LÉKAŘSKÁ FAKULTA MASARYKOVY UNIVERSITY
Interní hematoonkologická klinika LF MU a FN Brno
Centrum molekulární biologie a genové terapie

Zpracování dat

Boris Tichý
30.10.2012

EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

Převod obrazové informace na numerická data
Adresace
Segmentace
Extrakce intenzit

Příprava dat
Kontrola kvality
Normalizace
Filtry

Analýza dat
Class discovery vs. Class prediction
Supervised vs. Unsupervised
Integrace s dalšími zdroji informací

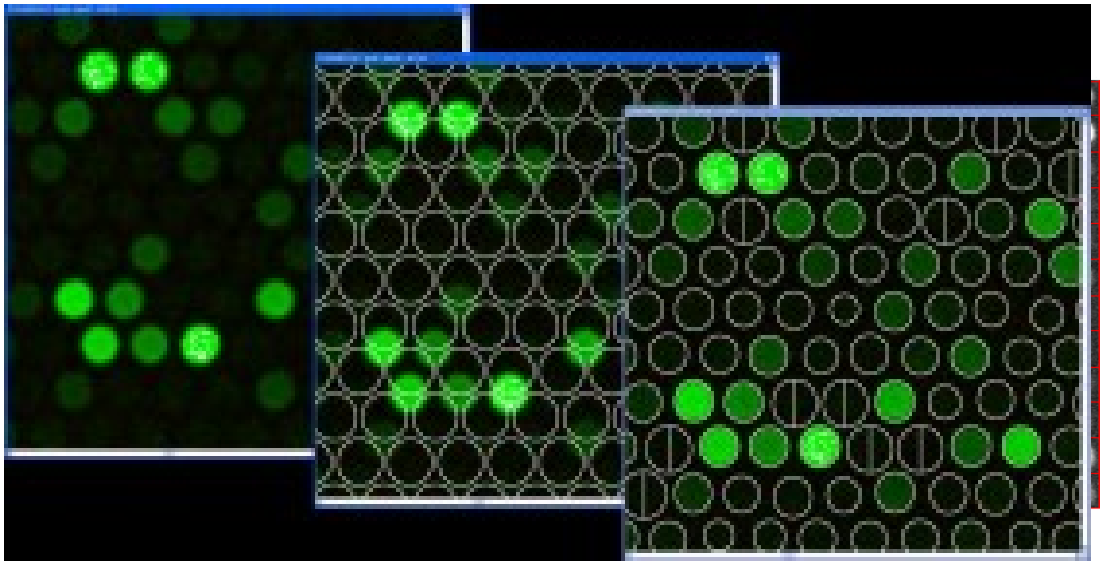
Meta-analýza

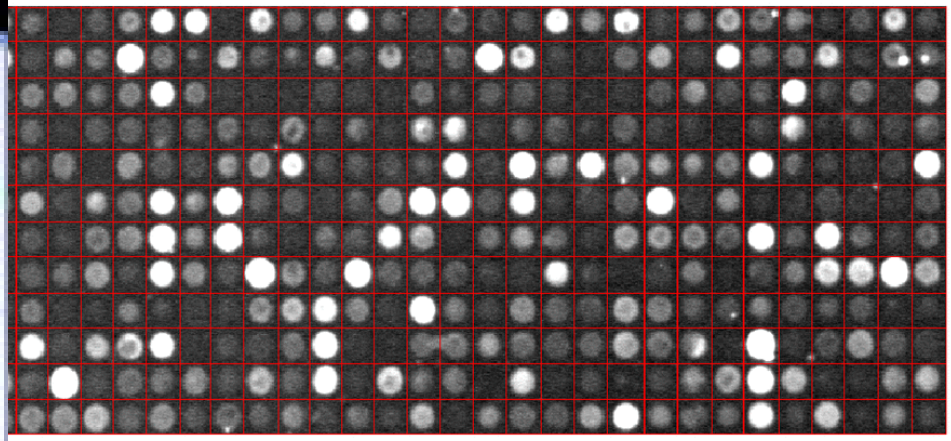
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERSITAS MASARYKIANA BRUNENSIS
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

Zpracování obrazu

Adresace (gridding)

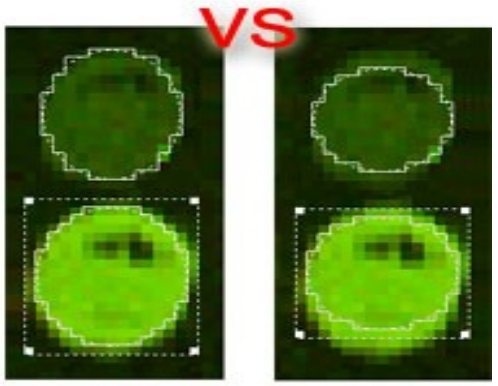
Určení pozice jednotlivých spotů
Většinou úprava předem daných očekávaných pozic
Pozice spotů bývají součástí anotace dodané výrobcem
Různé automatické a poloautomatické algoritmy
Většinou možnost manuální úpravy





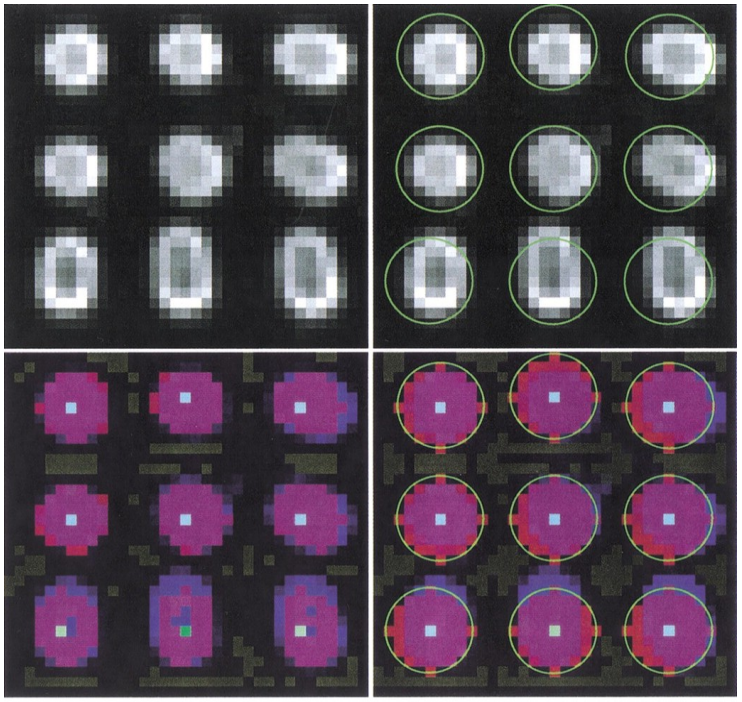
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

Zpracování obrazu

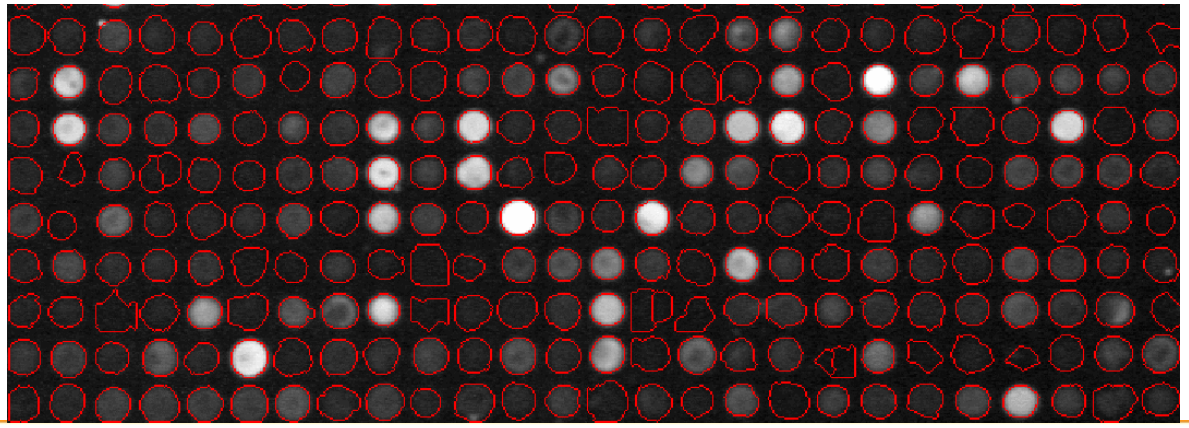


Segmentace

Klasifikace jednotlivých pixelů
Popředí
Pozadí
Různé algoritmy
Fixed circle – velmi jednoduchý, neadaptivní
Adaptive circle
Histogram
Adaptive shape – např. Watershed algoritmus
Shluková analýza



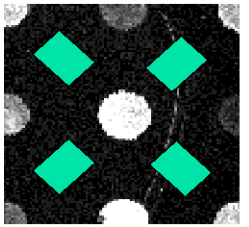
Spot default segmentation Segmentation using circles



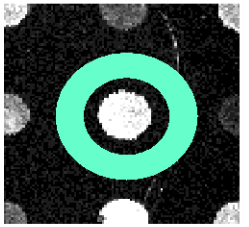
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERSITAS MASARYKIANA BRUNENSIS
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

Zpracování obrazu

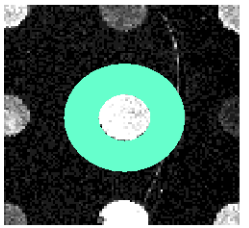
Extrakce intenzit



Různé vyjádření
Medián
Průměr
Poměr (červená/zelená)
Průměr poměrů jednotlivých pixelů



Výpočet parametrů kvality
Poměr signál/šum
Rozptyl intenzit pixelů (SD)
Cirkularita
Posun od očekávané pozice



Affymetrix

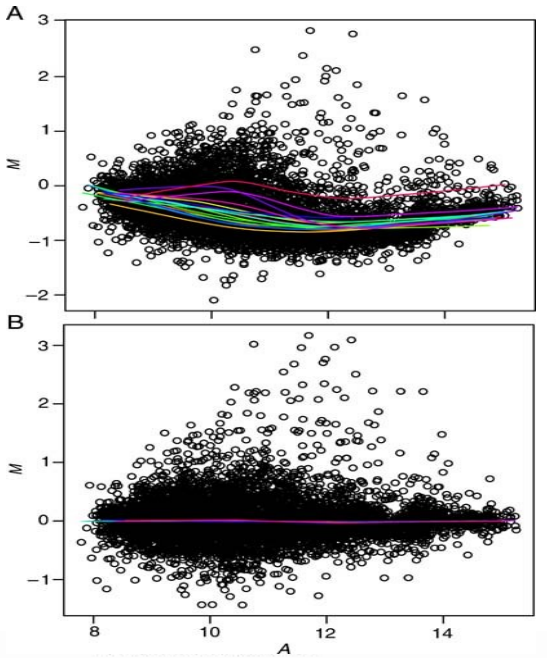
Probe sets = více sond pro jeden gen
Speciální metody pro výpočet výsledné hodnoty

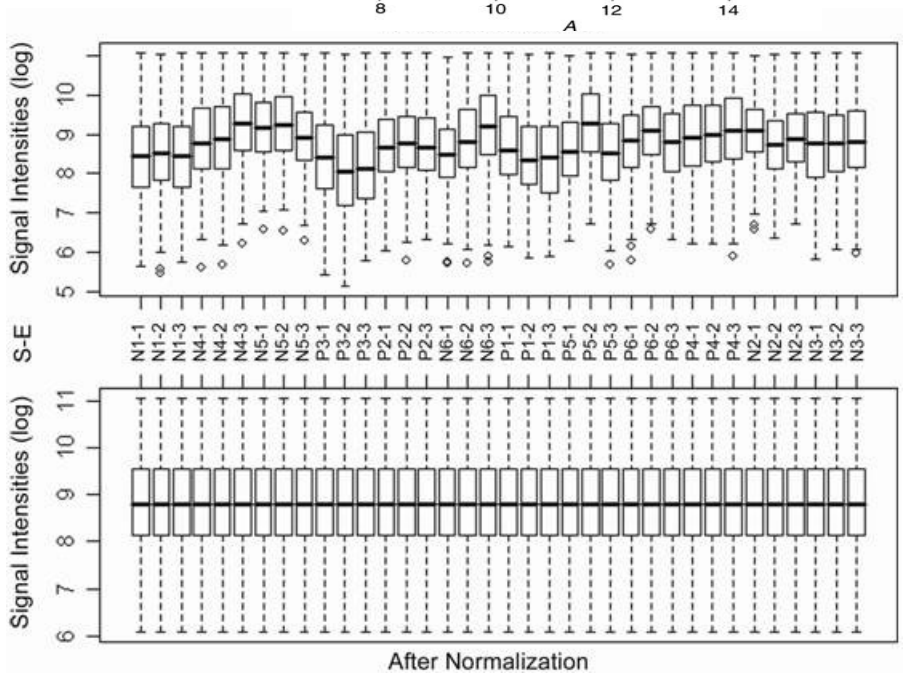
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERSITAS MASARYKIANA BRUNENSIS
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

Příprava dat

Normalizace

Dovoluje porovnávat naměřené hodnoty mezi sebou
V rámci jedné array
Mezi arrays
Jedna array
Mezi kanály (červená/zelená)
Mezi regiony
Mezi arrays
Dvoukanálové => poměry R/G
Jednokanálové => intenzity
Transformace
Log2 (poměry, intenzity)
Log10 (intenzity)





EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

Příprava dat

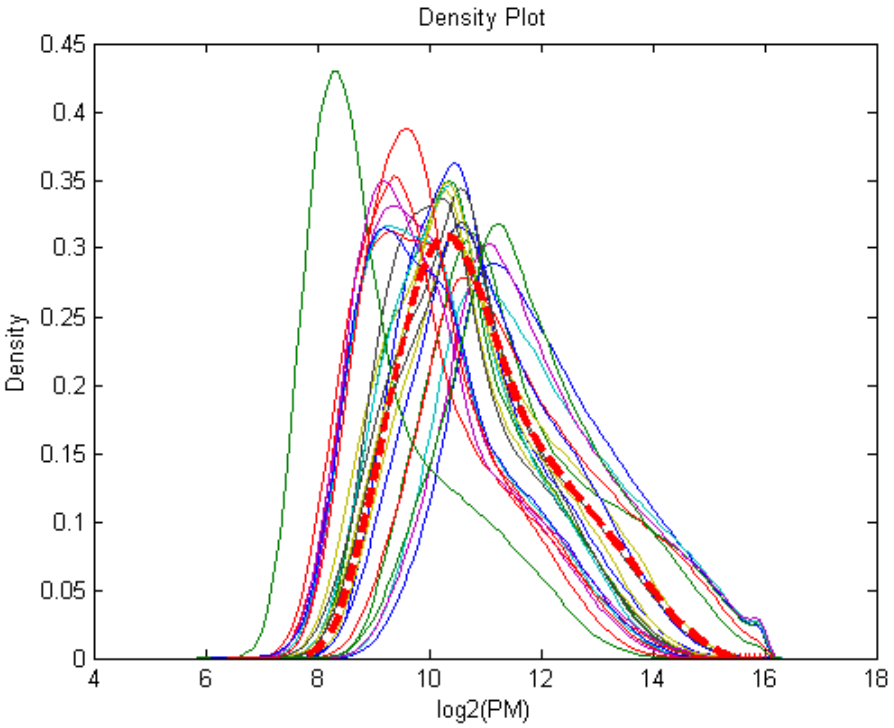
Normalizace

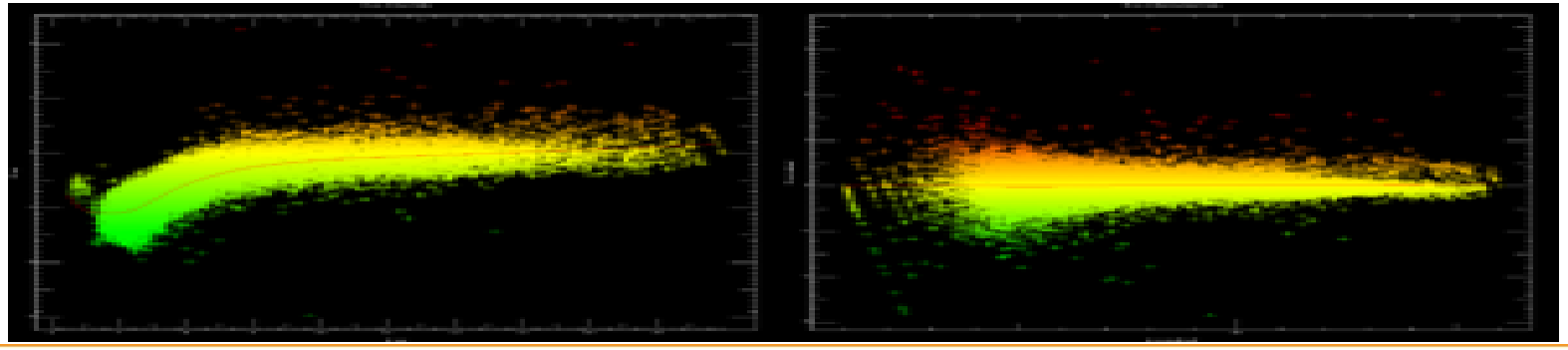
Metody
Lineární vs. nelineární

Vydělení průměrem
Scaling

Quantile
LOESS

Scaling s využitím housekeeping genů/spike-in kontrol





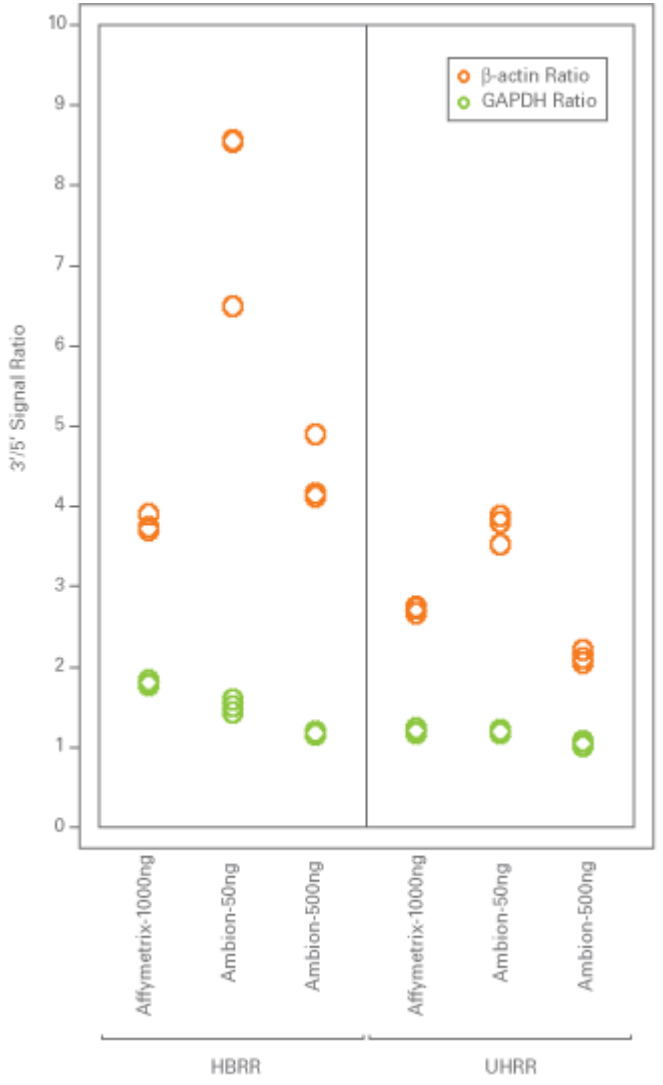
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERSITAS MASARYKIANA BRUNENSIS
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

Příprava dat

Kontrola kvality

Obecné
Pozadí
Distribuce intenzit
Distribuce poměrů

Specifické
Negativní, pozitivní kontroly
Spike-in
3'/5' poměr

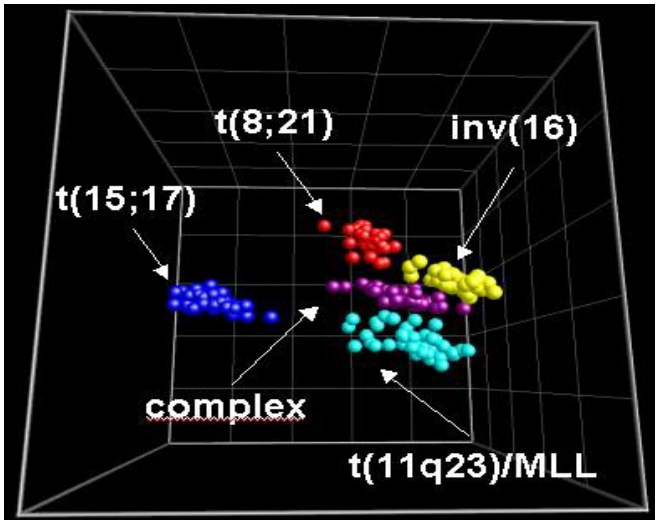


Filtry

Odstranění nekvalitních/nepotřebných dat

EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERSITAS MASARYKIANA BRUNENSIS
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

Analýza dat

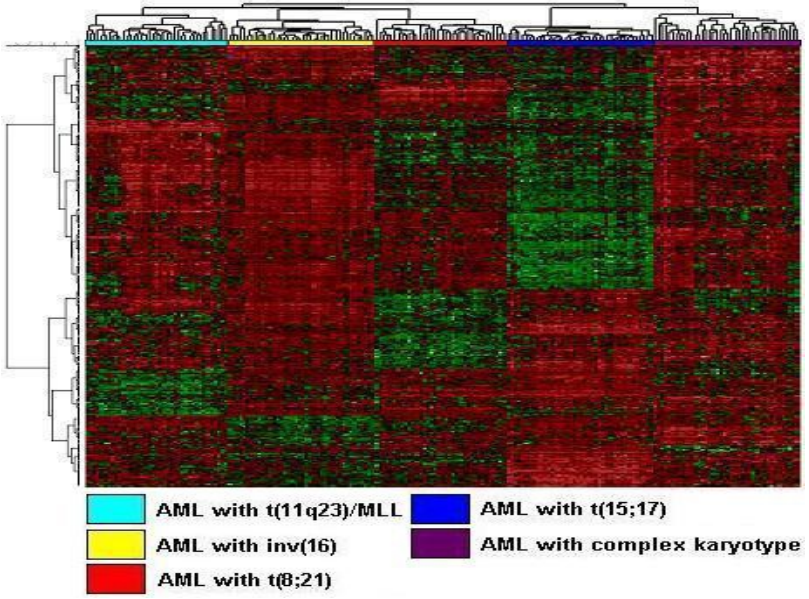


Class discovery

Nalezení vztahů v datech – skupiny vzorků, genů
Metody
Shluková analýza
    Hierarchické shlukování, k-means, self organising maps,..
Analýza hlavních komponent (PCA)

Class prediction

Klasifikace vzorků na základě expresních dat
Metody
Support vector machines
K-nearest neighbor
Neural networks
Decision trees
Nearest shrunken centroids

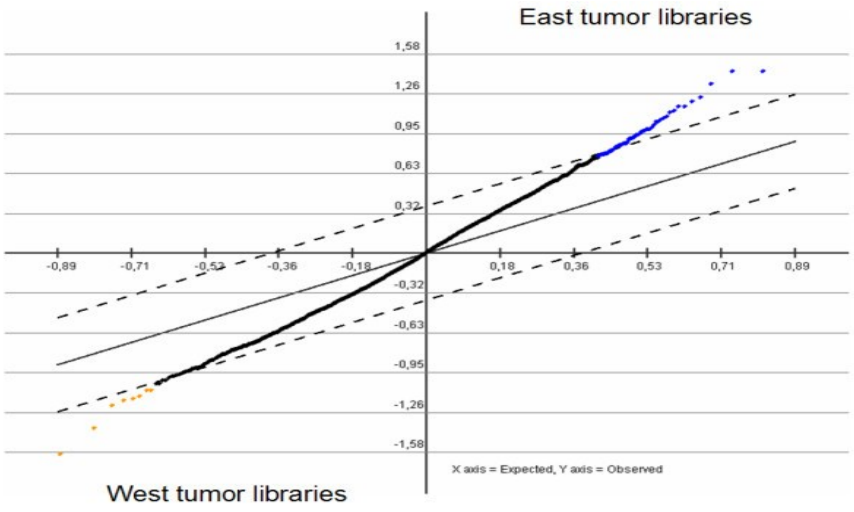


EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

Analýza dat

Unsupervised

Bez informace o vzorcích (genech)
(= Class discovery)



Supervised

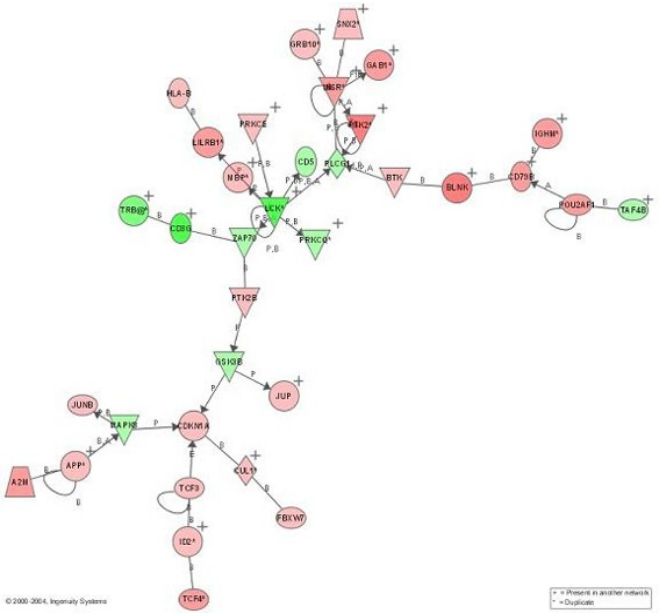
Vzorky (geny) předem rozděleny do skupin
Class prediction – algoritmy se učí na známých vzorcích
Statistické testy
2 a více skupin (vzorků) – hledání rozdílů (dif. exprimovaných genů)
t-test, ANOVA, neparametrické testy, SAM (significance analysis of microarrays)
! korekce mnohonásobného testování (Bonferoni,...), odhad chyby (FDR)

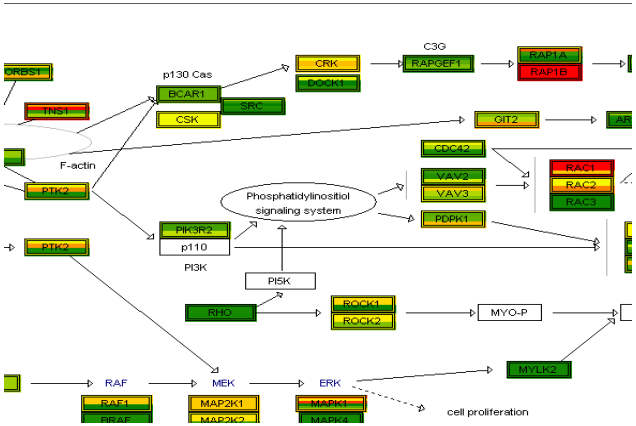
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERSITAS MASARYKIANA BRUNENSIS
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

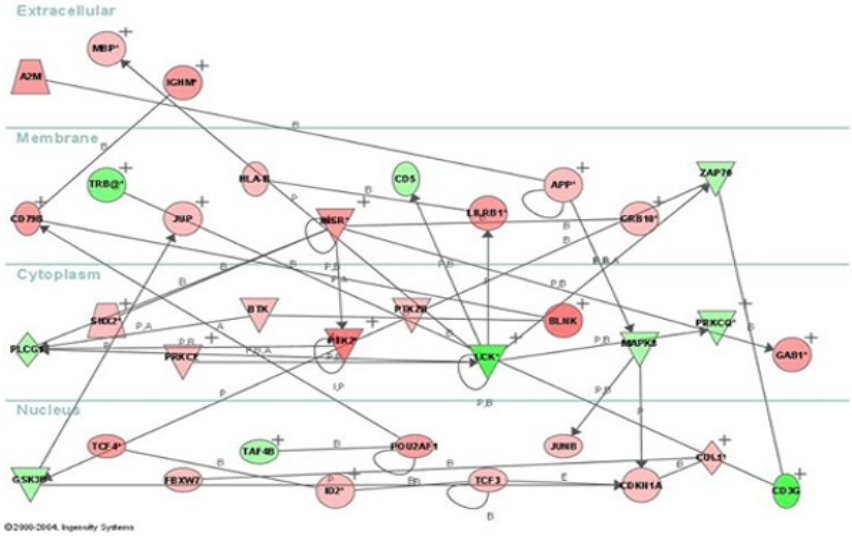
Analýza dat

Integrace s dalšími daty

Funkce genů
Gene Ontology, KEGG, BioCarta
Genové interakce, možné funkční vztahy
Transkripční faktory, microRNA motivy, publikace
Metody
Testy vyššího výskytu termínů mezi zvolenými geny (GSEA)
Sítě (Networks)
Dráhy (Pathways)







EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERSITAS MASARYKIANA BRUNENSIS
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky

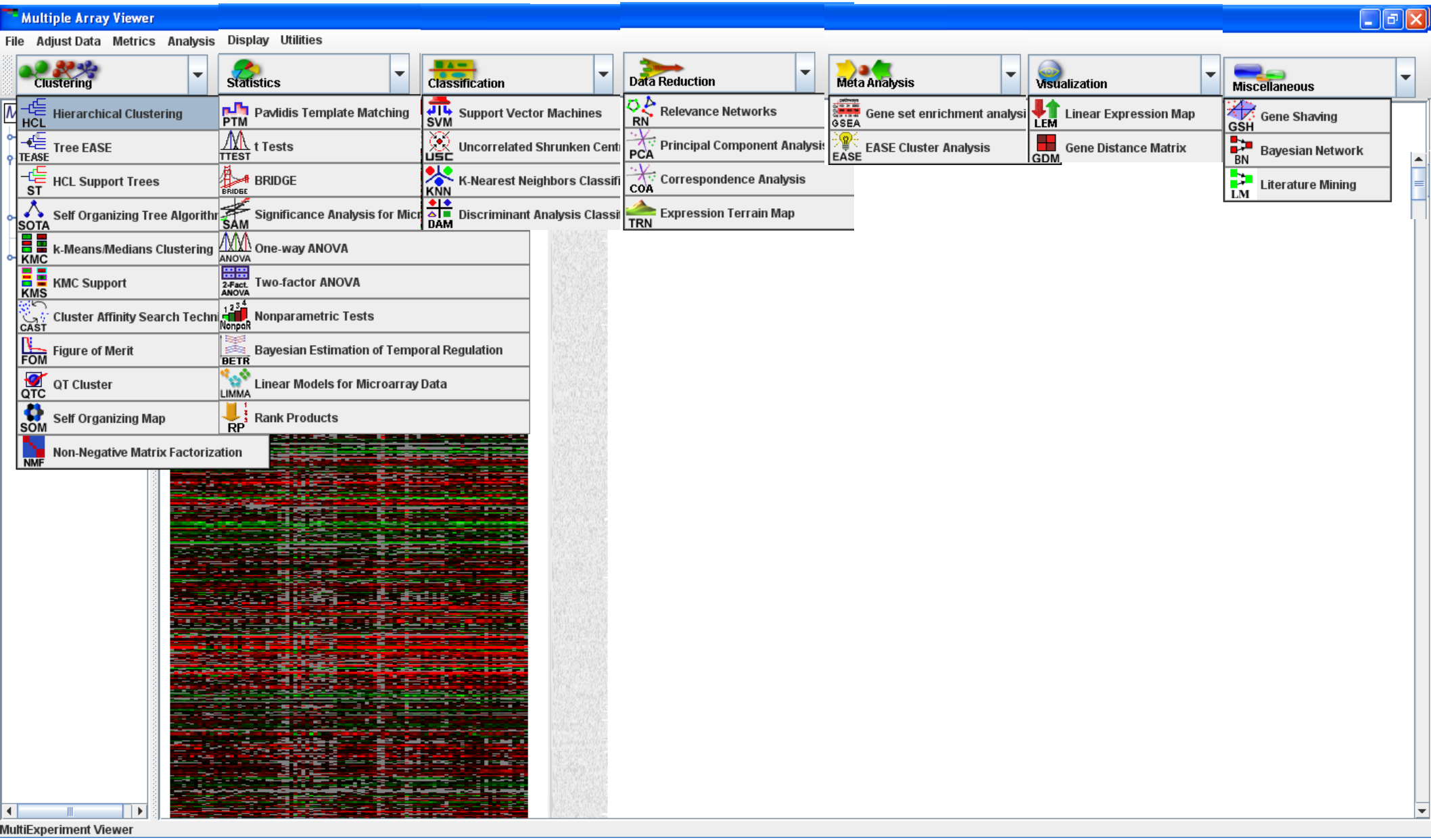
Analýza dat

Metaanalýza

Analýza více experimentů
Úložiště dat
GEO – gene expression omnibus (NCBI)
ArrayExpress (EBI)
Standardy
MAGE-ML – formát dat (jazyk)
MIAME – Minimum Information About Microarray Experiment
Informace umožňuje zopakování experimentu

Nedostatek informací o vzorcích
Porovnatelnost platforem, laboratoří

EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky



EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ
Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky