Int J Clin Exp Pathol 2014;7(1):255-263
www.ijcep.com /ISSN:1936-2625/IJCEP1310030

Original Article

Analysis of differentially expressed genes in colorectal adenocarcinoma with versus without metastasis by three-dimensional oligonucleotide microarray

Rita MAM Moura Franco[1], Marcelo M Linhares[1], Suzana S Lustosa[3], Ismael DGC Silva[1], Naiara CN Souza[2], Delcio Matos[1]

[1]UNIFESP - EPM (Federal University Of Sao Paulo), Escola Paulista de Medicina; [2]Piu XII Foundation - Barretos' Cancer Hospital, São Paulo, Brasil; [3]Research and Teaching Unit of Hospital Municipal Dr Munir Rafful, Volta Redonda, Rio de Janeiro, Brasil

Received October 11, 2013; Accepted November 19, 2013; Epub December 15, 2013; Published January 1, 2014

Abstract: Background: Our objective was to examine how the gene expression profile of tumor tissue correlates with lymph node metastasis in patients with advanced colorectal adenocarcinoma (CRAC). Methods: We studied 36 patients (20 men and 16 women, 22-90 years of age) treated for CRAC (classifications of T2, T3, or T4; histological grade of G1 or G2). Amplified tumor mRNA samples were exposed to 20,000 human sequence probes and digitized images of the hybridized samples were analyzed. Results: On average, 2389 probes were detected above the background, with an average correlation R value of 0.19 between data from different patient groups (with or without lymph node invasion, colon or rectal, with or without angio-lymphatic invasion, with or without recurrence). Lymph node metastasis had a statistically significant signature according to Significance Analysis of Microarrays (SAM) and parametric *t*-tests, with a false discovery rate (FDR) = 0.1% and p = 0.001, respectively. Cross-correlation of these two tests identified 102 transcripts as being potentially related to node metastases, with fold changes in the range of 2.182-12.960. Conclusion: We identified 102 differentially expressed genes related to the presence of lymph node metastases in patients with advanced colorectal cancer.

Keywords: Colorectal cancer, gene expression, metastasis, three-dimensional oligonucleotide microarray

Introduction

According to the American Cancer Society, mortality rates for colorectal cancer have decreased over the last two decades, owing to a decrease in the number of cases and improvements in early detection and treatment. When detected early, the 5-year survival rate approaches 90%; however, only 39% of cases are diagnosed early. When there is local or lymph node invasion, the 5-year survival rate falls to 68% [1].

The development of malignant tumors is thought to be the result of sequential changes in various oncogenes and tumor suppressor genes and thus to the proteins they encode [2]. Rarely is a single change in these genes sufficient to cause malignancy. Tumors often have different cytogenetic clones, which originate in cells initially transformed by a genetic change. This heterogeneity contributes to differences in clinical behavior and response to treatment, even in patients with the same histopathological characteristics or diagnosed with the same stage of cancer. Hence, studying carcinogenesis is considered to be of clinical importance to the development of effective treatments for these tumors. Perez et al. [3] noted as early as 1998 the importance of genetics and molecular biology to colorectal cancer. In particular, studies of oncogenes, tumor suppressor genes, and DNA repair genes may yield new perspectives on the diagnosis, treatment, prognosis and follow-up of patients.

Advances in gene expression techniques, such as DNA microarrays have made it possible to quantify genes on a large scale [4-6]. Moreover,

the CodeLink™ platform allows minute differences in gene expression to be detected with 95% confidence [4, 7].

Background

The objective of the present study was to examine potential correlations between tumor gene expression and lymph node metastasis in patients with advanced colorectal adenocarcinoma (CRAC). We performed a three-dimensional (3D) analysis of the expression of oligonucleotides (OGNs) whose hyperexpression has been implicated in lymph node metastasis.

Materials and methods

Setting and subjects

This study was conducted using a database of clinical and histopathological information and biological samples from patients with CRAC, who were treated by the Gastroenterology Surgery Unit of the Federal University of São Paulo (Universidade Federal de São Paulo, UNIFESP) between 2001 and 2008. This databank was developed prospectively using the protocol of the inter-institutional Clinical Genome Project for Cancer of the Foundation for Research Assistance of the State of São Paulo (Fundação de Amparo à Pesquisa no Estado de São Paulo, FAPESP) and the Ludwig Institute for Cancer Research. The project was analyzed and approved by the Medical Ethics Committee of UNIFESP-EPM.

This research reviewed and approved by the Ethics in Research Board UNIFESP, on October 26, 2001, with the number CEP 989/01, without restrictions.

Tumor samples for molecular study were collected by a pathologist in the operating room immediately after surgical removal of tumors. After identification, they were immersed in liquid nitrogen and transported for storage at -80°C.

Patients with a T2, T3, or T4 clinical-pathological classification (advanced cancer) were considered for inclusion. The exclusion factors were having received neoadjuvant radio- and chemotherapy. All included patients had low-grade (G1 or G2) tumors. The clinical and histopathological characteristics of the final study sample (N = 36 patients) are detailed in the Results.

Experimental design

We compared gene expression between patients with and without lymph node metastases, using the complete sample of 36 patients. We also performed gene expression comparisons between the following subgroups of patients: with versus without angio-lymphatic invasion; localized tumors in the colon versus in the rectum; and with versus without tumor recurrence.

RNA extraction

Total RNA was extracted from the tissue specimens with TRIzol® (monophasic phenol solution and guanidine thiocyanate; Life Technologies), following the manufacturer's instructions, purified in silicon columns (Qiagen, Valencia, CA), and evolved in DEPC-treated water. We determined the concentration and purity of each sample was determined by absorption readings at 260 and 280 nm in a Spectronic Genesys 5 spectrophotometer (Spectronic Instruments, Inc., USA). The RNA samples were subjected to electrophoresis in agarose gels with formaldehyde under distilled conditions; the presence of bands corresponding to ribosomal RNA 18 and 28S was considered confirmatory of RNA integrity.

Gene expression analysis

Gene expression was analyzed using the CodeLink™ microarray platform with CodeLink™ UniSet Human commercial microarrays (Amersham Biosciences, Piscataway, NJ) containing 20,000 OGNs, with 30 base pairs each, with each of these having a unique access number in GenBank [8]. CodeLink™ software was used to analyze images obtained by a digital scanner; the program attributes a signal-to-noise ratio (SNR) to each probe by calculating a spot average, divided by the background level (median signal produced by empty spaces between spots), at 1.5 times the standard deviation of the background. The spots were labeled as present (flag G) when the SNR was ≥ 1 or as absent (flag L) when the SNR was < 1. In addition, flags are also noted on the basis of other considerations, such as a contaminated spot or contaminated background (C), signal

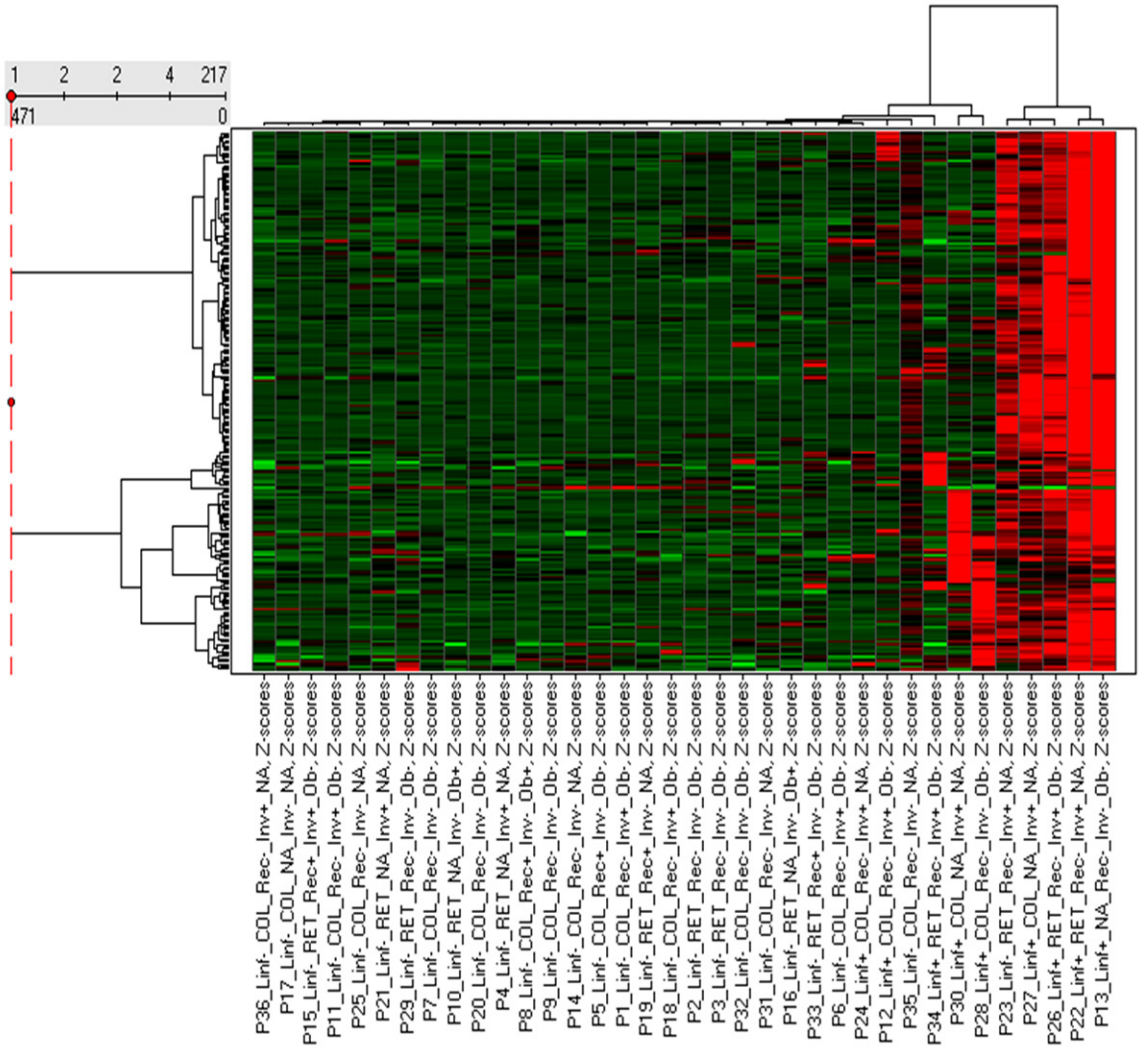

Figure 1. The 217 genes indicated by t-tests to be differentially expressed in patients with versus and without lymph node metastases at p = 0.001. Degree of gene expression dendrograms produced by the software are shown along the x- and y-axes. Rows correspond to individual genes. Columns correspond to individual patients, with the 9 patients with lymph node metastasis placed together on the right end of the figure. Note the concentration of red color on the right of the graph, showing hyperexpression of genes in patients with lymph node metastasis.

saturation (S), irregular form or irregular profile (I), spots removed by the manufacturer (M) or by the user (X). Positive and negative control probes were used to evaluate hybridization efficiency. Along with a set of housekeeping genes, the slides included 68 bacterial probes and 18 positive controls, to allow us to monitor synthesis of cDNA and cRNA, as well as 50 negative controls from the Instituto de Química of USP-SP. Some probes were unique, while others mapped to various genes; this information was important for measuring the reliability of our results and selecting which genes were to be the object of the closest analysis [9].

Reverse transcriptase polymerase chain reaction (RT-PCR) for mRNA amplification

One microgram samples of the total RNA was used to synthesize cDNA. RT-PCR products were purified in the QIAquick® column (Qiagen, Valencia, CA). Complementary RNA (cRNA) was generated by in vitro transcription, using T7 RNA polymerase with Biotin-11-UTP (Perkin Elmer, Boston, MA). The first cDNA strand was generated using SuperScript™ reverse transcriptase, with oligo-dT primers, which have a T7 promoter sequence for RNA polymerase attached at their ends (**Figure 1**). The second

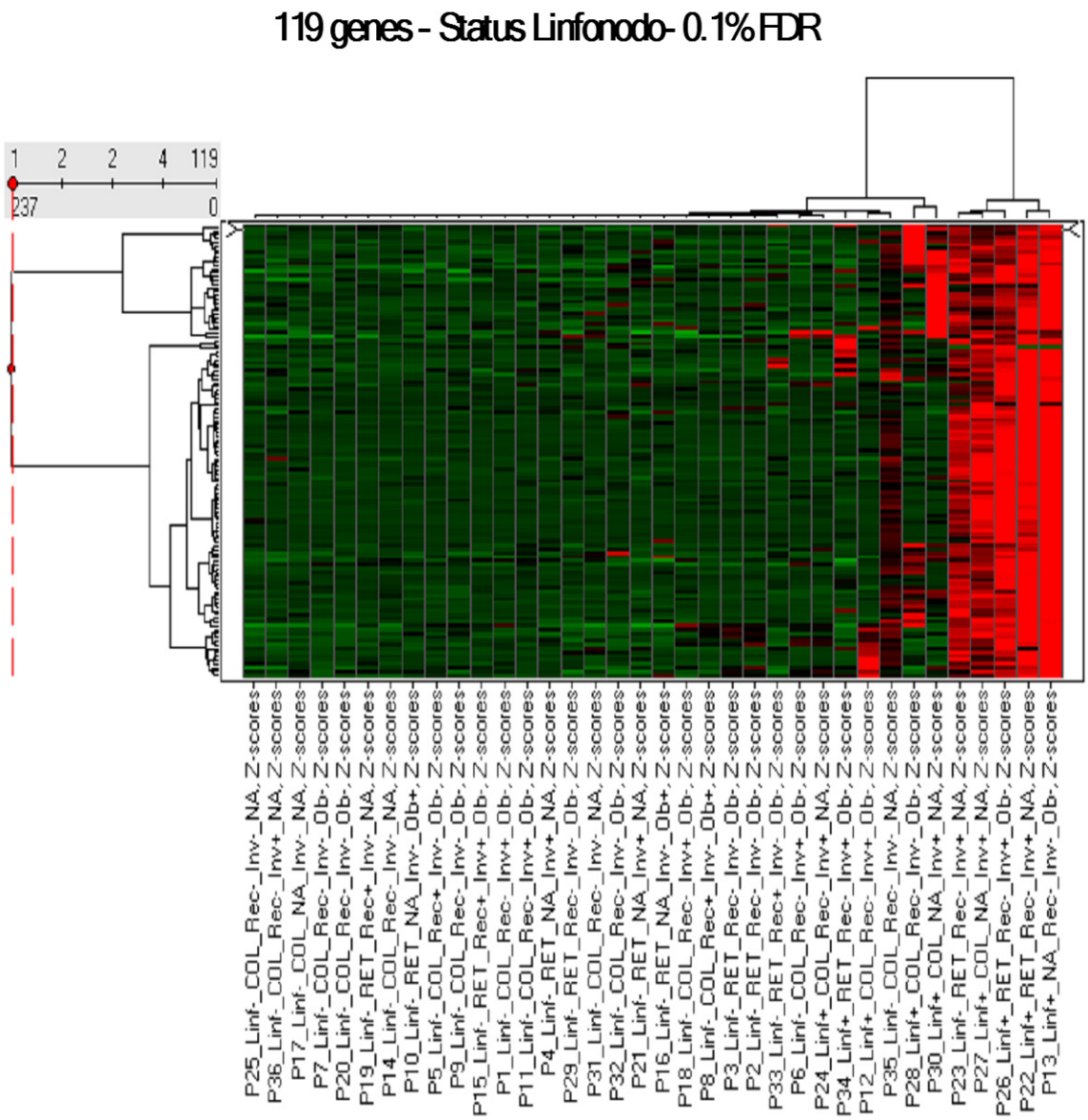

Figure 2. The 119 genes differentially expressed in patients with and without lymph node metastases according to SAM (FDR 0.1%). The layout of this graph follows that of **Figure 1**.

strand of cDNA was synthesized using DNA polymerase I from Escherichia coli. The cRNA was purified in an RNeasy® column (Qiagen) and quantified by ultraviolet spectrophotometry.

Hybridization

cRNA samples were injected into the microcams of the microarray slides for hybridization, for 18 h at 37°C, with a specific buffer, in an Innova™ 4080 incubator (New Brunswick Scientific, Edison, NJ), at 300 rotations per minute.

Post-hybridization, image capture and analysis

The slides were washed in 0.75× TNT buffer [1× TNT: 0.1 mol/L Tris-HCl (pH 7.6), 0.15 mol/L NaCl, and 0.05% Tween20] at 46°C for 1 h and then incubated with streptavidin-Alexa 647 (Molecular Probes, Eugene, OR) at room temperature for 30 min in the dark. The slides were then washed twice in 1× TNT for 5 min per wash, and then placed in 0.05% Tween20. The slides were dried by centrifugation and stored in the dark. Images were captured and digitized with a GenePix® scanner (Axon, Arlington, TX)

and analyzed using CodeLink™ Expression Scanning Software.

Microarray data analysis

The gene expression data were normalized to the intensity median of each slide and the normalized data were exported to Microsoft Excel® spreadsheets. Differential gene expression based on clinical and histopathological characteristics was determined using the parametric t-test [10] and the Significance Analysis of Microarrays (SAM) statistical approach [11].

The data were grouped hierarchically and visualized with SportFire software (TIBCO Inc., Somerville, MA). In the graphs generated, green color denotes less expression of the gene and red color greater expression, with the intensities of these colors being equivalent to levels of expression. We applied the chi-square test to evaluate differences between subgroups. The significance level adopted for our data analysis was $p < 0.05$.

Results

Patient characteristics

Specimens from 36 patients [16 females (44.4%) and 20 males (55.6%)] with CRAC were included in our analysis. These patients ranged in age from 22 to 90 years old (mean, 61.3 years). Most (33/36) of the patients had T3 class tumors, though 1/36 was classified as T2 and 2/36 were classified as T4.

One-fourth of the patients (9/36) presented with lymph node invasion, and three-fourths (27/26) did not. The number of lymph nodes dissected range from 2 to 24 (mean, 23 nodes). Angio-lymphatic invasion was present in 16/36 (44.4%) of the patients, and absent in 20/36 (55.6%) of the patients. A majority of the patients (21/36) had cancerous lesions only in the colon, while slightly more than a third (14/36) had cancerous lesions only in rectum and a single patient (1/36) had cancerous lesions in both the rectum and the colon. Several patients (5/29; 17.2%) experienced recurrence, though most (20/29; 82.8%) had not. With respect to follow-up, 22/36 (61.1%) of the patients had been followed for at least 12 months since removal of their tumors and 14/36 (38.9%) had been followed for less than 12 months at the time the study was completed. Among those who were followed for more than 12 months, the range of time beyond 12 months was 15-83 months (mean, 41 months). Three (13.6%) of the 22 patients who were followed for at least 12 months died during follow-up.

Differential gene expression

On average, 2,389 probes were detected above the background per sample and, out of the total number of slides analyzed, on average, 12% of the probes had reliable intensities.

As shown in **Figure 1**, 217 genes were found to be hyperexpressed in patients with neoplastic lymph node infiltration versus those without ($p = 0.001$, t-test). When we applied the SAM statistical test, we found 119 differentially expressed genes, with false discovery rate (FDR) of 0.1%. The distribution of the intensity of expression of these genes, individually for each patient, can be seen in **Figure 2**. Patients with lymph node metastasis had hyperexpression of the relevant genes.

As shown in **Figure 3**, cross-correlation of the two statistical tests (SAM and parametric t-test) identified 102 differentially expressed genes, that is, genes with a statistically significant expression signature (FDR = 0.1% and $p = 0.001$). Patients with lymph node metastases hyperexpressed these genes (see Supplemental Table 1 for details). An analysis of the global expression of these genes revealed an average fold-change of 5.667 (range, 2.182-12.960). That is, these 102 genes were expressed at a level that was 2.182 to 12.960 times greater than in patients with metastases than in those without lymph node metastases.

In addition, parametric t tests indicated that there were 95 genes that were differentially expressed in tumors from the colon versus tumors from the rectum ($p = 0.01$). The average fold change between tumors from these two places was 0.66 (range, 0.13-2.31). Parametric t tests also indicated that there were 29 genes that were differentially expressed in relation to the presence or absence of recurrence ($p = 0.01$). The average fold change between tumors from patients with versus without recurrence was 2.89 (range, 0.24-11.45). Finally, parametric t tests indicated that there were 37 genes

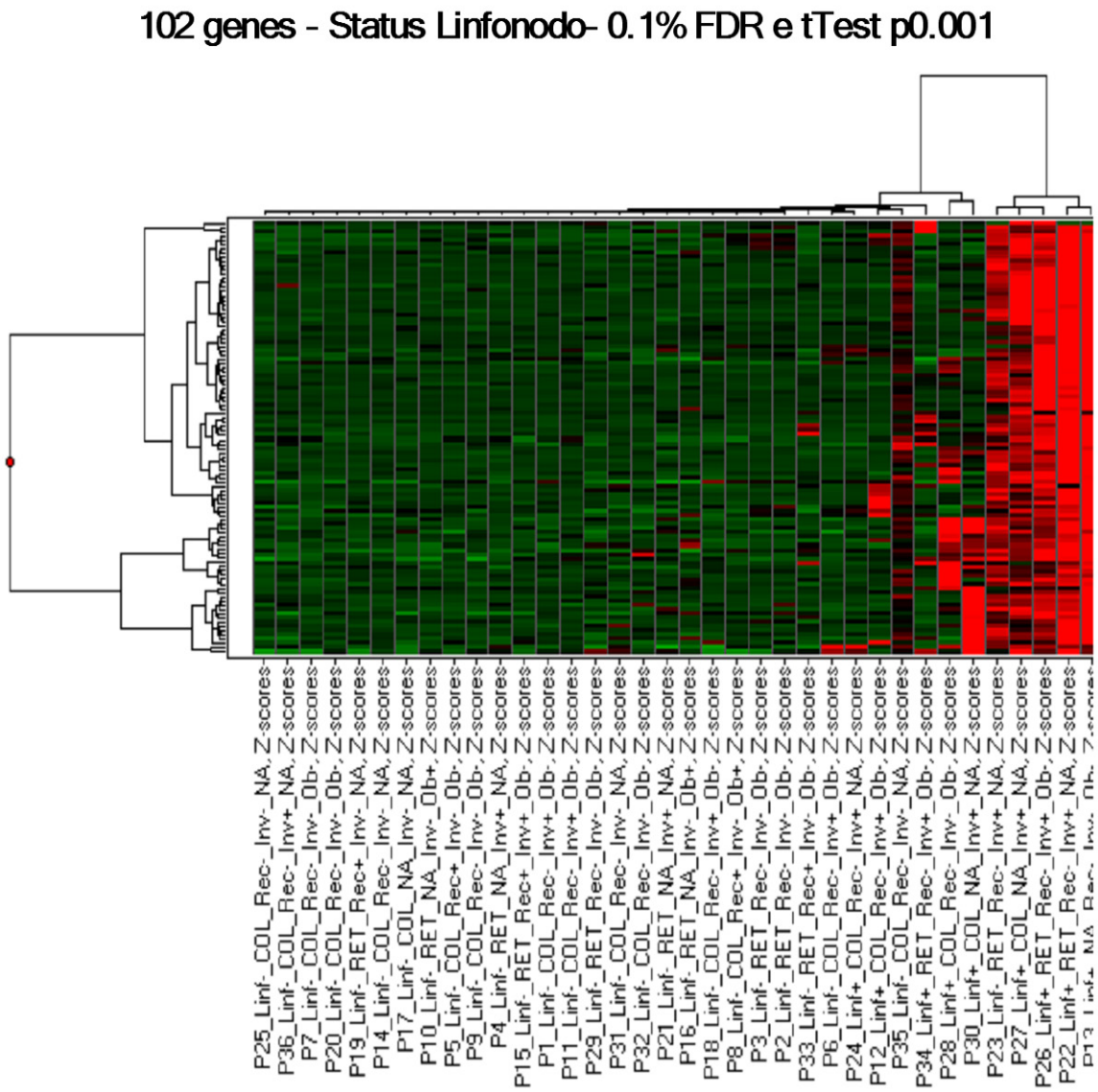

Figure 3. The 102 genes confirmed by cross-validation of t-tests (p = 0.001) and SAM (FDR 0.1%) to be differentially expressed in patients with versus without lymph node metastases. The layout of this graph follows that of **Figure 1**.

that were differentially expressed between patients with versus without angio-lymphatic invasion (p = 0.01). Thus, 37 genes were differentially expressed between patients with and without angio-lymphatic invasion, with an average difference in expression of 1.056 (range, 0.388-2.112).

Discussion

In the present analysis of gene expression in CRAC tumor samples using the CodeLink™ microarray platform, we identified 102 differentially expressed genes related to the presence of lymph node metastases by cross-correlating the results of two statistical tests, namely SAM and parametric t-tests. Parametric t-tests further pointed to 37 genes that appeared to be differentially expressed with versus without angio-lymphatic invasion, 95 genes that appeared to be differentially expressed in tumors from the colon versus from the rectum, and 29 that appeared to be differentially expressed between patients with versus without recurrence.

Among the transcribed genes with significant differential expression greater than 10 times in

lymph-node metastasized patient samples versus non-metastasized patient samples (t-test and SAM), we consider the following to be particularly noteworthy: UXT (ubiquitously-expressed transcript); CHCHD2 (coiled-coil-helix-coiled-coil-helix domain containing 2); FAN3D (family with sequence similarity 3, member D); IGJ (immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides); IFITM3 (interferon Induced transmembrane protein 3); ITN2C (integral membrane protein 2C); MRTS35 (mitochondrial ribosomal protein S35); PRAP1 (proline-rich acidic protein 1) and CLCA1 (chloride channel, calcium activated, family member 1).

Microarrays and the Codelink™ platform

Microarray technology is widely used in oncology to elucidate the biological mechanisms of oncogenesis, to discover new medicines, and to develop predictors of outcome, with an aim toward developing individualized treatments for patients [12, 13]. Microarrays are a powerful means of examining an enormous quantity of transcriptions of various genes at the same time. They are electronic systems that analyze DNA fragments and identify the intensity of gene action and metabolic activities. Linking the microarray technique with computer programs and statistical tools has provided important knowledge about areas such as gene expression, pathways mediating cell responses, and tumor classification. Researchers obtain a visual "map" of genetic organization in which hyperexpressed and hypoexpressed genes appear in different colors, while genes that are not differentially expressed appear an intermediate color [14].

Microarrays do have some limitations, such as pitfalls related to image acquisition, variability, classification errors in repeated measurements, and limitations in sensitivity [4]. In addition, it is difficult to compare datasets obtained via different platforms. Studies comparing different platforms have highlighted difficulties in reproducing data both within and across platforms [15-17].

Microarrays employ OGN or DNA probe hybridization to measure the expression of thousands of genes in a single hybridization experiment [11]. Enormous quantities of data are generated, necessitating methods to determine whether observed differences in gene expression are actually significant. Although analysis of microarray data grouping may yield coherent patterns of gene expression, it provides little information about statistical significance. Methods based on conventional tests do not address the probability that a difference in gene expression has occurred by accident. A $p = 0.01$ criterion for significance, which in the context of experiments evaluating a small number of genes may be acceptable, could lead to the identification of 200 genes by chance in an experiment examining 20,000 genes. This problem led Tusher et al. [11] to develop the SAM statistical method which is specifically adapted for analyzing microarray data.

SAM can identify genes with significant changes in their expression by assimilating a group of specific genes extrapolated by t-tests. Each gene is given a base point for changes in its genetic expression in relation to the standard deviation of repeated measurements for that gene. Genes with scores above a specified threshold are considered to have potentially significant changes in their expression. The percentage of those genes expected to be identified by accident is the FDR. To estimate the FDR, hypothetical genes are identified by an analysis of permutations in their measurements. The limit can be adjusted to identify fewer or greater numbers of genes, and FDRs are calculated for each group. The introduction of SAM has been an important advance given that conventional methods of analysis used FDRs that were between 60 and 80%.

In our study, the estimated FDR for our comparison of samples from patients with versus without lymph node metastases was only 0.1%. In other words, of the 102 genes identified as having altered expression, it is likely that none of them were identified by chance. That is to say, we can have strong confidence that these identified genes have actually undergone biological changes.

For the additional clinical and histopathological data analyzed in this study (location of the tumor, angio-lymphatic invasion, recurrence, and specific mortality), we did not apply the SAM test. These comparisons had *p* values near 0.01 in parametric t-tests, which we did not consider sufficiently reliable given that when 20,000 genes are being examined, ~200

genes (1% of 20,000) could be tagged as showing significant differences purely by chance. Nevertheless, it is possible that these genes could be important in relation to metastasis of colorectal cancer. Thus, these genes, or subgroups of them, should be evaluated in the future by real-time RT-PCR and immunohistochemistry of new samples.

The guidelines we followed in our study were as follows:

- We used p = 0.001 rather than 0.01 since the number of genes selected by chance would have been unacceptably high in a platform involving 20,000 genes.

- We observed gene expression in pre-established groups rather than classifying groups according to gene expression.

- We did not form the groups using methods that guaranteed only the best results.

For these reasons, we believe that the results of our microarray study are reliable given the available data. However, we also believe that there is a need for these data to be validated since this is the first study of colorectal cancer, to our knowledge, to use microarrays with the CodeLink™ platform.

In addition to the aforementioned strengths of this work, one additional fact further enhances our findings. Prior studies examining this topic have considered differences in gene expression between groups on the order of 2- or 3-fold to be significant. In our study, significant fold differences were in the range of 2.182-12.960 (mean, 5.667-fold). For example, expression of the LAMC-2 gene was 7.7-fold greater in patients with lymph node metastases than in those without metastases.

Several research groups have identified groups of genes that exhibit a progressive increment of expression favoring changes leading to metastasis. If we start from the premise that microarrays enable us to study thousands of genes and from there develop studies using superarrays (for which few genes have been analyzed, but which are extremely specific with respect to the function and development of carcinogenesis in colorectal cancer), we can confirm that we are on the right path for the development of this line of research.

Conclusion

Together with prior works, this study offers a valuable approach for revealing gene expression profiles that allow markers of aggressiveness to be identified. In other types of cancer, such as breast and blood cancers, research programs have reached a more advanced stage; however, in the field of colorectal cancer research, this project is a pioneering one in terms of pointing to the expression of 102 genes that may be involved in carcinogenesis using microarrays, and correlating the findings with histopathological characteristics.

Acknowledgements

This manuscript was partly supported by FAPESP (Support Research Foundation Of State Of São Paulo) - PROCESS NUMBER: 2004/10701-7. FAPESP: the role of the funding body was the provision of financial resources to purchase CodeLink™ microarray platform.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Rita MAM Moura Franco, Department of Surgery - UNIFESP - EPM (Federal University Of Sao Paulo) - Escola Paulista de Medicina Medicine and Pathology, Rua Apeú, 111 - Jardim da Saúde, São Paulo, Brasil. CEP: 04288-150. Tel: 5511/3729-1106; E-mail: ritmoura1@gmail.com; Marcelo Moura Linhares, UNIFESP - EPM (Federal University Of Sao Paulo), Escola Paulista de Medicina, São Paulo, Brasil. E-mail: marbet@uol.com.br

References

[1] American Cancer Society. Cancer facts & figures 2008 [Internet]. Atlanta: American Cancer Society; 2008 [cited 2013 Oct 10]. Available from: http://www.cancer.org/acs/groups/content/@nho/documents/document/2008cafffinalsecuredpdf.pdf.

[2] Croce CM. Oncogenes and cancer. N Engl J Med 2008; 358: 502-511.

[3] Perez RO, Habr-Gama A, Jacob CE, Sousa AHS Jr, Picolo MM, Pécora RA. [Genetics of colorectal cancer: principles for the surgeon]. Rev Bras Colo-proctol 1998 Jan-Mar; 18: 5-10. [cited 2013 Oct 10]. Portuguese. Available from: http://www.jcol.org.br/pdfs/18_1/01.pdf.

[4] Ramakrishnan R, Dorris D, Lublinsky A, Nguyen A, Domanus M, Prokhorova A, Gieser L, Touma E, Lockner R, Tata M, Zhu X, Patterson M, Shippy R, Sendera TJ and Mazumder A. An assessment of Motorola CodeLink microarray performance for gene expression profiling applications. Nucleic Acids Res 2002; 30: e30.
[5] Shalon D, Smith SJ and Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 1996; 6: 639-645.
[6] Stremmel C, Wein A, Hohenberger W and Reingruber B. DNA microarrays: a new diagnostic tool and its implications in colorectal cancer. Int J Colorectal Dis 2002; 17: 131-136.
[7] Otsuka AY, Andrade PM, Villanova FE, Borra RC and Silva ID. Human endometrium mRNA profile assessed by oligonucleotide three-dimensional microarray. Gynecol Endocrinol 2007; 23: 527-534.
[8] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J and Wheeler DL. GenBank: update. Nucleic Acids Res 2004; 32: D23-26.
[9] Diez D, Alvarez R and Dopazo A. Codelink: an R package for analysis of GE healthcare gene expression bioarrays. Bioinformatics 2007; 23: 1168-1169.
[10] Pan W. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics 2002; 18: 546-554.
[11] Tusher VG, Tibshirani R and Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001; 98: 5116-5121.
[12] Dupuy A and Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 2007; 99: 147-157.
[13] Liefers GJ and Tollenaar RA. Cancer genetics and their application to individualised medicine. Eur J Cancer 2002; 38: 872-879.
[14] Conselho de Informações sobre Biotecnologia. [Biochip mostra no computador a atividade genética]. Biotech 2004 Feb; 2: 4. [cited 2013 Oct 10]. Portuguese. Available from: http://www.cib.org.br/pdf/biotech05.pdf.
[15] de Reynies A, Geromin D, Cayuela JM, Petel F, Dessen P, Sigaux F and Rickman DS. Comparison of the latest commercial short and long oligonucleotide microarray technologies. BMC Genomics 2006; 7: 51.
[16] Held GA, Duggar K and Stolovitzky G. Comparison of Amersham and Agilent microarray technologies through quantitative noise analysis. OMICS 2006; 10: 532-544.
[17] Severgnini M, Bicciato S, Mangano E, Scarlatti F, Mezzelani A, Mattioli M, Ghidoni R, Peano C, Bonnal R, Viti F, Milanesi L, De Bellis G and Battaglia C. Strategies for comparing gene expression profiles from different microarray platforms: application to a case-control experiment. Anal Biochem 2006; 353: 43-56.

Supplemental Table 1. Genes (n = 102) differentially expressed by patients with and without lymph node metastases

Idx	ACCN#	Probe_Nar	Description	Average Linf -	Average Linf +	Fold Chan	t-Test ALL Linf	Score(d) -	q-value(%)	102 genes - Stat
5461	NM_00176	GE57904	CD1C antigen, c polypeptide (CD1C)	1,267	5,500	4,058	0,001	4,156	0	1
17920	H80132.1	GE86308	yu11b08s1 Soares fetal liver spleen 1NFLS cDNA clone IMAGE:2334	1,586	7,118	4,489	0,000	3,744	0	2
2881	NM_01762	GE55334	hypothetical protein FLJ20014 (FLJ20014)	1,035	4,770	4,612	0,001	3,517	0	3
5242	NM_00418	GE57689	Smcy homolog, X-linked (mouse) (SMCX)	9,314	35,903	3,855	0,001	3,689	0	4
20054	R94272.1	GE88442	yq38f01r1 Soares fetal liver spleen 1NFLS cDNA clone IMAGE:19807	2,158	8,072	3,740	0,000	3,651	0	5
10543	NM_00184	GE63370	ciliary neurotrophic factor receptor (CNTFR), transcript variant 2	3,023	15,167	4,109	0,000	3,891	0	6
11761	NM_19896	GE80148	amino-terminal enhancer of split (AES), transcript variant 1	1,957	10,560	5,140	0,000	4,002	0	7
8022	NM_00299	GE60412	syndecan 1 (SDC1), transcript variant 2	1,259	4,668	3,330	0,000	3,523	0	8
10524	NM_01652	GE63351	chromosome 9 open reading frame 78 (C9orf78)	2,598	16,746	5,978	0,000	3,482	0	9
16551	AK127475	GE84939	cDNA FLJ45567 fis, clone BRTHA3010135, highly similar to LAR pro	1,939	11,152	5,489	0,001	3,455	0	10
9172	AK026549	GE61995	cDNA: FLJ22896 fis, clone KAT04996, highly similar to HSU47105 H	2,335	10,059	4,308	0,001	3,455	0	11
10442	AI024862.1	GE63269	ov39g03x1 Soares_testis_NHT cDNA clone IMAGE:1639732 3' simil	2,611	19,725	7,556	0,001	3,563	0	12
11529	NM_00055	GE79916	hemoglobin, alpha 1 (HBA1)	1,763	16,963	9,621	0,001	3,739	0	13
5893	NM_00418	GE58325	ubiquitously-expressed transcript (UXT), transcript variant 2	9,690	100,457	10,367	0,001	3,670	0	14
5078	NM_00088	GE57528	IMP (inosine monophosphate) dehydrogenase 1 (IMPDH1), transcript	1,511	13,581	8,048	0,000	3,643	0	15
3856	AK096998	GE56316	cDNA FLJ39679 fis, clone SMINT2010068	1,479	5,455	3,499	0,000	3,463	0	16
6904	U66047.1	GE59326	clone Z'3-1 placenta expressed mRNA from chromosome X	1,122	5,844	4,968	0,000	3,926	0	17
7174	NM_00356	GE59596	phospholipase A2, group X (PLA2G10)	2,161	18,179	8,412	0,000	3,973	0	18
6784	NM_00573	GE59209	actin related protein 2/3 complex, subunit 2, 34kDa (ARPC2), transcr	2,556	23,652	8,587	0,000	3,939	0	19
5117	NM_00667	GE57567	HLA complex P5 (HCP5)	1,878	13,112	6,525	0,000	3,718	0	20
8545	NM_01613	GE61372	coiled-coil-helix-coiled-coil-helix domain containing 2 (CHCHD2)	10,859	111,278	10,248	0,001	3,662	0	21
9870	NM_01221	GE62693	integrin, alpha 11 (ITGA11), transcript variant 2	1,691	9,201	5,441	0,000	4,265	0	22
8244	NM_00178	GE61077	CD97 antigen (CD97), transcript variant 2	1,577	8,429	5,344	0,001	3,514	0	23
4210	NM_01807	GE56674	hypothetical protein FLJ10357 (FLJ10357)	1,021	4,266	3,959	0,000	3,566	0	24
19237	NM_13880	GE87625	family with sequence similarity 3, member D (FAM3D)	6,318	69,430	11,342	0,000	4,268	0	25
20324	NM_01817	GE88712	golgi phosphoprotein 3-like (GOLPH3L)	1,595	7,378	4,627	0,001	3,507	0	26
5496	NM_00173	GE57937	carbonic anhydrase I (CA1)	1,891	19,068	9,653	0,001	3,768	0	27
8113	NM_14464	GE60502	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha	1,453	18,826	12,960	0,001	3,752	0	28
2785	NM_00602	GE55238	transforming growth factor beta 1 induced transcript 4 (TGFB1I4), tran	1,747	15,447	8,195	0,001	3,456	0	29
8586	NM_02040	GE61410	polycythemia rubra vera 1 (PRV1)	2,287	12,237	5,351	0,001	3,491	0	30
2318	NM_01515	GE54771	PHD finger protein 3 (PHF3)	1,621	5,952	3,301	0,000	3,494	0	31
2293	NM_00473	GE54746	VAMP (vesicle-associated membrane protein)-associated protein B a	2,254	9,285	3,632	0,000	3,941	0	32
9482	NM_00490	GE62305	peroxiredoxin 6 (PRDX6)	1,648	4,007	2,431	0,000	3,482	0	33
685	NM_00422	GE53137	cofactor required for Sp1 transcriptional activation, subunit 2, 150kDa	1,105	3,806	3,184	0,000	3,454	0	34
1915	NM_00568	GE54368	phenylalanine-tRNA synthetase-like, beta subunit (FARSLB)	2,001	15,624	7,807	0,000	4,116	0	35
18401	INCYTE UI	GE86789	INCYTE UNIQUE	1,547	8,383	5,419	0,000	3,710	0	36
14193	NM_02103	GE82581	interferon induced transmembrane protein 3 (1-8U) (IFITM3)	4,466	53,997	11,343	0,001	3,471	0	37
9915	NM_01771	GE62738	membrane-spanning 4-domains, subfamily A, member 12 (MS4A12)	1,485	11,758	7,919	0,001	3,623	0	38
18676	NM_15371	GE87064	tubulin tyrosine ligase (TTL)	1,279	6,396	4,521	0,000	3,523	0	39
9544	NM_00021	GE62367	integrin, alpha 6 (ITGA6)	1,219	6,152	5,048	0,001	3,502	0	40
14809	NM_03092	GE83197	integral membrane protein 2C (ITM2C)	2,492	28,008	11,237	0,000	4,016	0	41
2155	NM_02182	GE54608	mitochondrial ribosomal protein S35 (MRPS35), nuclear gene encodi	2,833	32,463	11,745	0,001	3,683	0	42
4721	NM_00281	GE57181	proteasome (prosome, macropain) 26S subunit, non-ATPase, 8 (PSM	5,462	42,342	7,752	0,001	3,634	0	43
3667	AY128643	GE56124	PMEPA1 variant A protein mRNA	1,129	10,592	9,381	0,000	3,806	0	44
17294	NM_14520	GE85682	proline-rich acidic protein 1 (PRAP1)	4,101	46,593	11,360	0,001	3,712	0	45
12615	NM_00128	GE81002	chloride channel, calcium activated, family member 1 (CLCA1)	1,125	13,401	10,782	0,000	3,464	0	46
17036	INCYTE UI	GE85424	INCYTE UNIQUE	2,260	22,699	9,982	0,000	4,609	0	47
3196	NM_01798	GE55649	LIM and senescent cell antigen-like domains 2 (LIMS2)	1,564	14,372	8,953	0,000	4,487	0	48
16295	AF085835.	GE84683	full length insert cDNA clone YI41B09	1,269	6,609	5,210	0,000	3,740	0	49
2828	NM_02117	GE55281	LSM2 homolog, U6 small nuclear RNA associated (S cerevisiae) (LS	2,084	10,982	5,269	0,001	3,487	0	50
2338	NM_00631	GE54791	connector enhancer of kinase suppressor of Ras 1 (CNKSR1)	1,613	6,453	4,002	0,000	3,596	0	51
3464	NM_05402	GE55917	ankylosis, progressive homolog (mouse) (ANKH)	1,016	3,382	3,385	0,000	3,654	0	52
1403	NM_00700	GE53855	transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila) (TLE	1,013	3,511	3,464	0,000	3,571	0	53
7961	NM_00556	GE60355	laminin, gamma 2 (LAMC2), transcript variant 1	2,479	19,092	7,700	0,001	3,755	0	54
19444	NM_00317	GE87832	synaptophysin (SYP), mRNA	1,783	7,731	4,390	0,000	3,944	0	55
14678	NM_02506	GE83066	hypothetical protein FLJ23420 (FLJ23420)	1,696	11,511	7,018	0,000	3,950	0	56
18192	NM_00100	GE86580	olfactory receptor, family 56, subfamily B, member 1 (OR56B1)	1,886	12,863	6,960	0,000	4,111	0	57
19642	INCYTE UI	GE88030	INCYTE UNIQUE	1,549	10,263	6,716	0,001	3,753	0	58
10836	NM_01599	GE79223	complement component 1, q subcomponent, alpha polypeptide (C1Q	1,913	12,696	6,636	0,001	3,639	0	59
8763	NM_03231	GE61586	polymerase (DNA-directed), delta interacting protein 3 (POLDIP3), tra	1,068	4,805	4,138	0,000	4,000	0	60
15111	NM_03273	GE83499	interleukin 17 receptor C (IL17RC), transcript variant 3	1,991	9,767	4,986	0,000	4,038	0	61
4952	NM_01480	GE57409	engulfment and cell motility 1 (ced-12 homolog, C elegans) (ELMO1)	5,526	13,511	2,445	0,000	3,872	0	62
13417	NM_01248	GE81805	hyaluronan-mediated motility receptor (RHAMM) (HMMR), transcript	1,360	6,883	4,652	0,000	3,738	0	63
12339	W90162.1	GE80726	zh77g11s1 Soares_fetal_liver_spleen_1NFLS_S1 cDNA clone IMAGE	3,427	21,339	5,733	0,000	3,596	0	64
11827	NM_00682	GE80214	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation p	1,825	11,037	6,118	0,001	3,734	0	65
12020	NM_01435	GE80407	POU domain, class 2, transcription factor 3 (POU2F3)	1,852	10,019	5,265	0,001	3,735	0	66
6164	NM_01451	GE58596	upstream binding protein 1 (LBP-1a) (UBP1)	1,033	5,957	5,034	0,000	4,348	0	67
12781	NM_00261	GE81169	peroxisome biogenesis factor 10 (PEX10), transcript variant 2	1,029	3,351	3,227	0,000	3,629	0	68
14370	NM_02394	GE82758	cytochrome P450, family 4, subfamily F, polypeptide 12 (CYP4F12)	1,109	3,555	3,207	0,000	3,896	0	69
6805	NM_00315	GE59228	stromal interaction molecule 1 (STIM1)	1,771	6,888	3,888	0,000	4,409	0	70
7035	NM_00581	GE59457	glycoprotein A33 (transmembrane) (GPA33)	0,740	2,604	3,517	0,000	3,818	0	71
19940	BU159559	GE88328	AGENCOURT_7859553 NIH_MGC_67 cDNA clone IMAGE:6141023	1,342	10,913	8,369	0,000	5,010	0	72
16396	AK023086	GE84784	cDNA FLJ13024 fis, clone NT2RP3000865	1,352	5,253	3,544	0,000	3,873	0	73
1735	NM_17078	GE54188	zinc ribbon domain containing, 1 (ZNRD1), transcript variant a	1,067	2,741	2,569	0,000	3,583	0	74
11331	NM_01456	GE79718	orthodenticle homolog 1 (Drosophila) (OTX1)	0,993	4,123	4,193	0,000	3,610	0	75
6548	NM_00024	GE58977	MHC class II transactivator (MHC2TA)	1,234	6,478	4,805	0,000	5,005	0	76
17152	NM_08081	GE85540	oxoglutarate (alpha-ketoglutarate) receptor 1 (OXGR1)	1,116	3,413	3,059	0,000	3,608	0	77
3631	NM_01501	GE56085	amine oxidase (flavin containing) domain 2 (AOF2)	1,087	2,817	2,591	0,000	3,749	0	78
14117	NM_02015	GE82505	chromosome 15 open reading frame 24 (C15orf24), mRNA	1,067	3,938	3,689	0,000	3,532	0	79
5333	NM_00683	GE57777	inner membrane protein, mitochondrial (mitofilin) (IMMT)	0,928	2,791	3,007	0,000	4,145	0	80
15197	NM_03292	GE83585	hypothetical protein MGC15875 (MGC15875)	1,208	5,966	4,939	0,000	3,685	0	81
3943	AK124939	GE56403	cDNA FLJ42949 fis, clone BRSTN2006583	1,454	5,577	3,812	0,000	4,087	0	82
18760	INCYTE UI	GE87148	INCYTE UNIQUE	1,726	8,152	4,722	0,001	3,601	0	83
2165	NM_00206	GE54618	glutaredoxin (thioltransferase) (GLRX)	1,130	8,119	6,505	0,001	3,609	0	84
20067	NM_01638	GE88455	three prime repair exonuclease 1 (TREX1), transcript variant 1	1,741	7,824	4,493	0,001	3,470	0	85
6047	NM_01228	GE58479	KCNE1-like (KCNE1L)	1,849	8,749	4,562	0,000	4,012	0	86
567	NM_01427	GE53019	mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltran	0,954	4,204	4,405	0,000	3,834	0	87
10677	NM_02109	GE79064	peptide YY, 2 (seminalplasmin) (PYY2)	0,878	6,859	7,476	0,000	3,876	0	88
17133	AK021734	GE85521	cDNA FLJ11672 fis, clone HEMBA1004733	1,597	9,304	5,826	0,000	3,902	0	89
8165	NM_01430	GE60554	hypothetical protein HSPC117 (HSPC117)	1,275	8,232	5,592	0,000	3,750	0	90
15235	NM_03320	GE83623	zinc finger protein 101 (ZNF101)	1,318	4,970	3,771	0,000	3,653	0	91
9685	NM_00288	GE62508	Ran GTPase activating protein 1 (RANGAP1)	1,312	6,284	4,878	0,000	4,046	0	92
15168	NM_03285	GE83556	coronin 6 (CORO6)	1,057	3,376	3,205	0,000	4,351	0	93
13785	NM_01755	GE82173	hypothetical protein DKFZp434H2215 (DKFZp434H2215)	1,715	12,607	6,701	0,000	4,593	0	94
3675	NM_03146	GE56133	SH3 domain binding glutamic acid-rich protein like 2 (SH3BGRL2)	0,976	3,907	4,000	0,000	3,566	0	95
17977	NM_02103	GE86365	KIAA1404 protein (KIAA1404)	1,046	4,680	4,473	0,000	3,511	0	96
13337	NM_19804	GE81725	nudix (nucleoside diphosphate linked moiety X)-type motif 6 (NUDT6)	1,568	8,366	5,337	0,001	3,501	0	97
7116	NM_00419	GE59538	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxyl	1,273	4,630	3,534	0,000	3,578	0	98
3491	AK022090	GE55944	cDNA FLJ12028 fis, clone HEMBB1001850	0,973	3,229	3,225	0,000	3,736	0	99
12650	AF031174.	GE81038	Ig-like membrane protein (IGSF3) mRNA	2,040	8,250	3,798	0,000	3,757	0	100
4264	BC057843	GE56728	hypothetical protein FLJ25339, mRNA (cDNA clone MGC:71843 IMA	1,466	3,239	2,209	0,000	3,759	0	101
9763	NM_00259	GE62586	procollagen C-endopeptidase enhancer (PCOLCE)	0,976	2,129	2,182	0,000	3,527	0	102