

High-density molekulárne biologické dáta a ich analýza

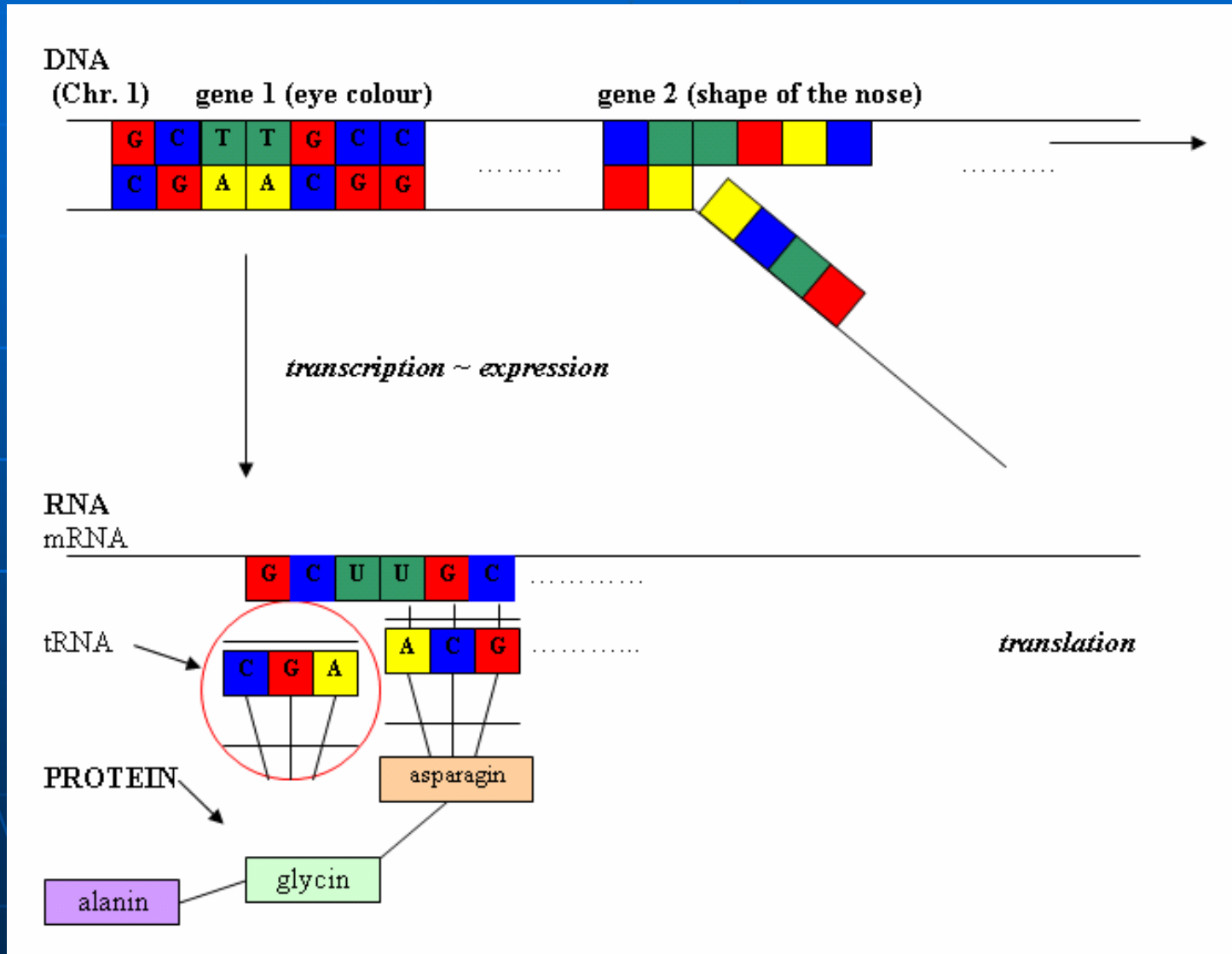
Eva Budinská

Seminár IBA

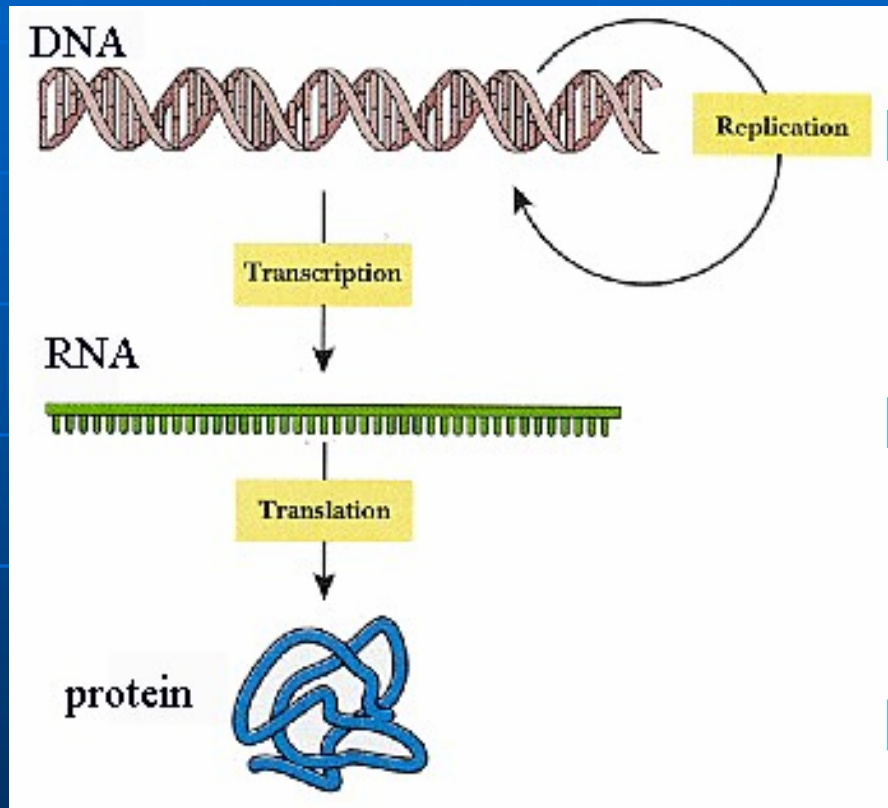
7.10.2008



Molekulárne biologické dáta I.



Molekulárne biologické dáta II.



1. Počet kópií génov

2. Hladina exprese génu
(výskyt RNA)

3. Výskyt, množstvo a štruktúra
proteínov

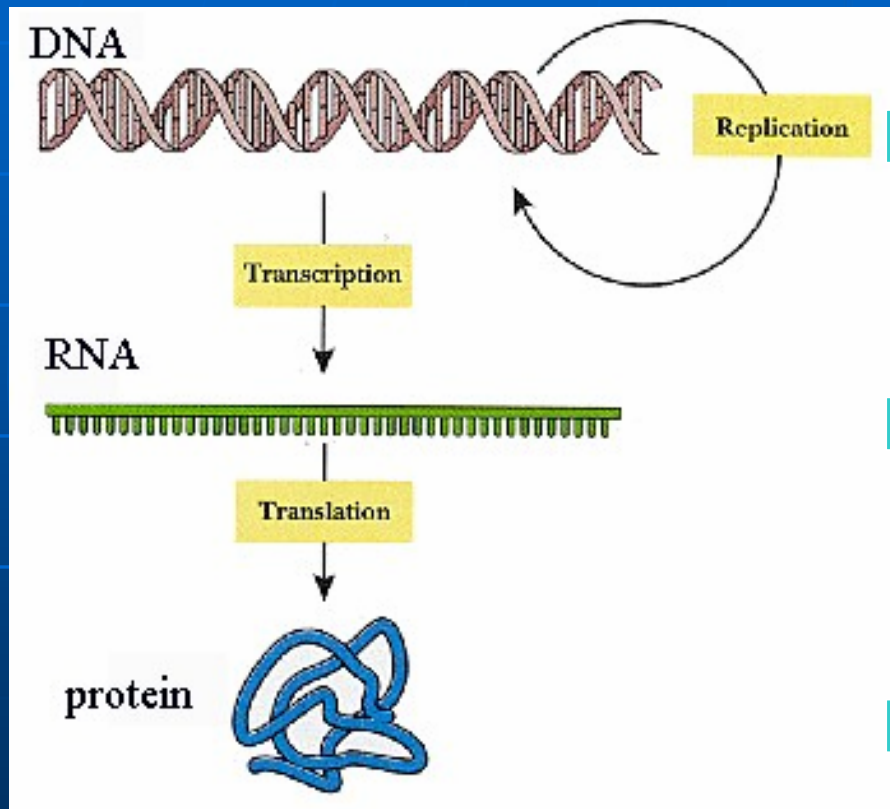
High-density dáta

- stovky až desaťtisíce informácií o jedinej biologickej vzorke (tkanivo pacienta, bunky pletiva rastliny, kmeň baktérie...)

High – density molekulárne biologické dáta I.

- informácie o stovkách až desaťtisícoch štruktúrnych alebo funkčných molekúl na jednu biologickú vzorku (tkanivo pacienta, bunky pletiva rastliny, kmeň baktérie...)
- získavané špeciálnymi technikami

High – density molekulárne biologické dáta II.

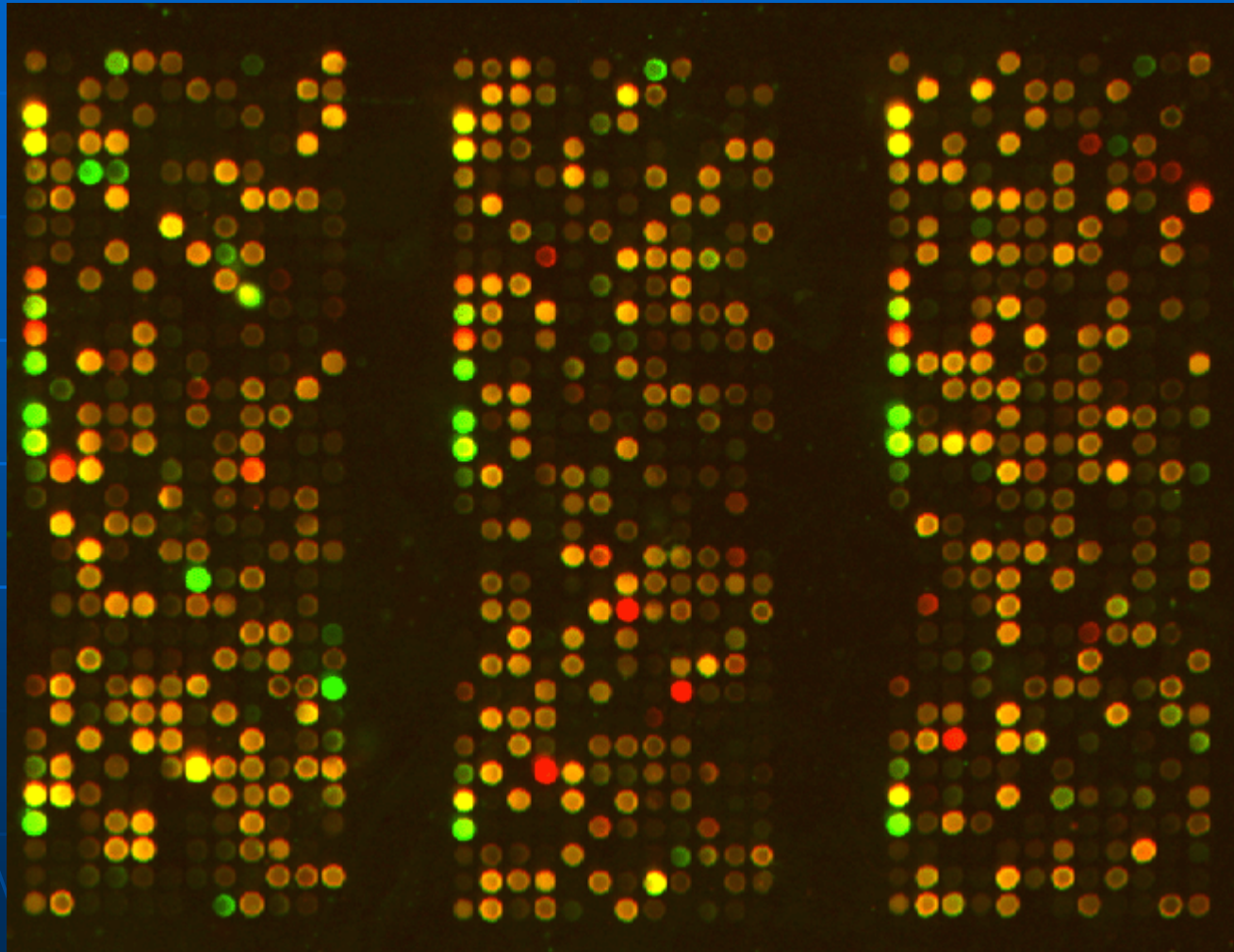


1. Počet kópií génov
(CGH microarraye)

2. Hladina expresie génu
(Expresné microarraye)

3. Výskyt, množstvo a štruktúra
proteínov
(Proteínové microarraye,
SELDI TOF spektrometria ...)

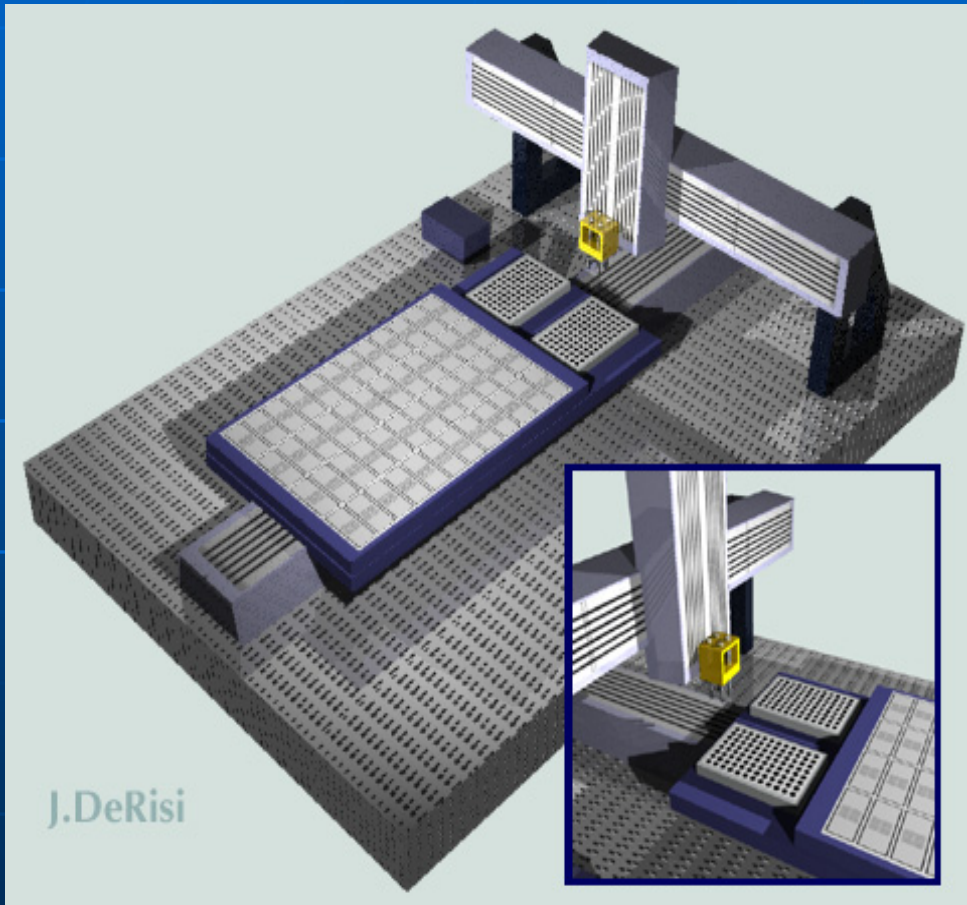
Microarray



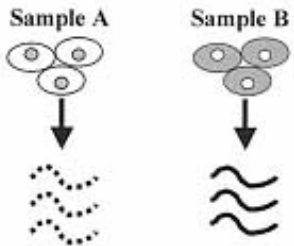
Využitie

1. v diagnostike, epidemiológii, klasifikácii nádorov
 - až 90% ľudských nádorov - genetická nestabilita: zmeny v počte chromozómov, mnohonásobné génové amplifikácie a delecie, zmeny v génovej expresii
 - => klinická rozdielnosť histologicky podobých nádorov
 - potreba komplexnej analýzy nádorov na úrovni DNA, RNA i proteínov
2. štúdie rozlišujúce jednotlivé druhy (baktérie...)
3. konštrukcia fylogenetických stromov

Príprava microarray sklíčka



A. RNA Isolation

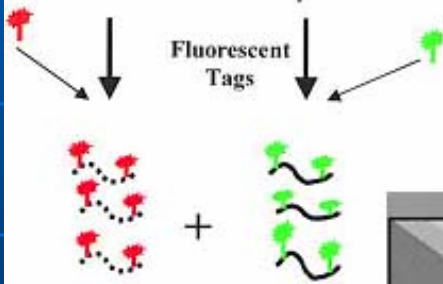


B. cDNA Generation

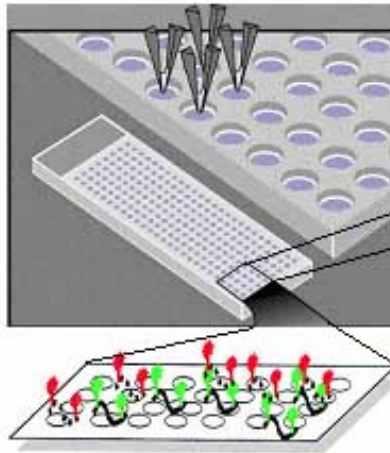
C. Labeling of Probe

Reverse Transcriptase

Fluorescent
Tags

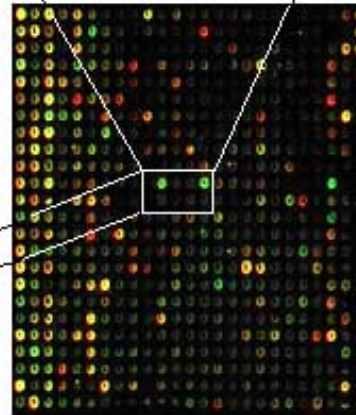
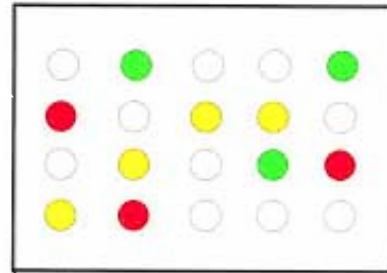


D. Hybridization to Array



E. Imaging

● Sample A > B
● Sample B > A
● Sample A = B



F. Analýza obrazu (snímanie intenzít jednotlivých kanálov)

Dátový súbor:

tisíce riadkov (génov)
X desiatky stĺpcov

- číselné hodnoty intenzít testovanej a referenčnej RNA (+ hodnoty pozadí...)
- kontrola kvality spotov
- ...

Ďalšia analýza

1. úpravy datového súboru
2. určenie odlišných génov
3. klasifikácia, predikcia....

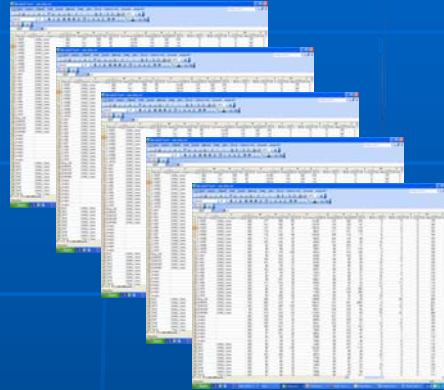
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Unique position	ID	Chromosome	Mb positio	SES end	Plate info	Block	Column	Row	Name	X	Y	Dia.
2	44	RP11-195a8	1	37581779	37726637	NK12C1	26	11	19	44	8600	35890	140
3	44	RP11-195a8	1	37581779	37726637	NK12C1	26	10	19	44	8370	35890	140
4	44	RP11-195a8	1	37581779	37726637	NK12C1	26	12	19	44	8820	35890	140
5	102	RP11-124d4	1	87374825	87558032	NK12B12	4	7	19	102	16600	8970	120
6	102	RP11-124d4	1	87374825	87558032	NK12B12	4	9	19	102	17060	8970	130
7	102	RP11-124d4	1	87374825	87558032	NK12B12	4	8	19	102	16830	8970	120
8	154	RP11-145H4	1	1.52E+08	1.52E+08	NK12G5	26	11	20	154	8600	36110	150
9	154	RP11-145H4	1	1.52E+08	1.52E+08	NK12G5	26	13	20	154	9040	36110	140
10	154	RP11-145H4	1	1.52E+08	1.52E+08	NK12G5	26	12	20	154	8820	36110	150
11	187	RP11-1122M	1	1.83E+08	1.83E+08	NK12F10	20	7	20	187	16690	27120	130
12	187	RP11-1122M	1	1.83E+08	1.83E+08	NK12F10	20	6	20	187	16460	27120	130
13	187	RP11-1122M	1	1.83E+08	1.83E+08	NK12F10	20	5	20	187	16240	27120	130
14	196	RP11-66fB	1	1.89E+08	1.9E+08	NK12C2	18	10	19	196	8330	26880	130
15	196	RP11-66fB	1	1.89E+08	1.9E+08	NK12C2	18	11	19	196	8560	26890	130
16	196	RP11-66fB	1	1.89E+08	1.9E+08	NK12C2	18	12	19	196	8780	26880	130
17	236	RP11-845b6	1	2.27E+08	2.27E+08	NK12C3	10	10	19	236	8330	17960	140
18	236	RP11-845b6	1	2.27E+08	2.27E+08	NK12C3	10	10	19	236	8330	17960	140
19	236	RP11-845b6	1	2.27E+08	2.27E+08	NK12C3	10	12	19	236	8780	17960	150
20	236	RP11-845b6	1	2.27E+08	2.27E+08	NK12C3	10	12	19	236	8780	17960	150
21	236	RP11-845b6	1	2.27E+08	2.27E+08	NK12C3	10	11	19	236	8550	17960	140
22	236	RP11-845b6	1	2.27E+08	2.27E+08	NK12C3	10	11	19	236	8550	17960	140
23	320	RP11-1084a2	2	47485695	47697380	NK11F10	24	7	20	320	16660	31610	130
24	320	RP11-1084a2	2	47485695	47697380	NK11F10	24	6	20	320	16440	31610	130
25	320	RP11-1084a2	2	47485695	47697380	NK11F10	24	5	20	320	16220	31610	130
26	323	RP11-460n15	2	47854784	48034160	NK12H8	4	12	20	323	17720	9190	130
27	323	RP11-460n15	2	47854784	48034160	NK12H8	4	11	20	323	17500	9190	130
28	323	RP11-460n15	2	47854784	48034160	NK12H8	4	13	20	323	17940	9190	130
29	324	RP11-3g11	2	47946940	48102089	NK12H7	12	11	20	324	17540	18150	130
30	324	RP11-3g11	2	47946940	48102089	NK12H7	12	12	20	324	17760	18160	140
31	324	RP11-3g11	2	47946940	48102089	NK12H7	12	13	20	324	17990	18160	140
32	361	RP11-232j18	2	71372264	71537932	NK11F4	8	20	19	361	19530	13430	130
33	361	RP11-232j18	2	71372264	71537932	NK11F4	8	1	20	361	15250	13660	130
34	361	RP11-232i18	2	71372264	71537932	NK11F4	8	19	19	361	19290	13430	130

Príklad

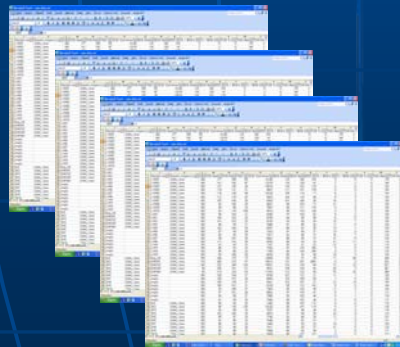
- Golub et al.(1998)
- Cieľ: Nájsť odlišne exprimované gény medzi podtypmi leukémie :
AML a ALL

Finálny dátový súbor

AML skupina



ALL skupina



riadok ~ gén
stĺpec ~ pacient

Microsoft Excel - BreastCancerAll.2

	A	B	C	D	E	F	G	H	I	J
	GeneID	Reporter name	117A	117B	123A	123B	125A	125B	138A	138
1	HSI001A01	HISTI1HC	1,777863621	1,498141614	0,312009636	-0,865553809	0,413427451	1,816147903	0,056904331	-0,0
2	HSI001A02	TAF11	0,256294553	0,363418528	0,379327176	0,311856954	0,419228718	-0,018800963	0,887658927	0,7
3	HSI001A03	NDUFC1	-0,453696857	-0,662449411	-0,205996814	-0,44287382	-0,632914343	-0,660918097	-0,391079926	-0,0
4	HSI001A04	SMCHD1	-0,243293629	0,295755589	-0,32141845	-0,877129231	0,112863779	0,446898061	0,382009142	0,6
5	HSI001A05	CR1L	2,36736488	1,757569614	1,493341934	2,146968064	3,258798183	4,146356701	1,174688618	2,7
6	HSI001A06	RGS2	0,894668283	1,576990652	2,000830136	0,177083704	0,328513017	1,153821628	-0,430994873	0,0
7	HSI001A07	MARF1	0,063204616	-0,251650027	-0,037516441	-0,071927848	-0,088472315	0,714017788	-0,603380403	0,0
8	HSI001A08	RIF1	-1,035395358	-0,597618118	-0,842048231	-0,333608944	-0,757768183	0,107537423	-0,552971777	0,0
9	HSI001A09	PRKAR1A	-0,312906681	0,903133397	-0,292124185	0,083872417	-0,75879236	0,819206247	-0,376241539	0,9
10	HSI001A10	RNF11	0,182274718	1,266396447	-0,03050399	0,096125579	-0,660121977	0,26392072	-0,762314596	0,4
11	HSI001A11	TNRC15	0,277119138	0,369293265	0,219858236	1,055130156	-0,151141992	0,882080858	-0,192741635	0,4
12	HSI001A12	RP11-301H17.1	-0,236985198	-0,2774671	0,062289917	0,547534322	-0,325565709	0,050133068	-0,332141361	0,0
13	HSI001B01	HDHD1A	-1,448739796	-1,427995838	-0,968385615	-1,144525017	-0,66976125	-0,452429773	-1,837768482	-1,4
14	HSI001B02	CD59	-0,181040975	0,024735645	-0,289595941	-0,693527013	-0,387395951	-0,180710213	-1,163419692	-0,7
15	HSI001B04	PIAS4	-1,125275376	-1,136551385	-1,678951039	-1,457463996	-1,457773178	-1,057856186	-1,320349016	-0,6
16	HSI001B05	OLFM1	1,337913142	-0,04331583	-0,998851046	-1,162263	-0,515078134	-1,402104824	-1,630167503	-1,7
17	HSI001B06	CACNA1I	-0,035725334	-0,192057838	-0,371683837	-0,774231543	-0,548358323	-0,539405912	0,003271304	-0,0
18	HSI001B07	GGA2	0,0639877235	-0,347913196	0,086640297	-0,058059331	0,08367469	0,33003719	-0,419614154	-0,0
19	HSI001B09	R3HDM1	-1,283873947	-0,51229348	-0,908368055	-0,591837637	-0,959277524	-0,580718778	-1,143809272	-0,9
20	HSI001B10	CLSPN	0,147019804	-0,178245198	-0,276177744	-1,220611452	-0,269190553	-1,737246983	-0,574205651	-1,1
21	HSI001B11	CALM3	-0,5371439	-1,19151506	-1,336076572	-1,492489971	-1,057042205	-2,599276331	-0,996461191	-1,7
22	HSI001B12	TRRAP	-0,945438274	-0,80570667	-0,902179278	-0,213094467	-0,778939642	-0,353523952	-0,131579844	0,3
23	HSI001C02	SFRS3	-0,995336771	0,88339825	-0,508137264	-0,777766453	-0,868692945	-0,618775625	-0,796440769	-0,4
24	HSI001C03	BAT2	-0,090094673	-1,994727357	-1,137113265	-0,708524595	-1,020030635	-2,568699984	-0,794103649	-1,0
25	HSI001C04	TRAF3IP2	-1,11213323	-0,737149218	-1,703117746	-1,83789235	-1,542585897	-1,105920749	-1,317314617	-1,0
26	HSI001C05	LAPTM5	2,791649297	1,962636109	1,706590717	2,852623068	3,580434805	4,301670566	1,564435766	2,8
27	HSI001C06	KIAA0146	-0,127196236	-0,472680155	-0,318892698	0,14108207	-0,598182089	-0,237611168	-0,204982088	0,0
28	HSI001C07	BLMH	-0,561548851	-0,868182319	-0,080542952	-0,218454597	-0,476097166	0,292222795	-0,734977724	-0,4

Analýza of microarray experimentu



Prínos matematického biológa

- pochopenie biologickej podstaty veci i štatistických postupov a metód
- konzultácie lekárom a biológom k postupu pri designe experimentu
- tvorba či zlepšovanie algoritmov a metód pre analýzu týchto dát
- praktická analýza dát pre
 - klinické účely
 - účely experimentálne

Konkrétne I.

- vývin metódy pre analýzu zašumených CGH arrayí
 - Eva Budinská (2008) Bio-mathematical modeling of tumor genome profiling data. Dissertation thesis. Masaryk University, Brno, 118 pp (in English)
 - Budinska,E., Gelnarova,E. & Schimek,M.G. (2008) MSMAD: A computationally efficient method for the analysis of noisy array CGH data. Under review
- vývin metódy pre simultánnu analýzu viacerých CGH arrayí
 - Picard,F., Lebarbier,E., Budinska,E., Radvanyi,F. and Robin,S. (2008) A statistical framework for the joint segmentation of multiple array CGH profiles. Under review.
- vývin metodiky pre meta-analýzu dát z microarray experimentov
- praktická analýza dát pre
 - klinické účely
 - účely experimentálne

Konkrétne II.

- vývin metodiky pre meta-analýzu dát z microarray experimentov
- praktická analýza dát pre
 - klinické účely
 - účely experimentálne
 - Matiasovicova,J, Adams,P., Barrow,P.A., Hradecka,H., Malcova,M., Karpiskova,R., Budinska,E., Pilousova,L. (2007) Identification of putative ancestors of the multidrug-resistant Salmonella enterica serovar typhimurium DT104 clone harboring the Salmonella genomic island 1. Archives of microbiology, 187, 415-424.
 - Šebková,A., Karasová,D., Crhánová,M., Budinská,E., Rychlík,I. (2008) aro mutations in Salmonella enterica cause defects in cell wall and outer membrane integrity. Journal of Bacteriology, American Society for Microbiology, May 2008, 190, 3155-3160.
 - Brozkova,K, Budinska,E., Bouchal,P., Hernychova,L., Knoflickova,D., Valik,D., Vyzula,R., Vojtesek,B., Nenutil,R.(2008) Surface-enhanced laser desorption/ionization time-of-flight proteomic profiling of breast carcinomas identifies clinicopathologically relevant groups of patients similar to previously defined clusters from cDNA expression. Breast Cancer Research, 10:R48
 - Vranová,V., Nečesalová,E., Kuglík,P., Cejpek,P., Pešáková,M., Budinská,E., Relichová,J., Veselská,R. (2007) Screening of genomic imbalances in glioblastoma multiforme using high-resolution comparative genomic hybridization. Oncology Reports, 17, 457-464.