

VYBRANÉ APLIKACE TECHNOLOGIE cDNA MICROARRAYS V ONKOLOGICKÉM VÝZKUMU

SELECTED APPLICATIONS OF cDNA MICROARRAYS TECHNOLOGY IN ONCOLOGICAL RESEARCH

KOUTNÁ I., KRONTORÁD P., SVOBODA Z., KOZUBEK M.

CENTRUM ANALÝZY BIOMEDICÍNSKÉHO OBRAZU, FAKULTA INFORMATIKY MU, BRNO

Souhrn

V této publikaci přinášíme přehled vědeckých prací, ve kterých nám technologie cDNA microarrays umožnila nové náhledy do procesů maligní transformace. Díky této technologii dnes můžeme sledovat v jeden okamžik expresní aktivitu tisíců genů. S pomocí vhodných nástrojů můžeme naměřené hodnoty dávat do souvislostí, které byly ještě nedávno prakticky nemyslitelné. V naší laboratoři jsme vyvinuli nové nástroje a postupy pro analýzu microarrays a pro zpracování výsledků do podoby použitelné v onkologickém výzkumu. Pro sledování některých charakteristik je nutné použít vizualizační nástroje. Například pro sledování aktivity určitých oblastí chromosomů jsou velmi vhodné transkripční mapy (dále TM). Implementovali jsme vlastní software pro generování TM a díky kombinaci jeho výstupů s dosavadními znalostmi jsme formulovali závěry našich experimentů. Intuitivním rozšířením TM jsme dosáhli mnoha dalších vizualizací např. hustoty aktivních genů, kumulativní exprese oblastí atp. Tyto nástroje by měly v budoucnosti pomáhat klinikům při diagnostice zhoubných malignit a určování následné léčby.

Klíčová slova: microarrays, leukémie, karcinom tlustého střeva, ex vivo diferenciace

Summary

In this paper we bring a scientific review of studies in which cDNA microarrays technology allowed us to create new insights into malignant transformation processes. We can monitor parallel expression activity of thousands of genes using this technology. Using appropriate tools we can evaluate biological correlations that were practically impossible few years ago. We have developed new tools and techniques transforming the data into form suitable for oncological research. Special visualizations are necessary for some types of studies. E.g. transcription maps (TMs) are a proper tool for studying activity of chromosomal regions. We have implemented our own software for generating TMs. On the basis of combination of its outputs and knowledge gained so far we have formulated our conclusions. Using intuitive TM extensions we have obtained many useful visualizations, e.g. density of active genes, cumulative expression etc. These tools are supposed to help clinicians while forming diagnoses of malignant diseases and while treatment planning.

Keywords: microarrays, leukemia, colorectal carcinoma, ex vivo differentiation

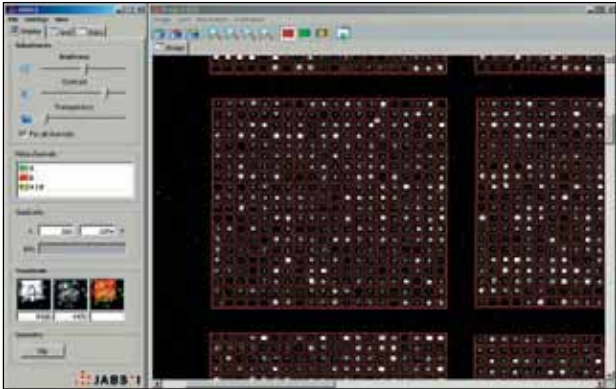
Materiál a metody

Na základě zkušeností s různými druhy microarrays jsme se pro naše experimenty rozhodli používat skla SS-H19K (Clinical Genomic Centre GCC, Toronto, Ontario, Canada). Tato skla nesou celkem 19008 EST. Takto vysoký počet EST nám umožňuje sledovat aktivitu téměř celého genomu v širokém spektru typů vzorků: vzorky odebrané z leukemických pacientů, nádory tlustého střeva atp. Náš výzkum se primárně orientuje na sledování změn genové exprese během maligních transformací a diferenciací [1 – 4]. Pracujeme s různými druhy stabilizovaných buněčných linií a se vzorky odebranými z pacientů. Odebíráme např. nádorovou tkáň tlustého střeva, kostní dřeň a krev leukemických pacientů. Experimenty využívající cDNA microarrays sestávají z následujících kroků:

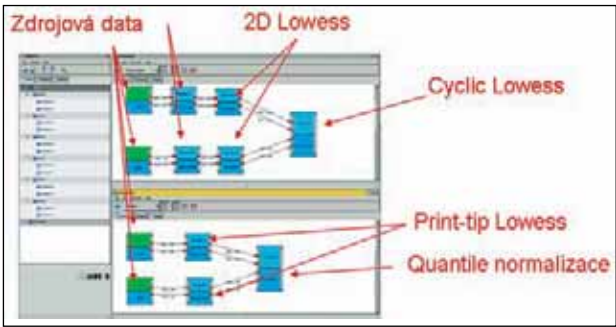
- izolace a purifikace RNA, amplifikace
- reverzní transkripce
- hybridizace
- skenování
- obrazová analýza, normalizace a statistické vyhodnocení.

Obraz pořízený skenováním sklíček analyzujeme pomocí vlastního software, který jsme v rámci výzkumu orientovaném na analýzu obrazu vyvinuli tak, aby co nejlépe vyhovoval našim potřebám. Mřížku nastavujeme poloautomaticky a provádíme velmi pečlivou kontrolu adresace (obr. 1). Další kroky jsou již plně automatické, volíme pouze použité metody. Náš software obsahuje velmi moderní segmentační a QC algoritmy kopírující vývoj v této oblasti [5]. Na výstup pro každý spot zapisujeme průměrnou intenzitu popředí, odhad intenzity lokálního pozadí a měření kvality obrazového vzorku. Numerické hodnoty získané pomocí našeho software dále normalizujeme opět ve vlastním programu. Implementovali jsme množství moderních normalizačních algoritmů a přehledné nástroje vizualizace tak, abychom mohli s ohledem na design experimentu co nejlépe transformovat vstupní data. Náš software obsahuje within-slide i between-slide normalizační metody (obr. 2) a umožňuje celý normalizační proces zapouzdřit do velmi přehledného a kompaktního prostředí. Pokud porovnáme hodnoty pouze jednotlivých skel, používáme nejčastěji LOWESS [6] within-slide normalizaci mezi jednotlivými

subgridy, která velmi kvalitně vyrovnává rozdíly v intenzitách způsobené především různým zabudováváním používaných fluorescenčních barviv. Některé experimenty vyžadují porovnávání hodnot z několika skel mezi sebou. V těchto případech volíme nejčastěji dvoufázovou normalizaci. Nejprve jednotlivá skla normalizujeme pomocí vhodné within-slide normalizace a poté aplikujeme between-slide normalizaci, nejčastěji metodu Quantile [7]. Normalizované hodnoty dále podrobujeme statistické analýze pro odhalení význačných genů nebo oblastí. Pro specifické aplikace dále implementujeme vizualizační nástroje, které nám přehledně zobrazují pozorované charakteristiky.



Obrázek 1.: Ukázka našeho software pro analýzu obrazu JABS.I. Na obrázku je vidět prostředí s definovanou mřížkou. Pro další informace o tomto software nás můžete kontaktovat.



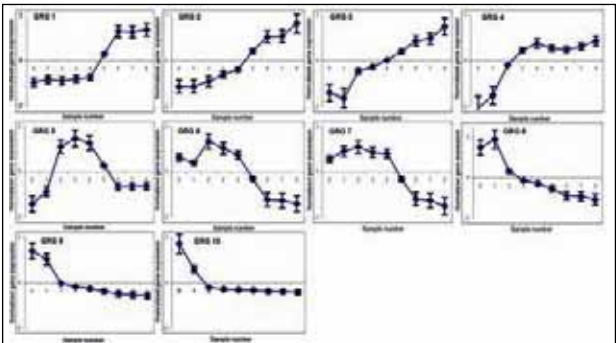
Obrázek 2.: Příklad definice normalizačního výpočtu v našem software JABS.N. Uživatel definuje výpočet graficky, pomocí schématu. Výpočet je tak velmi přehledně znázorněn.

Výzkum v oblasti leukemických onemocnění

Rádi bychom demonstrovali vhodnost použití technologie cDNA microarrays pro sledování časově závislých procesů, transkripční aktivity, diferenciálních a jiných transformací. Za tímto účelem popíšeme naše experimenty spojené s analýzou diferenciálních procesů. Studujeme kinetiku a rozdíly mezi vybranými diferenciacemi. V oblasti stabilizovaných buněčných linií provádíme intenzivní výzkum diferenciálních procesů u HL-60 a K562 buněk a také na buňkách CD34+ [8, 9].

Kinetické studie

Pro hledání rozdílů v časovém průběhu exprese jednotlivých genů používáme výše zmíněná vysokohustotní microarrays skla. Pro sledování změn genové exprese nám sloužily 3 modelové diferenciace: monocytární a granulocytární diferenciace HL-60 a megakaryocytární diferenciace buněk K562 [10 – 13]. Na základě získaných microarrays dat jsme rozdělili geny do 10 skupin (GRG) pro každý druh diferenciace. Kriteřiem byla časová kinetika změn genové exprese. Tyto skupiny jsme seřadili od 1 do 10 podle charakteristického průběhu podle následujícího klíče (obr. 3)



Obrázek 3.: Charakteristické časové průběhy genové exprese nalezené shlukovou analýzou naměřených hodnot pro všechny 3 typy diferenciace.

- rychlá up-regulace
- pomalá up-regulace
- up-regulace následovaná down-regulací
- pomalá down-regulace
- rychlá down-regulace

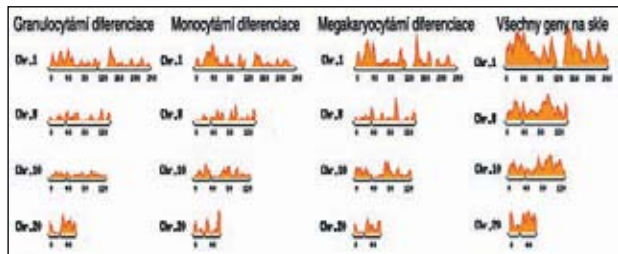
Počty genů v GRG se pohybovaly okolo 100. Za účelem porovnání těchto skupin mezi různými druhy diferenciace jsme vypočítali jejich průniky a zanesli je do tabulek. V nich můžeme pozorovat jistý řád poukazující na nenáhodné rozložení genů ve skupinách. Pro větší přehlednost jsme každé políčko obarvili tím červeněji čím více genů patří do příslušného průniku. Všimneme si červených „ostrůvků“ znamenajících velký průnik vzhledem k podobnému časovému průběhu. Diagonální políčka nám potvrzují existenci oblastí aktivních během všech druhů diferenciace (tab.1).

Megakaryocytární diferenciace										
Shluky	1	2	3	4	5	6	7	8	9	10
1	29	5	13	19	32	3	33	13	17	4
2	18	51	23	46	47	1	5	46	24	1
3	10	35	27	4	76	20	4	9	61	2
4	19	3	1	55	11	1	48	19	1	3
5	26	73	43		137	11	33	71	48	11
6	0	1	10	0	3	29	1	2	3	0
7	47	2	13	22	58	1	47	2	16	5
8	32	20	7	56	21	1	47	41	9	13
9	22	69	73	26	110	10	10		115	6
10	4	3	2	3	5	0	4	5	13	7

Tabulka 1.: Tabulka průniků jednotlivých shluků (1-10 GRG) pro megakaryocytární a monocytární diferenciace. Takovéto tabulky byly vygenerovány pro všechny kombinace diferenciálních studií. Intenzita červené barvy napomáhá k lepší přehlednosti tabulky.

Na druhou stranu červené ostrůvky ležící mimo diagonálu jsou dokladem existence oblastí, které jsou sice aktivní během různých diferenciací ale jejich aktivita se v čase pro různé diferenciace liší. Implementovali jsme vizualizační nástroj, který zobrazuje výpočet relativní hustoty exprese (weighted area regulation, WAR) vzhledem k pozici na chromosomu (obr. 4) a nástroj barevně odlišující geny jednotlivých deseti skupin. Díky němu jsme měli možnost sledovat shlukování regulovaných genů uvnitř RIDGE (regions of increased gene expression) oblastí a periodické změny kinetiky genové regulace. Získali jsme tak zcela nový pohled na dynamiku genové regulace a změnu struktury chromatinu. Pokud se zaměříme na oblasti hustě obsazené regulovanými geny, můžeme pozorovat častý jev periodického opakování genů z oblastí 1-10 na chromosomech (postupný nástup). Detailní zpracování hodnot a grafů WAR potvrdilo, že regulované geny se na chromosomech nevyskytují náhodně. Například shluky genů regulovaných v granulocytech a monocytů nacházíme na prvních 10Mbp chromosomu 7. Oproti tomu v této oblasti není žádný shluk u megakaryocytů. V oblasti okolo 90Mbp můžeme na chromosomu 8 vidět 35% genů regulovaných v megakaryocytech oproti 10% v granulocytech. Hus-

tota výskytu regulovaných genů obecně nekoreluje s hustotou genů danou obsazením sklíčka. Například pouze 5 genů ze 60 je regulovaných u monocytů v oblasti okolo 30Mbp chromosomu 20. Výskyt regulovaných genů v různých RIDGE [14, 15] oblastech není rovnoměrný a jednotlivé diferenciální procesy mají své individuální aktivní podoblasti uvnitř RIDGE.



Obrázek 4.: Hustota rozložení regulovaných genů pro různé druhy diferenciace. Díky této vizualizaci můžeme sledovat oblasti relativně (vzhledem k celkovému počtu genů v oblasti) bohaté na regulované geny a současně i s vysokým počtem (absolutně) genů. V prvním až třetím sloupci jsou data spojená s diferenciálními procesy, čtvrtý sloupec vyjadřuje hodnoty pro celé sklíčko (počty zkoumaných genů v daných oblastech).

Naše výsledky potvrdily [16], že nejsou pouze oblasti se striktně zvýšenou nebo sníženou hladinou exprese, ale že na některých polohách chromosomů existují oblasti specifické pro určité buněčné procesy. Výsledky poukazující na poziční shlukování genů během různých druhů diferenciálního procesu představují velký potenciál pro budoucí techniky diagnostického onkologického onemocnění.

Hledání reziduálních rozdílů v buněčných populacích

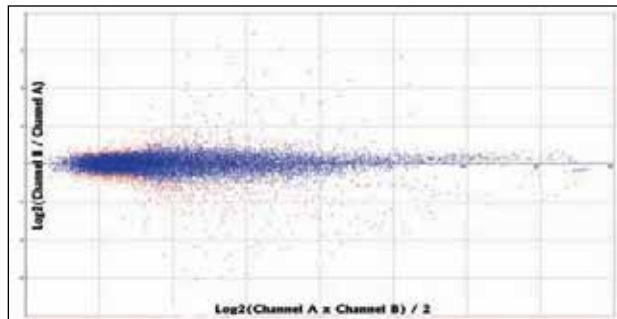
Velmi zajímavé z hlediska uplatnění microarrays v klinické diagnostice se také jeví studie hledající minimální resp. maximální shody dvou zkoumaných vzorků. V práci věnované tomuto problému [9] jsme porovnávali expresní profily CD34+ pozitivních buněk, selektovaných z periferní krve stimulovaných pacientů s non-Hodgkinovým lymfomem, dvěma způsoby selekce (CD34+ a Lin-). Buňky jsme porovnávali během ex vivo [17] diferenciace do granulocytů. Naším cílem bylo nalezení stupně shody dvou odlišně získaných vzorků. Prokázali jsme, že obě dvě selekční metody poskytují srovnatelné buněčné populace, které nemění svůj expresní status ani během dvoutýdenní ex vivo diferenciace do granulocytů.

Pro tuto studii jsme vyvinuli vlastní microarrays metodiku porovnávání vzorků. Jednoduché porovnávání hodnot je příliš zatížené systematickou chybou, především v oblasti nižších intenzit (variance se více projeví). Metoda využívající z-score [18, 19] je navržena tak, aby výsledek nebyl tímto faktem příliš ovlivněn. Námi navržená metodika se sestává nejprve z within-slide normalizace na každém sklíčku. Hodnoty ze vzájemně porovnávaných sklíčků dále normalizujeme between-slide normalizací. Poté je možné jejich hodnoty porovnávat. Z-score metodou nalezneme rozdílné hodnoty a na výstup vracíme procento zastoupení rozdílných hodnot, který reprezentuje rozdílnost (resp. shodu) vzorků aplikovaných na porovnávaných sklech (obr. 5).

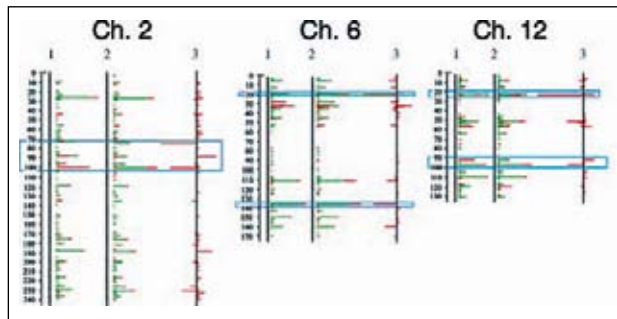
Výzkum v oblasti nádoru tlustého střeva

Experimenty zabývající se nádory tlustého střeva vedeme pomocí porovnávání zdravé a nádorové tkáně [20]. Detekujeme geny s významným rozdílem v expresi během různých stádií onemocnění. Naše cíle jsou jak v identifikaci důležitých genů, tak ve vyvíjení moderního diagnostického přístupu využívajícího srovnávací TM. Identifikovali jsme 195 důležitých genů s významným rozdílem (164 down a 31 up-regulovaných) v expresi mezi zdravou a nádorovou tkání. Pacienty jsme na základě našich studií rozdělili do dvou skupin podle výskytu metastáz v regionálních uzlinách. Provedli jsme funkční analýzu [21] identifikovaných genů. Ta rozdělila geny do kategorií dle jejich funkce a významu. Srovnávací TM nám ukázaly některé společné regiony u různých pacientů a také

regiony, které se mezi pacienty lišily [22]. Tento přístup umožňuje relativně detailní analýzu tkáně nádoru tlustého střeva. Na jejím základě je možné od sebe odlišovat pacienty s různými stádii Duke podle společných a individuálních regionů ve srovnávacích mapách. Výsledky microarrays analýz přenesené do podoby TM jsme studovali s cílem vyjasnit poziční vztahy mezi regulovanými geny. Tyto vztahy mohou objasnit vliv remodelace chromatinu na genovou expresi. Uvedené výsledky mohou doplňovat současnou diagnostiku a pomoci při plánování léčby [23] (obr. 6).



Obrázek 5.: Ukázka MA grafu pro porovnávání dvou buněčných populací (v tomto případě populace buněk získaných pomocí CD34+ a Lin-selekce na začátku ex vivo diferenciace). Modré tečky reprezentují geny s velmi malým rozdílem v expresi, červené tečky naopak geny s významnou odchylkou. Shoda pro tento graf činí 98,3%.



Obrázek 6.: Transkripční mapy jako diagnostický nástroj aplikovaný na data pacientů. V prvním sloupci je naměřená genová exprese pacientů bez metastáz, ve druhém sloupci jsou data pacientů s metastázami. Třetí sloupec reprezentuje rozdíl mezi prvním a druhým sloupcem.

Závěr

Technologie cDNA microarrays se nám osvědčila jako velmi silný nástroj při výzkumu v oblasti onkologických onemocnění a to především na hledání nových vztahů ve změně vyjádření genů, či rodin genů v průběhu maligní transformace. Díky velkému počtu současně zkoumaných genů jsou studie prováděné za pomoci této technologie velmi robustní a málo zatížené výběrovou chybou. Za pomoci stejných druhů sklů je možné sledovat široké spektrum jevů. V této publikaci jsme např. zmínili studie zabývající se leukemickými buňkami a buňkami tkáně tlustého střeva. Díky interdisciplinární povaze naší skupiny jsme vyvinuli nástroje, které nám transformovaly data do přehledné formy. Příkladem jsou transkripční mapy, které dávají do souvislosti genovou expresi s pozicí genu na chromosomu. Také nástroje funkční analýzy microarrays dat poskytují čím dál sofistikovanější výsledky umožňující propojit změny v expresním statusu s jednotlivými metabolickými dráhami [24]. Podobné nástroje je možné vytvářet i pro další studie a tím ještě více rozšířit užitečnost cDNA microarrays.

Poděkování

Tato práce byla podpořena grantem Grantové agentury České Republiky GAČR (301/04/P136) a grantem interní grantové agentury ministerstva zdravotnictví IGA MZ (1A/8241-3).

Literatura

1. Cremer, T., Cremer, C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet.* 2: 292-301.
2. Kozubek, S., Lukášová, E., Jirsová, P., Koutná, I., Kozubek, M., Ganová, A., Bártová, E., Falk, M., Paseková, R. 2002. 3D Structure of the human genome: order in randomness *Chromosoma*. Volume 111, Issue 5, Dec 2002:321 – 331.
3. Lukášová, E., Kozubek, S., Kozubek, M., Falk, M., Amrichova, J. 2002. The 3D structure of human chromosomes in cell nuclei. *Chromosome Res.* 10(7):535-48.
4. E. Bártová, P. Jirsová, M. Fojtová, K. Souček, S. Kozubek. 2003. Chromosomal territory segmentation in apoptotic cells. *Cellular and Molecular Life Science*. Volume 60, Number 5, pages: 979 – 990.
5. MacAulay, C., Palcic, B. 1988. A comparison of some quick and simple threshold selection methods for stained cells. *Analyt. Quant. Cytol. Histol.* 10:134-138.
6. Yang, Y.H. and Thorne, N. 2003. Normalization for Two-color cDNA Microarray Data. *Science and Statistics. A Festschrift for Terry Speed, D. Goldstein (eds.)*, IMS Lecture Notes, Monograph Series, Vol 40, pp. 403-418.
7. Bolstad, B.M., Irizarry, R.A., strand, M. and Speed, T.P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. Jan 22;19(2):185-93.
8. Van Epps DE, Bender J, Lee W, et al. Harvesting, Characterization, and Culture of CD34+ Cells from Human Bone Marrow, Peripheral Blood, and Cord Blood. *Blood Cells* 1994;20: 411- 423.
9. Koutná I. and Klabusay M., Kohutova V., Krontorád P., Svoboda Z., Kozubek M., Mayer J. Evaluation of CD34+ and Lin- selected cells from peripheral blood stem cell grafts of patients with lymphoma during differentiation in culture ex vivo using a DNA microarray technique. *Experimental Hematology*. 2006 Jul;34(7):832-40.
10. Lee, K.H., Chang, M.Y., Ahn, J.I., Yu, D.H., Jung, S.S., Choi, J.H., Noh, Y.H., Lee, Y.S., Ahn, M.J. 2002. Differential gene expression in retinoic acid-induced differentiation of acute promyelocytic leukemia cells, NB4 and HL-60 cells. *Biochem Biophys Res Commun*. Sep; 296(5):1125-33.
11. Bartova E, Kozubek S, Jirsova P, Kozubek M, Gajova H, Lukasova E, Skalníková M, Ganova A, Koutna I, Hausmann M (2002) Nuclear topography and gene activity in human differentiated cells. *J Struct Biol* 139:76–89.
12. Depraetere, S., Joniau, M. 1994. Polar agents with differentiation-inducing capacity prime myelomonocytic cell lines to lipopolysaccharide-induced cytotoxicity: the role of endogenous tumor necrosis factor. *Leukemia*. Nov;8(11):1951-9.
13. Depraetere, S., Vanhaesebroeck, B., Fiers, W., Willems, J., Joniau, M. 1995. Polar agents with differentiation inducing capacity potentiate tumor necrosis factor-mediated cytotoxicity in human myeloid cell lines. *J Leukoc Biol.* Jan;57(1):141-51.
14. Caron, H., Schaik, van B., Mee, van der M., Baas, F., Riggins, G., Sluis, van P., Hermus, M.C., Asperen, van R., Boon, K., Voute, P.A., Heisterkamp, S., Kampen, van A., Versteeg, R. 2001. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science*. 291:1289-1292.
15. Versteeg, R., Schik, B.D.C., Batenburg, M.F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H.J., Kampen, A.H.C. van (2003) The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res* 13: 1998-2004 .
16. Koutná, I. , Krontorád, P., Svoboda, Z., Bartová E., Kozubek, M., Kozubek, S. 2006. New insights into gene positional clustering and its properties supported by large-scale analysis of various differentiation pathways. *Genomics* (2006), doi:10.1016/j.ygeno.2006.07.013.
17. Sato N, Sawada K, Koizumi K, et al. In Vitro Expansion of Human Peripheral Blood CD34+ Cells. *Blood* 1993;82(12): 3600-3609.
18. Quackenbush J. Microarray data normalization and transformation. *Nat. Genet. Suppl.* 2002;32: 496-501.
19. Cheadle C, Vawater MP, Freed WJ, Becker KG. Analysis of microarray data using Z score transformation. *J. Mol. Diagn.* 2003;5(2): 73-81.
20. Cotran RS, Kumar V, Collins T. Robbins Pathologic Basis of Disease. Philadelphia: W.B. Saunders Company, 1999.
21. Abiko Y, Hiratsuka K, Kiyama-Kishikawa M, Tsushima K, Ohta M, Sashihara H. Profiling of differentially expressed genes in human gingival epithelial cells and fibroblasts by DNA microarray. *J Oral Sci*, 2004; 46: 19–24.
22. E Jansová, I Koutná, P Krontorád, Z Svoboda, S Křivánková, J Žaloudík, M Kozubek and S Kozubek. Comparative transcriptome maps: a new approach to the diagnosis of colorectal carcinoma patients using cDNA microarrays. *Clinical Genetics*, 2006, Volume 69, page 218 – 227.
23. Liefers GJ, Tollenaar RAEM. Cancer genetics and their application to individual medicine. *Eur J Cancer*, 2002; 38:872–879.
24. Mlecnik B, Scheideler M, Hackl, Hartler J, Sanchez-Cabo F , Trajanoski Z. PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res.* 2005, 33: W633-W637.