

Získávání expresních dat
Uplatnění v medicíně
Analýza expresních dat
Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

IV107 Bioinformatika I

Přednáška 8

Katedra informačních technologií
Masarykova Univerzita Brno

Jaro 2008

Předchozí týden

- ▶ Alternativní struktury DNA

- ▶ křížová
- ▶ triplex
- ▶ tetraplex

- ▶ Tandemová a jiná opakování (Xlandscape)

- ▶ Interakce DNA/protein (TRANSFAC, TESS)

Egr-1 early growth response 1 (induces apoptosis,
consensus WTGCGTGGGCGK)

GATATACGG Msx-1 (aka Hox-7; muscle segment
homeobox-like 1; CNGTAWNTG)

- ▶ Interakce protein/protein (DIP, PIP)

- ▶ Fylogenetické stromy (program PHYLIP)

- ▶ Matice párových vzdáleností
- ▶ UPGMA, NJ, max parsimony, ML
- ▶ format Newick A(B(D,E),C)

Získávání expresních dat

Uplatnění v medicíně

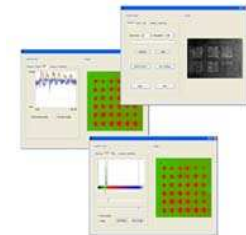
Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

Cesta k údajům o expresi



Získávání expresních dat

Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

Získávání expresních dat

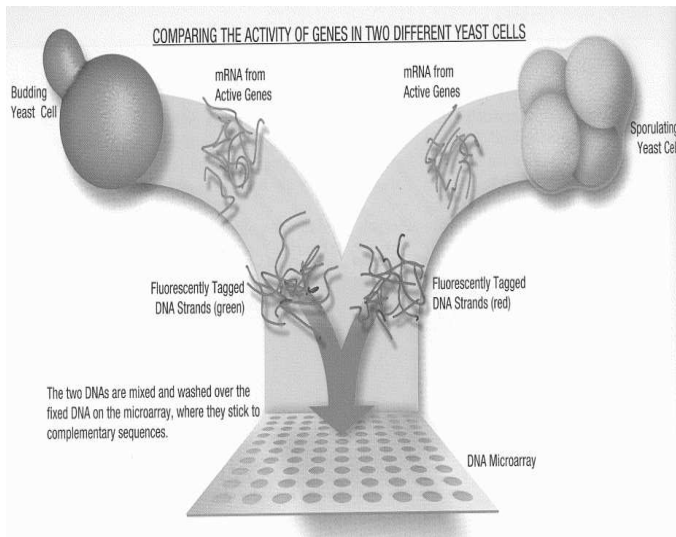
Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štápení proteinů a hmotnostní
spektra



Získávání expresních dat

Uplatnění v medicíně

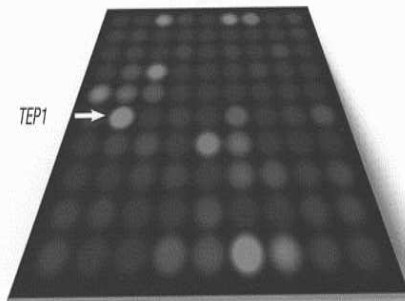
Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

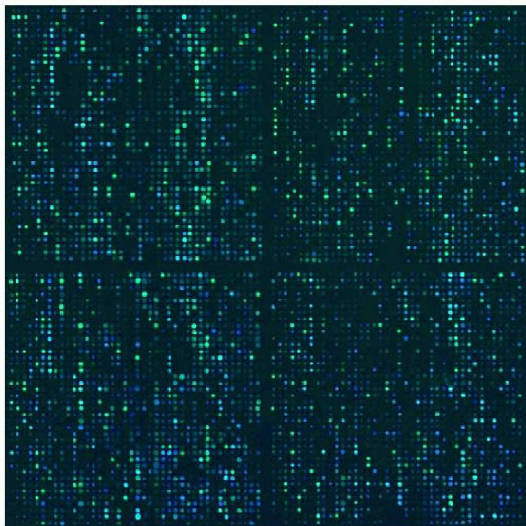
Štěpení proteinů a hmotnostní
spektra

A scanner measures the fluorescence of each dye separately.



The higher ratio of red to green fluorescence at the spot labeled *TEP1*, for instance, shows that the *TEP1* gene is expressed more highly in sporulating than in budding yeast cells.

Nasnímaný obraz z DNA čipu vypovídá o přítomnosti DNA určité sekvence ve vzorku



Získávání expresních dat

Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

Získávání expresních dat

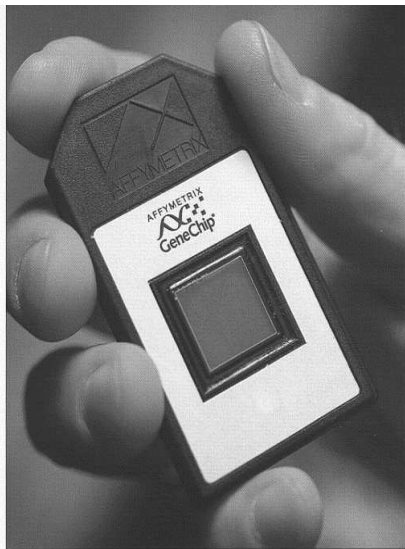
Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra



*Factory-made DNA chips such as this one
are produced with techniques normally
used in the semiconductor industry.*

Získávání expresních dat

Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

Uplatnění v medicíně

Rakovina nekontrolované dělení buněk v důsledku
mutace (onkogeny, supresory nádorů).

Měření genové exprese nachází uplatnění v diagnostice
rakoviny.

- ▶ o jaký druh rakoviny se jedná ?
- ▶ jaká je prognóza pro danou diagnózu ?

Získávání expresních dat

Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

Rakovina krvi nebo kostní dřeně leukemie

- ▶ Myelogenní, lymfocytická, akutní, chronická ?
- ▶ Acute Myelogenous Leukemia (AML) Acute Lymphocytic Leukemia (ALL)
- ▶ Není produkováno dostatečné množství krvinek
- ▶ Anemie → slabost, infekce, snížená srážlivost krve
- ▶ AML: 10,000 případů ročně
- ▶ ALL: 3,500 dospělí / 2,400 novorozenci
- ▶ AML nebo ALL ?

Získávání expresních dat

Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

- ▶ **Diagnóza ve specializované laboratoři**
 - ▶ ALL: 58
 - ▶ AML: 14
- ▶ **Léčba: chemoterapie, transplantace kostní dřeně**
 - ▶ ALL: corticosteroidy, vincristin, methotrexat, L-asparaginasa
 - ▶ AML: daunorubicin, cytarabin
- ▶ **Správná diagnóza je velice důležitá pro nasazení správné léčby**
- ▶ **Microarray může podporovat rozhodování**

Získávání expresních dat

Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štápení proteinů a hmotnostní
spektra

Příklad sady genů, které korelují s prognózou



Analýza expresních dat

- ▶ měření s nízkým počtem opakování
- ▶ rozpoznávání obrazu
- ▶ převod na matici číselných hodnot
- ▶ normalizace
- ▶ statistika jednotlivých pozic a genů (průměrná hodnota, rozptyl, eliminace extrémních hodnot, atd.)
- ▶ kontrast: dvě sady experimentálních podmínek
- ▶ specifika: $n \gg p$
- ▶ PCA, LDA, shluková analýza (clustering), strojové učení
- ▶ analýza v kontextu funkce (GO, KEGG)
- ▶ analýza v kontextu sekvence (blízké motivy a geny, zvláště promotor)
- ▶ analýza časových a prostorových řad (koregulované geny)

Získávání expresních dat

Uplatnění v medicíně

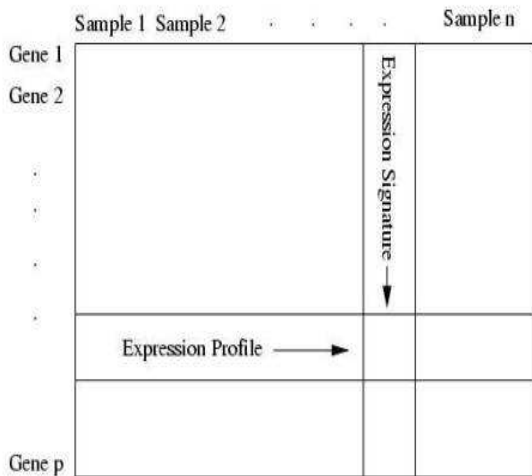
Analýza expresních dat

Příklad použití microarray dat v klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní spektra

Co je to expresní profil?



Získávání expresních dat

Uplatnění v medicíně

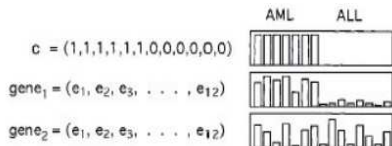
Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štápení proteinů a hmotnostní
spektra

Jak zjišťujeme podobnost dvou expresních profilů či jejich schopnost klasifikovat vzorek?



Expresní profil je chápán jako množina bodů v multidimenzionálním prostoru, popř. jako sada postupných výběrů z populace genů či zdroj informací o vzorku

$c \cdot gene_1$ vektorový součin (cos uhlu, který zvírají)

$d(c, gene)$ euklidovská vzdálenost (postihuje i absolutní hodnoty exprese)

$corr(c, gene)$ korelace

$MI(c, gene)$ vzájemná informace

Získávání expresních dat

Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v klasifikaci nádorů

Příště

Štápení proteinů a hmotnostní spektra

Získávání expresních dat

Uplatnění v medicíně

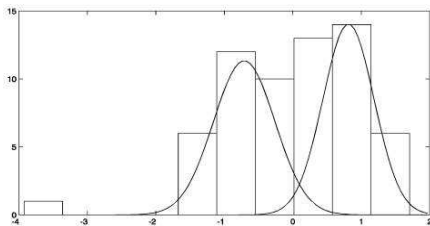
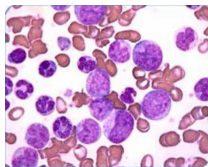
Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

Problém smíšených dat



Získávání expresních dat

Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

Multiclass cancer diagnosis using tumor gene expression signatures (Ramaswamy et al., 2001)

- ▶ V praxi se k diagnostice používá soubor klinických a histologických vyšetření.
- ▶ Je možné je nahradit molekulárním vyšetřením ?
 - ▶ 218 nádorů a 90 zdravých tkání
 - ▶ microarray (exprese genů)
 - ▶ 5 000 000 zjištěných hodnot
 - ▶ metody strojového učení (CLUSTER, SOM, SVM)
 - ▶ klasifikace (OVA = one v. all)
 - ▶ úspěšnost 78%

Získávání expresních dat

Uplatnění v medicíně

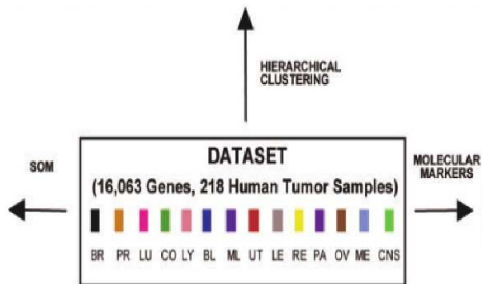
Analýza expresních dat

**Příklad použití microarray dat v
klasifikaci nádorů**

Příště

Štěpení proteinů a hmotnostní
spektra

Třídy a metody pro klasifikaci



Získávání expresních dat

Uplatnění v medicíně

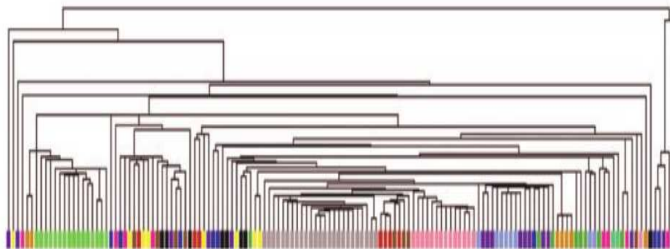
Analýza expresních dat

**Příklad použití microarray dat v
klasifikaci nádorů**

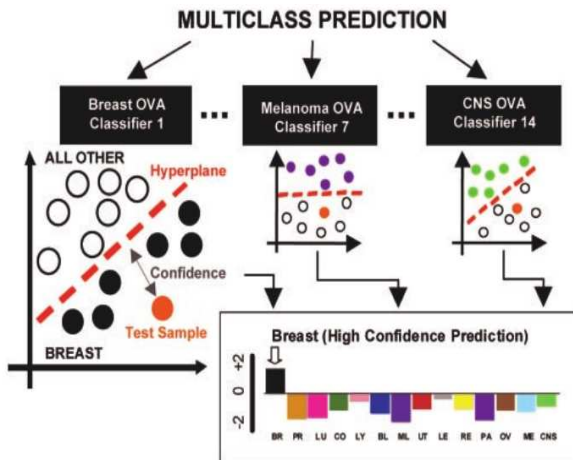
Příště

Štěpení proteinů a hmotnostní
spektra

Hierarichické aglomerativní zhlukování



Jak funguje klasifikace do více tříd



Získávání expresních dat
Uplatnění v medicíně
Analýza expresních dat
Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

Získávání expresních dat

Uplatnění v medicíně

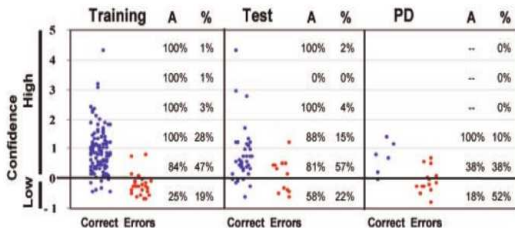
Analýza expresních dat

**Příklad použití microarray dat v
klasifikaci nádorů**

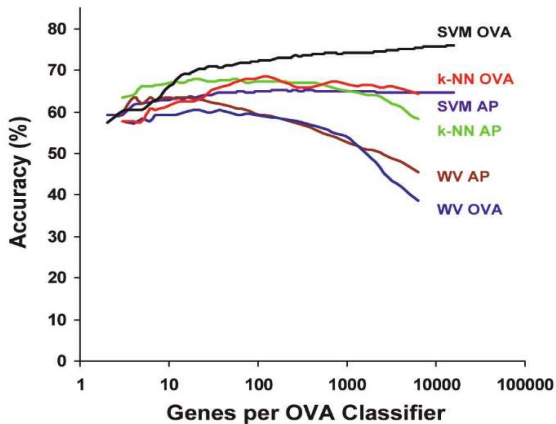
Příště

Štěpení proteinů a hmotnostní
spektra

Kvalita predikcí s různými stupni spolehlivosti



Počet genů potřebný k úspěšné klasifikaci



Získávání expresních dat

Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štěpení proteinů a hmotnostní
spektra

Získávání expresních dat
Uplatnění v medicíně

Analýza expresních dat

Příklad použití microarray dat v
klasifikaci nádorů

Příště

Štápení proteinů a hmotnostní
spektra

Spolehlivost klasifikace

Train

Predicted Class

	BL	BR	CNS	CO	LE	LU	LY	ME	ML	OV	PA	PR	RE	UT	<i>n</i>
BL	63%							13%		13%					8
BR		88%													8
CNS			100%												16
CO				75%						13%					8
LE					100%										24
LU	13%	13%				50%				13%					8
LY							100%								16
ME								100%							8
ML				13%					63%						8
OV	13%								13%	38%			13%	25%	8
PA	13%	13%						13%			63%				8
PR						13%			13%			75%			8
RE		13%							13%	13%			63%		8
UT														88%	8
<i>n</i>	8	11	16	10	24	5	16	10	8	8	6	7	6	9	144

Získávání expresních dat
Uplatnění v medicíně
Analýza expresních dat
Příklad použití microarray dat v
klasifikaci nádorů

Příště

**Štěpení proteinů a hmotnostní
spektra**

Příště Štěpení proteinů a hmotnostní spektra

Outline

Dodatek

Dodatek

For Further Reading

Dodatek

For Further Reading

For Further Reading
X