C7188 Úvod do molekulární medicíny 4/12

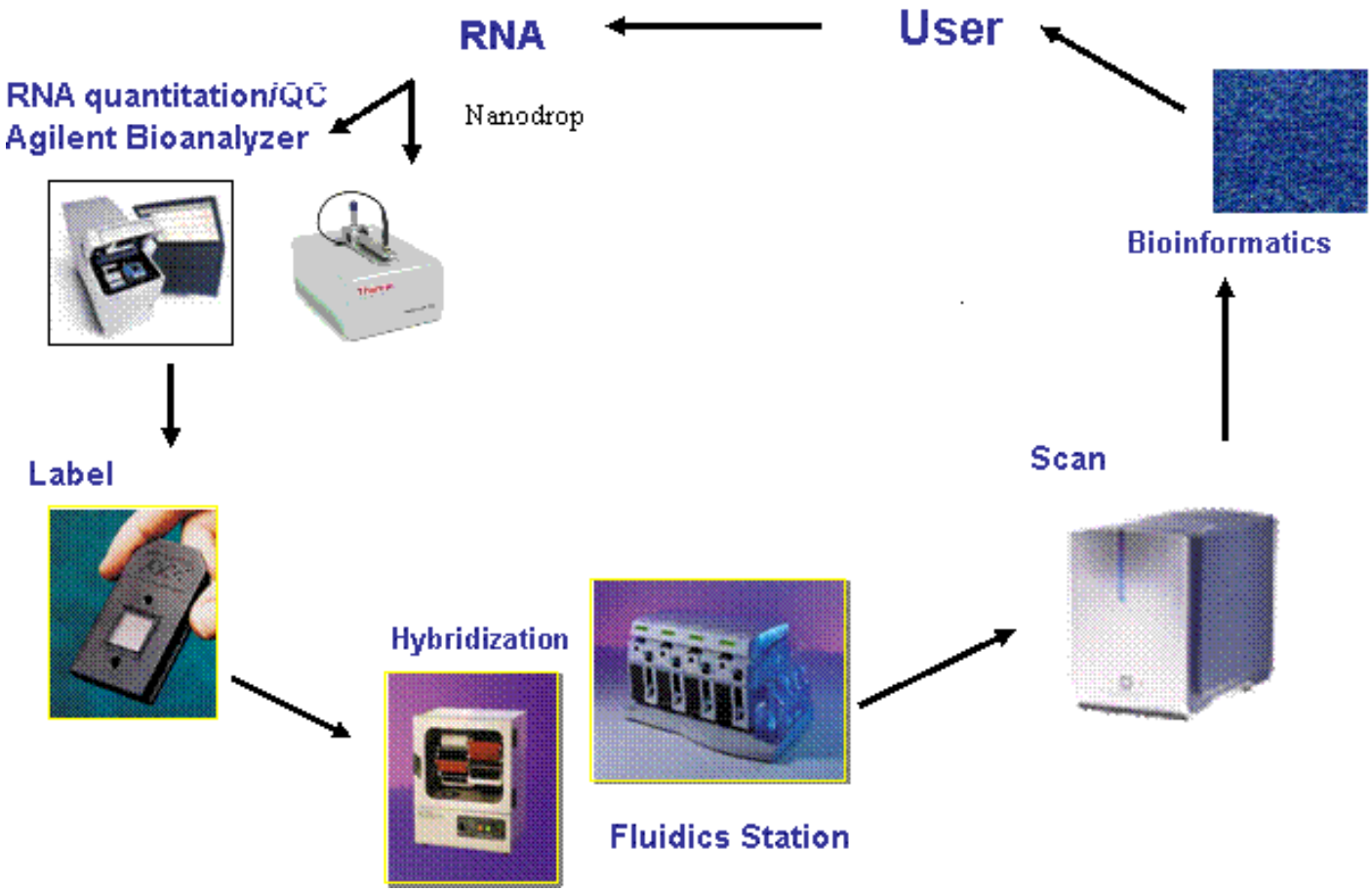
Moderní metodické přístupy v molekulární medicíně II

GENOMIKA II

Ondřej Slabý, Ph.D.
Masarykův onkologický ústav
Univerzitní centrum buněčné imunoterapie
Lékařská fakulta Masarykovy univerzity

© Ondřej Slabý, 2011

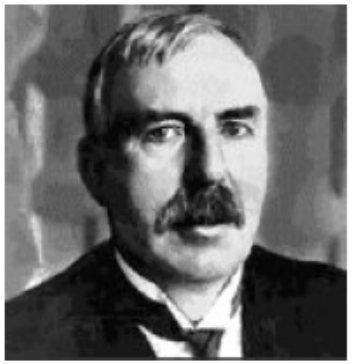
Microarray sample processing workflow

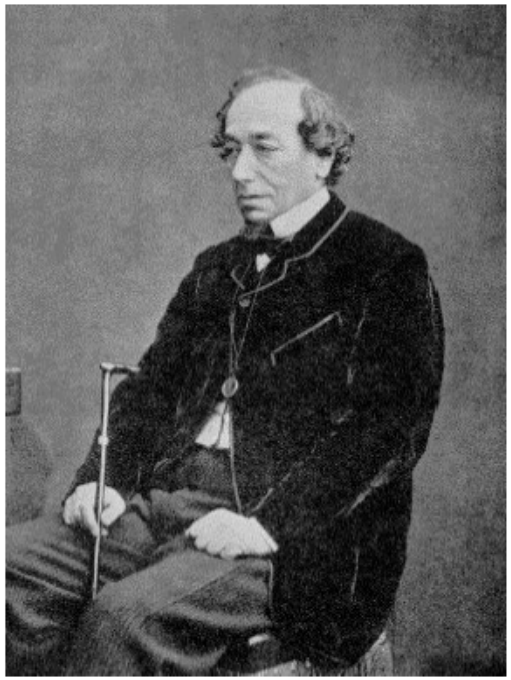


© Ondřej Slabý, 2009

“If your experiment needs statistics, you should have done a better experiment.”

Ernest Rutherford





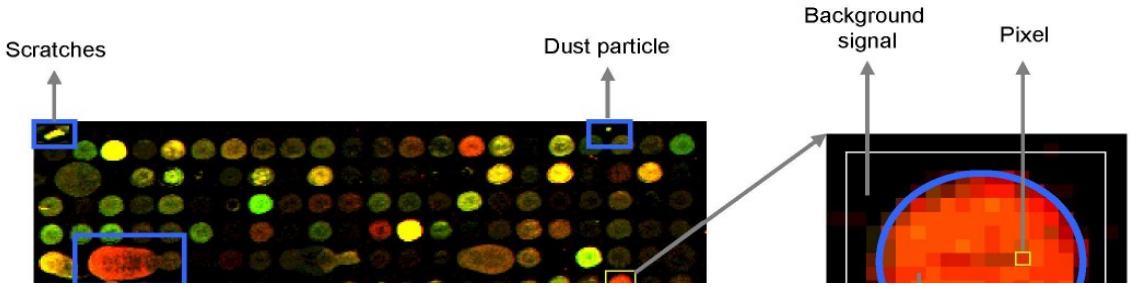
“There are three types of lies: lies, damn lies, and statistics!”

Benjamin Disraeli

© Ondřej Slabý, 2009

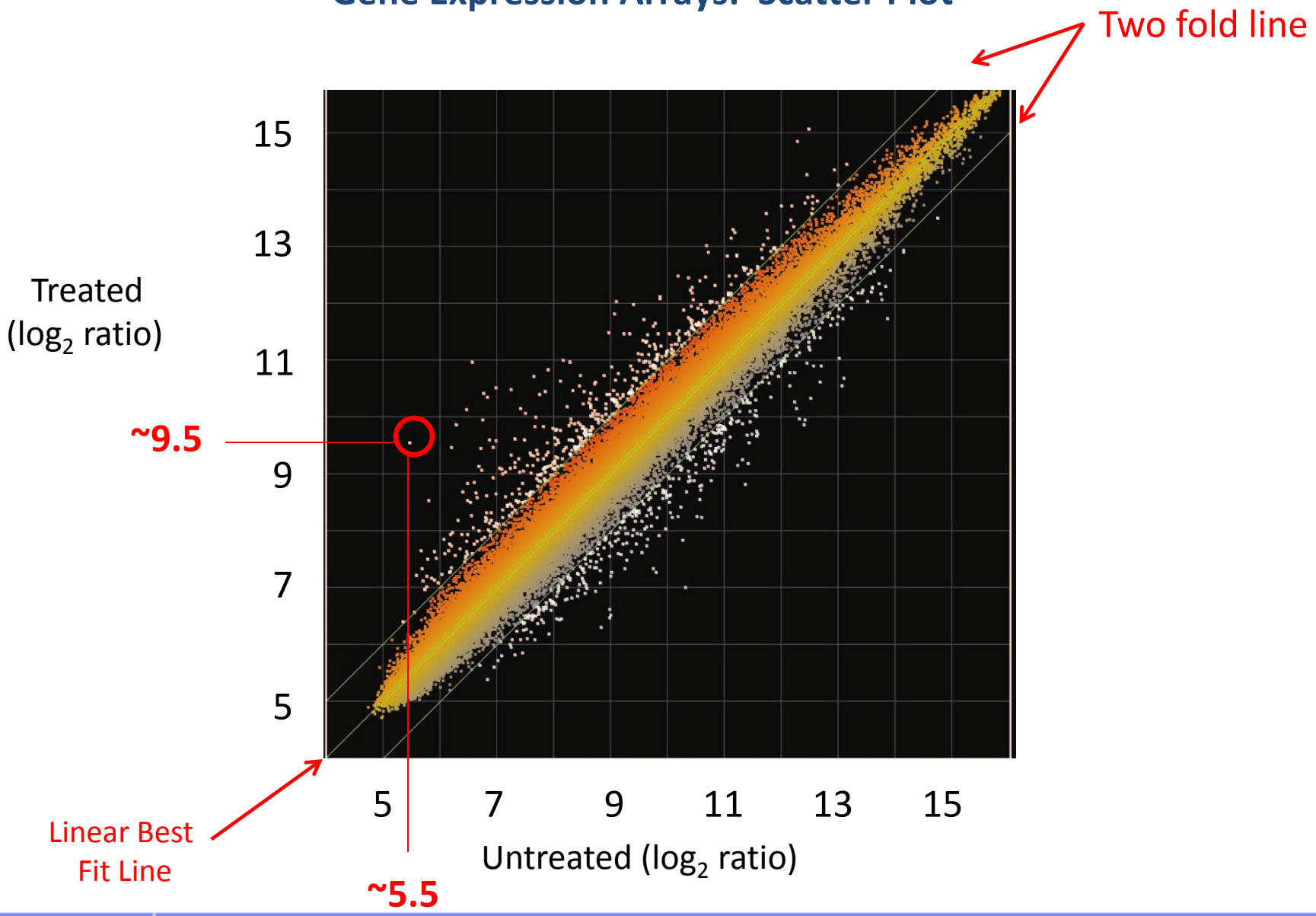
Analýza obrazu

Odečtení pozadí od „foreground“
– background substraction
(griding,local background, median, empty spot,…)
SNR – „signal-to-noise ratio“, odstup intenzity popředí od šumu pozadí
-převod obrazové informace na numerická data



© Ondřej Slabý, 2009

Gene Expression Arrays: Scatter Plot



© Ondřej Slabý, 2009

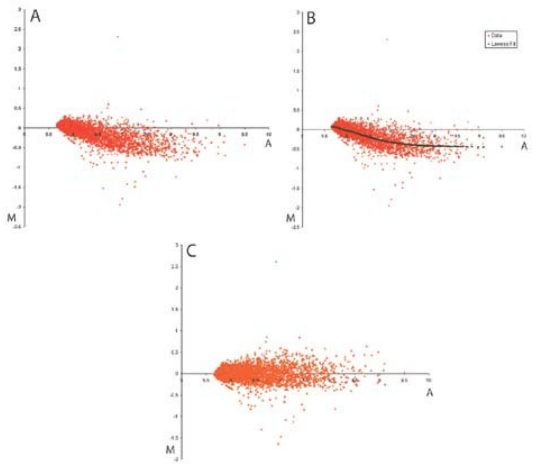
Normalizace dat

ukázky normalizačních metod

Základem většiny normalizačních metod je předpoklad, že počet genů se změněnou mírou exprese je výrazně nižší než těch, jejichž exprese se nemění.

-minimalizace vlivu „nebiologických“ zdrojů variability

-kompenzace nelinearity dat mezi jednotlivými čipy a uvnitř daného čipu



Pozitivní kontroly:

Referenční geny s „konstantní“ expresí ve tkáních

kontrolní genetický materiál (referenční vzorek)

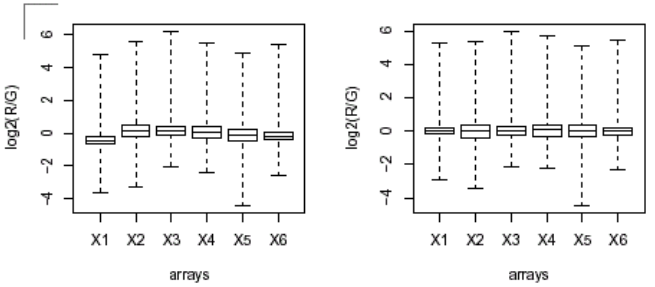
kontroly účinnosti hybridizace – artreficiální sekvence

Negativní kontrola:

-pozadí hybridizace

-Affymetrix - mutace v jednom nukleotidu sondy

Mean, median normalization



Subtract common mean or median:

old values: $y_{gi} = \log_2(R_{gi}/G_{gi})$ on array i

new values: $y_{gi}^{(n)} = y_{gi} - \text{mean}_i$

© Ondřej Slabý, 2009

Normalizace dat Cy3 a Cy5

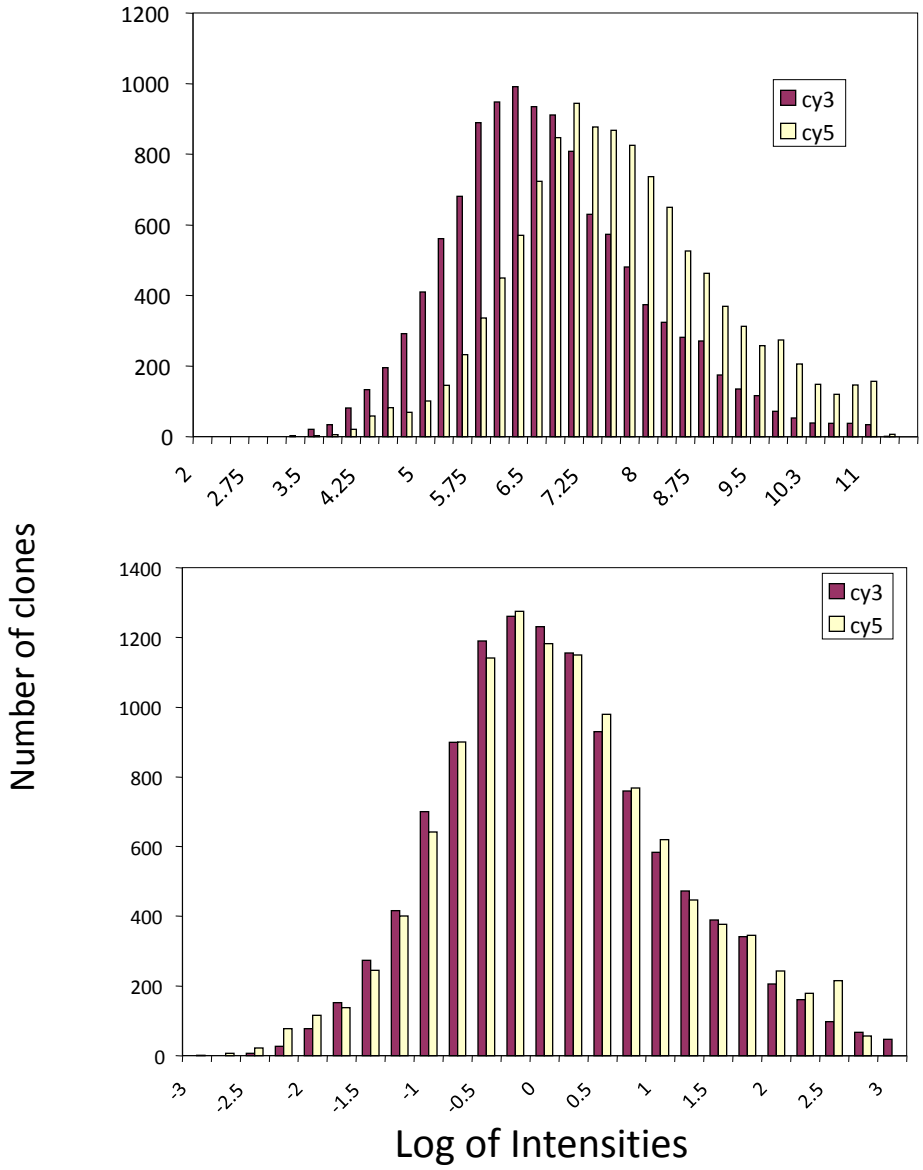
vybalancování rozdílného signálu ze značení Cy3 a Cy5

Jeho nerovnoměrnost může být způsobena:

-rozdílnou inkorporací barviva do NK

-rozdílným množstvím mRNA

-odlišnými parametry při skenování

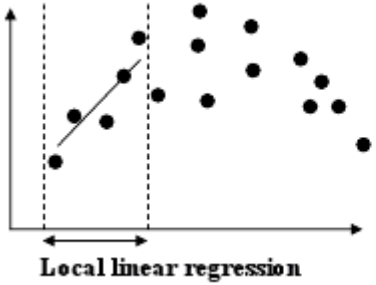


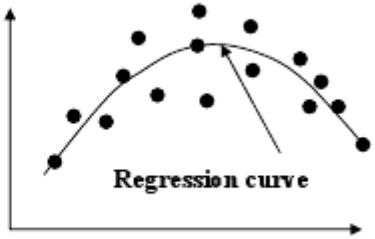
© Ondřej Slabý, 2009

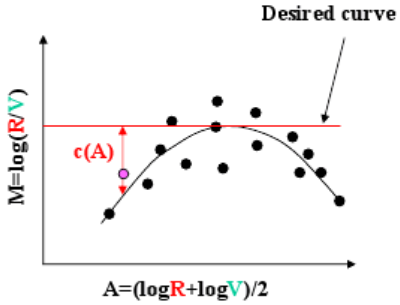
Normalizace Lowess: princip

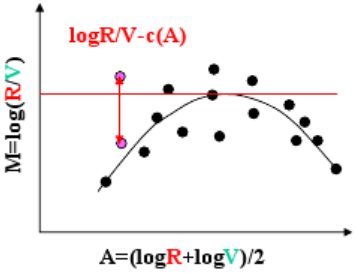
Loess (or lowess) : Locally WEighted Scatterplot Smoothing (vyhlazování)
Normalizace závislá na intezitě signálu
Místní lineární regrese

- $\log(R/V) \rightarrow \log(R/V) - c(A)$

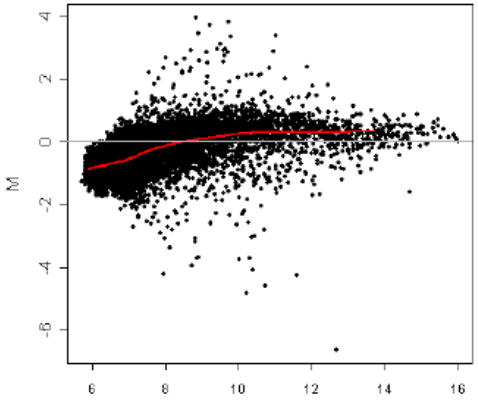




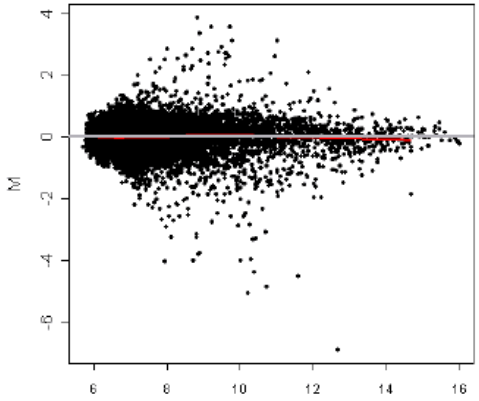




Before normalization



After normalization



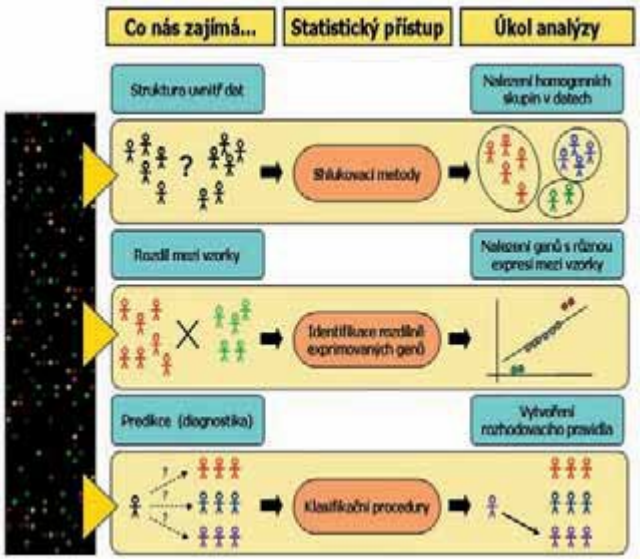
© Ondřej Slabý, 2009

Statistická analýza dat

Základní dělení metod je na „unsupervised“ (využívají jako vstup pouze naměřené hodnoty bez informací o vzorcích, klastrovací metody) a „supervised“ (využívají i další známé informace o vzorcích, identifikace biologicky významných genů, klasifikační metody)

Další skupina metod propojuje informaci o genové expresi s dostupnými informacemi o biologické funkci genů (analýzy signálních drah, informace z databáze GO „gene ontology“, a další)

1) Identifikace biologicky významných genů

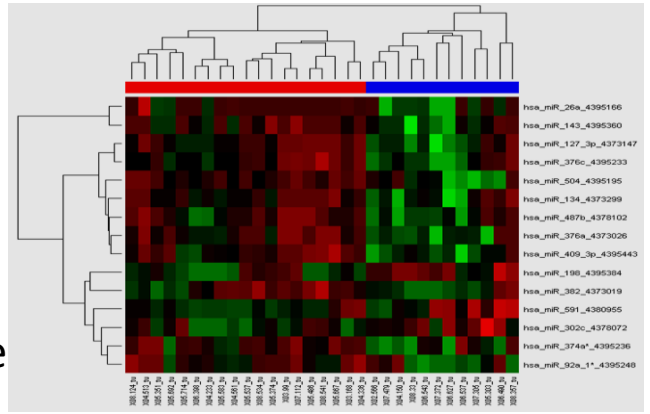


-geny s reprodukovatelnou signifikantně rozdílnou expresí mezi jednotlivými podmínkami experimentu
-poměr exprese v jednotlivých experimentech (fold change)
-t-test (test rozdílnosti průměrů exprese v jednotlivých skupinách)
-neparametrické testy (Mann-Whitney, Wilcoxon test)
-Significance Analysis of Microarrays (SAM)
-Multifaktoriální ANOVA (nejsignifikantnější geny pro dané skupiny)

Hranice významnosti (1% = 1 ze 100, u čipu je to 200 falešně pozitivních z 20000, Bonferoniho korekce)

© Ondřej Slabý, 2009

2) Ukázky multidemenzionálních metod analýzy čipových dat
Shlukovací analýzy



Shlukova analýza je jednou z nejpoužívanějších vícerozměrných statistických metod

Jedna se o explorativni techniku, ktera se používá zejmena v případech, kdy nemame žadne *a priori znalosti* o struktuře uvnitř dat.

každý gen je reprezentován vektorem jehož souřadnice, jsou hodnoty exprese genu v jednotlivých experimentech (v jednotlivých vzorcích), vzdálenost je měřena mezi vektory.

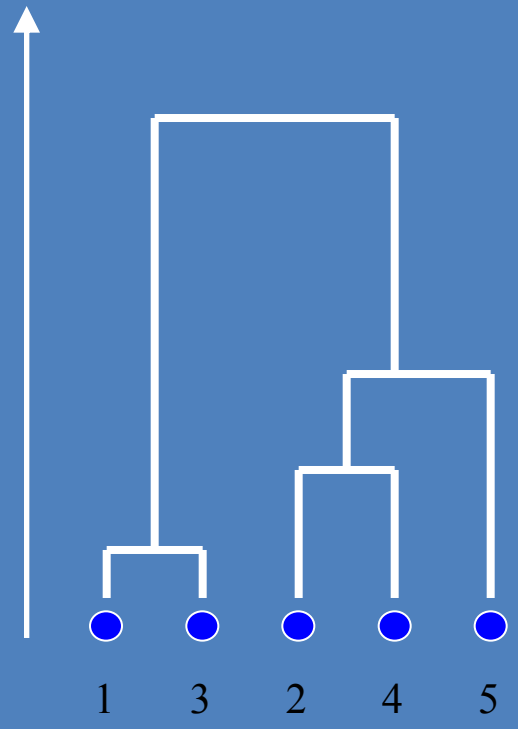
Ukolem shlukovacích metod je tedy najit v datech skupiny prvků (shluky) tak, že prvky jednotlivých skupin budou v jistem smyslu více podobné než prvky z jiných skupin, tzn. nalezené skupiny prvků budou co nejvíce homogenní
Snažíme se nalézt mezi zkoumanými geny (resp. biologickými vzorky) skupinky genů (resp. biologických vzorků), ktere vykazují za specifických podminek nebo u daného fenotypu, podobné chování.

© Ondřej Slabý, 2009

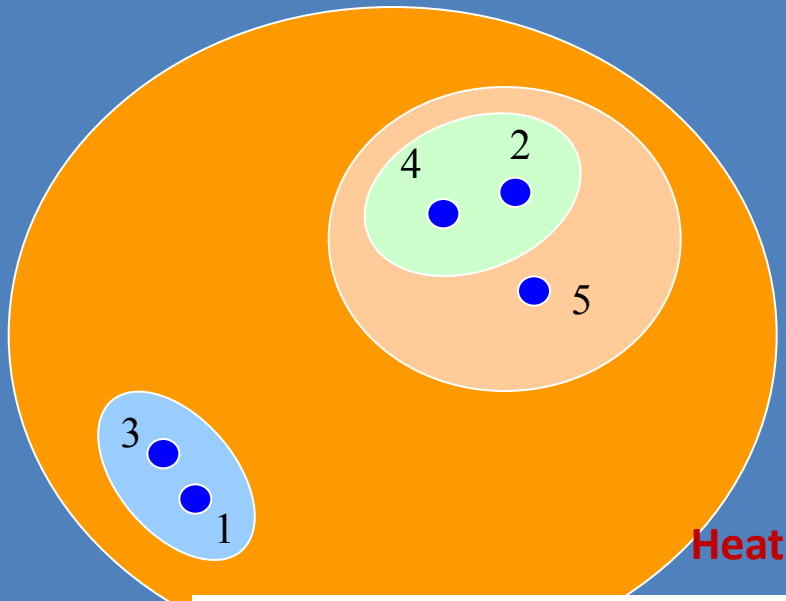
Hierarchické klastrování

-kalkulace vzdálenosti mezi všemi geny a nalezení nejmenší. K ní se seskupí všechny jí podobné a vytvoří se klastr.
-po vytvoření X počtu clusterů se hledají vzdálenosti mezi klastry (hierarchical clustering)
-počet klastrů není omezen

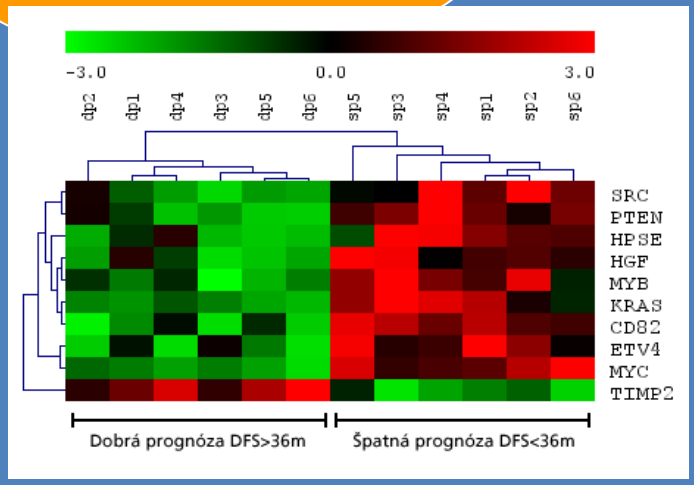
Vzdálenost mezi jednotlivými klastry



Dendrogram



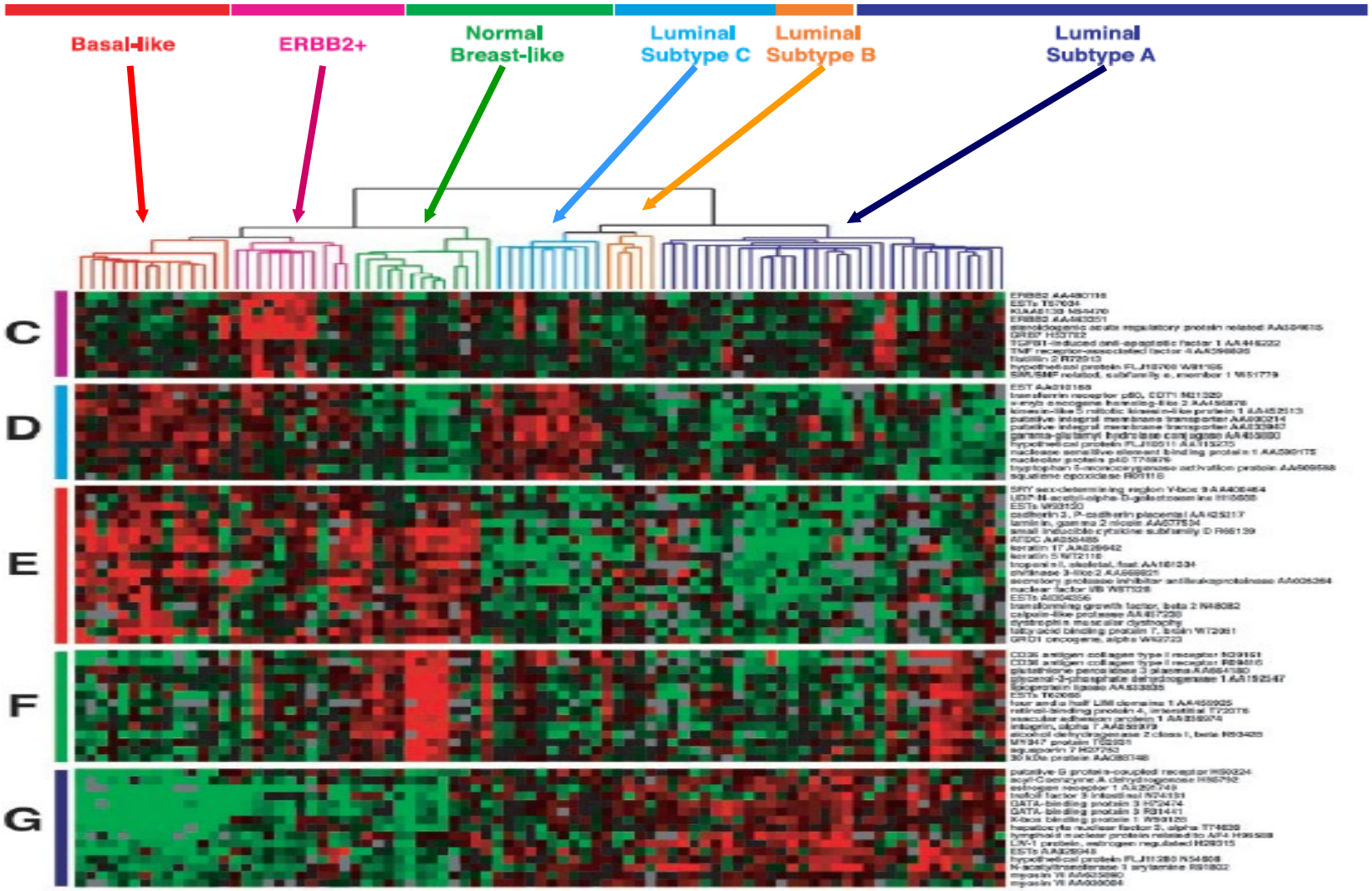
Heat map



Podobnost je vyjádřena hierarchickým stromem – dendrogram s teplotní mapou „heat-map“

© Ondřej Slabý, 2009

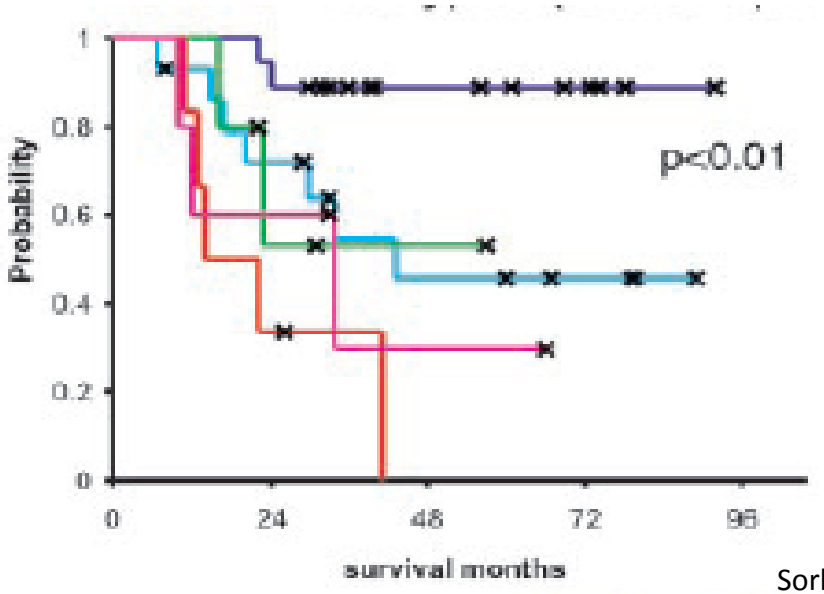
Klastrová analýza u karcinomu prsu

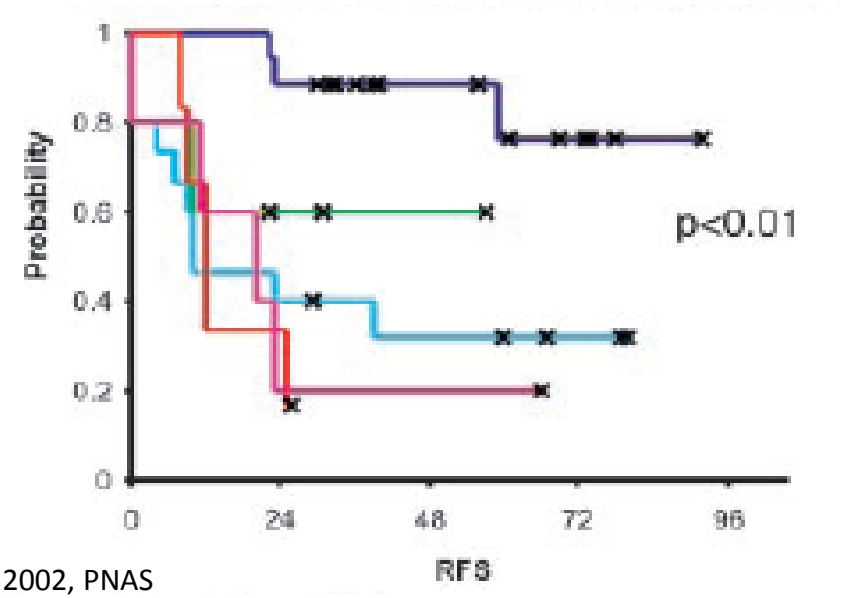


Sorlie et al, 2002, PNAS

© Ondřej Slabý, 2009

	Basal-like	ERBB2+	Normal Breast-like	Luminal Subtype C	Luminal Subtype B	Luminal Subtype A
ER	-/+	-/+	-/+	+/++	++	+++
HER2	-	+++	- / +	- / +	-	- / +
p53mut	82%	71%	33%	80%	40%	13%
CK 5/6, 17	+++	+/-	+++	-	-	-
CK 8/18	-	-	+	+/++	+/++	+++
c-myb / ost.	+++	+++	-/+	+++	-	-





Sorlie et al, 2002, PNAS

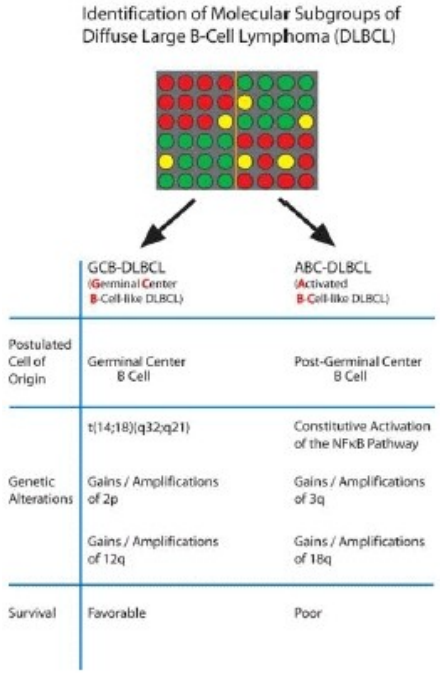
Lum A, Lum B+C, NorB-like, Basal, ERBB2+

© Ondřej Slabý, 2009

3) Klasifikační metody

Jejich podstatou je znalost informace o charakteru vzorku tzn. supervised přístup. Principem klasifikačních metod v analýze dat z DNA čipů je vytvoření rozhodovacího pravidla, ktere by na zakladě naměřených hodnot genové exprese umožňovalo přiřazení pacienta do jedne z předem definovaných tříd (například zdravý, nemocný). Z toho je zřejmé, že by se „dobré“ rozhodovací pravidlo založené na expresních datech mohlo zařadit po bok stavajících diagnostických metod a výrazně tak přispět ke zpřesnění diagnostiky závažných onemocnění (klasifikační stromy, Support Vector Machines (SVM), metoda k-nejbližších sousedů,..)

MOLEKULÁRNÍ KLASIFIKACE NÁDORŮ:
precizní klasifikace je základem léčebného úspěchu, současné metody jsou založeny na morfologii, imunohistochemii, genetice a klinické odpovědi
řada diagnostických nejasností (heterogenita)
ČIPY:
-identifikace nových jednotek na podkladě profilu genové exprese
-reklasifikace stávajících jednotek
-identifikace skupin či jednotlivých genů „markerů“ specifických pro dané jednotky



© Ondřej Slabý, 2009

Predikce metastatického potenciálu u pacientek s časnými stádii mamárního karcinomu

Van't Veer et al. (Nature, 2002)

96 sporadických mamárních karcinomů

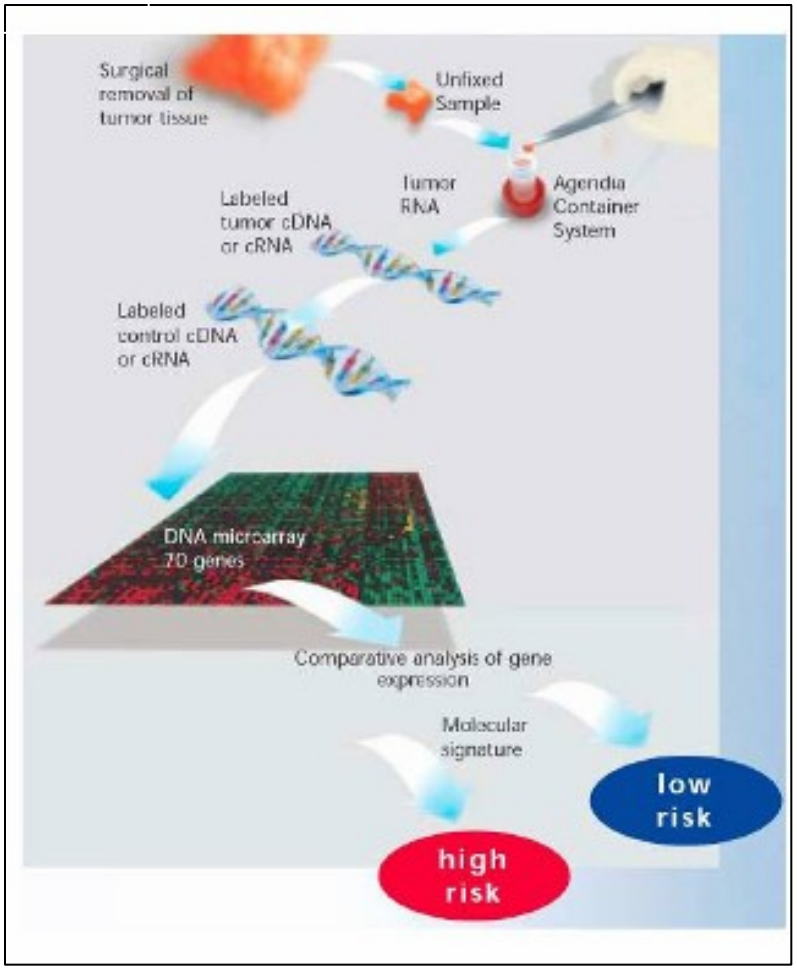
46 pacientek se špatnou prognózou (do 5 let se nevyvinuly vzdálené metastáze)

50 pacientek s dobrou prognózou (do 5 let se nevyvinuly vzdálené metastáze)

5852 genů se signifikantním rozdílem v expresi mezi skupinami

70 genů nejvíce korelujících s klinickým stavem použila pro klasifikaci

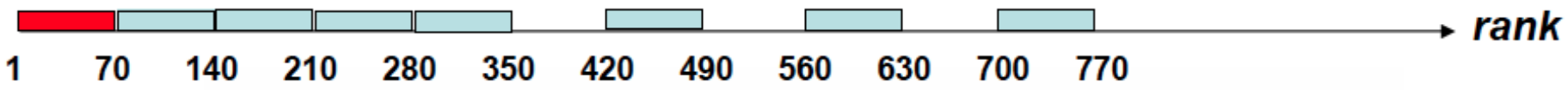
AGENDIA



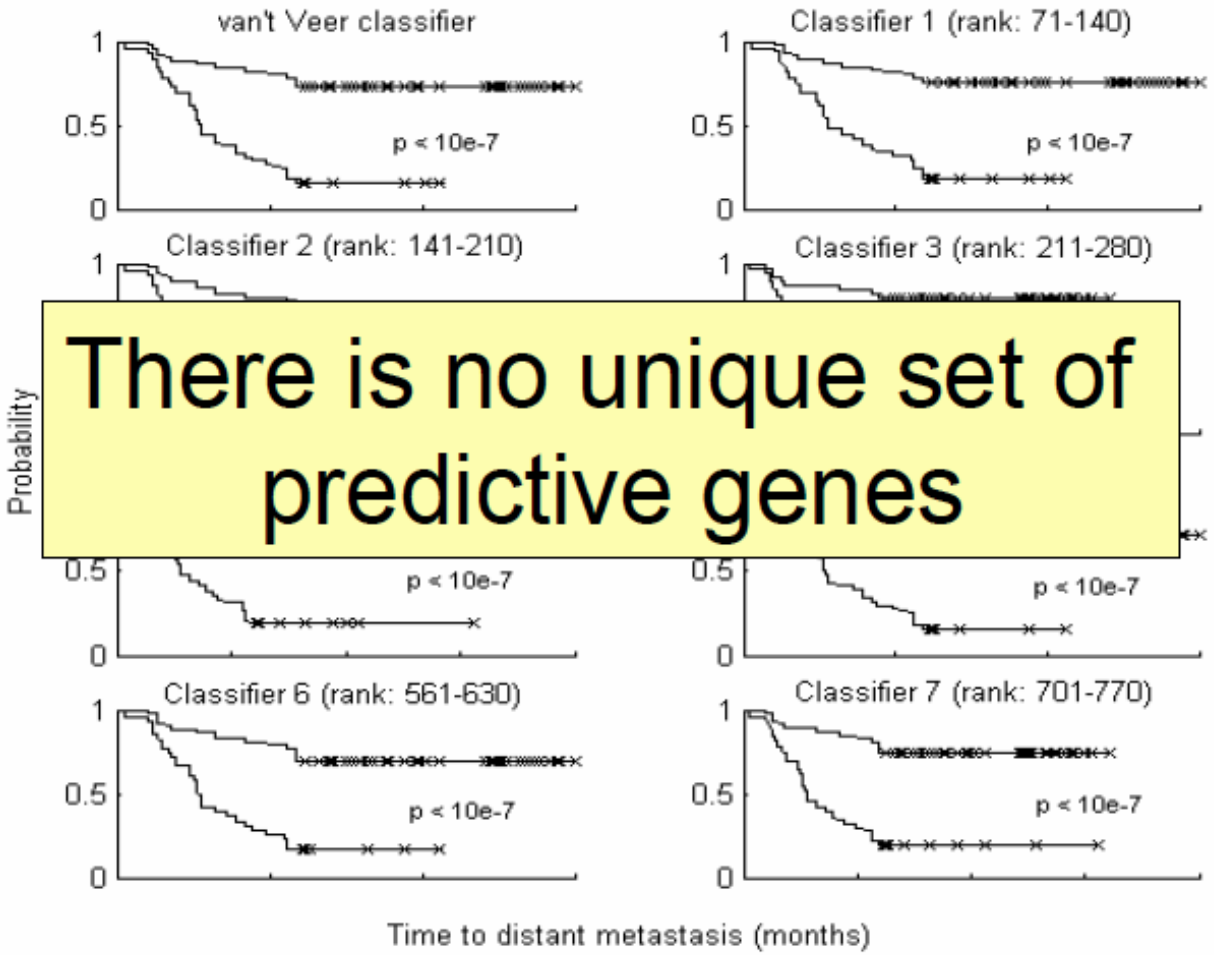
Sada 70 genů - patent

© Ondřej Slabý, 2009

Many sets of 70 genes can be used to predict time to distance metastasis



Van't Veer



There is no unique set of predictive genes

Ein-Dor et al, Bioinformatics 2005

Results for Breast Cancer Data

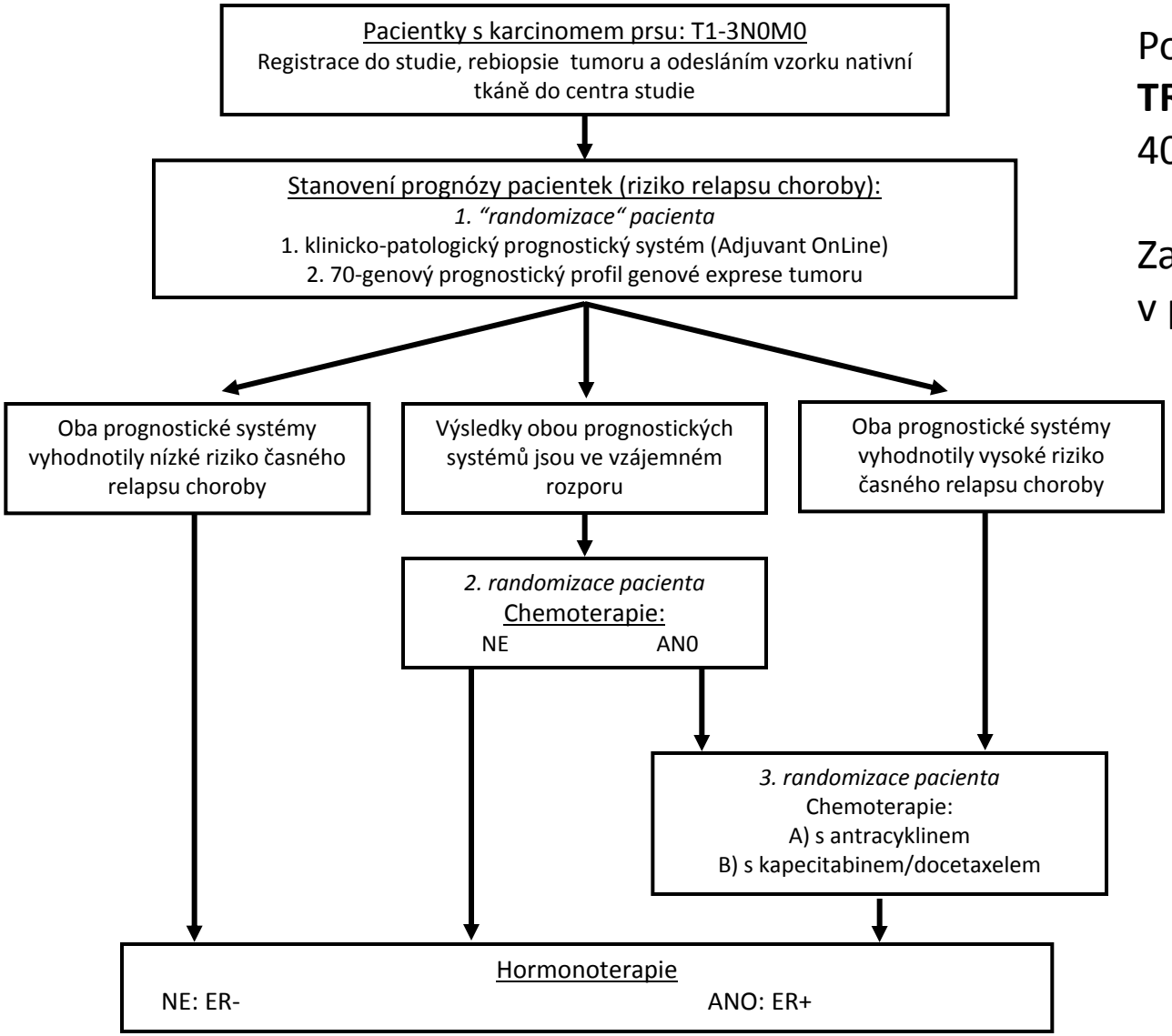
- For a typical overlap of 50% between two lists of 70 genes, more than 2300 patients are needed.

- The expected overlap between van't Veer's list and another list produced from similar experiment is less than 2%

Ein-Dor et al ***PNAS 2006***

PROTOKOL KLINICKÉ STUDIE „MINDACT“

(Microarray In Node negative Disease may Avoid ChemoTherapy)



Rozpočet ~ 20 mil. EUR

Pořádá síť
TRANSBIG
40 institucí 21 zemí

Zařazení 5000 pacientek
v prvních 3 letech

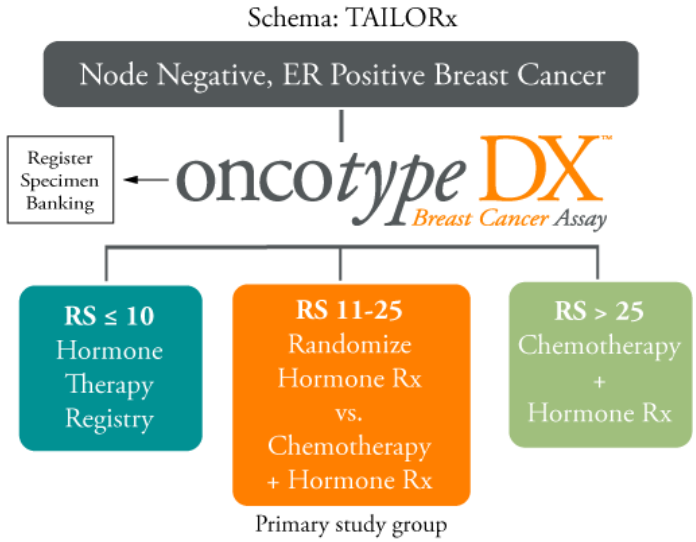
EORTC Protocol 10041 – BIG 3-04 - http://www.eortc.be/services/unit/mindact)

© Ondřej Slabý, 2009

Roche AmpliChip CYP450

- Approved for clinical use in US and EU
- Gene variations of CYP2D6 and CYP2C19
  - Metabolism of ~25% of all prescription drugs
  - Determine phenotype:
    poor, intermediate, extensive, or ultrarapid metabolizer
- Intended to be an aid for physicians
  - Individualized treatment and dosage

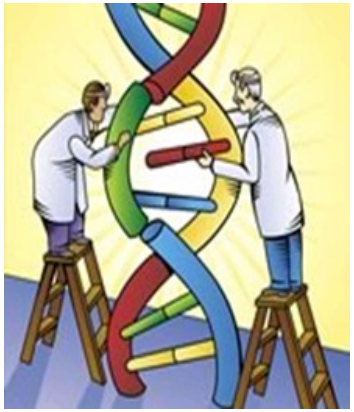
www.roche.com
www.amplichip.us



© Ondřej Slabý, 2009

Take home

Analýza čipových dat – pozadí, normalizace
Analýza čipových dat - identifikace biologicky významných genů
Analýza čipových dat -ukázky multidemenzionálních metod analýzy čipových dat - *Shlukovací analýzy*
Analýza čipových dat – klasifikační metody
Molekulární klasifikace nádorových onemocnění – ukázky
Aplikace čipových technologií do klinické praxe – studie MINDACT, Agendia, Roche AmpliChip CYP450
mikroRNA: nová úroveň regulace genové exprese – biogeneze a biologická funkce
mikroRNA v patogeneze nádorových onemocnění
mikroRNA jako biomarkery (SNP, tkáňové, sérové)
mikroRNA jako terapeutické cíle
mikroRNA čipy

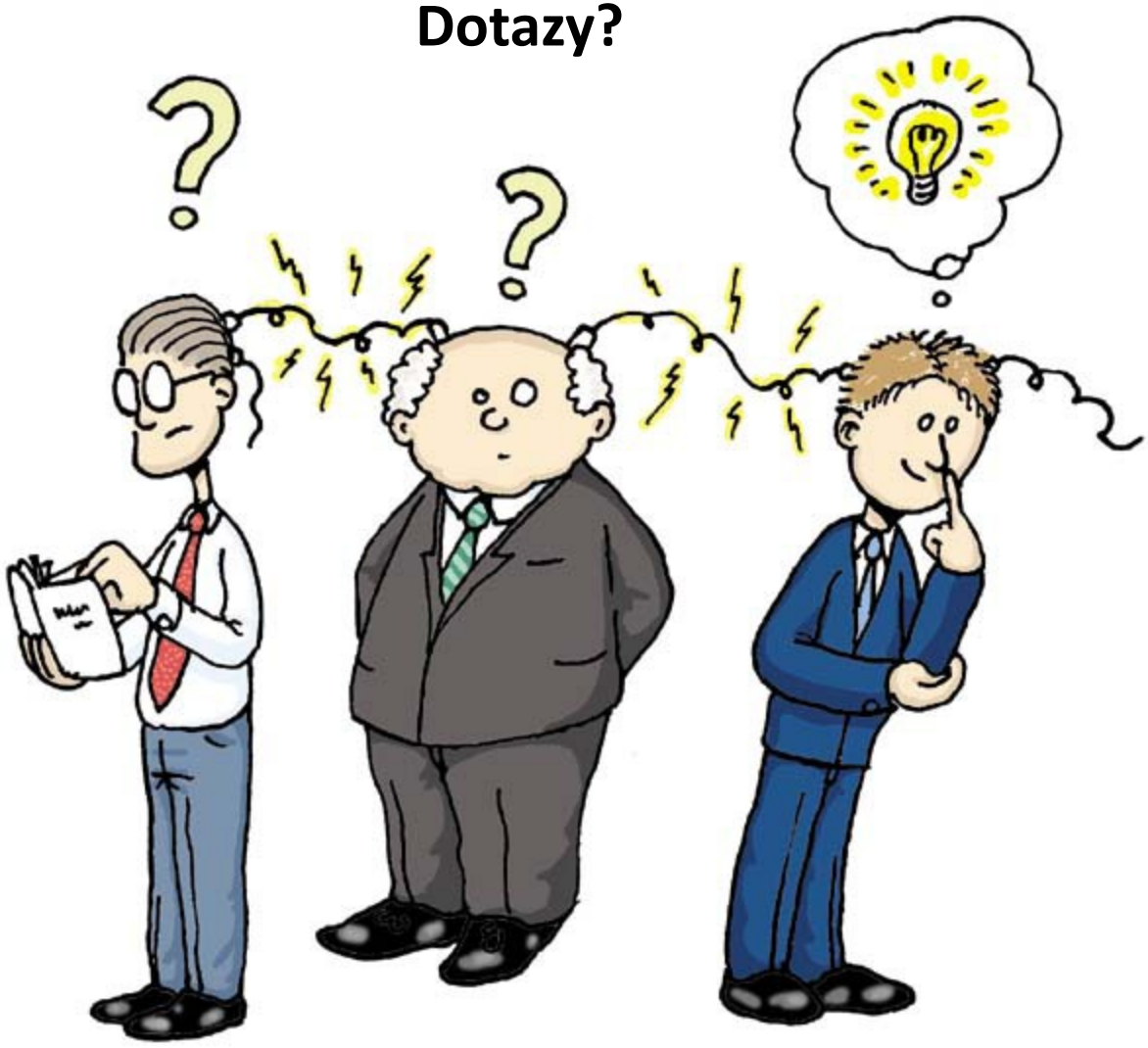


Náplň příští přednášky

Moderní metodické přístupy v molekulární medicíně II – proteomika (dvojrozměrná elektroforéza, hmotnostní spektrometrie, proteinové čipy), využití proteomiky v diagnostice nádorových onemocnění
Molekulární epidemiologie – definice a vymezení oboru, identifikace molekulárních rizikových faktorů vzniku a rozvoje onemocnění, analýza vztahu molekulárních faktorů a vlivů prostředí na rozvoj nádorového onemocnění, význam molekulární epidemiologie u karcinomu plic a kolorektálního karcinomu

© Ondřej Slabý, 2009

Dotazy?



© Ondřej Slabý, 2009