Cancer Science

WILEY | JCA JAPANESE CANCER ASSOCIATION

ORIGINAL ARTICLE OPEN ACCESS

Clinical Research

Single-Cell Analysis Identified a Recurrent Malignant-Associated Transcriptional State in Colorectal Cancer

Sadahiro Iwabuchi[1,2] | Kentaro Sawada[3,4] | Motohiro Kojima[5,6] | Miho Ozawa[3] | Hiroya Taniguchi[3,7] | Hideaki Bando[3] | Kazumi Abe[8] | Yuta Kuze[8] | Kiyomi Imamura[8] | Yukie Kashima[8] | Yoshihiko Hirohashi[9] | Atsushi Tajima[2] | Toshihiko Torigoe[9] | Yutaka Suzuki[8] | Takayuki Yoshino[3] | Shinichi Hashimoto[1]

[1]Department of Molecular Pathophysiology, Institute of Advanced Medicine, Wakayama Medical University, Wakayama-shi, Wakayama, Japan | [2]Department of Bioinformatics and Genomics, Graduate School of Medical Sciences, Kanazawa University, Kanazawa-shi, Ishikawa, Japan | [3]Department of Gastroenterology and Gastrointestinal Oncology, National Cancer Center Hospital East, Kashiwa-shi, Chiba, Japan | [4]Department of Medical Oncology, Kushiro Rosai Hospital, Kushiro-shi, Hokkaido, Japan | [5]Pathology Division, Research Center for Innovative Oncology, National Cancer Center Hospital East, Kashiwa-shi, Chiba, Japan | [6]Department of Surgical Pathology, Kyoto Prefectural University of Medicine & Graduate School of Medical Science, Kyoto-shi, Kyoto, Japan | [7]Department of Clinical Oncology, Aichi Cancer Center, Nagoya-shi, Aichi, Japan | [8]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Science, The University of Tokyo, Kashiwa-shi, Chiba, Japan | [9]Department of Pathology, Sapporo Medical University School of Medicine, Sapporo-shi, Hokkaido, Japan

Correspondence: Sadahiro Iwabuchi (iwabuchi@staff.kanazawa-u.ac.jp)

Received: 29 October 2025 | **Revised:** 13 January 2026 | **Accepted:** 20 January 2026

Keywords: colorectal cancer | FOXQ1 | KCNQ1OT1 | RASSF6 | single cell RNA sequencing

ABSTRACT

Colorectal cancer (CRC) exhibits substantial molecular heterogeneity, yet detailed single-cell transcriptomic data from Japanese patients remain limited. Here, we performed single-cell RNA sequencing (scRNA-seq) on tumor tissues and adjacent normal tissues from four Japanese CRC patients to characterize malignant cell populations. We identified a common cancer-associated cell cluster observed across all four cases, characterized by high expression of the long noncoding RNA KCNQ1 opposite strand/antisense transcript 1 (*KCNQ1OT1*). Further analysis revealed a marked decrease in Ras Association Domain Family Member 6 (*RASSF6*) expression within this cluster, despite clinical datasets reporting that elevated *RASSF6* expression correlates with poor overall survival in Western cohorts. Reanalysis of publicly available bulk RNA-seq data from independent East Asian cohorts confirmed that *RASSF6* expression was consistently reduced in tumor tissues compared with adjacent normal colon tissues, while *KCNQ1OT1* expression showed no apparent differences across cohorts. Collectively, our single-cell analysis delineates a recurrent cancer-associated transcriptional state observed within this cohort and highlights distinct gene expression features that may be underappreciated by bulk or population-averaged analyses, underscoring the value of single-cell profiling for refining molecular interpretations of CRC.

1 | Introduction

Epidemiological studies indicate that colorectal cancer (CRC) is the third most diagnosed cancer and the second leading cause of cancer-related deaths worldwide, with an estimated 1.9 million new cases and 930,000 deaths in 2020 [1]. A large-scale analysis of over 80,000 surgically treated CRC patients revealed that women have significantly better 5-year cancer-specific survival than men, despite similar tumor characteristics [2]. Recent efforts in Japanese populations, including the JPHC cohorts, have shown that incorporating GWAS-identified genetic markers into lifestyle-based risk

Abbreviations: CRC, colorectal cancer; OS, overall survival; PFS, progression-free survival; scRNA-seq, single-cell RNA sequencing; UMAP, Uniform Manifold Approximation and projection.

models enhances CRC risk prediction in Japanese men, underscoring the importance of population-specific polygenic risk score (PRS) modeling [3]. GWAS including Japanese cohorts have identified novel susceptibility loci, such as chromosome 10p14, highlighting unique genetic risk architectures in East Asian CRC populations [4]. Although GWAS and PRS studies have improved our understanding of genetic susceptibility to CRC in Japanese populations, many aspects of tumor biology such as transcriptomic differences between tumor and normal tissues or the presence of recurrent cancer-associated transcriptional states, remain unexplored. These features may differ from those observed in Western cohorts.

CRC is a highly heterogeneous malignancy, with wide variation across molecular subtypes, tumor microenvironments, and clinical outcomes. Recent advances in single cell RNA sequencing (scRNA-seq) and spatial transcriptomics have enabled unprecedented resolution of this heterogeneity, uncovering novel epithelial subtypes [5], cancer-associated fibroblast states [6], and immune cell dynamics associated with metastasis and therapy response [7, 8]. However, most of these studies have been conducted in Western or Chinese cohorts [5–12], or employed in vivo or organoid-based models [13, 14], limiting insights into East Asian-and particularly Japanese-CRC biology. Furthermore, large-scale multiomics and polygenic studies have revealed differences in risk allele distribution and functional impact between individuals of European and East Asian ancestry, with susceptibility loci enriched in East Asian cohorts identified in CRC [15, 16]. Ethnic-specific gene–environment interactions have also been reported in other malignancies, such as lung cancer [17]. These findings emphasize the importance of population-specific molecular profiling, including single-cell analyses, to fully understand CRC pathogenesis and progression.

In this study, we present the first scRNA-seq analysis of paired tumor and adjacent normal tissues from four Japanese CRC patients. We identified a recurrent cancer-associated transcriptional state enriched across all cases and discovered several differentially expressed genes that may contribute to CRC progression.

2 | Materials and Methods

2.1 | Patient Characteristics

Tumor and adjacent normal colon tissues were obtained from four Japanese CRC patients who underwent surgical resection at the National Cancer Center Hospital East between 2019 and 2020. All tumors were histologically diagnosed as moderately differentiated tubular adenocarcinomas, with primary lesions located in the sigmoid colon ($n=2$), rectum ($n=2$) (Table 1). The cohort consisted of two female and two male patients, aged 43–63 years. All tumors exhibited invasion beyond the muscularis propria, with pathological T stages ranging from pT3 (subserosal or adventitial invasion). Budding grades ranged from BD1 to BD3. Lymphovascular invasion was present in three of four cases, and venous invasion was observed in all cases (v1a-v3). Lymph node metastasis (N0, pN1a or pN1b) was identified in all patients, and distant metastasis was observed in Patient #2 with liver metastasis (M1a). Serum levels of carcinoembryonic antigen (CEA, ng/mL) and carbohydrate antigen 19–9 (CA19-9, U/mL) were measured preoperatively (approximately 1 month before surgery), on the day of surgery, and longitudinally during postoperative follow-up (up to 36 months). These markers were used to assess recurrence or disease progression.

2.2 | Single Cell Analysis

Detailed protocol for preparation of single cells was based on our previously published method for tissue dissociation in human cancer samples [18]. Briefly, resected tissues were placed on ice

TABLE 1 | Clinicopathological features of four Japanese patients with CRC.

	#1	#2	#3	#4
Age	43	60	47	63
Sex	Female	Female	Male	Male
Tumor location	Sigmoid	Sigmoid	Rectum	Rectum
Pathological diagnosis	Tubular adenocarcinoma	Tubular adenocarcinoma	Tubular adenocarcinoma	Tubular adenocarcinoma
Budding	BD1	BD1	BD3	BD1
Lymph vascular invasion	Ly0	Ly1a	Ly1a	Ly1a
Vascular invasion	V1a	V3	V1a	V1a
Pathological T	T3	T3	T3	T3
Pathological N	N1b	N1a	N1b	N0
Pathological M	M0	M1a	M0	M0
Metastatic organ	None	Liver	None	None
Pathological Stage	IIIb	IVa	IIIb	IIa (UICC 8th)
Tumor diameter (mm)	10	25	86	40

in cold PBS, finely minced, and centrifuged to remove the supernatant. Fragments were incubated in RPMI with 10% FBS and Liberase TH for 30 min at room temperature. The suspension was passed through a 70-μm strainer, and residual aggregates were gently ground using a syringe plunger. Cells were isolated using a two-layer Percoll gradient (20%/60%) and centrifuged at 1000×*g* for 20 min. Then red blood cells were lysed, and viable cells were counted. A total of 10,000 viable cells were then prepared for downstream analysis. The scRNA-seq libraries were constructed using the Chromium Single Cell 3′ Reagent Kits v3 (10x Genomics, Pleasanton, CA, USA) according to the manufacturer's instructions and as previously described in report [19]. Sequencing libraries were constructed and sequenced using the Illumina NextSeq 550 (150-cycle High Output kit) and NovaSeq 6000 (S1 flow cell, 100 cycles) platforms (Illumina, San Diego, CA, USA), with a paired-end configuration of 28 bp (Read 1), 8 bp (i7 index), 8 bp (i5 index), and 90 bp (Read 2).

2.3 | Data Analysis

Processing of raw sequencing data was performed using Cell Ranger v4.0.0 (10x Genomics). FASTQ files were generated using cellranger mkfastq and read alignment and gene counting were performed using cellranger count with default parameters. Quality control was conducted using Cell Ranger's standard cell-calling algorithm. No additional manual filtering based on gene count, UMI count, or mitochondrial read fraction was applied. Mitochondrial read-based filtering was not implemented by default in the Cell Ranger pipeline and was therefore not used in this study. All samples passed standard Cell Ranger quality control criteria. For multisample integration, filtered feature-barcode matrices from eight samples were aggregated using Cell Ranger Aggr with default depth normalization to generate a combined dataset. Analysis shown in Figures 1, 4, and 5 were performed using this

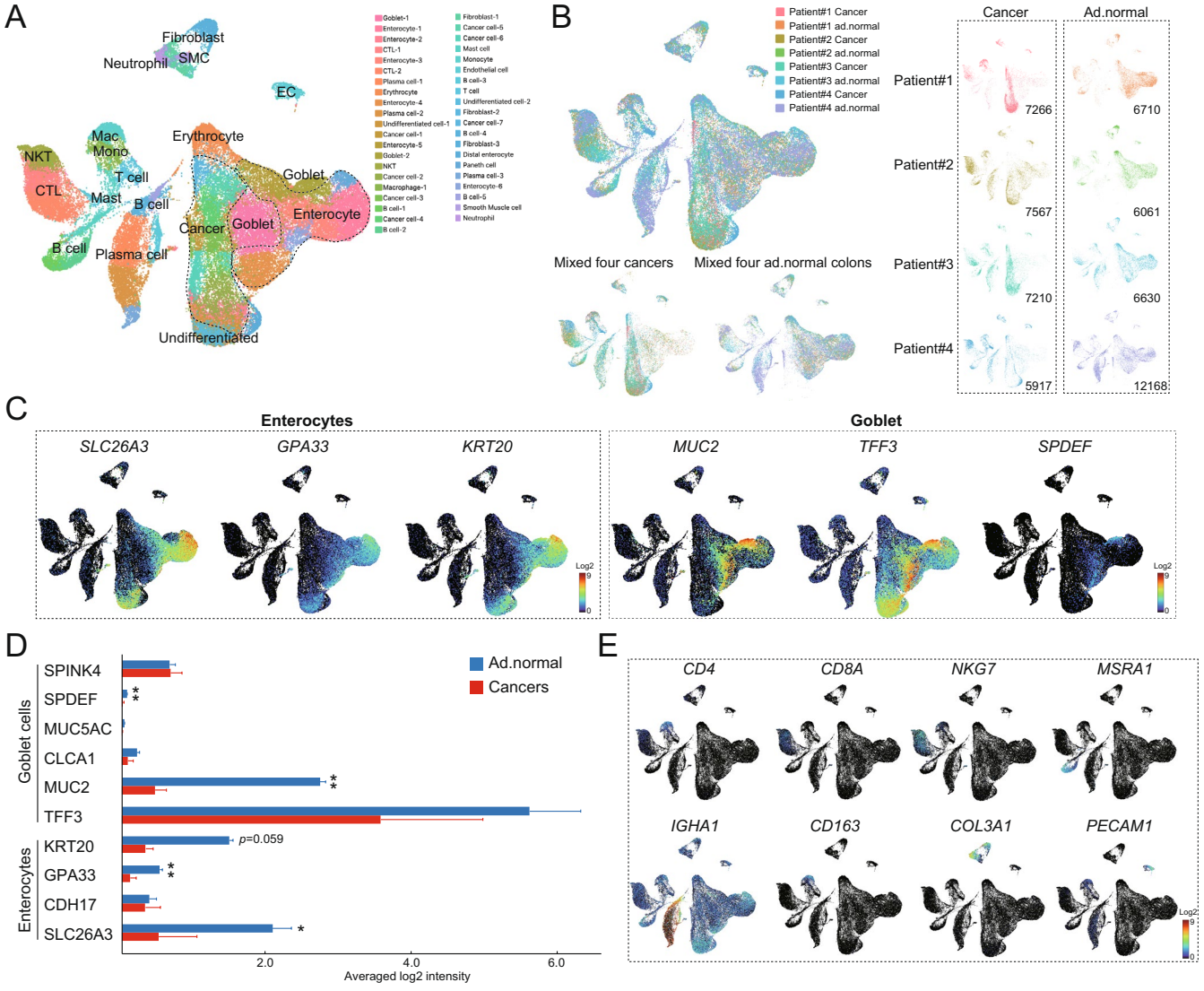
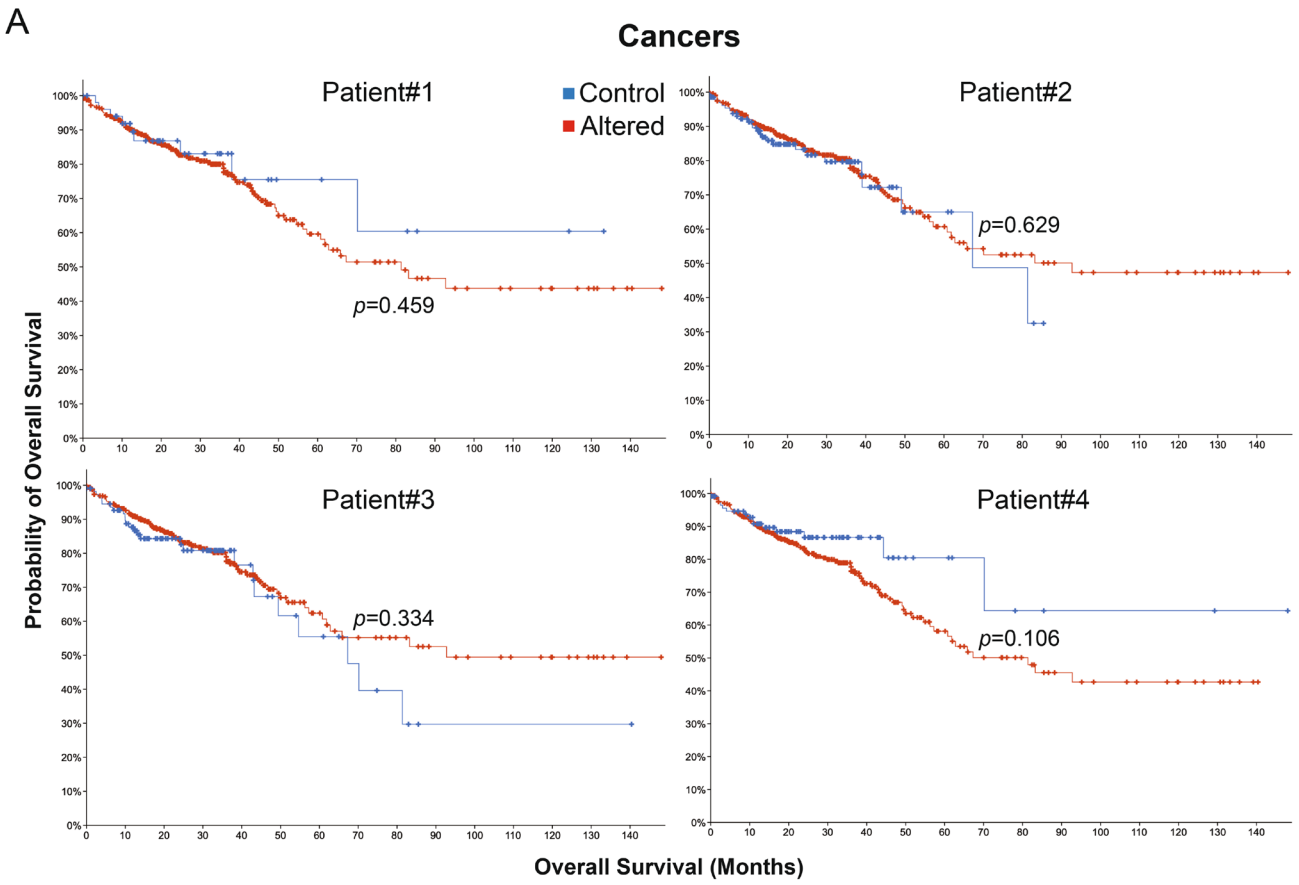


FIGURE 1 | Single-cell landscape of four CRC patients. (A) UMAP showing 41 clusters. CTL, Cytotoxic T-cell; EC, Endothelial cell; Mac, Macrophage; Mast, Mast cell; Mono, Monocyte; NKT, Natural killer T-cell; SMC, Smooth muscle cell. (B) Comparison of adjacent normal colon (ad. normal) vs. tumor tissues per patient. Lower panels show cancer vs. adjacent normal colon across four patients. The numbers shown below each UMAP plot indicate the number of cells included in the analysis. (C) Feature plots of enterocyte/goblet markers. Colors indicate $\log_2$ expression (red: High, blue: Low). (D) Bar plots of averaged $\log_2$ expression in normal (blue) vs. cancer (red) for representative Goblet cells and Enterocytes. $^*p < 0.05$, $^{**}p < 0.01$ (unpaired Student's *t*-test). (E) Feature plots showing expression of nonparenchymal cell marker genes.



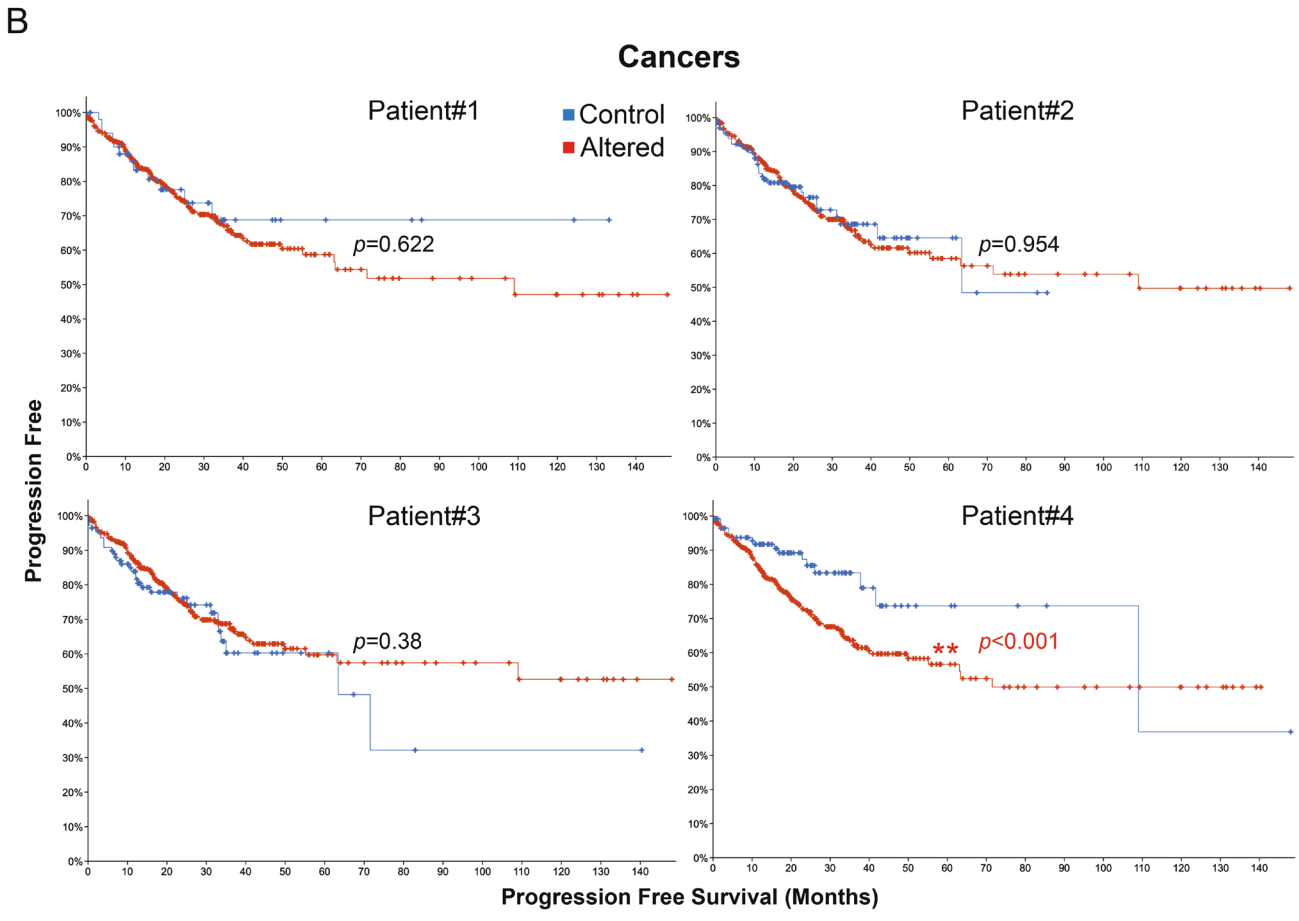


FIGURE 2 | Legend on next page.

FIGURE 2 | Kaplan–Meier survival analysis of patient-specific cancer gene sets. Overall survival (A) and progression-free survival (PFS) (B) showed poor prognosis only in Patient #4 (red: High expression, blue: Control). Example: 49 upregulated genes (e.g., *IGHG4*, *REG1A*, *KCNQ1OT1*) in Patient #4 were applied. *p*-values by log-rank test (**$p < 0.01$).

aggregated dataset. In total, 59,529 cells were retained after aggregation, with a median of 762 genes and 2867 UMIs per cell, and 81.6% of reads mapping to cells. Aggregation using Cell Ranger Aggr performs subsampling to equalize sequencing depth across samples and does not exclude additional cells beyond those filtered during cell calling. Consequently, samples with higher initial sequencing depth, such as Patient #4, exhibited a lower fraction of reads retained after normalization, reflecting depth normalization rather than reduced data quality. Comprehensive quality control metrics for the aggregated scRNA-seq dataset, including read depth, read retention, and normalization statistics, are provided in Table S1. Because Patient #4 was selected for detailed downstream analyses (Figure 6), tumor and adjacent normal colon tissues from this patient were additionally analyzed without aggregation. This separate analysis preserves the original sequencing depth and quality metrics that may be attenuated by depth normalization during aggregation. Quality control metrics for the tumor and adjacent normal colon tissues from Patient #4, including cell numbers, sequencing depth, gene and UMI counts, and mapping statistics, are provided in Tables S2 and S3. Most downstream analyses-including clustering, UMAP visualization [20], and differential expression-were conducted using Loupe Browser v8.1.2 (10x Genomics) with default settings. For trajectory inference, we referenced Monocle3 v1.3.1 following the standard workflow with default parameters. Somatic variant analysis, including large gap read mapping and validation of variants, was performed using CLC Genomics Workbench (v25.0.3; QIAGEN, Aarhus, Denmark) with default settings. Using Loupe Browser, sample-specific gene expression profiles can be extracted for each Patient ID. These profiles can be treated in a pseudo-bulk-like manner by averaging expression values for the genes of interest within each sample. The resulting values were then used to compare expression levels between tumor- and normal-derived cells for statistical analysis.

2.4 | Survival Analysis Using Public Databases

Gene sets specifically upregulated in each patient's tumor tissue were identified based on UMAP analysis using Loupe Browser, which utilizes Cell Ranger's statistical methods. For each cluster, differentially expressed genes were obtained in Loupe Browser by selecting "Run Differential Expression" → "To entire dataset" or "Between selected cluster(s) themselves," depending on the experimental design. The resulting "Differential Expression Output" feature table provided Benjamini-Hochberg adjusted *p*-values, and genes with $p < 0.05$ were defined as significantly enriched. These significantly enriched gene sets were then individually submitted to publicly available survival analysis platforms. Overall survival (OS) and progression-free survival (PFS) were analyzed using the TCGA PanCancer Atlas, Colorectal Adenocarcinoma dataset (mRNA data, RNA Seq v2 RSEM) via cBioPortal for Cancer Genomics v6.3.7 (https://www.cbioportal.org/). In addition, KM plotter v1.0 (http://kmplot.com/analysis/) was used to evaluate the reproducibility of the survival associations. Patient or cluster-specific gene sets were submitted individually, and KM curves were generated to assess the prognostic impact of altered (elevated) versus normal expression levels. Among the enriched genes, some were absent or had insufficient sample numbers in publicly available datasets. These genes were therefore excluded, and the average OS and PFS curves for each cluster were obtained. For individual gene-level survival analyses, the top five genes from each dataset were selected and their OS and PFS outcomes were presented.

2.5 | Statistical Analysis

All statistical analyses were performed using R (v4.2.0). For comparison between two groups (e.g., normal colon vs. CRC), the unpaired Student's *t*-test was used. For comparisons involving three groups (e.g., Figure 7), Welch's one-way ANOVA was applied to account for unequal variances, followed by post hoc pairwise *t*-tests without assuming equal variance. *p*-values less than 0.05 were considered statistically significant, and multiple comparisons were adjusted using Bonferroni correction.

3 | Results

3.1 | Identification of Common Cancer Cell States in Colorectal Cancer Tissue

We obtained 59,529 high-quality cells from four colorectal tumor samples (including both adjacent normal and cancer sites) by scRNA-seq. Cells passing quality control ranged from 5917 to 12,168 per sample. UMAP analysis grouped merged cells into 41 clusters (Figure 1A), with adjacent normal sites showing similar clusters across patients and cancer sites displaying heterogeneity (Figure 1B). Cluster identities were assigned using heat maps (Figures S1, S2 and Table S4) and referenced to published CRC scRNA-seq datasets [7, 10–12, 21]. Enterocyte and goblet cell populations were reduced in cancer sites, while Tregs, monocytes, and stem-like cells were increased (Figures 1C–E and S3).

To investigate whether cancer-specific cellular and molecular features identified by scRNA-seq were associated with clinical outcomes, we performed survival analyses using each sample-specific gene set. Only the cancer tissue-derived gene set from Patient #4 was associated with shorter PFS, whereas no significant association with OS was observed (Figure 2). Among gene sets derived from normal tissues, only the Patient #1-specific set showed a negative correlation with PFS, suggesting a potential protective signature (Figure S4). Postoperative serum markers were also assessed: CEA levels generally decreased across patients, whereas CA19-9 levels exhibited variable patterns (Figure 3). In Patient #4, cancer tissue showed persistent

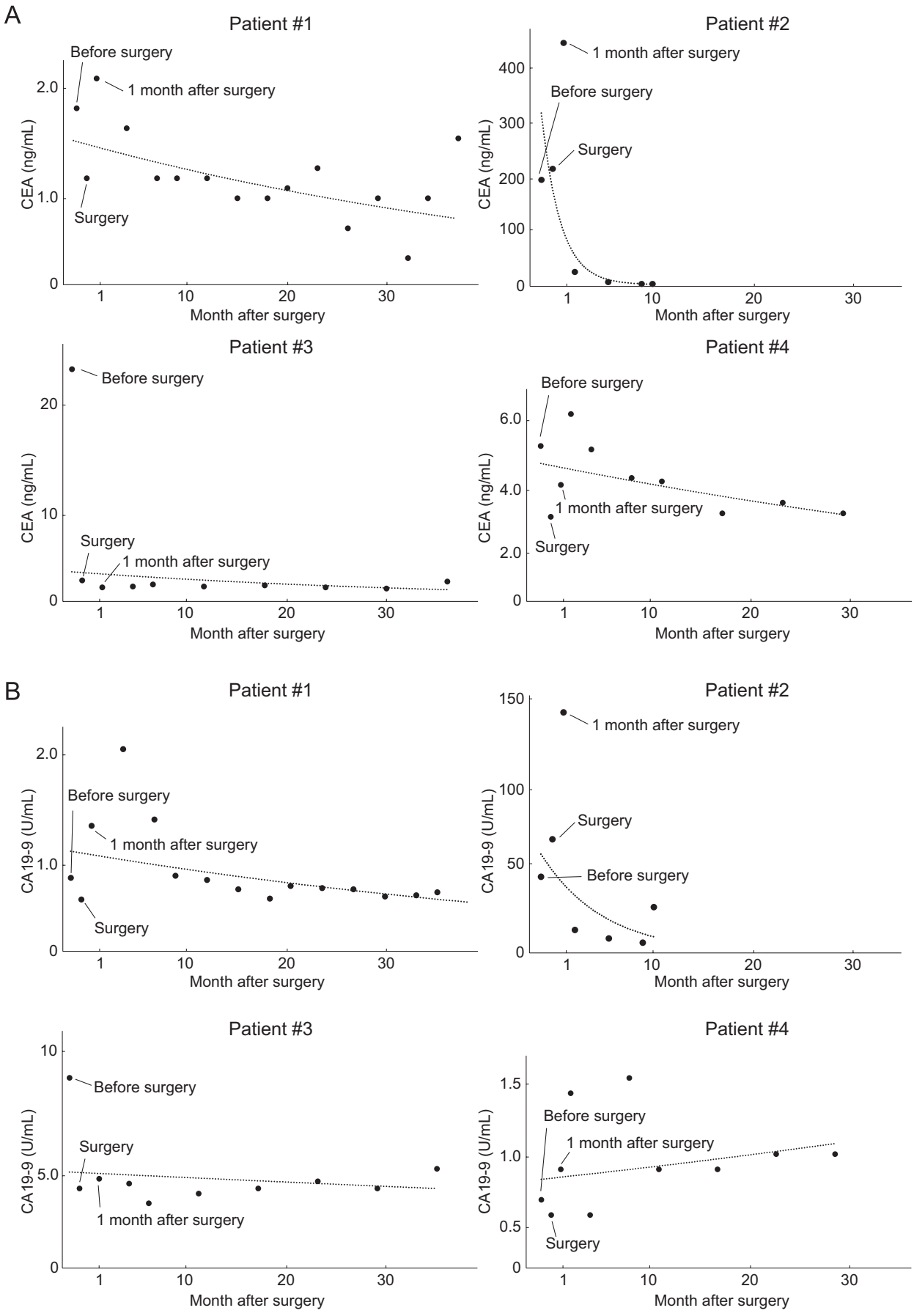


FIGURE 3 | Legend on next page.

FIGURE 3 | Changes of CEA and CA19-9 levels before and after surgery in each patient. The transition of CEA levels (ng/mL) in blood (A) or CA19-9 (U/mL) in blood (B) were shown approximately 1 month preoperatively, on the day of surgery, and from one to a maximum of 36 months postoperatively. The dotted lines in the graph indicated approximate curves, and each black dot represented a respective measurement value.

postoperative CA19-9 elevation, qualitatively reflecting the trend observed for its cancer site-derived gene set and shorter PFS, although this observation should be interpreted with caution.

To further dissect cancer-specific cell states, cells derived from the four tumor sites were reclustered to emphasize cancer-specific clusters. Tumor-derived cells were reclustered to identify six cancer-associated cell states, further refined into eight clusters, including small B-cell-like populations likely reflecting contamination (Figure 4A–E and Figures S5, S6). Patient contributions per cluster were assessed (Figure 4C). Through this stepwise clustering approach, we identified a common cancer cell cluster observed across all four cases, representing a recurrent cancer-associated transcriptional state in tumor tissues. To investigate the origin of the B cell-like populations, we performed pseudo-time trajectory analysis, which suggested that these cells did not arise from cancer cells (Figure S7), indicating that they likely represent nonmalignant contaminating cells rather than a differentiation trajectory. Nevertheless, these cells were retained in the subsequent reanalysis of the common cancer-associated cluster, which included contributions from all four patients. This reanalysis subdivided a common cancer cell cluster into five populations: *FABP1*^{high} cancer cells, goblet cells, KCNQ1 opposite strand/antisense transcript 1 (*KCNQ1OT1*)^{high} cancer cells, necrotic cells, and B-cell-like cells (Figure 5A,B). *KCNQ1OT1* expression was strong in all patients except Patient #3 and particularly high in Patient #4 (Figure S8). Among top-expressed genes, only *KCNQ1OT1* correlated with shorter PFS (Figure 5C,D). In contrast, for the *ANKRD28* highly expressed B-cell-like population, neither the top five genes nor individual gene expression showed any association with PFS (Figure 5E), indicating that this population likely represented nonmalignant contaminating B cells rather than cancer-associated cells. These results demonstrate that scRNA-seq can identify a common cancer cell cluster across patients and highlight the *KCNQ1OT1*^{high} cancer population as a potential contributor to shorter PFS in this cohort.

3.2 | Detailed Single-Cell Profiling of Patient #4 Uncovers Cancer-Associated Transcriptional States Associated With Shorter PFS

Detailed analysis focused on Patient #4, whose cancer tissue-derived cells predominantly constituted the *KCNQ1OT1*^{high} cancer-associated transcriptional state associated with shorter PFS. Because aggregation using Cell Ranger Aggr involves depth normalization that may attenuate sample-specific sequencing characteristics, Patient #4 was additionally analyzed as an independent dataset. This approach allowed more detailed examination of cancer- and adjacent normal-derived cell states while preserving the original sequencing depth and quality metrics (Tables S2, S3). A total of 18,085 cells (12,168 normal, 5917 tumor) were analyzed, with adjacent normal tissue-derived cells aiding accurate identification of tumor-specific cell states. The cells were classified into 26 populations (Figures 6A,B and S9). Notably, a cancer-associated cell cluster transcriptionally similar to tumor cells, designated as "Cancer cell-2," was identified even within adjacent normal tissue (within the dotted square labeled "Reclustering" in Figure 6B). These cells exhibited transcriptional features resembling tumor-derived cells; however, this observation does not imply definitive malignancy and may reflect early tumor-associated transcriptional alterations or microenvironmental influences. Reclustering of this cancer-associated cell cluster yielded five subsets (Figures 6C and S10). The *SAA1*^{high} cells and B cell-like cells were enriched in adjacent normal tissue, whereas *KRT17*^{high}, *UBE2C* ^{high}, and *PLCG2*^{high} cell clusters were predominantly detected in tumor tissue (Figure 6D). These results suggest that adjacent normal areas also harbored cells exhibiting tumor-associated transcriptional features. *KCNQ1OT1* expression was specific to cancer site and nearly absent in adjacent normal regions (Figure 6E–G), confirming that this gene was preferentially expressed in CRC.

To gain further insight into the molecular heterogeneity of this cancer-associated transcriptional staten, we performed reclustering of the *KCNQ1OT1*^{high} cluster as shown in Figure 6G. Six clusters were identified (Figure 6H and Figures S11, S12). Among these, *HOXD9*^{high} and *PROM1*^{high} clusters correlated with prolonged OS (Figures 6I and S13). The *HOXD9*^{high} cells further showed elevated forkhead box Q1 (*FOXQ1*), hes family bHLH transcription factor 1 (*HES1*), Ras Association Domain Family Member 6 (*RASSF6*), and *KCNQ1OT1* expression, all associated with prolonged OS in CRC (Figure 6J). By contrast, *PROM1*^{high} cells expressed cancer-associated genes such as *TM4SF20* and *TIMP1*, but their expression levels were not markedly elevated and no significant correlation for OS (Figures 6J and S14). We next examined whether the representative genes of the *HOXD9*^{high} cluster identified as potentially associated with prolonged OS in Patient #4, were also detectable within the merged dataset comprising cancer cell clusters from all four patients. As shown in Figure S15, feature plot analysis demonstrated partial expression of these genes across the dataset; however, their widespread distribution among diverse cell types did not show clear enrichment within any specific cluster. The *KCNQ1OT1*^{high} population was linked to unfavorable clinical features and prognosis, demonstrating that single-patient scRNA-seq can reveal CRC-related genes missed in multipatient analyses. Clinical data (Table 1) showed some discrepancies, highlighting scRNA-seq's ability to detect molecular features not evident from clinical observation. In addition, variant analysis of all candidate genes in Patient #4 tissue further revealed that only *RASSF6* harbored somatic mutations, which were uniquely presented in tumor tissue. Multiple heterozygous SNVs clustered in exon 1 (chr4:71755224-71778259), suggesting a potential role for *RASSF6* mutations in CRC progression, although functional validation will be required (Figure S16). A schematic overview summarizing the multistep reclustering strategy and dataset transitions was provided in Figure S17.

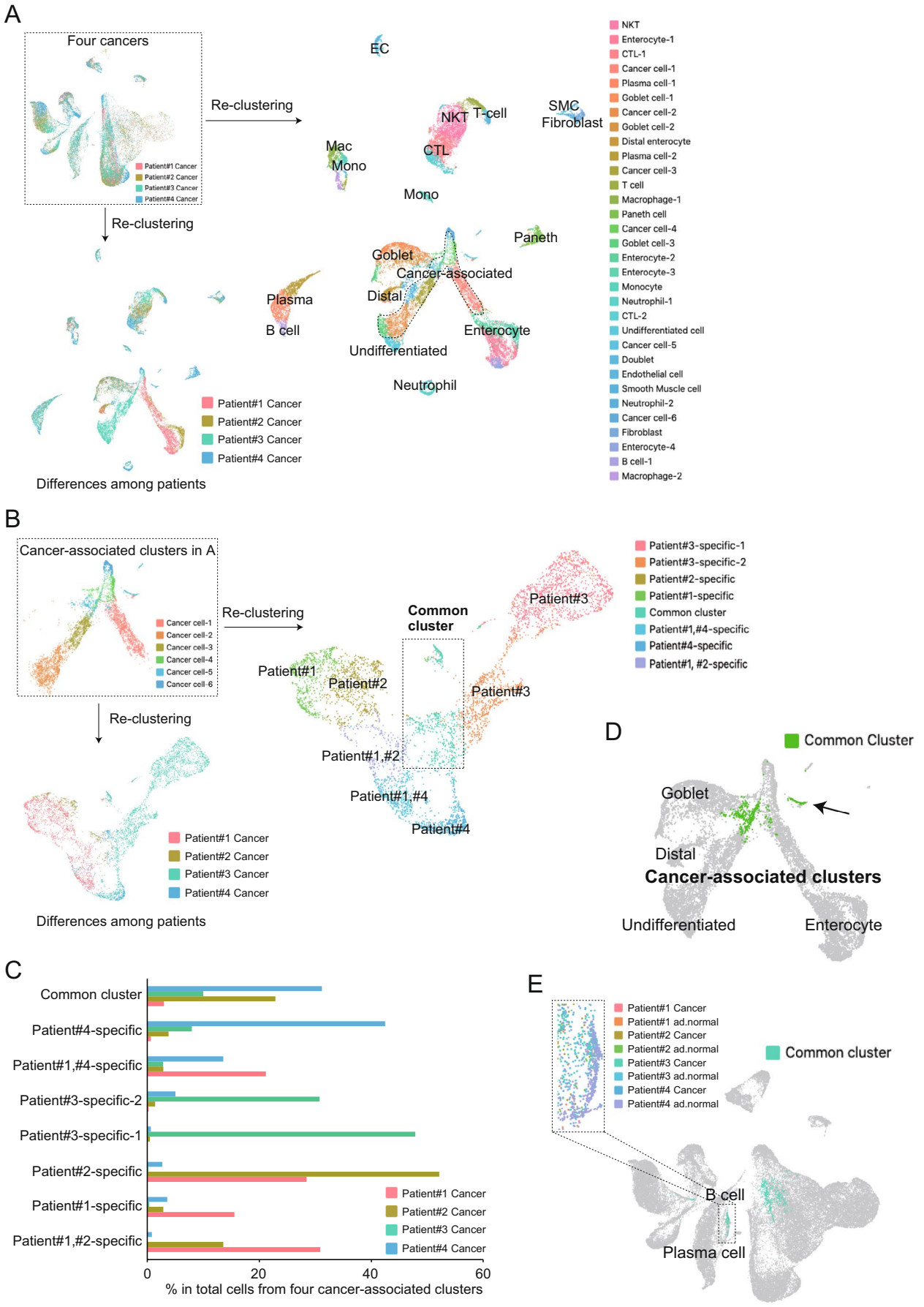


FIGURE 4 | Legend on next page.

FIGURE 4 | Reclustering of tumor-derived cells from four CRC cases. (A) Tumor tissues from four cases were reanalyzed. The upper-left panel shows tumor-derived regions extracted from the eight samples (including normal tissues in Figure 1). The lower-left panel shows UMAPs colored by patient after reanalysis. The right panel shows the 32 identified cell clusters (Distal = distal enterocyte, Plasma = plasma cell). (B) After reanalysis of the four tumor tissues, only cancer cell clusters were further reclustered. The lower-left shows UMAPs colored by sample, and the right shows the eight resulting clusters. The dotted outline marks a common cancer-associated cluster defined by shared distribution pattern. (C) Proportion of cancer clusters within total cells. Colors indicate Patient #1 (red), #2 (yellow-ocher), #3 (green), and #4 (sky blue). Values sum to 100% per patient to account for differences in total cell number. Patient #4 showed the highest proportion in its patient-specific cluster, followed by the common cancer-associated cluster. (D) Spatial location of the common cancer-associated cluster in the reclustered tumor datasets (highlighted in green). (E) Location of the common cancer-associated cluster mapped back to the original dataset of four adjacent normal (ad. normal) and tumor tissues (green). The inset (upper left) shows an enlarged view of the B-cell-like clusters. Colors reflect annotated clusters.

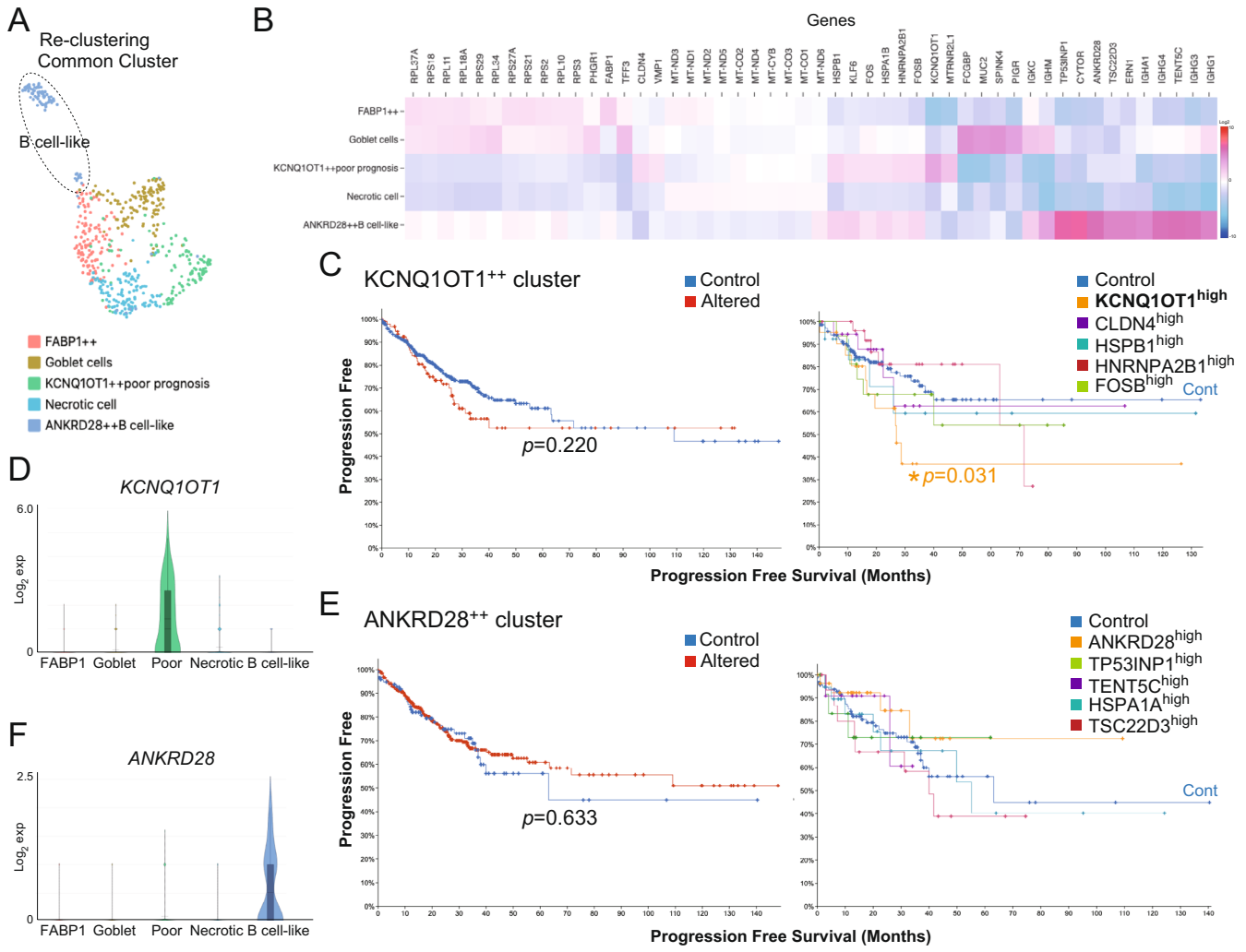


FIGURE 5 | Reclustering of a common cancer-associated cluster identifies a cell state associated with clinical outcome in CRC. (A) The common cancer-associated cluster was divided into five cell populations. (B) Heatmap of representative genes in each cluster. The name of cluster was decided by the gene expression profile. Left color bar shows the higher to lower expression levels in each gene. (C) PFS in clinical study using five marker genes of *KCNQ1OT1* cluster was analyzed by using Human Protein Atlas database. The averaged PFS of five genes did not have any significance. Only higher *KCNQ1OT1* expressed CRC patients (yellow line) tended to be poor prognosis of CRC. (D) The violin plot of *KCNQ1OT1* genes in the five cell clusters. The *Y*-axis indicates the $\log_2$ scale of intensity for gene expression (E) PFS using five marker genes of higher *ANKRD28* expressed B-cell-like cluster. No significant differences in PFS were observed compared with the reference CRC group. (F) The violin plot of *ANKRD28* expression in five cell clusters.

3.3 | Validation of Candidate Biomarker Genes Using External Bulk RNA-Seq Datasets

To validate the expression of candidate genes identified in our scRNA-seq analysis, we first performed a pseudo-bulk analysis by aggregating cells from four patients. For each patient, tumor and adjacent normal tissue-derived cells were averaged to generate eight pseudo-bulk samples. Expression levels of selected genes were compared between tumor and adjacent normal tissue pseudo-bulk samples (Figure 7A,B). Among these, *RASSF6*

expression was significantly lower in cancer tissue, whereas *HES1* expression was higher. *RASSF6* expression was consistently lower in tumor tissues across patients (Figure 7B), contrasting previous reports linking higher expression with poor clinical outcomes. To confirm the reproducibility of these findings, we examined publicly available bulk RNA-seq datasets based on availability of raw FASTQ data, CRC samples, and Asian patient cohorts. Analyzed datasets included GSE142279 (10 paired normal and tumor samples), GSE156451 (72 paired samples [22, 23]), and GSE158559 (18 paired samples including

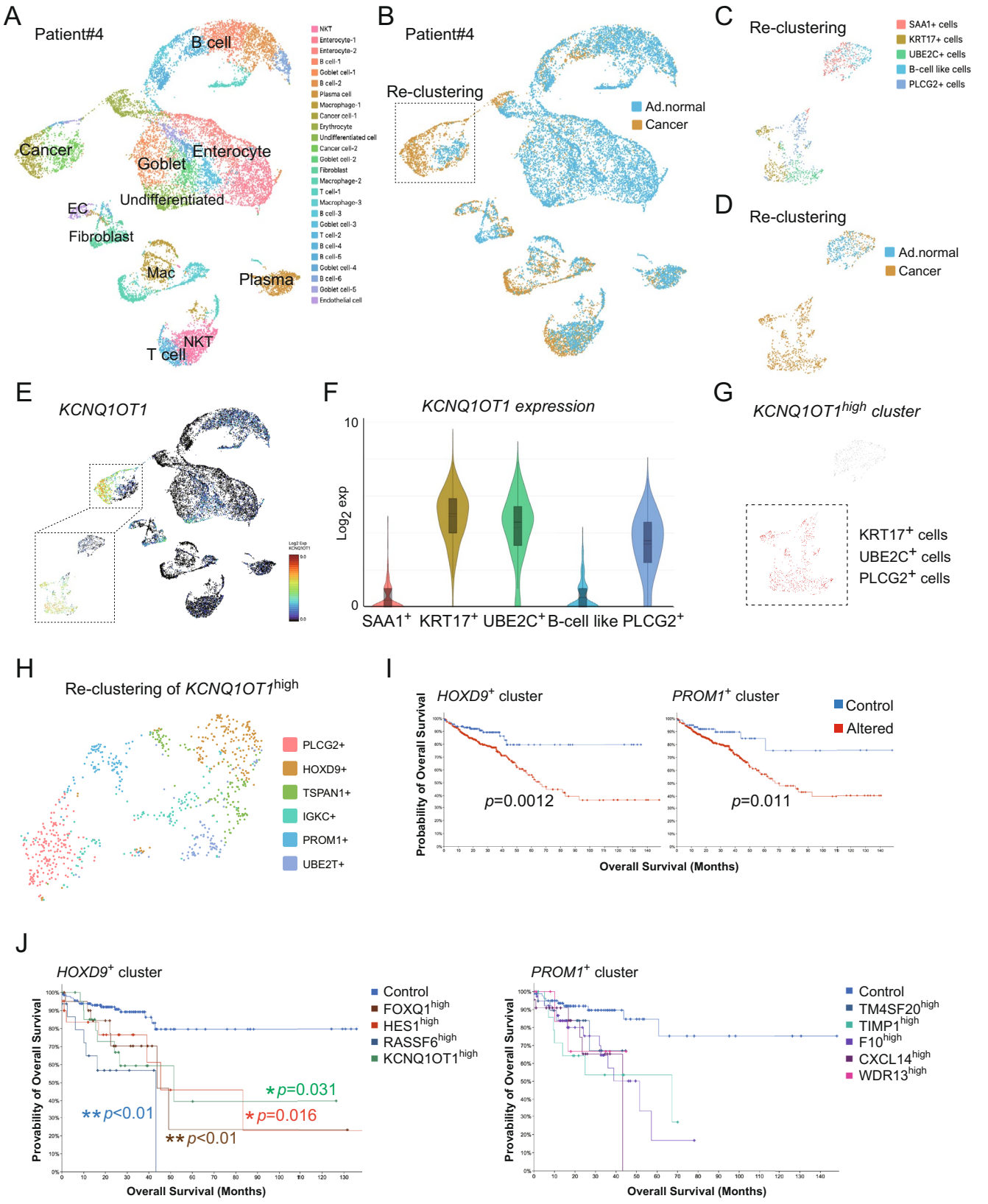


FIGURE 6 | Legend on next page.

FIGURE 6 | Single-cell landscape of CRC Patient #4. (A) A total of 26 clusters were identified. Cluster names are shown in the inset on the right. (B) Cell-type annotation in normal vs. tumor colon tissues for panel A. The area outlined by a dotted square indicates the cancer cell population. Ad. normal, Adjacent normal colon tissue. (C) Reclustering of the cancer-associated cell cluster identified in A, yielding five subpopulations. (D) Cell-type annotation of the five subpopulations in adjacent normal vs. tumor tissues. (E) Feature plot of *KCNQ1OT1* expression in patient #4. Colors represent $\log_2$ expression (inset, right). The dotted square indicates the cancer cell population, with reclustering shown in the inset. (F) Violin plot of *KCNQ1OT1* expression across the five reclustered subpopulations. *Y*-axis shows $\log_2$ values. (G) Distribution of *KCNQ1OT1* expression within the cancer-associated cluster (corresponding to the dotted region in E). (H) Reclustering of *KCNQ1OT1*^{high} cells (*KRT17*$^{+}$, *UBE2C*$^{+}$, *PLCG2*$^{+}$), producing six subpopulations. (I) Overall survival analysis in a clinical cohort using the top five marker genes in the *HOXD9*/*PROM1* cluster. The average OS of these five genes showed significance compared to control. (J) Progression-free survival using representative markers of the *HOXD9*^{high} or *PROM1*^{high} clusters. PFS differed significantly compared with the reference CRC group (*$p < 0.05$, **$p < 0.01$ vs Control).

adjacent normal colon tissue, primary CRC, and liver metastases [24]). Raw counts were normalized to TPM values, and expression levels of candidate genes were compared between tumor and adjacent normal tissues. Stage-specific expression analysis was performed using the independent Korean CRC cohort (KAP2306082). The expression of *KCNQ1OT1*, *FOXQ1*, and *HES1* was consistently upregulated in tumor tissues across datasets, supporting their relevance in CRC. In contrast, *RASSF6* showed a consistent trend of downregulation in tumor samples. Notably, in the metastatic setting (Figure 7E), only *HOXD9* and *RASSF6* were significantly downregulated in liver metastases compared to primary tumors and adjacent normal tissues. Finally, stage-specific expression analysis using an independent Korean CRC cohort (KAP2306082) [8] showed no significant differences in the expression of the selected genes across disease stages. While *FOXQ1* expression was significantly higher in Stage III compared to Stage II, no difference was observed between Stage III and Stage IV. These results suggest that, in the analyzed cohorts, the expression of these genes is not clearly associated with CRC stage progression (Figure 7F).

4 | Discussion

The scRNA-seq of four CRC and adjacent normal tissues from Japanese patients revealed a recurrent cancer-associated transcriptional state observed across all cases. Focused analysis identified five representative genes: *KCNQ1OT1*, *FOXQ1*, *HES1*, *HOXD9*, and *RASSF6*, enriched in specific cancer-associated cell states and associated with clinical outcome-associated features in our dataset and publicly available bulk RNA-seq data. Interestingly, while Western cohorts reported that high *RASSF6* expression correlates with poorer OS, *RASSF6* was suppressed in the analyzed East Asian cohorts.

The long noncoding RNA *KCNQ1OT1* interacts with microRNAs and chromatin-modifying complexes to regulate gene expression in CRC tissues [25–28]. It is consistently upregulated in CRC and linked to tumor size and clinical stage [25, 26], and promotes proliferation and metastasis via the PI3K/AKT pathway, while knockdown induces apoptosis and cell cycle arrest [26, 27]. In our dataset, *KCNQ1OT1* was highly expressed in Patient #4's tumor, who had the poorest predicted OS (Figure 6E), consistent with previous reports from East Asian cohorts [26, 28]. However, *KCNQ1OT1* was also detected in nonmalignant cells such as plasma cells, goblet cells, and enterocytes, raising questions about biomarker specificity. Public datasets showed no significant increase in liver metastases (Figure 7E), suggesting a limited role in metastasis. These results indicate that, although KCNQ1OT1 plays a role in CRC biology, its clinical utility as a diagnostic or therapeutic target may be constrained by non-specific expression. In addition, aggregation-based analyses enabled the identification of recurrent cancer-associated transcriptional states across patients in this study. At the same time, sample-level reanalysis, as performed for Patient #4, allowed more detailed characterization of these states by preserving sequencing depth and quality features that may be normalized during aggregation.

FOXQ1 is a protumorigenic transcription factor in CRC, promoting growth, angiogenesis, and apoptosis resistance via p21 regulation [29]. Pan-cancer analyses further link *FOXQ1* with increased mutational burden, microsatellite instability, immune infiltration, and metastasis [30, 31]. In our dataset, *FOXQ1* was enriched in a cancer-associated transcriptional state and showed higher expression in tumor sites ($p = 0.06$; Figure 7A). Public RNA-seq data confirmed up to 150-fold increases in tumors and higher expression in liver metastases ($p = 0.12$). These findings align with prior reports linking FOXQ1 to epithelial-to-mesenchymal transition and metastatic competence in CRC, supporting its potential as a biomarker across the analyzed cohorts. Importantly, there was no strong evidence of cohort-dependent differences (i.e., no divergence between Asian and non-Asian cohorts), suggesting FOXQ1 may serve as a broadly applicable CRC biomarker.

HES1, a Notch downstream effector, promotes CRC invasion via STAT3-MMP14 and its high expression correlates with metastasis and poor survival [32], while knockdown induces senescence and reduces invasiveness. HES1 also interacts with lncRNAs such as GNAS-AS1 to sustain stemness [33], and our previous organoid study showed HES1 as a key regulator of early cluster growth [14]. However, in the present cohort (Figure 7E), *HES1* expression did not differ between adjacent normal, primary, and metastatic sites, suggesting context-dependent roles and that *HES1* elevation may not universally correlate with metastasis in the analyzed cohorts.

RASSF6 has been identified as a tumor suppressor gene in several cancers, including CRC [34, 35]. However, analysis of TCGA CRC cohorts revealed that higher *RASSF6* expression correlates with poorer OS, representing a seemingly contradictory finding. In our case, *RASSF6* expression was consistently downregulated in cancer cell clusters compared to adjacent normal tissues. This trend was further validated by bulk RNA-seq data from independent cohorts, showing significant downregulation in both

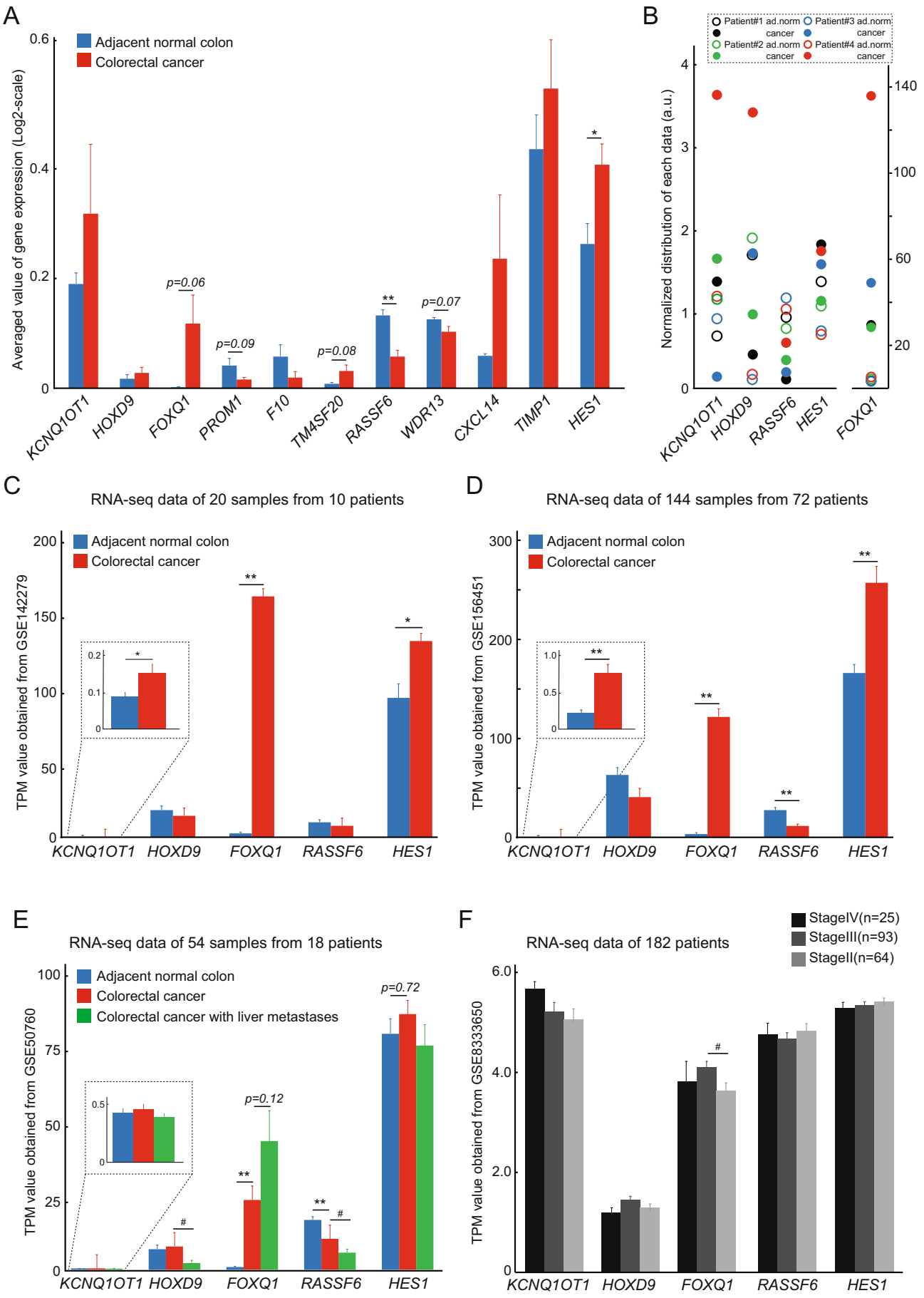


FIGURE 7 | Legend on next page.

FIGURE 7 | Validation of selected genes using scRNA-seq and publicly available bulk RNA-seq datasets in CRC. (A) Pseudo-bulk analysis of scRNA-seq data from four colorectal cancer patients (same dataset used in scRNA-seq). The average expression levels (mean ± SE) of selected genes are shown for tumor (red) and adjacent normal tissue (blue). Gene expression values were $\log_2$-transformed. Statistical significance was assessed using a two-tailed Student's *t*-test (*$p < 0.05$, **$p < 0.01$). (B) Distribution plots of representative five genes in paired normal (open) and tumor (filled) regions across the four patients. Each patient is color-coded (black, green, blue, and red). (C) Bulk RNA-seq analysis of 10 colorectal cancer patients (normal and tumor tissues) retrieved from GSE142279. The expression levels (mean ± SE) of the five selected genes are shown. *$p < 0.05$, **$p < 0.01$ by Student's *t*-test. (D) Bulk RNA-seq data of 72 patients (normal and tumor samples, total $n = 144$) from GSE156451. Expression levels of the same five genes are shown. Statistical significance is indicated as $p < 0.01$. (E) Bulk RNA-seq analysis of 54 samples (normal, primary tumor, and liver metastasis) from 18 patients in GSE158559. Blue: Normal; Red: Primary tumor; Green: Liver metastasis. Statistical analysis was performed using Welch's *t*-test. **$p < 0.01$ vs. normal, #$p < 0.05$ vs. cancer. (F) Stage-specific analysis of five genes using RNA-seq data from Korean colorectal cancer cohort (KAP2306082). Samples were categorized by cancer stage (II, III, IV) and plotted as mean ± SE. Welch's *t*-test was used for statistical comparison. #$p < 0.05$ vs. Stage III.

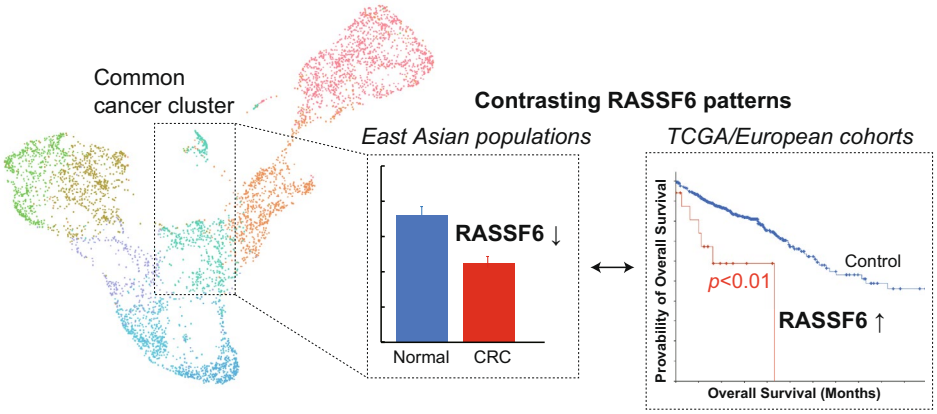


FIGURE 8 | Schematic overview of key cohort-associated findings revealed by single-cell analysis of CRC. This schematic summarizes the main findings of the study, including the identification of a recurrent cancer-associated transcriptional state and consistent downregulation of *RASSF6* within cancer cell state. The figure highlights the contribution of single-cell transcriptomic analysis in resolving cell state-specific molecular alterations that may be obscured in bulk tissue analyses. This schematic also serves as the Graphical Abstract.

primary tumors and metastatic lesions (Figure 7E). Previous studies support the tumor-suppressive role of RASSF6. Chen et al. reported somatic loss-of-function mutations in RASSF6 in sporadic CRC patients and demonstrated that *RASSF6* overexpression inhibited CRC cell proliferation, migration, and invasion, whereas its knockdown promoted malignant phenotypes [34]. Zhou et al. [35] elucidated that RASSF6 suppresses CRC progression by inhibiting the Wnt/β-catenin signaling pathway and EMT, with overexpression leading to increased E-cadherin and decreased mesenchymal markers such as Snail and matrix metalloproteinases, thus reducing metastatic potential. Our mutation analysis of matched tumor and adjacent normal tissue scRNA-seq data revealed a tumor-specific SNVs in exon 1 of *RASSF6* (Figure S15), suggesting that its downregulation may be driven not only by transcriptional repression but also by genetic inactivation. Since exon 1 encodes part of the Ras-association domain, such mutations likely result in loss of RASSF6 function. Interestingly, while TCGA data (predominantly Western populations) suggest that high *RASSF6* expression correlates with poor prognosis, the present cohort shows the opposite trend. This discrepancy may reflect population-specific differences in genetic or epigenetic regulation, or environmental influences. An important consideration in interpreting these findings is the difference in cellular resolution between bulk and single-cell transcriptomic analyses. Publicly available datasets such as TCGA and other bulk RNA-seq studies are derived from tumor tissues, which comprise a mixture of malignant epithelial cells and diverse nonmalignant components, including immune and stromal cells. As a result, gene expression signals specific to cancer cells may be masked or diluted in bulk analyses. In contrast, our single-cell approach enabled the isolation of cancer cell-enriched clusters, allowing direct comparison of *RASSF6* expression between malignant epithelial cells and normal epithelial counterparts. This analysis revealed consistent downregulation of *RASSF6* specifically within cancer cell states, a feature that may not have been readily apparent in bulk tumor-level analyses. Notably, reanalysis of independent bulk RNA-seq datasets from East Asian cohorts, although also derived from tumor tissues, showed concordant downregulation of *RASSF6*, supporting the robustness of this observation. These findings suggest that the apparent discrepancy between prior TCGA-based survival associations and the tumor-suppressive role of *RASSF6* may, at least in part, reflect differences in cellular composition and analytical resolution rather than true biological contradiction. Our results highlight how single-cell transcriptomic profiling can uncover clinically relevant cancer cell-intrinsic alterations that have remained underappreciated in existing bulk transcriptomic resources. To summarize these findings and highlight the key cohort-associated transcriptional features of CRC revealed by our study, we present a schematic overview in Figure 8, which also serves as the Graphical Abstract.

In summary, our findings demonstrate consistent downregulation of *RASSF6* in the present cohort, accompanied by

mutational evidence, and underscore the importance of single-cell transcriptomics in revealing cell state-specific molecular alterations. RASSF6 may represent a promising biomarker candidate, warranting further validation across independent cohorts.

Author Contributions

Sadahiro Iwabuchi: conceptualization, formal analysis, methodology, software, visualization, writing – original draft. **Kentaro Sawada:** data curation, investigation, resources, writing – original draft. **Motohiro Kojima:** conceptualization, resources, supervision. **Miho Ozawa:** data curation, investigation, resources. **Hiroya Taniguchi:** conceptualization, data curation, investigation, resources. **Hideaki Bando:** data curation, investigation, resources. **Kazumi Abe:** data curation, investigation, resources. **Yuta Kuze:** investigation, methodology, software. **Kiyomi Imamura:** investigation, methodology, software. **Yukie Kashima:** methodology, software, supervision. **Yoshihiko Hirohashi:** project administration, resources, supervision, writing – review and editing. **Atsushi Tajima:** software, supervision, writing – review and editing. **Toshihiko Torigoe:** project administration, resources, supervision, writing – review and editing. **Yutaka Suzuki:** conceptualization, methodology, project administration, software, supervision, writing – review and editing. **Takayuki Yoshino:** conceptualization, funding acquisition, project administration, resources, supervision, writing – review and editing. **Shinichi Hashimoto:** conceptualization, funding acquisition, project administration, supervision, writing – original draft, writing – review and editing.

Acknowledgments

We sincerely thank the patients and the staff at the National Cancer Center East Hospital for their cooperation in sample collection. In addition, we gratefully acknowledge the valuable technical assistance provided by Yu Kamide and Saki Kurokawa.

Funding

This work was supported by the CREST program of the Japan Science and Technology Agency (JST), Japan, under the research area "Development of integrated single-cell analysis platforms to elucidate biological control mechanisms," with the project title "Construction of tissue microenvironment information by single-cell gene expression analysis." The project was led by Shinichi Hashimoto and funded from FY2015 to FY2020.

Ethics Statement

This study was approved by the Institutional Review Board of the National Cancer Center, Japan on January 22, 2019 (Research Protocol No. 2018-345). The study was conducted in accordance with the Declaration of Helsinki.

Consent

All patients provided written informed consent prior to sample collection.

Conflicts of Interest

Dr. Toshihiko Torigoe is an Editorial Board Member of Cancer Science. Other authors declare no conflicts of interest.

Data Availability Statement

Due to patient privacy and the small sample size, raw scRNA-seq FASTQ files cannot be publicly released, as clinical information could potentially identify individuals and patient consent did not include germline disclosure. A curated subset of processed data will be made available upon reasonable request to the corresponding author. A list of representative marker genes in each patient's tissue was provided in Table S4.

References

1. E. Morgan, M. Arnold, A. Gini, et al., "Global Burden of Colorectal Cancer in 2020 and 2040: Incidence and Mortality Estimates From GLOBOCAN," *Gut* 72 (2023): 338–344.

2. K. Kotake, M. Asano, H. Ozawa, H. Kobayashi, and K. Sugihara, "Gender Differences in Colorectal Cancer Survival in Japan," *International Journal of Clinical Oncology* 21 (2016): 194–203.

3. M. Iwasaki, S. Tanaka-Mizuno, A. Kuchiba, et al., "Inclusion of a Genetic Risk Score Into a Validated Risk Prediction Model for Colorectal Cancer in Japanese Men Improves Performance," *Cancer Prevention Research* 10 (2017): 535–541.

4. Y. Takahashi, K. Sugimachi, K. Yamamoto, et al., "Japanese Genome-Wide Association Study Identifies a Significant Colorectal Cancer Susceptibility Locus at Chromosome 10p14," *Cancer Science* 108 (2017): 2239–2247.

5. F. Wang, J. Long, L. Li, et al., "Single-Cell and Spatial Transcriptome Analysis Reveals the Cellular Heterogeneity of Liver Metastatic Colorectal Cancer," *Science Advances* 9 (2023): eadf5464.

6. H. Zheng, H. Liu, Y. Ge, and X. Wang, "Integrated Single-Cell and Bulk RNA Sequencing Analysis Identifies a Cancer Associated Fibroblast-Related Signature for Predicting Prognosis and Therapeutic Responses in Colorectal Cancer," *Cancer Cell International* 21 (2021): 552.

7. Q. Zhang, Y. Liu, X. Wang, C. Zhang, M. Hou, and Y. Liu, "Integration of Single-Cell RNA Sequencing and Bulk RNA Transcriptome Sequencing Reveals a Heterogeneous Immune Landscape and Pivotal Cell Subpopulations Associated With Colorectal Cancer Prognosis," *Frontiers in Immunology* 14 (2023): 1184167.

8. J. Sheng, N. Li, S. Li, et al., "A Dynamic Molecular Landscape in Colorectal Cancer Progression at Single-Cell Resolution," *Journal of Translational Medicine* 23 (2025): 723.

9. M. Uhlen, P. Oksvold, L. Fagerberg, et al., "Towards a Knowledge-Based Human Protein Atlas," *Nature Biotechnology* 28 (2010): 1248–1250.

10. D. Chisanga, S. Keerthikumar, M. Pathan, et al., "Colorectal Cancer Atlas: An Integrative Resource for Genomic and Proteomic Annotations From Colorectal Cancer Cell Lines and Tissues," *Nucleic Acids Research* 44 (2016): D969–D974.

11. S. Bian, Y. Hou, X. Zhou, et al., "Single-Cell Multiomics Sequencing and Analyses of Human Colorectal Cancer," *Science* 362 (2018): 1060–1063.

12. R. Wang, J. Li, X. Zhou, et al., "Single-Cell Genomic and Transcriptomic Landscapes of Primary and Metastatic Colorectal Cancer Tumors," *Genome Medicine* 14 (2022): 93.

13. R. Wang, Y. Mao, W. Wang, et al., "Systematic Evaluation of Colorectal Cancer Organoid System by Single-Cell RNA-Seq Analysis," *Genome Biology* 23 (2022): 106.

14. Y. K. Lin, R. Coppo, K. Onuma, et al., "Growth Pattern of De Novo Small Clusters of Colorectal Cancer Is Regulated by Notch Signaling at Detachment," *Cancer Science* 115 (2024): 3648–3659.

15. C. Fernandez-Rozadilla, M. Timofeeva, Z. Chen, et al., "Deciphering Colorectal Cancer Genetics Through Multi-Omic Analysis of 100,204 Cases and 154,587 Controls of European and East Asian Ancestries," *Nature Genetics* 55 (2023): 89–99.

16. J. Xin, M. Du, D. Gu, et al., "Risk Assessment for Colorectal Cancer via Polygenic Risk Score and Lifestyle Exposure: A Large-Scale

Association Study of East Asian and European Populations," *Genome Medicine* 15 (2023): 4.

17. M. Zhu, J. Lv, Y. Huang, et al., "Ethnic Differences of Genetic Risk and Smoking in Lung Cancer: Two Prospective Cohort Studies," *International Journal of Epidemiology* 52 (2023): 1815–1825.

18. S. Iwabuchi and S. Hashimoto, "Single-Cell Transcriptome Analysis in Tumor Tissues," in *Transcriptome Analysis*, ed. M. Blumenberg (IntechOpen, 2019).

19. G. X. Zheng, J. M. Terry, P. Belgrader, et al., "Massively Parallel Digital Transcriptional Profiling of Single Cells," *Nature Communications* 8 (2017): 14049.

20. E. Becht, L. McInnes, J. Healy, et al., "Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP," *Nature Biotechnology* 37 (2018): 38–44.

21. R. Chai, Y. Zhao, Z. Su, and W. Liang, "Integrative Analysis Reveals a Four-Gene Signature for Predicting Survival and Immunotherapy Response in Colon Cancer Patients Using Bulk and Single-Cell RNA-Seq Data," *Frontiers in Oncology* 13 (2023): 1277084.

22. Q. L. Li, X. Lin, Y. L. Yu, et al., "Genome-Wide Profiling in Colorectal Cancer Identifies PHF19 and TBC1D16 as Oncogenic Super Enhancers," *Nature Communications* 12 (2021): 6407.

23. X. Lin, J. D. Chen, C. Y. Wang, et al., "Cooperation of MLL1 and Jun in Controlling H3K4me3 on Enhancers in Colorectal Cancer," *Genome Biology* 24 (2023): 268.

24. X. Zhu, X. Tian, L. Ji, et al., "A Tumor Microenvironment-Specific Gene Expression Signature Predicts Chemotherapy Resistance in Colorectal Cancer Patients," *npj Precision Oncology* 5 (2021): 7.

25. Y. Bian, G. Gao, Q. Zhang, et al., "KCNQ1OT1/miR-217/ZEB1 Feedback Loop Facilitates Cell Migration and Epithelial-Mesenchymal Transition in Colorectal Cancer," *Cancer Biology & Therapy* 20 (2019): 886–896.

26. Q. Duan, L. Cai, K. Zheng, et al., "lncRNA KCNQ1OT1 Knockdown Inhibits Colorectal Cancer Cell Proliferation, Migration and Invasiveness via the PI3K/AKT Pathway," *Oncology Letters* 20 (2020): 601–610.

27. G. Liu, L. Shi, B. Wang, et al., "Role of Oncogenic Long Noncoding RNA KCNQ1OT1 in Colon Cancer," *Oncology Research* 32 (2024): 585–596.

28. L. Wang, K. B. Cho, Y. Li, G. Tao, Z. Xie, and B. Guo, "Long Noncoding RNA (lncRNA)-Mediated Competing Endogenous RNA Networks Provide Novel Potential Biomarkers and Therapeutic Targets for Colorectal Cancer," *International Journal of Molecular Sciences* 20 (2019): 5758.

29. H. Kaneda, T. Arao, K. Tanaka, et al., "FOXQ1 Is Overexpressed in Colorectal Cancer and Enhances Tumorigenicity and Tumor Growth," *Cancer Research* 70 (2010): 2053–2063.

30. G. Pizzolato, L. Moparthi, S. Soderholm, C. Cantu, and S. Koch, "The Oncogenic Transcription Factor FOXQ1 Is a Differential Regulator of Wnt Target Genes," *Journal of Cell Science* 135 (2022): jcs260082.

31. W. D. Yang, Z. H. Zhang, M. Y. Zhao, et al., "P300-Dependent Acetylation of the FOXQ1 Complex Activates Super-Enhancers to Promote Colorectal Cancer Proliferation and Metastasis," *Communications Biology* 8 (2025): 1016.

32. M. T. Weng, P. N. Tsao, H. L. Lin, et al., "Hes1 Increases the Invasion Ability of Colorectal Cancer Cells via the STAT3-MMP14 Pathway," *PLoS One* 10 (2015): e0144322.

33. Y. Zhang, L. Zheng, X. Lao, et al., "Hes1 Is Associated With Long Non-Coding RNAs in Colorectal Cancer," *Annals of Translational Medicine* 7 (2019): 459.

34. E. Chen, F. Yang, H. He, et al., "Decreased Level of RASSF6 in Sporadic Colorectal Cancer and Its Anti-Tumor Effects Both In Vitro and In Vivo," *Oncotarget* 7 (2016): 19813–19823.

35. R. Zhou, L. Qiu, X. Liu, et al., "RASSF6 Downregulation Promotes the Epithelial-Mesenchymal Transition and Predicts Poor Prognosis in Colorectal Cancer," *Oncotarget* 8 (2017): 55162–55175.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Heatmap of differentially expressed genes (DEGs) in each cellular subcluster. **Figure S2:** Heatmap of representative DEGs in each cluster from each adjacent normal or cancer tissue. **Figure S3:** Summary of representative marker gene expression identified to nonparenchymal cells. **Figure S4:** Prediction of survival curves for each patient. **Figure S5:** Heatmap of DEGs in each cellular subcluster. **Figure S6:** Heatmap of DEGs in each cluster reanalyzed by cancer cell clusters from four different patient. **Figure S7:** A pseudo-time trajectory analysis in cancer cell clusters from four CRC patient. **Figure S8:** Expression of KCNQ1OT1 in colorectal cancer samples from four patients. **Figure S9:** Heatmap of DEGs in each cluster reanalyzed by samples from Patient #4. **Figure S10:** Heatmap of DEGs in each cluster reanalyzed by cancer cell clusters in Patient #4. **Figure S11:** Heatmap of DEGs in each cluster reanalyzed by KCNQ1OT1 highly expressed cell clusters. **Figure S12:** Reclustering of KCNQ1OT1high cancer cells in Patient #4. **Figure S13:** OS analysis of representative tumor cell clusters using TCGA colorectal cancer data (via cBioPortal). **Figure S14:** Expression levels of representative genes in HOXD9+ and PROM1+ cell populations. **Figure S15:** Expression of HES1, FOXQ1, and RASSF6in CRC cell clusters. **Figure S16:** Visualization of cancer-specific SNVs in the exon 1 region of RASSF6 in Patient #4. **Figure S17:** Schematic overview of the multistep reclustering workflow. **Table S1:** Quality control metrics for the aggregated scRNA-seq dataset. **Table S2:** Quality control metrics for scRNA-seq data in tumor tissue from Patient #4. **Table S3:** Quality control metrics for scRNA-seq data in adjacent normal colon tissue of Patient #4. **Table S4:** Representative genes identified by scRNA-seq in each tissue.