MUNI | RECETOX

Detekce biomarkerů z omics experimentů

- Mgr. Eva Budinská, PhD
- RECETOX
- eva.budinska@recetox.muni.cz

- podzim 2024

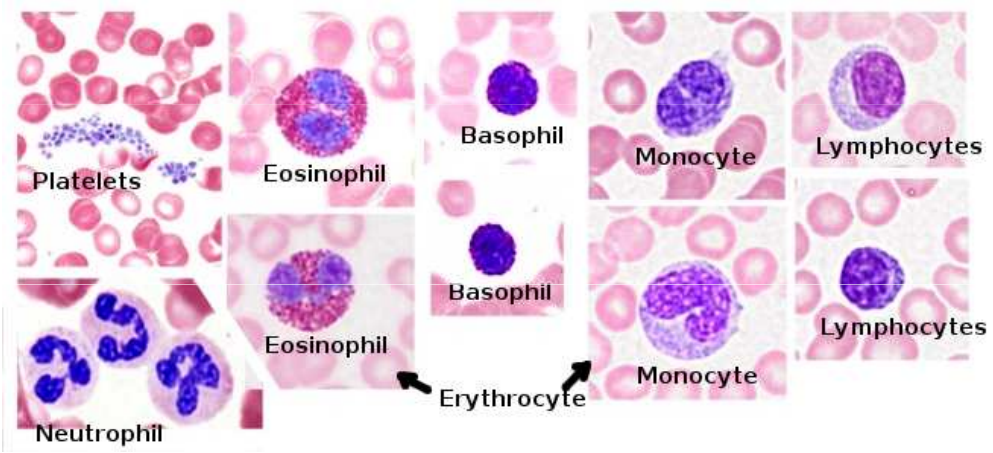
MUNI | RECETOX

Odhad proporce různých typů buněk ve vzorku

(dekonvoluce)

MUNI | RECETOX

Motivace



Cíl:

- Zjistit proporci **různých typů buněk** ve vzorku (např. imunitní buňky, epiteliální buňky, stromální buňky).
- Získat náhled na **heterogenitu** vzorků a **biologické procesy**, které ovlivňují genovou expresi.

Význam:

- Pochopení mikroprostředí tkání (např. nádorových tkání).
- Analýza imunitní odpovědi (např. podíl T-buněk nebo makrofágů).
- Interpretace výsledků genové exprese ve směsných vzorcích (bulk RNA-seq).

Příklad:

- Identifikace nádorových infiltrujících lymfocytů (TILs) z RNA-seq dat.

MUNI | RECETOX

Princip dekonvoluce genové exprese

Hlavní koncept:

Každý typ buňky má unikátní vzor genové exprese (tzv. signatura).
Genová exprese směsného vzorku je kombinací expresí z jednotlivých typů buněk podle jejich proporcí.

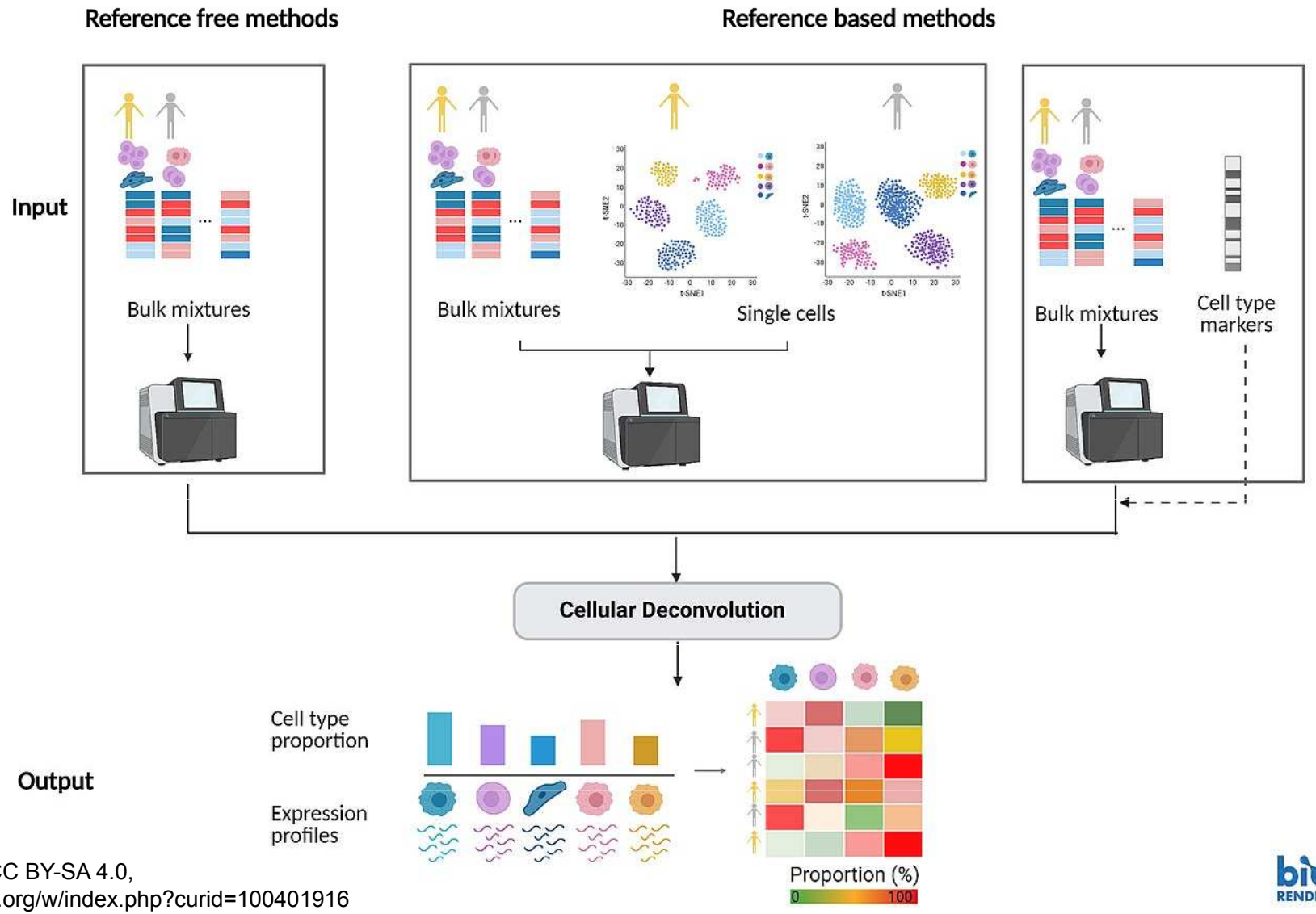
Základní model:

$$E = C \cdot S$$

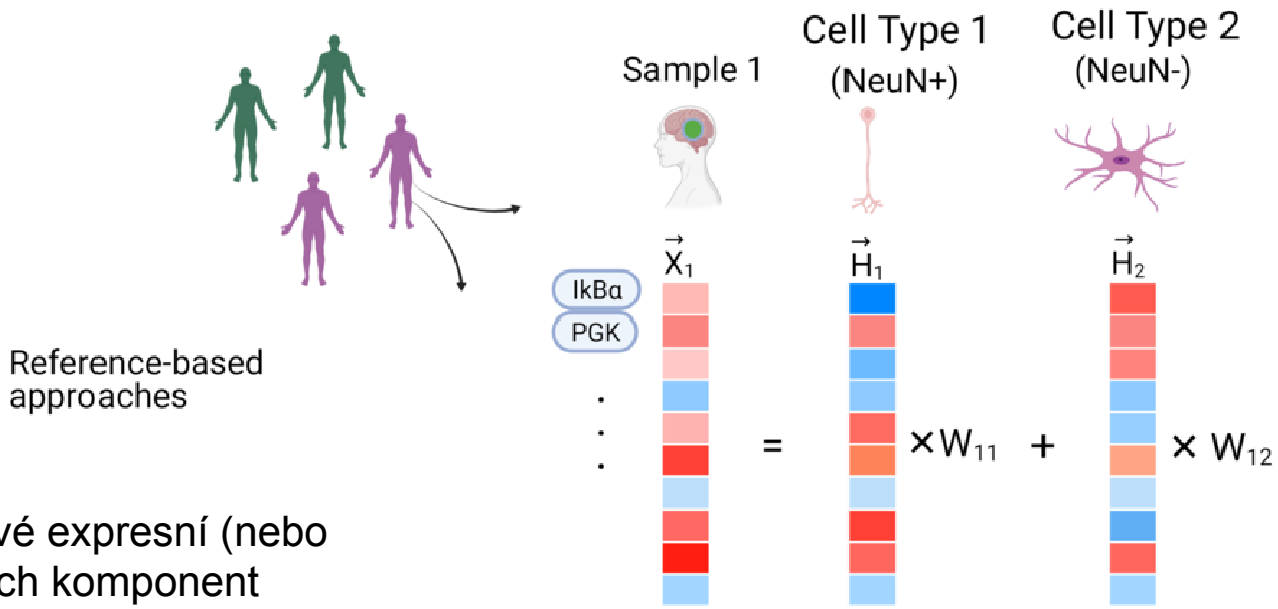
Kde:
E: Matice naměřených expresních dat (bulk RNA-seq).
C: Matice proporcí buněčných typů.
S: Matice signatur genové exprese specifických pro buněčné typy

MUNI | RECETOX

Metody



Metody založené na referenci (supervised)



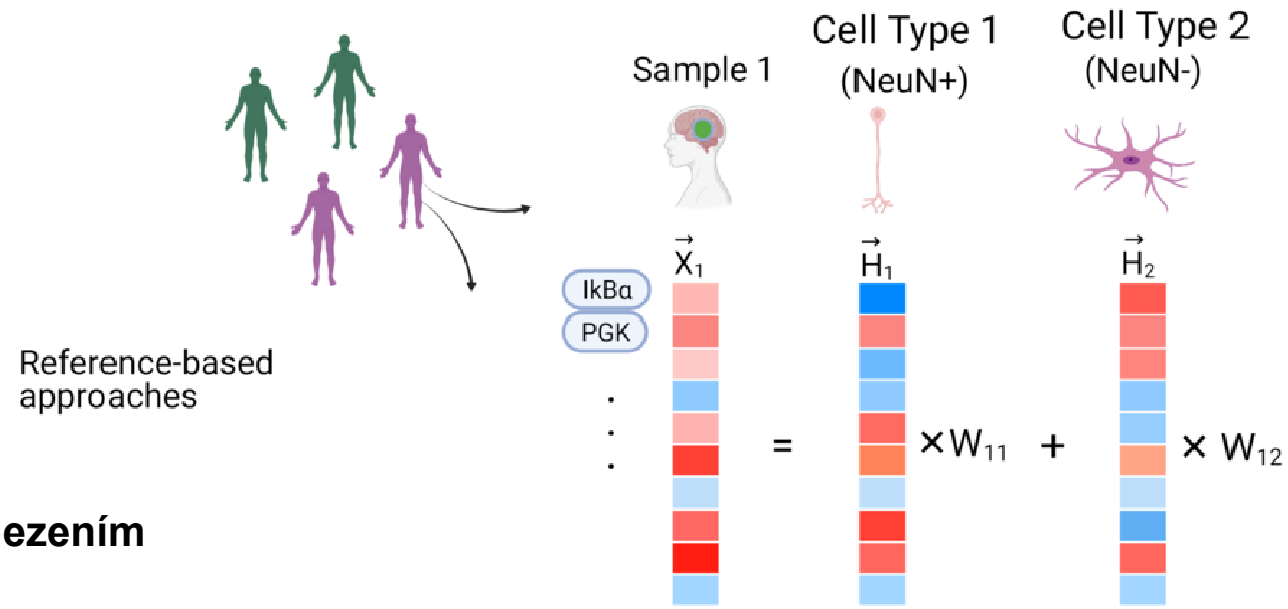
- Vyžadují **předem známé** genové expresní (nebo metylační atd) profily jednotlivých komponent

- Obvykle jsou známé hlavně **markerové geny**

Příklady markerových genů:

- Imunitní buňky: **CD3E** (T-buňky), **CD19** (B-buňky), **CD68** (makrofágy)
- Stromální buňky: **COL1A1** (fibroblasty)
- Epiteliální buňky: **EPCAM**, **KRT18**

Metody založené na referenci (supervised)



- Obvykle metody **regrese** s **omezením**

Nevýhody

- Výsledek závisí na kvalitě markerových genů a znalostní databáze
- Celý signál se rozloží pouze do buněk, které máme ve slovníku

MUNI | RECETOX

Jak vznikají referenční signatury buněk

Referenční signatury jsou **vzorové profily genové exprese (metylace)**, které reprezentují unikátní expresní charakteristiky specifické pro jednotlivé buněčné typy.

Typicky jde o seznam genů a jejich úrovní exprese (metylace)

1. **Izolace jednotlivých buněk:**
   - Fluorescenční třídění buněk (FACS)
   - Mikrodisekce
   - Jednobuněčná RNA-seq (scRNA-seq)
   - Buněčné kultury

2. **Analýza profilů**
   - Jednobuněčná RNA-seq (scRNA-seq)
   - RNAseq, microarray, ….

3. **Derivace markerových genů**

Jak vznikají referenční signatury buněk

Referenční signatury jsou **vzorové profily genové exprese (metylace)**, které reprezentují unikátní expresní charakteristiky specifické pro jednotlivé buněčné typy.

Typicky jde o seznam genů a jejich úrovní exprese (metylace)

1. **Izolace jednotlivých buněk:**
   - Fluorescenční třídění buněk (FACS)
   - Mikrodisekce
   - Jednobuněčná RNA-seq (scRNA-seq)
   - Buněčné kultury

2. **Analýza profilů**
   - Jednobuněčná RNA-seq (scRNA-seq)
   - RNAseq, microarray, ….

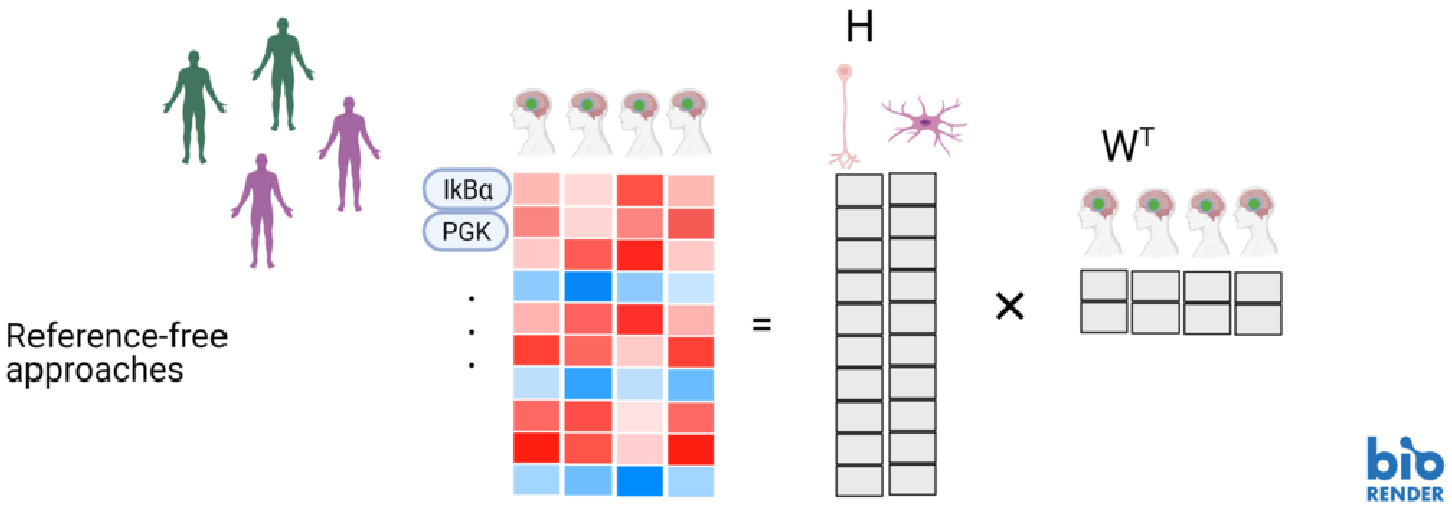
3. **Derivace markerových genů**

4. **Validace**

PROBLÉMY:
- Heterogenita buněk: Buněčné typy nejsou homogenní a mohou mít různé stavy (např. aktivované vs. klidové buňky).
- Specifičnost tkáně: Některé signatury se mohou měnit podle kontextu (např. makrofágy v plicích vs. ve slezině).
- Technické artefakty: Rozdíly mezi platformami (scRNA-seq vs. bulk RNA-seq) mohou ovlivnit přesnost.

MUNI | RECETOX

Metody bez referenčních profilů (unsupervised)



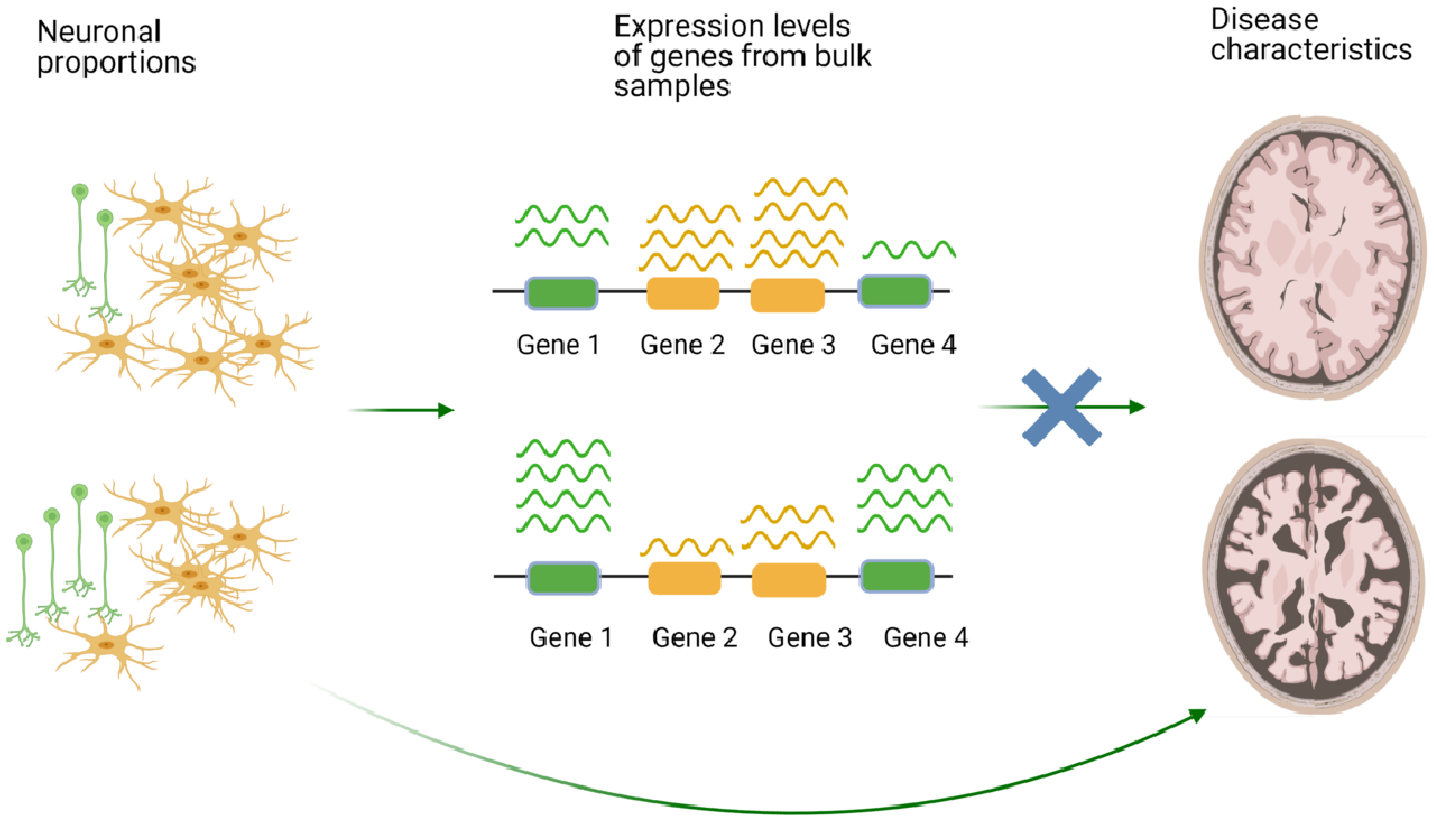
- Nemají referenční signatury odhadují tyto signatury i složení zaráz

- Nejčastější přístupy:
  - **Non-negative matrix factorization**
  - **Bayesovské metody**

MUNI | RECETOX

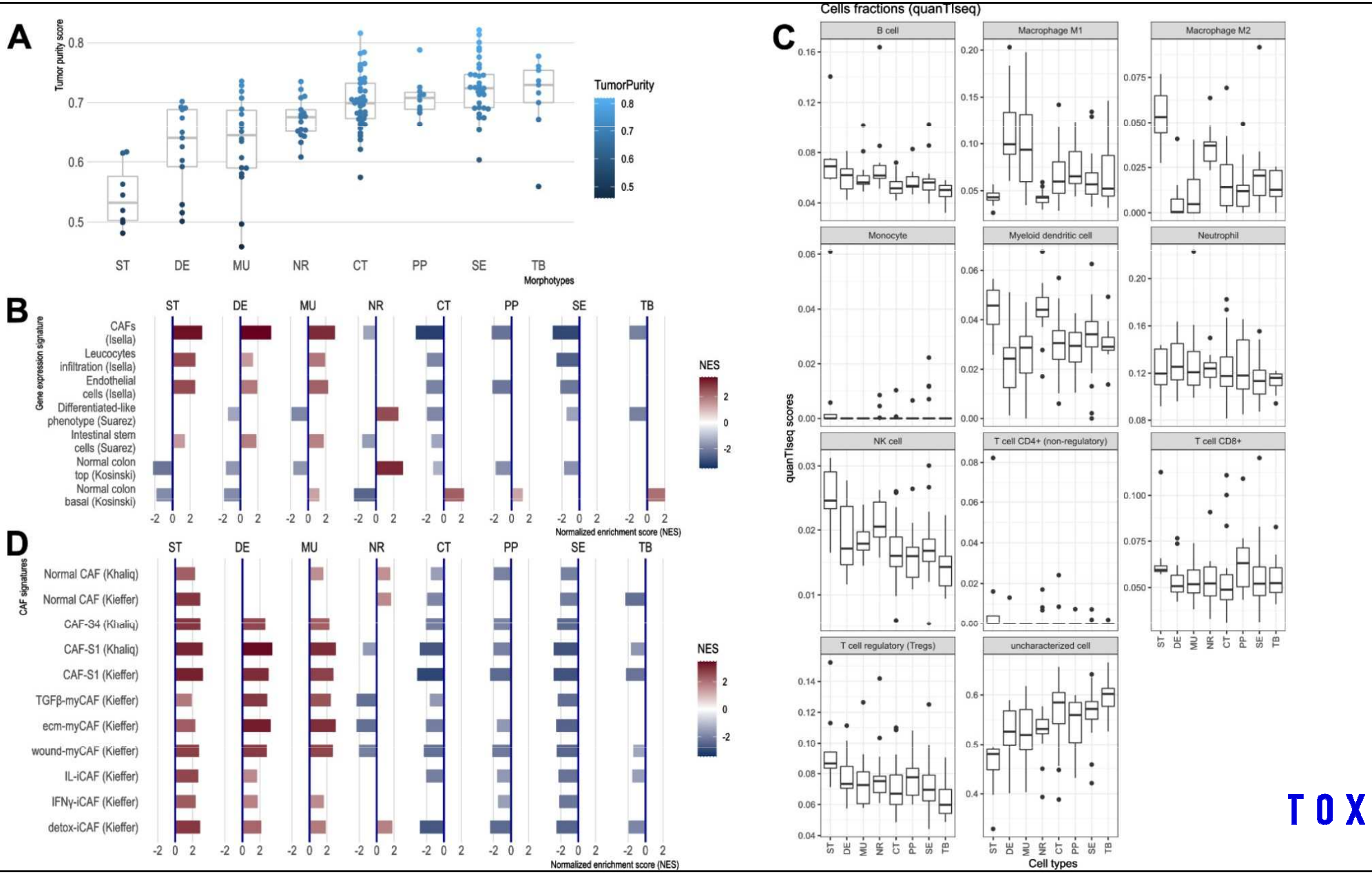
Příklad

Matoucí vliv podílu buněčných typů může vést k chybným asociacím mezi genovou expresí v mozkové kůře a klinickou patologií Alzheimerovy choroby.



Patrick E, Taga M, Ergun A, Ng B, Casazza W, Cimpean M, et al. (August 2020). "Deconvolving the contributions of cell-type heterogeneity on cortical gene expression". *PLOS Computational Biology*. **16** (8): e1008120.

MUNI | RECETOX


A
Tumor purity score
0.8
0.7
0.6
0.5
ST
DE
MU
NR
CT
PP
SE
TB
Morphotypes
TumorPurity
0.8
0.7
0.6
0.5
B
Gene expression signature
CAFs (Isella)
Leucocytes infiltration (Isella)
Endothelial cells (Isella)
Differentiated-like phenotype (Suarez)
Intestinal stem cells (Suarez)
Normal colon top (Kosinski)
Normal colon basal (Kosinski)
-2 0 2
Normalized enrichment score (NES)
NES
2
0
-2
D
CAF signatures
Normal CAF (Khaliq)
Normal CAF (Kieffer)
CAF-S4 (Khaliq)
CAF-S1 (Khaliq)
CAF-S1 (Kieffer)
TGFβ-myCAF (Kieffer)
ecm-myCAF (Kieffer)
wound-myCAF (Kieffer)
IL-iCAF (Kieffer)
IFNγ-iCAF (Kieffer)
detox-iCAF (Kieffer)
Normalized enrichment score (NES)
NES
C
Cells fractions (quanTIseq)
B cell
Macrophage M1
Macrophage M2
Monocyte
Myeloid dendritic cell
Neutrophil
NK cell
T cell CD4+ (non-regulatory)
T cell CD8+
T cell regulatory (Tregs)
uncharacterized cell
quanTIseq scores
Cell types

TOX

Nejčastější metody

Zkratka metody	Název metody	Typ metody	Vstupní data	Rok publikace
CIBERSORT[23]	Robust enumeration of cell subsets from tissue expression profiles	Reference based	Gene expression	2018
CDSeq[24]	A complete deconvolution method for dissecting tissue heterogeneity	Reference free	Gene expression	2019
FARDEEP[25]	Fast and robust deconvolution of expression profiles	Reference based	Gene expression	2019
UNDO[26]	Unsupervised deconvolution of tumor-stromal mixed expressions	Reference free	Gene expression	2015
dtangle[27]	Accurate and robust cell type deconvolution	Reference based	Gene expression	2019
EPIC[28]	Estimating the proportions of different cell types from bulk gene expression data	Reference based	Gene expression	2017
BSEQ-sc[29]	Deconvolution of bulk sequencing experiments using single cell data	Reference based	Gene expression	2016
MuSiC[18]	Cell-type Identification by estimating relative subsets of RNA transcripts	Reference based	Gene expression	2019
SCDC[30]	Bulk gene expression deconvolution by multiple single-Cell RNA sequencing references	Reference based	Gene expression	2020
DWLS[31]	Gene expression deconvolution using dampened weighted least squares	Reference based	Gene expression	2019
deconvSeq[32]	Deconvolution of cell mixture distribution in sequencing data	Reference based	Gene expression	2019
Bisque[19]	Decomposition of bulk expression with single-cell sequencing	Reference based	Gene expression	2020
TOAST[33]	Tools for the analysis of heterogeneous tissues	Reference free	DNA methylation	2019
Houseman[9]	Reference-free deconvolution of DNA methylation data and mediation by cell composition effects	Reference based	DNA methylation	2016
methylCC[34]	Technology-independent estimation of cell type composition using differentially methylated regions	Reference based	DNA methylation	2019
BayesCCE[35]	Bayesian framework for estimating cell-type composition from DNA methylation	Reference free	DNA methylation	2018

Srovnání metod

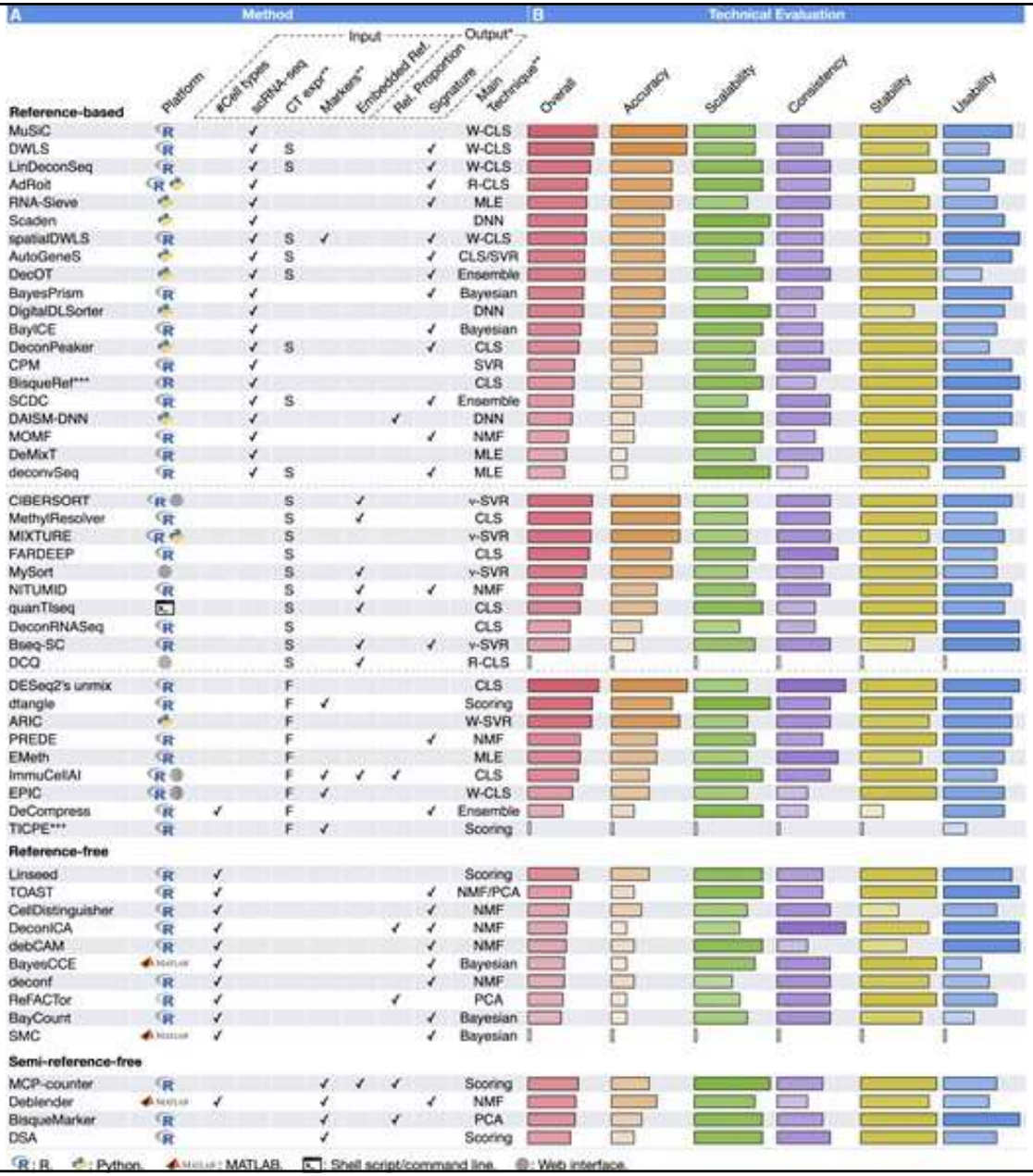


Figure 3. Key characteristics and technical evaluation of cellular deconvolution methods. (A) Method characterization according to implementation, input, output, embedded reference and the underlying algorithm.
(B) Performance assessment based on five criteria:
the accuracy of the predicted cell type proportions,
the scalability in analyzing large input sizes,
the stability (opposite of crash rate and other errors),
the consistency of the predicted cell type proportions using different initializations
usability as code quality and ease of use.

*Abbreviations: S: signature matrix; F: full cell-type expression matrix; PCA: principal component analysis; NMF: non-negative matrix factorization; CLS: constrained least squares; SVR: support vector regression; MLE: maximum likelihood estimation; DNN: deep neural network; ensemble: combination of multiple methods; scoring: enrichment using marker sets. W prefix: weighted. R prefix: regularized. ***BisqueRef requires scRNA data of at least two subjects as input. TICPE requires cancer cell expression, normal cell expression, immune cell expression and marker gene sets as input.

Nucleic Acids Res, Volume 52, Issue 9, 22 May 2024, Pages 4761–4783, https://doi.org/10.1093/nar/gkae267

Další čtení

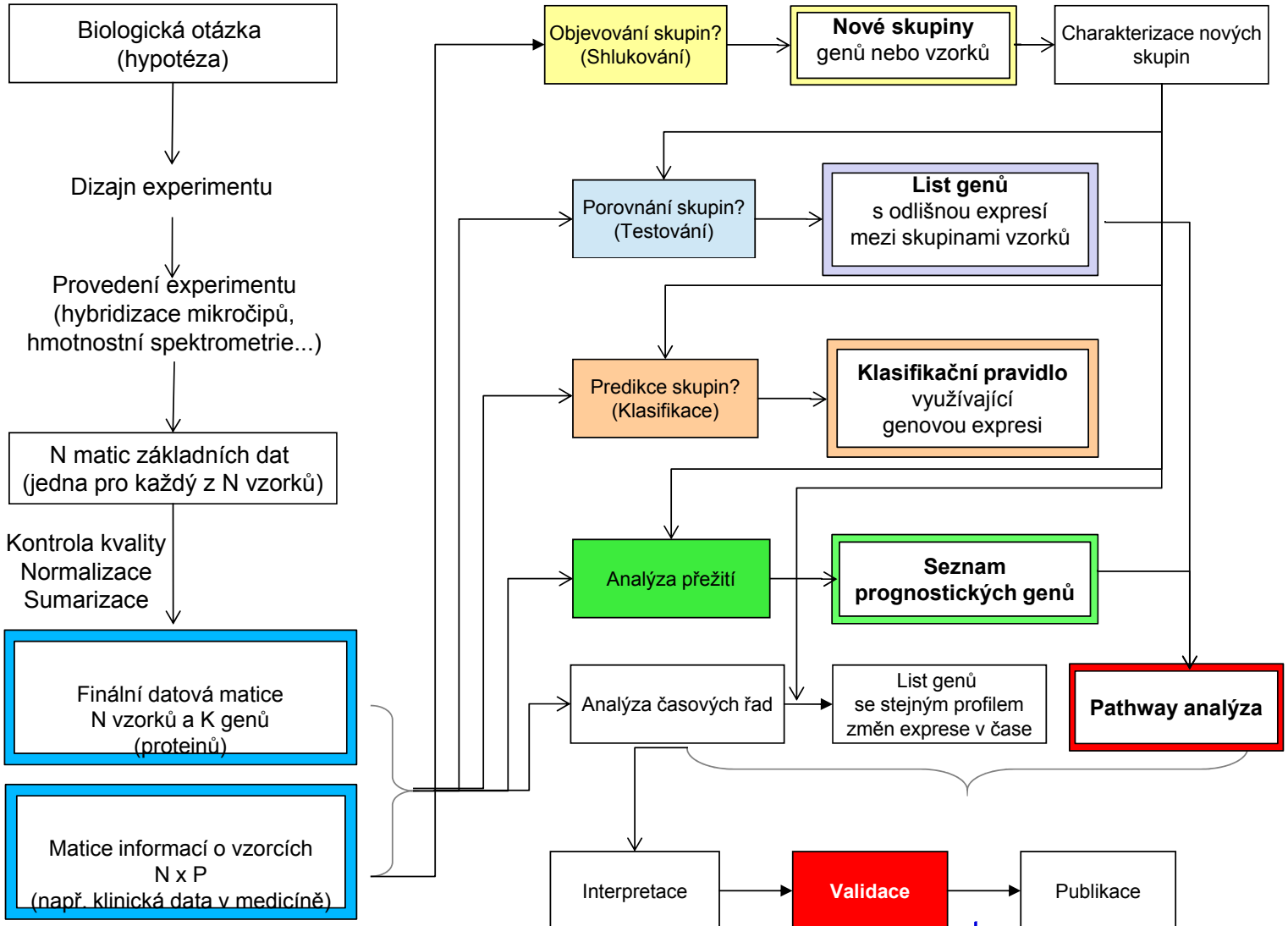
- Fourteen years of cellular deconvolution: methodology, applications, technical evaluation and outstanding challenges | Nucleic Acids Research | Oxford Academic
- Comprehensive evaluation of deconvolution methods for human brain gene expression | Nature Communications

Analýza genových sad

(pathway analýza)

MUNI | RECETOX

Jak se hledá potenciální biomarker v omics datech



MUNI | RECETOX

Motivace

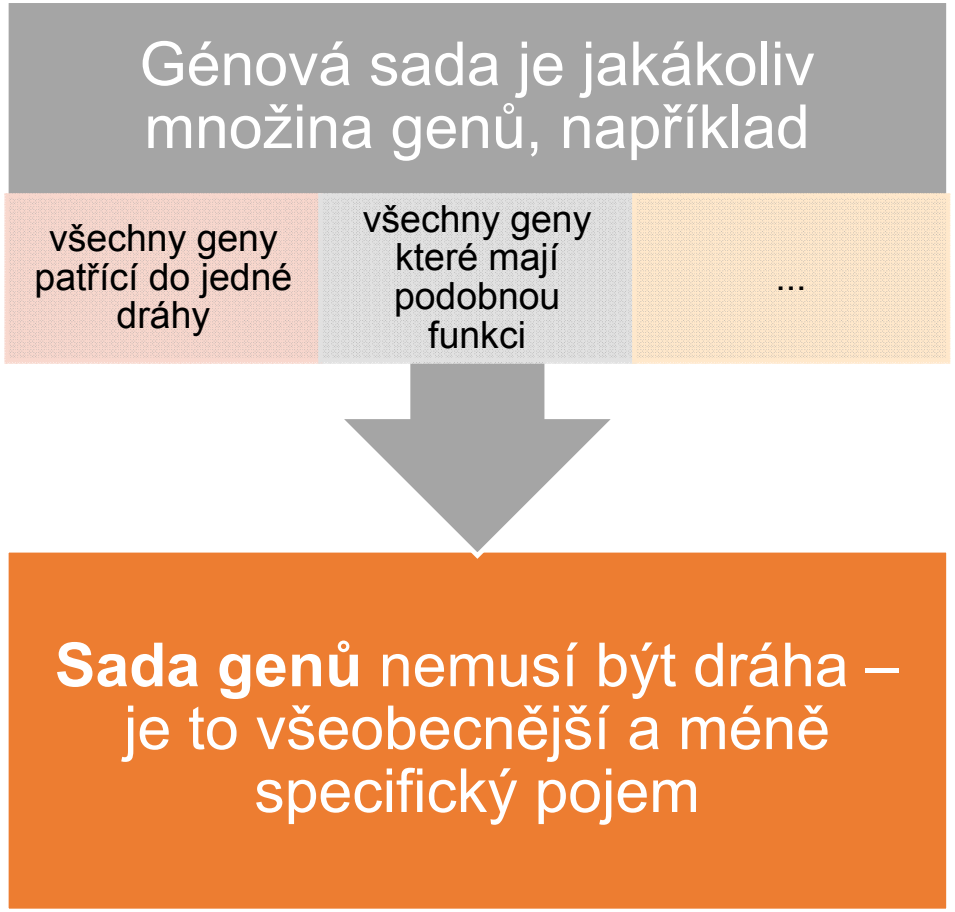
- Geny, proteiny a další molekuly jsou navzájem propojené ve velké spleti různých signálních, metabolických a různych jiných drah
- Potřebujeme zjistit, jaké dráhy jsou zasažené naším experimentálním protokolem (liší se v mezi skupinami)

Jak na to?

- Seznam molekul můžeme **ad-hoc vložit** do existující databáze drah a podívat se kam patří (KEGG, MsigDB....)
    - nevýhoda – nemáme statistickou významnost
- **Provedeme analýzu genových sad** (pathway analýzu)

- **Předpoklad všech těchto analýz:** operují s již definovanými skupinami genů

Genová sada vs dráha



Cíl analýzy genových sad

- Cíl je přiřadit každé genové sadě, případně dráze jedno číslo - skóre, a nebo p-hodnotu, abychom mohli odpovědět na otázku:

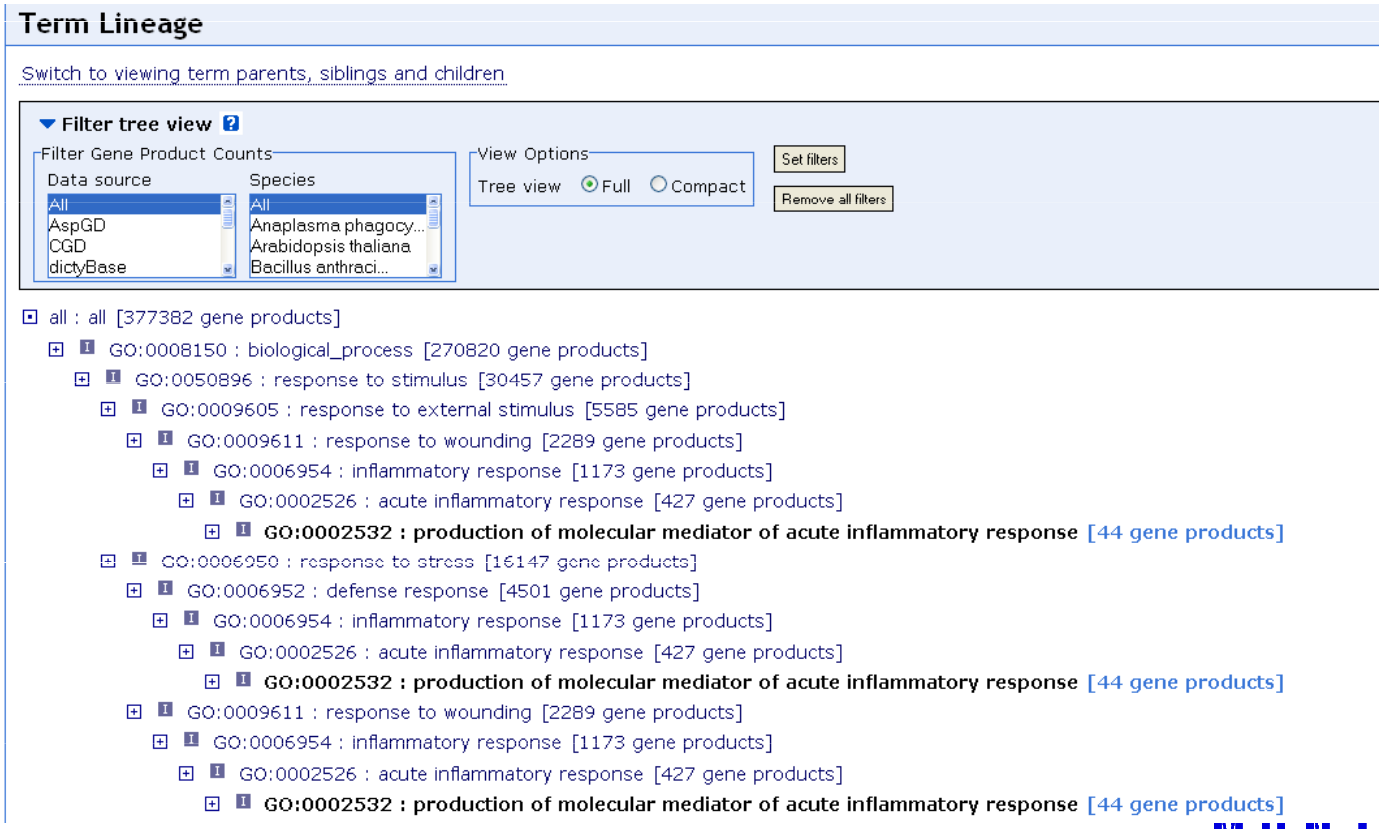
  Kolik genů je v sadě(pathway) odlišně exprimovaných a je to dostatečně statisticky významné, abychom mohli říct, že je tato dráha specifická jen pro naše porovnávané skupiny?

Databáze genových sad (pathways)

Gene Ontology (GO) databáze

- http://www.geneontology.org/
- Hierarchická databáze
- Rodičovské uzly: obecnější termíny
- Potomci uzlů: víc specifické
- Na konci hierarchie jsou molekuly (geny/proteiny)
- Na vrcholu jsou 3 rodičovské uzly:
  - Biologické procesy
  - Molekulární funkce
  - Buněčné složky

GO databáze

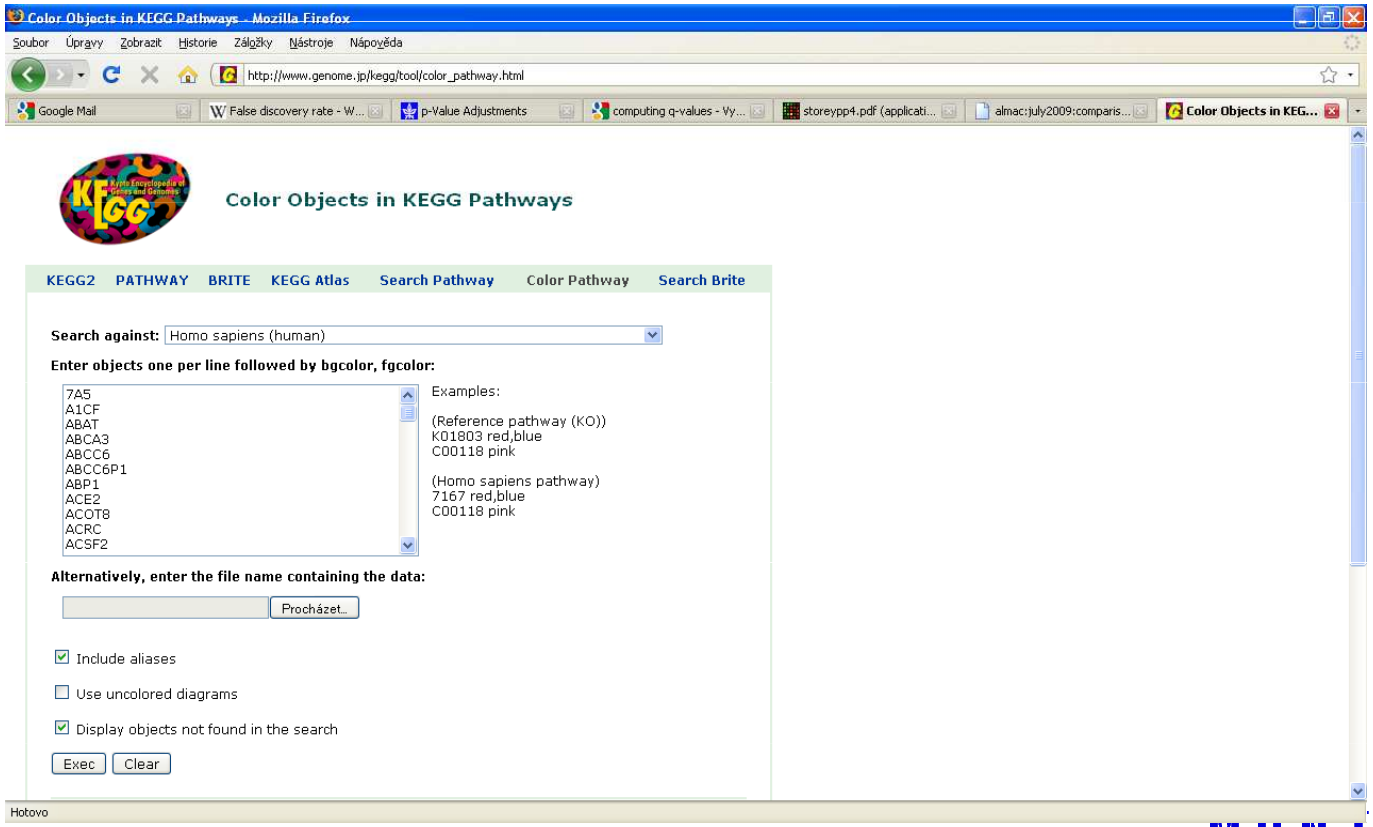


MUNI | RECETOX

KEGG pathway databáze

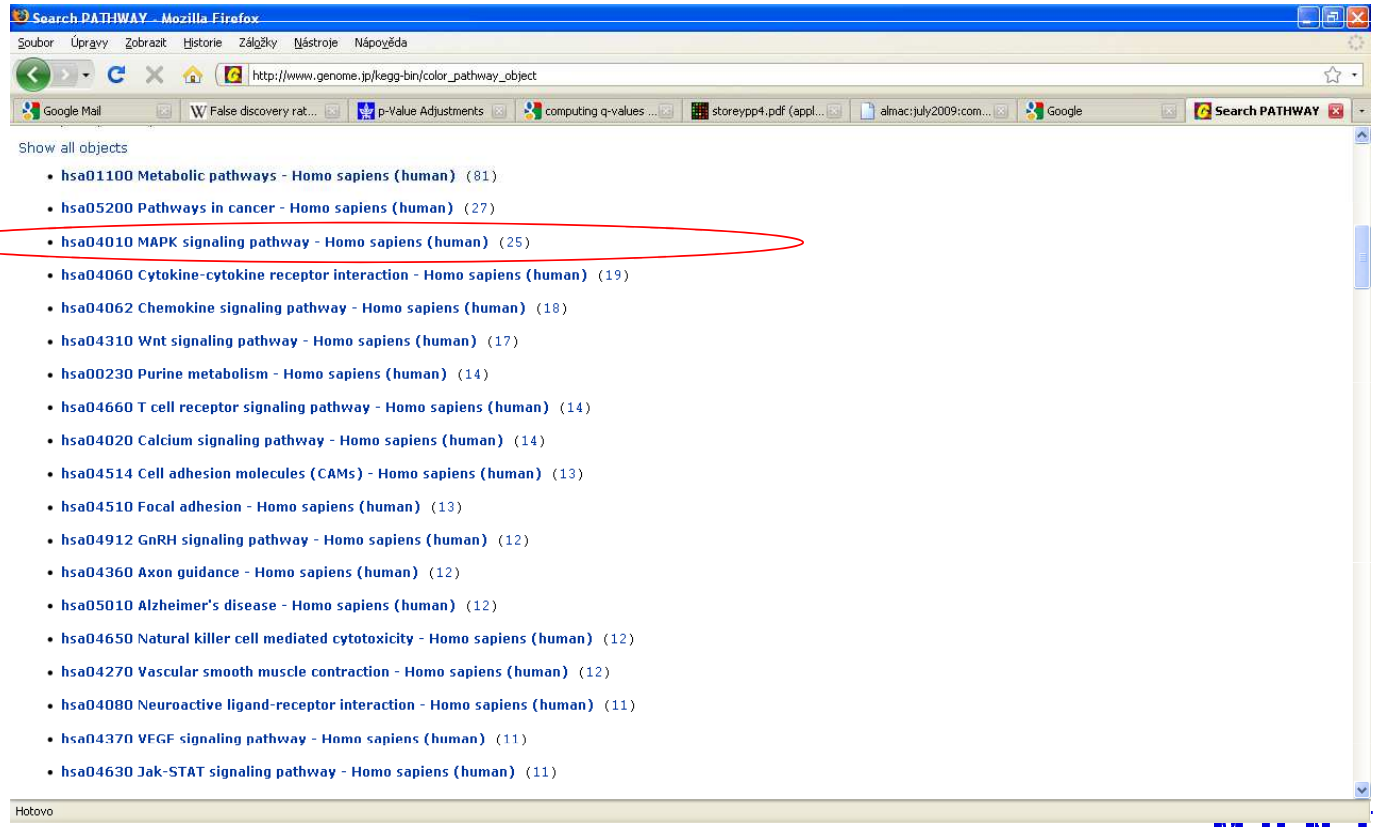
- KEGG = Kyoto Encyclopedia of Genes and Genomes
- http://www.genome.jp/kegg/pathway.html
- Více informací než GO, máme tu již vztahy mezi geny a genovými produkty
- Detailní informáce jen pro některé organizmy a procesy
- Využívá hlavně ověřené poznatky, nemůže ji kdokoliv změnit
- Proto se tu nenachází všechny geny (obvykle tak třetina až polovina z hledaných)
- Aktualizovaná databáze není volně přístupná

KEGG

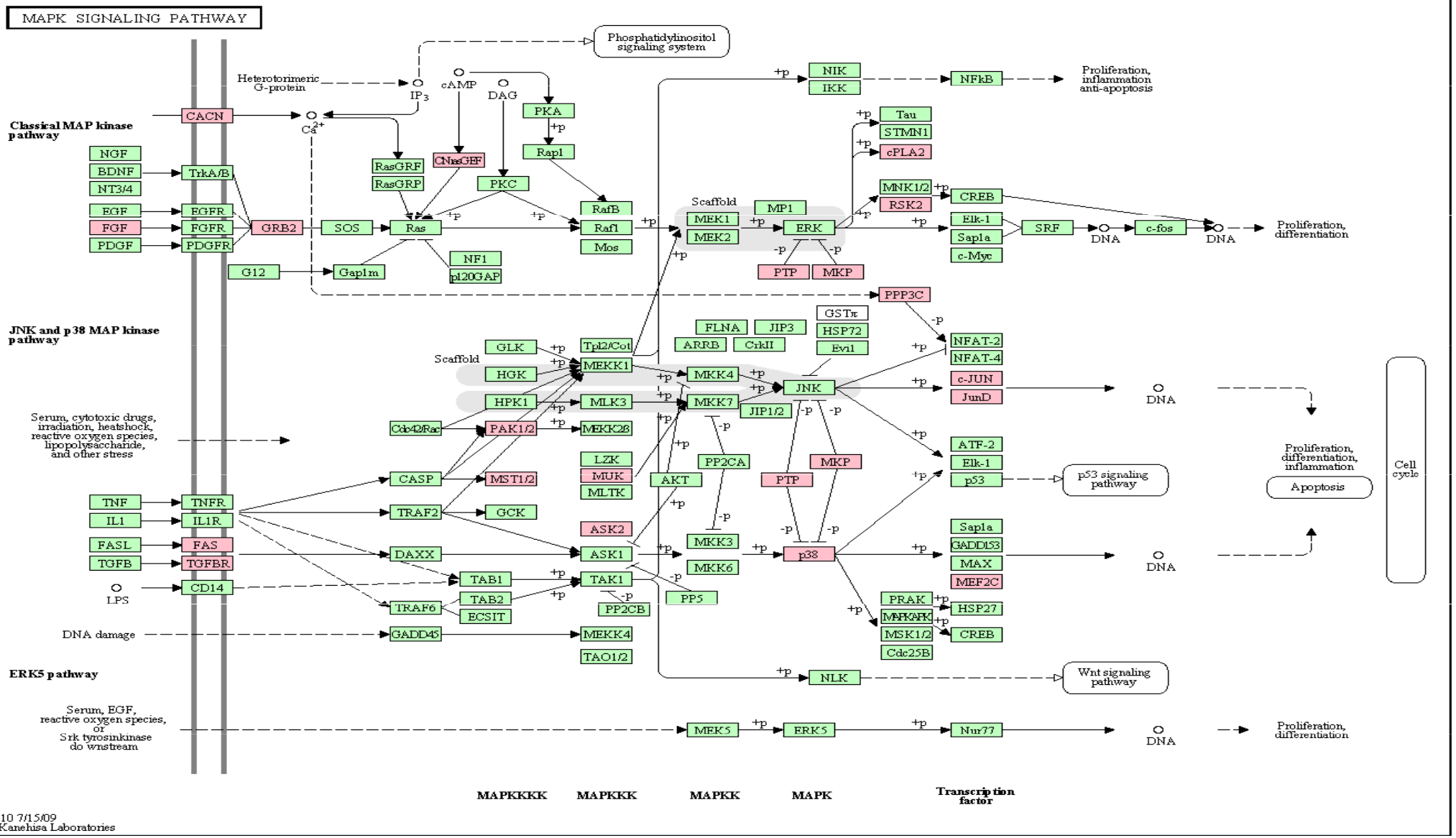


MUNI | RECETOX

KEGG



MUNI | RECETOX


MAPK SIGNALING PATHWAY
Phosphatidylinositol signaling system
Heterotorimeric G-protein
IP3
cAMP
DAG
PKA
Rap1
Classical MAP kinase pathway
CACN
Ca2+
NGF
BDNF
NT3/4
TrkA/B
EGF
FGF
PDGF
EGFR
FGFR
PDGFR
GRB2
SOS
Ras
RasGRF
RasGRP
CNasGEF
PKC
RafB
Raf1
Mos
G12
Gap1m
NF1
p120GAP
Scaffold
MP1
MEK1
MEK2
ERK
PTP
MKP
NIK
IKK
NFkB
Proliferation, inflammation anti-apoptosis
Tau
STMN1
cPLA2
MNK1/2
RSK2
CREB
Elk-1
Sap1a
c-Myc
SRF
c-fos
DNA
Proliferation, differentiation
PPP3C
JNK and p38 MAP kinase pathway
FLNA
JIP3
ARRB
CrkII
GSTπ
HSP72
Evi1
NFAT-2
NFAT-4
GLK
HGK
HPK1
Tpl2/Cot
MEKK1
MLK3
MEKK2/3
Cdc42/Rac
PAK1/2
MKK4
MKK7
JIP1/2
JNK
c-JUN
JunD
Serum, cytotoxic drugs, irradiation, heatshock, reactive oxygen species, lipopolysaccharide, and other stress
CASP
MST1/2
LZK
MUK
MLTK
PP2CA
AKT
PTP
MKP
ATF-2
Elk-1
p53
p53 signaling pathway
Proliferation, differentiation, inflammation
Apoptosis
Cell cycle
TNF
IL1
FASL
TGFB
TNFR
IL1R
FAS
TGFBR
CD14
LPS
TRAF2
GCK
DAXX
TAB1
TAB2
TRAF6
ECSIT
ASK2
ASK1
TAK1
PP2CB
MKK3
MKK6
PP5
p38
Sap1a
GADD153
MAX
MEF2C
PRAK
MAPKAPK
MSK1/2
Cdc25B
HSP27
CREB
DNA damage
GADD45
MEKK4
TAO1/2
ERK5 pathway
NLK
Wnt signaling pathway
Serum, EGF, reactive oxygen species, or Src tyrosinkinase do wnstream
MEK5
ERK5
Nur77
MAPKKKK
MAPKKK
MAPKK
MAPK
Transcription factor
04010 7/15/09
(c) Kanehisa Laboratories

KEGG pathway databáze

Poklikání na jednotlivé uzly zobrazí víc informací o jednotlivých genech:

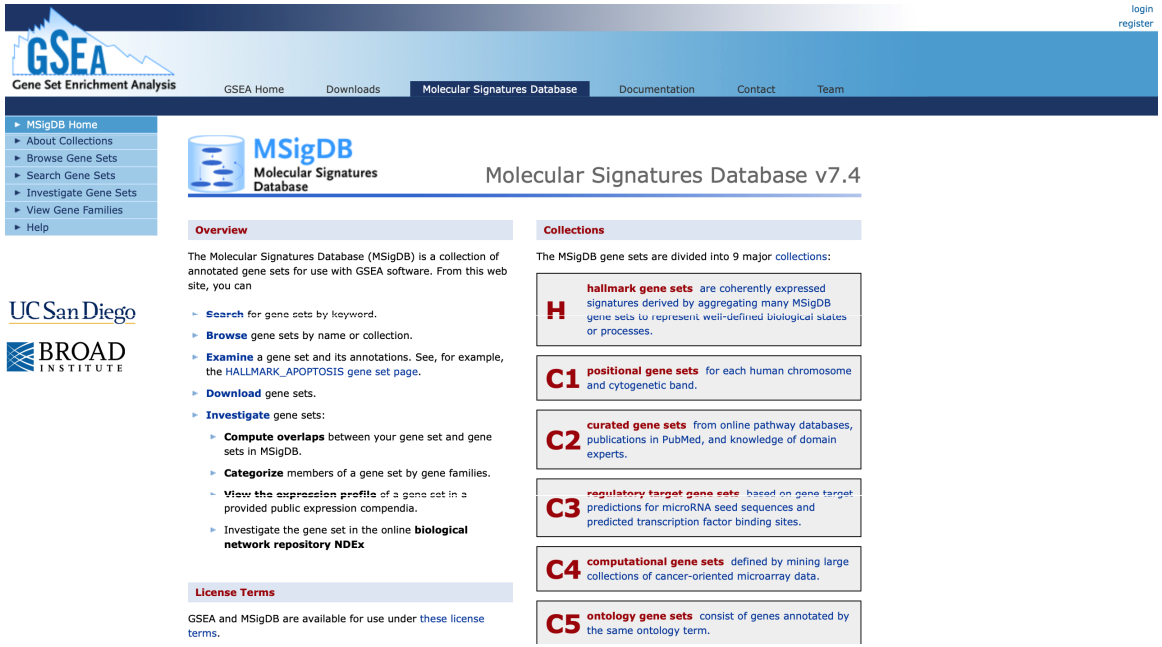
Všechny ostatní dráhy do kterých patří gen	Identifikátory daného genu v různých jiných databázích	Odkaz na literaturu z které byly informace čerpané, případně další důležité články	Informaci o sekvenci

Je možné zabarvit jednotlivé geny podle rozdílných barev

MUNI | RECETOX

MsigDB databáze

- https://www.gsea-msigdb.org/gsea/msigdb



MUNI | RECETOX

Metody analýzy genových sad

MUNI | RECETOX

Rozdělení metod

Podle toho s jakou informací pracují na

- *metody dělící hranice* – berou do úvahy jen informaci "významný" vs. "nevýznamný" gen
- *metody celého seznamu genů* – pracují přímo se všemi *p*-hodnotami (i nevýznamnými!) a teda s pořadím

Podle skupiny molekul které analyzují na:

- *uzavřené* – analýza jen v rámci genů v sadě
- *kompetitivní* – porovnání se všemi geny experimentu

Nové metody pracují i s topologií dráhy

MUNI | RECETOX

Dělení metod dle skupiny molekul které analyzují

Uzavřené vs. kompetitivní I.

Uzavřená metoda používá jen hodnoty genů z dané množiny:

- H_0 : “Žádné geny z genové množiny nejsou odlišně exprimované”

Kompetitivní test porovnává geny v genové množině s ostatními geny v experimentu

- H_0 : “Geny v genové množině nejsou víc odlišně exprimované než ostatní geny v experimentu”

MUNI | RECETOX

Příklad

Datový soubor 12 639 genů.
Z nich $p<0.05$ má 1272 genů

96 genů v genové sadě, z toho 8 má p-hodnoty < 5%

Kolik odlišně exprimovaných genů očekáváme náhodně?

Datový soubor 12 639 genů.
Z nich p<0.05 má 1272 genů

96 genů v genové sadě, z toho 8 má p-hodnoty < 5%

Kolik odlišně exprimovaných genů očekáváme náhodně?

Náhodně očekáváme 96 x 5% = 4.8 významných genů

Jaká je pravděpodobnost pozorování **8** a více významných genů?

Vhodné testy:

Binomický test (p = 0.1079)

Příklad, uzavřená metoda dělící hranice

Hypergeometrický nebo binomický test?

Kritérium	Binomický test	Hypergeometrický test
Reálné použití	Opakované pokusy s neměnnou šancí na výsledek	Biologická data s konečným počtem vzorků
Příklad aplikace	Určení podílu konkrétních výsledků v pokusech	Výběr genů bez vrácení, analýza nadměrného výskytu
Velikost populace	Teoreticky neomezená populace	**Konečná** populace
Typ vzorkování	S náhradou	**Bez náhrady**
Zohlednění závislosti vzorků	Vzorky jsou nezávislé	Zohledňuje vzájemnou závislost vzorků
Typický kontext použití	Opakované nezávislé pokusy (např. házení mincí)	**Obohacení genových sad** (gene set enrichment analysis)
Pravděpodobnost	Zůstává konstantní	**Mění se s každým výběrem**

Příklad, uzavřená metoda dělící hranice

Hypergeometrický test... pravděpodobnost výběru každého dalšího genu se mění s každým dalším výběrem

	V GS	Není v GS
Význ	8	1264
Nevýzn	88	11279

```
x <- 8  # Počet úspěchů ve vzorku (geny s p < 0,05 v sadě)
m <- 1272 # Celkový počet úspěchů v populaci (všechny geny s p < 0,05)
n <- 11367 # Celkový počet neúspěchů v populaci (všechny ostatní geny)
k <- 96 # Velikost vzorku (genová sada)
```

Výpočet hypergeometrické pravděpodobnosti

```
p_value <- phyper(q = x - 1, m = m, n = n, k = k, lower.tail = FALSE)
```

p_value **0.7627**

Příklad, kompetitivní metoda dělící hranice

Hypergeometrický nebo Fisherův test?

	V GS	Není v GS
Význ	8	1264
Nevýzn	88	11279

p = 0.7627 (Fisherův test – jednostranný)

- 1272 z 12639 genů je odlišně exprimovaných v tomto datovém souboru (to je zhruba 10%)
- V množině náhodně vybraných 96 genů očekáváme tedy 96 x 10% = 9.6 významných genů
- p-hodnotu vypočítáme z kontingenční tabulky pomocí Fisherova nebo Chi-kvadrát testu

Příklad, kompetitivní metoda dělící hranice

Dělení metod podle toho s jakou informací pracují

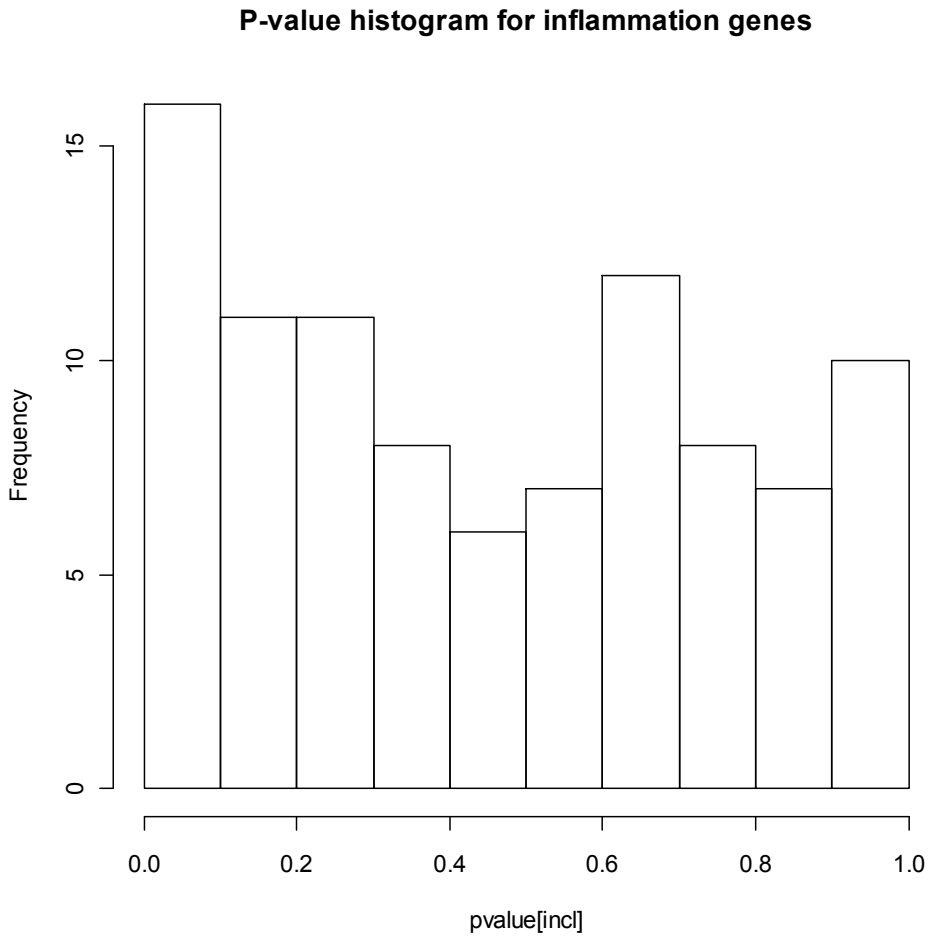
Metody dělící hranice vs. metody celého seznamu

- Dvě předchozí metody byly závislé na dělících hranicích – cut-offs a tedy závislé na N
- V případě, že řekneme, že gen je pro nás významný již na 10% FDR, výsledek se změní!
- Dále ztrácíme informaci tím, že redukujeme p-hodnotu na binární proměnné (významné/nevýznamné)
- Je rozdíl vědět jestli statisticky nevýznamné geny v naší množině jsou významné na hranici významnosti a nebo vůbec ne

MUNI | RECETOX

Metoda celého seznamu genů: *uzavřená*

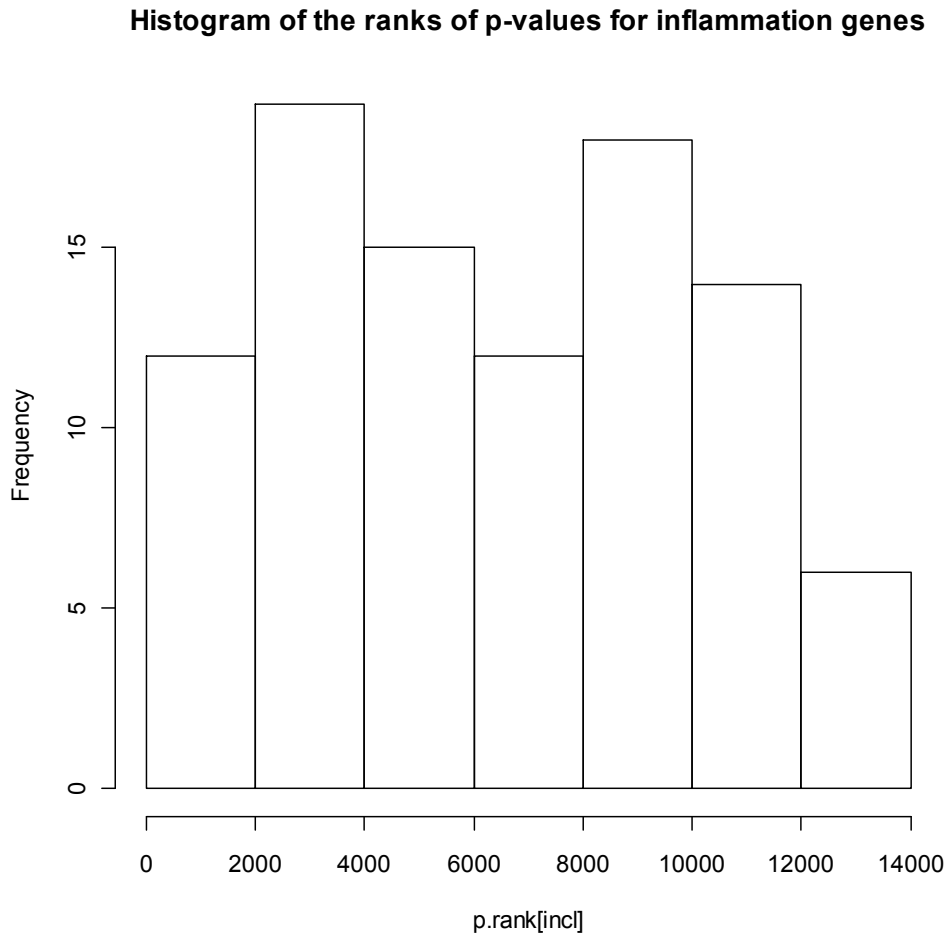
- Můžeme studovat **rozložení p-hodnot v genové sadě**
- V případě, že žádné geny nejsou odlišně exprimované, mělo by se jednat o uniformní rozložení
- Pík vlevo indikuje významnost některých genů
- Aplikujeme Kolmogorův-Smirnovův test pro porovnání rozložení
- p = 8.2%, není velmi významné
- Je to **uzavřená** metoda, protože používáme jen geny z genové sady



MUNI | RECETOX

Metoda celého seznamu genů: *kompetitivní*

- Alternativně se můžeme dívat na rozložení **pořadí** p-hodnot
- Toto by byla kompetitivní metoda, protože porovnáváme naši genovou sadu s ostatními geny v experimentu
- Opět můžeme aplikovat KS test
- p=85.1%, velmi nevýznamné



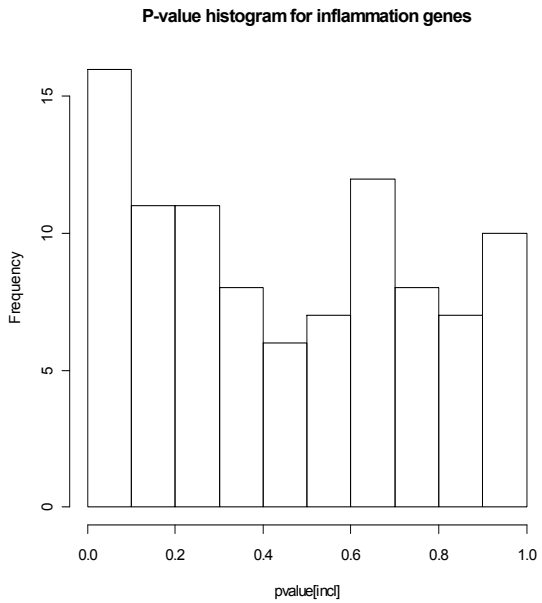
MUNI | RECETOX

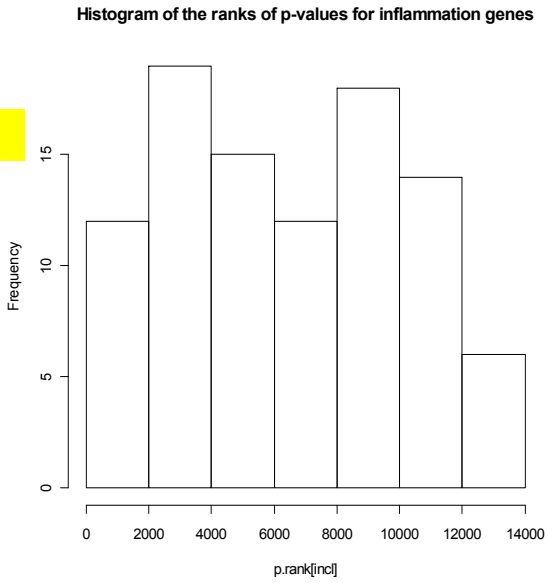
1272 z 12639 genů je odlišně exprimovaných
z toho 8 v genové sadě o 96 genech

Metoda celého seznamu genů: *uzavřená*

	p-hodnota	pořadí
Gen A	0.001	1
Gen H	0.001	2
Gen Z	0.031	3
Gen G	0.024	4
.	.	.
Gen M	0.024	62
.	.	.
Gen O	0.049	1272
Gen J	0.351	5843
.		
Gen L	0.454	7390
.		
Gen B	0.752	10287
.		
.		
Gen C	0.989	12639

Metoda celého seznamu genů: *kompetitivní*



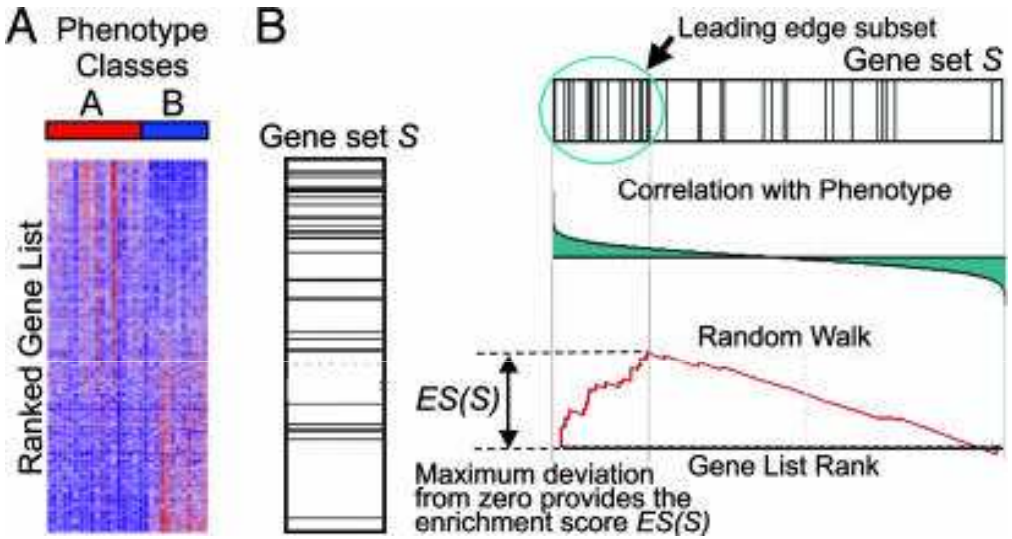


ozložení p-hodnot v genové sadě

MUNI | RECETOX

GSEA

- Najznámější je GSEA – gene set enrichment analysis (analýza obohacení genové sady)
- Počítá se na seřazených p-hodnotách a sleduje se, zda jsou geny z genové sady náhodně rozložené v tomto seřazeném listě, a nebo se vyskytují v horních, významných pozicích
- Postup: 1. Výpočet skóre obohacení (ES)
    - 2. Odhad významnosti ES (p-hodnota) na základě permutačního testu
    - 3. Upravení p-hodnot na problém mnohonásobného porovnávání



MUNI | RECETOX

Uzavřené vs. kompetitivní II.

- Výsledky kompetitivních testů závisí na počtu testovaných genů (např. genů na microarray sklíčku a předcházejícím filtrování)
  - Na malém mikročipovém sklíčku, kde jsou změněné všechny geny, kompetitivní metoda nenajde žádné odlišně exprimované množiny genů.
- Kompetitivní metody dávají méně významných výsledků než metody uzavřené

Další aspekty

Směr změny

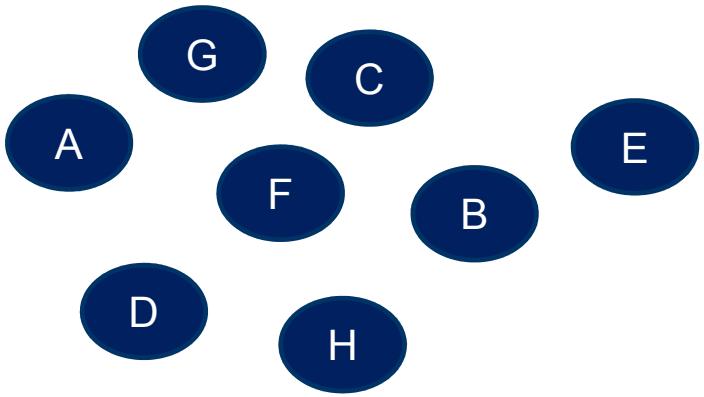
- Pokud chceme zjistit **směr** změny, musíme zopakovat analýzu pro jednostranný test
  - jen up-regulované
  - jen down-regulované

Mnohonásobné testování

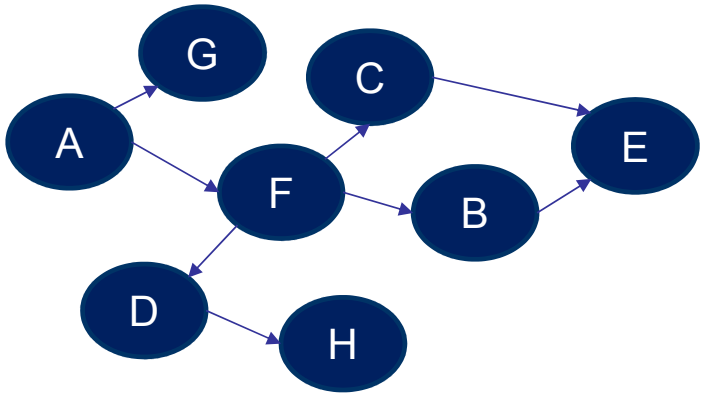
- Stejně jako u testování hypotéz na genech mezi skupinami, i pokud máme velký počet genových sad!
- FDR je trochu komplikované, protože genové množiny se překrývají
- Bonferroniho korekce vždy funguje

Topologie

Bez topologie



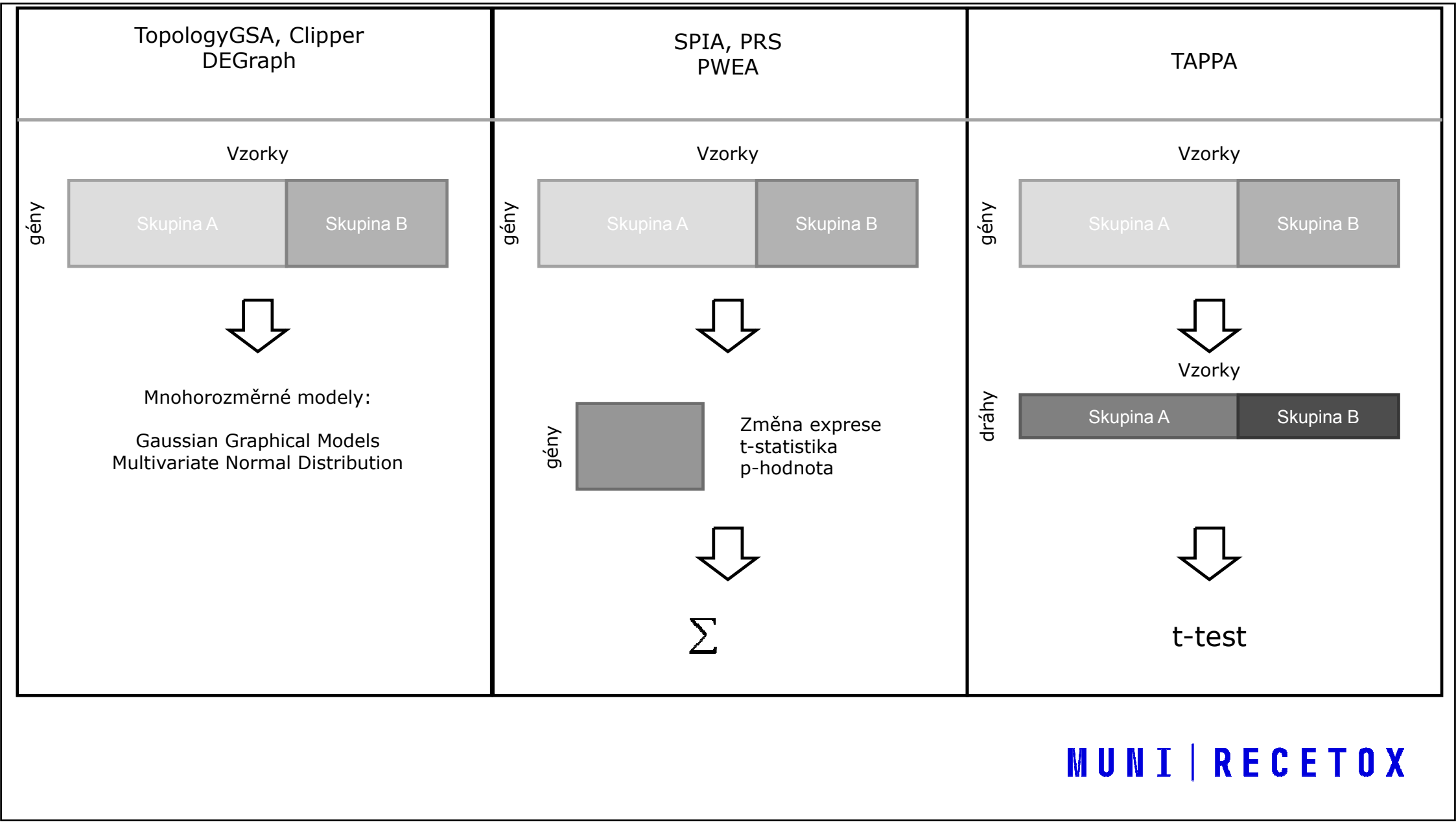
S topologií



MUNI | RECETOX

Topologie využívaná různě

- Cíl:
  - změna průměrné exprese, korelace, topologie
- Jednotka zájmu:
  - dráha, modul, cesta, geny
- Topologie známá dopředu a nebo odhadovaná z dat
- Celková síť a nebo individuální dráhy



MUNI | RECETOX

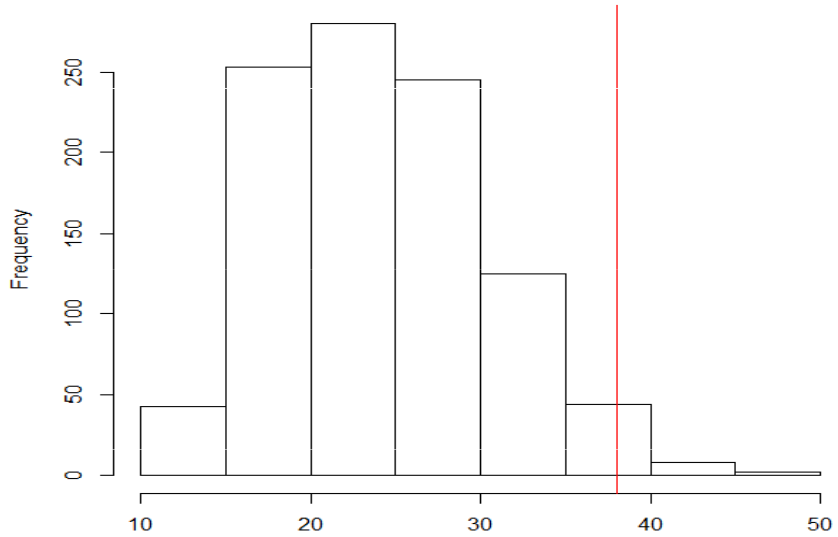
Příklad – topologie uzavřená metoda dělící hranice

- 96 genů v dráze, z toho 8 má p-hodnoty < 5%
- Je exprese dráhy změněná?

- Využití topologické informace:
  - Definujeme statistiku
  - $s = \sum_{i=1}^{n} w_i d_i$
  - n – počet genů v dráze
  - i – index pro gen
  - w_i – počet interakcí genu *i*
  - d_i – 1 – pokud je gen *i* odlišně exprimovaný, 0 jinak

Příklad – topologie uzavřená metoda dělící hranice

- Z 8 odlišně exprimovaných genů:
  - 2 interagují s 10 geny v dráze
  - 3 interagují s 5 geny v dráze
  - 3 interagují s jedním genem v dráze
- s = 2*10 + 3*5 + 3*1 = 38
- Opakovaně v dráze náhodně vybíráme 8 genů a získáme rozdělení statistik, které porovnáme s první statistikou.



- $p = \sum_{i=1}^{N} (s_{náhodne} \geq s_{pozorované})/N$
- N=počet náhodných výberov
- p=0.028, významné

- Z 8 odlišně exprimovaných genů:
  - 2 interagují s 10 geny v dráze
  - 3 interagují s 5 geny v dráze
  - 3 interagují s jedním genem v dráze
- s = 2*10 + 3*5 + 3*1 = 38
- Opakovaně v dráze náhodně vybíráme 8 genů a získáme rozdělení statistik, které porovnáme s první statistikou.

MUNI | RECETOX

PLOS ONE | PUBLISH ABOUT BROWSE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

A critical comparison of topology-based pathway analysis methods

Ivana Ihnatova, Vlad Popovici, Eva Budinska

Published: January 25, 2018 • https://doi.org/10.1371/journal.pone.0191154

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0191154

MUNI | RECETOX

Pozor na korelace mezi geny!

- Všechny testy, které jsme probírali předpokládají, že geny uvnitř skupin jsou nezávislé
    - To je ale velmi nepravděpodobné!
- Pokud jsou geny korelované, tak p-hodnoty jednotlivých testů (např. Fisherův test) budou nesprávné
    - Vyřešíme permutačními metodami
        - Popřehazujeme skupiny **vzorků**
        - Zopakujeme analýzu
        - Porovnáme hodnoty s pozorovanými daty

Pozor na průniky mezi dráhami

- 250 KEGG drah pro H. Sapiens
  - najčastěji zastoupené geny

PIK3CD	PIK3CG	PIK3R2	PIK3CA	MAPK3	MAPK1
70	70	70	71	78	79

Další studijní materiály a SW

- Hana Imrichová*: Možnosti propojení výsledku genomických experimentů s gene ontology online databázemi pro tvorbu metabolických sítí*, Masarykova Univerzita,2010,Bakalárska práca
- Ihnatova et al. A critical comparison of topology-based pathway analysis methods, PLoS One, 2018
- R balíky: PGSEA, GSA,ToPASeq, gage, DOSE, phenoTest, limma, GOstats
- MSigDB – web http://www.broadinstitute.org/gsea/msigdb/index.jsp
- Gorilla: http://cbl-gorilla.cs.technion.ac.il/
- DAVID: https://david.ncifcrf.gov/