

Differences in gene expression between individuals with multiple primary and single primary malignancies

G.P. STATHOPOULOS¹ and A. ARMAKOLAS²

¹A' Oncology Department, Errikos Dunant Hospital, Athens; ²Laboratory of Experimental Physiology, Athens Medical School, Athens, Greece

Received July 1, 2009; Accepted August 14, 2009

DOI: 10.3892/ijmm_00000272

Abstract. Cytogenetic and molecular studies have identified imbalanced chromosomal regions leading to the characterization of several candidate genes. Differences in gene expression were examined in the blood by whole genome microarray analysis among individuals with double or single primary malignancies and healthy individuals. Twenty-four individuals with at least two primary malignancies of the breast and/or colon and/or ovary were compared with 32 individuals with single breast, colon or ovarian cancer. The single malignancy group had a median duration of disease of 9 years (range 5-23 years). Validation was obtained by examining each patient separately with quantitative real-time reverse-transcriptase polymerase chain reaction (RT-PCR) analysis for the determined genes. Overall a large number of genes were determined to be deregulated. From the classifiers built, a 9-probe signature was determined between second primary and single tumor patients. Four other genes were determined to be repressed ($p < 1 \times 10^{-4}$) in individuals with two primary malignancies when compared with individuals with a single malignancy and also when comparing single malignancies and healthy subjects. The levels of gene deregulation were confirmed by validation with quantitative RT-PCR analysis. Functional analysis, suggested that these genes are associated with protein biosynthesis and folding, inhibition of apoptosis and intracellular signalling via GTP cascade. The outcome of the present study was 13 genes had a statistically significant difference in expression between individuals with double primary malignancies compared to individuals with single primary malignancies. Nine of those were confirmed by the classifier analysis.

Introduction

Multiple primary malignancies are defined by two or more malignant tumors which have a different primary site and

often a different histology. It appears that the incidence of two or more primary malignancies has been increasing over the last decades (1). This increased incidence might be attributed to the improvement of diagnostic tools, treatment modalities and the prolongation of survival time of patients with first primary tumors (2,3). During recent years, many carcinogens directly or indirectly related to cancer have been identified. The great majority of these carcinogens are environmental but genetic defects have also been involved (4). Of breast and colorectal cancers, ~10% are considered to be hereditary (5). An established example is the BRCA1 and BRCA2 genes which are the major causes of hereditary breast cancer mainly and also of ovarian cancer. The genetic defect in the majority of familial and hereditary breast cancers has not been clarified (2,6). Familial adenomatous polyposis due to mutations in adenomatous polyposis coli (APC) and hereditary nonpolyposis colorectal cancer (HNPCC) syndrome caused by mutations in the mismatch-repair genes MLH1, MSH2, MSH6 and PMS2, are the underlying causes of hereditary colorectal cancer in many families (5). The genetic defects have not yet been clarified with regard to most families whose members are at an older age at the onset of hereditary colorectal cancer (7,8). The mutation of CHEK2 400 delC has been associated with a 2-fold increased risk of breast cancer (9). A subtype of familial breast cancer and colorectal cancer has been identified (10). CHEK2 400 delC was determined at a frequency of 21% in families with the hereditary breast and colorectal cancer phenotype (11). An investigation of CHEK2 400 delC in Swedish women with double primary breast and colorectal cancers suggested that development of these two tumor types was not sufficient to recommend the mutation analysis of CHEK2 as it was found to occur at a low frequency (5). Research in identifying genes related to breast cancer development has rendered some data. It was shown that chromosomal imbalances correlate with altered gene expression (12). Cytogenetic and molecular studies have identified imbalanced chromosomal regions leading to the characterization of several candidate genes, such as MYC at 8q10, CCND1, EMS1, PRAD1 at 4q9 and many others (13-18). Genomic imbalances were found by matrix-comparative genomic hybridization (CGH) of 22 cases of advanced breast cancer. With 51 amplicons, genome database information defined 42 candidate genes, some of which were overexpressed; these data are correlated with advanced stage breast cancer (19). The early detection of breast cancer was attempted by gene expression in peripheral blood in a study

Correspondence to: Dr G.P. Stathopoulos, A' Oncology Department, Errikos Dunant Hospital, Semitelou 2A, 11528 Athens, Greece
E-mail: dr-gps@ath.forthnet.gr

Key words: microarrays, double primary malignancies, qRT-PCR, second primary tumors, gene expression

which identified 32 genes that correctly predicted the diagnostic class in at least 82% of the samples (20).

Multiple primary malignancies have been shown to have a certain gene expression. The synchronous appearance of endometrial and ovarian cancer is seen in significantly younger women than in women with either primary endometrial or ovarian cancer alone (21-23). This detection led to the hypothesis that a specific germ-line genetic predisposition or a sporadic gene target may play a role in a small group of patients in this category.

Based on data suggesting that certain genetic factors predispose to cancer as well as to second primary tumor development, the present study investigated the gene expression profiles of patients with one primary tumor versus patients with two primary tumors. RNA samples from the blood of patients with double tumors, single tumors and from healthy control subjects were studied by gene expression analysis using the whole genome microarray platform AB500 from Applied Biosystems. Gene expression profiles were derived from the microarray normalized raw data by biostatistical and classification analysis.

Materials and methods

Patients. Peripheral blood was collected from 56 patients with cancer and 10 healthy individuals. All of the individuals participating in the study were age-matched. Our study group consisted of 10 patients with two primary tumors (group AB) mostly (>90%) within a 5-year time period. Fourteen of these individuals (group A) presented with a combination of breast and/or colon and/or ovarian cancers, whereas the remaining 10 (group B), presented with a first primary tumor similar to group A, i.e. breast or colon or ovarian cancer, and a second primary tumor that was located elsewhere (Table I). The 4 control groups used in this study consisted of 10 patients with colon cancer (group C), 13 with breast cancer (group D), 9 with ovarian cancer (group E) and 10 healthy individuals (group N); AB, all double cancers; CDE, all single cancers; N, healthy controls (Table I). All patients with single malignancies had 9 years of median duration of disease (range, 5-23 years). Informed consent was obtained from all study participants.

Sample preparation

RNA collection. RNA was obtained by peripheral blood from patients and healthy individuals by using the PAXgene Blood RNA system (Qiagen) according to the manufacturer's instructions. In order to avoid epithelial cell contamination of our samples, the first 5 ml of blood in each case after vein puncture, was discarded.

RNA concentration and purity were determined on the NanoDrop ND-1000 spectral photometer (peqlab), whereas RNA integrity was determined by capillary electrophoresis of total RNA samples, using the 2100 Bioanalyzer (Agilent Technologies).

Preparation of labelled cRNA. Total RNA (950 ng) was introduced into an RT-IVT reaction. The RNA was reverse-transcribed into ds-cDNA and then converted into labelled cRNA by *in vitro* transcription (Nano-Amp RT-IVT Labeling Kit, Applied Biosystems) using digoxigenin-conjugated UTP nucleotides (Roche Diagnostics).

Table I. Types of tumors in patients with double primary malignancies and single primary malignancies.

Primary	Second primary	No. Pts	Category
Patients with two primary tumors			
Breast	Gastrointestinal	5	AB*
	Gynecological	3	AB*
	Lymphoma	2	AB*
	Uterus	1	AB*
Ovary	Genitourinary	1	AB*
	Gastrointestinal	1	AB*
Colon	Genitourinary (prostate)	4	AB*
	Gynecological	4	AB*
	Breast	1	AB*
	Gastrointestinal	1	AB*
	NSCLC	1	AB*
Total		24	
Patients with a single tumor			
Breast		13	D
Colon		10	C
Ovary		9	E
Total		32	

All CDE category patients had at least a 5-year follow-up prior to examination. Of the AB cases, 70-83% presented with a second primary tumor within a 5-year period. *AB, group A which includes 2 of the following malignancies, breast, ovarian, colorectal as primary and second primary tumors and group B, which includes one of the aforementioned tumors and tumors of a different kind.

Microarray measurement. Digoxigenin-labeled cRNA (10 µg) samples were fragmented and hybridized for 16 h to Human Genome Survey Microarrays V2.0 (Applied Biosystems) at a temperature of 55°C. After several washings, steps of increasing stringency, an anti-digoxigenin-antibody conjugated to alkaline phosphatase (Roche Diagnostics) was added. Substrate was applied and chemiluminescence and fluorescence signals were detected on an AB500 microarray reader. Reagents from the Chemiluminescence Detection Kit (Applied Biosystems) were used in this procedure. Human Genome Survey Microarrays V2.0 feature numerous control probes and 3,698 target gene-specific 60-mer probes detecting >19,000 different genes.

Bioinformatics data analysis. The number of cases selected for each group of individuals was considered adequate to draw statistical conclusions with microarray analysis (24).

Software for bioinformatics data analysis. The software platform R (Version 2.4.0) and the Bioconductor packages ABarray (25) and Limma (26) as well as the software Spotfire Decision Site for Functional Genomics 9.0 were used for bio-statistical data analysis and visualization.

Normalization. Normalization compensates for varying global signal intensities of the microarrays and adjusts them to a



level, thus making the individual microarrays suitable for downstream analysis. Quantile normalization was applied to the data set (27).

Averaging and fold-change calculation. The quantile normalized, log₂ transformed data were averaged within the subgroups A, B, C, D, E and N. The quantile normalized, log₂ transformed data were also averaged within the category 'double cancer' (group AB), and within the category 'single cancer' (group CDE). The extent and direction of differential expression between the groups are presented in a fold-change value on the log scale (see Statistical significance below).

Correlation analysis. The similarity of global gene expression profiles of the different patient blood samples was assessed by correlation analysis. For this purpose, MA plots were generated from the quantile normalized data.

Statistical evaluation. By applying functions from the bioconductor package, Limma, linear models were fitted to the data to fully model the systematic part of the data.

To analyze the comparisons of interest, design matrices were created which assign the samples to subgroups A through N and to categories AB, CDE and N. A linear model was then fitted for every gene and contrast matrices were generated to extract statistical parameters for the different comparisons. All possible comparisons among participating groups in this study were determined.

The contrasts were extracted after fitting the linear model and were tested for statistical significance using an ANOVA-like test (moderated F-Test, empirical Bayes method) (28).

To correct for multiple testing, the FDR-based method was applied (29). The p-value adjusted for multiple testing is usually several orders of magnitude higher than the unadjusted p-value.

To allow for identification of 'significant' comparison(s), a decision matrix was created. This matrix includes all comparisons within the contrast matrix and is filled with values 0, 1 and -1. Value 1 indicates that the respective comparison achieves statistical significance in the sense of upregulation, and value -1 indicates significant downregulation. Value 0 specifies a non-significant result of the test.

To identify genes with a robust differential expression among the comparator groups, several parameters were taken into account.

Robust detection. A probe was only considered if it was detectable [signal/noise (S/N) ≥ 3] in at least two thirds (66.9%) of the samples of at least one comparator group. Furthermore, a probe was excluded if its signal value was flagged (flag value ≥ 852) in one third or more of the samples of at least one comparator group.

Annotation/biological relevance. A probe was only considered if its sequence was annotated as 'current' in the latest annotation file (Gene Expression Omnibus GSE925). All sequences annotated as 'pseudogenes' or 'obsolete' were discarded.

Statistical significance. A probe was only classified as induced in a specific comparison if the adjusted p-value was < 0.1 and

its decision matrix cell contained the value 1. Conversely, a probe was classified as repressed if the adjusted p-value was < 0.1 and the decision matrix cell contained a value of -1.

No fold change criterion was used for filtering, i.e. even genes with a very small, but still reproducible and thus significant differential expression, were retrieved. This relatively low stringency was applied to allow for identification of as many deregulated genes as possible, including those with a weak significance. Nevertheless, more stringent criteria were also applied in each subsequent comparison and genes that presented a highly statistically significant value ($< 1 \times 10^{-4}$) were selected.

Hierarchical clustering. Hierarchical clustering of a set of pre-selected genes, based on statistical significance, was performed (supervised clustering) with the purpose of visualizing differential expression between samples. The quantile normalized, log transformed signal values were used as the basis for hierarchical clustering, and the clustering method 'complete linkage' was applied within the software Spotfire Decision Site for Functional Genomics 9.0.

The identification of differentially regulated processes and pathways for all of the genes obtained as well as the enrichment of biological classes and pathways in comparison to the whole genome was carried out (<http://www.pantherdb.org>).

As a matter of principle, the percentage distributions shown in the pie charts (data not shown) do not indicate if a certain biological process, molecular function or pathway is in fact overrepresented or if they just reflect the average percentage of the respective class in the human genome. To elucidate this question, the induced gene IDs were compared to all gene IDs represented on the Human AB500 Microarray V2.0; ([http://www.pantherdb.org/Tools/Gene Express Data Analysis/Compare gene lists](http://www.pantherdb.org/Tools/GeneExpressDataAnalysis/CompareGeneLists)) and we calculated whether a specific class was overrepresented when compared to the distribution of all genes on the microarray. A p-value is given for the significance of this enrichment, based on the binomial test (30). From our experience, p-values $\sim 10^{-6}$, are regarded as a sign of manifest enrichment in the context of a Panther analysis for biological processes and molecular functions. For Panther Pathway enrichment analysis, we found that the threshold can be slightly increased to p-values $\sim 10^{-6}$.

Classification analysis. Classification analysis was performed using the support vector machine (SVM) paradigm (31). We performed SVM on the highest quality subset of probes measured in the Human Genome Survey Microarray V2.0. This subset was determined as those probes that had an S/N ratio of ≥ 3 and a flag value of ≥ 852 , for all samples tested. This quality filter was passed by a total of 4,121 probes. The quantile-normalized expression levels of these probes were then standardized using the function scale within R 2.5.1., resulting in a set of 4,121 variables of which all had a mean of 0 and a standard deviation of 1, thus giving each probe an equal chance to enter as a factor into a classifier.

To build the classifiers for the various combinations of outcomes compared, the following procedure was chosen: i) application of an ANOVA analysis on the standardized quantile normalized probe expression levels and sorting the probes by p-values from the analysis, ii) from the subset of

Table II. Results of the pair-wise comparison among cancer groups and healthy individuals.

	vs CDE	Gene (Gene ID)	X CDE	P-value	vs N	Gene (Gene ID)	X N	p-value	
AB	1307↑	RAP1GA1 (5909)	3.38	0.001	2531↑	CA1 (759)	3.55	0.006	
		CA1 (759)	2.89	0.024		CA7 (766)	3.16	0.02	
		CA7 (766)	2.87	0.006		RAP1GA1 (5909)	3.11	0.001	
	1411↓	BPGM (669)	2.87	0.01	2375 ↓	hCG2040151	0.159	0.005	
		CTNNAL1 (8727)	2.51	0.02		hCG1791834.2	0.204	0.002	
		FCRLM1 (84824)	0.51	0.03		PRO (1073)	0.205	0.003	
CDE		BLK (640)	0.52	0.01	975↑	FLT (32322)	2.46	0.01	
		IGHD (3495)	0.52	0.06		SF3B4 (10262)	2.35	6.92x10 ⁻⁵	
						ZNF705A (440077)	2.19	0.008	
						LY6E (4061)	2.19	0.006	
						IFITM3 (10410)	2.13	0.07	
						1370↓	hCG2040151	0.19	0.005
						hCG1791834.2	0.25	0.002	
						RPL22L1 (200916)	0.28	0.0001	
						SUB1 (10923)	0.28	3.3x10 ⁻⁷	

Genes presenting the highest linear deregulation are quoted here.

the 300 probes with the lowest p-values, choice of a further subset by means of a reversible direction Monte-Carlo (32) adding, sampling and reduction method developed especially for this study, iii) quality of classifiers was judged by the cross-validation error, which was aimed to be minimized. The number of cross-validations for each step was set to half the effective sample size, i.e. the number of samples with class designations being part of the combination of the outcome investigated. At this point it should be noted that the accuracy of prediction in the total set of data was 100% in all cases (i.e. the training error was 0 in all cases).

The relative importance of a probe was calculated as follows. First, the cross-validation error for the total classifier was recorded, then in turn, each of the probes was left out of the classifier. For each of these reduced classifiers, the cross-validation error, which was expected and indeed found to be increased in every case, was recorded. The cross-validation errors of the reduced classifiers were then normalized such that the probe with the largest increase in cross-validation error was assigned a relative importance of 1.00 with all the other probes having increasingly lower importance slowly approaching 0 for the least important variables. This normalization was done by dividing each of the probe-specific increases in error by the maximum increase in error. These normalized values are designated as relative importance.

Out of a total of 65 classifiers that were built, the three most important denominated by the diagnostic categories used in their building, are presented here. These are, AB vs CDE; AB vs N; CDE vs N.

Validation analysis. The levels of expression of the genes determined by the classification analysis of groups AB vs CDE, AB vs N, CDE vs N, together with the genes that yield a high

statistical significance after the comparisons of AB vs CDE were validated by quantitative real-time RT-PCR analysis.

Reverse transcription (RT). The High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) was used for reverse transcription of total RNA into single stranded cDNA with the aid of random hexamer primers according to the manufacturer's instructions. Total RNA (2 µg) was introduced into the reaction. Two separate RT reactions were carried out for each sample to generate enough cDNA for the analysis of all 20 assays.

Quantitative real-time PCR. Diluted cDNA templates (1.5 µl of 1:1) were analyzed in each real-time PCR reaction in a total reaction volume of 20 µl. TaqMan® Gene Expression Master Mix (Applied Biosystems) and the TaqMan® assays were used for real-time PCR amplification (note, for real-time PCR, the TaqMan assays were employed separately and not in a pool, and their final concentration in the reaction was 1x).

The amplifications were run in 284-well format on an AB9900HT instrument (Applied Biosystems). The software SDS 2.2 was used for instrument control, data acquisition and primary analysis. The plates were run in the format Relative Quantification ($\Delta\Delta Ct$), and the following temperature profile was used: 50°C/2 min, 95°C/10 min, (95°C/0:3 min, 60°C/1 min) 40 cycles PCR cycle number was set to 40, and all PCR reactions were run in triplicate. To control for cross-contamination and to rule out non-specific amplification and detection, no template controls ('NTC', H₂O instead of cDNA) were included on every microplate and for each TaqMan assay. The relative expression levels were determined in each case in relation to two housekeeping genes the β -actin (ACTB) and the glyceraldehyde-3 phosphate dehydrogenase (GAPDH).



A, Enriched biological processes

	vs CDE		vs N	
	Induced	Repressed	Induced	Repressed
AB	Nitrogen utilization	Protein Metabolism Protein modification Nucleic acid metabolism	Signal transduction Developmental processes Cell structure and motility	Protein metabolism Protein modification Nucleic acid metabolism Protein trafficking Oxidative phosphorylation Protein targeting and localization
CDE			Anterior/posterior patterning Mesoderm development Cell cycle control Segment specification	Protein metabolism Protein modification Nucleic acid metabolism Oxidative phosphorylation General vesicle transport

B, Enriched molecular functions

	vs CDE		vs N	
	Induced	Repressed	Induced	Repressed
AB	No significant enrichment	Nucleic acid binding Transferase Isomerase Synthase Synthetase	Signalling molecule Structural protein	Nucleic acid binding Chaperone Isomerase Synthase Synthetase
CDE			Other transcription factor	Nucleic acid binding Chaperone Small GTPase Synthase Synthetase Other proteases

C, Enriched pathways

	vs CDE		vs N	
	Induced	Repressed	Induced	Repressed
AB	No significant enrichment	Heme biosynthesis UPP Apoptosis signalling pathway	No significant enrichment	Parkinson's disease UPP
CDE			TGF β signalling pathway Insulin/IGF pathway Protein kinase B Signalling cascade	UPP

Results

Microarray raw data. Raw and normalized microarray data are included in Gene Expression Omnibus GSE925.

MA plots. The similarity of the expression profiles of samples within each subgroup was analyzed by creating MA plots and determining the Pearson's correlation coefficient r (data not shown).

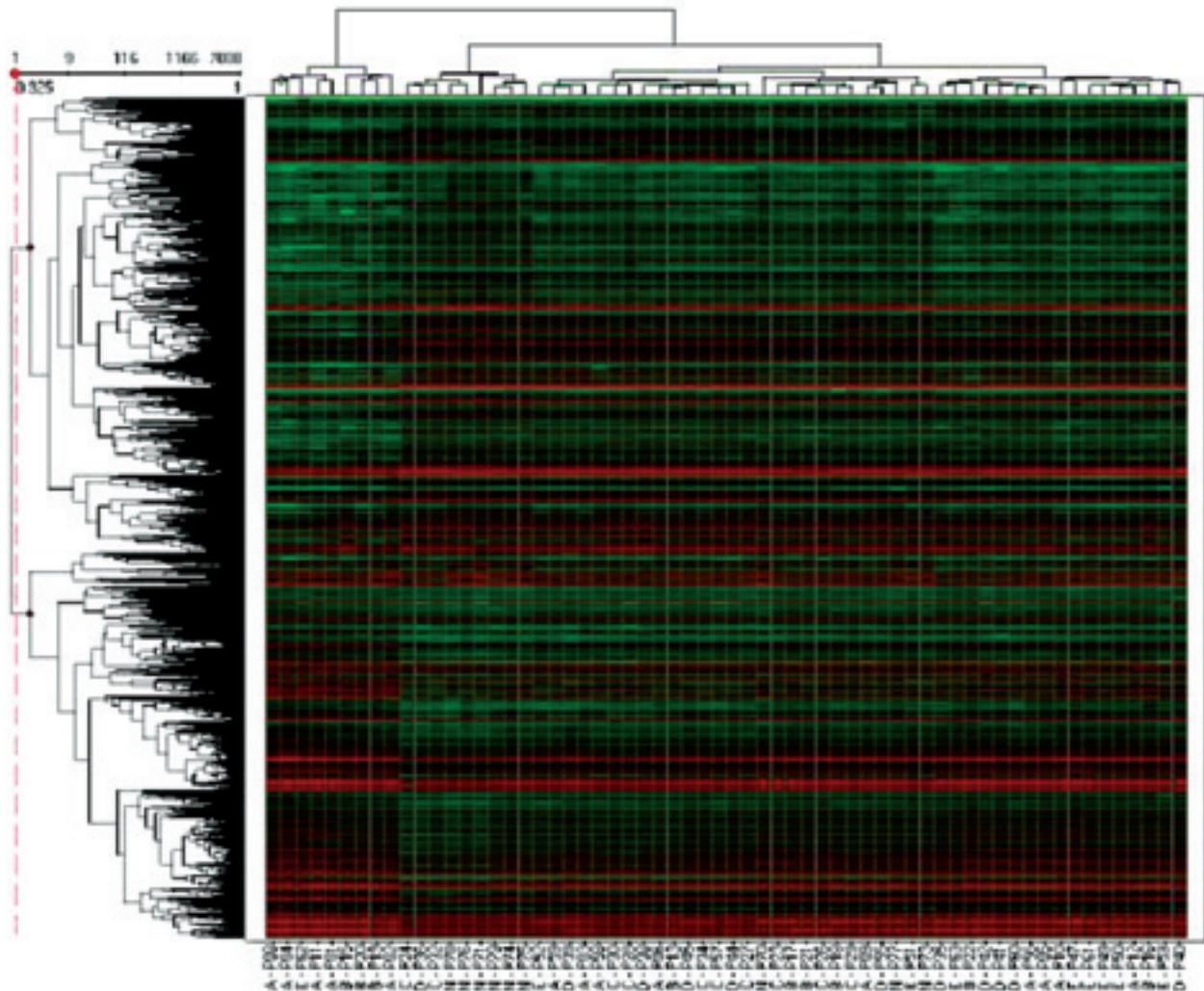


Figure 1. HeatMap 7088IDs. The subgroup and patient IDs are indicated for each sample and the five major sub-clusters are labelled. The leftmost cluster is composed mainly of samples from subgroup A and B (double cancers), however, other samples from subgroups A and B are more or less evenly distributed over the clusters. This indicates that the samples in the leftmost cluster make the strongest contribution to the differential expression between the subgroups and categories. The cluster two contains the majority of control samples (group N).

Thorough examination of the MA plots of subgroup B revealed that sample P14 (patient's individual number) deviates rather strongly from the other samples in the same subgroup and is considered as an outlier. Sample 14 was therefore excluded from further analysis. In-depth analysis of the MA plots of subgroup D revealed that samples P32 and P54 show a relatively strong deviation from the other samples in the same subgroup and are considered as outliers. Both samples were therefore excluded from further analysis.

Sample N92 (patient's individual number) was identified as an outlier and was excluded from further analysis. After exclusion of the four outlier samples described above, all data were re-normalized and the complete statistical analysis was based on the new data set comprising 62 samples.

Differential gene expression. The results of the pair-wise comparisons presented in the majority shows a vast number of deregulated probes (Table II). The probes that presented the greatest deviation are examined here. The comparison of groups A and B presented a fairly small deviation which in the

Panther analysis did not yield a statistically significant difference regarding gene up- and downregulation and the functional analyses.

Probe selection using the adjusted p-value threshold <0.1 , yielded a large number of probes in every comparison. By increasing the stringency of the selection to $p < 1 \times 10^{-4}$, four probes were determined to be repressed by comparing the individuals of group AB and CDE (data not shown); these probes corresponded to the following genes: a) HCG2040681 (unknown), b) HNRPC, c) SET and d) HSPE1. None of the induced probes for the same comparison fulfilled the high stringency selection criteria.

The genes induced in each pair-wise comparison were then subjected to functional analysis. Their participation in biological processes, their molecular functions and their role in signalling or metabolic pathways were assessed based on their annotations in the Panther classification system (Table III). An analogous approach was used for all other pair-wise comparisons in this report. The HeatMap obtained by our study is shown in Figs. 1 and 2.

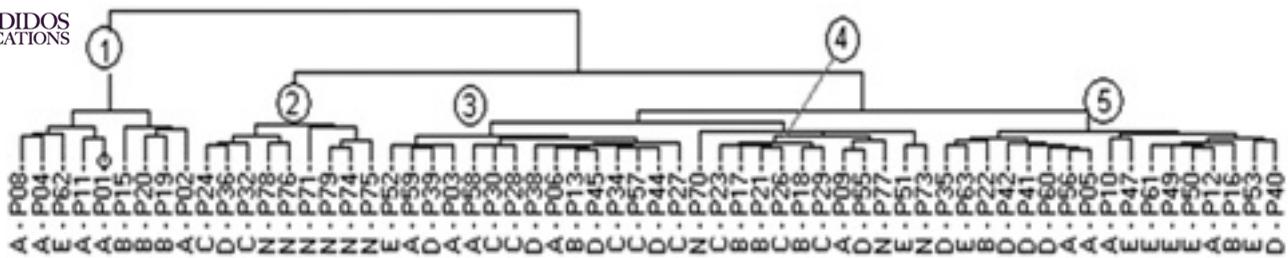


Figure 2. HeatMap 7088IDs in detail.

Table IV. Probe classifier analysis.

A, Individuals with two primary tumors (AB) vs individuals with one primary tumor (CDE)

Probe	Primary gene ID	Relative importance	P-value	Gene symbol	Gene name	Cytoband
113600	54516	1.00	0.0063	MTRF1L	mitochondrial translational release factor 1-like	6q25-q26
120515	150684	0.94	0.006	COMMD1	copper metabolism (Murr1) domain containing 1	2p15
112738	345630	0.47	0.051	LOC345630	Null	5q34
104890	84890	0.71	0.0006	C10orf22	chromosome 10 open reading frame 22	10q21.3
113628	56910	0.47	0.013	STARD7	START domain containing 7	2q11.2
120859	115703	0.35	0.003	SNX26	sorting nexin 26	19q13.12
119802	hCG19010.4	0.35	0.004	Null	Null	Null
122149	9531	0.12	0.0018	BAG3	BCL2-associated athanogene 3	10q25.2-q26.2
115163	91574	0.12	0.049	FLJ38663	Null	12q24.31

B, Individuals with two primary tumors (AB) vs healthy individuals (N)

Probe	Primary gene ID	Relative importance	P-value	Gene symbol	Gene name	Cytoband
118159	347918	1.00	0.0009	FLJ33915	Null	12q24.33
110926	254170	0.62	0.0040	FBXO33	F-box protein 33	14q21.1

C, Individuals with one primary tumor (CDE) vs healthy individuals (N)

Probe	Primary gene ID	Relative importance	P-value	Gene symbol	Gene name	Cytoband
104427	56942	1	0.00003	ENY2	enhancer of yellow 2 homolog (<i>Drosophila</i>)	8q23.1
107464	84640	0.33	0.0053	USP38	ubiquitin specific peptidase 38	Null
112937	hCG39909.2	0.22	0.0010	Null	Null	Null

Results of classification analysis. The number of factors varied from 2 to 9 which is comparatively low, resulting in a fairly efficient classification process. The training accuracy was 100% in every case and the cross validation accuracy was 100% in most of the cases. The corresponding genes and

their functional characteristics were determined. The probes being part of the classifier were ranked by their relative importance for each of the classifiers (results not presented). The probes obtained in each classifier, including the details of the corresponding genes, are shown analytically in Table IV.

Validation. The deregulation levels after the comparison of groups AB vs CDE, of the 9 genes determined by the classification analysis (Table IVa), as well as that of the four genes whose expression difference presented $p < 1 \times 10^{-4}$, were validated by quantitative RT-PCR. It was found that all 9 genes determined by classification analysis presented a similar direction of deregulation (overexpression or downregulation) with the one determined by the microarray analysis and of similar magnitude. This was not the case for the four genes, in which case although the two groups presented differences amongst them, the direction of deregulation was not the same and with the exception of the SET gene, the others presented an overexpression in the group CDE instead of downregulation.

Discussion

In the present study, the expression signatures of individuals with two primary tumors (group AB), including colon, breast and ovarian malignancy as one of the two primaries, were examined and compared with individuals with a single primary tumor, either breast, colon or ovarian (group CDE). Additionally, the individuals of group AB were also compared with each tumor category separately, colon (group C), breast (group D) and ovarian (group E), as well as to healthy individuals.

Microarray findings of the group of individuals with two primary carcinomas compared to the group of individuals with one primary carcinoma, led to the determination of a large number of deregulated genes.

We evaluated certain genes whose volume of deregulation may be of importance in the development of second primary tumors. The criterion of the latter selection was a linear expression value of < 0.53 with respect to repression and > 2.5 with respect to induction. The transcripts that presented the highest downregulation in group AB, when compared to group CDE, are FCRLM1, BLK and IHGD. These mostly correspond to immunologically relevant proteins and they seem to be associated with the proper functioning of B cells. The FCRLM1 (Fc receptor-like and mucin-like kinase) was found to be one of the most repressed genes in the AB group. This receptor has a fundamental role as an immunomodulatory agent in healthy and subverted B lineage cells and has been identified as being involved in malignancies (33). BLK is a member of the SRC family of proto-oncogenes and on the basis of its preferential expression in B-lymphoid cells, it functions in a signal transducing pathway specific to this lineage and it has been considered as a candidate for carcinogenesis (34). IHGD is the heavy chain of immunoglobulin D which is also localized on the surface of naïve B cells and it functions as an antigen receptor in apparent redundancy with IgM (35).

Concerning the functional analysis of deregulated genes, three pathways were found to be deregulated in group AB when compared to group CDE. The first is the heme biosynthesis pathway. It has been suggested that the activation of the heme catabolic pathway has beneficiary effects on cancer prevention through the action of bilirubin, a by-product of heme catabolism (36).

The ubiquitin proteasome pathway (UPP) was also down-regulated. The oncogenes of several known malignancies (for example, colon cancer), encode for ubiquitin ligase components and alterations in the UPP may lead to cancer (37). The third is

the apoptosis signalling pathway (Table III). In the comparison of group AB vs healthy individuals (N), by taking as a criterion the amount of linear expression of each statistically significant transcript, a 2.89-fold increase was observed in the transcripts of catenin, which is the major machinery for cell-cell adhesion in humans (38). Additionally, for both groups AB and CDE, when compared to N, the highest mRNA induction observed was that of the Rap1-GAP and the transcripts of carbonic anhydrases I and VI. The former was identified as an inactivator of Rap1 activity, a putative endogenous antagonist of Ras proteins. Similarly to catenin, the Rap1-GAP, regulates integrin-mediated and other cell adhesion processes (39). The primary function of the latter, as was shown in animals, is to interconvert carbon dioxide and bicarbonate in order to maintain the acid-base balance in the blood and other tissues, and to help transport carbon dioxide out of tissues (40). The fact that catenin is overexpressed in patients with 2 primary malignant tumors in comparison to healthy individuals and to individuals with one malignant tumor, may indicate a lower potential for metastasis when compared to individuals with one primary malignancy.

The genes that presented the greatest repression when group CDE was compared to healthy individuals, but not in AB vs N were: a) Antisense oligonucleotides against STK1, the human homologue of FMS related tyrosine kinase 3 (Flt 3), inhibit hematopoietic colony formation suggesting that Flt 3 may function as a growth factor receptor in hematopoietic and progenitor cells and it has also been associated with childhood leukemia (41) and b) the splicing factor 3b, known to be localized in the nucleus and involved in RNA splicing, binds BMPR-IA and specifically inhibits BMP-mediated osteochondral cell differentiation and it was found to be involved in prostate and ovarian cancer (42). Regarding the functional analysis of the repressed genes resulting from the comparison of individuals with two primary malignancies (group AB) and all of the individuals with one malignancy (group CDE) when compared to the group of healthy individuals (AB vs N, CDE vs N), it was determined that both cancer groups AB and CDE showed the UPP repressed in comparison to group N and that the AB patient group showed a repression in the Parkinson's disease pathway that was not observed in the CDE patient group and which has not previously been associated with cancer.

The classification analysis was performed in order to develop a useful prognostic tool that will allow us to determine healthy individuals or individuals with colon breast or ovarian cancer who are prone to developing a second primary tumor. Four out of the nine genes (COMMD1, Loc155630, c10orf1, hCG5010.4) are of unknown biological function and relation to cancer. The remaining five genes, MTRF1L, FLJ28663, have been related to protein biosynthesis and have not as yet been related with cancer development. STARD9 gene has been related to lipid and fatty acid transport and was found to be overexpressed in human trophoblast cells in choriocarcinoma (43). BAG 3 gene was found to be involved with protein folding and it also has an established anti-apoptotic function in different tumor cell lines. The high levels of BAG3 protein seen in some epithelial cancer cell lines may be relevant to mechanisms of tumor invasion and metastasis. It was also suggested that BAG3 is involved in the cell cycle regulation (27,45). SNX26 belongs to a family of small GTPases (Rho



and although it is involved in intracellular signalling and cell growth (46), it has not yet been associated to cancer development. The results of the classification analysis, were confirmed with qRT-PCR validation.

Regarding the genes that presented $p < 1 \times 10^{-4}$ (groups AB vs CDE), the gene ID HCG2040681 represents a new, unclassified gene that, to our knowledge, is the first time it has been associated with cancer.

HNRPC, the eukaryotic CPN10 homolog is essential for mitochondrial protein biogenesis, together with CPN60. It binds to CPN60 in the presence of Mg-ATP and suppresses the ATPase activity of the latter and it has been shown to be upregulated in the case of breast cancer (47).

SET (template activator factor-I β) is a highly conserved nuclear phosphoprotein that is ubiquitously expressed (48). SET has been suggested to regulate G2/M transition by modulating cyclin B-CDK1 activity (49) in addition SET is overexpressed in solid tumors of the breast, stomach, uterus and rectum (50).

Heparanase HSPE 1 is a proangiogenic growth factor that together with VEGF is considered to be involved in the metastatic biology of ovarian cancer (51). These four genes are downregulated in the single cancer group compared to the normal individuals and they are further downregulated in the individuals with second primary cancer when compared to the individuals with single primary malignancies.

In conclusion, the outcome of the present study was 9 genes with a statistically significant difference in expression between individuals with double primary malignancies compared to individuals with single primary malignancies, detected by whole genome microarray analysis. Nine of those were confirmed by the classifier analysis. That was confirmed by quantitative RT-PCR validation. Regarding the biological function known, 8 out of these 13 genes are defining a direct or indirect relation to cancer development.

Acknowledgements

The microarray and statistical analysis were done by IMG Co. in Martinsreid, Germany, www.imgm.com. We would like to thank the University of Athens, Medical School for their financial support.

References

- Liu YY, Chen YM, Yen SH, Tsai CM and Peng RP: Multiple primary malignancies involving lung cancer - clinical characteristics and prognosis. *Lung Cancer* 35: 189-194, 2002.
- Watanabe S, Kodama T, Shimosato Y, Arimoto H, Sugimura T, Suemasu K and Shiraiishi M: Multiple primary cancers in 5,456 autopsy cases in the National Cancer Center of Japan. *Natl Cancer Inst* 72: 1021-1027, 1984.
- Engeland A, Bjorge T, Haldorsen T and Tretlis S: Use of multiple primary cancers to indicate associations between smoking and cancer incidence: an analysis of 500,000 cancer cases diagnosed in Norway during 1953-1993. *Int J Cancer* 70: 401-407, 1997.
- Curtis RE, Boic JD, Kleinerman RA, Flannery JT and Fraumeni JF Jr: Multiple primary cancers in Connecticut and Denmark. *NCI Monogr* 68: 219-242, 1985.
- Isinger A, Bhat M, Borg A and Nilbert M: CHEK2 1100 delC in patients with metachronous cancers of the breast and the colorectum. *BMC Cancer* 15: 6-64, 2006.
- Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM and Ding W: A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266: 66-71, 1994.
- De la Chapelle A: Genetic predisposition to colorectal cancer. *Natl Rev Cancer* 4: 769-780, 2004.
- Peltomaki P and Vasen H: Mutations associated with HNPCC predisposition - update of ICG - HNPCC/INSIGHT mutation database. *Dis Markers* 20: 269-276, 2004.
- CHEK2 Breast Cancer Case-Control Consortium. CHEK2-1100 delC and susceptibility to breast cancer: a collaborative analysis involving 10,680 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet* 74: 1175-1182, 2004.
- Lynch HT, Krush AJ, Lemon HM, Kaplan AR, Condit PT and Bottomley RH: Tumor variation in families with breast cancer. *JAMA* 222: 1631-1635, 1972.
- Meijers-Heijboer H, Wijnen J, Vasen H, Wasielewski M, Wagner A, Hollestelle A, Elstrodt F, van den Bos R, de Snoo A, Fat GT, Brekelmans C, Jagmohan S, Franken P, Verkuiljen P, van den Ouweland A, Chapman P, Tops C, Möslein G, Burn J, Lynch H, Klijn J, Fodde R and Schutte M: The CHEK2 1100 delC mutation identifies families with a hereditary breast and colorectal cancer phenotype. *Am J Hum Genet* 72: 1308-1314, 2003.
- Brisson O: Gene amplification and tumor progression. *Biochim Biophys Acta* 1155: 25-41, 1993.
- Richard F, Pacyna-Gengelbach M, Schlüns K, Fleige B, Winzer KJ, Szymas J, Dietel M, Petersen I and Schwendel A: Patterns of chromosomal imbalances in invasive breast cancer. *Int J Cancer* 89: 305-310, 2000.
- Climent J, Martinez-Climent JA, Blesa D, Garcia-Barchino MJ, Saez R, Sánchez-Izquierdo D, Azagra P, Lluch A and Garcia-Conde J: Genomic loss of 18p predicts an adverse clinical outcome in patients with high-risk breast cancer. *Clin Cancer Res* 8: 3863-3869, 2002.
- Verhoeven E, Mooi WJ and Michalides RJ: Identification and cloning of two overexpressed genes, U21B31/PRAD1 and EMS1, within the amplified chromosome 11q13 region in human carcinomas. *Oncogene* 7: 355-361, 1992.
- Bärlund M, Monni O, Kononen J, Cornelison R, Torhorst J, Sauter G, Kallioniemi OP and Kallioniemi A: Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. *Cancer Res* 60: 5340-5344, 2000.
- Cuny M, Kramer A, Courjal F, Johannsdottir V, Iacopetta B, Fontaine H, Grenier J, Culine S and Theillet C: Relating genotype and phenotype in breast cancer: an analysis of the prognostic significance of amplification at eight different genes of loci and p53 mutations. *Cancer Res* 60: 1077-1083, 2000.
- Bärlund M, Monni O, Weaver JD, Kauraniemi P, Sauter G, Heiskanen M, Kallioniemi OP and Kallioniemi A: Cloning of BCAS3 (17q23) and BCAS4 (20q13) genes that undergo amplification, overexpression, and fusion in breast cancer. *Genes Chromosom Cancer* 35: 311-317, 2002.
- Nessling M, Richter K, Schwaenen C, Roerig P, Wrobel G, Wessendorf S, Fritz B, Bentz M, Sinn HP, Radlwimmer B and Lichter P: Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer* 65: 429-447, 2005.
- Sharma P, Sahni NS, Tibshirani R, Skaane P, Urdal P, Berghagen H, Jensen M, Kristiansen L, Moen C, Sharma P, Zaka A, Arnes J, Sauer T, Akslén LA, Schlichting E, Børresen-Dale AL and Lönneborg A: Early detection of breast cancer on gene-expression patterns in peripheral blood cells. *Breast Cancer Res* 7: R634-R644, 2005.
- Gitsch G, Hanzal E, Jensen D and Hacker NF: Endometrial cancer in premenopausal women 45 years and younger. *Obstet Gynecol* 85: 504-508, 1995.
- Pearl ML, Johnson CM, Frank TS and Roberts JA: Synchronous dual primary ovarian and endometrial carcinomas. *Int J Gynaecol Obstet* 42: 305-312, 1993.
- Falkenberry SS, Steinhoff MM, Gordinier M, Rappoport S, Gajewski W and Granai CO: Synchronous endometrioid tumors of the ovary and endometrium. A clinicopathologic study of 22 cases. *J Reprod Med* 41: 713-718, 1996.
- Dobbin K and Simon R: Sample size determination in microarray experiments for class comparison and prognostic classification. *Biostatistics* 6: 27-38, 2005.
- Yongming AS: <http://bioconductor.org/packages/1.9/bioc/html/ABarray.html>.
- Smyth GK and Limma: Linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds). 1st edition. Springer, New York, pp397-420, 2005.

27. Bolstad BM, Irizarry RA, Astrand M and Speed TP: A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. *Bioinformatics* 19: 185-193, 2003.
28. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments, *Statistical Applications in Genetics and Molecular Biology (SAGMB)* 3 No. 1, Article 3, 2004.
29. Benjamini Y and Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57: 289-300, 1995.
30. Cho R and Campbell M: Transcription, genomes, function. *Trends Genet* 16: 409-415, 2000.
31. Cortes C and Vapnik V: Support-vector network Machine Learning 20: 1-25, 1995.
32. Robert, C. P and Casella G. *Monte Carlo Statistical Methods* (2nd ed). Springer, New York, 2004.
33. Du X, Nagata S, Ise T, Stetler-Stevenson M and Pastan I: FCRL1 on chronic lymphocyte leukemia, hairy cell leukemia and B cell non Hodgkin lymphoma as a target of immunotoxins. *Blood* 111: 338-342, 2008.
34. Malek SN, Dordai DI, Reim J, Dintzis H and Desiderio S: Malignant transformation of early lymphoid progenitors in mice expressing an activated Blk tyrosine kinase. *Proc Natl Acad Sci USA* 95: 7351-7356, 1998.
35. Tschumper RC, Geyer SM, Campbell ME, Kay NE, Shanafelt TD, Zent CS, Nowakowski GS, Call TG, Dewald GW and Jelinek DF: Immunoglobulin diversity gene usage predicts unfavorable outcome in a subset of chronic lymphocytic leukemia patients. *J Clin Invest* 118: 306-315, 2008.
36. Vitek L and Schwertner HA: The Heme catabolic pathway and its protective effects on oxidative stress mediated diseases. *Adv Clin Chem* 42: 1-57, 2007.
37. Mani A and Gelmann EP: The Ubiquitin-Proteasome pathway and its role in cancer. *J Clin Oncol* 23: 4776-4789, 2005.
38. Yanagisawa M, Huveldt D, Kreinest P, Lohse CM, Cheville JC, Parker AS, Copland JA and Anastasiadis PZ: p120 catenin isoform switch affects Rho activity, induces tumor cell invasion and predicts metastatic disease. *J Biol Chem* 283: 18344-18354, 2008.
39. Jeon TJ, Lee DJ, Lee S, Weeks G and Firtel RA: Regulation of Rap1 activity by RapGAP1 controls cell adhesion at the front of chemotaxing cells. *J Cell Biol* 179: 833-842, 2007.
40. Pastorekova S, Zatovicova M and Pastorek J: Cancer-associated carbonic anhydrases and their inhibition. *Curr Pharm Des* 14: 685-698, 2008.
41. Small D, Levenstein M and Kim E: STK-1, the human homolog of Flk-2/Flt-3, is selectively expressed in CD34+ human bone marrow cells and is involved in the proliferation of early progenitor/stem cells. *Proc Nat Acad Sci USA* 91: 459-463, 1994.
42. Watanabe H, Shionyu M, Kimura T, Kimata K and Watanabe H: Splicing factor 3b subunit 4 binds BMPR-IA and inhibits osteochondral cell differentiation. *J Biol Chem* 282: 20728-20738, 2007.
43. Angeletti S, Rena V, Nores R, Fretes R, Panzetta-Dutari GM and Genti-Raimondi S: Expression and localization of StarD7 in trophoblast cells. *Placenta* 29: 396-404, 2008.
44. Gentilella A, Passiatore G, Deshmane S, Turco MC and Khalili K: Activation of BAG3 by Egr-1 in response to FGF-2 in neuroblastoma cells. *Oncogene* 27: 5011-5018, 2008.
45. Iwasaki M, Homma S, Hishiya A, Dolezal SJ, Reed JC and Takayama S: BAG3 regulates motility and adhesion of epithelial cancer cells. *Cancer Res* 67: 10252-10259, 2007.
46. Liu H, Nakazawa T, Tezuka T and Yamamoto T: Physical and functional interaction of Fyn Tyrosine kinase with a brain-enriched Rho GTPase-activating protein TCGAP. *J Biol Chem* 281: 23611-23619, 2006.
47. Zucchi I, Mento E, Kuznetsov VA, Scotti M, Valsecchi V, Simionati B, Vicinanza E, Valle G, Pilotti S, Reinbold R, Vezzoni P, Albertini A and Dulbecco R: Gene expression profiles of epithelial cells microscopically isolated from a breast-invasive ductal carcinoma and a nodal metastasis. *Proc Nat Acad Sci USA* 101: 18147-18152, 2004.
48. Adachi Y, Pavlakis GN and Copeland TD: Identification and characterization of SET a nuclear phosphoprotein encoded by the translocation break point in acute adifferentiated leukaemia. *J Biol Chem* 269: 2258-2262, 1994.
49. Canela N, Rodriguez-Villarupla A, Estanyol JM, Diaz C, Pujol MJ, Agell N and Bachs O: The SET protein regulates G2/M transition by modulating cyclin B-cyclin dependent kinase 1 activity. *J Biol Chem* 278: 1158-1164, 2003.
50. Ozbek U, Kandilci A, van Baal S, Bonten J, Boyd K, Franken P, Fodde R and Grosveld GC: SET-CAN the product of t(9;9) in acute undifferentiated leukaemia, causes expansion of early haematopoietic progenitors and hyperproliferation of stomach mucosa in transgenic mice. *Am J Pathol* 171: 654-666, 2007.
51. Ralph S, Brenchley PE, Summers A, Rosa DD, Swindell R and Jayson GC: Heparenase gene haplotype (CGC) is associated with stage of disease in patients with ovarian carcinoma. *Cancer Sci* 98: 844-849, 2007.