EUROPEAN JOURNAL OF CANCER 44 (2008) 1761–1769

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Expression profiling of angiogenic genes for the characterisation of colorectal carcinoma

Alessandro Carrer[a], *Serena Zacchigna*[a], *Alessandro Balani*[b], *Valentina Pistan*[b], *Adelino Adami*[b], *Fabio Porcelli*[b], *Monica Scaramucci*[b], *Mauro Roseano*[b], *Angelo Turoldo*[b], *Maria Cristina Prati*[c], *Matteo Dell'Omodarme*[c], *Nicolo' de Manzini*[b], *Mauro Giacca*[a,d,*]

[a]*Molecular Medicine Laboratory, International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park, Padriciano 99, 34012 Trieste, Italy*
[b]*Department of Surgery and Anesthesiology, Faculty of Medicine, University of Trieste, Italy*
[c]*Scuola Normale Superiore and INFN, Sezione di Pisa, Italy*
[d]*On leave of absence from the Department of Biomedicine, Faculty of Medicine, University of Trieste, Italy*

ARTICLE INFO

Article history:
Received 3 May 2008
Accepted 21 May 2008
Available online 23 July 2008

Keywords:
Angiogenesis
Cancer
Chemokines
Expression profile
Lymphangiogenesis

ABSTRACT

The development of new blood and lymphatic vessels is a crucial event for cancer growth, metastatic spread and relapse after therapy. In this work, the expression levels of chemokines, angiogenic and angiostatic factors and their receptors were determined in paired mucosal and tumour samples of patients with colorectal carcinoma and correlated with clinical and histological parameters by advanced multivariate analyses. The most important predictors to discriminate between tumour and paired normal mucosa turned out to be the levels of expression of plexin-A1 and stromal cell-derived factor 1 (SDF-1), the former overexpressed and the latter downregulated in tumours. The levels of osteopontin and Tie-2 transcripts discriminated between the presence and absence of lymph node infiltration, the former overexpressed in the presence of infiltration whilst the latter providing a protective role. These results add support to the notion that the expression levels of selected genes involved in new blood and lymphatic vessel formation represent trustable biomarkers of tumour development and invasion and contribute to the identification of novel molecular classifiers for colorectal carcinoma.

1. Introduction

Colorectal cancer is one of the leading cancer-related causes of death, responsible for approximately 400,000 deaths each year worldwide.[1] Surgery remains the only curative treatment to date and, despite adequate intervention, disease recurs in nearly half of the patients within 5 years, mainly because of undetected metastatic spread.[1,2] Thus, a better understanding of the molecular mechanisms underlying tumour development and metastasis is essential for both diagnostic and prognostic purposes.

Angiogenesis is a crucial event in tumour progression and dissemination.[3] When solid tumours grow, their core becomes hypoxic and activates a transcriptional cascade leading to new blood vessel formation,[4] which is associated with poor prognosis and relapse of the disease.[3] Amongst the genes responsible of tumour-associated angiogenesis, the members of the vascular endothelial growth factor (VEGF)

* *Corresponding author:* Address: Molecular Medicine Laboratory, International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park, Padriciano 99, 34012 Trieste, Italy Tel.: +39 040 375 7324; fax: +39 040 375 7380.
E-mail address: giacca@icgeb.org (M. Giacca).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2008.05.014

family (VEGF-A, B, C, D, E and PlGF – placental growth factor) play a key role.[5] Whereas it is generally agreed that vascular endothelial growth factor receptor 2 (VEGFR-2)/KDR/Flk-1 is the major receptor mediating the angiogenic effects of VEGF-A,[5] the role of VEGFR-1/Flt-1 (a receptor shared by VEGF-A, VEGF-B and PlGF) is less understood and variably proposed as an inhibitor or as an inducer of angiogenesis; recently, a specific role of Flt-1 in mediating vascular permeability, as well as mononuclear and tumour cell recruitment has been suggested.[5,6] VEGF-C and VEGF-D are key regulators of lymphatic sprouting, mainly through their interaction with VEGFR-3.[7] It has become increasingly apparent that VEGF receptors act in concert with other molecules, belonging to the integrin, cadherin, neuropilin and Ephrin receptor families.[8,9] Consistently, different ephrins and semaphorins (neuropilin ligands), originally identified as axonal guidance cues, have recently been shown to affect vascular development.[10]

In addition to VEGF, other growth factors participate to the angiogenic process, including members of the fibroblast growth factor (FGF) family[11] and angiopoietins, which control vessel stability and maturation by interacting with their Tie receptors.[12–15]

Chemokines also regulate both tumour angiogenesis and lymphangiogenesis, either by directly stimulating endothelial cell proliferation, as is the case of interleukin-8 (IL-8),[16] or by mediating mononuclear cell recruitment (e.g. monocyte chemotactic protein-1 (MCP-1)[17]) and retention (e.g. stromal cell-derived factor 1 (SDF-1)[8]). Finally, growing vessels are invariably accompanied by extracellular matrix remodelling by matricellular proteins, such as thrombospondin-1 and osteopontin.[19–21]

As all these molecules regulate angiogenesis in experimental animal tumours, their use as clinical biomarkers of cancer stage and disease progression is particularly attractive. However, a reliable assay to clearly define their contribution to tumour progression in humans has never been established.[22] Several of the above-mentioned factors have been studied by immunohistochemistry, which is notoriously poorly quantitative and only detects a few markers at a time.[23,24] On the other hand, microarrays provide genome-wide signatures expression profiles in cancer,[24] but the ensuing information is often redundant, poorly reproducible and difficult to translate into useful clinical practise.[25,26]

Here we wanted to assess whether colorectal carcinoma might be characterised and classified according to the expression levels of a series of selected, candidate genes, previously associated to tumour angiogenesis and lymphangiogenesis.

2. Materials and methods

2.1. Patients selection and sample collection

Consecutive patients scheduled to undergo surgical resection of tumour mass at the Department of Surgery of the Azienda Ospedaliero-Universitaria 'Ospedali Riuniti di Trieste' of Trieste, Italy were recruited for this study, with no specific criteria for exclusion. All patients were enroled according to protocols approved by the Ethical Committee of the Azienda Ospedaliero-Universitaria 'Ospedali Riuniti di Trieste', after written informed consent was obtained. No patients enroled in the study had received neither chemotherapy nor radiotherapy prior to surgical resection. The tumour mass was entirely removed and samples frozen at –80 °C in Solution D.[27] Apparently healthy mucosa samples were harvested for each patient, at least 15 cm away from the tumour mass.

2.2. Clinical variables and tumour staging

For each patient, tumour paraffin inclusions were sectioned and analysed by an independent group of pathologists. Tumour diagnosis and staging were determined according to commonly accepted criteria, including TNM and Dukes classification, UICC staging, grade, vascular and lymphatic extraparietal invasion, as routinely determined in all patients with colorectal cancer.

2.3. RNA extraction and reverse transcription

RNA extraction was performed according to established procedures.[27] Briefly, tissues in Solution D were homogenised for RNA extraction by phenol/chloroform, followed by isopropanol precipitation. Pellets were resuspended in DEPC-water.

All RNA samples were treated with DNAse I for 15 min at 37 °C. Accurate quantification of RNA was performed by means of RiboGreen® reagent (Molecular Probes, Eugene, OR, USA) according to manufacturer's instructions. One μg extracted RNA was mixed with the *in vitro*-transcribed murine RNA coding for RANTES (regulated upon transcription, normally T-expressed, and presumably secreted) (7 ng) as a normaliser and reverse transcribed with random hexameric primers using a commercially available kit (Invitrogen, Carlsbad, CA).

2.4. Quantitative PCR amplification

Quantification of the levels of expression of the panel of genes under investigation was performed by real-time PCR using TaqMan technology. Multiplex amplifications were carried out in an ABI Prism 7000 Sequence Detection System, using the Applied Biosystems Assay-on-Demand of Assay-by-design shown in Table 2. Data were normalised to the amount of total RNA added to each reaction. The relative expression levels were calculated according to the following equations:

$$\Delta C_T = C_T(\text{target}) - C_T(\text{normaliser});$$

$$\text{Comparative expression level (i.e.difference between tumour and mucosa)} = 2^{-\Delta\Delta CT}$$

2.5. Statistical analysis

Statistical analysis was performed on standardised data by a series of techniques,[28] which, in particular, included linear discriminant analysis (LDA) and logistic regression. Through LDA, linear combinations of genes are found, defined as linear discriminant functions, which discriminate amongst different clinical variables, assuming values on separate sets of real numbers. The functions are chosen in such a way that the ratios of between-groups to within-group variances are

maximised. Leave-one-out cross-validation[29] was used to compute the error rates (ER) of the LDA models.

In the model of logistic regression, the covariates are the levels of gene expression and the dependent binary variable separates different clinical variables in two groups.

Support vector machines (SVM) with recursive feature elimination[30] were used as an alternative, independent technique to validate results. Statistical analysis was carried out using R 2.2.0 (R Development Core Team 2005). Results were considered statistically significant if $P < 0.05$.

3. Results

3.1. Experimental strategy

Seventy-eight patients with colorectal carcinoma, scheduled to undergo surgical resection, were enroled in this study. The major clinical features of the patients are shown in Table 1, along with the histological characterisation of their tumours. During surgery, different samples were harvested from the tumour, together with a biopsy of matched, apparently healthy mucosa at a distance of at least 15 cm from the tumour border. From these samples, we assessed, by real-time quantitative polymerase chain reaction (PCR), the expression levels of 20 genes coding for angiogenic, lymphangiogenic or angiostatic factors and their receptors (VEGF-A, VEGF-C, VEGF-D, PlGF, VEGFR-1, VEGFR-2, VEGFR-3, angiopoietin-1 (Ang-1), Ang-2, Tie-2, FGF2, FGF-4, FGFR-1, FGFR-3, osteopontin, thrombospondin-1, semaphorin 3A, neuropilin-1, plexin-A1, plexin-B1), seven chemokines and their receptors (MCP-1, RANTES, IL-8, SDF-1α, C-C chemokine receptor type 2 (CCR2), CCR3, CCR5) and two housekeeping genes (glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and ribosomal 18S RNA). A complete list of gene full names and acronyms is reported in Table 2.

Table 1 – Characteristics of patients and tumour classification

Classification	Parameter	No. of patients
Sex, age	Male (average age)	40 (71)
	Female (average age)	38 (70)
Tumour localisation	Colon	61
	Rectum	17
Stage of tumour (T)	T = 1	6
	T = 2	7
	T = 3	37
	T = 4	28
Lymph nodal infiltration (N)	N = 0	46
	N = 1	14
	N = 2	12
	N = 3	6
Presence of metastasis (M)	M = 0	62
	M = 1	16
Grade (G)	G = 1	17
	G = 2	46
	G = 3	14
Dukes	D = a	10
	D = b	30
	D = c	22
	D = d	16
Global stage	S = 1	12
	S = 2	29
	S = 3	21
	S = 4	16
Extra-parietal lymphatic infiltration (EPLI)	Present	22
	Absent	52
Extra-parietal vascular infiltration (EPVI)	Present	18
	Absent	56
Lymphatic infiltration	Present	35
	Absent	43
Invasive front	Infiltrating	33
	Expanding	45

3.2. Standardisation of gene expression analysis

Given the expected variability in tumour sampling, the comparative analysis of gene expression requires reliable normalisation. We therefore quantified total RNA by fluorimetry (see Section 2) and compared three normalisation procedures. We first evaluated the expression levels the housekeeping gene GAPDH and unexpectedly observed that the expression of GAPDH was constantly higher in the tumour (2.7-fold average, $P < 0.01$) as compared to the paired mucosa samples (Suppl. Fig. 1), which rendered its use inappropriate for standardisation purposes. A similar variability was not observed for rRNA 18S expression; however, the much higher levels of this transcript compared to those of the other 27 genes assessed in the study, rendered its utilisation untrustworthy. We, therefore, decided to rely upon the addition of an external RNA molecule as a normaliser for the real-time PCR quantifications.

To tackle the issue of intra-tumour variability, we assessed whether the position of the biopsy, either at the centre or at the periphery of the primary tumour, affected the gene expression profile. In a selected group of patients ($n = 10$), three samples were harvested, one from the mucosa, one from the periphery and one from the core of the tumour. In these samples, we analysed the expression levels of a subset of genes, for which the variation between tumour samples and matched mucosae was particularly evident and common to most patients (Suppl. Fig. 2). All these genes were expressed at similar levels in both tumour regions ($P > 0.05$ in all cases; Fig. 1), indicating that a single biopsy can be considered representative of the whole mass, at least for the set of genes considered in this study.

3.3. Global changes in average gene expression levels

The analysis of the expression levels of the analysed genes provided a molecular profile for each patient. Two transcripts (CCR2 and FGF-4) were only sporadically detected and were therefore excluded from subsequent statistical analysis.

As an initial approach, we compared the expression levels of each gene in the tumour and in the normal tissue from the same patient. We found that a few genes had the tendency to be overexpressed in the tumour. These included IL-8 (100-fold average increase), PlGF (40-fold increase), osteopontin (30-fold increase) and Ang-2 (10-fold increase). In contrast, others appeared downregulated in tumours, including VEGF-D (which in several cases was undetectable in tumours, whilst always present in the matched mucosa), SDF-1 (10-fold average

Table 2 – Genes analysed in the study

Gene family	Gene	GeneBank Acc. No.	TaqMan assay
Vascular Endothelial Growth Factor (VEGF) and receptors	VEGF-A	NM_001025366	Hs00173626_m1
	VEGF-C	NM_005429	Hs00153458_m1
	VEGF-D	NM_004469	Hs00189521_m1
	PlGF (Placental Growth Factor)	NM_002632	Hs00182176_m1
	VEGFR-1 (or Flt-1)	NM_002019	Hs00176573_m1
	VEGFR-2 (or Flk-1)	NM_002253	Hs00176676_m1
	VEGFR-3 (or Flt-4)	NM_002020	Hs00176607_m1
Angiopoietins and receptors	Ang-1	NM_001146	Hs00375822_m1
	Ang-2	NM_001147	Hs00169867_m1
	Tie-2	NM_000459	Hs00176096_m1
Matrix-associated proteins	Osteopontin	NM_000582	Hs00167093_m1
	Thrombospondin	NM_003246	Hs00170236_m1
Chemokines and receptors	IL-8 (interleukin-8)	NM_000584	Hs00174103_m1
	SDF-1 (or CXCL-12)	NM_000609	Hs00171022_m1
	MCP-1 (or CCL-2)	NM_002982	Hs00234140_m1
	RANTES (or CCL-5)	NM_002985	Hs00174575_m1
	CCR2	NM_000647	Hs00174150_m1
	CCR3	NM_001837	Hs00266213_m1
	CCR5	NM_000579	Hs00152917_m1
Fibroblast Growth Factors (FGFs) and receptors	FGF-2	NM_002006	Hs00266645_m1
	FGF-4	NM_002007	Hs00173564_m1
	FGFR-1	NM_000604	Hs00241111_m1
	FGFR-3	NM_000142	Hs00179829_m1
Semaphorins and receptors	Sema 3A	NM_006080	Hs00173810_m1
	NP-1 (Neuropilin-1)	NM_001024628	Assay-by-design
	Plexin A1	NM_032242	Hs00413698_m1
	Plexin B1	NM_002673	Hs00367063_m1
Housekeeping genes	r18S	NM_022551	
	GAPDH	NM_002046	

The gene family, gene name and abbreviation, GenBank accession number and TaqMan assay (Applied Biosystem) catalogue number are shown.

Angiopoietin-1, Ang-1; angiopoietin-2, Ang-2; C-C chemokine receptor type 2, CCR2; C-C chemokine receptor type 3, CCR3; C-C chemokine receptor type 5, CCR5; fibroblast growth factor-2, FGF-2; fibroblast growth factor-4, FGF-4; fibroblast growth factor receptor-1, FGFR-1; fibroblast growth factor receptor-3, FGFR-3; glyceraldehyde-3-phosphate dehydrogenase, GAPDH; interleukin-8, IL-8; monocyte chemotactic protein-1, MCP-1; neuropilin-1, NP-1; osteopontin (alias SSP-1, always referred as 'osteopontin' in the text); placental growth factor, PlGF; Plexin A1, PlA1; Plexin B1, PlB1; regulated upon activation, normally T-expressed, and presumably secreted, RANTES; risomboal 18S RNA, r18S; semaphorin 3A, Sema 3A; stromal cell-derived factor 1, SDF-1; tunica internal endothelial cell kinase, alias TEK tyrosine kinase, Tie-2; thrombospondin-1, Thrombospondin; vascular endothelial growth factor A, VEGF-A; vascular endothelial growth factor C, VEGF-C; vascular endothelial growth factor D, VEGF-D; vascular endothelial growth factor receptor-1, VEGFR-1; vascular endothelial growth factor receptor-2, VEGFR-2 and vascular endothelial growth factor receptor-3, VEGFR-3.

decrease) and RANTES (7-fold decrease). The whole set of results is shown in Suppl. Fig. 2, along with the distribution of the ratios obtained between all paired values.

3.4. *Multigenic analysis for tumour classification*

The molecular variables (levels of gene expression of all analysed genes) and the clinical and histological variables (age, gender, localisation of the primary tumour, tumour-node-metastasis (TNM) stage, histotype, grade, front of infiltration, vascular and lymphatic invasion, Dukes stage) were submitted to multigenic analysis using complementary statistical approaches, in order to define possible molecular predictors of the biological behaviour of the tumour. Only patients, whose complete clinical and molecular dataset were available, were considered for statistical evaluation.

3.4.1. *Discrimination between tumour tissue and normal mucosa*

A comparison was initially made between the levels of gene expression in tumour tissue (T) and in normal mucosa (M) in 78 eligible patients. When analysed by logistic regression, the expression levels of SDF-1, plexin-A1 and IL-8 turned out to be the most relevant predictors of the two categories of samples, being the expression of plexin-A1 and IL-8 higher in tumour tissue whilst that of SDF-1 higher in normal mucosa ($P < 0.0001$). The best linear discriminant analysis (LDA) model again contained SDF-1 and plexin-A1 together with Ang-2. The ensuing discriminant function (reported in the legend of Fig. 2A) indicated that SDF-1 was the most relevant gene, showing a clear, protective effect. The error rate of the model (namely, the possibility to use the discriminant function to classify a new tissue), was 0.078, a result significantly

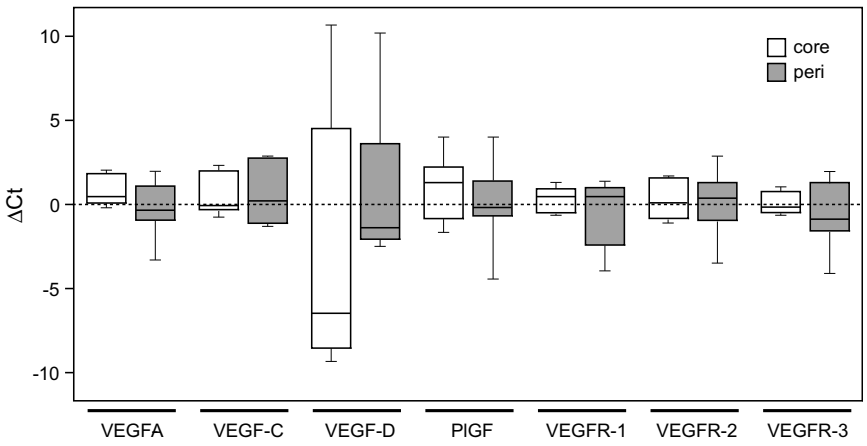

Fig. 1 – Homogeneity of gene expression within the tumour mass. The expression levels of the indicated genes were assessed in RNA extracted from samples taken in the tumour centre (core) or periphery (peri) (n = 10). Each box plot shows the percentile distribution of values; horizontal lines, from top to bottom, mark the 10th, 25th, 50 th, 75th and 90th percentile.

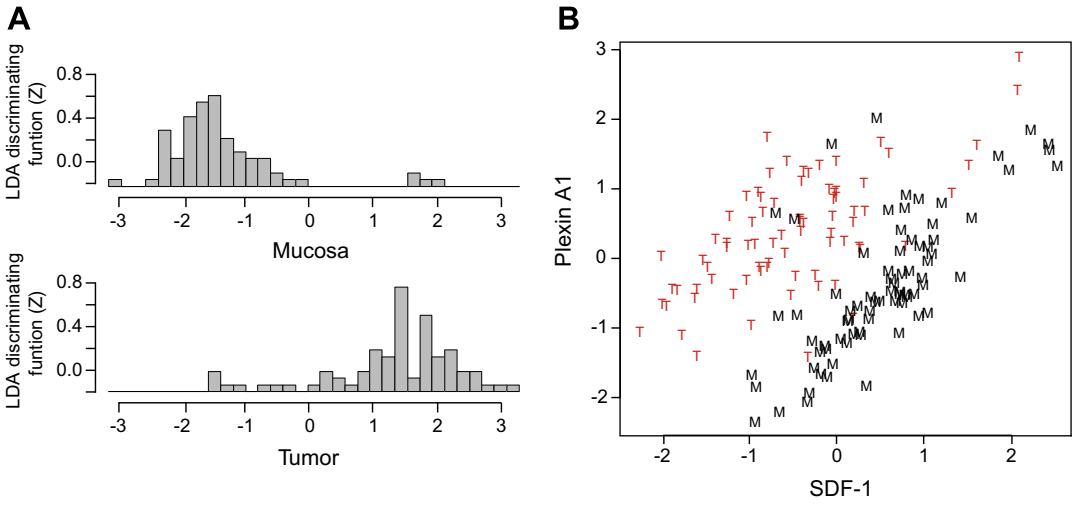

Fig. 2 – Discrimination between tumour tissue (T) and normal mucosa (M). (A) Frequency diagrams of the mucosa and tumour samples for the LDA discriminating function Z (Z = –1.5140 stromal cell-derived factor 1 (SDF-1) + 0.7468 angiopoietin-2 (Ang-2) + 0.7366 plexin-A1; error rate = 0.078 using leave-one-out cross-validation). A new tissue can be classified according to its Z value. The threshold between tumour and normal mucosa is Z = ln(78/78) = 0, and Z > 0 for a tumour tissue. (B) Scatterplot of the levels of expression of the SDF-1 and plexin-A1 genes in the sample, along with the indication of whether the sample was mucosa (M, black) or tumour (T, red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

better than the error rate of the null model, which was 0.50 (95% confidence interval 0.42–0.58). Fig. 2A reports the frequency diagrams of the mucosa and tumour samples for the LDA discriminating function, which show a clear separation of the two types of tissue. A scatterplot of the levels of expression of SDF-1 and plexin-A1, with the indication of whether the sample was mucosa or tumour, is shown in Fig. 2B.

3.4.2. *Molecular discrimination of the clinical variables N, M, EPLI, EPVI, Dukes, G*

In order to find new discriminating molecular markers possibly predicting the different clinical features of tumours, the entire tumour sample dataset (excluding the normal mucosae) was analysed according to both LDA and logistic regression. In a first instance, lymph node invasion (N) was studied assuming N = 0 in absence of node infiltration (n = 42), and N = 1 in the case of infiltrated lymph nodes (n = 31), regardless of their number. The logistic regression model identified osteopontin and RANTES (P = 0.005) as predictors; osteopontin was higher in N = 1, whilst RANTES was higher in N = 0. Osteopontin, along with Tie-2 and Ang-2, also appeared as a predictor in the LDA discriminating function (Table 3).

A further level of complexity was introduced by also considering normal mucosae. In this way, the sample size was doubled, and the resulting models contained two discriminating functions, the first one separating tumours from matched

Table 3 – Summary of the results for the best discriminating models of linear discriminant analysis (LDA) and logistic regression (Log Regr)

Lymph node invasion **N**	Sample size *n* = 73 (42 N = 0, 31 N = 1)
LDA: Z = –1.1453 Tie-2 + 0.8854 Ang-2 + 0.8840 **osteopontin**	
Z′ = ln(42/31) ≈ 0.3037; N = 1 if Z > Z′ ER = 0.32, ER_0 = 0.42 (0.31–0.55)	
Log Regr *P* = 0.005: **osteopontin** ↑N = 1, RANTES ↓N = 1	
Metastasis **M**	Sample size *n* = 70 (55 M = 0, 15 M = 1)
LDA: Z = 0.9923 **VEGFR-3**	
Z′ = ln(55/15) ≈ 1.2993; M = 1 if Z > Z′ ER = 0.22, ER_0 = 0.21 (0.13–0.33)	
Log Regr *P* = 0.041: **VEGFR-3** and VEGF-A ↑M = 1; RANTES ↓M = 1	
Extra-parietal lymphatic infiltration **EPLI**	Sample size n = 58 (44 EPLI = 0, 14 EPLI = 1)
LDA: Z = –1.2072 **MCP1** + 1.0179 **Ang-2** + 0.9472 FGFR-1	
Z′ = ln(44/14) ≈ 1.1451; EPLI = 1 if Z > Z′ ER = 0.17, ER_0 = 0.24 (0.14–0.37)	
Log Regr *P* = 0.0004: **Ang-2** and osteopontin ↑EPLI = 1; **MCP1** and IL-8 ↓EPLI = 1	
Extra-parietal vascular infiltration **EPVI**	Sample size *n* = 60 (46 EPVI = 0, 14 EPVI = 1)
LDA: Z = 1.2337 FGFR-1 – 0.9552 **RANTES** – 0.6942 FGFR-3	
Z′ = ln(46/14) ≈ 1.1896; EPVI = 1 if Z > Z′ ER = 0.22, ER_0 = 0.23 (0.13–.36)	
Log Regr *P* = 0.0003: VEGF-C and FGF-2 ↑EPVI = 1; **RANTES** and **FGFR-3** ↓EPVI = 1	
Dukes stage **D**	Sample size *n* = 64 (32 D = 0 (stages a and b), 32 D = 1 (stages c and d))
LDA: Z = –1.1318 **RANTES** + 0.9359 **VEGFR-3**	
Z′ = ln(32/32) = 0; D = 1 if Z > Z′ ER = 0.35, ER_0 = 0.50 (0.37–0.63)	
Log Regr *P* = 0.003: **VEGFR-3**, CCR3 and osteopontin ↑D = 1; **RANTES**, IL-8, FGFR-3 ↓D = 1	
Grading **G**	Sample size *n* = 69 (15 G = 0, 54G = 1)
LDA: Z = 1.0742 **VEGFR-1**	
Z′ = ln(15/54) ≈ –1.2809; G = 1 if Z > Z′ ER = 0.19, ER_0 = 0.22 (0.13–0.33)	
Log Regr *P* = 0.001: **VEGFR-1** ↑G = 1	

The following clinical variables were considered: (N), between N = 0 (absence of node infiltration) and N = 1 (node infiltration); (M), between M = 0 (no metastasis) and M = 1 (metastasis); extra-parietal lymphatic infiltration (EPLI), between EPLI = 0 (no infiltration) and EPLI = 1 (infiltration); extra-parietal vascular infiltration (EPVI), between EPVI = 0 (no infiltration) and EPVI = 1 (infiltration); Dukes stage (D), between D = 0 (including Dukes stages a and b together) and D = 1 (stages c and d); grade (G), between G = 0 (G grade 0) and G = 1 (G grade 2 and 3 together). No statistical analysis was performed to discriminate according to tumour size (T) since the sample was severely unbalanced toward the high T scores (5 T = 0, 67 T = 1, 2 and 3 together).

Legend: Z = LDA discriminating function; Z′ = threshold value for Z; ER = error rate of the LDA model; ER_0 = error rate of the null model, followed in parentheses by the 95% confidence interval; *P* = *P* value of the logistic regression model; (A ↑X = 1) = the gene expression of A is higher in X = 1; (A ↓X = 1) = the gene expression of A is lower in X = 1. Common genes between LDA and Log Regr are in boldface.

normal mucosae whilst the second one addressing a specific tumour characteristic. For the N variable again, LD1 discriminated between tumours (regardless of N state) and mucosae (arbitrary labelled as N2), whereas LD2 distinguished between N = 0 and N = 1. As shown in the scatterplot of Fig. 3A, the overall variance was mainly explained by the discriminant function LD1 (96%, with only 4% due to LD2). This means that an excellent separation between the normal mucosa N = 2 and the tumour tissue N = 0, 1 can be obtained according to LD1, which consists of a linear combination of SDF-1, plexin-A1 and Ang-2. In contrast, LD2 discriminated less well between the levels N = 0 and N = 1 of the tumour. It is, however, clear that the centroid of the N = 1 dots has a lower LD2 value than the centroid of the N = 0 points. The best predictors for LD2 were osteopontin, Tie-2 and IL-8.

A similar approach was then applied to the remaining clinical variables, including metastasis (M), extra-parietal lymphatic infiltration (EPLI), extra-parietal vascular infiltration (EPVI), Dukes stage (D) and grade (G). The results of these analyses are presented in Table 3 and Fig. 3B–F.

Collectively, both LDA and logistic regression highlighted an important discriminating role for a few genes, indicated in bold in Table 3. This is the case of osteopontin for lymph node invasion and VEGFR-3 and RANTES for Dukes stage. These common genes, identified in the above models, were further confirmed by a third, independent technique, support vector machines (SVM), which can also be applied to strongly correlated variables and cases of non-normal multivariate distributions.[30]

4. Discussion

Our analysis entailed the assessment of expression levels of 27 genes in paired mucosa and colorectal carcinoma samples from 78 patients. From the methodological point of view, the introduction of an exogenous RNA normaliser allowed us to generate more robust results compared to the use of both GAPDH and risomboal 18S RNA (r18S) housekeeping genes. Another critical methodological issue addressed in our study was the analysis of homogeneity of gene expression within the same tumour mass. By comparing samples from the centre and the periphery of the tumour, we indeed found similar gene expression levels, supporting the reliability of a single tumour biopsy as representative of the whole tumour.

A striking result came from the comparison of paired samples of tumour and normal mucosa. Several genes were consistently upregulated in the tumours, including IL-8, PlGF, osteopontin and Ang-2; in contrast, VEGF-D, SDF-1 and RANTES were clearly downregulated. When the whole dataset was analysed by multigenic analyses, we could discriminate the normal mucosa from tumours based on the respective expression profile. Both logistic regression and LDA indicated

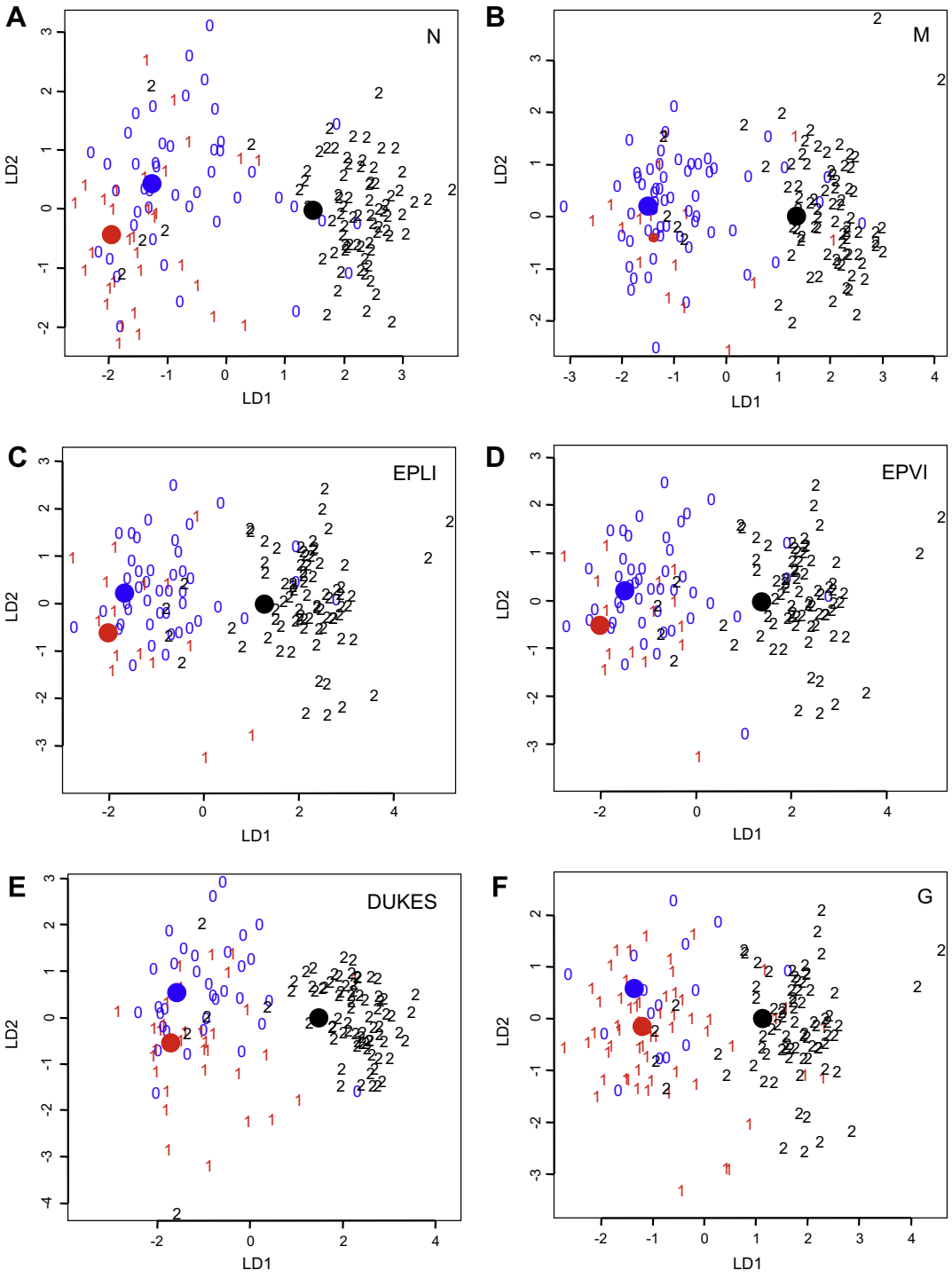

Fig. 3 – Results of multigenic analysis. For each variable (N, M, EPLI, EPVI, DUKES, G) the scatterplot in the proper discriminant plane (LD1, LD2) is displayed. The centroids are represented as filled circles. (A) Discrimination for lymph node infiltration, with N = 0 (no lymph node infiltration), N = 1 (lymph node infiltration) and N = 2 (normal mucosa). (B) Discrimination for the presence of metastasis, with M = 0 (no metastasis), M = 1 (metastasis), M = 2 (normal mucosa). (C) Discrimination for lymphatic vessel infiltration, with EPLI = 0 (no infiltration), EPLI = 1 (infiltration), EPLI = 2 (normal mucosa). (D) Discrimination for vascular infiltration, with EPVI = 0 (no infiltration), EPVI = 1 (infiltration), EPVI = 2 (normal mucosa). (E) Discrimination for Dukes' stage, with D = 0 (Dukes stages a + b), D = 1 (Dukes stages c + d) and D = 2 (normal mucosa). (F) Discrimination for tumour Grade, with G = 0 (G grade 1), G = 1 (G grades 2 and 3), G = 2 (normal mucosa).

plexin-A1 and SDF-1 as the most important predictors, the former being overexpressed and the latter downregulated in tumours. Discriminant analysis, based on these two genes together with Ang-2, generated an LD1 function that attributed positive values to tumours and negative values to matched mucosa (Fig. 2A). By using this function, the specimens

grouped into two clusters, concordant with pathological reports (cancer versus normal tissue). The reliability of the model could be further increased by the inclusion of a higher number of variables, although this would add complexity to the model (data not shown). Thus, gene expression profiling of a few selected genes can discriminate between normal and tumour colon tissue.

Type-A plexins, although originally identified as semaphorin receptors in axon guidance, have been recently shown to take part in both tumour angiogenesis,[31] and, directly, in the control of tumour cell growth.[32] The chemokine stromal cell-derived-factor-1 (SDF-1) has recently emerged for its ability to retain myeloid cells at the sites of neo-angiogenesis, in order to promote the formation of structured arterial vessels.[18] In cancer, this chemokine has been shown to stimulate metastatic spread.[33] Conversely, other studies have shown that its expression is unchanged or even lowered in advanced cancers.[34,35] In our samples, SDF was constantly downregulated in tumours, its levels being a major predictor to distinguish tumours from normal mucosa (Fig. 2). To reconcile these apparently contradictory findings, we can speculate that, during the early phases, SDF-1 production might sustain angiogenesis and tumour growth, whereas, in later stages, a lower expression might be selected along tumour development to avoid recruitment of leukocytes with anti-tumoural activity. Further experiments will tackle this issue more specifically.

When the whole dataset was analysed by multigenic analysis, the expression of osteopontin and Tie-2 appeared to discriminate about lymph node infiltration, the former being overexpressed in the presence of infiltrated nodes, whilst the latter providing a protective role. Osteopontin is an integrin-binding protein, previously proposed as a marker of tumour progression in various cancers,[36–38] including colon carcinoma.[39] Our finding of a 30-fold overexpression of osteopontin and its ability to discriminate between patients with and without node infiltration is perfectly consistent with these previous results.[39,40] On the other hand, Tie-2 is a receptor tyrosine kinase, expressed on vascular endothelium, where it binds Ang-1 and Ang-2.[41] Although the effect of angiopoietins on tumour angiogenesis is still debated, Ang-1 is generally considered a pro-angiogenic factor, especially in the presence of VEGF, whereas Ang-2 seems to destabilise the neo-vasculature, thus being anti-angiogenic.[42] Our finding that low levels of Tie-2 were particularly helpful to identify colon cancers with nodal invasion might imply an inverse correlation between local invasion and efficient neovessel formation. Consistently, we found that the presence of extra-parietal lymphatic infiltration was well discriminated by Ang-2, which was more expressed by invasive tumours. Additionally, when observing the average values of expression in the whole dataset, Ang-2 appeared significantly upregulated in tumours, whereas Ang-1 was downregulated.

The follow-up of this cohort of patients will eventually indicate whether some of the analysed variables might be of value for prognostic purposes. In this respect, it is worth mentioning that the levels of expression of VEGFR-3 permitted discrimination between low (a + b) and high (c + d) Dukes stage, which essentially considers local invasion. The levels of expression of VEGFR-3 have already been correlated with reduced survival in gastric, lung, cervical and prostate carcinomas.[43]

In conclusion, our findings support the notion that the expression levels of selected pro-angiogenic genes might provide trustable biomarkers of tumour development and invasion.[44] Notably, genes identified by both logistic regression and LDA were further confirmed by a third independent statistical technique (SVM). Our statistical analysis revealed that the identified genes associated with tumour behaviour could be validated on the whole dataset. In other terms, once a given gene is identified, it can also be used as a predictor of the variable of interest, thus representing a possible valuable biomarker for diagnostic purposes.

Conflict of interest statement

None declared.

Acknowledgments

This work was supported by a grant from 'Fondazione CR Trieste', Trieste, Italy. The authors are grateful to Suzanne Kerbavcic for precious editorial assistance.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2008.05.014.

REFERENCES

1. Greenlee RT, Hill-Harmon MB, Murray T, Thun M. Cancer statistics. *CA Cancer J Clin* 2001;**51**(1):15–36.
2. Cascinu S, Georgoulias V, Kerr D, Maughan T, Labianca R, Ychou M. Colorectal cancer in the adjuvant setting: perspectives on treatment and the role of prognostic factors. *Ann Oncol* 2003;**14**(Suppl. 2):ii25–9.
3. Carmeliet P. Angiogenesis in life, disease and medicine. *Nature* 2005;**438**(7070):932–6.
4. Blouw B, Song H, Tihan T, et al. The hypoxic response of tumors is dependent on their microenvironment. *Cancer Cell* 2003;**4**(2):133–46.
5. Ferrara N, Kerbel RS. Angiogenesis as a therapeutic target. *Nature* 2005;**438**(7070):967–74.
6. Vogel C, Bauer A, Wiesnet M, et al. Flt-1, but not Flk-1 mediates hyperpermeability through activation of the PI3-K/Akt pathway. *J Cell Physiol* 2007;**212**(1):236–43.
7. Alitalo K, Carmeliet P. Molecular mechanisms of lymphangiogenesis in health and disease. *Cancer Cell* 2002;**1**(3):219–27.
8. Bussolino F, Serini G, Mitola S, Bazzoni G, Dejana E. Dynamic modules and heterogeneity of function: a lesson from tyrosine kinase receptors in endothelial cells. *EMBO Rep* 2001;**2**(9):763–7.
9. Gale NW, Yancopoulos GD. Growth factors acting via endothelial cell-specific receptor tyrosine kinases: VEGFs, angiopoietins, and ephrins in vascular development. *Genes Dev* 1999;**13**(9):1055–66.
10. Gale NW, Baluk P, Pan L, et al. Ephrin-B2 selectively marks arterial vessels and neovascularization sites in the adult, with

expression in both endothelial and smooth-muscle cells. *Dev Biol* 2001;**230**(2):151–60.
11. Battler A, Scheinowitz M, Bor A, et al. Intracoronary injection of basic fibroblast growth factor enhances angiogenesis in infarcted swine myocardium. *J Am Coll Cardiol* 1993;**22**(7):2001–6.
12. Arsic N, Zentilin L, Zacchigna S, et al. Induction of functional neovascularization by combined VEGF and angiopoietin-1 gene transfer using AAV vectors. *Mol Ther* 2003;**7**(4):450–9.
13. Asahara T, Chen D, Takahashi T, et al. Tie2 receptor ligands, angiopoietin-1 and angiopoietin-2, modulate VEGF-induced postnatal neovascularization. *Circ Res* 1998;**83**(3):233–40.
14. Thurston G, Rudge JS, Ioffe E, et al. Angiopoietin-1 protects the adult vasculature against plasma leakage. *Nat Med* 2000;**6**(4):460–3.
15. Zacchigna S, Tasciotti E, Kusmic C, et al. In vivo imaging shows abnormal function of VEGF-induced vasculature. *Hum Gene Ther* 2007;**18**(6):515–24.
16. Charalambous C, Pen LB, Su YS, Milan J, Chen TC, Hofman FM. Interleukin-8 differentially regulates migration of tumor-associated and normal human brain endothelial cells. *Cancer Res* 2005;**65**(22):10347–54.
17. Gerszten RE, Garcia-Zepeda EA, Lim YC, et al. MCP-1 and IL-8 trigger firm adhesion of monocytes to vascular endothelium under flow conditions. *Nature* 1999;**398**(6729):718–23.
18. Grunewald M, Avraham I, Dor Y, et al. VEGF-induced adult neovascularization: recruitment, retention, and role of accessory cells. *Cell* 2006;**124**(1):175–89.
19. Loewer A, Lahav G. Cellular conference call: external feedback affects cell-fate decisions. *Cell* 2006;**124**(6):1128–30.
20. Kalluri R, Zeisberg M. Fibroblasts in cancer. *Nat Rev Cancer* 2006;**6**(5):392–401.
21. Rangaswami H, Bulbule A, Kundu GC. Osteopontin: role in cell signaling and cancer progression. *Trend Cell Biol* 2006;**16**(2):79–87.
22. Allen WL, Johnston PG. Have we made progress in pharmacogenomics? The implementation of molecular markers in colon cancer. *Pharmacogenomics* 2005;**6**(6):603–14.
23. Komori T, Takemasa I, Higuchi H, et al. Identification of differentially expressed genes involved in colorectal carcinogenesis using a cDNA microarray. *J Exp Clin Cancer Res* 2004;**23**(3):521–7.
24. Williams NS, Gaynor RB, Scoggin S, et al. Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. *Clin Cancer Res* 2003;**9**(3):931–46.
25. Millenaar FF, Okyere J, May ST, van Zanten M, Voesenek LA, Peeters AJ. How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. *BMC Bioinformatics* 2006;**7**:137.
26. Pylatuik JD, Fobert PR. Comparison of transcript profiling on Arabidopsis microarray platform technologies. *Plant Mol Biol* 2005;**58**(5):609–24.
27. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal Biochem* 1987;**162**(1):156–9.
28. Armitage P, Berry G, Matthews JNS. *Statistical methods in medical research*. Oxford, UK: Blackwell; 2002.
29. Venables WN, Ripley DB. *Modern applied statistics with S*. New York, NY, USA: Springer; 2002.
30. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002;**46**(1–3):389–422.
31. Guttmann-Raviv N, Kessler O, Shraga-Heled N, Lange T, Herzog Y, Neufeld G. The neuropilins and their role in tumorigenesis and tumor progression. *Cancer Lett* 2006;**231**(1):1–11.
32. Nguyen QD, Rodrigues S, Rodrigue CM, et al. Inhibition of vascular endothelial growth factor (VEGF)-165 and semaphorin 3A-mediated cellular invasion and tumor growth by the VEGF signaling inhibitor ZD4190 in human colon cancer cells and xenografts. *Mol Cancer Ther* 2006;**5**(8):2070–7.
33. Kryczek I, Wei S, Keller E, Liu R, Zou W. Stromal derived factor (SDF-1/CXCL12) and human tumor pathogenesis. *Am J Physiol Cell Physiol* 2006;**292**(3):C987–95.
34. Romero JM, Aptsiauri N, Vazquez F, et al. Analysis of the expression of HLA class I, proinflammatory cytokines and chemokines in primary tumors from patients with localized and metastatic renal cell carcinoma. *Tissue Antigens* 2006;**68**(4):303–10.
35. Rubie C, Frick VO, Wagner M, et al. Chemokine expression in hepatocellular carcinoma versus colorectal liver metastases. *World J Gastroenterol* 2006;**12**(41):6627–33.
36. Tuck AB, O'Malley FP, Singhal H, et al. Osteopontin expression in a group of lymph node negative breast cancer patients. *Int J Cancer* 1998;**79**(5):502–8.
37. Chambers AF, Wilson SM, Kerkvliet N, O'Malley FP, Harris JF, Casson AG. Osteopontin expression in lung cancer. *Lung Cancer* 1996;**15**(3):311–23.
38. Thalmann GN, Sikes RA, Devoll RE, et al. Osteopontin: possible role in prostate cancer progression. *Clin Cancer Res* 1999;**5**(8):2271–7.
39. Agrawal D, Chen T, Irby R, et al. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *J Natl Cancer Inst* 2002;**94**(7):513–21.
40. Yeatman TJ, Chambers AF. Osteopontin and colon cancer progression. *Clin Exp Metastasis* 2003;**20**(1):85–90.
41. Kobayashi H, Lin PC. Angiopoietin/Tie2 signaling, tumor angiogenesis and inflammatory diseases. *Front Biosci* 2005;**10**:666–74.
42. Armulik A, Abramsson A, Betsholtz C. Endothelial/pericyte interactions. *Circ Res* 2005;**97**(6):512–23.
43. Su JL, Yen CJ, Chen PS, et al. The role of the VEGF-C/VEGFR-3 axis in cancer progression. *Br J Cancer* 2006;**96**(4):541–5.
44. Dalton WS, Friend SH. Cancer biomarkers – an invitation to the table. *Science* 2006;**312**(5777):1165–8.