A five-gene signature as a potential predictor of metastasis and survival in colorectal cancer

J Hao

J Chen

H Sui

X Si-Ma

et al

Let us know how access to this document benefits you

Recommended Citation

Hao, J., Chen, J., Sui, H., Si-Ma, X., & et al. (2010) 'A five-gene signature as a potential predictor of metastasis and survival in colorectal cancer', *The Journal of Pathology*, 220(4), pp. 475-489. Wiley: Available at: https://doi.org/10.1002/path.2668

2010-03

A fivegene signature as a potential predictor of metastasis and survival in colorectal cancer

Hao, J

http://hdl.handle.net/10026.1/10314

Original Paper

A five-gene signature as a potential predictor of metastasis and survival in colorectal cancer

Jun-Mei Hao,[1−3][†] Juan-Zhi Chen,[2,3][†] Hong-Mei Sui,[1] Xue-Qing Si-Ma,[1] Guang-Qiu Li,[1] Chao Liu,[1] Ji-Liang Li,[4,5] Yan-Qing Ding[1−3]* and Jian-Ming Li[1−3]*

[1]*Department of Pathology, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, People's Republic of China*
[2]*Department of Pathology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, People's Republic of China*
[3]*Guangdong Provincial Key Laboratory of Molecular Tumour Pathology, Guangzhou 510515, People's Republic of China*
[4]*School of Biotechnology, Southern Medical University, Guangzhou 510515, People's Republic of China*
[5]*Cancer Research UK Molecular Oncology Laboratories, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK*

**Correspondence to:
Jian-Ming Li and Yan-Qing Ding,
Department of Pathology and
Guangdong Provincial Key
Laboratory of Molecular Tumour
Pathology, Southern Medical
University, Guangzhou 510515,
People's Republic of China.
E-mail: lixinyue@fimmu.com or
dyq@fimmu.com*

*†These authors contributed
equally to this work.*

*No conflicts of interest were
declared.*

Abstract

To understand the molecular mechanisms of metastasis and prognosis of colorectal cancer (CRC), we isolated single cell-derived progenies (SCPs) from SW480 cells *in vitro* and compared their metastatic potential in an orthotopic CRC tumour model *in vivo*. Two groups of SCPs with the capability of high and low metastasis, respectively, were obtained. By analysing the gene expression profiles of high (SCP51), low (SCP58) metastatic SCPs, and their parental cell line (SW480/EGFP), we demonstrated that 143 genes were differentially expressed either between SCP51 and SCP58 or between SCP58 and SW480/EGFP. Gene-annotation enrichment analysis of DAVID revealed 80 genes in the top ten clusters of the analysis (gene enrichment score >1). Of the 80-gene set, 32 genes are potentially involved in metastasis, as revealed by Geneclip. Five putative metastatic genes (*LYN, SDCBP, MAP4K4, DKK1*, and *MID1*) were selected for further validations. Immunohistochemical analysis in a cohort of 181 CRC clinical samples showed that the individual expression of LYN, MAP4K4, and MID1, as well as the five-gene signature, was closely correlated with lymph node metastasis in CRC patients. More importantly, the individual expression of LYN, MAP4K4, SDCBP, and MID1, as well as the five-gene signature, was significantly correlated with overall survival in CRC patients. Thus, our five-gene signature may be able to predict metastasis and survival of CRC in the clinic, and opens new perspectives on the biology of CRC.
Copyright © 2009 Pathological Society of Great Britain and Ireland. Published by John Wiley & Sons, Ltd.

Keywords: single cell-derived progenies; gene signature; prognosis; metastasis; colorectal cancer

Introduction

Metastasis is the main cause leading to death of colorectal cancer (CRC) patients. However, the molecular mechanisms of metastasis of CRC remain largely unknown. The discovery of novel biomarkers has relevance for the early diagnosis, prevention, and possible target therapy of CRC patients, as well as providing new biological insights. Microarray profiling has been shown to be a valuable technique to uncover biomarkers or prognosis factors for patients with different types of cancer [1–5]. A gene expression signature, defined as the expression pattern of a particular set of genes, has been widely used in identification of the specific cancer type or subtype, prediction of the outcome of patients, and selection of the appropriate therapeutic treatment [6–8].

Two common strategies based on the gene expression profiling technique have been used to identify novel biomarkers associated with many characteristics of cancer. Profiling of human tissue can distinguish the gene expression between normal and tumour tissues from patients. This approach usually results in large numbers of such genes and only a small proportion of these genes may be clinically useful. The second strategy is based on the notion that only a few cells from a heterogeneous population have metastatic capability. Therefore, metastatic variants can be isolated from the primary bulk of the tumour. *In vivo* selection has been used to discern genetic determinants of metastasis. Using this approach, Massague and co-workers have reported gene signatures mediating breast cancer metastasis to bone [9], lung [10], and brain [11]. The molecular signatures derived from mouse tumour

models could successfully predict the progression and prognosis of breast cancer.

In this study, we isolated single cell-derived progenies (SCPs) with different metastatic potentials from the SW480 cell line *in vitro* and compared their gene expression profiles *in vivo*. We identified a 32-gene set as the candidate genes associated with metastasis of CRC by cross-comparing, functional analysis and gene annotation, and text mining. In a cohort of 181 CRC clinical samples, we validated the expression of a five-gene signature and each of the five genes in CRC. We found that the five-gene signature may be able to predict metastasis and patient survival of CRC.

Materials and methods

Cell culture

The human CRC cell line SW480 was obtained from the Cell Bank at the Chinese Academy of Sciences (Shanghai, China). SW480/EGFP was established from SW480 by transfection of pEGFP-N1 plasmid. Tumour cells were cultured in RPMI 1640 medium (Gibco, Grand Island, NY, USA) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT, USA) and 100 U/ml penicillin/streptomycin (Gibco), and incubated in a humidified chamber with 5% CO_2 at 37 °C. SW480/EGFP cells were diluted in medium at 2–10 cells/ml. 100 μl of the dilution was added to each well of a 96-well plate. Wells containing a single clone derived from one cell were identified under a microscope (Figure 1A). Clones were then expanded in 10-cm tissue culture plates.

Cell proliferation assay

Tumour cells were prepared at a concentration of 1×10^3 cells/ml. Aliquots (200 μl) of the cells were dispensed in triplicate into 96-well plates. After indicated days, 20 μl of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (5 mg/ml; Promega) was added to each well. Four hours later, the supernatant was removed and 150 μl of dimethyl sulphoxide (Sigma, St Louis, MO, USA) was added. After 15-min incubation, the absorbance value (OD) was measured at 570 nm in a micro-plate reader. Four independent experiments were performed.

Tumour xenografts

Xenograft tumours were generated by subcutaneous injection of 5×10^6 cells on the flanks of nude mice. Orthotopic implantation of the tumour fragment obtained from the subcutaneous tumour to the caecum was performed (Supporting information, Supplemental experimental procedures) as described previously [12]. To observe tumour growth, fluorescence emitted by tumour cells was imaged by a whole-body green fluorescence protein (GFP) imaging system (Lighttools, Encinitas, CA, USA). Images were then analysed to

calculate the tumour area and tumour growth using IPP5.0 software (Cybermetics Co, Silver Spring, MD, USA). To dissect tumour metastasis, various organs were collected from the tumour-bearing mice. The samples was fixed in 10% neutral buffered formalin, embedded in paraffin, and stained with haematoxylin and eosin.

Microarray analysis

Nine samples, three from parental SW480/EGFP, three from SCP51 with high metastatic potential, and three from SCP58 with low metastatic potential, were collected. Total RNA was extracted from the samples using a Qiagen RNeasy kit (Qiagen, Chatsworth, CA, USA). The Affymetrix Human Genome U133 Plus 2.0 Array GeneChips (Affymetrix, Santa Clara, CA, USA) were used for microarray analysis. Hybridization and scanning of the genechips were performed by the CapitalBio Biotechnology Company (Beijing, China) according to the standard Affymetrix protocol (Supporting information, Supplementary experimental procedures). The data have been deposited in the public Gene Expression Omnibus repository (Accession number GSE16934). Significance analysis of microarrays (SAM) was performed to identify the genes expressed differentially between SCP51 and SCP58, or between SCP58 and SW480/EGFP. Common differential genes were obtained by cross-comparing the two data sets.

Functional analysis and gene annotation

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) was used to discover enriched function-related gene groups by gene-annotation enrichment analysis (http://david.abcc.ncifcrf.gov/). The genes in the top ten clusters (gene enrichment score >1) from the gene-annotation enrichment analysis were selected. Geneclip [13], a software program for clustering gene lists by literature profiling and constructing gene co-occurrence networks, was used to acquire a primary profile of known genes related to metastasis in colorectal cancer, melanoma, and other cancers.

Reverse transcription-polymerase chain reaction (RT-PCR)

RT-PCR was performed to measure the expression of LYN, SDCBP, MAP4K4, MID1, and DKK1 at the mRNA level in fresh colorectal tumour samples. Total RNA was extracted from the tumour using Trizol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. RNA was treated with DNase and used for cDNA synthesis with random hexamers. The primers used for amplifying LYN, SDCBP, MAP4K4, MID1, and DKK1 have been described in the Supporting information, Supplemental experimental procedures.

Western blot analysis

Mouse monoclonal antibodies to DKK1, MAP4K4, and MID1, and rabbit polyclonal antibodies to LYN and SDCBP were purchased from Abcam (Cambridge, UK). Procedural details of western blotting have been described in the Supporting information, Supplemental experimental procedures.

Immunohistochemical staining and scoring

Formalin-fixed, paraffin-embedded colorectal tumour samples obtained from patients at the Nanfang Hospital of Southern Medical University (Guangzhou, China) were used in our clinicopathological investigation. Individual consent on use of the clinical sample for research purposes was obtained from each patient and approved by the Ethical Committee of Southern Medical University. Histological classification and clinicopathological staging of the tumours were performed according to the Japanese General Rules for Clinical and Pathological Studies on Cancer of the Colon, Rectum, and Anus, along with the International Union Against Cancer Classification. Immunohistochemical staining was performed using a Dako EnVision System (Dako, Carpentaria, CA, USA) following the protocol recommended by the manufacturer to study the expression of LYN, SDCBP, MAP4K4, MID1, and DKK1 in 181 clinical samples. Mouse anti-MAP4K4 monoclonal antibody (1 : 100; Sigma-Aldrich, St Louis, USA), rabbit anti-human MID1 polyclonal antibody (1 : 50; Abnova, Taiwan), rabbit anti-human SDCBP (1 : 25), LYN (1 : 50), and DKK1 (1 : 10) polyclonal antibodies (Abcam, Cambridge, UK) were used in the study. The stained tumours were reviewed and scored separately by two pathologists blinded to the patients' clinical characteristics. Scoring of the tumour staining was performed as described previously [14,15]. Briefly, the intensity of staining was scored as negative (0), weak (1), medium (2), or strong (3). The extent of staining, defined as the percentage of positive staining areas in relation to the whole tumour section, was scored as 0 (0%), 1 (1–25%), 2 (26–50%), 3 (51–75%), or 4 (76–100%). An overall expression score, ranging from 0 to 12, was calculated by multiplying the score of intensity and that of extent.

For statistical analysis, a final score number ≥ 4 was considered high expression of the protein of interest. The score for the five-gene signature was calculated from the value of each gene in the signature. Since up-regulation of LYN, SDCBP, and MAP4K4 contributed to the progression of CRC, high expression was scored as 1, while low expression was scored as 0 for these genes. However, since down-regulation of MID1 and DKK1 was also associated with CRC progression, for these genes low expression was scored as 1, while high expression was scored as 0. The sum of the scores for the five proteins was regarded as the score of the five-gene signature. For statistical analysis, a final score ≥ 3 was considered to be positive expression of the five-gene signature.

Statistical analysis

Statistical analyses were performed using the SPSS software program (version 12.0; SPSS Inc, Chicago, IL, USA) and the results were confirmed by statisticians in the Department of Health Statistics, Southern Medical University. The chi-square test was used to assess differences in metastasis and peritoneal dissemination of tumour cells. The Mann–Whitney U-test was used to analyse the relationship between the expression of LYN, SDCBP, MAP4K4, MID1, DKK1 or the five-gene signature and clinicopathological characteristics of the patients. Survival curves for the patients with different expression of LYN, SDCBP, MAP4K4, MID1, and DKK1 were plotted using the Kaplan–Meier method and compared using the log-rank test. The significance of various survival-related variables was assessed using the Cox proportional hazards model in a multivariate analysis. *In vitro* proliferation assay, *in vivo* tumour growth assay, cell motility assay, and metastatic potentials of SCPs were tested using ANOVA. Statistical significance was concluded where $p \leq 0.05$ and is denoted in the figures by an asterisk.

Results

Isolation and metastatic characterization of single cell-derived progenies

A total of 59 SCPs were isolated from SW480/EGFP. Subcutaneous implantation of SCPs on flanks of nude mice showed that 29 SCPs (SCP3, SCP5, SCP8, SCP9, SCP10, SCP11, SCP12, SCP15, SCP20, SCP21, SCP23, SCP27, SCP28, SCP30, SCP31, SCP32, SCP33, SCP34, SCP35, SCP36, SCP37, SCP38, SCP43, SCP44, SCP50, SCP51, SCP55, SCP57, and SCP58) exhibit the capacity of tumourigenicity to produce a visible tumour (Figures 1B and 1C). SCPs that formed a subcutaneous tumour in ≥ 3 out of seven mice were selected for further study. The ability of an individual SCP to grow as a tumour under the skin of nude mice was shown to be significantly different ($F = 33.446$, $p = 0.000$) (Supporting information, Supplementary Table 1).

To mimic human colorectal cancer, seven SCPs were orthotopically implanted into the caecum individually. Dissection of the metastasis in several commonly metastasized sites including the peritoneal cavity, lymph node, and liver demonstrated that SCPs could be categorized into two groups based on their metastatic potential. SW480/EGFP, SCP12, SCP21, and SCP51 exhibited high metastatic potential, while SCP11, SCP28, and SCP58 had low metastatic potential (Figure 2A). To confirm the metastatic characteristics, orthotopic xenografts were performed again for SCP21, SCP51, SCP28, and SCP58. As shown in
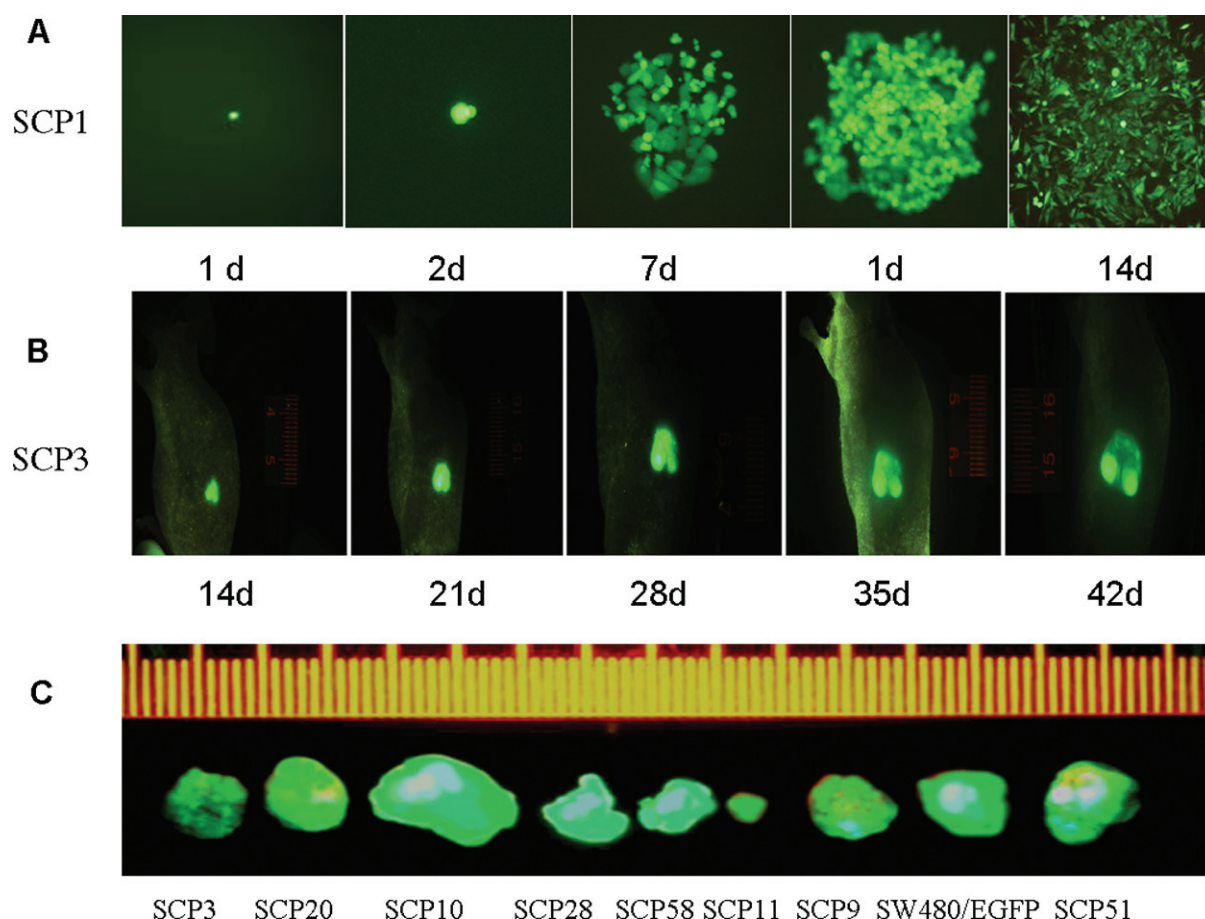
Figure 1. Single cell-derived progenies (SCPs) were isolated from SW480/EGFP. (A) *In vitro* growth and clonal formation of SCP1 observed under a fluorescent microscope. (B) Whole-body optical images of the SCP3 tumour that grew subcutaneously on the flanks of nude mice. (C) Tumour samples of SCP3, SCP20, SCP10, SCP28, SCP58, SCP11, SCP9, SCP51, and SW480/EGFP collected after subcutaneous growth on the flanks of nude mice for 5 weeks

Figure 2B, SCP21 and SCP51 were the high metastatic variants, whereas SCP28 and SCP58 were the low metastatic variants. The enhanced metastasis of SCP51 in lymph node and liver is demonstrated in Figures 2C and 2D.

In vitro proliferation, *in vivo* tumour growth, and migration of SCP51, SCP58, and SW480/EGFP

Next, we assessed the proliferation and migration capabilities of the SCPs. As shown in Figures 3A and 3B, both *in vitro* and *in vivo* SCP51 grew faster than SW480/EGFP, while SCP58 proliferated more slowly than SW480/EGFP. *In vitro* wound-healing migration assay showed that after 72 h, SCP51 cells filled in about 80% of the scratched area, whereas SCP58 cells filled in only about 20% of the scratched area (Figure 3C). Cell cycle analysis of SCP51, SCP58, and SW480/EGFP cells by flow cytometry demonstrated that the cell number of the G1 phase and the proliferative index (the percentage of S + G2 nuclei in the DNA histogram) of SCP51 were significantly increased compared with those of SW480/EGFP or SCP58 (Supporting information, Supplementary Table 2).

Identification of 32 candidate genes with metastatic potential

To investigate the molecular signature underlining the enhanced metastasis, we then carried out microarray analysis using Affymetrix human genechips. Several approaches were used to identify the candidate genes mediating CRC metastasis (Figure 4A). First, we screened the genes expressed differentially between SCP51 and SCP58 (SCP51 versus SCP58), or between SCP58 and SW480/EGFP (SW480/EGFP versus SCP58), using SAM with a false-discovery rate of 4%. A total of 439 genes (derived from 612 probes), with 229 up-regulated and 210 down-regulated, were found to be differentially expressed between SCP51 and SCP58 (Supporting information, Supplementary Table 3). Similarly, a total of 313 genes (from 359 probes), with 153 up-regulated and 160 down-regulated, were differentially expressed between SW480/EGFP and SCP58 (Supporting information, Supplementary Table 4). Next, we cross-compared the 439 genes expressed differentially between SCP51 and SCP58 with the 313 genes expressed differentially between SCP58 and SW480/EGFP. 143 genes were found to have metastatic potential (Figure 4B and Supporting information,

Table 1. Correlation of clinicopathological features of the patients with CRC and the expression of LYN, SDCBP, MAP4K4, MID1, DKK1, and the five-gene signature

	LYN expression			SDCBP expression			MAP4K4 expression			MID1 expression			DKK1 expression			Five-gene signature expression		
	High (%)	Low (%)	p	High (%)	Low (%)	p	High (%)	Low (%)	p	High (%)	Low (%)	p	High (%)	Low (%)	p	High (%)	Low (%)	p
Gender																		
Male	44 (40.4)	65 (39.6)	0.990	69 (63.3)	40 (36.7)	0.392	42 (38.5)	67 (61.5)	0.181	74 (67.9)	35 (32.1)	0.580	85 (78.0)	24 (22.0)	0.500	34 (31.2)	75 (68.8)	0.621
Female	29 (40.3)	43 (39.7)		41 (56.9)	31 (43.1)		35 (48.6)	37 (51.4)		46 (63.9)	26 (36.1)		53 (73.6)	19 (26.4)		25 (34.7)	47 (65.3)	
Age (years)																		
≤60	33 (35.9)	59 (64.1)	0.215	52 (56.5)	40 (43.5)	0.235	36 (39.1)	56 (60.9)	0.347	59 (64.1)	33 (35.9)	0.532	67 (72.8)	25 (27.2)	0.273	26 (28.3)	66 (71.7)	0.207
>60	40 (44.9)	49 (55.1)		58 (65.2)	31 (34.8)		41 (46.1)	48 (53.9)		61 (68.5)	28 (31.5)		71 (79.8)	18 (20.2)		33 (37.1)	56 (62.9)	
Site																		
Caecum or colon	38 (50.7)	37 (49.3)	0.017	48 (64.0)	27 (36.0)	0.456	37 (49.3)	38 (50.7)	0.121	53 (70.7)	22 (29.3)	0.297	60 (80.0)	15 (20.0)	0.319	27 (36.0)	48 (64.0)	0.413
Rectum	35 (33.0)	71 (67.0)		62 (58.5)	44 (41.5)		40 (37.7)	66 (62.3)		67 (63.2)	39 (36.8)		78 (73.6)	28 (26.4)		32 (30.2)	74 (69.8)	
Size (cm in diameter)																		
<5	20 (38.5)	32 (61.5)	0.745	34 (65.4)	18 (34.6)	0.421	19 (36.5)	33 (63.5)	0.301	40 (76.9)	12 (23.1)	0.056	39 (75.0)	13 (25.0)	0.804	14 (26.9)	38 (73.1)	0.303
≥5	53 (41.1)	76 (58.9)		76 (58.9)	53 (41.1)		58 (45.0)	71 (55.0)		80 (62.0)	49 (38.0)		99 (76.7)	30 (23.3)		45 (34.9)	84 (65.1)	
Tumour differentiation																		
Well	16 (30.2)	37 (69.8)	0.019	32 (60.4)	21 (39.6)	0.896	15 (28.3)	38 (71.7)	0.002	42 (79.2)	11 (20.8)	0.014	40 (75.5)	13 (24.5)	0.813	9 (17.0)	44 (83.0)	0.002
Moderate	37 (40.2)	55 (59.8)		57 (62.0)	35 (38.0)		40 (43.5)	52 (56.5)		58 (63.0)	34 (37.0)		70 (76.1)	22 (23.9)		33 (35.9)	59 (64.1)	
Poor	20 (55.6)	16 (44.4)		21 (58.3)	15 (41.7)		22 (61.1)	14 (38.9)		20 (55.6)	16 (44.4)		28 (77.8)	8 (22.2)		17 (47.2)	19 (52.8)	
Depth of tumour invasion																		
Inter-serosa	19 (28.8)	47 (71.2)	0.015	41 (62.1)	25 (37.9)	0.828	20 (30.3)	46 (69.7)	0.006	41 (62.1)	25 (37.9)	0.213	55 (83.3)	11 (16.7)	0.199	16 (24.2)	50 (75.8)	0.051
Serosa	37 (45.1)	45 (54.9)		49 (59.8)	33 (40.2)		38 (46.3)	44 (53.7)		54 (65.9)	28 (34.1)		58 (70.7)	24 (29.3)		29 (35.4)	53 (64.6)	
Adjacent organs	17 (51.5)	16 (48.5)		20 (60.6)	13 (39.4)		19 (57.6)	14 (42.4)		25 (75.8)	8 (24.2)		25 (75.8)	8 (24.2)		14 (42.4)	19 (57.6)	
Lymph node metastasis																		
Negative	30 (32.3)	63 (67.7)	0.023	52 (55.9)	41 (44.1)	0.170	31 (33.3)	62 (66.7)	0.010	69 (74.2)	24 (25.8)	0.021	75 (80.6)	18 (19.4)	0.154	20 (21.5)	73 (78.5)	0.001
Positive	43 (48.9)	45 (51.1)		58 (65.9)	30 (34.1)		46 (52.3)	42 (47.7)		51 (58.0)	37 (42.0)		63 (71.6)	25 (28.4)		39 (44.3)	49 (55.7)	
Dukes' stage																		
a + b	21 (27.6)	55 (72.4)	0.003	42 (55.3)	34 (44.7)	0.155	23 (30.3)	53 (69.7)	0.007	56 (73.7)	20 (26.3)	0.056	60 (78.9)	16 (21.1)	0.468	14 (18.4)	62 (81.6)	0.001
c + d	52 (49.5)	53 (50.5)		69 (65.7)	36 (34.3)		53 (50.5)	52 (49.5)		63 (60.0)	42 (40.0)		78 (74.3)	27 (25.7)		45 (42.9)	60 (57.1)	
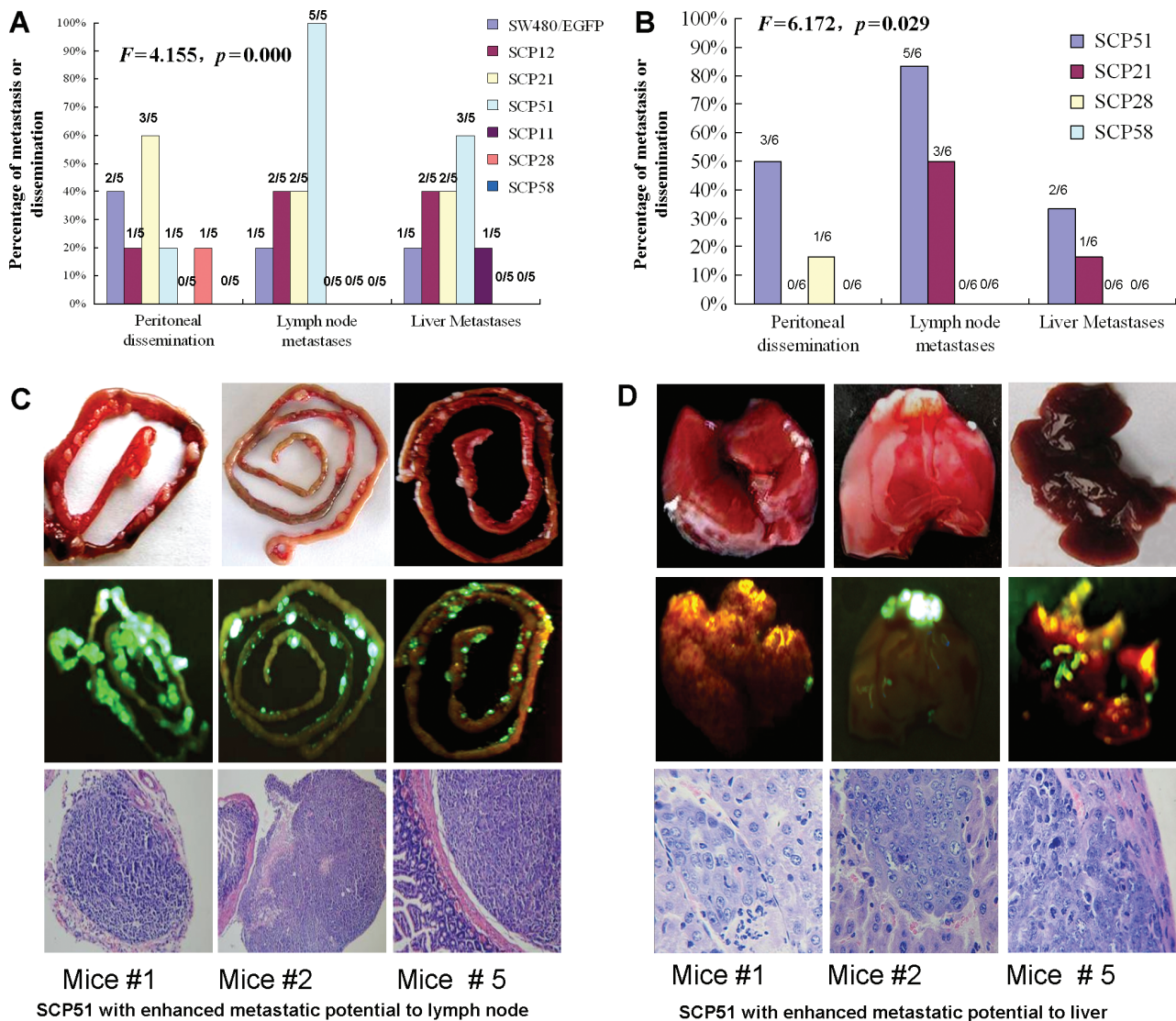
Figure 2. Spontaneous metastasis of SCPs in lymph node and liver, and peritoneal dissemination in orthotopic xenograft models. (A) Spontaneous metastasis and peritoneal dissemination of SCP12, SCP21, SCP51, SCP11, SCP28, SCP58, and SW480/EGFP. (B) Spontaneous metastasis and peritoneal dissemination of SCP51, SCP21, SCP28, and SCP58. (C) Whole-body optical images of lymphatic metastasis of SCP51 in the orthotopic mouse model. Lymphatic metastasis was confirmed by histological analyses. (D) Whole-body optical images of liver metastasis of SCP51 in the orthotopic mouse model. Liver metastasis was confirmed by histological examinations

Supplementary Table 5). We then further analysed these genes using the gene-annotation enrichment analysis of DAVID (gene enrichment score >1) and found that 80 genes were enriched in the top ten clusters (Supporting information, Supplementary Tables 6 and 7). Finally, we analysed the potential functions of these 80 genes by using Geneclip, a text-mining program.

Thirty-two genes were found to be closely associated with metastasis. These genes are known to be involved in several important pathways, such as transcription regulation (*SMARCA2*, *FOXO1A*, *HOXB8*, etc), development (*SEMA3A, MSX1, COL6A3*, etc), cell motility (*MARCKS, JAG1, THBS1*, etc), and signal transduction (*LYN, VDR, SAV1*, etc) (Supporting information, Supplementary Table 8). As shown in Figure 4C, *SCP51, SCP58, and SW480/EGFP* were clearly clustered into three

subgroups based on their metastatic potential using the 32-gene set. The literature profiling and constructing gene co-occurrence networks of the 32-gene set by the Geneclip analysis suggested that these 32 genes had important roles in metastasis (Figure 4D). A primary co-occurrence network between the 32-gene set and metastasis, colorectal cancer, melanoma, and other cancers further highlighted the potential of these genes in tumour metastasis (Figure 4E).

Validation of the expression of LYN, SDCBP, MAP4K4, DKK1, and MID1 in the primary tumours

We selected LYN, SDCBP, MAP4K4, DKK1, and MID1 from the 32-gene set for further investigations since their roles in CRC were unclear. Both
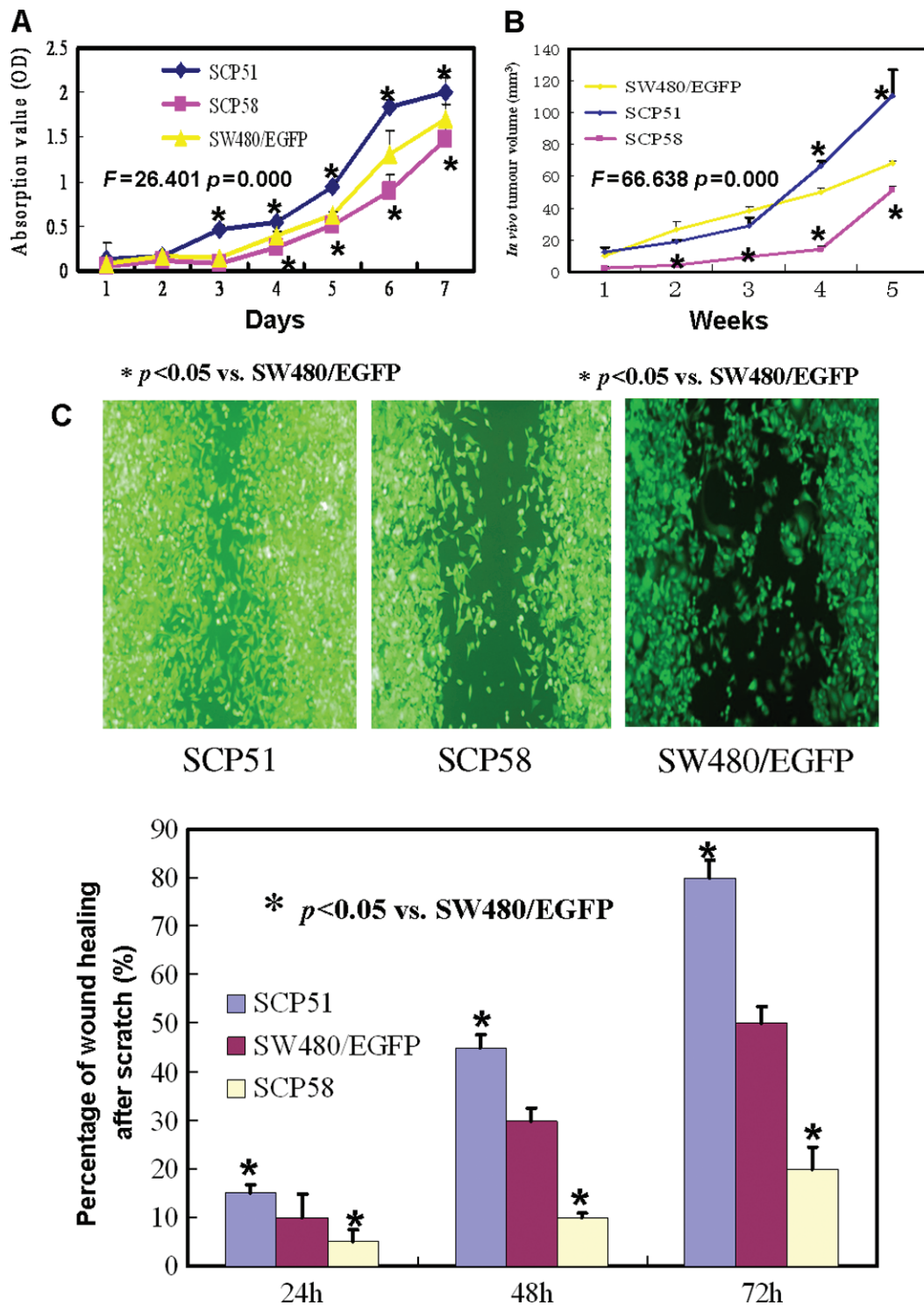
Figure 3. *In vitro* proliferation, *in vivo* tumour growth, and migration of SCP51, SCP58, and SW480/EGFP. (A) *In vitro* growth of SCP51 and SCP58 cells and SW480/EGFP. The values represent the absorbance value (OD) detected by MTT assay. (B) Subcutaneous growth of the tumours of SCP58, SCP51, and SW480/EGFP on the flanks of nude mice. Each value represents the mean ± SD of the tumour volume. (C) Cell migration of SCP51, SCP58, and SW480/EGFP for 72 h in a scratch wound-healing assay

RT-PCR and western blotting were used to validate their expression in fresh CRC samples. As shown in Figure 5, compared with that of the paired normal tissues, the expression of LYN and SDCBP was increased in most of the tumours at both mRNA and protein levels. In contrast, the expression of DKK1 was decreased in most of the tumours. For MID1, the expression of mRNA and protein of 75 kD, but not 34 kD, was decreased in most of the tumours. MID1 is known to be associated with microtubules as a homodimer [16], and thus the 75 kD protein is presumably the homodimer. The mRNA expression of MAP4K4 was increased in the primary tumours, consistent with the array result; however, the protein level

Table 2. Summary of univariate and multivariate Cox regression analysis of overall survival duration

Variable	Univariate			Multivariate		
	HR	CI (95%)	p	HR	CI (95%)	p
Gender	−0.066	−0.507–0.375	0.770			
Age	0.012	−0.004–0.028	0.151			
DKK1	−0.376	−0.850–0.098	0.120			
SDCBP	0.446	−0.013–0.905	0.056			
LYN	0.719	0.290–1.148	0.001*			
MAP4K4	0.477	0.048–0.906	0.029*			
Differentiation	0.559	0.251–0.867	0.000*	0.317	0.979–1.926	0.066
Dukes' stage	1.025	1.710–4.538	0.000*	0.669	1.172–3.251	0.010*
MID1	−0.712	−1.145––0.279	0.001*	−0.498	0.391–0.945	0.027*
Five-gene signature	1.183	0.752–1.614	0.000*	0.888	1.550–3.812	0.000*

* $p < 0.05$

was decreased in most of the primary tumours, suggesting post-transcriptional modifications of MAP4K4 in the tumour.

Correlation of gene expression with clinicopathological characteristics of the CRC patients

To further investigate the role of *LYN*, *SDCBP*, *MAP4K4*, *DKK1*, and *MID1* in colorectal tumour metastasis, we measured the protein levels of these genes in 181 archived paraffin-embedded CRC samples using immunohistochemical staining (Figure 6) and analysed the correlation of the clinicopathological characteristics of these patients with the expression of *LYN*, *SDCBP*, *MAP4K4*, *DKK1*, *MID1*, and the five-gene signature. As summarized in Table 1, there were no significant associations of either individual expression of *LYN*, *SDCBP*, *MAP4K4*, *DKK1*, and *MID1*, or the expression of the five-gene signature with age or gender of the patients ($p > 0.05$). However, we found that expression of *LYN*, *MAP4K4*, *MID1*, and the five-gene signature correlated closely with tumour differentiation and lymph node metastasis (Table 1). In addition, expression of *LYN*, *MAP4K4*, and the five-gene signature was significantly associated with Dukes' stage ($p < 0.01$). Moreover, expression of *LYN* and *MAP4K4* correlated significantly with tumour invasion ($p < 0.05$), although the five-gene signature did not correlate well with tumour invasion ($p = 0.051$).

Survival analysis and log-rank test results

To investigate the prognostic value of *MAP4K4*, *LYN*, *SDCBP*, *DKK1*, *MID1*, and the five-gene signature for CRC, we evaluated the association of individual expression of *MAP4K4*, *LYN*, *SDCBP*, *DKK1*, and *MID1*, and expression of the five-gene signature, with the survival rate using Kaplan–Meier analysis with the log-rank test. As shown in Figures 7A–7E, expression of *MAP4K4*, *LYN*, *SDCBP*, and *MID1* was found to correlate significantly with the overall survival of patients. Notably, the five-gene signature

also correlated highly with the overall survival of patients (Figure 7F). Clearly, the correlation of the five-gene signature with patient survival was much better than any single gene (see Figures 7A–7F).

To determine whether the expression of the five-gene signature acts as an independent prognostic factor for CRC, we performed multivariate survival analysis of the expression of the five-gene signature with numerous factors such as gender, age, tumour differentiation, Dukes' classification, and individual expression of the five genes. As shown in Table 2, Dukes' classification and the expression of MID1 and the five-gene signature were independent prognostic factors for CRC patients, of which the five-gene signature was the best prognostic factor ($p < 0.001$).

Discussion

In this study, we isolated SCPs from the SW480 cell line *in vitro* and assessed their ability to grow subcutaneously on the flanks of nude mice. Less than 50% of the SCPs (29/59) were capable of forming visible tumours *in vivo* and the growth traits of the SCPs were shown to be significantly different. Two groups of SCPs that exhibited the capability of either high or low metastasis were separated by orthotopic implantation of SCP cells in nude mice. Because the difference of gene profiling between the high and the low metastatic SCPs may be clinically relevant, we performed microarray analyses to excavate the molecules that drove the metastasis.

We used several approaches to identify the genes involved in CRC metastasis. First, we screened the genes expressed differentially between the high and the low metastatic SCPs. We then cross-compared the differential genes expressed between SCP51 and SCP58 with those found between SCP58 and SW480/ EGFP, and obtained the common genes associated with CRC metastasis. Moreover, we picked the enriched genes from the common genes by using the gene-annotation enrichment analysis of DAVID. Finally, we used Geneclip, text-mining software developed in our laboratory, to find the genes associated
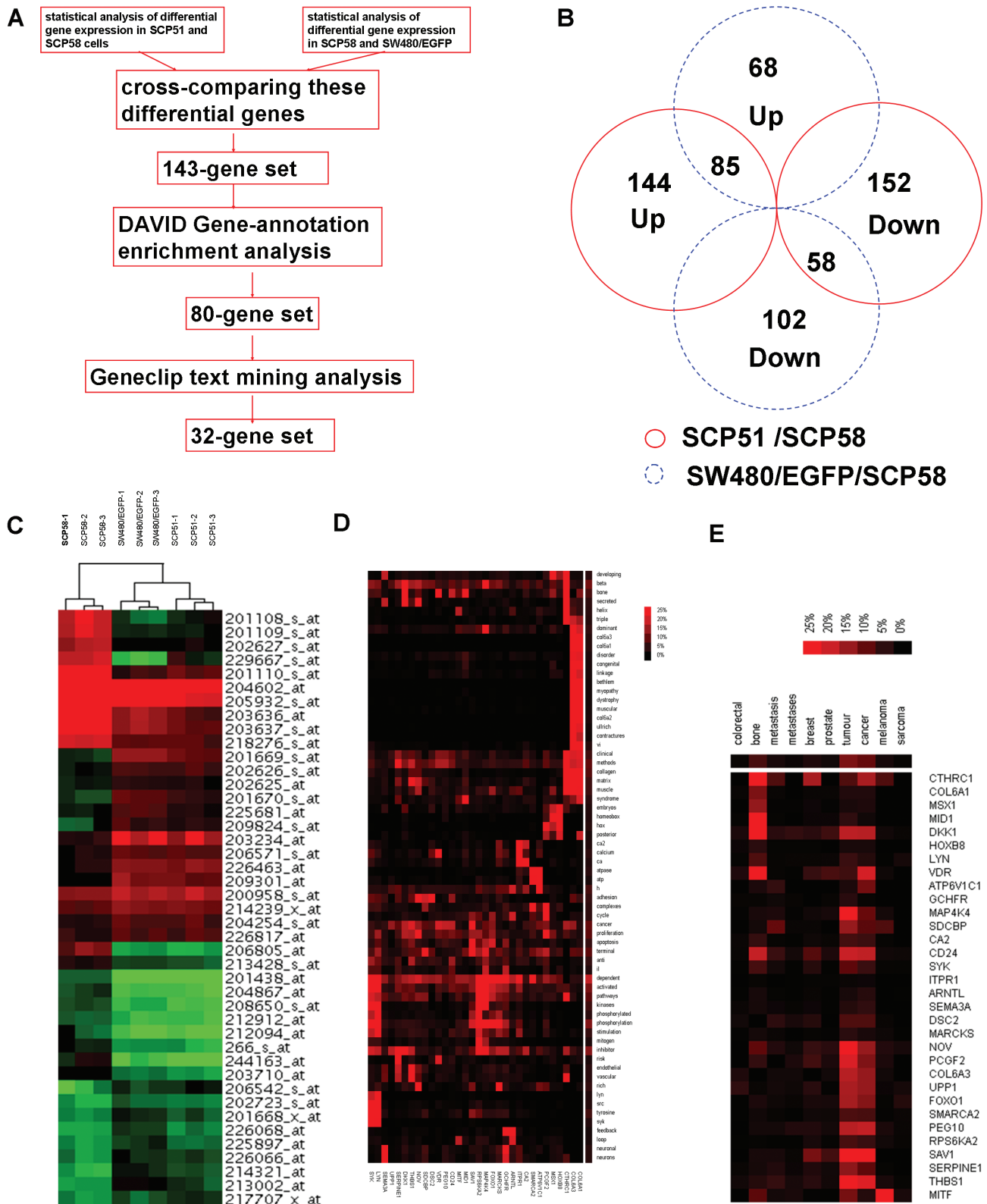
Figure 4. 32 candidate genes with metastatic potential were identified by microarray analyses. (A) A workflow of the strategy for identification of the genes associated with CRC metastasis. (B) Venn diagrams of the common differential genes identified by cross-comparing the genes expressed differentially between SCP51 and SCP58 cells, or between SCP58 and SW480/EGFP. (C) Hierarchical cluster diagram of 32 differentially expressed genes in SCP51, SCP58 cells, and SW480/EGFP. (D) Hierarchical cluster diagram of the 32-gene set based on literature profiling and constructing gene co-occurrence networks using default indicators of Geneclip. (E) A primary co-occurrence network between the 32-gene set and metastasis, colorectal cancer, melanoma, and other cancers by the Geneclip analysis

with metastasis. We successfully reduced the number of candidate genes from 143 to 32. The strategy is straightforward and without any complex calculations, and is thus adaptable for use in other studies.
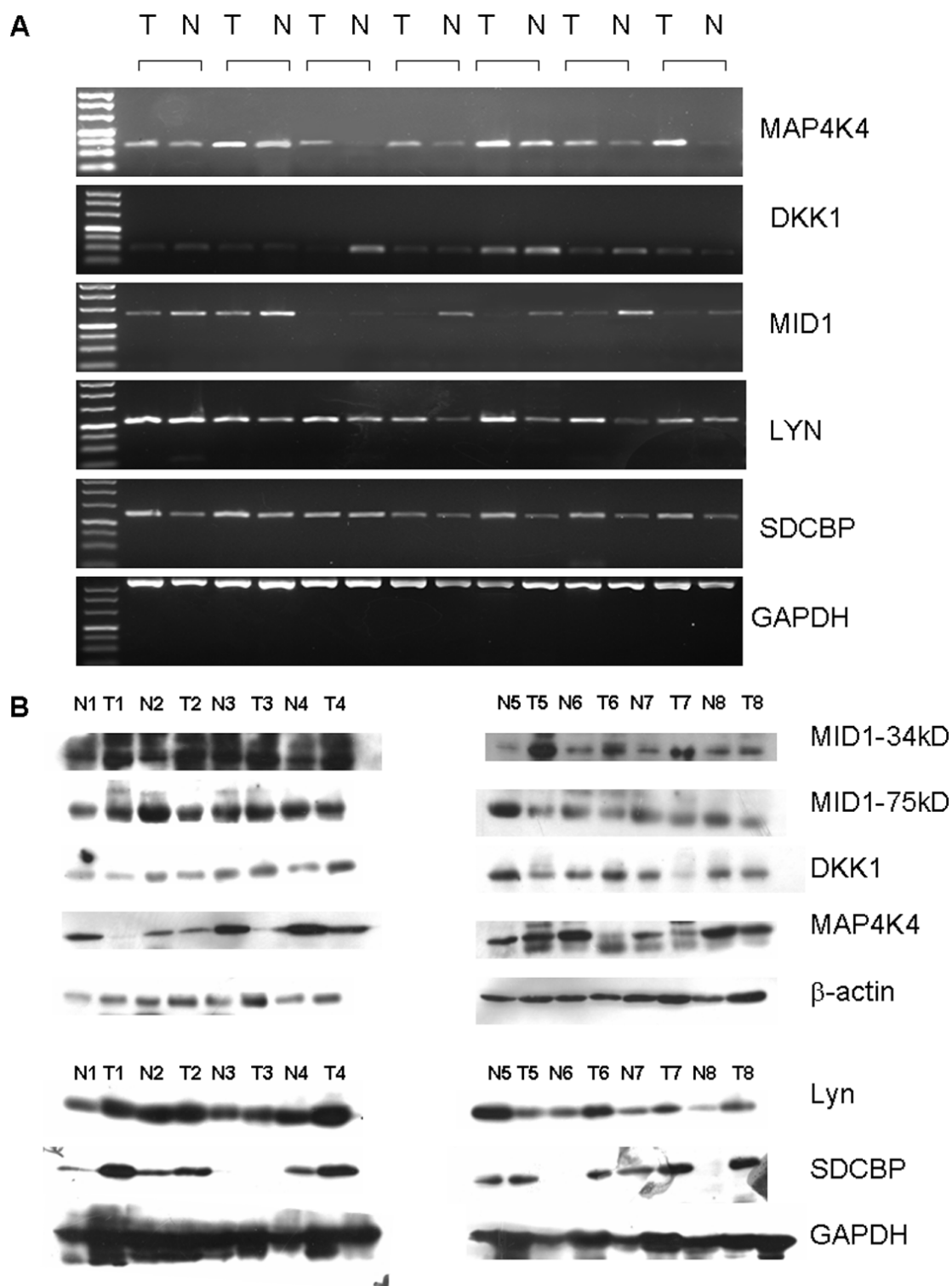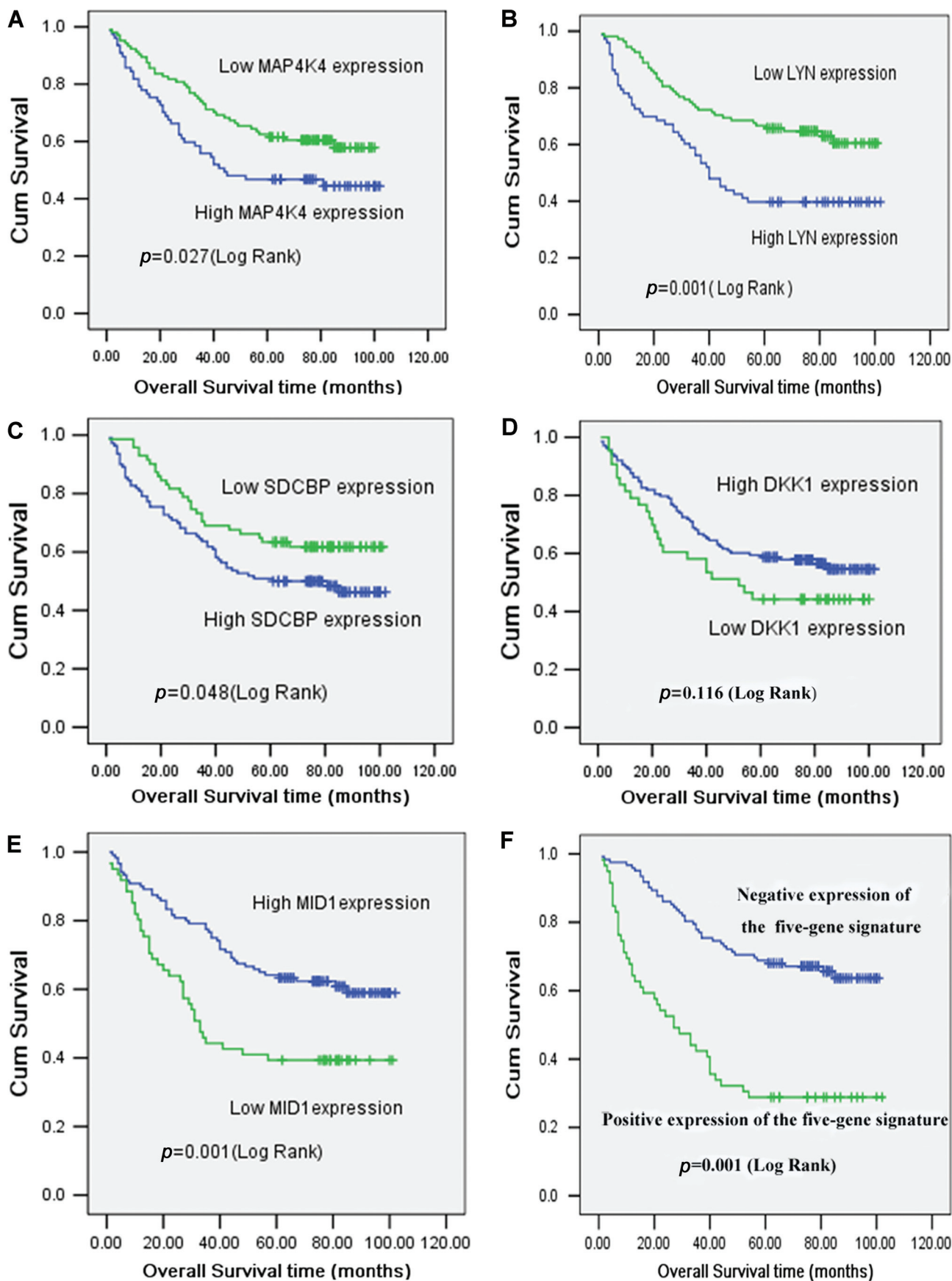
Figure 5. Expression of LYN, SDCBP, MAP4K4, DKK1, and MID1 in tumours. (A) Expression of the genes at the mRNA level detected by RT-PCR. (B) Expression of the genes at the protein level detected by western blotting. Paired tissues from different patients: N, normal tissue; T, colorectal tumour tissue

Of the 32 candidate genes, some have been well studied in CRC or metastasis, while others are poorly understood in CRC or metastasis. It has been noted that CA2 [17], CD24 [18,19], DKK1 [20], DSC2 [21], HOXB8 [22], LYN [23], SEMA3A [24], THBS1 [25], and VDR [26] are related with CRC; ARNTL, DKK1, DSC2, LYN, UPP1, and VDR are involved in the regulation of adhesion, growth, sensitivity or resistance to anti-cancer drugs of CRC. In addition, COL6A3 [27], EVI1 [28], and MARCKS [29] are known to be associated with the progression of breast cancer, melanoma or prostate cancer. We selected LYN, SDCBP, MAP4K4, DKK1, and MID1 for profound investigations because their roles in CRC metastasis are unclear. However, it is known that DKK1 is involved in embryonic development by inhibition
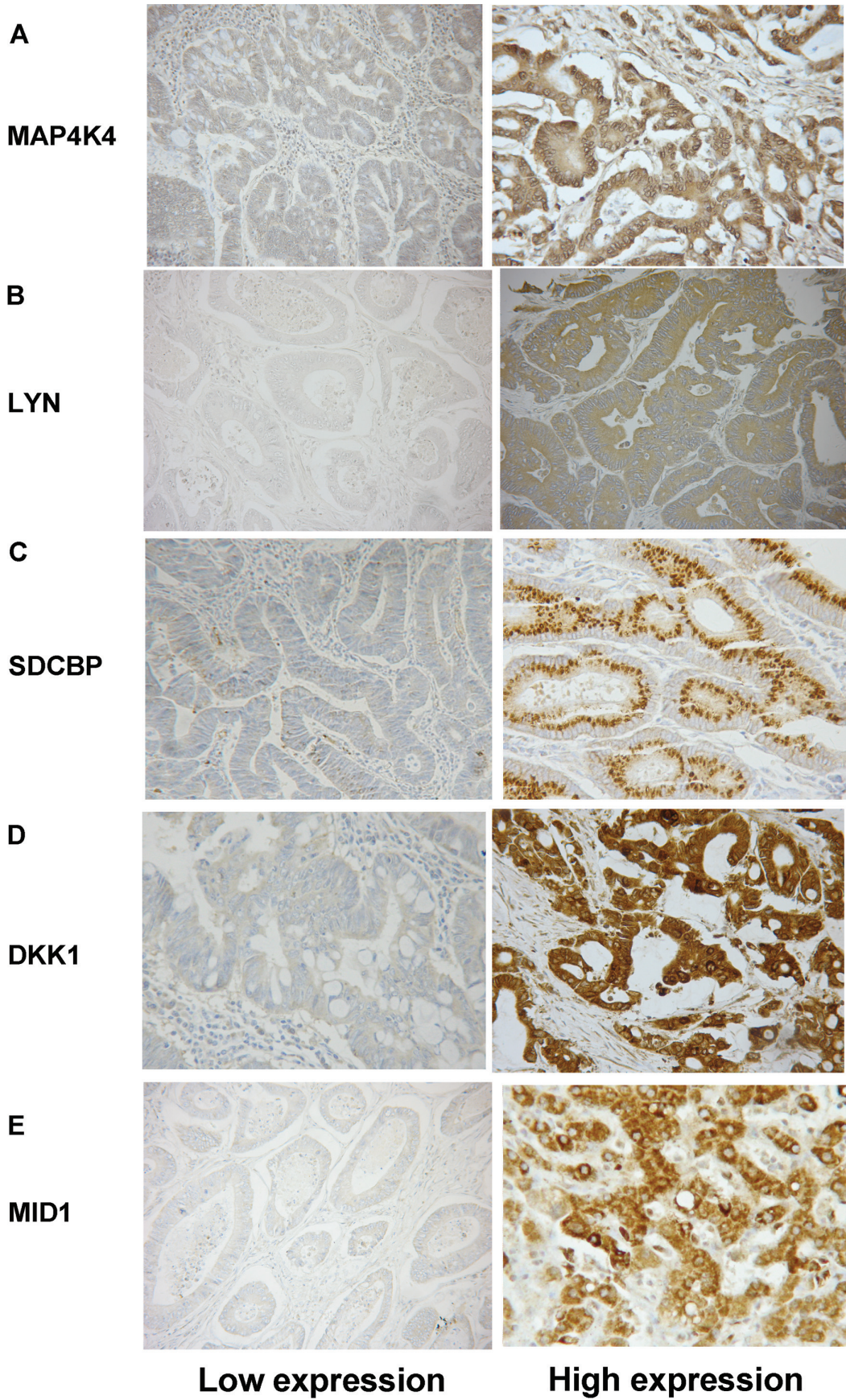
Figure 6. Immunohistochemical analysis of MAP4K4, LYN, SDCBP, DKK1, and MID1 on colorectal tumours. Low expression in some tumour tissues is presented in the panels on the left and high expression in tumour tissues in the panels on the right. Original magnification: 400×

Figure 7. Correlations of the expression of LYN, SDCBP, MAP4K4, DKK1, MID1, and the five-gene signature with Kaplan–Meier overall survival of CRC patients. The log-rank test was used to calculate *p* values

of the WNT signalling pathway [30]; DKK1 has been studied in colon [30,31] and gastric cancers [32]. LYN is a Src family protein tyrosine kinase expressed preferentially in haematopoietic cells and B cells. It was shown that activation of LYN stimulated cell proliferation and inhibited apoptosis. The expression of LYN has also been reported in solid tumours including prostate cancer [33], pancreatic cancer [34], colon carcinoma [35], and Kaposi's sarcoma [36]. LYN induced chemoresistance in colon cancer cell lines via the PI3-K/PKB-mediated cell survival pathway [35]. MID1 is a member of the tripartite motif (TRIM) family. It is involved in the formation of multi-protein structures, acting as anchor points to microtubules [37]. SDCBP (also called mda-9, syntenin) contains tandemly repeated PDZ domains that bind the cytoplasmic C-terminal domains of a variety of transmembrane proteins, which affect cytoskeletal-membrane organization, cell adhesion, protein trafficking, and activation of transcription factors. SDCBP was up-regulated in metastatic melanomas [38,39] and breast and gastric carcinomas [40]. Recently, SDCBP was shown to be an essential gene in the control of melanoma progression and metastasis [39]. MAP4K4 is a member of the serine/threonine protein kinase family. It is known that MAP4K4 specifically activates MAPK8/JNK [41] and is involved in cell migration and invasion in ovarian cancer [42] and breast cancer [43].

In this study, we found that LYN and SDCBP were up-regulated, while DKK1 and MID1 were down-regulated in colorectal tumours. The expression of LYN, MAP4K4, and the five-gene signature were closely associated with metastasis and Dukes' stage of CRC. In addition, the expression of MAP4K4, LYN, SDCBP, MID1, and the five-gene signature correlated significantly with the overall survival of CRC patients. Moreover, the expression of MID1 and the five-gene signature were independent prognostic factors for CRC patients, of which the five-gene signature was the best prognostic factor. Our findings provided new insights into the role of these genes, particularly the five-gene signature in CRC progression.

As the genes of interest function in various molecular pathways, their promotion of CRC metastasis would be achieved by means of regulating numerous signalling pathways. However, the fact that the five-gene signature correlates more closely with the metastasis and overall survival of CRC patients suggests that the concerted effects of many genes, the regulation of several pathways, and the cross-interactions of the pathways would be more important in CRC metastasis and patient survival. It has been reported that the combination of some genes in the metastasis signature led to a more aggressive metastatic activity of the breast cancer cell population [10]. Clearly, how the five-gene signature regulates the metastasis of CRC needs to be further investigated. siRNA-mediated knockdown in highly metastatic CRC cell lines and virus-mediated knock-in in poorly metastatic CRC cell lines of individual genes and the gene combination will be our future work. Further investigations with additional patient cohorts and delineation of the specific roles of these genes in the metastatic process may lead to a better understanding of the metastatic mechanisms and the development of a novel therapeutic strategy.

Acknowledgements

References

Note: Reference 44 is cited in the Supporting information to this article.

1. Swanton C, Caldas C. Molecular classification of solid tumours: towards pathway-driven therapeutics. *Br J Cancer* 2009;**100:**1517–1522.
2. Acharya CR, Hsu DS, Anders CK, Anguiano A, Salter KH, Walters KS, et al. Gene expression signatures, clinicopathological features, and individualized therapy in breast cancer. *J Am Med Assoc* 2008;**299:**1574–1587.
3. Garcia-Escudero R, Paramio JM. Gene expression profiling as a tool for basic analysis and clinical application of human cancer. *Mol Carcinog* 2008;**47:**573–579.
4. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;**415:**530–536.
5. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. *J Clin Invest* 2004;**113:**913–923.
6. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nature Genet* 2003;**33:**49–54.
7. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347:**1999–2009.
8. Kallioniemi A. Molecular signatures of breast cancer — predicting the future. *N Engl J Med* 2002;**347:**2067–2068.
9. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 2003;**3:**537–549.
10. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. *Nature* 2005;**436:**518–524.
11. Bos PD, Zhang XH, Nadal C, Shu W, Gomis RR, Nguyen DX, et al. Genes that mediate breast cancer metastasis to the brain. *Nature* 2009;**459:**1005–1009.
12. Wang S, Zhou J, Wang XY, Hao JM, Chen JZ, Zhang XM, et al. Down-regulated expression of SATB2 is associated with metastasis and poor prognosis in colorectal cancer. *J Pathol* 2009;**219:**114–122.
13. Huang ZX, Tian HY, Hu ZF, Zhou YB, Zhao J, Yao KT. GenCLiP: a software program for clustering gene lists by literature profiling and constructing gene co-occurrence networks related to custom. *BMC Bioinformatics* 2008;**9:**308.
14. Masunaga R, Kohno H, Dhar DK, Ohno S, Shibakita M, Kinugasa S, et al. Cyclooxygenase-2 expression correlates with tumor neovascularization and prognosis in human colorectal carcinoma patients. *Clin Cancer Res* 2000;**6:**4064–4068.
15. Soumaoro LT, Uetake H, Higuchi T, Takagi Y, Enomoto M, Sugihara K. Cyclooxygenase-2 expression: a significant prognostic

indicator for patients with colorectal cancer. *Clin Cancer Res* 2004;**10**:8465–8471.

16. Winter J, Lehmann T, Krauss S, Trockenbacher A, Kijas Z, Foerster J, *et al*. Regulation of the MID1 protein function is fine-tuned by a complex pattern of alternative splicing. *Hum Genet* 2004;**114**:541–552.

17. Ki DH, Jeung HC, Park CH, Kang SH, Lee GY, Lee WS, *et al*. Whole genome analysis for liver metastasis gene signatures in colorectal cancer. *Int J Cancer* 2007;**121**:2005–2012.

18. Sagiv E, Memeo L, Karin A, Kazanov D, Jacob-Hirsch J, Mansukhani M, *et al*. CD24 is a new oncogene, early at the multi-step process of colorectal cancer carcinogenesis. *Gastroenterology* 2006;**131**:630–639.

19. Sagiv E, Starr A, Rozovski U, Khosravi R, Altevogt P, Wang T, *et al*. Targeting CD24 for treatment of colorectal and pancreatic cancer by monoclonal antibodies or small interfering RNA. *Cancer Res* 2008;**68**:2803–2812.

20. Aguilera O, Pena C, Garcia JM, Larriba MJ, Ordonez-Moran P, Navarro D, *et al*. The Wnt antagonist DICKKOPF-1 gene is induced by 1alpha,25-dihydroxyvitamin D3 associated to the differentiation of human colon cancer cells. *Carcinogenesis* 2007;**28**:1877–1884.

21. Khan K, Hardy R, Haq A, Ogunbiyi O, Morton D, Chidgey M. Desmocollin switching in colorectal cancer. *Br J Cancer* 2006;**95**:1367–1370.

22. Vider BZ, Zimber A, Hirsch D, Estlein D, Chastre E, Prevot S, *et al*. Human colorectal carcinogenesis is associated with deregulation of homeobox gene expression. *Biochem Biophys Res Commun* 1997;**232**:742–748.

23. Chen WS, Kung HJ, Yang WK, Lin W. Comparative tyrosine-kinase profiles in colorectal cancers: enhanced arg expression in carcinoma as compared with adenoma and normal mucosa. *Int J Cancer* 1999;**83**:579–584.

24. Nguyen QD, Rodrigues S, Rodrigue CM, Rivat C, Grijelmo C, Bruyneel E, *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**:2070–2077.

25. Rashid A, Shen L, Morris JS, Issa JP, Hamilton SR. CpG island methylation in colorectal adenomas. *Am J Pathol* 2001;**159**:1129–1135.

26. Evans SR, Nolla J, Hanfelt J, Shabahang M, Nauta RJ, Shchepotin IB. Vitamin D receptor expression as a predictive marker of biological behavior in human colorectal cancer. *Clin Cancer Res* 1998;**4**:1591–1595.

27. Tilman G, Mattiussi M, Brasseur F, van Baren N, Decottignies A. Human periostin gene expression in normal tissues, tumors and melanoma: evidence for periostin production by both stromal and melanoma cells. *Mol Cancer* 2007;**6**:80.

28. Brooks DJ, Woodward S, Thompson FH, Dos Santos B, Russell M, Yang JM, *et al*. Expression of the zinc finger gene EVI-1 in ovarian and other cancers. *Br J Cancer* 1996;**74**:1518–1525.

29. Li T, Li D, Sha J, Sun P, Huang Y. MicroRNA-21 directly targets MARCKS and promotes apoptosis resistance and invasion in prostate cancer cells. *Biochem Biophys Res Commun* 2009;**383**:280–285.

30. Gonzalez-Sancho JM, Aguilera O, Garcia JM, Pendas-Franco N, Pena C, Cal S, *et al*. The Wnt antagonist DICKKOPF-1 gene is a downstream target of beta-catenin/TCF and is downregulated in human colon cancer. *Oncogene* 2005;**24**:1098–1103.

31. Aguilera O, Fraga MF, Ballestar E, Paz MF, Herranz M, Espada J, *et al*. Epigenetic inactivation of the Wnt antagonist DICKKOPF-1 (DKK-1) gene in human colorectal cancer. *Oncogene* 2006;**25**:4116–4121.

32. Sato H, Suzuki H, Toyota M, Nojima M, Maruyama R, Sasaki S, *et al*. Frequent epigenetic inactivation of DICKKOPF family genes in human gastrointestinal tumors. *Carcinogenesis* 2007;**28**:2459–2466.

33. Goldenberg-Furmanov M, Stein I, Pikarsky E, Rubin H, Kasem S, Wygoda M, *et al*. Lyn is a target gene for prostate cancer: sequence-based inhibition induces regression of human tumor xenografts. *Cancer Res* 2004;**64**:1058–1066.

34. Fu Y, Zagozdzon R, Avraham R, Avraham HK. CHK negatively regulates Lyn kinase and suppresses pancreatic cancer cell invasion. *Int J Oncol* 2006;**29**:1453–1458.

35. Bates RC, Edwards NS, Burns GF, Fisher DE. A CD44 survival pathway triggers chemoresistance via lyn kinase and phospho-inositide 3-kinase/Akt in colon carcinoma cells. *Cancer Res* 2001;**61**:5275–5283.

36. Munshi N, Ganju RK, Avraham S, Mesri EA, Groopman JE. Kaposi's sarcoma-associated herpesvirus-encoded G protein-coupled receptor activation of c-jun amino-terminal kinase/stress-activated protein kinase and lyn kinase is mediated by related adhesion focal tyrosine kinase/proline-rich tyrosine kinase 2. *J Biol Chem* 1999;**274**:31863–31867.

37. Tao H, Simmons BN, Singireddy S, Jakkidi M, Short KM, Cox TC, *et al*. Structure of the MID1 tandem B-boxes reveals an interaction reminiscent of intermolecular ring heterodimers. *Biochemistry* 2008;**47**:2450–2457.

38. Helmke BM, Polychronidis M, Benner A, Thome M, Arribas J, Deichmann M. Melanoma metastasis is associated with enhanced expression of the syntenin gene. *Oncol Rep* 2004;**12**:221–228.

39. Boukerche H, Su ZZ, Prevot C, Sarkar D, Fisher PB. mda-9/Syntenin promotes metastasis in human melanoma cells by activating c-Src. *Proc Natl Acad Sci U S A* 2008;**105**:15914–15919.

40. Koo TH, Lee JJ, Kim EM, Kim KW, Kim HD, Lee JH. Syntenin is overexpressed and promotes cell migration in metastatic human breast and gastric cancer cell lines. *Oncogene* 2002;**21**:4080–4088.

41. Collins CS, Hong J, Sapinoso L, Zhou Y, Liu Z, Micklash K, *et al*. A small interfering RNA screen for modulators of tumor cell motility identifies MAP4K4 as a promigratory kinase. *Proc Natl Acad Sci U S A* 2006;**103**:3775–3780.

42. Urzua U, Roby KF, Gangi LM, Cherry JM, Powell JI, Munroe DJ. Transcriptomic analysis of an *in vitro* murine model of ovarian carcinoma: functional similarity to the human disease and identification of prospective tumoral markers and targets. *J Cell Physiol* 2006;**206**:594–602.

43. Wang T, Lee K, Rehman A, Daoud SS. PRIMA-1 induces apoptosis by inhibiting JNK signaling but promoting the activation of Bax. *Biochem Biophys Res Commun* 2007;**352**:203–212.

44. Rashidi B, Gamagami R, Sasson A, Sun FX, Geller J, Moossa AR, *et al*. An orthotopic mouse model of remetastasis of human colon cancer liver metastasis. *Clin Cancer Res* 2000;**6**:2556–2561.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article.

Supplemental experimental procedures.

Table S1. Tumour growth under subcutaneous tissue of nude mice of SCPs including SCP3, SCP9, SCP10, SCP20, SCP28, SCP51, SCP58, and SW480/EGFP (mean $\pm$ SD, $n = 3$).

Table S2. Cell cycles of SCP51, SCP58, and their parent SW480/EGFP cells detected by flow cytometry (mean $\pm$ SD, $n = 3$).

Table S3. The differential genes between SCP51 and SCP58 cells (SCP51 versus SCP58).

Table S4. The differential genes between SW480/EGFP and SCP58 (SW480/EGFP versus SCP58).

Table S5. Common genes found as the candidates for CRC metastasis when cross-comparing differential genes expressed between SCP51 and SCP58 cells with the differential genes found between SCP58 and SW480/EGFP.

Table S6. Genes found in the top ten clusters using gene-annotation enrichment analysis of DAVID.

Table S7. Fold changes in the genes of the top ten clusters using gene-annotation enrichment analysis of DAVID in SCP51, SCP58 and SW480/EGFP cells.

Table S8. A 32-gene set associated with metastasis of CRC using cross-comparing, functional analysis and gene annotation, text mining.