Stromal gene expression defines poor prognosis subtypes in colorectal cancer

Alexandre Calon[1], Enza Lonardo[1], Antonio Berenguer-Llergo[1], Elisa Espinet[1,2], Xavier Hernando-Momblona[1], Mar Iglesias[3], Marta Sevillano[1], Sergio Palomo-Ponce[1], Daniele V.F. Tauriello[1], Daniel Byrom[1], Carme Cortina[1], Clara Morral[1], Carles Barceló[1], Sebastien Tosi[1], Antoni Riera[1], Camille Stephan-Otto Attolini[1], David Rossell[1,4], Elena Sancho[1] and Eduard Batlle[1,5]

1. Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain.
2. Present address: Division of Stem Cells and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany
3. Department of Pathology, Hospital Universitari del Mar, Universitat Autònoma de Barcelona, Barcelona, Spain
4. Present address: Department of Statistics, University of Warwick, UK
5. Institució Catalana de Recerca i Estudis Avançats (iCREA).

Correspondence should be addressed to:

Eduard Batlle (eduard.batlle@irbbarcelona.org)

ABSTRACT

Recent molecular classifications of Colorectal Cancer (CRC) based on global gene expression profiles have defined subtypes displaying resistance to therapy and poor prognosis. Upon evaluation of these classification systems we discovered that their predictive power arises from genes expressed by stromal cells rather than by epithelial tumor cells. Bioinformatic and immunohistochemical analyses identify stromal markers that associate robustly with disease relapse across the various classifications. Functional studies indicate that cancer-associated fibroblasts (CAFs) increase the frequency of tumor initiating cells, an effect that is dramatically enhanced by transforming growth factor beta (TGF-beta) signaling. Likewise, we find that all poor-prognosis CRC subtypes share a gene program induced by TGF-beta in tumor stromal cells. Using patient-derived tumor organoids and xenografts, we show that the use of TGF-beta signaling inhibitors to block the crosstalk between cancer cells and the microenvironment halts disease progression.

INTRODUCTION

About 40-50% of CRC patients with locally advanced disease (AJCC stages II to III) show resistance to therapy and develop recurrent cancer over the course of their treatment. Current CRC staging based on histopathology and imaging has a limited ability to predict prognosis, thus leading to attempts to elaborate molecular classifications. Several independent studies have recently proposed CRC subtypes based on distinct global gene expression profiles[1-4]. While these studies differed regarding the number of tumor subtypes identified, they all concurred in that poor patient outcome in CRC is associated with the expression of stem cell and mesenchymal genes in CRC cells[1-4]. This stem-like/mesenchymal CRC subtype represents a particular class of highly aggressive CRC. These findings pave the way to improve the patient staging system and to identify molecules and signaling pathways associated with CRC metastasis and recurrence that could be targeted therapeutically.

RESULTS

Expression of poor-prognosis genes in tumor cell types

We studied the expression patterns of genes associated with disease relapse in CRC. To this end, we interrogated the three CRC patient datasets used to establish the above mentioned molecular classifications and identified all the genes significantly associated with decreased disease-free survival intervals after surgery in each dataset (hazard ratio for cancer recurrence HR>1, $P<0.01$) (Fig. 1a, grey). A significant proportion of the genes found in this analysis were upregulated in the poor-prognosis molecular subtypes (Fig. 1a, red). De Sousa E Melo et al., distinguished three CRC classes named CCS1, CCS2 and CCS3[1]. Patients belonging to the CCS3 group have higher risk of recurrence after tumor resection than the other subtypes[1]. We identified 160 genes upregulated in CCS3 (>2 fold, $P<0.05$) compared to CCS2 and CCS1 subtypes that were positively associated with disease relapse (HR>1 and $P<0.01$; Fig. 1a; Supplementary Table 1). Sadanandam et al. defined 5 distinct molecular CRC subtypes including a subset of tumors that adopt expression programs similar to that of intestinal stem cells[2]. Patient bearing Stem-like CRCs that had not been treated with chemotherapy after the surgery had lower disease free survival[2]. We found 286 genes positively associated with disease relapse that were upregulated (>2 fold, $P<0.05$) in the Stem-like tumor group compared to the other 4 subtypes (Fig. 1a; Supplementary Table 1). Marisa et al. defined 6 molecular subtypes including one, named C4, which showed borderline

3

association with an increased risk of disease relapse[3]. We found 98 genes upregulated (>2 fold, P<0.05) in the C4 subtype that were positively associated with disease relapse (Fig. 1a; Supplementary Table 1). All these transcriptomic datasets were obtained by profiling whole tumor samples. Thus, the expression of each gene could be potentially contributed by epithelial tumor cells, stromal cells or by both populations. To investigate these possibilities, we analyzed the poor-prognosis genesets in a transcriptomic dataset of CRC and normal mucosa samples, in which epithelial cells and stromal cells had been microdissected by laser capture and profiled separately (n=13)[5]. This analysis revealed that the global set of genes positively associated with recurrence in the different patient datasets (all HR>1, P<0.01; grey boxes), as well as the gene subset upregulated in the poor-prognosis subtypes (red boxes), were significantly upregulated in the microdissected tumor stroma (Str) compared to epithelial tumor areas (Ep) (Fig. 1b) and to lower extent in the stroma of the normal mucosa compared to the colonic epithelium (Fig. 1b). The upregulation of poor prognosis genesets in the tumor stroma was consistent across multiple thresholds of differential gene expression and significance (Supplementary Table 2).

We next analyzed the cell type specific expression of the poor prognosis genesets. To this end, we isolated various cell types from dissociated human primary CRC samples (n=14) and performed global transcriptomic analysis of each population (Fig. 2a)[6]. Figure 2b shows differential expression of cell type-specific marker genes in the purified fractions, which confirmed enrichment in epithelial cancer cells (EpCAM+), leukocytes (CD45+), endothelial cells (CD31+) and cancer-associated fibroblasts (FAP+) respectively. The subsequent comparative analysis demonstrated that the global set of genes positively associated with recurrence (All HR>1, P<0.01) in each CRC dataset was very significantly upregulated in stromal cell populations compared to epithelial tumor cells (Fig. 2c). Particularly, the expression of these genes was more elevated in cancer-associated fibroblasts (CAFs – FAP+), followed by endothelial cells (CD31+) and leukocytes (CD45+) (Fig. 2c). This expression pattern was even more evident for the subset upregulated in the poor-prognosis CRC subtypes (Fig. 2c). Altogether, these results indicate that tumor associated stromal cells contribute a major proportion of the transcriptome positively associated with poor prognosis in the three molecular CRC classifications proposed.

Contribution of stromal genes to define CRC subtypes

The above observations prompted us to study the relative contribution of epithelial

4

and stromal genes to the identification of distinct molecular subtypes. Sadanandam and colleagues applied a 786 gene signature to distinguish between 5 patient subtypes[2]. For each probeset mapping these genes we annotated: association with cancer recurrence (HR), significant expression in epithelial or stromal fraction from CRC microdissected samples and significant upregulation in particular tumor cell populations. We performed hierarchical clustering on these annotated probesets to explore their contribution to each molecular subtype in the GSE14333 dataset (Fig 3). This analysis revealed several gene clusters, including two composed largely by genes upregulated in epithelial tumor cells (Fig. 3a, Epithelial Enriched) that were differentially expressed across the distinct CRC subtypes. Another cluster contained genes characteristic of leukocytes (CD45+ cells) that was particularly upregulated in CRCs of the inflammatory subtype (Fig. 3a, Leukocyte Enriched). The gene cluster most highly expressed in the Stem-like poor prognosis subtype contained a large proportion of genes upregulated in cancer-associated fibroblasts (Fig. 3a, CAF Enriched). We then reclassified the CRC patient cohort either considering only the expression of genes included in the epithelial cluster or excluding genes belonging to the CD45 and FAP clusters. These two analyses resulted in the misclassification of either 42.1% or 34.5% of the patients belonging to the Stem-like subtype into good prognosis subtypes respectively (Supplementary Table 3). Therefore, expression of epithelial genes is not sufficient to define accurately the CRC molecular subtypes. We reached an equivalent conclusion from the analyses of De Sousa E Melo et al.[1] and Marisa et al.[3] classifications (Supplementary Table 3).

As predicted from the data included in Figures 1 and 2, the majority of genes included in the CAF cluster associated positively with cancer recurrence (HR>1). We noticed that expression levels of these genes were heterogeneous amongst patients belonging to good prognosis subtypes (Goblet-like, Enterocyte-like, TA-like and Inflammatory subtypes; Fig. 3a). Remarkably, we found a linear association between the average expression of this CAF gene cluster and the risk of relapse after therapy in good prognosis subtype patients (Fig. 3b). This finding thus enabled further stratification of patients belonging to good prognosis subtypes into high and low risk of relapse depending on high (GP-HiC) or low (GP-LoC) CAF gene cluster expression (Fig. 3c). Of note, the disease-free survival progression of the GP-HiC subgroup across all 4 good prognosis subtypes overlapped with that of the Stem-like poor prognosis subtype (Fig. 3c). Similarly, analysis of the molecular classification proposed by De Sousa E Melo[1], showed that elevated expression of CAF gene-enriched clusters identified a subset of patients belonging to good prognosis

subtypes that behaved as the CCS3 poor prognosis subtype (Supplementary Fig. 1). A similar trend was found in the Marisa et al. classification[3] yet the associations with relapse were of borderline significance in this case (Supplementary Fig. 2). From these observations, we conclude that elevated levels of the CAF gene program identify poor prognosis patients across CRC molecular subtypes.

Analysis of poor prognosis genes by immunohistochemistry

83 genes included in the poor prognosis genesets were common to at least two of the three molecular classifications (Fig. 4a and Supplementary Table 4). We interrogated the Human Protein Atlas Database[7] and analyzed tissue patterns of proteins encoded by poor prognosis genes. This database included immunohistochemistry (IHC) data for antibodies against 76 out of 83 proteins encoded by these genes. 3 antibodies (4%) showed no detectable staining in CRC samples and therefore we focused on 73 proteins displaying discrete IHC patterns (Supplementary Fig. 3a and Supplementary Table 4). Of these, only 2 antibodies (3%) marked exclusively epithelial tumor cells. In contrast, 31% stained solely the tumor stroma in CRC samples whereas the remaining (62%) labeled both stromal and epithelial tumor cells (Supplementary Fig. 3a and Supplementary Table 4). We performed a more detailed analysis for 4 poor prognosis genes that displayed either stromal specific (CALD1, POSTN) or mixed epithelial-stromal (FAP, IGFBP7) patterns in CRC according to the Human Protein Atlas Database. Bioinformatic analyses indicated that the mRNA of all 4 genes was upregulated in CAFs and other stromal cell populations compared to epithelial tumor cells (Supplementary Fig. 3b). In addition, the antibodies selected against these 4 proteins detected single bands of the expected molecular weights by western-blot in extracts of normal colon fibroblasts (Fig. 5e). We performed IHC on n=79 tissue samples belonging to Stage I, II and III CRC patients which confirmed the *in silico* predictions. Figure 4b shows overall staining intensity in stromal and epithelial compartments assessed under the microscope by an expert pathologist (M.I). POSTN and CALD1 were exclusively expressed in the stromal cells whereas IGFBP7 and FAP showed higher expression in the stroma compared to epithelial tumor cells (Fig. 4b and examples in 4c). We assessed association between intensity of either stromal or epithelial staining and disease free survival after tumor resection in this patient cohort (Fig. 4d-e; Supplementary Tables 5-6). This analysis showed that elevated expression of CALD1, FAP or IGFPB7 in stromal cells predicted robustly shorter disease-free intervals, either as a linear (Fig. 4d) or categorized variable (Fig. 4e). Of note, the expression of these stromal proteins was a prognostic factor independent of the main

clinical variables including AJCC stage, treatment with adjuvant chemotherapy, age, tumor location and gender (Supplementary Table 5). In contrast, epithelial expression of IGFBP7 or FAP was not associated positively with recurrence (Supplementary Table 6). Overall, these results suggest that good and poor prognosis CRC subtypes can be identified based on the expression levels of a small subset of stromal proteins.

TGF-beta signaling in CRC subtypes

TGF-beta signaling was identified as one of the biological processes enriched in the poor-prognosis molecular CRC subtypes[1-3]. We confirmed elevated expression of overall TGF-beta levels (Fig. 5a) and particularly of *TGFB1* and *TGFB3* in CCS3, Stem-like and C4 subtypes (Supplementary Fig. 4a and Supplementary Table 7). We have recently reported that genes upregulated by TGF-beta in stromal cells are robust predictors of cancer recurrence and metastasis in CRC[6]. Several of the stromal TGF-beta response signatures[6] (TBRS – see Supplementary Table 8 for gene lists) predicted disease relapse in the three CRC patient cohorts analyzed, independently of the main clinical parameters (Supplementary Table 9). We thus explored the association of these TBRSs with the proposed molecular classifications of CRC. Our analyses also revealed that the genes induced by TGF-beta in normal colonic Fibroblasts (F-TBRS) were upregulated in all poor prognosis CRC subtypes (Fig. 5b and Supplementary Table 7). We obtained equivalent results using TBRS derived from Macrophages and T-Cells (T-TBRS and Ma-TBRS) (Supplementary Fig. 4b-c).

The average expression of the CCS3, Stem-like and C4 poor prognosis genesets was tightly correlated with combined expression of *TGFB1* and *TGFB3* in the three CRC patient cohorts (Fig. 5c). To investigate whether these genesets were directly regulated by TGF-beta, we profiled the global expression of normal colon fibroblasts before and after induction with TGF-beta. GSEA indicated that TGF-beta treatment elevated the expression of these poor-prognosis genesets in fibroblasts (Fig. 5d). We also analyzed the expression of the markers used in Figure 4 to identify poor-prognosis patients by IHC. The levels of these proteins (CALD1, POSTN, FAP, IGFBP7) were upregulated by TGF-beta in colon fibroblasts (Fig. 5e). Altogether, these data show that TGF-beta signaling in stromal cells is a defining feature of poor-prognosis CRC subtypes.

Analysis of TGF-beta signaling in the tumor stroma

The above correlations prompted us to analyze the interactions of CRC cells with their microenvironment and specifically with fibroblasts through TGF-beta (Fig. 6). To functionally dissect this effect without the interference of TGF-beta signaling in epithelial cancer cells, we used the CRC cell line HT29-M6, which carries mutations in *SMAD4* that inactivate the response to TGF-beta[6]. We enforced TGF-beta signaling in the tumor microenvironment by engineering HT29-M6 to secrete active TGF-β1. HT29-M6$^{TGF-\beta}$ cells displayed no autocrine responses to TGF-β1 secretion (Supplementary Fig. 5a-b)[6]. HT29-M6^{Con} or HT29-M6$^{TGF-\beta}$ cells were not tumorigenic upon subcutaneous injection in immunodeficient mice at low numbers (Fig. 6a). In contrast, co-inoculation of HT29-M6$^{TGF-\beta}$ cells with normal colonic fibroblasts enhanced very significantly the frequency of xenograft formation (Fig. 6a). We analyzed the fate of co-inoculated fibroblasts by engineering them to express eGFP constitutively. At experimental end points, we detected a small proportion of GFP-labeled cells in HT29-M6$^{TGF-\beta}$–derived xenografts, which ranged between 1-5% of all stroma cells (n=18. Example in Fig. 6b). Therefore, the co-inoculated fibroblasts enhanced the capacity of HT29-M6$^{TGF-\beta}$ to engraft in recipient mice yet they did not contribute extensively to form the tumor stroma.

Knockdown of TGFBR1 (shTBRI) or expression of dominant negative TGFBR2 (DNR) in fibroblasts blocked the compounded effect of TGF-beta on xenograft formation (Fig. 6c). We obtained equivalent results using KM12L4a$^{TGF-\beta}$ cells (Supplementary Fig. 5c), a CRC cell line which carries biallelic *TGFBR2* loss of function mutations and that neither responds to TGF-beta (Supplementary Fig. 5a-b)[6,8]. Importantly, addition of wild-type or TGF-beta pathway defective fibroblasts did not modify xenograft growth rates (Supplementary Fig. 6a-b). Moreover, tumors displayed equivalent histological traits and were formed by similar proportions of epithelial and stromal cells (Supplementary Fig. 6c-e). We also co-inoculated shTBRI fibroblasts with control HT29-M6 cells. In these experiments, both shCon fibroblasts and fibroblasts lacking a TGF-beta response (shTBRI) showed minor capacity to support tumor formation (Supplementary Fig. 7a). In a second approach, we co-inoculated HT29-M6^{Con} cells with normal colon fibroblasts that had been pre-treated *in vitro* with TGF-beta for 96 hours prior inoculation (Fig. 6d). TGF-beta upregulated the expression of poor prognosis genes in these fibroblasts (Fig. 5d-e) while it slowed down their proliferation (Supplementary Fig. 7b). TGF-beta pre-stimulated fibroblasts reduced tumor latency and increased engraftment of control HT29-M6

cells (Fig. 6d). In this setting, transient activation of fibroblasts prior to inoculation suggests that TGF-beta-activated stroma operates at the initial phase of tumor formation.

It has been shown that the frequency of tumor initiating cells (TICs) is a surrogate for cancer stem cell activity[9,10]. This tumor cell population is believed to mediate disease relapse and metastasis[11], which are the major drivers of poor prognosis in CRC. We measured the impact of TGF-beta-activated fibroblasts on tumor initiating cells (TIC) frequency by performing limiting dilution assays (Supplementary Fig. 7c). Groups of mice were injected subcutaneously with HT29-M6[Con] or HT29-M6[TGF-β] cells at different cell concentrations with or without normal colon fibroblasts and tumor initiation was assessed. We calculated that HT29-M6[Con] cells contained 1 TIC every 1.08×10^5 cells (Fig. 6e and Supplementary Fig. 7d). Secretion of TGF-beta increased TIC frequency 5-fold over the level of control cells upon inoculation, which was roughly equivalent to that obtained upon co-inoculation of HT29-M6[Con] with normal colon fibroblasts (Fig. 6e and Supplementary Fig. 7d). Remarkably, co-inoculation of HT29-M6[TGF-β] cells with fibroblasts increased the TIC frequency to 1 every 493 cells, which represents a 200-fold increase compared to HT29-M6[Con] cells alone (Fig. 6e and Supplementary Fig. 7d). These findings reveal a synergistic as opposed to an additive effect of TGF-beta and fibroblasts on TIC frequency (Fig. 6f; P=0.0001). In addition, we calculated tumor size over time starting from the day that tumors were palpable yet we did not find significant differences in the growth rates of xenografts arising in the four experimental conditions (Supplementary Fig. 5d). We thus conclude that TGF-beta signaling in fibroblasts acts by specifically enhancing the tumor-initiating potential of CRC cells.

Analysis of TGF-beta signaling in epithelial tumor cells
In addition to regulating gene programs in stromal cells, TGF-beta may possibly also control the behavior of epithelial CRC cells that carry a wild-type pathway. TGF-beta is a well-established inducer of epithelial-to-mesenchymal transition (EMT)[12] and therefore it could elevate the expression of mesenchymal genes in CRCs, contributing to the acquisition of a more aggressive phenotype. An important limitation to test this hypothesis is the fact that virtually all CRC cell lines available have lost the response to TGF-beta as a consequence of inactivating mutations in TGF-beta pathway components[13-15]. To overcome this obstacle, we took advantage of a recently developed methodology that enables *in vitro* propagation of primary CRCs as 3D primary cultures (Fig. 7a)[16,17]. In these conditions, CRC cells grew

indefinitely as epithelial structures lacking stromal cells called tumor organoids (Fig. 7b and data not shown). We expanded organoids from 8 primary CRCs. Supplementary Table 10 details the main clinic-pathological features of the CRC of origin for each tumor organoid culture. Exome sequencing analysis confirmed that these organoids displayed distinct combinations of mutations in the main driver pathways (Supplementary Table 10). The organoid technology allowed us to study for first time primary CRC samples that display a TGF-beta response. In 4 tumor organoids (#2, #3, #4 and #6) addition of TGF-beta induced robust cytostasis that included decreased organoid forming capacity (Fig. 7c) and reduced growth rates (Supplementary Fig. 8). This response was blocked by LY2157299, a TGFBR1-specific inhibitor[18]. In these 4 tumor organoids, treatment with TGF-beta upregulated the CDK4/6 inhibitors *CDKN1A, CDKN2A* and *CDKN2B* whereas it downregulated *KI67* and *MYC* (Fig. 7e) indicating a tumor suppressor response similar to that previously shown for keratinocytes[19,20]. Exome sequencing of organoid #5 revealed no apparent alteration in genes encoding for TGF-beta pathway components yet this organoid neither halted proliferation nor upregulated expression of cell cycle inhibitors in the presence of TGF-beta (Fig. 7c-e and Supplementary Fig. 8). Tumor organoids #7 and #8 displayed SMAD4 loss-of-function alterations in homozygosis (Supplementary Table 10) and TGF-beta did not modify their growth rates (Fig. 7c-e and Supplementary Fig. 8). Tumor organoid #1 carried TGFBR2 loss of function mutations in the two alleles (Supplementary Table 10) and was completely insensitive to the action of TGF-beta (Fig. 7c-e and Supplementary Fig. 8). Of note, even after 1 week in the presence of TGF-beta, none of the tumor organoids acquired mesenchymal traits and they all remained epithelial as shown by the maintenance of a cuboidal morphology and basolateral E-cadherin staining (Fig. 7d). There were minimal changes in the expression of epithelial markers (*CDH1, CLDN1, EPCAM*) or EMT master genes (*SNAIL, TWIST1, ZEB1*)[12] (Fig. 7e). Therefore, CRCs appear to be largely resilient to the induction of EMT by TGF-beta, which nonetheless triggered a robust cytostatic response in TGF-beta pathway wild-type tumor cells.

Use of TGF-beta signaling inhibitors to block metastasis

The ultimate stage of CRC progression is metastatic disease, which is caused by disseminated cancer cells that hold the capacity to initiate a new tumor within a foreign tissue[11], mainly liver and lungs. To assess the ability of the different tumor organoids to generate metastasis, we inoculated them as dissociated cells through the spleen of immunodeficient mice (Fig. 8). Despite thousands of tumor cells

entering the portal vein and reaching the liver within minutes in this experimental model, most tumor organoids produced few or no metastases (Fig. 8a). The exceptions were organoids #1 and #7, both of which expressed elevated TGF-beta levels (Fig. 8b) and were highly metastatic (Fig. 8a and pictures in Fig. 8c). LY2157299[18] is a TGFBR1-specific inhibitor that is currently in phase II clinical trials for the treatment of hepatocellular carcinoma[21]. Treatment of mice with this drug reduced the number of metastases formed by these two tumor organoids (Fig. 8c and 8d). Reversibly, enforced secretion of active TGF-beta1 by organoid #5 massively increased the metastatic burden (Fig. 8d-e). Analysis of the kinetics of metastasis using luciferase-labeled organoids indicated that LY2157299 acted by blocking the capacity of tumor cells to thrive in the liver over the colonization phase (i.e. during the first few days after inoculation) (insets in Fig. 8c). Of note, tumor organoids #1 and #7 exhibited biallelic inactivation of the *TGFBR2* and *SMAD4* genes respectively (Supplementary Table 10) and both were insensitive to TGF-beta (Fig. 7 and Supplementary Fig. 8). Therefore, reduced metastatic capacity caused by LY2157299 treatment in these tumors can only be attributed to inhibition of TGF-beta signaling in the tumor microenvironment. Indeed, we observed decreased p-SMAD2 staining and reduced expression of TGF-beta target genes CALD1 and POSTN in the stroma of the liver metastases upon LY2157299 treatment (Fig. 8f). Remarkably, despite the *in vitro* cytostatic effect imposed by TGF-beta in tumor organoids with an intact TGF-beta pathway, we did not observe enhancement of metastasis *in vivo* upon treatment of mice with LY2157299 (Fig. 8d; #2 and #6). LY2157299 promoted neither initiation nor growth of TGF-beta-responsive tumor organoids upon subcutaneous inoculation in mice (Supplementary Fig. 9).

DISCUSSION

Recently developed molecular classification systems offer great opportunities to improve the stratification and treatment of CRC patients[1-3]. The finding that poor-prognosis subtypes are characterized by elevated expression of mesenchymal genes has led to the speculation that EMT may be responsible for their aggressiveness[1-3]. Our analyses reveal that elevated expression of mesenchymal genes associated with poor prognosis in CRC samples is mainly contributed by tumor-associated stromal cells rather than by epithelial tumor cells. This discovery does not invalidate the possibility that individual tumor cells undergo EMT, particularly at invasion fronts. Yet, it argues against a generalized expression of a mesenchymal poor-prognosis gene program in epithelial CRC cells. Sadanandam et al. proposed that the poor prognosis subtype is characterized by upregulation of a gene program similar to that

of intestinal stem cells. This conclusion was based on the expression of genes upregulated in microdissected crypt bottoms compared to crypt tops in stem cell-like tumors[22]. However, the signature of crypt bottoms used in this study also contained mesenchymal genes expressed by pericryptal fibroblasts, the presumptive stem cell niche cells, which were microdissected together with the epithelial cells[22]. Our bioinformatic and IHC analyses indicate that several poor-prognosis genes are expressed by stromal cells in the normal mucosa (Fig. 1b and Fig. 4c), which may suggest that the CRC microenvironment reproduces some features of the crypt niche. We propose that, to a large extent, the various molecular classifications published so far distinguish good and poor prognosis CRCs based on distinct features of the tumor stroma. Importantly, even those patients classified within good prognosis subtypes that expressed elevated levels of the CAF program displayed increased risk of relapse and disease-free survival intervals, similar to those of poor prognosis subtype patients. These observations could help unify the multiple CRC subtypes identified in each molecular classification. Our conclusions are further supported by various studies that link elevated expression of particular stromal-specific genes with poor outcome in CRC[23, 24-27].

We showed that a common feature of all poor-prognosis subtypes is elevated TGF-beta expression. Our functional data underscore a dual role for the TGF-beta pathway in epithelial versus stromal tumor cells. TGF-beta signaling slows-down proliferation of epithelial CRC cells without triggering EMT. A large fraction of CRCs avert this block by losing sensitivity to TGF-beta via mutations and possibly through additional mechanisms. On the other hand, TGF-beta in CAFs boosts the tumor-initiating capacity of CRC cells, a property that is connected with increased metastatic potential and ability to regenerate the disease after therapy[11]. Indeed, our data suggest that TGF-beta target genes are an important fraction of the poor prognosis genesets . We show that pharmacological inhibition of TGF-beta signaling in the tumor microenvironment prevents metastasis formation by patient-derived tumor organoids. These findings corroborate our previous results demonstrating a dependency of TGF-beta signaling in stromal cells during metastasis[6]. Our work further warrants the development of anti-TGF-beta therapies for the treatment of poor prognosis CRCs. The observation that LY2157299 did not boost xenograft growth even in TGF-beta responsive tumor organoids indicates that the use of this inhibitor may be safe for a wide range of CRC patients.

METHODS

Methods and any associated references are available in the online version of the paper.

CONTRIBUTIONS

A.C, E.L. and E.E. designed, planned and performed experiments, and analyzed the results. X.H. and S.P.P. provided crucial assistance with *in vivo* experiments. M.S. performed immunohistochemistry. M.I. scored the TMA. S.T. developed an algorithm to quantify xenograft images. D.V.F.T, C.C, C.B. and C.M. performed experiments and/or analyzed results. D.V.F.T, D.B. and A.R. synthesized the LY2157299 inhibitor. A.B, C.S.-O, and D.R. designed and performed biostatistical analyses. E.B. conceptualized and supervised the project, analyzed results and wrote the manuscript, with the assistance of ES.

COMPETING FINANCIAL INTERESTS

Authors declare no competing financial interests

FIGURE LEGENDS

Figure 1. Genesets defining poor prognosis CRC subtypes are expressed in the tumor stroma.

(**a**) Venn diagrams illustrate the overlap (red) between genes predicting relapse (All HR>1, $P<0.01$; grey) and the genes upregulated in the poor prognosis groups of their respective datasets (white). FDR and P-values for intersections assuming hypergeometric distributions are; FDR=13.1% and $P<2.22e-16$ for CCS3∩HR, FDR=9.4% and $P<2.22e-16$ for Stem-like∩HR, FDR=21.4% and $P<2.22e-16$ for C4∩HR. For the GSE14333 cohort, only patients that did not received chemotherapy were considered. (**b**) Mean expression (Z-score) of genes that predict relapse (grey panels) in the three studies in Laser Capture Microdissected (LCM) tumor epithelial (Ep) and stromal (Str) compartments or stroma and epithelial cells from normal colonic mucosa. Red boxes indicate equivalent analyses for the genesets defining the CCS3, Stem-like or C4 groups that hold predictive power for relapse (****:$P<0.0001$ for tumor stroma vs. tumor epithelium, * is $P<0.05$ in normal stroma *vs*. normal epithelium). Whiskers in boxplots extend to minimum and maximum values.

Figure 2. Genesets defining poor prognosis CRC subtypes are predominantly expressed in cancer-associated fibroblasts (CAF).

(**a**) Scheme depicting purification of cell populations from disaggregated primary CRC samples, enriching for the indicated cell types. (**b**) Heat map showing expression levels of epithelial (*CDH1, EPCAM, KRT20*), leukocyte (*CD2, CD19, PTPRC*), endothelial (*CDH5, ENG, VWF*) and cancer-associated fibroblasts marker genes (*DCN, PDPN, FAP*) in each FACS-purified cell population. Data represent normalized, centered and scaled Affymetrix probe intensities in a log2 scale. (**c**) *Z*-score means of genes that predict relapse in the three depicted studies within the

different human tumor cell types (*** indicates $P<0.001$). Whiskers in boxplots extend to minimum and maximum values.

Figure 3. High levels of genes characteristic of CAFs identify poor prognosis patients across CRC subtypes.

(**a**) Clustering analysis of the 786-gene signature used to classify CRC patients into subtypes by Sadanandam et al.[2] in the GSE14333 cohort. We allowed unsupervised hierarchical clustering of the 786 genes whereas we enforced the classification of patients into subtypes. Data shows normalized, centered and scaled Affymetrix probesets intensities in a log2 scale. HR lane represents the hazard ratio for the corresponding genes in a log2 scale. Genes significantly upregulated in microdissected epithelial or stromal compartment[5] are depicted in epithelial (LCM) and stroma (LCM) lanes respectively. Genes specifically upregulated in epithelial, endothelial, leukocyte or FAP cell population are represented in EpCAM, CD31, CD45 and FAP lanes respectively. Patients across the entire cohort with high average expression (Z-scores>0) of the CAF-enriched gene cluster are marked in green (GP-HiC) (**b**) Smooth estimate of HR (+1SD - for one standard deviation) shows higher risk of relapse for patients in good prognosis subtypes presenting higher average expression of the CAF cluster (GP-HiC group). Dashed lines indicate 95% confidence bands. (**c**) Kaplan–Meier curve shows recurrence-free survival of patients within good prognosis cancer subtypes presenting low expression levels of the CAF cluster geneset (blue, GP-LoC), and of patients within good prognosis cancer subtypes but presenting high expression levels of the CAF cluster geneset (yellow, GP-HiC), both compared to the Stem-like poor prognosis subtype (red, Stem-like). HR and P-values are indicated.

Figure 4. Identification of poor prognosis patients by IHC.

(**a**) Venn diagram illustrates the set of genes shared between CCS3 and Stem-like (*CALD1, FAP, POSTN*) and between Stem-like and C4 poor-prognosis signatures (*IGFBP7*). (**b**) Quantification of TMA analysis (79 patients) of common stromal genes in both tumor epithelium and stroma, scored from 0 to 3, as units of intensity (UI). Values are mean ± s.e.m. (**c**) Immunostaining of CALD1, FAP, POSTN and IGFBP7 in representative human normal colon and tumor samples (S: stroma; T: tumor). Scale bars: 200 μm. (**d**) HR (+1UI; increase in recurrence risk for every unit of intensity) as a smooth function of protein intensity based on their TMA score of CALD1, FAP and IGFBP7. Indicated are 95% confidence bands (red dashed lines). HRs for continuous model and corresponding *P*-value are indicated. (**e**) Kaplan–Meier curves show recurrence-free survival of patients with different levels of stromal protein intensity (L, low; M, medium; H, high) of CALD1, FAP and IGFBP7. HRs and corresponding *P*-values are indicated (lower intensity category as reference, blue line).

Figure 5. High levels of TGF-beta and F-TBRS expression characterize poor prognosis CRC subtypes.

(**a**) Overall *TGFB* mRNA expression levels in the three CRC patient datasets as *Z*-score mean computed on *TGFB1*, *TGFB2* and *TGFB3* selected probesets ($P < 0.001$ for all comparisons involving CCS3, Stem-like or C4 subtypes. See Supplementary Table 7 for pair-wise comparisons). (**b**) *Z*-score mean of F-TBRS in the three datasets for each molecular subtype ($P < 0.001$ for all comparisons involving CCS3, Stem-like or C4 subtypes. See Supplementary Table 7 for pair-wise comparisons). Poor prognosis groups for each dataset are depicted in red in (**a**) and (**b**). Whiskers represent upper and lower quartiles. (**c**) Correlation between *TGFB1+TGFB3* mRNA levels and poor-prognosis signatures in the three datasets. Spearman's correlations

(R) and *P*-values are indicated. (**d**) Gene set enrichment analyses (GSEA) of poor-prognosis genesets in fibroblasts control or treated with TGF-beta (ES: enrichment score; NES: normalized enrichment score; FDR: false discovery rate). (**e**) Levels of CALD1, FAP, POSTN and IGFBP7 proteins (upper panels) in fibroblasts (FIB) untreated (Con), treated with recombinant TGFB1, treated with LY2157299 (LY) and in HT29-M6 cells. Western blots were performed with the antibodies used for IHC analyses in Fig. 4. Bottom panel shows Actin protein levels as normalization controls.

Figure 6. TGF-beta activated fibroblasts promote tumor initiation.

(**a**) Disease-free survival (DFS) plots for mice co-inoculated with HT29-M6$^{TGF\beta}$ cells (10^3 cells, n=12 inoculations; 10^2 cells, n=6 inoculations) plus fibroblasts (FIB, 5x10^4 cells) compared to mice inoculated with fibroblast alone (5x10^4 cells, n=6 tumor cell inoculations), HT29-M6^{Con} alone (10^3 cells, n=6 inoculations) or HT29-M6$^{TGF\beta}$ alone (10^3 cells, n=20 inoculations; 10^2 cells, n=6 inoculations). (**b**) GFP immunostaining of a representative tumor derived from HT29-M6$^{TGF\beta}$ cells co-inoculated with GFP-expressing fibroblasts. Arrows point to GFP+ fibroblasts. Inset: magnification of a GFP+ fibroblast. Scale bars =100 µm. (**c**) Disease-free survival plots for mice after co-inoculation of 6x10^3 HT29-M6$^{TGF\beta}$ cells together with control fibroblasts (FIB-shCon, n=20; FIB-Con, n=14), co-inoculated with FIB-shTBRI (n=14) or FIB-DNR (n=20) compared with HT29-M6$^{TGF\beta}$ alone (n=14 inoculations). ** is *P*<0.01 and **** is *P*<0.0001 (**d**) Disease-free survival plots of mice co-inoculated with HT29-M6 cells (15x10^3) plus 5x10^4 FIB pre-stimulated during 4 days with recombinant TGFB1 (n=24 inoculations), plus untreated fibroblasts (n=40) or HT29-M6 cells alone (n=14).

**** is *P*<0.0001. (**e**) Estimated frequency of tumor initiating cells (TICs) calculated using ELDA (see online method). ***: *P*<0.001 for comparisons against TGFB + FIB. 95% confidence intervals are indicated. (**f**) Lines indicate the percentage of mice developing tumors (solid lines) according to estimations provided by a logistic

regression model (see online methods) including the interaction term (likelihood-ratio test, *P*=0.0001). Dashed lines are 95% confidence bands.

Figure 7. TGF-beta induces a cytostatic response but not EMT in the tumor organoids with wild-type pathway

(**a**) Colorectal cancer stem cells (CoCSCs) were isolated from fresh patient biopsies and cultured in Matrigel and niche factors. Under *in vitro* optimal culture conditions CoCSCs consistently form dysplastic organoid structures. (**b**) Confocal images for Phalloidin (red) and nuclei (blue) of organoid derived from 8 patients. TGF-beta pathway status is indicated. Scale bars: 50 µm. (**c**) Tumor organoid formation capacity in the presence or absence of TGFB1 recombinant protein (5 ng mL^{-1}) or LY2157299 TBRI inhibitor (1 µM). Values are mean ± s.d. (n≥6 independent organoid cultures). (**d**) Confocal images for E-cadherin (green) of organoids treated or untreated with recombinant TGFB1 for 7 days. Scale bars: 50 µm. (**e**) qRT-PCR analysis of cell cycle, EMT and epithelial genes in organoids treated or untreated for 7 days with recombinant TGFB1. Data are normalized to *PP1A* and are presented as fold change *versus* untreated cells (n≥6 independent tumor organoid cultures).

Figure 8. TGF-beta activated stroma promotes tumor initiation, an effect reversed by its chemical inhibition with LY2157299.

(**a**) Intra-splenic injection of #2, #6, #5, #7 or #1 (2x10^6 cells: n=5 mice; 0.3x10^6 cells: n=5 mice). Table shows disease-free survival and incidence of distant metastasis after inoculation (na: not available). (**b**) Relative expression levels of *TGFB1, 2* and *3* genes measured by qRT-PCR in the indicated tumor organoid cultured *in vitro*. Values are mean ± s.d.; n=3 cultures per condition. (**c, e**). Normalized bioluminescence over time from mice inoculated intra-splenic (**c**) with #1 or #7 cells and treated with LY2157299 (red line; ·#1 n=5; #7 n=6) or with vehicle (Con, blue line; #1 n=5; #7 n=6) or (**e**) with #5-TGFB1-inducible cells (2x10^6 cells) and treated

with doxycycline (yellow line; n=6 mice) or with sucrose (blue line; n=6 mice). Signal intensities were normalized to day 0, which was set arbitrarily to 100. Values are mean ± s.e.m. (*: P<0.05; **: P<0.01). Macroscopic pictures of representative livers are shown. Arrows point to metastatic nodules. (**d**) Quantification of liver metastases at time of sacrifice in (**c, e**). Values are mean ± s.e.m. (**f**) H&E and immunostaining for P-SMAD2, POSTN and CALD1 in representative liver metastases obtained from mice treated with carrier (Con; upper panel) or with LY2157299 (LY; lower panel) (S: stroma; T: tumor; Scale bars: 100 μm).

REFERENCES

1. De Sousa, E.M.F. *et al.* Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med* **19**, 614-8 (2013).
2. Sadanandam, A. *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* **19**, 619-25 (2013).
3. Marisa, L. *et al.* Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* **10**, e1001453 (2013).
4. Sadanandam, A. *et al.* Reconciliation of classification systems defining molecular subtypes of colorectal cancer: interrelationships and clinical implications. *Cell Cycle* **13**, 353-7 (2014).
5. Nishida, N. *et al.* Microarray analysis of colorectal cancer stromal tissue reveals upregulation of two oncogenic miRNA clusters. *Clin Cancer Res* **18**, 3054-70 (2012).
6. Calon, A. *et al.* Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. *Cancer Cell* **22**, 571-84 (2012).
7. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* **28**, 1248-50 (2010).
8. Mouradov, D. *et al.* Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res* **74**, 3238-47 (2014).
9. O'Brien, C.A., Pollett, A., Gallinger, S. & Dick, J.E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106-110 (2007).
10. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111-115 (2007).
11. Oskarsson, T., Batlle, E. & Massague, J. Metastatic stem cells: sources, niches, and vital pathways. *Cell Stem Cell* **14**, 306-21 (2014).
12. Thiery, J.P., Acloque, H., Huang, R.Y. & Nieto, M.A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871-90 (2009).
13. Markowitz, S.D. & Bertagnolli, M.M. Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med* **361**, 2449-60 (2009).
14. Grady, W.M. & Markowitz, S.D. Genetic and epigenetic alterations in colon cancer. *Annu Rev Genomics Hum Genet* **3**, 101-28 (2002).
15. Markowitz, S. *et al.* Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**, 1336-8 (1995).
16. Jung, P. *et al.* Isolation and in vitro expansion of human colonic stem cells. *Nat Med* **17**, 1225-7 (2011).
17. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762-72 (2011).
18. Bueno, L. *et al.* Semi-mechanistic modelling of the tumour growth inhibitory effects of LY2157299, a new type I receptor TGF-beta kinase antagonist, in mice. *Eur J Cancer* **44**, 142-50 (2008).
19. Seoane, J., Le, H.V. & Massague, J. Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature* **419**, 729-34 (2002).
20. Seoane, J. *et al.* TGFbeta influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b. *Nat Cell Biol* **3**, 400-8 (2001).

21.   Giannelli, G., Villa, E. & Lahn, M. Transforming growth factor-beta as a therapeutic target in hepatocellular carcinoma. *Cancer Res* **74**, 1890-4 (2014).

22.   Kosinski, C. *et al.* Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc Natl Acad Sci U S A* **104**, 15418-23 (2007).

23.   Berdiel-Acer, M. *et al.* A 5-gene classifier from the carcinoma-associated fibroblast transcriptomic profile and clinical outcome in colorectal cancer. *Oncotarget* **5**, 6437-52 (2014).

24.   O'Shannessy, D.J. *et al.* Influence of tumor microenvironment on prognosis in colorectal cancer: Tissue architecture-dependent signature of endosialin (TEM-1) and associated proteins. *Oncotarget* **5**, 3983-95 (2014).

25.   Franci, C. *et al.* Snail1 protein in the stroma as a new putative prognosis marker for colon tumours. *PLoS One* **4**, e5595 (2009).

26.   Ngan, C.Y. *et al.* Quantitative evaluation of vimentin expression in tumour stroma of colorectal cancer. *Br J Cancer* **96**, 986-92 (2007).

27.   Calon, A., Tauriello, D.V. & Batlle, E. TGF-beta in CAF-mediated tumor growth and metastasis. *Semin Cancer Biol* **25**, 15-22 (2014).

ONLINE METHODS

Description of CRC transcriptomic datasets

We used three Affymetrix datasets publicly available in NCBI's Gene Expression Omnibus[28] with available clinical data and follow-up information: GSE14333[29], GSE33113[30], and GSE39582[3]. GSE14333 contains a pool of 290 CRC patients treated at two different hospitals; Peter MacCallum Cancer Center (Australia) and H. Lee Moffitt Cancer Center (USA). This dataset was used to define the molecular classification by Sadananman et al.[2]. GSE33113 dataset includes disease free survival information for 90 AJCC stage II patients collected in the Academic Medical Center in Amsterdam (The Netherlands) and it was used to define the molecular classification by De Sousa E Melo et al.[1]. Finally, GSE39582 includes expression and clinical data for 566 CRC patients collected for the Cartes d'Identité des Tumeurs (CIT) program, from the French Ligue Nationale Contre le Cancer (http://cit.ligue-cancer.net). This dataset was used to build the CRC classification by Marisa et al.[3]. All stage 1, 2 and 3 (but not stage 4) patients in each cohort were taken into consideration for analyses. See Supplementary Note for details about data processing.

Expression profile of poor-prognosis molecular subtypes

To characterize each poor-prognosis subtype (CCS3, Stem-like and C4), we defined a transcriptomic profile on the same data originally used to define their corresponding classification (GSE33113, GSE14333 and GSE39582, respectively). Genes included in these profiles were required to meet stringent statistical criteria: at least 2-fold overexpression in the poor-prognosis subtype and FDR<0.01.

Association of poor-prognosis subtypes with clinical outcome

Association of gene expression with relapse was assessed in the CRC transcriptomic datasets using a univariate Cox proportional hazards model. For Sadananman et al. signature, an association with recurrence was reported only in patients not treated with adjuvant chemotherapy. Therefore, this analysis was restricted to this group of patients in dataset GSE14333. When assessing overlap with subtypes profiles, risk genes (Hazard Ratio>1, HR>1) showing a *P*-value less than 0.01 were considered. A gene was located in the intersection when some of their probesets were found in both profiles. Significance of the overlapping genes was assessed assuming a hypergeometric distribution and median and percentiles 2.5% and 97.5% of False

Discovery Rate (FDR) were computed in each case, assuming that occurrences in the intersection were due to randomness.

Expression of poor-prognosis subtypes in sub-population cells

Three GEO datasets were used to characterize the subtypes gene profiles according to their specific expression in tumoral cell sub-populations: GSE39395[6], GSE39396[6], and GSE35602[5]. In GSE39395 and GSE39396 and as we described previously[6], Fluorescence Activated Cell Sorting (FACS) was used to separate the following populations from 14 fresh CRC samples; [CD45(+), EpCAM(-), CD31(-), FAP(-)], [CD45(-) EpCAM(+), CD31(-), FAP(-)], [CD45(-), EpCAM(-), CD31(+), FAP(-)] and [CD45(-), EpCAM(-), CD31(-), FAP(+)]. GSE35602 contains transcriptomic data of epithelial cells and stromal cells microdissected from 13 CRC tissue samples and 4 adjacent morphologically normal colorectal mucosae (>5 cm from the tumor). The signatures derived from the poor-prognosis subtypes were summarized and evaluated in these samples. Population groups were then compared using Kruskal and Mann-Whitney tests. In addition, we explored the sensitivity of these results to the thresholds used to define the subtypes and recurrence signatures (see Supplementary Note for details).

Contribution of genes expressed by epithelial or stromal cells to the molecular classification of CRC

Gene signatures defining the molecular classifications derived by Sadanandam et al[2] (786 genes), De Sousa E Melo et al[1] (146 Affymetrix HG U133 Plus 2.0 probesets) and Marisa et al[3] (1459 Affymetrix HG U133 Plus 2.0 probesets) were retrieved from their original publications. All these probesets were annotated according to their corresponding estimation of recurrence risk (HR for continuous variables), their possible up-regulation in microdissected epithelial or stromal compartment in dataset GSE35602 (fold-change>1.5, *P*-value<0.05 after Benjamini and Yekutieli correction[31]) and their eventual specific upregulation in epithelial, endothelial, leukocyte or FAP cell population according to datasets GSE39395 and GSE39396 (minimum fold-change>1.5 compared to any other cell population; all raw *P*-values < 0.05).

For each classification, a hierarchical clustering was performed on probesets using correlation distance and Ward's agglomeration method. Clustering results were shown in a heatmap in which samples were enforced to be grouped according to

their molecular subtype. CAFs, leukocytes and epithelial enriched clusters were identified according to the annotation of the probesets that were part of them.

To evaluate the role of epithelial and stromal genes in the definition of the molecular subtypes, we trained a PAM classifier[32] to predict these subtypes using the probesets included in the subtype signatures as predictors. For each classification, this analysis was performed using a) all probesets in the profile; b) all probesets excluding those being part of the CAF or leukocyte enriched clusters; c) epithelial enriched clusters only, when it existed. Global and subtype specific error rates were computed in each case.

For each molecular classification, CAF clusters were summarized and their prognosis values were evaluated after excluding the poor-prognosis molecular subtype from the dataset. In addition, patients not harboring the poor-prognosis subtype were classified according to the expression shown in the CAF cluster geneset (Low: less than zero; High: greater than zero). Differences in relapse risk between these groups and the poor-prognosis subtype were then assessed. In all cases, association with recurrence was analyzed using Kaplan-Meier estimates and Cox proportional hazards models (see Supplementary Note for further details).

Association of poor-prognosis subtypes with TGFB

To evaluate the association of *TGFB* with poor-prognosis subtypes (CCS3, Stem-like and C4), we assessed the association of *TGFB1*, *TGFB2* and *TGFB3* expression with each molecular classification in datasets GSE33113, GSE14333 and GSE39582 respectively. We described previously which probesets in Affymetrix Human Genome U133 Plus 2.0 correlate better with measurements of *TGFB1*, *TGFB2* and *TGFB3* obtained by qRT-PCR data[6]. Selected probesets were *TGFB1*: 203085_s_at, *TGFB2*: 220406_at, *TGFB3*: 209747_at. Due to the generalized low expression intensity of *TGFB2*, overall expression of *TGFB* was summarized using only *TGFB1* and *TGFB3* expression. Subtypes comparisons were performed for *TGFB1*, *TGFB3* and overall *TGFB* expression levels using standard linear models and two-sided t-tests.

Association of poor-prognosis subtypes with stromal TGFB response signatures (TBRS)

Generation of TGFB response signatures have been previously described Insert Ref with Endonote style. To assess association with poor-prognosis subtypes, we

summarized each TBRS signature in datasets GSE33113, GSE14333 and GSE39582 separately. Subtypes comparison were conducted using standard linear models and two-sided t-tests.

Correlation of poor-prognosis subtypes with TGF-beta

Correlation between *TGFB1/TGFB3* expression and the poor-prognosis subtypes (CCS3, Stem-like and C4) was independently assessed in the data where they were originally described (GSE33113, GSE14333 and GSE39582, respectively). Similarities between subtypes and *TGFB* profiles summaries were evaluated via Spearman correlation. Summaries of signatures were centered and scaled before computing the Spearman's correlation.

Tissue microarray construction

Formalin-fixed, paraffin-embedded tissue blocks of colorectal adenocarcinomas were retrieved from the archives of the Servei de Patologia from Hospital del Mar. Samples were obtained under informed consent and approval of the Tumor Bank Committees according to Spanish ethical regulations. The study followed the guidelines of the Declaration of Helsinki and patient´s identity of pathological specimens remained anonymous in the context of this study. Different areas of invasive carcinoma and, when possible, different histological tumor patterns (cribriform, mucinous, poorly-differentiated); adenomatous lesions from the same surgical sample, when available; and normal mucosa, located far from the carcinoma, were identified on corresponding hematoxylin-eosin-stained slides.

Immunohistochemistry

Immunostainings were carried out using 4 μm tissue sections according to standard procedures. Briefly, after antigen retrieval, samples were blocked with Peroxidase-Blocking Solution (Dako, S202386) for 10 min at RT, and then primary antibodies were incubated overnight. Slides were washed with EnVision FLEX Wash Buffer (Dako, K800721) and the corresponding secondary antibody was incubated with the sample for 45 min at RT. Samples were developed using 3,3′-diaminobenzidine, counterstained with hematoxylin and mounted. Antibodies against CALD1 (HPA008066, Sigma), POSTN (HPA012306, Sigma) and IGFBP7 (HPA002196, Sigma) were used at 1:200 dilution. Anti-EpCAM (AF960, R&D Systems) and anti-FAP (MABS1004, Vitatex) were used at 1:50 dilution.

Association of markers with recurrence in immunohistochemistry data

IHC was performed on a Tissue Micro Arrays (TMA) of stage I, II and III CRCs samples patients (n = 79) from Hospital del Mar distributed in two independent TMAs. TMAs contained two to five representative samples per tumor. The TMAs were stained with a panel of antibodies (see Supplementary Information). Qualitative categorization of the samples according to staining in stromal vs. epithelial cancer cells was performed by an expert pathologist (M.I) in a blinded fashion in respect of the clinical information. An intensity score from 0 to 3 (units of intensity or UI) was given for each sample and average intensity score per tumor was calculated for each patient. For tumor-associated stroma, all cell types stained were taken into consideration.

We assessed the association between tumor relapse and protein intensity staining (scored from 0 to 3) using Kaplan-Meier estimates and Cox proportional hazard models. Staining values were evaluated as continuous (whenever it was possible) and categorical covariates (see Supplementary Note). As these samples were distributed across two independent TMAs, this information was included as an adjusting covariate in the analyses. In addition, Cox models were fitted adjusting by AJCC stage, treatment with adjuvant chemotherapy, age, tumor location and gender (see Supplementary Note for further details).

Western Blot

Western-blot was carried out according to standard procedures. Antibodies against POSTN (HPA012306, Sigma), FAP (MABS1004, Vitatex) and IGFBP7 (HPA002196, Sigma) were used at 1:250 dilution. Anti-CALD1 (HPA008066, Sigma) and anti-Beta-Actin (ab20272, Abcam) were used respectively at 1:1000 and 1:30000 dilutions.

Detection of poor prognosis subtypes markers in The Human Protein Atlas

Qualitative categorization of the samples according to staining in stromal vs. epithelial cancer cells was performed by an expert pathologist (M.I) in a blinded fashion. For tumor-associated stroma, all cell types stained were taken into consideration.

Culture of tumor organoids

Human biological samples used to expand tumor organoids were obtained from individuals treated at Hospital del Mar and Hospital de la Santa Creu i Sant Pau, under informed consent and approval of the Tumor Bank Committees according to Spanish ethical regulations. The study followed the guidelines of the Declaration of

Helsinki and patient´s identity of pathological specimens remained anonymous in the context of this study. Colorectal cancer stem cells (CRC-SCs) were identified and purifed using the same protocol that we developed for normal colon stem cells (CoSC)[16] and later used to isolate stem cell-like cells from CRC samples[33]. Briefly, CRC cells with high EphB2 levels were FACS sorted from dissociated tumors and cultured embedded in Matrigel (Basement Membrane Matrix Low Concentration, BD) with a simplified version of the normal CoSC-specific media described by Jung et al.[16] (Advanced DMEM/F12; 10 mM HEPES; 1x Glutamax; 1x B-27 without retinoic acid; 1x N-2; 20 ng mL^{-1} bFGF; 50 ng mL^{-1} EGF; 1 µM LY2157299 and 10 µM Y-27632). In these conditions, EphB2-high cells expanded tumor organoids that we could propagate indefinitely whereas EphB2-medium and low cells did not give rise to tumorids. Tumor organoids #7 and G#8, a kind gift of Giorgio Stassi (University of Palermo), were obtained from dissociation of whole CRCs in suspension as described elsewhere[34]. They were cultured with the media described above upon arrival to our laboratory. All cells were tested weekly for mycoplasma contamination and resulted negative.

Cell lines

Cell lines were maintained in DMEM supplemented with 10% FBS. KM12L4a, HT29-M6, CCD-18co fibroblasts (FIB)[35,36] cells control or TGF-beta secreting have been described elsewhere. For bioluminescent tracking, cancer cells were infected with a fusion protein reporter construct encoding red fluorescent protein (mCherry) and firefly luciferase. For cell tracking, fibroblasts were infected with a construct encoding green fluorescent protein (GFP). Knockdown experiments were conducted using shRNA targeting *TBR1* (Sigma Aldrich). Non-targeting shRNA sequence was used as control (Sigma Aldrich).   The dominant-negative mutant form of the type II TGF-beta receptor (DNR)[37] cloned in pcDNA3 vector was provided by Dr. Lalage Wakefield (NCI, Maryland). All genetically manipulated cells were produced via lentiviral infection. All cell lines were tested weekly for mycoplasma contamination and resulted negative.

Fibroblasts treatments

For association with each poor-prognosis subtype, a differential expression analysis was performed by comparing colonic fibroblasts (FIB) treated or not with TGF-beta1 for 8 hours. (FC>2, Benjamini-Yekutieli *P*-value<0.05). This geneset was obtained from an Affymetrix Human Genome U133 Plus 2.0 expression dataset including control fibroblasts samples and fibroblasts stimulated with recombinant TGF-beta1

(n=2). Genes were ranked according to the t-statistics derived from this comparison. This ranking was then tested for enrichment of genes found to be highly upregulated in each poor-prognosis subtype and associated with prognosis using GSEA[38]. These data were submitted in GEO and are accessible through GEO series accession number GSE64192.

For *in vivo* co-injections pre-stimulation of fibroblasts before subcutaneous injection into nude mice was performed as follows: cells were cultured in DMEM 5% FBS in the presence or absence of human recombinant TGF-beta1 (5 ng mL^{-1}) for a total of 96 hours. Media were changed every 24 hours.

Quantification of stromal and epithelial tumor composition with Masson's Trichrome staining

Masson's trichrome staining was carried out using 4 µm tissue sections according to standard procedures (Sigma, ref: HT15). *Fiji Trainable Weka segmentation* (http://fiji.sc/Trainable_Weka_Segmentation), an ImageJ plugin based on the *Weka* [39] machine learning Java library, was used to classify images based on local colorimetric, textural and structural features in the neighborhood of each pixel. Images were processed with a custom macro (http://adm.irbbarcelona.org/image-j-fiji). Pairwise comparisons were explored using a two-sided t-test.

Orthotopic Mouse studies

All experiments with mouse models were approved by the animal care and use committee of the Barcelona Science Park (CEEA-PCB) and the Catalan Government. Cells were injected subcutaneously or intra-splenically[40] in 5 to 6 weeks old females NSG or Swiss nude mice (Jackson Labs), which were followed until sacrifice. Sample size was predetermined empirically according to previous experience using the same strains and treatments. Generally, we used n>6 mice per genotype and condition. We ensured that experimental groups were balanced in terms of animal age, sex and weight. The experiments were not randomized. Animals were caged together and treated in the same way. Neither the technician nor the investigator could distinguish them during the experiment or when assessing outcomes. The general condition of animals was monitored using animal fitness and weight controls throughout the experiment. When deteriorating clinical alterations were observed, the animals were excluded from the study and sacrificed. Subcutaneous tumor appearance was assessed by palpation. Tumor volume was measured twice a week until sacrifice. Tumorigenesis after intra-splenic injection was

assessed by bioluminescent imaging (see section: Bioluminescent imaging and analysis).

Estimation of TIC frequency and assessment of synergistic interactions

Serial dilutions of control cancer cells (HT29-M6 Con) or cancer cells with forced TGF-beta1 expression (HT29-M6 TGF-beta) were co-injected subcutaneously with or without fibroblasts (FIB: 5 x 10^4). The proportion of mice showing development of tumor was modeled using a logistic regression model in which TGF-beta status, fibroblasts inoculation and dose were considered as covariates. Dose was included in the model in logarithm scale. An interaction term was also included in the model to assess whether a synergistic effect between TGF-beta and fibroblasts existed on tumor development. A Likelihood Ratio Test (LRT) was used to assess the significance of the corresponding interaction term as opposed to an additive effect of TGF-beta and fibroblasts. For each condition, the interaction model was used to estimate the logit functional relationships between the proportion of mice developing tumors and the number of injected cells, as well as their 95% confidence bands. For each sample condition, the proportion of Tumor Initiating Cells (TICs) and its 95% confidence interval was estimated using ELDA[41], a methodology that assumes the Poisson single hit model. A LRT test was used to assess pairwise differences of TICs between conditions.

Pharmacological TGF-beta inhibition by LY2157299 treatment *in vivo*

Mice were treated twice a day with a dose of 4mg *per os*, starting three days before cancer cell inoculation and continued until the end of the experiment. Control mice were treated with vehicle. The drug was synthesized in house and prepared as previously described[6].

Exome Sequencing Analysis

Genomic DNA from each tumor organoid (3 µg, quantitated by Qubit fluorometer DNA Hs Assay) was fragmented and Exome capture was performed using Nimblegen Sequence Capture EZ Human Exome Library v2.0. Libraries were sequenced on a HiSeq 2000 platform (Illumina), 2x 100bp paired end (PE) Sequencing. Fastqc files were generated by FastQC software v0.10.1.

SNP and copy number calling in exome sequencing data

Samples were aligned to the human reference genome version GRCh37[42] using the BWA aligner software[43] with default parameters. Preprocessing included removing of

duplicate reads[44] and base quality recalibration using BaseRecalibrator from the Genome Analyzer Toolkit[45]. Local realignment was performed around INDELs (IndelRealigner from[45]). The UnifiedGenotyper algorithm from[45] was used to call SNPs in the merged file (see[46,47] for details). SNPs were filtered with parameters recommended in the Best Practices of the GATK website[48], removing candidates that fulfilled at least one of the following criteria: QD<2.0, FS>60.00, MQ<40.0 and DP<10. Variants were annotated using the dbSNP database version 13, the UCSC human RefGene69 and the software snpEFF[49,50].

Copy number alterations discovery

Total counts per probe in the enrichment array were computed for each sample after removing duplicates using the Rsamtools package[51] for R[52]. Gene annotations for the probes were downloaded using the UCSC browser website[53]. GC content was computed for each probe using the package BSgenome.Hsapiens.UCSC.hg19[54], and the rtracklayer package[55]. Probe counts were normalized by exon length. Normalized counts were used to correct for GC content sequencing biases. A generalized additive model was fitted to the data through the function gam in the mgcv package[56]. Normalized and GC content corrected counts were used to compute gene level copy number estimates by calculating the mean across exon estimates.

The genomes were segmented based on corrected counts using the segment function from the DNAcopy R package[57]. Segments with mean lower than -3.5 (higher than 0.5) were considered deletions (amplifications). The rest were classified as diploid.

REFERENCES (for online methods)

28. Barrett, T. & Edgar, R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods Enzymol 411, 352-69 (2006).

29. Jorissen, R.N. et al. Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer. Clin Cancer Res 15, 7642-7651 (2009).

30. de Sousa, E.M.F. et al. Methylation of cancer-stem-cell-associated wnt target genes predicts poor prognosis in colorectal cancer patients. Cell Stem Cell 9, 476-85 (2011).

31. Benjamini, Y., Krieger, A.M. & Yekutieli, D. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93, 491-507 (2006).

32.   Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99, 6567-72 (2002).

33.   Merlos-Suarez, A. et al. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell Stem Cell 8, 511-24 (2011).

34.   Lombardo, Y. et al. Bone morphogenetic protein 4 induces differentiation of colorectal cancer stem cells and increases their response to chemotherapy in mice. Gastroenterology 140, 297-309 (2011).

35.   Morikawa, K., Walker, S.M., Jessup, J.M. & Fidler, I.J. In vivo selection of highly metastatic cells from surgical specimens of different primary human colon carcinomas implanted into nude mice. Cancer Res 48, 1943-8 (1988).

36.   Morikawa, K. et al. Influence of organ environment on the growth, selection, and metastasis of human colon carcinoma cells in nude mice. Cancer Res 48, 6863-71 (1988).

37.   Bottinger, E.P., Jakubczak, J.L., Haines, D.C., Bagnall, K. & Wakefield, L.M. Transgenic mice overexpressing a dominant-negative mutant type II transforming growth factor beta receptor show enhanced tumorigenesis in the mammary gland and lung in response to the carcinogen 7,12-dimethylbenz-[a]-anthracene. Cancer Res 57, 5564-70 (1997).

38.   Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545-50 (2005).

39.   Cunningham, I.H.W.E.F.L.T.M.H.G.H.S.J. Weka: Practical Machine Learning Tools and Techniques with Java Implementations. Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems (1999).

40.   Warren, R.S., Yuan, H., Matli, M.R., Gillett, N.A. & Ferrara, N. Regulation by vascular endothelial growth factor of human colon cancer tumorigenesis in a mouse model of experimental liver metastasis. J Clin Invest 95, 1789-97 (1995).

41.   Hu, Y. & Smyth, G.K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J Immunol Methods 347, 70-8 (2009).

42.   Lander, E.S. et al. Initial sequencing and analysis of the human genome. Nature 409, 860-921 (2001).

43.   Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-9 (2009).

44.   O'Brien, C.A. et al. ID1 and ID3 regulate the self-renewal capacity of human colon cancer-initiating cells through p21. Cancer Cell 21, 777-92 (2012).

45.   McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-303 (2010).

46.   DePristo, M.A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43, 491-8 (2011).

47.   Pathmanathan, S., Hamilton, E., Atcheson, E. & Timson, D.J. The interaction of IQGAPs with calmodulin-like proteins. Biochem Soc Trans 39, 694-9 (2011).

48.   Clevers, H. & Batlle, E. EphB/EphrinB receptors and Wnt signaling in colorectal cancer. Cancer Res. 66, 2-5 (2006).

49.   Cingolani, P. snpEff: Variant effect prediction. (2012).

50.   Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6, 80-92 (2012).

51.   Morgan, M., Pages, H. & Obenchain, V. Rsamtools: Binary alignment (BAM), variant call (BCF), or tabix file import. R package version 1.16.1,.

52.   R Core Team; R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2014).

53.   Kent, W.J. et al. The human genome browser at UCSC. Genome Res 12, 996-1006 (2002).

54.   Team, T. BSgenome.Hsapiens.UCSC.hg19: Full genome sequences for Homo sapiens (UCSC version hg19). R package version 1.3.1000.

55.   Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics 25, 1841-2 (2009).

56.   Wood, S.N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the ROyal Statistical SOciety (B) 73, 3-36 (2011).

57.   Seshan, V.E. & Olshen, A. DNAcopy: DNA copy number data analysis. R package version 1.38.1.

Figure1

a

De Sousa E Melo et al.
GSE33113

All *HR* > 1   *CCS3*-upregulated
P < 0.01   > 2 fold *P* < 0.05

543   160   431

CCS3 ∩ HR

Sadanandam et al.
GSE14333

All *HR* > 1   *Stem-like*-upregulated
P < 0.01   > 2 fold *P* < 0.05

612   286   313

Stem-like ∩ HR

Marisa et al.
GSE39582

All *HR* > 1   *C4*-upregulated
P < 0.01   > 2 fold *P* < 0.05

470   98   668

C4 ∩ HR

b

Z-score mean

All *HR* > 1
P < 0.01

CCS3 ∩ HR

Ep Str   Ep Str
normal   tumor

All *HR* > 1
P < 0.01

Stem-like ∩ HR

Ep Str   Ep Str
normal   tumor

All *HR* > 1
P < 0.01

C4 ∩ HR

Ep Str   Ep Str
normal   tumor

Figure 2

a

CRC Patient → Primary tumor n = 14

- EpCAM
- CD45
- CD31
- FAP

b

DCN
PDPN
FAP

VWF
CDH5
ENG

PTPRC
CD2
CD19

KRT20
CDH1
EPCAM

EpCAM   CD45   CD31   FAP

c

GSE33113

All HR > 1
P < 0.01

CCS3 ∩ HR

GSE14333

All HR > 1
P < 0.01

Stem-like ∩ HR

GSE39582

All HR > 1
P < 0.01

C4 ∩ HR

Z-score mean

EpCAM CD45 CD31 FAP

Figure 3

a

HiC

| Goblet-like | TA | Inflammatory | Enterocyte | Stem-like |

CAF Enriched

Leukocyte Enriched

Epithelial Enriched

HR

Epithelial (LCM)
Stroma (LCM)
EpCAM
CD31
CD45
FAP

Expression (log 2)

−1.5   −1.0   −0.5   0   0.5   1.0   1.5

HR (log 2)

−1.0   −0.5   0   0.5   1.0

b

CAF cluster expression
on good prognosis subtypes
(GOB + TA + INFL + ENT)

log Hazard ratio

GP-LoC   GP-HiC

HR (+1SD) = 2.20
P = 0.003

CAF cluster *Z*-score mean (SD)

−2   −1   0   1

c

— Good prognosis subtypes-Low CAF cluster (GP-LoC)
— Good prognosis subtypes-High CAF cluster (GP-HiC)
— Stem cell-like poor prognosis subtype (Stem-like)

% Recurrence free

GP-LoC
n = 98

GP-HiC
n = 61

Stem-like
n = 38

HR

Stem-like vs GP-LoC = 3.82	*P* = 0.001
Stem-like vs GP-HiC = 1.19	*P* = 0.626
GP-HiC vs GP-LoC = 3.22	*P* = 0.003

Recurrence (months)

0   20   40   60   80   100   120

Figure 4

Figure 5

Figure 6

a

CRC cells (TGFβ MUT)

TGFβ

TGFβ + FIB 5x10⁴

DFS / Time (days)

P = 0.0017

FIB
HT29-M6 10³

HT29-M6 TGFβ
- 10³
- 10³ + FIB
- 10²
- 10² + FIB

b GFP

HT29-M6 TGFβ + FIB-GFP

c

CRC cells (TGFβ MUT)

TGFβ

TGFβ + FIB

TGFβ + FIB-KD

DFS / Time (days)

**

****

HT29-M6 TGFβ
HT29-M6 TGFβ + FIB-shCon
HT29-M6 TGFβ + FIB-shTBRI
HT29-M6 TGFβ + FIB-Con
HT29-M6 TGFβ + FIB-DNR

d

CRC cells (TGFβ MUT)

+ FIB

+ pre-treated FIB

DFS / Time (days)

****

HT29-M6
HT29-M6 + FIB
HT29-M6 + FIB pre-treated

e

CRC cells (TGFβ MUT)

Con

+ FIB

TGFβ

+ FIB

1 TIC in every (x10³)

***
***
***
***

f

P = 0.0001

Mice with tumors (%) / Injected cells x10³

HT29-M6
- Con
- TGFβ
- Con + FIB
- TGFβ + FIB

Figure 7

a

CRC Patient → Primary tumor → CoCSCs → In vitro expansion → Patient derived tumor organoids

b

#1 (MUT) #2 (WT)
#3 (WT) #4 (WT)
#5 (WT) #6 (WT)
#7 (MUT) #8 (MUT)

c

#1 #2 #3 #4
#5 #6 #7 #8

Number of tumoroids

Con TGFβ LY

d

#1 #2 #3 #4 #5 #6 #7 #8

Control

TGFβ (7days)

e

		#1	#2	#3	#4	#5	#6	#7	#8
Cell cycle	*CDKN1A*	1.3	8.2	2.5	3.7	1.4	2.9	ND	-5.0
	CDKN2A	1.3	65.5	4.8	1.7	1.3	9.1	2.5	1.4
	CDKN2B	1.0	42.0	4.0	21.5	ND	6.4	3.2	3.1
	MYC	1.1	-2.5	-2.5	1.4	1.4	-3.3	-3.3	1.7
	KI67	1.3	-5.0	1.4	-2.5	1.2	-2.5	1.1	1.7
EMT	*SNAIL*	1.7	2.0	1.0	2.0	1.1	1.4	1.1	1.1
	TWIST1	2.4	4.4	1.5	1.7	ND	1.4	1.0	ND
	ZEB1	1.3	1.1	1.4	ND	ND	ND	-100	ND
Epithelial	*CDH1*	1.3	1.3	1.1	1.4	1.1	2.0	1.0	1.0
	EPCAM	1.3	1.5	1.3	1.2	3.1	1.5	1.4	1.0
	CLAUDIN1	1.1	1.5	1.3	1.4	2.0	1.4	1.1	1.3

Fold change (TGFβ / Con)

≥ 2.5
≥ 7
≤ -2.5
ND (not detectable)

Figure 8

a

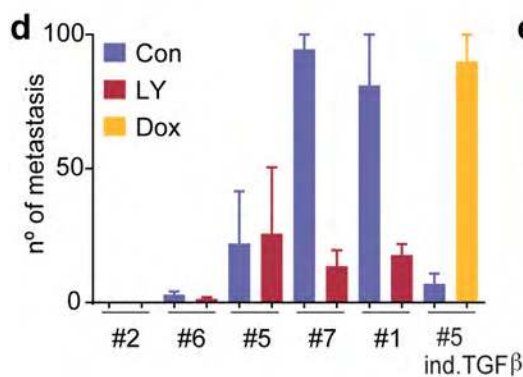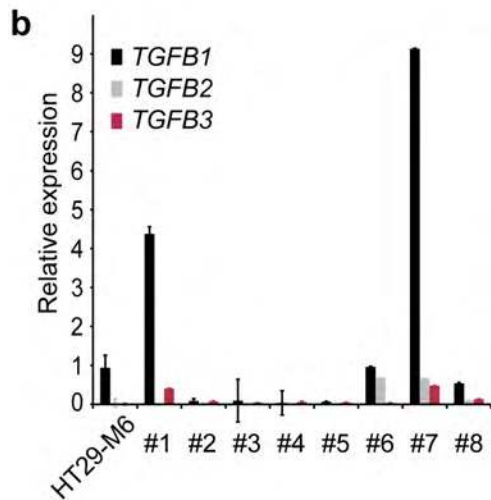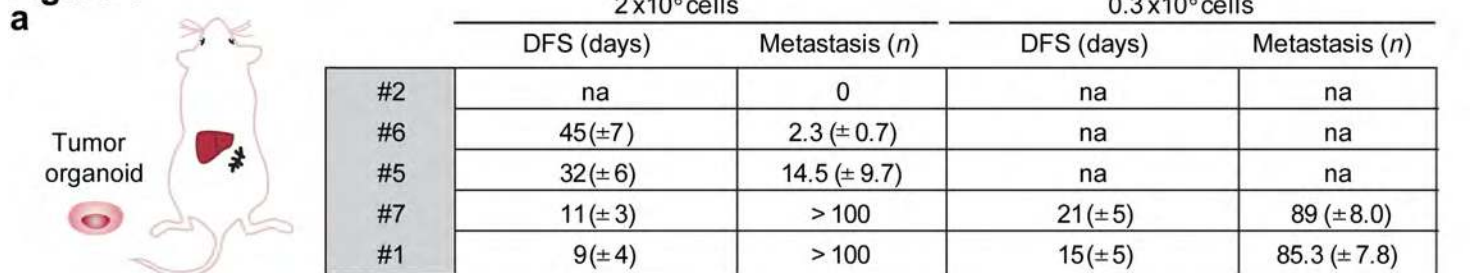|	2 × 10⁶ cells		0.3 × 10⁶ cells	
	DFS (days)	Metastasis (n)	DFS (days)	Metastasis (n)
#2	na	0	na	na
#6	45 (±7)	2.3 (± 0.7)	na	na
#5	32 (± 6)	14.5 (± 9.7)	na	na
#7	11 (± 3)	> 100	21 (±5)	89 (±8.0)
#1	9 (± 4)	> 100	15 (±5)	85.3 (± 7.8)

Statistics and general methods

Gene differential expression analysis was carried out using linear models and empirical Bayes shrinkage[1] as implemented in *limma* R package[2]. Benjamini and Yekutieli correction[3] was applied for multiple contrast adjustment. To summarize the signatures analyzed in this work, *Z*-scores were computed for each gene and each sample, which were then averaged across all genes included in the profile. When their assumptions were met, signature differences between groups were assessed using linear models and t-tests for pairwise comparisons; otherwise, Kruskal and Mann-Whitney tests were applied, respectively.

When evaluating survival differences, Kaplan-Meier estimates and Cox proportional hazard models were used. A Likelihood Ratio Test (LRT) was applied to assess significance of covariates while Wald tests were used for pairwise comparisons between groups. For continuous variables, we obtained smooth estimates of the relationship between the covariate and recurrence hazard ratio (HR) and their corresponding 95% confidence interval bands, as implemented in the R package *phenoTest*[4]. In addition, the linear gain in risk due to a unit increase in the continuous variable was also reported. Gene signature summaries were centered and scaled when assessing their association with recurrence.

Sample sizes were chosen empirically following previous experience in the assessment of experimental variability. We used a standard log rank test to calculate mouse survival.

If not explicitly indicated, statistical significance was defined at the 5% level. Most of the analyses in this work were performed using tools from R[5] and Bioconductor[6].

Processing of microarray data

Transciptomic datasets corresponding to colorectal cancer (CRC) patients (GSE33113[7], GSE14333[8] and GSE39582[9]) and laser microdissected tumors (GSE35602[10]) were downloaded from GEO[11] repository. Preprocessed series matrixes originally provided by the authors were used in the analyses. GSE39582 raw cel files were also processed using RMA normalization method as implemented in Bioconductor R package *affy*[12] and corrected by scanner batch using a linear model in which recurrence, age at diagnosis, gender, tumor location, stage, treatment with adjuvant chemotherapy (Yes/No) and missmatch repair status (MMS)

were also included as covariates. This version of GSE39582 dataset was used to perform the disease-free survival analyses of CAF cluster derived signature for Marisa et al classification (see Contribution of genes expressed by epithelial or stromal cells to the molecular classification of CRC in the Online Methods section). Expression data from tumor sub-population cells (GSE39395 and GSE39396)[13] and from colonic fibroblasts (FIB) treated or not with TGF-beta1 (deposited in GEO with accession number GSE64192) were also processed using RMA.

Annotation of Affymetrix microarray probesets was performed using Bioconductor R package *hgu133plus2.db*[14]. Probesets from dataset GSE35602 (Agilent-014850 Whole Human Genome Microarray 4x44K G4112F) were mapped to their corresponding gene symbols using the annotation available in GEO for this platform. Unavoidably, some genes mapped by Affymetrix were missing in GSE35602 and, therefore, they were removed when both platforms were used in the same analysis.

Expression of poor-prognosis subtypes in sub-population cells
Three GEO datasets were used to characterize the subtypes gene profiles according to their specific expression in tumoral cell sub-populations: GSE39395, GSE39396, and GSE35602. In order to explore the sensitivity of these results to the thresholds used to define the subtypes and recurrence signatures, we computed de HR means and the FC upregulation means of micro dissected tumor stroma compared to epithelial tumor areas in dataset GSE35602 for probesets selected at different cutoffs of these two parameters. HR means and FCs means of selected probesets were compared to those of the rest of probesets not included in the signature using a t-test.

Contribution of genes expressed by epithelial or stromal cells to the molecular classification of CRC
In order to characterize the CRC subtypes being studied in this work, expression signatures defining the molecular classifications were retrieved from their original paper and represented as a heatmap at probeset level. De Sousa E Melo et al.[15] and Marisa et al.[9] reported their profiles as a list of selected Affymetrix probesets (148 and 1459, respectively). Sadanandam et al.[16] provided a list of genes (786) with no probeset specification. Because of that, we included in Sadandandam profile all probesets mapping to any gene of that list (1963 probesets in total).

In the heatmap representation, only genes with available annotation based on risk to relapse and upregulation in stroma or epithelial cells were included. Therefore, genes mapped by Affymetrix probesets not found in dataset GSE35602 were excluded from this analysis. The final list of probesets for clustering included 138 probesets for De Sousa E Melo et al. signature, 1938 probesets from Sadanandam et al. and 1286 probesets from Marisa et al. molecular classification.

In the probeset clustering, clusters definition was based on the height between branches in the dendrogram as well as on the probeset annotation based on risk to relapse and upregulation in stroma or epithelial cell.

Dataset GSE33113 included a low number of relapses within good-prognosis subtypes according to De Sousa E Melo et al. classification (CCS1 and CCS2, n=6). Therefore and to increase the statistical power, we evaluated the prognosis association of the corresponding CAF cluster signature in dataset GSE14333. For doing so, we retrieved the classification provided in [15] for this public dataset.

Processing of dataset GSE39582 included a batch effect correction that did not take into account the relapse information as variable of interest[9]. Due to that, associations of gene expression with prognosis can potentially be attenuated in this dataset. For this reason, we downloaded the raw cel files from this dataset and processed as previously described (section Description of CRC transcriptomic datasets). All analyses assessing association of CAF cluster signature with recurrence were performed on this reprocessed version of GSE39582.

To evaluate the impact of cell specific genes in the molecular classification, we used a PAM[17] classifier to classify subtypes using their signature probesets as predictors. We chose this classification technique as it was used by Sadanandam et al. and De Sousa E Melo et al. as a final step of their signature definition and, therefore, it provided a way to compare the performance of our analyses. In addition, a similar centroid based method was also used in Marisa et al. to build their profile. For this analysis and accordingly to [16], Sadanandam et al. signature was summarized at the gene level using the most variable probeset within each gene (standard deviation). To evaluate the impact of the number of predictors in the results when evaluating a subset of probesets, 100 runs of these analyses were conducted using the same number of randomly selected probesets from within the corresponding signature. The

mean of the error rate (global and by subtype) was used as a measure of classification performance and compared to those observed in the real scenarios.

Analysis of TMAs by IHC

To assess the association between tumor relapse and protein intensity staining, we computed a smooth estimate of the relationship between staining intensity and recurrence hazard ratio. Additionally, we computed Kaplan-Meier survival curves for patients with low, medium and high intensity staining. Whenever was possible, these groups were defined using the corresponding intensity tertiles as cutoffs. As epithelium measures showed a high proportion of zeros, intensities were categorized as "zero", "between zero and M" and "greater than M", where M was the median intensity computed for non-zero values in that measure. Statistical significance was assessed introducing each set of protein values as continuous or as categorical covariate in a Cox proportional hazards model, for which corresponding HRs were computed.

As these samples were distributed across two independent Tissue Microarrays (TMAs), this information was included as an adjusting covariate in the Cox models and was used to adjust the Kaplan-Meier curves (estimated from the Cox model) and the smoothing estimate of the HRs. Moreover, categorizations of staining intensities were performed separately in each TMA.

REFERENCES for SUPPLEMENTARY NOTE

1. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
2. Smyth, G.K. Limma: linear models for microarray data. in *In: Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds. Gentleman, R., Carey, V., Dudoit, S., Irizarry, R. & Huber, W.) 397-420 (Springer, New York, 2005).
3. Benjamini, Y., Krieger, A.M. & Yekutieli, D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491-507 (2006).
4. E, P. phenoTest: Tools to test association between gene expression and phenotype in a way that is efficient, structured, fast and scalable. We also provide tools to do GSEA (Gene set enrichment analysis) and copy number variation. *R package v 1.12.0* (2013).
5. Team, R.C. A language and environment for statistical computing. (2014).
6. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).

7.   de Sousa, E.M.F. *et al.* Methylation of cancer-stem-cell-associated wnt target genes predicts poor prognosis in colorectal cancer patients. *Cell Stem Cell* **9**, 476-85 (2011).

8.   Jorissen, R.N. et al. Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer. Clin Cancer Res 15, 7642-7651 (2009).

9.   Marisa, L. *et al.* Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* **10**, e1001453 (2013).

10.   Nishida, N. *et al.* Microarray analysis of colorectal cancer stromal tissue reveals upregulation of two oncogenic miRNA clusters. *Clin Cancer Res* **18**, 3054-70 (2012).

11.   Barrett, T. & Edgar, R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol* **411**, 352-69 (2006).

12.   Irizarry, R.A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-64 (2003).

13.   Calon, A. *et al.* Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. *Cancer Cell* **22**, 571-84 (2012).

14.   Carlson M. *hgu133plus2.db: Affymetrix Human Genome U133 Plus 2.0 Array annotate on data (chip hgu133plus2)*. R package version 3.0.0.

15.   De Sousa, E.M.F. *et al.* Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med* **19**, 614-8 (2013).

16.   Sadanandam, A. *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* **19**, 619-25 (2013).

17.   Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* **99**, 6567-72 (2002).