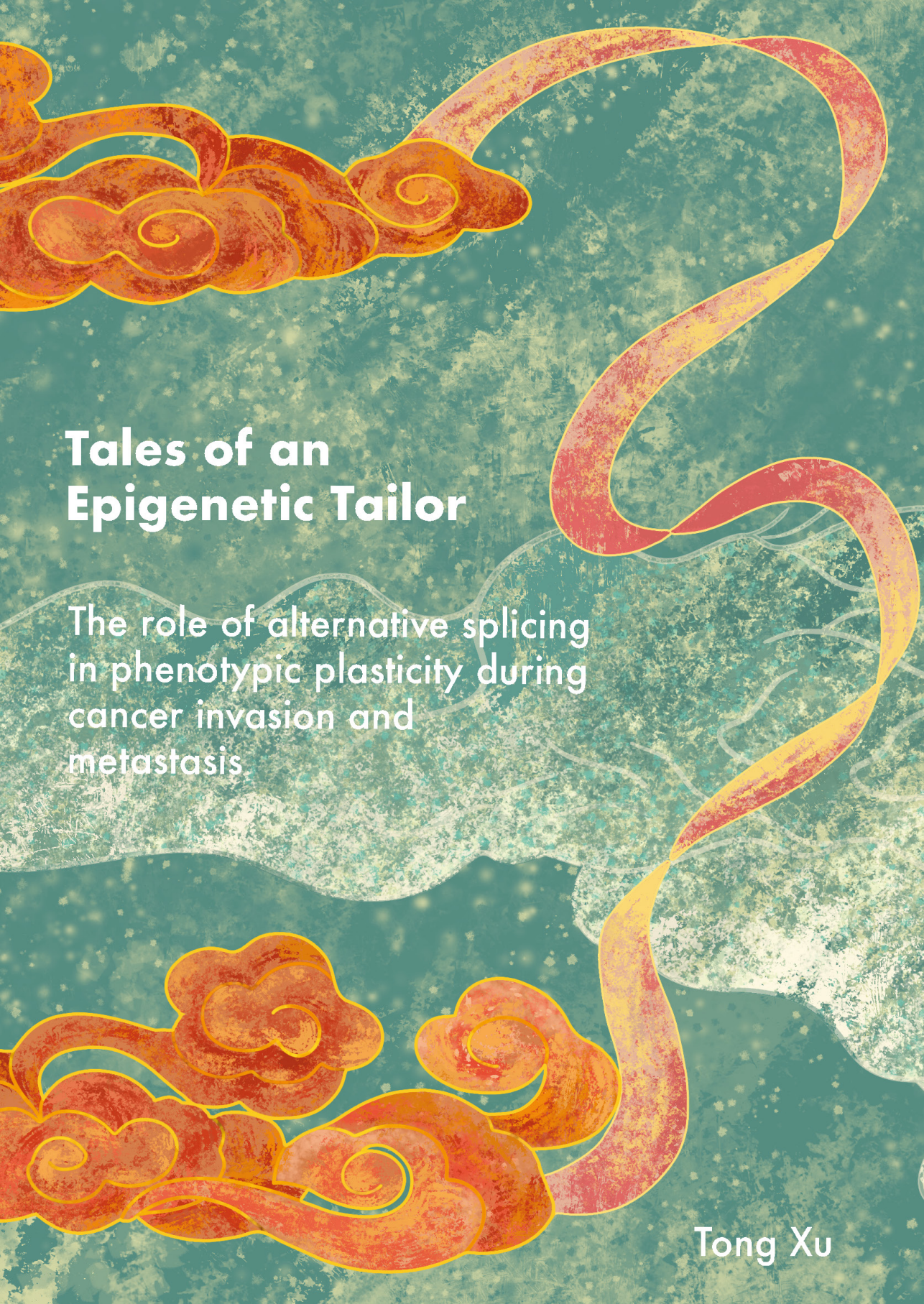


Tales of an Epigenetic Tailor

The role of alternative splicing in phenotypic plasticity during cancer invasion and metastasis

Tong Xu

Tales of an Epigenetic Tailor:

The role of alternative splicing in phenotypic plasticity during cancer invasion and metastasis

Tong Xu

COLOFON

Author: Tong Xu
ISBN 978-94-6506-215-0
Cover design: Tong Xu
Printed by: Ridderprint BV, Ridderkerk

The research described in this thesis was conducted at the Department of Pathology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, the Netherlands.

Tales of an Epigenetic Tailor:
The role of alternative splicing in phenotypic plasticity during cancer invasion and metastasis

Verhalen van een Epigenetische Kleermaker:
De rol van alternatieve splicing in fenotypische plasticiteit bij kankerinvasie en metastase

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
Rector Magnificus

Prof. dr. A.L. Bredenoord

and in accordance with the decision of Doctorate Board.
The public defence shall be held on

Tuesday 17 September 2024 at 15.30hrs

Tong Xu
Born in Shenyang, China.

Erasmus University Rotterdam

Doctoral committee

Supervisor:	Prof. dr. R. Fodde
Other members:	Prof. dr. G. W. Jenster
	Prof. dr. W. M. van Weerden
	Prof. dr. M.A.G.G. Vooijs

CONTENT

CHAPTER 1

General Introduction

1. The two phases of cancer evolution: genetic initiation and epigenetic progression

Cancer is a complex and multifaceted disease that arises from a single (epi)genetic event to gradually progress into a potentially fatal condition. This process involves intricate cellular and molecular mechanisms and encompasses different phases. Understanding the cellular and molecular mechanisms which underlie its progression towards malignancy is essential for cancer diagnosis, prognosis, and treatment[1].

Cancer onset begins with the occurrence of somatic genetic changes in a single cell. These are genomic DNA mutations or DNA modifications caused by a broad spectrum of factors, including exposure to carcinogens and/or other cytotoxic environmental agents[2]. These genetic alterations are selected upon the advantage they provide to the nascent cancer cell thus allowing its clonal expansion and the formation of a cluster of abnormal cells. Promoting factors include hormones, inflammation, and additional (epi)genetic alterations[3,4]. In general, tumor progression at the primary site is earmarked by sequential clonal expansion, where each new mutation will provide additional selective advantages. As the cluster of abnormal cells grows, these somatic mutations will progressively confer one or more of the so-called "hallmarks of cancers"[5]. Hence, one can say that this first phase of the evolution of a tumor is strictly "Darwinian", meaning that the somatic alterations are selected by the cancer cell because of the selective advantage they provide. However, following this phase, the initially benign tumor start to spread to nearby tissues, i.e. it becomes malignant[6]. Tumor cells invade through the basement membrane and infiltrate the surrounding stromal microenvironment. The invasive behavior of cancer cells allows them to further migrate to distant organ sites where they will seed and form metastases. As explained here below, this second phase of tumor evolution is mainly driven by transient and reversible epigenetic changes rather than DNA mutations.

Local invasion and distant metastasis represent the clinically most relevant aspects of cancer progression. Here, cancer cells spread from the primary tumor and migrate through the bloodstream or lymphatic system to eventually seed in distant organs or tissues. The so-called "invasion-metastasis cascade" involves different obstacles that the disseminated cancer cells need to overcome to form a secondary lesion in a distant organ (Figure 1). To this aim, cancer cells adopt different strategies to be able to become motile, locally invade the TME, intravasate lymphatic or blood vessels, survive in circulation by evading the surveillance of the immune system[7], extravasate the vessel in the proximity of the distant organ where they will eventually seed and form a metastasis. Moreover, during this process cancer cells develop resistance to chemo-, radio- and targeted therapies[8] which will result in treatment failure and in loco-regional and distant recurrence[9].

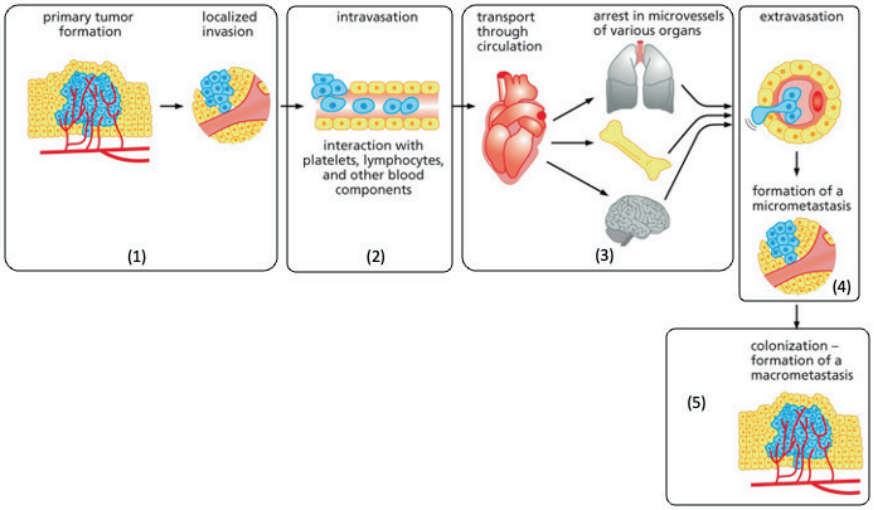


Figure 1. Schematic representation of the invasion-metastasis cascade.

The process unfolds in distinct stages: (1) infiltration of epithelial cells from the primary tumor into the adjacent stroma and extracellular matrix, (2) intravasation of cancer cells into the bloodstream, (3) dissemination to distant sites, (4) formation of micrometastases, and (5) establishment of secondary tumors in distant organs.

Comparative genomic and gene expression analysis of matched primary tumors with the corresponding metastases showed the absence of main differences as far as driver gene mutations are concerned[10]. This so-called "*progression puzzle*" indicates that the transient and reversible mechanisms that underlie this second phase of tumor progression (i.e. local and systemic dissemination, and metastasis) are more likely to be of transient, epigenetic nature.

A range of epigenetic processes, including DNA methylation, histone modification, and microRNA activity, orchestrate phenotype changes in cancer metastasis[11]. These epigenetic mechanisms can induce changes in physical DNA structure, lead to DNA sequence retention, and in the context of cancer, silence genes related to tumor suppression[12]. DNA methylation, for instance, triggers gene mutations or affects the promoter regions of tumor suppressor genes, causing their silencing due to CpG island promoter hypermethylation[13,14]. Moreover, alternative splicing is recognized as another contributor to oncogenesis, enhancing proteome diversity and influencing key carcinogenic pathways, thus underscoring the multifaceted nature of gene regulation in cancer[15].

This "biphasic" nature of cancer progression underscores the existence of two distinct subpopulations of malignant cells, stationary and migrating CSCs, that coexist in the tumor microenvironment. Migrating CSCs present challenges for effective treatment due to their aggressive nature, characterized by rapid invasion and metastasis[16,17]. Furthermore, migrating CSCs exhibiting EMP characteristics are highly plastic and can form distant metastases. On the other hand, stationary CSCs primarily exhibit an epithelial nature and are highly proliferative, serving as

a source of "fuel" for surrounding cancer cells. Understanding the existence of these different subpopulations within the same cancer type is of significant importance in cancer biology, progression, and therapeutic strategies[18] [16].

In summary, onset and malignant progression of a carcinoma are distinct phases featuring different underlying cellular and molecular mechanisms. While genetic changes alone can not provide a solution to the progression puzzle[19], the transient and reversible nature of the adaptive changes disseminated cells undergo points to the pivotal role played by epigenetic regulation in the clinically most relevant hallmark of cancer, namely "*activating invasion and metastasis*".

2. Phenotypic Plasticity and EMT in Cancer

In order to shed light on the epigenetic mechanisms which underlie the invasion-metastasis cascade, we need to introduce the concept of phenotypic plasticity in cancer, here defined as the ability of one genotype to produce more than one phenotype when exposed to different environments. Phenotypic plasticity plays an important role in homeostasis and has been reported across a broad spectrum of species in the animal kingdom, usually associated with the organism's regenerative capacity[20]. In cancer, this plasticity endows cells with the ability to transiently and temporarily alter their cellular identity so to overcome the multiple obstacles present along the route to metastasis[21]. Epithelial-mesenchymal transitions (EMT) and the reverse process, mesenchymal-epithelial transition (MET), are the quintessential features of phenotypic plasticity, and are often collectively described as epithelial-mesenchymal plasticity (EMP), which plays a central role in the "*activating invasion and metastasis*" hallmark of cancer and is thought to represent a relevant target for future therapeutic interventions[22].

As mentioned above, EMP is not exclusive to cancer: it plays important functional roles in embryonic development, wound healing, tissue regeneration, and fibrosis[22]. EMP encompasses the interplay of stemness and plasticity between epithelial and mesenchymal cells. However, it is not a binary system where cells transit between a fully epithelial (E) and mesenchymal (M) states. Instead, it progresses sequentially through intermediary stages, each with distinct degrees of plasticity. In fact, cancer cells in a fully E or M state tend to be less plastic that the intermediate quasi-mesenchymal or quasi–epithelial cells that usually display weaker cell-cell adhesion and are more metastable during invasion and localization[17]. Recent research has identified cancer cells co-expressing both E- and M-specific genes, usually referred to as " hybrid" or "partial EMT" (pEMT), capable of shifting between epithelial and mesenchymal states[23,]. pEMT cells constitute a small subpopulation of cancer cells that is extremely challenging to identify because the lack of specific markers and the impossibility to implement lineage tracing experiments to demonstrate their central role in cancer metastasis[24] As pEMT cells are likely to possess the highest metastatic potential and cellular plasticity[25], their multiomics profiling and the identification of specific markers is of great basic and translational relevance[26,27].

In conclusion, phenotypic plasticity underlies much of the complexity of malignant progression where cancer cells adapt their form and function in response to changing environments. The subpopulation of cancer cells with quasi-mesenchymal characteristics, constantly transitioning between epithelial and mesenchymal states, play a crucial role in tumor metastasis, invasion, and drug resistance, impacting cancer treatments and in particular the resistance to chemotherapy.

3. Alternative Splicing and RNA-binding proteins (RBPs): Shaping Cancer Phenotypes

Alternative splicing (AS) describes how the splicing of a single primary transcript (pre-mRNA) can result in a variety of different mature mRNAs. This enables the expression of protein isoforms with different structural and functional properties from a single coding gene[28]. Pre-mRNA transcripts are alternatively spliced to produce mRNA isoforms with different stability or coding potential, thus generating proteins with different protein-protein interactions, subcellular localization, and stimulated or inhibited activity[29]. About 92–94% of human multi-exon genes are alternatively spliced to produce different protein isoforms. Pre-mRNA alternative splicing constitutes a form of epigenetic regulation in cancer. Several AS modalities have been reported, including skipped exons (SE), alternative 3′-splice sites (A3SS), alternative 5′-splice sites (A5SS), mutually exclusive exons (MXE), and retained intron (RI)(Figure 2). Among these, the most frequent in mammalian cells is SE, where one or more exons are entirely skipped[30]. Furthermore, alternative splicing encompasses a broad definition that includes alternative promoter usage. There exists a functional correlation between alternative promoters and alternative splicing, with both promoter and transcription factors influencing alternative splicing. Moreover, genes with multiple alternative promoters tend to have a greater number of alternative splicing variants. This underscores a positive correlation between alternative promoters and alternative splicing[31].

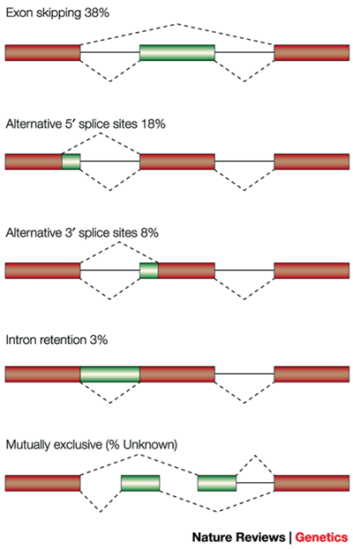


Figure 2. Alternative Splicing.

Schematics depicting the various modalities of alternative splicing, including, from top to bottom, exon skipping, alternative 5′ splice site, alternative 3′ splice site, intron retention, and mutually exclusive exons. Red: exons are not spliced; Green: alternatively spliced exons. Introns are shown in solid lines, dashed lines represent alternative splicing events[30].

Three theories have been put forth to explain the mechanisms underlying alternative splicing: exon shuffling, exoneration of intron, and transformation of a constituent exon to another exon. It has been observed that exons have higher nucleosome occupancy compared to introns. The removal of an intron involves a chemical mechanism where a phosphodiester is excised from RNA polynucleotide chains to form new ones. This process can be broken down into two phases.

First, a 2′-5′ phosphodiester bond is generated between the intron's 5′ nucleotide and an internal adenosine residue upstream of the intron's 3′ end. Following this, two spliced exons combine, and the introns are released as a lariat product[15]. Nucleosomes slow down exon synthesis by retarding RNA polymerase II[32] , i.e. the enzyme that synthesizes pre-mRNAs encompassing introns and exons. Pre-mRNA AS has been shown to play important functional roles not only during embryogenesis[33] but also in tumor onset and cancer development, and as such may be the source of potential therapy targets[34].

RNA-binding proteins (RBPs) play key roles in alternative splicing. In general, AS requires the synergistic effects of multiple, rather than individual RBPs, to regulate the expression of splicing targets[35]. As different RBPs recognize different RNA sequences, the concept of 'splicing code' was introduced. Deciphering the splicing code requires a comprehensive catalog of RBP binding sites and their associated cis-acting binding sites. The latter play pivotal roles in deciding which introns or exons are retained or spliced. Cis-acting elements that control alternative splicing are referred to as exonic splicing enhancers/silencers (ESEs/ESSs) and intronic splicing enhancers/silencers (ISEs/ISSs) (Figure 3)[36]. These enhancers and silencers consist of sequences approximately 5-10 nt in length and are specifically recognized by the RBPs binding domain.

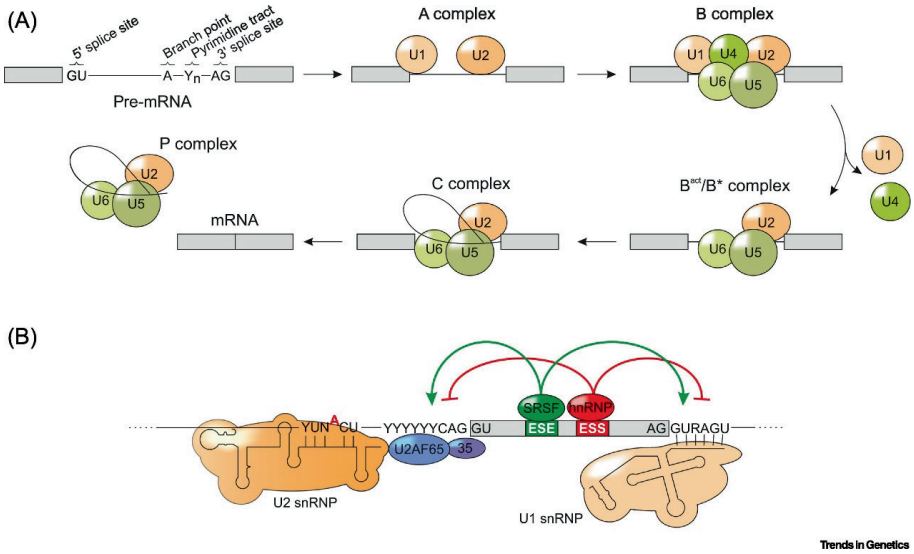


Figure 3. Splicing mechanism and its control.

(A) Schematic illustration of how the small nuclear ribonucleoprotein (snRNP) particles are assembled to form the spliceosome. The A complex is created when U1 binds to the 5′ splice site and U2 binds to the branch point. The B complex is created when U4, U5 and U6 join the A complex. The spliceosome is activated by the removal of U1 and U4. The C complex is created following the conclusion of the first splicing phase. The spliced mRNA product is released from the post-splicing P complex following the second splicing process. (B) RNA-binding proteins control splicing. By promoting the interaction of U2AF or the U1 snRNP with the RNA, SR proteins SRSF-bound exonic splicing enhancer (ESE) can induce splicing at both upstream and downstream splice sites. hnRNP proteins bind to exonic splicing silencers (ESS), which inhibit upstream and downstream splice sites by competing with SR proteins.

The spliceosome, composed of five small nuclear ribonucleic protein (snRNP) particles (U1, U2, U4, U5, and U6) and other associated proteins and polypeptides, catalyzes the removal of introns[37]. U1 and U2 have individual functions as single proteins, while U4 and U6 normally function as hetero-dimers and, with U5 as hetero-trimers. The different E, A, B, C and H complexes assemble the spliceosome on the pre-mRNA under the guidance of sequences located at the ends of the introns[38] (Figure 2A). The initiation of spliceosome assembly is triggered by U1 snRNP recognizing the 5′ splice site through binding of U1 snRNA's with 6–8 nucleotides at the 5′ end of the intron. The recognition of 3' splice site is accomplished by three independent elements recognized by three proteins, U2AF1, U2AF2, and SF1. These three sequence-interacting proteins can undergo mutated in cancer progression. The ATP-dependent binding of U2 snRNP around branch sites is then followed by U2 snRNA recruitment, leading to complex A formation (Figure 2B). One or multiple RBPs recognize specific RNA sequences, resulting in alternative splicing of variants and the generation of isoform proteins[39].

Moreover, RBPs have been demonstrated to function as co-transcriptional factors and perform various non-alternative splicing (non-AS) functions, extending beyond their role in alternative splicing alone. The regulation of alternative splicing is not only contingent on the interaction of RBPs with their target sequences in pre-mRNA; rather, it is linked to transcription factors. Transcription factors can impact alternative splicing by influencing promoter identity and occupation[40].

It has been shown that alternative splicing represents a key 'epigenetic' mechanism in EMT and MET in cancer, as various splicing factors and AS targets contribute to these processes[41] (Figure 3). Splicing factors can be divided into several groups based on binding profiles and functions, ranging from splicing regulators, 3′-splice site factors, SR proteins, to hnRNPs[35]. Most splicing regulators have been found to regulate downstream target genes and promote EMT. For example, the -Quaking (QKI), a conserved member of the STAR (signal transduction and activation of RNA) family of RBP proteins, is crucial for embryonic and postnatal development[42]. QKI mediates alternative splicing of a broad spectrum of downstream genes and promotes mesenchymal splicing patterns in different cancers[43]. QKI is both essential and sufficient for establishing an intermediate mesenchymal cell state and enhancing tumorigenicity in breast cancer. Additionally, QKI plays a causal role in controlling EMT by orchestrating the splicing and function of the actin-binding protein FLNB[44].

The RNA binding FOX (RBFOX) family of proteins encompasses three paralogs in mammals, RBFOX1, RBFOX2, and RBFOX3, each of which plays a role in regulating tissue-specific alternative splicing[45]. These three members of the RBFOX family share a core RNA-recognition motif that recognizes the consensus sequence (U)GCAUG located in the lateral introns of the target exons. RBFOX proteins interact with the upstream region of an alternative exon while excluding its downstream portion, thus promoting exon skipping[46]. Among these three RBFOX proteins, RBFOX2 is a crucial regulator of tissue-specific alternative splicing in both healthy and cancerous tissues, including its involvement in EMT[47]. RBFOX2 plays a crucial role in influencing the splicing of mesenchymal and epithelial cells, although its role as a regulator of epithelium-specific splicing is still a matter of debate when compared with other RBP families such as epithelial splicing regulatory proteins (ESPRs, see below). Alterations in the alternative splicing of genes such as Cortactin (*Cttn*), Par-3 Family Cell Polarity Regulator (*Pard3)*, and dynamin 2 (*Dnm2*) are correlated with increased RBFOX2 expression levels, while alternative splicing of genes like STE20 like kinase

(*Slk)*, and TSC complex subunit 2 (*Tsc2)* is induced by the downregulation or complete loss of RBFOX2[47]. Recent studies have revealed that RBFOX2 suppresses the ESRP1 expression by binding to the transcription process of Slug[48].

The SR (splicing regulator) family of splicing factors consists of 12 members (SRSF1-12) which share similar domains. SRSF1 and SRSF2 enhance exon splicing in a context-sensitive manner, while SRSF3 and SRSF4 exhibit slightly higher exon splicing than intronic region splicing. SR proteins recognize splicing elements in pre-RNA and can either promote or inhibit splicing events by recruiting and assembling spliceosomes[35]. Aberrant splicing resulting from abnormal SR protein production in tumors encourages tumor cell proliferation, migration, and resistance to apoptosis[49-51] (Figure 4).

Apart from the splicing factors mentioned above, the epithelial splicing regulatory proteins (ESRPs) 1 and 2 play a central role in regulating alternative splicing of a broad spectrum of target genes during EMT[43]. Both ESRP1 (RBM35a) and ESRP2 (RBM35b) belong to the hnRNP family[52]. ESRP1, identified as a potential master regulator of EMT in somatotroph adenomas, demonstrated transiently decreased levels of E-cadherin and alterations in the expression of EMT-associated genes upon inhibition of its expression in a growth hormone-producing rat pituitary cell line (GH3)[53]. The diminished expression of ESRP1 results in the production of the HMENA11A isoform, prompting a mesenchymal phenotype and fostering cell invasion under hypoxia[48].Recent findings indicate that FLNB in prostate cancer can undergo alternative splicing by both ESRP1 and ESRP2[54]. However, the splicing mechanism and the functional roles of FLNB isoforms in prostate cancer remain to be fully understood.CD44 represents an exemplary ESRP EMT target as its conversion from the variant (CD44v) to the standard subtype (CD44s) is necessary for EMT and malignant progression in many types of cancer including breast[55] and lung cancer[56] (Figure 3).

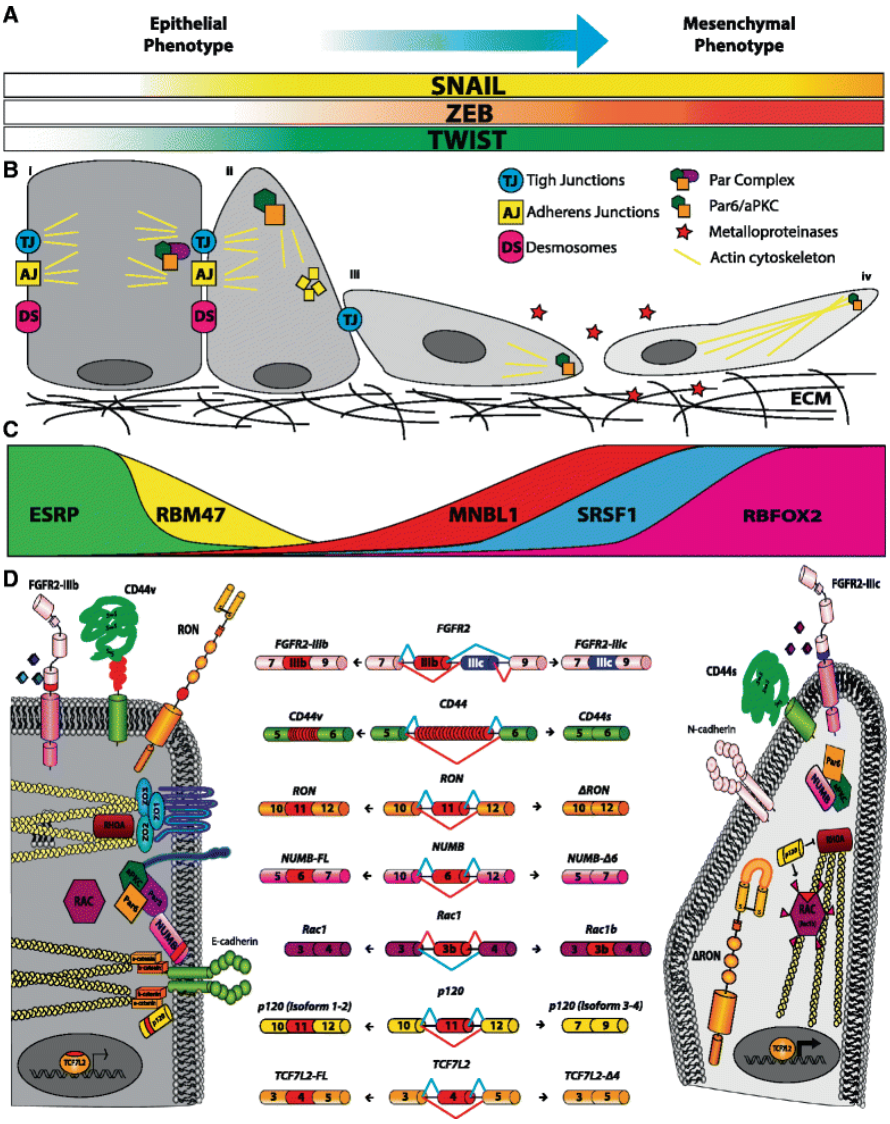


Figure 4. EMT RBPs.

(A) EMT-transcription factors (EMT-TFs) are regulated during the epithelial to mesenchymal phenotypic changes. (B) Cell junction progressively loosen as EMT progress. (C) Differentially expressed RBPs during EMT. (D) AS targets regulated by the above RBPs. Red: spliced exons [41,57].

Over the past two decades, alternative splicing in humans has provided novel insights into the genetic causes of a broad spectrum of pathologies, including malignancies[39,58]. Alternative splicing contributes to the establishment and regulation of phenotypic traits in different cell and tissue types. Likewise, it affects growth and malignant progression in cancer, as well as the cellular plasticity of EMT[41] (Figure 3). With the development of global sequencing and analysis technologies,

numerous important upstream SR proteins, RBPs, and downstream functional alternative splicing target genes have been identified in tumorigenesis. The hallmarks of alternative splicing in cancer include promoting unrestricted replication, suppressing growth inhibitors, activating invasion and metastasis, promoting angiogenesis, and deregulating cellular energetics[57].

Human tissue consists of various cell types, including stem cells, muscle cells, blood cells, and nerve cells. Numerous superimposed splicing mechanisms exist, which can make it challenging to associate specific proteins with particular RNA sequence motifs. This complexity limits our capacity to predict splicing regulation solely based on sequence features and tissue splicing data using *in silico* approaches. However, sequencing of RBP-binding site by high-throughput analysis combined with other sequencing methods such as crosslinking and immunoprecipitation (CLIP), systematic evolution of ligands by exponential enrichment (SELEX) and chromatin immunoprecipitation(ChIP)[59,60]. CLIP, or CLIP-seq, is a molecular biology technique that integrates UV crosslinking with immunoprecipitation, aiming to pinpoint RNA binding sites of proteins on a transcriptome-wide scale and enhance our understanding of post-transcriptional regulatory networks[60]. This method has led to the identification of RNA-binding proteins which not only play crucial roles during embryonic development (e.g., MBNL, PTBP2, RBFOX1, and RBFOX2)[61,62], but also promote tumor metastasis and invasion by modulating AS at downstream target genes closely related to EMT (ESRPs, RBMs, QKI, and hnRNPs)[44,55,63]. In addition to CLIP, other protocols, such as genomic SELEX, provide an experimental method for the condition-independent identification of protein-binding RNAs. This involves generating RNA libraries from genomic DNA, subjecting them to rounds of selection based on affinity to a target protein, and cloning and sequencing high-affinity ligands, enabling the discovery of protein-binding RNAs and the definition of whole protein–nucleic acid networks[59]. With the advent of faster sequencing methods, yet undiscovered splicing patterns will be identified.

In summary, RNA-binding proteins have been identified as key players in regulating alternative splicing events related to EMT and phenotypic plasticity in cancer. They can modulate the expression of EMT-related genes thereby affecting the behavior and characteristics of cancer cells. The dysregulation of these RNA-binding proteins and their impact on alternative splicing has been implicated in promoting cancer progression, metastasis, and treatment resistance by influencing the phenotypic plasticity of cancer cells. The elucidation of these mechanisms is of relevance for our fundamental understanding of splicing and holds potential for the identification of novel therapeutic targets and strategies to manage cancer.

4 . Metastatic Routes in Colon Cancer and Ovarian Cancer

It is evident that the complex nature of cancer presents various challenges at different stages of cancer development. A profound comprehension of these complexities is vital for both the diagnosis and effective treatment of the disease. In this thesis, I explored alternative splicing as one of the main mechanisms underlying a pivotal hallmark of cancer progression, i.e. the invasion and dissemination of cancer cells to distant organs and tissues, a phenomenon that significantly impacts patient outcomes.

We focused our investigations on AS as an essential EMT-driving mechanism in two specific types of malignancies, namely colon and ovarian cancer, selective because of their very distinct metastatic routes. The rationale for this choice was based on the assumption that distinct RBPs and AS target genes are expected to underlie the activation of E->M and M->E transitions and

phenotypic plasticity in these cancer types. Colon cancer is one of the deadliest cancers worldwide, with approximately 50% of CRC patients developing liver metastasis[64]. Colon cancer cells detach from the primary mass to disperse into the surrounding TME and eventually reach distant organs where to form metastatic lesions through hematogenous routes[65], much going through the multiple steps of the above mentioned invasion-metastasis cascade. Liver, lungs, lymph nodes, and peritoneum are the most common sites for colon cancer metastasis, while the brain and bones are less frequently affected. As predicted by the progression puzzle[19], studies have shown that the genomic distinctions between primary and metastatic tumors are relatively minimal, with both primary and metastatic cancers sharing over 70% of high-frequency somatic single-nucleotide variations[66].

In sharp contrast to colon and many other cancer types, ovarian cancer is unique in that no physical barrier is present between the primary cancer and the preferred metastatic site, namely the abdominal cavity. The posterior surface of the ovary is directly exposed to the peritoneal cavity where cells detached from the ovarian cancer mass can disseminate in the absence of any major physical barrier preventing them to do so. This process is often referred to as transcoelomic dissemination, and it involves exfoliation due to progressive loosening of cell-cell junctions from the primary tumor, survival and transport in the peritoneal fluid, followed by metastatic colonization of the organs within the peritoneal cavity[67]. A key step for the successful metastasis formation by ovarian cancer cells is their attachment and productive interactions with the mesothelial cells covering peritoneal organs for the establishment of metastatic tumors.

By taking advantage of the different metastatic routes between colon and ovarian cancer, we studied and compared the respective EMT-related RBPs and AS gene targets. Understanding the complexities of metastasis is crucial for advancing cancer diagnosis and treatment strategies.

Aims and Outline of this Thesis.

In this thesis, the role of alternative splicing in epithelial-to-mesenchymal transition (EMT), stemness, and phenotypic plasticity in colon and ovarian cancer metastasis and invasion has been explored. The aim is to identify alternative splicing targets associated with epithelial-to-mesenchymal and mesenchymal-to-epithelial transitions during colon and ovarian cancer metastasis. This was achieved by comparing quasi-mesenchymal and -epithelial subpopulations of colon and ovarian cancer cells identified within conventional immortalized cell lines. To validate our findings and elucidate the underlying mechanisms and function of the newly identified alternative splicing targets, mouse models, and cancer tissues and ascites from colon and ovarian cancer patients have been employed.

In **Chapter 2**, we first developed a general approach, based on colon cancer cell lines, to isolate quasi-mesenchymal subpopulations from the epithelial bulk by FACS and characterized them in terms of heightened invasiveness, resistance to chemotherapy, and metastatic potential. In **Chapter 3**, by taking advantage of this approach, we analyzed differentially expressed RBPs and alternatively spliced genes between the quasi-mesenchymal and epithelial subpopulations of colon cancer cells. The newly identified RBPs and AS targets contribute to the enhancement of phenotypic plasticity and the promotion of malignant behaviors in colon cancer, downstream of EMT.

In **Chapter 4**, following a similar approach using ovarian cancer cell lines, we have identified RNA-binding proteins and isoforms targets, unique to this gynecological malignancy when

compared with colon cancer. These studies have led to the identification of Tropomyosin 1 (*TPM1*), whose isoforms Tpm1.8/9 are likely to be instrumental in promoting invasion and metastasis by affecting lamellipodia formation and motility of ovarian cancer cells.

Last, in **Chapter 5**, we extended our analysis of immortalized colon cancer cell lines as unique models to study the different subpopulation of cancer stem cells, namely the resident CSCs that fuel tumor growth at the primary sites and the migrating CSCs which underlie local invasion and metastasis. In **Chapter 6**, we discuss the challenges associated with leveraging alternative splicing targets in clinical therapeutic approaches for cancer treatment.

References

1 Sethi, N. & Kang, Y. Unravelling the complexity of metastasis - molecular understanding and targeted therapies. Nat Rev Cancer 11, 735-748, doi:10.1038/nrc3125 (2011).

2 Nguyen, D. X. & Massague, J. Genetic determinants of cancer metastasis. Nat Rev Genet 8, 341-352, doi:10.1038/nrg2101 (2007).

3 Clayton, P. E., Banerjee, I., Murray, P. G. & Renehan, A. G. Growth hormone, the insulin-like growth factor axis, insulin and cancer risk. Nat Rev Endocrinol 7, 11-24, doi:10.1038/nrendo.2010.171 (2011).

4 Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. Cancer-related inflammation. Nature 454, 436-444, doi:10.1038/nature07205 (2008).

5 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. Cell 144, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

6 De Palma, M., Biziato, D. & Petrova, T. V. Microenvironmental regulation of tumour angiogenesis. Nat Rev Cancer 17, 457-474, doi:10.1038/nrc.2017.51 (2017).

7 Ogino, S., Galon, J., Fuchs, C. S. & Dranoff, G. Cancer immunology--analysis of host and tumor factors for personalized medicine. Nat Rev Clin Oncol 8, 711-719, doi:10.1038/nrclinonc.2011.122 (2011).

8 Mina, L. A. & Sledge, G. W., Jr. Rethinking the metastatic cascade as a therapeutic target. Nat Rev Clin Oncol 8, 325-332, doi:10.1038/nrclinonc.2011.59 (2011).

9 Boumahdi, S. & de Sauvage, F. J. The great escape: tumour cell plasticity in resistance to targeted therapy. Nat Rev Drug Discov 19, 39-56, doi:10.1038/s41573-019-0044-1 (2020).

10 Martinez-Jimenez, F. et al. Pan-cancer whole-genome comparison of primary and metastatic solid tumours. Nature 618, 333-341, doi:10.1038/s41586-023-06054-z (2023).

11 Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. Nat Med 19, 1438-1449, doi:10.1038/nm.3336 (2013).

12 Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. Nat Rev Genet 3, 415-428, doi:10.1038/nrg816 (2002).

13 Yoo, C. B. & Jones, P. A. Epigenetic therapy of cancer: past, present and future. Nat Rev Drug Discov 5, 37-50, doi:10.1038/nrd1930 (2006).

14 Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet 8, 286-298, doi:10.1038/nrg2005 (2007).

15 Bonnal, S. C., Lopez-Oreja, I. & Valcarcel, J. Roles and mechanisms of alternative splicing in cancer - implications for care. Nat Rev Clin Oncol 17, 457-474, doi:10.1038/s41571-020-0350-x (2020).

16 Brabletz, T., Jung, A., Spaderna, S., Hlubek, F. & Kirchner, T. Opinion: migrating cancer stem cells - an integrated concept of malignant tumour progression. Nat Rev Cancer 5, 744-749, doi:10.1038/nrc1694 (2005).

17 Bakir, B., Chiarella, A. M., Pitarresi, J. R. & Rustgi, A. K. EMT, MET, Plasticity, and Tumor Metastasis. Trends Cell Biol 30, 764-776, doi:10.1016/j.tcb.2020.07.003 (2020).

18 Du, W. & Elemento, O. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. Oncogene 34, 3215-3225, doi:10.1038/onc.2014.291 (2015).

19 Bernards, R. & Weinberg, R. A. A progression puzzle. Nature 418, 823,

doi:10.1038/418823a (2002).
20 Sommer, R. J. Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges. Genetics 215, 1-13, doi:10.1534/genetics.120.303163 (2020).
21 Gupta, P. B., Pastushenko, I., Skibinski, A., Blanpain, C. & Kuperwasser, C. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. Cell Stem Cell 24, 65-78, doi:10.1016/j.stem.2018.11.011 (2019).
22 Williams, E. D., Gao, D., Redfern, A. & Thompson, E. W. Controversies around epithelial-mesenchymal plasticity in cancer metastasis. Nat Rev Cancer 19, 716-732, doi:10.1038/s41568-019-0213-x (2019).
23 <Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer.pdf>.
24 Simeonov, K. P. et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell 39, 1150-1162 e1159, doi:10.1016/j.ccell.2021.05.005 (2021).
25 Pastushenko, I. & Blanpain, C. EMT Transition States during Tumor Progression and Metastasis. Trends Cell Biol 29, 212-226, doi:10.1016/j.tcb.2018.12.001 (2019).
26 Sacchetti, A. et al. Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. Elife 10, doi:10.7554/eLife.61461 (2021).
27 Pastushenko, I. et al. Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. Nature 589, 448-455, doi:10.1038/s41586-020-03046-1 (2021).
28 Faustino, N. A. & Cooper, T. A. Pre-mRNA splicing and human disease. Genes Dev 17, 419-437, doi:10.1101/gad.1048803 (2003).
29 Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470-476, doi:10.1038/nature07509 (2008).
30 Ast, G. How did alternative splicing evolve? Nat Rev Genet 5, 773-782, doi:10.1038/nrg1451 (2004).
31 Xin, D., Hu, L. & Kong, X. Alternative promoters influence alternative splicing at the genomic level. PLoS One 3, e2377, doi:10.1371/journal.pone.0002377 (2008).
32 Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 11, 345-355, doi:10.1038/nrg2776 (2010).
33 McIlwain, D. R. et al. Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. Proc Natl Acad Sci U S A 107, 12186-12191, doi:10.1073/pnas.1007336107 (2010).
34 Zhang, Y., Qian, J., Gu, C. & Yang, Y. Alternative splicing and cancer: a systematic review. Signal Transduct Target Ther 6, 78, doi:10.1038/s41392-021-00486-7 (2021).
35 Fu, X. D. & Ares, M., Jr. Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet 15, 689-701, doi:10.1038/nrg3778 (2014).
36 Blencowe, B. J. Alternative splicing: new insights from global analyses. Cell 126, 37-47, doi:10.1016/j.cell.2006.06.023 (2006).
37 Jurica, M. S. & Moore, M. J. Pre-mRNA Splicing. Molecular Cell 12, 5-14, doi:10.1016/s1097-2765(03)00270-3 (2003).
38 Matlin, A. J., Clark, F. & Smith, C. W. Understanding alternative splicing: towards a cellular code. Nature reviews. Molecular cell biology 6, 386-398, doi:10.1038/nrm1645 (2005).
39 Xiong, H. Y. et al. The human splicing code reveals new insights into the genetic determinants of disease. Science 347, 1254806-1254806, doi:10.1126/science.1254806

(2014).
40 Kornblihtt, A. R. Promoter usage and alternative splicing. Curr Opin Cell Biol 17, 262-268, doi:10.1016/j.ceb.2005.04.014 (2005).
41 Pradella, D., Naro, C., Sette, C. & Ghigna, C. EMT and stemness: flexible processes tuned by alternative splicing in development and cancer progression. Mol Cancer 16, 8, doi:10.1186/s12943-016-0579-2 (2017).
42 Zong, F. Y. et al. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. PLoS Genet 10, e1004289, doi:10.1371/journal.pgen.1004289 (2014).
43 Yang, Y. et al. Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition. Molecular and cellular biology 36, 1704-1719, doi:10.1128/MCB.00019-16 (2016).
44 <An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer.pdf>. doi:10.7554/eLife.37184.001
10.7554/eLife.37184.002.
45 Sun, S., Zhang, Z., Fregoso, O. & Krainer, A. R. Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. Rna 18, 274-283, doi:10.1261/rna.030486.111 (2012).
46 <A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG..pdf>.
47 Braeutigam, C. et al. The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. Oncogene 33, 1082-1092, doi:10.1038/onc.2013.50 (2014).
48 Ahuja, N. et al. Hypoxia-induced TGF-beta-RBFOX2-ESRP1 axis regulates human MENA alternative splicing and promotes EMT in breast cancer. NAR Cancer 2, zcaa021, doi:10.1093/narcan/zcaa021 (2020).
49 Zhou, Z. & Fu, X. D. Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma 122, 191-207, doi:10.1007/s00412-013-0407-z (2013).
50 <Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis.pdf>.
51 Kurokawa, K. et al. Downregulation of serine/arginine-rich splicing factor 3 induces G1 cell cycle arrest and apoptosis in colon cancer cells. Oncogene 33, 1407-1417, doi:10.1038/onc.2013.86 (2014).
52 Lu, X. et al. The function and regulatory mechanism of RNA-binding proteins in breast cancer and their future clinical treatment prospects. Front Oncol 12, 929037, doi:10.3389/fonc.2022.929037 (2022).
53 Osorio, J. Pituitary gland: ESRP1--a regulator of epithelial-mesenchymal transition in somatotroph adenomas? Nat Rev Endocrinol 8, 444, doi:10.1038/nrendo.2012.94 (2012).
54 Munkley, J. et al. Androgen-regulated transcription of ESRP2 drives alternative splicing patterns in prostate cancer. Elife 8, doi:10.7554/eLife.47678 (2019).
55 Brown, R. L. et al. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. The Journal of clinical investigation 121, 1064-1074, doi:10.1172/JCI44540 (2011).
56 Yae, T. et al. Alternative splicing of CD44 mRNA by ESRP1 enhances lung colonization of

metastatic cancer cell. Nature communications 3, 883, doi:10.1038/ncomms1892 (2012).

57 Oltean, S. & Bates, D. O. Hallmarks of alternative splicing in cancer. Oncogene 33, 5311-5318, doi:10.1038/onc.2013.533 (2014).

58 Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. Nature Reviews Molecular Cell Biology 18, 437-451, doi:10.1038/nrm.2017.27 (2017).

59 Lorenz, C., von Pelchrzim, F. & Schroeder, R. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. Nat Protoc 1, 2204-2212, doi:10.1038/nprot.2006.372 (2006).

60 Porter, D. F. et al. easyCLIP analysis of RNA-protein interactions incorporating absolute quantification. Nat Commun 12, 1569, doi:10.1038/s41467-021-21623-4 (2021).

61 <MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation.pdf>. doi:10.1038/ncomms3480.

62 <The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation.pdf>. doi:10.7554/eLife.01201.001
10.7554/eLife.01201.002.

63 Bonomi, S. et al. HnRNP A1 controls a splicing regulatory circuit promoting mesenchymal-to-epithelial transition. Nucleic acids research 41, 8665-8679, doi:10.1093/nar/gkt579 (2013).

64 Zeineddine, F. A. et al. Survival improvement for patients with metastatic colorectal cancer over twenty years. NPJ Precis Oncol 7, 16, doi:10.1038/s41698-023-00353-4 (2023).

65 Nagtegaal, I. D. & Schmoll, H. J. Colorectal cancer: What is the role of lymph node metastases in the progression of colorectal cancer? Nat Rev Gastroenterol Hepatol 14, 633-634, doi:10.1038/nrgastro.2017.122 (2017).

66 Reiter, J. G. et al. Minimal functional driver gene heterogeneity among untreated metastases. Science 361, 1033-1037, doi:10.1126/science.aat7171 (2018).

67 Lengyel, E. Ovarian cancer development and metastasis. Am J Pathol 177, 1053-1064, doi:10.2353/ajpath.2010.100105 (2010).

CHAPTER 2

Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer

Elife 10, doi:10.7554/eLife.61461 (2021)

RESEARCH ARTICLE

Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer

Andrea Sacchetti[1†], Miriam Teeuwssen[1†], Mathijs Verhagen[1†], Rosalie Joosten[1], Tong Xu[1], Roberto Stabile[1], Berdine van der Steen[2], Martin M Watson[1‡], Alem Gusinac[1], Won Kyu Kim[3], Inge Ubink[4], Harmen JG Van de Werken[5], Arianna Fumagalli[6], Madelon Paauwe[7], Jacco Van Rheenen[8], Owen J Sansom[7,9], Onno Kranenburg[4], Riccardo Fodde[1]*

[1]Department of Pathology, Erasmus MC, Rotterdam, Netherlands; [2]Department of Otorhinolaryngology, Head and Neck Surgery, Erasmus University Medical Center, Erasmus MC, Rotterdam, Netherlands; [3]Natural Product Research Center, Korea Institute of Science and Technology, Gangneung, Republic of Korea; [4]Department of Surgical Oncology, Cancer Centre, University Medical Centre Utrecht, Utrecht, Netherlands; [5]Cancer Computational Biology Center and Department of Urology; Erasmus University Medical Center, Rotterdam, Netherlands; [6]Princess Máxima Center for Pediatric Oncology, Utrecht, Netherlands; [7]Cancer Research UK Beatson Institute, Glasgow, United Kingdom; [8]Department of Molecular Pathology, Oncode Institute, Netherlands Cancer Institute, Amsterdam, Netherlands; [9]Institute of Cancer Sciences, University of Glasgow, Glasgow, United Kingdom

***For correspondence:** r.fodde@erasmusmc.nl

[†]These authors contributed equally to this work

Present address: [‡]Gastrointestinal Translational Research Unit, Department of Gastrointestinal Surgery at Stavanger University Hospital, Stavanger, Norway

Competing interest: See page 24

Funding: See page 24

Received: 26 July 2020
Accepted: 17 May 2021
Published: 26 May 2021

Reviewing Editor: Margaret C Frame, University of Edinburgh, United Kingdom

Abstract Phenotypic plasticity represents the most relevant hallmark of the carcinoma cell as it bestows it with the capacity of transiently altering its morphological and functional features while *en route* to the metastatic site. However, the study of phenotypic plasticity is hindered by the rarity of these events within primary lesions and by the lack of experimental models. Here, we identified a subpopulation of phenotypic plastic colon cancer cells: EpCAM^{lo} cells are motile, invasive, chemo-resistant, and highly metastatic. EpCAM^{lo} bulk and single-cell RNAseq analysis indicated (1) enhanced Wnt/β-catenin signaling, (2) a broad spectrum of degrees of epithelial to mesenchymal transition (EMT) activation including hybrid E/M states (partial EMT) with highly plastic features, and (3) high correlation with the CMS4 subtype, accounting for colon cancer cases with poor prognosis and a pronounced stromal component. Of note, a signature of genes specifically expressed in EpCAM^{lo} cancer cells is highly predictive of overall survival in tumors other than CMS4, thus highlighting the relevance of quasi-mesenchymal tumor cells across the spectrum of colon cancers. Enhanced Wnt and the downstream EMT activation represent key events in eliciting phenotypic plasticity along the invasive front of primary colon carcinomas. Distinct sets of epithelial and mesenchymal genes define transcriptional trajectories through which state transitions arise. pEMT cells, often earmarked by the extracellular matrix glycoprotein SPARC together with nuclear ZEB1 and β-catenin along the invasive front of primary colon carcinomas, are predicted to represent the origin of these (de)differentiation routes through biologically distinct cellular states and to underlie the phenotypic plasticity of colon cancer cells.

Introduction

Cancers of epithelial origin such as breast, prostate, pancreas, lung, and colon carcinomas are thought to develop from normal tissues through a multistep sequence of genetic events from benign precursor lesions to increasingly more malignant stages. This is exemplarily illustrated by the adenoma-carcinoma sequence in colon cancer where a stepwise buildup of genetic alterations in specific oncogenes and tumor suppressor genes underlies tumor initiation and progression (***Fearon and Vogelstein, 1990***). These alterations result in well-defined cellular changes largely reflecting the so-called 'hallmarks of cancer,' which provide different selective advantages to the developing tumor and represent essential requirements for carcinoma formation at the primary site (***Hanahan and Weinberg, 2000***). However, with regard to the capacity to disseminate through the tumor microenvironment and establish metastases in distant organ sites, epigenetic changes, rather than genetic mutations, underlie what is the most clinically relevant hallmark of cancer, namely phenotypic plasticity (***Varga and Greten, 2017***; ***Teeuwssen and Fodde, 2019***).

Malignant cells, and in particular those responsible for local dissemination and distant metastasis, are often endowed with the capacity to undergo transient and reversible morphological and functional changes. In particular, epithelial to mesenchymal transition (EMT), that is, the progressive loss of epithelial features and the acquirement of a more migratory and mesenchymal phenotype (***Nieto et al., 2016***), is regarded as a crucial event in tumor cell invasion and dissemination at the primary site. EMT bestows cancer cells with stem-like plastic characteristics (***Mani et al., 2008***) needed to acquire quasi-mesenchymal features at the invasive front of the primary tumor, disseminate and attain therapy resistance, and to revert back to more epithelial states (mesenchymal to epithelial transition [MET]) at the organ site of metastasis (***Brabletz et al., 2005***). Epigenetic activation and silencing of EMT-inducing transcription factors (EMT-TFs) underlies the transient nature of these cellular modifications (***Skrypek et al., 2017***). Notwithstanding these ground rules, a very broad spectrum of molecular and cellular routes underlies EMT and the resulting phenotypic plasticity in a context-dependent fashion (***Cook and Vanderhyden, 2020***).

The 'migrating cancer stem cell' (mCSC) model has been first proposed for colon cancer by T. Brabletz (***Brabletz et al., 2005***), also as a solution to the so-called 'β-catenin paradox' (***Fodde and Brabletz, 2007***). In the majority of sporadic colorectal cancer cases, the rate-limiting loss of the *APC* tumor suppressor is predicted to lead to nuclear β-catenin translocation and full-blown Wnt signaling activation. Notwithstanding these predictions, tumor cells with nuclear β-catenin represent only a small minority of the primary lesion and tend to cluster non-randomly at the invasive front of colon carcinomas where they gain mesenchymal features to detach and disseminate into the adjacent stromal tissues. In view of these observations, it is plausible that cues secreted from the tumor microenvironment elicit EMT downstream of full-blown Wnt signaling activation, earmarked by nuclear β-catenin, in a subset of cells located at the invasive front (***Brabletz et al., 2005***; ***Fodde and Brabletz, 2007***). However, the molecular and cellular mechanisms underlying Wnt and EMT activation at the invasive front of colon cancers are yet largely unknown also due to a lack of robust in vitro and in vivo models.

Previously, it was shown that human immortalized breast cancer cell lines encompass different subpopulations of cells with distinct phenotypic states and functional characteristics maintained in a dynamic equilibrium through stochastic transitions between states (***Gupta et al., 2011***). Similar observations were made in oral squamous carcinoma cell lines where distinct CSC phenotypes are present: whereas non-EMT CSCs are proliferative and retain epithelial characteristics, the EMT-competent CSC fraction is (quasi)mesenchymal and of increased cellular motility (***Biddle et al., 2011***). As such, conventional immortalized cancer cell lines may offer a valid model to elucidate the mechanisms underlying phenotypic plasticity in cancer and to identify novel EMT/CSC-related therapeutic targets.

Here, we identified and extensively characterized a subpopulation of quasi-mesenchymal colon cancer cells endowed with phenotypic plasticity that underlie local invasion and distant metastasis, and whose expression signature is predictive of reduced disease-free survival among colon cancer patients.

Results

Conventional colon cancer cell lines encompass a subpopulation of ZEB1-driven quasi-mesenchymal, highly metastatic, and chemo-resistant cells

To assess whether, as observed for breast cancer (***Gupta et al., 2011***), commonly employed colon cancer cell lines encompass distinct differentiated and more stem-like subpopulations of cancer cells, we first analyzed a broad panel of cell lines by FACS with antibodies directed against different (cancer) stem cell markers (CD44, CD133, CD24, ALDEFLUOR) in combination with the epithelial marker EpCAM. As shown in ***Figure 1A*** and ***Figure 1—figure supplement 1*** for HCT116 and SW480, the CD44/EpCAM combination best highlighted the presence of distinct subpopulations with a predominant $CD44^{high}EpCAM^{high}$ cluster and a minority of $CD44^{high}EpCAM^{low}$ cells. The same $CD44^{high}EpCAM^{low}$ cells were identified at similarly variable percentages in other commonly employed colon cancer cell lines (***Figure 1—source data 1***). Here, we mainly focused on the HCT116 and SW480 lines as they are representative of the two main colon cancer subtypes earmarked by microsatellite instability (MIN) and chromosomal instability (CIN; also known as microsatellite stable [MSS]), respectively (***Lengauer et al., 1997***). For the sake of clarity and readability, from this point on we will refer to the $CD44^{high}EpCAM^{low}$ and $CD44^{high}EpCAM^{high}$ subpopulations of colon cancer cells as $EpCAM^{lo}$ and $EpCAM^{hi}$, respectively.

To characterize the $EpCAM^{lo}$ and $EpCAM^{hi}$ colon cancer cells, they were sorted by FACS and shown to have distinct epithelial- ($EpCAM^{hi}$) and mesenchymal-like ($EpCAM^{lo}$) morphologies (***Figure 1B***). Likewise, $EpCAM^{lo}$ cells were shown to have significantly increased migratory and invasive ability when compared with their $EpCAM^{hi}$ counterparts (***Figure 1—figure supplement 2A***).

RT-qPCR analysis of the sorted subpopulations revealed differential expression of EMT-associated marker genes and transcription factors. Significantly reduced mRNA levels of *EPCAM* and E-cadherin (*CDH1*), and increased expression of vimentin (*VIM*) and of the EMT-transcription factor *ZEB1* were observed in $EpCAM^{lo}$ cells from both lines, as also confirmed by immunofluorescence (IF) analysis (***Figure 1—figure supplement 2B–D***). Accordingly, knockdown of *ZEB1* expression by shRNA dramatically decreased the percentage of $EpCAM^{lo}$ cells in both lines (***Figure 1—figure supplement 3A–C***).

Expression of *ZEB1* (and *ZEB2*) has been shown to be regulated by the miR-200 superfamily of microRNAs that target specific 3′UTR sequences (***Brabletz and Brabletz, 2010***). RT-qPCR analysis of sorted cells revealed significantly reduced expression levels of all five miR-200 members in $EpCAM^{lo}$ cells from both cell lines (***Figure 1—figure supplement 3D***), that is, in agreement with the observed increase in *ZEB1* expression. Proliferation and cell cycle analysis indicated decreased mitotic activity in $EpCAM^{lo}$ cells from both cell lines (***Figure 1—figure supplement 3E,F*** and ***Figure 1—source data 2***).

In view of the well-established correlation between EMT and therapy resistance (***Dean et al., 2005***), $EpCAM^{lo}$ cells were cultured in the presence of oxaliplatin and 5-fluorouracil (5-FU) and their viability compared with that of $EpCAM^{hi}$ and bulk cells by metabolic activity assay (MTT). $EpCAM^{lo}$ cells showed increased viability at all tested oxaliplatin (***Figure 1—figure supplement 4A,C***, left panels) and 5-FU (***Figure 1—figure supplement 4B,D***, left panels) concentrations. Likewise, regrowth assays revealed that $EpCAM^{lo}$ cells from both cell lines are able to re-enter the cell cycle at a broad range of oxaliplatin and 5-FU concentrations when compared with $EpCAM^{hi}$ cells (***Figure 1—figure supplement 4***, right panels).

Last, to assess in vivo their capacity to form metastatic lesions in the liver, HCT116 and SW480 bulk and $EpCAM^{hi/lo}$ sorted cells were injected in the spleen of immune-incompetent recipient mice. $EpCAM^{lo}$ cells from both lines resulted in significantly more liver metastases than with $EpCAM^{hi}$ and bulk cells (***Figure 1C,D***). Notably, immunohistochemistry (IHC) analysis of the resulting liver metastases revealed a heterogeneous pattern of intracellular β-catenin, with membranous and cytoplasmic localization in cells from within the center of the lesion, and nuclear β-catenin accumulation in cells localized in the periphery, thus recapitulating what is observed in primary colon carcinomas (***Fodde and Brabletz, 2007***; ***Kirchner and Brabletz, 2000***; ***Figure 1—figure supplement 5A***). FACS analysis of the $EpCAM^{lo}$-derived liver metastases revealed predominant epithelial features with a vast majority of $EpCAM^{hi}$ cells (>99%), thus highlighting their striking plasticity and the key role played by MET in metastasis formation (***Figure 1—figure supplement 5B*** and ***Figure 1—source data 3***).

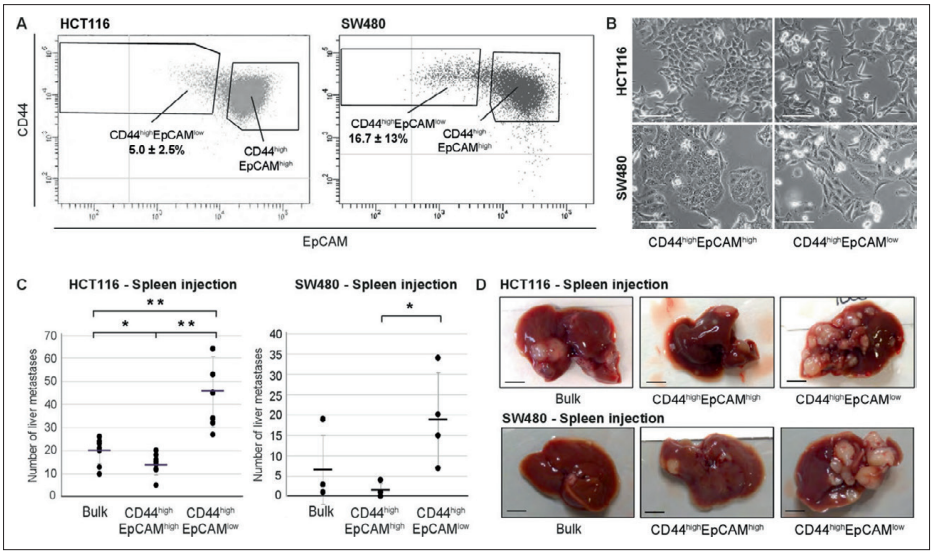


Figure 1. Identification and characterization of EpCAM^{lo} cells in colon cancer cell lines. (A) Flow cytometric analysis of the colon cancer cell lines HCT116 (left panel) and SW480 (right panel) with antibodies directed against CD44 and EpCAM. EpCAM/CD44-positive and -negative regions (gray quadrants) were defined as in Figure supplement 1 using multiple isotype controls and are shown by the quadrants in the plots. Notably, both HCT116 and SW480 revealed a continuum of different EpCAM and CD44 expression levels with a large CD44^{high}EpCAM^{high} (EpCAM^{hi}) cluster followed by a tail of gradually decreasing EpCAM and increasing CD44 levels. By applying specific gates, cells were divided in a large EpCAM^{hi} cluster, together with a considerably smaller CD44^{high}EpCAM^{low} (EpCAM^{lo}) subpopulation. To ensure good separation from the large EpCAM^{hi} cluster and maximal sorting purity, EpCAM^{lo} cells were gated as CD44^{hi} events ≤ 60% of the EpCAM fluorescence intensity of the left border of the EpCAM^{hi} gate and sorted from ≤50% of that value. Variable percentages of EpCAM^{lo} cells were found to feature the HCT116 (5.0% ± 2.5%) and SW480 (16.7% ± 13%) cell lines, respectively. For the sake of simplicity, gates are shown in the figure only if they encompass sizeable percentages of cells. Graphs show representative analysis of one experiment. (**B**) Phase-contrast microscopy images of sorted EpCAM^{hi} and EpCAM^{lo} cells from HCT116 (upper images) and SW480 (lower images) cells. While EpCAM^{hi} cells formed compact colonies with characteristic epithelial morphology, EpCAM^{lo} cells showed a more spindle- and mesenchymal-like appearance. Scale bar: 100 µm. (**C**) Intrasplenic injection of bulk, EpCAM^{hi}, and EpCAM^{lo} cells from HCT116 (left panel) and SW480 (right panel). For each transplantation experiment, 2 × 10^4 cells were injected in the spleen of a recipient NSG mouse. 4 (HCT116) and 8 (SW480) weeks after injection, mice were sacrificed and individual tumors counted. Single and double asterisks indicate significant differences ($p<0.05$ and $p<0.01$, respectively). HCT116: bulk (n = 8), EpCAM^{hi} (n = 9), and EpCAM^{lo} (n = 7). SW480: bulk (n = 4), EpCAM^{hi} (n = 4), and EpCAM^{lo} (n = 4). (**D**) Images of mouse livers 4 (HCT116) and 8 (SW480) weeks after orthotopic injection with 10^4 cells. Scale bar: 5 mm.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. EpCAM^{lo} cells among colon cancer cell lines.

Source data 2. Cell cycle analysis of EpCAM^{hi} and EpCAM^{lo} cells in HCT116 and SW480.

Source data 3. Quantification of EpCAM$^{hi/lo}$ percentages of all liver metastases as determined by FACS.

Figure supplement 1. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: FACS analysis.

Figure supplement 2. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: migration/invasion and EMT analysis.

Figure supplement 3. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: EMT and cell cycle analysis.

Figure supplement 4. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: chemoresistance.

Figure supplement 5. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: invasive and metastatic behaviour.

In order to validate the role of *ZEB1*-driven EMT in the establishment of the EpCAM^{lo} subpopulation of colon cancer cells in a cell line-unrelated in vivo model mimicking the course of events observed in colon cancer patients, we employed mouse intestinal organoids carrying specific mutations at the *Apc*, *Kras*, and *Tp53* genes (*Apc*$^{fl/fl}$::*Kras*$^{G12D/+}$::*Trp53*$^{fl/R172H}$; AKP) (***Fumagalli et al., 2017***). Orthotopic transplantation of AKP organoids results in the establishment of primary carcinomas in the caecum and subsequent metastasis development at distant organ sites, mainly liver and lungs (***Fumagalli et al., 2018***). We further modified the AKP organoids by tagging them with GFP and a

click beetle luciferase (***Hall et al., 2018***) and by making them doxycycline-inducible for the expression of mouse *Zeb1* (AKP-Z) (***Figure 1—figure supplement 5C***). Upon orthotopic transplantation of the AKP-Z organoids and subsequent establishment of the primary tumor in the caecum, mice were administered doxycycline for 1 week in the drinking water to induce *Zeb1* expression. FACS analysis of the primary tumor revealed an increase in EpCAM^{lo} cells from 4.8% in the non-induced tumors up to 22–76% upon dox-induction of *Zeb1* expression (***Figure 1—figure supplement 5D***). As expected, only a marginal increase of lung and liver metastases was observed in AKP-Z transplanted mice upon continuous dox administration in the drinking water for 8 weeks when compared with control (no dox) animals (***Figure 1—figure supplement 5E***), likely to result from the continuous induction of *Zeb1* expression and the consequent inhibition of METs essential for metastasis formation (***Brabletz et al., 2005***).

Overall, the results show the presence within colon cancer cell lines of EMT-driven, quasi-mesenchymal and therapy-resistant EpCAM^{lo} cells with increased invasive and metastatic capacity. *ZEB1* expression underlies in vivo the establishment and maintenance of the subpopulation of EpCAM^{lo} colon cancer cells, thereby contributing to increased dissemination along the invasion-metastasis cascade.

EpCAM^{lo} colon cancer cells are maintained in equilibrium with EpCAM^{hi} through stochastic state transitions

To further investigate the plasticity of EpCAM^{lo} cancer cells, we assessed their capacity to differentiate into the more epithelial EpCAM^{hi} type and reconstitute the heterogeneous composition of the parental cell lines. Sorted EpCAM$^{lo/hi}$ cells from HCT116 and SW480 were grown separately under conventional culture conditions and analyzed by FACS at different time points. As shown in ***Figure 2A***, while the majority of EpCAM^{lo} cells from both cell lines revert to the epithelial phenotype, only a minority of EpCAM^{hi} cells switches to the more mesenchymal state. Accordingly, *CDH1* and *EPCAM* expression was significantly increased in 'late' (e.g., cultured for >60 and >100 days, respectively) vs. 'early' EpCAM^{lo} cells (e.g., collected after <7 days of culture), whereas *VIM* and *ZEB1* expression was decreased in 'late' vs. 'early' EpCAM^{lo} cells (***Figure 2—figure supplement 1A***). In addition, the migratory capacity of 'late' EpCAM^{lo} HCT116 cells was reduced to levels comparable with those of 'early' EpCAM^{hi} cells (***Figure 2—figure supplement 1B***).

To exclude cross-contamination between subpopulations, single EpCAM$^{hi/lo}$ cells from both cell lines were sorted into 96-well dishes, cultured up to 70–80 days, and analyzed by FACS at intermediate and end time points. As shown in ***Figure 2B*** and ***Figure 2—source data 1***, the majority of EpCAM^{lo} single cells were capable of generating substantial percentages of EpCAM^{hi} progeny to eventually recapitulate the heterogeneous composition of the parental cell lines (e.g., ***Figure 2B***, clones 2F12 and 2E10). A minority of the cells, however, appears to have lost this plasticity and retains, even after extended culture, the EpCAM^{lo} phenotype (e.g., ***Figure 2B***, clone 2H12). In contrast, the majority of EpCAM^{hi} single cells retained their epithelial features with <1% switching to the EpCAM^{lo} state (***Figure 2B*** and ***Figure 2—source data 1***).

Based on these results, a two-state Markov model was developed to estimate the average probabilities to transition from one state to the other. First, the FACS data was employed to estimate the average doubling time of cells in both populations; slightly increased doubling times were reported for EpCAM^{lo} compared to EpCAM^{hi} cells in both lines (HCT116: 1.09 vs. 1.00 days; SW480: 1.86 vs. 1.68 days). Next, we employed least-square optimization to estimate the transition probabilities that best fit the observed population dynamics. The fitted model predicts that both subpopulations have a high probability to retain their cell identity, with minor though significant likelihood to transit to the other state. Of note, EpCAM^{lo} cells show a higher transition probability compared to EpCAM^{hi} (8.1- and 4.0-fold in HCT116 and SW480, respectively) (***Figure 2C***).

Due to the observed differences in doubling times between the two states, subclones with a lower EpCAM$^{hi>lo}$ transition probability will experience a slight growth advantage, which will become prevalent in the long run. We ran a simulation of this effect by starting from a culture with multiple subclones having distinct transition probabilities; the results indicate that subclones with lower plasticity gain dominance within a few months (***Figure 2—figure supplement 2***). Consequently, especially in the long run, the percentage of EpCAM^{lo} cells will decrease as observed in late cultures (***Figure 1—source data 1***).

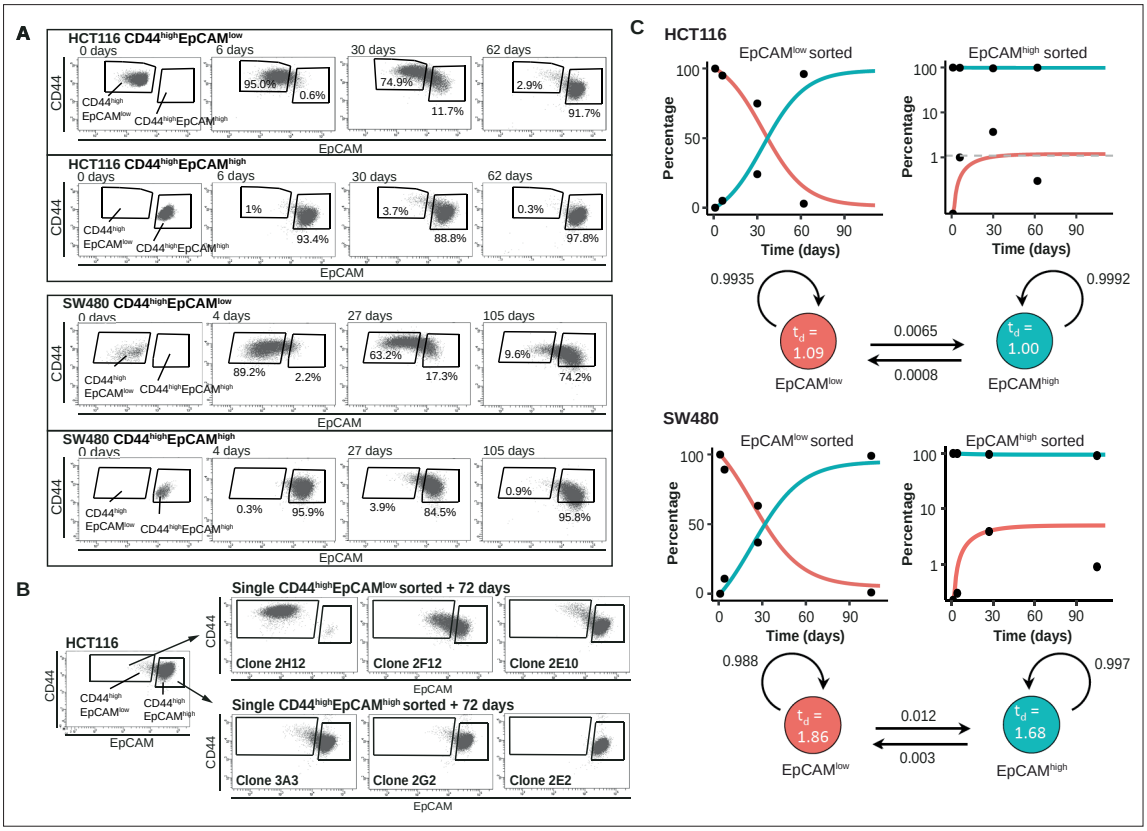


Figure 2. Phenotypic plasticity maintains EpCAM^{lo} and EpCAM^{hi} cells in a stochastic equilibrium. (**A**) Analysis of plasticity of EpCAM^{hi} and EpCAM^{lo} cells from HCT116 (upper panel) and SW480 (lower panel). EpCAM^{hi} and EpCAM^{lo} cell fractions were sorted and plated in culture. At different time points, as indicated, cells were reanalyzed by flow cytometry for their levels of CD44 (*y*-axis) and EpCAM (*x*-axis) expression. (**B**) Flow cytometric analysis of single CD44^{hi}EpCAM^{hi} and CD44^{hi}EpCAM^{lo} HCT116 cells sorted by FACS and cultured for 72 days. Three representative individual single-cell clones per cell fraction are shown. (**C**) Dynamics of the EpCAM^{hi} and EpCAM^{lo} subpopulations from the HCT and SW480 cell lines as measured by FACS (% of total) over time. Under each graph, a schematic shows the estimated transition probabilities from the fitted two-state Markov model.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Phenotypic plasticity maintains EpCAM^{lo} and EpCAM^{hi} cells in a stochastic equilibrium: clonal analysis.

Figure supplement 1. Further characterization of phenotypic plasticity in EpCAM^{lo} and EpCAM^{hi} cells.

Figure supplement 2. Simulation of the HCT116 two-state Markov model with a non-homogenous starting population.

Overall, the above results highlight the high plasticity and stem-like features of EpCAM^{lo} cells in their ability to acquire epithelial features and reconstitute the heterogeneous composition of the parental cell line, independently of external factors other than the conventional culture conditions here employed.

Enhanced Wnt signaling activation underlies EMT and the establishment of EpCAM^{lo} colon cancer cells

In order to elucidate the mechanisms underlying plasticity and EMT in EpCAM^{lo} colon cancer cells, RNAseq analysis was performed on the sorted subpopulations from the HCT116 and SW480 lines. Multidimensional scaling (MDS) showed a separation in the second dimension of EpCAM^{hi} and EpCAM^{lo}

cells in both cell lines (***Figure 3A***). A total of 152 and 353 differentially regulated genes were identified between the EpCAM^{hi} and EpCAM^{lo} cells in HCT116 and SW480, respectively (p. adjusted<0.01, when applying a $\log_2$ fold change <–1.5 and > 1.5). Among these, only a relatively few (n = 34) were common to both lines (***Figure 3—source data 1***). However, Ingenuity Pathway Analysis (IPA) revealed that genes differentially expressed in each cell line reflect common molecular and cellular functions including cellular motility, cellular assembly and organization, and drug metabolism (***Figure 3—figure supplement 1***). Nonetheless, IPA of the combined EpCAM^{lo} expression profiles highlighted significant associations with EMT regulation, Wnt/β-catenin signaling, human ESC pluripotency, IL8 signaling, and colorectal cancer metastasis (***Figure 3B*** and ***Figure 3—figure supplement 1***).

The activation of canonical Wnt signaling in EpCAM^{lo} colon cancer cells is of interest in view of the fact that both cell lines harbor mutations (loss and gain of APC and β-catenin function in SW480 [***Nishisho et al., 1991***] and HCT116 [***Ilyas et al., 1997***], respectively) predicted to result in the constitutive activation of the pathway. Notwithstanding the latter, Wnt appears to be increased in EpCAM^{lo} cells, possibly due to the epigenetic activation/inhibition of synergistic/antagonistic loci. In view of these observations and of the established functional link between Wnt and EMT (***Lamouille et al., 2014***; ***Ghahhari and Babashah, 2015***), we evaluated whether Wnt signaling 'super-activation' in the already Wnt-ON HCT116 and SW480 cell lines could expand the relative size of the EpCAM^{lo} subpopulations. Indeed, upon treatment with the glycogen synthase 3β (GSK3β) inhibitor CHIR99021 (Chiron), a robust Wnt signaling activation was observed in both cell lines by TopFLASH reporter assay (***Figure 3C***). FACS analysis of the treated cell lines showed that the enhancement of Wnt signaling led to an approximately threefold increase of the EpCAM^{lo} population in the HCT116 cell line, though not in SW480 (***Figure 3D***). However, IF analysis showed that Chiron treatment was consistently accompanied by an increase in *ZEB1* expression in both cell lines, in agreement with the role of *ZEB1* as a downstream Wnt target (***Sánchez-Tilló et al., 2011***; ***Sánchez-Tilló et al., 2015***; ***Figure 3E***).

To further validate the role of Wnt in the establishment of the EpCAM^{lo} subpopulation, we took advantage of the recently generated isogenic subclones of the HCT116 cell line with distinct β-catenin genotypes, namely wild type (HCT116-WT), hetero- (HCT116-P), and homozygous (HCT116-MT) for the Ser45del mutant allele (***Kim et al., 2019***). FACS analysis of these cell lines revealed a progressive increase in the EpCAM^{lo} subpopulation from 3.9% in HCT116-WT to 7.9% in HCT116-P and 28.7% in HCT116-MT cells (***Figure 3F***). The observed increase in EpCAM^{lo} cells in the isogenic HCT116 lines matches their morphological features ranging from a distinct epithelial-like morphology in HCT116-WT to a progressively increased spindle-shaped and scattered patterns in HCT116-P and -MT cells, as originally reported by ***Kim et al., 2019***.

In the majority of colon cancers, nuclear accumulation of β-catenin is exclusively observed at the invasive front where tumor cells are more exposed to growth factors and cytokines from the stromal microenvironment likely to further enhance Wnt signaling in a localized fashion, thus triggering EMT, invasion, and distant metastasis (***Brabletz et al., 2005***; ***Fodde and Brabletz, 2007***). We analyzed the invasive front of a small cohort of colon carcinomas by IHC with antibodies directed against β-catenin and ZEB1 in consecutive sections. As shown in ***Figure 3—figure supplement 2***, co-localization of nuclear β-catenin and ZEB1 expression was found in 5 out of 25 cases investigated.

Overall, the results highlight the key role played by enhanced Wnt signaling activation in establishing and maintaining the EpCAM^{lo} subpopulation of colon cancer cells through *ZEB1* upregulation and EMT induction.

EpCAM^{lo} cells are associated with the CMS4 group of patients with shorter disease-free and overall survival

Distinct recurrent gene expression patterns underlie the recently proposed classification of human colon cancers in four consensus molecular subtypes (CMS1–4) (***Guinney et al., 2015***). Of these, the mesenchymal CMS4 subtype has the greatest propensity to form metastases. While fibrosis is a hallmark of CMS4 and a dominant contributor of mesenchymal gene expression, the cancer cells themselves can also express genes reflecting a (quasi-)mesenchymal state. Accordingly, the CMS4 subtype was identified in tumor-derived organoids and cell lines, suggesting that CMS4 is an intrinsic property of the mesenchymal colon cancer cell (***Vellinga et al., 2016***). Therefore, we asked whether expression of the signatures derived from the RNAseq analysis of EpCAM^{lo} cells would correlate with the CMS classification of human colon cancers and cell lines. To this end, we employed a compiled dataset

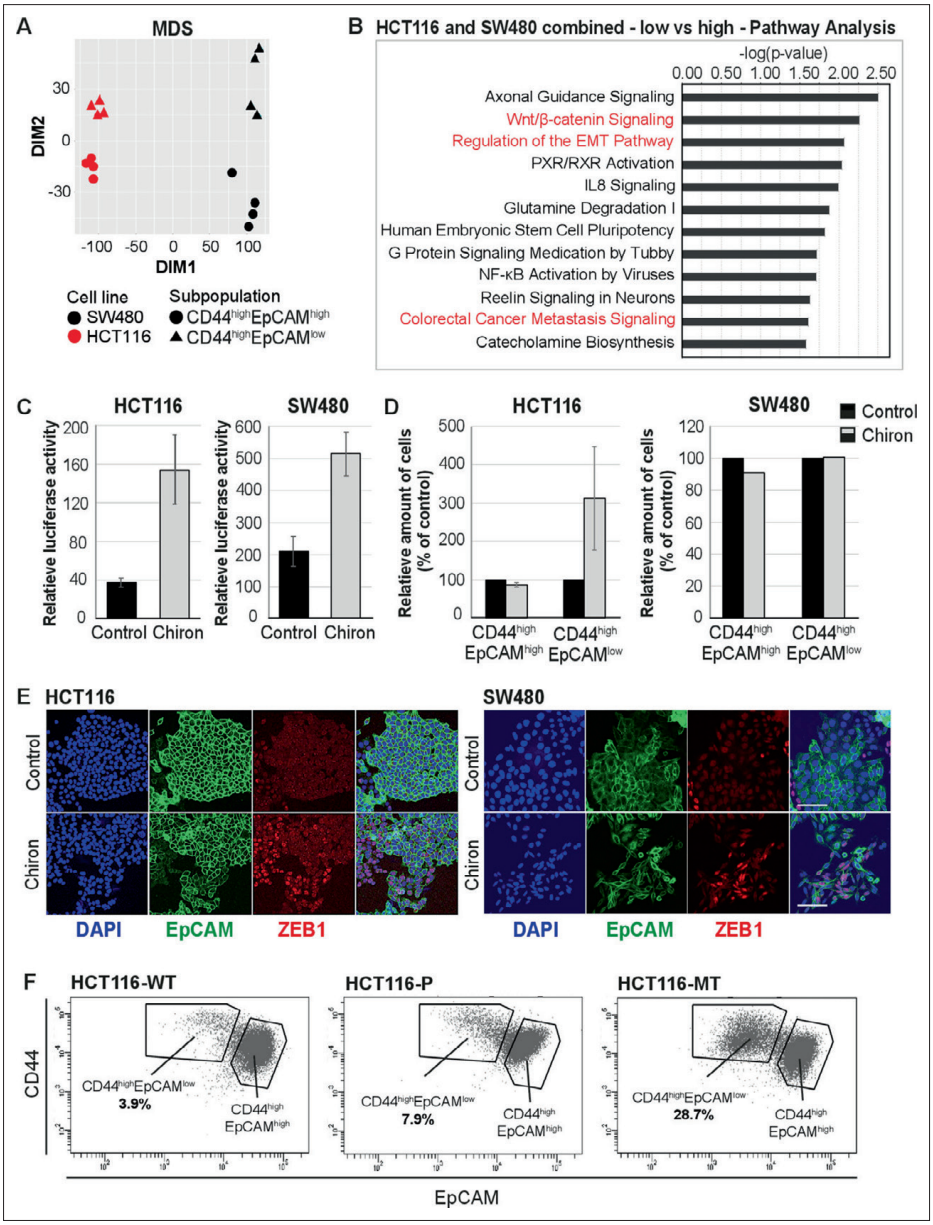


Figure 3. RNAseq analysis reveals enhanced Wnt signaling in EpCAM^{lo} cells. (**A**) Multidimensional scaling analysis of RNAseq profiles of EpCAM^{hi} and EpCAM^{lo} cells from the HCT116 and SW480 lines. Red: HCT116, black: SW480, circle: EpCAM^{hi}, triangle: EpCAM^{lo}. (**B**) Ingenuity Pathway Analysis (IPA) of the HCT116 and SW480 expression profiles from the multicell line analysis (p adjusted value <0.01; $\log_2$ fold change <–1.5 and >1.5). Red marked pathways highlight the enhanced involvement of pathways involved in epithelial to mesenchymal transition, Wnt signaling, and the formation of colon cancer metastasis in the EpCAM^{lo} subpopulation compared to EpCAM^{hi} cells. (**C**) TOP-Flash luciferase reporter analysis of Wnt signaling activity in colon cancer cell lines HCT116 and SW480 upon treatment with 4 μM Chiron for 3 days. Each bar represents the mean ± SD of two independent experiments.

Figure 3 continued on next page

Figure 3 continued

(**D**) Flow cytometric analysis using antibodies directed against CD44 and EpCAM of control and 4 µM Chiron—figure supplemented HCT116 (**A**) and SW480 (**B**) cultures. Graphs show percentage of cells within the CD44^{hi}EpCAM^{hi} and CD44^{hi}EpCAM^{lo} gates relative to the control. Each bar represents the mean ± SD of two independent experiments. (**E**) Immunofluorescence analysis of control and Chiron-treated HCT116 (left panel) and SW480 (right panel) cells. After 3 days of treatment, cells were fixed with 4% paraformaldehyde and stained with antibodies against EpCAM (green) and ZEB1 (red). Nuclei were visualized by DAPI staining of DNA (blue). Scale bar: 100 µm. (**F**) Flow cytometric analysis of three HCT116 cell lines with differential β-catenin mutation status, a parental HCT116 (HCT116-P) cell line harboring one WT and one mutant allele (Ser45 del), and two HCT116-WT and HCT116-MT cell lines harboring one WT or one mutant allele, respectively, generated by disruption of the other allele in the parental cell line (***Kim et al., 2019***).

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. List of differentially expressed genes in EpCAM^{lo} vs. EpCAM^{hi} cells in HCT116 and SW480.

Figure supplement 1. Gene ontology (GO) of molecular and cellular functions in HCT116 (upper panel), SW480 (middle), and the combined (bottom panel) gene lists.

Figure supplement 2. Hematoxylin and eosin (H&E) and immunohistochemistry (IHC) analyses with antibodies directed against beta-catenin and ZEB1 in consecutive sections of colon cancers from three unrelated patients with sporadic colon cancer.

encompassing expression data relative to 3232 human colon cancers classified as CMS1–4 (***Guinney et al., 2015***). Expression of the HCT116 and SW480 signatures was highly correlated with each other (***Figure 4A***), with the CMS4 signature genes (n = 143) (***Figure 4A–B***), and with the expression signature of colon cancer cell lines previously classified as mesenchymal-like (CCS3) (***De Sousa et al., 2013***; ***Figure 4—figure supplement 1A***).

Next, we used both HCT116 and SW480 signatures from the bulk RNAseq to generate high, intermediate, and low expression subgroups in the CMS cohort through k-means clustering (***Figure 4—figure supplement 1B***). High expression of the EpCAM^{lo} signatures from both cell lines identifies a group of primary colon cancer patients with a high propensity to form metastases and with significantly shorter overall survival (***Figure 4C***).

Overall, the above data strongly link the expression of the EpCAM^{lo} signatures derived from common colon cancer cell lines to the CMS4 subtype and to shorter disease-free and overall survival.

scRNAseq analysis of EpCAM^{lo} colon cancer cells reveals high heterogeneity and partial EMT intermediate stages

To further elucidate the heterogeneity and molecular mechanisms underlying the phenotypic plasticity of EpCAM^{lo} colon cancer cells, single-cell RNA sequencing (scRNAseq) was performed using chromium controller (10 X Genomics) technology on sorted subpopulations from HCT116 and SW480. More than 1000 cells were analyzed from each subpopulation and sequenced to a depth of approximately 50,000 reads each with the MiSeq System (Illumina). After dimension reduction with tSNE, the EpCAM^{lo} and EpCAM^{hi} cells clustered in separate groups in HCT116 (***Figure 5A***), whereas the SW480 cells showed a partial overlap between the two subpopulations and a distinct EpCAM^{hi} cluster, identified as the non-adherent subpopulation within the SW480 cell line ('spheres') (***Hirsch et al., 2014***; ***Yi et al., 2020***; ***Figure 5—figure supplement 1A***). Substantial overlap between the EpCAM$^{hi/lo}$ subpopulations was retained upon subsequent exclusion of the non-adherent cells in SW480, attributed to additional variance in genes independent of EpCAM$^{hi/lo}$ differences (***Figure 5—figure supplement 1B***). Next, the dimension reduction was repeated in supervised fashion by taking advantage of the publicly available EMT gene list from the Nanostring nCounter PanCancer Progression Panel (n = 180; 107 of which were found to overlap with the scRNAseq data) (***Figure 5—source data 1***) . Using this 'EMT signature,' the two subpopulations were clearly resolved both in HCT116 and SW480 (***Figure 5B***).

To rank and order the cells along the E-M axis, epithelial (E; n = 51) and mesenchymal (M; n = 56) gene sets were defined from the EMT signature based on the provided annotations (Nanostring) and evaluated by Gene Set Variation Analysis (GSVA). Next, an EMT score was computed by subtracting the E score from the M score (EMT = M – E), while a partial EMT score was obtained based on the co-expression of the E and M gene sets (pEMT = min[E,M]). The expression of genes previously found to be differentially expressed between EpCAM$^{hi/lo}$ by bulk RNAseq was then evaluated over the course of EMT progression (***Figure 5C***). While HCT116 showed a clear transition from EpCAM^{hi} to EpCAM^{lo} with a minority of 'in-between' cells earmarked by high pEMT scores, in SW480 the transition appeared to be accompanied by higher levels of intercellular variance. We continued to use

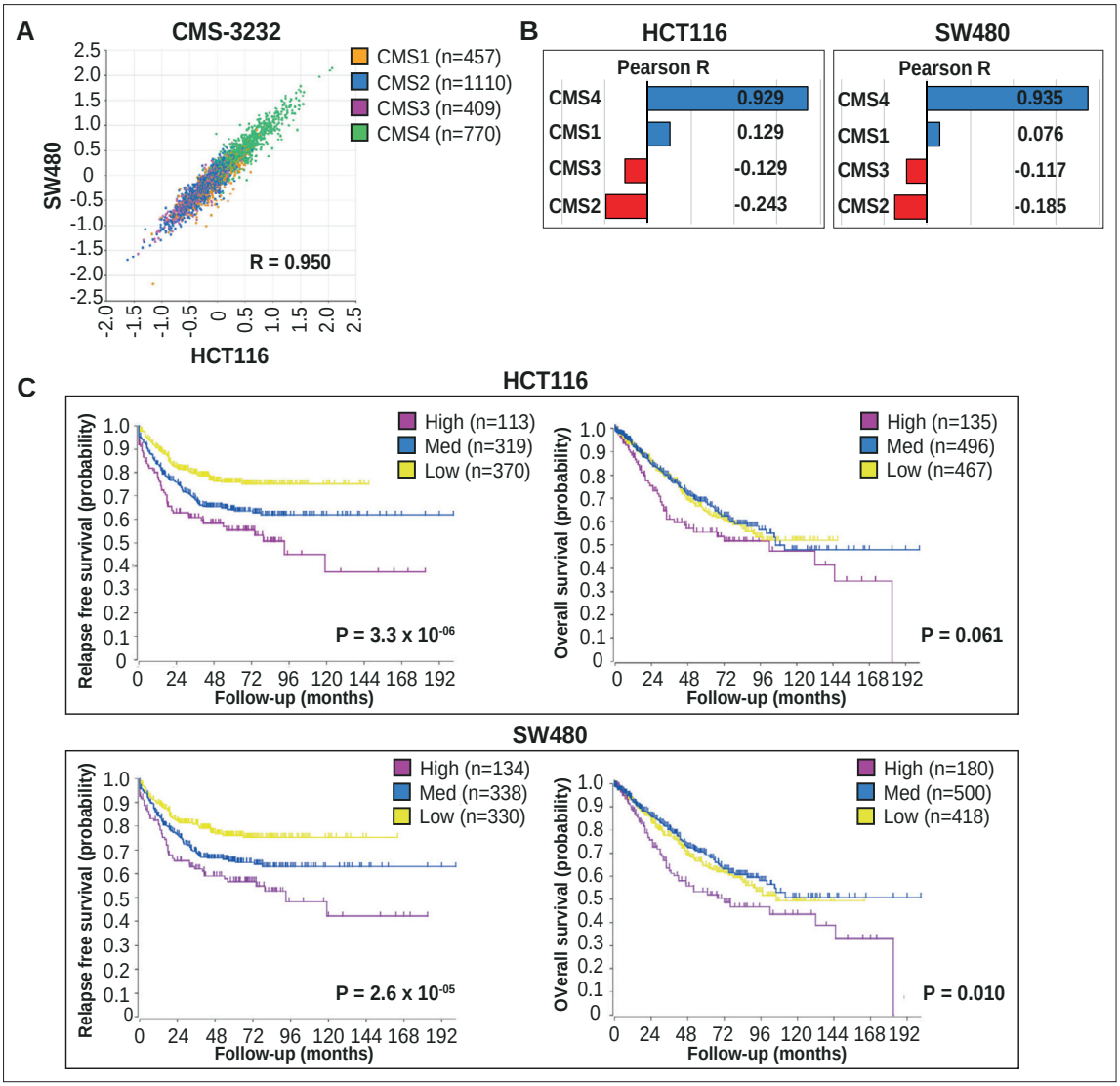


Figure 4. EpCAM^{lo} gene expression profiles correlate with CMS4 colon cancer patients with shorter disease-free and overall survival. (**A**) Correlation of meta-gene expression values of the signatures derived from EpCAM^{lo} HCT116 and SW480 cells in the consensus molecular subtype (CMS)3232 human colon cancer cohort (***Guinney et al., 2015***). (**B**) Correlation of meta-gene expression values of the signatures derived from EpCAM^{lo} HCT116 and SW480 cells with expression of CMS classifier genes positively identifying each of the four molecular subtypes. (**C**) Kaplan– Meier analysis. The gene sets identifying the EpCAM^{lo} cells from both HCT116 and SW480 cell lines were used to cluster the tumors in the CMS3232 cohort into high (purple), intermediate (blue), and low (yellow) expression groups by k-means clustering. The Kaplan–Meier method was subsequently used to assess significant differences in relapse-free (left panels) and overall (right panels) survival between the generated subgroups.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. EpCAM^{lo} expression signatures correlate with CCS3 cell lines and CMS4 colon cancer patients.

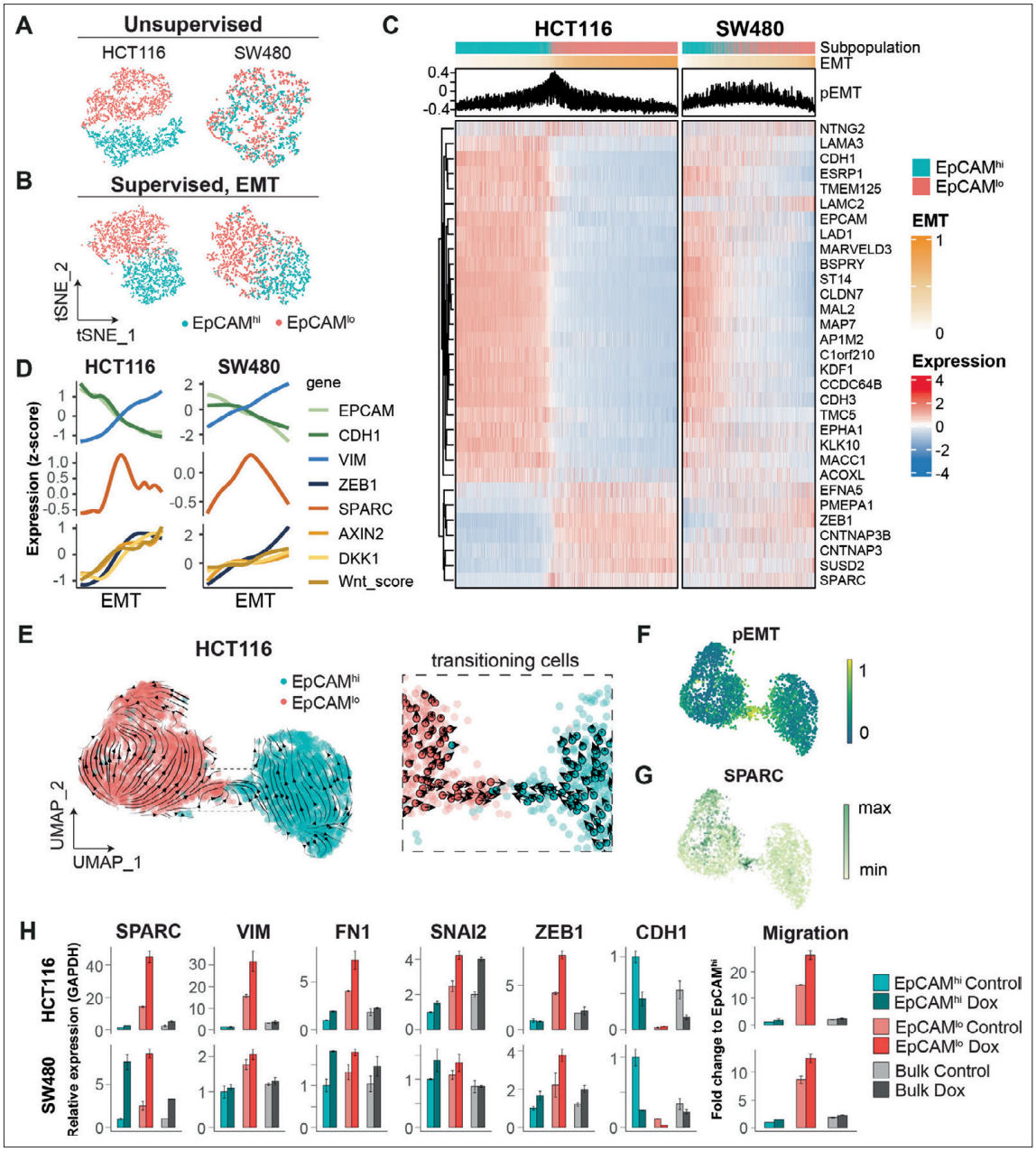


Figure 5. scRNAseq analysis of EpCAM^{lo} cells reveals specific markers of partial EMT cells. (**A**) tSNE of HCT116 and SW480 cells based on the variable expressed genes across EpCAM^{lo} and EpCAM^{hi} populations. (**B**) tSNE of HCT116 and SW480 cells based on genes from the epithelial to mesenchymal transition (EMT) signature (N = 107). (**C**) Heatmap of differentially expressed genes between EpCAM^{lo} and EpCAM^{hi} populations. Cells were ranked according to their EMT score (EMT = M – E). (**D**) Gene expression trends projected over the EMT axis. Expression values were imputed with MAGIC,

Figure 5 continued on next page

Figure 5 continued

scaled by their z-score, and smoothened by general additive models to visualize the gene expression trend. (**E**) RNA velocity analysis of the HCT116 scRNAseq data. Cells from both populations were moving in their respective state (left panel) with a minority population of transitioning cells (right panel). (**F**) Projection of the pEMT score on the UMAP embedding of the HCT116 cell line. (**G**) The expression of SPARC on the UMAP embedding of the HCT116 cell line. (**H**) Left panels: qPCR analysis of SPARC overexpression in the subpopulations of HCT116 and SW480. Gene expression values are relative to GAPDH and normalized to the EpCAM^{hi} subpopulation. Right panel: quantification of the transwell migration assay upon overexpression of SPARC.

The online version of this article includes the following source data and figure supplement(s) for figure 5:

Source data 1. EMT and CMS4 classifiers.

Figure supplement 1. Further scRNAseq analysis of EpCAM^{lo} cells: EMT/MET transcriptional trajectory.

the EMT axis in both cell lines to study gene expression trends. As expected, epithelial expression of *EPCAM* and *CDH1* gradually decreased over the EMT axis, while mesenchymal genes such as *ZEB1* and *VIM*, and Wnt target genes such as *AXIN2* and *DKK1*, showed opposing behavior. More generically, the level of Wnt signaling ('Wnt.score'), measured by the activation of the WNT_PATHWAY from the PID database, increased over the EMT axis (***Figure 5D***). We evaluated the expression trends of other genes upregulated in EpCAM^{lo} cells and observed that *SPARC* (osteonectin) peaked in between the extremities of the EMT axis, indicative of a partial EMT state (***Figure 5D***).

To further characterize the transition process between EpCAM^{lo} and EpCAM^{hi}, we analyzed the HCT116 scRNAseq data by RNA velocity, an algorithm that predicts the future cell state of individual cells on a short-term timescale (usually hours), based on the ratio between spliced and unspliced mRNA. Consistent with the fitted Markov model, the majority of cells were moving in their respective state, and both states included source and sink points that could elicit or inhibit plasticity (***Figure 2—figure supplement 1C***). On a population level, EpCAM^{lo} cells were more likely to transit to EpCAM^{hi} cells than vice versa (by partition-based graph abstraction or PAGA; ***Figure 2—figure supplement 1D***). Within the EpCAM^{lo} subpopulation, *CD44* and mesenchymal genes such as *VIM* and *ZEB1* were expressed with different dynamics (***Figure 5—figure supplement 1C***). In fact, only a small population of cells was captured in the process of transitioning to the other identity (***Figure 5E***). This apparently plastic population, earmarked by the opposing velocity arrows, showed the highest pEMT score (***Figure 5F***), as well as high *SPARC* expression (***Figure 5G***).

Lastly, we performed cluster analysis on the HCT116 scRNAseq data. Unsupervised clustering using shared neighbor (SSN) modularity optimization revealed the presence of distinct subclusters (n = 8): three of EpCAM^{hi} origin and five EpCAM^{lo} (***Figure 5—figure supplement 1D***). To further investigate the gene expression patterns underlying the different EMT states, we performed a cluster analysis where the EMT signature genes are grouped with k-means according to their average expression in the unsupervised clusters (***Figure 5—figure supplement 1E***). This analysis revealed four sets of genes expressed in different combinations throughout the different clusters. Notably, gene set mes1, including *ZEB1*, *VIM*, and *SNAI2*, is expressed throughout the EpCAM^{lo} clusters, while gene set mes2, including *SPARC*, *FN1*, and *TWIST1*, is mostly expressed in the pEMT EpCAM^{lo} cluster (#7). Of note, the partial EMT clusters #6 and #7 showed distinct expression of these gene sets indicating alternative activation of specific arrays through which partial EMT cells arise (***Figure 5—figure supplement 1F***).

As *SPARC* earmarks pEMT states (***Figure 5D***), we overexpressed it in the HCT116 and SW480 cell lines and observed a marked increase in the expression of EMT-TFs (*ZEB1* and *SNAI2*) and the corresponding up- and downregulation of mesenchymal and epithelial markers in EpCAM^{lo} cells and a corresponding increase in their motility and invasive capacity (***Figure 5H***).

Taken together, our in silico analysis shows substantial heterogeneity within the EpCAM^{lo} subpopulation across colon cancer cell lines. Among EpCAM^{lo} cells, a minority exhibits partial EMT and underlies the stochastic EpCAM$^{lo->hi}$ and EpCAM$^{hi->lo}$ transitions. Of note, pEMT is shown here not only as an '*in-between*' state earmarked by the co-expression of E- and M-specific genes, but also by specific genes like *SPARC* whose expression peaks at pEMT states.

Identification of EpCAM^{lo} colon cancer cells in primary colorectal tumors

As shown above, the EpCAM^{lo} gene signatures derived from both cell lines show high concordance with the CMS4 subtype and are strongly associated with poor survival. Previously, it has been questioned to what extent the CMS4 profile reflects a *bona fide* clinical identity rather than being a representation of contamination from the tumor microenvironment (**Calon et al., 2015**; **Isella et al., 2015**). We addressed this issue by evaluating the CMS4 signature (here referred to as '*CMS4_RF*') (**Figure 5—source data 1**) originally developed by **Guinney et al., 2015** and by taking advantage of a recent scRNAseq study on colon cancer resections from 29 patients (**Lee et al., 2020**). Based on the latter study, as depicted in **Figure 6A**, the CMS4_RF signature shows the highest association with normal and tumor stromal cells. At the bulk RNAseq level, using the larger (n = 3232 patients) cohort of colon tumors from **Guinney et al., 2015**, the CMS4_RF signature is clearly enriched among the CMS4 tumors (**Figure 6B**) and is likewise associated with poor survival (**Figure 6C**). Hence, the CMS4_RF signature reflects the presence of tumor-associated stromal cells, likely to be enriched in patients with decreased overall survival. To question whether the presence of a subset of *bona fide* tumor cells in quasi-mesenchymal state (EpCAM^{lo}) may also represent a feature of colon cancers with poor prognosis, we derived a distinct signature ('*CMS4_TC*') (**Figure 5—source data 1**) by selecting genes correlated (Pearson > 0.3) with the CMS4_RF signature within the tumor epithelial fraction from the **Lee et al., 2020** study. Accordingly, the CMS4_TC signature shows the highest association with epithelial tumor cells (**Figure 6D**). Of note, at the bulk RNA level, the CMS4_TC signature shows increased association with both CMS1 and CMS4 tumors (**Figure 6E**) and outperforms the CMS4_RF signature in stratifying patients on survival (**Figure 6F**). The same CMS4_TC signature can clearly stratify patients from CMS groups other than CMS4, and in particular CMS1 and CMS3, on survival (**Figure 6—figure supplement 1**).

Finally, in order to identify EpCAM^{lo} cells in primary colorectal tumors, we took advantage of the scRNAseq data from **Lee et al., 2020** to annotate the tumor cells with highest association to the CMS4_TC signature as mesenchymal-like (Mes-like = 9.7% of all tumor cells). In agreement with the observations made at the bulk RNA level, tumor cells from CMS1 and CMS4 patients showed the highest association to the CMS4_TC signature (**Figure 6G**). Differential expression between the mesenchymal-like and bulk tumor cells revealed lower expression of *EPCAM*, while higher expression of *CD44* and mesenchymal markers such as *VIM*, *COL1A1*, *MMP7*, *ECM1*, and *SPARC* (**Figure 6H**), the latter previously identified to peak at intermediate EMT levels and predicted by RNAvelocity to earmark a subpopulation of plastic cells transitioning in between states (**Figure 5D–G**).

Discussion

The progression from primary tumor to metastasis still represents an unsolved puzzle as genomic and gene expression profiles of primary cancers and their matched metastases are in general strikingly similar (**Bernards and Weinberg, 2002**). Colon cancer provides a unique example for this conundrum: although it is well-established that a sequence of gene mutations underlies the adenoma-carcinoma sequence at the primary site, no alterations have been identified in genes potentially able to underlie local invasion and distant metastasis. Hence, the capacity to metastasize could already be predetermined by specific mutations acquired by tumor cells at relatively early stages of tumorigenesis. However, this does not yet explain why and how only rare cells within the primary lesion acquire a metastatic phenotype that endows them with increased cell motility and with the capacity to invade the stromal microenvironment to eventually home at distant organ sites and form metastases. From this perspective, phenotypic plasticity likely represents a relevant mechanism for metastasizing colon carcinoma cells to transiently and reversibly change their cellular identity along the invasion-metastasis cascade (**Varga and Greten, 2017**; **Teeuwssen and Fodde, 2019**).

The vast majority of colon cancers are initiated by mutations at the *APC* gene predicted to lead to constitutive Wnt signaling by β-catenin nuclear translocation. However, IHC analysis has shown that nuclear accumulation of β-catenin is exclusively observed at the invasive front where tumor cells are exposed to growth factors and cytokines secreted by the stromal microenvironment likely to further enhance Wnt signaling in a localized fashion (**Fodde and Brabletz, 2007**). As different levels of Wnt/β-catenin signaling are associated with distinct cellular outcomes (**Gaspar and Fodde, 2004**),

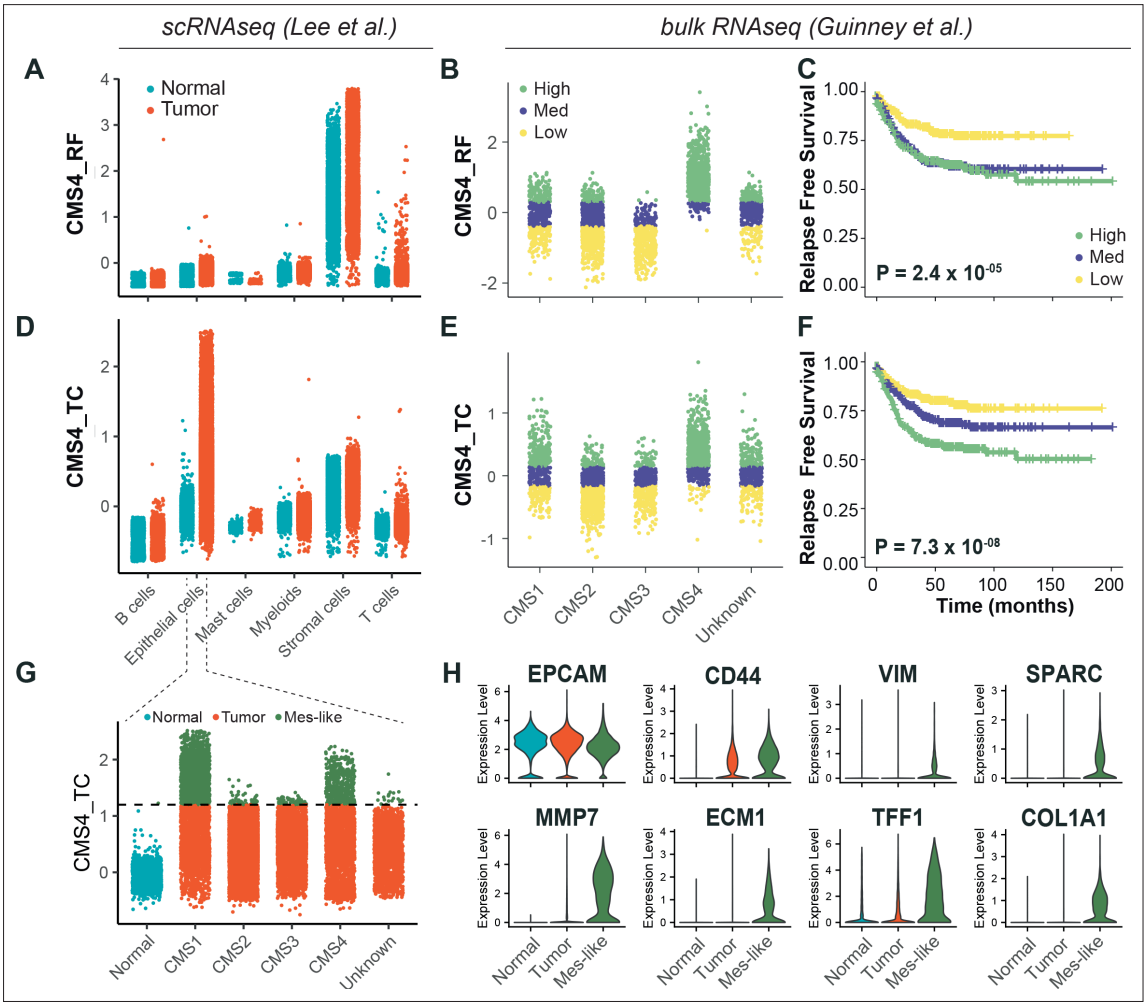


Figure 6. Identification of EpCAM^{lo} cells in primary colorectal tumors. (**A**) Expression of the CMS4_RF signature in the scRNAseq data from Lee et al. (N = 91,103 cells) indicates high association to the stromal cells. (**B**) Expression of the CMS4_RF signature in the bulk RNAseq data from Guinney et al. (N = 3232 tumors) shows association to CMS4 tumors. Tumors were grouped in three equal groups according to their association with the CMS4_RF signature. (**C**) Kaplan–Meier plot of the three CMS4_RF groups shows significant differences in relapse-free survival. (**D**) Expression of the CMS4_TC signature in the scRNAseq data reveals high association to the tumor epithelial cells. (**E**) Expression of the CMS4_TC signature in the bulk RNAseq data shows association to CMS1 and CMS4 tumors. Tumors were grouped in three equal groups according to their association with the CMS4_TC signature. (**F**) Kaplan–Meier plot of the three CMS4_TC groups shows significant differences in relapse-free survival. (**G**) Expression of the CMS4_TC signature in the fraction of epithelial cells (N = 24,707 cells). Cells from CMS1 and CMS4 tumors show the highest association to the CMS4_TC signature and were annotated as mes-like. (**H**) Violin plots of normal, tumor, and mes-like tumor cells showing expression patterns across different genes.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Survival analysis using the CMS4_RF and CMS4_TC signatures across the different consensus molecular subtypes.

it is plausible to think that the 'full-blown' activation of this pathway, earmarked by nuclear β-catenin, may trigger EMT and endow phenotypic plasticity in a minority of colon cancer cells located at the

invasive front. The transient and reversible nature of these cellular modifications is likely driven by epigenetic changes at EMT-TFs.

Here, we have taken advantage of the coexistence of biologically distinct cellular phenotypes within commonly employed colon cancer cell lines to study the cellular and molecular mechanisms that underlie phenotypic plasticity and the acquisition of invasive and metastatic capacity. While the bulk of the cancer cells have a characteristic epithelial phenotype (here referred to as EpCAM^{hi}), a minority of cells with mesenchymal morphology and plastic features (EpCAM^{lo}) coexist in a dynamic equilibrium through stochastic state transitions with their more committed counterpart. Similar observations have been previously made in breast cancer and oral squamous cell carcinoma cell lines (**Gupta et al., 2011**; **Biddle et al., 2011**), which underlines their relevance for the study of phenotypic plasticity in a broad spectrum of malignancies.

As predicted by their morphology, EpCAM^{lo} cells feature increased invasive and metastatic capacity and a distinct gene expression profile when compared to their epithelial counterpart. In particular, EMT activation resulting from enhanced Wnt signaling and *ZEB1* upregulation are distinctive features of the EpCAM^{lo} transcriptome. Of note, *ZEB1* plays many pleiotropic roles ranging from the modulation of oncogenic and tumor-suppressive pathways, cell-fate determination, stemness, and cell plasticity (**Caramel et al., 2017**) and is as such likely to be a determinant in colon cancer invasion and metastasis. From this perspective, the recent debate on EMT as an essential requirement for metastasis (**Maheswaran and Haber, 2015**) likely mirrors the complexity of transcription factors and their downstream targets involved in these processes. Moreover, a recent scRNAseq comparative study of various time-course EMT models has revealed very limited overlap among differentially expressed genes indicative of the vastly context-dependent nature of these processes (**Cook and Vanderhyden, 2020**). The here presented data emphasize the key role played by EMT and its reverse MET in colon tumor cell dissemination and distant metastasis formation.

The observed role of enhanced Wnt signaling in EpCAM^{lo} cells is of interest in view of the presence of *APC* and *CTNNB1* (β-catenin) gene mutations in these cell lines (**Nishisho et al., 1991**; **Ilyas et al., 1997**). Further enhancement of Wnt signaling is likely to result in EMT activation through *ZEB1* upregulation and its conversion from a repressor to activator state (**Sánchez-Tilló et al., 2011**). On its turn, increased *ZEB1* expression was also shown to enhance Wnt, possibly in an autocrine stimulatory circuit (**Sánchez-Tilló et al., 2015**). Accordingly, Wnt 'super-activation' by GSK3β inhibition (i.e., Chiron) in the *APC*/*CTNNB1*-mutated colon cancer cell lines results in a significant expansion of EpCAM^{lo} cells and increased *ZEB1* expression. Vice versa, *ZEB1* downregulation by shRNA leads to a dramatic contraction of the EpCAM^{lo} subpopulation. Notably, isogenic subclones of HCT116 carrying wild type, heterozygous, and homozygous mutations at the β-catenin locus (**Kim et al., 2019**) show a progressive increase in the size of the EpCAM^{lo} subpopulation. These observations highlight the relevance of Wnt dosages in the downstream activation of EMT.

Transcriptional regulation of EMT-TFs has been shown to be controlled by miRNAs binding to specific 3′UTR sequences. In particular, members of the miR-200 family inhibit EMT and cancer cell motility by suppressing Wnt/β-catenin signaling and directly interfering with the expression of EMT-TFs and in particular of *ZEB1/2* (**Brabletz and Brabletz, 2010**; **Ghahhari and Babashah, 2015**). Accordingly, we have shown that the expression of different members of the miR-200 family is downregulated in EpCAM^{lo} colon cancer cells when compared with their epithelial counterpart. Moreover, the RKO cell line, composed entirely of EpCAM^{lo} cells, was previously reported to be characterized by epigenetic downregulation of miR-200a/b/c, leading to high *ZEB1* expression levels (**Davalos et al., 2012**) in the absence of Wnt-activating mutations (**Sparks et al., 1998**). As the epigenetic activation/silencing of the miR-200 locus is dynamically regulated by common EMT-inducers such as TGFβ and Wnt (**Ghahhari and Babashah, 2015**; **Davalos et al., 2012**), it is likely that phenotypic plasticity is triggered in vivo in cells located at the invasive front of colon carcinomas as a result of a complex interplay between their specific mutation spectra and epigenetic profiles modulated by different signaling cues from the microenvironment. The consistent presence of cells with nuclear co-localization of ZEB1 and β-catenin located at the invasive front of patient-derived colon carcinomas validates these observations. Other EMT-TFs like ZEB2 are likely to be upregulated in ZEB1-negative cases (**Kahlert et al., 2011**).

The clinical relevance of our results, originated from the analysis of immortalized cancer cell lines, came from the bioinformatic mining of bulk and scRNAseq data from patient-derived colon cancers.

First, EpCAM^{lo} expression signatures are highly correlated with the CMS4 signature (**Guinney et al., 2015**) and with shorter disease-free and overall survival. The mesenchymal CMS4 subtype accounts for approximately 25% of the cases and is associated with the greatest propensity to form metastases. While the CMS4 signature was initially attributed to the presence of tumor-associated stroma and not to a cancer cell-intrinsic property (**Calon et al., 2015**; **Isella et al., 2015**), it has been shown that *bona fide* tumor cells from stroma-rich cancers have a markedly mesenchymal gene expression pattern (**Vellinga et al., 2016**). By taking advantage of the Lee et al. scRNAseq study (**Lee et al., 2020**), we derived a tumor-specific signature (*CMS4_TC*) by selecting epithelial-specific genes from the original CMS4_RF signature that outperforms other commonly employed prognostic and predictive markers (e.g., *BRAF* mutation and MSI; *data not shown*) in stratifying CMS1, 3, and 4 colon cancer patients based on overall survival. These results highlight the relevance of our study in the identification of a quasi-mesenchymal cellular state with plastic, invasive, and metastatic properties predictive of poor prognosis in colon cancer patients regardless of their CMS classification.

Recent studies on the role of EMT in eliciting phenotypic plasticity in cancer cells have highlighted the relevance of intermediate cellular stages co-expressing both epithelial and mesenchymal genes for tumor progression and metastasis (**Jolly et al., 2015**; **Aiello et al., 2018**; **Pastushenko et al., 2018**). These hybrid E/M or partial EMT cells are thought to be endowed with increased invasive and metastatic capacity. Our scRNAseq analysis of EpCAM^{lo} colon cancer cells has revealed not only fully mesenchymal but also hybrid E/M subclusters, the latter predicted in silico to underlie the observed transcriptional heterogeneity. Of note, transcriptional activation of specific arrays of E- (epi1/2) and M- (mes1/2) genes accompanies the transition between cellular states. Partial EMT cells, predicted by RNAvelocity to transit between states, are characterized not only by the co-expression of E- and M-specific genes at intermediate levels, but also by increased expression of specific genes like *SPARC* (Secreted Protein Acidic and Rich in Cysteine, also known as osteonectin), encoding for a matricellular protein involved in the modulation of cell–cell and cell–matrix interactions and known as a prognostic marker in colon cancer (**Kim et al., 2013**). The role of SPARC in cancer is controversial as it has been shown to promote EMT and metastasis, but also to encompass tumor-suppressive functions in a context-dependent fashion (**Podhajcer et al., 2008**). Importantly, SPARC triggers EMT through direct cell-to-cell contact and upregulation of other EMT-inducers like fibronectin (FN1) (**Takigawa et al., 2017**), reminiscent of the interactions occurring between parenchymal and stromal cells at the invasive front in colon cancer (**Fodde and Brabletz, 2007**) where pEMT is expected to underlie plasticity and invasion through the ECM. Although SPARC is unique in its expression peaking at pEMT states and earmarking the transition between EpCAM^{lo} and EpCAM^{hi} in both cell lines and in mes-like cells from patient-derived colon cancers, it seems unlikely that specific genes exist that can independently elicit pEMT. Instead, tumor-specific and context-dependent activation of subset of genes with distinct functions (e.g., *SPARC*, *FN1*, *MMP7*, *ZEB1*, *ECM1*) synergistically promoting, for example, collective cell migration upon interaction with the stromal microenvironment, may represent a more likely scenario.

The metastable and plastic features of EpCAM^{lo} cells were further highlighted by their striking capacity of giving rise to distant metastases reminiscent of the primary tumors both in the prevalence of EpCAM^{hi} cells and in the distinct patterns of β-catenin intracellular localization between the periphery and center of the lesions. These observations are in agreement with the key role of MET for the onset of distant metastasis (**Brabletz et al., 2005**). The admittedly marginal increase in liver and lung metastases upon continuous induction of *ZEB1* expression is likely to result from MET inhibition. From this perspective, our results provide support to the 'migrating CSC' (**Brabletz et al., 2005**) and 'β-catenin paradox' (**Fodde and Brabletz, 2007**) models and shed light on the cellular and molecular mechanisms underlying Wnt super-activation, (p)EMT induction, and cell dissemination along the invasive front of colon carcinomas.

The characterization of the cellular and molecular mechanisms underlying phenotypic plasticity in colon cancer cells along the invasion-metastasis cascade has great relevance for the development of future diagnostic and therapeutic approaches based on circulating tumor cells (CTCs). Although current detection methodologies mainly rely on their epithelial features, more recent studies have indicated that CTCs with less pronounced epithelial characteristics are likely to underlie metastasis at distant organ sites. In breast and oral cancer, hybrid E/M tumor cells feature the highest degree of phenotypic plasticity coupling the capacity of undergoing EMT/MET with therapy resistance and

other stem cells features (***Yu et al., 2013***; ***Biddle et al., 2016***). Further single-cell RNAseq and functional analyses of extensive cohorts of EpCAM^{lo} cells from matched primary tumors, CTCs, and the corresponding distant metastases will open new avenues for diagnostic and therapeutic applications.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Cell line (*Homo sapiens*)	HCT116 (adult colorectal carcinoma)	ECACC	Cat# 91091005, RRID:CVCL_0291	For sorting of subpopulations; see Materials and methods, section Flow cytometry and sorting
Cell line (*Homo sapiens*)	HCT116-WT, -P, and -MT (wild type, hetero- and homozygous for the b-catenin Ser45del mutant allele)	Author:A.K. (***Kim et al., 2019***)		
Cell line (*Homo sapiens*)	SW480 (adult colorectal carcinoma)	ECACC	Cat# 87092801, RRID:CVCL_0546	For sorting of subpopulations, see Materials and methods, section Flow cytometry and sorting
Biological sample (*Mus musculus*)	APC-Kras-P53 (AKP) organoids (*Apc$^{fl/fl}$::Kras$^{G12D/+}$::Trp53$^{fl/R172H}$*)	Author: J.vR (***Fumagalli et al., 2017***; ***Fumagalli et al., 2018***)		
Biological sample(*Mus musculus*)	AKP-Zeb1 (AKP-Z) (dox-inducible *Zeb1* expression)	This study		See Materials and methods, section Construction of Zeb1 inducible vector
Antibody	Anti-mouse CD44-APC (rat monoclonal)	BD Pharmingen	Cat# 559250, RRID:AB_398661	FACS (1 µg/10^6 cells)
Antibody	Anti-human EpCAM-FITC (mouse monoclonal)	GeneTex	Cat# GTX30708, RRID:AB_1240769	FACS (1 µg/10^6 cells)
Antibody	Anti-human EpCAM (mouse monoclonal)	Santa Cruz Biotechnology	Cat# sc-66020, RRID:AB_2098654	IF (1:250)
Antibody	Anti-human ZEB1 (rabbit polyclonal)	Santa Cruz Biotechnology	Cat# sc-25388, RRID:AB_2217979	IF (1:200)
Antibody	Anti-human β-catenin (mouse monoclonal)	BD Biosciences	Cat# 610154, RRID:AB_397555	IHC (1:500)
Antibody	Anti-human ZEB1 (rabbit polyclonal)	Sigma-Aldrich	Cat# HPA027524, RRID:AB_1844977	IHC (1:200)
Chemical compound, drug	5-fluorouacil	Sigma-Aldrich	Cat# F6627	1.5–100 µg/mL
Chemical compound, drug	Oxaliplatin	Sigma-Aldrich	Cat# O9512	1.25–20 µg/mL
Commercial assay or kit	TruSeq Sample Preparation	Illumina	Cat# 15026495F	v.2
Commercial assay or kit	Single Cell 3´ Reagent Kit Protocol	10XGenomics	Cat# CG00052	v2 chemistry
Software, algorithm	CellRanger	10XGenomics	RRID:SCR_017344	Version 2.1.1
Software, algorithm	R	Seurat, GSVA, MAGIC (***Stuart et al., 2019***; ***van Dijk et al., 2018***; ***Hanzelmann et al., 2013***)	RRID:SCR_007322, RRID:SCR_021058	Version 4.0.4
Software, algorithm	Python	Velocyto, scVelo (***La Manno et al., 2018*** ; ***Bergen et al., 2020***)	RRID:SCR_018167, RRID:SCR_018168	Version 3.8.3

Cell cultures

Human colon cancer cell lines were acquired from the European Collection of Authenticated Cell Cultures (ECACC) and cultured in DMEM medium (11965092, Thermo Fisher Scientific) (supplemented with 10% heat-inactivated fetal bovine serum [FBS; Thermo Fisher Scientific], 1% penicillin/streptomycin [penicillin: 100 U/mL, streptomycin: 100 µg/mL; 15140122 Thermo Fisher Scientific]) in humidified atmosphere at 37°C and 5% CO_2. The identity of each cell line was confirmed by DNA fingerprinting with microsatellite markers (Amelogenin, CSF1PO, D13S317, D16S539, D5S818, D7S820, THO1, TPOX, vWA, D8S1179, FGA, Penta E, Penta D, D18S51, D3S1358, D21S11) and compared with the analogous data provided by ATCC, EACC, and https://web.expasy.org/cellosaurus/ (data not shown). The HCT116-P, HCT116-WT, and HCT116-MT cell lines were kindly given by the laboratory of Hoguen Kim from the Yonsei University College of Medicine in Seoul, Korea. All employed cell lines tested negative for *Mycoplasma*.

Apc$^{fl/fl}$*::Kras*$^{G12D/+}$*::Trp53*$^{fl/R172H}$ (AKP) organoids were grown in 50 µL droplets of Matrigel (Corning) covered with Advanced DMEM-F12 medium (12634028, Thermo Fisher Scientific) supplemented with 1× GlutaMAX (35050-061, Thermo Fisher Scientific), HEPES (15630-056, Thermo Fisher Scientific), gentamicin (15750060, Thermo Fisher Scientific), 100 ng/mL mEgf (PMG8041, Thermo Fisher Scientific), 2% B27 supplement (12587010, Thermo Fisher Scientific), mNoggin (self-produced according to protocol), and 100 µM Y-27632 (Y0503, Sigma Aldrich).

Chiron (CHIR99021, 4423, Tocris, dissolved at 20 mM in DMSO), 4 µM CHIR99021 was added to the culture medium every 48 hr for 1 week.

Flow cytometry and sorting

For flow cytometry and sorting purposes, cells were detached from the culture dishes using trypsin-EDTA (15400054, Thermo Fisher Scientific) and resuspended in staining buffer (PBS supplemented with 4% FCS). Antibody staining was performed on ice for 30'' with CD44-APC (clone IM7, 559250, BD Pharmingen) and EpCAM-FITC (ESA-214, GTX30708, GeneTex). Cells were then washed and resuspended in PBS 4% FCS. Flow cytometry analysis and cell sorting were carried out with a FACSAria III Cell Sorter (BD Biosciences).FSC-H versus FSC-W and SSC-H versus SSC-W were employed to eliminate cell aggregates and ensure single-cell sorting. Gates were defined as depicted and specified in ***Figure 1***. The DAPI nuclear dye (D9542, Sigma-Aldrich) was used at 1 µg/mL to exclude dead cells. FITC and GFP were analyzed using 488 nm laser and 502 LP and 530/30 BP filters; APC and Sytox Red with a 633 nm laser and a 660/20 BP filter; BV421 using a 405 nm laser and a 450/40 BP filter; BV785 with a 405 nm laser and a 750 LP and 780/60 BP filter; PE with a 461 nm laser and a 582/15 BP filter.

For the clonogenicity assays, single cells were sorted into 96-well plates at 1 cell/well in culture medium. After sorting, the plates were checked under the microscope to confirm the presence of a single cell per well. Upon confluency, each single clone was FACS-analyzed for CD44/EpCAM at different time points.

For the flow cytometry analysis of mouse liver metastases and primary AKP-Z organoid-derived tumors, freshly resected tissue samples were cut into small pieces, washed three times with ice-cold PBS, and subsequently digested in Advanced DMEM-F12 (12634028, Thermo Fisher Scientific) containing 50 mg/mL of collagenase A (11088793001, Roche) for 30' at 37°C, with vigorous pipetting every 10'. Residual tissue fragments were further treated with TrypLE Express (12605-010, Thermo Fisher Scientific), 0,25% Trypsin (Thermo Fisher Scientific), and 0.1 mg/mL DNaseI (DN25, Sigma-Aldrich) for 45' at 37°C. Samples were then washed and filtered in sequence through a 100 and 40 µm cell strainer (Corning). The HCT116-derived cells from the liver metastasis were identified and analyzed using the following antibodies for 30' on ice: HLA-A, B, C-biotin (clone W6/32, 311434, Biolegend), Streptavidin-PE-Dazzle 549 (Biolegend), rat anti-mouse CD45-BV421 (clone 30-F11, 563890, BD Biosciences), rat anti-mouse CD31-BV421 (clone 390, 563356, BD Biosciences), rat anti-mouse TER-199-BV421 (clone TER-119, 563998, BD Biosciences), CD44-APC (clone IM7, 559250, BD Pharmingen) ,and EpCAM-FITC (ESA-214, GTX30708, GeneTex). AKP-Z tumor-derived cells were stained using the following antibodies: rat anti-mouse CD45-BV421 (clone 30-F11, 563890, BD Biosciences), rat anti-mouse CD31-BV421 (clone 390, 563356, BD Biosciences), rat anti-mouse TER-199-BV421 (clone TER-119, 563998, BD Biosciences), anti-mouse EpCAM-PE (clone G8.8, Biolegend), anti-mouse/human

CD44-APC (clone!M7, BD Biosciences), and DAPI. Intrinsic GFP expression was also used to select for organoid-derived cells.

Cell proliferation assay and cell cycle analysis

For proliferation assays, sorted cells were seeded in 24-well dishes (2500 cells/well) and cultured in conventional medium. Cells were harvested and counted at the indicated time points. For cell cycle analysis, sorted cells were centrifuged, fixed in 70% ethanol, and stained with 50 µg/mL propidium iodide (PI) (Sigma) and 0.2 mg/mL RNAse A (10109142001, Sigma-Aldrich). The stained cells were then FACS-analyzed using a 655 LP and a 695/40 BP filter.

Cell viability assays

For the chemoresistance assays, cells were seeded in 96-well plates at 15,000 cells/well and left overnight to adhere. Three technical replicates were plated per tested condition. Both oxaliplatin (Sigma-Aldrich) and 5-fluorouacil (5-FU; Sigma-Aldrich) were dissolved in DMSO. HCT116 cells were incubated for 3 or 5 days with 5-FU and oxaliplatin, respectively, whereas SW480 cells were treated with 5-FU or oxaliplatin for 7 days (calibrated on the response of the parental line in terms of viability). After removal of the chemotherapeutic drug, cells were washed with PBS and left to regrow in standard culture medium. Cell viability was assessed using the MTT (3-(4,5-dimethylthiazol-2-yl)–2,5-diph enyltetrazolium bromide; Sigma-Aldrich) assay at time 0, that is, upon removal of the drug, and after regrowth (range 1–36 days). Briefly, cells were incubated at 37°C, 5% CO_2 for 3 hr in culture medium supplemented with 0.45 mg/mL MTT. The 96-well plates were then centrifuged at 1000 rpm for 5′ and the culture medium removed. MTT formazan precipitates were solubilized with DMSO. O.D. reading was performed at 595 nm with a Microplate Reader (Model 550, Bio-Rad). Background measurements were subtracted from each data point. MTT-based results were also validated by manual count of trypan blue-positive cells using cells harvested from additional 96 wells. At least two biological experiments were performed for each individual cell line and drug.

Selection of oxaliplatin and 5-FU concentrations and administration times for chemo-resistance assays was as follows. In a predefined range of concentrations (1.5–100 µg/mL for 5-FU and 0.6–50 µg/mL for oxaliplatin) based on previous data, the maximal concentration and length of treatment were selected as the combination able to induce cell death in approximately 80% of $EpCAM^{lo}$ cells. From that concentration, the curve was scaled down by 50% at least three times to reach the lowest concentration to be tested. The minimal treatment length was 72 hr, while the maximal duration was 1 week. Oxaliplatin was considerably slower than 5-FU in its cytotoxicity with HCT116 cells. The difference was less relevant in SW480 cells where numerous cells, although irreversibly damaged, remained viable for several days. For the latter reason, treatment was longer in SW480 cells.

Migration and invasion assays

For the 'transwell' (modified Boyden chamber) migration and invasion assays, cells were starved for 24 hr in DMEM supplemented with 1% FBS. For each tested subpopulation, 1.0×10^6 cells were harvested, resuspended in 1% FBS medium, and plated in transwell tissue culture inserts (8 µm polycarbonate membrane, 24 well, 3428; Corning). For the invasion assays, 1.5 µg/mL Matrigel dissolved in serum-free medium was pipetted on top of the transwell membrane and left to set at 37°C for several hours before plating the cells. DMEM medium with 10% FBS was then placed in the lower compartment of the transwell chamber. For both assays, three technical replicates were plated for each subpopulation. Plates were incubated for 24 hr at 37°C/5% CO_2, after which the cells and membranes were fixed in methanol and stained with crystal violet. The non-migrated cells on the upper surface of the membrane were removed with a cotton-wool bud and the cells migrated to the lower surface of the membrane counted. For both assays, two independent experiments were performed.

RNA isolation and qRT-PCR

RNA was isolated from cells using TRIzol Reagent (15596018, Thermo Fisher Scientific) according to the manufacturer's instructions. RNA concentration was measured by NanoDrop. Subsequently, reverse transcription into cDNA was conducted using the High-Capacity cDNA Reverse Transcription Kit (4368814, Life Technologies) according to the manufacturer's instructions. RT-qPCR was performed with Fast SYBR Green Master Mix (4385617; Thermo Fisher Scientific). PCR amplification and detection

were implemented with the 7400 Fast Real-Time PCR System. Expression levels were normalized for the endogenous *GAPDH* reference gene. All PCR primers here employed are listed below:

Gene	Forwardprimer	Reverseprimer
GAPDH	5′-ACCCAGAAGACTGTGGATGG-3′	5′-TCTAGACGGCAGGTCAGGTC-3′
EPCAM	5′-GCAGCTCAGGAAGAATGTG-3′	5′-CAGCCAGCTTTGAGCAAATGAC-3′
CDH1	5′-TGCCCAGAAAATGAAAAAGG-3′	5′-GTGTATGTGGCAATGCGTTC-3′
VIM	5′-GAGAACTTTGCCGTTGAAGC-3′	5′-GCTTCCTGTAGGTGGCAATC-3′
CDH2	5′-CAACTTGCCAGAAAACTCCAGG-3′	5′-ATGAAACCGGGCTATCTGCTC-3′
FN1	5′-CAGTGGGAGACCTCGAGAAG-3′	5′-TCCCTCGGAACATCAGAAAC-3′
ZEB1	5′-GCACAACCAAGTGCAGAAGA-3′	5′-CATTTGCAGATTGAGGCTGA-3′
ZEB2	5′-TTCCTGGGCTACGACCATAC-3′	5′-TGTGCTCCATCAAGCAATTC-3′
TWIST1	5′-GTCCGCAGTCTTACGAGGAG-3′	5′-GCTTGAGGGTCTGAATCTTGCT-3′
SNAI1	5′-GCGAGCTGCAGGACTCTAAT-3′	5′-CCACTGTCCTCATCTGACA-3′
SNAI2	5′-GGGGAGAAGCCTTTTTCTTG-3′	5′-TCCTCATGTTTGTGCAGGAG-3′
FOXC2	5′-GCCTAAGGACCTGGTGAAGC-3′	5′-TTGACGAAGCACTCGTTGAG-3′

miRNA

For miRNA experiments, sorted cell populations from HCT116 and SW480 cell lines were cultured up till 70% confluency in 6-well-multiwell plates. Total RNA was then isolated using the Trizol (Ambion) protocol. For *miR-200-family (miR-200a, miR-200b, miR-200c, miR-141, miR-429)*, reverse transcription was performed using the TaqMan MicroRNA RT Kit (Applied Biosystems), followed by qRT-PCR using TaqMan MicroRNA assay (Thermo Fisher Scientific). *U6* snRNA was used as tanendogenous control. The expression of *miR-200-family* was analyzed by the ΔCT method. The experiment was repeated three times.

shRNA

In order to knock down *ZEB1* expression, lentiviral-inducible shRNA vectors encompassing control or *ZEB1* sequences were developed. Cloning was performed according to the manufacturer's instructions (***Wiederschain et al., 2009***). Briefly, Tet-pLKO-puro (gift from D. Wiederschain, Addgene plasmid # 21915) was digested with AgeI and EcoRI and isolated by gel purification (QIAEX II Gel Extraction Kit, Qiagen). The sequences of the control and *ZEB1* shRNA were as follows: shRNA control (shCT) (Addgene sequence #1864): top 5′-CCGGCCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGGTTTTTTG-3′, bottom: 5′-AATTCAAAAAACCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG-3′, shZEB1-A: target sequence (Broad Institute): 5′-GCTGCCAATAAGCAAACGATT-3′ oligo sequence: top: 5′-CCGGGCTGCCAATAAGCAAACGATTCTCGAGAATCGTTTGCTTATTGGCAGCTTTTT-3′, bottom: 5′-AATTAAAAAGCTGCCAATAAGCAAACGATTCTCGAGAATCGTTTGCTTATTGGCAGC-3′, shZEB1-B: target sequence (Broad Institute): 5′-GTCTGGGTGTAATCGTAAATT-3′ oligo sequence: top 5′-CCGGGTCTGGGTGTAATCGTAAATTCTCGAGAATTTACGATTACACCCAGACTTTTT-3′, bottom 5′- AATTAAAAAGTCTGGGTGTAATCGTAAATTCTCGAGAATTTACGATTACACCCAGAC-3′.

The 'top' and 'bottom' labels indicate the oligonucleotide that were annealed in 0.1 M NaCl, 10 mM Tris HCl, pH 7.4, after incubation at 95°C followed by a cooling down step until room temperature is reached. The digested vector (200 ng) was ligated with 1 µL of the oligonucleotides (0.45 nmol/µL) using T4 DNA ligase (Promega) for 3 hr at room temperature. 'One Shot Stbl3' chemically competent *Escherichia coli* (Thermo Fisher Scientific) were transformed with the ligation product. Individual colonies were inoculated for mini-prep cultures, and the presence of the insert was checked using the restriction enzyme XhoI and by sequencing.

For lentivirus production, the shCT or the pool of the two *ZEB1* shRNA constructs were packaged into second-generation virus particles using psPAX2 (Addgene plasmid # 12260; gift from Dr. Didier Trono) and pMD2.G (Addgene plasmid # 12259) into HEK293T. Virus particles were titrated with

the HCT116 cell line and a MOI of 0.5 was employed to produce the shRNA-inducible HCT116 and SW480 cell lines. Lentivirus-infected cells were selected in medium containing 1 µg/mL puromycin (Dulbecco). shRNA induction was implemented by using 1 µg/mL doxycycline for 72 hr. The extent of *ZEB1* downregulation was assessed by RT-qPCR.

Construction of Zeb1-inducible vector

A pORF mZeb1 plasmid (Cat no. ORF062179) encompassing the cDNA of the *Zeb1* gene was employed for the construction of the lentiviral vector. The *Zeb1* gene was cloned into the pEN_TmiRc3 plasmid, gift from Iain Fraser (California Institute of Technology, CA). The *Zeb1* insert was transferred into a pSLIK-Hygro plasmid (#25737; Addgene, USA) by Gateway technology. All newly generated plasmids were sequence-verified by LGC Genomics (LGC Genomics GmbH, Germany). The pSLIK-Hygro plasmid was transiently transfected by Fugene HD (Promega) in HEK293T cells together with the VSV-G, MD, and REV packaging plasmids. After 2 days, the culture medium was collected and viral particles were concentrated by ultracentrifugation. AKP organoids were transduced with the concentrated virus. After 24 hr, the transduced cells were selected by hygromycin B (Thermo Fisher Scientific, The Netherlands) (60 µg/mL) for 7 days. Expression of *Zeb1* was confirmed by qRT–PCR upon stimulation with doxycycline hyclate.

Animal experiments

All protocols involving animals were approved by the Dutch Animal Experimental Committee and conformed to the Code of Practice for Animal Experiments in Cancer Research established by the Netherlands Inspectorate for Health Protections, Commodities and Veterinary Public health (The Hague, the Netherlands, 1999). Animals were bred and maintained in the Erasmus MC animal facility (EDC) under conventional specific pathogen-free (SPF) conditions.

Spleen transplantation assays were implemented on 6- to 8-week-old NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) male and female mice anesthetized intraperitoneally with ketamine (Ketalin, 0.12 mg/mL) and xylazine (Rompun, 0.61 mg/mL). Carpofen (Rimadyl, 5 mg/mL) was given subcutaneously as analgesia. The spleen was exteriorized through a left lateral flank incision and 2.5×10^4 HCT116 and SW480 cells, resuspended in 50 µL of PBS, were injected into the spleen parenchyma using an insulin syringe. Fifteen minutes after injection, a splenectomy was performed with a fine tip cautery pen in order to remove spilled cells and ensure hemostasis. The peritoneum and skin were then sutured in two layers. Mice injected with HCT116 cells were sacrificed 4 weeks after injection for tumor collection; mice injected with SW480 cells were killed 8 weeks after spleen transplantation. Upon liver resection, individual macroscopic metastatic lesions were counted, dissected, and fixed in 4% paraformaldehyde (4% PFA). Residual liver tissue and the lungs were also fixed in 4% PFA for further immunohistochemical analyses.

Orthotopic transplantation of intestinal mouse organoids was performed as previously described (***Fumagalli et al., 2018***). In brief, the day before transplantation AKP-Zeb1 organoids containing a GFP and click beetle luciferase vector (ATG-1929, a gift from Keith Wood, Addgene plasmid #108712) were collected and mechanically dissociated into small clumps. About 100,000 cells were plated in 10 µL drops neutralized Rat Tail High Concentrate Type I Collagen (Corning, Cat. no. 354249) and let to recover overnight at 37°C in Advanced DMEM-F12 medium (12634028, Thermo Fisher Scientific) 100 ng/mL mEgf (PMG8041, Thermo Fisher), 2% B27 supplement (12587010, Thermo Fisher), mNoggin (self-produced according to protocol), and 100 µM Y-27632 (Y0503, Sigma Aldrich). Caecum transplantation assays were implemented on 6- to 8-week-old male NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice anesthetized intraperitoneally with ketamine (Ketalin, 0.12 mg/mL) and xylazine (Rompun, 0.61 mg/mL). Carpofen (Rimadyl, 5 mg/mL) was given subcutaneously as analgesia. The caecum was exposed through a midline abdominal incision and a collagen drop containing tumor cells was surgically transplanted in the caecal submucosa. The peritoneum and skin were then sutured in two layers. Tumor growth was monitored by abdominal palpation. Mice were sacrificed 6–8 weeks after transplantation. Upon collection, primary caecal tumors were single-cell digested and further analyzed by FACS. Mice used for quantification of liver and lung metastases were injected with Luciferin-D and imaged with an IVIS Spectrum imaging system (Caliper Life Sciences, Hopkinton, MA, USA). After imaging tissues were fixed and cut into 500 µm slices using a Fibrotome, processed for IHC, stained for β-catenin visualize tumor cells, scanned using a NanoZoomer, and counted using NDP view software.

Immunofluorescence analysis

Coverslips containing a monolayer of cancer cells were fixed for 30′ in 4% PFA at 37°C and washed twice with PBS. Cells were first permeabilized for 15 min at room temperature with 0.2% of Triton X-100 and then incubated in blocking buffer (5% milk powder in PBS-Tween) for 1 hr at room temperature. Cells were then exposed overnight at 4°C to primary antibodies against EpCAM (mouse, 1:250; sc-66020; Santa Cruz Biotechnology) and ZEB1 (rabbit, 1:200; sc-25388, Santa Cruz Biotechnology). After washing twice with PBS-Tween, coverslips were incubated for 1 hr at room temperature in blocking buffer containing the following secondary antibodies: Goat anti-Rabbit Alexa Fluor 594 conjugate (1:250, #A-11037, Life Technologies) and Donkey anti-Mouse Alexa Fluor 488 conjugate (1:250, #A-21202, Life Technologies). Cells were counterstained with DAPI to visualize the nuclei. Coverslips were mounted in VECTAHIELD HardSet Antifade Mounting Medium (#H-1400, Vector Labs) and imaged with a Zeiss LSM-700 confocal microscope. Images were processed with ImageJ (U.S. National Institutes of Health, Bethesda, MD, USA).

Immunohistochemistry analysis

Tissues from animal experiments were fixed overnight in 4% PFA and embedded in paraffin. Paraffin blocks containing human colon cancer tissue were obtained from the Department of Pathology at the Erasmus Medical Center in Rotterdam. 4 µm sections were mounted on slides. IHC was performed using the EnVision Plus-HRP system (Dako) and antibodies directed against β-catenin (1:200, 610154, BD Biosciences) and ZEB1 (1:200, HPA027524, Sigma-Aldrich). Briefly, paraffin-embedded sections were dewaxed with Xylene and hydrated in 100 and 70% ethanol. Antigen retrieval was performed using pressure cooker pretreatment in a citrate buffer (pH 6.0) for ZEB1 and in a Tris-EDTA buffer (pH 9.0) for the anti-human-mitochondria and anti-β-catenin antibodies. Subsequently, slides were incubated at room temperature in 3% hydrogen peroxidase for 15′ to block endogenous peroxidase activity. Tissue sections were washed and blocked with 5% BSA in PBS-Tween for 1 hr to then be incubated with the primary antibodies overnight at 4°C. Slides were washed twice with PBS-Tween and incubated with Rabbit EnVision+ System HRP (K4001, Dako) or Mouse EnVision+ System HRP (K4007, Dako) for 30′. Subsequently, signal detection was done and tissues were counterstained with Mayer's Hematoxylin. Dehydration was performed by incubation in 70 and 100% ethanol followed by Xylene before sleds were mounted using Pertex (00811, Histolab).

For the IHC analysis of patient-derived colon cancers, all paraffin blocks were collected anonymously from the archives of the Department of Pathology of the Erasmus MC.

TOP-Flash reporter assay

For the β-catenin/TCF reporter assay (TOP-Flash reporter assay), cells were plated on 48-well dishes and cultured in medium with or without 4 µM CHIR99021. After 48 hr, when 70% confluence was reached, cells were transfected by Fugene HD (Promega) with 125 ng of the TOP-Flash or FOP-Flash reporter constructs together with 25 ng of the Renilla luciferase vector for normalization purposes. Luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega) 24 hr post-transfection. Luminescence was measured using a GloMax Luminometer.

Next-generation sequencing (NGS): RNAseq

RNA quality and quantity was evaluated on a 2100 Bio-analyzer (Agilent) using the Agilent RNA 6000 Pico Kit. RNA samples were further processes according to the TruSeq Sample Preparation v.2 Guide (Illumina) and paired end-sequenced on the HiSeq 2,500 (Illumina).

Illumina paired-end reads of 76 bases were trimmed by removing the TrueSeq adapter sequences using Trimmomatic (v.0.33) (***Bolger et al., 2014***). Subsequently, the reads were mapped in a two-pass procedure to the human reference genome build hg38 with the RNA-seq aligner STAR (v2.4.2a) (***Dobin et al., 2013***) and the *Homo sapiens* GENCODE v23 annotation (***Harrow et al., 2012***). Raw counts were summed with the summarize overlaps function with union mode from the Bioconductor Genomic Alignments package (***Gentleman et al., 2004***) (v1.14.0). Genes were called differentially expressed with a generalized linear model using a negative binomial distribution with correcting for cell lines in multi-cell line comparisons. DESeq2 (v1.16.1) was used to perform these calculations (***Love et al., 2014***). We applied a Wald test to identify statistically significant differently expressed genes. p-values were adjusted using the Benjamini–Hochberg (***Benjamini and Hochberg, 1995***) correction

based on which a threshold value was set at <0.01. MDS was performed after the read counts were normalized with blind variance stabilizing log2 transformation function of DESeq2. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) gene enrichment analyses were carried out as described previously (***Meinders et al., 2015***). R (v 3.4.0) (R Core Team, 2017; https://www.R-project.org/) was employed for statistical analysis and visualization of the data.

Bioinformatics analysis bulk RNAseq

For pathway analysis, generated RNAseq datasets were uploaded into Ingenuity Pathway Analysis software (Qiagen). For other bioinformatics analyses, the generated datasets were uploaded into the R2 Genomics Analysis and Visualization Platform (http://r2.amc.nl). First, we used the 'differential expression between groups' option to identify the hundred genes with highest expression in the EpCAM^{lo} fraction in both the HCT116 and SW480 cells. These genes were saved as separate gene sets. Expression values of all genes in both gene sets were then assessed in the CMS3232 composite cohort (***Guinney et al., 2015***), as well as in a large cohort of profiled cell lines originating from the Broad institute. This yielded single meta-gene expression values per tumor or cell line. These gene set expression values were then stored as separate tracks and compared using the 'relate two tracks' option, yielding Pearson r values and accompanying p values.

The 'relate two tracks' option was also used to compare the two gene sets identifying the EpCAM^{lo} cells with gene sets positively identifying the four different molecular subtypes within the published 273-gene CMS classifier, again yielding r values and corresponding p-values.

The gene sets identifying the EpCAM^{lo} cells from both cell lines were also used to cluster the tumors in the CMS3232 cohort into low, intermediate, and high expression groups by k-means clustering. The Kaplan–Meier method was subsequently used to assess significant differences in survival between the generated subgroups. As all tumors in this cohort had previously been assigned to specific CMS subgroups, we then analyzed the contribution of each CMS subtype to each of the generated low, intermediate, and high expression subgroups.

Single-cell RNAseq

Cell lines were brought to 60–70% confluency before the start of the experiment. For each sample, between 5×10^4 and 1×10^5 EpCAM^{lo} and EpCAM^{hi} cells were FACS sorted and processed using the 10 X Genomics Chromium Single Cell Controller. Samples were deep-sequenced (Illumina) to a depth ranging 49–65k reads/cells. Gene-cell matrices were obtained by conversion of the raw data using the Cell Ranger pipeline. Filtered gene-cell matrices were merged in R and processed for downstream analysis using the Seurat package (***Stuart et al., 2019***). Dimension reduction was performed using PCA, tSNE, and UMAP. Epithelial and mesenchymal scores were computed using the Rmagic (imputation) (***van Dijk et al., 2018***) and GSVA (***Hanzelmann et al., 2013***) (scoring) packages. RNA velocity analysis was done in Python using Velocyto (***La Manno et al., 2018***) and scVelo (***Bergen et al., 2020***) packages.

Statistical analysis

For each experiment, data are shown as mean ± SD. IBM SPSS Statistics software was used for data analysis. The Mann–Whitney U test was used to analyze the difference between two groups of quantitative variables; α-value was set at 5%.

Acknowledgements

This study has been made possible by funding to RF from the Dutch Digestive Foundation (FP 15-09), Dutch Cancer Society, and the Erasmus MC (EMCR 2015-8090). Also, by funding to MP and OS from the Cancer Research UK core funding to the CRUK Beatson Institute (A17196) and to OS (A21139). The authors are grateful to Drs. R Smits, M Trerotola, S Alberti, and Catherine Winchester for their advice, to Dr. Eric Bindels for assistance with RNAseq analysis, and to 10 X Genomics for their support through their Pilot Award Grant Program.

Additional information

Competing interests

Andrea Sacchetti: The authors of this manuscript do not have any competing financial interests in relation to the work described.. Martin M Watson: The other author declares that no competing interests exist.

Funding

Funder	Grant reference number	Author
Maag Lever Darm Stichting	FP 15-09	Riccardo Fodde
KWF Kankerbestrijding	EMCR 2015-8090	Riccardo Fodde
Cancer Research UK	A17196	Owen J Sansom
Cancer Research UK	A21139	Owen J Sansom

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Andrea Sacchetti, Miriam Teeuwssen, conceptualization, formal-analysis, Investigation, Methodology, validation, visualization, writing-original-draft; Mathijs Verhagen, conceptualization, formal-analysis, Investigation, software, visualization, writing-original-draft; Rosalie Joosten, Investigation, Methodology, project-administration, validation; Tong Xu, Roberto Stabile, Berdine van der Steen, Martin M Watson, Alem Gusinac, Investigation, Methodology; Won Kyu Kim, Madelon Paauwe, Resources; Inge Ubink, Investigation, Resources; Harmen JG Van de Werken, Investigation, Methodology, Resources; Arianna Fumagalli, Methodology, Resources; Jacco Van Rheenen, Onno Kranenburg, Methodology, Resources, Writing – review and editing; Owen J Sansom, Resources, Writing – review and editing; Riccardo Fodde, conceptualization, data-curation, formal-analysis, funding-acquisition, project-administration, Resources, supervision, visualization, writing-original-draft, Writing – review and editing

Author ORCIDs

Mathijs Verhagen http://orcid.org/0000-0003-3126-8379
Harmen JG Van de Werken http://orcid.org/0000-0002-9794-1477
Owen J Sansom http://orcid.org/0000-0001-9540-3010
Riccardo Fodde http://orcid.org/0000-0001-9839-4324

Ethics

All protocols involving animals were approved by the Dutch Animal Experimental Committee and were conformed to the Code of Practice for Animal Experiments in Cancer Research established by the Netherlands Inspectorate for Health Protections, Commodities and Veterinary Public health (The Hague, the Netherlands, 1999).

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.61461.sa1
Author response https://doi.org/10.7554/eLife.61461.sa2

Additional files

Supplementary files

- Transparent reporting form

Data availability

The RNA-sequencing data from this study have been submitted to the Gene Expression Omnibus (GEO) database under the accession number GSE154927 and GSE154930 for the bulk and single-cell RNAseq data, respectively.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Verhagen MP, Teeuwssen MJ, Sacchetti A, van de Werken HG, Fodde R	2020		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE154927	NCBI Gene Expression Omnibus, GSE154927
Verhagen MP, Teeuwssen MJ, Sacchetti A, van de Werken HG, Fodde R	2020		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE154930	NCBI Gene Expression Omnibus, GSE154930

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Hong Y, Etlioglu HE, Pomella V, van den Bosch B, Vanhecke J, Tejpar S, Boeckx B, Lambrechts D	2020		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144735	NCBI Gene Expression Omnibus, GSE144735
Hong Y, Lee H, Cho YB, Park W	2020		https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132465	NCBI Gene Expression Omnibus, GSE132465
Guinney J, Dienstmann R, Wang X	2014		https://www.synapse.org/#!Synapse:syn2623706/wiki/67246	Synapse, 10.7303/syn2623706

References

Aiello NM, Maddipati R, Norgard RJ, Balli D, Li J, Yuan S, Yamazoe T, Black T, Sahmoud A, Furth EE, Bar-Sagi D, Stanger BZ. 2018. EMT subtype influences epithelial plasticity and mode of cell migration. *Developmental Cell* **45**: 681-695.. DOI: https://doi.org/10.1016/j.devcel.2018.05.027, PMID: 29920274

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* **57**: 289–300. DOI: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology* **38**: 1408–1414. DOI: https://doi.org/10.1038/s41587-020-0591-3, PMID: 32747759

Bernards R, Weinberg RA. 2002. A progression puzzle. *Nature* **418**: 823. DOI: https://doi.org/10.1038/418823a, PMID: 12192390

Biddle A, Liang X, Gammon L, Fazil B, Harper LJ, Emich H, Costea DE, Mackenzie IC. 2011. Cancer stem cells in squamous cell carcinoma switch between two distinct phenotypes that are preferentially migratory or proliferative. *Cancer Research* **71**: 5317–5326. DOI: https://doi.org/10.1158/0008-5472.CAN-11-1059, PMID: 21685475

Biddle A, Gammon L, Liang X, Costea DE, Mackenzie IC. 2016. Phenotypic Plasticity Determines Cancer Stem Cell Therapeutic Resistance in Oral Squamous Cell Carcinoma. *EBioMedicine* **4**: 138–145. DOI: https://doi.org/10.1016/j.ebiom.2016.01.007, PMID: 26981578

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120. DOI: https://doi.org/10.1093/bioinformatics/btu170, PMID: 24695404

Brabletz T, Jung A, Spaderna S, Hlubek F, Kirchner T. 2005. Opinion: migrating cancer stem cells - an integrated concept of malignant tumour progression. *Nature Reviews. Cancer* **5**: 744–749. DOI: https://doi.org/10.1038/nrc1694, PMID: 16148886

Brabletz S, Brabletz T. 2010. The ZEB/miR-200 feedback loop--a motor of cellular plasticity in development and cancer. *EMBO Reports* **11**: 670–677. DOI: https://doi.org/10.1038/embor.2010.117, PMID: 20706219

Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, Sevillano M, Palomo-Ponce S, Tauriello DVF, Byrom D, Cortina C, Morral C, Barceló C, Tosi S, Riera A, Attolini CS-O, Rossell D, Sancho E, Batlle E. 2015. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nature Genetics* **47**: 320–329. DOI: https://doi.org/10.1038/ng.3225, PMID: 25706628

Caramel J, Ligier M, Puisieux A. 2017. Pleiotropic Roles for ZEB1 in Cancer. *Cancer Research* **78**: 30–35. DOI: https://doi.org/10.1158/0008-5472.CAN-17-2476, PMID: 29254997

Cook DP, Vanderhyden BC. 2020. Context specificity of the EMT transcriptional response. *Nature Communications* **11**: 2142. DOI: https://doi.org/10.1038/s41467-020-16066-2, PMID: 32358524

Davalos V, Moutinho C, Villanueva A, Boque R, Silva P, Carneiro F, Esteller M. 2012. Dynamic epigenetic regulation of the Microrna-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. *Oncogene* **31**: 2062–2074. DOI: https://doi.org/10.1038/onc.2011.383, PMID: 21874049

De Sousa EMF, Wang X, Jansen M. 2013. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nature Medicine* **19**: 614–618. DOI: https://doi.org/10.1038/nm.3174

Dean M, Fojo T, Bates S. 2005. Tumour stem cells and drug resistance. *Nature Reviews. Cancer* **5**: 275–284. DOI: https://doi.org/10.1038/nrc1590, PMID: 15803154

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. Star: Ultrafast Universal RNA-SEQ aligner. *Bioinformatics* **29**: 15–21. DOI: https://doi.org/10.1093/bioinformatics/bts635, PMID: 23104886

Fearon ER, Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. *Cell* **61**: 759–767. DOI: https://doi.org/10.1016/0092-8674(90)90186-i, PMID: 2188735

Fodde R, Brabletz T. 2007. Wnt/beta-catenin signaling in cancer stemness and malignant behavior. *Current Opinion in Cell Biology* **19**: 150–158. DOI: https://doi.org/10.1016/j.ceb.2007.02.007, PMID: 17306971

Fumagalli A, Drost J, Suijkerbuijk SJE, van Boxtel R, de Ligt J, Offerhaus GJ, Begthel H, Beerling E, Tan EH, Sansom OJ, Cuppen E, Clevers H, van Rheenen J. 2017. Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *PNAS* **114**: E2357–E2364. DOI: https://doi.org/10.1073/pnas.1701219114, PMID: 28270604

Fumagalli A, Suijkerbuijk SJE, Begthel H, Beerling E, Oost KC, Snippert HJ, van Rheenen J, Drost J. 2018. A surgical orthotopic organoid transplantation approach in mice to visualize and study colorectal cancer progression. *Nature Protocols* **13**: 235–247. DOI: https://doi.org/10.1038/nprot.2017.137, PMID: 29300390

Gaspar C, Fodde R. 2004. APC dosage effects in tumorigenesis and stem cell differentiation. *The International Journal of Developmental Biology* **48**: 377–386. DOI: https://doi.org/10.1387/ijdb.041807cg, PMID: 15349813

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, et al. 2004. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology* **5**: R80. DOI: https://doi.org/10.1186/gb-2004-5-10-r80, PMID: 15461798

Ghahhari NM, Babashah S. 2015. Interplay between microRNAs and WNT/beta-catenin signalling pathway regulates epithelial-mesenchymal transition in cancer. *European Journal of Cancer* **51**: 1638–1649. DOI: https://doi.org/10.1016/j.ejca.2015.04.021, PMID: 26025765

Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa E Melo F, Missiaglia E, Ramay H, Barras D, Homicsko K, et al. 2015. The consensus molecular subtypes of colorectal cancer. *Nature Medicine* **21**: 1350–1356. DOI: https://doi.org/10.1038/nm.3967, PMID: 26457759

Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, Lander ES. 2011. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**: 633–644. DOI: https://doi.org/10.1016/j.cell.2011.07.026, PMID: 21854987

Hall MP, Woodroofe CC, Wood MG, Que I, Van't Root M, Ridwan Y, Shi C, Kirkland TA, Encell LP, Wood KV, Löwik C, Mezzanotte L. 2018. Click beetle luciferase mutant and near infrared naphthyl-luciferins for improved bioluminescence imaging. *Nature Communications* **9**: 132. DOI: https://doi.org/10.1038/s41467-017-02542-9, PMID: 29317625

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell* **100**: 57–70. DOI: https://doi.org/10.1016/s0092-8674(00)81683-9, PMID: 10647931

Hanzelmann S, Castelo R, Guinney J. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**: 7. DOI: https://doi.org/10.1186/1471-2105-14-7, PMID: 23323831

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, et al. 2012. GENCODE: The reference human genome annotation for the ENCODE project. *Genome Research* **22**: 1760–1774. DOI: https://doi.org/10.1101/gr.135350.111, PMID: 22955987

Hirsch D, Barker N, McNeil N, Hu Y, Camps J, McKinnon K, Clevers H, Ried T, Gaiser T. 2014. LGR5 positivity defines stem-like cells in colorectal cancer. *Carcinogenesis* **35**: 849–858. DOI: https://doi.org/10.1093/carcin/bgt377, PMID: 24282287

Ilyas M, Tomlinson IP, Rowan A, Pignatelli M, Bodmer WF. 1997. Beta-catenin mutations in cell lines established from human colorectal cancers. *PNAS* **94**: 10330–10334. DOI: https://doi.org/10.1073/pnas.94.19.10330, PMID: 9294210

Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, Mellano A, Senetta R, Cassenti A, Sonetto C, Inghirami G, Trusolino L, Fekete Z, De Ridder M, Cassoni P, Storme G, Bertotti A, Medico E. 2015. Stromal contribution to the colorectal cancer transcriptome. *Nature Genetics* **47**: 312–319. DOI: https://doi.org/10.1038/ng.3224, PMID: 25706627

Jolly MK, Boareto M, Huang B, Jia D, Lu M, Ben-Jacob E, Onuchic JN, Levine H. 2015. Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Frontiers in Oncology* **5**: 155. DOI: https://doi.org/10.3389/fonc.2015.00155, PMID: 26258068

Kahlert C, Lahes S, Radhakrishnan P, Dutta S, Mogler C, Herpel E, Brand K, Steinert G, Schneider M, Mollenhauer M, Reissfelder C, Klupp F, Fritzmann J, Wunder C, Benner A, Kloor M, Huth C, Contin P, Ulrich A,

Koch M, et al. 2011. Overexpression of zeb2 at the invasion front of colorectal cancer is an independent prognostic marker and regulates tumor invasion in vitro. *Clinical Cancer Research* **17**: 7654–7663. DOI: https://doi.org/10.1158/1078-0432.CCR-10-2816, PMID: 22042972

Kim JY, Jeong D, Ahn TS, Kim HJ, Park DS, Park SY, Bae SB, Lee S, Lee SS, Lee MS, Cho HD, Baek MJ. 2013. Expression of secreted protein acidic and rich in cysteine in the stroma of a colorectal carcinoma is associated with patient prognosis. *Annals of Coloproctology* **29**: 93–99. DOI: https://doi.org/10.3393/ac.2013.29.3.93, PMID: 23862126

Kim WK, Kwon Y, Jang M, Park M, Kim J, Cho S, Jang DG, Lee WB, Jung SH, Choi HJ, Min BS, Il Kim T, Hong SP, Paik YK, Kim H. 2019. β-catenin activation down-regulates cell-cell junction-related genes and induces epithelial-to-mesenchymal transition in colorectal cancers. *Scientific Reports* **9**: 18440. DOI: https://doi.org/10.1038/s41598-019-54890-9, PMID: 31804558

Kirchner T, Brabletz T. 2000. Patterning and nuclear beta-catenin expression in the colonic adenoma-carcinoma sequence Analogies with embryonic gastrulation. *The American Journal of Pathology* **157**: 1113–1121. DOI: https://doi.org/10.1016/s0002-9440(10)64626-3, PMID: 11021815

La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundström E, Castelo-Branco G, Cramer P, et al. 2018. RNA velocity of single cells. *Nature* **560**: 494–498. DOI: https://doi.org/10.1038/s41586-018-0414-6, PMID: 30089906

Lamouille S, Xu J, Derynck R. 2014. Molecular mechanisms of epithelial-mesenchymal transition. *Nature Reviews. Molecular Cell Biology* **15**: 178–196. DOI: https://doi.org/10.1038/nrm3758, PMID: 24556840

Lee H-O, Hong Y, Etlioglu HE, Cho YB, Pomella V, Van den Bosch B, Vanhecke J, Verbandt S, Hong H, Min J-W, Kim N, Eum HH, Qian J, Boeckx B, Lambrechts D, Tsantoulis P, De Hertogh G, Chung W, Lee T, An M, et al. 2020. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nature Genetics* **52**: 594–603. DOI: https://doi.org/10.1038/s41588-020-0636-z, PMID: 32451460

Lengauer C, Kinzler KW, Vogelstein B. 1997. Genetic instability in colorectal cancers. *Nature* **386**: 623–627. DOI: https://doi.org/10.1038/386623a0, PMID: 9121588

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550. DOI: https://doi.org/10.1186/s13059-014-0550-8, PMID: 25516281

Maheswaran S, Haber DA. 2015. Cell fate: Transition loses its invasive edge. *Nature* **527**: 452–453. DOI: https://doi.org/10.1038/nature16313, PMID: 26560026

Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, Campbell LL, Polyak K, Brisken C, Yang J, Weinberg RA. 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**: 704–715. DOI: https://doi.org/10.1016/j.cell.2008.03.027, PMID: 18485877

Meinders M, Kulu DI, van de Werken HJG, Hoogenboezem M, Janssen H, Brouwer RWW, van Ijcken WFJ, Rijkers E-J, Demmers JAA, Krüger I, van den Berg TK, Suske G, Gutiérrez L, Philipsen S. 2015. SP1/SP3 transcription factors regulate hallmarks of megakaryocyte maturation and platelet formation and function. *Blood* **125**: 1957–1967. DOI: https://doi.org/10.1182/blood-2014-08-593343, PMID: 25538045

Nieto MA, Huang RY, Jackson RA, Thiery JP. 2016. Emt: 2016. *Cell* **166**: 21–45. DOI: https://doi.org/10.1016/j.cell.2016.06.028, PMID: 27368099

Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horii A, Koyama K, Utsunomiya J, Baba S, Hedge P. 1991. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* **253**: 665–669. DOI: https://doi.org/10.1126/science.1651563, PMID: 1651563

Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, Van Keymeulen A, Brown D, Moers V, Lemaire S, De Clercq S, Minguijón E, Balsat C, Sokolow Y, Dubois C, De Cock F, Scozzaro S, Sopena F, Lanas A, D'Haene N, et al. 2018. Identification of the tumour transition states occurring during EMT. *Nature* **556**: 463–468. DOI: https://doi.org/10.1038/s41586-018-0040-3, PMID: 29670281

Podhajcer OL, Benedetti LG, Girotti MR, Prada F, Salvatierra E, Llera AS. 2008. The role of the matricellular protein sparc in the dynamic interaction between the tumor and the host. *Cancer Metastasis Reviews* **27**: 691–705. DOI: https://doi.org/10.1007/s10555-008-9146-7, PMID: 18542844

Sánchez-Tilló E, de Barrios O, Siles L, Cuatrecasas M, Castells A, Postigo A. 2011. β-catenin/tcf4 complex induces the epithelial-to-mesenchymal transition (EMT)-activator zeb1 to regulate tumor invasiveness. *PNAS* **108**: 19204–19209. DOI: https://doi.org/10.1073/pnas.1108977108, PMID: 22080605

Sánchez-Tilló E, de Barrios O, Valls E, Darling DS, Castells A, Postigo A. 2015. Zeb1 and TCF4 reciprocally modulate their transcriptional activities to regulate wnt target gene expression. *Oncogene* **34**: 5760–5770. DOI: https://doi.org/10.1038/onc.2015.352, PMID: 26387539

Skrypek N, Goossens S, De Smedt E, Vandamme N, Berx G. 2017. Epithelial-to-Mesenchymal Transition: Epigenetic Reprogramming Driving Cellular Plasticity. *Trends in Genetics* **33**: 943–959. DOI: https://doi.org/10.1016/j.tig.2017.08.004, PMID: 28919019

Sparks AB, Morin PJ, Vogelstein B, Kinzler KW. 1998. Mutational analysis of the Apc/beta-catenin/tcf pathway in colorectal cancer. *Cancer Research* **58**: 1130–1134 PMID: 9515795.

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive integration of single-cell data. *Cell* **177**: 1888-1902.. DOI: https://doi.org/10.1016/j.cell.2019.05.031, PMID: 31178118

Takigawa H, Kitadai Y, Shinagawa K, Yuge R, Higashi Y, Tanaka S, Yasui W, Chayama K. 2017. Mesenchymal stem cells induce epithelial to mesenchymal transition in colon cancer cells through direct cell-to-cell contact. *Neoplasia* **19**: 429–438. DOI: https://doi.org/10.1016/j.neo.2017.02.010, PMID: 28433772

Teeuwssen M, Fodde R. 2019. Cell heterogeneity and phenotypic plasticity in metastasis formation: The case of colon cancer. *Cancers* **11**: 1368. DOI: https://doi.org/10.3390/cancers11091368, PMID: 31540068

van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, Burdziak C, Moon KR, Chaffer CL, Pattabiraman D, Bierie B, Mazutis L, Wolf G, Krishnaswamy S, Pe'er D. 2018. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**: 716-729.. DOI: https://doi.org/10.1016/j.cell.2018.05.061, PMID: 29961576

Varga J, Greten FR. 2017. Cell plasticity in epithelial homeostasis and tumorigenesis. *Nature Cell Biology* **19**: 1133–1141. DOI: https://doi.org/10.1038/ncb3611, PMID: 28945230

Vellinga TT, den Uil S, Rinkes IHB, Marvin D, Ponsioen B, Alvarez-Varela A, Fatrai S, Scheele C, Zwijnenburg DA, Snippert H, Vermeulen L, Medema JP, Stockmann HB, Koster J, Fijneman RJA, de Rooij J, Kranenburg O. 2016. Collagen-rich stroma in aggressive colon tumors induces mesenchymal gene expression and tumor cell invasion. *Oncogene* **35**: 5263–5271. DOI: https://doi.org/10.1038/onc.2016.60, PMID: 26996663

Wiederschain D, Wee S, Chen L, Loo A, Yang G, Huang A, Chen Y, Caponigro G, Yao Y-M, Lengauer C, Sellers WR, Benson JD. 2009. Single-vector inducible lentiviral RNAi system for oncology target Validation. *Cell Cycle* **8**: 498–504. DOI: https://doi.org/10.4161/cc.8.3.7701, PMID: 19177017

Yi H, Li G, Long Y, Liang W, Cui H, Zhang B, Tan Y, Li Y, Shen L, Deng D, Tang Y, Mao C, Tian S, Cai Y, Zhu Q, Hu Y, Chen W, Fang L. 2020. Integrative multi-omics analysis of a colon cancer cell line with heterogeneous wnt activity revealed RUNX2 as an epigenetic regulator of EMT. *Oncogene* **39**: 5152–5164. DOI: https://doi.org/10.1038/s41388-020-1351-z, PMID: 32535615

Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM, Concannon KF, Donaldson MC, Sequist LV, Brachtel E, Sgroi D, Baselga J, Ramaswamy S, Toner M, Haber DA, Maheswaran S. 2013. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**: 580–584. DOI: https://doi.org/10.1126/science.1228522, PMID: 23372014

Figures and figure supplements

Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer

Andrea Sacchetti *et al*

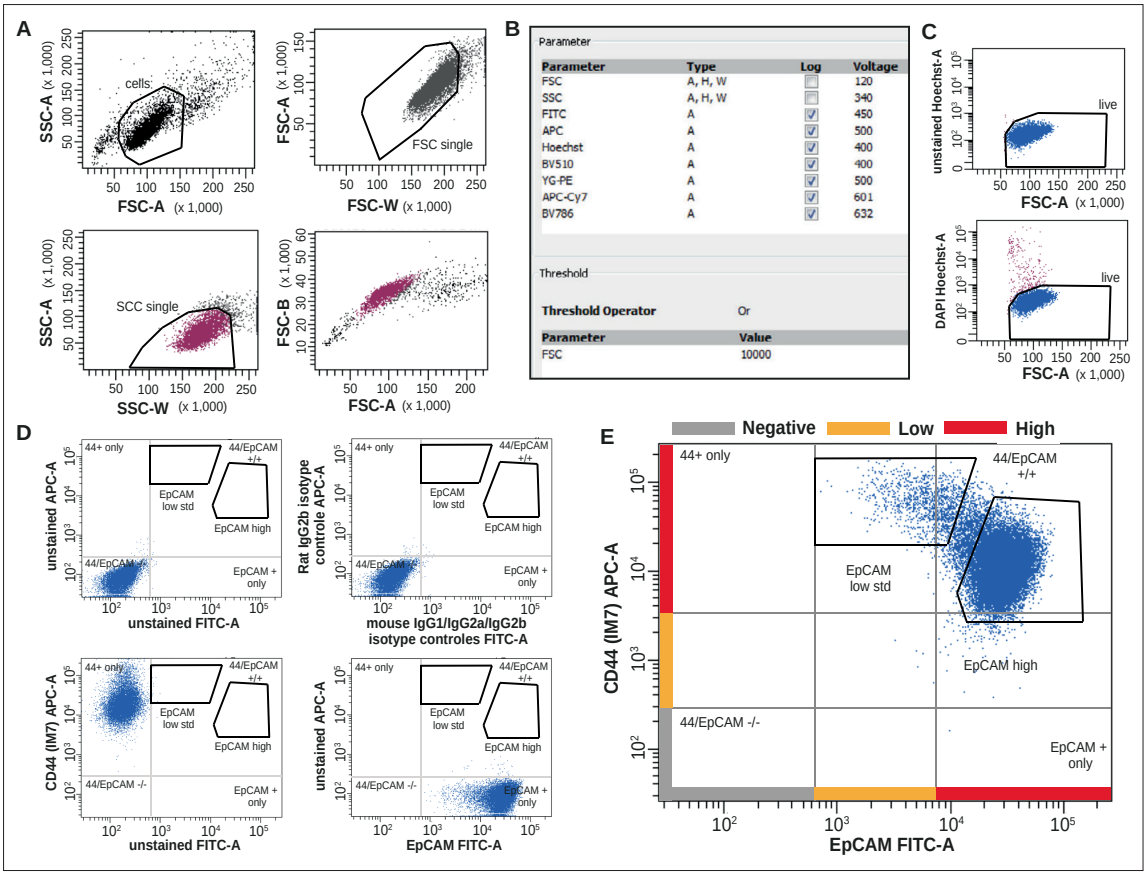


Figure 1—figure supplement 1. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: FACS analysis. (**A**) FSC-A/SSC-A, FSC-W/FSC-A, and SSC-W/SSC-A single-cell gates (confirmed by gating on FSC-A/FSC-H). Purity of sorted single cells was confirmed by microscopy. (**B**) Acquisition parameters used for FACS analysis. (**C**) Gating strategy employed to select for live cells by DAPI staining (top: unstained cells; bottom: after DAPI staining). (**D**) Top: unstained and isotype controls, showing absence of relevant unspecific antibody binding. Bottom: fluorescence minus one (FMO) samples, showing the absence of uncompensated spillover between fluorescence channels. For the sake of clarity, a quadrant is placed to show the relative position of unstained and single stains in both cells lines. For the EpCAM-FITC antibody, a mouse IgG1-FITC isotype control S. Cruz sc-2855 was used; for the CD44-APC antibody, a Rat IgG2a-APC isotype control S. Cruz sc-2895 was used. Similar results were obtained with SW480 cells (not shown). (**E**) Full EpCAM/CD44 staining, also showing the rationale behind the definition of high and low referred to EpCAM and CD44 levels. Despite the differences in CD44 levels between EpCAM^{hi} and EpCAM^{lo} cells, for the sake of simplicity the populations were defined 'low' for a defined marker if they were mainly covering a region within the 1st log above the negative gate, high if they were above the 1st log. FSC-H: forward-scatterheight; FSC-W: forward-scatter width.

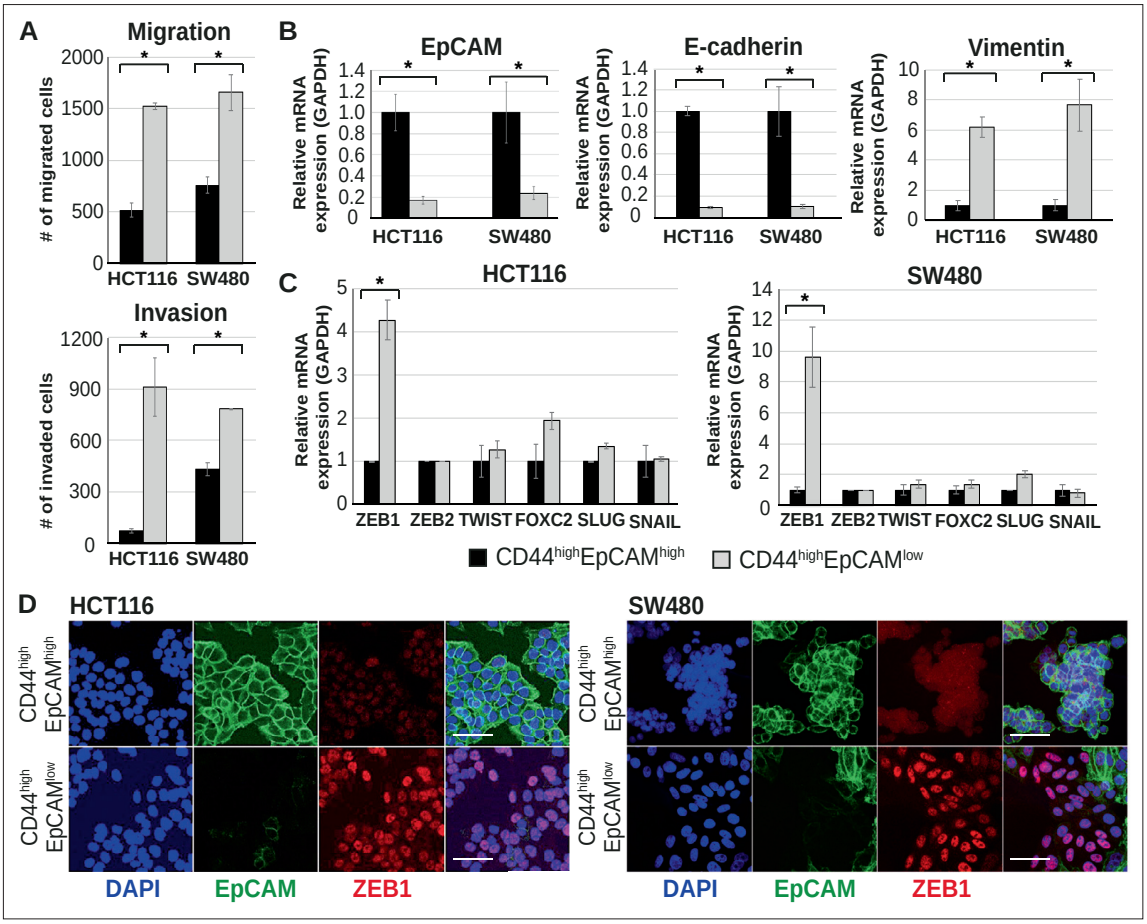


Figure 1—figure supplement 2. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: migration/invasion and EMT analysis. (**A**) Top panel: transwell migration assay of EpCAM^{hi} (black bar) and EpCAM^{lo} (gray bar) cells from the HCT116 and SW480 lines. 10^5 cells were plated on TC-coated membrane in triplicate and left overnight before counting the number of migrated cells on the bottom side of the membrane. Each bar represents the mean ± SD of two independent experiments. Asterisks indicate significant differences ($p<0.05$). Bottom panel: invasion assay of EpCAM^{hi} (black bar) and EpCAM^{lo} (gray bar) cells from the HCT116 and SW480 lines. 10^5 cells were plated in triplicate on top of an extracellular matrix-coated membrane and left overnight before counting the number of cells migrated to other side of the membrane. Each bar represents the mean ± SD of two independent experiments. Asterisks indicate significant differences ($p<0.05$). (**B**) RT-qPCR expression analysis of epithelial (*EPCAM* and *CDH1*) and mesenchymal (*VIM*) markers in sorted EpCAM^{hi} (black bars) and EpCAM^{lo} (gray bars) from the HCT116 and SW480 lines. *GAPDH* was employed for normalization purposes. Each bar represents the mean ± SD of three independent experiments. Asterisks indicate significant differences ($p<0.05$). (**C**) RT-qPCR expression analysis of epithelial to mesenchymal transition transcription factors (*ZEB1, ZEB2, TWIST, FOXC2, SLUG*, and *SNAIL*) in EpCAM^{hi} (black bars) and EpCAM^{lo} (gray bars) cells. Left panel: HCT116. Right panel: SW480. *GAPDH* was employed for normalization. Each bar represents the mean ± SD of three independent experiments. Asterisks indicate significant differences ($p<0.05$). (**D**) Immunofluorescence (IF) analysis of EpCAM^{hi} and EpCAM^{lo} cells. Cells were sorted and directed plated on cover slips. After 4 days, cells were fixed with 4% paraformaldehyde and stained with antibodies against EpCAM (green) and ZEB1 (red). Nuclei were visualized by DAPI staining of DNA (blue). Scale bar: 50 μm.

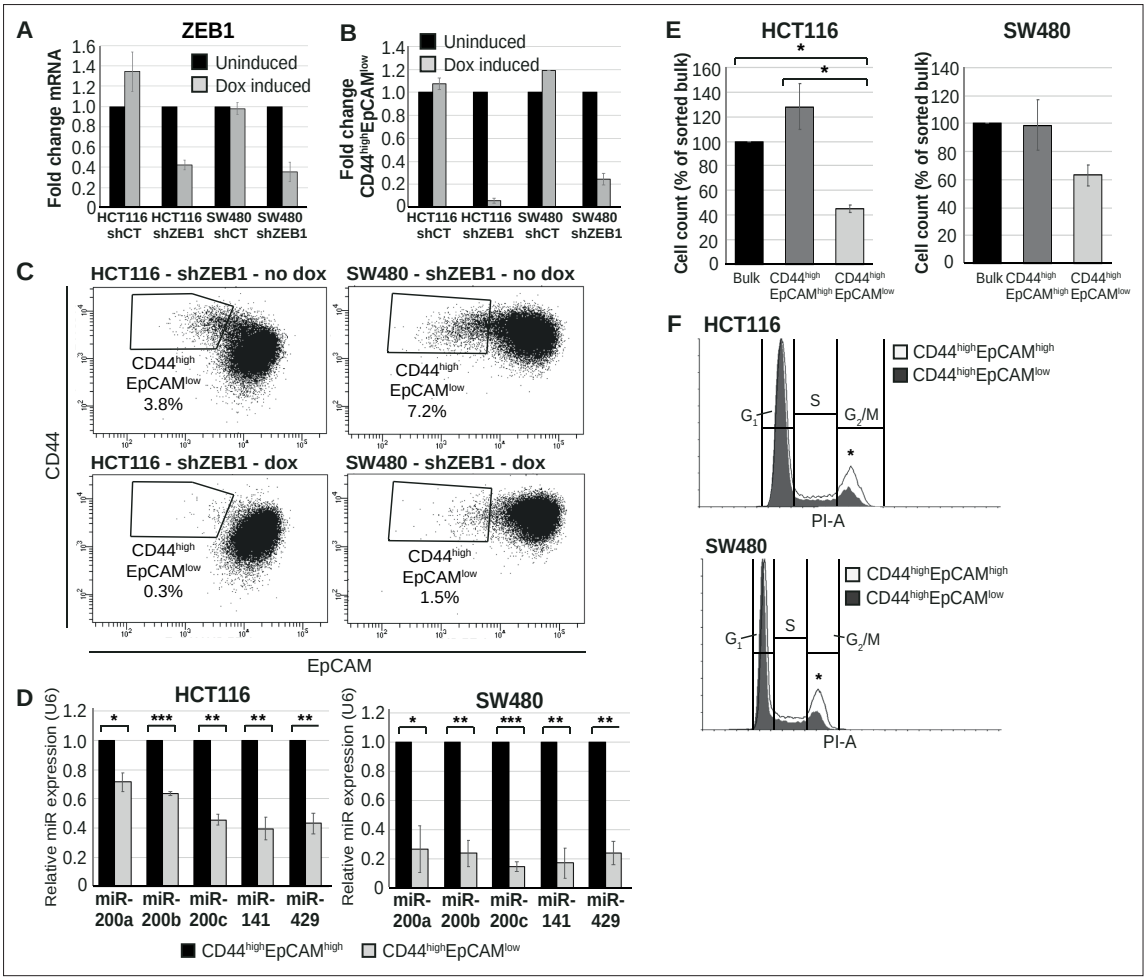


Figure 1—figure supplement 3. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: EMT and cell cycle analysis. (**A**). qRT-PCR expression analysis of *ZEB1* in HCT116 and SW480 transduced with an inducible control (shCT) or ZEB1-shRNA (shZEB1) construct. shRNA expression was induced with 1 μg/mL of doxycycline. Each bar represents the mean ± SD of three independent experiments. (**B**) Bar graph of flow cytometric analysis (see **B**). Each bar represents the relative mean ± SD of three independent experiments. (**C**) Representative analysis of the flow cytometric analysis of the shCT- and shZEB1-transfected HCT116 and SW480 cell lines using antibodies against CD44 and EpCAM. Cells were induced with 1 μg/mL doxycycline for 72 hr before analysis. (**D**) RT-qPCR expression analysis of the members of the miRNA 200 family (miR-200a, miR-200b, miR-200c, miR-141, and miR-429) in EpCAM^{hi} (black bars) and EpCAM^{lo} (gray bars) cells. Upper panel: HCT116. Bottom panel: SW480. *U6* was employed for normalization. Each bar represents the mean ± SD of three independent experiments. Single asterisks indicate significant differences of $p<0.05$, double asterisks of $p<0.01$, and triple asterisks of $p<0.001$. (**E**) Cell proliferation assay. Sorted bulk, EpCAM^{hi} and EpCAM^{lo} cells were seeded in triplicate in plates and cultured in conventional medium. HCT116 and SW480 cells were harvested and number of cells was counted at 4 and 11 days, respectively. Each bar represents the mean ± SD three independent experiments. (**F**) Cell cycle analysis of EpCAM^{hi} and EpCAM^{lo} cells in HCT116 (upper panel) and SW480 (lower panel). Cell fractions were sorted and plated in culture. After 72 hr, cells were fixed and stained with propidium iodide. Cell cycle distribution was assayed by flow cytometry. Graphs show representative analysis of one experiment. Tables demonstrate average and standard deviation of three independent experiments. White graph: EpCAM^{hi}; gray graph: EpCAM^{lo}. Asterisks show the significant ($p<0.05$) differences between EpCAM^{hi} and EpCAM^{lo} cells in G_1 and G_2M-phases.

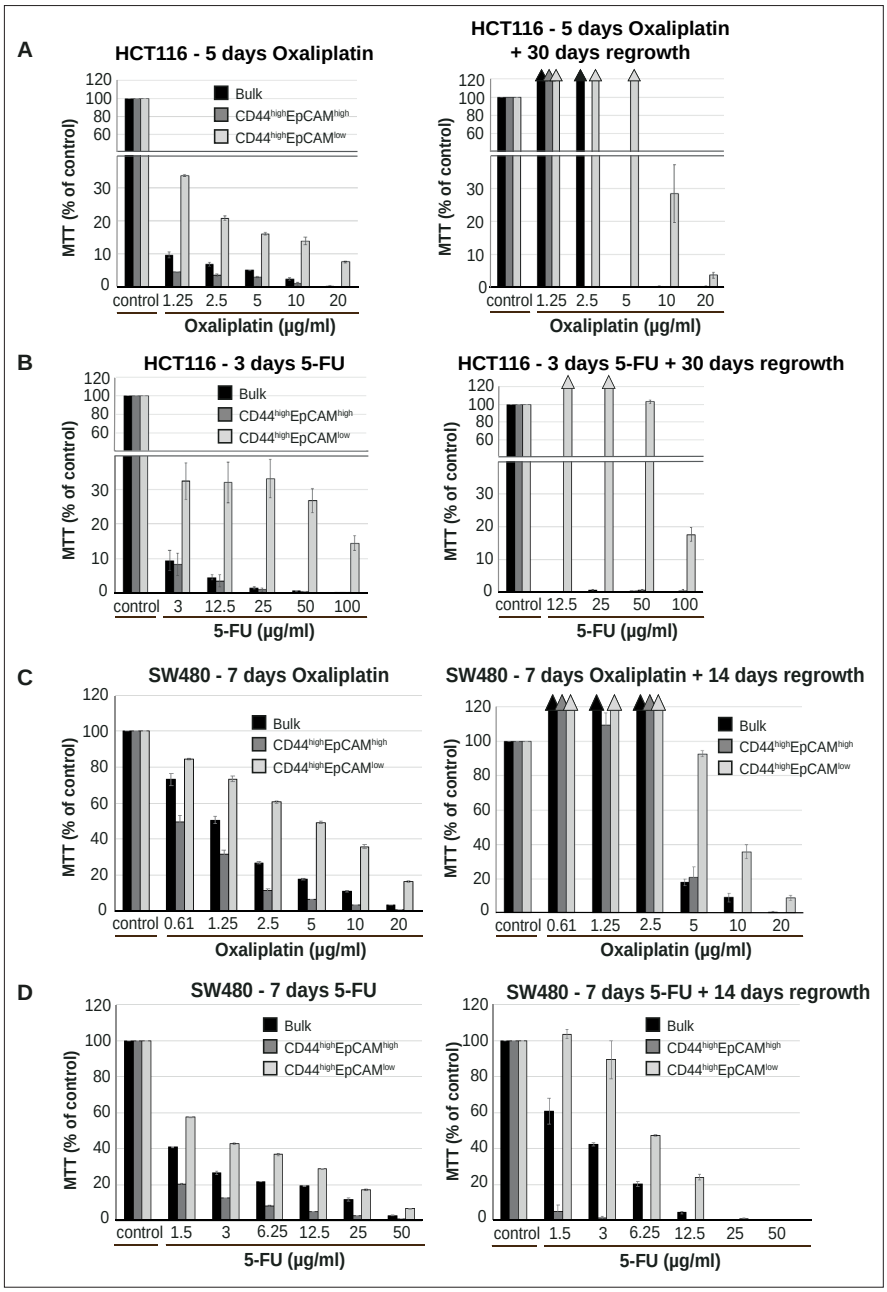


Figure 1—figure supplement 4. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: chemoresistance. Bulk, EpCAM^{hi}, and EpCAM^{lo} HCT116 and SW480 cells were sorted and plated to recover and expand for 4 days.10^5 cells per well were plated in triplicate in a 96-well plate and left to attach. After 24 hr, standard cell culture medium was changed for medium containing chemotherapeutic drug oxaliplatin or 5-fluorouracil (5-FU). HCT116 cells were left with 5-FU and oxaliplatin for respectively 3 and 5 days. For each drug, SW480 cells were treated for 7 days. After removal of

Figure 1—figure supplement 4 continued on next page

Figure 1—figure supplement 4 continued

the chemotherapeutic drug, cells were washed with PBS and left to regrow in standard culture medium. Cell viability was assessed using the MTT upon removal of the drug, and after regrowth (ranging 13–30 days). O.D. reading was performed at 595 nm with a Microplate Reader. Background measurements were subtracted from each data point. All experiments were performed at least in duplicate for each individual cell line and drug. (**A, B**) Chemo-sensitivity assays using oxaliplatin (**A**) or 5-FU (**B**) in bulk (black bars), EpCAM^{hi} (dark gray bars) and EpCAM^{lo} (light gray bars) HCT116 cells. Bars represent the average ± SD of the O.D. corrected for the background measurements. (**C, D**) Chemo-sensitivity assays using oxaliplatin (**C**) or 5-FU (**D**) in bulk (black bars), EpCAM^{hi} (dark gray bars) and EpCAM^{lo} (light gray bars) SW480 cells. Bars represent the average ± SD of the O.D. corrected for the background measurements.

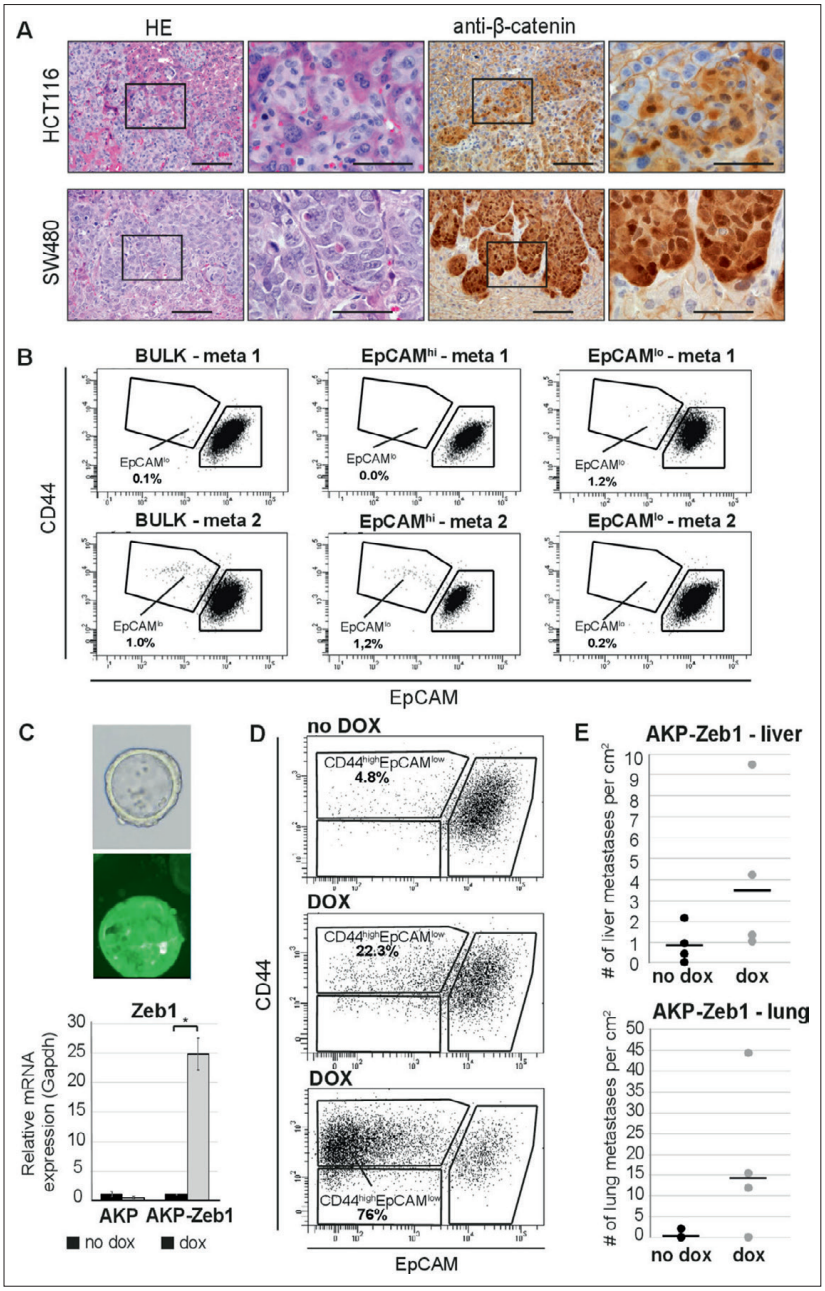


Figure 1—figure supplement 5. Further characterization of EpCAM^{lo} cells in colon cancer cell lines: invasive and metastatic behaviour. (**A**) Hematoxylin and eosin (H&E; first two panels) and immunohistochemistry (IHC) with antibody directed against β-catenin (third and fourth panels) in liver metastasis obtained 4 or 8 weeks after intrasplenic injection with HCT116 (upper panels) and SW480 (lower panels) cells, respectively. Second and fourth

Figure 1—figure supplement 5 continued on next page

Figure 1—figure supplement 5 continued

panels show zoom of the marked area in respectively first and third panels. Scale bar first and third panels: 100 μm. Scale bar second and fourth panels: 50 μm. (**B**) FACS analysis of liver metastases obtained by spleen injection of HCT116 bulk, EpCAM^{hi}, and EpCAM^{lo} cells. Upper panel: representative FACS plots. (**C**) Microscopic (top, left panel) analysis of AKP-Z organoids tagged with GFP and click beetle luciferase. The graph (bottom, left panel) shows the RT-qPCR expression analysis of *Zeb1* in AKP (left) and AKP-Z (right) organoids upon in vitro doxycycline treatment for 48 hr. Black bars: no doxycycline treatment; gray bars: 1 μg/mL doxycycline. Each bar represents the mean ± SD of three independent experiments. (**D**) Upon establishment of a primary tumor in the caecum, transplanted mice were administered doxycycline in the drinking water to induce Zeb1 expression. FACS analysis of the primary tumor was performed 1 week after the start of the doxycycline. The panels show representative FACS plots of a control and two dox-treated primary tumors. (**E**) Quantification of the number of lung and liver metastases in uninduced (black; n = 4) and dox-induced (gray; n = 5) AKP-Z transplanted mice. Liver tissue was cut into 500 μm slices, processed for IHC, stained for β-catenin to visualize tumor cells, scanned using a NanoZoomer, and counted using NDP view software. The area of tissue analyzed was used to normalize the data.

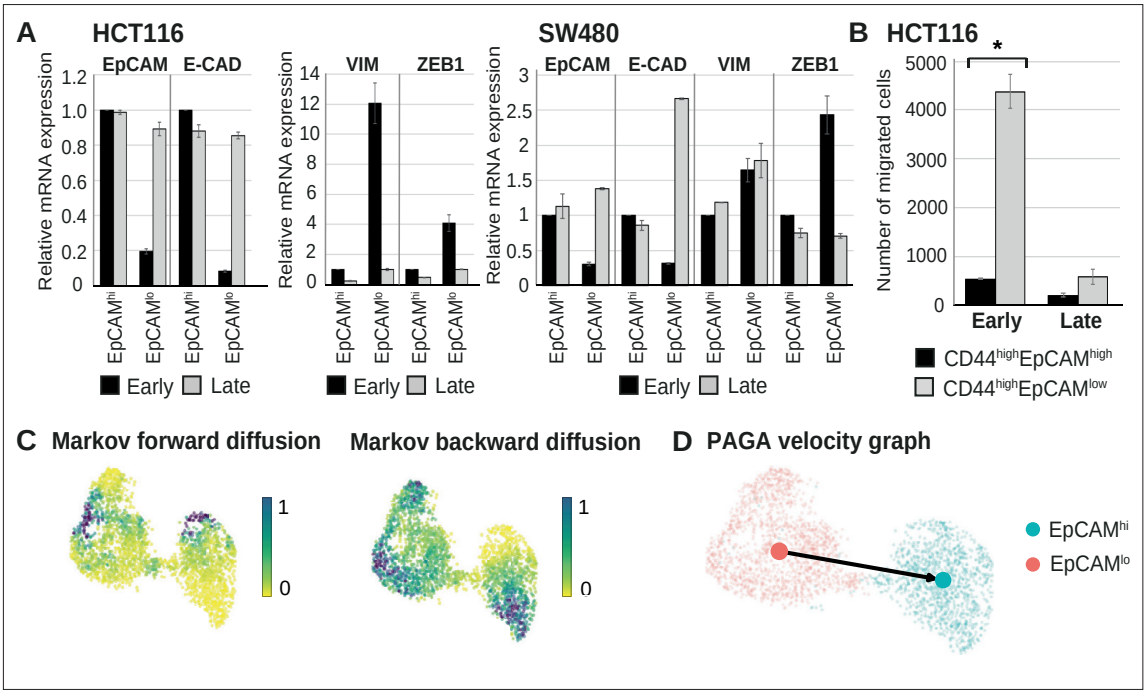


Figure 2—figure supplement 1. Further characterization of phenotypic plasticity in EpCAM^{lo} and EpCAM^{hi} cells. (**A**) RT-qPCR expression analysis of *EPCAM*, *CDH1*, *VIM*, and *ZEB1* in 'early' (e.g., cells that were used shortly after FACS sorting) and 'late' (e.g., cells that were cultured for an extended period of time before performing the experiment) sorted EpCAM^{hi} and EpCAM^{lo} cell from HCT116 and SW480 cell lines. *GAPDH* was employed for normalization purposes. Each bar represents the mean ± SD of two independent experiments. (**B**) Transwell migration assay of 'early' and 'late' EpCAM^{hi} (black bar) and EpCAM^{lo} (gray bar) cultures in HCT116. 10^5 cells were plated in triplicate on TC-coated membrane and left overnight before counting the number of migrated cells on the bottom side of the membrane. Each bar represents the mean ± SD of two independent experiments. Asterisks indicate significant differences ($p<0.05$). (**C**) Analysis of the HCT116 scRNAseq data as a Markov diffusion process. Markov forward (left) and backward (right) diffusion indicating the presence of sink and source points in both EpCAM^{hi} and EpCAM^{lo} populations. (**D**) Partition-based graph abstraction velocity graph mapping out the direction of velocity on a subpopulation level in HCT116.

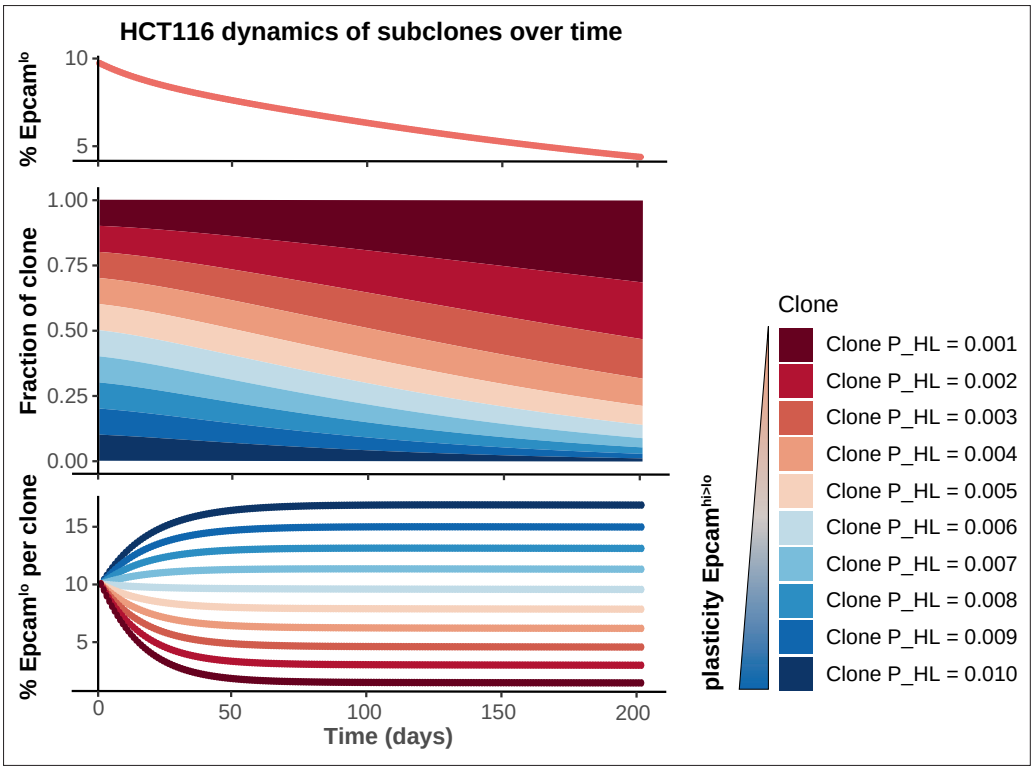


Figure 2—figure supplement 2. Simulation of the HCT116 two-state Markov model with a non-homogenous starting population. The starting population was defined as 10 subclones with varying transition probability P_HL at equal proportion. The Markov model was run till day 200 and shows that the total EpCAM^{lo} population decreases over time (top panel) because clones with lower plasticity will gain dominance in the long run (lower panels).

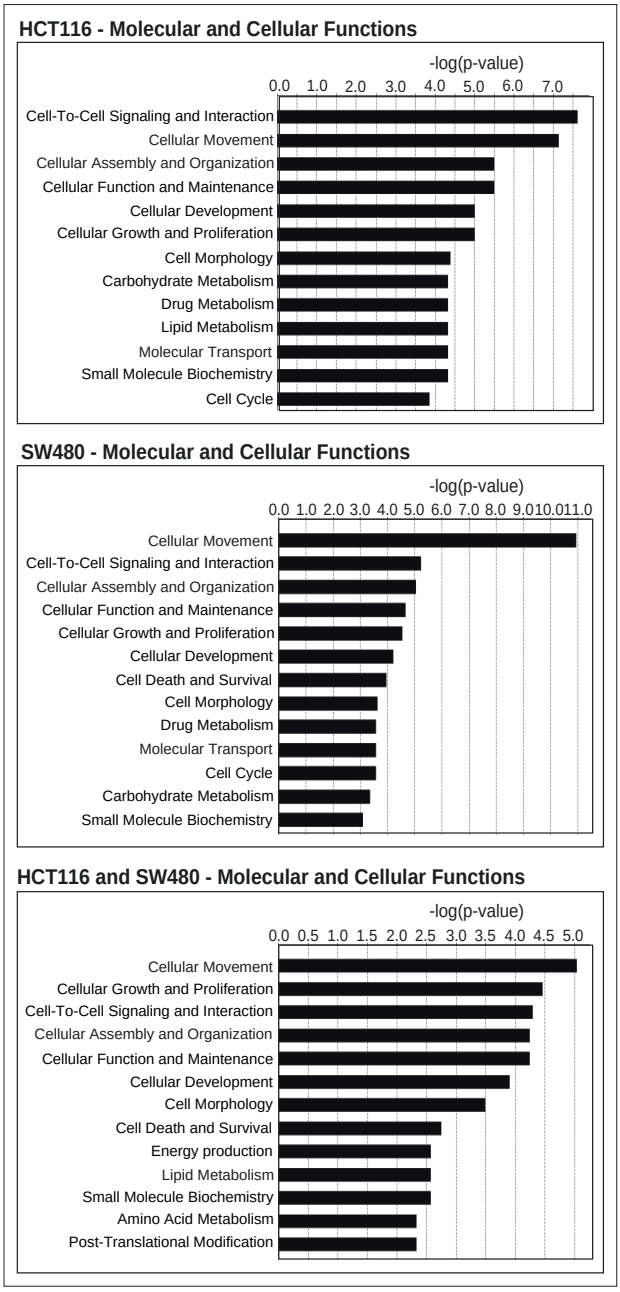


Figure 3—figure supplement 1. Gene ontology (GO) of molecular and cellular functions in HCT116 (upper panel), SW480 (middle), and the combined (bottom panel) gene lists.

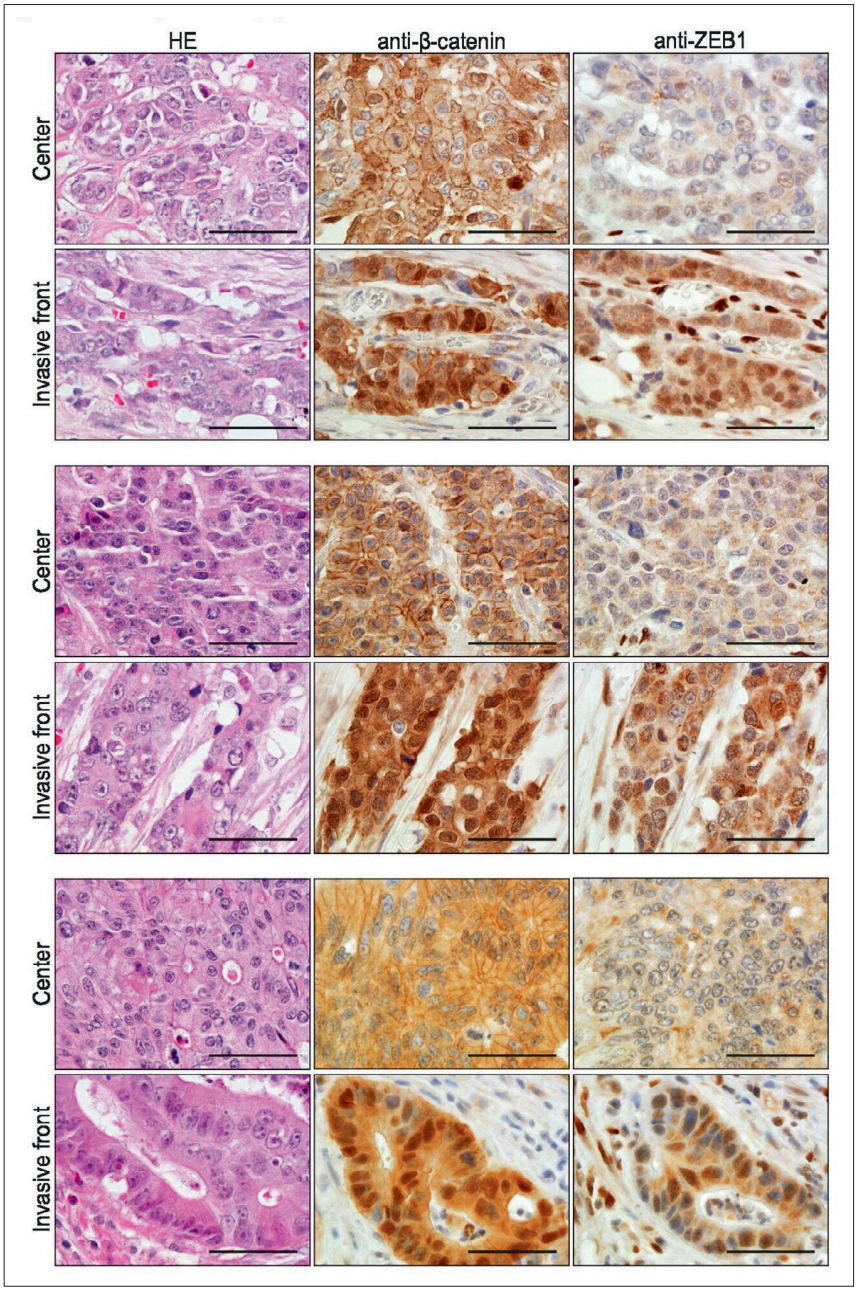


Figure 3—figure supplement 2. Hematoxylin and eosin (H&E) and immunohistochemistry (IHC) analyses with antibodies directed against beta-catenin and ZEB1 in consecutive sections of colon cancers from three unrelated patients with sporadic colon cancer. For each case, areas from the tumor center and the invasive front are shown. Co-localization of nuclear β-catenin and ZEB1 expression was found in 5 out of 25 cases investigated. Scale bar: 50 μm.

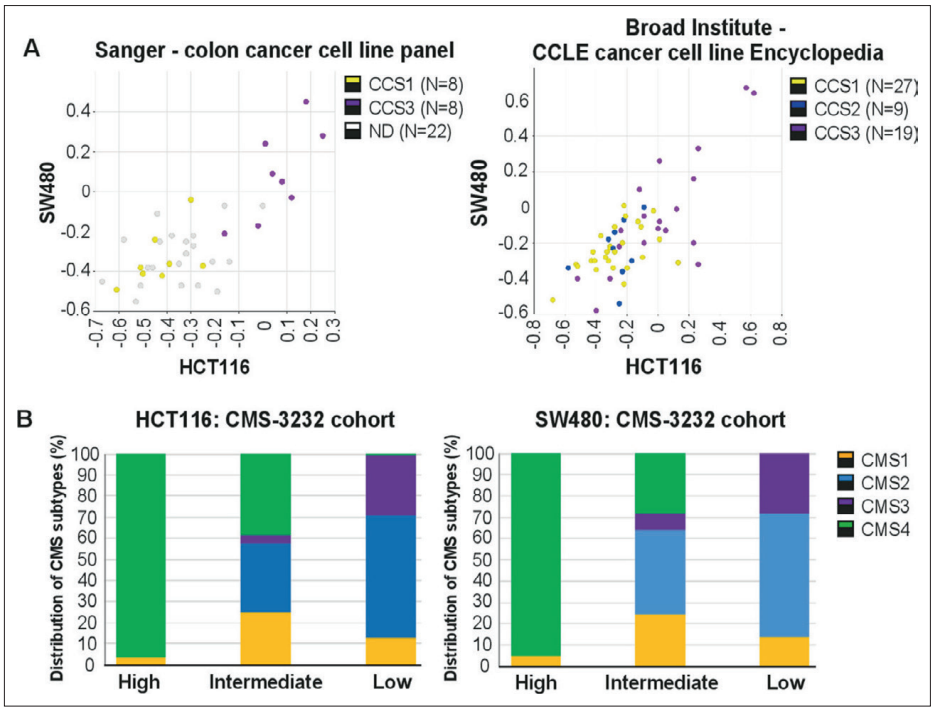


Figure 4—figure supplement 1. EpCAM^{lo} expression signatures correlate with CCS3 cell lines and CMS4 colon cancer patients. (**A**) Correlation of meta-gene expression values of the signatures derived from EpCAM^{lo} HCT116 and SW480 cells in two cohorts of colon cancer cell lines in the absence of a stromal component. CCS1: epithelial-like tumor lines; CCS2: tumor lines with microsatellite instability. CCS3: mesenchymal-like tumor cell lines (***De Sousa et al., 2013***). (**B**) The gene sets identifying the EpCAM^{lo} cells from both HCT116 and SW480 were used to cluster the tumors in the consensus molecularsubtype (CMS)3232 cohort into high, intermediate, and low expression groups by k-means clustering. All tumors in this cohort had previously been assigned to specific CMS subgroups, and then analyzed the contribution of each CMS subtype to each of the generated high, intermediate, and low expression subgroups.

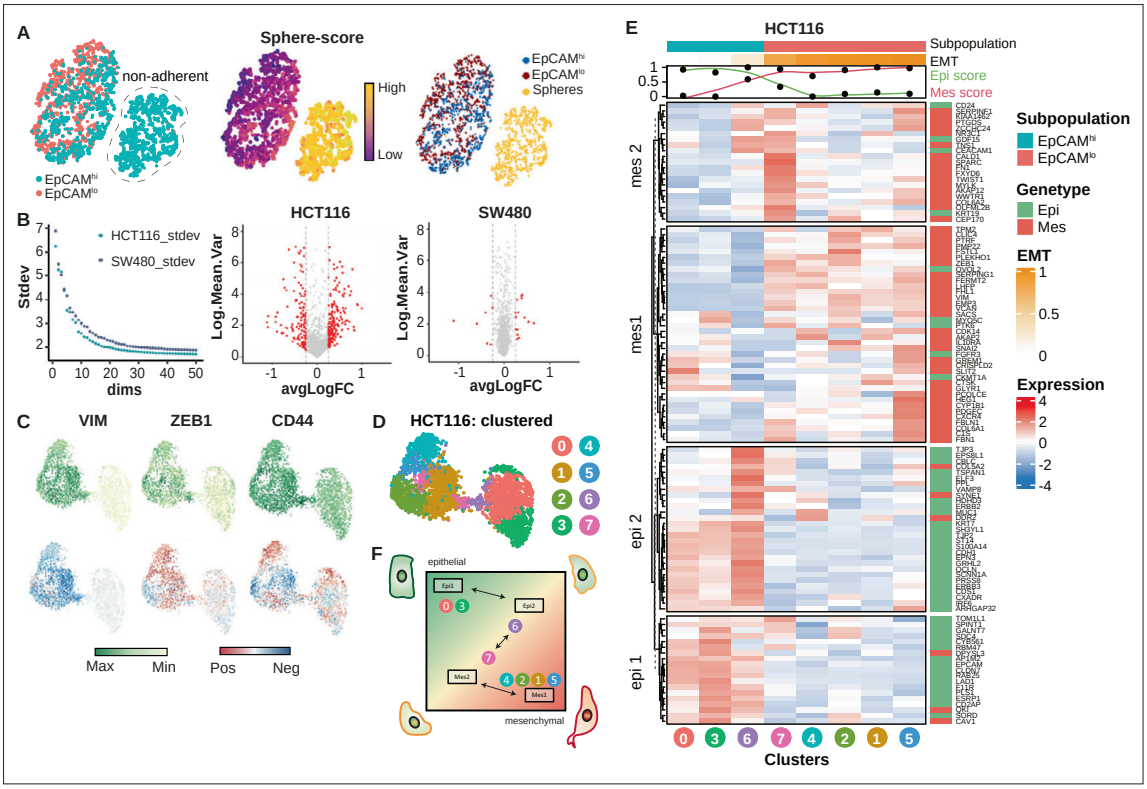


Figure 5—figure supplement 1. Further scRNAseq analysis of EpCAM^{lo} cells: EMT/MET transcriptional trajectory. (**A**) tSNE of the SW480 cell line indicating an additional subpopulation in the EpCAM^{hi} population (left panel). Using a signature derived from the bulk RNAseq, this population was identified as the 'sphere' population (middle panel) and annotated to be excluded for further analysis (right panel). (**B**) Left panel: the SW480 cell line, after exclusion of the 'sphere' population, contains slightly higher variability compared to the HCT116 cell line, as evidenced by the variance of the top 50 principal components. Right panel: while in HCT116 most of the variable expressed genes are differentially expressed between the EpCAM^{hi} and EpCAM^{lo} population, this is not the case in SW480, where most of the highly variable genes do not differ between the two populations. (**C**) Top panels: expression values of *VIM*, *ZEB1*, and *CD44* on the UMAP embedding of the HCT116 cell line. Lower panels: projection of the RNA velocity direction of the same genes. (**D**) HCT116 UMAP embedding annotated with the eight unsupervised clusters. (**E**) Heatmap of HCT116 with expression values of the epithelial to mesenchymal transition (EMT) signature averaged by the eight clusters. Clusters were ranked according to their EMT score, and genes were clustered in four distinct gene sets using k-means clustering. (**F**) Schematic diagram showing a transcriptional trajectory with distinct gene arrays through which pEMT cells arise.

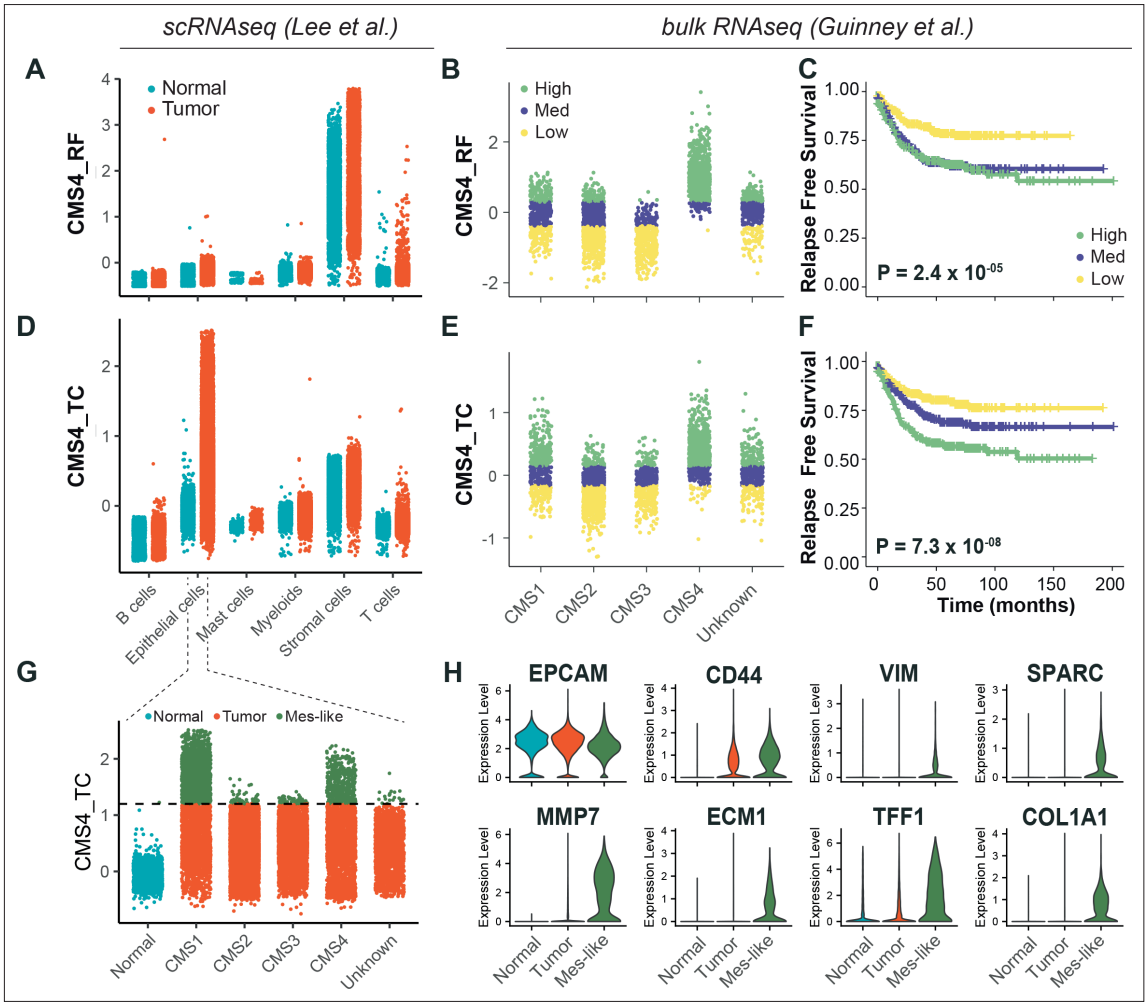


Figure 6. Identification of EpCAM^{lo} cells in primary colorectal tumors. (**A**) Expression of the CMS4_RF signature in the scRNAseq data from Lee et al. (N = 91,103 cells) indicates high association to the stromal cells. (**B**) Expression of the CMS4_RF signature in the bulk RNAseq data from Guinney et al. (N = 3232 tumors) shows association to CMS4 tumors. Tumors were grouped in three equal groups according to their association with the CMS4_RF signature. (**C**) Kaplan–Meier plot of the three CMS4_RF groups shows significant differences in relapse-free survival. (**D**) Expression of the CMS4_TC signature in the scRNAseq data reveals high association to the tumor epithelial cells. (**E**) Expression of the CMS4_TC signature in the bulk RNAseq data shows association to CMS1 and CMS4 tumors. Tumors were grouped in three equal groups according to their association with the CMS4_TC signature. (**F**) Kaplan–Meier plot of the three CMS4_TC groups shows significant differences in relapse-free survival. (**G**) Expression of the CMS4_TC signature in the fraction of epithelial cells (N = 24,707 cells). Cells from CMS1 and CMS4 tumors show the highest association to the CMS4_TC signature and were annotated as mes-like. (**H**) Violin plots of normal, tumor, and mes-like tumor cells showing expression patterns across different genes.

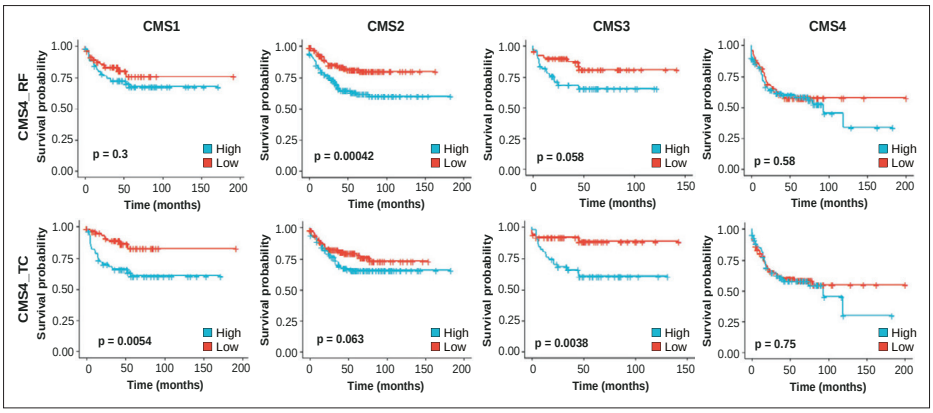


Figure 6—figure supplement 1. Survival analysis using the CMS4_RF and CMS4_TC signatures across the different consensus molecular subtypes. Patients were divided into two equal groups based on their association to the signatures and Kaplan–Meier plots were made based on relapse-free survival.

CHAPTER 3

The SW480 cell line as a model of resident and migrating colon cancer stem cells

Manuscript submitted to iScience.

The SW480 cell line as a model of resident and migrating colon cancer stem cells

Mathijs P. Verhagen, Tong Xu[#], Roberto Stabile[#], Rosalie Joosten, Francesco A. Tucci[$], Marco Trerotola[@], Saverio Alberti[%], Andrea Sacchetti*, and Riccardo Fodde*

Department of Pathology, Erasmus University Medical Center, Rotterdam, The Netherlands.

[#*]shared contributions

[$]Present address: European Institute of Oncology IRCCS, and School of Pathology, University of Milan, Milan, Italy.

[@]Department of Medical, Oral and Biotechnological Sciences, University of Chieti-Pescara, Italy.

[%]Department of Biomedical Sciences, University of Messina, Italy.

*to whom correspondence should be addressed at:

Erasmus MC, Department of Pathology, PO Box 2040, 3000 CA Rotterdam, The Netherlands

https://labriccardofodde.nl

e-mail: r.fodde@erasmusmc.nl; a.sacchetti@erasmusmc.nl

tel.: +31 10 7043896

Abstract

Intra-tumor heterogeneity, i.e. the presence of diverse cell types and subpopulations within tumors, presents a significant obstacle in cancer treatment due to its negative consequences for resistance to therapy and disease recurrence. However, the mechanisms which underlie intra-tumor heterogeneity and result in the plethora of different cancer cells within a single lesion remain poorly understood. Here, we leverage the SW480 cell line as a model system to investigate the molecular and functional diversity of colon cancer cells. Through a combination of fluorescence activated cell sorting (FACS) analysis and transcriptomic profiling, we identify three distinct subpopulations, namely resident cancer stem cells (rCSCs), migratory CSCs (mCSCs), and high-relapse cells (HRCs). These subpopulations show varying Wnt signaling levels and gene expression profiles mirroring their stem-like properties. Examination of publicly available spatial transcriptomic data confirms the presence of these subpopulations in patient-derived cancers and reveals their distinct spatial distribution relative to the tumor microenvironment.

Introduction

Intra-tumor heterogeneity is one of the main determinants of resistance to therapy and relapse[1]. Hence, understanding the mechanisms that drive heterogeneity is crucial for our understanding of cancer. In this regard, phenotypic plasticity is increasingly recognized as the main driver of intra-tumor heterogeneity[2,3] and as a major determinant of resistance to therapy, as tumor cells exploit plasticity as mechanism to adapt to changes in the environment.

In colon cancer, several studies have reported on the identification of cancer stem cells (CSC) earmarked by specific membrane-bound antigens such as *CD133*, *CD44*, and *CD166*, used to enrich for cells with tumor-propagating capacity in limiting dilution transplantation assays[4,5-8]. These experimental approaches mainly rely on self-renewal and pluripotency rather than on their metastatic potential. Thomas Brabletz and collaborators originally postulated on the existence of a distinct CSC type, named migratory CSCs (mCSCs)[9], usually characterized by the expression of markers such as *ZEB1* and *RUNX2*[10] and/or the reduced expression of epithelial markers E-cadherin (*CDH1*)[11] and EpCAM[12], associated with epithelial-to-mesenchymal transitions and plasticity (EMT/EMP)[13]. More recently, the *EMP1* (epithelial membrane protein 1) gene was identified as a marker for high-relapse cells (HRCs) that efficiently metastasize and correlate with poor patient outcome[14]. Hence, colon cancer cells with metastatic capacity appear to exhibit a distinct set of markers compared to the more conventional CSCs endowed with tumor-fueling capacity[9].

Although immortalized cancer cell lines have represented a powerful tool in cancer research, they are usually not considered equally useful as a source of CSCs. However, in 2011 Gupta and collaborators demonstrated how, even within cell lines, phenotypically distinct subpopulations of cancer cells can be found which are kept in balance through stochastic state transitions[15]. As such, they still represent a valuable model to study phenotypic plasticity, as also shown by our previous characterization of quasi-mesenchymal mCSCs ($CD44^{hi}$/$EpCAM^{lo}$) in colon and ovarian cancer cell lines together with their validation in patient-derived samples[12,16,17].

Here, we employed the SW480 cell line as a model system to study an additional subpopulation of resident colon cancer stem cells, (rCSCs, here referred to as 'spheres') thought to fuel primary tumor growth because of their enhanced self-renewal and tumor-propagating capacity[18-20]. These partially adhesive, sphere-like cells are distinct from the bulk of the adherent ($EpCAM^{hi}$) SW480 cells. By integrating data from the analysis of the SW480 sphere subpopulation and patient-derived spatial transcriptomic data, we now show that colon cancers exhibit a mixture of distinct stem-like subpopulations located within discrete and specific niches.

Results

SW480 encompasses three morphologically distinct subpopulations with discrete Wnt levels

We previously characterized the CD44^{high}EpCAM^{low} (EpCAM^{lo}) and CD44^{high}EpCAM^{high} (EpCAM^{hi}) subpopulations in the HCT116 and SW480 colon cancer cell lines[12]. While doing so, we identified an additional subpopulation of EpCAM^{hi} cells in the SW480 cell line (depicted in green in Figure 1A), though not in HCT116, with a broad range (high-to-low) of CD44 expression values. Previously, these cells have been referred to as "spheres" due to their characteristic morphology, scarce adhesion to the plates, and ability to grow as tumor-spheroids in culture[10,18-20]. Attempts to enrich the SW480 spheres have been implemented in various ways, ranging from the collection of floating, non-adhesive cells in culture dishes[18,21], or by sorting based either on their relatively low level of Wnt signaling upon transfection with fluorescent reporters[10], or on the increased levels of CSC markers such as CD133 (*PROM1*)[19]. Despite these efforts, a comprehensive molecular and phenotypic profile of the SW480 sphere subpopulation to allow its identification and isolation by fluorescence-activated cell sorting (FACS) is still lacking.

To this aim, we first employed a set of antibodies directed against surface (cancer) stem cell markers and adhesion molecules in combination with EpCAM and CD44. As shown in Figure 1A, the addition of an antibody against TROP2 (trophoblast cell surface antigen 2; also known as TACSTD2 or tumor associated calcium signal transducer 2) contributed to resolve SW480 cells into three more distinct subpopulations, namely the spheres (EpCAM^{hi}TROP2^{neg}) and the adherent bulk, the latter to be further be subdivided in EpCAM^{hi} (CD44^{hi}EpCAM^{hi}TROP2^{pos}) and EpCAM^{lo} (CD44^{hi}EpCAM^{lo}TROP2^{pos}) cells (Figure 1B, Supplementary Figure 1A-C). In agreement with Wang et al.[19], CD133 expression was rarely observed in adherent cells (0.5 and 0.8% in EpCAM^{lo} and EpCAM^{hi}, respectively), though it was clearly present among sphere cells (10.1%)(Supplementary Figure 1D). Subsequent sorting of the three subpopulations by FACS revealed distinct morphologies, with EpCAM^{hi} cells growing in cobblestone-like colonies, EpCAM^{lo} cells as spindle-like cells, and spheres in multi-layered, compact spheroid-structures (Figure 1C).

To confirm that the SW480 subpopulations have distinct functional identities, we employed a lentiviral Wnt reporter based on the 7xTcf-eGFP cassette, that also ensures stable mCherry expression to distinguish infected from non-infected cells[22]. In accordance with the Yi et al. study[10], sphere cells displayed lower Wnt signaling activity when compared to adherent cells, as measured by the GFP/mCherry signal ratio (Figure 1D). Among the adherent bulk, EpCAM^{lo} cells showed the highest level of Wnt signaling activity, as also confirmed by immunofluorescence (Figure 1E) and by TOP-flash reporter assay (Figure 1F). The latter confirms that, as predicted by the loss of APC function characteristic of the SW480 cell line, all three subpopulations are Wnt-ON, albeit at variable quantitative levels. Hence,

accurate isolation of three subpopulations in SW480 confirms their distinct morphologies and Wnt signaling levels.

The three subpopulations in SW480 represent distinct transcriptional states

In order to characterize the expression profiles of the distinct SW480 colon cancer cell subpopulations, we sorted them by FACS using the aforementioned established conditions and analyzed their transcriptomes by RNA sequencing. Principal component analysis (PCA) revealed differences between spheres and adherent in the first principal component (accounting for 89% of the variance), while differences between the adherent EpCAM$^{hi/lo}$ subpopulations became notable in the second principal component (6% variance)(Figure 2A). We then performed several rounds of differential expression analyses by comparing each subpopulation individually to the other subpopulations, which resulted in a total of N = 3963 differentially expressed genes. Clustering of the differentially expressed genes revealed gene sets that were either expressed in two subpopulations (cluster 4, 5 in Figure 2B), or specific to one of the subpopulations (cluster 1, spheres; cluster 2, EpCAM^{lo}; and cluster 3, EpCAM^{hi}; Figure 2B, and Supplementary Table 1). Closer examination of the gene sets revealed typical (cancer) stem cell markers in cluster 1 (spheres) such as *MET*, *PROM1* (CD133), *LGR5* and *MYC*. The EpCAM$^{hi/lo}$ adherent subpopulations shared gene expression patterns (cluster 5), but differed in the degree of EMT (e.g. illustrated by *ZEB1*, *SPARC* and *MMP7* in EpCAM^{lo} cluster 2). EpCAM^{hi}-specific genes (cluster 3) included *CEACAM1*, *EMP1*, *LAMC2* and *TSPAN1*, previously reported as HRC markers[14]. Evaluation of the signatures characteristic for these identities confirmed their association with the distinct subpopulations (Figure 2C, Supplementary Table 2); additional qPCR analyses validated differences in key markers discriminating the populations (Figure 2D). Overall, it appears that the 3 SW480 subpopulations reflect distinct aspects of the cancer stem cell phenotype: the *Lgr5*$^{+}$ resident CSCs represented by the spheres, the EpCAM^{lo} migratory CSCs, and the EpCAM^{hi} HRCs. Of note, although constitutive Wnt signaling activation earmarks all 3 subpopulations, its enhanced levels in EpCAM^{lo} cells underlie EMT and the acquisition of migratory and invasive features. In this regard, whether the two metastatic subpopulations, EpCAM^{lo} and HRCs, represent distinct routes to the colonization of distant organs or are connected through stochastic state transitions, is unclear.

Phenotypic plasticity and epigenetic barriers across the colon cancer subpopulations

Since the three subpopulations appear to be maintained in phenotypic equilibrium in the SW480 cell line, we subsequently questioned their individual plasticity and dynamics. As reported in our previous study, when sorted by FACS and subsequently cultured, EpCAM$^{hi/lo}$ subpopulations restore the original

homeostatic equilibrium after several passages[12]. The same approach was here employed to investigate whether similar scenario applies to the spheres and their capacity to transit towards more adherent subpopulations. As shown in Figure 3A, spheres were sorted and cultured for up to 10 weeks during which they retained their original cellular identity. Likewise, the adherent fraction did not give rise to sphere-like cells after prolonged culture. Hence, plasticity between SW480 adherent and sphere cells appears to be very limited under conventional culture conditions.

Next, we tested whether sphere cells can be converted by modulating the expressing of specific transcription factors known to be differentially expressed among the three subpopulations. To this aim, we first focused on RUNX2, previously shown to increase the metastatic properties of the SW480 cell line by enhancing Wnt signaling and EMT[10]. First, we validated by RT-qPCR analysis the differential *RUNX2* expression between spheres and adherent subpopulations in the parental SW480 cell line (Figure 3B). As shown in Figure 3C, upon *RUNX2* ectopic expression, *CD44* levels dramatically increased in the spheres while a decrease in *EPCAM* is observed both in EpCAM^{hi} and sphere cells, as also confirmed by FACS (Figure 3D, Supplementary Figure 3). Of note, a significant increase in *TROP2* expression was also seen in the spheres. Of note, the sphere-specific increase in CD44 and decrease in EpCAM is indicative of a partial shift towards the adherent subpopulations (Figure 3D,E).

The RNAseq analysis of the SW480 cell subpopulations revealed a noticeable increase in *ARID1A* (AT-rich interactive domain-containing protein 1A) expression in the spheres (Figure 4A). This gene encodes for a subunit of the SWI/SNF chromatin remodeling complex. A more detailed analysis of the expression of all known SWI/SNF subunits suggests that, of the three broad SWI/SNF subfamilies namely canonical BAF (cBAF), polybromo-associated BAF (PBAF), and the GLTSCR1- or GLTSCR1L- and BRD9-containing (GBAF) complexes, cBAF appears to be upregulated in the sphere subpopulation (Supplementary Figure 2). In light of these observations, we set to experimentally evaluate the effect of specific SWI/SNF perturbations on the spheres by culturing them in the presence of the ARID1A inhibitor inhibitor BRD-K98645985 (from here on referred to as ED98) [23]. As shown in Figure 4B, ARID1A inhibition induced a clear morphological change resulting in the appearance of spindle-like adherent cells. RT-qPCR analysis of the ED98-treated spheres revealed a specific decrease in *LGR5* expression and the increase of the EMT marker *VIM* (Vimentin), whereas EpCAM did not show any significant variation (Figure 4C).

Overall, these results indicate that inhibition of ARID1A is sufficient to trigger the partial conversion from spheres to adherent cells. Therefore, whereas the EpCAM$^{hi\leftrightarrow lo}$ conversion seems to occur via stochastic state transitions even under normal culture conditions, the sphere-to-adherent state transition appears to be safeguarded by the SWI/SNF (cBAF) complex.

Identification and spatial organization of the subpopulations in colon cancers

To assess and validate the presence and relevance of the SW480 subpopulations in patient-derived malignancies, we interrogated a panel of spatial transcriptomic studies on colon cancers and liver metastases[24-26]. We first compiled the data from the different studies resulting in a total set of 66,281 spots across 23 samples and 15 different patients (Supplementary Figure 4A-B, Supplementary Table 3). Dimension reduction and clustering resolved tumor areas from transcriptionally distinct cell types belonging to the microenvironment (Figure 5A). Subsequent sub-clustering of the tumor-specific areas revealed five tumor entities with distinct expression profiles (Figure 5B). Out of these, three showed expression profiles reminiscent of the SW480 subpopulations (Figure 5C), namely the spheres, reminiscent of resident CSCs, earmarked by expression of genes such as *ASCL2*, *LGR5*, and *MYC*; the quasi-mesenchymal EpCAM^{lo} (migrating CSCs), characterized by EMT-related markers (i.e. *VIM, SPARC, ZEB1)*; and the EpCAM^{hi} high-relapse cells (HRC), displaying expression of markers such as *EMP1, MAL2 and TACSTD2*. The remaining tumor entities expressed markers of secretory cells (e.g. *MUC2, SPINK4, TFF2*) and those usually upregulated in inflammatory bowel disease (e.g. *CCL20, LCN2, CD24*), respectively, and were accordingly labeled as “secretory” and “inflammatory” (Figure 5D). Looking at the individual tumors, the relative proportion and representation of the distinct entities appear to be variable (Figure 5E), the rCSC/sphere-like fraction being the most abundant and stable across the samples (35.9±14.7%), followed by the EMT segment (28.6 ± 15.8%). The HRC subpopulation made up 10.9±8.3% of the tumors, while the secretory and inflammatory areas contributed to 16.6±25.4% and 8.1±7.9%, respectively. No clear differences in abundance were observed between the primary and metastatic samples although, admittedly, the sample size was rather limited to discern differences between the lesions.

Last, in order to investigate the spatial organization of the different cell clusters, we first grouped the data into micro-neighborhoods (see Methods). We next clustered the neighborhoods based on their transcriptional profile (Figure 6A, Supplementary Figure 4C-D). Using this approach, neighborhoods could be classified into different types of cancer niches, and labeled based on their composition of tumor- and stromal cells (Figure 6B). Of note, from the different tumor areas, the EpCAM^{lo}-like tumor cells were most frequently observed in the tumor front (40.5% of its area). The other HRCs- and sphere-like tumor areas appeared in patches throughout the tumors. Interestingly, the sphere-like cells spatially connected distinct tumor entities, suggestive of their plastic potential (Figure 6C).

Discussion

In this study, we outline a molecular definition by which the SW480 cell line can be dissected into three separate entities with distinct morphologies and transcriptomic states. Notably, the so-called 'spheres' express a set of Wnt-related genes (e.g. *LGR5, MYC, ASCL2*) reminiscent of the profile of the Wnt-ON intestinal stem cells in homeostasis. Somewhat counterintuitively, the Wnt signaling levels characteristic of the spheres was found to be lower when compared to the adherent, and in particular to the EpCAM^{lo} subpopulations. This observation illustrates the complex, dosage-dependent relationship between Wnt activity and its downstream effects[27], and underlines the role of enhanced Wnt signaling in the activation of epithelial-mesenchymal plasticity[10,12]. The consequences of *RUNX2* overexpression in SW480 cells, i.e. increased Wnt signaling and EMT activation, underlie the partial transition from spheres into more adherent-like states.

Under normal culture conditions, the sphere 'state' appears to be epigenetically fixed which prevents it from transitioning into adherent cells. This is reminiscent of intestinal crypt homeostasis where the *Lgr5*$^{+}$ ISCs divide symmetrically and stochastically adopt stem or transient-amplifying fates following patterns of neutral drift dynamics[28]. Our results suggest that *in vitro*, the sphere phenotype is stabilized by the expression of specific SWI/SNF subunits such as ARID1A and of transcription factors (e.g. PROX1) previously shown to reinforce a ISC-like state[21,29]. Of note, *ARID1A* acts as a tumor suppressor gene in colon cancer and loss-of-function mutations are found in approx. 10% of the cases, often in association with the MSI-high status[30]. Apart from somatic mutations, silencing by DNA promoter methylation was also found to be common among sporadic colon cancers[31]. Accordingly, ARID1A inhibition by ED98 results in the partial transition from spheres to a more adherent-like morphology and expression profile.

Next, the transition from the adherent EpCAM^{hi} state into the quasi-mesenchymal EpCAM^{lo} cells is likely to be underlined by enhanced Wnt signaling and the activation of other EMT-inducing pathways. From this perspective, it is of interest to observe that both adherent SW480 subpopulations, namely the EpCAM^{hi} high-relapse cells (HRCs) and the EMT-competent EpCAM^{lo} cells, have been described as highly metastatic[12,14]. This raises the question on whether they denote different steps along the invasion-metastasis cascade, or represent two distinct metastatic routes. Additional studies are needed to clarify these issues in the near future. The present results are suggestive of a two-step sequence where HRCs may derive from resident CSCs (spheres) due to (epi)genetic perturbations of the SWI/SNF chromatin remodeling complex. Additional somatic alterations in genes leading to increase in Wnt signaling such as *RUNX2* or *PROX1* will then underlie epithelial-mesenchymal plasticity and the acquisition of motile and invasive capacity.

Admittedly, our study is centered on a single immortalized colon cancer cell line which limits its translational applicability. Nonetheless, the fact that cancer cell lines encompass distinct stem-like identities makes them an attractive *in vitro* model to study cell-autonomous mechanisms, likely to be governed *in vivo* by tumor-TME interactions, thought to be the main drivers of heterogeneity and plasticity. As shown here, interrogation of colon cancer spatial transcriptomics data from recent studies[24-26] has revealed how the expression profile signatures of the SW480 subpopulations mirror distinct cellular states in clinical specimens. While these analyses are admittedly correlative, they reveal a number of cellular states recurrently observed with distinct spatial organization in patient-derived colon cancers. Whereas sphere- and EpCAM^{hi}-like tumor areas appeared in patches, EpCAM^{lo}-like tumor cells exhibited a more scattered pattern along the tumor margins.

Taken together, our study contributes to our understanding of cellular plasticity in colon cancer. The SW480 cell line displays an admixture of distinct cellular morphologies kept in equilibrium through stochastic state transitions driven by cell-autonomous mechanisms. In clinical specimens, the same mechanisms are regulated by interactions with the tumor microenvironment and with intra-tumor niches, as also reflected by the spatial organization of the distinct stem-like cell identities. In the near future, single-cell spatial profiling techniques will deepen our understanding of phenotypic plasticity in colon cancer and its relevance for prognosis and response to therapy.

Limitations of the study

Clearly, the main limitation of this study lies on the fact that it is entirely based on a single, immortalized colon cancer cell line, namely SW480. Indeed, the presence of at least 3 distinct subpopulations of colon cancer cells has been the main reason why this line was selected. Also, in a previous publication[1], we already have shown how one of these SW480 subpopulations, the quasi-mesenchymal EpCAM^{hi}CD44^{lo}, represents the phenotypic plastic cells that underlie local invasion and distant metastasis in colon cancer. As such, it appears that the SW480 molecular signatures are reminiscent of clinical colon cancer specimens.

In the present study, we extended our analysis to the so-called sphere cells, a likely proxy for resident CSCs. Clearly, it should be stated that a single immortalized cell line cannot possibly capture the complexity and heterogeneity of the carcinomas from which they are derived also in the absence of the microenvironment that is now established as a main determinant of the identity and plasticity of cancer cells. Intra-tumor heterogeneity is determined *in vivo* by a combination of cell-autonomous and paracrine mechanisms, the latter triggered from the micro-environment. Within immortalized cancer cell lines, the

same mechanisms have become entirely cell-autonomous but are nonetheless worth studying as a proxy of the patient-derived carcinoma.

Methods

Cell Culture

The human colon cancer cell line SW480, obtained from the American Type Culture Collection (ATCC), was cultured in DMEM medium (Thermo Fisher Scientific) with 10% FCS (Thermo Fisher Scientific), 1% penicillin/streptomycin (Thermo Fisher Scientific, #15140122), and 1% glutamine (Gibco, #25030024), in humidified atmosphere at 37°C with 5% CO_2. The identity of the cell line and sorted subpopulations were confirmed by DNA fingerprinting with microsatellite markers (Amelogenin, CSF1PO, D13S317, D16S539, D5S818, D7S820, THO1, TPOX, vWA, D8S1179, FGA, Penta E, Penta D, D18S51, D3S1358, D21S11) and compared with the analogous data provided by ATCC, EACC, and https://web.expasy.org/cellosaurus/ (data not shown).

Plasmid transfection and lentiviral transduction

Stable transfection of the *RUNX2* overexpression plasmid (kind gift from Liang Fang, Addgene #52962) was performed using FuGENE HD transfection reagents (Promega, #E2311) according to the manufacturer's protocol. Geneticin (Gibco, #10131035) was employed at a concentration of 800 µg/ml to select for transfected cells. The level of overexpression was assessed by qPCR 24h and 72h post transfection.

Nusse Wnt reporter

The 7TGC (Addgene #24304) plasmid was transfected using Fugene HD transfection reagent (promega E2311) together with packaging vectors pMD2.G (Addgene #12259) and psPAX2 (Addgene #12260). 24 hours after transfection medium was collected and filtered. SW480 cells were transduced with the virus containing supernatant. Cells were selected based on mCherry expression.

TOP-Flash reporter assay

For the β-catenin/TCF reporter assay (TOP-Flash reporter assay), cells were plated on 48-well dishes and cultured in DMEM medium. After 24 hrs, when 70% confluence was reached, cells were transfected by Fugene HD (Promega) with 125 ng of the TOP-Flash or FOP-Flash reporter constructs together with 25 ng of the Renilla luciferase vector for normalization purposes. As a control HEK293 cells were cultured with L-cell or Wnt-conditioned media. Luciferase activity was measured using the Dual-GLO Luciferase Reporter Assay System (Promega) 48 hrs post-transfection. Luminescence was measured using a GloMax Luminometer.

qRT-PCR and PCR analyses

Total RNA was isolated using TRIzol reagent (Thermo Fisher Scientific, #15596018) and was reverse-transcribed using high-capacity cDNA reverse transcription kit (Life Technologies, #4368814), according to the manufacturer's instructions. qRT-PCR was implemented using the Fast SYBR Green Master Mix (Thermo Fisher Scientific) on an Applied Biosystems StepOne Plus Real-Time Thermal Cycling Research with three replicates per group. Relative gene expression was determined by normalizing the expression of each target gene to GAPDH. Results were analyzed using the 2-(ΔΔCt) method. qRT-PCR and PCR primers are listed in the table below.

Gene	**Forward primer**	**Reverse primer**
GAPDH	5'-ACCCAGAAGACTGTGGATGG-3'	5'-TCTAGACGGCAGGTCAGGTC-3'
LGR5	5'- GGAAATCATGCCTTACAGAGC-3'	5'- CCTGGGGAAGGTGAACACT-3'
EPCAM	5'- GCAGCTCAGGAAGAATGTG-3'	5'- CAGCCAGCTTTGAGCAAATGAC-3'
VIMENTIN	5'- GAGAACTTTGCCGTTGAAGC-3'	5'- GCTTCCTGTAGGTGGCAATC-3'
EMP1	5'- GTGTTCCAGCTCTTCACCATGG -3'	5'- GGAATAGCCGTGGTGATACTGC-3'
ZEB1	5'- GCACAACCAAGTGCAGAAGA -3'	5'- CATTTGCAGATTGAGGCTGA -3'
CD44	5'- TACAGCATCTCTCGGACGGA -3'	5'- CACCCCTGTGTTGTTTGCTG -3'
RUNX2	5'- CCCTGAACTCTGCACCAAGT -3'	5'- CCCAGTTCTGAAGCACCTGA -3'
TROP2	5'- CGGCAGAACACGTCTCAGAA -3'	5'- GCCCTGGAATAGAGACTCGC -3'

Flow cytometry and antibody staining

Cells were harvested using trypsin-EDTA (Thermo Fisher Scientific, #15400054), resuspended in PBS with 4% FCS, and stained on ice for 30' with the selected panel of antibodies, each at the concentration of 5μg/ml. The following antibodies were used: CD44-APC (clone IM7, 559250, BD Pharmingen), EpCAM-FITC (ESA-214, GTX30708, GeneTex), EpCAM-PerCP-Cy5.5 (clone 9C4, 324214, Biolegend), EpCAM-Pacific Blue (clone 9C4, 324217 Biolegend), Trop2-BV786 (743277 optibuild, BD Pharmingen), and Trop2-BV510 (563244, BD Pharmingen), CD133/2-PE (clone 293C3, 130-090-853. Miltenyi),. Trop2- Alexa546 (clone T16). The latter was produced as reported in Ambrogi et al.[32], CD133/2-PE (clone 293C3, 130-090-853. Miltenyi) and conjugated using an Alexa Fluor™ 546 Antibody Labeling Kit (A20183. Termofisher). After staining, cells were washed twice by centrifugation at 1200 rpm for 5' and resuspension in PBS with 4% FCS. Flow cytometric analysis and sorting were carried out with a FACSAria III Cell Sorter (BD Biosciences, New Jersey, USA). Sequential gating on FSC-A versus FSC-W and SSC-A versus SSC-W were employed to eliminate doublets and aggregates and ensure single-cell sorting. Dead cells were excluded by gating out the fraction positive to the nuclear dye DAPI (Sigma-Aldrich, #D9542), used at 0.5 μg/mL. All the

preliminary gating strategies for live/dead and doublet discrimination were performed as reported in Sacchetti et al.[12]. Additional gates, FMO samples, and compensation controls, were defined as specified in Suppl. Fig 1 and Suppl. Figure 3. FITC and GFP were detected using a 488 nm laser and 502 LP and 530/30 BP filters; APC was detected with a 633 nm laser and a 660/20 BP filter; DAPI and Pacific Blue were detected using a 405 nm laser and a 450/40 BP filter; BV786 was detected with a 405 nm laser and a 750 LP and 780/60 BP filter; PE, Alexa-546 and mCherry were detected with a 461 nm laser and a 582/15 BP filter. Cell sorting was performed using a 85μm nozzle and a pressure of 45psi. Sorting purity, tested directly on the sorted fraction, was generally higher than 99.9%. To ensure absolute purity of the sorted fractions, plasticity experiments were performed with cells sorted twice.

Compound Treatment

SW480 cells were divided into the adherent and Sphere fractions by FACS sorting. Afterwards, 25,000 cells were plated in 6-well Multi-well Cell plates with 2 mL of complete medium. The following day, the cells were treated with the following inhibitors for 48 hours: E(Z)h2 inhibitors (H3K27me3), GSK-126 (50uM); HDAC inhibitors: Trichostatin A (which turned out to be toxic for the cells at every concentration we tried: 1, 10 and 50 nM); BRG-1 inhibitor (BAFi, 50nM) and ED98 (Baficillin; ARID1A inhibitor 10uM). The cells were cultured in a 37°C, 5% CO2 incubator. 1% DMSO treated cells were also taken along as negative control. There was no effect on assay readout at DMSO. After 48hours incubation with the selected compounds, pictures were taken and RNA isolated for further characterization.

Top-Flash assay

For the β-catenin/TCF reporter assay (TOP-Flash reporter assay), cells were plated on 48-well dishes. After 48h, when 70% confluence was reached, cells were transfected by Fugene HD with 125 ng of the TOP-Flash or FOP-Flash reporter constructs together with 25 ng of the Renilla luciferase vector for normalization purposes. Luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega, #E1910) 24h post-transfection. Luminescence was measured using a GloMax Luminometer (Promega, #9100-102).

RNA seq analysis

TrueSeq adapter sequences were removed with Trimmomatic (v.0.33). Next, reads were aligned with STAR (v.2.4.2.a)[33] to the human reference genome (hg38) with GENCODE (v23) annotations. A count table was generated with FeatureCounts (v.2.0.3)[34] and downstream analysis was done in R with DESeq2 (1.36.0)[35]. Counts were normalized with variance stabilizing transformation (VST) and principal

component analysis (PCA) was performed on the top500 variably expressed features. Differential expression (DE) analysis was done by comparing each subpopulation to another using a Wald test with Benjamini-Hochberg adjustments for multiple testing. DE genes were selected based on p_adj < 0.05 and log2FoldChange > 1.5. For heat map visualization, DE genes were z-score normalized and clustered with k-means (k = 5). Signatures were evaluated with a gene set enrichment analysis (GSVA, v1.44.5)[36] and GSVA scores were compared with ANOVA to test for significance.

Analysis of spatial transcriptomics data

Pre-processed 10X Visium data sets[24-26] were imported in R and analyzed with Seurat (v4.3.0)[37]. Batch correction was performed with 2000 integration anchors using the reciprocal PCA (RPCA) method. Next, dimension reduction was performed with Uniform Manifold Approximation and Projection (UMAP; min.dist = 0.2, n.neighbors = 100, spread = 2) based on the first 50 principal components. Unsupervised clustering was done with the Leiden method (resolution = 0.2). Clusters of spots were annotated according to the expression of marker genes and by comparison with the previously published annotations. Tumor spots were sub-clustered with FindSubCluster (Louvain algorithm, resolution = 0.25), and annotated according to the expression of markers from the SW480 subpopulations. Signature scores were derived by averaging the Markov Affinity-based Graph Imputation of Cells (MAGIC)[38] imputed values after a z-score normalization to equalize gene weights. Neighborhood analysis was done with the SF package[39] by aggregating the expression profile of neighboring spots in honeycomb grids of 4x4 to create local niches. Subsequently, UMAP dimension reduction was performed and niches were clustered with the Leiden method (resolution = 0.3). Niches were annotation according to the composition of respective cell types. Spatial patterns were visualized with the SpatialDimPlot function in Seurat.

Data availability

RNA sequencing data of sorted SW480 spheres and bulk cells has been deposited in the gene expression omnibus and is publicly accessible with identifier GSE253110. Expression profiling of the EpCAM^{hi} and EpCAM^{lo} subpopulations can be accessed with identifier GSE154927. The spatial transcriptomics data sets used in this study are publicly available and can be accessed from Mendely data with identifier ys6j8bndby[24], from Zenodo using identifier 7744244[25], and from cancerdiversity.asia[26].

Figure legends

Figure 1. FACS-based molecular dissection of SW480 into three distinct subpopulations. **A**. The combination of fluorescent antibodies against EPCAM and TROP2 efficiently separates Spheres (turquoise, left) from the adherent subpopulations (right) in SW480 cells. **B**. Antibodies against EPCAM and CD44 can further separate EpCAMlo (red) cells from EpCAMhi cells (yellow) within the adherent population, and show heterogeneous CD44 levels inside the sphere population[12]. The quadrants indicate the regions that are negative, positive for the single markers, and double positive as shown in Suppl. Figure 1. The average percentage of SW480 adherent and spheres in standard and relatively young (<1.5 months) cultures is reported in the plots**.** **C**. Bright field pictures showing the morphology of the parental cell line and FACS-enriched subpopulations. Scale bar: 100 μm **D**. Comparison of Wnt activity with the Nusse reporter by FACS. Values denote relative Wnt activity, i.e. GFP normalized by the mCherry intensity. **E**. Confocal imaging of FACS-enriched subpopulations with the Nusse reporter. Scale bar: 50 μm **F**. Bar plot denoting results of TOPFlash assays for the distinct subpopulations of SW480.

Figure 2. Transcriptomic analyses of the SW480 subpopulation. **A**. Principal component analysis (PCA) plot showing the differences between the bulk and FACS-enriched subpopulations of SW480. **B**. Heat map denoting the z-score normalized expression values of the differentially expressed genes. Genes were clustered into distinct groups with k-means (k = 5) clustering. **C**. Evaluation of signatures for high-relapse cells (HRC), cancer-stem cells (CSC) and epithelial-to-mesenchymal transition (EMT) with gene set variation analysis. **D**. Bar plot indicating expression level as measured by qPCR. Markers represent genes for HRC (*EMP1*), CSC (*LGR5*) and EMT (*ZEB1*).

Figure 3. Effect of RUNX2 overexpression on the identity of SW480 subpopulations. **A**. Longitudinal FACS analysis of sorted adherent and sphere subpopulations. **B**. Analysis by qPCR of the *RUNX2* mRNA. N = 3 replicates. **C**. Bar plots showing the results of qPCRs for *RUNX2*, *CD44*, *EPCAM* and *TROP2*. Expression values were normalized against GAPDH and against the bulk control samples with the $2^{-\Delta\Delta Ct}$ method. **D**. FACS analysis showing the effect of 1w overexpression of RUNX2 on the SW480 subpopulations. **E**. Quantification of FACS result by comparison of mean fluorescent intensity (MFI) across different subpopulations. Asterisks denote significant levels of the adjusted p value after correcting for multiple testing with Tukey post-hoc test.

Figure 4. Effect of ARID1A inhibitor on the identity of SW480 spheres. **A**. RNAseq expression levels of ARID1A across the distinct SW480 subpopulation. **B**. Bright field images showing the effect of ARID1A

inhibitor ED98 on the morphology of spheres. Scale bar: 100 µm **C**. Bar plots showing the result of ED98 treatment on the expression levels of EPCAM, LGR5 and VIM as measured with qPCR analysis. Expression values were normalized against GAPDH and against the adherent control samples with the 2–ΔΔCt method. P values denote significant levels of the adjusted p value after correcting for multiple testing with Tukey post-hoc test.

Figure 5. Identification of subpopulations in colon cancer. **A**. Uniform Manifold Approximation and Projection (UMAP) dimension reduction plot displaying the distinct cell types across an integrated data set of three Visium spatial transcriptomics studies. **B**. UMAP plot showing the sub-clustering results of the tumor areas. Five distinct transcriptomic states were identified. **C**. Evaluation of signatures for EMT, HRC and CSC on the UMAP plot of the tumor area. **D**. Heat map showing the markers of the SW480 subpopulations, as well as additional markers for the identified 'secretory' and 'inflammatory' tumor areas. Vales denote z-score average expression values across the tumor sub-clusters. **E**. Stacked bar plot showing the relative fraction of cell types from the micro-environment (top) and tumor entities (bottom) throughout the samples of the data sets.

Figure 6. Spatial organization of tumor subpopulations in colon cancer. **A**. Uniform Manifold Approximation and Projection (UMAP) dimension reduction plot showing the expression profiles of local neighborhoods. Neighborhoods were clustered and annotated into different cancer niches. **B**. Heat map displaying the fraction of distinct cell types (rows) across the different cancer niches (columns). **C**. Spatial plots showing four examples of colon cancer. Tumor areas were projected on the H&E and colored according to their transcriptomic entity.

Supplementary Figure Legends

Supplementary Figure 1. Staining and compensation controls relative to Figure 1. **A**. EpCAM FITC vs Trop2-BV546; **B**. EpCAM FITC vs CD44-APC; **C**. CD44-APC vs Trop2 Alexa546. Unstained = mix of isotype controls. **D**. FACS plot showing the three different subpopulations in SW480 (left) with their corresponding expression of CD133 (right).

Supplementary Figure 2. Heat map denoting the z-score normalized expression value of the RNAseq data for the different SWI/SNF complex subunits across the different subpopulations.

Supplementary Figure 3. **A**. FACS Staining and compensation controls relative to Fig. 3A for (**A**) EPCAM and TROP2; (**B**) CD44 and EPCAM; (**C**) CD44 and TROP2; (**D-F**) Staining and compensation controls relative

to Fig. 3D (RUNX2 overexpression experiment) with separate staining for EPCAM (**D**), TROP2 (**E**) and CD44 (**F**). vs. the mCherry/RUNX2 reporter. Shown are plots relative to bulk SW480/RUNX2.

Supplementary Figure 4. A. UMAP plot showing merged data set on the Visium spot level. **B**. UMAP plot after integration with the reciprocal PCA (RPCA) method. **C**. UMAP plot showing merged data set on neighborhood level. **D**. UMAP plot of the neighborhoods after integration with the reciprocal PCA (RPCA) method.

References

1. Marusyk, A., Janiszewska, M., and Polyak, K. (2020). Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. Cancer Cell *37*, 471-484. S1535-6108(20)30147-1 [pii]10.1016/j.ccell.2020.03.007.2. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discov *12*, 31-46. 12/1/31 [pii]10.1158/2159-8290.CD-21-1059.

3. Househam, J., Heide, T., Cresswell, G.D., Spiteri, I., Kimberley, C., Zapata, L., Lynn, C., James, C., Mossner, M., Fernandez-Mateos, J., et al. (2022). Phenotypic plasticity and genetic control in colorectal cancer evolution. Nature *611*, 744-753. 10.1038/s41586-022-05311-x [pii]5311 [pii]10.1038/s41586-022-05311-x.

4. Munro, M.J., Wickremesekera, S.K., Peng, L., Tan, S.T., and Itinteang, T. (2018). Cancer stem cells in colorectal cancer: a review. J Clin Pathol *71*, 110-116. jclinpath-2017-204739 [pii]10.1136/jclinpath-2017-204739.

5. Ishimoto, T., Nagano, O., Yae, T., Tamada, M., Motohara, T., Oshima, H., Oshima, M., Ikeda, T., Asaba, R., Yagi, H., et al. (2011). CD44 variant regulates redox status in cancer cells by stabilizing the xCT subunit of system xc(-) and thereby promotes tumor growth. Cancer Cell *19*, 387-400. S1535-6108(11)00050-X [pii]10.1016/j.ccr.2011.01.038.

6. O'Brien, C.A., Pollett, A., Gallinger, S., and Dick, J.E. (2007). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature *445*, 106-110. nature05372 [pii]10.1038/nature05372.

7. Schepers, A.G., Snippert, H.J., Stange, D.E., van den Born, M., van Es, J.H., van de Wetering, M., and Clevers, H. (2012). Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. Science *337*, 730-735. science.1224676 [pii]10.1126/science.1224676.

8. Dalerba, P., Dylla, S.J., Park, I.K., Liu, R., Wang, X., Cho, R.W., Hoey, T., Gurney, A., Huang, E.H., Simeone, D.M., et al. (2007). Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A *104*, 10158-10163. 0703478104 [pii]6559 [pii]10.1073/pnas.0703478104.

9. Brabletz, T., Jung, A., Spaderna, S., Hlubek, F., and Kirchner, T. (2005). Opinion: migrating cancer stem cells - an integrated concept of malignant tumour progression. Nat Rev Cancer *5*, 744-749. nrc1694 [pii]10.1038/nrc1694.

10. Yi, H., Li, G., Long, Y., Liang, W., Cui, H., Zhang, B., Tan, Y., Li, Y., Shen, L., Deng, D., et al. (2020). Integrative multi-omics analysis of a colon cancer cell line with heterogeneous Wnt activity revealed RUNX2 as an epigenetic regulator of EMT. Oncogene *39*, 5152-5164. 10.1038/s41388-020-1351-z [pii]10.1038/s41388-020-1351-z.

11. Beerling, E., Seinstra, D., de Wit, E., Kester, L., van der Velden, D., Maynard, C., Schafer, R., van Diest, P., Voest, E., van Oudenaarden, A., et al. (2016). Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. Cell Rep *14*, 2281-2288. S2211-1247(16)30137-1 [pii]10.1016/j.celrep.2016.02.034.

12. Sacchetti, A., Teeuwssen, M., Verhagen, M., Joosten, R., Xu, T., Stabile, R., van der Steen, B., Watson, M.M., Gusinac, A., Kim, W.K., et al. (2021). Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. Elife *10*. 61461 [pii]10.7554/eLife.61461.

13. Cook, D.P., and Vanderhyden, B.C. (2022). Transcriptional census of epithelial-mesenchymal plasticity in cancer. Sci Adv *8*, eabi7640. 10.1126/sciadv.abi7640.
14. Canellas-Socias, A., Cortina, C., Hernando-Momblona, X., Palomo-Ponce, S., Mulholland, E.J., Turon, G., Mateo, L., Conti, S., Roman, O., Sevillano, M., et al. (2022). Metastatic recurrence in colorectal cancer arises from residual EMP1(+) cells. Nature *611*, 603-613. 10.1038/s41586-022-05402-9 [pii]10.1038/s41586-022-05402-9.
15. Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633-644. S0092-8674(11)00824-5 [pii]10.1016/j.cell.2011.07.026.
16. Xu, T., Verhagen, M., Joosten, R., Sun, W., Sacchetti, A., Munoz Sagredo, L., Orian-Rousseau, V., and Fodde, R. (2022). Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behavior in colon cancer. Elife *11*. 82006 [pii]10.7554/eLife.82006.
17. Xu, T., Verhagen, M.P., Teeuwssen, M., Sun, W., Joosten, R., Sacchetti, A., Ewing-Graham, P.C., Jansen, M., Boere, I.A., Bryce, N.S., et al. (2024). Tropomyosin1 isoforms underlie epithelial to mesenchymal plasticity, metastatic dissemination, and resistance to chemotherapy in high-grade serous ovarian cancer. Cell Death Differ. 10.1038/s41418-024-01267-9 [pii]10.1038/s41418-024-01267-9.
18. Hirsch, D., Barker, N., McNeil, N., Hu, Y., Camps, J., McKinnon, K., Clevers, H., Ried, T., and Gaiser, T. (2014). LGR5 positivity defines stem-like cells in colorectal cancer. Carcinogenesis *35*, 849-858. bgt377 [pii]10.1093/carcin/bgt377.
19. Wang, Y., Zhou, L., Qing, Q., Li, Y., Li, L., Dong, X., and Xiao, B. (2019). Gene expression profile of cancer stem-like cells in the SW480 colon adenocarcinoma cell line. Oncol Rep *42*, 386-398. 10.3892/or.2019.7146.
20. Xiong, B., Ma, L., Hu, X., Zhang, C., and Cheng, Y. (2014). Characterization of side population cells isolated from the colon cancer cell line SW480. Int J Oncol *45*, 1175-1183. 10.3892/ijo.2014.2498.
21. Petrova, T.V., Nykanen, A., Norrmen, C., Ivanov, K.I., Andersson, L.C., Haglund, C., Puolakkainen, P., Wempe, F., von Melchner, H., Gradwohl, G., et al. (2008). Transcription factor PROX1 induces colon cancer progression by promoting the transition from benign to highly dysplastic phenotype. Cancer Cell *13*, 407-419. S1535-6108(08)00060-3 [pii]10.1016/j.ccr.2008.02.020.
22. Fuerer, C., and Nusse, R. (2010). Lentiviral vectors to probe and manipulate the Wnt signaling pathway. PLoS One *5*, e9370. 09-PONE-RA-15047R1 [pii]10.1371/journal.pone.0009370.
23. Marian, C.A., Stoszko, M., Wang, L., Leighty, M.W., de Crignis, E., Maschinot, C.A., Gatchalian, J., Carter, B.C., Chowdhury, B., Hargreaves, D.C., et al. (2018). Small Molecule Targeting of Specific BAF (mSWI/SNF) Complexes for HIV Latency Reversal. Cell Chem Biol *25*, 1443-1455 e1414. S2451-9456(18)30270-8 [pii]10.1016/j.chembiol.2018.08.004.
24. Qi, J., Sun, H., Zhang, Y., Wang, Z., Xun, Z., Li, Z., Ding, X., Bao, R., Hong, L., Jia, W., et al. (2022). Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer. Nat Commun *13*, 1742. 10.1038/s41467-022-29366-6 [pii]29366 [pii]10.1038/s41467-022-29366-6.

25. Valdeolivas, A.A., B.; Giroud, N. et al. (2023). Charting the Heterogeneity of Colorectal Cancer Consensus Molecular Subtypes using Spatial Transcriptomics. BioRXiv. https://doi.org/10.1101/2023.01.23.525135.
26. Wu, Y., Yang, S., Ma, J., Chen, Z., Song, G., Rao, D., Cheng, Y., Huang, S., Liu, Y., Jiang, S., et al. (2022). Spatiotemporal Immune Landscape of Colorectal Cancer Liver Metastasis at Single-Cell Level. Cancer Discov *12*, 134-153. 2159-8290.CD-21-0316 [pii]10.1158/2159-8290.CD-21-0316.
27. Albuquerque, C., Breukel, C., van der Luijt, R., Fidalgo, P., Lage, P., Slors, F.J., Leitao, C.N., Fodde, R., and Smits, R. (2002). The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. Hum Mol Genet *11*, 1549-1560. 10.1093/hmg/11.13.1549.
28. Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., and Clevers, H. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134-144. S0092-8674(10)01064-0 [pii]10.1016/j.cell.2010.09.016.
29. Moorman, A.R., Cambuli, F., E.K., B., and al., e. (2023). Progressive plasticity during colorectal cancer metastasis. BioRXiv. https://doi.org/10.1101/2023.08.18.553925.
30. Chou, A., Toon, C.W., Clarkson, A., Sioson, L., Houang, M., Watson, N., DeSilva, K., and Gill, A.J. (2014). Loss of ARID1A expression in colorectal carcinoma is strongly associated with mismatch repair deficiency. Hum Pathol *45*, 1697-1703. S0046-8177(14)00162-2 [pii]10.1016/j.humpath.2014.04.009.
31. Ye, J., Zhou, Y., Weiser, M.R., Gonen, M., Zhang, L., Samdani, T., Bacares, R., DeLair, D., Ivelja, S., Vakiani, E., et al. (2014). Immunohistochemical detection of ARID1A in colorectal carcinoma: loss of staining is associated with sporadic microsatellite unstable tumors with medullary histology and high TNM stage. Hum Pathol *45*, 2430-2436. S0046-8177(14)00351-7 [pii]10.1016/j.humpath.2014.08.007.
32. Ambrogi, F., Fornili, M., Boracchi, P., Trerotola, M., Relli, V., Simeone, P., La Sorda, R., Lattanzio, R., Querzoli, P., Pedriali, M., et al. (2014). Trop-2 is a determinant of breast cancer survival. PLoS One *9*, e96993. PONE-D-13-51525 [pii]10.1371/journal.pone.0096993.
33. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21. bts635 [pii]10.1093/bioinformatics/bts635.
34. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930. btt656 [pii]10.1093/bioinformatics/btt656.
35. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. s13059-014-0550-8 [pii]550 [pii]10.1186/s13059-014-0550-8.
36. Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7. 1471-2105-14-7 [pii]10.1186/1471-2105-14-7.

37. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587 e3529. S0092-8674(21)00583-3 [pii]10.1016/j.cell.2021.04.048.

38. van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716-729 e727. S0092-8674(18)30724-4 [pii]10.1016/j.cell.2018.05.061.

39. Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal *10:1, 439-446*.

Acknowledgements

We thank all members from the Fodde laboratory for helpful discussions. This study was financially supported by the Dutch Cancer Society (KWF; project no. 11407) and the China Scholarship Council (CSC; project no. 201806300047).

Author Contributions

M.P.V., T.X, R.S., R.J., and A.S. designed and performed experiments. M.P.V. analyzed sequencing data sets. A.S. optimized and implemented the FACS strategy. M.P.V. and R.F. wrote the manuscript. R.F. supervised the study.

Competing Interests

The authors declare no competing interests.

Figure 1

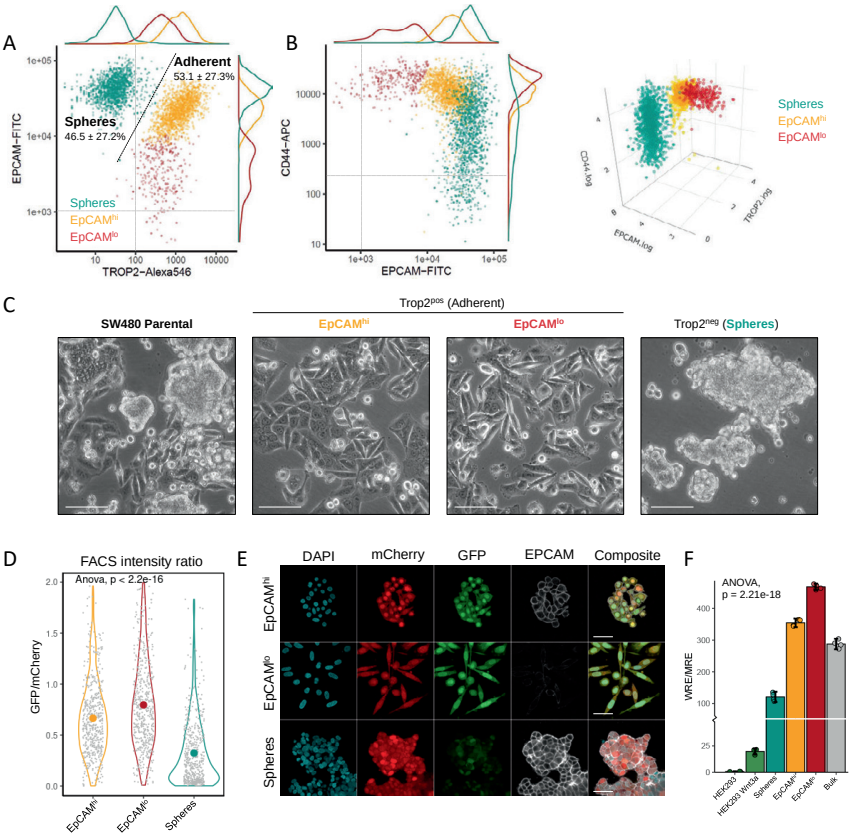

A
Adherent
53.1 ± 27.3%
Spheres
46.5 ± 27.2%
Spheres
EpCAMhi
EpCAMlo
EPCAM−FITC
TROP2−Alexa546
B
CD44−APC
EPCAM−FITC
Spheres
EpCAMhi
EpCAMlo
C
SW480 Parental
Trop2pos (Adherent)
EpCAMhi
EpCAMlo
Trop2neg (Spheres)
D
FACS intensity ratio
Anova, p < 2.2e-16
GFP/mCherry
EpCAMhi
EpCAMlo
Spheres
E
DAPI
mCherry
GFP
EPCAM
Composite
EpCAMhi
EpCAMlo
Spheres
F
ANOVA,
p = 2.21e-18
WRE/MRE
HEK293
HEK293 Wnt3A
Spheres
EpCAMhi
EpCAMlo
Bulk

Figure 2

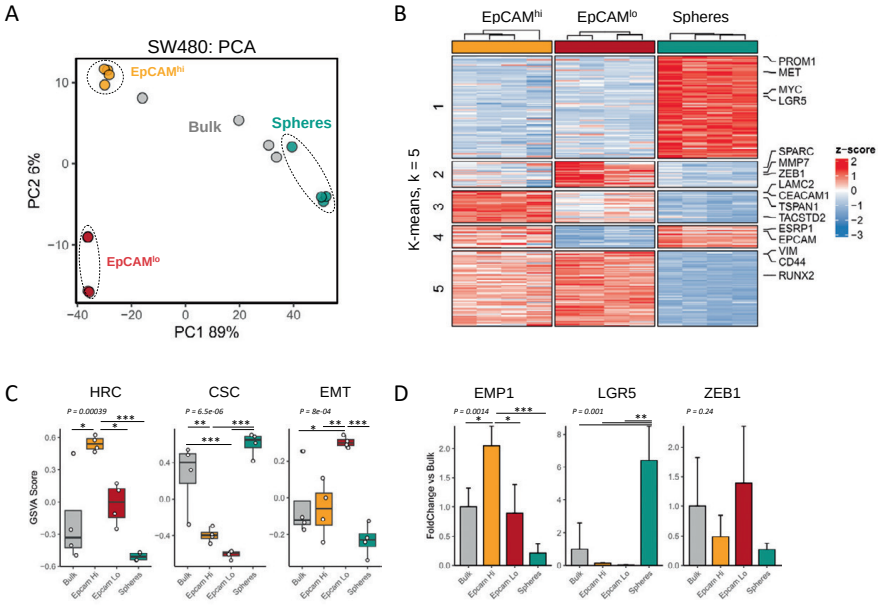

A
SW480: PCA
EpCAMhi
Bulk
Spheres
EpCAMlo
PC2 8%
PC1 89%
B
EpCAMhi
EpCAMlo
Spheres
K-means, k = 5
PROM1
MET
MYC
LGR5
SPARC
MMP7
ZEB1
LAMC2
CEACAM1
TSPAN1
TACSTD2
ESRP1
EPCAM
VIM
CD44
RUNX2
z-score
C
HRC
CSC
EMT
GSVA Score
Bulk
Epcam Hi
Epcam Lo
Spheres
D
EMP1
LGR5
ZEB1
FoldChange vs Bulk

Figure 3

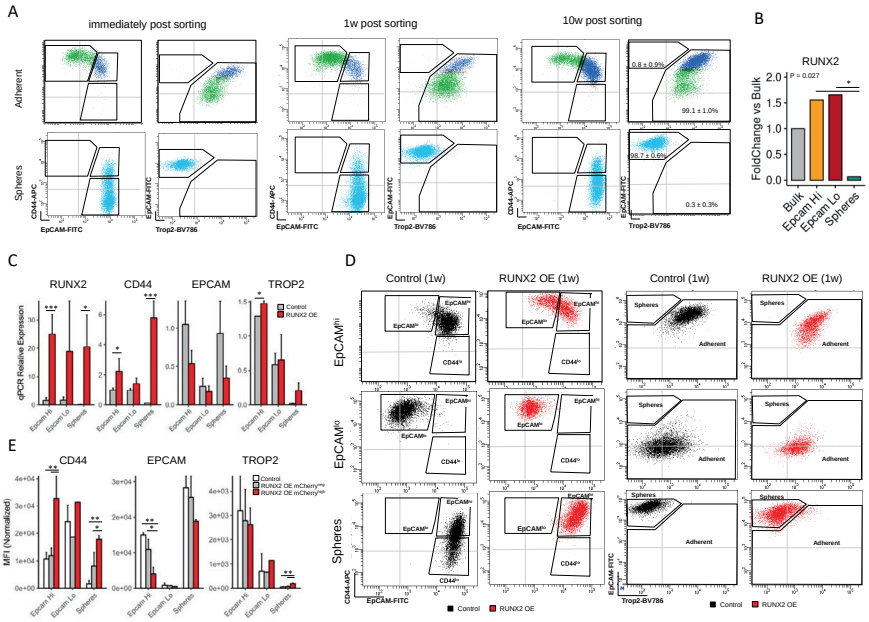

A
immediately post sorting
1w post sorting
10w post sorting
Adherent
Spheres
EpCAM-FITC
Trop2-BV786
CD44-APC
0.8 ± 0.9%
99.1 ± 1.0%
98.7 ± 0.6%
0.3 ± 0.3%
B
RUNX2
P = 0.027
FoldChange vs Bulk
Bulk
Epcam Hi
Epcam Lo
Spheres
C
RUNX2
CD44
EPCAM
TROP2
qPCR Relative Expression
Control
RUNX2 OE
D
Control (1w)
RUNX2 OE (1w)
EpCAMHi
EpCAMLo
Spheres
Adherent
E
CD44
EPCAM
TROP2
MFI (Normalized)

Figure 4

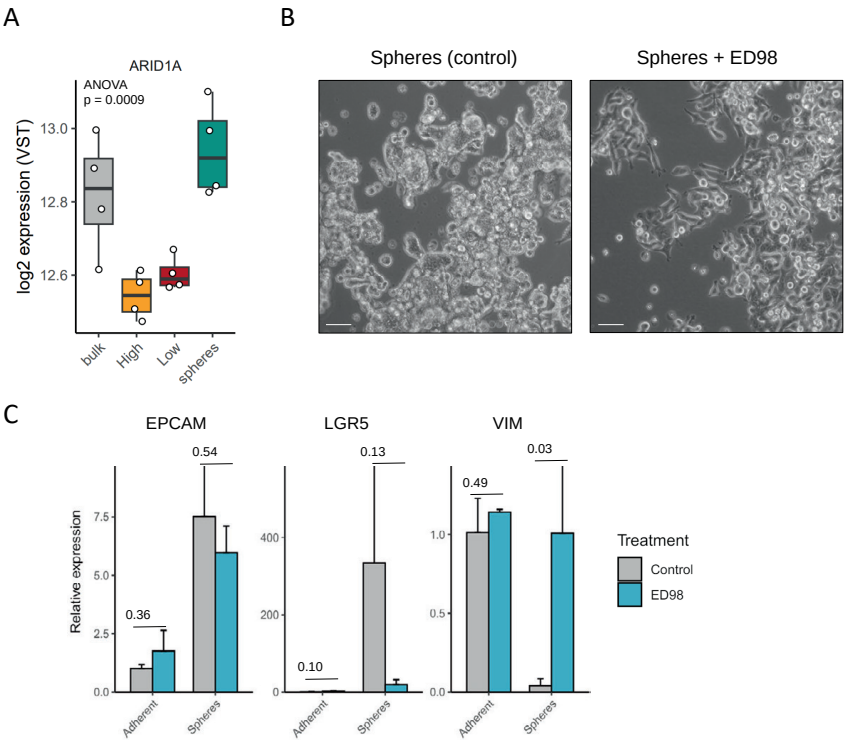

A
ARID1A
ANOVA
p = 0.0009
log2 expression (VST)
13.0
12.8
12.6
bulk
High
Low
spheres
B
Spheres (control)
Spheres + ED98
C
EPCAM
LGR5
VIM
Relative expression
0.36
0.54
0.10
0.13
0.49
0.03
Adherent
Spheres
Treatment
Control
ED98

Figure 5

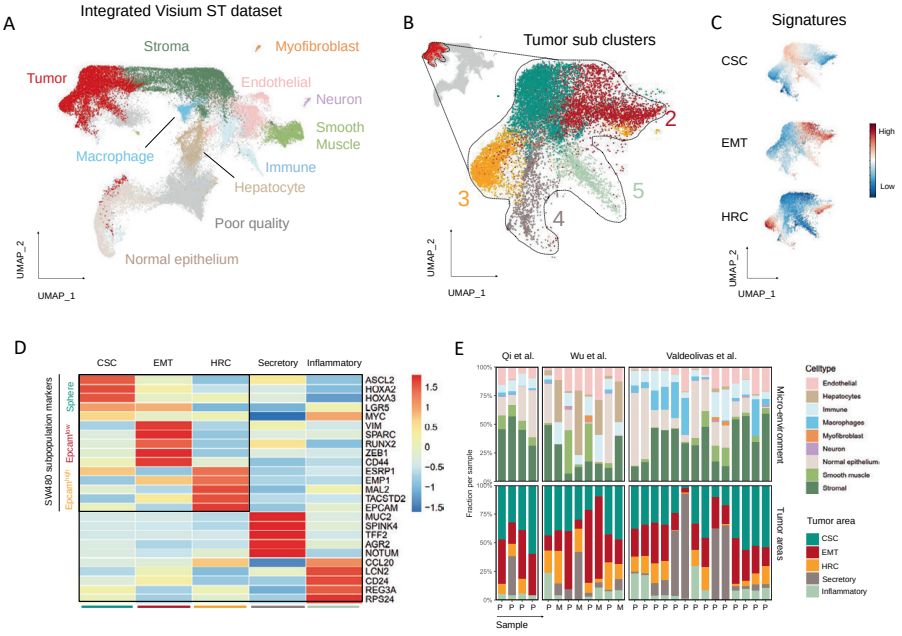
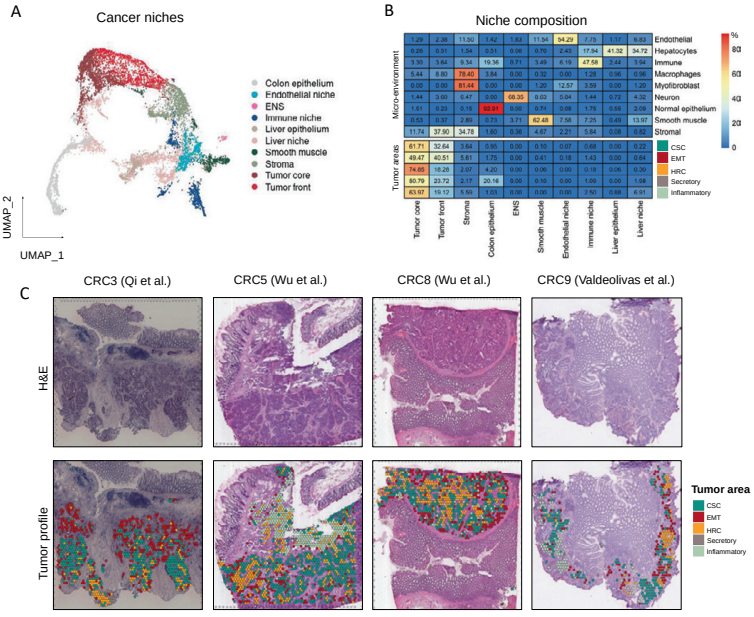
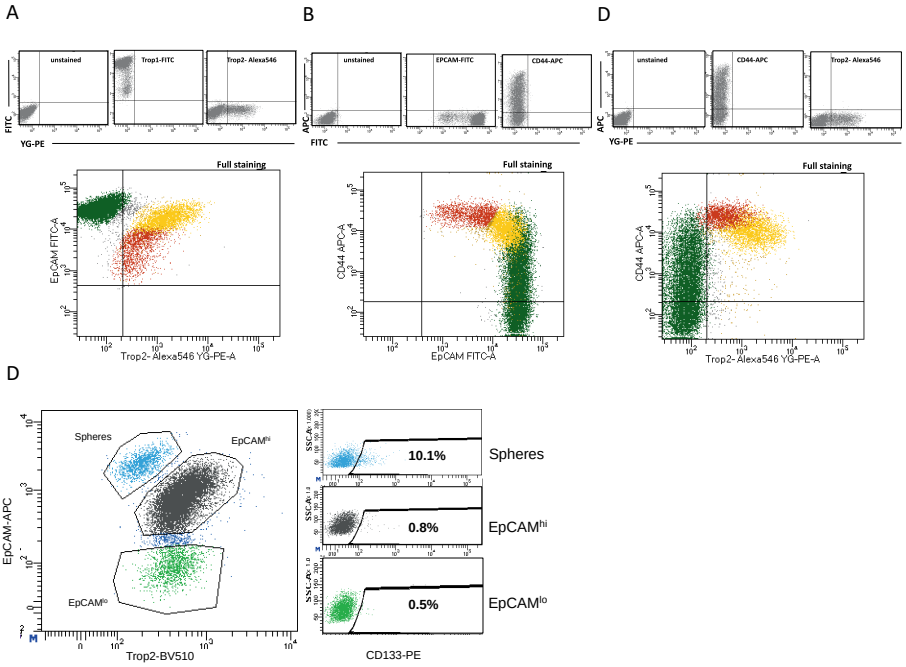

A
Integrated Visium ST dataset
Stroma
Myofibroblast
Tumor
Endothelial
Neuron
Smooth Muscle
Macrophage
Immune
Hepatocyte
Poor quality
Normal epithelium
UMAP_2
UMAP_1
B
Tumor sub clusters
1
2
3
4
5
UMAP_2
UMAP_1
C
Signatures
CSC
EMT
HRC
High
Low
UMAP_2
UMAP_1
D
CSC
EMT
HRC
Secretory
Inflammatory
SW480 subpopulation markers
Sphere
Epcam^low
Epcam^high
ASCL2
HOXA2
HOXA3
LGR5
MYC
VIM
SPARC
RUNX2
ZEB1
CD44
ESRP1
EMP1
MAL2
TACSTD2
EPCAM
MUC2
SPINK4
TFF2
AGR2
NOTUM
COL20
LCN2
CD24
REG3A
RPS24
1.5
1
0.5
0
−0.5
−1
−1.5
E
Qi et al.
Wu et al.
Valdeolivas et al.
Fraction per sample
100%
75%
50%
25%
0%
Micro-environment
Tumor areas
Celltype
Endothelial
Hepatocytes
Immune
Macrophages
Myofibroblast
Neuron
Normal epithelium
Smooth muscle
Stromal
Tumor area
CSC
EMT
HRC
Secretory
Inflammatory
P P P P P M P M P M P P P P P P P P P P P P P P P P P M
Sample

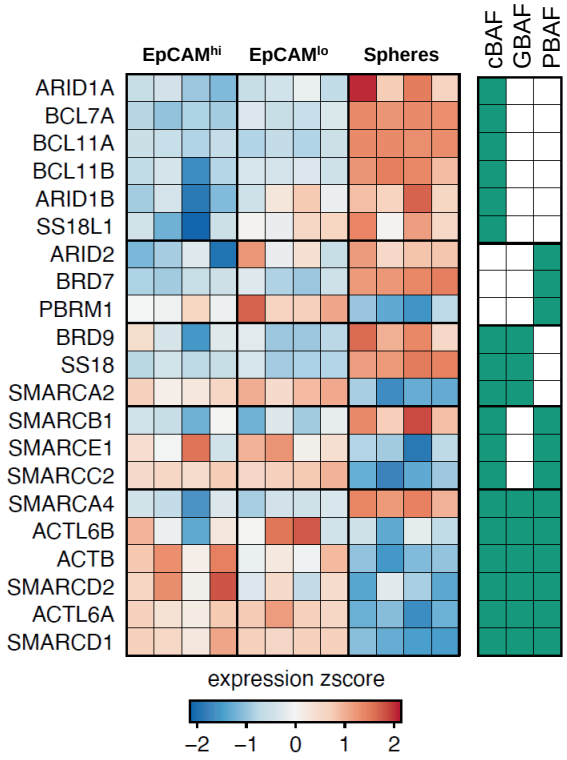
Figure 6


A
Cancer niches
Colon epithelium
Endothelial niche
ENS
Immune niche
Liver epithelium
Liver niche
Smooth muscle
Stroma
Tumor core
Tumor front
UMAP_2
UMAP_1
B
Niche composition
Micro-environment
Tumor areas
Endothelial
Hepatocytes
Immune
Macrophages
Myofibroblast
Neuron
Normal epithelium
Smooth muscle
Stromal
CSC
EMT
HRC
Secretory
Inflammatory
Tumor core
Tumor front
Stroma
Colon epithelium
ENS
Smooth muscle
Endothelial niche
Immune niche
Liver epithelium
Liver niche
C
CRC3 (Qi et al.)
CRC5 (Wu et al.)
CRC8 (Wu et al.)
CRC9 (Valdeolivas et al.)
H&E
Tumor profile
Tumor area
CSC
EMT
HRC
Secretory
Inflammatory

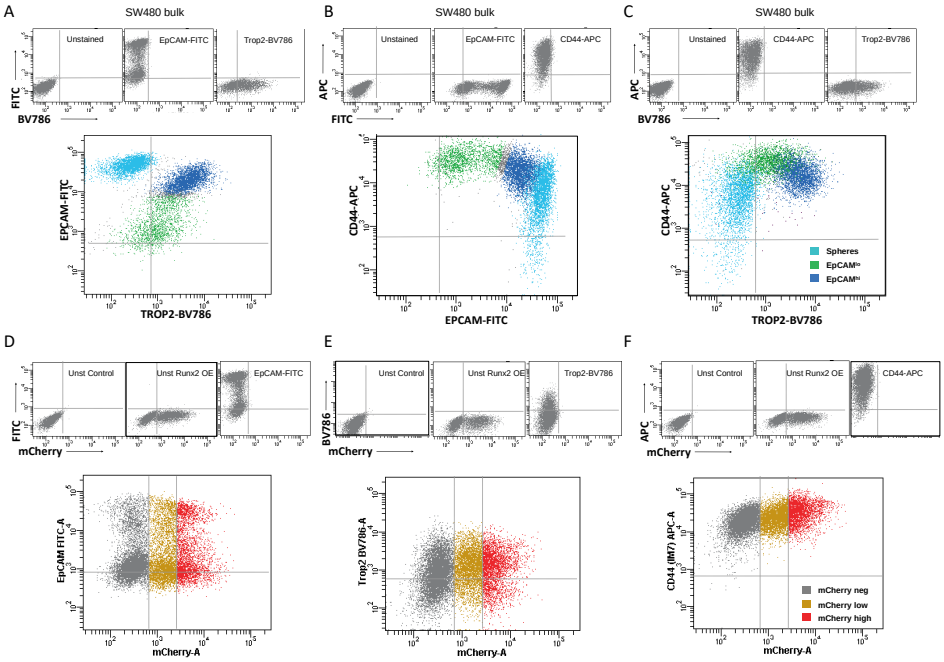
Supplementary Figure 1



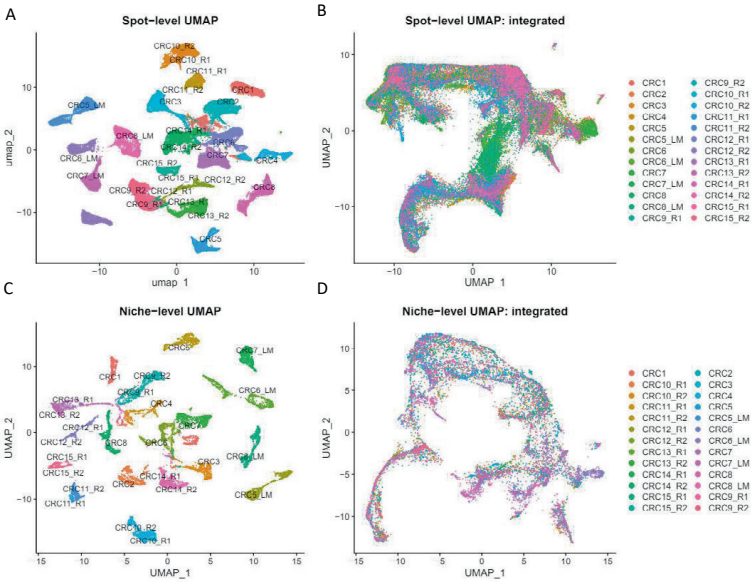
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



CHAPTER 4

Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behavior in colon cancer

Elife. 11:e82006. (2022)

RESEARCH ARTICLE

Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behavior in colon cancer

Tong Xu[1], Mathijs Verhagen[1], Rosalie Joosten[1], Wenjie Sun[2], Andrea Sacchetti[1], Leonel Munoz Sagredo[3,4], Véronique Orian-Rousseau[3], Riccardo Fodde[1]*

[1]Department of Pathology, Erasmus University Medical Center, Rotterdam, Netherlands; [2]Laboratory of Genetics and Developmental Biology, Institute Curie, Paris, France; [3]Institute of Biological and Chemical Systems - Functional Molecular Systems (IBCS FMS), Karlsruhe Institute of Technology, Karlsruhe, Germany; [4]Faculty of Medicine, University of Valparaiso, Valparaiso, Chile

Abstract Phenotypic plasticity allows carcinoma cells to transiently acquire the quasi-mesenchymal features necessary to detach from the primary mass and proceed along the invasion-metastasis cascade. A broad spectrum of epigenetic mechanisms is likely to cause the epithelial-to-mesenchymal (EMT) and mesenchymal-to-epithelial (MET) transitions necessary to allow local dissemination and distant metastasis. Here, we report on the role played by alternative splicing (AS) in eliciting phenotypic plasticity in epithelial malignancies with focus on colon cancer. By taking advantage of the coexistence of subpopulations of fully epithelial (EpCAM^{hi}) and quasi-mesenchymal and highly metastatic (EpCAM^{lo}) cells in conventional human cancer cell lines, we here show that the differential expression of *ESRP1* and other RNA-binding proteins (RBPs) downstream of the EMT master regulator *ZEB1* alters the AS pattern of a broad spectrum of targets including *CD44* and *NUMB*, thus resulting in the generation of specific isoforms functionally associated with increased invasion and metastasis. Additional functional and clinical validation studies indicate that both the newly identified RBPs and the CD44s and NUMB2/4 splicing isoforms promote local invasion and distant metastasis and are associated with poor survival in colon cancer. The systematic elucidation of the spectrum of EMT-related RBPs and AS targets in epithelial cancers, apart from the insights in the mechanisms underlying phenotypic plasticity, will lead to the identification of novel and tumor-specific therapeutic targets.

***For correspondence:** r.fodde@erasmusmc.nl

Competing interest: The authors declare that no competing interests exist.

Funding: See page 19

Preprinted: 01 March 2022
Received: 19 July 2022
Accepted: 07 November 2022
Published: 08 November 2022

Reviewing Editor: Lynne-Marie Postovit, University of Alberta, Canada

Editor's evaluation

This fundamental study provides a valuable analysis of the splicing landscape in colon cancer cells that have properties intermediate between those typically found in primary cancers ("epithelial") and those that are spreading by metastasis ("mesenchymal"). The strength of evidence provided is solid and convincing and supports current ideas that changes in the way that RNA from particular genes is processed plays a key role in cancer spread.

Introduction

Colon cancer still represents one of the major causes of cancer-related morbidity and mortality worldwide. Apart from its high incidence, the adenoma-carcinoma sequence along which colon cancer progresses has served as a classic model to elucidate the underlying genetic alterations representative of virtually all of the hallmarks of cancers (***Hanahan, 2022***), possibly with the only exception of

'*activating invasion and metastasis (unlocking phenotypic plasticity; non-mutational epigenetic reprogramming)*'. As also reported in other epithelial cancers, the several steps of the invasion-metastasis cascade are not caused by genetic alterations but rather by transient morphological and gene expression changes of epigenetic nature (**Bernards and Weinberg, 2002**; **Reiter et al., 2018**). In this context, epithelial-mesenchymal transition (EMT) and its reverse mesenchymal-epithelial transition (MET) likely represent the main mechanisms underlying local dissemination and distant metastasis (**Thiery et al., 2009**; **Brabletz et al., 2005**). EMT is triggered at the invasive front of the primary colon carcinoma in cells earmarked by nuclear β-catenin and enhanced Wnt signaling, as the result of their physical and paracrine interactions with the microenvironment (**Fodde and Brabletz, 2007**). The acquisition of quasi-mesenchymal features allows local invasion and dissemination through the surrounding stromal compartment. Of note, EMT/MET should not be regarded as binary processes in view of the existence of metastable hybrid E/M states (partial EMT [pEMT]) endowed with phenotypic plasticity and likely to underlie the reversible morphological and functional transitions necessary to successfully complete the invasion-metastasis cascade (**Teeuwssen and Fodde, 2019**).

The molecular basis of the epigenetic changes underlying EMT and MET is likely to encompass a broad spectrum of mechanisms ranging from chromatin remodeling and histone modifications to promoter DNA methylation, non-coding RNAs (e.g. microRNAs), and alternative splicing (AS). The inclusion/exclusion of specific exons in mature mRNAs results in different protein isoforms with distinct biological functions. AS occurs in 92–94% of human genes leading to enriched protein density (**Wang et al., 2008**; **Blencowe, 2006**). Several sequence-specific RNA-binding proteins (RBPs) have been identified which bind pre-mRNAs to control AS in context-dependent fashion (**Fu and Ares, 2014**). Multiple cancer-specific AS variants have been found to underlie progression and metastasis (**Kahles et al., 2018**). Likewise, AS has been suggested to play key roles in EMT/MET (**Roy Burman et al., 2021**; **Oltean and Bates, 2014**) and phenotypic plasticity (**Biamonti et al., 2019**) in cancer by expression changes in RBP-encoding genes and their consequences for the modulation of downstream AS targets.

The *ESRP1* (epithelial splicing regulatory protein 1) gene encodes for an epithelial-specific RBP and splicing regulator shown to play a central role in EMT by modulating AS of EMT-associated genes including *FGFR2*, Mena, *CD44*, and p120-catenin (**Thiery et al., 2009**). Relevant to the present study, ESRP1 was reported to regulate the EMT from CD44v (variable) to CD44s (standard) isoforms in breast and lung cancer progression (**Brown et al., 2011**; **Yae et al., 2012**). As for colon cancer, whether ESRP1 regulates AS of CD44 and other target genes downstream of EMT/MET activation during invasion and metastasis is yet poorly understood.

Recently, we identified and thoroughly characterized subpopulations of $CD44^{hi}/EpCAM^{lo}$ cells (here referred to as $EpCAM^{lo}$) that coexist within immortalized colon cancer cell lines with their epithelial counterparts ($CD44^{hi}/EpCAM^{hi}$; for brevity $EpCAM^{hi}$) through stochastic state transitions governed by phenotypic plasticity and pEMT (**Sacchetti et al., 2021**). Accordingly, $EpCAM^{lo}$ cells feature highly invasive and metastatic capacities. Here, we took advantage of these in vitro models of phenotypic plasticity to test the hypothesis according to which AS driven by upstream RBPs underlie EMT (and MET). Among the identified AS targets, specific CD44 and NUMB isoforms were shown to play specific and unexpected roles in stemness and cancer. Moreover, we provide an extensive list of additional EMT-related RBPs and AS targets and show that many are conserved in other epithelial malignancies. Likewise, RBPs and AS targets differentially expressed among distinct carcinoma types are likely to reflect the distinct modalities through which these malignant cells metastasize.

Results

Differential expression of RBPs in the quasi-mesenchymal and highly metastatic $EpCAM^{lo}$ colon cancer cells affects AS of a broad spectrum of downstream target genes

As previously reported, the $EpCAM^{lo}$ subpopulation of colon cancer cells is earmarked by increased expression of the *ZEB1* transcription factor, responsible for EMT activation and for their quasi-mesenchymal and highly metastatic phenotype (**Sacchetti et al., 2021**). It has been established that in breast and pancreatic cancer *ZEB1*-driven EMT downregulates the expression of the RBP and splicing regulator *ESRP1* as part of a self-enforcing feedback loop (**Preca et al., 2015**). Accordingly,

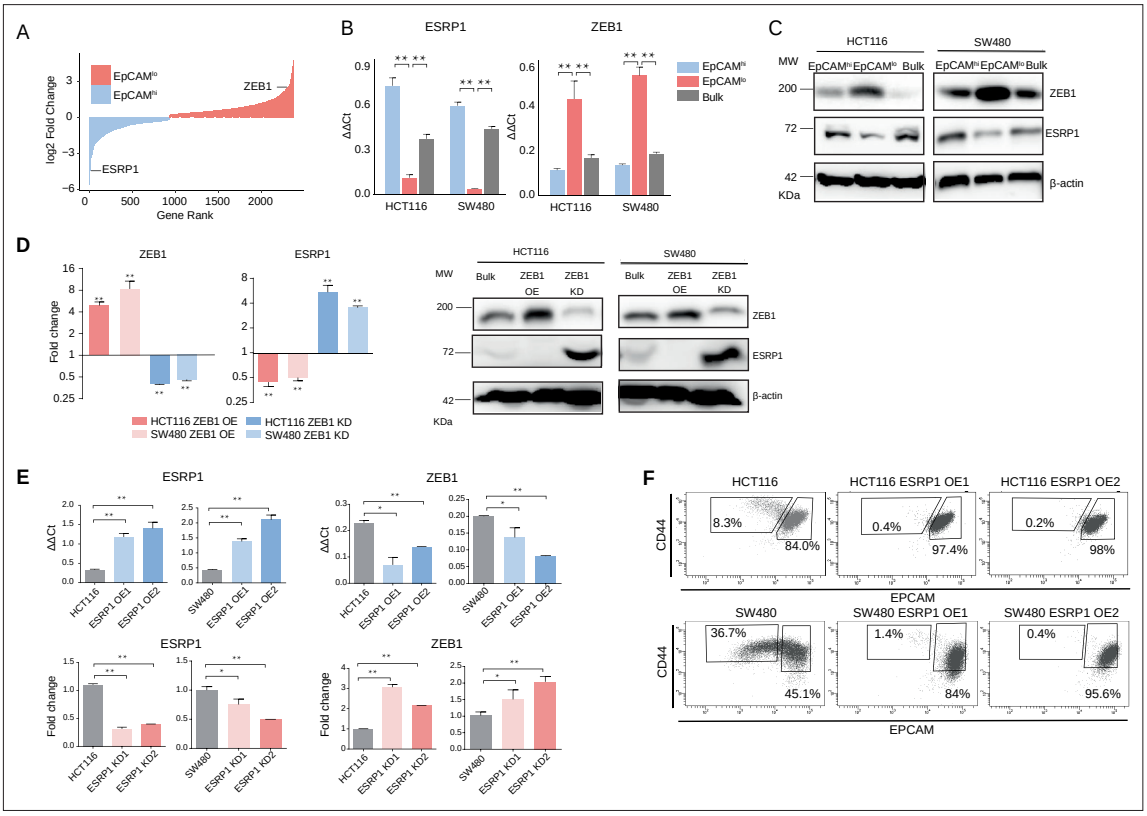


Figure 1. *ZEB1* and *ESRP1* differential expression in quasi-mesenchymal and highly metastatic EpCAM^{lo} colon cancer cells. (**A**) Gene rank plot showing differentially expressed genes between EpCAM^{hi} and EpCAM^{lo} with combined analysis of HCT116 and SW480. (**B**) RT-qPCR *ESRP1* and *ZEB1* expression analysis of HCT116 and SW480 EpCAM^{hi}, EpCAM^{lo}, and bulk subpopulations. *GAPDH* expression was used as control (means ± SEM, n=3). **=p < 0.01. (**C**) ESRP1 and ZEB1 western analysis in HCT116 and SW480 EpCAM^{hi}, EpCAM^{lo}, and bulk fractions. β-Actin was used as loading control. (**D**) RT-qPCR and western analysis of *ZEB1* and *ESRP1* expression in *ZEB1*-OE and -KD HCT116 and SW480 cells. Expression values were normalized in each sample with those from the parental HCT116 and SW480 cell lines. HCT116 and SW480 cells transduced with the sh*ZEB1* lentivirus were induced by 1 µg/mL doxycycline for 72 hr. Expression values were normalized with those from non-induced cells; *GAPDH* expression was employed as control (means ± SEM, n=3). *=p < 0.05, **=p < 0.01. β-Actin was used as loading control. (**E**) RT-qPCR *ZEB1* and *ESRP1* expression analysis in *ESRP1*-OE and -KD HCT116 and SW480 cells. Two independent *ESRP1*-OE clones were selected for each cell line. Expression values were normalized in each sample with those from the parental HCT116 and SW480 cell lines. HCT116 and SW480 cells transduced with the sh*ESRP1* lentivirus were induced by 1 µg/mL doxycycline for 72 hr. Two independent clones were selected for each cell line. Expression values were normalized with those from non-induced cells; *GAPDH* expression was employed as control (means ± SEM, n=3). *=p < 0.05, **=p < 0.01. (**F**) CD44/EpCAM FACS analysis of HCT116 and SW480 EpCAM^{lo} and EpCAM^{hi} subpopulations in ESRP1-OE cells. Two independent clones are shown for each cell lines.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. Original files and labelled bands of western blots in ***Figure 1C–D***.

Figure supplement 1. *ESRP1* and RNA-binding proteins (RBPs) functional and expression analysis in cell lines and patient-derived colon cancers.

among the top differentially expressed genes (DEGs) between EpCAM^{lo} and EpCAM^{hi} in SW480 and HCT116 colon cancer cells, *ESRP1* was found to be downregulated both at the RNA and at the protein level in the quasi-mesenchymal subpopulation where *ZEB1* expression is upregulated (***Figure 1A–C***). Gain- and loss-of-function analyses of both genes confirmed the inter-dependence of their expression levels in both cell lines (***Figure 1D–E***). Of note, *ESRP1* overexpression in the HCT116 and SW480 cell lines resulted in the dramatic reduction of their EpCAM^{lo} subpopulations and the expansion of the

epithelial bulk (EpCAM^{hi}), as shown by FACS analysis (***Figure 1F***, ***Figure 1—figure supplement 1A***). However, *ESRP1* knockdown (KD) gave rise to less clear and extremely variable results among the individual clones analyzed by FACS, in particular in the SW480 cell line. More coherent and representative results were obtained with the pools of the KD transfections (***Figure 1—figure supplement 1B***).

These results suggest that RBPs other than ESRP1 are likely to be involved in the AS regulation of the EpCAM^{lo} colon cancer subpopulation. Indeed, by taking advantage of the RBPDB database (***Cook et al., 2011***), we found that, apart from *ESRP1*, consistent differential expression in the quasi-mesenchymal subpopulation of both cell lines was observed for *ESRP2*, *RBM47*, *MBNL3* (downregulated) and *NOVA2*, *MBNL2* (upregulated). Other RBPs were found to be differentially expressed though in only one of the two cell lines (***Figure 1—figure supplement 1C***). In validation of the clinical relevance of the RBPs found to be differentially expressed between the EpCAM$^{hi/lo}$ subpopulations derived from the SW480 and HCT116 cell lines, the RBP-coding genes *QKI*, *RBM24*, and *MBNL2* (up in EpCAM^{lo}), and *ESRP1/2* and *RBM47* (down in EpCAM^{lo}) were found to be respectively up- and down-regulated in the consensus molecular subtype 4 (CMS4) of colon cancers, responsible for ~25% of the cases and earmarked by poor prognosis and a pronounced mesenchymal component (***Figure 1—figure supplement 1D***; ***Guinney et al., 2015***).

Differentially spliced target genes between EpCAM^{lo} and EpCAM^{hi} colon cancer cells from the SW480 and HCT116 cell lines were selected based on exon skip splicing events with ΔPSI (differential percentage spliced in) values >10%. The PSI value ranges from 0 to 1 and is a measurement of the percentage of isoform with an alternative exon included (***Schafer et al., 2015***). This resulted in a large and rather heterogeneous group of alternative spliced targets (n=1495; ***Supplementary file 1a***) with no clear enrichment in any specific gene ontology class (data not shown). In order to identify differentially spliced target genes in RBP-specific fashion, we took advantage of RNAseq data sets from previous *ESRP1*-, *ESRP2*-, *RBM47*-, and *QKI*-KD studies in different cancer cell lines and compared them with our own AS data relative to the EpCAM$^{hi/lo}$ colon cancer subpopulations (***Sacchetti et al., 2021***; ***Figure 2A*** and ***Figure 2—figure supplement 1***). A total of 32 common skipped exons events in 20 genes were identified between EpCAM^{lo} colon (both cell lines) and *ESRP1* KD H358 lung cancer cells (***Yang et al., 2016***; ***Figure 2A***). More extensive lists of common *ESRP1* AS events and target genes were obtained when the SW480 and HCT116 cell lines were individually compared with the lung cancer study (***Supplementary file 1B-C***). As for the AS targets of RBPs other than *ESRP1*, based on the available RNAseq data from KD studies of *ESRP2* (in the LNCaP cell line ***Nieto et al., 2016***), *RBM47* (H358 ***Yang et al., 2016***), and *QKI* (CAL27; GEO Accession: GSM4677985), several common and unique genes were found (***Figure 2—figure supplement 1*** and ***Supplementary file 2***). Notably, four EMT-related genes (*CTNND1* ***Hernández-Martínez et al., 2019***, LSR ***Shimada et al., 2021***, *SLK* ***Conway et al., 2017***, and *TCF7L2* ***Karve et al., 2020***) were common to all RBP KD studies analyzed (***Figure 2—figure supplement 1***).

The CD44s and NUMB2/4 ESRP1-specific AS isoforms are preferentially expressed in EpCAM^{lo} colon cancer cells

From the newly generated lists of RBP-specific AS targets, we selected *CD44* and *NUMB* for further analysis, based both on their *ESRP1*-specific AS patterns and on their well-established roles in EMT, stemness/differentiation, and cancer progression.

CD44, a transmembrane cell surface glycoprotein, has been show to play key roles in inflammatory responses and in cancer metastasis (***Orian-Rousseau, 2015***). The *CD44* gene encompasses 20 exons of which 1–5 and 16–20 are constant and exist in all isoforms. In contrast, exons 6–14, also referred to as variants exons v2-v10, are alternatively spliced and often deregulated in cancer (***Orian-Rousseau, 2015***). The *NUMB* gene and its protein product have been involved in a broad spectrum of cellular phenotypes including cell fate decisions, maintenance of stem cell niches, asymmetric cell division, cell polarity, adhesion, and migration. In cancer, NUMB is a tumor suppressor that regulates, among others, Notch and Hedgehog signaling (***Pece et al., 2011***). The mammalian *NUMB* gene encodes for four isoforms, ranging from 65 to 72 KD, differentially encompassing two key functional domains, that is, the amino-terminal phosphotyrosine-binding domain, and a C-terminal proline-rich region domain (***Pece et al., 2011***).

Based on the above ΔPSI-based AS analysis, decreased expression of CD44v (variable) isoforms was observed in EpCAM^{lo} and *ESRP1*-KD cells, accompanied by increased CD44s (standard) isoform

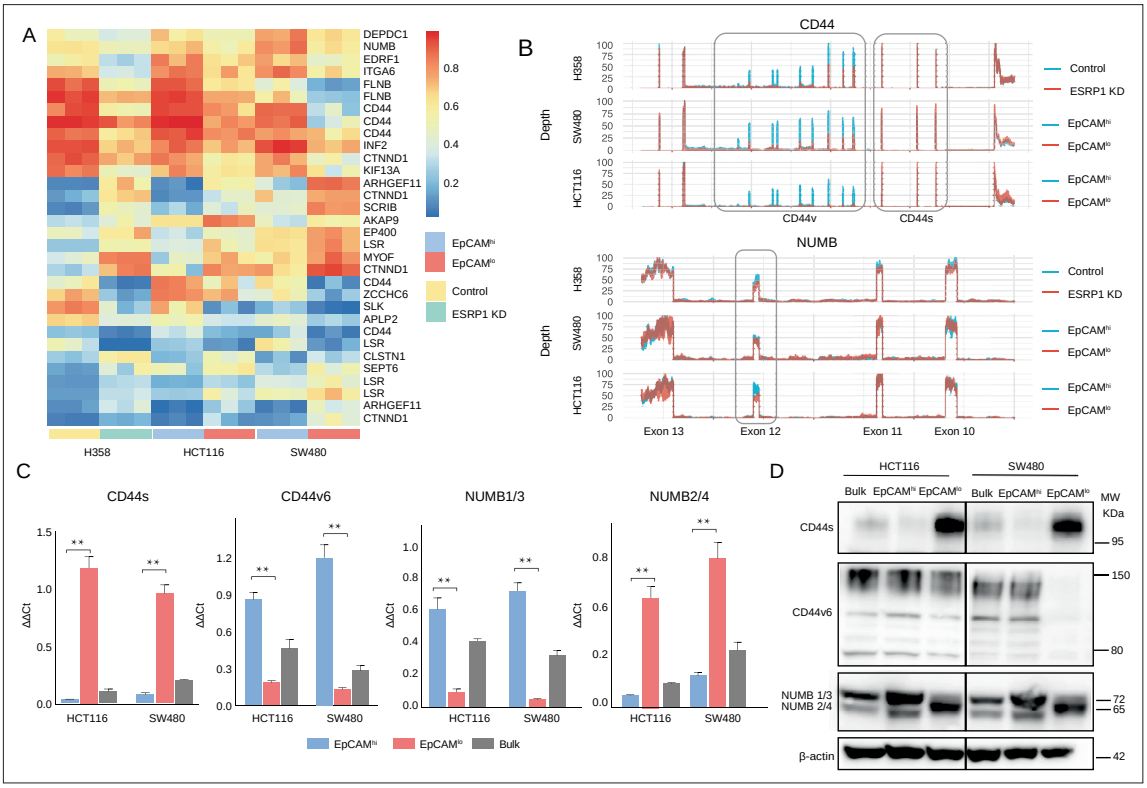


Figure 2. *ESRP1* downregulation in EpCAM^{lo} colon cancer cells affects alternative splicing (AS) of *CD44* and *NUMB* among a broad spectrum of downstream target genes. (**A**) Heatmap of common AS events between RNAseq data from a previous *ESRP1*-KD study in human non-small cell lung cancer cells (H358) (***Yang et al., 2016***) and our own HCT116 and SW480 EpCAM^{hi} and EpCAM^{lo} RNAseq data (***Sacchetti et al., 2021***). The gene list on the right of the heatmap encompasses AS variants earmarked by ΔPSI (differential percentage spliced in) > 0.1. (**B**) *CD44* and *NUMB* exon peak plots relative to the AS analysis of the RNAseq data obtained from a previous *ESRP1*-KD study in human non-small cell lung cancer cells (H358; upper graph) (***Yang et al., 2016***) and from our own HCT116 (middle graph) and SW480 (lower graph) EpCAM$^{hi/lo}$ analysis (***Sacchetti et al., 2021***). Each peak plot depicts the expression of specific exons; the height of each peak is indicative of the expression level of the specific exons. CD44v: CD44 exons v2 to v10. CD44v and CD44s, and NUMB exon 12 is highlighted by gray rectangles. (**C**) RT-qPCR expression analysis of *CD44*s, *CD44v6*, *NUMB1/3*, and *NUMB2/4* isoforms in HCT116 and SW480 EpCAM^{hi}, EpCAM^{lo}, and bulk subpopulations. Expression of the constitutive *CD44* and *NUMB* exons was employed to normalize the results (means ± SEM, n=3). **=p < 0.01. (**D**) Western analysis of CD44s, CD44v6, and NUMB isoforms in HCT116 and SW480 EpCAM^{hi}, EpCAM^{lo}, and bulk subpopulations. Please note that the molecular weight of CD44v6 is expected to range between 80 and 150 kDa (***Azevedo et al., 2018***, ***Ponta et al., 2003***). β-Actin was used as loading control.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Original files and labelled bands of western blots in ***Figure 2D***.

Figure supplement 1. *ESRP1/2-*, *RBM47-*, and *QKI*-regulated alternative splicing (AS) targets.

expression (***Figure 2B***). Likewise, the NUMB2/4 isoforms appear to be preferentially expressed in EpCAM^{lo} and *ESRP1*-KD, accompanied by decreased NUMB1/3 expression (***Figure 2B***, ***Figure 2—figure supplement 1B***). RT-qPCR and western analyses validated these in silico data: CD44s and NUMB2/4 isoforms were preferentially expressed in EpCAM^{lo} colon cancer cells, in contrast with the increased CD44v and NUMB1/3 levels in EpCAM^{hi} cells (***Figure 2C–D***). In view of its previously suggested role in invasion and metastasis (***Todaro et al., 2014***), we focused on the CD44v6 isoform.

As reported above, AS events at the *NUMB* and *CD44* genes correlate with decreased ESRP1 expression. To confirm this observation, we up- and downregulated *ESRP1* in the SW480 and HCT116

cell lines. The dox-inducible shRNA vector used for the KD studies reduces ESRP1 expression by 5- to 10-fold (***Figure 1D–E***) and resulted in the upregulation of the CD44s and NUMB2/4 isoforms at the mRNA and protein level in both cell lines (***Figure 3A–B*** and ***Figure 3—figure supplement 1A–B***). Likewise, *ESRP1* overexpression led to an increase in the CD44v6 and NUMB1/3 isoforms, found in association with the bulk of epithelial colon cancer cells (***Figure 3C–D*** and ***Figure 3—figure supplement 1C–D***).

Transcriptional and functional consequences of the CD44s and NUMB2/4 isoforms on colon cancer invasion and metastasis

In order to elucidate the functional contribution exerted by the newly identified CD44s and NUMB2/4 isoforms on the overall invasive and metastatic capacities of colon cancer cells, we first ectopically expressed each of them (individually and in combination for NUMB1/3 and 2/4) in the HCT116 and SW480 cell lines (***Figure 3—figure supplement 1E–H***), and analyzed their consequences in vitro by cell proliferation, transwell migration assay, RT-qPCR, western, FACS, and RNAseq, and in vivo by spleen transplantation. A significant increase in migratory capacity (***Figure 3—figure supplement 2A–B***), comparable to that of EpCAM^{lo} cells sorted from the parental lines, was observed in SW480 and HCT116 upon overexpression of the CD44s and NUMB2/4 isoforms (***Figure 3—figure supplement 2A–B***). Likewise, ectopic expression of the single NUMB2 or -4 isoforms resulted in increased migration rates when compared with NUMB1 and -3. In contrast, overexpression of CD44v6 and NUMB1/3, normally prevalent in the epithelial bulk (EpCAM^{hi}) of both cell lines, did not affect their migratory properties (***Figure 3—figure supplement 2A–B***).

In agreement with the migration assays, overexpression of CD44s and NUMB2/4 results in the significant upregulation of the EMT transcription factors (EMT-TFs) *ZEB1*, accompanied by the up- and downregulation regulation of mesenchymal and epithelial markers such as *VIM* (vimentin), *CDH1* (E-cadherin), and *EpCAM*, respectively (***Figure 3—figure supplement 2C***). Of note, expression of *ESRP1*, the main upstream splicing regulator of both CD44 and NUMB, was also decreased in CD44s- and NUMB2/4-OE cells, in confirmation of the self-enforcing feedback loop that characterize its interaction with ZEB1 and EMT activation (***Preca et al., 2015***). In agreement with the well-established regulation of Notch signaling by NUMB isoforms (***Pece et al., 2011***), established Notch target genes and were accordingly up- (*HES1*, *HEY1*) and downregulated (*ID2*) upon overexpression of NUMB2/4 (***Figure 3—figure supplement 2D***).

FACS analysis was then employed to evaluate the overall effect of the ectopic expression of the specific CD44 and NUMB isoforms on the relative percentages of the EpCAM$^{hi/lo}$ subpopulations in the HCT116 and SW480 cell lines. As shown in ***Figure 4A***, CD44s overexpression led to a dramatic increase of the EpCAM^{lo} subpopulation at the expenses of EpCAM^{hi} cells. The opposite effect was observed with CD44v6, that is, the enlargement of the EpCAM^{hi} gate and the corresponding decrease of EpCAM^{lo} cells. As for NUMB, ectopic expression of NUMB2/4 significantly increased the relative proportion of EpCAM^{lo} cells while reducing the size of the EpCAM^{hi} subpopulation, while the opposite was observed with NUMB1/3 (***Figure 4B–C***). Of note, the single NUMB2 and NUMB4 isoforms appear dominant in their capacity to enlarge the HCT116 and SW480 EpCAM^{lo} subpopulations, respectively. The same was true for NUMB1 and NUMB3 in the consequences of their ectopic expression in reducing the size of the HCT116 and SW480 EpCAM^{lo} fractions, respectively (***Figure 4B–C***). In agreement with the RTqPCR analysis of EMT markers, CD44s overexpression negatively affected overall proliferation rates in both cell lines, whereas the opposite was observed upon CD44v6 expression (***Figure 4—figure supplement 1A–B***). Likewise, NUMB1/3 expression positively affected proliferation rates in HCT116 and SW480, whereas the NUMB2/4 isoforms exert the opposite effects. In both cases, synergistic effects were observed upon co-expression of NUMB1/3 and 2/4, when compared to the individual isoforms (***Figure 4—figure supplement 1C–D***).

In order to assess the in vivo consequences of the ectopic expression of the CD44 and NUMB isoforms on the capacity of colon cancer cells to form metastatic lesions in the liver, parental HCT116 and SW480 cells and their CD44s-, CD44v6-, NUMB1/3-, and NUMB1/4-overexpressing (OE) counterparts were injected in the spleen of immune-incompetent recipient mice. In agreement with the in vitro results, overexpression of both NUMB2/4 and CD44s isoforms significantly increased the multiplicity of liver metastases, whereas CD44v6 and NUMB1/3 did not differ from the parental controls (***Figure 4D–E***).

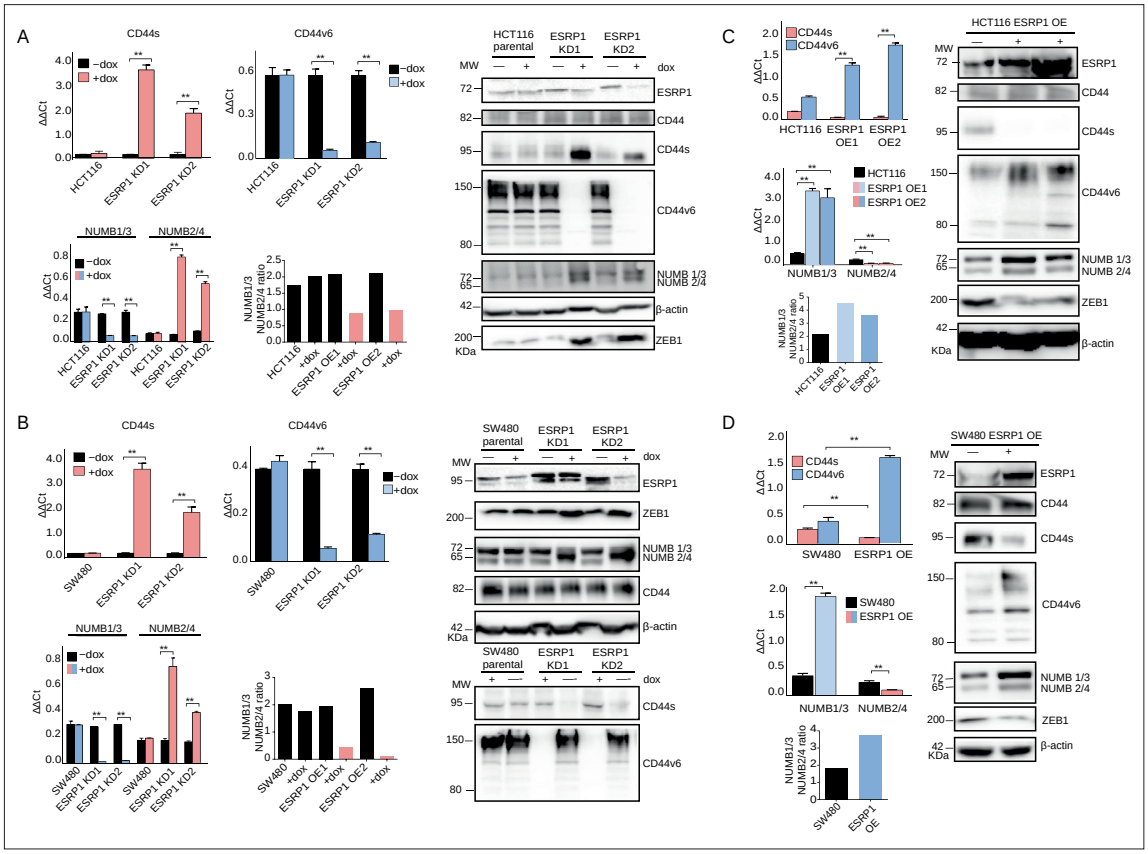


Figure 3. *ESRP1* differential expression regulates *CD44* and *NUMB* alternative splicing (AS) isoforms expression. (**A**) RT-qPCR (left histogram panels) and western (right panel) analysis of CD44 and NUMB isoforms expression in *ESRP1*-KD (sh*ESRP1*-transduced) HCT116 cells. Two independent HCT116 *ESRP1*-KD clones were employed. Cells were induced with 1 µg/mL doxycycline for 72 hr before analysis. Expression of the constitutive *CD44* and *NUMB* exons was employed to normalize the results (means ± SEM, n=3). **=p < 0.01. The ratio of NUMB1/3 and NUMB2/4 bands was quantified by ImageJ and shown in bar plot. Please note that the molecular weight of CD44v6 is expected to range between 80 and 150 kDa (***Azevedo et al., 2018***, ***Ponta et al., 2003***). β-Actin was used as loading control for western blots. (**B**) RT-qPCR (left histogram panels) and western (right panel) analysis of CD44 and NUMB isoforms expression in *ESRP1*-KD (sh*ESRP1*-transduced) SW480 cells. Two independent SW480 *ESRP1*-KD clones were employed. Cells were induced with 1 µg/mL doxycycline for 72 hr before analysis. Expression of the constitutive *CD44* and *NUMB* exons was employed to normalize the results (means ± SEM, n=3). **=p < 0.01. The ratio of NUMB1/3 and NUMB2/4 bands was quantified by ImageJ and shown in bar plot. Please note that the molecular weight of CD44v6 is expected to range between 80 and 150 kDa (***Azevedo et al., 2018***, ***Ponta et al., 2003***). β-Actin was used as loading control for western blots. (**C**) RT-qPCR (left histogram panels) and western (right panel) analysis of CD44 and NUMB isoforms expression in *ESRP1*-OE HCT116 cells. Two independent HCT116 *ESRP1*-OE clones were employed. Expression of the constitutive *CD44* and *NUMB* exons was employed to normalize the results (means ± SEM, n=3). **=p < 0.01. The ratio of NUMB1/3 and NUMB2/4 bands was quantified by ImageJ and shown in bar plot. Please note that the molecular weight of CD44v6 is expected to range between 80 and 150 kDa (***Azevedo et al., 2018***, ***Ponta et al., 2003***). β-Actin was used as loading control for western blots. (**D**) RT-qPCR (left histogram panels) and western (right panel) analysis of CD44 and NUMB isoforms expression in *ESRP1*-OE SW480 cells. Expression of the constitutive *CD44* and *NUMB* exons was employed to normalize the results (means ± SEM, n=3). **=p < 0.01. The ratio of NUMB1/3 and NUMB2/4 bands was quantified by ImageJ and shown in bar plot. Please note that the molecular weight of CD44v6 is expected to range between 80 and 150 kDa (***Azevedo et al., 2018***, ***Ponta et al., 2003***). β-Actin was used as loading control for western blots.

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. Original files and labelled bands of western blots in ***Figure 3A***.

Source data 2. Original files and labelled bands of western blots in ***Figure 3B***.

Figure 3 continued on next page

Figure 3 continued

Source data 3. Original files and labelled bands of western blots in ***Figure 3C***.

Source data 4. Original files and labelled bands of western blots in ***Figure 3D***.

Figure supplement 1. *ESRP1*, *CD44*, and *NUMB* isoforms analysis in overexpressing and knockdown (KD) colon cancer cell lines.

Figure supplement 1—source data 1. Original files and labelled bands of PCR gels in ***Figure 3—figure supplement 1A***.

Figure supplement 1—source data 2. Original files and labelled bands of PCR gels in ***Figure 3—figure supplement 1B***.

Figure supplement 1—source data 3. Original files and labelled bands of PCR gels in ***Figure 3—figure supplement 1C***.

Figure supplement 1—source data 4. Original files and labelled bands of PCR gels in ***Figure 3—figure supplement 1***.

Figure supplement 1—source data 5. Original files and labelled bands of western blots in ***Figure 3—figure supplement 1***.

Figure supplement 1—source data 6. Original files and labelled bands of western blots in ***Figure 3—figure supplement 1***.

Figure supplement 2. *CD44* and *NUMB* isoform-specific expression affects cell migration and Notch signaling activation.

Next, in order to elucidate the signaling pathways and molecular and cellular mechanisms triggered by the CD44 isoforms, we analyzed by RNAseq HCT116 and SW480 cells ectopically expressing CD44s and CD44v6. After dimension reduction with principal component analysis (PCA), the samples separated by group (i.e. CD44s-OE, CD44v6-OE, and controls) (***Figure 5A***). Notably, the CD44s-OE samples showed most distinct expression in both cell lines when compared to the parental and CD44v6-OE cell lines. In HCT116, the CD44v6 samples shared most similarity with the CD44s samples, while in SW480, the CD44v6 samples were most similar to the parental cell line. Thus, we observed both an isoform-independent effect, presumably as the result of the ectopic CD44 expression (and most dominantly visible in HCT116), and an isoform-dependent effect as depicted by the separation of CD44s and CD44v6 samples (***Figure 5A***). As expected, differential expression analysis of the CD44s and v6 isoforms overexpressing samples compared with the parental cell lines revealed an overall upregulation of gene expression (***Figure 5—figure supplement 1A***). Next, in order to identify which genes are specifically upregulated by the different CD44 isoforms, we performed differential expression analysis between the CD44s samples and the CD44v6 samples. To this aim, we employed k-means clustering on the scaled expression values to separate genes specific for the CD44s isoform (e.g. *SPARC, ZEB1, VIM*), the CD44v6 isoform (e.g. *IL32, TACSTD2, CSF2*), and genes that were indiscriminative for the CD44v6 isoform or the parental cell lines (e.g. *MAL2, ESRP1, CDH1*) (***Figure 5B***). Finally, to identify the most distinct differences in signaling pathways and GO functional categories, we performed a gene set enrichment analysis (GSEA) by comparing the CD44s- with the CD44v6-OE samples in the individual cell lines. Among the significantly altered pathways (normalized enrichment score [NES] >1, pval <0.05), EMT was the only one upregulated in CD44s vs. CD44v6 in both cell lines (***Figure 5C–D***). Additional pathways and GO categories activated by CD44s appeared to be cell line specific, for example, Wnt β-catenin signaling (HCT116) and oxidative phosphorylation (SW480). Of note, the detailed GSEA evidenced how several inflammatory (TNF/NFκB; IL6/JAK/STAT3; IFα/γ; ILK2/STAT5) and signaling (KRAS, MYC, E2F) pathways were common to both CD44s and v6, presumably as the result of the ectopic CD44 expression, regardless of the isoform (***Figure 5—figure supplement 1B***).

Increased ZEB1 and decreased ESRP1 expression correlate with the NUMB2/4 and CD44s isoforms and with poor overall survival

In order to assess the clinical relevance of the results obtained with the SW480 and HCT116 cell lines, we analyzed RNAseq data from patient-derived colon cancers available from the public domain and the scientific literature. To this aim, the TCGA Splicing Variants Database (TSVdb; http://www.tsvdb.com/) was employed to integrate clinical follow-up data with RBP and AS expression profiles obtained from The Cancer Genome Atlas (TCGA) project and from the ***Guinney et al., 2015***, study on the classification of human colon cancers into four consensus molecular subtypes (CMS1-4). The main limitation of this approach is the low representation of quasi-mesenchymal (EpCAM^{lo}-like) subpopulations in bulk RNAseq preparations and the masking effect that the majority of epithelial (EpCAM^{hi}-like) cancer cells are likely to cause. To identify tumors enriched in EpCAM^{lo}-like cells, we first stratified them based on *ZEB1* expression (*ZEB1* >8.6: ZEB1^{hi}; ZEB1 <8.3. ZEB1^{lo}; 8.2<ZEB1<8.6: Intermediate). Subsequently, we used *ESPR1* expression levels to further define the tumors into *ZEB1*^{hi}*ESRP1*^{lo}

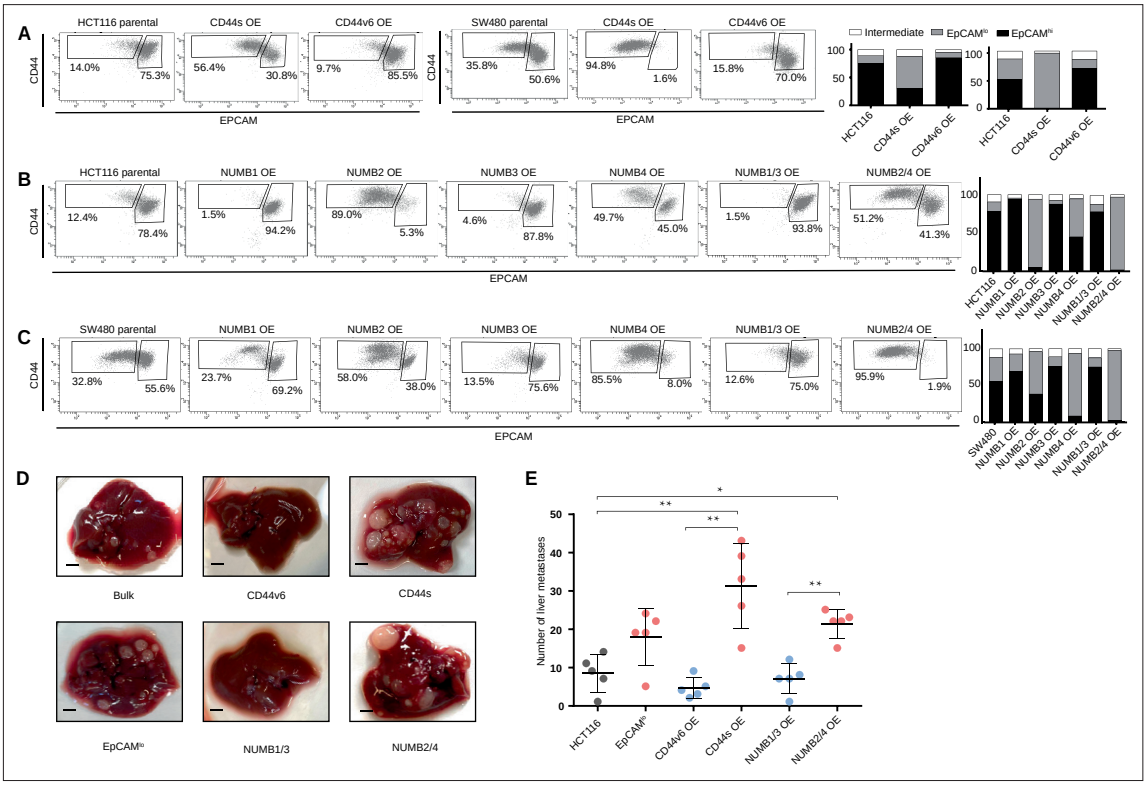


Figure 4. *CD44* and *NUMB* alternative splicing (AS) isoforms have opposite functions in quasi-mesenchymal and epithelial colon cancer cells and their capacity to metastasize the liver. (**A**) CD44/EpCAM FACS analysis of EpCAM^{lo} and EpCAM^{hi} subpopulations in CD44s-OE (left) and CD44v6-OE HCT116 and SW480 cell lines. The bar charts on the right depict the percentages of EpCAM^{lo} and EpCAM^{hi} cells. The subpopulation of cells mapping in between, but yet outside, the CD44^{hi}EpCAM^{hi} and CD44^{hi}EpCAM^{lo} gates, is here labelled as 'intermediate'. (**B**) and (**C**) CD44/EpCAM FACS analysis of EpCAM^{lo} and EpCAM^{hi} subpopulations in NUMB1-4 OE HCT116 and SW480 cells. The bar charts on the right depict the percentages of EpCAM^{lo} and EpCAM^{hi} cells. (**D**) Macroscopic images of livers from mice spleen-injected with CD44s-, CD44v6-, NUMB2/4-, and NUMB1/3-OE HCT116 cells. HCT116 EpCAM^{lo} and bulk cells were used as positive control. Scale bar: 5 mm. (**E**) Liver metastasis multiplicity after intrasplenic injection of CD44s-, CD44v6-, NUMB2/4-, and NUMB1/3-OE HCT116 cells. For each transplantation experiment, 5×10^4 cells were injected in the spleen of recipient NSG mouse. Six weeks after injection, mice were sacrificed and individual tumors counted. (means ± SEM, n=5) *=p < 0.05; **=p < 0.01.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. CD44 and NUMB isoforms regulate colon cancer cell proliferation.

(*ESRP1* <11.8; hereafter referred to as *ZEB1*^{hi}), *ZEB1*^{lo}*ESRP1*^{hi} (*ESRP1* >11.6; hereafter referred to as *ZEB1*^{lo}). Tumors with intermediate *ZEB1* expression levels and tumors with *ESRP1* expression levels outside these thresholds were defined as intermediate (***Figure 6A***). Kaplan-Meier analysis showed that *ZEB1*^{hi} tumors have an overall decreased survival probability (p=0.045) (***Figure 6B***). Next, we compared the expression of CD44 and NUMB isoforms across the *ZEB1*$^{hi/lo}$ tumors. Notably, while no significant differences were observed based on the expression level of the whole CD44 and NUMB genes, significant differences were found for their specific isoforms (***Figure 6C***). Analysis of the specific isoforms expression across the different CMS (***Guinney et al., 2015***) revealed elevated CD44s and NUMB2/4 expression in the CMS4 subtype, known to be enriched in mesenchymal lineages in tumor and TME cells, and strongly associated with poor survival and the greatest propensity to form distant metastases (***Figure 6D***). Likewise, the majority of the *ZEB1*^{hi} group was composed of the CMS4

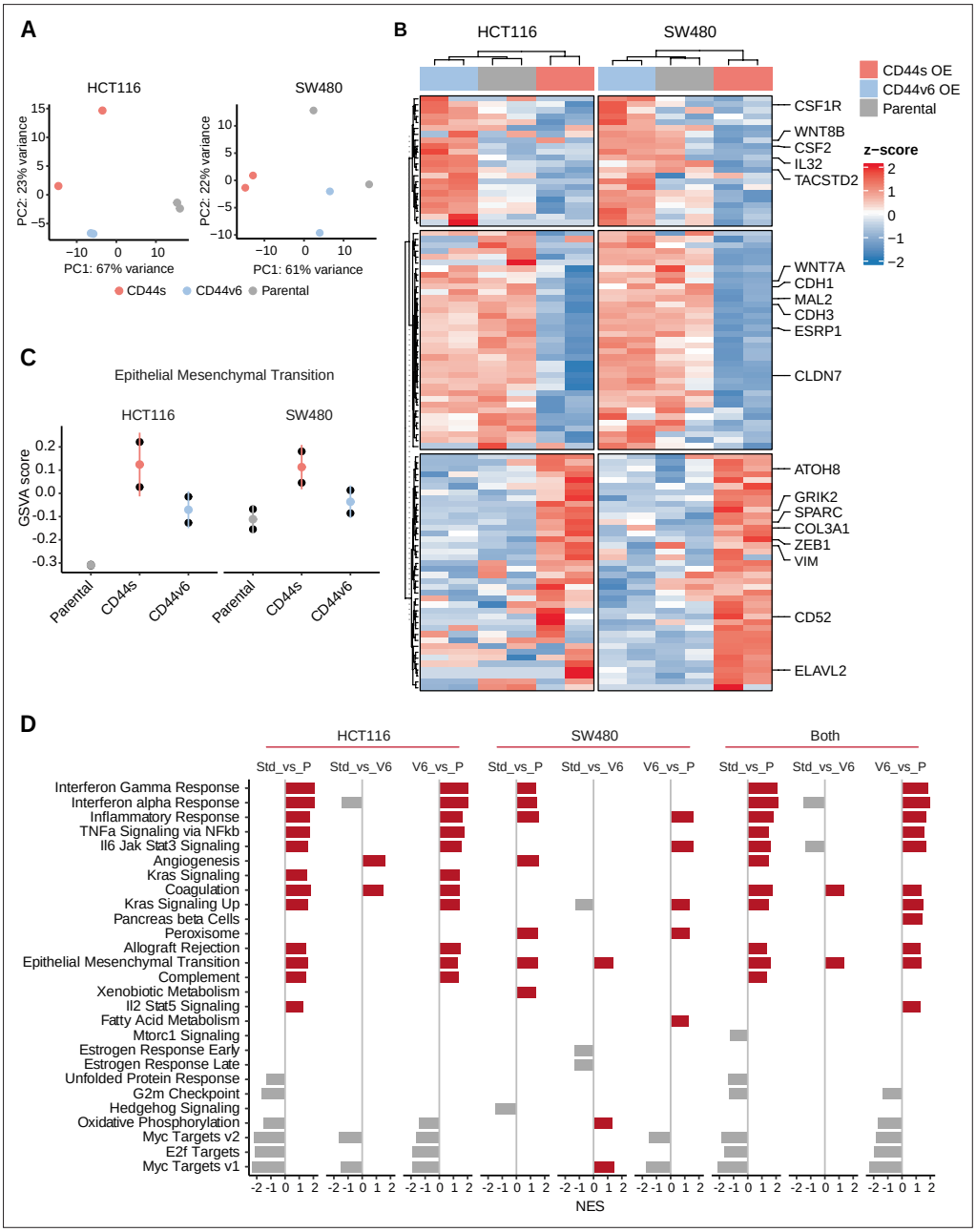


Figure 5. RNAseq analysis of CD44s- and CD44v6-expressing colon cancer cells reveals a broad spectrum of downstream alternative splicing (AS) targets and biological functions. (**A**) Principal component analysis (PCA) of RNAseq profiles from CD44s- and CD44v6-OE HCT116 and SW480 cell lines. (**B**) Heatmap of differentially expressed gene among HCT116 and SW480 CD44s-OE, CD44v6-OE, and parental cells. (**C**) Gene set enrichment analysis (GSEA) of epithelial-mesenchymal transition (EMT) in expression profiles from HCT116 and SW480 parental, CD44s-OE, and CD44v6-OE cells.

Figure 5 continued on next page

Figure 5 continued

Normalized enrichment score (NES) >1, and pval <0.05. (**D**) GSEA of HCT116 and SW480 expression profiles in parental, CD44s-OE, CD44v6-OE cells compared with each other. Plots show only significantly altered pathways, with NES >1, and pval <0.05.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Gene enrichment and pathway analysis of CD44s- and CD44v6-overexpressing (OE) colon cancer cells.

subtype (72%), while the *ZEB1*^{lo} group was mainly contributed by CMS2 (49%) and CMS3 tumors (31%), with few CMS4 tumors (1%) (***Figure 6E***).

Next, we correlated the expression of CD44s/v6 isoforms in patient-derived colon tumors with the DEGs identified in the isoform-overexpressing cell lines (***Figure 7A***). While overall *CD44* expression correlated with both isoforms, the DEGs from the CD44s-OE samples showed specific correlation with CD44s expression in patient-derived tumors (e.g. *SPARC, ZEB1*), the DEGs from the CD44v6 samples correlated with CD44v6 but not with CD44s (e.g. *KDF1, ESRP1*).

Last, we correlated the CD44 and NUMB isoforms expression in patient-derived colon cancers with functional signatures obtained by averaging the scaled expression levels for each of the hallmark sets (***Liberzon et al., 2015***). The CD44s and NUMB2/4 isoforms showed overall similar correlating hallmarks and pathways. However, the same was not true when compared to the CD44v6- and NUMB1/3-associated functional signatures. Here, most invasion/metastasis-relevant hallmarks (e.g.

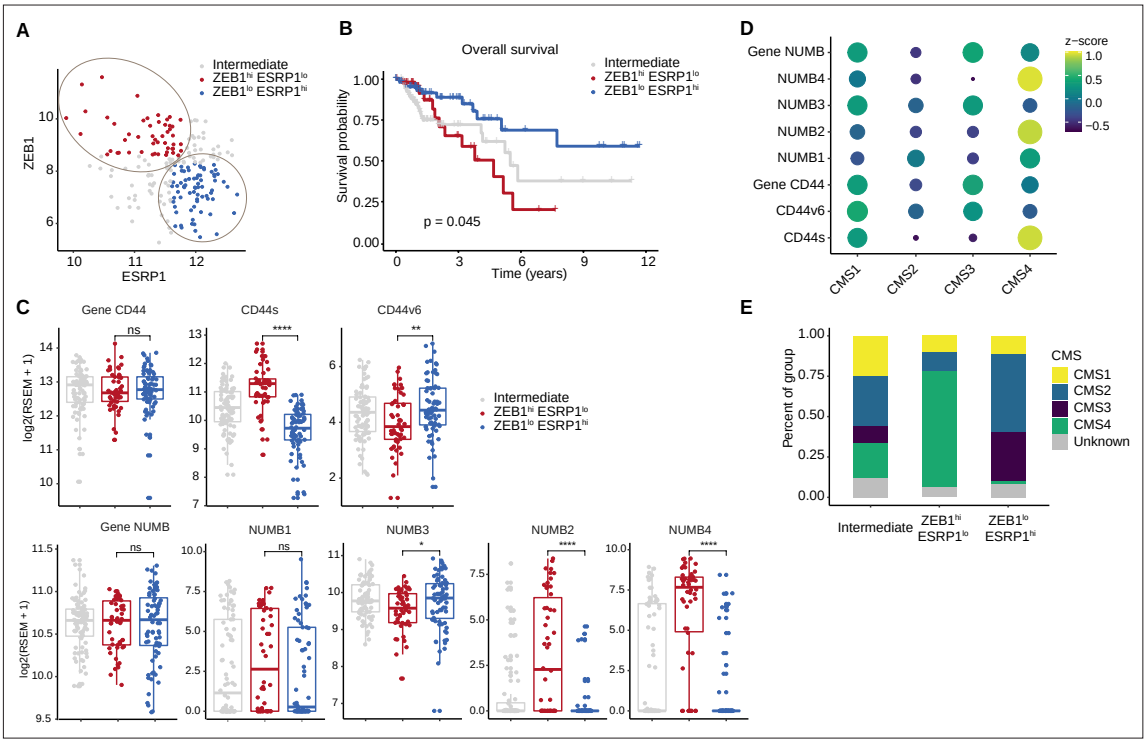


Figure 6. Increased *ZEB1* and decreased *ESRP1* expression correlate with the NUMB2/4 and CD44s isoforms and with poor overall survival. (**A**) RNAseq data from The Cancer Genome Atlas (TCGA) were subdivided into three groups based on *ZEB1* and *ESRP1* expression level: *ZEB1*^{hi}*ESRP1*^{lo} (*ZEB1*^{hi}, red dots), *ZEB1*^{lo}*ESRP1*^{hi} (*ZEB1*^{lo}, blue dots), and intermediate (gray dots). (**B**) Kaplan-Meier analysis of overall survival in the *ZEB1*^{hi}*ESRP1*^{hi} and *ZEB1*^{lo}*ESRP1*^{lo} patient groups. (**C**) Box plots showing CD44 and NUMB gene and isoforms expression across the *ZEB1*^{hi}*ESRP1*^{lo}, *ZEB1*^{lo}*ESRP1*^{hi}, and intermediate patient groups. (**D**) Dot plot analysis of the z-score scaled expression values of CD44s, CD44v6, NUMB1-4 isoforms across the four colon cancer consensus molecular subtypes (CMS). (**E**) Stacked bar plot showing the composition of the CMS across the *ZEB1*$^{hi/lo}$ and intermediate patient groups.

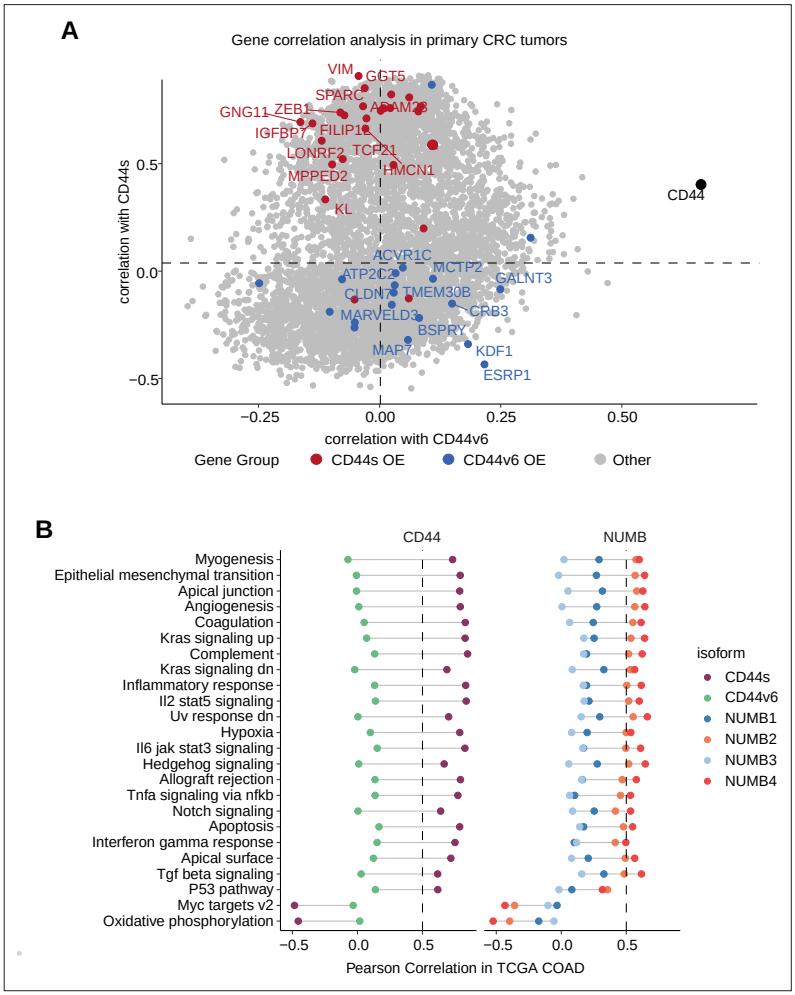


Figure 7. Gene and pathway correlation analyses of CD44 and NUMB isoforms in patient-derived colon cancers. (**A**) Gene correlation analysis showing the correlation of gene expression with CD44s and CD44v6 isoform expression in the The Cancer Genome Atlas (TCGA) patient cohort. Differentially expressed genes from CD44s- (red) and CD44v6-OE (blue) RNAseq data are highlighted. (**B**) Pathway correlation analysis showing the correlation of pathway activity CD44 and NUMB isoform expression in the TCGA patient cohort.

The online version of this article includes the following figure supplement(s) for figure 7:

Figure supplement 1. *CD44* and *NUMB* isoforms expression in EpCAM$^{hi/lo}$ ovarian and cervical cancer cells.

EMT, angiogenesis, apical junctions) showed a positive correlation with CD44s and NUMB2/4, though not with CD44v6 and NUMB1/3 (***Figure 7B***).

In sum, we confirmed a switch in isoform expression (CD44v6 vs. CD44s and NUMB1/3 vs. NUMB2/4) as a function of *ESRP1* and *ZEB1* expression in colon cancer. Expression of the EpCAM^{lo}-specific isoforms (CD44s and NUMB2/4) is elevated in CMS4 tumors overall survival.

Upregulation of the NUMB2/4 and CD44s isoforms is common to quasi-mesenchymal cells from cancers other than colon

In order to assess whether the preferential expression of the NUMB2/4 and CD44s isoforms is specific to the modalities of local invasion and distant metastasis characteristic of colon cancer, we interrogated expression profiling data previously obtained by comparing epithelial and quasi-mesenchymal subpopulations from ovarian (OV90) and cervical (SKOV6) cancer cell lines (*manuscript in preparation*). Ovarian cancer, because of the distinct anatomical localization of the primary lesion, metastasizes the abdominal cavity with very different modalities than colon cancer, namely by peritoneal dissemination rather than local dissemination into the stroma microenvironment followed by intra- and extravasation of the portal blood stream (***Adam and Adam, 2004***; ***Goswami et al., 2009***). On the other hand, metastasis in carcinoma of the cervix occurs both by lymphatic and by hematogenous spread to the lung, liver, and bones. We asked whether, notwithstanding the distinctive patterns of metastatic spread, the CD44s and NUMB2/4 isoforms were preferentially expressed in the corresponding EpCAM^{lo} RNAseq profiles. To this aim, EpCAM$^{hi/lo}$ subpopulations from OV90 and SKOV6 were sorted and analyzed by RNAseq and RT-qPCR, similar to our previous study on colon cancer (***Sacchetti et al., 2021***). As shown in ***Figure 7—figure supplement 1***, both NUMB2/4 and CD44s isoforms appear to be upregulated in the OV90 and SKOV6 cell lines, as also validated by RT-qPCR.

Discussion

The capacity to invade the tumor microenvironment and to form distant metastases undoubtedly represents the most clinically relevant hallmark of epithelial cancer cells. However, the complexity and diversity of the obstacles that carcinoma cells encounter along the invasion-metastasis cascade require transient and reversible changes that cannot be explained by the de novo acquisition of genetic alterations. Instead, epigenetic (non-mutational) modifications underlie phenotypic plasticity, that is, the capacity of cancer cells with a given genotype to acquire more than one phenotype in a context-dependent fashion (***Varga and Greten, 2017***). EMT and MET are central to the phenotypic plasticity characteristic of metastasizing carcinoma cells and are prompted by a broad spectrum of epigenetic mechanisms ranging from chromatin remodeling by histone modifications, DNA promoter methylation, non-coding RNAs, and AS (***Dixit et al., 2016***). Here, we have taken advantage of our previous identification of phenotypic plastic and highly metastatic EpCAM^{lo} colon cancer cells (***Sacchetti et al., 2021***) to characterize the genome-wide AS events that accompany EMT/MET state transitions between the epithelial bulk (EpCAM^{hi}) and the quasi-mesenchymal subpopulation.

In view of the central role played by RBPs in eliciting AS, we first identified RBP-coding genes differentially expressed between the EpCAM^{lo} and EpCAM^{hi} fractions of two commonly employed colon cancer cell lines, representative of the chromosomal- and microsatellite-instable subtypes (SW480, CIN; HCT116, MIN) (***Lengauer et al., 1997***). The *ESRP1/2* genes (***Warzecha et al., 2009***), the '*splicing masterminds*' of EMT (***Tavanez and Valcárcel, 2010***; ***Warzecha et al., 2010***), were found among the top downregulated RBP-coding genes in EpCAM^{lo} colon cancer cells, as part of a self-enforcing feedback loop with the EMT-TF *ZEB1* (***Preca et al., 2015***). Accordingly, *ZEB1* upregulation in EpCAM^{lo} colon cancer cells is invariably accompanied by *ESRP1/2* downregulation, and *ZEB1*^{hi}/*ESRP1*^{lo} colon cancers, predominantly belonging to the mesenchymal CMS4 subgroup, have a significantly worse survival outcome when compared with *ZEB1*^{lo}/*ESRP1*^{hi} patients.

Apart from *ESRP1*, several other RBP-coding genes were found to be differentially expressed between epithelial and quasi-mesenchymal colon cancer cells. Whereas the majority of RBP-coding DEGs, like *ESRP1*, appear to be downregulated upon EMT induction (*ESRP1/2*, *RBM14/19/47*, *MBNL3*, *HNRPAB/PF*, *USAF2*), others were activated in the quasi-mesenchymal EpCAM^{lo} fraction (*NOVA2*, *MBNL2*, *QKI*, *SRSF5*, *HNRNPH*, *RBM24/43*). Accordingly, in patient-derived colon cancers stratified according to their consensus molecular signature, the same *QKI*, *RBM24*, and *MBNL2* genes were found to have increased expression in CMS4 tumors, known for their pronounced mesenchymal composition and poor prognosis (***Guinney et al., 2015***). Of note, the mesenchymal nature of CMS4 tumors has previously been questioned as these lesions often feature pronounced infiltration from the surrounding microenvironment, the extent of which might cover their true cellular identity other than representing a mere contamination from the tumor microenvironment (***Calon et al., 2015***; ***Isella et al., 2015***). As shown in our previous study (***Sacchetti et al., 2021***), the EpCAM^{lo} cells do represent

bona fide quasi-mesenchymal colon cancer cells, enriched among CMS4 cases, and likely responsible for their poor prognosis. The observed upregulation of RBPs such as quaking (*QKI*) is caused by the presence in its 3'UTR of target sequences of the miR-200 family of microRNAs (**Pillman et al., 2018**; **Kim et al., 2019**). The latter is analogous to the regulation of the expression of the EMT-TF *ZEB1* gene, whose activation during EMT is regulated by the same microRNA family (**Brabletz and Brabletz, 2010**). Accordingly, the significantly reduced levels of all five miR-200 members in EpCAM^{lo} cells (**Sacchetti et al., 2021**) underly the coordinated upregulation of both *ZEB1* and *QKI*.

The here observed *RBM47* downregulation in CMS4 colon cancers is in agreement with a previous report on its decreased protein expression during EMT in association with metastasis in a cohort of primary CRCs (**Rokavec et al., 2017**). On the other hand, the increased expression of other RBP-coding genes such as *RBM24* and *MBNL2* (muscleblind-like 2) in CMS4 tumors and in EpCAM^{lo} cells is in sharp contradiction with their alleged tumor suppressing roles in colon and other cancers (**Xia et al., 2021**; **Lin et al., 2021**). Of note, MBNL2 regulates cancer migration and invasion through PI3K/AKT-mediated EMT (**Lin et al., 2021**) and its overexpression in breast and cancer cell lines inhibits their metastatic potential (**Zhang et al., 2019b**). In contrast to *MBNL2*, *MBNL3*, a distinct member of the muscleblind family, is downregulated in EpCAM^{lo} colon cancer cells, similar to what reported in prostate cancer by **Lu et al., 2015**. *NOVA2*, a member of the Nova family of neuron-specific RBPs, was also upregulated in the quasi-mesenchymal cells from both cell lines, possibly as the result of the differential expression of miR-7-5p (**Xiao, 2019**), as previously shown in non-small cell lung (**Xiao, 2019**) and prostate (**Lu et al., 2015**) cancer. The identification the AS targets downstream of specific RBPs in quasi-mesenchymal cancer cells from different malignancies will likely clarify these apparent contradictions and shed light on the functional roles of distinct members of the splicing machinery in EMT and metastasis.

The spectrum of AS target genes downstream of the RBPs differentially expressed in EpCAM^{lo} colon cancer cells appears extremely broad when it comes to specific cellular processes or signaling pathways. Nonetheless, comparison of our RNAseq data with KD studies of specific RBPs from the public domain (*ESRP1/2* **Nieto et al., 2016**, *RBM47* **Yang et al., 2016**, and *QKI*; GEO Accession: GSM4677985) allowed us to identify common and unique AS target genes associated with specific downstream effectors. By following this admittedly imperfect approach, the top four AS targets common to all of the above-mentioned RBPs notwithstanding their up- or downregulation in EpCAM^{lo} colon cancer cells, that is, *CTNND1* (δ- or p120-catenin), *LSR* (lipolysis stimulated lipoprotein receptor), *SLK* (STE20-like kinase), and *TCF7L2* (transcription factor 7-like 2, or TCF4) are known regulators and effectors of EMT (**Hernández-Martínez et al., 2019**; **Shimada et al., 2021**; **Conway et al., 2017**; **Karve et al., 2020**), thus pointing to the central role played by AS in the regulation of EMT in the malignant evolution of colon cancer.

Here, we have focused on CD44 and NUMB as two ESRP1-specific AS target genes with well-established functional roles in EMT and in cancer invasion and metastasis. The CD44s and NUMB2/4 isoforms appear to be specifically expressed in quasi-mesenchymal colon cancer cells both from the immortalized cell lines and from patient-derived tumors, with a striking enrichment in the CMS4 subgroup of colon cancer patients. In contrast, the CD44v6 and NUMB1/3 isoforms are preferentially expressed in the epithelial bulk of the tumor. The latter, as far as CD44v6 is concerned, contrasts what previously reported by **Todaro et al., 2014**, where this specific isoform was found to earmark the colon cancer stem cells (CSCs) which underlie metastasis. CD44v6 and other 'variable' CD44 isoforms (CD44v4-10) earmark *Lgr5*$^{+}$ intestinal stem cells (ISCs), that is, the cells of origin of intestinal tumors, and accordingly promote adenoma formation in vivo (**Zeilstra et al., 2008**; **Zeilstra et al., 2014**; **Misra et al., 2009**). A plausible explanation for the discordant results lies in the epithelial nature of the models employed in the above study and in the requirement of both EMT and MET for the completion of the invasion-metastasis cascade (**Brabletz et al., 2005**). By employing tumor spheres and freshly sorted CD133^{+} tumor cells, Todaro et al. focused on epithelial CSCs where, as observed in normal ISCs, the CD44v6 isoform is predominantly expressed, and is necessary for EMT to occur upon interaction with c-MET (**Todaro et al., 2014**). The CD44v6 isoform is required for c-MET activation by hepatocyte growth factor (HGF, or scatter factor) (**Orian-Rousseau et al., 2002**) and as such plays an essential role in triggering EMT at the invasive front where tumor cells are exposed to these TME-secreted factors. Our own immunoprecipitation studies confirmed that CD44v6 but not CD44s binds to cMET in response to HGF stimulation (*data not shown*). Therefore, HGF/SF stimulation of

colon cancer cells along the invasive front will trigger the acquisition of quasi-mesenchymal characteristics and the AS-driven switch from CD44v6 to CD44s, the latter unable to bind HGF and as such controlling the extension of EMT activation. The reverse switch will take place upon the activation of the MET necessary for the colonization of the distal metastatic site. From this perspective, both CD44 isoforms are essential for the completion of the invasion-metastasis cascade.

The functional relevance of the CD44s isoforms has been highlighted in malignancies other than colon cancer, namely in prostate (**Lu et al., 2015**) and breast cancer where it activates, among others, PDGFRβ/Stat3 and Akt signaling to promote EMT and CSC traits (**Brown et al., 2011**; **Zhang et al., 2019a**). GO analysis of the RNAseq profiles from colon cancer cells ectopically expressing CD44s highlighted a broader spectrum of signaling pathways likely to underlie EMT. Accordingly, analysis of RNAseq data from primary colon cancers stratified for their CD44s expression revealed an equally broad spectrum of downstream EMT-related biological processes. Of note, among the DEGs identified upon CD44s ectopic expression which correlate with *ZEB1*^{hi}/*ESRP1*^{lo} (and CMS4) colon cancers, the *SPARC* gene, a pEMT marker in the EpCAM$^{hi/lo}$ state transitions (**Sacchetti et al., 2021**), was found.

Expression of NUMB2/4 isoforms both in cells lines and in patient-derived colon tumors is associated with signaling pathways and GO categories largely overlapping with those linked to CD44s (and CD44v6 with NUMB1/3), possibly suggesting synergism between AS at these genes. Accordingly, NUMB is involved in a broad spectrum of cellular phenotypes in homeostasis and in cancer where it mainly function as a tumor suppressor (**Pece et al., 2011**). NUMB inhibits EMT by suppressing the Notch signaling pathway. As such, downregulation of NUMB can induce an EMT phenotype in isoform-specific fashion. Analysis of colon cancer cells individually overexpressing each of the four isoforms revealed an increased basal Notch signaling in NUMB2 and -4, as shown by the expression of the 'universal' targets *HES1* and *HEY1*. Instead, ectopic expression of NUMB1/3 resulted in increased transcriptional levels of the more atypical Notch signaling target *ID2*. Although the functional consequences of the NUMB2/4 (and 1/3) isoforms on Notch regulation of EMT are yet unclear, it seems plausible that the complex network of AS targets activated downstream of the RBP-coding DEGs, including CD44, NUMB, and many others as shown here, will eventually lead to the 'just-right' level of plasticity needed to allow both the 'mesenchymalization' during local invasion and systemic dissemination, and the reacquisition of epithelial features at the distant site of metastasis.

Overall, it appears that AS substantially contributes to the epigenetic mechanisms that underlie EMT/MET in cancer metastasis. From this perspective, several aspects of our study are novel: first, the identification of colon cancer-specific AS target genes paralleled by the corresponding RBPs which, when stratified according to the CMS classification of colon cancers, reveal notable differences and consequences on patients' survival. Moreover, the results of the functional analysis of AS at the CD44 gene contrast what previously reported (**Todaro et al., 2014**) and shed new light on the relevance of the standard and v6 isoforms in the migrating CSC model (**Brabletz et al., 2005**). Comparison of the RBP/AS analysis among colon, cervical, and ovarian cancer highlights how, although the majority of AS targets are common to different types of malignancies in RBP-specific fashion, notable differences also exist possibly in reflection of the specific modalities of local dissemination and distal metastasis formation in different cancers. Also, the use of immortalized cell lines for the analysis of epithelial and quasi-mesenchymal tumor cell subpopulations represents an original approach yet based on an 'old-fashioned' laboratory reagent (**Sacchetti et al., 2021**). Finally, the systematic elucidation of the RBPs and AS targets which underlie phenotypic plasticity in different types of cancer will provide novel tumor-specific targets for therapeutic intervention based on small molecule inhibitors and even RNA vaccination.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Cell line (*Homo sapiens*)	HCT116 (adult colorectal carcinoma)	ECACC	Cat# 91091005, RRID:CVCL_0291	

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Cell line (*Homo sapiens*)	SW480 (adult colorectal carcinoma)	ECACC	Cat# 87092801, RRID:CVCL_0546	
Transfected construct (*Homo sapiens*)	Human-ESRP1 shRNA	Horizon	Cat# V3THS_400802	Lentiviral construct to transfect express the shRNA
Antibody	Anti-human ZEB1 (rabbit monoclonal)	Cell Signaling	Cat# 3396, RRID:AB_1904164	WB (1.1000)
Antibody	Anti-human ESRP1 (rabbit polyclonal)	Thermo Fisher	Cat# PA5-11520, RRID:AB__2899836	WB (1.1000)
Antibody	Anti-human CD44s (mouse monoclonal)	Thermo Fisher	Cat# MA5-13890, RRID:AB_10986810	WB (1.100)
Antibody	Anti-human CD44v6 (mouse monoclonal)	Abcam	Cat# ab78960, RRID:AB_1603730	WB (1.1000)
Antibody	Anti-human NUMB (rabbit monoclonal)	Cell Signaling	Cat# 2756, RRID:AB_2534177	WB (1.1000)
Antibody	Anti-human B-actin (rabbit monoclonal)	Cell Signaling	Cat# 8457,R RID:AB_10950489	WB (1.2000)
Antibody	Anti-mouse CD44-APC (rat monoclonal)	BD Pharmingen	Cat# 559250, RRID:AB_398661	FACS (1 µg/10^6 cells)
Antibody	Anti-human EpCAM-FITC (mouse monoclonal)	GeneTex	Cat# GTX30708, RRID:AB_1240769	FACS (1 µg/10^6 cells)
Recombinant DNA reagent	ESRP1 cDNA ORF Clone (human)	Sino Biological	Cat# HG13708-UT	
Recombinant DNA reagent	pcDNA empty vector (plasmid)	Gift from Ron Smits		
Recombinant DNA reagent	pcDNA-human-CD44s (plasmid)	Gift from Véronique Orian-Rousseau		
Recombinant DNA reagent	pUC57-human-CD44v6 (plasmid)	Gift from Véronique Orian-Rousseau		
Recombinant DNA reagent	pcDNA-human-NUMB1 (plasmid)	Gift from Salvatore Pece		
Recombinant DNA reagent	pcDNA-human-NUMB2 (plasmid)	Gift from Salvatore Pece		
Recombinant DNA reagent	pcDNA-human-NUMB3 (plasmid)	Gift from Salvatore Pece		
Recombinant DNA reagent	pcDNA-human-NUMB4 (plasmid)	Gift from Salvatore Pece		
Recombinant DNA reagent	shZEB1	***Sacchetti et al., 2021***	Cat# 1864	
Recombinant DNA reagent	pSLIK-Hygro	Addgene	Cat# 25737	
Software, algorithm	R	Seurat, GSVA, MAGIC (***Stuart et al., 2019***; ***van Dijk et al., 2018***; ***La Manno et al., 2018***)	RRID:SCR_007322, RRID:SCR_021058	Version 4.0.4
Software, algorithm	Python	Velocyto, scVelo (***La Manno et al., 2018***; ***Bergen et al., 2020***)	RRID:SCR_018167, RRID:SCR_018168	Version 3.8.3
Software, algorithm	STAR	***Dobin et al., 2013***	RRID:SCR_004463	
Software, algorithm	MISO	***Katz et al., 2010***.	RRID:SCR_003124	

Cell cultures

The human colon cancer cell lines HCT116 and SW480, obtained from the European Collection of Authenticated Cell Culture (ECACC), were cultured in DMEM (11965092, Thermo Fisher Scientific) with 10% FBS (Thermo Fisher Scientific), 1% penicillin/streptomycin (Thermo Fisher Scientific, 15140122), and 1% glutamine (Gibco, 25030024), in humidified atmosphere at 37°C with 5% CO_2. Both cell lines tested negative for mycoplasma. The identity of each cell line was confirmed by DNA fingerprinting (STR) with microsatellite markers (Amelogenin, CSF1PO, D13S317, D16S539, D5S818, D7S820, THO1, TPOX, vWA, D8S1179, FGA, Penta E, Penta D, D18S51, D3S1358, D21S11) and compared with the analogous data provided by ATCC, EACC, and https://web.expasy.org/cellosaurus/ (data not shown).

Plasmid transfection and lentiviral transduction

Stable transfection of the *ESRP1* (Sino Biological plasmid # HG13708-UT), *CD44s*, *CD44v6*, and NUMB1-4 (from VOR) expression plasmids was performed using FuGENE HD transfection reagent (Promega, E2311) according to the manufacturer's protocol and selected with Geneticin (Gibco, 10131035). As for the KD constructs, the *ESRP1*-shRNA plasmid (Horizon, V3THS_335722) was packaged by pPAX2 (Addgene # 12260) and pMD2.G (Addgene # 12259) into HEK293T. The virus-containing supernatant was collected 24 hr after transfection, filtered, and employed to infect the HCT116 and SW480 cell line. Selection was applied with 750 ng/mL puromycin (Invivogen, San Diego, CA, USA) or 800 µg/mL of Geneticin selection for 1–2 weeks. The efficiency of overexpression and KD was assessed by qPCR and western blot 48–72 hr after transfection.

RT-qPCR and PCR analyses

Total RNA was isolated using TRIzol reagent (Thermo Fisher Scientific, 15596018) and was reverse-transcribed using high-capacity cDNA reverse transcription kit (Life Technologies, 4368814), according to the manufacturer's instructions. RT-qPCR was performed using the Fast SYBR Green Master Mix (Thermo Fisher Scientific) on an Applied Biosystems StepOne Plus Real-Time Thermal Cycling Research with three replicates per group. Relative gene expression was determined by normalizing the expression of each target gene to GAPDH. Results were analyzed using the 2-(ΔΔCt) method. To validate isoform switches by RT-PCR, CD44-specific primers were as listed in ***Supplementary file 3***.

Western analysis

Cells were lysed in 2× Laemmli buffer containing 4% sodium dodecyl sulfate (SDS), 48% Tris 0.5 M pH 6.8, 20% glycerol, 18% H_2O, bromophenol blue and 10% 1 M DTT, and subjected to SDS-polyacrylamide gel electrophoresis (PAGE), followed by transfer onto polyvinylidene fluoride membranes (Bio-Rad). After blocking with 5% milk in TBS-Tween, the membranes were incubated with primary antibodies against ZEB1 (1.1000, Cell Signaling, #3396), ESRP1 (1.1000, Invitrogen, PA5-11520), CD44s (1.100, Invitrogen, MA5-13890), CD44v6 (1.1000, Abcam, VFF-7), NUMB (1.1000, Cell Signaling, C29G11), and β-actin (1.2000, Cell Signaling, 8547), followed by polyclonal goat anti-mouse/rabbit immunoglobulins horseradish peroxidase-conjugated secondary antibody (Dako) at appropriate dilutions. The signals were detected with Pierce ECT western blotting subtrade (Thermo) using Amersham AI600 (GE Healthcare, Chicago, IL, USA).

Flow cytometry analysis and sorting

Single-cell suspensions generated in PBS supplemented with 1% FBS were incubated with anti-EpCAM-FITC (1.20, Genetex, GTX30708), and anti-CD44-APC (1.20, BD Pharmingen, 559250) antibodies for 30 min on ice and analyzed on a FACSAria III Cell Sorter (BD Biosciences). $CD44^{hi}EpCAM^{hi}$and $CD44^{hi}EpCAM^{lo}$ HCT116 and SW480 cells were sorted and cultured in humidified atmosphere at 37°C with 5% CO_2 for 3–5 days before collecting RNA or protein, as previously described (***Sacchetti et al., 2021***). The subpopulation of cells mapping in between the $CD44^{hi}EpCAM^{hi}$ and $CD44^{hi}EpCAM^{lo}$ gates was labelled as intermediate and was further not employed for analysis.

MTT assay

For MTT assay, 2×10^3 HCT116, SW480 parental, CD44v6, CD44s, and NUMB1-4 OE cells were plated in 96-well plates and incubated at 37°C, 5% CO_2. Twenty-four hours later, in the culture medium was supplemented with 100 µL 0.45 mg/mL MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium

bromide; Sigma-Aldrich) and again incubated for 3 hr. The 96-well plates were then centrifuged at 1000 rpm for 5 min and the culture medium removed. MTT formazan precipitates were solubilized with DMSO. OD reading was performed at 595 nm with microplate reader (Model 550, Bio-Rad). Background measurements were subtracted from each data point. Experiments were performed in duplicate for each individual cell line and drug. Cell numbers were calculated every 24 hr for a 6-day period for proliferation analysis.

Cell migration assay

Migration assays were conducted with 8 µm pore PET transwell inserts (BD Falcon) and TC-treated multi-well cell culture plate (BD Falcon). 5×10^4 cells were seeded in 100 µL of serum-free growth medium in the top chamber. Growth medium containing 10% FBS was used as a chemoattractant in the lower chamber. After 24 hr, cells migrated to the lower chamber were fixed with 4% PFA, stained with 0.1% trypan blue solution, and counted under the microscope.

Mouse spleen transplantation

All mice experiments were implemented according to the Code of Practice for Animal Experiment in Cancer Research from the Netherlands Inspectorate for Health Protections, Commodities and Veterinary Public Health. Mice were fed in the Erasmus MC animal facility (EDC). NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice from 8 to 12 weeks of age were used for spleen transplantation. Anesthetics Ketamine (Ketalin, 0.12 mg/mL) and xylazine (Rompun, 0.61 mg/mL) were given intraperitoneally, while the analgesic Carpofen (Rimadyl, 5 mg/mL) was injected subcutaneously. 5×10^4 HCT116 and SW480 cells resuspended in 50 µL PBS were injected into the exposed spleen with an insulin syringe and left for 15 min before splenectomy. Transplanted mice were sacrificed after 4 and 8 weeks and analyzed for the presence of liver metastases.

AS analysis

The following public available RNASeq (SRA database) data relative to RBP KD studies were used: ESRP1-KD and RMB47-KD in the human non-small cell lung cancer cell line H358 (***Yang et al., 2016***) with accession ID SRP066789 and SRP066793; ESRP2-KD in the human prostate adenocarcinoma cancer cell line LNCaP (***Nieto et al., 2016***) with accession ID SRP191570; the QKI-KD in the oral squamous cell carcinoma cell line CAL27 datasets with accession number SRX8772405. Together with our own EpCAM$^{hi/lo}$ RNASeq data obtained from the colon cancer cell lines (***Sacchetti et al., 2021***), the sequencing reads were mapped to GRCh37.p13.genome by STAR (***Dobin et al., 2013***) (https://www.gencodegenes.org/human/release_19.html). MISO (***Katz et al., 2010***) was used to quantify AS events with annotation from https://miso.readthedocs.io/en/fastmiso/index.html#iso-centric. The MISO uses the alternative exon reads and adjacent conservative reads to measure the percentage of transcript isoform with specific exon included, termed PSI or Ψ. The PSI ranges from 0 (i.e. no isoform includes a specific alternative exon) to 1 (i.e. all of the isoforms detected comprise the alternative exon).

We removed alternative events with low expression of related transcript isoforms if less than three samples in a dataset had more than 10 informative reads to calculate the PSI. Next, we compared the PSI between RBPs KD and wild type in each cell line, as well as the PSI between EpCAM^{hi} and EpCAM^{lo} groups in the SW480 and HCT116 colon cancer cell lines. AS events were defined as differentially spliced events when the difference of mean PSI between two groups (ΔPSI) was >10%.

RNAseq analysis

RNA quality was first evaluated by NanoDrop and further purified by DNAse treatment followed by the TURBO DNA-free Kit protocol (Invitrogen). Samples were sequenced with the DNA nanoball (DNB) seq protocol (BGI) to a depth of 50 million reads per sample. Adapter sequences and low-quality sequences were filtered from the data using SOAPnuke software (BGI). Reads were aligned to the human reference genome build hg19 with the RNAseq aligner STAR (v2.7.9a) and the *Homo sapiens* GENCODE v35 annotation. Duplicates were marked with Sambamba (0.8.0) and raw counts were summed using FeatureCounts (subread 2.0.3). Downstream analysis was performed in R using the DESeq2 package (v1.30.1). After variance stabilizing transformation, PCA was performed on each cell line separately. DEGs were identified by comparing the different groups of ectopically expressing CD44 samples with a Wald test, and by selecting the genes with absolute log fold change

above 1.5 and padj <0.1. GSEA was performed with the Fsgsea package using the HallMark geneset from the molecular signature database, and by selecting significant pathways based on NES >1 and p-value <0.05.

RNAseq data from primary (patient-derived) colon cancers

Patient data from TCGA, with annotation of the CMS as described in ***Guinney et al., 2015***, were integrated with splicing data from the TSVdb (http://www.tsvdb.com/). For splicing analysis, RNAseq by expectation maximization values were log transformed and expression levels of each isoform (CD44std: isoform_uc001mvx, CD44v6: exon_chr11.35226059.35226187, NUMB1. isoform_uc001xny, NUMB2. isoform_uc001xoa, NUMB3: isoform_uc001xnz, NUMB4: isoform_uc001xob) were annotated to the patients. Isoform expression was compared in groups based on the CMS groups and tumor expression levels (*ZEB1*, *ESRP1*). Tumors were stratified on *ZEB1* expression levels using a log rank test top optimize overall survival differences (thresholds: 8.3, 8.6). Next, ESRP1 expression was used to purify the groups into $ZEB1^{hi}ESRP1^{lo}$ and $ZEB1^{lo}ESRP1^{hi}$ (thresholds: 11.6, 11.8). Survival analysis was done using the Kaplan-Meier method with the survival and survminer packages in R. Correlation analysis was done by computing the Pearson correlation between the isoforms and whole gene expression levels as processed in ***Guinney et al., 2015***. Likewise, association between isoform expression and pathway activity was evaluated by computing the Pearson correlation between the isoforms and the average scaled expression values of the pathways, as defined in the HallMark gene set from the molecular signature database (***Liberzon et al., 2015***).

Acknowledgements

We are grateful to Dr Juan Valcarcel (CRG, Barcelona, Spain) for his critical reading of the manuscript.

Additional information

Funding

Funder	Grant reference number	Author
China Scholarship Council	201806300047	Tong Xu

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Tong Xu, Resources, Validation, Investigation, Visualization, Methodology, Writing - original draft; Mathijs Verhagen, Data curation, Software, Formal analysis, Validation, Methodology; Rosalie Joosten, Resources, Supervision, Methodology; Wenjie Sun, Formal analysis, Methodology; Andrea Sacchetti, Investigation, Methodology; Leonel Munoz Sagredo, Investigation, Methodology, Writing – review and editing; Véronique Orian-Rousseau, Conceptualization, Methodology, Writing – review and editing; Riccardo Fodde, Conceptualization, Resources, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Methodology, Writing - original draft, Project administration, Writing – review and editing

Author ORCIDs

Tong Xu http://orcid.org/0000-0002-7046-5917
Mathijs Verhagen http://orcid.org/0000-0003-3126-8379
Riccardo Fodde http://orcid.org/0000-0001-9839-4324

Ethics

All mice experiments were implemented according to the Code of Practice for Animal Experiment in Cancer Research from the Netherlands Inspectorate for Health Protections, Commodities and Veterinary Public Health. Permit number AVD1010020171344.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.82006.sa1

Author response https://doi.org/10.7554/eLife.82006.sa2

Additional files

Supplementary files

- Supplementary file 1. List of alternative splicing targets in ESRP1 knocking down H358 line (a), HCT116 (b), and SW480 (c) EpCAM^{lo} and EpCAM^{hi} subpopulation, filtered by ΔPSI (differential percentage spliced in) > 0.1.
- Supplementary file 2. List of alternative splicing targets in *ESRP1*-KD in the H358 cell line, *ESRP2*-KD in LNCaP, *RBM47*-KD in H358 line, QKI-KD in CAL27, and HCT116 and SW480 EpCAM^{lo} and EpCAM^{hi} subpopulation, filtered by ΔPSI (differential percentage spliced in) > 0.1.
- Supplementary file 3. Lists of primer sequences used for RT-PCR analysis.
- Supplementary file 4. Differential expressed gene lists from the RNAseq analysis HCT116 CD44s- and CD44v6-OE cells.
- Supplementary file 5. Differential expressed gene lists from the RNAseq analysis SW480 CD44s- and CD44v6-OE cells.
- Supplementary file 6. List of gene set enrichment analysis (GSEA) in CD44s OE vs. CD44v6 OE vs. parental HCT116 and SW480 cells.
- Supplementary file 7. List of gene set variation analysis (GSVA) in CD44s OE vs. CD44v6 OE vs. parental HCT116 and SW480 cells.
- MDAR checklist

Data availability

The RNA-sequencing data from this study have been submitted to the Gene Expression Omnibus (GEO) database under the accession number GSE192877. Other data referenced in this study are publicly available and can be accessed from the GEO using GSE154927, GSE154730 and Synapse using identifier syn2623706 .

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Xu T, Verhagen MP, Fodde R	2022	CD44s and CD44v6 overexpressed RNAseq profiles of colon cancer cell lines HCT116 and SW480	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE192877	NCBI Gene Expression Omnibus, GSE192877

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Sacchetti A	2020	CD44highEpCAMhigh and CD44highEpCAMlow RNAseq profiles of colon cancer cell lines HCT116 and SW480, in triplicate	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE154927	NCBI Gene Expression Omnibus, GSE154927
Kang K, Hong JH, Ahn Y, Ko YH	2020	Transcriptome datasets of the Quaking (QKI) gene knock-down human oral squamous cell carcinoma (OSCC) cells	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi=GSE154730	NCBI Gene Expression Omnibus, GSE154730
Guinney J	2015	Colorectal Cancer Subtyping Consortium (CRCSC)	https://doi.org/10.7303/syn2623706	Synapse, 10.7303/syn2623706

References

Adam RA, Adam YG. 2004. Malignant ascites: past, present, and future. *Journal of the American College of Surgeons* **198**:999–1011. DOI: https://doi.org/10.1016/j.jamcollsurg.2004.01.035, PMID: 15194082

Azevedo R, Gaiteiro C, Peixoto A, Relvas-Santos M, Lima L, Santos LL, Ferreira JA. 2018. Cd44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. *Clinical Proteomics* **15**:22. DOI: https://doi.org/10.1186/s12014-018-9198-9, PMID: 29983670

Bergen V, Lange M, Peidli S. 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**:1408–1414. DOI: https://doi.org/10.1038/s41587-020-0591-3

Bernards R, Weinberg RA. 2002. A progression puzzle. *Nature* **418**:823. DOI: https://doi.org/10.1038/418823a, PMID: 12192390

Biamonti G, Infantino L, Gaglio D, Amato A. 2019. An intricate connection between alternative splicing and phenotypic plasticity in development and cancer. *Cells* **9**:E34. DOI: https://doi.org/10.3390/cells9010034, PMID: 31877720

Blencowe BJ. 2006. Alternative splicing: new insights from global analyses. *Cell* **126**:37–47. DOI: https://doi.org/10.1016/j.cell.2006.06.023, PMID: 16839875

Brabletz T, Jung A, Spaderna S, Hlubek F, Kirchner T. 2005. Opinion: migrating cancer stem cells-an integrated concept of malignant tumour progression. *Nature Reviews. Cancer* **5**:744–749. DOI: https://doi.org/10.1038/nrc1694, PMID: 16148886

Brabletz S, Brabletz T. 2010. The ZEB/mir-200 feedback loop -- a motor of cellular plasticity in development and cancer? *EMBO Reports* **11**:670–677. DOI: https://doi.org/10.1038/embor.2010.117, PMID: 20706219

Brown RL, Reinke LM, Damerow MS, Perez D, Chodosh LA, Yang J, Cheng C. 2011. Cd44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *The Journal of Clinical Investigation* **121**:1064–1074. DOI: https://doi.org/10.1172/JCI44540, PMID: 21393860

Calon A, Lonardo E, Berenguer-Llergo A, Espinet E, Hernando-Momblona X, Iglesias M, Sevillano M, Palomo-Ponce S, Tauriello DVF, Byrom D, Cortina C, Morral C, Barceló C, Tosi S, Riera A, Attolini CS-O, Rossell D, Sancho E, Batlle E. 2015. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nature Genetics* **47**:320–329. DOI: https://doi.org/10.1038/ng.3225, PMID: 25706628

Conway J, Al-Zahrani KN, Pryce BR, Abou-Hamad J, Sabourin LA. 2017. Transforming growth factor β-induced epithelial to mesenchymal transition requires the ste20-like kinase SLK independently of its catalytic activity. *Oncotarget* **8**:98745–98756. DOI: https://doi.org/10.18632/oncotarget.21928, PMID: 29228724

Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR. 2011. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Research* **39**:D301–D308. DOI: https://doi.org/10.1093/nar/gkq1069, PMID: 21036867

Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N, Regev A. 2016. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**:1853–1866.. DOI: https://doi.org/10.1016/j.cell.2016.11.038, PMID: 27984732

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. DOI: https://doi.org/10.1093/bioinformatics/bts635, PMID: 23104886

Fodde R, Brabletz T. 2007. Wnt/Beta-Catenin signaling in cancer stemness and malignant behavior. *Current Opinion in Cell Biology* **19**:150–158. DOI: https://doi.org/10.1016/j.ceb.2007.02.007, PMID: 17306971

Fu XD, Ares M. 2014. Context-Dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews. Genetics* **15**:689–701. DOI: https://doi.org/10.1038/nrg3778, PMID: 25112293

Goswami S, Philippar U, Sun D, Patsialou A, Avraham J, Wang W, Di Modugno F, Nistico P, Gertler FB, Condeelis JS. 2009. Identification of invasion specific splice variants of the cytoskeletal protein Mena present in mammary tumor cells during invasion in vivo. *Clinical & Experimental Metastasis* **26**:153–159. DOI: https://doi.org/10.1007/s10585-008-9225-8, PMID: 18985426

Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa E Melo F, Missiaglia E, Ramay H, Barras D, Homicsko K, et al. 2015. The consensus molecular subtypes of colorectal cancer. *Nature Medicine* **21**:1350–1356. DOI: https://doi.org/10.1038/nm.3967, PMID: 26457759

Hanahan D. 2022. Hallmarks of cancer: new dimensions. *Cancer Discovery* **12**:31–46. DOI: https://doi.org/10.1158/2159-8290.CD-21-1059, PMID: 35022204

Hernández-Martínez R, Ramkumar N, Anderson KV. 2019. P120-catenin regulates WNT signaling and EMT in the mouse embryo. *PNAS* **116**:16872–16881. DOI: https://doi.org/10.1073/pnas.1902843116, PMID: 31371508

Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, Mellano A, Senetta R, Cassenti A, Sonetto C, Inghirami G, Trusolino L, Fekete Z, De Ridder M, Cassoni P, Storme G, Bertotti A, Medico E. 2015. Stromal contribution to the colorectal cancer transcriptome. *Nature Genetics* **47**:312–319. DOI: https://doi.org/10.1038/ng.3224, PMID: 25706627

Kahles A, Lehmann K-V, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, Stegle O, Kohlbacher O, Sander C, Cancer Genome Atlas Research Network, Rätsch G. 2018. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* **34**:211–224.. DOI: https://doi.org/10.1016/j.ccell.2018.07.001, PMID: 30078747

Karve K, Netherton S, Deng L, Bonni A, Bonni S. 2020. Regulation of epithelial-mesenchymal transition and organoid morphogenesis by a novel tgfbeta-TCF7L2 isoform-specific signaling pathway. *Cell Death Dis* **11**:704. DOI: https://doi.org/10.1038/s41419-020-02905-z

Katz Y, Wang ET, Airoldi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* **7**:1009–1015. DOI: https://doi.org/10.1038/nmeth.1528, PMID: 21057496

Kim EJ, Kim JS, Lee S, Lee H, Yoon J-S, Hong JH, Chun SH, Sun DS, Won HS, Hong SA, Kang K, Jo JY, Choi M, Shin DH, Ahn Y-H, Ko YH. 2019. QKI, a mir-200 target gene, suppresses epithelial-to-mesenchymal transition and tumor growth. *International Journal of Cancer* **145**:1585–1595. DOI: https://doi.org/10.1002/ijc.32372, PMID: 31026342

La Manno G, Soldatov R, Zeisel A. 2018. RNA velocity of single cells. *Nature* **560**:494–498. DOI: https://doi.org/10.1038/s41586-018-0414-6

Lengauer C, Kinzler KW, Vogelstein B. 1997. Genetic instability in colorectal cancers. *Nature* **386**:623–627. DOI: https://doi.org/10.1038/386623a0, PMID: 9121588

Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The molecular signatures database (msigdb) hallmark gene set collection. *Cell Systems* **1**:417–425. DOI: https://doi.org/10.1016/j.cels.2015.12.004, PMID: 26771021

Lin G, Li J, Cai J, Zhang H, Xin Q, Wang N, Xie W, Zhang Y, Xu N. 2021. Rna-Binding protein MBNL2 regulates cancer cell metastasis through mir-182-MBNL2-AKT pathway. *Journal of Cancer* **12**:6715–6726. DOI: https://doi.org/10.7150/jca.62816, PMID: 34659561

Lu Z, Huang Q, Park JW, Shen S, Lin L, Tokheim CJ, Henry MD, Xing Y. 2015. Transcriptome-wide landscape of pre-mrna alternative splicing associated with metastatic colonization. *Molecular Cancer Research* **13**:305–318. DOI: https://doi.org/10.1158/1541-7786.MCR-14-0366, PMID: 25274489

Misra S, Hascall VC, De Giovanni C, Markwald RR, Ghatak S. 2009. Delivery of CD44 shrna/nanoparticles within cancer cells: perturbation of hyaluronan/cd44v6 interactions and reduction in adenoma growth in apc min/+ MICE. *The Journal of Biological Chemistry* **284**:12432–12446. DOI: https://doi.org/10.1074/jbc.M806772200, PMID: 19246453

Nieto MA, Huang RY-J, Jackson RA, Thiery JP. 2016. EMT: 2016. *Cell* **166**:21–45. DOI: https://doi.org/10.1016/j.cell.2016.06.028, PMID: 27368099

Oltean S, Bates DO. 2014. Hallmarks of alternative splicing in cancer. *Oncogene* **33**:5311–5318. DOI: https://doi.org/10.1038/onc.2013.533, PMID: 24336324

Orian-Rousseau V, Chen L, Sleeman JP, Herrlich P, Ponta H. 2002. Cd44 is required for two consecutive steps in HGF/c-Met signaling. *Genes & Development* **16**:3074–3086. DOI: https://doi.org/10.1101/gad.242602, PMID: 12464636

Orian-Rousseau V. 2015. Cd44 acts as a signaling platform controlling tumor progression and metastasis. *Frontiers in Immunology* **6**:154. DOI: https://doi.org/10.3389/fimmu.2015.00154, PMID: 25904917

Pece S, Confalonieri S, R Romano P, Di Fiore PP. 2011. NUMB-ing down cancer by more than just a Notch. *Biochimica et Biophysica Acta* **1815**:26–43. DOI: https://doi.org/10.1016/j.bbcan.2010.10.001, PMID: 20940030

Pillman KA, Phillips CA, Roslan S, Toubia J, Dredge BK, Bert AG, Lumb R, Neumann DP, Li X, Conn SJ, Liu D, Bracken CP, Lawrence DM, Stylianou N, Schreiber AW, Tilley WD, Hollier BG, Khew-Goodall Y, Selth LA, Goodall GJ, et al. 2018. MiR-200/375 control epithelial plasticity-associated alternative splicing by repressing the RNA-binding protein quaking. *The EMBO Journal* **37**:e99016. DOI: https://doi.org/10.15252/embj.201899016, PMID: 29871889

Ponta H, Sherman L, Herrlich PA. 2003. Cd44: from adhesion molecules to signalling regulators. *Nature Reviews. Molecular Cell Biology* **4**:33–45. DOI: https://doi.org/10.1038/nrm1004, PMID: 12511867

Preca B-T, Bajdak K, Mock K, Sundararajan V, Pfannstiel J, Maurer J, Wellner U, Hopt UT, Brummer T, Brabletz S, Brabletz T, Stemmler MP. 2015. A self-enforcing cd44s/ZEB1 feedback loop maintains EMT and stemness properties in cancer cells. *International Journal of Cancer* **137**:2566–2577. DOI: https://doi.org/10.1002/ijc.29642, PMID: 26077342

Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, Tokheim CJ, Brown A, DeBlasio RM, Niyazov J, Zucker A, Karchin R, Kinzler KW, Iacobuzio-Donahue CA, Vogelstein B, Nowak MA. 2018. Minimal functional driver gene heterogeneity among untreated metastases. Science **361**:1033–1037. DOI: https://doi.org/10.1126/science.aat7171, PMID: 30190408

Rokavec M, Kaller M, Horst D, Hermeking H. 2017. Pan-Cancer EMT-signature identifies RBM47 down-regulation during colorectal cancer progression. *Scientific Reports* **7**:4687. DOI: https://doi.org/10.1038/s41598-017-04234-2, PMID: 28680090

Roy Burman D, Das S, Das C, Bhattacharya R. 2021. Alternative splicing modulates cancer aggressiveness: role in EMT/metastasis and chemoresistance. *Molecular Biology Reports* **48**:897–914. DOI: https://doi.org/10.1007/s11033-020-06094-y, PMID: 33400075

Sacchetti A, Teeuwssen M, Verhagen M, Joosten R, Xu T, Stabile R, van der Steen B, Watson MM, Gusinac A, Kim WK, Ubink I, Van de Werken HJ, Fumagalli A, Paauwe M, Van Rheenen J, Sansom OJ, Kranenburg O, Fodde R. 2021. Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. *eLife* **10**:e61461. DOI: https://doi.org/10.7554/eLife.61461, PMID: 34036938

Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. 2015. Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). Current Protocols in Human Genetics **87**:11. DOI: https://doi.org/10.1002/0471142905.hg1116s87, PMID: 26439713

Shimada H, Kohno T, Konno T, Okada T, Saito K, Shindo Y, Kikuchi S, Tsujiwaki M, Ogawa M, Matsuura M, Saito T, Kojima T. 2021. The roles of tricellular tight junction protein angulin-1/lipolysis-stimulated lipoprotein receptor (LSR) in endometriosis and endometrioid-endometrial carcinoma. *Cancers* **13**:6341. DOI: https://doi.org/10.3390/cancers13246341, PMID: 34944960

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive integration of single-cell data. *Cell* **177**:1888–1902. DOI: https://doi.org/10.1016/j.cell.2019.05.031, PMID: 31178118

Tavanez JP, Valcárcel J. 2010. A splicing mastermind for EMT. *The EMBO Journal* **29**:3217–3218. DOI: https://doi.org/10.1038/emboj.2010.234, PMID: 20924395

Teeuwssen M, Fodde R. 2019. Cell heterogeneity and phenotypic plasticity in metastasis formation: the case of colon cancer. *Cancers* **11**:E1368. DOI: https://doi.org/10.3390/cancers11091368, PMID: 31540068

Thiery JP, Acloque H, Huang RYJ, Nieto MA. 2009. Epithelial-Mesenchymal transitions in development and disease. *Cell* **139**:871–890. DOI: https://doi.org/10.1016/j.cell.2009.11.007, PMID: 19945376

Todaro M, Gaggianesi M, Catalano V, Benfante A, Iovino F, Biffoni M, Apuzzo T, Sperduti I, Volpe S, Cocorullo G, Gulotta G, Dieli F, De Maria R, Stassi G. 2014. Cd44V6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. *Stem Cell* **14**:342–356. DOI: https://doi.org/10.1016/j.stem.2014.01.009, PMID: 24607406

van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, Burdziak C, Moon KR, Chaffer CL, Pattabiraman D, Bierie B, Mazutis L, Wolf G, Krishnaswamy S, Pe'er D. 2018. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**:716–729. DOI: https://doi.org/10.1016/j.cell.2018.05.061, PMID: 29961576

Varga J, Greten FR. 2017. Cell plasticity in epithelial homeostasis and tumorigenesis. *Nature Cell Biology* **19**:1133–1141. DOI: https://doi.org/10.1038/ncb3611, PMID: 28945230

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**:470–476. DOI: https://doi.org/10.1038/nature07509, PMID: 18978772

Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. 2009. Esrp1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Molecular Cell* **33**:591–601. DOI: https://doi.org/10.1016/j.molcel.2009.01.025, PMID: 19285943

Warzecha CC, Jiang P, Amirikian K, Dittmar KA, Lu H, Shen S, Guo W, Xing Y, Carstens RP. 2010. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *The EMBO Journal* **29**:3286–3300. DOI: https://doi.org/10.1038/emboj.2010.195, PMID: 20711167

Xia RM, Liu T, Li WG, Xu XQ. 2021. Rna-Binding protein RBM24 represses colorectal tumourigenesis by stabilising PTEN mRNA. *Clinical and Translational Medicine* **11**:e383. DOI: https://doi.org/10.1002/ctm2.383, PMID: 34709758

Xiao H. 2019. MiR-7-5p suppresses tumor metastasis of non-small cell lung cancer by targeting NOVA2. *Cellular & Molecular Biology Letters* **24**:60. DOI: https://doi.org/10.1186/s11658-019-0188-3, PMID: 31832068

Yae T, Tsuchihashi K, Ishimoto T, Motohara T, Yoshikawa M, Yoshida GJ, Wada T, Masuko T, Mogushi K, Tanaka H, Osawa T, Kanki Y, Minami T, Aburatani H, Ohmura M, Kubo A, Suematsu M, Takahashi K, Saya H, Nagano O. 2012. Alternative splicing of CD44 mrna by ESRP1 enhances lung colonization of metastatic cancer cell. *Nature Communications* **3**:883. DOI: https://doi.org/10.1038/ncomms1892, PMID: 22673910

Yang Y, Park JW, Bebee TW, Warzecha CC, Guo Y, Shang X, Xing Y, Carstens RP. 2016. Determination of a comprehensive alternative splicing regulatory network and combinatorial regulation by key factors during the epithelial-to-mesenchymal transition. *Molecular and Cellular Biology* **36**:1704–1719. DOI: https://doi.org/10.1128/MCB.00019-16, PMID: 27044866

Zeilstra J, Joosten SPJ, Dokter M, Verwiel E, Spaargaren M, Pals ST. 2008. Deletion of the WNT target and cancer stem cell marker CD44 in apc(min/+) mice attenuates intestinal tumorigenesis. *Cancer Research* **68**:3655–3661. DOI: https://doi.org/10.1158/0008-5472.CAN-07-2940, PMID: 18483247

Zeilstra J, Joosten SPJ, van Andel H, Tolg C, Berns A, Snoek M, van de Wetering M, Spaargaren M, Clevers H, Pals ST. 2014. Stem cell cd44v isoforms promote intestinal cancer formation in apc(min) mice downstream of wnt signaling. *Oncogene* **33**:665–670. DOI: https://doi.org/10.1038/onc.2012.611, PMID: 23318432

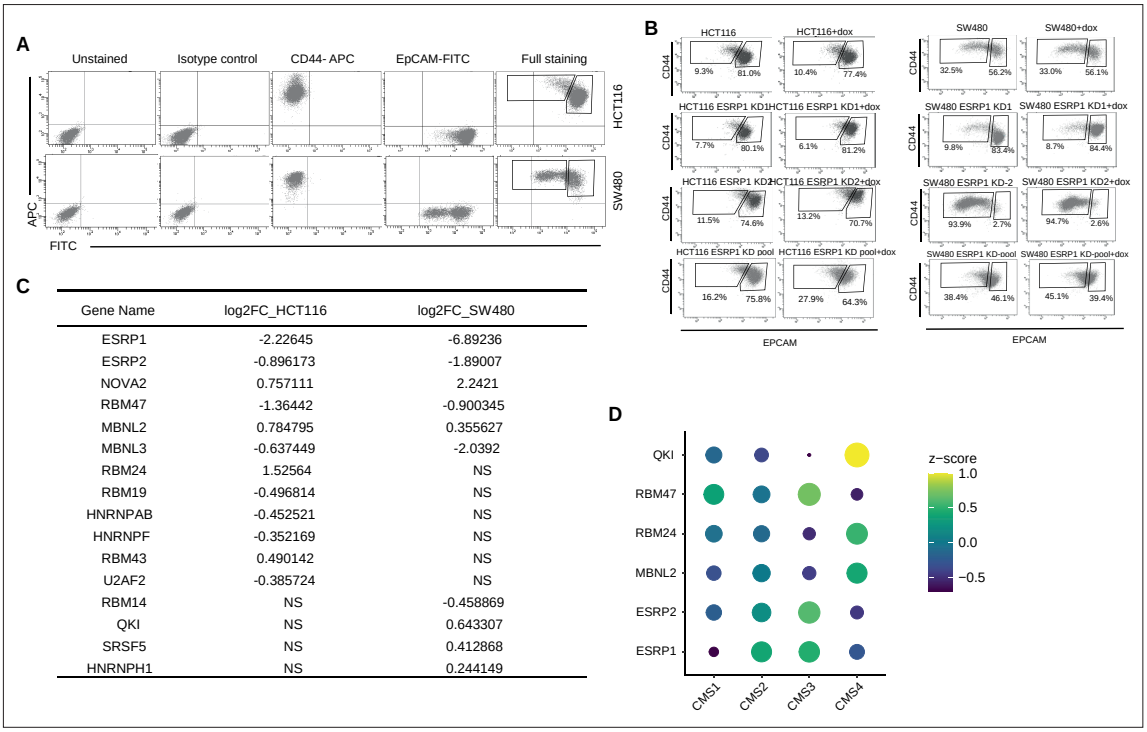
Zhang H, Brown RL, Wei Y, Zhao P, Liu S, Liu X, Deng Y, Hu X, Zhang J, Gao XD, Kang Y, Mercurio AM, Goel HL, Cheng C. 2019a. Cd44 splice isoform switching determines breast cancer stem cell state. *Genes & Development* **33**:166–179. DOI: https://doi.org/10.1101/gad.319889.118, PMID: 30692202

Zhang J, Zheng Z, Wu M, Zhang L, Wang J, Fu W, Xu N, Zhao Z, Lao Y, Xu H. 2019b. The natural compound neobractatin inhibits tumor metastasis by upregulating the RNA-binding-protein MBNL2. *Cell Death & Disease* **10**:554. DOI: https://doi.org/10.1038/s41419-019-1789-5, PMID: 31320607

Figures and figure supplements

Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behavior in colon cancer

Tong Xu *et al*.



C

Gene Name	log2FC_HCT116	log2FC_SW480
ESRP1	-2.22645	-6.89236
ESRP2	-0.896173	-1.89007
NOVA2	0.757111	2.2421
RBM47	-1.36442	-0.900345
MBNL2	0.784795	0.355627
MBNL3	-0.637449	-2.0392
RBM24	1.52564	NS
RBM19	-0.496814	NS
HNRNPAB	-0.452521	NS
HNRNPF	-0.352169	NS
RBM43	0.490142	NS
U2AF2	-0.385724	NS
RBM14	NS	-0.458869
QKI	NS	0.643307
SRSF5	NS	0.412868
HNRNPH1	NS	0.244149

Figure 1—figure supplement 1. *ESRP1* and RNA-binding proteins (RBPs) functional and expression analysis in cell lines and patient-derived colon cancers. (**A**) FACS isotype and compensation controls in the analysis of the HCT116 and SW480 cell lines. The gates relative to the EpCAM$^{hi/lo}$ subpopulations are specifically designed for the HCT116 and SW480 cell lines, as shown for the full staining. For the sake of simplicity and readability, the quadrants showing negative, single positive, and double positive regions have not been repeated in the figures encompassing FACS analyses. (**B**) CD44/EpCAM FACS analysis of EpCAM^{lo} and EpCAM^{hi} subpopulations in ESRP1-KD (*shESRP1*-transduced) HCT116 and SW480 cells. Cells were induced with 1 µg/mL doxycycline for 72 hr before analysis. (**C**) List of RBPs differentially expressed between EpCAM^{lo} and EpCAM^{hi} subpopulation in SW480 and HCT116. The RBPs' list was from reference (***Fu and Ares, 2014***). (**D**) Dot plot analysis of the z-score scaled RBPs' expression values across the four colon cancer consensus molecular subtypes (CMS; annotated according to ***Guinney et al., 2015***). RNAseq data were obtained from the COAD (COlon-ADenoma) tumors of The Cancer Genome Atlas (TCGA) deposited in the TCGA Splicing Variants Database (TSVdb) (n=206 primary tumors).

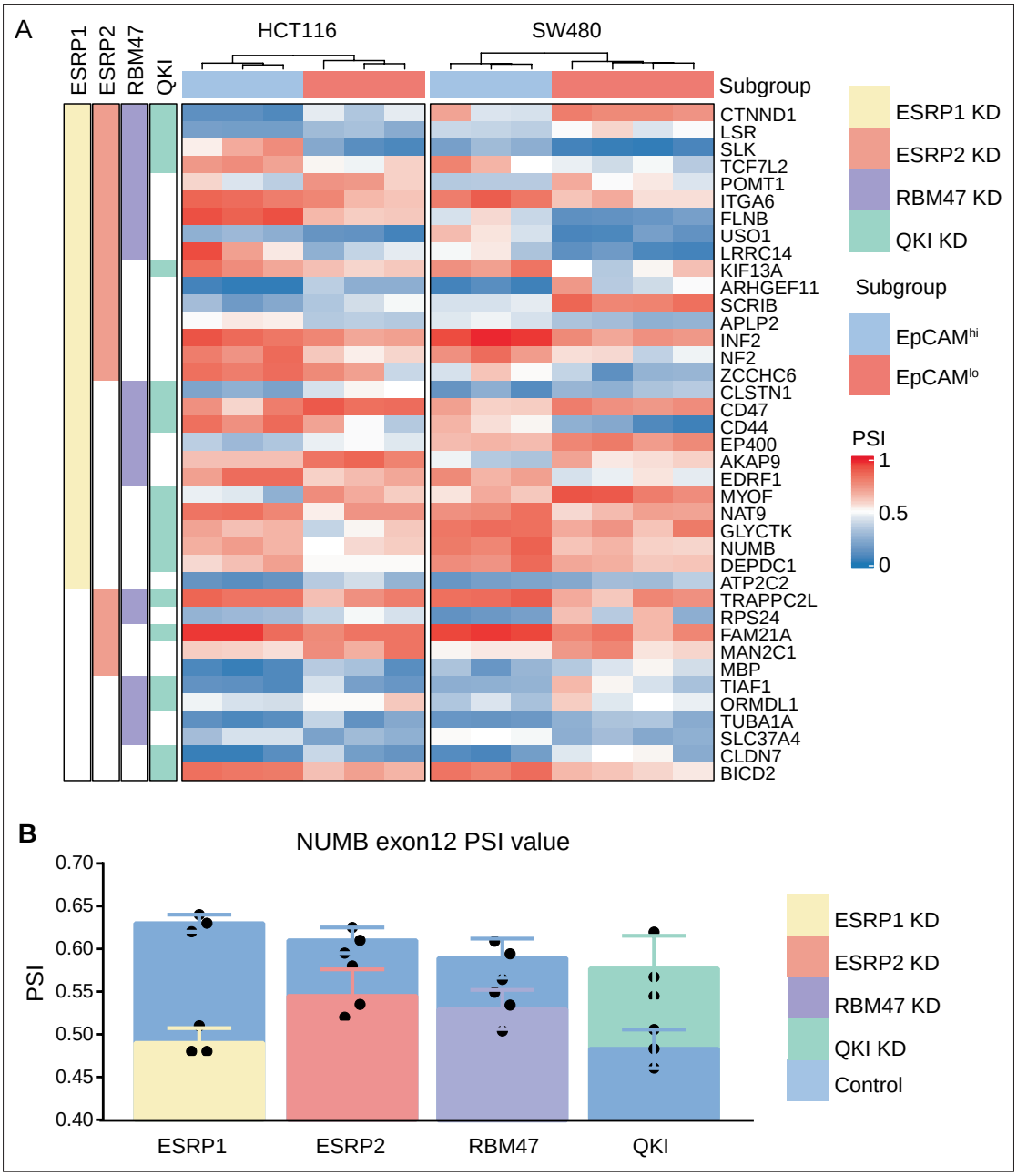


Figure 2—figure supplement 1. *ESRP1/2-*, *RBM47-*, and *QKI*-regulated alternative splicing (AS) targets. (**A**) Heatmap of the AS events observed by comparing previously published RNAseq data from RBP-KD studies (ESRP1-KD in H358 ***Yang et al., 2016***, ESRP2-KD in LNCaP ***Nieto et al., 2016***, RBM47-KD in H358 ***Yang et al., 2016***, and QKI-KD in CAL27; GEO Accession: GSM4677985) with our own HCT116/SW480 EpCAM$^{hi/lo}$ RNAseq data (***Sacchetti et al., 2021***). Shared AS targets between RBPs KD cells, and HCT116/SW480 EpCAM$^{hi/lo}$ subpopulations are shown. The gene list on the

Figure 2—figure supplement 1 continued on next page

Figure 2—figure supplement 1 continued

right side of the heatmap encompasses variants earmarked by ΔPSI (differential percentage spliced in) >0.1. The colored bars on the left of the heatmap shows if there are variants spliced by different RNA-binding proteins (RPBs). Color in white means AS is not involved in. (**B**) PSI value of NUMB exon 12 between ESRP1 KD, ESRP2 KD, RBM47 KD, QKI KD, and control cells.

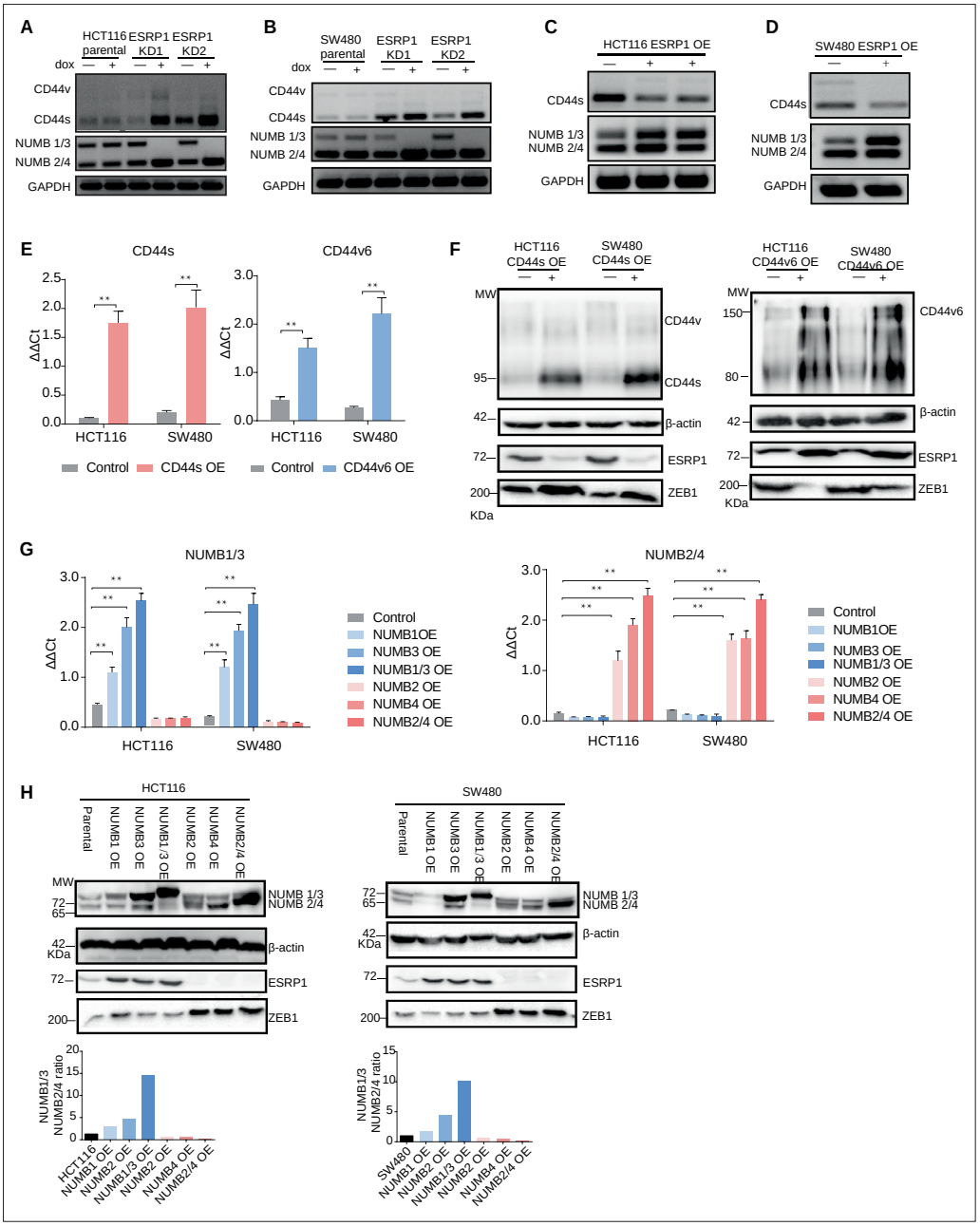


Figure 3—figure supplement 1. *ESRP1*, *CD44*, and *NUMB* isoforms analysis in overexpressing and knockdown (KD) colon cancer cell lines. RT-PCR analysis of CD44 and NUMB isoforms expression in HCT116 (**A**) and SW480 (**B**) ESRP1-KD (*shESRP1*-transduced) cells, and in HCT116 (**C**) and SW480 (**D**) *ESRP1*-OE cells. Cells were induced with 1 μg/mL doxycycline for 72 hr before RNA isolation. *GAPDH* was used as control. (**E**) RT-qPCR analysis of CD44s and CD44v6 expression in HCT116 and SW480 CD44s- (left), and CD44v6-OE (right) cells. Expression of the constitutive *CD44* exons was

Figure 3—figure supplement 1 continued on next page

Figure 3—figure supplement 1 continued

employed to normalize the results (means ± SEM, n=3). **=p < 0.01. (**F**) Western analysis of CD44, ESRP1, and ZEB1 expression in HCT116 and SW480 CD44s- (left), and CD44v6-OE (right) cells. Please note that the molecular weight of CD44v6 is expected to range between 80 and 150 kDa (***Azevedo et al., 2018***, ***Ponta et al., 2003***). β-Actin was used as loading control for western blots. (**G**) RT-PCR analysis of NUMB1-4 isoforms expression in HCT116 and SW480 NUMB1-4 OE cells. Expression of the constitutive *NUMB* exons was employed to normalize the results (means ± SEM, n=3). **=p < 0.01. (**H**) Western analysis of NUMB1-4 isoforms, ESRP1, and ZEB1 expression in HCT116 and SW480 NUMB1-4 OE cells (upper panels). The ratio of NUMB1/3 and NUMB2/4 bands was quantified by ImageJ and shown in bar plot (lower histogram panels). β-Actin was used as loading control for western blots.

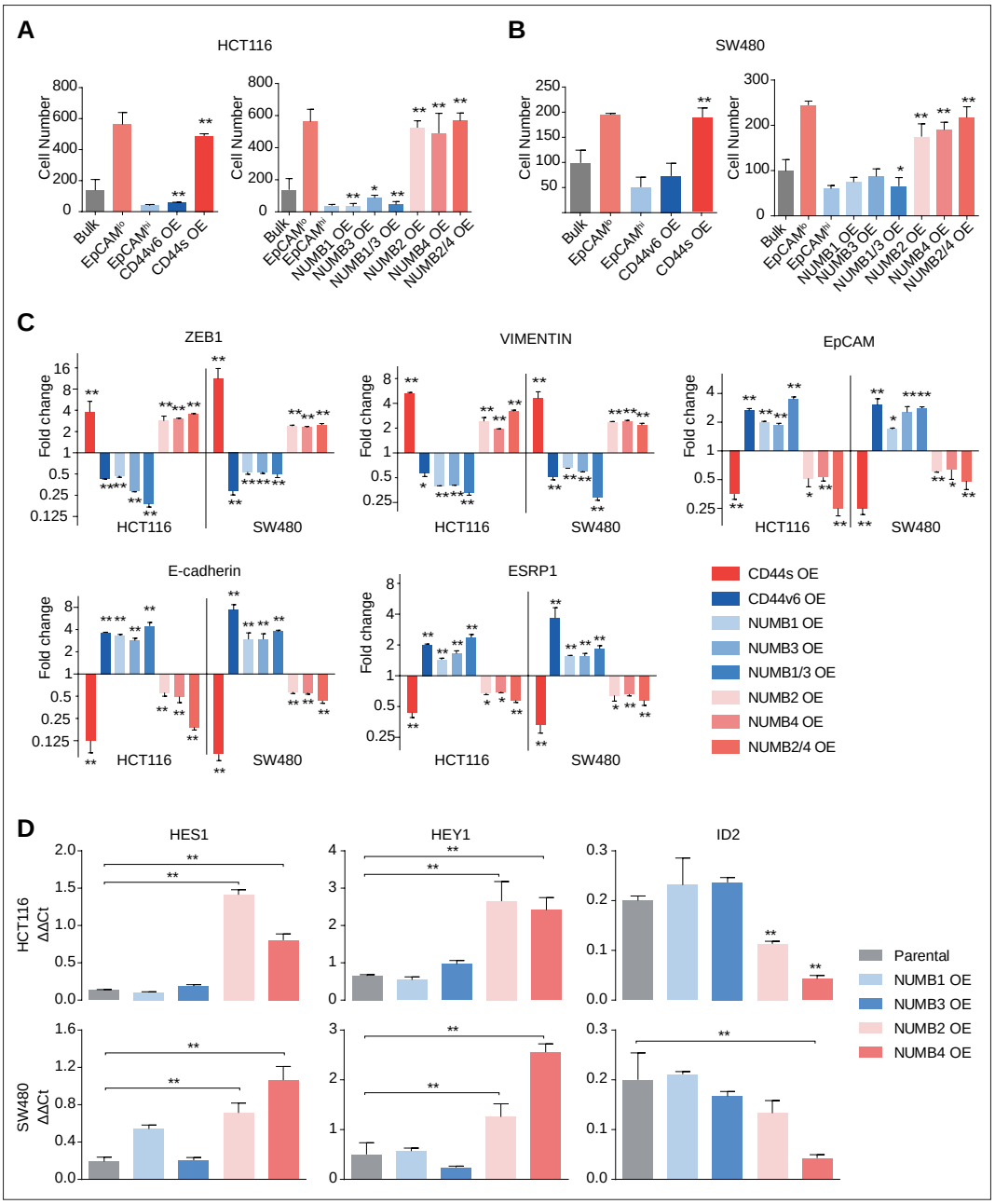


Figure 3—figure supplement 2. *CD44* and *NUMB* isoform-specific expression affects cell migration and Notch signaling activation. (**A**) Migration assay analysis of HCT116 CD44s-, CD44v6-, and NUMB1/4-OE cells. EpCAM^{lo} and EpCAM^{hi} cells were used as controls. Each bar represents the mean ± SD of cells migrated to the bottom of the transwell from two independent experiments. (n=3) (**B**) Migration assay analysis of SW480 CD44s-, CD44v6-, and NUMB1/4-OE cells. EpCAM^{lo} and EpCAM^{hi} cells were used as controls. Each bar represents the mean ± SD of cells migrated to the bottom of

Figure 3—figure supplement 2 continued on next page

Figure 3—figure supplement 2 continued

the transwell from two independent experiments. (n=3) (**C**) RT-qPCR analysis of EMT-TFs in HCT116 and SW480 CD44s-, CD44v6-, and NUMB1/4-OE cells. *GAPDH* expression was used as control, normalized with the HCT116 or SW480 parental in each sample (means ± SEM, n=3). Increased gene expression is depicted by red bars, whereas downregulation – when compared with parental cells – is shown by blue bar. (**D**) RT-qPCR analysis of the Notch signaling pathway markers *HES1*, *HEY1*, and *ID2* in HCT116 and SW480 NUMB1/4-OE cells. *GAPDH* expression was used as control (means ± SEM, n=3). *=$p < 0.05$; **=$p < 0.01$.

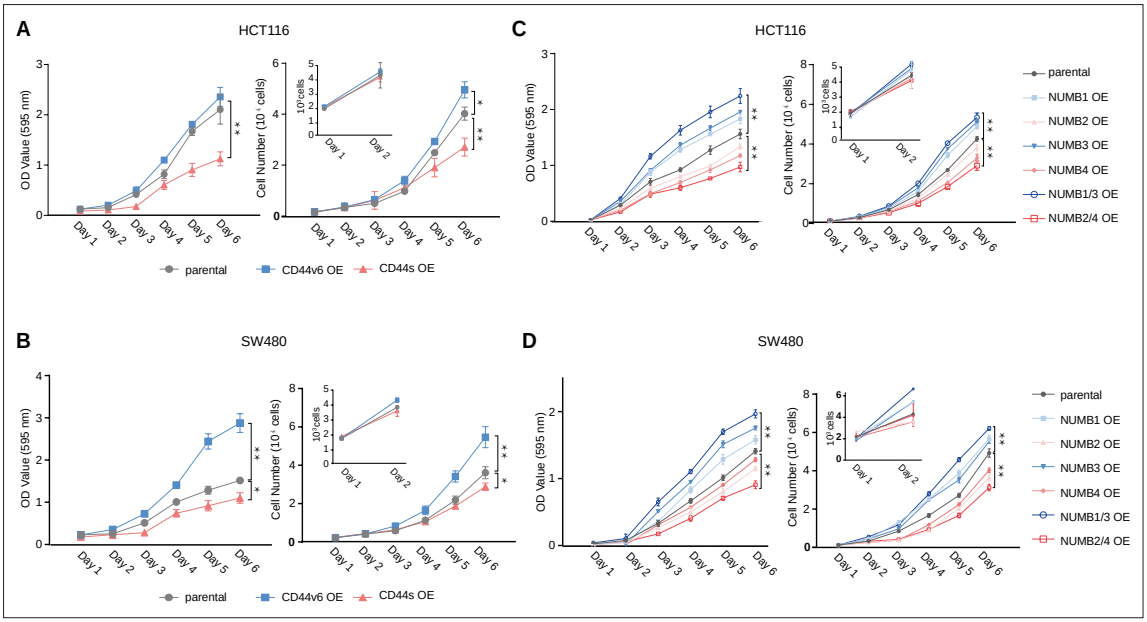


Figure 4—figure supplement 1. CD44 and NUMB isoforms regulate colon cancer cell proliferation. (**A**) Proliferation assays of HCT116 CD44s- and CD44v6-OE cells. Both OD values and cell multiplicities are shown from day 1 to 6 (means ± SEM, n=3). *=p < 0.05; **=p < 0.01. (**B**) Proliferation assays of SW480 CD44s- and CD44v6-OE cells. Both OD values and cell multiplicities are shown from day 1 to 6 (means ± SEM, n=3). *=p < 0.05; **=p < 0.01. (**C**) Proliferation assays of HCT116 NUMB1/4-OE cells. Both OD values and cell multiplicities are shown from day 1 to 6 (means ± SEM, n=3). **=p < 0.01. (**D**) Proliferation assays of SW480 NUMB1/4-OE cells. Both OD values and cell multiplicities are shown from day 1 to 6 (means ± SEM, n=3). **=p < 0.01.

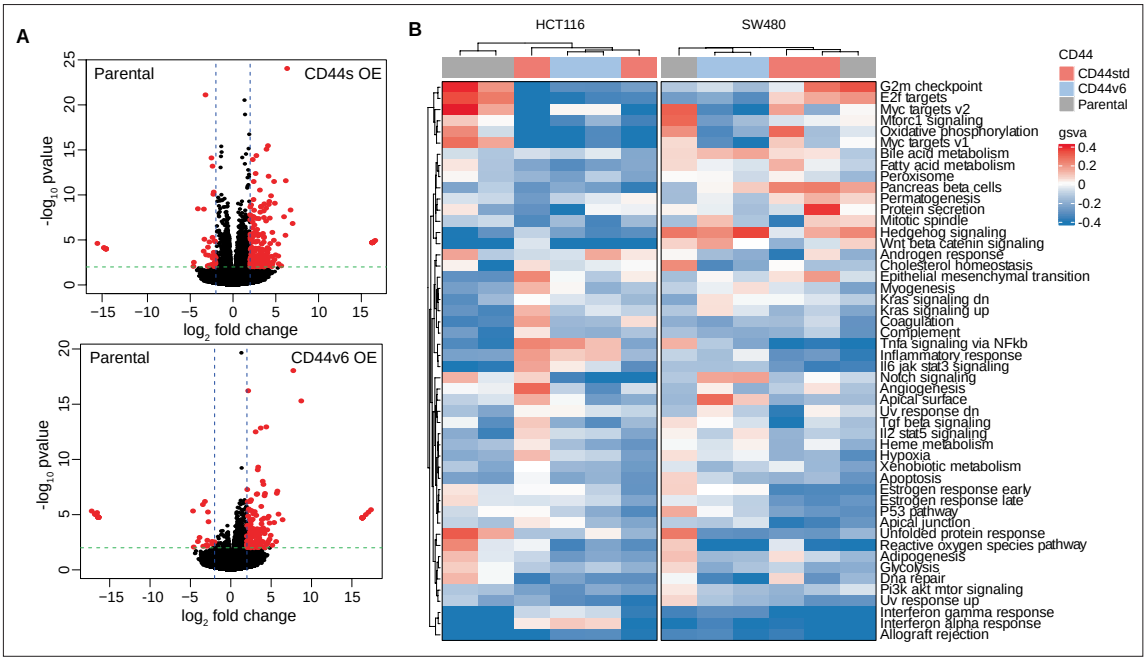


Figure 5—figure supplement 1. Gene enrichment and pathway analysis of CD44s- and CD44v6-overexpressing (OE) colon cancer cells. (**A**) Volcano plots showing differentially expressed genes (absLFC >2, pval <0.01, red) by comparing parental cell lines to the CD44s- and CD44v6-OE samples in both cell lines. (**B**) Gene set enrichment analysis of parental, CD44s- and CD44v6-OE cells compared with each other in HCT116 and SW480 cells shown in heatmap, respectively. Only significantly altered pathways (normalized enrichment score [NES] >1, and pval <0.05) are shown.

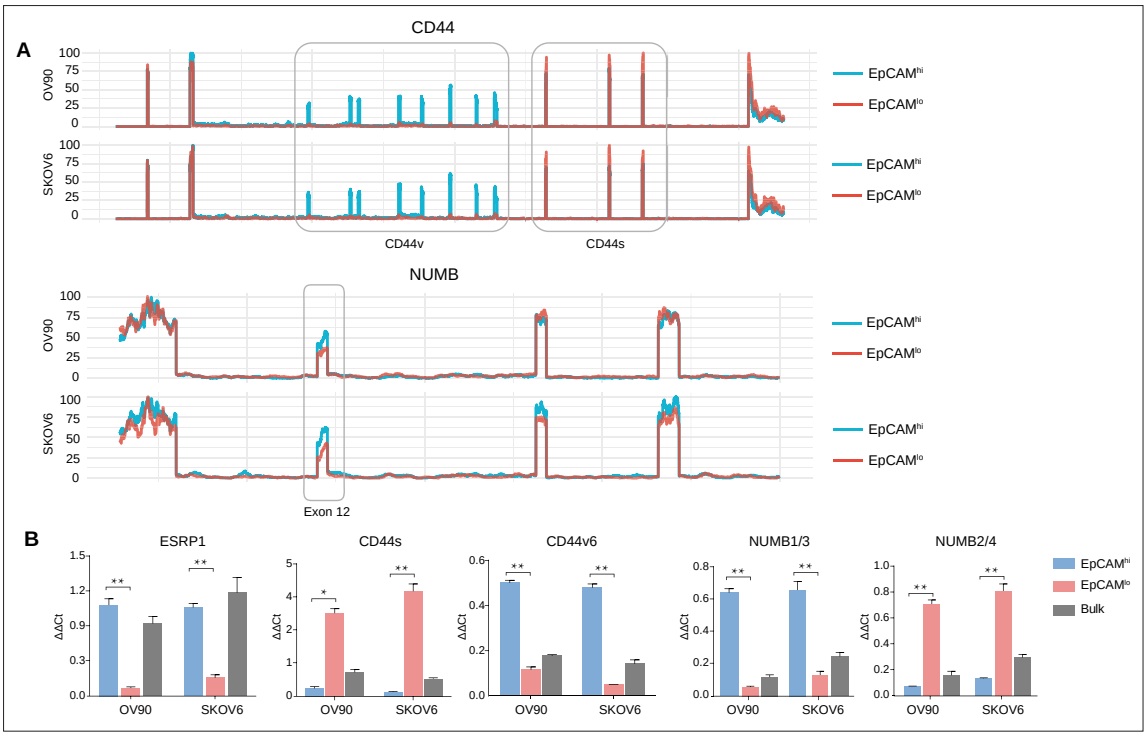


Figure 7—figure supplement 1. *CD44* and *NUMB* isoforms expression in EpCAM$^{hi/lo}$ ovarian and cervical cancer cells. (**A**) *CD44* and *NUMB* exon chromosome sites information from alternative splicing (AS) analysis in the ovarian and cervical cancer cell lines OV90 and SKOV6. Exon peak plot depicts the expression of different exons in the three groups; peak height is indicative of the expression level of specific exons. CD44v: CD44 exons v2 to v10. CD44v and CD44s, and NUMB exon 12 are highlighted by gray rectangles. (**B**) RT-qPCR expression analysis of *ESRP1*, *CD44s*, *CD44v6*, *NUMB1/3*, and *NUMB2/4* isoforms in EpCAM^{hi}, EpCAM^{lo}, and bulk subpopulations in OV90 and SKOV6 ovarian cancer cell lines. *GAPDH* expression was used as control (means ± SEM, n=3). **=p < 0.01.

CHAPTER 5

Tropomyosin1 isoforms underlie epithelial to mesenchymal plasticity, metastatic dissemination, and resistance to chemotherapy in high-grade serous ovarian cancer

Cell Death Differ, doi:10.1038/s41418-024-01267-9 (2024)

CDDpress

www.nature.com/cdd

ARTICLE OPEN

Check for updates

Tropomyosin1 isoforms underlie epithelial to mesenchymal plasticity, metastatic dissemination, and resistance to chemotherapy in high-grade serous ovarian cancer

Tong Xu[1], Mathijs P. Verhagen[1], Miriam Teeuwssen[1,6], Wenjie Sun[2], Rosalie Joosten[1], Andrea Sacchetti[1], Patricia C. Ewing-Graham[1], Maurice P. H. M. Jansen[3], Ingrid A. Boere[3], Nicole S. Bryce[4,7], Jun Zeng[5], Herbert R. Treutlein[5,8], Jeff Hook[4], Edna C. Hardeman[4], Peter W. Gunning[4] and Riccardo Fodde[1]✉

Phenotypic plasticity, defined as the ability of individual cells with stable genotypes to exert different phenotypes upon exposure to specific environmental cues, represent the quintessential hallmark of the cancer cell en route from the primary lesion to distant organ sites where metastatic colonization will occur. Phenotypic plasticity is driven by a broad spectrum of epigenetic mechanisms that allow for the reversibility of epithelial-to-mesenchymal and mesenchymal-to-epithelial transitions (EMT/MET). By taking advantage of the co-existence of epithelial and quasi-mesenchymal cells within immortalized cancer cell lines, we have analyzed the role of EMT-related gene isoforms in the regulation of epithelial mesenchymal plasticity (EMP) in high grade serous ovarian cancer. When compared with colon cancer, a distinct spectrum of downstream targets characterizes quasi-mesenchymal ovarian cancer cells, likely to reflect the different modalities of metastasis formation between these two types of malignancy, i.e. hematogenous in colon and transcoelomic in ovarian cancer. Moreover, upstream RNA-binding proteins differentially expressed between epithelial and quasi-mesenchymal subpopulations of ovarian cancer cells were identified that underlie differential regulation of EMT-related isoforms. In particular, the up- and down-regulation of RBM24 and ESRP1, respectively, represent a main regulator of EMT in ovarian cancer cells. To validate the functional and clinical relevance of our approach, we selected and functionally analyzed the Tropomyosin 1 gene (*TPM1*), encoding for a protein that specifies the functional characteristics of individual actin filaments in contractile cells, among the ovarian-specific downstream AS targets. The low-molecular weight *Tpm1.8/9* isoforms are specifically expressed in patient-derived ascites and promote invasion through activation of EMT and Wnt signaling, together with a broad spectrum of inflammation-related pathways. Moreover, *Tpm1.8/9* expression confers resistance to taxane- and platinum-based chemotherapy. Small molecule inhibitors that target the *Tpm1* isoforms support targeting Tpm1.8/9 as therapeutic targets for the development of future tailor-made clinical interventions.

Cell Death & Differentiation; https://doi.org/10.1038/s41418-024-01267-9

INTRODUCTION

Epithelial ovarian cancer (EOC) is the leading cause of death amongst gynecologic malignancies due to its high case-to-fatality ratio [1, 2]. EOC generally becomes manifest at advanced disease stages, i.e. when metastases have already spread to pelvic organs (stage II), the abdomen (stage III), or beyond the peritoneal cavity (stage IV) [3]. Based on the underlying genetic defects, two main EOC subtypes have been recognized [4]. Type I tumors are slow growing, mostly restricted to the ovary, and thought to arise from well-differentiated precursor lesions called "borderline" tumors. They are further subdivided into low-grade serous, mucinous, clear cell, and endometrioid subtypes. Mutations in *KRAS*, *BRAF*, *PTEN*, and *CTNNB1* (β-catenin) earmarks type I EOCs, often together with a relatively stable karyotype. High-grade serous (HGSOC) and undifferentiated carcinomas are type II EOCs and are frequently characterized by *TP53* mutations and by aneuploidy [4]. HGSOC represents the most malignant and common ovarian cancer type accounting for up to 70% of all cases with poor prognosis and survival [5].

Of note, EOC is the only cancer type where no physical barrier exists between primary lesion and the main metastatic site, i.e. the intraperitoneal cavity. The dissemination of ovarian cancer cells results in their adhesion to intra-abdominal organs and the peritoneum [6] eventually leading to ascites accumulation due to the obstruction of lymphatic vessels [7].

Epithelial-to-mesenchymal (EMT) and mesenchymal-to-epithelial transitions (MET) are thought to underlie local dissemination and distant metastasis in the majority of epithelial

[1]Department of Pathology, Erasmus University Medical Center, Rotterdam, The Netherlands. [2]Institut Curie, Laboratory of Genetics and Developmental Biology, Paris, France. [3]Department of Medical Oncology, Erasmus University Medical Center, Rotterdam, The Netherlands. [4]School of Biomedical Sciences, Faculty of Medicine and Health, The University of New South Wales, Sydney, New South Wales, Australia. [5]Computist Bio-NanoTech, Scoresby, VIC 3179, Australia. [6]Present address: Elisabeth-TweeSteden Ziekenhuis (ETZ), Tilburg, The Netherlands. [7]Present address: The Victor Chang Cardiac Research Institute, Darlinghurst, NSW, Australia. [8]Present address: Sanoosa Pty. Ltd, Moonee Ponds, VIC 3039, Australia. ✉email: r.fodde@erasmusmc.nl

Received: 24 July 2023 Revised: 1 February 2024 Accepted: 6 February 2024
Published online: 16 February 2024

malignancies, including ovarian cancer where, as mentioned above, the absence of a physical barrier would suggest alternative shedding mechanisms [8]. EMT/MET are regulated by a broad spectrum of epigenetic mechanisms involving chromatin remodeling due to histone methylation/acetylation, non-coding RNAs, promoter DNA methylation, and post-transcriptional mechanisms such as alternative splicing (AS) [9]. AS occurs in the majority of human genes and represents a major determinant of protein diversity [10]. A broad spectrum of RNA-binding proteins (RBPs) regulate splicing by recognizing specific sequences within (pre) mRNA and are known to play a role in AS through their differential expression among different tissues in homeostasis and disease [11–14]. RBP-driven alternative splicing is thought to play a crucial role in cancer metastasis, as well as in EMT/MET [15]. Among the various RBPs, ESRP1 (epithelial splicing regulatory protein 1) was shown to regulate splicing of multiple downstream targets during EMT in different cancer types [16].

Previously, we have shown that *ESRP1*, together with other RBP-coding genes, are differentially expressed between epithelial and quasi-mesenchymal colon cancer cells as the result of their role in the regulation of alternative splicing downstream of EMT [17]. Among several AS targets, the *CD44* and *NUMB* genes were shown to play relevant functional roles to promote EMT and phenotypic plasticity in colon cancer. Here, we have employed a similar approach to study the role of alternative splicing and alternative promoter usage in EOC metastasis. Systematic analysis of the downstream EMT isoform targets only revealed a small overlap with what was previously identified in colon cancer [17]. Accordingly, the majority of the targets were ovarian cancer-specific as illustrated by the Tropomyosin 1 (*TPM1*) gene that encodes for various actin-binding protein isoforms that specify the functional characteristics of individual actin filaments in muscle and non-muscle cells [18]. *TPM1* was previously shown to suppress tumor development in multiple cancer types through distinct mechanisms [19, 20]. Here, we established that the low molecular weight isoforms *Tpm1.8/9* play a key role in EMT, Wnt signaling, and chemo-resistance in ovarian cancer and represent a candidate target for future therapies as shown by the in vitro results obtained both by isoform-specific siRNA and small molecule inhibitors.

RESULTS

A subpopulation of quasi-mesenchymal cells co-exists with epithelial cells in high-grade serous ovarian cancer

Following the experimental strategy previously adopted to identify subpopulations of quasi-mesenchymal cells in immortalized colon cancer cell lines [17, 21], we analyzed the HGSOC cell lines OV90, SKOV3, COV504, and CAOV3 by FACS with CD44 and EpCAM antibodies. As shown in Fig. 1A, distinct distributions of $CD44^{hi}EpCAM^{hi}$ and $CD44^{hi}EpCAM^{lo}$ cells (from here on referred to as $EpCAM^{hi}$ and $EpCAM^{lo}$, respectively) earmarked each line albeit in different percentages. Upon sorting and short-term culturing, $EpCAM^{lo}$ cells showed a mesenchymal-like morphology, in contrast with the epithelial appearances of their $EpCAM^{hi}$ counterparts (Fig. 1B). Accordingly, $EpCAM^{lo}$ cells showed increased migration and invasion capacity in trans-well assays (Fig. 1C).

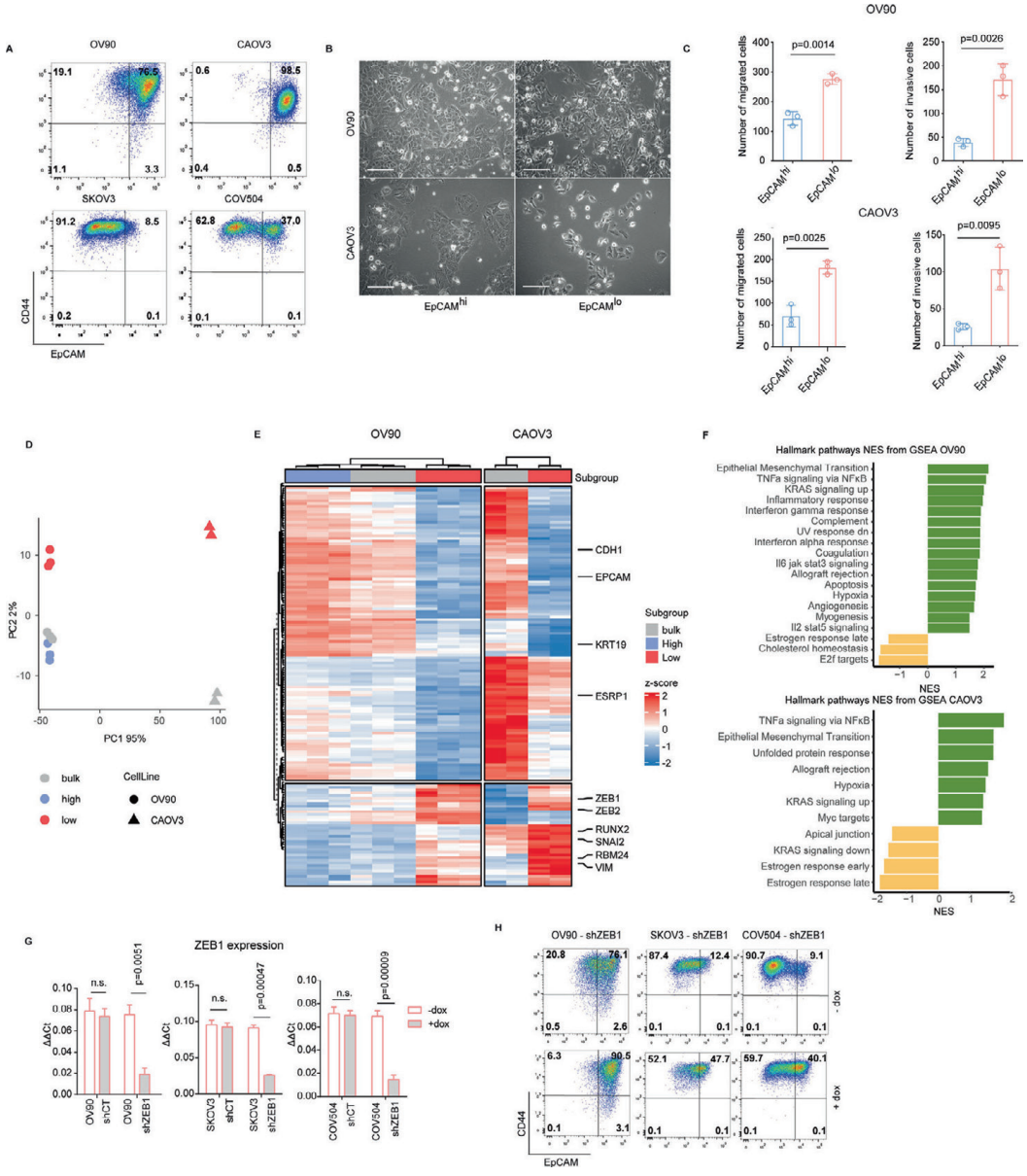
In order to elucidate the global gene expression profiles of the two distinct subpopulations of ovarian cancer cells, RNAseq analysis was carried out on the $EpCAM^{hi/lo}$ cells sorted by FACS from the OV90 and CAOV3 lines. Principal component analysis (PCA) by multidimensional scaling (MDS) revealed a clear separation of the quasi-mesenchymal $EpCAM^{lo}$ cells from their epithelial counterpart in the second dimension (Fig. 1D). Unsupervised clustering of the RNAseq data highlighted differentially expressed genes between the $EpCAM^{lo}$ and $EpCAM^{hi}$ subpopulations from the two cell lines (Fig. 1E). Among these genes, different EMT transcription factors (EMT-TFs; *ZEB1/2*, *RUNX2*, *SNAI2*), EMT target genes (*CDH1*, *VIM*), and EMT-related RBPs (*ESRP1*, *RBM24*) earmarked $EpCAM^{lo}$ cells in both OV90 and CAOV3. Accordingly, pathway analysis (PA) of the genes whose expression earmarks the quasi-mesenchymal ovarian cancer cells revealed significant associations with EMT, KRAS signaling, and several TNF- and interferon-related inflammatory pathways (Fig. 1F). Pathways with a normalized enrichment score (NES) > 1 and P value < 0.05 were labeled as significant. To confirm the central role played by EMT in the establishment of the $EpCAM^{lo}$ identity as indicated by the PA analysis, *ZEB1* expression was downregulated by shRNA in the OV90, SKOV3, and COV504 cell lines (Fig. 1G). As expected, FACS analysis of these cells shows that knockdown of the differentially expressed EMT-TF results in a substantial reduction of the $EpCAM^{lo}$ subpopulation (Fig. 1H).

To demonstrate the clinical relevance of the above data obtained from immortalized cancer cell lines, publicly available single-cell RNAseq data from 42 patient-derived ovarian cancers [22] were employed to search for quasi-mesenchymal cells reminiscent of the $EpCAM^{lo}$ subpopulation. To this aim, we took advantage of the $EpCAM^{lo}$ signature derived from the above RNAseq analysis of the OV90 and CAOV3 cell lines (Supplementary Table 1). After classifying patient-derived cancer cells according to the $EpCAM^{lo}$ signature (Fig. 2A), approx. 60% were found to cluster into one of the three subpopulations of HGS ovarian cancer cells defined in the original study by Vazquez-Garcia et al. [22] as EMT-like (labeled as #2, #4, and #6) (Fig. 2B). Of note, a substantial fraction of the $EpCAM^{lo}$-like cells fell outside the three clusters indicative of large heterogeneity of cellular identities reminiscent of the quasi-mesenchymal state in the cell lines. As shown in Fig. 2C, cells from within the $EpCAM^{lo}$-like clusters were more represented in metastases and ascites (i.e. non-adnexa including omentum, peritoneum, etc.) when compared with primary tumors. We then evaluated module scores from the Hallmark gene sets across the $EpCAM^{lo}$-like clusters in primary ovarian cancers (adnexa), ascites, and metastases (non-adnexa). As shown in Fig. 2D, E, EMT-like signature and inflammatory signaling pathways (TNFα, IL6/Jak/Stat3, IFN, and TGFβ) were significantly upregulated in $EpCAM^{lo}$-like clusters, in particular in cluster #4 in ascites and metastatic lesions.

Collectively, these results indicate that, HGSOC cell lines encompass subpopulations of quasi-mesenchymal cells endowed with increased motility and invasive capacity, and characterized by EMT-TFs the expression of which is central to their cellular identity. Similar subpopulations of ovarian cancer cells are found in patient-derived primary and metastatic lesions, and in malignant ascites.

Differential expression of RNA-binding proteins underlies alternative splicing of a subset of target genes in quasi-mesenchymal ovarian cancer cells

Recently, we showed that several RNA-binding proteins (RBPs) known to be involved in alternative splicing are differentially expressed between $EpCAM^{lo}$ and $EpCAM^{hi}$ cells in colon cancer cell lines and play significant functional roles in controlling E-to-M and M-to-E transitions and phenotypic plasticity during local dissemination and distant metastasis [17]. By following an analogous approach, we have identified several differentially expressed RBPs likely to play an active role in the regulation of EMT-related alternative isoforms between epithelial- and mesenchymal-like subpopulations of ovarian cancer cells. As shown in Fig. 3A, *ESRP1* and *RBM24* were respectively down- and upregulated in $EpCAM^{lo}$ cells in both cell lines. Moreover, other RBPs (i.e. *ESRP2*, *RBM47*, *RBMS3*, and *QKI*) were differentially expressed in one of the cell lines examined. The $EpCAM^{lo}$-specific up- and downregulation of *RBM24* and *ESRP1*, respectively, was validated by RTqPCR and western analysis of the subpopulations sorted from the OV90, CAOV3 and COV504 cell lines (Fig. 3B, C).



Next, we analyzed differentially spliced genes by MISO (Mixture of Isoforms) [23] and filtered the results by selecting ΔPSI (differential Percentage Spliced In) values > 10% (Supplementary Tables 2 and 3), and by comparing them with the corresponding lists of isoform targets previously found in colon cancer cell lines [17]. As shown in Supplementary Fig. 3 Supplement 1, the large majority of ovarian cancer targets was not found in colon cancer. Among the ovarian-specific targets with a known function in EMT ($n = 39$), diverse cellular components, e.g. extracellular matrix, focal adhesion, and the actin cytoskeleton, and cellular processes, e.g. ECM organization, integrin- and TGFβ-mediated signaling, and cell migration were represented (Supplementary Fig. 3, Supplement 1). In particular, the presence of Tropomyosin 1 (*TPM1*), a member of a broad family of actin-binding proteins, was noteworthy both because of the high ΔPSI values in OV90 and CAOV3 cells and because of its known cellular function, i.e. in the cytoskeleton of non-muscle cells and in the contractile system of striated and smooth muscles [24, 25]. Although a direct causative role between *TPM1* and EMT has not been reported yet, previous studies have shown that TGF-β signaling increases expression of

Fig. 1 A subpopulation of quasi-mesenchymal cells co-exists with epithelial cells in HGS ovarian cancer cell lines. A FACS analysis of the ovarian cancer cell lines OV90, CAOV3, SKOV3, and COV504 with antibodies directed against CD44 and EpCAM. EpCAM/CD44 positive and negative areas were defined as previously described [17, 21] using multiple isotype controls and are shown by the quadrants in the plots. Using specific gates, cells were separated in $CD44^{hi}EpCAM^{hi}$ and $CD44^{hi}EpCAM^{lo}$ subpopulations. The percentages of $CD44^{hi}EpCAM^{lo}$ and $CD44^{hi}EpCAM^{hi}$ cells within each cell line are depicted in each quadrant. Notably, as previously observed for SW480 and HCT116, the ovarian cancer cell lines revealed a continuum of different EpCAM and CD44 expression levels with a large $EpCAM^{hi}$ (or $EpCAM^{lo}$ as in the case of SKOV3) cluster followed by a tail of gradually decreasing (increasing for SKOV3) EpCAM levels. By applying the indicated gates, cells were sorted into $EpCAM^{hi}$ and $EpCAM^{lo}$ subpopulations. Graphs show representative analysis from an individual experiment. **B** Phase contrast microscopy images of sorted $EpCAM^{hi}$ and $EpCAM^{lo}$ cells from $EpCAM^{hi}$ and $EpCAM^{lo}$ OV90 and CAOV3 sorted cells. Scale bar: 100 μm. While $EpCAM^{hi}$ cells show characteristic epithelial morphology, $EpCAM^{lo}$ cells showed a more spindle- and mesenchymal-like appearance. Scale bar: 100 μm. **C** Transwell migration assay (upper graph) and invasion assay (lower graph) of $EpCAM^{hi}$ (blue bar) and $EpCAM^{lo}$ (red bar) OV90 and CAOV3 cells. Each bar symbolizes the mean ± SD. **D** Principal component analysis (PCA) of the RNAseq profiles of $EpCAM^{hi}$ and $EpCAM^{lo}$ cells from the OV90 and CAOV3 lines. **E** Heatmap of common differentially expressed gene among $EpCAM^{hi/lo}$ and bulk subpopulations from the OV90 and CAOV3 cell lines (abs LFC > 1.5, *P* value < 0.01). Complete-linkage hierarchical clustering with split by k-means (k = 2) clustering was used. **F** Hallmark pathways based on the Gene Set Enrichment Analysis (GSEA) of OV90 and CAOV3 $EpCAM^{lo}$ expression profiles compared with $EpCAM^{hi}$. Plots show only significantly expressed pathways, with a normalized enrichment score (NES) > 1 and *P* value < 0.05. **G** RT-qPCR expression analysis of *ZEB1* in OV90, SKOV3 and COV504 transduced with an inducible control (shCT) and with a *ZEB1*-shRNA (shZEB1) construct. shRNA expression was induced with 1 μg/mL of doxycycline. GAPDH expression was used as control. Each bar represents the mean ± SD. *P* value is indicated. **H** FACS analysis of the OV90, SKOV3 and COV504 cell lines transfected with the sh*ZEB1* and control constructs using antibodies against CD44 and EpCAM. Cells were induced with 1 μg/mL doxycycline for 72 h prior to the FACS analysis. The percentages of $EpCAM^{lo}$ and $EpCAM^{hi}$ cells within each cell line are depicted in each quadrant.

high-molecular weight tropomyosins together with the formation of actin stress fibers thus affecting cell motility and invasion [26, 27].

Notably, the *TPM1* isoform pattern observed in ovarian cancer involved exons 1a/2a, which earmark the *Tpm1.6/7* isoforms upregulated in $EpCAM^{hi}$ cells, and exon 1b, featuring the *Tpm1.8/9* isoforms upregulated in $EpCAM^{lo}$, as validated both by RT-qPCR and western blot analysis (Fig. 3D, E). Of note, the western results relative to *Tpm1.8/9* clearly match the RNA-based analysis, whereas the same is less evident for *Tpm1.6/7* that, at the protein level, seem to undergo more subtle variations. As a further validation of these results, HGS ovarian cancer cell lines exclusively encompassing $EpCAM^{hi}$ (PEA2) or $EpCAM^{lo}$ (PEA1) cells (Fig. 3F), solely expressed the *Tpm1.6/7* and *Tpm1.8/9* isoforms, respectively (Fig. 3G).

In order to establish a cause-effect relationship between the differential expression of *ESRP1* and *RBM24* in $EpCAM^{lo}$ ovarian cancer cells and the observed downstream AS targets using *TPM1* as a model, we performed RBP knockdown and overexpression assays in OV90 and COV504 cell lines. First, as predicted by their unique $EpCAM^{hi/lo}$ distribution, the PEA1 and PEA2 cell lines exclusively express *RBM24* and *ESRP1*, respectively (Fig. 3G, right panel). Accordingly, both *ESRP1* knockdown and *RBM24* ectopic expression in the OV90 and COV504 cell lines resulted in the up- and downregulation of the *Tpm1.8/9* and *Tpm1.6/7* isoforms, respectively, both at the RNA and protein levels (Fig. 4A, Supplementary Fig. 4-Suppl. 1). Vice versa, *RBM24* knockdown and *ESRP1* ectopic expression resulted in the up- and downregulation of the *Tpm1.6/7* and *Tpm1.8/9* isoforms, respectively (Fig. 4B, Supplementary Fig. 4-Suppl. 1). Of note, the observed changes in RBP expression were also accompanied by a pronounced increase of the $EpCAM^{lo}$ subpopulation as shown by FACS analysis (Fig. 4C, Supplementary Fig. 4-Suppl. 2). Concurrent *RBM24* overexpression and *ESRP1* knockdown had significant effects not only on the *TPM1* isoform shift but, more importantly, on the relative percentages of the $EpCAM^{hi/lo}$ subpopulations when compared with the single RBP gain- and loss-of-function assays (Fig. 4, Supplementary Fig. 4-Suppl. 2).

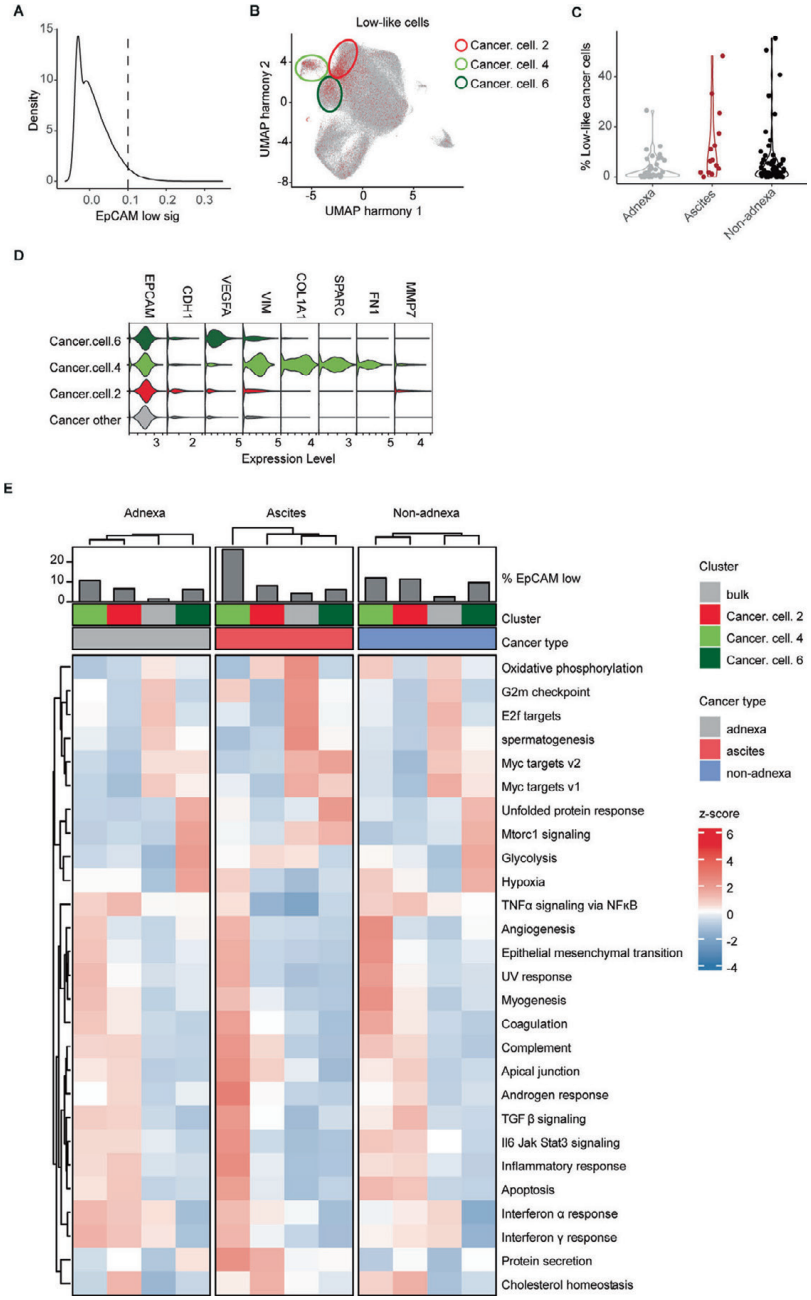
RBPs have been known to play essential roles not only in alternative splicing but also in the co-transcriptional regulation of gene promoters. As the differentially expressed *Tpm1.6/7 and 8/9* isoforms are more likely to result from alternative first exon selection rather than alternative splicing, we performed an in silico analysis of both TPM1a/b promoters using the FANTOM database [28]. No ESRP1 or RBM24 binding motifs were detected. However, the same approach indicated potential interaction with the EMT-TFs SNAI1/2, ZEB1, and RUNX2 (Supplementary Fig. 4-Suppl. 3A, B).

To further study the alternative usage of the TPM1a (Tpm1.6/7) and TPM1b (Tpm1.8/9) promoters, we cloned both promoters into luciferase constructs to assess their transcriptional activity by transient reporter assays. As shown in Supplementary Fig. 4-Suppl. 3C, D, an increase in reporter activity was observed from the TPM1a promoter in the sorted $EpCAM^{hi}$ and in ESRP1-overexpressing OV90 cells. Likewise, increased reporter activity was observed from the TPM1b promoter in the sorted $EpCAM^{lo}$ and in RBM24-overexpressing OV90 and COV504 cells (albeit marginally). Conversely, decreased reporter activity was observed from the TPM1a promoter in sorted OV90 and COV504 $EpCAM^{hi}$ cells upon ESRP1-knockdown. Similarly, decreased reporter activity was observed from the TPM1b promoter in sorted OV90 and COV504 $EpCAM^{lo}$ cells upon RBM24 knockdown (Supplementary Fig. 4-Suppl. 3E-H).

Overall, these results showed that epithelial and quasi-mesenchymal subpopulations of HGS ovarian cancer cells are earmarked by the differential expression of specific RBPs known to be involved in alternative splicing and co-transcriptional regulation of alternative promoters. Accordingly, differential isoform patterns of a subset of EMT-related target genes characterize the $EpCAM^{lo}$ cells. While some of the target gene isoforms are shared with those previously identified in colon cancer, the majority appear to be specific for ovarian cancer. Among these, *TPM1* is of interest in view of both its function as an actin-binding cytoskeletal protein in various cell types [18] and of its previously reported role as a tumor suppressor in breast cancer [29].

Transcriptional and functional consequences of TPM1 isoforms on quasi-mesenchymal ovarian cancer cells

To assess the functional relevance of the specific *TPM1* isoforms, their ectopic expression was induced and validated by RT-qPCR and western blot analysis in multiple ovarian cancer cell lines (OV90, COV504, PEA1 and PEA2) (Fig. 5A, B). Cell viability and proliferation assays indicated that *Tpm1.6/7* and *Tpm1.8/9* expression was significantly associated with increased and decreased rates of cell division, respectively (Fig. 5C). Of note, ectopic expression of *Tpm1.8/9* in OV90 and COV504 cells resulted in a complete growth arrest within 6 days. Moreover, trans-well assays clearly showed that *Tpm1.8/9* overexpression resulted in significantly increased migratory and invasive features when compared with *Tpm1.6/7* (Fig. 5D). *Tpm1.8/9*-overexpressing (OE) ovarian cancer cells appeared to invade the collagen layer collectively, as narrow linear strands with "leader" and "follower" cells (Fig. 5E). The latter observation is of interest since in non-muscle cells *Tpm1.8/9* are specifically expressed in lamellipodia, i.e. the membrane protrusions found at



the leading edge driven by branched as well as unbranched filaments composed of actin and *Tpm1.8/9* that promote cell motility [30, 31]. IF analysis with *TPM1* isoform-specific antibodies confirmed the co-localization of *Tpm1.8/9* and ARP2, a specific marker for lamellipodia, at the edge of ovarian cancer cells; instead, *Tpm1.6/7* were mainly localized in the cytoplasm (Fig. 5F).

In order to carry out a more comprehensive study of the transcriptional and functional consequences of the ectopic expression of the specific *TPM1* isoforms, RNAseq analysis was carried out on the OV90 overexpressing cells. Unsupervised hierarchical clustering and principal component analyses confirmed the distinct transcriptional identity of the OV90 parental,

Fig. 2 EpCAM^{lo} cells co-exist with epithelial cells in high-grade serous ovarian cancer. A Density plot shows the distribution of patient-derived ovarian cancer cells (scRNAseq data from Vazquez-Garcia et al. [22]) earmarked by expression of the EpCAM^{lo} signature. The threshold was set as ≥0.1 for the subsequent analyses. The EpCAM^{lo} signature was defined by genes that were identified as upregulated EpCAM^{lo} cells in both CAOV3 and OV90. EpCAM^{lo} upregulated genes were selected after differential expression analysis with EpCAM^{hi}/bulk cells using the cut-offs log2FC > 1.5 and padj <0.05. **B** UMAP plot of patient-derived ovarian cancer cells. Cells positive for the EpCAM^{lo} signature are highlighted in red and show enrichment within three clusters (#2, #4, and #6) labeled as EMT-like in the Vazquez-Garcia et al. study [22]. Please note that a substantial fraction of the EpCAM^{lo}-like cells appears to fall outside these clusters and is distributed throughout the UMAP. **C** Violin plots showing the distribution of EpCAM^{lo}-like cells (according to z-score) in different anatomical localization of ovarian cancers (adnexa, ascites, and non-adnexa). **D** Violin plots showing the expression levels of EMT-like signature of single cell cluster 2, 4 and 6. **E** Heatmap of hallmark gene sets across the EpCAMlo-like clusters in primary ovarian cancers (adnexa), ascites, and metastases (non-adnexa) based on the GSEA of three EpCAM^{lo}-like clusters and bulk (p val <0.05). Complete-linkage hierarchical clustering was used.

Tpm1.6/7-OE, and *Tpm1.8/9*-OE cells (Fig. 6A, B). Gene set enrichment analysis (GSEA) was then performed to allow the identification of specific signaling pathways and gene ontology functions characteristic of each of the above sample groups. As shown in the heatmap of Fig. 6C, *Tpm1.8/9* overexpression resulted in the pronounced activation of Hedgehog, Wnt/β-catenin, TGF-β, and Notch signaling, i.e. pathways known to be involved in the induction and regulation of EMT. Accordingly, EMT was also found to earmark these cells, as also shown by the distinct expression patterns of several EMT-related genes between *Tpm1.6/7*- and *Tpm1.8/9*-OE samples (Fig. 6D).

To functionally validate the activation of the canonical Wnt/β-catenin signaling pathway, we implemented TopFLASH reporter assays, as previously described [32]. As shown in Fig. 6E, a ~10-fold increase in luciferase activity was observed upon ectopic expression of the individual *Tpm1.8* and *1.9* isoforms. Likewise, Wnt signaling activity was significantly reduced upon *Tpm1.8/9* siRNA-driven knockdown.

Last, to evaluate the in vivo consequences of the ectopic expression of the specific tropomyosin isoforms on their invasive and metastatic capacity, intraperitoneal (IP) injections were performed with *Tpm1.6/7*- and *Tpm1.8/9*-OE cells to model late-stage ovarian cancer with intra-abdominal dissemination and ascites formation. As illustrated in Fig. 6F, G by IVIS imaging and quantification, the results showed that, somewhat surprisingly, IP transplantation of *Tpm1.6/7*-OE cells resulted in increased ascites volumes when compared with *Tpm1.8/9*-OE cells. Accordingly, pancreas and omental metastatic lesions were observed in mice injected IP with *Tpm1.6/7*-OE cells, whereas, no metastases were observed with *Tpm1.8/9*-OE cells (Fig. 6H). These apparently contradictory results are explained by MET repression in ovarian cancer cells constitutively expressing the EMT-inducing *Tpm1.8/9* isoforms. As both E-to-M and M-to-E transitions are necessary for ovarian cancer cell dissemination and metastasis formation, the constitutive induction of EMT in *Tpm1.8/9*-OE cells prevents them from efficiently colonizing the abdominal cavity. As a confirmation of this, "tumoroids" were derived and cultured from the ascites formed in the animals transplanted with ovarian cancer cells overexpressing the *Tpm1.6/7* or *Tpm1.8/9* isoforms. As shown in Fig. 6I, the morphology of these floating cell aggregates derived from *Tpm1.6/7*-OE cells was strikingly structured with multilayered epithelial cells and often with an internal lumen. In sharp contrast, tumoroids from mice transplanted with *Tpm1.8/9*-OE cells lacked any kind of structure and appeared as random aggregates of floating cells. Accordingly, colony formation analysis of single cells from the ascites-derived tumoroids showed that overexpression of the *Tpm1.6/7* results in a substantially increased colonization capacity when compared with *Tpm1.8/9*-OE cells (Fig. 6J).

Overall, these results show that the *TPM1* isoforms confer specific functional features on ovarian cancer cells. In particular *Tpm1.8/9* isoforms are strongly associated with EMT-inducing and inflammatory signaling pathways likely to underlie 'transcoelomic' dissemination of ovarian cancer cells and the formation of ascites. However, their ectopic overexpression is also a potential source of artifacts as EMT is equally essential as is MET in the formation of intra-abdominal metastases.

TPM1.8/9 isoforms confer resistance to taxane- and platinum-based chemotherapy and are expressed in ovarian cancer patients-derived ascites

To investigate the role played by *TPM1* alternative isoforms in women suffering from ovarian cancer, patient-derived tumor tissues were examined by immunohistochemistry (IHC) and in situ hybridization (ISH) analyses with isoform-specific antibodies and oligonucleotide probes. Among the high- and low-grade serous tumors analyzed ($n = 13$), *Tpm1.6/7* appeared to be consistently expressed in primary and metastatic lesions (Fig. 7A). In contrast, *Tpm1.8/9* expression was virtually undetectable above background levels, with few cases showing patchy and enhanced staining. Although based on an admittedly limited number of tumors, these observations seem to suggest that, while the *Tpm1.6/7* isoforms are mainly expressed in the bulk of epithelial tumor cells, their *Tpm1.8/9* counterparts are only rarely observed possibly in association with late and chemo-resistant stages of the disease.

The formation of intra-abdominal ascites predominantly occurs in stage III and IV ovarian cancer patients due to the spreading of tumor cells to the peritoneum and the obstruction of lymphatic vessels. As such, ascites-derived tumor cells may encompass quasi-mesenchymal and chemo-resistant cell types en route to the metastatic colonization of abdominal organs. Hence, we evaluated whether the *Tpm1.8/9* isoforms are transiently expressed in ovarian cancer patient-derived ascites when compared with *Tpm1.6/7*. We collected ascites samples from $n = 13$ patients and sorted the cellular content by FACS according to the [CD45(Lin)$^-$/CD90$^+$] and [CD45(Lin)$^-$/CD90$^-$] gates, encompassing immune/stromal and cancer cells, respectively (Fig. 7B, C). Total RNA was then extracted from the sorted cells for RTqPCR analysis of *TPM1* isoform expression. As shown in Fig. 7D, increased *Tpm1.8/9* expression was observed in the ascites-derived cancer cells (CD90$^-$) when compared with *Tpm1.6/7*. Among the immune/stromal cells (CD90$^+$), *Tpm1.6/7* expression levels were increased when compared with *Tpm1.8/9* (Fig. 7D).

Last, we employed expression profiles from the Cancer Genome Atlas (TCGA) project and the TCGA Splicing Variants Database (TSVdb; http://www.tsvdb.com/) and integrated the clinical follow-up data with the expression of *TPM1* (whole gene) and its isoforms. As shown in Supplementary Fig. 7 Suppl. 1, Kaplan–Meier analysis was borderline significant for the *TPM1* gene and its 1.6/7 isoforms, whereas the p values for *Tpm1.8/9* were both >0.05. The latter is not surprising in view of the very low expression level of these low-molecular weight isoforms in primary ovarian cancer from which the TGCA data are derived.

In view of these results pointing at the specific *Tpm1.8/9* upregulation in ovarian cancer ascites, i.e. at late and recurrent disease stages often in association with chemo-resistance and poor prognosis, we asked whether the same *TPM1* isoforms may confer resistance to the platinum- and taxane-based therapies commonly employed in the clinical management of ovarian cancer. We therefore cultured the OV90, COV504, and CAOV3 parental cell line in the presence of two distinct chemotherapeutic agents, namely cisplatin and paclitaxel, and then analyzed *TPM1* isoform expression in cells surviving the treatment. RTqPCR

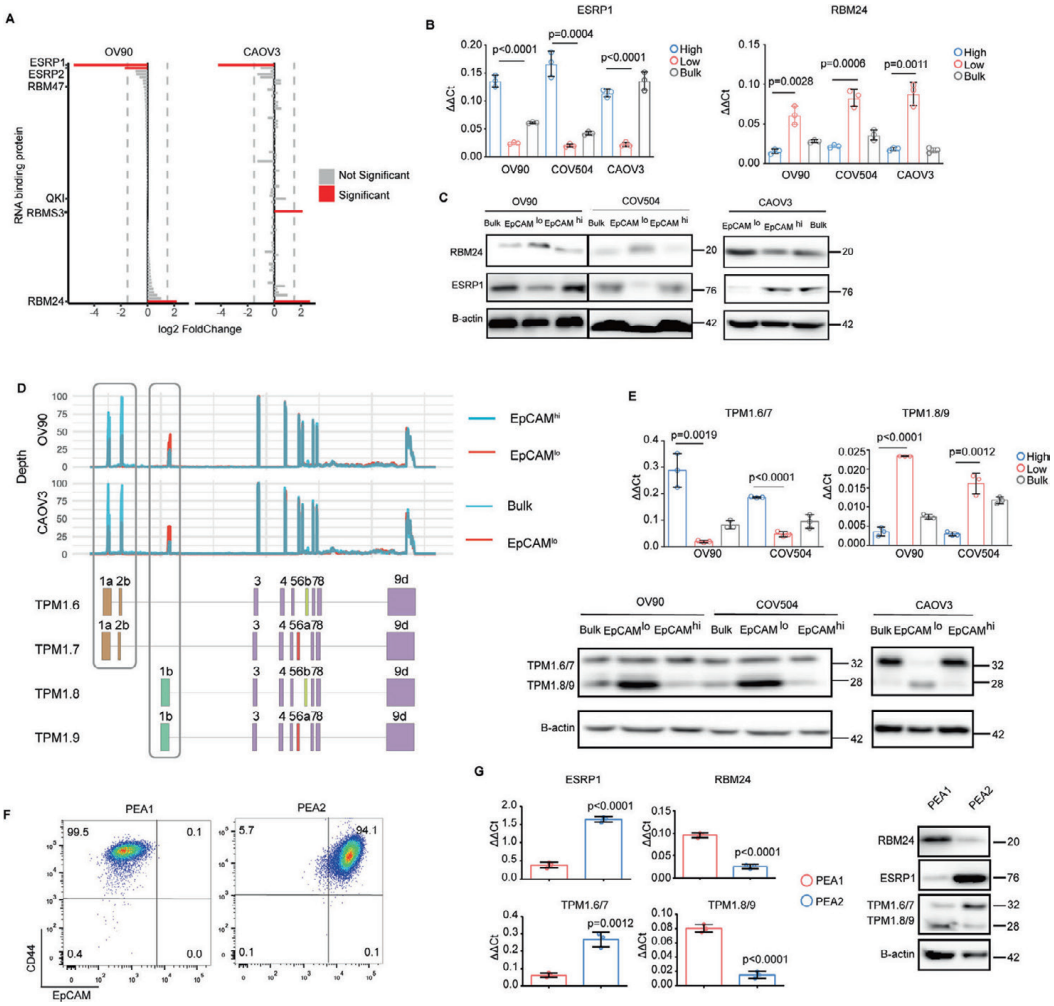


Fig. 3 Differential expression of RBPs *ESRP1* and *RBM24* regulates *TPM1* isoforms in HGS ovarian cancer. A Fold change analysis of RBPs [original list from reference [14]] differentially expressed between EpCAM^{lo} and EpCAM^{hi} cells in OV90 and CAOV3. Red bar indicates RBPs with significant differential expression (log2-fold change >2, and *P* value < 0.05). The dotted line represents the 1.5 absolute fold change cut-off. **B** RT-qPCR analysis of *ESRP1* and *RBM24* expression in OV90, CAOV3, and COV504 EpCAM$^{hi/lo}$ and bulk cells. GAPDH expression was employed as control (means ± SD, $n = 3$). *P* values are indicated. **C** Western blot analysis of ESRP1 and RBM24 expression in OV90, CAOV3, and COV504 EpCAM$^{hi/lo}$ and bulk cells. β-actin was used as loading control for western blots. **D**. *TPM1* exon peak plots relative to the AS analysis of RNAseq data obtained from OV90 EpCAM$^{hi/lo}$ and CAOV3 bulk/EpCAM^{lo} analysis. Each peak indicates the expression of specific exons; the height of each peak is indicative of the expression level of the specific exons. The exon-intron structure of the corresponding *TPM1* isoforms is depicted below the exon peak plot. Exon 1a and 2b are specific to *Tpm1.6/7* isoforms (brown), while exon 1b is specific to *Tpm1.8/9* (green). Exon 6a (red) is only present in the *Tpm1.7* and *Tpm1.9* isoforms; whereas, exon 6b (yellow) earmarks the *Tpm1.6* and *Tpm1.8* isoforms. With the exception of exon 6a and 6b, exons 3 to 9d are present in all *TPM1* isoforms. **E** RT-qPCR (histogram panels) and western analysis of *TPM1* isoform expression in OV90, CAOV3, and COV504 EpCAM$^{hi/lo}$ and bulk cells. *GAPDH* expression was employed as control. (Means ± SD, $n = 3$). *P* value is indicated. β-actin was used as loading control for western blots. **F** CD44/EpCAM FACS analysis of the ovarian cancer cell lines PEA1 and PEA2. EpCAM/CD44 positive and negative areas were identified using multiple isotype controls. The percentages of cells within each quadrant are indicated. **G** RT-qPCR (left histogram panels) and western (right panel) analysis of ESRP1, RBM24, and *TPM1* isoform expression in PEA1 and PEA2 ovarian cancer cell line; *GAPDH* expression was employed as control. (Means ± SD, $n = 3$). *P* values are indicated. β-actin was used as loading control for western blots.

analysis revealed a dramatic downregulation of *Tpm1.6/7* expression with both agents when compared with the untreated parental cells. In contrast, *Tpm1.8/9* expression appeared to increase in cells surviving both the cisplatin- and paclitaxel-treatment (Fig. 7E). Dose-response curves with OV90, COV504, and CAOV3 cells ectopically expressing the individual isoforms confirmed that *Tpm1.8/9*-OE cells displayed a 40-fold higher resistance ($IC_{50} = 3.266/3.414$ μM) than *Tpm1.6/7* OE cells ($IC_{50} =$

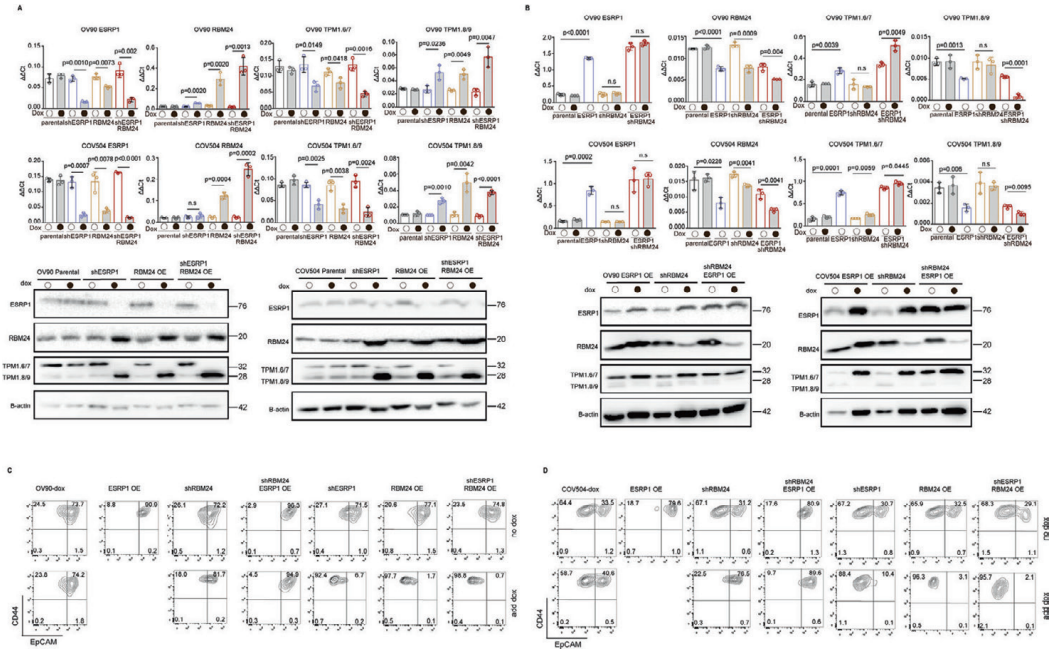


Fig. 4 The RBPs *ESRP1* and *RBM24* synergistically regulate *TPM1* isoforms and the relative proportion of EpCAM$^{hi/lo}$ cells. A RT-qPCR (histogram panels) and western (lower panel) analysis of ESRP1, RBM24, and *TPM1* isoform expression in *RBM24*-OE (overexpressing) and sh*ESRP1*-KD (knockdown) OV90 and COV504 ovarian cancer cell line; *GAPDH* expression was employed as control (Means ± SD, $n = 3$). β-actin was used as loading control for western blots. **B** RT-qPCR (histogram panels) and western (lower panel) analysis of ESRP1, RBM24, and *TPM1* isoform expression in *ESRP1*-OE (overexpressing) and sh*RBM24*-KD (knockdown) OV90 and COV504 ovarian cancer cell line; *GAPDH* expression was employed as control (Means ± SD, $n = 3$). β-actin was used as loading control for western blots. **C** CD44/EpCAM FACS analysis of *RBM24*-OE/KD and *ESRP1*-OE/KD OV90 cells. Cells were induced with 1 μg/mL doxycycline for 72 h before analysis. The relative percentages of EpCAM^{lo} and EpCAM^{hi} cells are indicated in each quadrant. **D** CD44/EpCAM FACS analysis of *RBM24*-OE/KD and *ESRP1*-OE/KD COV504 cells. Cells were induced with 1 μg/mL doxycycline for 72 h before analysis. The relative percentages of EpCAM^{lo} and EpCAM^{hi} cells are indicated in each quadrant.

0.078/0.043 μM) upon cisplatin treatment in OV90 (Fig. 7F, Supplementary Fig. 7 Suppl. 2). The same was also observed with paclitaxel ($IC_{50} = 1.28/1.66$ μM for *Tpm1.8/9*-OE; $IC_{50} = 0.017/0.015$ μM for *Tpm1.6/7*-OE cells). To validate this observation, we developed siRNA assays to selectively downregulate the *Tpm1.6/7* and *Tpm1.8/9* isoforms in parental OV90, COV504, and CAOV3 cells and assess their chemo-resistance. As shown in Fig. 7G, H and Supplementary Fig. 7 Suppl. 2, the specific downregulation of the *Tpm1.8/9* isoforms, validated both at the RNA and protein levels, reduced the cisplatin- and paclitaxel-specific IC_{50}, while the opposite was true for the siRNA-driven knockdown of *Tpm1.6/7*. Furthermore, increased resistance to cisplatin and paclitaxel was observed in EpCAM^{lo}-sorted cells upon ZEB1 knockdown (by SH) in the OV90 and COV504 cell lines (Supplementary Fig. 7 Suppl. 2).

TPM1.8/9-specific small molecule inhibitors for ovarian cancer therapy

Given the newly uncovered role played by the *Tpm1.8/9* isoforms in ovarian cancer and the broad spectrum of consequences at the cellular and molecular level that their specific inhibition may exert on cell motility and proliferation, EMT/MET, several oncogenic signaling pathways including Wnt, and therapy resistance, the development of small molecule antagonists may provide novel tools in the clinical management of late-stage ovarian cancer. Differences in the N-and C-termini between tropomyosin isoforms provide an opportunity to develop compounds that preferentially target specific isoforms [33–35]. Compounds that target the C-terminus of TPM3.1 have been shown to inhibit the function of this isoform both in vitro and in vivo by incorporating into the overlap junction between adjacent dimers in actin/Tpm3.1 co-polymers [36, 37]. Compounds targeting the N-terminus show similar activity [38]. Based on the differences in the N-terminal sequences of exon 1b in the *TPM1*, *TPM3*, and *TPM4* genes, compounds were developed that target *Tpm4.2* [39] and *Tpm1.8/1.9* (*European patent application No. 23187348.0*). Virtual screening of compound libraries with the N-terminal model of Tpm1.8/9 was used to identify compounds as potential binders. Targeting of *Tpm1.8/9* by virtual screening identified 6 candidate compounds that were tested for biological activity in human fibroblasts and OV90 cells based on their ability to disperse the target *Tpm1.8/9* away from actin-containing structures in the lamellipodium (Supplementary Fig. 8 Suppl. 1). Two compounds, *Tpm1.8/9*-1 (PubChem CID 6494468) and *Tpm1.8/9*-3 (PubChem CID 18973468) (hereafter for brevity referred to as compounds #1 and #3) were identified with activity in the low micro molar range. Of note, the effective concentration of an inhibitor is influenced by its affinity for the target, metabolism, and other factors, rather than the concentration of the target in the cell. The compounds do not show interaction with the target in cell-free biochemical assays, compatible with either a slow on-rate or fast off-rate resulting in a low affinity interaction with the tropomyosin dimer followed by incorporation into a high affinity binding state in the overlap junction during polymerization. This proposed model is based on previous studies of the incorporation of Tpm3.1 inhibitors into the overlap junction during polymerization.

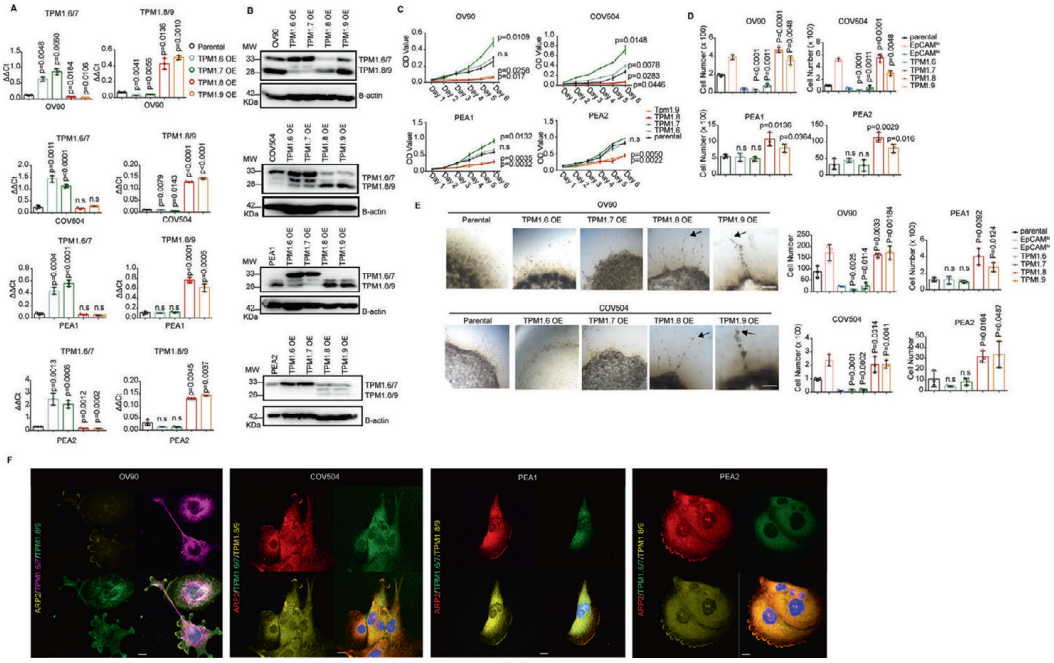


Fig. 5 Ectopic expression of *Tpm1.6/7* and *Tpm1.8/9* isoforms results in increased migration and invasion, and decreased cell proliferation. A RT-qPCR analysis of OV90, COV504, PEA1, and PEA2 ovarian cancer cell lines transduced to ectopically express the *Tpm1.6/7*-OE and *Tpm1.8/9*-OE isoforms; GAPDH expression was employed as control (Means ± SD, $n = 3$). *P* values are relative to the comparison with the parental cell lines. **B** Western analysis of OV90, COV504, PEA1, and PEA2 ovarian cancer cell lines transduced to ectopically express the *Tpm1.6/7*-OE and *Tpm1.8/9*-OE isoforms. β-actin was employed as loading control. **C** Proliferation assays of OV90, COV504, PEA1, and PEA2 ovarian cancer cell lines transduced to ectopically express the *Tpm1.6/7*-OE and *Tpm1.8/9*-OE isoforms. O.D. values are shown from day 1 to 6 (Means ± SD, $n = 3$). *P* values are relative to the comparison with the parental cell lines. **D** Transwell migration assay of OV90, COV504, PEA1, and PEA2 ovarian cancer cell lines transduced to ectopically express the *Tpm1.6/7*-OE and *Tpm1.8/9*-OE isoforms. 5×10^4 cells were plated on TC-coated membranes and left O/N. The number of cells that migrated to the lower side of the membrane were counted and plotted (Means ± SD, $n = 3$). *P* values are relative to the comparison with the parental cell lines. **E** Confocal images of OV90 and COV504 parental and *Tpm1.6/7/8/9*-OE cells seeded on collagen layers and incubated for 6 days. As indicated by the arrows, *Tpm1.8/9*-OE cells appear to invade the collagen layer collectively as narrow linear strands with "leader" and "follower" cells. Scale bar: 250 μm. The number of cells invading the collagen was quantified and plotted (Means ± SD, $n = 3$). *P* values are relative to the comparison with the parental cell lines. Plots relative to the PEA1 and PEA2 ovarian cancer cell lines were also calculated (bottom). **F** Immunofluorescence analysis of OV90, COV504, PEA1 and PEA2 parental cells with antibodies directed against ARP2, *Tpm1.6/7* and *Tpm1.8/9*. Nuclei were visualized by DAPI staining of DNA. Scale bar: 5 μm.

In order to functionally validate the two compounds, we first determined their most effective concentration in the 0 to 10 μM range both on the OV90 parental cell line and on its sorted EpCAM^{lo} subpopulation. RTqPCR and western analysis showed that compounds #1 and #3 do not affect *Tpm1.6/7* or *Tpm1.8/9* expression at either the RNA or protein level (Fig. 8A, B). This is not unexpected since displacement of *Tpm1.8/9* from actin filaments to the soluble pool in cells does not result in *Tpm1.8/9* turnover [40]. However, the expression of EMT-related genes was affected by both compounds: *ZEB1* and *VIM* expression was suppressed, whereas, EpCAM was increased, also in agreement with the expected MET-inducing effects of *Tpm1.8/9* antagonists (Fig. 8A, B). Furthermore, the treated OV90 and CAOV3 bulk and EpCAM^{lo} cells showed a 2–5-fold reduction of the cisplatin- and paclitaxel-specific IC_{50} values when compared with untreated cells (Fig. 8C–F). Last, compounds #1 and #3 dramatically reduced Wnt/β-catenin signaling as shown by TopFLASH reporter assays (Fig. 8G). As expected, the Wnt-inhibiting effects of the compounds are more clearly illustrated by the EpCAM^{lo} cells because of their Wnt- and EMT-hi transcriptional profiles when compared with the parental OV90 and COV504 cell line where a majority of EpCAM^{hi} (Wnt-lo) and EpCAM^{lo} cells coexist.

Overall, these in vitro results suggest the potential for the future development of novel therapeutic strategies for high-grade serous ovarian cancer centered around the *Tpm1.8/9* isoforms.

DISCUSSION

It is generally accepted that epithelial-mesenchymal plasticity (EMP), i.e. the transient and reversible identity conferred to cancer cells by EMT/MET processes, underlies tumor progression, local invasion, distant metastases, therapy resistance, and immune evasion. As elegantly shown by Cook and Vanderhyden [41, 42], EMT is a highly variable process with a very broad spectrum of upstream signals from the TME, intracellular regulatory mechanisms, and downstream effectors, the identity of which largely depends on the tumor type and its micro- and macro-environment. The same is true when it comes to the nature of the epigenetic mechanisms that underlie EMP [43]. We have recently shown that the differential expression of RNA-binding proteins between the epithelial tumor bulk and subpopulations of quasi-mesenchymal colon cancer cells underlies alterative splicing at a variety of target genes known to play functional roles in EMT, metastasis, and resistance to chemotherapy [17, 21]. Here, we

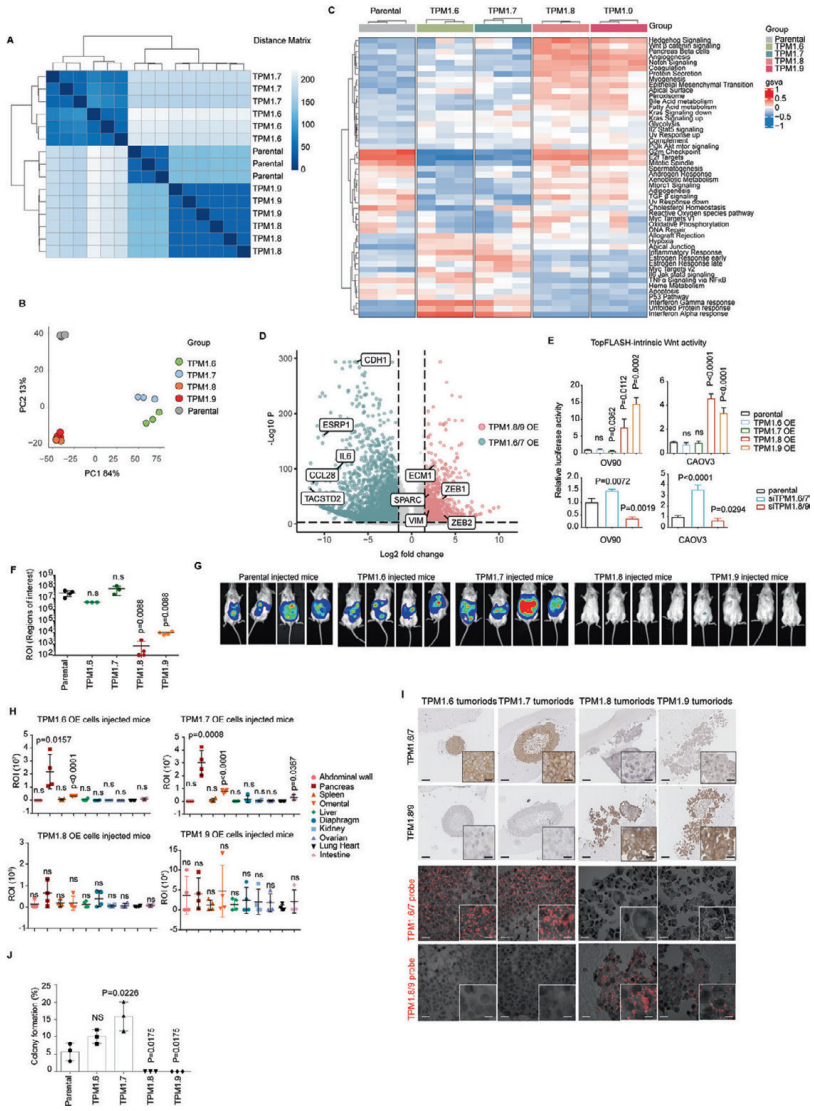
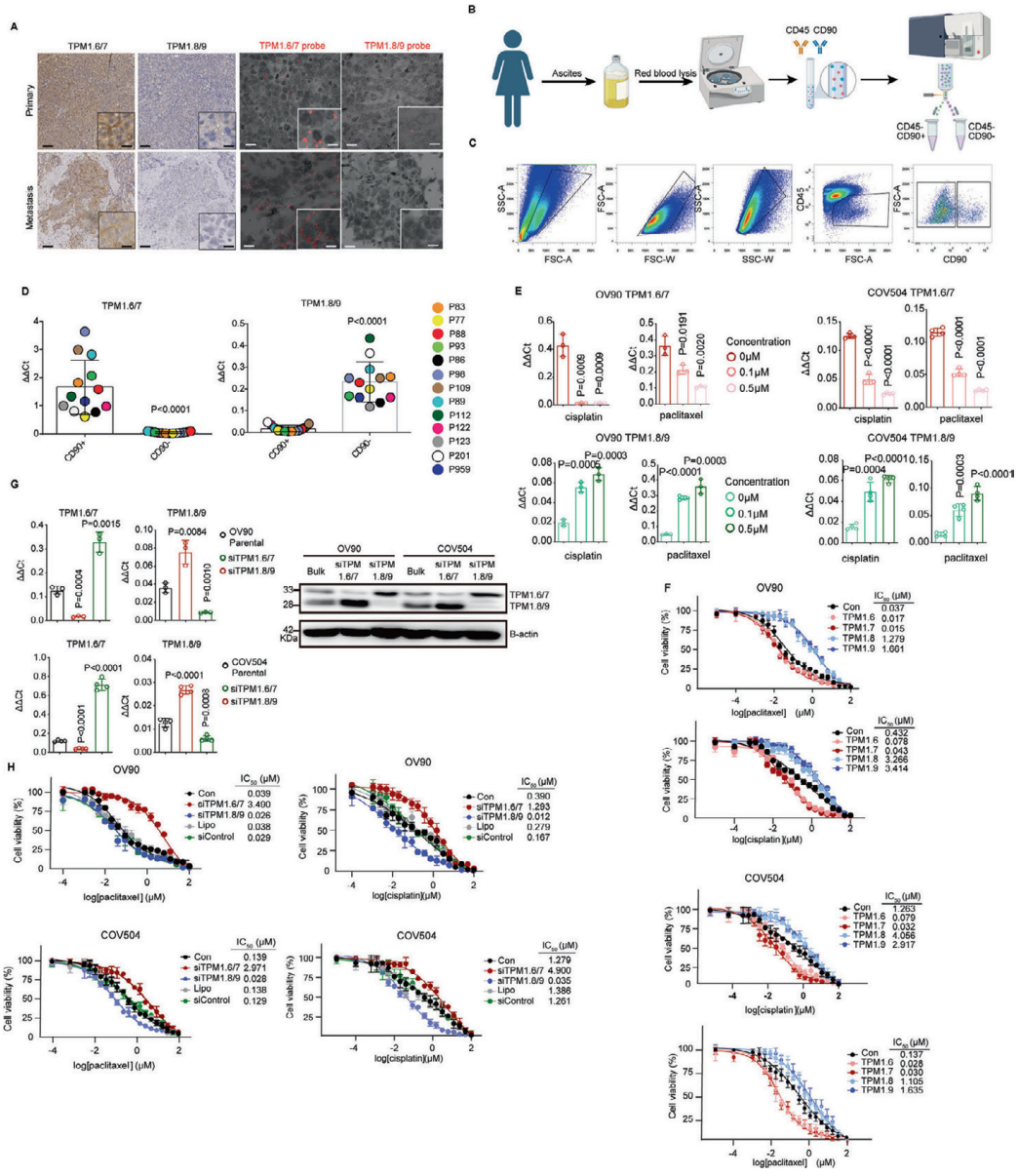


Fig. 6 RNAseq analysis revealed *TPM1* isoforms function in Wnt pathway and contribute to metastasis in vivo. A Hierarchical clustering of the RNAseq data relative to *Tpm1.6/7/8/9*-OE OV90 cells. Complete-linkage hierarchical clustering was used. **B** Principal component analysis (PCA) of RNAseq profiles from parental and *Tpm1.6/7/8/9*-OE OV90 cells. **C** Hallmarks pathways based on the Gene Set Enrichment Analysis (GSEA) of parental and *Tpm1.6/7/8/9*-OE OV90 cells. The heatmap only includes significantly altered pathways, with NES > 1, and *P* value < 0.05. Complete-linkage hierarchical clustering was used. **D** Volcano plots showing differentially expressed genes between *Tpm1.6/7*-OE (left, green) and *Tpm1.8/9*-OE (pink, right) OV90 cells (abs LFC > 1.5, *P* value < 0.01). **E** TOP-Flash luciferase reporter analysis of Wnt signaling activity in *Tpm1.6/7/8/9*-OE (upper histogram) and upon knockdown by siRNA of *Tpm1.6/7* and *Tpm1.8/9* in OV90 and CAOV3 cells. *P* values are relative to the comparison with the parental cell lines (Means ± SD, $n = 3$-4). **F** Quantification of IVIS bioluminescence signals obtained from NSG mice injected IP with *Tpm1.6/7/8/9*-OE OV90 cells. Recipient animals were sacrificed 5 wk after injection. Four mice were analyzed for each type of transplanted cells. Data are presented as mean values ± SD. Y-axis meaning ROI (regions of interests) are user-defined areas within IVIS optical imaging. **G** Examples of IVIS bioluminescence signals from *Tpm1.6/7/8/9*-OE cell-injected mice at 5 wk after injection. The Spectrum in vivo imaging system was employed. For in vivo imaging purposes, mice were injected IP with D-luciferin (150 mg kg^{-1}). **H** IVIS bioluminescence signals relative to specific organs from mice transplanted with the *TPM1* isoform-OE cells. Y-axis meaning ROI (regions of interests) are user-defined areas within IVIS optical imaging. **I** *Tpm1.6/7* and *Tpm1.8/9* IHC (upper panels) and ISH (BaseScope; lower panels) analyses of tumoroids derived from ascetic fluids from mice transplanted with the *TPM1* isoform-OE cells. Scale bars: 100 μm (large panels) and 15 μm (inlets) for IHC; 20 μm and 5 μm (inlets) for ISH. **J** Colony formation assay relative to cells derived from tumoroids obtained from mice transplanted with the *TPM1* isoform-OE cells (Means ± SD, $n = 3$). Y-axis means the percentage of tumoroids single cells form into colonies.



applied a similar strategy toward the identification and functional characterization of genes whose isoform patterns are altered during EMT/MET in ovarian cancer. Comparison of alternatively expressed gene isoforms between epithelial and quasi-mesenchymal colon and ovarian cancer cells revealed few common targets with many ovarian-specific events. This possibly reflect of the distinct modalities of local dissemination and metastatic colonization characteristic of these types of carcinoma. Whereas most colon cancer metastases follow a hematogenous route, unique for ovarian cancer is the 'transcoelomic' dissemination of tumor cells and the formation of ascites fluid in the abdominal and pelvic cavity which provide a favorable tumor microenvironment (TME) for the disseminated cancer cells. Nonetheless, previous studies have indicated that EMT does contribute to ovarian cancer progression and to chemotherapy resistance [44]. Hence, even though through distinct cellular and molecular mechanisms, EMT does play a key role in ovarian cancer metastasis and chemo-resistance. Gene ontology analysis of the ovarian cancer specific isoform targets (Supplementary Fig. 3, Supplement 1B-E) revealed an extremely broad spectrum of

Fig. 7 *Tpm1.8/9* isoforms are enriched in malignant ascites from ovarian cancer cells and confer resistance to platinum- and taxane-based therapies. A Examples of IHC (left panels) and ISH (BaseScope; right panels) analyses of patient-derived ovarian cancers with antibodies (IHC) and oligonucleotides probes (ISH) specific for the *Tpm1.6/7* and *Tpm1.8/9* isoforms. Ovarian cancer tissues were obtained from a primary tumor and a metastasis (without chemotherapy). Scale bar: 100 μm and 15 μm (inlets) for IHC; 20 μm and 5 μm (inlets) for ISH. **B** Schematic flowchart of the analysis of ascites from late-stage ovarian cancer patients. **C** FACS analysis and sorting strategy of $CD45^{-}CD90^{+}$ and $CD45^{-}CD90^{-}$ cells from patient-derived ascites. From left to right: FSC-A/SSC–A, FSC-W/FSC-A, SSC-W/SSC-A, FSC-A/$CD45^{-}$ and $CD45^{-}CD90^{+/-}$ single cell gates. **D** RT-qPCR analysis of *Tpm1.6/7* and *Tpm1.8/9* expression in sorted $CD45^{-}CD90^{+}$ and $CD45^{-}CD90^{-}$ cells from patient-derived ascites ($n = 13$) sorted by FACS; GAPDH expression was employed as control (Means ± SD). **E** RT-qPCR analysis of *Tpm1.6/7* and *Tpm1.8/9* expression in OV90 and COV504 cells exposed to cisplatin and paclitaxel cells; GAPDH expression was employed as control (Means ± SD, $n = 3$-4). **F** Dose-response curves relative to *Tpm1.6/7/8/9*-OE cells grown in the presence of different concentrations of paclitaxel and cisplatin (log scale and cell viability on the x and y axis, respectively). IC_{50} values were calculated from biological triplicates to quintuplicates for each experiment (Means ± SD, $n = 3$–5). **G** RT-qPCR (left histogram panels) and western (right) analysis of TPM1 isoform expression in si*Tpm1.6/7* and si*Tpm1.8/9* knockdown OV90 and COV504 cells; GAPDH expression was employed as control (Means ± SD, $n = 3$-4). β-actin was employed as loading control for the western blots. **H** Dose-response curves of si*Tpm1.6/7* and si*Tpm1.8/9* knockdown OV90 and COV504 cells cultured in the presence of different concentrations of paclitaxel (left) and cisplatin (right) concentrations (log scale and cell viability on the x and y axis, respectively). IC_{50} values were calculated from biological triplicates to quintuplicates for each experiment (Means ± SD, $n = 3$–5).

biological processes, molecular functions, and cellular components likely to collectively contribute to the transition to quasi-mesenchymal ovarian cancer cells capable of local invasion and distant metastatic colonization.

Previously, AS targets likely to contribute to ovarian cancer progression have been reported such as *BCL2L12* [45] and *ECM1* [46]. Here, among the EMT-related isoforms, we selected *TPM1* because of its function in the regulation of cell motility through cytoskeletal modifications [30, 31], and its alleged role as a tumor suppressor and even oncogene in multiple cancer types [47, 48]. *TPM1* isoforms are found in multiple tissues [30, 49]. Our results establish a direct causative relation between the *TPM1.8/9* isoforms and the activation of EMT in ovarian cancer cells. Whether the RBM24 and ESRP1 RBPs, differentially expressed between epithelial and quasi-mesenchymal cells, also play a direct role as co-transcriptional factors of the TPM1a/b promoters, is at present unclear. Analysis of the FANTOM database, while failing to identify known RBP binding sites, did reveal the presence of known EMT-TFs. The results obtained with the luciferase reporter constructs are admittedly inconclusive due to the confounding role of EMT upon ectopic expression or downregulation of the ESRP1 and RBM24 RBPs. Future ChIP or CUT&RUN analyses will elucidate the mechanisms underlying the regulation of Tpm1.6/7 and Tpm1.8/9 isoforms directly and primarily by EMT-TFs or with the RBPs as co-transcriptional factors.

Because of their localization to the lamellipodia and functional role in cell motility, the *TPM1.8/9* isoforms are likely to facilitate dissemination from the primary tumor to the intra-abdominal cavity, as also shown by their enrichment in patient-derived ascites. Allegedly as a consequence of their EMT-inducing capacity, ectopic expression of the low-molecular weight *TPM1* isoforms confers resistance to taxane- and platinum-based chemotherapy. Accordingly, *Tpm1.8/9* are also found to be expressed at very low levels, if any, in primary ovarian cancers and their metastases albeit increased in malignant ascites.

Notwithstanding the above, our attempts to provide in vivo evidence for the metastatic capacity of ovarian cancer cells overexpressing the *Tpm1.8/9* isoforms failed to show any increase when compared with *Tpm1.6/7*. This apparently contradictory result can be explained by the transient and reversible nature of EMT along the multistep events that underlie dissemination and metastasis. As previously proposed by Thomas Brabletz and collaborators [50], while the acquisition of quasi-mesenchymal characteristics is required for local invasion and systemic dissemination, METs are equally essential for the colonization of distant organs. In the OV90 cells overexpressing the EMT-inducing *Tpm1.8/9* isoforms, MET is inhibited thus negatively affecting their metastatic potential. The presence of tumor cells with a strongly compromised colony formation capacity in the malignant ascites from recipient mice transplanted IP with *Tpm1.8/9*-OE cells supports the hypothesis according to which MET suppression in these cells negatively affects their metastatic potential. Another indication of the potential off-target artifacts caused by the non-physiological expression levels of *TPM1* isoforms in these cells became apparent in the analysis of their RNAseq profiles. While several inflammation pathways were upregulated together with *Tpm1.8/9* in the $EpCAM^{lo}$ ovarian cancer cells when compared with their epithelial counterpart, ectopic *Tpm1.6/7* expression resulted in the activation of similar inflammation-related pathways, e.g. IL6/Jak/Stat3, IFN, and TNF, when compared with *Tpm1.8/9* OE cells.

Because of the apparent multifunctional role of *TPM1* isoforms in ovarian cancer malignancy and resistance to therapy, *Tpm1.8/9* forms a potentially relevant therapeutic target. As shown here, the development of small molecule inhibitors will likely prevent EMT and reduce cell motility, and simultaneously inhibit the activation of key signal transduction pathways such as Wnt, known to play a central role in ovarian cancer stemness, EMT, and chemoresistance [51]. Last, combined treatment with conventional taxane- and platinum-based chemotherapies may increase their therapeutic efficacy by antagonizing chemoresistance.

MATERIALS AND METHODS

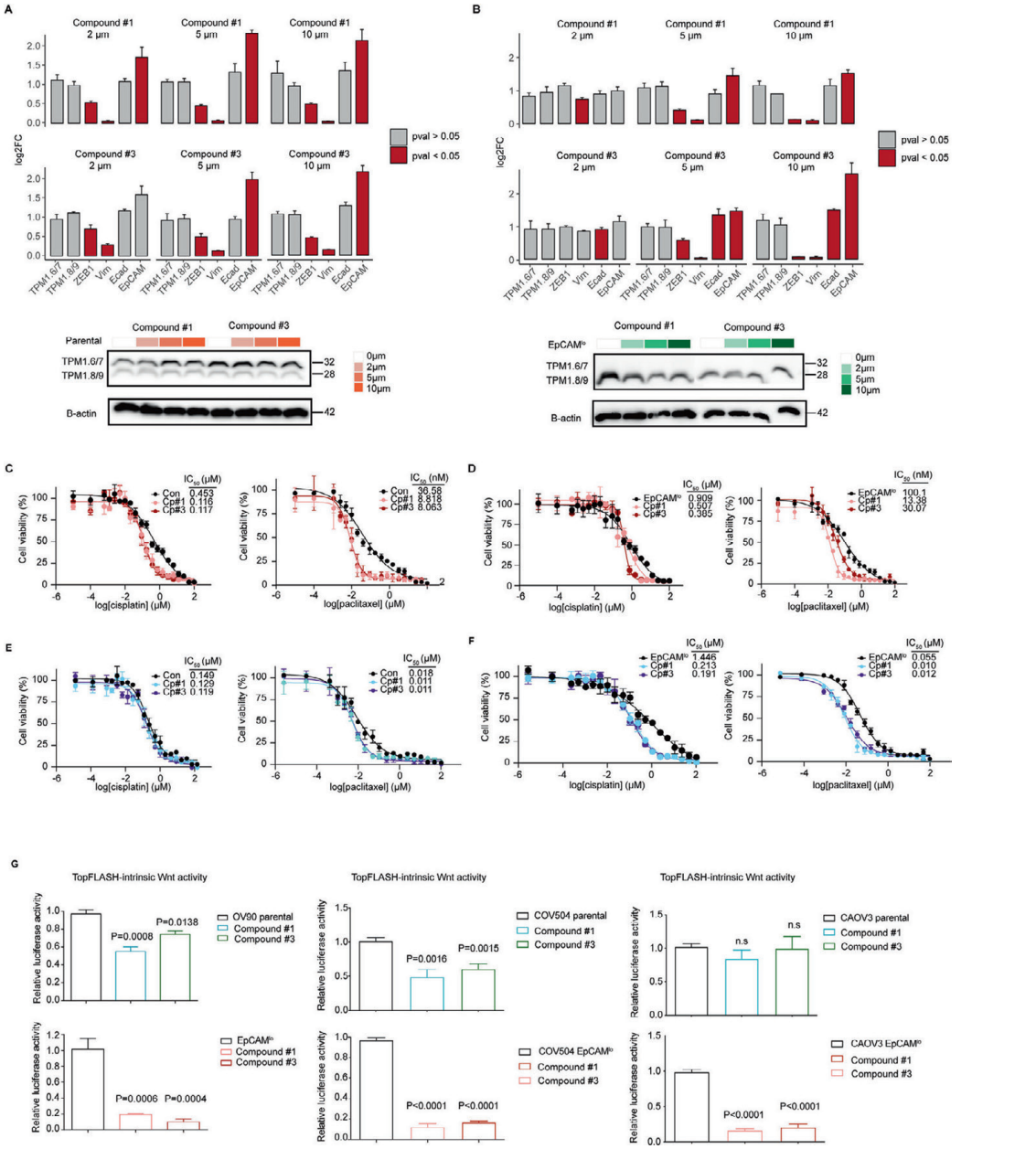
Cell cultures

The human ovarian cancer cell line OV90, obtained from the American Type Culture Collection (ATCC), was cultured in a 1:1 mixture of MCDB 105 medium (M6395; Sigma Aldrich containing 1.5 g/L sodium bicarbonate) and Medium 199 (31150022; Thermo Fisher Scientific containing 2.2 g/L sodium bicarbonate) supplemented with 15% heat inactivated fetal bovine serum (FBS; #16140071, Thermo Fisher Scientific) and 1% Penicillin/Streptomycin (Pen/Strep; penicillin: 100 U/mL, streptomycin: 100 μg/mL; 15140122 Thermo Fisher Scientific). CAOV3 (ATCC), SKOV3 [European Collection of Authenticated Cell Cultures (ECACC) via Sigma], COV504 (ECACC), HEK293T (ATCC) cell lines were cultured in DMEM medium (11965092, Thermo Fisher Scientific) supplemented with 10% heat inactivated FBS, 2 mM L-glutamine (200 mM; 25030081; Thermo Fisher Scientific), and 1% Pen/Strep. PEA1 (ECACC) and PEA2 (ECACC) cell lines were cultured in RPMI 1640 medium (61870036, Thermo Fisher Scientific) with 10% FBS, 1% Pen/Strep, and 2 mM L-glutamine.

The identity of each cell line was confirmed by DNA fingerprinting with microsatellite markers (Amelogenin, CSF1PO, D13S317, D16S539, D5S818, D7S820, THO1, TPOX, vWA, D8S1179, FGA, Penta E, Penta D, D18S51, D3S1358, D21S11) and compared with the analogous data provided by ATCC, EACC, and https://web.expasy.org/cellosaurus/ (data not shown).

Plasmid transfection and lentiviral transduction

cDNAs encoding *Tpms 1.6, 1.7, 1.8* and *1.9* were excised from the bacterial expression vectors pGEX or pET [from P.W.G. [52]] and cloned into the mammalian expression vector pcDNA3.1(+). Stable transfections of *ESRP1* (Sino Biological plasmid # HG13708-UT) and *Tpm1.6/7/8/9* expression vectors were performed with the FuGENE HD transfection reagent (Promega, E2311) according to the manufacturer's protocol, and selected with Geneticin (#10131035, Thermo Fisher Scientific).



The inducible pSLIK-RBM24 vector was constructed using pDONR 233-RBM24 as entry plasmid (Horizon, #OHS6084) following Gateway Cloning instructions (11791-020, Thermo Fisher Scientific). The inducible shZEB1 lentiviral vector was obtained as described in our previous study [21]. For all of the above inducible vectors, *ESRP1* (Horizon, V3THS_335722), sh*RBM24* (Horizon, V3SH11240-225117283) lentiviral constructs were packaged by psPAX2 (Addgene # 12260) and pMD2.G (Addgene # 12259) into HEK293T cells. The virus-containing supernatant was collected 24 h after transfection, filtered, and used to infect the OV90 and COV504 cell lines. Selection was applied with 750 ng/mL puromycin (#ant-pr-1, InvivoGen) or 800 μg/mL Geneticin for 1–2 wk. Validation of ectopic expression and knockdown of target genes was conducted 72 h after transfection by qPCR and western blot.

Luciferase reporter constructs

The promoter regions of TPM1 exons 1a and 1b were used, as reported by Savill et al. [53, 54]. Amplified PCR products were digested with XhoI and

Fig. 8 Small molecule inhibitors directed against *Tpm1.8/9* isoforms antagonize their effects on EMT, Wnt signaling, and resistance to chemotherapy. A Upper panels: RT-qPCR analysis of *TPM1* isoforms and EMT-related gene expression in OV90 parental cells cultured for 24 h in the presence of compound #1 or #3 at 0, 2, 5, and 10 μM. The values were calculated by normalizing with the untreated cells. *P* values < 0.05 are shown by red bars while gray bars indicate lower values; GAPDH expression was employed as control (Means ± SD, $n = 3$). Lower panels: western analysis of *TPM1* isoform expression in OV90 parental cells cultured for 24 h in the presence of compound #1 or #3 at 0, 2, 5, and 10 μM. B-actin was used as loading control for western blots. **B** RT-qPCR analysis of *TPM1* isoforms and EMT-related gene expression in OV90 EpCAM^{lo} cells cultured for 24 h in the presence of compound #1 or #3 at 0, 2, 5, and 10 μM. The values were calculated by normalizing with the untreated cells. *P* values < 0.05 are shown by red bars while gray bars indicate lower values; GAPDH expression was employed as control (Means ± SD, $n = 3$). Lower panels: western analysis of *TPM1* isoform expression in OV90 EpCAM^{lo} cells cultured for 24 h in the presence of compound #1 and #3 at 0, 2, 5, and 10 μM. β-actin was used as loading control for western blots. **C** Dose-response curves of parental OV90 cells treated with compound #1 or #3 in the presence of different paclitaxel and cisplatin concentrations. IC_{50} values were calculated from biological triplicates for each experiment (Means ± SD, $n = 3$). **D** Dose-response curves of OV90 EpCAM^{lo} cells treated with compound #1 or #3 in the presence of different paclitaxel and cisplatin concentrations. IC_{50} values were calculated from biological triplicates for each experiment (Means ± SD, $n = 3$). **E** Dose-response curves of parental CAOV3 cells treated with compound #1 or #3 in the presence of different paclitaxel and cisplatin concentrations. IC_{50} values were calculated from biological quadruplicates for each experiment (Means ± SD, $n = 4$). **F** Dose-response curves of CAOV3 EpCAM^{lo} cells treated with compound #1 or #3 in the presence of different paclitaxel and cisplatin concentrations. IC_{50} values were calculated from biological quadruplicates for each experiment (Means ± SD, $n = 4$). **G** TOP-Flash luciferase reporter analysis of Wnt signaling activity in OV90, COV504 and CAOV3 parental (upper panel) and EpCAM^{lo} (lower panel) cells treated with compound #1 or #3 (Means ± SD, $n = 3$–5).

SacI, and then ligated into XhoI- and SacI-digested pGL4.10 luciferase reporter plasmid (Promega, Southampton, UK). Constructs were validated by sequencing.

siRNA transfection

siRNA target sequence for human *Tpm1.6/1.7* was 5′-AAGCTGGAGCTGG-CAGAGAAA-3′ (codons 70-76 of *Tpm1.6/1.7* in exon 2b) and for human *Tpm1.8/1.9* was 5′-CGAGAGGAAGCTGAGGGAGAC-3′ (codons 38-44 of *Tpm1.8/1.9* in exon 1b). *Tpm* siRNAs (Horizon Discovery, Waterbeach UK) and control siRNA (#4390843, Thermo Fisher Scientific) were transfected by Lipofectamine RNAiMAX (#13778150, Invitrogen) according to the manufacturer's instructions. Seventy-two hrs. following siRNA transfection, cells were collected for RNA and protein analysis. To evaluate the effect of gene knockdown on drug resistance, cells were seeded 24 h. after siRNA transfection at a density of 5000 cells/well, followed by 3 days of incubation in the presence of cisplatin (#PHR1624, Sigma-Aldrich) or paclitaxel (#S1150, Selleck Chemicals).

RT-qPCR and PCR analyses

Total RNA was isolated using the TRIzol reagent (Thermo Fisher Scientific, 15596018) followed by reverse transcription using high-capacity cDNA reverse transcription kit (Life Technologies, 4368814), according to the manufacturer's instructions. RT-qPCR was performed using the Fast SYBR Green Master Mix (#4368708, Thermo Fisher Scientific) on an Applied Biosystems StepOne Plus Real-Time Thermal Cycling Research device, with three replicates for each analysis. Relative gene expression was determined by normalizing the expression of each target gene to that of *GAPDH*. Results were analyzed using the 2-(ΔΔCt) method. RT-qPCR primers are listed in Supplementary Table 4.

Western analysis

Cells were lysed in 2X Laemmli buffer (4% SDS, 48% Tris 0.5 M pH6.8, 20% glycerol, 18% H_2O, bromophenol blue and 10% 1 M DTT), and subjected to SDS-PAGE, followed by transfer onto polyvinylidene fluoride (PVDF) membranes (Bio-Rad). After blocking with 5% milk in TBS-Tween, the membranes were incubated with primary antibodies directed against ESRP1 (1:1000, # PA5-25833, Thermo Fisher Scientific), RBM24 (1:100, #18178-1-AP, Proteintech), *TPM1* (1:1000, #MBS127505, MyBioSource) and β-actin (1:2000, #4970, Cell Signaling). The secondary Ab's were a goat anti-mouse immunoglobulins/HRP (1:10000, #P0448, DAKO), and a goat anti-rabbit immunoglobulins/HRP (1:10000, #P0161, DAKO). Detection was by the Pierce ECT western blotting substrate (#34578, Thermo Fisher Scientific) using the Amersham AI600 imager (GE Healthcare).

Flow cytometry analysis and sorting

Single-cell suspensions in PBS supplemented with 1% FBS were incubated with anti-EpCAM-FITC (1:20, #GTX30708, Genetex), and anti-CD44-APC (1:20, #559250, BD Pharmingen) antibodies for 30 min on ice and analyzed on a FACSAria III Cell Sorter (BD Biosciences). CD44^{hi}EpCAM^{hi} and CD44^{hi}EpCAM^{lo} OV90 and CAOV3 cells were sorted and incubated in a humidified atmosphere at 37 °C with 5% CO_2 for 3–5 days before RNA or protein were collected as described above. See Fig. 1 for a more detailed protocol and gating specifications.

Patient-derived ascites were first washed 1-2 times with 1 × RBC lysis buffer [150 mM NH_4Cl (#7173-51-5, Sigma Aldrich), 10 mM $KHCO_3$ (#298-14-6, Sigma Aldrich), 100 μM EDTA (#60-00-4, Sigma Aldrich)] to remove erythrocytes. The cell pellets were then labeled with anti-CD90 (Brilliant® Violet 421; 1:20, #328122, Clone:5E10, BioLegend), anti-CD45-APC (1:20, #304037, Clone: HI30, BioLegend), and SYTOX™ Red (1:1000, #S34859, Thermo Fisher Scientific) antibodies, and sorted by FACS. RNA from CD45$^-$CD90$^-$ and CD45$^-$CD90$^+$ cells was isolated directly after FACS sorting.

Cell proliferation assays

To analyze cell proliferation rates, 2×10^3 ovarian cancer parental cells and *Tpm1.6/7/8/9*-OE cells were plated into 96-well plates and incubated at 37 °C, 5% CO_2. After 24 h. (day 1), cells were incubated at 37 °C, 5% CO_2 for 3 h. in culture medium supplemented with 0.45 mg/mL MTT [(3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Sigma-Aldrich]. The 96-well plates were then centrifuged at 1000 rpm for 5 min and the culture medium removed. O.D. reading was performed at 595 nm with a microplate reader (Model 550, Bio-Rad). Background measurements were subtracted from each data point. Experiments were performed in triplicate for each individual cell line.

Cell migration and invasion assays

Migration assays were conducted using 8 μm pore PET Transwell inserts (#353097, BD Falcon™) and TC-treated multi-well cell culture plates (#353047, BD Falcon™). 5×10^4 cells were seeded in the upper chamber with 100 μL of serum-free culture medium. Culture medium supplemented with 10% FBS was used as a chemoattractant in the lower chamber. After 24 h, cells that migrated to the lower chamber were fixed with 4% PFA (#9713.9010, VWR Chemicals), stained with 0.1% Trypan Blue solution (#15250061, Thermo Fisher Scientific) and counted using a microscope.

The invasion assays were conducted as described here above with the only addition of 200 μL Matrigel mixed with 10–20 μl of a 0.01 M Tris (pH 8.0)/0.7% NaCl solution on top of the Transwell-Clear insert and incubated at 37 °C for 2 h. After having removed the excess liquid, 5×10^4 cells were added to the well in 150 μL serum-free medium. Next, 0.5 mL of culture medium supplemented with 10% FCS was added. After 24 h, the cells that invaded the lower chamber were fixed with 4% PFA and stained with 0.1% Trypan Blue solution. Cells were counted using a microscope.

200 μL collagen mix [7.5% 10 ×PBS, 57% collagen, 0.1% 1 M NaOH, 9.5% H_2O and 25% culture medium] were plated into scaffolds and incubated at 37 °C for 1 h. 5×10^4 cells were added on top of the solid collagen and incubated at 37 °C. After 1 week, the entire scaffold with cells and collagen was fixed with 4% PFA and embedded in paraffin. 4 μm sections were mounted and counterstained with Hematoxylin. Slides were dehydrated and mounted in Pertex (#00811, Histolab).

Immunohistochemistry – cultured cells

0.5-1×10^4 cells were plated into 24-well plates containing glass cover slips coated with 0.2% gelatin. After 6–24 h, culture medium was removed, ice-

cold methanol added to each well and plates incubated for 20 min at 4 °C. Cells were washed twice with PBS, and incubated in 0.2% Triton for 20 min with rotation. Cells were blocked in 2% FBS in PBS for 1 h. Primary antibodies for *Tpm1.6/7* (1:200, from P.W.G.), *Tpm1.8/9* (1:200, from P.W.G.), Arp2 (1:200, #ab47654, Abcam) were added and incubated O/N at 4 °C. Cells were washed twice with PBS, and secondary antibodies [goat anti-rat Alexa Fluor 488 conjugate (1:250, #A10528, Life Technologies); goat anti-rabbit Alexa Fluor 546 conjugate (1:250, #A11035, Life Technologies); goat anti-mouse Alexa Fluor 647 conjugate (1:250, #A32728, Life Technologies)] and DAPI (#D1306, Thermo Fisher Scientific) added. Slides were mounted using VECTASHIELD® (#H100010, VECTOR laboratories) and cells imaged using a LSM-700 (Zeiss) with 20×, 40×, 63× lenses. Images were analyzed using ImageJ.

Immunochemistry – patient samples

Formalin-fixed, paraffin embedded (FFPE) ovarian cancer patient tissue blocks were obtained from the Department of Pathology, Erasmus Medical Center, Rotterdam. The average fixation time was 1–2 y. 4 µm sections were mounted on slides, dewaxed with Xylene (#28979.294, VWR Chemicals) and hydrated. Antigen retrieval was performed in Tris-EDTA buffer (pH 9.0) using a pressure cooker procedure. Slides were incubated in 3% hydrogen peroxidase (#95321, Sigma Aldrich) at RT for 10 min and blocked with 5% milk (#115363, Millipore) in PBS-Tween (#P1379, Sigma Aldrich) for 30 min. Immunohistochemistry was performed using antibodies directed against *Tpm1.6/7* (1:100), *Tpm1.8/9* (1:100) and mitochondria (1:100, #MAB1273, Sigma Aldrich) followed by the EnVision Plus-HRP system (Dako). Slides were incubated with primary antibodies O/N at 4 °C, washed twice with PBS-Tween and incubated with Rat EnVision+ System-HRP (#P0405, Dako) or Mouse EnVision+ System-HRP (#K4007, Dako) for 30 min. Slides were counterstained with Hematoxylin (#MHS16, Sigma Aldrich).

In vivo study

Mouse experiments were performed according to the Code of Practice - Animal Experiments in Cancer Research, Netherlands Inspectorate for Health Protection, Commodities and Veterinary Public Health, and the Animal Experiment Committee (DEC). 6–8-week-old NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) female mice were used. Animal randomization was employed. 50 µl PBS containing 1 ×10^5 OV90 cells overexpressing *Tpm1.6/7/8/9* isoforms was injected IP into each mouse. 150 mg/kg D-luciferin (#L2916, Invitrogen) was injected IP for bioluminescence signal. After 10 min of isoflurane-induced anesthesia (#B506, Zoetis), the bioluminescence of the mouse was measured using the IVIS Spectrum imaging system (Caliper Life Science, Hopkinton, MA) and bioluminescence analyzed using LIVINGIMAGE 4.4 software (Caliper Life Science). Ascites was obtained by syringe, and tumoroids collected and washed with RBC lysis buffer. Mice were sacrificed and tissue fixed in 4% PFA for further analysis.

Ethics

The Dutch Animal Experimental Committee granted approval for all protocols related to animal research, ensuring adherence to the Code of Practice for Animal Experiments in Cancer Research as outlined by the Netherlands Inspectorate for Health Protections, Commodities, and Veterinary Public Health (The Hague, the Netherlands, 1999).

Mouse tumoroid culture

Mouse ascites with tumoroids were collected. Tumoroids were washed with RBC lysis buffer 1-2 times, then plated in 24-well ultra-low attachment surface plates (#33019010, Corning). Tumoroids were cultured in a 1:1 mixture of OV90 medium and advanced DMEM/F12 medium (#2322978; Thermo Fisher Scientific) containing 4% B27 (#A1895601, Life Technologies), 2% N-2 supplement (#11520536, Thermo Fisher Scientific) and 0.04% EGF (#PMG8045, Invitrogen).

BaseScope assay

BaseScope assays were performed following the guidelines from ACD (Advanced Cell Diagnostics, Newark, CA). 4 µm sections were cut onto Superfrost plus slides (#10149870, Thermo Fisher Scientific) and stored O/N at RT. Sections were baked for 1 h at 60 °C before deparaffinizing in xylene and 100% ethanol. Sections were dried for 5 min at 60 °C, incubated in hydrogen peroxide at RT for 10 min, underwent target retrieval for 15 min at 100 °C, and protease treatment for 30 min at 40 °C. BaseScope probes were added and slides incubated in an oven for 2 h at 40 °C before adding reagents AMP1 (30 min at 40 °C), AMP2 (30 min at 40 °C), AMP3 (15 min at 40 °C), AMP4 (30 min at 40 °C), AMP5 (30 min at 40 °C), AMP6 (15 min at RT), AMP7 (30 min at RT) and AMP8 (15 min at RT). Fast Red A and B was added to slides and incubated for 10 min at RT, then counterstained with Gill's hematoxylin. Slides were dried for 15 min at 60 °C, then mounted in VectaMount permanent mounting medium (H5000, Vector labs). Images were taken using an LSM-700 with 40× lens.

Chemoresistance and IC_{50} measurement

Cells were seeded in 96-well plates at 5000 cells/well and left O/N to adhere. 3–5 biological replicates were plated per tested condition. Both cisplatin (#PHR1624, Sigma-Aldrich) and paclitaxel (#S1150, Selleck Chemicals) were dissolved in DMSO (#D2650, Sigma Aldrich). Cells were incubated for 3 days with cisplatin and paclitaxel. After removal of the chemotherapeutic drug, cells were washed with PBS and left to re-grow in standard culture medium for 1 day. Cell viability was assessed using the MTT as described previously [55]. Absolute viability values were converted to percentage viability versus DMSO control treatment, then non-linear fit of log(inhibitor) versus response was performed in GraphPad Prism v7.0 to obtain an IC_{50} values.

TOP-Flash reporter assay

For the β-catenin/TCF reporter assay (TOP-Flash reporter assay), cells were plated on 48-well dishes. After 48 h, when 70% confluence was reached, cells were transfected by Fugene HD with 125 ng of the TOP-Flash or FOP-Flash reporter constructs together with 25 ng of the Renilla luciferase vector for normalization purposes. Luciferase activity was measured using the Dual-Luciferase Reporter Assay System (#E1910, Promega) 24 h post-transfection. Luminescence was measured using a GloMax Luminometer (#9100-102, Promega).

Transient transfection and dual-luciferase reporter assay

OV90 and COV504 cells (1 ×10^4 cells) were transfected with 0.3 pmol of the luciferase reporter plasmid and 0.01 pmol of the Renilla control plasmid using FuGENE6 transfection reagent, following the manufacturer's instructions. Luciferase activity was measured 24 h post-transfection using the Dual-Luciferase Reporter Assay System (#E1910, Promega). Luminescence was measured with a GloMax Luminometer (#9100-102, Promega).

Identification of compounds targeting Tpm1.8/1.9

A model of the N-terminus of human Tpm1.8 (identical for Tpm1.9) containing the region of greatest diversity between the four TPM genes (residues 4–16) was constructed. A virtual Screening of a zinc library was performed to identify docking hits to Tpm1.8. The screening protocol consisted of: 1) multiple copy simultaneous search "MCSS" of functional groups such as benzene, pyridine, pyrimidine, pyrazine and phenol around the N-terminus of human Tpm1.8; 2) pharmacophores such as aromatic rings were derived from the distribution of the minima of these fragments; 3) searching the zinc database library (version 2016) based on the pharmacophores, reducing the virtual library of 2553 compounds; 4) docking of these 2553 compounds onto the N-terminus of Tpm1.8 and the best conformations of best overlay with the fragment minima and binding energies to the target were selected. All the calculations were carried out using the software QuCBit [56, 57]. Six compounds were selected and purchased from suppliers. Mouse embryo fibroblasts were exposed to each of the six compounds or vehicle alone for 24 h, fixed and stained for Tpm1.8/1.9 using isoform specific antibodies (Brayford et al.). Control cells show strong enrichment of Tpm1.8/1.9 in the lamellipodium (Brayford et al.). Two of the compounds, Tpm1.8/9-1 (PubChem CID 6494468) and -3 (PubChem CID 18973468), prevented enrichment in the lamellipodium of mouse (not shown) and human fibroblasts at 10 µM (Supplementary Fig. 8 Suppl. 1). They were selected for further studies.

Alternative splicing analysis

EpCAMhi/lo RNASeq data was obtained from the ovarian cancer cell lines OV90 and CAOV3 and the sequencing reads mapped to GRCh37.p13. genome by STAR [58] (https://www.gencodegenes.org/human/release_19.html). MISO [23] was used to quantify AS events with annotation from https://miso.readthedocs.io/en/fastmiso/index.html#iso-centric. The MISO [23] uses the alternative exon reads and adjacent conservative reads to measure the percentage of transcript isoform with

specific exon included, termed Percentage Spliced In (PSI or Ψ). The PSI ranges from 0 (i.e. no isoform includes a specific alternative exon) to 1 (i.e. all of the isoforms detected comprise the alternative exon). GRCh37.p13 was utilized while MISO does not support the GRCh38.p14 assembly.

We removed alternative events with low expression of related transcript isoforms if less than 3 samples in a dataset had more than 10 informative reads to calculate the PSI. Next, we compared the PSI between EpCAM^{hi} and EpCAM^{lo} groups in the OV90 and CAOV3 ovarian cancer cell lines. AS events were defined as differentially spliced events when the difference of mean PSI between two groups (Δpsi; differential Percentage Spliced In) was >10%.

RNA seq analysis of subpopulations in OV90 and CAOV3

RNA was isolated from sorted populations with Trizol reagent. Libraries were prepared with the TruSeq RNA sample prep kit v2 (Erasmus MC, Biomics). Samples were sequenced with Illumina HiSeq 2000 and adapter sequences were removed with Trimmomatic (v.0.33). Subsequently, the reads were mapped in a two-pass procedure to the human reference genome build hg38 with the RNA-seq aligner STAR (v2.4.2a) [58] using the Homo sapiens GENCODE v23 annotations. Raw counts were imported in DESeq2 (v1.36.0) and normalized with a variance stabilizing transformation (VST) [59]. Differential expressed genes were identified by comparing Epcamlow versus Epcamhigh/bulk samples using absLogFC >1.5 and padj <0.05, and visualized with the ComplexHeatmap [60] package (v2.12.1) after a z-score scaling. Pathways activity was evaluated using gene set enrichment analysis (fgsea v1.22.0) on the hallmark gene sets from the Molecular Signature data base [61, 62]. Complete-linkage hierarchical clustering with split by k-means (k = 2) clustering was used.

RNA seq analysis of TPM1 OE in OV90

Paired end mRNA sequencing was performed with the DNA Nanoball sequencing (DNBseq) technology till a depth of 25 M reads per sample (BGI Genomics, Shen Zhen). Adapter trimming and quality filtering was performed using the SOAPnuke pipeline (BGI Genomics). Clean FASTQ files were aligned to the GRCh37 reference genome with RSEM (v1.3.3) [63] using the STAR aligner (v2.7.9a) [58]. Gene level data was imported with tximport [64] (v1.24.0) and downstream analysis was performed using DESeq2 (v1.36.0) [59]. Counts were normalized with a variance stabilizing transformation (VST). Gene set activity was evaluated with a gene set variation analysis (GSVA, v1.44.5) [65] using the Hallmark gene set from the Molecular Signature database [61] and visualized with the ComplexHeatmap [60] package (v2.12.1). Principal component analysis was computed using the top 500 genes with highest row variance. Differential expression analysis was performed by comparing the TPM1.6/7 samples with the TPM1.8/9 samples and results were displayed with a volcano plot using EnhancedVolcano (v1.14.0). Complete-linkage hierarchical clustering was used.

scRNAseq analysis of ovarium cancer cells

Publicly available data from Vázquez-García et al. were retrieved from CellxGene portal [22]. Downstream analysis was performed in Seurat (v4.3.0) [66]. An EpCAM^{lo} signature was evaluated with AddModuleScore based on the previously identified upregulated gene list ($N = 38$). A threshold (>0.1) was used to annotate cells with the highest association to the EpCAM^{lo} signature ("low-like cells", 4% of cancer cells). Next, low-like cells were visualized on the integrated UMAP embedding from Vázquez-García et al. and pathway activity of clusters encompassing low-like cells were visualized with ComplexHeatmap [60]. After approval of a data transfer agreement, FASTQ files from Izar et al. [67] were downloaded from the TerraBio repository and processed with RSEM using the STAR aligner to the hg19 human reference genome with isoform annotation from UCSC [58, 63]. Files were imported with tximport [64] (v1.24.0) and cells were selected that contained at least 500 different genes (nFeature_RNA > 500). Cells were clustered (kmeans, k = 4) according to the percentage of their respective TPM1 isoform expression and subsequent analysis was performed in Seurat (v4.3.0) [66], where TPM1.7 expressing cells were compared to TPM1.9 with FindMarkers. A gene set enrichment analysis was performed using the Hallmark gene set and pathways were filtered according to similar activity in the OV90 cell line and patient data. Complete-linkage hierarchical clustering was used.

Survival analysis of TPM1 in TCGA OVCA

RSEM processed data from the TCGA cohort was downloaded from the tsvDB [68]. Data was log2 transformed and survival analysis was performed with the survival package. Survival curves were generated with the survminer package for the whole TPM1 gene, TPM1.7 (isoform_uc002alk) and TPM1.9 (isoform_uc002alt) based on clinical data on overall survival [69].

Statistical analysis

For statistical comparison, we performed unpaired *t test*. Statistical analyses were performed using Prism 7 software (GraphPad). Data with statistical significance are as indicated. Information on replicates, independent experiments and statistical test can be found in the Fig. Legends. Analysis tools were run with all parameters as default unless otherwise stated.

DATA AVAILABILITY

RNA sequencing data has been deposited to the Gene Expression Omnibus (GEO) and can be accessed using the following identifiers: GSE192920 (subpopulations in OV90 and CAOV3), GSE231560 (TPM1 OE in OV90). Other single cell RNA sequencing data used in this study are publicly available and can be accessed from GEO for the SmartSeq2 data [GSE146026 [67]] and from Synapse for the 10X Genomics data [syn25569736 [22]].

REFERENCES

1. Jayson GC, Kohn EC, Kitchener HC, Ledermann JA. Ovarian cancer. Lancet. 2014;384:1376–88.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68:7–30.
3. Bast RC Jr, Hennessy B, Mills GB. The biology of ovarian cancer: new opportunities for translation. Nat Rev Cancer. 2009;9:415–28.
4. Kurman RJ, Shih IeM. Pathogenesis of ovarian cancer: lessons from morphology and molecular biology and their clinical implications. Int J Gynecol Pathol. 2008;27:151–60.
5. Bowtell DD, Böhm S, Ahmed AA, Aspuria P-J, Bast RC, Beral V, et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. Nat Rev Cancer. 2015;15:668–79.
6. Naora H, Montell DJ. Ovarian cancer metastasis: integrating insights from disparate model organisms. Nat Rev Cancer. 2005;5:355–66.
7. Kipps E, Tan DSP, Kaye SB. Meeting the challenge of ascites in ovarian cancer: new avenues for therapy and research. Nat Rev Cancer. 2013;13:273–82.
8. Lengyel E. Ovarian cancer development and metastasis. Am J Pathol. 2010;177:1053–64.
9. Samatov TR, Tonevitsky AG, Schumacher U. Epithelial-mesenchymal transition: focus on metastatic cascade, alternative splicing, non-coding RNAs and modulating compounds. Mol Cancer. 2013;12:107.
10. Blencowe BJ. Alternative splicing: new insights from global analyses. Cell. 2006;126:37–47.
11. Bechara EG, Sebestyen E, Bernardis I, Eyras E, Valcarcel J. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. Mol Cell. 2013;52:720–33.
12. Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, et al. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nat Neurosci. 2012;15:1488–97.
13. Ohno G, Ono K, Togo M, Watanabe Y, Ono S, Hagiwara M, et al. Muscle-specific splicing factors ASD-2 and SUP-12 cooperatively switch alternative pre-mRNA processing patterns of the ADF/cofilin gene in Caenorhabditis elegans. PLoS Genet. 2012;8:e1002991.
14. Fu XD, Ares M Jr. Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet. 2014;15:689–701.
15. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. Oncogene. 2014;33:5311–8.
16. Wang S, Sun Z, Lei Z, Zhang HT. RNA-binding proteins and cancer metastasis. Semin Cancer Biol. 2022;86:748–68.
17. Xu T, Verhagen M, Joosten R, Sun W, Sacchetti A, Munoz Sagredo L, et al. Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behavior in colon cancer. Elife. 2022;11:e82006.
18. Gunning PW, Hardeman EC. Tropomyosins. Curr Biol. 2017;27:R8–R13.
19. Mlakar V, Berginc G, Volavsek M, Stor Z, Rems M, Glavac D. Presence of activating KRAS mutations correlates significantly with expression of tumour suppressor genes DCN and TPM1 in colorectal cancer. BMC Cancer. 2009;9:282.
20. Zhu S, Si ML, Wu H, Mo YY. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). J Biol Chem. 2007;282:14328–36.

21. Sacchetti, A. et al. Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. Elife 10, https://doi.org/10.7554/eLife.61461 (2021)
22. Vazquez-Garcia I, Uhlitz F, Ceglia N, Lim JLP, Wu M, Mohibullah N, et al. Ovarian cancer mutational processes drive site-specific immune evasion. Nature. 2022;612:778–86.
23. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods. 2010;7:1009–15.
24. Pittenger MF, Kazzaz JA, Helfman DM. Functional properties of non-muscle tropomyosin isoforms. Curr Opin Cell Biol. 1994;6:96–104.
25. Groger H, Callaerts P, Gehring WJ, Schmid V. Gene duplication and recruitment of a specific tropomyosin into striated muscle cells in the jellyfish Podocoryne carnea. J Exp Zool. 1999;285:378–86.
26. Zheng Q, Safina A, Bakin AV. Role of high-molecular weight tropomyosins in TGF-beta-mediated control of cell motility. Int J Cancer. 2008;122:78–90.
27. Bakin AV, Safina A, Rinehart C, Daroqui C, Darbary H, Helfman DM. A critical role of tropomyosins in TGF-beta regulation of the actin cytoskeleton and cell motility in epithelial cells. Mol Biol Cell. 2004;15:4682–94.
28. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science. 2005;309:1559–63.
29. Bharadwaj S, Prasad GL. Tropomyosin-1, a novel suppressor of cellular transformation is downregulated by promoter methylation in cancer cells. Cancer Lett. 2002;183:205–13.
30. Brayford S, Bryce NS, Schevzov G, Haynes EM, Bear JE, Hardeman EC, et al. Tropomyosin Promotes Lamellipodial Persistence by Collaborating with Arp2/3 at the Leading Edge. Curr Biol. 2016;26:1312–8.
31. Cagigas ML, Bryce NS, Ariotti N, Brayford S, Gunning PW, Hardeman EC. Correlative cryo-ET identifies actin/tropomyosin filaments that mediate cell-substrate adhesion in cancer cells and mechanosensitivity of cell proliferation. Nat Mater. 2022;21:120–8.
32. Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, et al. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science. 1997;275:1784–7.
33. Stehn JR, Schevzov G, O'Neill GM, Gunning PW. Specialisation of the tropomyosin composition of actin filaments provides new potential targets for chemotherapy. Curr Cancer Drug Targets. 2006;6:245–56.
34. Stehn JR, Haass NK, Bonello T, Desouza M, Kottyan G, Treutlein H, et al. A novel class of anticancer compounds targets the actin cytoskeleton in tumor cells. Cancer Res. 2013;73:5169–82.
35. Currier MA, Stehn JR, Swain A, Chen D, Hook J, Eiffe E, et al. Identification of Cancer-Targeted Tropomyosin Inhibitors and Their Synergy with Microtubule Drugs. Mol Cancer Ther. 2017;16:1555–65.
36. Bonello TT, Janco M, Hook J, Byun A, Appaduray M, Dedova I, et al. A small molecule inhibitor of tropomyosin dissociates actin binding from tropomyosin-directed regulation of actin dynamics. Sci Rep. 2016;6:19816.
37. Janco M, Rynkiewicz MJ, Li L, Hook J, Eiffe E, Ghosh A, et al. Molecular integration of the anti-tropomyosin compound ATM-3507 into the coiled coil overlap region of the cancer-associated Tpm3.1. Sci Rep. 2019;9:11262.
38. Hardeman E, Gunning P, E. E, inventors; TroBio Therapeutics, Pty Ltd., assignee. Sulfonamide compounds and the use thereof in the treatment of cancer. 2021. Patent: WO2021072487
39. Hardeman E, Gunning P, Eiffe E, inventors; TroBio Therapeutics, Pty Ltd, assignee. Substituted indole compounds and the use thereof. 2022. Patent: WO2020037079A1
40. Meiring JCM, Bryce NS, Wang Y, Taft MH, Manstein DJ, Liu Lau S, et al. Co-polymers of Actin and Tropomyosin Account for a Major Fraction of the Human Actin Cytoskeleton. Curr Biol. 2018;28:2331–7.e5.
41. Cook DP, Vanderhyden BC. Context specificity of the EMT transcriptional response. Nat Commun. 2020;11:2142.
42. Cook DP, Vanderhyden BC. Transcriptional census of epithelial-mesenchymal plasticity in cancer. Sci Adv. 2022;8:eabi7640.
43. Skrypek N, Goossens S, De Smedt E, Vandamme N, Berx G. Epithelial-to-Mesenchymal Transition: Epigenetic Reprogramming Driving Cellular Plasticity. Trends Genet. 2017;33:943–59.
44. Loret N, Denys H, Tummers P, Berx G. The Role of Epithelial-to-Mesenchymal Plasticity in Ovarian Cancer Progression and Therapy Resistance. Cancers. 2019;11:838.
45. Wang Z, Wang S, Qin J, Zhang X, Lu G, Liu H, et al. Splicing factor BUD31 promotes ovarian cancer progression through sustaining the expression of anti-apoptotic BCL2L12. Nat Commun. 2022;13:6246.
46. Yin H, Wang J, Li H, Yu Y, Wang X, Lu L, et al. Extracellular matrix protein-1 secretory isoform promotes ovarian cancer through increasing alternative mRNA splicing and stemness. Nat Commun. 2021;12:4230.
47. Varga AE, Stourman NV, Zheng Q, Safina AF, Quan L, Li X, et al. Silencing of the Tropomyosin-1 gene by DNA methylation alters tumor suppressor function of TGF-beta. Oncogene. 2005;24:5043–52.
48. Pan H, Gu L, Liu B, Li Y, Wang Y, Bai X, et al. Tropomyosin-1 acts as a potential tumor suppressor in human oral squamous cell carcinoma. PloS one. 2017;12:e0168900.
49. Hardeman EC, Bryce NS, Gunning PW. Impact of the actin cytoskeleton on cell development and function mediated via tropomyosin isoforms. Semin Cell Dev Biol. 2020;102:122–31.
50. Brabletz T, Jung A, Spaderna S, Hlubek F, Kirchner T. Opinion: migrating cancer stem cells - an integrated concept of malignant tumour progression. Nat Rev Cancer. 2005;5:744–9.
51. Teeuwssen M, Fodde R. Wnt Signaling in Ovarian Cancer Stemness, EMT, and Therapy Resistance. J Clin Med. 2019;8:1658.
52. Schevzov G, Whittaker SP, Fath T, Lin JJ, Gunning PW. Tropomyosin isoforms and reagents. Bioarchitecture. 2011;1:135–64.
53. Savill SA, Leitch HF, Daly AK, Harvey JN, Thomas TH. Polymorphisms in the tropomyosin TPM1 short isoform promoter alter gene expression and are associated with increased risk of metabolic syndrome. Am J Hypertens. 2010;23:399–404.
54. Savill SA, Leitch HF, Harvey JN, Thomas TH. Functional structure of the promoter regions for the predominant low molecular weight isoforms of tropomyosin in human kidney cells. J Cell Biochem. 2012;113:3576–86.
55. Sacchetti A, Teeuwssen M, Verhagen M, Joosten R, Xu T, Stabile R, et al. Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. Elife. 2021;10:e61461.
56. Zeng J, Treutlein HR. A method for computational combinatorial peptide design of inhibitors of Ras protein. Protein Eng. 1999;12:457–68.
57. Zeng J. Mini-review: computational structure-based design of inhibitors that target protein surfaces. Comb Chem High Throughput Screen. 2000;3:355–62.
58. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
59. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
60. Gu Z, Hubschmann D. Make Interactive Complex Heatmaps in R. Bioinformatics. 2022;38:1460–2.
61. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102:15545–50.
62. Gennady Korotkevich, V. S., Nikolay Budin, Boris Shpak, Maxim N. Artyomov, Alexey Sergushichev. Fast gene set enrichment analysis. https://doi.org/10.1101/060012 (2021).
63. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinforma. 2011;12:323.
64. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4:1521.
65. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinforma. 2013;14:7.
66. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573–87.e29.
67. Izar B, Tirosh I, Stover EH, Wakiro I, Cuoco MS, Alter I, et al. A single-cell landscape of high-grade serous ovarian cancer. Nat Med. 2020;26:1271–9.
68. Sun W, Duan T, Ye P, Chen K, Zhang G, Lai M, et al. TSVdb: a web-tool for TCGA splicing variants analysis. BMC Genomics. 2018;19:405.
69. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. https://doi.org/10.1007/978-1-4757-3294-8 (2001).

ACKNOWLEDGEMENTS

TX's appointment was kindly supported by the China Scholarship Council (201806300047). This work was supported by grants from the Australian National Health and Medical Research Council (NHMRC grants APP1079866, APP1100202) and The Kid's Cancer Project to PWG and ECH, and from the Dutch Cancer Society (KWF; grant EMCR 2015-7588) to RF.

AUTHOR CONTRIBUTIONS

TX conceived the experimental strategy, performed most of the experiments and wrote the first manuscript draft. M.V. contributed the in silico analysis and validation of the AS results in patient-derived (sc)RNAseq data. MT, RJ and AS contributed to the implementation of PCR, mouse, and FACS experiments. WS analyzed AS splicing in the RNAseq data. PCE-G, MPHMJ and IAB contributed to patient tissue preparation. NSB, JZ, HRT and JH contributed to compounds identification. ECH and PWG contributed to tropomyosin data analysis, tropomyosin materials support and critically revised the manuscript. RF conceived the experimental strategy and wrote the manuscript.

COMPETING INTERESTS

PWG and ECH receive funding from and are Directors of TroBio Therapeutics, a company commercializing anti-tropomyosin drugs. The remaining authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41418-024-01267-9.

Correspondence and requests for materials should be addressed to Riccardo Fodde.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

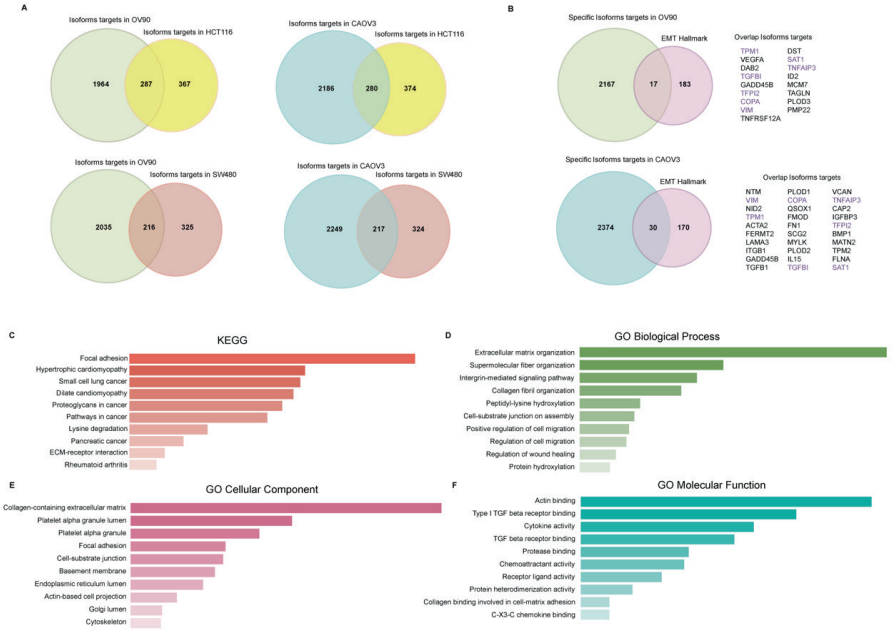


Fig. 3 - Supplement 1. Isoforms targets in HGSOC.

A. Venn diagrams showing the overlap between ovarian cancer and colon cancer isoforms targets list. The list of isoforms targets is depicted on the right of each Venn diagram. OV90, n=2251; CAOV3, n=2466; HCT116, n=654; SW480, n=541.

B. Venn diagrams showing the overlap between ovarian cancer-specific isoforms targets and the EMT hallmark gene list. The list of EMT-related isoforms targets is depicted on the right of each Venn diagram. OV90, n=17; CAOV3, n=30; n=7 overlapped between the two cell lines.

C. KEGG pathway analysis of the 40 EMT-related ovarian cancer-specific isoforms targets (and CAOV3-specific AS targets).

D-F. Enrichment of Gene Ontology (GO) processes (C. Biological; D. Cellular Components; E. Molecular Functions) relative to the 40 EMT-related ovarian cancer-specific AS targets. Selected GO terms are the top 10 significantly enriched filtered by p value<0.01.

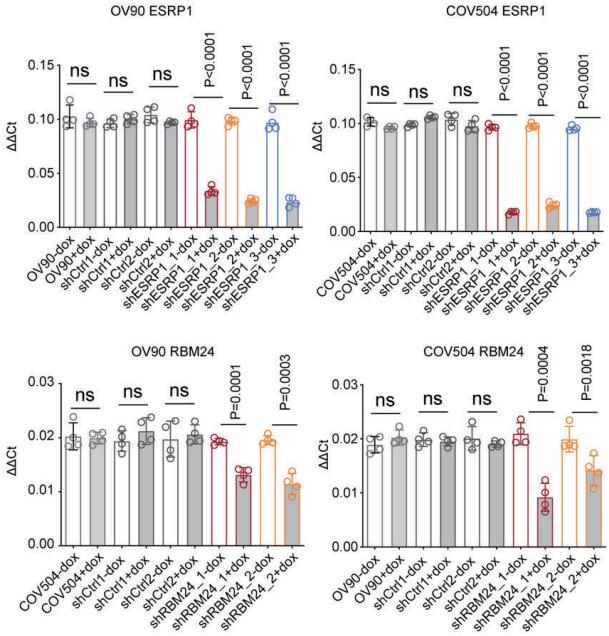


Fig. 4 - Supplement 1. Knockdown efficiency of shESRP1 and shRBM24 vectors

RT-qPCR analysis of ESRP1 and RBM24 expression in sh*RBM24-KD* (knockdown) and sh*ESRP1*-KD (knockdown) OV90 and COV504 ovarian cancer cell line; *GAPDH* expression was employed as control (Means±SD, n=4). There are two shControl vectors and 2-3 sh*RBM24* (knockdown) and sh*ESRP1*-KD (knockdown) used.

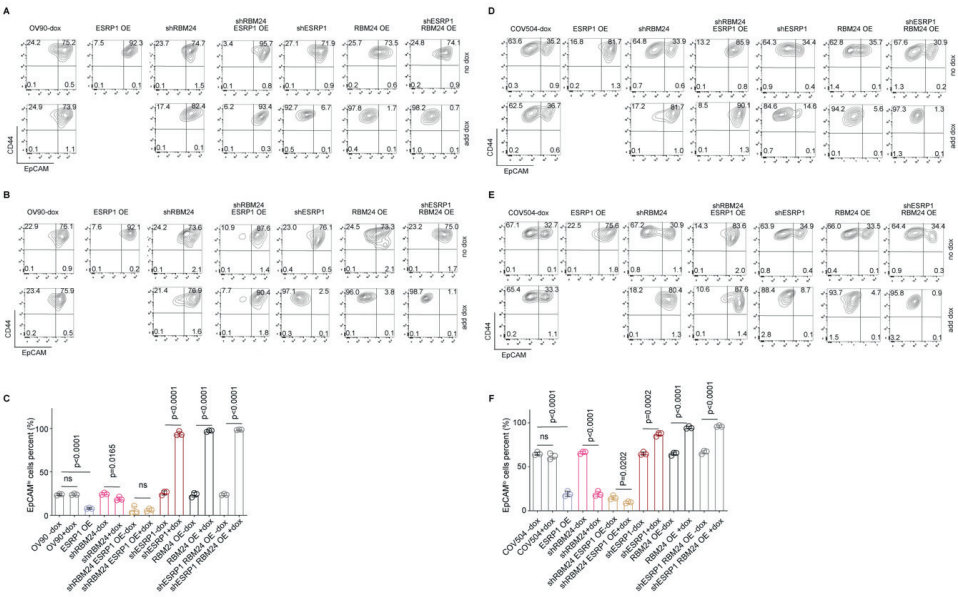


Fig. 4 - Supplement 2. CD44/EpCAM FACS analysis of RBM24-OE/KD and ESRP1-OE/KD cells.

A. CD44/EpCAM FACS analysis of *RBM24*-OE/KD and *ESRP1*-OE/KD OV90 cells (replica 2). Cells were induced with 1 µg/mL doxycycline for 72 h before analysis. The relative percentages of $EpCAM^{lo}$ and $EpCAM^{hi}$ cells are indicated in each quadrant.

B. CD44/EpCAM FACS analysis of *RBM24*-OE/KD and *ESRP1*-OE/KD OV90 cells (replica 3). Cells were induced with 1 µg/mL doxycycline for 72 h before analysis. The relative percentages of $EpCAM^{lo}$ and $EpCAM^{hi}$ cells are indicated in each quadrant.

C. Quantitative analysis of $EpCAM^{lo}$ percentage in FACS analysis of *RBM24*-OE/KD and *ESRP1*-OE/KD OV90 cells (replica 1-3). (Means±SD, n=3)

D. CD44/EpCAM FACS analysis of *RBM24*-OE/KD and *ESRP1*-OE/KD COV504 cells (replica 2). Cells were induced with 1 µg/mL doxycycline for 72 h before analysis. The relative percentages of $EpCAM^{lo}$ and $EpCAM^{hi}$ cells are indicated in each quadrant.

E. CD44/EpCAM FACS analysis of *RBM24*-OE/KD and *ESRP1*-OE/KD COV504 cells (replica 3). Cells were induced with 1 µg/mL doxycycline for 72 h before analysis. The relative percentages of $EpCAM^{lo}$ and $EpCAM^{hi}$ cells are indicated in each quadrant.

F. Quantitative analysis of $EpCAM^{lo}$ percentage in FACS analysis of *RBM24*-OE/KD and *ESRP1*-OE/KD COV504 cells (replica 1-3). (Means±SD, n=3)

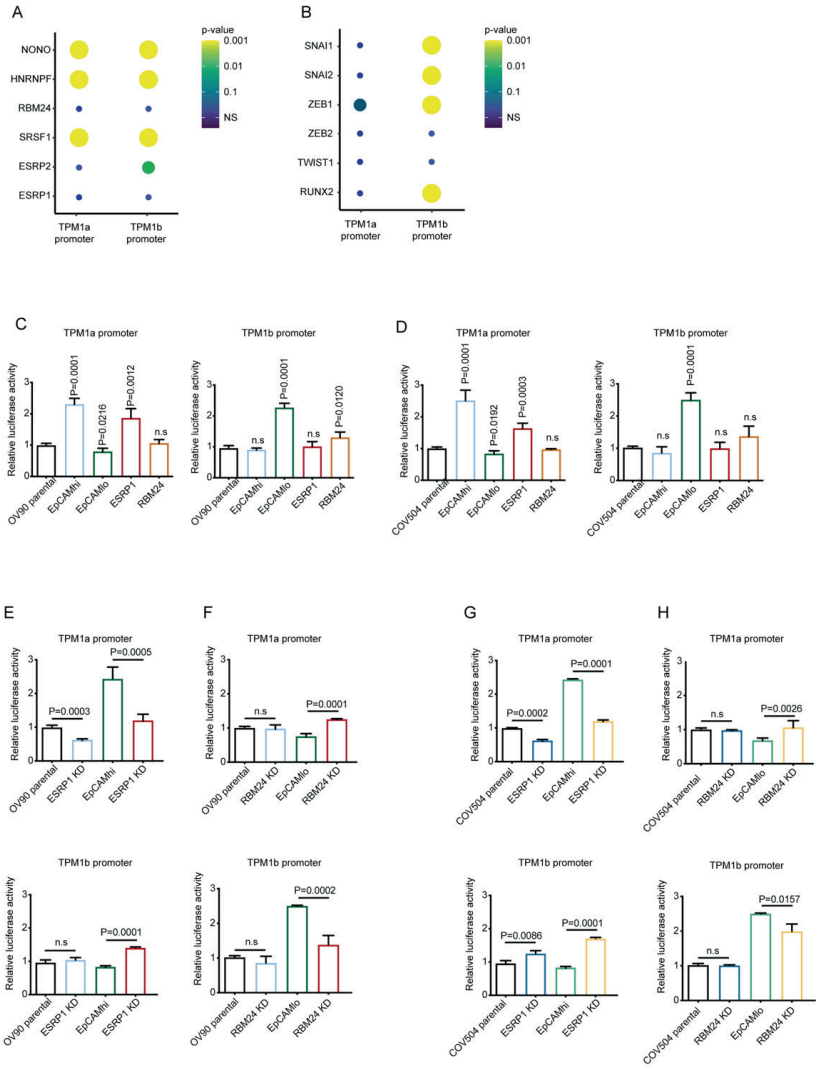


Fig. 4 - Supplement 3.

A-B. Detection of RBP (A) and EMT-TFs (B) binding motifs from the FANTOM database in the TPM1a/b promoter regions.

C-D. TPM1a/b reporter vectors were obtained by cloning genomic PCR products encompassing the two promoter sequences as defined in Sevill et al.[53,54] in a promoter-less Luciferase plasmid. Transient transfections were conducted on OV90 (A) and COV504 (B) ovarian cancer cell lines (EpCAM^{hi}, EpCAM^{lo}, and ESRP1- and RBM24-OE cells; Means±SD, n=4). P values are relative to the comparison with the parental cell lines.

E-H. TPM1a/b reporter vectors were obtained by cloning genomic PCR products encompassing the two promoter sequences as defined in Sevill et al.[53,54] in a promoter-less Luciferase plasmid. Transient transfections were conducted on OV90 (C-D) and COV504 (E-F) ovarian cancer cell lines (EpCAM^{hi}, EpCAM^{lo}, and ESRP1- and RBM24-KD cells; Means±SD, n=4). P values are determined by comparison with the parental, EpCAM^{hi} orEpCAM^{lo} cell line.

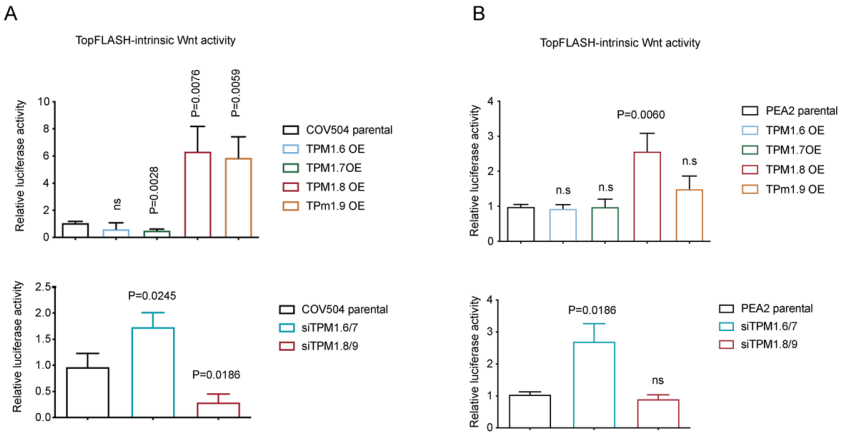


Fig. 6 - Supplement 1. TOP-Flash luciferase reporter analysis of Wnt signaling activity.

TOP-Flash luciferase reporter analysis of Wnt signaling activity in *Tpm1.6/7/8/9*-OE (upper histogram) and upon knockdown by siRNA of *Tpm1.6/7* and *Tpm1.8/9* in COV504 and PEA2 cells. P values are relative to the comparison with the parental cell lines (Means±SD, n=4-5).

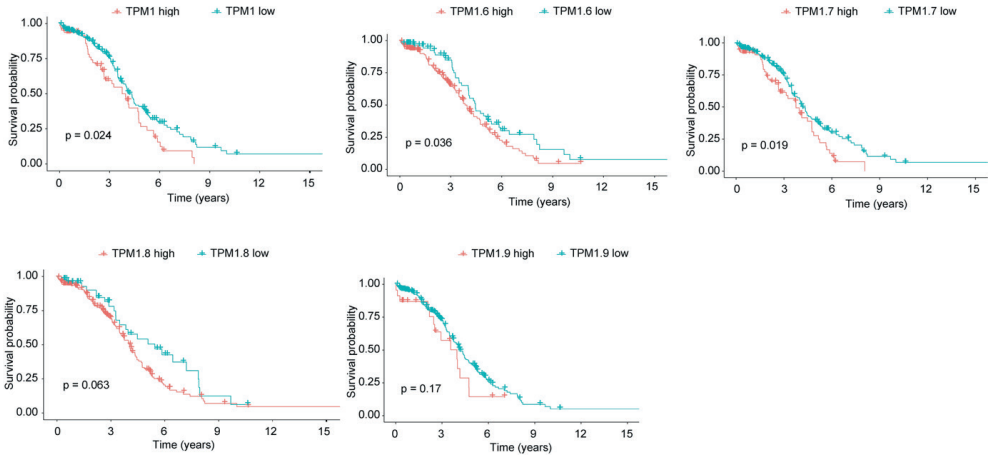


Fig. 7 - Supplement 1.
Kaplan-Meier analysis of overall survival based on the relative expression of the *TPM1* gene, and its Tpm1.6/7/8/9 isoforms.

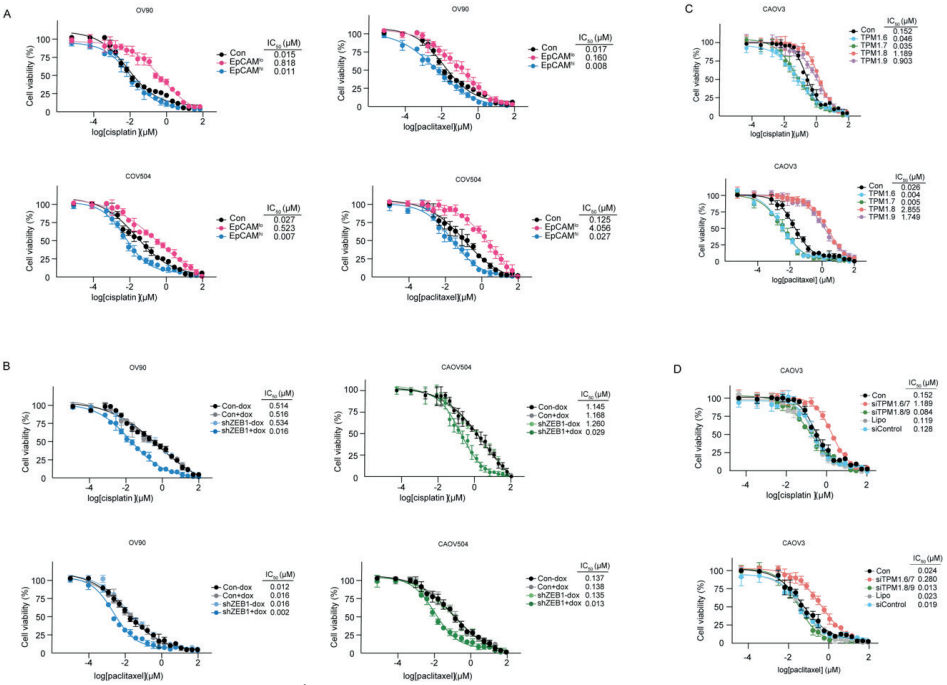


Fig. 7 - Supplement 2. EpCAM^{lo} , ZEB1 and *Tpm1.8/9* isoforms confer resistance to platinum- and taxane-based therapies.

A. Dose-response curves relative to EpCAM^{hi} and EpCAM^{lo} subpopulations of OV90 and COV504 grown in the presence of different concentrations of paclitaxel and cisplatin (log scale and cell viability on the x and y axis, respectively). IC_{50} values were calculated from biological quadruplicates for each experiment (Means±SD, n=4).

B. Dose-response curves relative to *ZEB1*-shRNA induced OV90 and COV504 grown in the presence of different concentrations of paclitaxel and cisplatin (log scale and cell viability on the x and y axis, respectively). IC_{50} values were calculated from biological quadruplicates for each experiment (Means±SD, n=4).

C. Dose-response curves relative to *Tpm1.6/7/8/9*-OE CAOV3 cells grown in the presence of different concentrations of paclitaxel and cisplatin (log scale and cell viability on the x and y axis, respectively). IC_{50} values were calculated from biological quadruplicates for each experiment (Means±SD, n=4).

D. Dose-response curves of si*Tpm1.6/7* and si*Tpm1.8/9* knockdown CAOV3 cells cultured in the presence of different concentrations of paclitaxel (left) and cisplatin (right) concentrations (log scale and cell viability on the x and y axis, respectively). IC_{50} values were calculated from biological quadruplicates for each experiment (Means±SD, n=4).

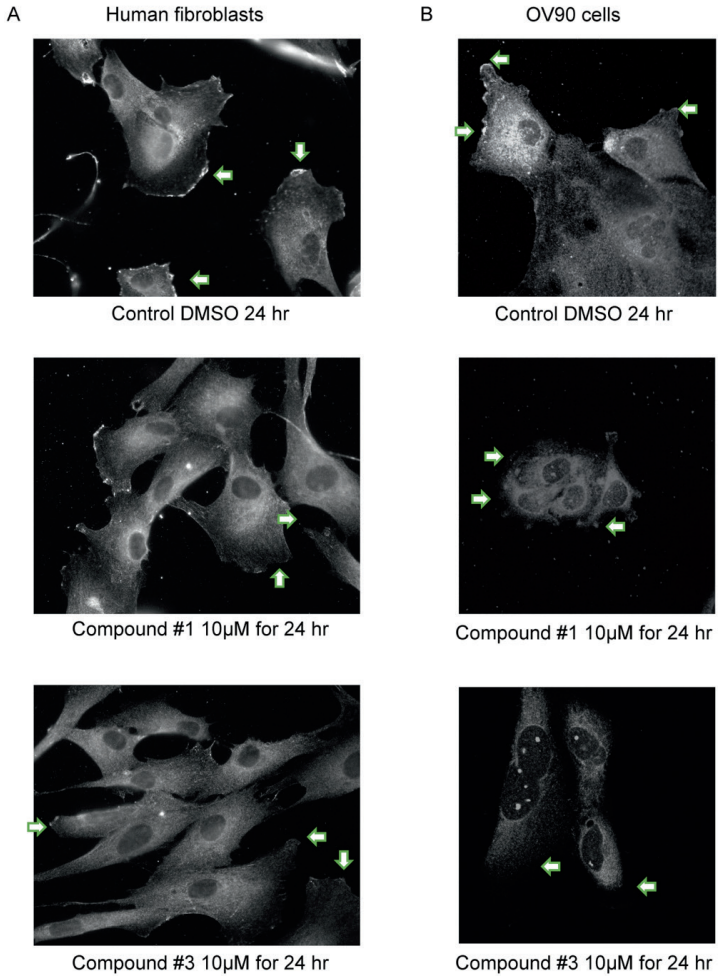


Fig. 8 - Supplement 1.

A. Compound # 1 and #3 cause the removal of Tpm1.8/1.9 from the lamellipodium of human fibroblasts. Human immortalized BJeH fibroblasts were plated onto glass coverslips for 24 h and then exposed to DMS0 (up) or 10 μM compound #1 or #3 (middle and down) dissolved in DMSO for a further 24 h. The cells were then fixed and Tpm1.8/1.9 visualized using a rat monoclonal antibody exactly as described in Brayford et al. (2016). Arrowheads identify lamellipodia in (up) control cells positive for Tpm1.8/9 and (middle and down) lamellipodia negative for Tpm1.8/9 in cells exposed to compound #1 or 3..

B. Compound #1 and #3 causes the removal of Tpm1.8/1.9 from the lamellipodium of OV90 cells. OV90 cells were plated onto glass coverslips for 24 h and then exposed to DMS0 (up) or 10 μM (middle) compound #1 and 3 (down) dissolved in DMSO for a further 24 h. Arrowheads identify lamellipodia in (up) control cells positive for Tpm1.8/9 and (middle and down) lamellipodia negative for Tpm1.8/9 in cells exposed to compound # 1 or #3.

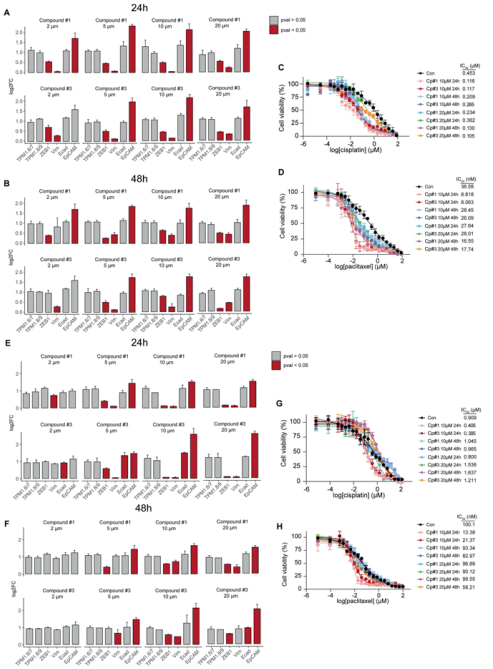


Fig. 8 - Supplement 2.

A. RT-qPCR analysis of *TPM1* isoforms and EMT-related gene expression in OV90 parental cells cultured for 24 hrs. in the presence of compound #1 and #3 at 0, 2, 5, 10 and 20 µM. The values were calculated by normalizing with the untreated cells. P values < 0.05 are shown by red bars while grey bars indicate lower values; GAPDH expression was employed as control (Means±SD, n=4).

B. RT-qPCR analysis of *TPM1* isoforms and EMT-related gene expression in OV90 parental cells cultured for 48 hrs. in the presence of compound #1 or #3 at 0, 2, 5, 10 and 20 µM. The values were calculated by normalizing with the untreated cells. P values < 0.05 are shown by red bars while grey bars indicate lower values; GAPDH expression was employed as control (Means±SD, n=4).

C. Dose-response curves of parental OV90 cells treated with compound #1 or #3 in the presence of different cisplatin concentrations. IC_{50} values were calculated from biological triplicates for each experiment (Means±SD, n=4).

D. Dose-response curves of parental OV90 cells treated with compound #1 or #3 in the presence of different paclitaxel concentrations. IC_{50} values were calculated from biological triplicates for each experiment (Means±SD, n=4).

E. RT-qPCR analysis of *TPM1* isoforms and EMT-related gene expression in OV90 EpCAM^{lo} cells cultured for 24 hrs. in the presence of compound #1 or #3 at 0, 2, 5, 10 and 20 µM. The values were calculated by normalizing with the untreated cells. P values < 0.05 are shown by red bars while grey bars indicate lower values; GAPDH expression was employed as control (Means±SD, n=4).

F. RT-qPCR analysis of *TPM1* isoforms and EMT-related gene expression in OV90 EpCAM^{lo} cells cultured for 48 hrs. in the presence of compound #1 or #3 at 0, 2, 5, 10 and 20 µM. The values were calculated by normalizing with the untreated cells. P values < 0.05 are shown by red bars while grey bars indicate lower values; GAPDH expression was employed as control (Means±SD, n=4).

G. Dose-response curves of OV90 EpCAM^{lo} cells treated with compound #1 or #3 in the presence of different cisplatin concentrations. IC_{50} values were calculated from biological triplicates for each experiment (Means±SD, n=4).

H. Dose-response curves of OV90 EpCAM^{lo} cells treated with compound #1 or #3 in the presence of different paclitaxel concentrations. IC_{50} values were calculated from biological triplicates for each experiment (Means±SD, n=4).

CHAPTER 6

General Discussion

1. Phenotypic Plasticity and Cancer Stem Cells

The "biphasic" nature of cancer progression highlights two forms of phenotypic plasticity and of cancer stem cells, namely resident and migrating cancer stem cells (rCSCs and mCSCs). Resident CSCs have an epithelial identity, continuously fuel tumor growth, and, just like normal adult stem cells, are earmarked by enhanced self-renewal and differentiation capacity. Migrating CSCs instead are responsible for local dissemination and distant metastases, are usually located at the tumor-host interface (the invasive front), and are generated from tumor cells through the activation of epithelial-mesenchymal transition (EMT)[1] by cues from the tumor microenvironment.

The main scope of this thesis is the comprehensive investigation of alternative splicing mechanisms underlying phenotypic plasticity, E->M and M->E transition, and resistance to therapy in colon and ovarian cancer. This included the characterization of the cancer cell subpopulations, the identifyication of AS regulatory factors, the validation of the targets, and the challenges and opportunities in leveraging alternative splicing for cancer treatment.

1.1 Immortalized Cancer Cell Lines as a Source of Resident and Migratory CSCs

In order to study this CSC dichotomy, we have taken advantage of a previous study showing that human immortalized breast cancer cell lines encompass various subpopulations with unique phenotypic states and functional traits that are maintained in a dynamic equilibrium through stochastic transitions between them[2]. In **Chapter 2**, Our initial characterization of similar subpopulations in colon and ovarian cell lines showed that while the majority of cancer cells exhibit a typical epithelial phenotype ($EpCAM^{hi}$), a small subset displays mesenchymal morphology and plastic characteristics ($EpCAM^{lo}$). These two populations coexist in a dynamic balance, transitioning between states through stochastic processes.

In **Chapter 3**, we identified three distinct subpopulations in the SW480 colon cancer cell line, namely $EpCAM^{hi}$ ($CD44^{hi}EpCAM^{high}$) likely to represent the epithelial bulk of the tumor, the quasi-mesenchymal $EpCAM^{lo}$ ($CD44^{hi}EpCAM^{lo}$), a proxy of migrating CSCs, and the so-called “spheres” subpopulation ($CD44^{med}EpCAM^{high}TROP2^{neg}$), here reported as model for resident CSCs. The sphere subpopulation is earmarked by expression of normal stem cell markers, including LGR5 and RUNX2.

Sphere cells exhibit properties reminiscent of epithelial stem cells in various cancer types[3-5]. They also show enhanced resistance to anti-cancer therapy[5]. Sphere cells are remarkably stable and hardly differentiate into the other SW480 subpopulations. Apart from the above mentioned stem cell markers, several others earmark the sphere subpopulation and can be utilized to isolate and further characterize them[3,5].

1.2. Migrating Stem Cells and Epithelial-Mesenchymal Plasticity

In the above discussion, I delved into the role of resident CSCs with sphere morphology. However, most of the experimental evidence encompassed in my thesis address the other facet of cancer stemness, namely the migrating CSCs. Migrating CSCs results from EMT, a process that underlies phenotypic plasticity by enhancing the motility and invasive capabilities of ‘bulk’ cancer cells crucial in cancer development and therapeutic resistance. The so-called epithelial-mesenchymal plasticity (EMP) bestows cancer cells with the ability to migrate, disseminate locally, and eventually establish metastatic sites in distant organs through lymphatic or hematogenous routes.

Recent research on the role of EMT in prompting phenotypic plasticity in cancer cells has

underscored the significance of intermediate stages with cells co-expressing both epithelial and mesenchymal genes, for tumour progression and metastasis[6,7]. Researchers found that partial/hybrid EMT (pEMT) is associated with collective migration, enhanced plasticity and metastatic potential, and therapy resistance in multiple cancer types[8-10]. Our investigation using single-cell RNA sequencing (scRNAseq) on EpCAM^{lo} colon cancer cells, we identified hybrid E/M cells. The EMT score was derived by subtracting the E score from the M score, while a pEMT score was determined through the co-expression of the E and M gene sets. Predicted by the RNAvelocity model to be responsible for the observed variation in gene expression, these hybrid cells co-express E- and M-specific genes at intermediate levels. Of note, they are earmarked by elevated expression of the Secreted Protein Acidic and Rich in Cysteine (SPARC) gene. SPARC is a gene that encodes for a matricellular protein responsible for regulating interactions between cells and the extracellular matrix[11]. It is a prognostic marker in colon cancer. The role of SPARC in cancer is complex and controversial. While it has been found to promote epithelial-mesenchymal transition EMT and metastasis, it also exhibits tumor-suppressive functions under specific conditions[12].

Apart from SPARC, other genes have been associated with partial EMT[7,12,13]. FAT atypical cadherin 1 (FAT1) was identified as a pEMT marker in squamous cell carcinomas (SCCs), with mutations found in skin squamous cell carcinoma and lung tumor mouse models. Fat1 deletion results in tumor initiation and malignant progression in mouse models. However, the allegedly epigenetic regulation of pEMT raises questions about whether mutated genes such as FAT1 can effectively serve as pEMT markers and warrants further discussion[13]. Other p-EMT marker genes include E-cadherin, Integrin a5, Sp1, p63, and Snai1/2 have been reported in multiple cancer types[14,15].

While these studies employed different genes to identify the epithelial and mesenchymal states in various cancer types [9,16], drawing conclusions about typical pEMT markers remains challenging. It is challenging to determine what constitutes an extreme epithelial or mesenchymal state. Although there are EMT related genes to identify cells in epithelial and mesenchymal states, the expression levels of these EMT related genes vary among different cancers or within the same cancer type from different patients. It is important to determine the expression patterns of genes associated with epithelial or mesenchymal states may vary depending on the stage of transition. Therefore, determining whether this identified state is already in the early or late stages of pEMT adds further complexity to the process of identifying pEMT markers.

2. Epigenetic regulation of Epithelial Mesenchymal Plasticity

Epigenetic regulation stands out as a critical player in epithelial mesenchymal plasticity (EMP), and cancer metastasis. It encompasses several distinct mechanisms including histone modification, DNA promoter methylation, microRNAs and long non-coding RNAs (lncRNA) regulation, and alternative splicing (AS)[17-20]. Alternative splicing occurs in approximately 94% of human genes, introducing diversity in protein structures derived from the same gene[21]. Please note that in this thesis the term alternative splicing is used in its broadest definition, including alternative promoter usage, as its regulation not only depends on splicing factors interacting with their pre-mRNA target sequences but is also coupled to transcription. As such, the mechanisms by which transcription affects AS through promoter identity and occupation are here also regarded as integral part of the epigenetic regulation underlying EMP.

Our findings in **Chapter 4**, illustrated in Figure 1, demonstrate that RNA-binding proteins (RBPs)-

mediated alternative splicing downstream the activation of transcription factors known to activate EMT (EMT-TFs). In the specific case of the EpCAM^{lo} subpopulation identified in colon cancer cell lines, activation of the EMT-TF ZEB1 suppresses the expression of the RBP ESRP1, thereby modulating the alternative splicing of several target genes among which NUMB and CD44. The alternative spliced NUMB and CD44 isoforms are a direct consequence of the dynamic equilibrium between EMT and MET in colon cancer cells and play distinct functional role in the acquisition of a quasi-mesenchymal phenotype and of invasive features (Figure 1).

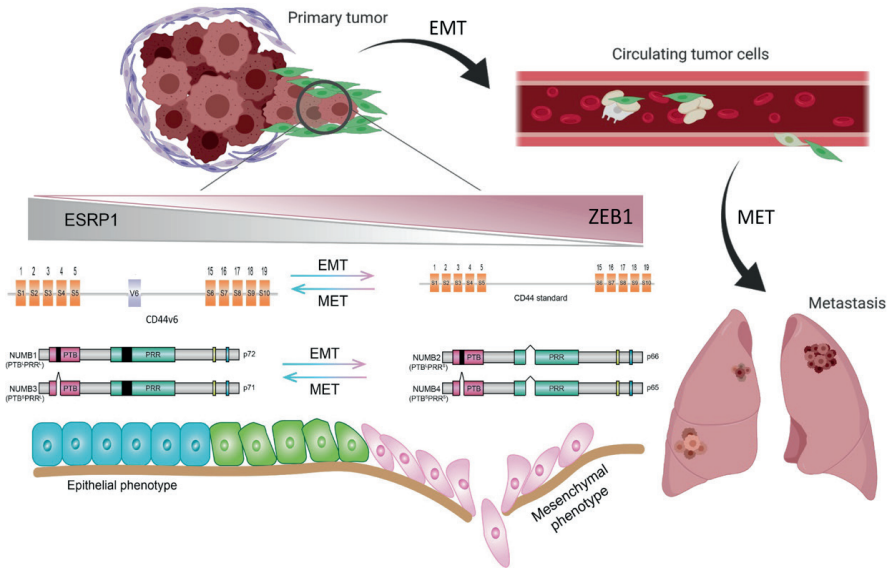


Figure 1 Alternative splicing promotes EMT, phenotypic plasticity and malignant behaviour in colon cancer.

Recent studies have uncovered that multiple upstream RBPs can regulate the same alternative splicing (AS) targets. As demonstrated in Chapter 4, several AS targets, including CTNND1, LSR, SLK, TCF7L2, and others, are alternatively spliced by the RBPs ESRP1/2, RBM47, and Quaking (QKI)[22]. Additionally, ESRP1 and ESRP2 were found to simultaneously regulate the splicing of the CTNND1 (p120 catenin) gene in human epithelial cell lines[23], while the RBP Zeppo1 was also shown to regulate splicing of the same CTNND1 gene during EMT in breast cancer[24,25].

The complexity of alternative splicing regulation is further underscored by the case of CD44, an AS target prevalent during EMT in multiple cancer types. In **Chapter 4**, it was revealed that CD44 undergoes alternative splicing not only through the differential ESRP1 expression in colon cancer, but also through the regulatory action of QKI and RBM47 in squamous cell carcinoma and non-small cell lung cancer (NSCLC). Other studies have implicated the RBPs CELF1 and ELAVL1 in the splicing of CD44 exons v7 to v10 across eight human cancer types[26]. Additionally, the RBP ZMAT3, induced by p53, was identified as a regulator of CD44 isoforms in colorectal cancer cells[27]. Additional complexity arises as the RBPs YB-1, Tra2-β1, hnRNPM, together with PCBP1, were shown

to jointly regulate CD44 AS in different cancer types[28] [29,30] [31]. In **Chapter 5**, we also demonstrated that Tropomyosin 1 gene (TPM1) isoforms are co-regulated by the simultaneous down- and up-regulation of ESRP1 and RBM24 upon EMT. Therefore, upon EMT/MET in local invasion and metastasis, the regulation of AS at multiple downstream targets occurs through complex mechanisms often involving different RBPs targeting the same genes.

In chapters 4 and 5, our results show that RBPs not only regulate alternative splicing downstream of EMT but can also promote EMT in a reinforcing feedback loop. In breast cancer, the fibronectin isoform encompassing the EDA domain, contribute to tumor cell invasion and the formation of metastases. The alternative splicing of fibronectin EDA isoforms is regulated by SNAIL1 and the RBP SRSF1. SNAIL1 primarily regulates the binding of SRSF1 to fibronectin exon 33, thereby forming fibronectin EDA isoforms[32]. As another example, cytosolic polyadenylation element–binding protein 2 (CPEB2) has two isoforms: the isoform CPEB2B contributes to EMT and metastasis *in vivo*, whereas CPEB2A does not possess this function. It was showed that the CPEB2B isoform induces TWIST1 expression by acting as a translational to promote EMT in breast cancer[33].

When it comes to specific events resulting in isoforms starting from alternative promoter, as here observed in the TPM1 HMW and LMW isoforms, the analysis of the underlying molecular mechanisms led to interesting scenarios. In chapter 5, luciferase assays of the TPM1a and TPM1b promoters, along with silico analysis based on the FANTOM[34] and RBP motif databases[35], revealed that alternative splicing, driven by the differential expression of RBPs such as ESRP1 and RBM24, and alternative promoter usage mediated by EMT-TFs, rather than directly by the aforementioned RBPs, are the main mechanisms underlying the differential expression of TPM1 HMW and TPM1 LMW isoforms in epithelial and quasi-mesenchymal ovarian cancer cells.

Future studies based on techniques such as Chromatin Immunoprecipitation (ChIP) and cross-linking and immunoprecipitation (CLIP) will help elucidating these mechanisms. CLIP is an antibody-based technique that uses UV radiation to cross-link RNA-binding proteins to the RNA they are bound to[36], while ChIP selectively enriches specific DNA-binding proteins along with their DNA targets. The integration of these two techniques may uncover whether the production of isoforms is regulated by RBPs in pre-mRNA levels or by variant promoter usage through upstream genes or transcription factors.

In summary, the epigenetic regulation of EMP is a complex process. This complexity arises due to the concurrent occurrence of a broad spectrum of regulatory mechanisms, including histone modification, DNA promoter methylation, microRNA, lncRNA regulation, and alternative splicing, which operate at different levels, i.e. genomic DNA, pre-mRNA, mRNA, and protein. And even when considering alternative splicing alone, numerous RBPs affects not only multiple EMT-related AS targets but also directly target EMT-TFs. The elucidation of the intricacies of epigenetic regulation in EMP will contribute unravelling the unpredictability of cancer metastasis and comprehending the complexities of the tumour microenvironment.

3. RBPs and their downstream targets: potential prognostic markers and therapeutic targets.

Recent research has revealed the significant influence of alternative splicing on diverse facets of tumour progression ranging from cell migration, angiogenesis, cellular metabolism, to response to cancer treatment, and in particular resistance to chemotherapy (Figure 2).

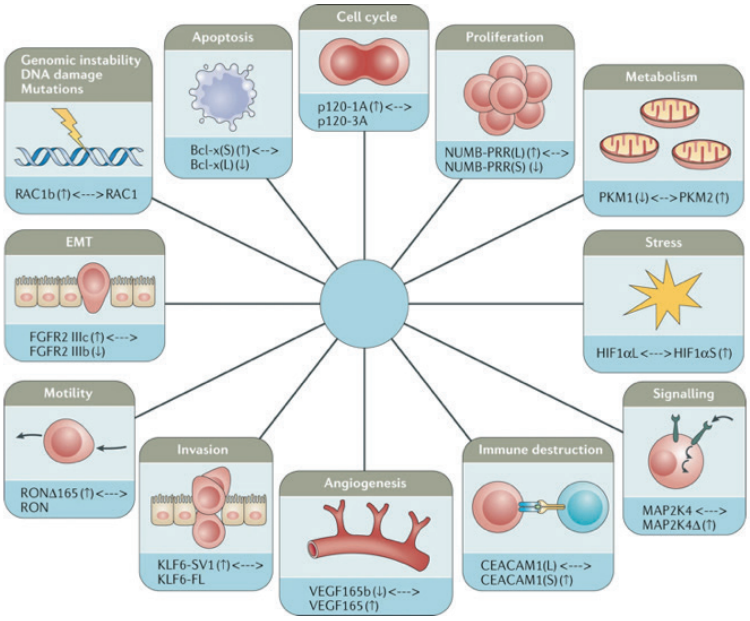


Figure 2. Alternate splicing deregulation impacts on tumour progression.
Arrows pointing up and down indicate whether different isoforms contribute positively or negatively to each tumour progression[37].

In view of their functional roles in alternative splicing and transcriptional regulation of several target genes in cancer, RBPs may represent on their own prognostic biomarkers and therapeutic targets. Several studies have shown that differential regulation of multiple RBPs is associated with resistance to chemotherapy. RBP polypyrimidine tract-binding protein 3 (PTBP3) have been reported to lead to hypoxia-induced chemoresistance in pancreatic cancer cells[38]. PTBP3 promotes metastasis and migration, as well as inducing EMT in other types of cancer[39,40]. Furthermore, a list of RBPs was revealed by genome-wide AS profiling across human cancers which are associated with specific drugs. The correlation of these RBPs with drug IC50 values was evaluated. Additionally, these RBPs form a dense network with drugs, such as pelitinib, masitinib, and omipalisib[41]. However, this study that proposes RBPs as therapeutic targets or anyhow related to above drugs, did not explore the function of RBPs in EMT.

Together with RBPs, other splicing factors involved in AS may represent potential therapeutic targets and/or clinical prognostic markers. Expression of SF3B1 (splicing factor 3b subunit 1) for example, has been associated with shorter overall survival in chronic lymphocytic leukemia (CLL)[42,43], mucosal melanoma[44], and breast cancer[45]. In T cell acute lymphoblastic leukemia (T-ALL), SF3B1 has been shown to perturb AS, inhibit survival, and impact therapeutic response[46].

In **Chapter 4**, our research showed that NUMB isoforms not only promote proliferation (NUMB1/3) but also induce EMT (NUMB2/4), invasion, and motility (NUMB2/4) in colon cancer. While studies on NUMB isoforms as therapeutic targets are limited, the NUMB gene plays a pivotal role in cancer therapy across various cancer types[47-49].

Therefore, RBPs, SFs, and their AS downstream targets represents potential therapeutic targets in a broad spectrum of different cancer types. RNA interference by short hairpin oligonucleotides

or the use of highly specific small molecule inhibitors capable of interfering with RBP-RNA interactions may pave the way for novel therapeutic approaches directed at RBPs. This approach aligns with the growing understanding of the intricate interplay between RBPs, alternative splicing, and malignant progression in various cancers.

4. AS-related signaling pathways in cancer progression.

In **Chapter 4 and 5**, the pathways associated with CD44, NUMB, and TPM1 isoforms, as identified through RNAseq analysis in colon and ovarian cancer cell lines include TGFα, Notch, Wnt β-catenin, and TGFβ signaling. Beyond these, other signaling pathways are relevant to specific AS targets across multiple cancer types.

As shown in **Chapter 4 and 5**, the Wnt β-catenin pathway is upregulated in both TPM1.8/9 OE ovarian cancer cells and CD44s OE colon cancer cell lines. GSK3β, a serine/threonine protein kinase, plays a pivotal role in regulating the Wnt/β-catenin pathway. Activation of GSK3β through Wnt signaling leads to the phosphorylation and subsequent degradation of its downstream target, β-catenin. In chronic myeloid leukemia, the novel exons 8/9 deleted isoform of GSK3β (m-GSK3β) lacks critical domains necessary for its function within this complex. Therefore, the presence of m-GSK3β impairs its ability to phosphorylate β-catenin, leading to its stabilization and enhanced activation. Lentiviral transduction of chronic phase (CP) CML progenitors with m-GSK3β leads to increased levels of activated β-catenin expression. Conversely, overexpression of full-length GSK3β reduces activated β-catenin expression and leukemic engraftment by BC progenitors in hematopoietic tissues.[50]. Through Wnt signaling, GSK-3β phosphorylates β-catenin at specific serine and threonine residues within exon 3[51] . In esophageal cancer metastasis, β-catenin exon 3 upregulates the expression of the oncogene c-Myc thus promote cancer cell migration[52].

Axin1 serves as a pivotal regulator of β-catenin degradation within the Wnt signaling pathway. It acts as a negative regulator and modulates the multiprotein complex responsible for β-catenin decrasing. Axin1 exists in two isoforms: Axin1v1, the full transcript with 11 exons encoding for 862 amino acids, and Axin1v2, lacking the 3' coding region (exon 8), resulting in a 36 amino acid product[53,54]. Additionally, Axin1v2 exhibits decreased interaction with protein phosphatase 2A(PP2A) compared to Axin1v1. This decreased interaction results in a reduced ability of Axin1v2 to promote the dephosphorylation of serine 62 (S62) on the Myc protein thus leading to its stabilization. Axin1v2 exhibits a higher ability to promote cancer metastasis than Axin1v1 in breast cancer cells[55]. Furthermore, the existence of another short Axin1 isoform with deletions of exons 1 and 2, as well as exons 4 and 5 was shown in hepatocellular carcinoma cell lines. Due to the deletion of essential binding domains, these isoforms are incapable of forming effective destruction complexes with β-catenin thus suppressing growth of hepatocellular and colorectal cancers[56]. Of note, aside from the short Axin1 isoform which decreased β-catenin expression, further research is needed to understand the correlation between Axin1 isoforms and the Wnt pathway.

Wnt ligands constitute a substantial class of secreted glycoproteins that intricately interact with Wnt receptors, governing the (dis)assembly of the β-catenin destruction complex. The alternative splicing of genes encoding for Wnt ligands typically occurs in the 5' and 3' regions as in Wnt2B/Wnt13, Wnt8a/d, and Wnt16 among others, through exon skipping and alternate promoter usage[57,58]. While functioning as morphogens to activate the canonical Wnt/β-catenin pathway during normal embryogenesis, the Wnt2B2/Wnt13A isoforms also play a role in malignant

transformation in gastric cancer[59]. Additionally, the overexpression of the Wnt16B variant in acute lymphoblastic leukemia is associated with the promotion of pre-B leukemic cell proliferation, whereas the Wnt16A isoform leads to decreased keratinocyte proliferation in basal cell carcinomas[60,61]. Moreover, the Wnt5A-short (Wnt5A-S) isoform has been identified as an inducer of cell growth in breast cancer, cervix cancer, and neuroblastoma cells, while the Wnt5A-long (Wnt5A-L) isoform exhibits tumor-suppressive activity in breast cancer, cervix cancer, and neuroblastoma cells. However, the same Wnt5A-L promotes tumorigenesis in melanomas, gastric, and pancreatic carcinomas[62-67].

Of note, many of these Wnt ligands isoforms affect the pathway by regulating β-catenin activity, including m-GSK3β, the short Axin1 isoform, and Wnt5A-S isoforms. However, the regulatory mechanisms and effects of Wnt signaling element isoforms on other Wnt members still require further exploration.

In addition to Wnt, we have also shown that the Notch pathway, mainly through NUMB, contributes to EMP in cancer metastasis. This was manifested by the distinct expression patterns of NUMB isoforms observed in the EpCAM^{lo} subpopulation in colon and ovarian cancer cells, as shown in **Chapters 4 and 5**. Notch signalling plays a crucial role in tumour development, controlling EMT, cell cycle, apoptosis, and proliferation across various cancer types[68]. The Notch pathway can function either as an oncogene or a tumour suppressor, in context-dependent fashion. For instance, Notch has a tumour suppressor role in medullary thyroid and prostate cancer, where overexpression of the Notch1 signalling pathway inhibits cancer cell growth[69,70]. On the other hand, it has also been implicated in promoting EMT and tumour aggressiveness in lung, oral, and breast cancers[71-73].

In summary, our exploration of AS-related signaling pathways has revealed a complex interplay that orchestrates cancer progression, as illustrated by the CD44, NUMB, and TPM1 isoforms affecting the Notch pathways, and by the splicing dynamics of critical members of the Wnt cascade including GSK3β, β-catenin, Axin1, and several ligands. The observed patterns of isoform expression at specific target genes emerge as key determinants in cancer development. The same isoforms are likely to represent both promising prognostic markers and therapeutic targets in a broad spectrum of malignancies, beyond colon and ovarian cancer.

5. The cancer molecular landscape: alternative splicing as a therapeutic frontier

5.1 RBPs and Variants Specific Compounds and Inhibitors

Currently, several strategies have been validated for the targeting of RBPs, splice factors, and downstream isoforms in cancer therapy: (1) small molecular compounds and drugs to inhibit the activity of splicing factors, RBPs, and their downstream gene targets; (2) RNA interference to degrade isoform-specific mRNAs; (3) isoform-specific antibodies conjugated to tumor cell toxins; and (4) splice-switching antisense oligonucleotides (SSOs) or synthetically modified oligonucleotides to block access to splicing factors[37,74,75] (Figure 3).

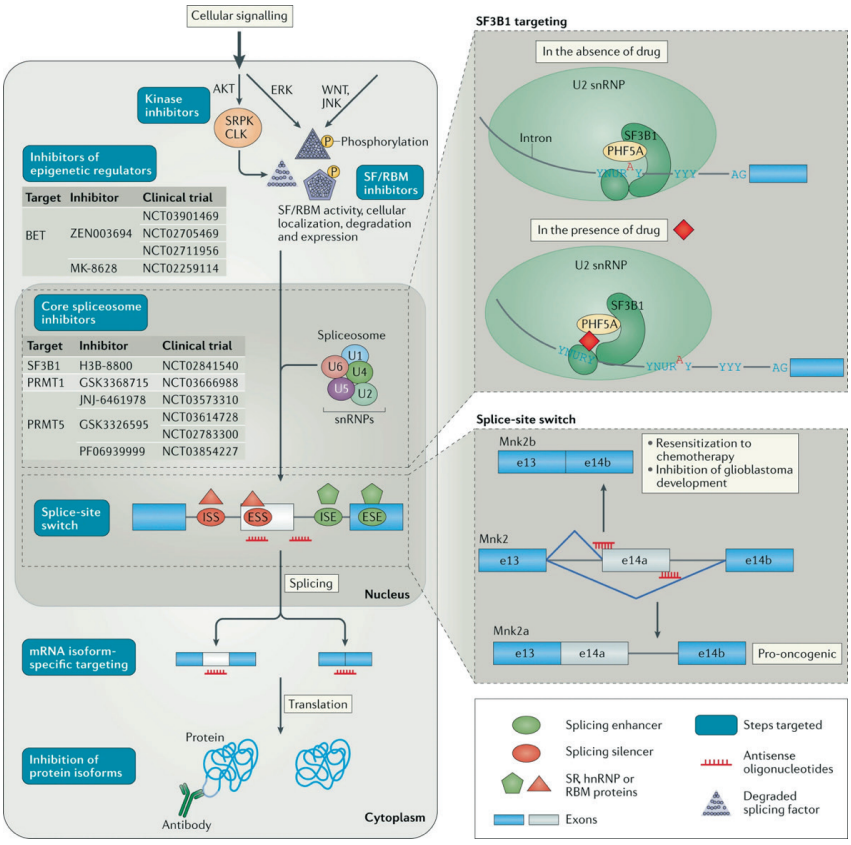


Figure 3: Methods for Controlling Cancer-Related Splicing Events.
Various strategies have focussed on cancer- and AS-related targets, including RNA-binding motif (RBM) proteins and splicing factors (SFs). Specific compounds bind SF3B1 and PHF5A in the U2 snRNP, induce splicing changes. The upper right inset shows how such a specific molecule may affect U2 snRNP function. Medications disrupting protein-protein or protein-RNA interactions, like antisense oligonucleotides and anti-isoform antibodies, induce mRNA isoforms or splice-site shifts (bottom right insets). An overview of the current ongoing trials involving small-molecule modulators targeting splicing elements and their targets can be found at ClinicalTrials.gov[37].

In **Chapter 5**, we have demonstrated that specific small molecule compounds tailor-made to target the TPM1.8/9 isoforms can effectively inhibit ovarian cancer cell resistance to cisplatin and paclitaxel. Likewise, inhibitors specifically targeting PLD exhibit potential in diminishing drug resistance and invasion in breast cancer cell lines. Notably, inhibitors targeting the PLD1.d311 isoform exhibit higher selectivity in vitro[76].

Kyle et al. identified isoform-specific inhibitors of the phosphoinositide 3-kinase (PI3K) gene class I PI3K, each with distinctive roles in breast cancer. Class I PI3K consists of a catalytic subunit (p110) and a regulatory subunit (p85), with the latter stabilizing p110 and suppressing its kinase

activity under basal conditions. The p110 subunit is further categorized into class IA enzymes, including p110α, p110β, and p110δ[77]. In mammals, p110α and p110β are widely expressed across various tissues, whereas p110δ is predominantly found in the immune system and is preferentially expressed in leukocytes[78,79]. Both p110α and p110β isoforms share a common domain structure and exhibit significant sequence homology[80]. In PTEN-negative breast tumor xenografts, the combination of p110α/δ and p110β isoform inhibitors demonstrated greater efficacy when compared with the p110α/β/δ inhibitor alone, as substantiated by both *in vivo* and *in vitro* studies. In the context of PTEN-negative tumor xenografts, pharmacodynamics biomarker analysis indicated that an inhibitor with limited effectiveness against the p110β isoform alone failed to block the PI3K pathway[81]. Additionally, another study uncovered the significant role of p110α-specific inhibitors in inhibiting the phosphorylation of both PKB (protein kinase B)/Akt and S6, contrasting with the effects of p110β inhibitors. Interestingly, p110α-specific inhibitors selectively reduced PKB/Akt phosphorylation in PTEN mutant cell lines, establishing a close connection to S6 phosphorylation[82]. These findings underscore the potential of isoform-specific PI3K inhibitors in precisely modulating signaling pathways for targeted and effective cancer treatment strategies. Isoform-specific PI3K inhibitors are being developed for clinical use with the aim of achieving targeted inhibition of specific isoforms while minimizing off-target side effects. The identification and characterization of isoform-specific PI3K inhibitors offer a promising approach for targeted cancer therapy, demonstrating superior efficacy and reduced off-target effects compared to broader inhibitors targeting multiple isoforms.

The advancement and evolution of metastatic prostate cancer have been notably impacted by alternative splicing variants of the androgen receptor (AR). AR is a nuclear receptor transcription factor responsible for mediating the cellular actions of androgens. It has been discovered that AR promotes EMT and the development of prostate cancer[83]. Among AR variants, splice variation 7 (AR-V7) has been extensively studied and associated with an increased probability of relapse and poorer overall survival outcomes. In clinical investigations, AR-V7 inhibitor compound has shown promise in re-sensitizing prostate cancer cell lines resistant to enzalutamide and abiraterone treatment[84,85]. Using siRNA against the RNA-binding protein hnRNP A1 and a BRD4 inhibitor to block the AR-V7-ZFX pathway could reduce AR-V7 expression, thus contributing to enzalutamide resistance in prostate cancer cells[86,87]. Another example Aminoacyl-tRNA synthetase interacting multifunctional protein 2(AIMP2) was found to have function in promoting proliferation, migration, and invasion of nasopharyngeal carcinoma cells[88]. AIMP2-DX2, a splicing variants of the tumour suppressor AIMP2, confers paclitaxel resistance in ovarian cancer[89]. A specific compound, BC-DXI01, targeting AIMP2-DX2, reduces AIMP2-DX2 expression at the RNA level and decreases tumour sizes and volumes by up to 60% compared to untreated mice[90].

Moreover, specific isoforms confer sensitivity to chemotherapy in various cancer types. For instance, BCL2 associated X (Bax) promotes MET in lung and breast cancer cells by binding to respiratory complex-I[91]. The Bax isoform BaxΔ2, observed in primary tumors and cell lines with microsatellite instability (MSI), induces cancer cell death without affecting mitochondria. The BaxΔ2 isoform selectively sensitizes MSI-H tumour cells (PC3, LS174T, SW1116, and 104-R1) to chemotherapeutic agents like the topoisomerase II inhibitor Adriamycin, proteasome inhibitor bortezomib, and endoplasmic reticulum stress inducer tunicamycin[92].

The gene encoding the epidermal growth factor receptor (EGFR) directs the synthesis of a receptor protein. This protein spans the cell membrane, with one end positioned inside the cell

and the other projecting outward from the cell surface. Mutations in EGFR have been identified in specimens from patients with NSCLC, making it a crucial therapeutic target for tumor treatment[93,94] . EGFR exhibits various isoforms, including EGFRvI (deletion of the N-terminus), EGFRvII (Δexons 14 and 15), EGFRvIII (Δexons 2–7), EGFRvIV (Δexons 25–27), and EGFRvV (Δexons 25–28)[95]. In glioblastoma multiforme (GBM) studies, EGFR isoform EGFRvIII confers resistance to ionizing radiation (IR), linked to proficient double-strand breaks repair in primary murine astrocytes[96]. However, NSCLC, gefitinib and erlotinib generally showed resistance in vitro to Ba/F3 cells transformed with the EGFRvIII isoform mutant, while the EGFR inhibitor HKI-272 rendered them responsive[97].

Another example of potential AS-related therapy target is survivin, also known as BIRC5 (baculoviral inhibitor of apoptosis repeat-containing 5), a gene belonging to the inhibitor of apoptosis (IAP) family that promotes EMT and cell migration in multiple cancer types[98,99] . BIRC5 exists in two isoforms: survivin-2B and survivin-ΔEx3. While survivin-ΔEx3 lacks exon 3, survivin-2B retains part of intron 2 as a cryptic exon. Survivin-ΔEx3 largely retains its anti-apoptotic properties, whereas survivin-2B exhibits reduced efficacy in inhibiting apoptosis. It has been shown that AS of the survivin isoform 2B by siRNA inhibits ovarian cancer cell resistance to docetaxel chemotherapy, both *in vivo* and *in vitro*. Progression-free survival data from 117 initial debulking surgeries on ovarian cancer patients supported the role of survivin isoforms in chemoresistance[100].

In summary, the growing research on AS across various cancer types underscores the potentials of isoform-specific small molecule inhibitors and/or interference oligonucleotides as therapeutic agents.

5.2 RNA virus and mRNA vaccines

Viruses can infect host cells and take full control of the host's fate. RNA viruses, including Orthomyxoviruses, the Hepatitis C Virus (HCV), Ebola, SARS, influenza, polio, measles, and other retroviruses such as HTLV-1 and HIV, are major players in human diseases. mRNA vaccines have been shown to confer strong immunity against various infectious diseases in animal models[101]. Infections of specific types of RNA viruses, i.e. human papillomavirus (HPV), are strongly associated with cancer[102].

The development of RNA-based vaccines during the Covid pandemics has sparked increased interest for their use in cancer therapy. Various approaches to mRNA cancer vaccines have been explored in pre-clinical and clinical trials with promising results in antigen-specific T cell responses and prolonged disease-free survival in various cancer types[103]. These vaccines can be designed to target tumor-associated antigens expressed specifically in cancer cells, including growth-associated factors or antigens specific to cancer metastasis cells. Neoantigens, along with their neoepitopes, have been utilized as mRNA vaccine targets, showcasing enhanced CD8+ T cells activity and decreased tumor growth in mouse models[104]. Another avenue involves self-replicating RNA (srRNA) cancer vaccines, where intramuscular vaccination of naked srRNA molecules has shown efficacy in eliciting anti-tumor responses and providing protection against colorectal carcinoma[105].

Overall, the current assessment of isoforms as a therapeutic strategy holds promise in the specific case of cancer and its molecular complexity. Our findings, detailed across various chapters, underscore the potential of isoform-specific interventions to overcome drug resistance and enhance treatment efficacy. Whether directly targeting specific isoforms involved in cancer cell

resistance to chemotherapy such as TPM1.8/9 and PLD1.d311, or modulating critical pathways like PI3K with isoform-specific inhibitors, the potential for precision medicine in cancer therapeutics is evident. The diverse examples of isoform-specific responses in different cancer types highlight the personalized nature of such interventions. As we delve deeper into the molecular mechanism of alternative splicing, it becomes increasingly apparent that unraveling these complexities holds the key to unlocking more effective and tailored cancer therapies. The integration of AS target-specific mRNA vaccination with complementary therapies, including traditional chemotherapy, radiotherapy, and compounds/inhibitors, has improved positive outcomes in preclinical cancer studies and shows potential for future clinical therapy.

6. Conclusion

In this thesis, we investigated the role of alternative splicing in shaping EMP and influencing therapeutic approaches within the realms of colon and ovarian cancer. Employing a multifaceted approach, we utilized cancer cell lines, mouse models, patient-derived tissues and ascites from colon and ovarian cancer patients, and extensive cancer patient data analyzed *in silico*. This comprehensive strategy allowed us to emphasize that alternative splicing serves as the underlying mechanism influencing the dynamics of stemness and invasive subpopulations within metastatic cancer cells. Moreover, this thesis advances tour understanding of alternative splicing's role in driving phenotypic plasticity, EMT/MET, and resistance to therapy in colon and ovarian cancer. By characterizing quasi-mesenchymal cancer cell subpopulations and identifying key regulatory factors, this thesis offers valuable insights into the mechanisms underlying metastasis and highlights novel potential therapeutic targets. Last, the validation of novel splicing variants underscores the potential for developing targeted therapies to disrupt invasive and chemoresistant phenotypes in these cancers.

References

1 Brabletz, T., Jung, A., Spaderna, S., Hlubek, F. & Kirchner, T. Opinion: migrating cancer stem cells - an integrated concept of malignant tumour progression. Nat Rev Cancer 5, 744-749, doi:10.1038/nrc1694 (2005).

2 Gupta, P. B. et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell 146, 633-644, doi:10.1016/j.cell.2011.07.026 (2011).

3 Ricci-Vitiani, L. et al. Identification and expansion of human colon-cancer-initiating cells. Nature 445, 111-115, doi:10.1038/nature05384 (2007).

4 Grimshaw, M. J. et al. Mammosphere culture of metastatic breast cancer cells enriches for tumorigenic breast cancer cells. Breast Cancer Res 10, R52, doi:10.1186/bcr2106 (2008).

5 Han, M. E. et al. Cancer spheres from gastric cancer patients provide an ideal model system for cancer stem cell research. Cell Mol Life Sci 68, 3589-3605, doi:10.1007/s00018-011-0672-z (2011).

6 Pastushenko, I. et al. Identification of the tumour transition states occurring during EMT. Nature 556, 463-468, doi:10.1038/s41586-018-0040-3 (2018).

7 Aiello, N. M. et al. EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. Dev Cell 45, 681-695 e684, doi:10.1016/j.devcel.2018.05.027 (2018).

8 Campbell, K. et al. Collective cell migration and metastases induced by an epithelial-to-mesenchymal transition in Drosophila intestinal tumors. Nat Commun 10, 2311, doi:10.1038/s41467-019-10269-y (2019).

9 Luond, F. et al. Distinct contributions of partial and full EMT to breast cancer malignancy. Dev Cell 56, 3203-3221 e3211, doi:10.1016/j.devcel.2021.11.006 (2021).

10 Papadaki, M. A. et al. Circulating Tumor Cells with Stemness and Epithelial-to-Mesenchymal Transition Features Are Chemoresistant and Predictive of Poor Outcome in Metastatic Breast Cancer. Mol Cancer Ther 18, 437-447, doi:10.1158/1535-7163.MCT-18-0584 (2019).

11 Zhu, X. et al. miR-186 regulation of Twist1 and ovarian cancer sensitivity to cisplatin. Oncogene 35, 323-332, doi:10.1038/onc.2015.84 (2016).

12 Berman, M., Mattheolabakis, G., Suresh, M. & Amiji, M. Reversing epigenetic mechanisms of drug resistance in solid tumors using targeted microRNA delivery. Expert Opin Drug Deliv 13, 987-998, doi:10.1080/17425247.2016.1178236 (2016).

13 Pastushenko, I. et al. Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. Nature 589, 448-455, doi:10.1038/s41586-020-03046-1 (2021).

14 Pastushenko, I. & Blanpain, C. EMT Transition States during Tumor Progression and Metastasis. Trends Cell Biol 29, 212-226, doi:10.1016/j.tcb.2018.12.001 (2019).

15 Pal, A., Barrett, T. F., Paolini, R., Parikh, A. & Puram, S. V. Partial EMT in head and neck cancer biology: a spectrum instead of a switch. Oncogene 40, 5049-5065, doi:10.1038/s41388-021-01868-5 (2021).

16 Liao, C. et al. Partial EMT in Squamous Cell Carcinoma: A Snapshot. Int J Biol Sci 17, 3036-3047, doi:10.7150/ijbs.61566 (2021).

17 De Craene, B. & Berx, G. Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer 13, 97-110, doi:10.1038/nrc3447 (2013).

18 Pistore, C. et al. DNA methylation variations are required for epithelial-to-mesenchymal transition induced by cancer-associated fibroblasts in prostate cancer cells. Oncogene 36, 5551-5566, doi:10.1038/onc.2017.159 (2017).

19 Pillman, K. A. et al. miR-200/375 control epithelial plasticity-associated alternative splicing by repressing the RNA-binding protein Quaking. EMBO J 37, doi:10.15252/embj.201899016 (2018).

20 Teng, L. et al. The pan-cancer lncRNA PLANE regulates an alternative splicing program to promote cancer pathogenesis. Nat Commun 12, 3734, doi:10.1038/s41467-021-24099-4 (2021).

21 Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470-476, doi:10.1038/nature07509 (2008).

22 Xu, T. et al. Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behavior in colon cancer. Elife 11, doi:10.7554/eLife.82006 (2022).

23 Warzecha, C. C., Shen, S., Xing, Y. & Carstens, R. P. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. RNA Biol 6, 546-562, doi:10.4161/rna.6.5.9606 (2009).

24 Talvinen, K. et al. Altered expression of p120catenin predicts poor outcome in invasive breast cancer. J Cancer Res Clin Oncol 136, 1377-1387, doi:10.1007/s00432-010-0789-8 (2010).

25 Slorach, E. M., Chou, J. & Werb, Z. Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization. Genes Dev 25, 471-484, doi:10.1101/gad.1998111 (2011).

26 David, G. et al. The RNA-binding proteins CELF1 and ELAVL1 cooperatively control the alternative splicing of CD44. Biochem Biophys Res Commun 626, 79-84, doi:10.1016/j.bbrc.2022.07.073 (2022).

27 Muys, B. R. et al. The p53-induced RNA-binding protein ZMAT3 is a splicing regulator that inhibits the splicing of oncogenic CD44 variants in colorectal carcinoma. Genes Dev 35, 102-116, doi:10.1101/gad.342634.120 (2021).

28 Chen, Q. et al. TGF-beta1 promotes epithelial-to-mesenchymal transition and stemness of prostate cancer cells by inducing PCBP1 degradation and alternative splicing of CD44. Cell Mol Life Sci 78, 949-962, doi:10.1007/s00018-020-03544-5 (2021).

29 Xu, Y. et al. Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. Genes Dev 28, 1191-1203, doi:10.1101/gad.241968.114 (2014).

30 Watermann, D. O. et al. Splicing factor Tra2-beta1 is specifically induced in breast cancer and regulates alternative splicing of the CD44 gene. Cancer Res 66, 4774-4780, doi:10.1158/0008-5472.CAN-04-3294 (2006).

31 Stickeler, E. et al. The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. EMBO J 20, 3821-3830, doi:10.1093/emboj/20.14.3821 (2001).

32 Franco-Valls, H. et al. Formation of an invasion-permissive matrix requires TGFbeta/SNAIL1-regulated alternative splicing of fibronectin. Breast Cancer Res 25, 143, doi:10.1186/s13058-023-01736-y (2023).

33 DeLigio, J. T., Lin, G., Chalfant, C. E. & Park, M. A. Splice variants of cytosolic polyadenylation element-binding protein 2 (CPEB2) differentially regulate pathways linked to cancer metastasis. J Biol Chem 292, 17909-17918, doi:10.1074/jbc.M117.810127 (2017).

34 Carninci, P. et al. The transcriptional landscape of the mammalian genome. Science 309, 1559-1563, doi:10.1126/science.1112014 (2005).

35 Giudice, G., Sanchez-Cabo, F., Torroja, C. & Lara-Pezzi, E. ATtRACT-a database of RNA-binding proteins and associated motifs. Database (Oxford) 2016, doi:10.1093/database/baw035 (2016).

36 Yee, B. A., Pratt, G. A., Graveley, B. R., Van Nostrand, E. L. & Yeo, G. W. RBP-Maps enables robust generation of splicing regulatory maps. RNA 25, 193-204, doi:10.1261/rna.069237.118 (2019).

37 Bonnal, S. C., Lopez-Oreja, I. & Valcarcel, J. Roles and mechanisms of alternative splicing in cancer - implications for care. Nat Rev Clin Oncol 17, 457-474, doi:10.1038/s41571-020-0350-x (2020).

38 Ma, J. et al. PTBP3 promotes malignancy and hypoxia-induced chemoresistance in pancreatic cancer cells by ATG12 up-regulation. J Cell Mol Med 24, 2917-2930, doi:10.1111/jcmm.14896 (2020).

39 Wu, Q. et al. PTBP3 promotes migration of non-small cell lung cancer through regulating E-cadherin in EMT signaling pathway. Cancer Cell Int 20, 172, doi:10.1186/s12935-020-01240-9 (2020).

40 Hou, P. et al. PTBP3-Mediated Regulation of ZEB1 mRNA Stability Promotes Epithelial-Mesenchymal Transition in Breast Cancer. Cancer Res 78, 387-398, doi:10.1158/0008-5472.CAN-17-0883 (2018).

41 Li, J. et al. Alternative splicing perturbation landscape identifies RNA binding proteins as potential therapeutic targets in cancer. Mol Ther Nucleic Acids 24, 792-806, doi:10.1016/j.omtn.2021.04.005 (2021).

42 Quesada, V. et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 44, 47-52, doi:10.1038/ng.1032 (2011).

43 Zhang, Z. et al. SF3B1 mutation is a prognostic factor in chronic lymphocytic leukemia: a meta-analysis. Oncotarget 8, 69916-69923, doi:10.18632/oncotarget.19455 (2017).

44 Quek, C. et al. Recurrent hotspot SF3B1 mutations at codon 625 in vulvovaginal mucosal melanoma identified in a study of 27 Australian mucosal melanomas. Oncotarget 10, 930-941, doi:10.18632/oncotarget.26584 (2019).

45 Fu, X. et al. SF3B1 mutation is a poor prognostic indicator in luminal B and progesterone receptor-negative breast cancer patients. Oncotarget 8, 115018-115027, doi:10.18632/oncotarget.22983 (2017).

46 Han, C. et al. SF3B1 homeostasis is critical for survival and therapeutic response in T cell leukemia. Sci Adv 8, eabj8357, doi:10.1126/sciadv.abj8357 (2022).

47 Belle, V. A., McDermott, N., Meunier, A. & Marignol, L. NUMB inhibition of NOTCH signalling as a therapeutic target in prostate cancer. Nat Rev Urol 11, 499-507, doi:10.1038/nrurol.2014.195 (2014).

48 Tosoni, D. et al. Pre-clinical validation of a selective anti-cancer stem cell therapy for Numb-deficient human breast cancers. EMBO Mol Med 9, 655-671, doi:10.15252/emmm.201606940 (2017).

49 Hristova, D. M. et al. NUMB as a Therapeutic Target for Melanoma. J Invest Dermatol 142, 1882-1892 e1885, doi:10.1016/j.jid.2021.11.027 (2022).

50 Abrahamsson, A. E. et al. Glycogen synthase kinase 3beta missplicing contributes to leukemia stem cell generation. Proc Natl Acad Sci U S A 106, 3925-3929, doi:10.1073/pnas.0900189106 (2009).

51 Yost, C. et al. The axis-inducing activity, stability, and subcellular distribution of beta-catenin is regulated in Xenopus embryos by glycogen synthase kinase 3. Genes Dev 10, 1443-1454, doi:10.1101/gad.10.12.1443 (1996).

52 Roth, M. J. et al. beta-Catenin splice variants and downstream targets as markers for neoplastic progression of esophageal cancer. Genes Chromosomes Cancer 44, 423-428, doi:10.1002/gcc.20251 (2005).

53 Parveen, N., Hussain, M. U., Pandith, A. A. & Mudassar, S. Diversity of axin in signaling pathways and its relation to colorectal cancer. Med Oncol 28 Suppl 1, S259-267, doi:10.1007/s12032-010-9722-x (2011).

54 Salahshor, S. & Woodgett, J. R. The links between axin and carcinogenesis. J Clin Pathol 58, 225-236, doi:10.1136/jcp.2003.009506 (2005).

55 Zhang, X. et al. Mechanistic insight into Myc stabilization in breast cancer involving aberrant Axin1 expression. Proc Natl Acad Sci U S A 109, 2790-2795, doi:10.1073/pnas.1100764108 (2012).

56 Satoh, S. et al. AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. Nat Genet 24, 245-250, doi:10.1038/73448 (2000).

57 Katoh, M., Kirikoshi, H., Saitoh, T., Sagara, N. & Koike, J. Alternative splicing of the WNT-2B/WNT-13 gene. Biochem Biophys Res Commun 275, 209-216, doi:10.1006/bbrc.2000.3252 (2000).

58 Struewing, I. T., Toborek, A. & Mao, C. D. Mitochondrial and nuclear forms of Wnt13 are generated via alternative promoters, alternative RNA splicing, and alternative translation start sites. J Biol Chem 281, 7282-7293, doi:10.1074/jbc.M511182200 (2006).

59 Katoh, M., Kirikoshi, H., Terasaki, H. & Shiokawa, K. WNT2B2 mRNA, up-regulated in primary gastric cancer, is a positive regulator of the WNT- beta-catenin-TCF signaling pathway. Biochem Biophys Res Commun 289, 1093-1098, doi:10.1006/bbrc.2001.6076 (2001).

60 Mazieres, J. et al. Inhibition of Wnt16 in human acute lymphoblastoid leukemia cells containing the t(1;19) translocation induces apoptosis. Oncogene 24, 5396-5400, doi:10.1038/sj.onc.1208568 (2005).

61 Teh, M. T. et al. Role for WNT16B in human epidermal keratinocyte proliferation and differentiation. J Cell Sci 120, 330-339, doi:10.1242/jcs.03329 (2007).

62 Bauer, M., Benard, J., Gaasterland, T., Willert, K. & Cappellen, D. WNT5A encodes two isoforms with distinct functions in cancers. PLoS One 8, e80526, doi:10.1371/journal.pone.0080526 (2013).

63 Weeraratna, A. T. et al. Wnt5a signaling directly affects cell motility and invasion of metastatic melanoma. Cancer Cell 1, 279-288, doi:10.1016/s1535-6108(02)00045-4 (2002).

64 Da Forno, P. D. et al. WNT5A expression increases during melanoma progression and correlates with outcome. Clin Cancer Res 14, 5825-5832, doi:10.1158/1078-0432.CCR-07-5104 (2008).

65 Liang, H. et al. Wnt5a inhibits B cell proliferation and functions as a tumor suppressor in hematopoietic tissue. Cancer Cell 4, 349-360, doi:10.1016/s1535-6108(03)00268-x (2003).

66 Ying, J. et al. WNT5A exhibits tumor-suppressive activity through antagonizing the Wnt/beta-catenin signaling, and is frequently methylated in colorectal cancer. Clin Cancer Res 14, 55-61, doi:10.1158/1078-0432.CCR-07-1644 (2008).

67 Bi, L. et al. Wnt5a involved in regulation of the biological behavior of hepatocellular carcinoma. Int J Clin Exp Pathol 7, 987-995 (2014).

68 Zhang, L. et al. Activation of Notch pathway is linked with epithelial-mesenchymal transition in prostate cancer cells. Cell Cycle 16, 999-1007, doi:10.1080/15384101.2017.1312237 (2017).

69 Shou, J., Ross, S., Koeppen, H., de Sauvage, F. J. & Gao, W. Q. Dynamics of notch expression during murine prostate development and tumorigenesis. Cancer Res 61, 7291-7297 (2001).

70 Kunnimalaiyaan, M., Vaccaro, A. M., Ndiaye, M. A. & Chen, H. Overexpression of the NOTCH1 intracellular domain inhibits cell proliferation and alters the neuroendocrine phenotype of medullary thyroid cancer cells. J Biol Chem 281, 39819-39830, doi:10.1074/jbc.M603578200 (2006).

71 Kar, R. et al. A "NOTCH" Deeper into the Epithelial-To-Mesenchymal Transition (EMT) Program in Breast Cancer. Genes (Basel) 10, doi:10.3390/genes10120961 (2019).

72 Yuan, X. et al. Notch signaling and EMT in non-small cell lung cancer: biological significance and therapeutic application. J Hematol Oncol 7, 87, doi:10.1186/s13045-014-0087-z (2014).

73 Ishida, T., Hijioka, H., Kume, K., Miyawaki, A. & Nakamura, N. Notch signaling induces EMT in OSCC cell lines in a hypoxic environment. Oncol Lett 6, 1201-1206, doi:10.3892/ol.2013.1549 (2013).

74 Zhang, Y., Qian, J., Gu, C. & Yang, Y. Alternative splicing and cancer: a systematic review. Signal Transduct Target Ther 6, 78, doi:10.1038/s41392-021-00486-7 (2021).

75 Ghigna, C., Valacca, C. & Biamonti, G. Alternative splicing and tumor progression. Curr Genomics 9, 556-570, doi:10.2174/138920208786847971 (2008).

76 Scott, S. A. et al. Design of isoform-selective phospholipase D inhibitors that modulate cancer cell invasiveness. Nat Chem Biol 5, 108-117, doi:10.1038/nchembio.140 (2009).

77 Vanhaesebroeck, B., Guillermet-Guibert, J., Graupera, M. & Bilanges, B. The emerging mechanisms of isoform-specific PI3K signalling. Nat Rev Mol Cell Biol 11, 329-341, doi:10.1038/nrm2882 (2010).

78 Chantry, D. et al. p110delta, a novel phosphatidylinositol 3-kinase catalytic subunit that associates with p85 and is expressed predominantly in leukocytes. J Biol Chem 272, 19236-19241, doi:10.1074/jbc.272.31.19236 (1997).

79 Vanhaesebroeck, B. et al. P110delta, a novel phosphoinositide 3-kinase in leukocytes. Proc Natl Acad Sci U S A 94, 4330-4335, doi:10.1073/pnas.94.9.4330 (1997).

80 Bi, L., Okabe, I., Bernard, D. J. & Nussbaum, R. L. Early embryonic lethality in mice deficient in the p110beta catalytic subunit of PI 3-kinase. Mamm Genome 13, 169-172, doi:10.1007/BF02684023 (2002).

81 Edgar, K. A. et al. Isoform-specific phosphoinositide 3-kinase inhibitors exert distinct effects in solid tumors. Cancer Res 70, 1164-1172, doi:10.1158/0008-5472.CAN-09-2525 (2010).

82 Torbett, N. E. et al. A chemical screen in diverse breast cancer cell lines reveals genetic enhancers and suppressors of sensitivity to PI3K isoform-selective inhibition. Biochem J 415, 97-110, doi:10.1042/BJ20080639 (2008).

83 Yuan, X. et al. Androgen receptor functions in castration-resistant prostate cancer and mechanisms of resistance to new agents targeting the androgen axis. Oncogene 33, 2815-2825, doi:10.1038/onc.2013.235 (2014).

84 Liu, C. et al. Niclosamide inhibits androgen receptor variants expression and overcomes enzalutamide resistance in castration-resistant prostate cancer. Clin Cancer Res 20, 3198-3210, doi:10.1158/1078-0432.CCR-13-3296 (2014).

85 Li, Y. et al. Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. Cancer Res 73, 483-489, doi:10.1158/0008-5472.CAN-12-3630 (2013).

86 Nadiminty, N. et al. NF-kappaB2/p52:c-Myc:hnRNPA1 Pathway Regulates Expression of Androgen Receptor Splice Variants and Enzalutamide Sensitivity in Prostate Cancer. Mol Cancer Ther 14, 1884-1895, doi:10.1158/1535-7163.MCT-14-1057 (2015).

87 Tummala, R., Lou, W., Gao, A. C. & Nadiminty, N. Quercetin Targets hnRNPA1 to Overcome Enzalutamide Resistance in Prostate Cancer Cells. Mol Cancer Ther 16, 2770-2779, doi:10.1158/1535-7163.MCT-17-0030 (2017).

88 Cao, Q., Zhang, J. & Zhang, T. AIMP2-DX2 Promotes the Proliferation, Migration, and Invasion of Nasopharyngeal Carcinoma Cells. Biomed Res Int 2018, 9253036, doi:10.1155/2018/9253036 (2018).

89 Choi, J. W. et al. Splicing variant of AIMP2 as an effective target against chemoresistant ovarian cancer. J Mol Cell Biol 4, 164-173, doi:10.1093/jmcb/mjs018 (2012).

90 Lee, H. S. et al. Chemical suppression of an oncogenic splicing variant of AIMP2 induces tumour regression. Biochem J 454, 411-416, doi:10.1042/BJ20130550 (2013).

91 Kim, E. M., Jung, C. H., Song, J. Y., Park, J. K. & Um, H. D. Pro-apoptotic Bax promotes mesenchymal-epithelial transition by binding to respiratory complex-I and antagonizing the malignant actions of pro-survival Bcl-2 proteins. Cancer Lett 424, 127-135, doi:10.1016/j.canlet.2018.03.033 (2018).

92 Haferkamp, B. et al. BaxDelta2 is a novel bax isoform unique to microsatellite unstable tumors. J Biol Chem 287, 34722-34729, doi:10.1074/jbc.M112.374785 (2012).

93 Kobayashi, S. et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med 352, 786-792, doi:10.1056/NEJMoa044238 (2005).

94 da Cunha Santos, G., Shepherd, F. A. & Tsao, M. S. EGFR mutations and lung cancer. Annu Rev Pathol 6, 49-69, doi:10.1146/annurev-pathol-011110-130206 (2011).

95 Rutkowska, A., Stoczynska-Fidelus, E., Janik, K., Wlodarczyk, A. & Rieske, P. EGFR(vIII): An Oncogene with Ambiguous Role. J Oncol 2019, 1092587, doi:10.1155/2019/1092587 (2019).

96 Mukherjee, B. et al. EGFRvIII and DNA double-strand break repair: a molecular mechanism for radioresistance in glioblastoma. Cancer Res 69, 4252-4259, doi:10.1158/0008-5472.CAN-08-4853 (2009).

97 Ji, H. et al. Epidermal growth factor receptor variant III mutations in lung tumorigenesis and sensitivity to tyrosine kinase inhibitors. Proc Natl Acad Sci U S A 103, 7817-7822, doi:10.1073/pnas.0510284103 (2006).

98 Li, C., Li, J., Wu, D. & Han, G. The involvement of survivin in insulin-like growth factor 1-induced epithelial-mesenchymal transition in gastric cancer. Tumour Biol 37, 1091-1096, doi:10.1007/s13277-015-3909-x (2016).

99 Tai, C. J. et al. Survivin-mediated cancer cell migration through GRP78 and epithelial-mesenchymal transition (EMT) marker expression in Mahlavu cells. Ann Surg Oncol 19, 336-343, doi:10.1245/s10434-011-1692-5 (2012).

100 Vivas-Mejia, P. E. et al. Silencing survivin splice variant 2B leads to antitumor activity in taxane--resistant ovarian cancer. Clin Cancer Res 17, 3716-3726, doi:10.1158/1078-0432.CCR-11-0233 (2011).

101 Poltronieri, P., Sun, B. & Mallardo, M. RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load. Curr Genomics 16, 327-335, doi:10.2174/1389202916666150707160613 (2015).

102 Bialkowski, L. et al. Intralymphatic mRNA vaccine induces CD8 T-cell responses that inhibit the growth of mucosally located tumours. Sci Rep 6, 22509, doi:10.1038/srep22509 (2016).

103 Pardi, N., Hogan, M. J., Porter, F. W. & Weissman, D. mRNA vaccines - a new era in vaccinology. Nat Rev Drug Discov 17, 261-279, doi:10.1038/nrd.2017.243 (2018).

104 Van der Jeught, K. et al. Intratumoral administration of mRNA encoding a fusokine consisting of IFN-beta and the ectodomain of the TGF-beta receptor II potentiates antitumor immunity. Oncotarget 5, 10100-10113, doi:10.18632/oncotarget.2463 (2014).

105 Ying, H. et al. Cancer therapy using a self-replicating RNA vaccine. Nat Med 5, 823-827, doi:10.1038/10548 (1999).

CHAPTER 7

Summary
Samenvatting

SUMMARY

The focus of this thesis is the study of the relative role of alternative splicing in shaping phenotypic plasticity, orchestrating EMT and influencing therapeutic approaches in the context of colon and ovarian cancer. **In Chapters 2 and 3**, our exploration revealed the existence of quasi-mesenchymal subpopulations of cancer cells, maintained through E-to-M and M-to-E state transitions, with migratory and chemotherapy-resistant features. We present the discovery and characterization of a subgroup of colon cancer cells exhibiting phenotypic flexibility, contributing to local infiltration and distant metastasis. These cells, identified as EpCAM^{lo}, demonstrate significant motility, invasiveness, resistance to chemotherapy, and metastatic potentialboth in vitro and in vivo. Examination through bulk and single-cell RNA sequencing reveals a diverse range of EMT activation levels within this population, coupled with traits reminiscent of stem cells, increased Wnt/β-catenin pathway activity, and a strong association with the CMS4 subtype, which correlates with poor prognosis in approximately 25% of colon cancer cases. Further analysis uncovers distinct sets of epithelial and mesenchymal genes orchestrating transcriptional shifts, thereby delineating pathways of cellular differentiation. Notably, in chapter 3 the identification of a distinct subpopulation of sphere-like cells, reminiscent of resident CSCs, further underscore the utility of immortalized cancer cell lines for the study on cancer stemness.

In **Chapter 4 and 5** we further unraveled the epigenetic mechanisms which underlie alternative splicing. The broad spectrum of AS-targets associated with EMT and with Wnt signaling connects phenotypic plasticity and chemotherapeutic resistance, i.e. the main metastatic traits of mCSCs. In Chapter 4, we demonstrate the impact of differential expression of ESRP1 and other RBPs, downstream effectors of the EMT master regulator ZEB1, on AS patterns. This regulation extends to a wide array of targets, including CD44 and NUMB, resulting in the generation of specific isoforms associated with invasion and metastasis features. Further validation studies, both functional and clinical, confirm the roles of these novel identified RBPs, as well as the CD44s and NUMB2/4 splicing variants, in promoting local invasion, distant metastasis, and correlating with diminished survival rates in colon cancer patients. In Chapter 5, we investigated the influence of EMT-related gene isoforms on the regulation of EMP in high-grade serous ovarian cancer. Comparing our findings with those in colon cancer, we observed a distinct spectrum of downstream targets in quasi-mesenchymal ovarian cancer cells, likely reflecting differences in metastatic mechanisms between these malignancies—hematogenous in colon cancer and transcoelomic in ovarian cancer. Furthermore, we identified upstream RBPs that are differentially expressed between epithelial and quasi-mesenchymal subpopulations of ovarian cancer cells, governing the regulation of EMT-related isoforms. The upregulation of RBM24 and the downregulation of ESRP1 emerged as key regulators of EMT in ovarian cancer cells. To validate the functional and clinical significance of our findings, we focused on the TPM1, which encodes a protein crucial for defining the functional properties of individual actin filaments in contractile cells, among the downstream AS targets specific to ovarian cancer. We found that the Tpm1.8/9 isoforms are selectively expressed in patient-derived ascites and promote invasion through the activation of EMT and Wnt signaling pathways, along with a wide range of inflammation-related pathways. Moreover, Tpm1.8/9 expression confers resistance to taxane- and platinum-based chemotherapy.

In Chapter 6, we discuss the challenges associated with leveraging alternative splicing targets in clinical therapeutic approaches for cancer treatment. The identification of specific compounds directed against members of the AS machinery opens novel avenues for the development of therapeutic tools and anticancer targets. The integration of cancer sequencing data with experimental validation holds the promise of unraveling concealed basic mechanisms and pharmaceutical targets, paving the way for innovative cancer treatments in the future.

SAMENVATTING

Het doel van deze thesis is het bestuderen van de relatieve rol van alternatieve splicing bij het vormgeven van fenotypische plasticiteit, het orkestreren van EMT en het beïnvloeden van therapeutische benaderingen in de context van darm- en ovariumkanker. In Hoofdstukken 2 en 3 onthulde ons onderzoek het bestaan van quasi-mesenchymale subpopulaties van kankercellen, in stand gehouden door E-naar-M en M-naar-E transitieprocessen, met migrerende en chemotherapie-resistente eigenschappen. We presenteren de ontdekking en karakterisering van een subgroep darmkankercellen die fenotypische flexibiliteit vertonen, bijdragend aan lokale infiltratie en verre metastase. Deze cellen, geïdentificeerd als EpCAMlo, vertonen aanzienlijke mobiliteit, invasiviteit, resistentie tegen chemotherapie en metastatische potentie zowel in vitro als in vivo. Onderzoek via bulk- en single-cell RNA-sequencing onthult een divers scala aan EMT-activatieniveaus binnen deze populatie, gekoppeld aan kenmerken die doen denken aan stamcellen, verhoogde activiteit van het Wnt/β-cateninepad en een sterke associatie met het CMS4-subtype, dat overeenkomt met een slechte prognose bij ongeveer 25% van de darmkankerpatiënten. Verder onderzoek onthult verschillende sets van epitheliale en mesenchymale genen die transciptionele verschuivingen orchestreren, waardoor paden van celdifferentiatie worden afgebakend. Met name in hoofdstuk 3 benadrukt de identificatie van een afzonderlijke subpopulatie van bolvormige cellen, die doen denken aan residente CSC's, de bruikbaarheid van geïmmortaliseerde kankercellijnen voor onderzoek naar kankerstamcellen.

In Hoofdstukken 4 en 5 hebben we verder de epigenetische mechanismen ontrafeld die ten grondslag liggen aan alternatieve splicing. Het brede spectrum van AS-doelen geassocieerd met EMT en met Wnt-signalering verbindt fenotypische plasticiteit en chemotherapieresistentie, namelijk de belangrijkste metastatische kenmerken van mCSC's. In Hoofdstuk 4 tonen we de impact aan van differentiële expressie van ESRP1 en andere RBPs, neerwaartse effectoren van de EMT-masterregulator ZEB1, op AS-patronen. Deze regulatie strekt zich uit tot een breed scala aan doelen, waaronder CD44 en NUMB, resulterend in de generatie van specifieke isoformen die geassocieerd zijn met invasie- en metastasekenmerken. Verdere validatiestudies, zowel functioneel als klinisch, bevestigen de rollen van deze nieuw geïdentificeerde RBPs, evenals de CD44s- en NUMB2/4-splicevarianten, bij het bevorderen van lokale invasie, verre metastase, en correlatie met verminderde overlevingspercentages bij darmkankerpatiënten. In Hoofdstuk 5 onderzochten we de invloed van EMT-gerelateerde genisoformen op de regulering van EMP bij ovariumkanker met hooggradige serieuze histologie. Bij het vergelijken van onze bevindingen met die in darmkanker, observeerden we een verschillend spectrum van neerwaartse doelen in quasi-mesenchymale ovariumkankercellen, waarschijnlijk weerspiegelend verschillen in metastatische mechanismen tussen deze maligniteiten - hematogeen bij darmkanker en transcoelomisch bij ovariumkanker. Bovendien identificeerden we opwaartse RBPs die differentieel tot expressie komen tussen epitheliale en quasi-mesenchymale subpopulaties van ovariumkankercellen, die de regulering van EMT-gerelateerde isoformen besturen. De upregulatie van RBM24 en de downregulatie van ESRP1 kwamen naar voren als belangrijke regulerende factoren van EMT in ovariumkankercellen. Om de functionele en klinische relevantie van onze bevindingen te valideren, richtten we ons op het TPM1, dat een eiwit codeert dat cruciaal is voor het

definiëren van de functionele eigenschappen van individuele actinefilamenten in contractiele cellen, onder de neerwaartse AS-doelen die specifiek zijn voor ovariumkanker. We ontdekten dat de Tpm1.8/9-isoformen selectief tot expressie komen in patiënt-afgeleide ascites en invasie bevorderen door activering van EMT- en Wnt-signaleringspaden, samen met een breed scala aan ontstekingsgerelateerde paden. Bovendien verleent de expressie van Tpm1.8/9 resistentie tegen op taxaan- en platina-gebaseerde chemotherapie.

In Hoofdstuk 6 bespreken we de uitdagingen die gepaard gaan met het benutten van alternatieve splicing-doelen in klinische therapeutische benaderingen voor kankerbehandeling. De identificatie van specifieke verbindingen gericht tegen leden van de AS-machinerie opent nieuwe wegen voor de ontwikkeling van therapeutische hulpmiddelen en antikankertargets. De integratie van kankersequencinggegevens met experimentele validatie belooft verborgen basismechanismen en farmaceutische doelen te ontrafelen, en legt zo de basis voor innovatieve kankerbehandelingen in de toekomst.

APPENDICES

Acknowledgements
List of Publications
PhD Portfolio
Curriculum Vitae

Acknowledgements

It has been an incredible adventure to do groundbreaking research and pursue a PhD degree in the Laboratory of Stem Cell and Cancer Research at Erasmus Medical Center. I would like to express my deepest gratitude to those who have guided, supported, and accompanied me throughout this journey.

First and foremost, I extend my heartfelt thanks to my supervisor, Prof. Riccardo Fodde. Riccardo, your guidance and support have been invaluable, not only in imparting a profound understanding of cancer biology but also in providing inspiration, motivation, and a sense of responsibility. Under your mentorship, I have honed my critical thinking skills and gained independence in my research endeavors. Using the metaphor of a "beacon," I acknowledge your role in leading me into the exciting and profound world of research and showing me the method to navigate its paths. I appreciate your realistic approach, not only in our research discussions but also in pointing out areas for improvement during my PhD studies. Working with you has been an unforgettable experience, and I have grown significantly in both scientific aspects and problem-solving.

Furthermore, I extend my deep appreciation to Prof. Peter Gunning and Prof. Edna Hardeman for the collaborative effort on the tropomyosin project. Our interactions during Zoom meetings have been enlightening, and I have thoroughly enjoyed and learned a great deal from our collaboration. Your motivation and extensive knowledge of tropomyosin and biology have consistently inspired me. The challenges I faced during the tropomyosin project were alleviated by the encouragement and support received through our collaboration.

Additionally, I express my sincere gratitude to Prof. Véronique Orian-Rousseau and Prof. Leonel Munoz Sagredo for your contributions to the CD44 project. Collaborating with both of you on the discussions of alternative splicing and CD44 has been a valuable experience, marking my first project in this setting. To be honest, I was still quite inexperienced in networking and scientific collaboration as a fresh PhD student. Thank you and Riccardo for generously guiding me on how to navigate collaborations. I have learned a great deal from our interactions, and the experience has been truly enjoyable, something I deeply appreciate.

I also want to express my gratitude to my co-supervisor, Dr. Ron Smits, for your meticulous and hands-on guidance in wet lab experiments. Ron, your careful instruction has been instrumental in my experimental work, and I appreciate your help to ensuring its success.

The work presented in this thesis has been critically assessed and approved by an outstanding committee to whom I am more than grateful: Prof. Guido Jenster, Prof. Ruud Delwel, and Prof. Peter Verrijzer.

Dr. Mark Schmitt, I still vividly recall our first online interview before embarking on my PhD journey here. Your rigorous approach to science left a lasting impression on me. I extend my sincere gratitude for your kind guidance when I initially joined the lab, unfamiliar with everything.

Rosalie Joosten, your assistance with my wet lab experiments has been invaluable. I couldn't have completed the experimental phase of my project without your support. My troubleshooting skills have improved significantly thanks to the lessons learned from you. I am grateful for your consistent presence whenever I encountered challenges in the lab. Additionally, I appreciate your organization and management for those amazing activities, which aided in my acclimation to Dutch culture and living environment.

To Mathijs Verhagen and Roberto Stabile, I am grateful for the opportunity to work alongside you as colleagues during my PhD. Our scientific discussions and project brainstorming sessions have been inspirational, and I have genuinely enjoyed our collaboration. Thank you so much for all the encouragement and support. I often reflect on how fortunate I am to have you as my colleagues in this journey toward obtaining my PhD. The time spent working with you guys will be one of my most unforgettable experiences. You are amazing!

Sun Wenjie, I express my heartfelt thanks for your outstanding analysis and contributions to our project. Your sense of responsibility and kindness extends not only to the project but also personally. Collaborating with you has been a truly enriching experience.

Miriam Teeuwssen, I am grateful for the guidance and collaboration we shared as colleagues and "supervisor." Working with you has been a valuable learning experience, and I appreciate your guidance and assistance, especially when I was not yet familiar with the wet experiment materials and protocols at the beginning of my PhD study.

Andrea Sacchetti, thank you for your contributions to the FACS sorting and engaging discussions on CD44. Though we didn't talk much, your reliability is evident, and I knew I could count on you never to miss any planned sorting. I feel fortunate to have had your technical support on my journey to obtaining my PhD degree.

Special thanks to all the former and present member of Fodde lab: Martine de Herdt, Berdine van der Steen, Emily Middendorp Guerra, Francesco Tucci, Mario Román Cabezas, Wiktoria Szymańska, Emilie de Vet, Danielle Seinstra, Alem Gusinac, Ting Chen, Yaser Atlasi, Daan te Rietmole. It has been a great experience to meet and work with all of you. I am grateful to have spent these wonderful study times with you, fascinating people in the Fodde lab.

Dr. Patricia Ewing-Graham, reviewing the histological aspects of cancer IHC slides with you has been a valuable experience. I have gained a wealth of professional knowledge on ovarian cancer pathology through our collaboration, making it an enjoyable and enlightening experience.

I extend my gratitude to my co-authors, Maurice Jansen, Ingrid Boere, Nicole Bryce, Jun Zeng and Herbert R. Treutlein for your hard work and significant contributions to the studies in this thesis. I look forward to future collaborations.

I would like to thank all other colleagues from JNI for their support, providing cell lines, and sharing

reagents when needed. Alex Nigg, Thank you for all those fancy microscope guide and teaching. Thierry van den Bosch, thank you for your helpful imformation and tips on Basescope and IHC staining. Corine M. Beaufort, thank you for sharing those amazing ovarian cancer cell lines with us.

This PhD journey would not have been possible without the financial support of the China Scholarship Council, to whom I am sincerely grateful.

I am fortunate to have a supportive family and friends who have been with me throughout the PhD journey.

Firstly, to my friends in the Netherlands, Mou Yuchan, Luo jie, Wang Wenxing, Shi Xinglan, You yueyue, Lu Tianqi and Rik Thank you for accompanying me on this adventure. Your support and camaraderie have been invaluable as we adapted to the new environment. You've made me never feel lonely, even though I am far away from China and couldn't come back during Covid. I am grateful to have met you guys and to have shared so many great times with you during my PhD.

Chen Qing, when I arrived in the Netherlands, you were the only friend I knew. Thank you for your assistance and the information you shared before and after I started my PhD. I appreciate your help when I applied for CSC, and I'm grateful to have had your support throughout my entire academic journey from Master to PhD.

Diya Gao, thank you for helping me adapt to the environment and for sharing important information to ease my transition into studying and living here. I was fortunate to have you as my landlord when I first came to the Netherlands, and your assistance helped me adapt to the new environment. To my kind roommates, Wang Yifan and Lv Qianting , thank you for your warm company when I first arrived alone.

Zhao Liyun, although we only met each other twice in the Netherlands, it was so sweet to have your visit and spend enjoyable moments together.

I would like to express my gratitude to all the wonderful people I met during my PhD, Zhang Peifen, Zhang Xuxi, Wang Xiangjun, Xu Yanning, Huang Wenqiu, Huang shaozhuo, Cao Wanlu, Du Zhaoyu, Xu Dawei , Qi Hongchao, Shi yuwei, Wu Yahong, Tang Jinchi, Cui Rui. You are wonderful, and it has been a pleasure to meet all of you here.

I would also like to express my gratitude to my friends in China, Qi Tuoya, Li Yankun, Zhang yujie, Wang Wanying, Tao Daijun, Yuan Mengke even though we haven't seen each other often in recent years. Your encouragement and kind words have been a significant source of motivation for me. The times spent apart have made me realize that you are the individuals who contribute to my personal growth and improvement. You hold a special place as my most important friends.

A special thanks to my Master supervisor, Prof. Gao MingQing, for providing mentorship and teaching fundamental skills and knowledge, leading me into the realm of scientific research. The techniques and theoretical knowledge I acquired from you have greatly assisted me in my further

studies.

I express my gratitude to my family members for their support during my PhD study. Your caring and encouragement from afar have been motivating. My parents, you have always been my strongest support. Mom, thank you for your selfless dedication, love, and support. I wouldn't have come this far without you. Dad, thank you for teaching me independence and resilience. These qualities I learned from you are the most precious treasures on my journey to obtaining a PhD.

Lastly, I extend my heartfelt thanks to the most important person, Ruizhi Deng. Thank you for supporting me in academia, life, mental health, humanity, and sophistication. I am so fortunate to have your company in my life. You made me a better person.

While I am happy to have obtained my PhD degree, the time and memories with all the mentioned individuals are the most precious treasures in my life. Thanks to the luck to have all of your support and meet with all of you in my life.

List of Publications

T. Xu, M. Verhagen, M. Teeuwssen, W. Sun, R. Joosten, A. Sacchetti, P. Ewing-Graham, E. Hardeman, P. Gunning, R. Fodde et al. Tropomyosin1 isoforms underlie epithelial to mesenchymal plasticity, metastatic dissemination, and resistance to chemotherapy in high-grade serous ovarian cancer. **Cell Death Differ**, doi:10.1038/s41418-024-01267-9 (2024).

T. Xu, M. Verhagen, R. Joosten, W. Sun, A. Sacchetti, L Muñoz-Sagredo, V Orian-Rousseau, R. Fodde. Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behaviour in colon cancer. **Elife** 11, doi:10.7554/eLife.82006 (2022).

Sacchetti, A., M. Teeuwssen, M. Verhagen, R. Joosten, **T. Xu**, R. Stabile, B. van der Steen, M. M. Watson, A. Gusinac, W. K. Kim, I. Ubink, H. J. G. van de Werken, A. Fumagalli, M. Paauwe, J. van Rheenen, O. Sansom, O. Kranenburg, R. Fodde, Phenotypic plasticity underlies local invasion and distant metastasis in colon cancer. **Elife** 10, doi:10.7554/eLife.61461 (2021).

M, Verhagen. **T, Xu**. R, Stabile. Sacchetti, A. R, Joosten. F, A. Tucci. R, Fodde. et al. The SW480 cell line as a model of resident and migrating colon cancer stem cells. **iScience** (2024) (Under Review)

PhD portfolio

Name PhD student:	Tong Xu
Erasmus MC Department:	Pathology
Research school:	Molecular Medicine
Supervisor:	Prof. Riccardo Fodde

PhD training

Courses	Year
Basic And Translational Oncology	2018
Scientific Writing	2019
Basic Course on R	2021

Conferences	Year
KWF Tumor Cell Biology Meeting	2018
5th International Cancer Symposium of the Cancer Research Center of Lyon (Poster presentation)	2022
KWF Tumor Cell Biology Meeting (Poster presentation)	2022

Seminars	Year
Monthly JNI Oncology Lectures	2018-2023
JNI Scientific Meetings	2018-2023
Annual Molecular Medicine Day	2018-2023
Pathology lab meeting (PALM)	2018-2023

Curriculum Vitae

Tong Xu was born in ShenYang, China. She obtained a Bachelor's Degree in Life Science and a Master's Degree in Animal Science in Xi'an. During her Master's thesis, she worked at the National Key Transgenic Laboratory with Prof. Gao MingQing. After completing her Master's Degree, Tong joined Prof. Dr. Riccardo Fodde's lab to pursue her PhD degree. During her PhD, her work primarily focused on the study of epigenetic regulation in epithelial mesenchymal plasticity in colon and ovarian cancer. In 2024, Tong joined the lab of Stephan Huveneers (Amsterdam UMC) to pursue a post-doctoral project on endothelial signaling in vascular malformations.