ORIGINAL ARTICLE

Claudin-7 indirectly regulates the integrin/FAK signaling pathway in human colon cancer tissue

Lei Ding[1,4], Liyong Wang[2,4], Leiming Sui[2], Huanying Zhao[2], Xiaoxue Xu[2], Tengyan Li[1], Xiaonan Wang[1], Wenjing Li[1], Ping Zhou[3] and Lu Kong[2]

The claudin family of proteins is integral to the structure and function of tight junctions. The role of claudin-7 (Cldn-7, *CLDN7*) in regulating the integrin/focal adhesion kinase (FAK)/ERK signaling pathway remains poorly understood. Therefore, we investigated differences in gene expression, primarily focusing on *CLDN7* and integrin/FAK/ERK signaling pathway genes, between colon cancer and adjacent normal tissues. Quantitative real-time reverse transcription–PCR and immunohistochemistry were utilized to verify the results of mRNA and protein expression, respectively. *In silico* analysis was used to predict co-regulation between Cldn-7 and integrin/FAK/ERK signaling pathway components, and the STRING database was used to analyze protein–protein interaction pairs among these proteins. Meta-analysis of expression microarrays in The Cancer Genome Atlas (TCGA) database was used to identify significant correlations between Cldn-7 and components of predicted genes in the integrin/FAK/ERK signaling pathway. Our results showed marked cancer stage-specific decreases in the protein expression of Cldn-7, Gelsolin, MAPK1 and MAPK3 in colon cancer samples, and the observed changes for all proteins except Cldn-7 were in agreement with changes in the corresponding mRNA levels. Cldn-7 might indirectly regulate *MAPK3* via *KRT8* due to *KRT8* co-expression with *MAPK3* or *CLDN7*. Our bioinformatics methods supported the hypothesis that Cldn-7 does not directly regulate any genes in the integrin/FAK/ERK signaling pathway. These factors may participate in a common network that regulates cancer progression in which the MAPK pathway serves as the central node.

INTRODUCTION

Colorectal cancer (CRC), the most common type of gastrointestinal cancer, is one of the leading causes of mortality worldwide.[1] Understanding the mechanisms underlying the growth and progression of CRC is essential for improving treatment. Several pathways have been implicated in colorectal carcinogenesis; for instance, the integrin/FAK/ERK pathway is activated in colon cancer cells.[2] Integrins constitute a family of heterodimeric glycoproteins involved in bidirectional cell signaling that participate in the regulation of gene transcription and cell shape, adhesion, migration, differentiation, survival and proliferation.[3,4] Studies also suggest that proton beam irradiation can inhibit metastatic potential, including cell adhesion and migration, by modulating the expression of genes involved in integrin trafficking and integrin-mediated signaling, which are necessary for tumor progression.[5,6] Claudins are cell adhesion molecules that localize to tight junctions. Claudin-7 (Cldn-7, *CLDN7*) is unique in that it has a stronger basolateral membrane distribution than other claudins, proteins that typically localize to apical tight junctions in the intestinal epithelium. Cldn-7 is highly expressed in the normal colon, and its knockdown in mice results in altered epithelial cell homeostasis and neonatal death. Notably, dysregulation of

epithelial homeostasis potentiates oncogenic transformation and tumor growth. However, the role of Cldn-7 in the regulation of colon tumorigenesis remains poorly understood.[7] Our recent study provides strong evidence that Cldn-7 performs a non-tight junction-related function in mediating cell-matrix interactions by interacting with integrin α2 at the basolateral membrane of normal intestinal epithelia;[8] however, the modulatory effect of Cldn-7 on the integrin/FAK pathway has not been reported to date. In the present study, tissue microarrays and gene chips were analyzed to detect the Cldn-7 expression levels at different stages of colorectal carcinoma, and bioinformatic analysis was performed to identify the relationships between Cldn-7 and proteins of the integrin/FAK signaling pathway. We also examined the expression of putatively related genes at the mRNA or protein level. The purpose of this study was to provide a theoretical basis for the relationship between Cldn-7 and the integrin/FAK signaling pathway in colon tumorigenesis.

MATERIALS AND METHODS

Ethical statement

Colon cancer specimens were obtained with informed consent from Beijing Shijitan Hospital of Capital Medical University with the approval of the Ethical

[1]Department of Oncology, Beijing Shijitan Hospital, Capital Medical University, Beijing, China; [2]Department of Biochemistry and Molecular Biology, Capital Medical University, Beijing, China and [3]Department of Bioinformatics, Biomedical Engineering Institute of Capital Medical University, Capital Medical University, Beijing, China
[4]These authors contributed equally to this work.
Correspondence: Dr L Ding, Department of Oncology, Beijing Shijitan Hospital, Capital Medical University, Beijing 100038, China
E-mail: dinglei1005@126.com
or Dr L Kong, Department of Biochemistry and Molecular Biology, Cancer Institute, Capital Medical University, No. 10 Xitoutiao, You An Men, Beijing 100069, China.
E-mail: konglu@ccmu.edu.cn
Received 29 October 2015; revised 14 March 2016; accepted 14 March 2016; published online 28 April 2016

Review Board of Investigation in Human Beings at Beijing Shijitan Hospital of Capital Medical University. This investigation was performed in accordance with the ethical standards and according to the Declaration of Helsinki Principles as well as national and international guidelines.

Human tissue samples and tissue microarray analysis

In the microarray study, we analyzed samples of six paired carcinoma and para-cancerous healthy tissues. Adjacent control samples were collected 5 cm from the cancer tissue from colon carcinoma patients undergoing surgery. The cancer tissue histology was examined independently by two histopathologists who categorized the degree of differentiation as well ($n = 1$), moderate ($n = 4$) or poor ($n = 1$). Consensus was reached in all cases. The carcinoma samples were classified according to the tumor node metastasis (TNM) staging system. The distribution of individuals included in the study according to gender and age is shown in Supplementary Table S1.

A human colon cancer tissue array was purchased from US Biomax (cat. no.: HCol-Ade180CS-01, US Biomax, Inc., Swampscott, MA, USA). The microarrays contained tissue cores of 90 paired colon cancer and para-cancerous samples (90 para-cancerous normal tissue cores; 17 stage 1 colon cancer tissue cores; 13 stage 2A cancer tissue cores; 8 stage 2B cancer tissue cores; 2 stage 2C cancer tissue cores; 27 stage 3B cancer tissue cores; 8 stage 3C cancer tissue cores; 7 stage 4A cancer tissue cores; and 8 stage 4B cancer tissue cores; clinical staging standards based on the AJCC, seventh edition). The information for patients who provided the samples for the tissue array is available at http://www.superchip.com.cn/product/detail_182.aspx.

RNA isolation

Tissue samples were frozen in liquid nitrogen until use. RNA was isolated using the Ambion RNA Isolation kit (cat. no. 1560, Ambion, Applied Biosystems, Foster City, CA, USA), and the quantity and quality of the isolated RNA were assessed using a NanoDrop 2000 system (Thermo Scientific, Boston, MA, USA). The RNA (A260) and protein (A280) concentrations and sample purity (A260/A280 ratio) were also measured. Only high-quality (A260/A280 ≥ 2), intact total RNA was examined for samples that showed a classic 18S and 28S ribosomal RNA band pattern based on bioanalyzer results.

Gene expression microarray and statistics

RNA (300 ng) was reverse transcribed into complementary DNA and labeled with biotin-UTP using an RNA amplification kit (cat. no. 1791, Ambion). Complementary RNA was quantified using a NanoDrop 2000, and complementary RNA quality was further evaluated using a bioanalyzer. Biotinylated complementary RNA was hybridized to the HumanHT-12 whole-human genome expression array (47 231 probes; Illumina, San Diego, CA, USA); the microarrays were washed and subsequently scanned using an Illumina BeadStation 500.

The raw signal intensities for the microarray data were processed and analyzed using GenomeStudio (2008, Illumina), and background subtraction and average normalization were performed using the average signal intensities; P-values were then calculated. To examine the integrin/FAK signaling pathway, differential gene expression between CRC tissues and matched para-cancerous control tissues from six cases was investigated using significance analysis of microarrays based on paired Student's t-tests. Significance analysis of microarray was applied individually to the microarray data sets for each of the 12 tissue sets under consideration. During each significance analysis of microarray test, 100 random iterations were performed to determine the false discovery rate (FDR) for each gene. The FDR was constrained to zero to identify significantly differentially expressed genes (DEGs). Genes absent across all samples were filtered out, and genes displaying a significant interaction ($P < 0.05$) were filtered in. Statistically significant DEGs between cancerous and para-cancerous tissues that passed all of the described filtering criteria (including a foldchange cut-off of > 1.5 and an FDR-adjusted P-value of < 0.05) were entered into the Gene Functional Annotation Tool available at the DAVID website (https://david.ncifcrf.gov/) using its official gene symbols. The Gene Ontology options GOTERM_BP_ALL and GOTERM_MF_ALL were selected, and a functional annotation chart was generated. To retrieve only significant categories, a maximum P-value of 0.05 was chosen. The microarray

data have been submitted to Gene Expression Omnibus (accession number GSE54986; http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE54986).

For functional enrichment analysis of selected genes, gene probe IDs were annotated using Entrez gene symbols (GenomeStudio Plus HumanHT--12_V4_0) and analyzed using DAVID Bioinformatics Resources 6.7 to identify significantly over-represented KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways in the data set; the default settings and Fisher's exact test were applied. The selected pathways displayed $P < 0.01$. These results produced lists of the genes involved in the enriched pathways and provided a connectivity map of each pathway. A model network based on these pathways was generated using KEGG (http://www.kegg.jp/). Gene expression heatmaps were generated using GenomeStudio.

Quantitative real-time reverse transcription–PCR

The expression levels of selected genes identified by gene expression profiling based on the microarray results were validated by quantitative real-time reverse transcription–PCR (qRT-PCR). For extraction of total RNA from tissues, the samples were homogenized using a TissueLyser II system (Qiagen, Düsseldorf, Germany), and total RNA was isolated using commercial RNeasy Fibrous Tissue kits (cat. no. 74704, Qiagen). The extracted RNA was treated with DNase I and then quantified using a NanoDrop 2000 (Thermo). Complementary DNA synthesis was performed with ~ 4 μg RNA in a total volume of 20 μl using a SuperScript III first-strand synthesis system (cat. no. 18080-051, Invitrogen, Carlsbad, CA, USA), random primers and Oligo(dT)20 according to the manufacturer's protocol. The resulting complementary DNA was used as a template for quantitative PCR, which was performed using the IQ5 system (Bio-Rad, Hercules, CA, USA) with primers specific to the CLDN7, ITGB1BP1, PTK2 (FAK), MAPK1 (ERK2), PTK2B (FAK2, PYK2), DIAPH1, GSN (gelsolin), RHOA, MYLK, TLN1 (Talin-1) and MAPK3 genes and the ACTB and GAPDH housekeeping genes and Power SYBR Green PCR Master Mix (cat. no. 4367659, Applied Biosystems, Foster City, CA, USA). To avoid amplification of any contaminating genomic DNA, the primers were designed to span at least one intron. GAPDH and ACTB were used as internal controls for normalization of mRNA expression levels. The amplification efficiencies of all primers were calculated from the standard curve slope. The primer sequences are listed in Supplementary Table S2. The quantitative PCR protocol was as follows: 10 min at 95 °C followed by 40 cycles of denaturation at 95 °C for 15 s and annealing at 60 °C for 1 min. Melting curve analysis was performed to examine the final PCR products. Relative expression values were obtained from Ct values detected using IQ5 analysis software according to the calibration curve method. The data were normalized to a factor calculated as the geometric mean of the relative mRNA concentrations of the two most stable housekeeping genes (GAPDH and ACTB); these analyses were performed using geNORM and NormFinder under our experimental conditions. All samples were quantified in triplicate.

Immunohistochemistry of colon cancer tissue array samples and statistics

The sections used for tissue microarray analysis were dehydrated in a graded ethanol series from 50 to 100% ethanol. The sections were washed with Tris-buffered saline (TBS; 5 min, three times) and incubated in retrieval buffer (10 mM citrate buffer, pH 6.0) in an autoclave at 120 °C for 10 min. The sections were then treated with 3% H_2O_2 in distilled water for 10 min and rinsed thoroughly in distilled water, washed in TBS, pH 7.6, for 5 min, followed by a 10 min incubation in 2% bovine serum albumin diluted in TBS. Subsequently, the sections were incubated with primary antibodies overnight at 4 °C. The optimized dilutions for antibodies were as follows: Cldn-7 (cat. no. H00001366-D01, Novus, Littleton, CO, USA) 1:500; ERK1/2 (cat. no. 9101S, CST, Boston, MA, USA) 1:800; Gelsolin (cat. no. NB110-10067, Novus) 1:1600; MYLK (cat. no. NBP2-32497, Novus) 1:150; PYK2 (cat. no. NBP1-61259, Novus) 1:6000; and Talin-1 (cat. no. NBP1-87820, Novus) 1:800. The next day, the sections were washed with TBS and incubated with the secondary antibody for 30 min at 37 °C. After an additional three washes for 5 min each in TBS, a streptavidin-biotin system and diaminobenzidine (cat. no. K-4011, DAKO, Glostrup, Denmark) were applied for 10 min. The slides were then rinsed in distilled water, followed by counterstaining with Mayer's hematoxylin for

90 s, rinsing in water for 5 min, and dehydration. Finally, all specimens were sealed with neutral glue. Images were captured using an Aperio ScanScope CS system (Vista, CA, USA). The relative distributions of the proteins of interest in the immunostained tissue array sections were semiquantitatively analyzed by a trained analyst using Positive Pixel Count (version 8.1, Vista, CA, USA) software. Only normal or tumor areas were selected and analyzed. The protein staining levels on IHC slides were collected as total optical density values per unit of positive area in one field of view. The final quantitative result of each protein level is the average value of six cube images: Positive density = total intensity of positive area/total positive area. Analysis of variance was adopted to statistically analyze the immunohistochemistry results, comparing cancer tissues of different stages with para-cancerous normal tissue. The mean ± s.d. values are displayed in the tables. The data were analyzed using SPSS, version 19.0 for Windows (SPSS, Inc., Chicago, IL, USA). $P < 0.05$ was considered to indicate a statistically significant difference.

In silico co-expression array database study

We predicted the association between Cldn-7, Gelsolin and MAPK1/3 using COXPRESdb version 6.0, located at http://coxpresdb.jp/,[9] and the STRING database (the Search Tool for the Retrieval of Interacting Genes/Proteins)[10] was used to predict protein–protein interaction (PPI) pairs. In addition, further validation was performed via a meta-analysis of data in the TCGA database to identify the association between *CLDN7* and integrin/FAK pathway-related genes.[11]

All expression data for COXPRESdb are based on Affymetrix GeneChips (Affymetrix Inc., Santa Clara, CA, USA) the information for which has been released by NCBI Gene Expression Omnibus. Using the default settings and entering Cldn-7, Gelsolin, MAPK1 and MAPK3 as gene symbols in the NetworkDrawer, co-expressed gene networks were drawn according to the clustering of pathways from the KEGG database. The similarity in the expression between *CLDN7* and all other genes in the database was calculated using Pearson's correlation coefficients; these values were then ranked. In addition, we extracted genes involved in the integrin/FAK signaling pathway from our microarray and used the WGCNA package to construct a weighted

gene co-expression network between CLDN7 and the integrin/FAK pathway. The analysis results were visualized using Cytoscape3.3.0 (The Cytoscape Collaboration, National Institute of General Medical Sciences (NIGMS), Bethesda, MD, USA).
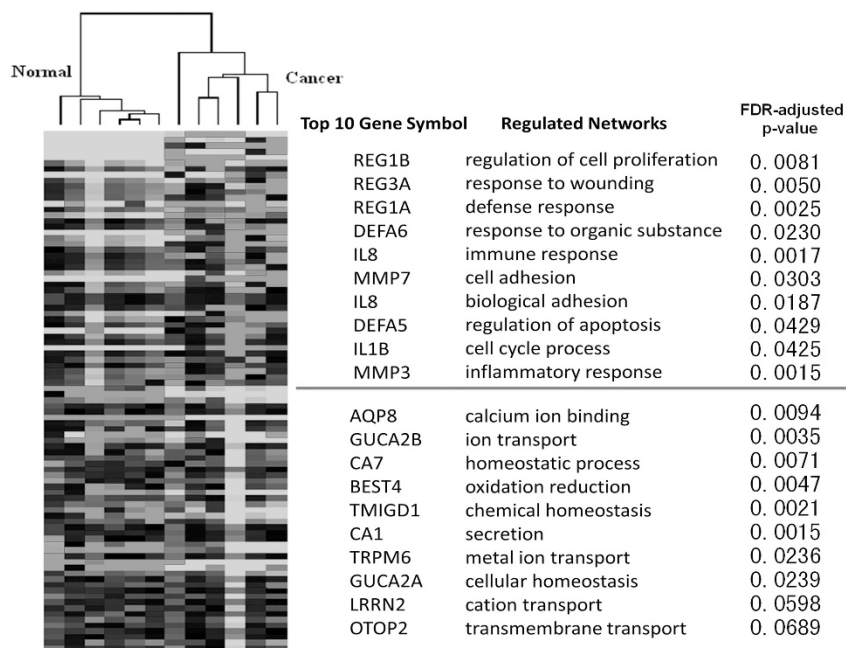
The STRING (http://string-db.org/) database is another database of known and predicted protein interactions that provides information regarding both experimental and predicted protein interactions; the database currently covers 9 643 763 proteins from 2031 organisms. The interactions include direct physical and indirect functional associations.[12] To examine protein interactions between Cldn-7 and components of the integrin/FAK signaling pathway in colon tumorigenesis, genes were mapped onto the STRING database to build PPI pairs. PPI pairs with reliability scores higher than 0.4 were screened to construct a PPI network of Cldn-7 and integrin/FAK signaling pathway components.

The TCGA database is freely and readily accessible.[13] It contained 33 cancer types and subtypes with matched tumor and normal tissues from 11 000 patients, including 10 rare cancers. We downloaded gene expression data on 262 colon cancer samples from the TCGA database (sample sizes of the data sets: batch 28 = 155; batch 29 = 7; batch 30 = 15; batch 36 = 33; batch 41 = 48; and batch 45 = 4) and performed a meta-analysis of the correlation coefficients using a DerSimonian-Laird approach in the metacor package of R.[14] The mean correlation (r) values, 95% confidence intervals (CIs) and P-values were calculated. To decrease the risk for bias in the meta-analysis, gene expression data generated using the same platform (Agilent G4502A-07, Agilent Technologies, Santa Clara, CA, USA) were selected. Pearson's correlation coefficients between each pair of genes was performed using SPSS 22. The alpha-level for the main test and for the CIs was 0.05. Significant correlations had an upper or lower limit of the 95% CI of more than 0.1 or less than − 0.1.

RESULTS

Differential gene expression patterns between CRCous tissues and matched para-cancerous normal tissues

To identify changes in gene expression in human colon carcinoma tissue, the expression of 18 401 human genes was profiled. Genes displaying an FDR-adjusted P-value of < 0.05 and a foldchange in the

Top 10 Gene Symbol	Regulated Networks	FDR-adjusted p-value
REG1B	regulation of cell proliferation	0. 0081
REG3A	response to wounding	0. 0050
REG1A	defense response	0. 0025
DEFA6	response to organic substance	0. 0230
IL8	immune response	0. 0017
MMP7	cell adhesion	0. 0303
IL8	biological adhesion	0. 0187
DEFA5	regulation of apoptosis	0. 0429
IL1B	cell cycle process	0. 0425
MMP3	inflammatory response	0. 0015
AQP8	calcium ion binding	0. 0094
GUCA2B	ion transport	0. 0035
CA7	homeostatic process	0. 0071
BEST4	oxidation reduction	0. 0047
TMIGD1	chemical homeostasis	0. 0021
CA1	secretion	0. 0015
TRPM6	metal ion transport	0. 0236
GUCA2A	cellular homeostasis	0. 0239
LRRN2	cation transport	0. 0598
OTOP2	transmembrane transport	0. 0689

Figure 1 Hierarchical clustering and Gene Ontology (GO) enrichment analyses of differentially expressed genes between colon cancer and para-cancerous normal tissues. The heatmap was generated using the top 10 up- or downregulated genes obtained from GenomeStudio. The top portion of the heatmap and pathways corresponds to upregulated genes, and the bottom portion corresponds to downregulation genes. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

expression of >1.5 in cancerous tissues compared with matched normal tissues were considered to be DEGs. A total of 790 genes were upregulated and 181 genes downregulated in the cancer tissues. The top 10 significantly upregulated or downregulated genes ($P<0.001$) are listed in Figure 1 and Supplementary Table S3. Gene Ontology and KEGG analyses identified several regulated networks of upregulated or downregulated genes, including (1) regulation of cell proliferation, (2) response to wounding, (3) defense response and (4) cell adhesion.

Focusing on functional changes with regard to cell adhesion during carcinogenesis, we chose to delineate associations between Cldn-7 and integrin/FAK pathway members in human colon carcinoma. The mRNA levels of *CLDN7* and candidate genes associated with the integrin/FAK pathway based on our expression microarray data were analyzed (Table 1). *CLDN7* mRNA showed no change in cancer tissue (foldchange = 0.63, $P=0.14$). Among 18 selected integrin/FAK pathway-associated genes, *GSN* (foldchange = 0.5, $P=0.031$) was differentially downregulated, whereas *DIAPH1* and *MYL2* were differentially upregulated. The other 15 genes displayed a foldchange >1 but with $P>0.05$. Furthermore, the mRNA levels of *CLDN7*, *ITGB1BP1*, *PTK2*, *MAPK1*, *PTK2B*, *DIAPH1*, *GSN*, *RHOA*, *MYLK*, *TLN1* and *MAPK3* were validated by qRT-PCR (Figure 2, Supplementary Table S4), showing a highly significant 3.8-fold reduction in *GSN* mRNA, a 1.8-fold reduction in *MAPK3* mRNA and a 1.5-fold reduction in *MAPK1* mRNA in the carcinoma tissues compared with the para-cancerous normal tissues. Although these changes in gene expression are accordance with the microarray results, the observed 2.3-fold reduction in *PTK2B* mRNA contrasts with the microarray results. In addition, the *CLDN7* mRNA level tended to be reduced in the carcinoma tissues (1.48 ± 0.78) compared with the para-cancerous normal tissues (1.84 ± 0.70; $P>0.05$).

Expression and localization of Cldn-7 protein and selection of candidate proteins in CRC samples

To verify that the protein levels changed in accordance with the mRNA levels and to investigate the significance of Cldn-7 and associated proteins in the integrin/FAK pathway during multi-stage colon carcinogenesis, immunohistochemical staining was performed on a multi-stage colon cancer tissue array using antibodies against Cldn-7, MYLK, GSN, MAPK1/3, PTK2 and TLN1.

Tissue sections stained with hematoxylin and eosin were divided into para-cancerous normal mucosa, stage 1, stage 2, stage 3 and stage 4 samples. The expression of three proteins, Cldn-7, GSN and MAPK1/3, clearly correlated with the advanced colon cancer stage (Figure 3), with expression decreasing in the colon cancer tissues at lower stages. Cldn-7 expression was detected around the cell membrane of the basal crypt region and the superficial crypt region in para-cancerous normal colorectal epithelia. Cldn-7 immunoreactivity gradually decreased as the cancer stage increased, as follows: para-cancerous normal, 0.22 ± 0.08 ($n=90$); stage 1, 0.19 ± 0.06 ($n=17$); stage 2, 0.13 ± 0.07 ($n=23$, vs para-cancerous normal, $P<0.01$); stage 3, 0.13 ± 0.07 ($n=35$, vs para-cancerous normal, $P<0.01$); and stage 4, 0.08 ± 0.09 ($n=15$, vs para-cancerous normal, $P<0.01$; Table 2). These results are in agreement with previous reports.[15,16] Gelsolin expression was detectable in the cytoplasm and the cytoskeleton of all colon samples examined, but no nuclear labeling was detected (Figure 3). Gelsolin expression was decreased in the cancerous tissues compared with the para-cancerous normal tissues: para-cancerous normal, 0.80 ± 0.08 ($n=90$); stage 1, 0.75 ± 0.09 ($n=17$); stage 2, 0.79 ± 0.09 ($n=23$, vs para-cancerous normal, $P<0.01$); stage 3, 0.73 ± 0.13 ($n=35$, vs para-cancerous normal, $P<0.01$); and stage 4, 0.78 ± 0.15 ($n=15$,

Table 1 Candidate genes (upregulated or downregulated genes) associated with the integrin/FAK pathway

Accession no.	Gene symbol	Chromosome	Definition	Foldchange in expression	FDR-adjusted P-value
NM_004763.2	ITGB1BP1	2	*Homo sapiens* integrin beta-1-binding protein 1 (ITGB1BP1), transcript variant 1, mRNA	1.5	0.083
NM_012278.1	ITGB1BP2	X	*Homo sapiens* integrin beta-1-binding protein 2 (melusin) (ITGB1BP2), mRNA	1.16	0.053
NM_014446.1	ITGB1BP3	19	*Homo sapiens* integrin beta-1-binding protein 3 (ITGB1BP3), mRNA	1.11	0.081
XM_942347.1	LOC652722		PREDICTED: *Homo sapiens* similar to protein tyrosine kinase 2 (PTK2) isoform a (LOC652722), mRNA	1.09	0.076
NM_153831.2	PTK2	8	*Homo sapiens* PTK2 (PTK2), transcript variant 2, mRNA	1.3	0.061
NM_173176.1	PTK2B	8	*Homo sapiens* PTK2 beta (PTK2B), transcript variant 3, mRNA	1.23	0.072
NM_006289.2	TLN1	9	*Homo sapiens* Talin-1 (TLN1), mRNA	1.02	0.069
NM_005219.3	DIAPH1	5	*Homo sapiens* diaphanous homolog 1 (*Drosophila*) (DIAPH1), transcript variant 1, mRNA	1.3	0.042
NM_002037.3	FYN	6	*Homo sapiens* FYN oncogene related to SRC, FGR, YES (FYN), transcript variant 2, mRNA	1.08	0.068
NM_198252.1	GSN	9	*Homo sapiens* Gelsolin (amyloidosis, Finnish type) (GSN), transcript variant 2, mRNA	0.5	0.031
NM_138957.2	MAPK1	22	*Homo sapiens* mitogen-activated protein kinase 1 (MAPK1), transcript variant 2, mRNA	1.05	0.057
NM_002746.1	MAPK3		*Homo sapiens* mitogen-activated protein kinase 3 (MAPK3), mRNA	0.63	0.024
NM_001040056.1	MAPK3	16	*Homo sapiens* MAPK3, transcript variant 2, mRNA	0.67	0.051
NM_000432.1	MYL2	12	*Homo sapiens* myosin, light polypeptide 2, regulatory, cardiac, slow (MYL2), mRNA	1.03	0.047
NM_005965.3	MYLK	3	*Homo sapiens* myosin light chain kinase (MYLK), transcript variant 8, mRNA	1.11	0.067
NM_002859.1	PXN	12	*Homo sapiens* paxillin (PXN), mRNA	1.04	0.089
NM_001664.2	RHOA	3	*Homo sapiens* Ras homolog gene family, member A (RHOA), mRNA	1.2	0.082
NM_001173.2	ARHGAP5	14	*Homo sapiens* Rho GTPase-activating protein 5 (ARHGAP5), transcript variant 2, mRNA	1.05	0.072

Abbreviation: FDR, false discovery rate.
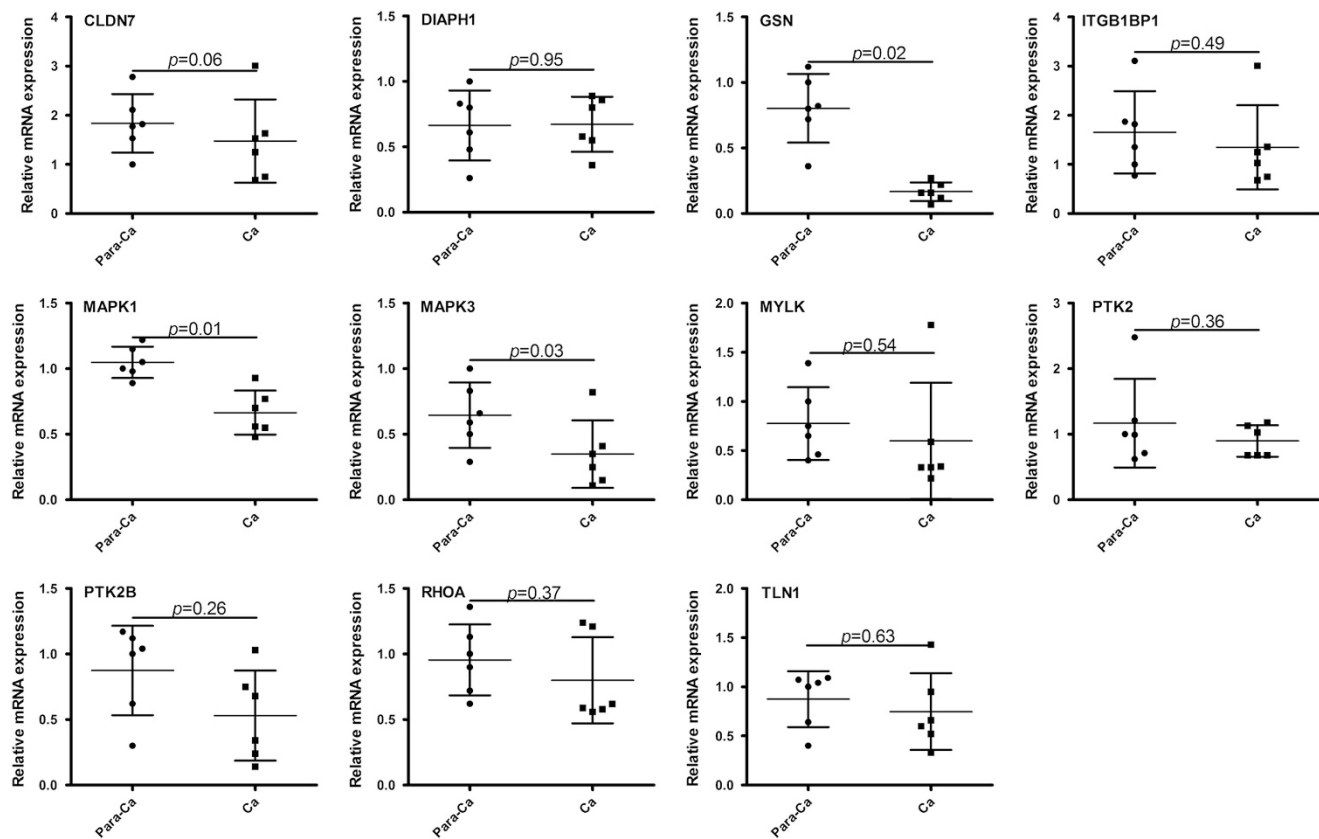
Figure 2 Claudin-7 mRNA levels and candidate genes associated with the integrin/FAK pathway in colorectal cancer tissue, as determined by quantitative real-time RT-PCR. Samples from para-cancerous normal tissues (filled circle) and cancerous tissues (filled square) were analyzed for *CLDN7, ITGB1BP1, PTK2, MAPK1, PTK2B, DIAPH1, GSN, RHOA, MYLK, TLN1* and *MAPK3* mRNA levels relative to *ACTB* or *GAPDH* mRNA levels. The horizontal line represents the mean values. *P*-values were calculated using paired Student's *t*-tests. Ca, cancerous tissues; para-Ca, para-cancerous normal tissues. $n = 6$.

vs para-cancerous normal, $P < 0.01$; Table 2). Semiquantitative analysis of the nuclear expression levels of MAPK1/3 revealed the following: para-cancerous normal, 0.07 ± 0.03; stage 1, 0.04 ± 0.03; stage 2, 0.02 ± 0.02; stage 3, 0.04 ± 0.04; and stage 4, 0.03 ± 0.02. Compared with the para-cancerous normal tissue, all colon cancer tissue stages displayed a *P*-value of < 0.01. The changes in the levels of these three proteins were in accordance with their respective changes at the mRNA level.

The expression of the remaining three proteins, Talin, MYLK and PTK2, did not correlate with the advanced stage of colon cancer (Figure 3). Talin expression was primarily localized to the cytoplasm, with no significant change at either the protein or mRNA level between colon cancer tissues at different stages and para-cancerous normal tissues. MYLK showed cytoplasmic staining, and its protein expression was increased in stage 1 cancerous tissue ($P < 0.05$), in contrast to its mRNA expression. PTK2 was detected at the cell membrane, and its expression at both the mRNA and protein levels displayed a decreasing trend in advanced stage colon cancer tissue but a significant increase ($P < 0.01$) in early-stage colon cancer tissue (Table 2, Figure 3).

***In silico* analysis of genes and pathways co-expressed with Cldn-7, GSN, MAPK1 and MAPK3 in CRC tissue**
To search for a network of co-expressed genes including Cldn-7, GSN, MAPK1 and MAPK3, an *in silico* analysis was performed using the COXPRESdb version 6.0 database.[9] We first investigated genes directly

associated with *CLDN7* with respect to their network and their KEGG pathways (Figure 4a). Many of these genes were found to be directly connected with *CLDN7*, including *SPINT1*, which is involved in the regulation of epithelial integrity and immunity. However, we did not find a direct interaction between *CLDN7* and *GSN, MAPK1* or *MAPK3* in the integrin/FAK pathway (Figure 4b). We identified the principal roles of MAPK1 and MAPK3 in the regulation of cancer progression and cell adhesion pathways (Figure 4c). In addition, we utilized the STRING database to construct a PPI network (Supplementary Figure S1). *CLDN7* and DEGs of the integrin/FAK/ERK signaling pathway were integrated into the network. The nodes represent the genes and the lines the corresponding PPI pairs. However, *CLDN7* failed to form a PPI pairs with any one gene from the integrin/FAK/ERK signaling pathway. To verify this prediction, we used a Pearson's correlation to evaluate co-expression between Cldn-7 and integrin/FAK/ERK signaling pathway genes from our microarray data. Significant correlations were found between *CLDN7* and *PTK2* ($P = 0.020$) and between *DIAPH1* and *RHOA* ($P = 0.002$), whereas no correlations were observed for the other pairs (Supplementary Table S5). Mammalian Diaph1 is a downstream effector of the Rho family that mediates Rho action at the actin cytoskeleton;[17,18] therefore, *DIAPH1* and *RHOA* are co-expressed. We also used the WGCNA package, a comprehensive collection of R functions for performing various aspects of weighted correlation network analysis (version 1.47 can be used with Windows R-3.1.3), to analyze *CLDN7* and genes of the integrin/FAK pathway identified by our microarray data, but no
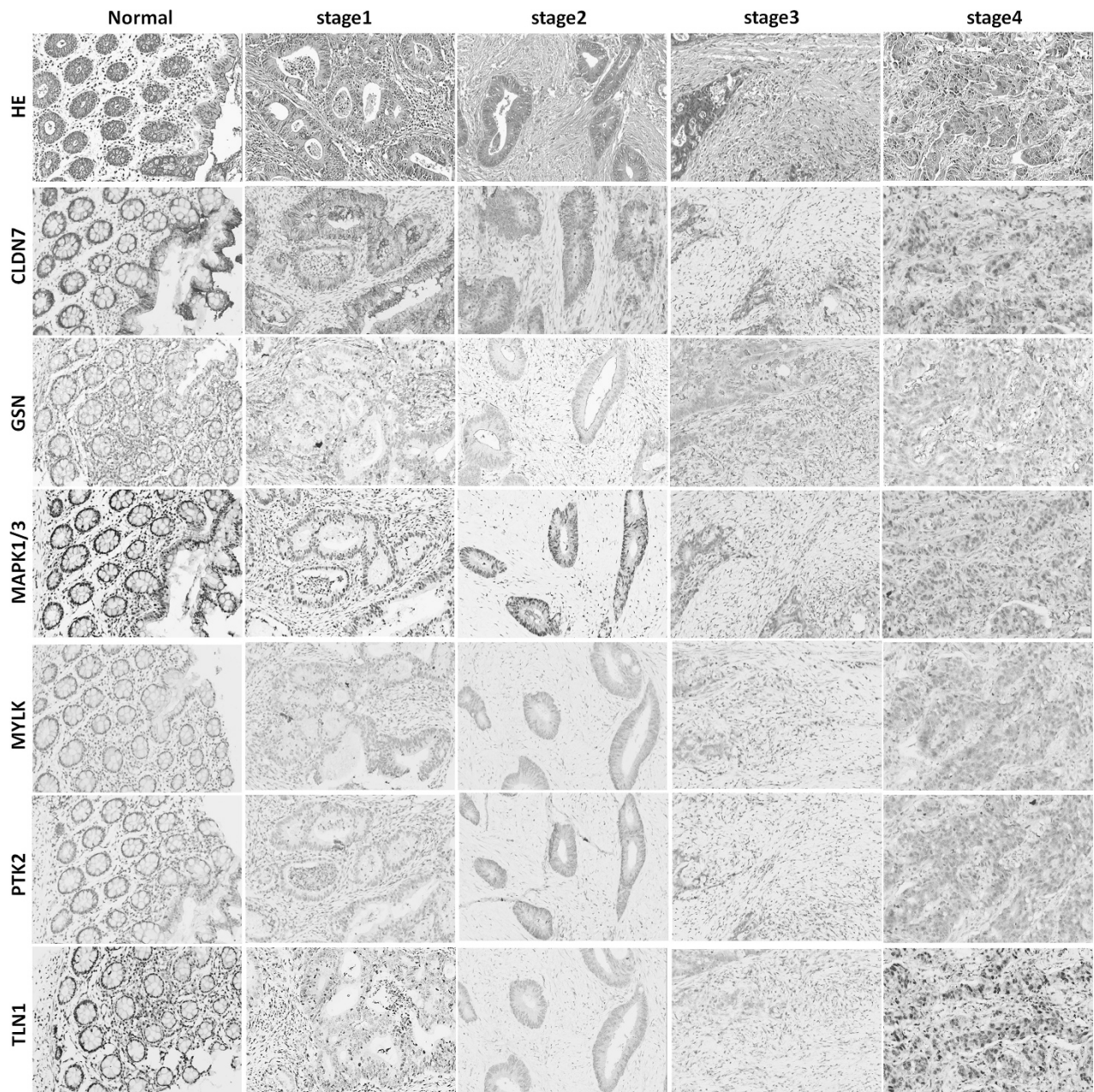
Figure 3 Immunohistochemical staining for claudin-7, Gelsolin, MAPK1/3, Talin, FAK and MYLK in different-stage colorectal carcinomas and adjacent normal tissue. The first row shows a tissue section stained with hematoxylin and eosin. Immunostaining for claudin-7, gelsolin and MAPK1/3 in samples of different histological grades. Images were obtained using an Aperio ScanScope. Magnification, ×100. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

co-expression between *CLDN7* and these genes was found (Supplementary Figure S2). We consider the discrepancies to be caused by the small sample number. Thus, further validation was performed in a larger cohort of patients. Meta-analysis of correlation coefficients was conducted using gene expression data from 262 colon cancer samples from the TCGA database. The meta-analysis approach identified significant correlations between the following gene pairs: *CLDN7* and *KRT8* (keratin 8; $r=0.311$; 95% CI, 0.194, 0.420), *CLDN7* and *ST14* ($r=0.333$; 95% CI, 0.217, 0.440), *MAPK3* and *KRT8* ($r=0.280$; 95% CI, 0.161, 0.391), and *MAPK3* and *RAB25* ($r=0.291$; 95% CI, 0.151, 0.420; Table 3). Although the *P*-values were <0.05, the lower or upper limits

of the 95% CI were near zero, suggesting no correlations between Cldn-7 and genes in the integrin/FAK/ERK signaling pathway: *CLDN7* and *PTK2* ($r=-0.212$; 95% CI, -0.328, 0.090), *CLDN7* and *PTK2B* ($r=0.0194$; 95% CI, 0.071, 0.311), *CLDN7* and *FYN* ($r=-0.153$; 95% CI, -0.272, -0.029), *CLDN7* and *GSN* ($r=0.233$; 95% CI, -0.060, 0.392), *CLDN7* and *MAPK3* ($r=0.186$; 95% CI, 0.001, 0.358) and *CLDN7* and *PXN* ($r=0.150$; 95% CI, -0.004, 0.297; Table 3).

DISCUSSION
On the basis of our microarray data results, DEGs between colorectal cancerous tissues and matched para-cancerous normal tissues were

Table 2 Statistical evaluation of the expression of Cldn-7 and related proteins in colon tumors via immunohistochemical analysis

	Quantitative analysis results (OD: mean ± s.d.)				
	Para-cancerous normal	Stage 1 cancer	Stage 2 cancer	Stage 3 cancer	Stage 4 cancer
Proteins name	n = 90	n = 17	n = 23	n = 35	n = 15
Cldn-7	0.22 ± 0.08	0.19 ± 0.06	0.13 ± 0.07**	0.13 ± 0.07**	0.08 ± 0.09**
MAPK1/3	0.07 ± 0.03	0.04 ± 0.03**	0.02 ± 0.02**	0.04 ± 0.04**	0.03 ± 0.02**
GSN	0.80 ± 0.09	0.75 ± 0.09*	0.79 ± 0.09	0.73 ± 0.13**	0.78 ± 0.15
MYLK	0.34 ± 0.16	0.46 ± 0.20*	0.37 ± 0.14	0.31 ± 0.14	0.33 ± 0.17
PTK2	0.23 ± 0.14	0.38 ± 0.21**	0.22 ± 0.12	0.16 ± 0.13	0.18 ± 0.10
TLN1	0.38 ± 0.15	0.46 ± 0.16	0.43 ± 0.15	0.38 ± 0.14	0.38 ± 0.19

Abbreviations: Cldn-7, claudin-7; OD, optical density.
*$P < 0.05$; **$P < 0.01$, vs para-cancerous normal tissue.

identified using thresholds (foldchange in expression of > 1.5 and an FDR-adjusted P-value < 0.05). The top 10 upregulated networks are primarily involved in cell proliferation and apoptosis, inflammation and immunity and cell adhesion. The crucial mechanism underlying colon cancer progression is the weakening of the endothelial barrier function in response to injury, and many signaling events can disrupt the intestinal barrier. Focal adhesions supported by the integrin/FAK pathway are the first barrier maintaining endothelial integrity.[3] Endothelial FAK, an immediate downstream target of integrin, is activated through complex intracellular cascades, leading to conformational changes in the endothelial cytoskeleton and in cell junctions.[19] Cldn-7 serves as another putative barrier based on our previous finding that Cldn-7 knockout mice have severe intestinal defects.[8] The claudin family, which are the major structural and functional components of tight junctions, in mammals includes at least 24 members; in particular, Cldn-7 is highly expressed in the basolateral membrane of the intestinal epithelia.[20] Another previous study revealed that Cldn-7 interacts with integrin α2 in the basolateral compartment of the intestinal epithelia.[21] Taken together, our findings demonstrated a putative association between Cldn-7 and the integrin/FAK pathway.
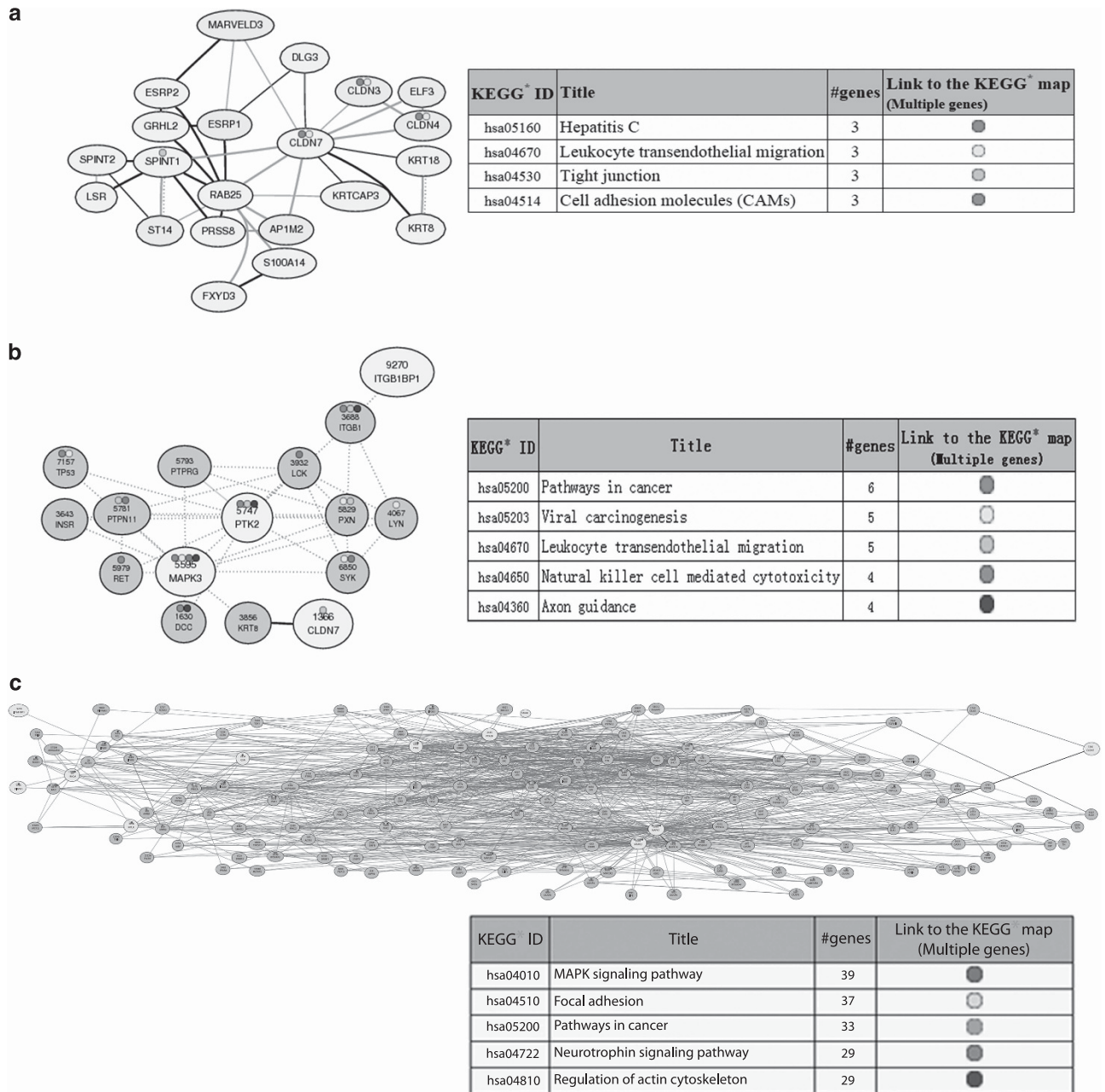
Our study demonstrated a gradual decrease in the levels of Cldn-7 from colon adenocarcinoma stages 1–4; however, no significant decrease in the level of *CLDN7* mRNA was observed by qRT-PCR. We thoroughly analyzed paired tissues from six patients. Interestingly, the mRNA level of *CLDN7* in a 64-year-old female patient with poorly differentiated colon cancer showed a slight increase between the cancerous and para-cancerous normal tissues. However, the levels of *CLDN7* mRNA in the samples from the other patients with well differentiated or moderately differentiated colon cancer displayed a significant downregulation (foldchange = 0.57, $P = 0.020$) in colon cancer tissues. This result agrees with the finding that the *CLDN7* mRNA level is decreased as an early event in colorectal carcinogenesis.[15] Our meta-analysis of correlation coefficients showed that *CLDN7* was significantly co-expressed with *ST14* and *KRT8*. Interestingly, co-expression of *CLDN7* and *ST14* (encoding matriptase) has been reported in colorectal carcinogenesis.[15] However, there are no reports on the association between *CLDN7* and *KRT8* (a type of intermediate filament).

On the basis of our microarray data, we selected up- and downregulated genes associated with the integrin/FAK pathway according to the KEGG database. On the basis of qRT-PCR, none of the genes upregulated according to the microarray results exhibited a significant change between colon cancer and para-cancerous normal

tissues. Alternatively, the levels of *Gelsolin*, *MAPK1* and *MAPK3* mRNA were confirmed to be significantly downregulated in the cancer tissues by qRT-PCR. According to the immunohistochemistry results from our tissue microarray analysis of 180 colon adenocarcinoma samples, we detected a decrease in the Gelsolin protein level beginning from an early colon cancer stage. Gelsolin is a well-known actin-binding protein involved in the interaction between the extracellular matrix and integrins via the regulation of integrin clustering.[22] Previous studies have reported that the expression of Gelsolin protein and mRNA is severely downregulated, with data demonstrating that Gelsolin downregulation is an early and nearly universal event in colon carcinogenesis and is associated with the transition from adenoma to carcinoma.[23,24] In line with our observations, the cytoplasmic distribution of Gelsolin gradually decreased from stage 1 to stage 4 in colon adenocarcinoma.

MAPK1 and MAPK2 are also referred to as ERK2 and ERK1, respectively. Interactions between integrin β1 and the collagen matrix regulate cell survival by associating with FAK and initiating MAPK/ ERK signaling.[19] MAPK signaling pathway members have been determined to be among the gene markers most strongly associated with CRC,[25] and seven MAPK genes have been identified as important for CRC.[26] In our study, we found that MAPK1 and MAPK3 were downregulated at the mRNA and protein levels in colon cancer in a stage-specific manner. Our meta-analysis of correlation coefficients showed that *MAPK3* significantly correlated with *RAB25*, which belongs to the Rab family of small GTPases, and with *KRT8*. The functional association with the ERK signaling pathway that mediates the effects of RAB25 on autophagy in ovarian cancer cells has been validated.[27] The association between ERK and KRT8 phosphorylation may have a role in the induction of cancer cell apoptosis by an acetone extract of Bupleurum scorzonerifolium.[28]

In addition, we evaluated FAK (also referred to as PTK2), MYLK and Talin-1 via qRT-PCR and immunohistochemistry. The mRNA and protein levels of FAK showed a tendency toward downregulation, but these results were not statistically significant, and there were no significant differences in MYLK or Talin-1 expression at the mRNA or protein level. However, there is some controversy concerning the reported findings for Talin-1. A previous report showed that the preoperative serum levels of Talin-1 were significantly increased in patients with colon cancer.[29] Talin-1 is a cytoskeletal protein, and its head region promotes the binding of integrin to the extracellular matrix. It has been demonstrated that Talin-1 expression increases cell adhesion, migration and invasion in prostate cancer.[30] Thus, further

a

KEGG* ID	Title	#genes	Link to the KEGG* map (Multiple genes)
hsa05160	Hepatitis C	3	●
hsa04670	Leukocyte transendothelial migration	3	○
hsa04530	Tight junction	3	◐
hsa04514	Cell adhesion molecules (CAMs)	3	●

b

KEGG* ID	Title	#genes	Link to the KEGG* map (Multiple genes)
hsa05200	Pathways in cancer	6	●
hsa05203	Viral carcinogenesis	5	○
hsa04670	Leukocyte transendothelial migration	5	○
hsa04650	Natural killer cell mediated cytotoxicity	4	●
hsa04360	Axon guidance	4	●

c

KEGG* ID	Title	#genes	Link to the KEGG* map (Multiple genes)
hsa04010	MAPK signaling pathway	39	●
hsa04510	Focal adhesion	37	○
hsa05200	Pathways in cancer	33	●
hsa04722	Neurotrophin signaling pathway	29	●
hsa04810	Regulation of actin cytoskeleton	29	●

Figure 4 *In silico* analysis showing the claudin-7 co-expressed gene network. The figure shows an image of the analysis performed using the COXPRESdb version 6.0 http://coxpresdb.jp/. Each gene is represented by a node. The lines between nodes indicate co-expression. The KEGG pathway is shown under the network. (**a**) *In silico* analysis showing genes directly associated with *CLDN7* with respect to their network and their KEGG pathways. (**b**) *In silico* analysis showing the claudin-7, integrin, MAPK3 and PTK2 co-expressed gene network. (**c**) *In silico* analysis showing the claudin-7 and integrin/FAK/MAPK pathway member co-expressed gene network. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

studies of Talin-1 are needed to identify the difference in the expression between serum and tissue samples.

We performed a bioinformatic and statistical analysis to evaluate the interaction between Cldn-7 and components of the integrin/FAK pathway. Furthermore, we conducted a meta-analysis as validation. The *in silico* prediction and experimental analysis indicated that Cldn-7 does not directly regulate any of the examined integrin/FAK pathway genes. Nonetheless, we cannot rule out the possibility that Cldn-7 indirectly associates with the integrin/FAK pathway to co-regulate cancer progression, cell adhesion and neurotrophin

signaling pathways (Figure 4). This result was verified using PPI data from the STRING database, Pearson's correlation analysis and WGCNA package analysis of our microarray data. The meta-analysis based on a large sample cohort showed that *CLDN7* may be co-expressed with *KRT8*, which may be co-expressed with *MAPK3*. These data suggested that *CLDN7* may be indirectly associated with *MAPK3* via *KRT8*. The MAPK pathway is the central node of this network, and our previous work demonstrated that Cldn-7 inhibits lung cancer cell migration and invasion via the ERK/MAPK signaling pathway in response to growth factor stimulation.[21] Further

Table 3 Meta-analysis of the correlations between *CLDN7* and integrin/FAK signaling pathway genes in TCGA colon cancer microarrays ($n=262$)

Co-expressed gene pairs	Correlation coefficients in different vectors						Mean of correlation coefficients	95% confidence interval (CI)	P-value	Evaluation of the correlation
	Batch 28 (n=155)	Batch 29 (n=7)	Batch 30 (n=15)	Batch 36 (n=33)	Batch 41 (n=48)	Batch 45 (n=4)				
CLDN7-PTK2	−0.171	−0.132	−0.187	−0.352	−0.258	−0.550	−0.212	−0.328 to −0.090	0.000	×
CLDN7-ITGB1BP1	−0.193	0.662	−0.503	0.092	−0.108	0.019	−0.112	−0.304 to −0.089	0.137	×
CLDN7-ITGB1BP2	−0.126	−0.698	0.329	−0.057	−0.144	−0.672	−0.110	0.032 to −0.248	0.064	×
CLDN7-ITGB1BP3	0.055	−0.203	0.355	0.049	0.182	−0.227	0.088	−0.037 to 0.210	0.084	×
CLDN7-PTK2B	0.147	0.059	0.056	0.142	0.409	0.538	0.194	0.071 to 0.311	0.001	×
CLDN7-TLN1	0.010	−0.502	0.152	0.319	−0.139	−0.045	0.023	−0.126 to 0.170	0.382	×
CLDN7-DIAPH1	0.079	0.104	−0.213	0.308	−0.285	−0.177	0.000	−0.210 to 0.210	0.499	×
CLDN7-FYN	−0.202	−0.540	0.042	0.055	−0.116	−0.738	−0.153	−0.272 to −0.029	0.008	×
CLDN7-GSN	0.226	0.507	0.551	0.351	0.019	−0.797	0.233	0.060–0.392	0.004	×
CLDN7-MAPK1	−0.018	0.237	−0.133	−0.065	0.017	−0.272	−0.020	−0.144 to 0.105	0.378	×
CLDN7-MAPK3	0.085	0.238	0.093	0.499	0.077	0.859	0.186	0.001 to −0.358	0.024	×
CLDN7-MYL2	−0.098	0.296	0.330	0.159	0.157	0.043	0.017	−0.114 to 0.148	0.400	×
CLDN7-MYLK	−0.072	−0.293	0.125	0.241	−0.250	−0.060	−0.059	−0.191 to 0.075	0.195	×
CLDN7-PXN	0.126	−0.239	0.650	0.042	0.129	0.288	0.150	−0.004 to 0.297	0.028	×
CLDN7-RHOA	−0.101	0.076	−0.028	0.086	0.160	0.286	−0.022	−0.146 to 0.103	0.367	×
CLDN7-ARHGAP5	−0.051	−0.321	−0.215	0.260	−0.245	−0.615	−0.068	−0.222 to 0.090	0.199	×
CLDN7-KRT8	0.273	−0.457	0.474	0.400	0.395	0.204	0.311	0.194 to 0.420	0.000	√
CLDN7-KRT18	0.155	−0.538	0.232	0.249	0.193	0.021	0.165	0.041 to 0.284	0.005	×
CLDN7-RAB25	0.159	0.166	0.259	0.228	0.163	0.215	0.174	0.050 to 0.292	0.003	×
CLDN7-ST14	0.400	0.161	0.507	0.222	0.118	0.564	0.333	0.217 to 0.440	0.000	√
MAPK3-KRT8	0.306	0.272	−0.108	0.472	0.135	0.675	0.280	0.161 to 0.391	0.000	√
MAPK3-KRT18	0.140	0.250	−0.142	0.300	0.047	0.505	0.133	0.009 to 0.254	0.018	×
MAPK3-RAB25	0.336	0.657	0.463	−0.036	0.308	−0.282	0.291	0.151 to 0.420	0.000	√
MAPK3-ST14	0.090	−0.056	−0.196	0.296	0.214	0.663	0.126	0.001 to 0.247	0.024	×

Abbreviations: TCGA, The Cancer Genome Atlas; √, significant correlation between the two genes; ×, no correlation between the two genes.
The alpha-level for the main test and the confidence intervals is 0.05. Significant correlations had an upper or lower limit of the 95% CI of >0.1 or $<−0.1$.

investigation is needed to elucidate the regulatory relationship between Cldn-7 and ERK/MAPK pathway members and to better improve clinical diagnosis and treatment.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA Cancer J. Clin.* **63**, 11–30 (2013).
2 Ou, J., Deng, J., Wei, X., Xie, G., Zhou, R., Yu, L. *et al.* Fibronectin extra domain A (EDA) sustains CD133(+)/CD44(+) subpopulation of colorectal cancer cells. *Stem Cell Res.* **11**, 820–833 (2013).
3 Giancotti, F. G. Integrin signaling: specificity and control of cell survival and cell cycle progression. *Curr. Opin. Cell Biol.* **9**, 691–700 (1997).
4 Lock, R. & Debnath, J. Extracellular matrix regulation of autophagy. *Curr. Opin. Cell Biol.* **20**, 583–588 (2008).
5 Ha, B. G., Park, J. E., Cho, H. J., Lim, Y. B. & Shon, Y. H. Inhibitory effects of proton beam irradiation on integrin expression and signaling pathway in human colon carcinoma HT29 cells. *Int. J. Oncol.* **46**, 2621–2628 (2015).
6 Luo, C. W., Wu, C. C. & Ch'ang, H. J. Radiation sensitization of tumor cells induced by shear stress: the roles of integrins and FAK. *Biochim. Biophys. Acta* **1843**, 2129–2137 (2014).
7 Bhat, A. A., Pope, J. L., Smith, J. J., Ahmad, R., Chen, X., Washington, M. K. *et al.* Claudin-7 expression induces mesenchymal to epithelial transformation (MET) to inhibit colon tumorigenesis. *Oncogene* **34**, 4570–4580 (2014).
8 Ding, L., Lu, Z., Foreman, O., Tatum, R., Lu, Q., Renegar, R. *et al.* Inflammation and disruption of the mucosal architecture in claudin-7-deficient mice. *Gastroenterology* **142**, 305–315 (2012).
9 Obayashi, T., Hayashi, S., Shibaoka, M., Saeki, M., Ohta, H. & Kinoshita, K. COXPRESdb: a database of coexpressed gene networks in mammals. *Nucleic Acids Res.* **36**, D77–D82 (2008).
10 Rock, A., Irwin, J., Dur, A., Parsons, T. & Parson, W. SAM: string-based sequence search algorithm for mitochondrial DNA database queries. *Forensic Sci. Int. Genet.* **5**, 126–132 (2011).
11 Reddy, R. B., Bhat, A. R., James, B. L., Govindan, S. V., Mathew, R., Dr, R. *et al.* Meta-analyses of microarray datasets identifies ANO1 and FADD as prognostic markers of head and neck cancer. *PLoS ONE* **11**, e0147409 (2016).
12 Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568 (2011).
13 Kong, J., Cooper, L. A., Wang, F., Gutman, D. A., Gao, J., Chisolm, C. *et al.* Integrative, multimodal analysis of glioblastoma using TCGA molecular data, pathology images, and clinical outcomes. *IEEE Trans. Biomed. Eng.* **58**, 3469–3474 (2011).
14 Laliberte, E., Wells, J. A., Declerck, F., Metcalfe, D. J., Catterall, C. P., Queiroz, C. *et al.* Land-use intensification reduces functional redundancy and response diversity in plant communities. *Ecol. Lett.* **13**, 76–86 (2010).
15 Bornholdt, J., Friis, S., Godiksen, S., Poulsen, S. S., Santoni-Rugiu, E., Bisgaard, H. C. *et al.* The level of claudin-7 is reduced as an early event in colorectal carcinogenesis. *BMC Cancer* **11**, 65 (2011).
16 Suren, D., Yildirim, M., Kaya, V., Alikanoglu, A. S., Bulbuller, N., Yildiz, M. *et al.* Loss of tight junction proteins (Claudin 1, 4, and 7) correlates with aggressive behavior in colorectal carcinoma. *Med. Sci. Monit.* **20**, 1255–1262 (2014).
17 Sewer, M. B. & Li, D. Regulation of adrenocortical steroid hormone production by RhoA-diaphanous 1 signaling and the cytoskeleton. *Mol. Cell. Endocrinol.* **371**, 79–86 (2013).
18 Lu, Q., Lu, L., Chen, W., Chen, H., Xu, X. & Zheng, Z. RhoA/mDia-1/profilin-1 signaling targets microvascular endothelial dysfunction in diabetic retinopathy. *Graefes Arch. Clin. Exp. Ophthalmol.* **253**, 669–680 (2015).
19 Saleem, S., Li, J., Yee, S. P., Fellows, G. F., Goodyer, C. G. & Wang, R. beta1 integrin/FAK/ERK signalling pathway is essential for human fetal islet cell differentiation and survival. *J. Pathol.* **219**, 182–192 (2009).
20 Lu, Z., Ding, L., Lu, Q. & Chen, Y. H. Claudins in intestines: Distribution and functional significance in health and diseases. *Tissue Barriers* **1**, e24978 (2013).
21 Lu, Z., Ding, L., Hong, H., Hoggard, J., Lu, Q. & Chen, Y. H. Claudin-7 inhibits human lung cancer cell migration and invasion through ERK/MAPK signaling pathway. *Exp. Cell Res.* **317**, 1935–1946 (2011).
22 Robins, P., Harris, A. L., Goldsmith, I. & Lindahl, T. Cross-linking of DNA induced by chloroethylnitrosourea is presented by O^6-methylguanine-DNA methyltransferase. *Nucleic Acids Res.* **11**, 7743–7758 (1983).
23 Radwanska, A., Litwin, M., Nowak, D., Baczynska, D., Wegrowski, Y., Maquart, F. X. *et al.* Overexpression of lumican affects the migration of human colon cancer cells through up-regulation of gelsolin and filamentous actin reorganization. *Exp. Cell Res.* **318**, 2312–2323 (2012).
24 Gay, F., Estornes, Y., Saurin, J. C., Joly-Pharaboz, M. O., Friederich, E., Scoazec, J. Y. *et al.* In colon carcinogenesis, the cytoskeletal protein gelsolin is down-regulated during the transition from adenoma to carcinoma. *Hum. Pathol.* **39**, 1420–1430 (2008).
25 Urosevic, J., Nebreda, A. R. & Gomis, R. R. MAPK signaling control of colon cancer metastasis. *Cell Cycle* **13**, 2641–2642 (2014).
26 Slattery, M. L., Lundgreen, A. & Wolff, R. K. Dietary influence on MAPK-signaling pathways and risk of colon and rectal cancer. *Nutr. Cancer* **65**, 729–738 (2013).
27 Liu, Y., Tao, X., Jia, L., Cheng, K. W., Lu, Y., Yu, Y. *et al.* Knockdown of RAB25 promotes autophagy and inhibits cell growth in ovarian cancer cells. *Mol. Med. Rep.* **6**, 1006–1012 (2012).
28 Chen, Y. L., Lin, S. Z., Chang, W. L., Cheng, Y. L. & Harn, H. J. Requirement for ERK activation in acetone extract identified from Bupleurum scorzonerifolium induced A549 tumor cell apoptosis and keratin 8 phosphorylation. *Life Sci.* **76**, 2409–2420 (2005).
29 Bostanci, O., Kemik, O., Kemik, A., Battal, M., Demir, U., Purisa, S. *et al.* A novel screening test for colon cancer: Talin-1. *Eur. Rev. Med. Pharmacol. Sci.* **18**, 2533–2537 (2014).
30 Zhang, W., Mao, Y. Q., Wang, H., Yin, W. J., Zhu, S. X. & Wang, W. C. MiR-124 suppresses cell motility and adhesion by targeting talin 1 in prostate cancer cells. *Cancer Cell Int.* **15**, 49 (2015).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)