Stromal VCAN expression as a robust prognostic biomarker for disease recurrence in stage II-III colorectal cancer

Stromal VCAN expression as a robust prognostic biomarker for disease recurrence in stage II-III colorectal cancer

（StageII-III 大腸癌の予後バイオマーカーとしての間質遺伝子 VCAN 発現）

Shun Chida

Department of Organ Regulatory Surgery, Fukushima Medical University School of Medicine

概要

大腸癌（結腸直腸癌）は、全癌において死亡・罹患数が最も上位の癌種のひとつであるが、ここ１０年で大腸癌診療は飛躍的に進歩した。すなわち内視鏡治療の高度化や、低侵襲な腹腔鏡手術の標準化、また分子標的薬を含む新規薬剤などの開発により生存期間の延長やQOLの改善に寄与している。急速に進歩を遂げつつある大腸癌診療において、治療の個別化が現在臨床で求められている。今回、我々が着目するのは外科的切除の対象となる大多数の大腸癌である。本邦では治癒切除後の5年生存率は約８０%と比較的良好と言えるが、再発した場合の根治は困難である。重要なのは、根治切除後にいかに再発を回避するかであり、この役割を担うのは術後補助化学療法である。補助化学療法により無病生存期間が延長するという明確なエビデンスが蓄積している。しかしながら恩恵を受け得る患者は一部のみである。本邦ではStage Ⅱの約１５%、Stage Ⅲの約３０%が再発すると言われるが、多くの患者は手術単独で治癒する可能性がある一方で、標準的な補助療法を行っても再発を来たす一定の患者が存在するからである。また化学療法には少なくない副作用と費用が不可避であり患者QOLおよび医療経済的に大きな負担である。現在、大腸癌の治療および予後は、TNM分類に基づいてステージングされているが国際標準として術後補助化学療法はStageⅢ症例のみが対象となりStageⅡ症例については議論の余地がある。現在求められるのは、術後補助化学療法が必要な再発リスクの高い症例を識別する新規の治療個別化バイオマーカーの開発である。

新規バイオマーカーの探求の発端に我々はNCBI(National Center for Biotechnology Information)が提供するマイクロアレイなどの網羅的遺伝子発現情報の公開データベースであるGEO(Gene Expression Omnibus)を用いた。個々のデータセットには無再発生存期間などの予後情報が付属している。まず２つの公開マイクロアレイデータセット(GSE41258, n=94,stageⅡ-Ⅲ)(GSE17538,n=145,stageⅡ-Ⅲ)を用いた。前者には正常大腸(n=54)と大腸癌(n=186)の遺伝子発現情報も含まれている。大腸癌特異的に発現する遺伝子かつ、２つの独立したデータセットにおいて大腸癌の無再発生存期間と関連するものを分析し８つの遺伝子を同定した。幾つかの遺伝子は主に間質で発現するとされるものであり、その局在を明らかにするためにレーザーマイクロダイセクション法にて大腸癌の上皮成分と間質成分を個別に解析した公開マイクロアレイデータセット(GSE35602)を用いた。８つのうち６遺伝子は大腸癌において癌上皮より癌間質に有意に高発現しており、さらにそのうち４遺伝子(COL4A2,COL4A1,VCAN,SERPINE1)は癌間質において正常粘膜・正常間質と比較しても有意に高発現しており、癌間質特異的遺伝子と考えられた。この組織レベルでの解析を細胞レベルにおいて検証するためにフローサイトメトリー法にて大腸癌の腫瘍関連細胞や上皮細胞、血管内皮細胞、血球細胞を個々に分析した公開マイクロアレイデータセット(GSE39396)を用い発現を解析した結果、VCANは腫瘍関連線維芽細胞に特異的に発現を認め、SERPINE1は主に血管内皮細胞での特異的な発現を認めた。近年、癌の進展において癌細胞だけでなく癌間質が極めて重要な役割を果たすことが報告されている。なかでも癌間質

における線維芽細胞は癌関連線維芽細胞(Cancer associated fibroblast,CAF)と呼ばれ癌の浸潤や転移、血管新生を促進する癌微小環境として注目されている。先述のVCANはコンドロイチン硫酸プロテオグリカンの１つであり、子宮癌や卵巣癌間質において癌細胞の増殖および転移に関連する可能性が報告されている。我々は新規の着眼点として、大腸癌間質遺伝子の一つVCANが予後予測バイオマーカーになり得るか検討した。

当科で手術を受けた３３８例の大腸癌患者から構成されたデータセットを用い大腸癌間質におけるVCAN発現を免疫染色法により予後との関連を検討した。染色強度をスコア化し２分化したところ染色陽性のものは全Stageにおいても、StageⅡ･Ⅲにおいても統計学的に有意に無再発生存期間の低下を認めた。さらに、結腸のみに限定した場合、StageⅡ･Ⅲ(P=0.0003)においてもStageⅡのみ(P=0.0029)においても非常に有意な無再発生存期間の低下を認めた。一方、直腸のみに限定した場合は何れのケースでも有意差は認めなかった。また臨床病理学的特徴を検討したところ、ステージ分類、T分類、N分類、M分類はVCAN発現に伴い無再発生存期間を有意に低下させた。またStageⅡ･Ⅲ症例での多変量解析でも無再発生存期間に関して、リンパ節転移の有無（HR2.37,P=0.009）に続いてVCAN発現（HR2.37,P=0.022）が統計学的有意差のある予後因子として示された。さらに結腸のみに限ったStageⅡ･Ⅲ症例での多変量解析では無再発生存期間に関して、VCAN発現（HR8.98,P=0.004）及びリンパ節転移の有無（HR3.58,P=0.008）が統計学的有意差のある予後因子として示された。またStageⅡ･Ⅲ大腸癌４５３例とStageⅡ大腸癌８９例によるそれぞれ独立した公開マイクロアレイデータセットを用いて検証を行ったところ、どちらもVCAN陽性が無再発生存期間を有意に低下させた（P=0.0334，P=0.0041）。

大腸癌の、特に結腸癌間質におけるVCAN発現が予後予測バイオマーカーになり得る可能性が示唆された。

Introduction

Colorectal cancer (CRC) is one of the leading cause of cancer-related death worldwide (1-3). Its heterogeneous nature associated with multiple molecular mechanisms, results in diverse clinical behaviors. Surgery with curative intent is the mainstay of treatment for locally advanced disease. However, 25~40% of patients with stage II-III CRC will develop tumor recurrence despite potential curative resection. This high rate of relapse provides the rationale for postoperative adjuvant chemotherapy. The TNM classification remains the most reliable prognostic indicator that is used to stratify patients who would benefit from adjuvant chemotherapy (3, 4). Although adjuvant chemotherapy can improve survival in patients with Stage III CRC, supported by large randomized studies, the routine use of adjuvant chemotherapy is not recommended for stage II patients (3, 5, 6). It is therefore critical to identify patients with stage II-III CRC at high risk of recurrence who may benefit from adjuvant chemotherapy as well as high-intensity postoperative surveillance plans. Conversely, 60-80% of patients with stage II-III disease who are cured by surgery alone could be spared from intensive postoperative management. In this respect, molecular features reflecting tumor characteristics with heterogeneous outcomes could help in optimizing patient selection for adjuvant chemotherapy. Biomarkers are needed to identify patients who are at risk of developing recurrence after curative surgery for precise personalized management.

Many studies have exploited microarray-based gene expression profiling to subclassify

tumor entities and to predict prognosis of CRC (7-11). However, none of the previously reported prognostic gene signatures has been applied in the current clinical settings (12, 13). This is possibly due to the low reproducibility of those signatures. Also, potential differences in data processing, RNA preparation and handling methods and various expression platforms may affect discrepant results. It is also surprising that there is virtually no overlap of genes between previously published gene signatures. In view of that, the present study initially attempted to identify a number of genes whose expression levels are predictive of disease recurrence with high reproducibility across multiple cohorts. Also, we consider that the use of formalin-fixed paraffin-embedded (FFPE) tissue samples could extend the practical utility of biomarkers in routine clinical diagnostics. To this end, we evaluated two publicly available microarray cohorts of stage II-III CRC to identify genes with associations with recurrence, followed by testing the robustness of genes using two additional independent datasets. The final goal of this study was to test whether the prognostic impact of mRNA transcript levels identified by comprehensive analyses of multiple cohorts were translated into protein expression biomarkers assessed by immunohistochemistry (IHC) using readily-available FFPE tissues.

Materials and Methods

Selection of candidate prognostic genes

All microarray data are publicly available from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo). For all analyses, the normalized expression values were obtained from each dataset and were not processed further. To identify candidate genes, we obtained two large series of colorectal cancer, including GSE41258 from the Memorial Sloan-Kettering Cancer Center, U.S.A., which was based on Affymetrix HG-U133A platform (14), and GSE17538 based on Affymetrix HG-U133 plus2.0 platform (7). We initially conducted differential expression analysis using GSE41258, in which 54 normal colon mucosa and 186 primary colorectal cancer tissues with stage I-IV disease were available. Class comparison analysis using BRB-ArrayTools (http://brb.nci.nih.gov/BRB-ArrayTools.html) identified differentially expressed probes between normal and tumor tissues with stringent p-values at <0.0001. We next focused on survival data in those two datasets. Because of our aim of identifying prognostic genes for stage II-III patients, survival analysis only used stage II-III patients with available relapse-free survival information. Among 186 primary tumors in GSE41258, we found 94 stage II-III tumors with survival data. GSE17538 consisted of expression data from the Moffit Cancer Center, U.S.A., (GSE17536) and the Vanderbilt Medical Center, U.S.A., (GSE17537), each of which had 111 and 34 stage II-III patients, respectively (n=145, in total). These datasets were each filtered to include only probes with variances in expression levels higher than 0.2, and then stage II-III patients were dichotomized into clearly defined high and low groups on the basis of the median expression value for each probe in

each cohorts separately. GSE41258 and GSE17538, in which GSE17536 and GSE17537 were combined, were each subjected to survival analysis by using BRB-ArrayTools.

Characterization and validation of candidate genes

After developing the list of candidate genes using the probe-based analysis as described above, we further attempted to validate our finding in multiple GEO datasets, including GSE35602 based on Agilent Whole Human Genome Microarray 4x44K G4112F (15), GSE39396 on Affymetrix HT HG-U133+PM (16), and GSE39582 and GSE33113 on Affymetrix HG-U133 plus2.0 (10, 17). Since these datasets each used various platforms comprised of different probe sets, we averaged the multiple probes corresponding to a single gene from further analysis.

GSE35602 consisted of expression profiling data for laser-microdissected epithelial and stromal areas from 4 normal mucosa as well as 13 colorectal cancer tissues (15). In dataset GSE39396, 4 distinct cell subpopulations were separated from 6 primary colorectal cancer samples using FACS cell sorting (16). Purified populations included CD45+EPCAM-CD31-FAP- (leukocytes), CD45-EPCAM+CD31-FAP- (endothelial cells), CD45-EPCAM-CD31+FAP- (endothelial cells) and CD45-EPCAM-CD31-FAP+ (cancer-associated fibroblasts). Expression levels of candidate genes in GSE39396 were median-centered and were subjected to an unsupervised hierarchical clustering for genes by the centroid linkage method using Cluster3.0 and were

visualized using Java TreeView (18). Finally, two recent transcriptome datasets of CRC, including GSE39582 and GSE33113, were used to validate the prognostic significance of candidate genes (10, 17). A total of 566 stage I-IV colon cancer, collected from seven centers in France, were analyzed in GSE39582. This dataset included 453 stage II-III colon cancer patients with relapse-free survival information, while GSE33113 consisted of 90 stage II colorectal cancer patients (one case lacked survival data) who treated in the Academic Medical Center, Netherland. Consistent with the initial survival analysis, patients were categorized as low or high based on the median expression of candidate genes within each cohort separately.

Patients and samples

Primary colorectal tumors from 368 consecutive patients who underwent surgical resection between 1990 and 2010 at the Department of Organ Regulatory Surgery in Fukushima Medical University hospital were enrolled. Tumors were classified according to the TNM classification of malignant tumors (UICC 7^{th} edition) (4). Ten patients were excluded because of missing or inadequate FFPE blocks. Twenty patients who received preoperative chemotherapy and/or radiotherapy before surgery were also excluded. In total, 338 patients with stage 0-IV CRC were eligible to be included in this study. Of these, 42 adjacent non-malignant mucosal tissues were also available for evaluation. Clinical and pathological information were retrospectively obtained by

review of medical records and the last follow-up was September 2015. For survival analysis, 272 patients with stage I-IV CRC who underwent potential curative resection (R0) were used. The primary end point of interest was relapse-free survival (RFS), which defined as time from the date of surgery to the date of first relapse. The median follow-up time was 61.2 months. The study was conducted in accordance with Declaration of Helsinki and was approved by the Institutional Review Board of Fukushima Medical University.

Immunohistochemistry

For immunohistochemistry (IHC), a suitable antibody was identified using the Human Protein Atlas database, in which antibody-based comprehensive proteomic data with images are available ( ) (19-21). Sections (4μm thick) were deparaffinized in xylene and rehydrated in a graded ethanol series. Endogenous peroxidases were blocked with 0.3 % hydrogen peroxide in methanol. For VCAN staining, antigens were retrieved by autoclave for 5 min at 10 mM citrate buffer solution (105℃, pH 6.0). Primary rabbit polyclonal anti-VCAN antibody (HPA004726, Prestige Antibodies® Powered by Atlas Antibodies, Sigma-Aldrich, Co. LLC. St. Louis, MO, USA) was incubated at a 1:500 dilution in 10 mM phosphate-buffered saline containing Tween 20 (Sigma-Aldrich) at 4° C overnight, and subsequently detected by a horseradish peroxidase–coupled anti-rabbit polymer (Envision, Dako, Heverlee, Belgium) followed by incubation with

diaminobenzidine (Dako). All sections were counterstained with Carrazzi's hematoxylin. Percentage positivity of stromal VCAN was scored as 0 (0%–5%), 1 (5%–25%), 2 (>25%) in tumor stromal area. Intensity of staining was scored as either 0 (negative), 1 (weak-moderate), 2 (strong). Scores were combined to generate a VCAN IHC score (min, 0; max, 4). Microscopic analyses were evaluated independently by two investigators who had no prior knowledge of the clinical data.

Statistical analysis

Fisher's exact test, Chi-square test, Student t-test and Mann-Whitney U test were used to determine differences in clinicopathological variables between high and low groups. Spearman's rank correlation was used to evaluate the correlations of immunohistochemical score or gene expression with TNM stage. Cumulative survival was estimated by the Kaplan-Meier method, and differences between 2 groups were analyzed by log-rank test. Univariate and multivariate models were computed using Cox proportional hazards regression. All statistical analyses were two-sided and were conducted using Graphpad Prism v6.0 (Graphpad Software Inc.) and SPSS Statistics version 22 (IBM Corporation). Differences were considered statistically significant at the levels of $P < 0.05$.

Results

Eight candidate genes selected by a genome-wide approach using two microarray cohorts.

To identify genes associated with disease relapse in CRC, we started with 3 independent comprehensive approaches utilizing 2 Affymetrix microarray datasets of CRC. First, GSE41258 was analyzed to search for Affymetrix probes with significantly altered expression in 186 primary tumors compared to 54 normal tissues (<0.0001) (14). Secondly, survival analysis was conducted in two cohorts separately. Since our aim was to develop biomarkers that can stratify stage II-III CRC patients with high risk of relapse, survival analysis only included 94 and 145 patients with stage II-III CRC in GSE41258 and GSE17538 datasets, respectively (Supplementary Table 1) (7, 14). In each dataset, available Affymetrix probes were dichotomized into high or low based on median, and then Cox hazard ratios and p-values for each probe were computed. Finally, a set of 9 probes, which correspond to 8 genes, were identified as candidates for further analysis (Table 1 and Supplementary Figure 1). Six of 8 genes, including COL4A2, COL4A1, SERPINE1, VCAN, NOTCH3 and GPR116, were each found to be upregulated in CRC compared to normal colon, and also associated with decreased relapse-free intervals in both cohorts. Whereas, 2 genes downregulated in CRC, including TTC38 and SULT1B1, were negatively associated with poor prognosis.

Genes associated with high risk of relapse are predominantly expressed in cancer stroma.

We noticed that the candidate gene list contained collagen genes, a proteoglycan-related

gene and a member of serpin family that are supposed to be expressed in stroma, raising a question of whether the origin of those genes are cancer cells themselves or stromal cells in the tumor microenvironment. Since the CRC datasets we used were obtained by macro-dissection of whole tumor samples, both epithelial and stromal cell components within tumor could potentially contribute to the expression profiles. To address this issue, we analyzed a dataset GSE35602, in which expression data were obtained separately from epithelial cells and stromal cells of normal mucosa as well as CRC samples by laser capture micro-dissection (15). As demonstrated in Figure 1, 6 of 8 genes, excluding SULT1B1 and TTC38, were found to be expressed predominantly in stromal components in both normal and cancer tissues. In fact, the expression of those 6 genes were significantly higher in cancer stroma than that of cancer epithelium. Furthermore, COL4A2, COL4A1, VCAN and SERPINE1 were each significantly upregulated in cancer stroma in comparison to normal stroma, indicating that those 4 genes were specifically expressed in cancer stroma. Therefore, it is considered that the expression levels of those genes in macro-dissected CRC samples depend primarily on stromal components. We next attempted to determine the cell-type specificity of each of the stromal genes we identified. We utilized a transcriptome dataset GSE39396 in which 4 distinct cell populations were isolated by FACS from dissociated primary CRC samples, consisting of expression data for epithelial cancer cells (Epcam+), leukocytes (CD45+), endothelial cells (CD31+) and cancer-associated fibroblasts (CAFs) (FAP+) (16). As shown in Figure 2, this

analysis confirmed the stromal origin of the genes. We found that the expression of COL4A1 and COL4A2 was expressed by endothelial cells and CAFs. Strikingly, the expression of VCAN and NOTCH3 was found specifically in CAFs, while SERPINE1 and GPR116 expression was mainly observed in endothelial cells. Among 4 genes upregulated in cancer stroma (Figure 1), the localization of COL4A2 and COL4A1 protein expression in CRC was well-characterized by previous studies (22, 23). However, it is unclear if the protein expression of VCAN and SERPINE1 is primarily localized to tumor stroma in CRC. To confirm the localization and expression of VCAN and SERPINE1 in protein levels, we performed immunohistochemistry (IHC) using a small series of normal colon mucosa and CRC tissues. For IHC analysis, suitable antibodies were identified using the Human Protein Atlas database. VCAN protein expression was exclusively found in cancer stroma in agreement with that of mRNA levels as demonstrated above, while it was undetectable in any of the normal epithelium, normal stroma, or cancer cells (Figure 3). In tumor stroma, VCAN staining was found in infiltrated fibroblasts and extracellular matrix, but not in immune cells or vessels. On the other hand, IHC for SERPINE1 showed diffuse equivocal staining pattern in the majority of CRC tissues (data not shown). Thus, immunohistochemical evaluation of SERPINE1 protein was omitted from further analysis.

Independent validation of prognostic significance of stromal genes in stage II-III CRC patients.

Recent studies of CRC have clearly shown that stromal cells in the tumor microenvironment, such as CAFs, contribute to transcriptional subtypes based on expression profiling of whole tumor samples (24, 25). Importantly, those stromal genes link to biologically and clinically aggressive tumor behavior, displaying the association with poor relapse-free survival. Those evidence seems consistent with our findings, as we discovered a unique set of stromal genes that were associated with disease relapse in two cohorts of stage II-III CRC. In an attempt to further validate our initial findings, we utilized additional microarray datasets of colon cancer in which RFS data were available (Supplementary Table 1). Among 6 stromal genes we identified, high expression of VCAN and SERPINE1 were each significantly associated with poor RFS in 453 patients with stage II-III colon cancer in GSE39582 (Figure 4). However, COL4A2, COL4A1, NOTCH3 and GPR116 had no significant prognostic impact in this cohort. An independent dataset GSE33113, consisted of 89 stage II colon cancer patients, was used for further validation of VCAN and SERPINE1, showing significant association of the 2 genes with RFS (Figure 5).

Prognostic impact of stromal VCAN protein by immunohistochemistry

We consider that the most straightforward way to examine stromal gene expression in the tumor would be by histopathology. Indeed, we validated the stromal localization of VCAN protein by IHC as described above (Figure 3). We also observed that stromal VCAN staining levels in CRC

varied considerably from tumor to tumor. This observation prompted us to further characterize the clinical significance of stromal VCAN protein using a larger set of CRC. Thus, we evaluated the expression of VCAN using IHC in a total of 338 FFPE samples obtained from stage 0-IV CRC as shown in Table 2, and representative images are demonstrated in Figure 6. Semiquantitative VCAN IHC was scored as 0 in 89 (26.3%), 1 in 14 (4.1%), 2 in 41 (12.1%), 3 in 111 (32.8%) and 4 in 83 (24.6%). By contrast, the majority of adjacent normal mucosa demonstrated negative VCAN staining (35 of 42, 83%) and the others showed score 1 (7 of 42, 17%). It is worth noting that stromal VCAN immunoreactivity score increased significantly along with tumor progression (Figure 7A). This finding was also confirmed by gene expression analysis in transcript levels (Figure 7B). Of 338 FFPE tissues used for IHC, 272 patients with stage I-IV CRC with curative resection (R0) were available for RFS analysis (Supplementary Figure 2A-B). In 272 stage I-IV patients, higher levels of stromal VCAN score was significantly associated with poorer RFS (Supplementary Figure 2C, Trend P=0.0005). We divided tumors into high or low stromal VCAN expression and we next evaluated its relationship with clinicopathological characteristics of CRC patients (Table 2). Tumors exhibiting high stromal VCAN expression were significantly associated with aggressive characteristics, including advanced stage (P<0.0001), deeper depth of invasion (P<0.0001), positive lymph node involvement (P<0.0001) and presence of distant metastasis (P=0.0002). Also, high stromal VCAN was more frequently observed in rectal cancer than that of proximal or distal colon

cancer (P=0.040). There was no significant association of VCAN with patient age, gender, or histological type. Tumors with high stromal VCAN expression was significantly associated with shorter RFS intervals than that of low VCAN expression in patients with stage I-IV CRC (P<0.0001, Supplementary Figure 3D). Its prognostic significance remained significant when the analysis was limited to stage II-III CRC patients (P=0.0154, Figure 8A). This is highly consistent with the results of our microarray analyses in which a total of 4 independent cohorts of stage II-III CRC were used for gene selection and validation.

Stromal VCAN was associated with disease recurrence in colon cancer

The vast majority of microarray cohorts we used were expression profiling of colon cancer (at least more than 93%), while only a fraction of them were from rectal cancer (Supplementary Table 1). Also, there are some differences between colon and rectal cancer in terms of the treatment approaches and the pattern of postoperative recurrence (3). Thus, we further analyzed the association of VCAN with prognosis in colon and rectal cancer separately. Strikingly, high expression of VCAN was significantly associated with shorter RFS only in patients with stage II-III colon cancer (P=0.0003, Figure 8B), but not in patients with rectal cancer (P=0.5099, Figure 8C), despite no significant survival difference between colon and rectal cancer (Supplementary Figure 2B). We observed that primary CRC tumors exhibiting high VCAN expression frequently resulted in distant

recurrence, but not in locoregional recurrence (P=0.0032, Supplementary Table 2). Since the majority who experienced locoregional recurrence were patients with rectal cancer (12 of 13 patients), our findings suggest that the relatively high frequency of locoregional recurrence in rectal cancer might confound the prognostic impact of VCAN. We were not able to perform stratified analysis by colon and rectal cancer in microarray cohorts, since those datasets mostly consisted of patients with colon cancer.

Stromal VCAN as independent prognostic factor in colon cancer

Given that no significant impact of VCAN was seen in patients with rectal cancer, further analysis was performed by focusing on colon cancer. In stratified analysis of stage II patients, colon cancer exhibiting high stromal VCAN expression showed significant worse RFS, while no relapse was observed in patients with low VCAN (P=0.0029, Figure 9A). Clear tendency of high VCAN expression with poor survival was also found in patients with stage III colon cancer, although it did not reach statistical significance due to small sample size (P=0.0736, Figure 9B). Univariate Cox analysis in stage II-III colon cancer revealed that presence of lymph node metastasis and T4 tumors were significantly associated with poor RFS, in agreement with previous studies (Table 3) (6). High stromal VCAN expression was also associated with shorter RFS (hazard ratio [HR] 9.09; 95% confidence interval [CI] 2.13 to 38.77; P=0.003). In the multivariate analysis, high stromal VCAN

was significantly associated with unfavorable RFS (HR 8.98; 95%CI 2.01 to 40.06; P=0.004), independent of clinical parameters, including lymph node metastasis, depth of invasion, the administration of adjuvant chemotherapy and other conventional factors. We also attempted to conduct univariate and multivariate analyses for VCAN mRNA expression using the largest microarray cohort of 453 patients with stage II-III colon cancer (GSE39582). Again, significant association of high VCAN expression with poor RFS was observed in multivariate analysis (HR 1.60; 95%CI 1.13 to 2.27; P=0.009) (Table 4). Notably, its prognostic value was independent of stage, receipt of adjuvant chemotherapy, and microsatellite status.

Discussion

In the present study, our comprehensive approaches using two CRC cohorts identified a list of genes with differential expression in tumor compared to non-tumor that were associated with RFS in stage II-III CRC. We have attempted to validate the prognostic significance of those candidate genes not only in mRNA levels, but also in protein levels by IHC. Although this study was based on retrospective analyses, our strategy included the use of multiple independent cohorts consisting of 781 patients in total with stage II-III CRC that were publicly available, followed by our large set of well-characterized CRC cohort with long term survival information. In all 5 cohorts we analyzed, the expression levels of VCAN transcript or protein were found to be significantly

associated with disease relapse in stage II-III CRC. Although it is known that genome-wide approaches, such as microarray analysis, could yield numerous false negative results, our sequential validation strategy by integrating currently available microarray data was successfully minimized such false discovery, leading to the identification of novel and robust prognostic genes in CRC with high reliability.

VCAN (versican) belongs to the family of large chondroitin sulfate proteoglycans and has hyaluronate binding properties. Versican has been known to play a role in the formation of tumor-specific extracellular matrix that can support cancer cell growth and metastasis in some solid cancers. Stromal VCAN expression has been proposed as a prognostic biomarker for worse disease outcome in several cancer types, including ovarian cancer, endometrial cancer, gastric gastrointestinal stromal tumors and oral squamous cell carcinoma (26-29). However, there have been only a few studies examining the clinical implication of stromal VCAN expression as a biomarker in CRC. Although one study reported the association of VCAN expression with favorable prognosis, they addressed its expression in epithelial cells (30). By contrast, our sequential validation strategy, in which we integrated multiple datasets of gene expression profiling from both macro-dissected and micro-dissected samples as well as IHC analysis with a validated antibody supported by the Human Protein Atlas, clearly indicates that VCAN is specifically localized at tumor stroma but not in epithelial cells. Hence, this study has provided the first evidence that VCAN is exclusively expressed

in tumor stroma and is associated with disease recurrence in CRC.

We have shown that the prognostic impact of VCAN expression is robust and that transcript data and IHC data from a variety of countries using different platforms provide highly consistent results. Noteworthy is the fact that VCAN was found to be exclusively expressed in cancer stroma, but not in tumor epithelial cells or normal mucosa. Most recently, two studies have highlighted the significant contribution of stromal components to a transcriptional subtype of CRC which is characterized by highly aggressive phenotype associated with adverse patient outcomes (24, 25). Gene expression profiles of macro-dissected whole tumor samples could consist of mixed signals originating from different types of cells, including epithelial tumor cells and tumor-associated stromal cells. Although several investigators previously reported molecular classifications (9, 10, 31), the above two studies have clearly shown that the highly aggressive subtypes defined by transcriptomic studies rely on the gene expression signature by tumor stromal cells, such as CAFs, rather than that of cancer cells. Hence, tumor stroma should not be considered contamination of gene expression data but it plays a major role in promoting tumorigenesis through the interaction with cancer cells. Calon et al. have revealed that CAFs can increase tumor aggressiveness enhanced by TGF-β signaling, while poor prognostic signatures share a stromal gene program that is induced by TGF-β (24). This concept seems highly consistent with our findings that VCAN was found to be specifically expressed by CAFs in the tumor microenvironment and its

expression levels increased significantly with tumor progression. The stepwise upregulation of VCAN from normal tissue, to early stage tumor through advanced tumor was observed in both mRNA and protein levels. Also, higher stromal VCAN expression was frequently found in primary CRC tumors that exhibited aggressive characteristics, including deeper depth of invasion, positive lymph node involvement and distant metastasis. The most recent and robust molecular subtyping of CRC reported by the CRC Subtyping Consortium (CRCSC), based on more than 4000 CRC samples of transcriptomic and genomic data, have clearly demonstrated a subtype (CMS4) which is characterized by stromal infiltration and prominent TGF-β activation (13). It seems that our discovery of stromal VCAN as poor prognostic indicator recapitulated this highly aggressive subtype. Indeed, CMS4 tumors not only exhibited stromal infiltration, overexpression of extracellular matrix proteins and higher admixture with non-cancer cells, but also had the highest rate of recurrence across all subtypes. Those findings consistently suggest that CAF-derived VCAN expression in CRC is involved in pro-tumorigenic program, possibly through the enhanced TGF-β signaling in the tumor microenvironment. In the meantime, in ovarian cancer cell lines, it was revealed that TGF-β can enhance cancer cell invasion by upregulating CAF-specific VCAN expression (32). Concerning the functional role in ovarian cancer and unfavorable prognostic impact of VCAN in several cancer types, the expression of VCAN in tumor stroma is not specific to CRC but might characterize aggressive behavior across solid cancer types.

The ultimate goal of this study was to develop a single gene-based prognostic biomarker for stage II-III CRC to help guide clinical decisions after curative surgery. The prognostic impact of VCAN expression, identified and validated by 4 independent microarray analyses using a total of 781 patients with stage II-III CRC, was finally verified on the protein levels assessed by IHC in 196 patients with stage II-III CRC. Despite its robust prognostic significance in CRC, our IHC analyses revealed that no impact of VCAN on survival of patients with rectal cancer. This might not be unexpected because we identified and validated the significance of VCAN expression using multiple microarray cohorts, in which the vast majority of samples were obtained from colon cancer patients. Considering potential differences in etiology, genetics, treatment strategies and outcomes between colon and rectal cancer, studying these two entities separately may be more appropriate. In contrast to rectal cancer, there was strong association of stromal VCAN with disease recurrence in colon cancer, independent of other clinical covariates. Strikingly, the subset analysis with only patients with stage II colon cancer indicated that tumors exhibiting high stromal VCAN was significantly associated with poor RFS. This suggests that VCAN IHC has potential to identify high risk patients with stage II colon cancer who may benefit from postoperative adjuvant chemotherapy.

Previous microarray studies reported multigene prognostic signatures which utilize dozens to hundreds of genes with complex classification models and usually require frozen tissues. We consider this might be an obstacle to clinical translation of such biomarkers, because the current

clinical practice relies heavily on FFPE tissues. This study focused on the identification of single genes with prognostic relevance, rather than to generate new multigene prognostic signatures. Immunohistochemical detection for stromal VCAN is simple and uses only a single gene that can be analyzed on readily-available archived FFPE tissues. This suggests that stromal VCAN by IHC can be potentially used as a practical prognostic biomarker for patients with CRC.

In conclusion, our genome-wide approaches with stringent validation strategy using multiple independent cohorts identified and validated the significant association between VCAN transcript expression and RFS in stage II-III CRC. Its independent prognostic impact was further verified on protein levels by immunohistochemical analyses in a large set of CRC by multivariate Cox analysis. It is worth noting that stromal VCAN expression efficiently discriminated stage II colon cancer patients with distinct survival outcomes. This highlights the potential of stromal VCAN expression in identifying stage II colon cancer patients at high risk of recurrence who may benefit from adjuvant chemotherapy.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 15K10143 and 25870582. Microarray analyses were performed using BRB-ArrayTools developed by Dr. Richard Simon and BRB-ArrayTools Development Team.

Figure Legends

Figure 1. Expression levels of 8 candidate genes in a transcriptomic dataset of laser-microdissected epithelial (Ep) and stromal (Str) areas from 4 normal mucosa as well as 13 colorectal cancer tissues (GSE35602). Most of the candidates were found predominantly in micro-dissected stromal components. COL4A2, COL4A1, VCAN and SERPINE1 were specifically expressed in cancer stroma. *P<0.05, **P<0.01, ***P<0.001.

Figure 2. Heat map depicting cell type-specific expression of the candidate stromal genes in a dataset of 4 distinct cell subpopulations from 6 primary colorectal cancer samples using FACS cell sorting (GSE39396). COL4A1 and COL4A2 were expressed by endothelial cells and cancer-associated fibroblasts (CAFs). The expression of VCAN and NOTCH3 was found specifically in CAFs, while SERPINE1 and GPR116 expression was predominantly observed in endothelial cells.

Figure 3. Immunohistochemistry for VCAN in colon cancer samples. The expression of VCAN protein was exclusively found in cancer stroma, while it was undetectable in any of the normal epithelium, normal stroma, or tumor epithelial cells.

Figure 4. Validation of the prognostic significance of 6 stomal genes in an independent dataset (GSE39582). Kaplan-Meier curves showing relapse-free survival of 453 patients with stage II-III colon cancer. Patients were dichotomized as high or low based on the median expression of candidate stromal genes, including VCAN (A), SERPINE1 (B), COL4A1 (C), COL4A2 (D), NOTCH3 (E) and GPR116 (F).

Figure 5. Validation of the prognostic significance of VCAN and SERPINE1 in an additional independent dataset (GSE33113). Kaplan-Meier curves showing relapse-free survival of 89 patients with stage II colon cancer. Patients were categorized as high or low based on the median expression of VCAN (A) and SERPINE1 (B).

Figure 6. Immunohistochemistry for VCAN ranged from negative to strong stromal staining in colorectal cancer tissues. Representative images of VCAN staining in colon tumor stroma are shown.

Figure 7. Upregulation of VCAN expression along with tumor progression. (A) Immunoreactivity score for stromal VCAN protein expression in normal and stage 0-IV colorectal cancer tissues. (B) VCAN mRNA expression levels in normal and stage I-IV colorectal cancer tissues in GSE41258.

Figure 8. The association of relapse-free survival with stromal VCAN expression by immunohistochemistry. Kaplan-Meier curves showing 196 patients with stage II-III colorectal cancer (A), and stratified analysis of stage II-III patients with colon cancer (B) and rectal cancer (C).

Figure 9. The association of relapse-free survival with stromal VCAN expression by immunohistochemistry focusing on colon cancer patients. Kaplan–Meier curves for patients stage II (A) and stage III (B) colon cancer.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015;65(1):5-29.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65(2):87-108.
3. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet. 2014;383(9927):1490-502.
4. Sobin LH, Gospodarowicz MK, Wittekind C, International Union against Cancer. TNM classification of malignant tumours. 7th ed. Chichester, West Sussex, UK ; Hoboken, NJ: Wiley-Blackwell; 2010. xx, 310 p. p.
5. Gray R, Barnwell J, McConkey C, Hills RK, Williams NS, Kerr DJ, et al. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. Lancet. 2007;370(9604):2020-9.
6. Dienstmann R, Salazar R, Tabernero J. Personalizing colon cancer adjuvant therapy: selecting optimal treatments for individual patients. J Clin Oncol. 2015;33(16):1787-96.

7. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology. 2010;138(3):958-68.

8. Oh SC, Park YY, Park ES, Lim JY, Kim SM, Kim SB, et al. Prognostic gene expression signature associated with two molecularly distinct subtypes of colorectal cancer. Gut. 2012;61(9):1291-8.

9. De Sousa E Melo F, Wang X, Jansen M, Fessler E, Trinh A, de Rooij LP, et al. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. Nat Med. 2013;19(5):614-8.

10. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. 2013;10(5):e1001453.

11. Agesen TH, Sveen A, Merok MA, Lind GE, Nesbakken A, Skotheim RI, et al. ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. Gut. 2012;61(11):1560-7.

12. Sanz-Pamplona R, Berenguer A, Cordero D, Riccadonna S, Solé X, Crous-Bou M, et al. Clinical value of prognosis gene expression signatures in colorectal cancer: a systematic review. PLoS One. 2012;7(11):e48877.

13. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015.

14. Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F, et al. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proc Natl Acad Sci U S A. 2009;106(17):7131-6.

15. Nishida N, Nagahara M, Sato T, Mimori K, Sudo T, Tanaka F, et al. Microarray analysis of colorectal cancer stromal tissue reveals upregulation of two oncogenic miRNA clusters. Clin Cancer Res. 2012;18(11):3054-70.

16. Calon A, Espinet E, Palomo-Ponce S, Tauriello DV, Iglesias M, Céspedes MV, et al. Dependency of colorectal cancer on a TGF-ß-driven program in stromal cells for metastasis initiation. Cancer Cell. 2012;22(5):571-84.

17. de Sousa E Melo F, Colak S, Buikhuisen J, Koster J, Cameron K, de Jong JH, et al. Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. Cell Stem Cell. 2011;9(5):476-85.

18. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863-8.

19. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.

20. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. 2010;28(12):1248-50.

21. Pontén F, Jirström K, Uhlen M. The Human Protein Atlas--a tool for pathology. J Pathol. 2008;216(4):387-93.

22. Ikeda K, Iyama K, Ishikawa N, Egami H, Nakao M, Sado Y, et al. Loss of expression of type IV collagen alpha5 and alpha6 chains in colorectal cancer associated with the hypermethylation of their promoter region. Am J Pathol. 2006;168(3):856-65.

23. Tanjore H, Kalluri R. The role of type IV collagen and basement membranes in cancer progression and metastasis. Am J Pathol. 2006;168(3):715-7.

24. Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. Nat Genet. 2015;47(4):320-9.

25. Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, et al. Stromal contribution to the colorectal cancer transcriptome. Nat Genet. 2015;47(4):312-9.

26. Setoguchi T, Kikuchi H, Yamamoto M, Baba M, Ohta M, Kamiya K, et al. Microarray analysis identifies versican and CD9 as potent prognostic markers in gastric gastrointestinal stromal tumors. Cancer Sci. 2011;102(4):883-9.

27. Ghosh S, Albitar L, LeBaron R, Welch WR, Samimi G, Birrer MJ, et al. Up-regulation of stromal versican expression in advanced stage serous ovarian cancer. Gynecol Oncol. 2010;119(1):114-20.

28. Pukkila M, Kosunen A, Ropponen K, Virtaniemi J, Kellokoski J, Kumpulainen E, et al. High stromal versican expression predicts unfavourable outcome in oral squamous cell carcinoma. J Clin Pathol. 2007;60(3):267-72.

29. Kodama J, Hasengaowa, Kusumoto T, Seki N, Matsuo T, Ojima Y, et al. Prognostic significance of stromal versican expression in human endometrial cancer. Ann Oncol. 2007;18(2):269-74.

30. de Wit M, Belt EJ, Delis-van Diemen PM, Carvalho B, Coupé VM, Stockmann HB, et al. Lumican and versican are associated with good outcome in stage II and III colon cancer. Ann Surg Oncol. 2013;20 Suppl 3:S348-59.

31. Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschleger S, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat Med. 2013;19(5):619-25.

32. Yeung TL, Leung CS, Wong KK, Samimi G, Thompson MS, Liu J, et al. TGF-β modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. Cancer Res. 2013;73(16):5016-28.

Table 1. Candidate genes

Affymetrix Probe ID	Gene Symbol	Description	GSE41258 Tumor vs Normal Fold-change	GSE41258 Tumor vs Normal *P*	GSE41258 RFS Hazard ratio	GSE41258 RFS *P*	GSE17538 RFS Hazard ratio	GSE17538 RFS *P*
211964_at	COL4A2	collagen, type IV, alpha 2	**1.34**	0.0001	**4.65**	0.006	**3.41**	0.003
211981_at	COL4A1	collagen, type IV, alpha 1	**1.96**	< 1e-07	**3.70**	0.011	**3.44**	0.002
211571_s_at	VCAN	versican	**2.67**	< 1e-07	**2.85**	0.032	**2.83**	0.008
211980_at	COL4A1	collagen, type IV, alpha 1	**1.79**	< 1e-07	**2.77**	0.037	**4.06**	0.001
202627_s_at	SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	**2.91**	< 1e-07	**2.76**	0.038	**3.70**	0.001
203238_s_at	NOTCH3	notch 3	**1.82**	< 1e-07	**2.67**	0.044	**3.93**	0.001
212950_at	GPR116	G protein-coupled receptor 116	**1.24**	0.0004	**2.65**	0.046	**6.57**	0.000
207601_at	SULT1B1	sulfotransferase family, cytosolic, 1B, member 1	**0.55**	< 1e-07	**0.38**	0.046	**0.47**	0.037
218272_at	TTC38	tetratricopeptide repeat domain 38	**0.72**	< 1e-07	**0.29**	0.017	**0.37**	0.008

Table 2. Clinical characteristics of colorectal cancer patients according to stromal VCAN protein expression

		Total (n=338)	Stromal VCAN expression		*P*
			High n=194 (%)	Low n=144 (%)	
Age	Mean±SD	66.9±11.7	66.2±12.1	68.0±11.2	0.155
Gender					0.934
	Male	201	115 (59.3)	86 (59.7)	
	Female	137	79 (40.7)	58 (40.3)	
Location					0.040
	Proximal colon	106	51 (26.3)	55 (38.2)	
	Distal colon	103	63 (32.5)	40 (27.8)	
	Rectum	129	80 (41.2)	49 (34.0)	
Histlogical differentiation					0.100
	Well	162	83 (42.8)	79 (54.9)	
	Moderately	146	95 (49.0)	51 (35.4)	
	Poorly	9	5 (2.6)	4 (2.8)	
	Mucinous	20	11 (5.7)	9 (4.6)	
	Other	1	0 (0.0)	1 (0.5)	
Stage(UICC)					<0.0001
	0	14	1 (0.5)	13 (9.0)	
	I	65	22 (11.3)	43 (29.9)	
	II	123	70 (36.1)	53 (36.8)	
	III	91	64 (33.0)	27 (18.8)	
	IV	45	37 (19.1)	8 (5.6)	
Tumor invasion					<0.0001
	Tis (m)	17	1 (0.5)	16 (11.1)	
	T1 (sm)	33	10 (5.2)	23 (16.0)	
	T2 (mp)	50	25 (12.9)	25 (17.4)	
	T3 (ss-a)	140	88 (45.4)	52 (36.1)	
	T4 (se-si/ai)	98	70 (36.1)	28 (19.4)	
Lymph node metastasis					<0.0001
	Absent	214	103 (53.1)	111 (77.1)	
	Present	121	88 (45.4)	33 (22.9)	
	not available	3	3 (1.5)	0 (0.0)	
Distant metastasis					0.0002
	Absent	293	157 (80.9)	136 (94.4)	
	Present	45	37 (19.1)	8 (5.6)	

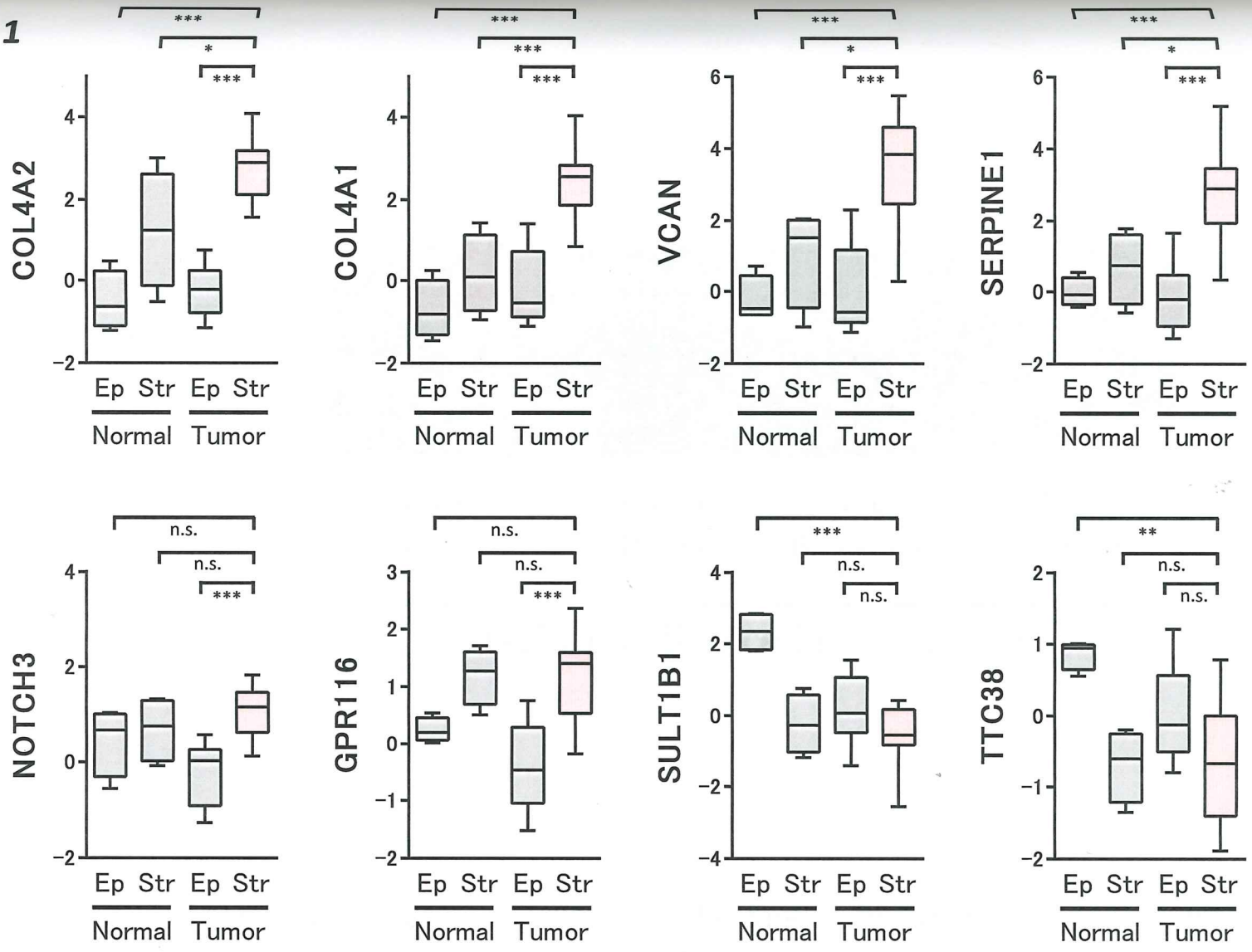
Table 3. Univariate and Multivariate Cox regression of stromal VCAN expression for relapse-free survival in Stage II-III colon cancer by immunohistochemistry

Variable		Univariate			Multivariate		
		HR	95% CI	*P*	HR	95% CI	*P*
Stromal VCAN	High (69) *vs* Low (52)	9.09	2.13 - 38.77	**0.003**	8.98	2.01 - 40.06	**0.004**
Lymph node metastasis	Present (44) *vs* Absent (77)	3.76	1.59 - 8.87	**0.003**	3.58	1.40 - 9.14	**0.008**
Tumor invasion	T4 (48) *vs* T1-3 (73)	2.61	1.13 - 6.03	**0.025**	2.07	0.83 - 5.16	0.118
Age	Continous	1.03	0.99 - 1.08	0.122	1.07	1.00 - 1.15	0.070
Gender	Female (50) *vs* Male (71)	1.15	0.50 - 2.62	0.743	1.47	0.62 - 3.52	0.384
Location	Distal colon (63) *vs* Proximal colon (58)	0.61	0.27 - 1.40	0.246	0.66	0.27 - 1.60	0.354
Histological differentiation	Poor-Others (14) *vs* Well-Moderate (107)	2.05	0.70 - 6.02	0.193	3.25	0.98 - 10.82	0.055
Adjuvant chemotherapy	Yes (63) *vs* No (55)	0.87	0.38 - 1.96	0.730	1.00	0.34 - 2.91	0.996

Table 4. Univariate and Multivariate Cox regression of VCAN expression for relapse-free survival in Stage II-III colon cancer by microarray analysis

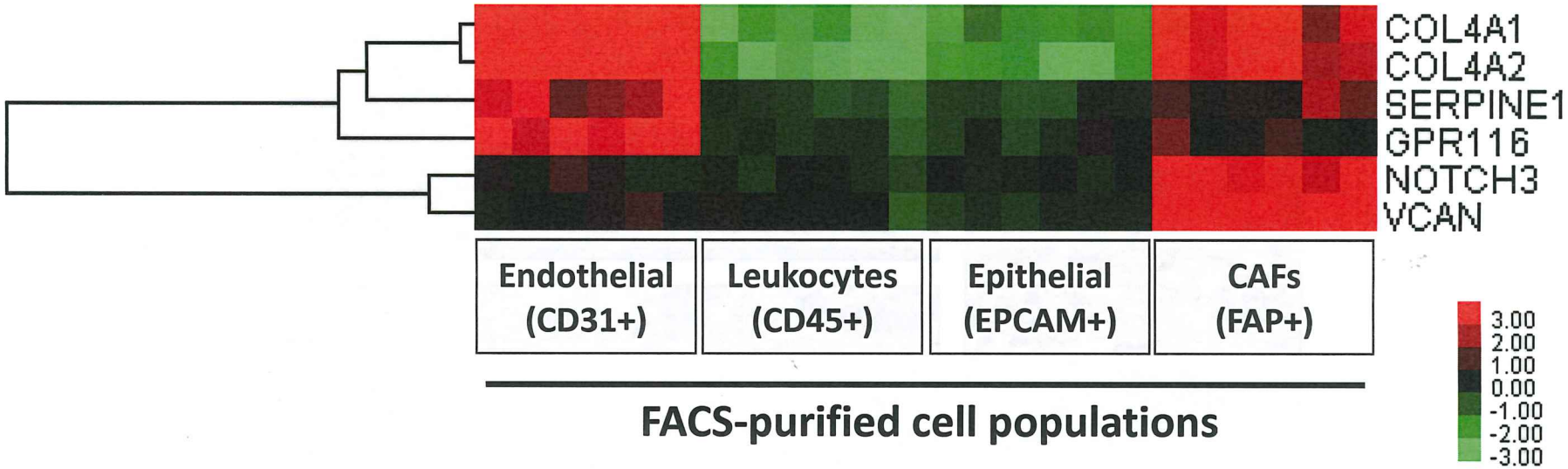
		Univariate			Multivariate		
Variable		HR	95% CI	*P*	HR	95% CI	*P*
VCAN	High (226) *vs* Low (227)	1.44	1.02 - 2.02	**0.036**	1.60	1.13 - 2.27	**0.009**
Stage	Stage III (200) *vs* Stage II (253)	2.05	1.45 - 2.89	**0.000**	1.99	1.29 - 3.08	**0.002**
Age	Continous	1.01	1.00 - 1.02	0.233	1.01	1.00 - 1.03	0.149
Gender	Female (198) *vs* Male (255)	0.76	0.54 - 1.07	0.120	0.79	0.55 - 1.14	0.206
Location	Distal colon (268) *vs* Proximal colon (185)	1.17	0.83 - 1.66	0.373	1.14	0.78 - 1.67	0.505
Microsatellite status	deficient MMR (60) *vs* proficient MMR (350)	0.51	0.27 - 0.97	**0.039**	0.50	0.25 - 0.99	**0.046**
Adjuvant chemotherapy	Yes (200) *vs* No (252)	1.56	1.11 - 2.19	**0.010**	0.89	0.57 - 1.38	0.591

Figure 1



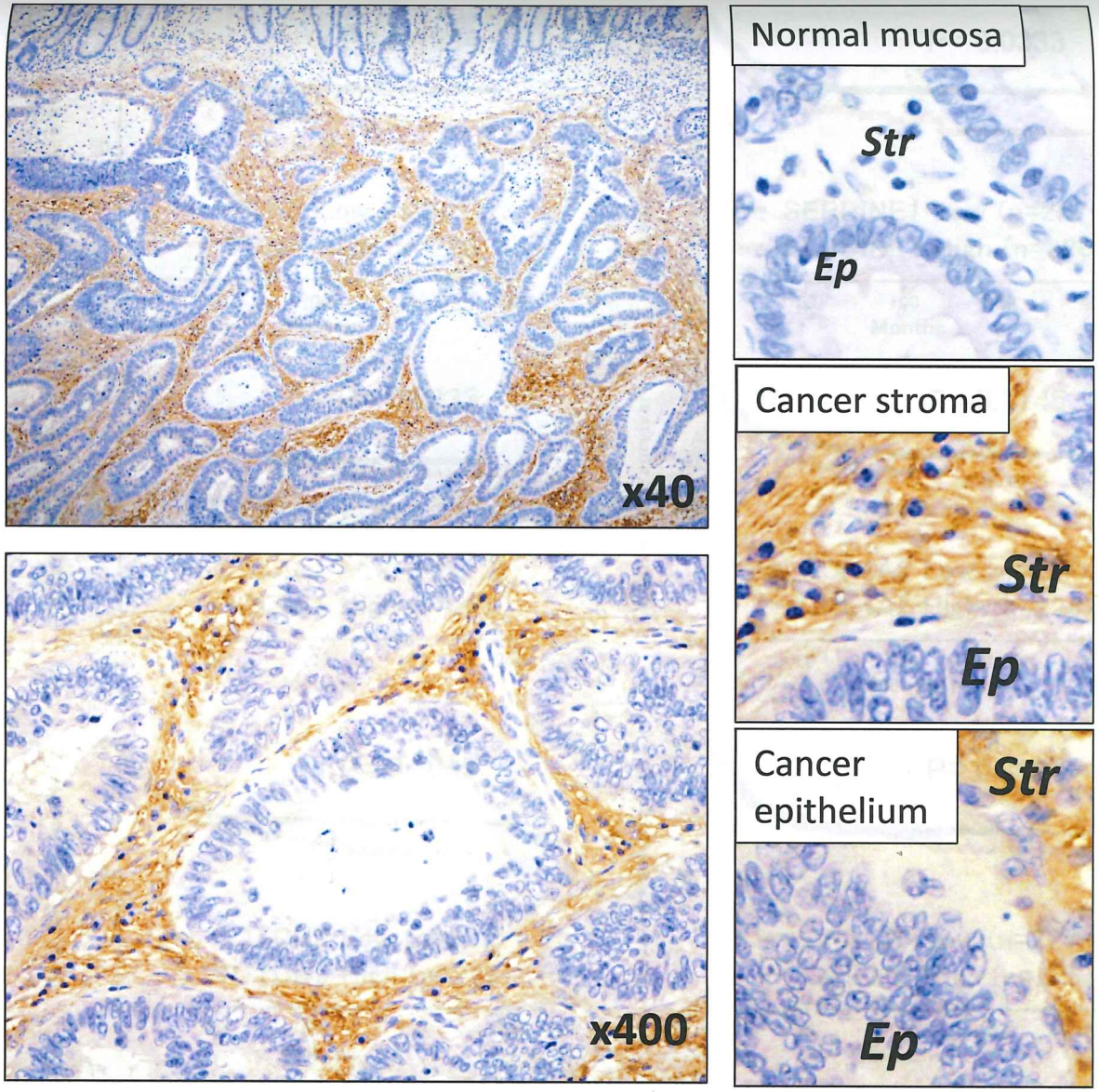
Expression levels of 8 candidate genes in a transcriptomic dataset of laser-microdissected epithelial (Ep) and stromal (Str) areas from 4 normal mucosa as well as 13 colorectal cancer tissues (GSE35602). Most of the candidates were found predominantly in micro-dissected stromal components. COL4A2, COL4A1, VCAN and SERPINE1 were specifically expressed in cancer stroma. *P<0.05, **P<0.01, ***P<0.001.

Figure 2



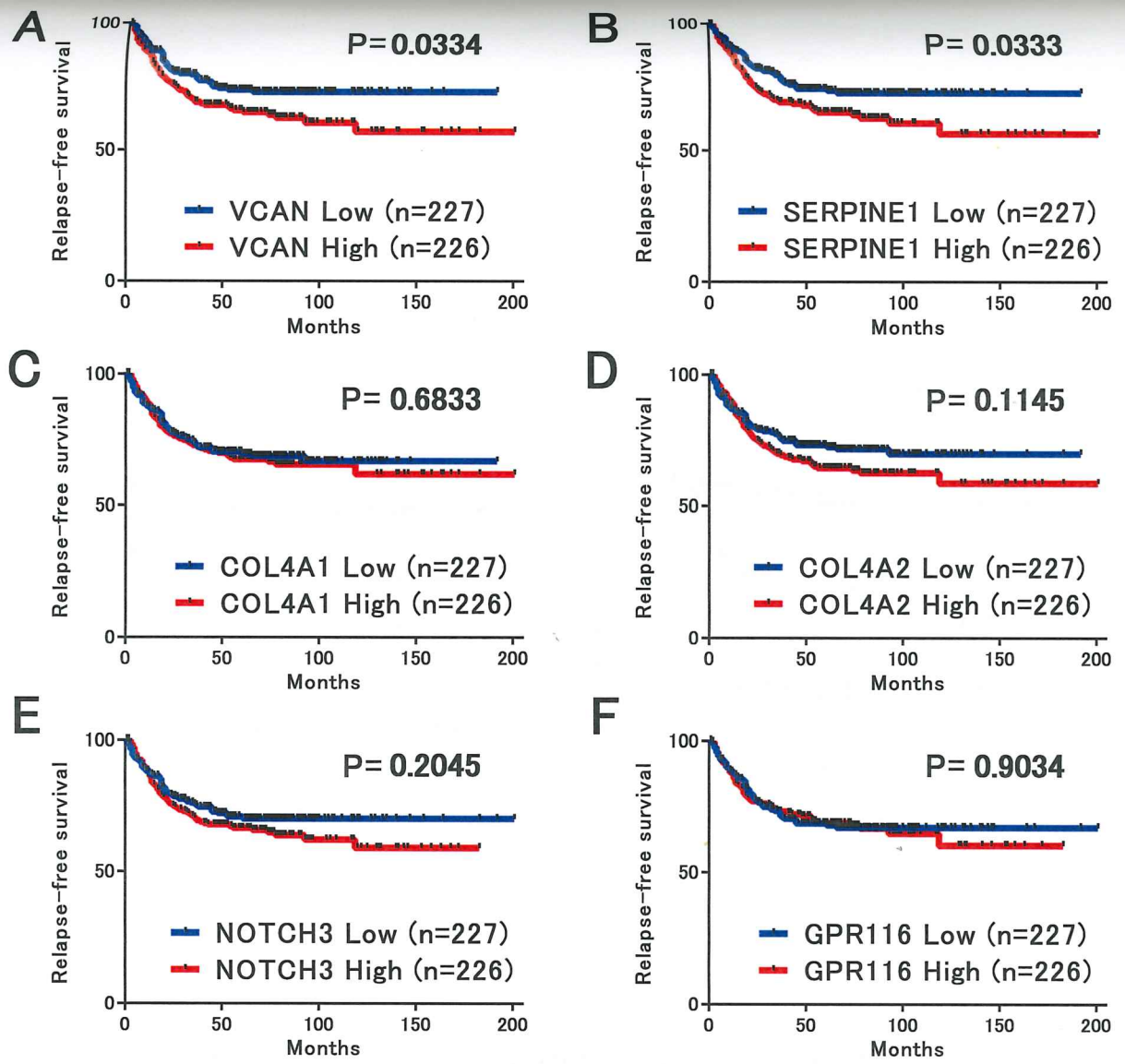
Heat map depicting cell type-specific expression of the candidate stromal genes in a dataset of 4 distinct cell subpopulations from 6 primary colorectal cancer samples using FACS cell sorting (GSE39396). COL4A1 and COL4A2 were expressed by endothelial cells and cancer-associated fibroblasts (CAFs). The expression of VCAN and NOTCH3 was found specifically in CAFs, while SERPINE1 and GPR116 expression was predominantly observed in endothelial cells.

Figure 3



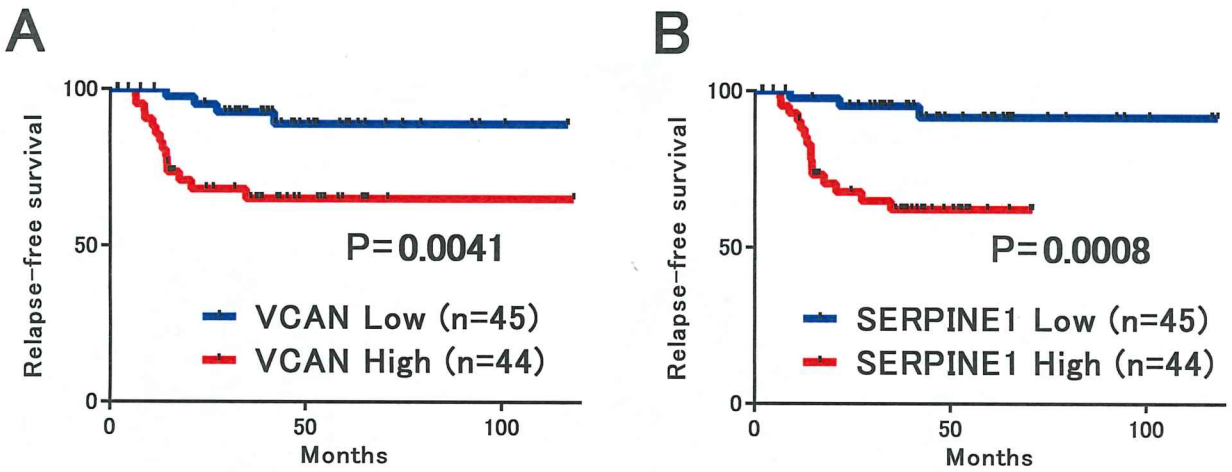
Immunohistochemistry for VCAN in colon cancer samples. The expression of VCAN protein was exclusively found in cancer stroma, while it was undetectable in any of the normal epithelium, normal stroma, or tumor epithelial cells.

Figure 4



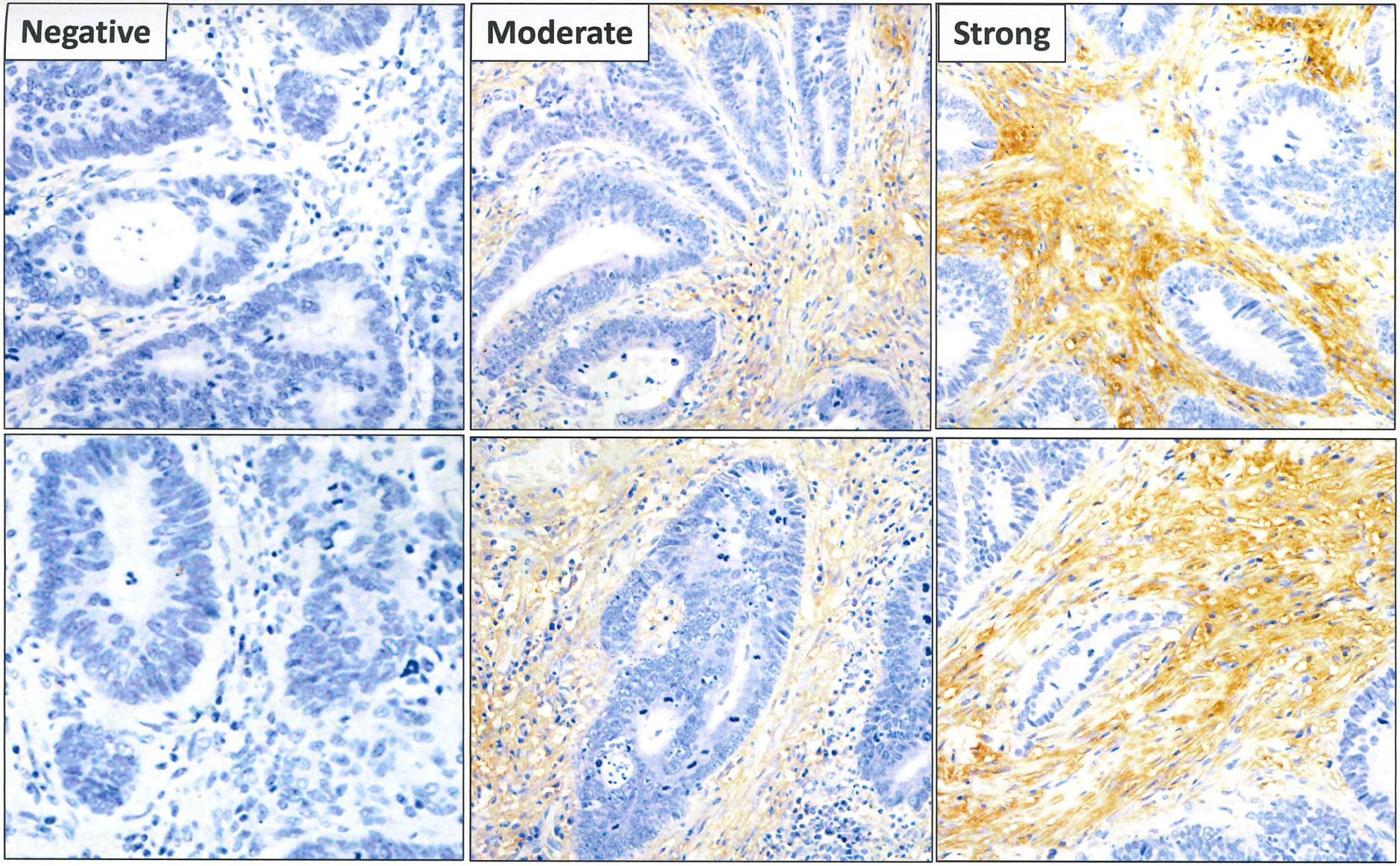
Validation of the prognostic significance of 6 stomal genes in an independent dataset (GSE39582). Kaplan-Meier curves showing relapse-free survival of 453 patients with stage II-III colon cancer. Patients were dichotomized as high or low based on the median expression of candidate stromal genes, including VCAN (A), SERPINE1 (B), COL4A1 (C), COL4A2 (D), NOTCH3 (E) and GPR116 (F).

Figure 5



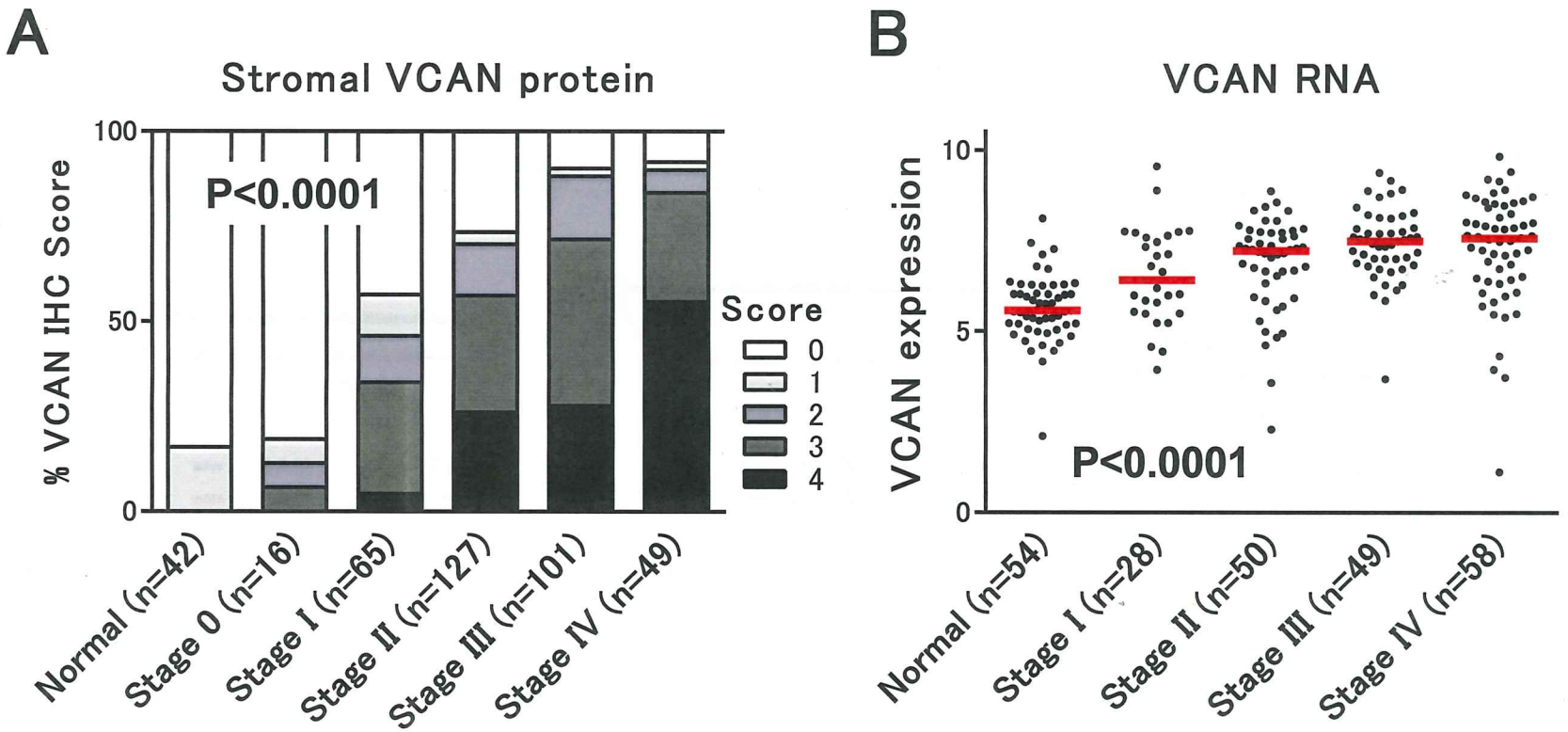
Validation of the prognostic significance of VCAN and SERPINE1 in an additional independent dataset (GSE33113). Kaplan-Meier curves showing relapse-free survival of 89 patients with stage II colon cancer. Patients were categorized as high or low based on the median expression of VCAN (A) and SERPINE1 (B).

Figure 6



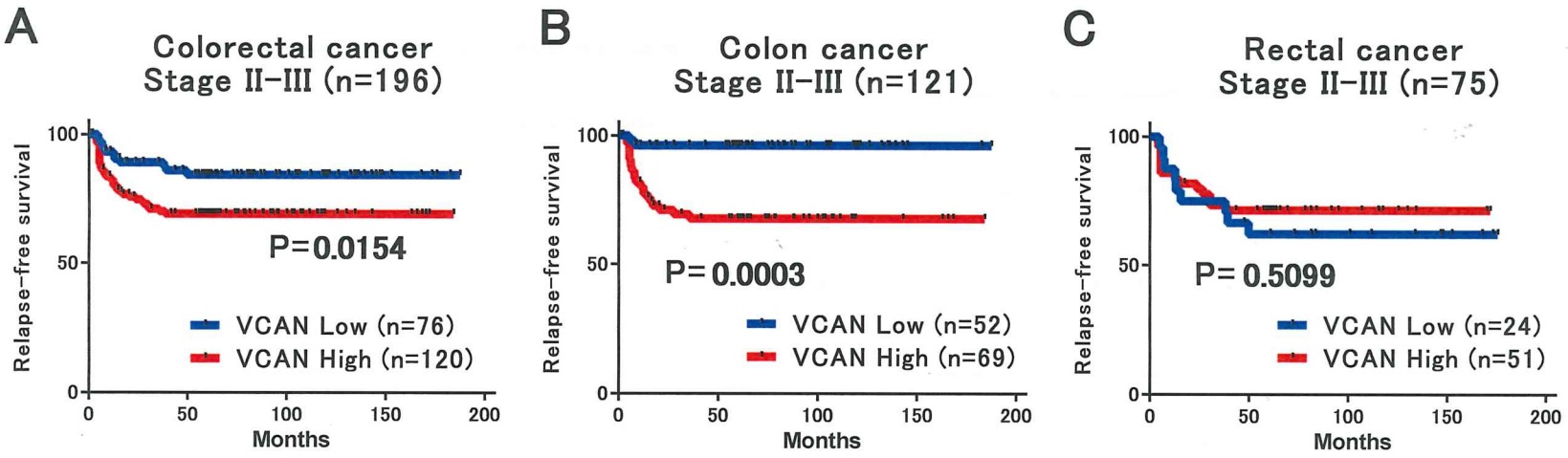
Immunohistochemistry for VCAN ranged from negative to strong stromal staining in colorectal cancer tissues. Representative images of VCAN staining in colon tumor stroma are shown.

Figure 7



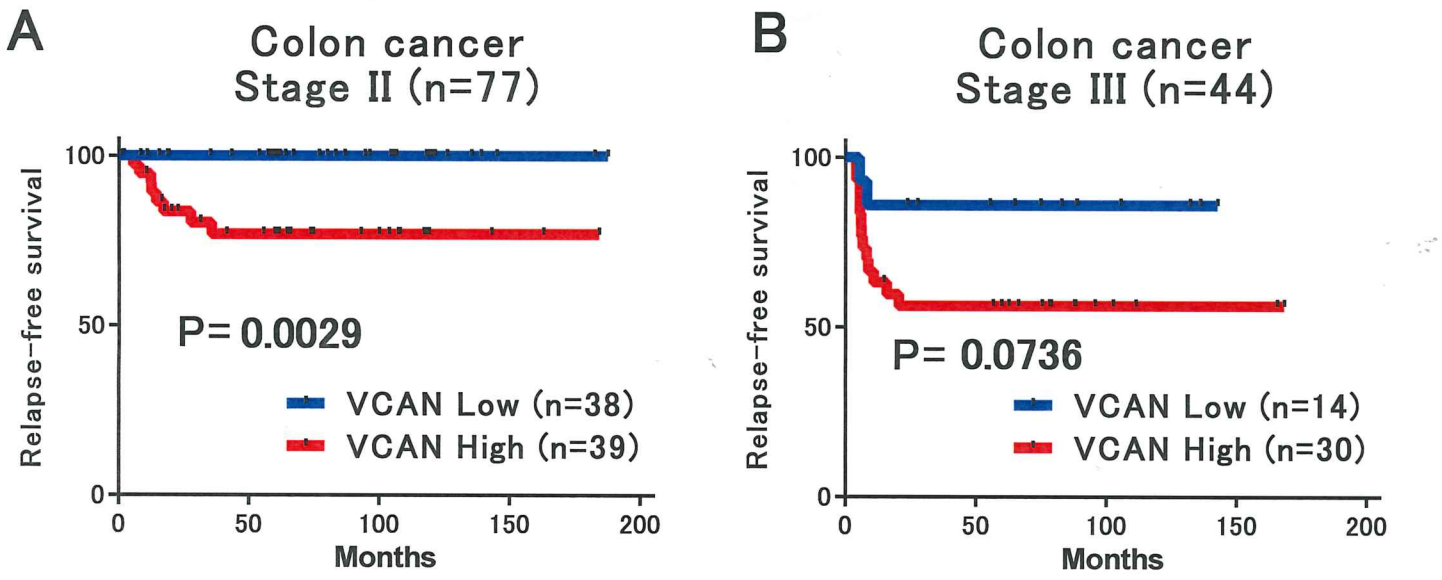
Upregulation of VCAN expression along with tumor progression. (A) Immunoreactivity score for stromal VCAN protein expression in normal and stage 0-IV colorectal cancer tissues. (B) VCAN mRNA expression levels in normal and stage I-IV colorectal cancer tissues in GSE41258.

Figure 8



The association of relapse-free survival with stromal VCAN expression by immunohistochemistry. Kaplan-Meier curves showing 196 patients with stage II-III colorectal cancer (A), and stratified analysis of stage II-III patients with colon cancer (B) and rectal cancer (C).

Figure 9



The association of relapse-free survival with stromal VCAN expression by immunohistochemistry focusing on colon cancer patients. Kaplan–Meier curves for patients stage II (A) and stage III (B) colon cancer.

Supplementary Table 1. Microarray datasets of stage II-III CRC with relapse-free survival information

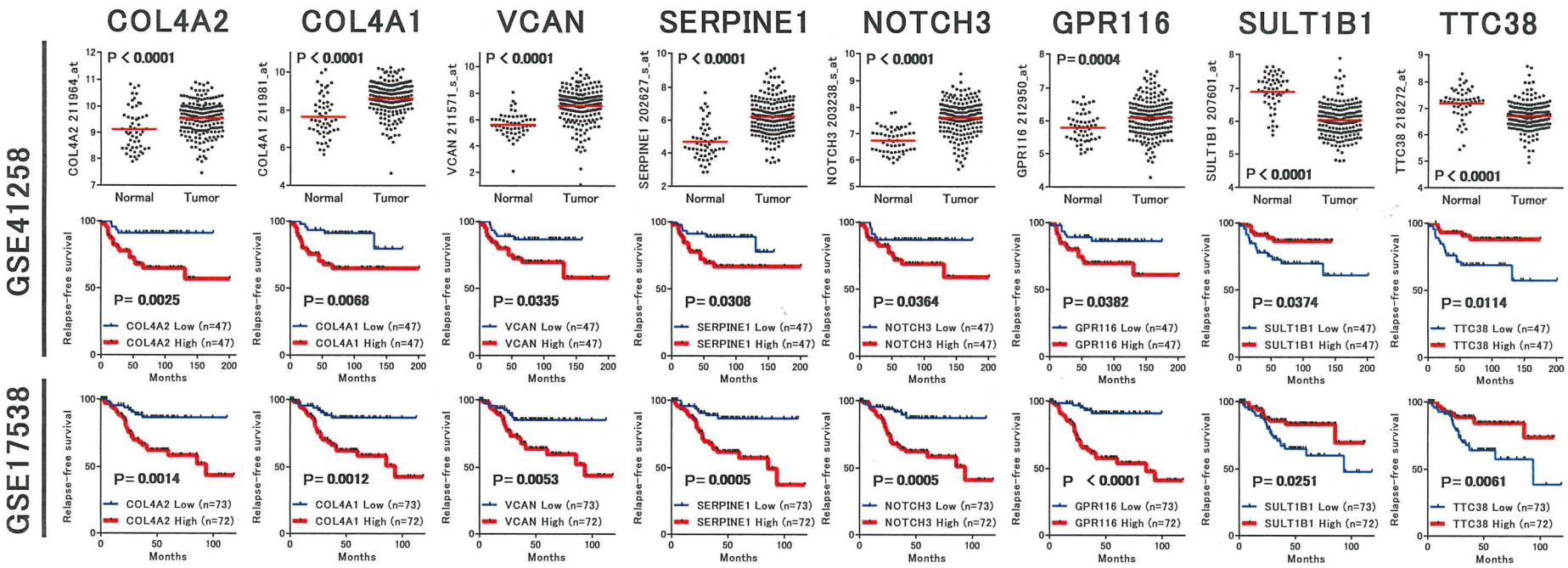
Purpose		*Gene selection*			*Validation 1*	*Validation 2*
		GSE41258	**GSE17538**		**GSE39582**	**GSE33113**
	Total (n=781)	**Total (n=94)**	**Moffit Total (n=111)**	**Vanderbilt Total (n=34)**	**Total (n=453)**	**Total (n=89)**
Age						
Mean±SD		63.8±13.0	64.6±13.8	64.5±12.6	67.5±13.0	52.5±6.4
Gender						
Male	417	48	53	19	255	42
Female	364	46	58	15	198	47
Location						
Colon	733	80	111	0	453	89
Rectum	14	14	0	0	0	0
Colon or rectum (not specified)	34	0	0	34	0	0
Stage						
II	460	48	55	15	253	89
III	321	46	56	19	200	0
Microsatellite status*						
MSI-High or dMMR	79	19	0	0	60	0
MSS/MSI-Low or pMMR	414	64	0	0	350	0
Unknown	288	11	111	34	43	89
Adjuvant therapy						
Yes	200	0	0	0	200	0
No	252	0	0	0	252	0
Unknown	329	94	111	34	1	89
Number of relapse	208	20	31	3	136	18

* deficient MMR (dMMR) and proficient MMR (pMMR) were defined in GSE39582.

Supplementary Table 2. Patterns of recurrence in patients with colorectal cancer after curative surgery

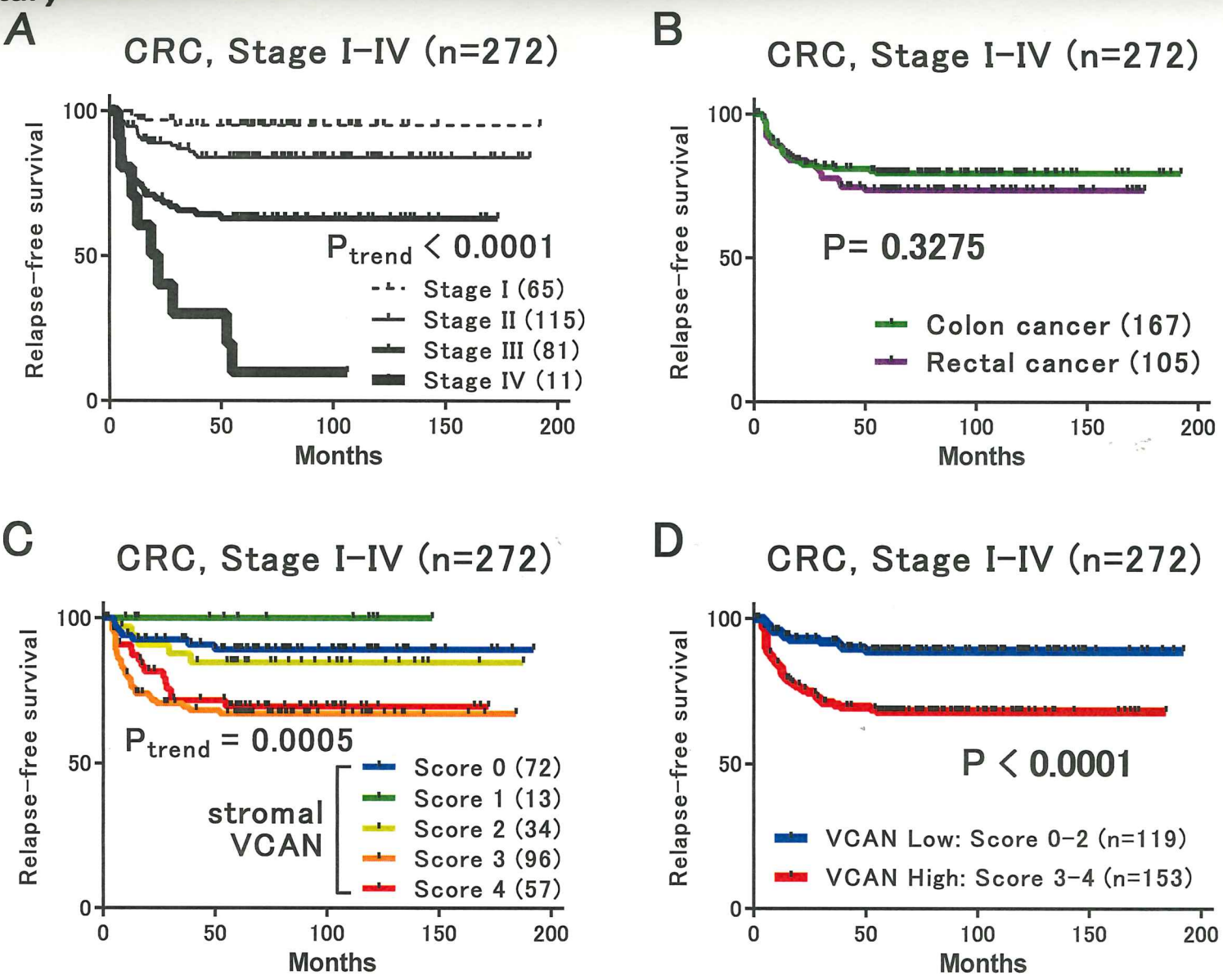
	Total (n=62)	Stromal VCAN expression		*P*
		High n=49 (%)	Low n=13 (%)	
Patterns of recurrence				0.0032
Locoregional	13	6 (12.2)	7 (53.8)	
Distant	49	43 (87.8)	6 (46.2)	

Supplementary Figure 1



6 of 8 genes, including COL4A1, COL4A2, SERPINE1, VCAN, NOTCH3 and GPR116, were found to be upregulated in CRC compared to normal colon, and also associated with poor relapse-free survival in two independent cohorts (GSE41258 and GSE17538). 2 genes downregulated in CRC, including TTC38 and SULT1B1, were negatively associated with poor prognosis.

Supplementary Figure 2



272 patients with stage I-IV CRC with curative resection were available for relapse-free survival (RFS) analysis. Kaplan-Meier curves of RFS stratified by stage (A), and by location (B). RFS curves for subgroups of stage I-IV patients according to stromal VCAN immunoreactivity score is shown (C,D).