

PA052: Úvod do systémové biologie

David Šafránek

13.10.2011



Obsah

Biologické sítě

Rekonstrukce biologických sítí

Ontologie biologických znalostí

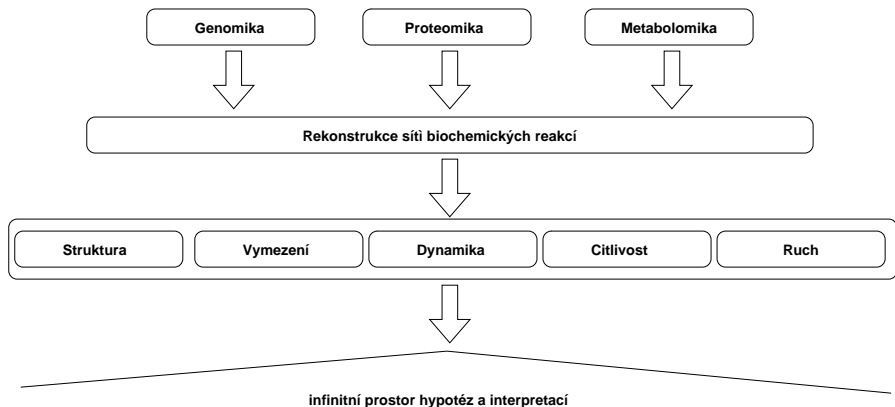
Obsah

Biologické sítě

Rekonstrukce biologických sítí

Ontologie biologických znalostí

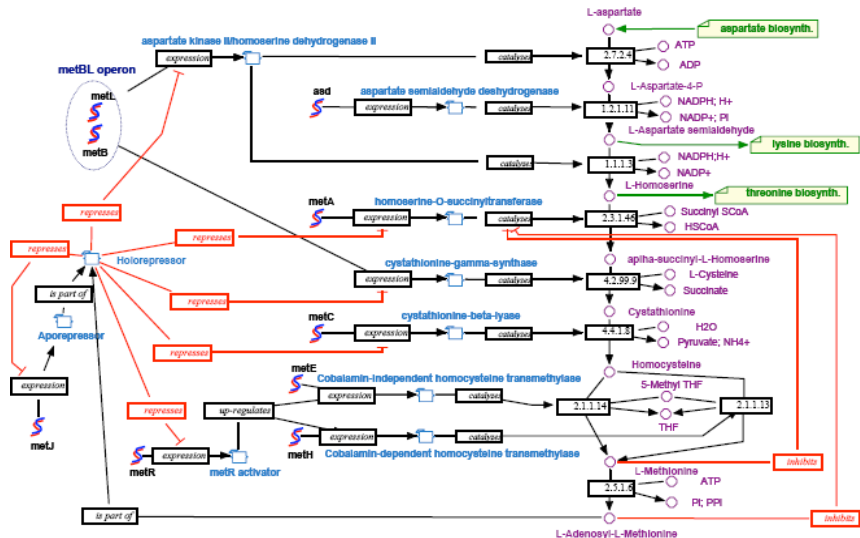
Rekonstrukce biologických sítí v kontextu SB



Biologické sítě a dráhy

- biochemická interakce molekul popsaná grafem
- uzly
 - molekuly/komplexy biochemických látek
 - biochemické reakce
- hrany
 - regulace (aktivace, represe, katalýza)
 - příslušnost k reakci (produkt, zdroj)
- dráhy — zaměřené na určitá specifika (látky, reakce)
 - typicky signální dráhy
- sítě — komplexní interakce
- různé úrovně abstrakce, různé notace, např. Kohn's diagrams
<http://www.nature.com/msb/journal/v2/n1/full/msb4100044.html>

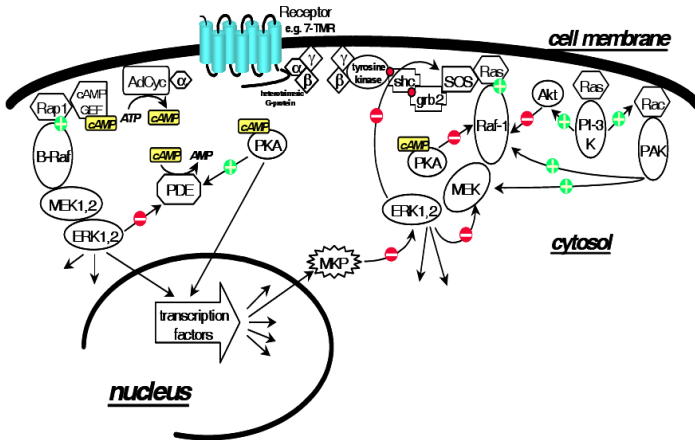
Příklad komponenty biologické sítě



Biologické sítě a dráhy

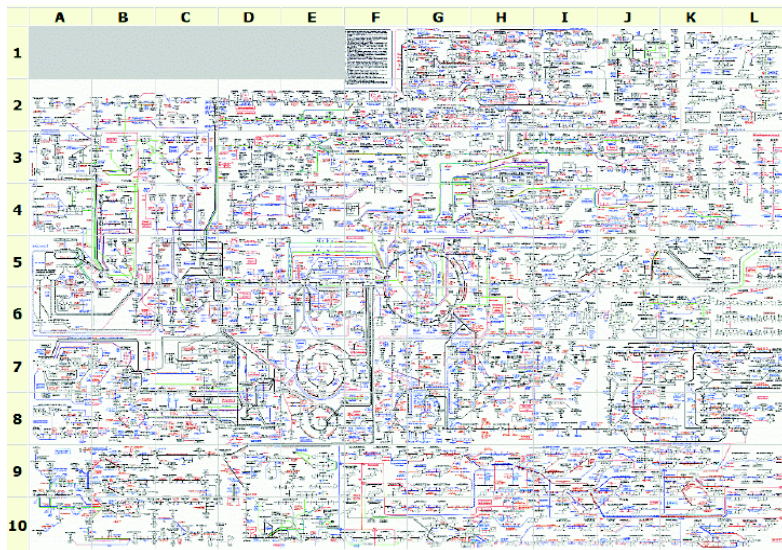
- komplexy vzájemně interagujících chemických reakcí
- klíčem ke studiu fyziologie organismu
- představují základní informaci pro tvorbu in silico modelu
- v průběhu evoluce může docházet k přidávání/ubírání hran

Biologické sítě a dráhy

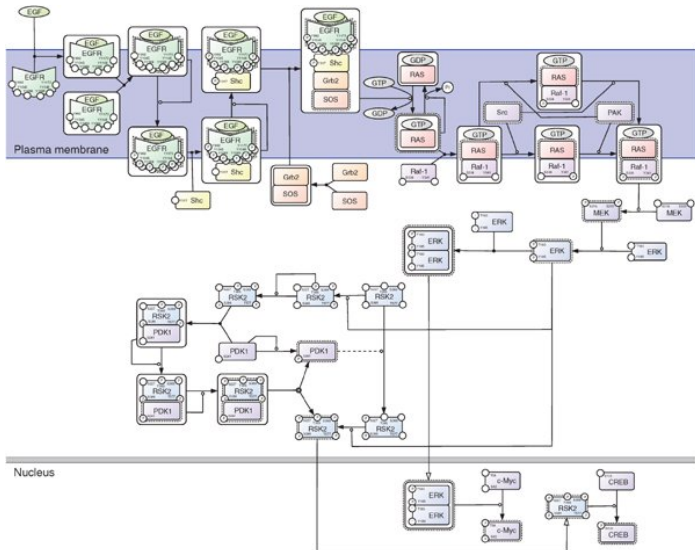


- neformální notace
- vyvíjejí se standardy — SBGN (podporuje např. CellDesigner)

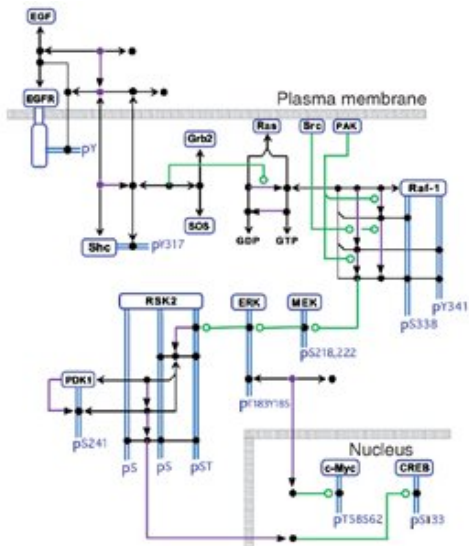
Metabolická síť – ad hoc diagram



Signální síť – SBGN



Signální síť – Kohnova mapa



Obsah

Biologické sítě

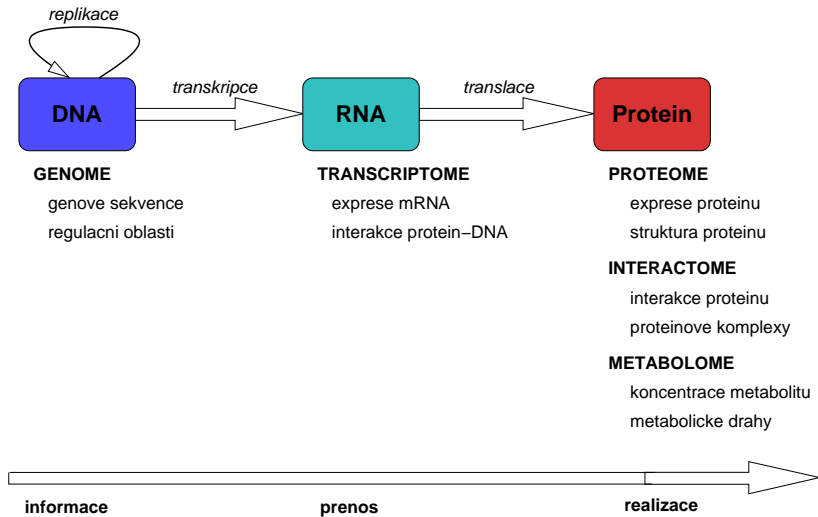
Rekonstrukce biologických sítí

Ontologie biologických znalostí

Původ biologických dat

- tradiční biochemie a genetika
 - výsledky redukcionistického období biologie
 - forma: vědecké články
 - získání dat: dolování znalostí z textů (text mining)
- low-throughput technologie
 - separace proteinů prostřednictvím antigenů
 - genové reportery
- high-throughput technologie
 - sekvenování genomu
 - měření exprese genů na DNA mikročipu
 - hmotnostní spektrometrie
 - proteomika a metabolomika

Rekonstrukce v kontextu centrálního dogmatu



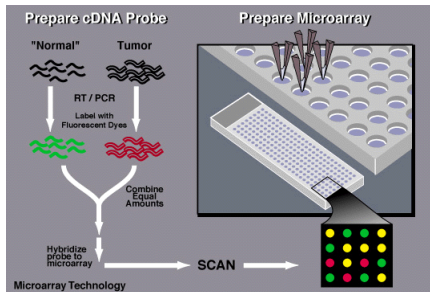
Data na úrovni RNA

- exprese mRNA
 - DNA microarray: high-throughput data (genome-scale expresní profily), typicky na populační úrovni
 - individuální informace pomocí genových reportérů (lze i na úrovni buňky)
 - srovnání s knock-out/knock-in mutacemi
 - dolování hypotéz o regulačních interakcích (shlukováním profilů)
 - transkriptom poskytuje hrubý odhad o expresi proteinů, nemusí být vypovídající (post-transkripční modifikace, regulace degradace mRNA, splicing)
- interakce protein-DNA
 - ChIP-chip (chromatin immunoprecipitation + microarray): přímá detekce vazeb transkripčních faktorů na DNA

K rekonstrukci regulační sítě je nezbytné integrovat data z úrovně genomu a transkriptomu.

Měření *mRNA* exprese

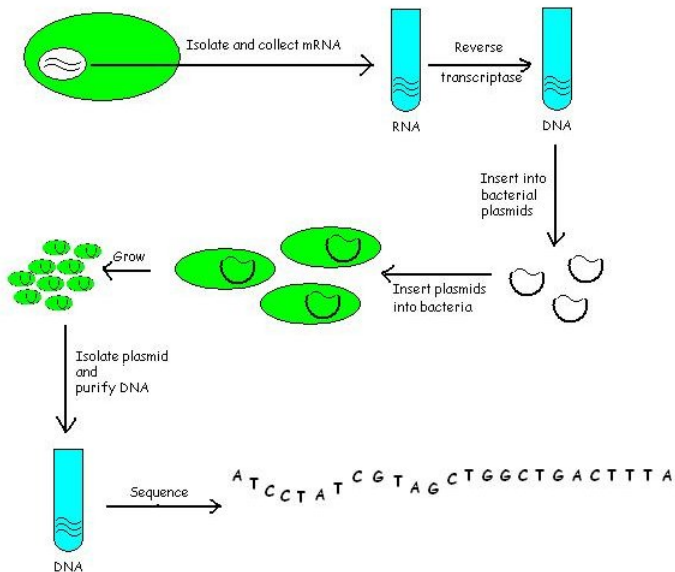
- nejpoužívanějším nástrojem je technologie DNA microarray
 - v daném okamžiku je paralelně nasamplována exprese všech genů v genomu příslušného organismu
 - postaveno na relativním srovnání minimálně dvou různých vzorků
 - exprese v přítomnosti vs. nepřítomnosti O_2
 - exprese při knock-outu určitého genu vs. normální stav
 - ...



Postup při DNA microarray experimentu

1. konstrukce čipu z cDNA knihovny (amplifikace a rozmístění sekvencí)
2. odběr celkové mRNA z experimentálních vzorků (typicky 2)
3. reverzní transkripce do cDNA asociované s fluorescenčním barvivem
4. hybridizace odebrané cDNA s cDNA na čipu
5. omytí čipu a oskenování výsledku
6. analýza dat
7. komerční čipy používají místo cDNA knihovny skupinu oligonukleotidů pro každý gen
⇒ pouze jeden vzorek mRNA je analyzován na jednom čipu (porovnání více identicky připravených čipů)

Reverzní transkripce cDNA a amplifikace



Využití Polymerase Chain Reaction (PCR)

- umožňuje replikaci určité části DNA (forma amplifikace)
- DNA je zahřátím rozdělena
- úsek DNA je označen párem oligonukleotidů (15-25 bazí)
 - při snížení teploty hybridizace oligonukleotidů s řetězcem DNA
 - doplnění zbývající sekvence DNA prostřednictvím RNA polymerázy
- <http://www.dnalc.org/resources/animations/pcr.html>
- lze využít i pro mRNA: RT-PCR (reverse transcription PCR)
 - reverzní transkripce mRNA do cDNA
 - amplifikace cDNA (PCR)

Databáze microarray dat

- Stanford Microarray Database – různé pohledy na data, filtrace
<http://smd.stanford.edu/cgi-bin/cluster/drpGetData.pl>
- ArrayExpress – statisticky zpracovaná data
<http://www.ebi.ac.uk/gxa/>
- Gene Expression Omnibus (GEO)
<http://www.ncbi.nih.gov/geo/>
- MUSC DNA Microarray Database
<http://proteogenomics.musc.edu/ma/>
- GenExpDB (E. Coli specifická data)
<http://genexpdb.ou.edu/>

GenExpDB

Welcome to the *E. coli* Community's Gene Expression Database (GenExpDB)

Hosted by the [University of Oklahoma](#)

Instructions: Search *E. coli* GenExpDB using either gene, b-number or location (multiple entries separate by comma or space).

Example: [edd](#) • [b3517](#) • [laca,lacy,lacz](#) • [416366 \(location\)](#) • [reset](#)

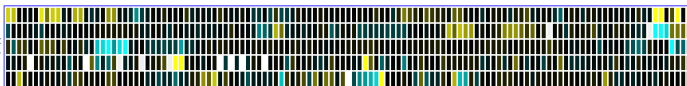
Search by: Operon Regulon Sigma Text Experiment

MultiFun

Query

Pearson

lacZ
b0344

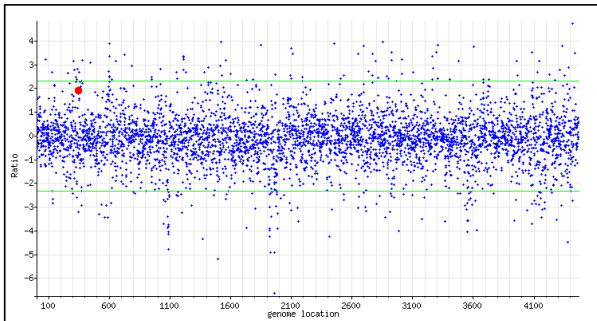


GSE15406

ppGpp0 GraL / ppGpp0 control (short term)

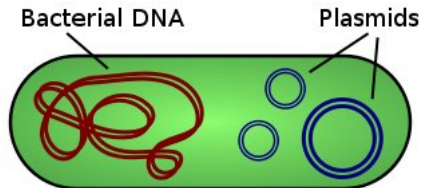
Data Count: 4368

[close](#)

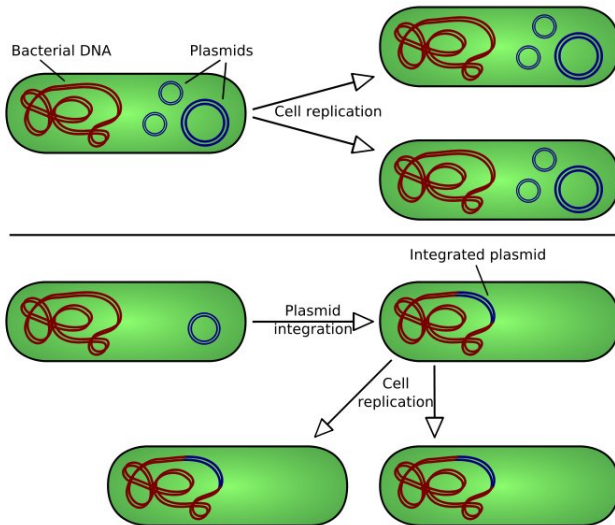


Měření *mRNA* exprese pomocí genových reportérů

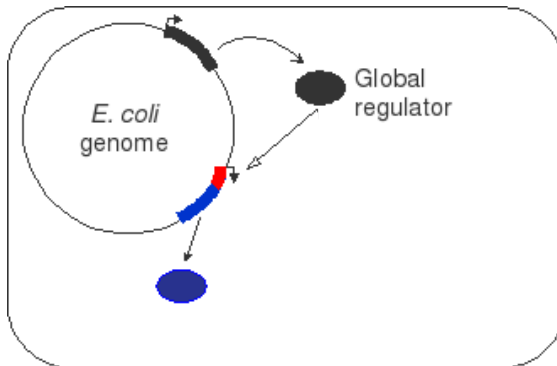
- genovým reporterem se rozumí umělá subsekvence DNA zavedená do buňky
 - formou persistentního syntetického plasmidu (externí molekula DNA zavedená do buňky)
 - modifikací chromozomu (prostřednictvím plasmidu)



Zavedení plasmidu do buňky

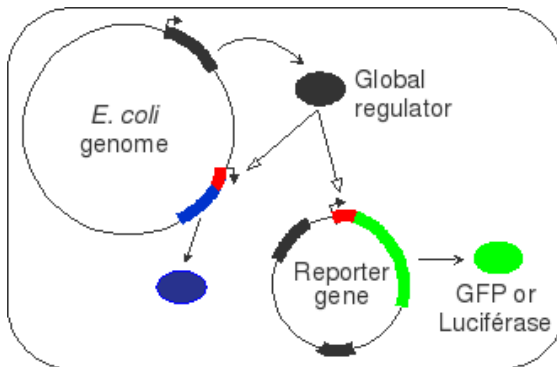


Schema aplikace genového reporteru



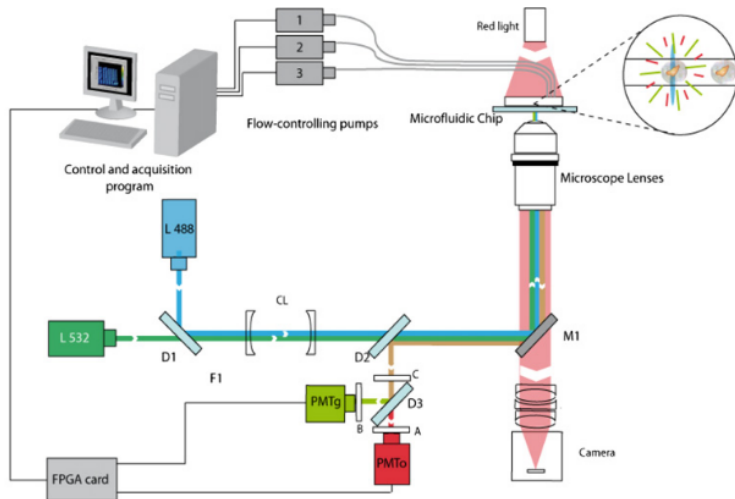
- genový reporter převádí aktivitu na určitém promotoru na měřitelný jev (nejčastěji fluorescence – GFP, luciferase)
- měření reportovaného jevu probíhá na populační úrovni, lze i single-cell (digital microfluidics)

Schema aplikace genového reporteru



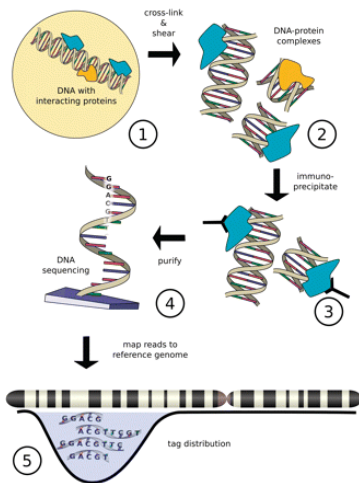
- genový reporter převádí aktivitu na určitém promotoru na měřitelný jev (nejčastěji fluorescence – GFP, luciferase)
- měření reportovaného jevu probíhá na populační úrovni, lze i single-cell (digital microfluidics)

Mikrofluidická soustava pro genové reportéry



Baret, Jean-Christophe; Beck, Yannick; Billas-Massobrio, Isabelle; Moras, Dino; Griffiths, Andrew D. Quantitative Cell-Based Reporter Gene Assays Using Droplet-Based Microfluidics. *Chemistry & biology* doi:10.1016/j.chembiol.2010.04.010 (volume 17 issue 5 pp.528 - 536)

Detekce vazeb protein-DNA



Szalkowski, A.M, and Schmid, C.D.(2010). Rapid innovation in ChIP-seq peak-calling algorithms is outdistancing benchmarking efforts. Briefings in Bioinformatics.

Data rekonstruovaných regulačních sítí

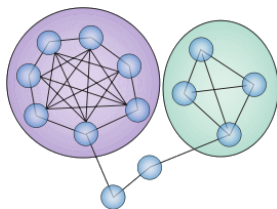
	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>S. cerevisiae</i> (ChIP-chip)
regulující geny	123	109	203
regulované geny	451	418	1296
regulační interakce	1468	945	3353

- *E. coli* – RegulonDB (<http://regulondb.ccg.unam.mx/>)
- *S. cerevisiae* –
http://web.wi.mit.edu/young/regulator_network/

Christopher T. Harbison, et al. Transcriptional regulatory code of a eukaryotic genome. Nature 431, 99-104 (2 September 2004). doi:10.1038/nature02800.

S Shen-Orr, R Milo, S Mangan & U Alon. Network motifs in the transcriptional regulation network of Escherichia coli. Nature Genetics, 31:64-68 (2002).

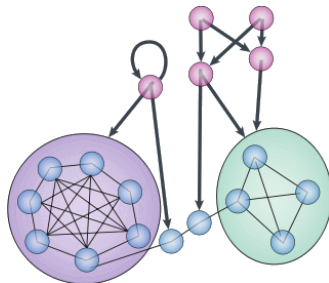
Úrovně pohledu na regulační interakce



koexpresní síť

uzly: geny

hrany: podobný expresní profil



genová regulační síť

uzly: geny a TF

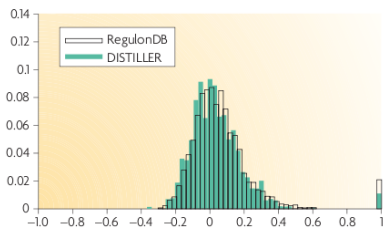
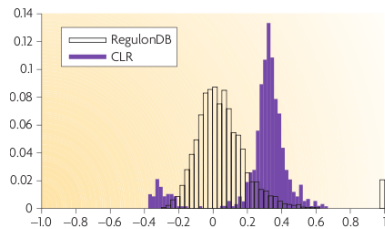
hrany: regulace

Problémy rekonstrukce

- uvažme *E. coli*
 - ⇒ 4500 genů, z toho 300 kódujících známý TF
- máme 4500×300 možností interakcí ...
- navíc dochází ke kombinatorickým regulacím
- nutno uplatnit optimalizační strategie:
 - shlukování expresních profilů (clustering)
 - integrace dat
 - metody řízené dotazy (query-driven reconstruction pomocí bayesovských závislostí)

Problémy rekonstrukce

Srovnání integračních a expresních metod

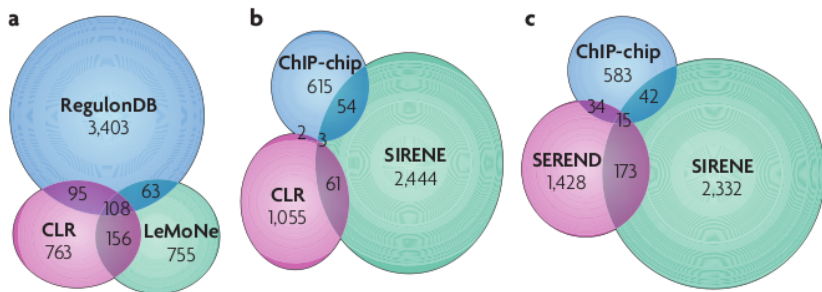


Riet De Smet & Kathleen Marchal. Advantages and limitations of current network inference methods. *Nature Reviews Microbiology* 8, 717-729 (October 2010). doi:10.1038/nrmicro2419

- DISTILLER – nástroj postavený na integrační metodě (integrace expresních a interakčních dat)
- CLR – nástroj postavený na výpočtu korelace expresních dat

Problémy rekonstrukce

Divergence metod

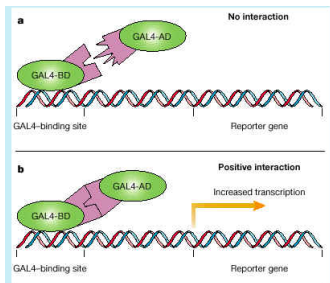


Riet De Smet & Kathleen Marchal. Advantages and limitations of current network inference methods. Nature Reviews Microbiology 8, 717-729 (October 2010). doi:10.1038/nrmicro2419

- LeMoNe – nástroj postavený na strojovém učení z expresních profilů (shlukování)
- SIRENE – řízené strojové učení
- SEREND – částečně řízené strojové učení + integrace dat

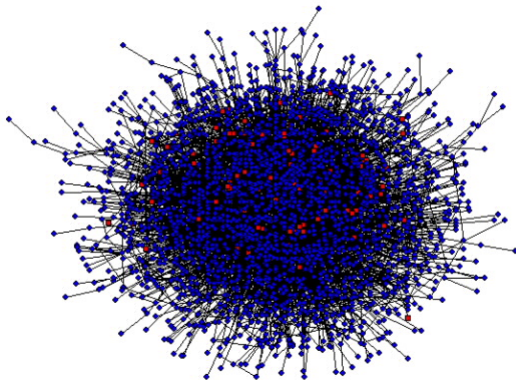
Data na úrovni proteomu a interaktomu

- exprese proteinů
 - proteinové chipy (problémové – např. proteinové interakce závislé na termodynamických podmínkách)
 - hmotnostní spektrometrie (separace chromatografií, ionizace a měření rychlosti pohybu mezi elektrodami)
- detekce interakcí protein-protein
 - proteinové chipy (univerzální použití)
 - Y2H (yeast two-hybrid system)



Data na úrovni proteomu a interaktomu

Sítě proteinových interakcí

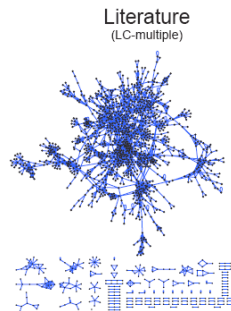
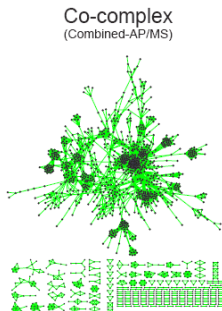
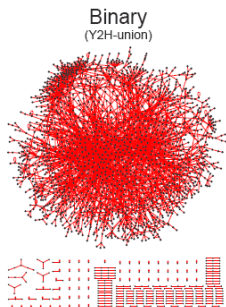


interakce proteinů člověka

zdroj: <http://www.estradalab.org/research/index.html>

Data na úrovni proteomu a interaktomu

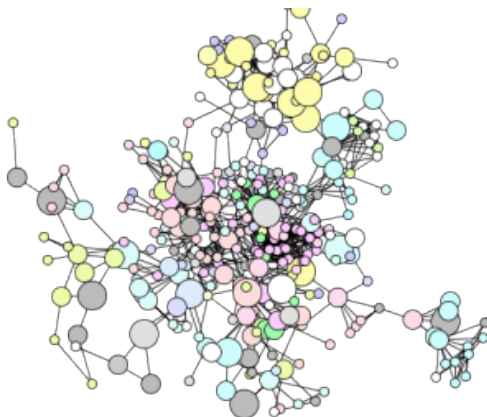
*Interakce proteinů *S. cerevisiae**



Yeast Interactome Database –
http://interactome.dfci.harvard.edu/S_cerevisiae/

Data na úrovni proteomu a interaktomu

*Interakce proteinů *E. coli**



Bacteriome – <http://www.compsysbio.org/bacteriome/>

Data na úrovni proteomu a interaktomu

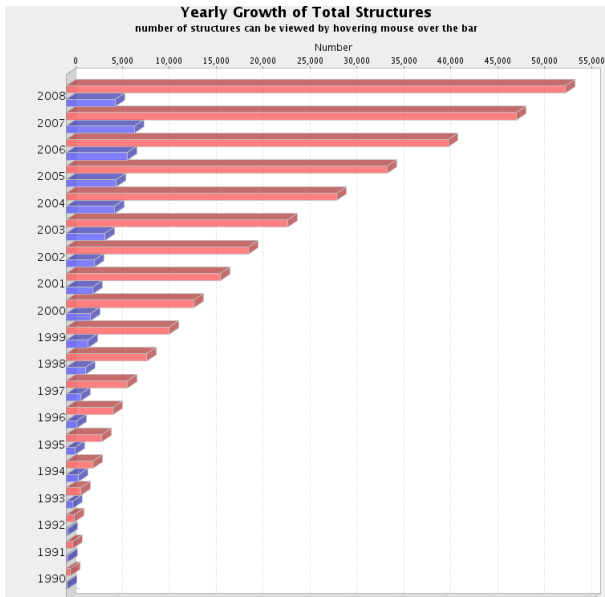
Další zdroje

- *C. elegans* – http://interactome.dfci.harvard.edu/C_elegans/
- *Drosophila* – <http://www.droidb.org/>

Proteinová databanka PDB

- databáze 3D strukturálních dat proteinů a nukleových kyselin
- založena 1971
- veřejný přístup
- <http://www.rcsb.org/pdb/>
- PDB formát souborů –
<http://www.wwpdb.org/documentation/format30>
- primární struktura (sekvence aminokyselin) – FASTA formát
- vyhledávání proteinů dle sekvencí – BLAST –
<http://www.ebi.ac.uk/Tools/blast2/>

Proteinová databanka PDB



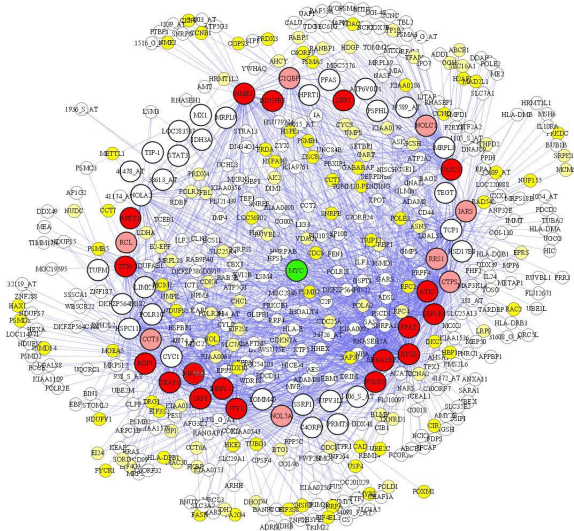
Data na úrovni metabolomu

- interakce mezi metabolity determinovány chemicky (proteiny ve funkci enzymů)
- měření metabolitů pomocí hmotnostní spektrometrie
- měření aktivity enzymů na úrovni proteomu/transkriptomu
- integrace s daty na nižších úrovních
- data získaná z literatury a genových sekvencí (pro daný organismus)
- data dostupná ve veřejných databázích
- problémy mohou vznikat díky chybné identifikaci sekvencí

Data na úrovni metabolomu

Organismus	# enz. genů	# metabolitů	# reakcí
<i>E. coli</i>	904	625	931
<i>S. cerevisiae</i>	750	646	1149
<i>Staphylococcus aureus</i>	619	571	640
<i>Haemophilus influenzae</i>	362	343	488
<i>Helicobacter pylori</i>	268	340	444

Metabolické sítě



zdroj: <http://www.di.unipi.it/~braccia/ToolCode/>

KEGG – Kyoto Encyclopedia of Genes and Genomes

- integrativní databáze genových produktů
- genový prostor
⇒ GENES, GENOME, ORTHOLOGY, Organisms
- chemický prostor
⇒ COMPOUND, GLYCAN, REACTION, ENZYME, LIGAND
- systémový prostor
⇒ PATHWAY, BRITE, DISEASE, DRUG
- každý pojem má jednoznačné ID
- zachycena ortologie (podobnost) genů
- genový prostor čerpá aktuálně z několika zdrojů

KEGG – ORGANISMS



GENOME: Escherichia coli K-12 MG1655

[Help](#)

Entry	T00007	Complete Genome
Name	eco, E.coli, ECOLI, 511145	
Definition	Escherichia coli K-12 MG1655	
Annotation	manual	Show organism
Taxonomy Lineage	TAX:511145 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia	
	Taxonomy	
Data source	RefSeq (Project:225)	
Original DB	Wisconsin Pasteur RegulonDB EcoGene ECOCYC	
Chromosome Sequence Length	Circular RS:NC_000913 4639675	
Statistics	Number of nucleotides:	4639675
	Number of protein genes:	4144
	Number of RNA genes:	175
Reference Authors	PMID: 9278503 Blattner FR, et al.	
Title	The complete genome sequence of Escherichia coli K-12.	
Journal	Science 277:1453-74 (1997)	

KEGG – ORGANISMS



[Brite menu]

▼ ▼ ▼ ▼ ▼

▶ **Eukaryotes** (248)

▼ **Prokaryotes** (1189)

▼ **Bacteria** (1099)

▼ **Gammaproteobacteria** (265)

▼ **Escherichia** (31)

eco Escherichia coli F-12 H01655
 ecj Escherichia coli F-12 K3119
 ecd Escherichia coli F-12 DB108
 ebx Escherichia coli F-12 MC4100(MuLac) H02952
 eee Escherichia coli O157:H7 EDL933 (EHEC)
 ees Escherichia coli O157:H7 Sakai (EHEC)
 eef Escherichia coli O157:H7 EC4115 (EHEC)
 etw Escherichia coli O157:H7 Tc14359 (EHEC)
 eoJ Escherichia coli O26:H11 11368 (EHEC)
 eoi Escherichia coli O111:H- 11128 (EHEC)
 eoh Escherichia coli O103:H2 12009 (EHEC)
 ecq Escherichia coli O127:H6 E2348/69 (EPEC)
 eok Escherichia coli O55:H7 CB9615 (EPEC)
 eec Escherichia coli O6:H2:H1 CFT073 (UPEC)
 ecp Escherichia coli O6:H15:H31 536 (UPEC)
 eei Escherichia coli O18:H1:H7 UT189 (UPEC)
 ecv Escherichia coli O1:H1:H7 (APEC)
 ecx Escherichia coli O9 HS (commensal)
 ecw Escherichia coli O139:H20 E24377A (ETEC)
 ecm Escherichia coli SBE3-5 (environmental)
 ecy Escherichia coli O152:H20 SE11 (commensal)
 ecr Escherichia coli O8 IAL1 (commensal)
 ecq Escherichia coli O81 ED1a (commensal)
 eck Escherichia coli 55989 (EPEC)
 ect Escherichia coli O7:H1 IA139 (ExpEC)
 eun Escherichia coli O17:H52:H18 URH026 (ExpEC)
 eoz Escherichia coli O45:H1:H7 S08 (ExpEC)
 ecl Escherichia coli C ATCC 8739
 ebr Escherichia coli B PHL606
 ebd Escherichia coli HL21-Gold(DK3)plysS AG
 efe Escherichia fergusonii

▶ **Salmonella** (16)

▶ **Yersinia** (13)

▶ **Shigella** (7)


▶ **Pectobacterium** (3)

KEGG – GENES



Escherichia coli K-12 MG1655: b1891

Help

Entry	b1891 CDS E.coli
Gene name	flhC, ECK1892, flaI, JW1880
Definition	DNA-binding transcriptional dual regulator with FlhD
Orthology	K02402 flagellar transcriptional activator FlhC
Pathway	eco02020 Two-component system eco02040 Flagellar assembly
Class	Environmental Information Processing; Signal Transduction; Two-component system [PATH: eco02020] Cellular Processes; Cell Motility; Flagellar assembly [PATH: eco02040] BRITE hierarchy
SSDB	Ortholog Paralog Gene cluster GFIT
Motif	Pfam: FlhC Motif
Other DBs	NCBI-GI: 16129843 NCBI-GeneID: 947280 RegulonDB: B1891 EcoGene: EG10319 ECOCYC: EG10319 UniProt: P0ABY7
Structure	PDB: 2AVU Thumbnails  Jmol

http://www.genome.jp/dbget-bin/www_bget?eco:b1891

KEGG – ORTHOLOGY

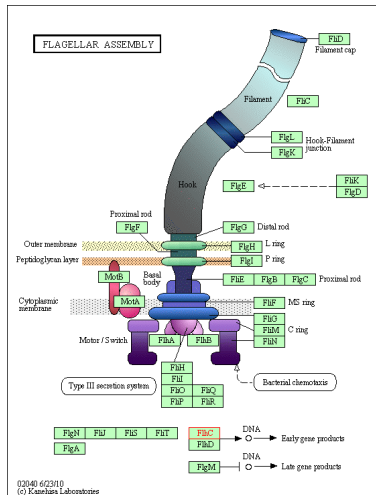


ORTHOLOGY: K02402

[Help](#)

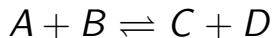
Entry	K02402 KO
Name	flhC
Definition	flagellar transcriptional activator FlhC
Pathway	ko02020 Two-component system ko02040 Flagellar assembly
Class	Environmental Information Processing; Signal Transduction; Two-component system [PATH: ko02020] Cellular Processes; Cell Motility; Bacterial motility proteins [BR: ko02035] Cellular Processes; Cell Motility; Flagellar assembly [PATH: ko02040] BRITE hierarchy
Genes	ECO: b1891 (flhC) ECJ: Jw1880 (flhC) EBW: BwG_1705 (flhC) ECE: Z2945 (flhC) ECS: ECs2601 ECF: ECH74115_2630 (flhC) ETW: ECSP_2465 (flhC) EOJ: EC026_2743 (flhC) EOI: EC0111_2477 (flhC) EOH: EC0103_2153 (flhC) ECG: E2348C_2014 (flhC) EOK: G2583_2344 (flhC) ECC: c2306 (flhC) ECP: ECP_1835 ECI: UT189_C2094 (flhC)

KEGG – PATHWAYS



Data (bio)chemických reakcí

- vstupy reakce – substráty (reaktanty)
- výstupy reakce – produkty

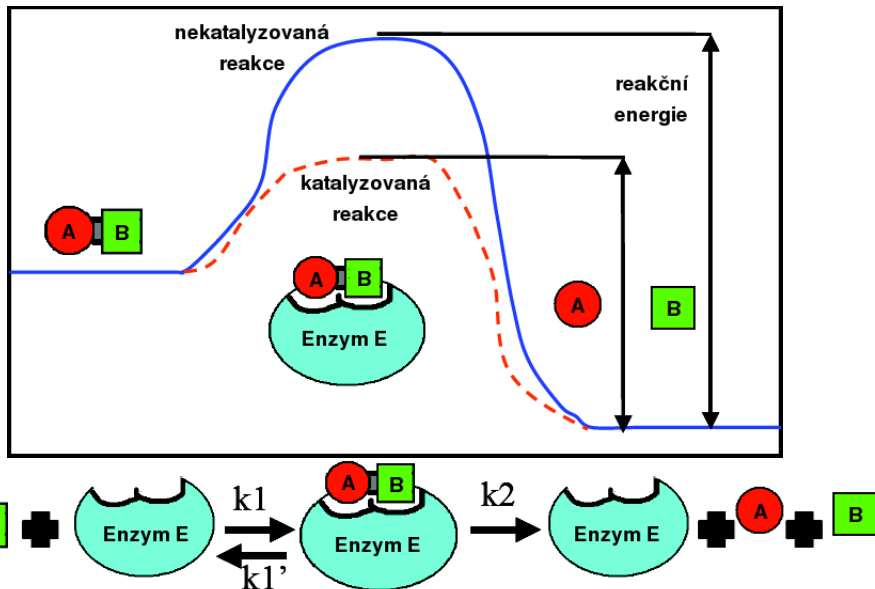


- řídicí látky (specifické proteiny — enzymy, koenzymy)
 - katalýza – urychlení/zpomalení reakce (aktivátory/inhibitory)
 - zesilování katalytických účinků enzymů koenzymy
 - např. katalytické urychlení rozpadu:



- pilířem je komplex katalytických reakcí – metabolismus
- metabolismus umožňuje zpracování a přeměnu energie

Energetický průběh reakce



Složitost metabolismu

- v e.coli je více než 1000 reakcí
- opakující se motivy
 - zpracování ATP (adenozintrifosfát) – zisk energie
 - omezené množství aktivních meziproductů
 - jen cca 100 molekul má zásadní funkci
 - omezené množství typů reakcí
- individuální reakce jsou typicky jednoduché (omezené množství substrátů/productů)

Klasifikace reakcí

- každá reakce identifikována EC číslem
- definováno dle názvosloví enzymů

<http://www.chem.qmw.ac.uk/iubmb/enzyme>

EC 3.2.1.23

typ reakce

druh substrátu

popis akceptoru

subidentifikace

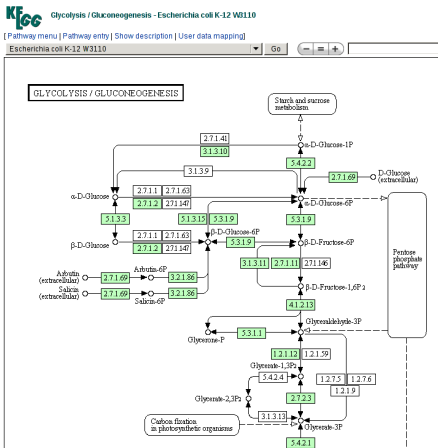
Klasifikace reakcí

- celkem 8 typů reakcí (enzymů)
 - oxydoreduktázy, transferázy, hydrolázy, ...
- EC 3.2.1.23 – β -galaktocidáza
 - hydrolýza inhibující tvorbu laktózy, součást galaktózové metabolické dráhy
 - http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/enzymes/GetPage.pl?ec_number=3.2.1.23

Databáze enzymů, metabolických drah

- BRENDA — <http://www.brenda-enzymes.org/>
⇒ zahrnuje kvantitativní data
- IntEnz — <http://www.ebi.ac.uk/intenz/>
⇒ zaměřeno na názvosloví
- KEGG — <http://www.genome.jp/kegg/>
⇒ schemata metabolických drah
<http://www.genome.jp/kegg/pathway.html>
- MetaCyc — <http://metacyc.org/>

KEGG – PATHWAYS



http:

//www.genome.jp/kegg-bin/show_pathway?org_name=ecj&mapno=00010&mapscale=&show_description=hide

Cvičení

Uvažujme sekvenci aminokyselin (FASTA formát):

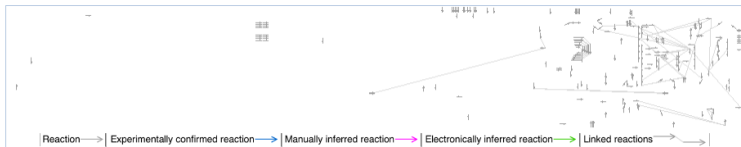
```
SVVEEHGQLSISNGELVNERGEVQLKGMSSHGLQWYGQ
FVNYESMKWLRDDWGINVFRAAMYTSSGGYIDDP
VKEKVKEAVEAAIDLDIYVIIDWHILSDNDPNIYK
EEAKDFFDEMSELYGDYPNVIYEIANEPNGSDVTW
GNQIKPYAEEVIPIIRNNDPNNIIIVGTGTWSQDV
HHAADNQLADPNVMYAFHFYAGTHGQNLRDQVDYA
LDQGAAIFVSEWGTSAAATGDGGVFLDEAQVWIDFM
DERNLSWANWLSLTHKDESSAALMPGANPTGGWTEA
ELSPSGTFVREKIRE
```

- najděte na <http://www.ebi.ac.uk/intenz/>
- prozkumejte relevantní dráhy na <http://www.genome.jp/kegg/>

Reactome

Reactome - a curated knowledgebase of biological pathways

The data displayed is for **Escherichia coli** Use the menu to change the species. Check for cross-species comparison.



Apoptosis	Axon guidance	Biological oxidations	Botulinum neurotoxicity
Cell junction organization	Cell Cycle Checkpoints	Cell Cycle, Mitotic	Chromosome Maintenance
DNA Repair	DNA Replication	Diabetes pathways	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.
Gene Expression	Hemostasis	HIV infection	Interactions of the immunoglobulin superfamily (IgSF) member proteins
Influenza infection	Integration of energy metabolism	Integrin cell surface interactions	Membrane Trafficking
Metabolism of amino acids and derivatives	Metabolism of carbohydrates	Metabolism of lipids and lipoproteins	Metabolism of nitric oxide
Metabolism of nucleotides	Metabolism of porphyrins	Metabolism of proteins	Metabolism of RNA
Metabolism of vitamins and cofactors	Muscle contraction	mRNA Processing	Myogenesis
Pyruvate metabolism and Citric Acid (TCA) cycle	Regulation of beta-cell development	Regulatory RNA pathways	Signaling by BMP
Signaling by EGFR	Signaling by FGFR	Signaling by GPCR	Signaling by PDGF
Signaling in Immune system	Signaling by insulin receptor	Signalling by NGF	Signaling by Notch

<http://www.reactome.org/>

Specificky zaměřené zdroje dat

- EcoCyc — <http://ecocyc.org> - E. coli K12
- SGD — <http://www.yeastgenome.org/>
- WORMBASE – http://www.sanger.ac.uk/Projects/C_elegans/WORMBASE/
- FlyBase — <http://flybase.org/>
- ...

Obsah

Biologické sítě

Rekonstrukce biologických sítí

Ontologie biologických znalostí

Databáze odborných článků

- PubMed Entrez – vyhledávání v odborných biologických/lékařských článcích

`http://www.ncbi.nlm.nih.gov/sites/entrez`

- Maintained by U.S. National Library of Medicine
- obsahuje okolo 15 mil. referencí
- zahrnuje články více než 4000 časopisů
- slovník pojmů MeSH

Ontologie genů a genových produktů

- nutnost systematicky uchopit genom, genové produkty a funkce (interaktom)
- akumulace biologických dat
 - decentralizovaný proces
 - paralelní proces
- problém: nejednoznačné popisy téhož objektu, procesu
 - např. proteolýza vs. (řízená degradace proteinů)
- od roku 1998 vývoj Gene Ontology
⇒ <http://geneontology.org>

Gene Ontology

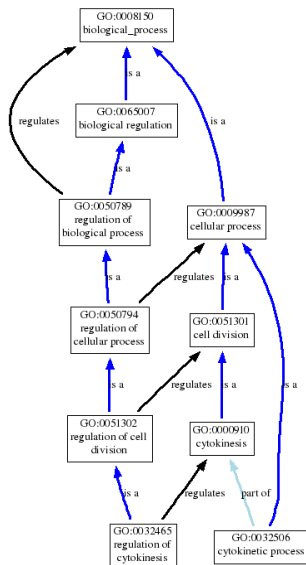
- cílem je ucelený systematický popis genových produktů a jejich funkcí
- ontologie představuje systematický slovník pojmů
 - hierarchický pohled na data včetně vazeb mezi nimi (DAG)
 - zachycení synonym
- GO obsahuje 3 kategorie (DAGy):
 - biological process
 - cellular component
 - molecular function

Gene Ontology

- respektuje standard OBO a OWL (W3C) pro strukturu a reprezentaci ontologií
 - <http://obofoundry.org/>
 - <http://www.w3.org/TR/owl-features/>
- každý uzel představuje jednoznačný pojem (množinu synonym)
⇒ jednoznačně reprezentován ID (tzv. GO termem)
- různé typy relací mezi uzly:
 - `part_of`, `is_a`, `located_in`, `derived_from`, ...
 - viz <http://obofoundry.org/ro/>

Gene Ontology





Příklad podgrafu GO



Nástroje pro Gene Ontology

- on-line i off-line vyhledávače v GO
- statistické testy na overreprezentaci dané množiny genů v pojmech GO
 - Gostat –
<http://gostat.wehi.edu.au/cgi-bin/goStat.pl>
 - DAVID – <http://david.abcc.ncifcrf.gov/>
 - BiNGO –
<http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html>
- mapování microarray dat na ontologický strom
 - eGOn – <http://www.genetools.no/>
 - High-Throughput GoMiner –
<http://discover.nci.nih.gov/gominer/htgm.jsp>
 - Meta Gene Profiler (MetaGP) – <http://metagp.ism.ac.jp/>
- ...

Literatura

-  Kitano, H. *Foundations of Systems Biology*. MIT Press, 2001.
-  Palsson, B. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, 2006.
-  Rigoutsos, I. and Stephanopoulos, G. *Systems Biology Volume II: Networks, Models, and Applications*. Oxford University Press, 2004.
-  Klipp, E. et al. *Systems Biology in Practise: Concepts, Implementation and Application*. Wiley-VCH, 2005.