5): 2623–2629.
126. Nepal S., Kim M. J., Subedi A., Lee E.-S., Yong C. S., Kim J.-A., Kang W., Kwak M.-K., Arya D. S. és Park P.-H. (2012). Globular adiponectin inhibits ethanol-induced apoptosis in HepG2 cells through heme oxygenase-1 induction. *Biochem Pharmacol*, 84(7): 974–983.
127. Nixon R. A. (2013). The role of autophagy in neurodegenerative disease. *Nat Med*, 19(8): 983–997.
128. Nussinov R., Tsai C.-J. és Csermely P. (2011). Allo-network drugs: harnessing allostery in cellular networks. *Trends Pharmacol Sci*, 32(12): 686–693.

129. Orchard S., Ammari M., Aranda B., Breuza L., Briganti L., Broackes-Carter F., Campbell N. H., 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 R. C., Meldal B., Melidoni A. N., Milagros M., Peluso D., Perfetto L., Porras P., Raghunath A., Ricard-Blum S., Roechert B., Stutz A., Tognolli M., 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(Database issue): D358–D363.
130. Ouwerkerk P. B. és Meijer A. H. (2001). *Current protocols in molecular biology*. John Wiley & Sons, Inc., Hoboken, NJ, USA. Fejezet: 12.12. Yeast One-Hybrid Screening for DNA-Protein Interactions.
131. Owuor E. D. és Kong A.-N. T. (2002). Antioxidants and oxidants regulated signal transduction pathways. *Biochem Pharmacol*, 64(5-6): 765–770.
132. Pang R. W. C. és Poon R. T. P. (2012). Cancer stem cell as a potential therapeutic target in hepatocellular carcinoma. *Curr Cancer Drug Targets*, 12(9): 1081–1094.
133. Papin J. A. és Palsson B. O. (2004). The JAK-STAT signaling network in the human b-cell: an extreme signaling pathway analysis. *Biophys J*, 87(1): 37–46.
134. Parrish J. R., Gulyas K. D. és Finley Russell L. J. (2006). Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol*, 17(4): 387–393.
135. Patil A., Nakai K. és Nakamura H. (2011). HitPredict: a database of quality assessed protein-protein interactions in nine species. *Nucleic Acids Res*, 39(Database issue): D744–D749.
136. Paz A., Brownstein Z., Ber Y., Bialik S., David E., Sagir D., Ulitsky I., Elkon R., Kimchi A., Avraham K. B., Shiloh Y. és Shamir R. (2012). SPIKE: a database of highly curated human signaling pathways. *Nucleic Acids Res*, 39(Database issue): D793–D799.
137. Peña-Llopis S., Vega-Rubin-de-Celis S., Schwartz J. C., Wolff N. C., Tran T. A. T., Zou L., Xie X.-J., Corey D. R. és Brugarolas J. (2011). Regulation of TFEB and v-ATPases by mTORC1. *EMBO J*, 30(16): 3242–3258.

138. Petherick K. J., Williams A. C., Lane J. D., Ordóñez-Morán P., Huelsken J., Collard T. J., Smartt H. J. M., Batson J., Malik K., Paraskeva C. és Greenhough A. (2013). Autolysosomal beta-catenin degradation regulates wnt-autophagy-p62 crosstalk. *EMBO J*, 32(13): 1903–1916.
139. Pi J., Leung L., Xue P., Wang W., Hou Y., Liu D., Yehuda-Shnaidman E., Lee C., Lau J., Kurtz T. W. és Chan J. Y. (2010). Deficiency in the nuclear factor e2-related factor-2 transcription factor results in impaired adipogenesis and protects against diet-induced obesity. *J Biol Chem*, 285(12): 9292–9300.
140. Pi J., Bai Y., Reece J. M., Williams J., Liu D., Freeman M. L., Fahl W. E., Shugar D., Liu J., Qu W., Collins S. és Waalkes M. P. (2007). Molecular mechanism of human Nrf2 activation and degradation: role of sequential phosphorylation by protein kinase CK2. *Free Radic Biol Med*, 42(12): 1797–1806.
141. Pires-daSilva A. és Sommer R. J. (2003). The evolution of signalling pathways in animal development. *Nat Rev Genet*, 4(1): 39–49.
142. Pirkkala L., Nykänen P. és Sistonen L. (2001). Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *FASEB J*, 15(7): 1118–1131.
143. Portales-Casamar E., Arenillas D., Lim J., Swanson M. I., Jiang S., McCallum A., Kirov S. és Wasserman W. W. (2009). The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Res*, 37(Database issue): D54–D60.
144. Portales-Casamar E., Thongjuea S., Kwon A. T., Arenillas D., Zhao X., Valen E., Yusuf D., Lenhard B., Wasserman W. W. és Sandelin A. (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res*, 38(Database issue): D105–D110.
145. Proikas-Cezanne T., Waddell S., Gaugel A., Frickey T., Lupas A. és Nordheim A. (2004). WIPI-1alpha (WIPI49), a member of the novel 7-bladed WIPI protein family, is aberrantly expressed in human cancer and is linked to starvation-induced autophagy. *Oncogene*, 23(58): 9314–9325.

146. Punta M., Coghill P. C., Eberhardt R. Y., Mistry J., Tate J., Boursnell C., Pang N., Forslund K., Ceric G., Clements J., Heger A., Holm L., Sonnhammer E. L. L., Eddy S. R., Bateman A. és Finn R. D. (2011). The pfam protein families database. *Nucleic Acids Res*, 40(Database issue): D290–D301.
147. Pyo J.-O., Yoo S.-M. és Jung Y.-K. (2013). The interplay between autophagy and aging. *Diabetes Metab Res*, 37(5): 333.
148. Qu H. és Fang X. (2013). A brief review on the human encyclopedia of DNA elements (ENCODE) project. *Genomics Proteomics Bioinformatics*, 11(3): 135–141.
149. Raghavachari B., Tasneem A., Przytycka T. M. és Jothi R. (2007). DOMINE: a database of protein domain interactions. *Nucleic Acids Res*, 36(Database issue): D656–D661.
150. Reece-Hoyes J. S., Deplancke B., Shingles J., Grove C. A., Hope I. A. és Walhout A. J. M. (2005). A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol*, 6(13): R110.
151. Reggiori F. (2012). Autophagy: new questions from recent answers. *ISRN Mol Biol*, 2012: 738718.
152. Reichelt M., Mellor K., Curl C., Stapleton D. és Delbridge L. (2013). Myocardial glycophagy — a specific glycogen handling response to metabolic stress is accentuated in the female heart. *J Mol Cell Cardiol*, 65: 67–75.
153. Ren S., Uversky V. N., Chen Z., Dunker A. K. és Obradovic Z. (2008). Short linear motifs recognized by SH2, SH3 and Ser/Thr kinase domains are conserved in disordered protein regions. *BMC Genomics*, 9(Suppl 2): S26.
154. Rhodes D. R., Tomlins S. A., Varambally S., Mahavisno V., Barrette T., Kalyana-Sundaram S., Ghosh D., Pandey A. és Chinnaiyan A. M. (2005). Probabilistic model of the human protein-protein interaction network. *Nat Biotechnol*, 23(8): 951–959.

155. Rual J.-F., Venkatesan K., Hao T., Hirozane-Kishikawa T., Dricot A., Li N., Berriz G. F., Gibbons F. D., Dreze M., Ayivi-Guedehoussou N., Klitgord N., Simon C., Boxem M., Milstein S., Rosenberg J., Goldberg D. S., Zhang L. V., Wong S. L., Franklin G., Li S., Albala J. S., Lim J., Fraughton C., Llamosas E., Cevik S., Bex C., Lamesch P., Sikorski R. S., Vandenhaute J., Zoghbi H. Y., Smolyar A., Bosak S., Sequerra R., Doucette-Stamm L., Cusick M. E., Hill D. E., Roth F. P. és Vidal M. (2005). Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, 437(7062): 1173–1178.
156. Russell R. C., Tian Y., Yuan H., Park H. W., Chang Y.-Y., Kim J., Kim H., Neufeld T. P., Dillin A. és Guan K.-L. (2013). ULK1 induces autophagy by phosphorylating beclin-1 and activating VPS34 lipid kinase. *Nat Cell Biol*, 15(7): 741–750.
157. Rustici G., Kolesnikov N., Brandizi M., Burdett T., Dylag M., Emam I., Farne A., Hastings E., Ison J., Keays M., Kurbatova N., Malone J., Mani R., Mupo A., Pedro Pereira R., Pilicheva E., Rung J., Sharma A., Tang Y. A., Ternent T., Tikhonov A., Welter D., Williams E., Brazma A., Parkinson H. és Sarkans U. (2012). ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res*, 41(Database issue): D987–D990.
158. Salih D. A. M. és Brunet A. (2008). FoxO transcription factors in the maintenance of cellular homeostasis during aging. *Curr Opin Cell Biol*, 20(2): 126–136.
159. Sangokoya C., Telen M. J. és Chi J.-T. (2010). microRNA miR-144 modulates oxidative stress tolerance and associates with anemia severity in sickle cell disease. *Blood*, 116(20): 4338–4348.
160. Schaefer C. F., Anthony K., Krupa S., Buchoff J., Day M., Hannay T. és Buetow K. H. (2010). PID: the pathway interaction database. *Nucleic Acids Res*, 37(Database issue): D674–D679.
161. Sengupta A., Molkenin J. D. és Yutzey K. E. (2009). FoxO transcription factors promote autophagy in cardiomyocytes. *J Biol Chem*, 284(41): 28319–28331.
162. Seo Y.-K., Jeon T.-I., Chong H. K., Biesinger J., Xie X. és Osborne T. F. (2011). Genome-wide localization of SREBP-2 in hepatic chromatin predicts a role in autophagy. *Cell Metab*, 13(4): 367–375.

163. Settembre C., Di Malta C., Polito V. A., Garcia Arencibia M., Vetrini F., Erdin S., Erdin S. U., Huynh T., Medina D., Colella P., Sardiello M., Rubinsztein D. C. és Ballabio A. (2011). TFEB links autophagy to lysosomal biogenesis. *Science*, 332(6036): 1429–1433.
164. Settembre C., Zoncu R., Medina D. L., Vetrini F., Erdin S., Erdin S., Huynh T., Ferron M., Karsenty G., Vellard M. C., Facchinetti V., Sabatini D. M. és Ballabio A. (2012). A lysosome-to-nucleus signalling mechanism senses and regulates the lysosome via mTOR and TFEB. *EMBO J*, 31(5): 1095–1108.
165. Smialowski P., Pagel P., Wong P., Brauner B., Dunger I., Fobo G., Frishman G., Montro-ne C., Rattei T., Frishman D. és Ruepp A. (2009). The negatome database: a reference set of non-interacting protein pairs. *Nucleic Acids Res*, 38(Database issue): D540–D544.
166. Smith B., Ceusters W., Klagges B., Köhler J., Kumar A., Lomax J., Mungall C., Neuhaus F., Rector A. L. és Rosse C. (2005). Relations in biomedical ontologies. *Genome Biol*, 6(5): R46.
167. Smyth G. K. (2005). Limma: linear models for microarray data. *Bioinformatics and computational biology solutions using R and bioconductor*. Szerk. Gentleman R., Carey V., Dudoit S., Irizarry R. és Huber W. Springer, New York: 397–420.
168. Snider J., Kittanakom S., Curak J. és Stagljar I. (2010). Split-ubiquitin based membra-ne yeast two-hybrid (MYTH) system: a powerful tool for identifying protein-protein interactions. *J Vis Exp*, (36): e1698.
169. Sreenivasaiah P. K., Rani S., Cayetano J., Arul N. és Kim D. H. (2012). IPAVS: integrated pathway resources, analysis and visualization system. *Nucleic Acids Res*, 40(Database issue): D803–D808.
170. Stelzl U., Worm U., Lalowski M., Haenig C., Brembeck F. H., Goehler H., Stroedicke M., Zenkner M., Schoenherr A., Koeppen S., Timm J., Mintzlaff S., Abraham C., Bock N., Kietzmann S., Goedde A., Toksöz E., Droege A., Krobitsch S., Korn B., Birchmeier W., Lehrach H. és Wanker E. E. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122(6): 957–968.

171. Straface E., Gambardella L., Brandani M. és Malorni W. (2013). Sex differences at cellular level: “cells have a sex”. *Sex and gender differences in pharmacology*. Vol. 214. Springer, Berlin, Heidelberg: 49–65.
172. Sun Z., Chin Y. E. és Zhang D. D. (2009). Acetylation of Nrf2 by p300/CBP augments promoter-specific DNA binding of Nrf2 during the antioxidant response. *Mol Cell Biol*, 29(10): 2658–2672.
173. 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(13): 1526–1539.
174. Szalay K. Z. és Csermely P. (2013). Perturbation centrality and turbine: a novel centrality measure obtained using a versatile network dynamics tool. *PLoS ONE*, 8(10): e78059.
175. Szalay-Beko M., Palotai R., Szappanos B., Kovacs I. A., Papp B. és Csermely P. (2012). ModuLand plug-in for cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics*, 28(16): 2202–2204.
176. Taniguchi C. M., Emanuelli B. és Kahn C. R. (2006). Critical nodes in signalling pathways: insights into insulin action. *Nat Rev Mol Cell Bio*, 7(2): 85–96.
177. Tasdemir E., Maiuri M. C., Galluzzi L., Vitale I., Djavaheri-Mergny M., D’Amelio M., Criollo A., Morselli E., Zhu C., Harper F., Nannmark U., Samara C., Pinton P., Vicencio J. M., Carnuccio R., Moll U. M., Madeo F., Paterlini-Brechot P., Rizzuto R., Szabadkai G., Pierron G., Blomgren K., Tavernarakis N., Codogno P., Cecconi F. és Kroemer G. (2008). Regulation of autophagy by cytoplasmic p53. *Nat Cell Biol*, 10(6): 676–687.
178. Taylor I. W., Linding R., Warde-Farley D., Liu Y., Pesquita C., Faria D., Bull S., Pawson T., Morris Q. és Wrana J. L. (2009). Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol*, 27(2): 199–204.
179. Taylor I. W. és Wrana J. L. (2012). Protein interaction networks in medicine and disease. *Proteomics*, 12(10): 1706–1716.
180. Thomson D. W., Bracken C. P. és Goodall G. J. (2011). Experimental strategies for microRNA target identification. *Nucleic Acids Res*, 39(16): 6845–6853.

181. Tran N. H., Choi K. P. és Zhang L. (2013). Counting motifs in the human interactome. *Nat Commun*, 4: 2241.
182. Tsesmetzis N., Couchman M., Higgins J., Smith A., Doonan J. H., Seifert G. J., Schmidt E. E., Vastrik I., Birney E., Wu G., D'Eustachio P., Stein L. D., Morris R. J., Bevan M. W. és Walsh S. V. (2008). Arabidopsis reactome: a foundation knowledgebase for plant systems biology. *Plant Cell*, 20(6): 1426–1436.
183. Tung Y.-T., Wang B.-J., Hu M.-K., Hsu W.-M., Lee H., Huang W.-P. és Liao Y.-F. (2012). Autophagy: a double-edged sword in alzheimer's disease. *J Biosci*, 37(1): 157–165.
184. UniProt Consortium (2012). Update on activities at the universal protein resource (UniProt) in 2013. *Nucleic Acids Res*, 41(Database issue): D43–D47.
185. Vaquerizas J. M., Kummerfeld S. K., Teichmann S. A. és Luscombe N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*, 10(4): 252–263.
186. Vastrik I., D'Eustachio P., Schmidt E., Joshi-Tope G., Gopinath G., Croft D., Bono B., Gillespie M., Jassal B., Lewis S., Matthews L., Wu G., Birney E. és Stein L. (2007). Reactome: a knowledge base of biologic pathways and processes. *Genome Biol*, 8(3): R39.
187. Vasudevan S., Tong Y. és Steitz J. A. (2007). Switching from repression to activation: microRNAs can up-regulate translation. *Science*, 318(5858): 1931–1934.
188. Vellai T., Takács-Vellai K., Sass M. és Klionsky D. J. (2009). The regulation of aging: does autophagy underlie longevity? *Trends Cell Biol*, 19(10): 487–494.
189. Venkatesan K., Rual J.-F., Vazquez A., Stelzl U., Lemmens I., Hirozane-Kishikawa T., Hao T., Zenkner M., Xin X., Goh K.-I., Yildirim M. A., Simonis N., Heinzmann K., Gebreab F., Sahalie J. M., Cevik S., Simon C., Smet A.-S., Dann E., Smolyar A., Vinayagam A., Yu H., Szeto D., Borick H., Dricot A., Klitgord N., Murray R. R., Lin C., Lalowski M., Timm J., Rau K., Boone C., Braun P., Cusick M. E., Roth F. P., Hill D. E., Tavernier J., Wanker E. E., Barabási A.-L. és Vidal M. (2008). An empirical framework for binary interactome mapping. *Nat Methods*, 6(1): 83–90.

190. Vergoulis T., Vlachos I. S., Alexiou P., Georgakilas G., Maragkakis M., Reczko M., Gerangelos S., Koziris N., Dalamagas T. és Hatzigeorgiou A. G. (2012). TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res*, 40(Database issue): D222–D229.
191. Via A., Gould C. M., Gemünd C., Gibson T. J. és Helmer-Citterich M. (2009). A structure filter for the eukaryotic linear motif resource. *BMC Bioinformatics*, 10(1): 351.
192. Vinayagam A., Zirin J., Roesel C., Hu Y., Yilmazel B., Samsonova A. A., Neumüller R. A., Mohr S. E. és Perrimon N. (2013). Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nat Methods*, 11(1): 94–99.
193. Walhout A. J. M. (2011). Gene-centered regulatory network mapping. *Methods Cell Biol*, 106: 271–288.
194. Wang X. és El Naqa I. M. (2008). Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics*, 24(3): 325–332.
195. Wei K., Wang P. és Miao C.-Y. (2012). A double-edged sword with therapeutic potential: an updated role of autophagy in ischemic cerebral injury. *CNS Neurosci Ther*, 18(11): 879–886.
196. Weis S., Toniazzo A., Ander B., Zhan X., Careaga M., Ashwood P., Wyse A., Netto C. és Sharp F. (2014). Autophagy in the brain of neonates following hypoxia–ischemia shows sex- and region-specific effects. *Neuroscience*, 256: 201–209.
197. Wen S., Niu Y., Lee S. O. 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(1): 31–40.
198. White E. és DiPaola R. S. (2009). The double-edged sword of autophagy modulation in cancer. *Clin Cancer Res*, 15(17): 5308–5316.
199. Wu C. C. és Yates J. R. (2003). The application of mass spectrometry to membrane proteomics. *Nat Biotechnol*, 21(3): 262–267.
200. Wu K. C., Liu J. és Klaassen C. D. (2012). Role of Nrf2 in preventing ethanol-induced oxidative stress and lipid accumulation. *Toxicol Appl Pharmacol*, 262(3): 321–329.

201. Wu L., Li X., Yang J., Liu Y., Fan X. és Cheng Y. (2013). CHD@ZJU: a knowledgebase providing network-based research platform on coronary heart disease. *Database*, 2013: bat047.
202. Xu J. és Li Y.-H. (2012). miRDeathDB: a database bridging microRNAs and the programmed cell death. *Cell Death Differ*, 19(9): 1571.
203. Xu P., Das M., Reilly J. és Davis R. J. (2011). JNK regulates FoxO-dependent autophagy in neurons. *Genes Dev*, 25(4): 310–322.
204. Yamamoto S., Sakai N., Nakamura H., Fukagawa H., Fukuda K. és Takagi T. (2011). INOH: ontology-based highly structured database of signal transduction pathways. *Database*, 2011: bar052.
205. Yang F., Li X., Sharma M., Sasaki C. Y., Longo D. L., Lim B. és Sun Z. (2002). Linking beta-catenin to androgen-signaling pathway. *J Biol Chem*, 277(13): 11336–11344.
206. Yang M., Yao Y., Eades G., Zhang Y. és Zhou Q. (2011). MiR-28 regulates Nrf2 expression through a Keap1-independent mechanism. *Breast Cancer Res Treat*, 129(3): 983–991.
207. Yang Z. és Klionsky D. J. (2010). Mammalian autophagy: core molecular machinery and signaling regulation. *Curr Opin Cell Biol*, 22(2): 124–131.
208. Yeager-Lotem E., Sattath S., Kashtan N., Itzkovitz S., Milo R., Pinter R. Y., Alon U. és Margalit H. (2004). Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc Natl Acad Sci USA*, 101(16): 5934–5939.
209. Yu N., Seo J., Rho K., Jang Y., Park J., Kim W. K. és Lee S. (2013). hiPathDB: a human-integrated pathway database with facile visualization. *Nucleic Acids Res*, 40(Database issue): D797–D802.
210. Zhang D. D. (2006). Mechanistic studies of the Nrf2-Keap1 signaling pathway. *Drug Metab Rev*, 38(4): 769–789.

211. Zhu C., Xu F., Wang X., Shibata M., Uchiyama Y., Blomgren K. és Hagberg H. (2006). Different apoptotic mechanisms are activated in male and female brains after neonatal hypoxia-ischaemia. *J Neurochem*, 96(4): 1016–1027.
212. Zuberi K., Franz M., Rodriguez H., Montojo J., Lopes C. T., Bader G. D. és Morris Q. (2013). GeneMANIA prediction server 2013 update. *Nucleic Acids Res*, 41(Web Server issue): W115–W122.

10. Saját publikációk jegyzéke

Kapcsolódó publikációk

1. Papp D., Lenti K., Módos D., Fazekas D., Dúl Z., **Türei D.**, Földvári-Nagy L., Nusinov R., Csermely P. és Korcsmáros T. (2012). The NRF2-related interactome and regulome contain multifunctional proteins and fine-tuned autoregulatory loops. *FEBS Lett*, 586(13): 1795–1802. (IF: 3,54)
2. Fazekas D.*, Koltai M.*, **Türei D.***, Módos D, Pálffy M, Dúl Z, Zsákai L, Szalay-Bekő 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: 3,15)
3. Kubisch J., **Türei D.**, Földvári-Nagy L., Dunai Z. A., 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(4): 252–261. (IF: 7,44)
4. **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: 2,84)
5. **Türei D.**, Földvári-Nagy L., Módos D., Fazekas D., Csermely P., Vellai T. és Korcsmáros T. (előkészületben). Autophagy regulatory network – an integrated resource to identify novel regulations and interactions that control autophagy.

* – megosztott első szerzőség

Egyéb publikációk

1. Hufnagel L., Gaál M., Ladányi M., Cs S., Petrányi G., Aczél D., **Türei D.** és Zimmerman D. (2005). Klímaváltozás potenciális hatásai magyarország rovarfaunájára. *VII. magyar biometriai és biomatematikai konferencia: összefoglalók*. Szerk. Szenteleki K. és Szilágyi K.: 21.
2. Hufnagel L., Sipkay C., Drégelyi-Kiss Á., Farkas E., **Türei D.**, Gergócs V., Petrányi G., Baksa A., Gimesi L., Eppich B., Dede L. és Horváth L. (2008). *Klímaváltozás: környezet–kockázat–társadalom*. Szerk. Hufnagel L. Szaktudás Kiadó Ház, Budapest. Fejezet: Klímaváltozás, biodiverzitás és közösségökológiai folyamatok kölcsönhatásai.
3. Ferenczy A. and Eppich B., Varga R. D., Bíró I., Kovács A., Petrányi G., Hirka A., Szabóky C., Isépy I., Priszter S., **Türei D.**, Gimesi L., Á G., Homoródi R. és Hufnagel L. (2009). Fenológiai jelenségek és meteorológiai indikátorok kapcsolatának összehasonlító elemzése rovar és növény adatsorok alapján. *LII. georgikon napok: gazdaságosság és/vagy biodiverzitás?* Szerk. Szenteleki K. és Szilágyi K. Keszthely: 35.
4. Vadadi-Fülöp C., **Türei D.**, Sipkay C., Verasztó C., Drégelyi-Kiss Á. és Hufnagel L. (2009). Comparative assessment of climate change scenarios based on aquatic food web modeling. *Environ Model Assess*, 14(5): 563–576. (IF: 0,97)
5. Ferenczy A., Eppich B., Varga R. D., Bíró I., Kovács A., Petrányi G., Hirka A., Szabóky C., Isépy I., Priszter S., **Türei D.**, Gimesi L., Á G., Homoródi R. és Hufnagel L. (2010). Comparative analysis of the relationship between phenological phenomena and meteorological indicators based on insect and plant monitoring. *Appl Ecol Env Res*, 8(4): 367–376. (IF: 0,38)
6. Verasztó C., Kiss K. T., Sipkay C., Gimesi L., Vadadi-Fülöp C., **Türei D.** és Hufnagel L. (2010). Long-term dynamic patterns and diversity of phytoplankton communities in a large eutrophic river (the case of river danube, hungary). *Appl Ecol Env Res*, 8(4): 329–349. (IF: 0,38)

7. Hufnagel L., Kúti Z., Hlaszny E., Reiczigel Z., Molnár M., Homoródi R., Flórián N., Gergócs V., **Türei D.** és Ladányi M. (2012). A klímaváltozás közösségökológiai hatásainak elemzése. *Fenntartható fejlődés, élhető régió, élhető települési táj; tudományos közlemények III.* Szerk. Szenteleki K. és Szilágyi K. Budapesti Corvinus Egyetem, Budapest: 7–24.
8. Komoly C., **Türei D.**, Csathó A. I., Pifkó D., Juhász M., Somodi I. és Bartha S. (2012). Fűvetés hatása a parlagfű (*Ambrosia artemisiifolia* L.) tömegességére egy tiszalparti fiatal parlagon. *Természetvédelmi Közlemények*, 18: 283–293.
9. **Türei D.** (2012). *A klímaváltozás hatása ökológiai folyamatokra és közösségekre.* Szerk. Hufnagel L. és Sipkay C. Budapesti Corvinus Egyetem, Budapest. Fejezet: Vízi és vizes élőhelyek specifikumai: 85–128.

11. Köszönetnyilvánítás

Ezúton köszönöm témavezetőmnek, Prof. Csermely Péternek, az MTA levelező tagjának a rendszeres konzultációkat, a publikációk és doktori munkám elkészítése során nyújtott értékes tanácsait, valamint munkám anyagi feltételeinek megteremtését.

Köszönöm Prof. Bánhegyi Gábornak, a Semmelweis Egyetem Orvosi Vegytani, Molekuláris Biológiai és Pathobiokémiai Intézet igazgatójának, és Prof. Mandl Józsefnek, az MTA rendes tagjának, a Molekuláris Orvostudományok Doktori Iskola vezetőjének, hogy doktori munkámat az általuk vezetett intézetben és doktori iskolában végezhettem.

Köszönettel tartozom a Semmelweis Egyetem LINK hálózatkutató csoportjának munkámat segítő tanácsaikért.

Köszönöm az ELTE NetBiol kutatócsoportjának a folyamatos együttműködést, számtalan elméleti és gyakorlati segítséget közös projektjeinkben, és a SignaLink 2 alapjául szolgáló irodalmi gyűjtésben, valamint a SignaLink 1 megalkotásában végzett munkájukat.

Szeretném köszönetemet kifejezni Dr. Korcsmáros Tamásnak, a NetBiol csoport vezetőjének, az elmúlt három év során nyújtott szakmai útmutatásért, a projektek koordinálásában mutatott felkészültségéért és rugalmasságáért, mely nélkül doktori munkám és publikációink nem lehettek volna sikeresek.

Külön köszönet illeti Fazekas Dávidot, a NetBiol csoport tagját, aki informatikai kérdésekben számtalan hasznos tanáccsal segített, és elkészítette az adatbázisok letöltő modulját.

Hálával tartozom Dr. Papp Diánának, aki a Semmelweis Egyetem Orvosi Vegytani Intézetében doktoranduszként elkészítette az NRF2 szabályozási hálózatot. Az ő munkája tette lehetővé az NRF2ome kifejlesztését.

Köszönöm Földvári-Nagy Lászlónak, a NetBiol csoport tagjának, az autofágia fehérjék közti kapcsolatok irodalmi gyűjtését, és az együttműködést az ezekre épülő ARN fejlesztése során.

Külön köszönet illeti Módos Dezsőt, aki a NetBiol csoport tagjaként statisztikai módszerekkel segített a predikciók eredményében megtalálni a hamis pozitívok és hamis negatívok

közti egyensúlyt, ezenkívül a mikro-RNS adatbázisok adatainak kezelésében is pótolhatatlan segítséget nyújtott.

Köszönöm Kubisch Jánosnak, a NetBiol csoport tagjának a 26. ábra elkészítését, és informatikai kérdésekben nyújtott hasznos tanácsait.

Köszönettel tartozom Farkas J. Illésnek és Dúl Zoltánnak a weboldalak tesztelése során észrevett hibákért.

A dolgozatomban bemutatott eredményeket a TÁMOP 4.2.1./B-09/1/KMR-2010-0003 és 4.2.2/B-10/1-2010-0013 számú pályázata, az OTKA K83314, K75334 és NK78012 számú, és az NKTH 5LET-08-2-2009-0041 pályázatainak támogatásával valósíthattam meg.

A munkám során felhasznált tudományos publikációk internetes elérhetőségét a Semmelweis Egyetem Központi Könyvtára biztosította.

Hálával gondolok a szabad szoftver mozgalomra, különösen az Apache, az Arch Linux, a Cytoscape Web, a GIMP, az igraph, az Inkscape, a jQuery, a L^AT_EX, a Linux kernel, a MySQL, a PHP, a Python és az R projekteket fejlesztő közösségekben részt vevő sok ezer programozóra. Ezen nagyszerű eszközök nélkül munkám informatikai része nem lett volna lehetséges.

Végül köszönöm családomnak, és a Lujza utcai kommuna lakóinak: Bezgődi Hajnalkának, Hódi Csillának, Kathrin Sharonnak, Krasznahorkai Emmának, Kremmer Saroltának és Lujzának, hogy mindvégig velem voltak, és támogattak munkám során.