UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Biologia

CORSO DI DOTTORATO DI RICERCA IN: Bioscienze e Biotecnologie

CURRICOLO: Biologia Cellulare

CICLO XXIX

BTBD7, a gene identified with a transposon based forward genetic screening, is important for colorectal cancer metastasis.

Tesi redatta con il contributo finanziario del Centro di Riferimento Oncologico di Aviano

Coordinatore: Ch.mo Prof. Paolo Bernardi

Supervisore: Ch.mo Prof. Paolo Bernardi

Co-Supervisore: Dr.ssa Milena Sabrina Nicoloso

Dottoranda : Eleonora Grisard

Ai miei nonni Liliana, Giovanna, Federico e Ugo, che sarebbero stati orgogliosi di me.

Table of Contents

RIASSUNTO

Il tumore del colon-retto è il secondo tumore più letale a causa della diffusione metastatica della lesione primaria. Un'ipotesi attuale è che la metastasi sia basata sulla transizione epitelio-mesenchimale (in inglese EMT, epithelial to mesenchymal transition), un processo biologico in cui le cellule epiteliali perdono gradualmente i loro caratteri epiteliali per convertirsi a un programma di tipo mesenchimale. I saggi *in vitro* che selezionano cellule EMT sono fondamentali per poter fare screening genetici che selezionano cellule che si sono convertite al programma EMT. Ad esempio, il saggio *in vitro* di anoikis si basa sulla crescita delle cellule in condizioni di mancanza di attacco alla matrice ed è stato usato per selezionare cellule con fenotipo più aggressivo; tuttavia alcuni tipi di cellule tumorali sono capaci di resistere a queste condizioni di crescita rafforzando i contatti cellula-cellula, un comportamento in netto contrasto con il fenotipo EMT. Un saggio *in vitro* sviluppato nel nostro laboratorio, denominato forced Single Cell Suspension assay (fSCS) si è rivelato più stringente rispetto al saggio *in vitro* di anoikis e seleziona cellule andate incontro alla EMT. La parte non codificante del genoma, nonostante abbia un ruolo fondamentale nella regolazione della EMT e nella metastasi (come avviene ad esempio per i miRNA), risulta molto meno studiata rispetto alla controparte codificante. I saggi genetici *in vitro* basati sull'uso di trasposoni interrogano il genoma in modo più randomico rispetto ad altri tipi di saggi (ad esempio quelli basati sull'uso di retrovirus). Per avvalerci di un saggio *in vitro* che consenta uno screening molto efficiente dei geni che regolano la EMT, abbiamo combinato il saggio di fSCS con uno screening genetico *in vitro* basato sull'uso del trasposone Sleeping Beauty in cellule di cancro colo-rettale HCT116. Abbiamo identificato un clone cellulare, TN4_20, che mostra le seguenti caratteristiche: maggior resistenza al saggio di fSCS, morfologia mesenchimale, espressione di marcatori della EMT (ad esempio Slug ↑, Twist ↑, Vimentin ↑, E-cadherin ↓, Has-2 ↑), e abilità di generare un maggior numero di colonie satelliti nel saggio di evasione in matrigel. Inoltre, in un esperimento pilota condotto *in vivo*, le cellule TN4_20, iniettate in topi mediante iniezione intra-cecale, formano metastasi a distanza. Dopo essere risaliti alle posizioni genomiche delle inserzioni del trasposone nel DNA genomico delle TN4_20, ci siamo focalizzati sull'inserzione localizzata nella 3'UTR (3' regione non tradotta) del gene BTBD7. Abbiamo scelto di

studiare questa inserzione perché BTBD7 è un noto regolatore di EMT e metastasi e perché questa inserzione del trasposone è localizzata nel sito bersaglio predetto del miR-23b, un miRNA con note funzioni anti-metastatiche. Abbiamo ipotizzato e dimostrato che il miR-23b bersaglia il gene BTBD7 e i nostri dati suggeriscono che l'inserzione del trasposone nella 3'UTR di BTBD7 interferisce con questa interazione. Inoltre, abbiamo dimostrato che l'interazione tra il miR-23b e BTBD7 è importante per la resistenza all' fSCS. I nostri risultati inoltre dimostrano che il silenziamento di BTBD7 interferisce con la resistenza all' fSCS sia in cellule HCT116 parentali che TN4_20, e che la over-espressione di un costrutto ectopico eGFP-Btbd7 in cellule HCT116 parentali conferisce resistenza all' fSCS e l'abilità di generare un maggior numero di colonie satelliti nel saggio di evasione in matrigel. Inoltre, l'over-espressione di eGFP-Btbd7 induce una riduzione dei livelli di trascritto e di proteina di E-caderina, e un aumento dei livelli di Vimentina, entrambi marcatori di EMT. In più, la over-espressione di eGFP-Btbd7 aumenta i livelli di trascritto e di proteina del fattore di trascrizione Zeb-1.

In una versione estesa del nostro saggio, attraverso l'esecuzione di round multipli di fSCS sia in cellule HCT116 parentali che in cellule HCT116 trasposte con il trasposone Piggybac (PB), abbiamo ottenuto gruppi, e non singoli cloni, di cellule resistenti all' fSCS. Abbiamo osservato che cellule che sopravvivono a ogni round di fSCS generano un maggior numero di colonie sopravviventi, le quali acquisiscono una morfologia più mesenchimale. Inoltre, le colonie di cellule sopravvissute all' fSCS mostrano ridotti livelli di espressione di E-caderina, aumentati livelli di espressione di Vimentina, e un aumentato numero di cellule con ridotta espressione di EpCAM, suggerendo che round multipli di fSCS determinano un arricchimento di cellule con tratti di EMT e staminalità. Inoltre, abbiamo osservato che cellule resistenti all'fSCS mostrano una maggiore resistenza al trattamento con 5-fluororacile (5-FU) e un aumentato potenziale metastatico *in vivo*. Infine, round ripetuti di fSCS arricchiscono due famiglie di miRNA, cioè miR-30 e miR-302, che sono già stati descritte avere un ruolo nella EMT e nella metastasi, e che potrebbero quindi regolare anche la resistenza all' fSCS.

SUMMARY

Colorectal cancer (CRC) is the second most lethal cancer because of the metastatic spread of the primary tumor. A current hypothesis is that metastasis relies on epithelial to mesenchymal transition (EMT), which is a biological process in which epithelial cells gradually loose epithelial features to switch to a mesenchymal program. *In vitro* assays that select EMT cells are fundamental to perform *in vitro* genetic screens that select cells switched to EMT program. For instance, *in vitro* anoikis assay consists in growing cells in low adherence conditions (loss of cell-matrix contacts) and has been used to select more aggressive tumor cells; however, some tumor cells survive to loss of cell-matrix contacts by strengthening cell-cell contacts, which is counteracting for cells that undergo to EMT. An assay that was developed in our lab and was named Forced Single Cell Suspension Assay (fSCS) is more stringent compared to *in vitro* anoikis and selects for cells that undergo EMT. The non-coding part of the genome, despite having a fundamental role in regulating EMT and metastasis (e.g miRNAs) is less studied respect to the protein-coding counterpart. Transposon based screens interrogate the genome more randomly than other screens (e.g retroviral based screens). To perform an *in vitro* assay that permits the high-throughput screening of EMT genes, we combined fSCS with an *in vitro* Sleeping Beauty (SB) transposon (TN) based screen in HCT116 CRC cells. We identified a cell clone, TN4_20, that shows the following features: greatest fSCS resistance, mesenchymal morphology, expression of EMT markers (e.g. Slug ↑, Twist ↑, Vimentin ↑, E-cadherin ↓, Has-2 ↑), and the ability to generate more satellite colonies in matrigel evasion assay. Moreover, in a pilot *in vivo* experiment, TN4_20 intra-caecal injected mice developed distant metastases compared to control. We retrieved the genomic position of TN insertions from TN4_20 genomic DNA, and we focused on the TN insertion located within the 3' UTR of BTBD7. We chose to study this insertion because BTBD7 is a known EMT and metastasis regulator and because, interestingly, this TN insertion locates within the predicted target site of miR-23b, a known anti-metastatic miRNA. We hypothesized and demonstrated that miR-23b targets BTBD7 gene, and our data suggest that TN insertion impairs miR-23b/BTBD7 interaction. Moreover, we demonstrated that the interaction between miR-23b and BTBD7 is important for fSCS resistance. We found that Btbd7 silencing impairs fSCS survival in HCT116 parental and in TN4_20, and that

the overexpression of ectopic eGFP-Btbd7 in HCT116 parental confers fSCS resistance and the ability to generate more satellite colonies in matrigel evasion assay. Moreover, the overexpression of ectopic eGFP-Btbd7 induces the down-regulation of E-cadherin at the mRNA and protein level, and the up-regulation of Vimentin, both markers of EMT. In addition, Btbd7 overexpression up-regulates Zeb-1 transcription factor mRNA and protein levels.

In an extended version of our TN- fSCS based screen, by performing sequential rounds of fSCS in both HCT116 Parental and Piggybac (PB) TN-cells, we obtained pools of fSCS resistant cells, instead of single clones. We observed that cells that survived to each round of fSCS generated more surviving colonies and acquired a greater scattered/mesenchymal morphology. Moreover, surviving colonies after fSCS showed decreased E-cadherin expression, increased Vimentin expression and increased number of cells with EpCAM low (dim), suggesting that multiple rounds of fSCS enriches for EMT/stem-cell traits. In addition, we observed that enriched fSCS resistant cells showed increased resistance to 5-fluoro-uracil (5FU) treatment and increased *in vivo* metastatic potential. Finally, repeated rounds of fSCS enrich for two families of miRNAs, miR-30 and miR-302 that were already shown to regulate EMT and metastasis and that may potentially regulate fSCS resistance.

List of abbreviations

3D: three dimensional

AJC: apical-junctional complex

CIMP: cPG island methylator phenotype

CIN: Chromosomal instability

CRC: Colorectal cancer

CTC: circulating tumor cells

CSC: cancer stem cell

DTC: disseminating tumor cells

ECM: extracellular matrix

EGF: epidermal growth factor

EGF-R: epidermal growth factor receptor

EMT: Epitelial mesenchymal transition

FAP: familial adenomatous polyposis

fSCS: forced Single Cell Suspension Assay

HNNPC: hereditary non-polyposis CRC

HGF: Hepatocyte growth factor

IGF: insulin growth factor

IGF-R: insulin growth factor receptor

IR: inverted repeats

lincRNA: long intergenic non-coding RNA

LINE: Long Interspersed Nuclear Elements

lncRNA: long non-coding RNA

LTR: long terminal repeats

MET: Mesenchymal epithelial transition

miRNA: microRNA

MMR: Mismatch repair

MSI: Microsatellite instability

MSS: Microsatellite stable

PCG: protein-coding gene

PDGF: Platelet-derived growth factor

piRNA: piwi-interacting RNA

polyA site: poly adenylation site

rRNA: ribosomal RNA

SA site: splicing acceptor site

SD site: splicing donor site

ShRNA: short hairpin RNA

SINE: Short Interspersed Nuclear Elements

siRNA: small interfering RNA

snoRNA: small nuclear RNA

TISC: tumor initiating stem cell

TGFβ: transforming growth factor β

TN: transposon or TE: transposable element

TNM: Tumor Node Metastasis

tRNA: transfer RNA

VEGF: Vascular Endothelial Growth Factor

1. INTRODUCTION

1.1 COLORECTAL CANCER

1.1.1 Incidence of Colorectal Cancer

In the Western world, colorectal cancer (CRC) is a relevant health problem, as it represents the third most frequent tumor in industrialized countries. CRC incidence among males results to be higher between the ages of 50 and 65 and ranks third after lung and prostate tumors. Instead, among females CRC incidence ranks second after breast cancer. Approximately one million new cases are registered worldwide per year, with 550,000 men and 470,000 women affected [1][2]. Considerably, in 2008 nearly 600,000 deaths related to CRC have been registered by the World Health Organization, thus reaffirming the importance of CRC as a public health problem [3].

In the Tumor-Node-Metastasis (TNM) staging system, "T" is used to describe how deeply the primary tumor has grown into the wall of the intestine, "N" is used to describe if cancer has spread to regional lymph nodes, and "M" is used to describe cancer that has spread to other organs of the body, most commonly liver and lungs. The overall stage of the tumor is determined by combining T, N and M and goes from 0 to IV, with higher numbers indicating more advanced cancers and providing the basis for therapeutic decisions.

The diagnosis of CRC usually occurs late, at advanced stages (stage IV), when the tumor has spread to distant sites. In the last years, a great effort has been accomplished for the improvement of adjuvant chemotherapies, surgical techniques, radiotherapy techniques as well as campaigns for primary and secondary prevention. However, it has been estimated that half of the patients recently diagnosed with CRC will progress to metastatic cancer and that the average survival will be 5 years for 50-60% of these patients [4].

Colorectal cancers are divided into three categories based on occurrence: sporadic (60-80%), familial (10-30%) and hereditary CRC (5-6%) (Figure 1). *Sporadic CRC* accounts for the vast majority of CRC and appears in individuals who do not carry germline predisposing mutations [5]. *Familial CRC* are less common and also for this category no associated gene has been identified. It has been estimated that the risk of

developing CRC is two to three times higher than in the general population if family members of primary consanguinity have suffered from sporadic colon cancer. In these cases, the risk to develop familial CRC in individuals who are genetically predisposed is determined by environmental factors [6]. In *Hereditary CRC* the inherited cancer risk of some CRC cancer–prone families is due to known genetic mutations. The two best defined hereditary forms of CRC can be distinguished by the presence or not of adenomatous polyps: 1) *Familial adenomatous polyposis* (FAP), which is caused by pathogenic variants of the APC (Adenomatous Polyposis Coli) gene; FAP is characterized by the presence of multiple polyps. If these polyps are not treated with preventive surgery, the risk to become malignant at the average age of 40 is higher [7]. 2) *Hereditary nonpolyposis CRC* (HNPCC) or *Lynch Syndrome* is the hereditary CRC variant not associated with polyposis. *Lynch Syndrome* in addition to *CRC* has also high risk for developing other non-digestive cancer and is caused by germline pathogenic variants in DNA mismatch repair (MMR) genes (e.g. hMLH1, hMSH2, hMSH6 and hPMS2, see next section) [8].

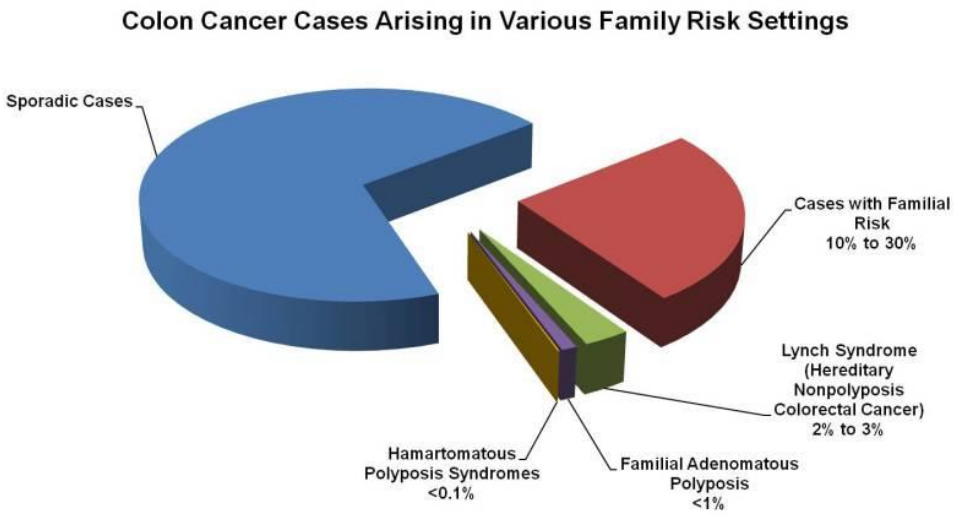


Figure 1. The figure represents the fractions of CRC cases arising in various family risk settings. Reprinted from *Gastroenterology*, Vol. 119, No. 3, Randall W. Burt, Colon Cancer Screening, Pages 837-853, Copyright (2000), with permission from Elsevier

1.1.2 Natural history of CRC

Colorectal tumors are predominantly epithelial-derived tumors and range from benign growths to invasive cancers (i.e. from adenomas to adenocarcinomas). Colorectal lesions can be classified into three groups: a) *Non-neoplastic/adenomatous polyps* (hyperplastic, juvenile, hamartomatous, inflammatory, and lymphoid polyps), which are not considered as cancer precursors lesions; b) *Neoplastic/adenomatous polyps and adenomas*, which are benign glandular tumors that naturally evolve into malignant tumors and c) *cancers* [9].

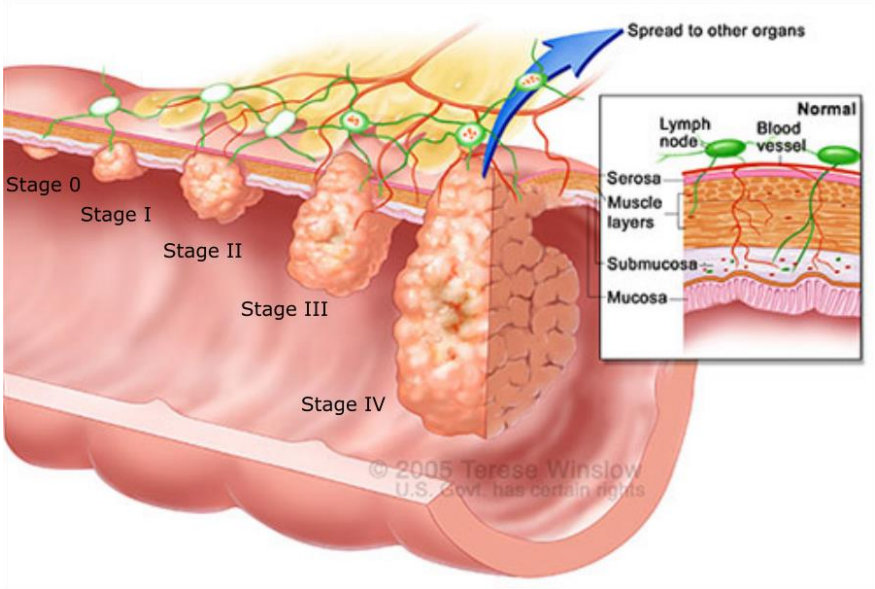


Figure 2: Schematic view of adenoma-carcinoma sequence in CRC. ©2005 Terese Winslow

Neoplastic polyps, in particular adenomatous polyps, can predispose to the development of CRC [10]. Adenomas have variable degrees of dysplasia, that range from low-grade (G1, well differentiated), intermediated grade (G2, moderately differentiated) and high-grade (G3, poorly differentiated, G4 undifferentiated). Adenomatous polyps with high grade dysplasia (G3, G4), high content of villous tissue within the polyp, and greater than 1 cm in diameter correlate with an increased risk of malignant progression [9][11].

The adenoma-carcinoma sequence is a stepwise process in which a benign polyp or adenoma develops in an advanced adenoma and then progresses to invasive cancer. CRC stages I and II are confined within the wall of the colon; CRC stage III invade submucosa, muscle layers and disseminate to regional lymph nodes; finally, CRC stage IV has metastasized to one or more distant sites [12][13][14] (Figure 2). Surgical excision solves stage I and II tumors and, with the use of adjuvant chemotherapy, more than 73% of cases of stage III disease [12][13][15]. Unfortunately, stage IV of CRC is usually incurable, although recent advances in chemotherapy have improved survival [12][13].

1.1.3 Molecular bases of CRC

From a molecular perspective, the adenoma-carcinoma sequence in CRCs is characterized by the accumulation of genetic abnormalities [16][17]. The loss of genomic stability facilitates the acquisition of multiple tumor-associated mutations, thereby contributing to the development of CRC [18]. The two major types of genomic instability that have been recognized as alternative mechanisms of CRC carcinogenesis are: *a) Chromosomal Instability (CIN)* and *b) Microsatellite instability (MSI)* [17][18][19]. Another type of instability harbored by CRCs is epigenomic instability. Epigenetic instability can be classified as global hypomethylation or as the *CpG island methylator phenotype (CIMP)*, defined as methylation at three or more specific marker loci [18].

CIN: Being present in approximately 70-80% of cases, CIN is the most common type of genomic instability in CRCs [19]. CIN is characterized by the presence of multiple structural or numerical chromosome alterations in tumor cells, which can determine the physical loss of a wild-type copy of a tumor-suppressor gene, such as APC, TP53 and SMAD4 [17]. A tight association between the clinicopathological changes occurring in adenoma-carcinoma sequence and the physical loss of tumor suppressor genes has been described in CIN CRCs [20]. The initial step of tumorigenesis leading to adenoma formation is the physical loss of APC gene and alteration of the oncogenic Wnt signaling pathway [17][19][20]. APC protein is a component of the oncoprotein β-catenin degradation complex, which degrades β-catenin and inhibits its nuclear localization. Loss of β-catenin degradation and its translocation to the nucleus are sufficient to trigger Wnt pathway; in fact, once in the nucleus, β-catenin binds to its nuclear partners (members of the T-cell factor–lymphocyte enhancer factor family) and becomes a transcription factor which regulates genes involved in cellular activation. In the absence

of functional APC, *β*-catenin gets rid of degradation and Wnt signaling is inappropriately and constitutively activated [17]. According to recent studies, microRNAs (miRNAs) represent an alternative mechanism for APC regulation in CRC: miR-135a and miR-135b decrease translation of APC *in vitro* and were also found upregulated in CRC patient samples with low APC levels [21]. Progression along the adenoma-carcinoma sequence occurs upon accumulation of additional mutations: intermediate adenomas present alterations in small GTPases (e.g. activation of KRAS); late adenomas present the additional loss of chromosome 18q carrying SMAD4 locus, a tumor suppressor that acts downstream of transforming growth factor-β (TGFβ). Additionally, SMAD4 can be regulated also at the post transcriptional level by miR-224, which is found increased in advanced CRCs [22]; finally, invasive cancers present mutations in TP53 [19] [20](Figure 3*, upper part).*

MSI: Microsatellites are tracts of tandemly repeated DNA motifs (2-5 bp in length), typically repeated 5-50 times and widely distributed throughout the genome. MSI tumors are characterized by abnormal microsatellite length, which can be measured at specific microsatellite marker sites. This is due to deficiency of mismatch repair (MMR) system that normally corrects errors occurring in DNA replication. If cells fail to recognize mismatches and repair them, the result is microsatellite slippage and elevated spontaneous mutation rate (mutator phenotype) [23]. In humans, seven MMR proteins work in sequential steps to repair DNA mismatches: hMLH1, hMLH3 (human mut L homologues), hMSH2, hMSH3, hMSH6 (human mut S homologues), hPMS1 and hPMS2 (human post-mitotic segregation).

Approximately 15% of CRCs develop through the MSI pathway. The defect can be inherited, which is the case in Lynch syndrome, or acquired, as in sporadic MSI tumors. In patients with Lynch syndrome, the MSI phenotype is caused by germline mutations in mismatch repair genes (mostly hMLH1 and hMSH2), whereas sporadic MSI CRCs are typically caused by epigenetic silencing of the hMLH1 gene [24][25]. MiRNAs appear to be important in the regulation of MSI in CRC: the overexpression of miR-155 significantly downregulates MMR core proteins (hMSH1, hMSH6 and hMLH1) by inducing an MSI in CRC; moreover, an inverse correlation between miR-155 and the expression of hMLH1 or hMSH2 was found in human CRC samples [26].

In MSI CRC, which can be sporadic or associated with Lynch Syndrome, loss of MMR genes occurs early in adenoma-carcinoma sequence. In addition, affection of Wnt signaling, with mutations of CTNNB1 encoding *β-catenin*, also occurs at early steps (see Figure 3, lower part) [19][24]. In intermediate adenomas, BRAF mutations occur in place of KRAS mutations; in this context, positive selection of tumor cells with mutated microsatellites in hMSH3 and hMSH6 genes has been shown to increase MSI status. Mutations that affect microsatellites in TGFβ receptor 2 (*TGFBR2*) [27] and insulin-like growth factor 2 receptor (*IGF2R*) produce an additional positive selection in late adenomas. Finally, mutations that affects microsatellites in *BAX*142, provide a mechanism of progression to carcinoma which is not dependent on p53 (see Figure 3, lower part) [19].

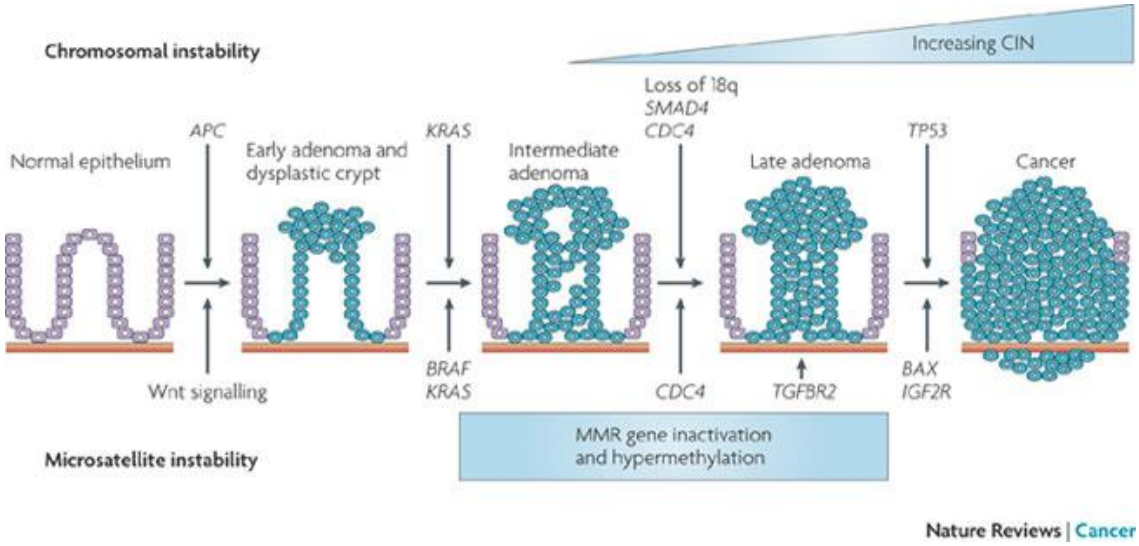


Figure 3: Upper part: adenoma-carcinoma model for CIN pathway; Lower part: adenoma-carcinoma sequence model for MSI pathway. Adapted from: *Walther et al, Genetic prognostic and predictive markers Nat Rev Cancer 2011* © 2011 Macmillan Publishers

1.2 METASTATIC PROGRESSION OF COLORECTAL CANCER

1.2.1 The metastatic cascade

Metastasis is a multistep process which causes more than 90% of cancer-related deaths. As mentioned above, metastasis is a relevant problem in CRC, since half of recently diagnosed CRC patients will progress to metastasis, with an average survival of five years for 50-60%[4]. In an anatomic perspective, a tumor that is still confined within the walls of intestine and has not invaded the basement membrane is a carcinoma in situ (CIS) or a carcinoma stage 0 (see figure 2). To become invasive and metastatic, a normal colon epithelial cell must undergo dramatic phenotypic and biochemical changes. This changes primarily involve motility and cell shape mechanisms, but also growth factor signaling and gene expression [28].

In CRC metastasis, similarly to other tumors, invasive cancer cells leave the primary tumor site, travel via the circulation to a distant tissue site, and form a secondary tumor [29]. This process can be divided in five steps, which are collectively termed the *metastatic cascade* (Figure 4):

Detachment and invasion: This step involves the detachment of tumor cells from the primary tumor and the invasion of the adjacent tissues.

The detachment of tumor cells from the epithelium requires the loss of contact with their extracellular matrix (ECM). In normal cells, the loss of cell-matrix contacts induces a programmed cell death named anoikis. By preventing epithelial cells to survive in the absence of ECM contacts, anoikis constitutes an important barrier to metastatic dissemination. Therefore, cancer cells to be able to metastasize must overcome anoikis. Moreover, it has been observed that tumor cells detaching from their site of origin undergo epithelial-mesenchymal transition (EMT), a process involving the loss of epithelial traits and the acquisition of a more mesenchymal and motile phenotype [30] [31]. This indicates a role for EMT in the initiation of the metastatic cascade.

To invade the underlying stroma, tumor cells must disrupt the basement membrane. To do so, they overexpress proteolytic enzymes, such as matrix metalloproteinases (MMPs), which promote the digestion of the basement membrane major components, i.e. laminin and collagen IV [28][32].

Intravasation: During intravasation, cancer cells penetrate the endothelial cell junctions and enter in the blood and/or lymphatic vessels. In this process, tumor cells must undergo dramatic shape changes and cytoskeletal remodeling, thus a critical role is played by cytoplasm elasticity and viscosity [32]. Recent findings indicate that the relative stiffness and deformation capability of cell organelles plays an important role in intravasation processes. The interphase nucleus is the largest organelle in the cell and is ten times stiffer than the cytoplasm, thus it is not surprising that its deformability is a rate-limiting step in the migration of tumor cells within the endothelium [33][34][35]. Moreover, it has been reported that oncogene-mediated alterations in cellular actomyosin contractility and small GTPase Rho activity can regulate tumor cell deformability and reduce cytoskeletal stiffness, enabling metastatic cells to move through tight spaces, such as between endothelial cells [36].

Circulation: In this step, tumor cells, referred to as circulating tumor cells (CTCs), travel via the blood or lymphatic circulation. Several physical and mechanical parameters, like blood flow, diameter of blood vessels and the interplay between intercellular adhesion and sheer force of the blood flow have the ability to influence the trajectory of CTCs [32]. Moreover, during their transit CTCs must withstand toxic conditions (*e.g.* high concentration of oxygen), immunological stress and collision with host cells (*e.g.* blood cells and endothelial cells) [28][32]. These constraints can affect CTCs survival and their ability to establish metastatic foci: the main part of CTCs die, indicating that metastatic spread is a very inefficient process and that a selection for particularly resistant and aggressive tumor cell takes place.

Extravasation: In this process CTCs arrest, leave the circulation and penetrate the endothelium. The arrest of CTCs occurs in two manners depending on vessel diameter: *a) physical occlusion* and *b) adhesion*. Physical occlusion occurs when the diameter of vessel is less than the diameter of the cell. With a diameter greater than 10μm, for example, epithelial tumors typically arrest in small vessels and capillaries with a diameter of <10 μm [32]. When they are in larger vessels, CTCs arrest by adhering to the vessel through the formation of specific bonds. P-L and E-selectin expressed on the surface of endothelial cells are particularly important for the arrest of CTCs and metastasis [37]; for example, CRC cells directly contact endothelial selectins through their surface CD44 variant isoforms, CEA (carcinoembryionic antigen) and PODXL (podocalyxin)[38] [39] [40]. Strikingly, metastasis of a CRC xenograft is markedly

inhibited in P-selectin knockout mice [37]. Moreover, the arrest of CRC cells also depends on the interaction between integrins expressed on tumor surface and their counter-receptors, ICAM1 and VCAM1 (Intercellular and Vascular adhesion molecule 1, respectively) on endothelial cells [41].

Colonization: CTCs that successfully penetrate the endothelium and then infiltrate the target organ are called disseminating tumor cells (DTCs). To colonize the target organ and form a secondary tumor, DTCs must survive to immune surveillance and adapt to a foreign microenvironment [28]. The location of metastatic site does not seem to be casual: CRC preferentially metastatizes to the liver and lungs [42]. The preferential localization of metastases at certain organs is explained only in part by the vessel anatomy and by the physics of blood flow. This was observed for the first time in 1889 by Steven Paget [43] who formulated the "seed and soil" theory. According to this theory, the so called "premetastatic niche" at a distant site — the soil — is be more conducive for DTCs — the seeds — than others, promoting the development of metastases. In addition, recent evidences indicate that growth factors secreted by the primary tumor itself, such as VEGF (Vascular Endothelial Growth Factor) and inflammatory chemokines, "initiate" pre-metastatic niches for the engraftment of tumor cells [44][45].

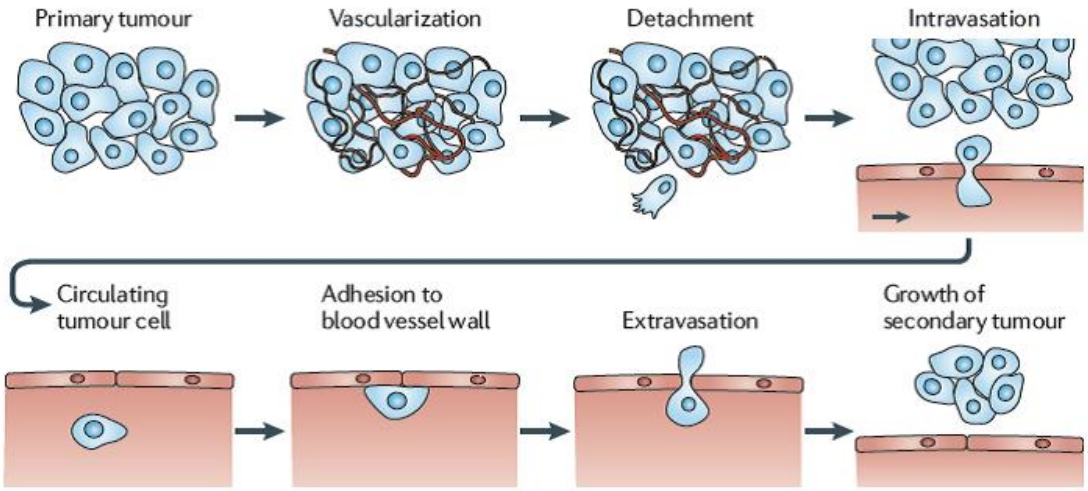


Figure 4: Schematic view of the metastatic cascade. Adapted from: Wirtz et al. The physics of cancer: the role of physical interactions and mechanical forces in metastasis. Nat Rev Cancer.

1.2.2 Anoikis: a barrier to metastasis

As mentioned above, the initial step of metastatic dissemination in carcinomas is the detachment of tumor cells from the epithelium (see Figure 4). To do this, cancer cells must lose their contacts with the neighboring cells and the ECM. The loss of cell-cell contacts and contacts with the ECM in a normal cell triggers a type of programmed cell death that has been termed "*anoikis*", a Greek word that means "*homelessness*" [30]. Thus, anoikis is critical in maintaining tissue organization and in preventing epithelial cells from colonizing elsewhere. Normal epithelial and endothelial cells are very sensitive to anoikis, and only upon oncogenic transformation they become anoikis-insensitive; whereas, normal fibroblast can acquire the capacity to survive the loss of ECM contacts by pro-survival signaling triggered by serum growth factors. [46].

The major cell death pathway involved in anoikis is apoptosis (Figure 5). Apoptosis is the most diffused type of programmed cell death in vertebrates, and is divisible into two pathways: extrinsic (death receptor-dependent) and intrinsic (mitochondria-dependent)[47]. Both pathways are initiated upon loss of cell ECM contact and ultimately cause activation of cysteine proteases, called caspases, which enzymatically destroy the cell in a manner that, unlike necrosis, avoids inflammation. The extrinsic pathway is triggered by the binding of membrane-bound death receptors such as Fas receptor, TNFR1 (tumor necrosis factor receptor 1), and TRAIL R1/R2 (TNF-related apoptosis-inducing ligand receptor-1 or-2), with soluble ligands such as Fas ligand, TNF- or TRAIL, respectively. When Fas is stimulated, the Fas-associated death domain (FADD) adapter molecule directly triggers procaspase-8 oligomerization and autocleavage into active caspase-8 monomers [48]. The intrinsic pathway, relies on the mitochondria and is triggered by cell stressors (e.g oxidative stress, DNA damage, UV radiation and viral infection). The Bcl-2 family of proteins intracellularly mediates this pathway. The common effect of these intrinsic factors is the increase of mitochondrial membrane permeability and the release of cytochrome *c*, that in turn triggers caspase-9 activation and cell execution via apoptosis [49].

Integrins are the main receptors and for cell-cell adhesions and for cell adhesion to ECM. Upon engagement with components of the ECM, integrins trigger signal transduction pathways that regulate proliferation, cell-survival/apoptosis, shape, polarity and differentiation [50]. In the primary tumor, integrins contact the ECM and

lead to: 1) local organization of cytoskeleton and 2) the formation of focal adhesion complexes. Focal adhesion kinase (FAK) is one of the major components of focal adhesion complexes, and its activation leads to cell survival by inducing pro-proliferative pathways, such as the mitogen-activated protein kinase (MAPK) or phosphatidylinositol-3-kinase (PI3K)-Akt. The detachment from the ECM causes the disruption of the actin cytoskeleton, which in turn results in: 1) deactivation of FAK, that leads to the attenuation of pro-survival pathways, and 2) release of pro-apoptotic Bcl2 family members, such as Bim (Bcl2-interacting mediator of cell death), Bmf (Bcl2-modifying factor) Bak (Bcl2-homologous antagonist/killer) and Baxx (Bcl2-like protein 4) that, freed from prior inhibition, induce mitochondrial mediated apoptosis [51][52]. Alternatively, integrin disruption leads to increased expression of Fas ligand receptors, thus activating extrinsic apoptosis via FADD-mediated caspase activation [51][53].

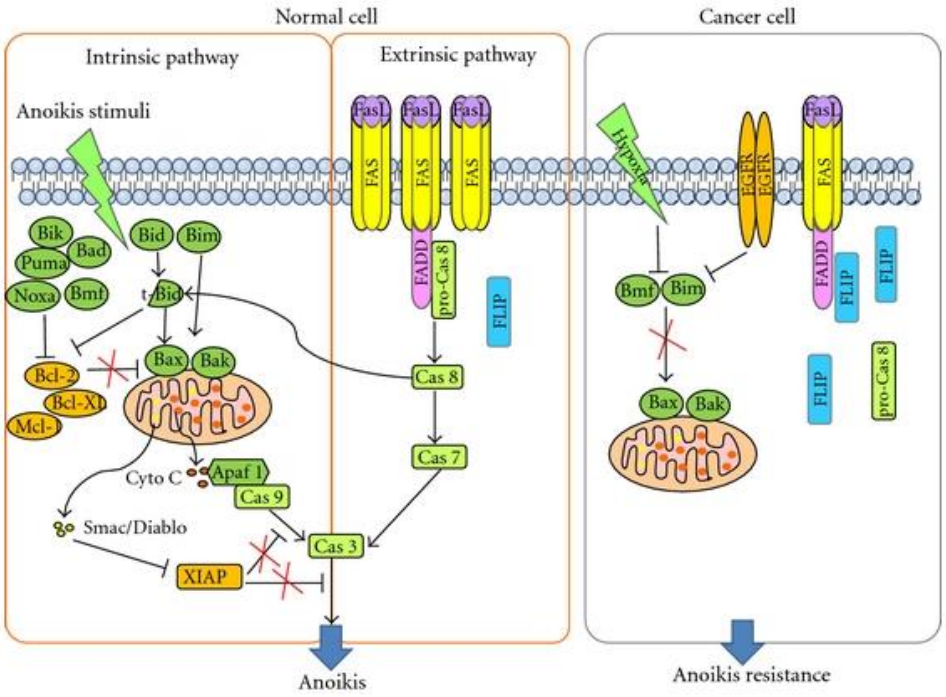


Figure 5: Left: anoikis regulation in normal cells by apoptosis; Right: mechanisms leading to anoikis resistance in cancer cells.

Autophagy is a process engaged by cells in response to nutrient deprivation, starvation, and other stress stimuli (including loss of ECM contacts) and can occur during anoikis. Autophagy is context-dependent and can lead to both cell death or cell survival in a context dependent manner: due to this ambiguity, the exact role of autophagy in anoikis regulation appears to be very complex and is still debated. During autophagy, cytoplasmic constituents that have been targeted for degradation are isolated from the rest of the cell and included into the autophagosome, a double-membrane vesicle. Then, the fusion of the autophagosome with a lysosome allows degradation and recycling of the contents [54]. Beclin 1 is one of the key molecules that initiates, by aggregating with PI3K and Vps34, the autophagic cascade. This includes the aggregation of autophagy proteins (Atg) first into an intermediate complex (Atg12-Atg5-Atg16) and then, upon activation and ligation of LC3/Atg8 (microtubule-associated-protein 1 light-chain3) to the phospholipid phosphatidylethanolamine (PE) via Atg3, Atg4 and Atg7, into a higher complex (Atg12-Atg5-Atg16-LC3-PE) that is ultimately responsible for the autophagosome formation. Another key regulator of autophagy is mTOR (mammalian Target Of Rapamycin) which is a potent inhibitor of Atg12-Atg5-Atg16 complex formation [55].

On one hand, autophagy has been described to act as promoter of cell-death in anoikis: indeed, during formation of three-dimensional (3D) mammary epithelial acini, central cells die as a result of the loss of contact with ECM; interestingly, numerous autophagic vacuoles are present in these dying cells [56]. On the other hand, numerous evidences suggest a requirement for autophagy during anoikis resistance and metastasis. For example, autophagy facilitates anoikis resistance of hepatocellular carcinoma (HCC) cells and increases HCC invasion and metastasis [57]. Moreover, in an organotypic model of invasion through a collagen matrix, the knockdown of the autophagic protein Atg12 decreased the invasion capability of glioma cells [58]. Consistent with his opposite finding, autophagy and apoptosis have been demonstrated to cross-talk, either positively or negatively during anoikis; for example, upon loss of ECM and cell-cell contacts apoptosis can impair autophagy through the pro-apoptotic protein Bax. In fact, activated Bax can promote the caspase dependent cleavage of the autophagic protein Beclin 1 [59]. In this case apoptosis wins over autophagy leading to anoikis cell death.

1.2.3 *In vitro* anoikis resistance

The dependence of cell growth on the anchorage to ECM was studied for the first time *in vitro* in 1978 by transferring epithelial cells from standard, adhesive cell culture dishes to culture dishes coated with different concentrations of polyHEMA (poly-2-hydroxyethylmethacrylate)[60]. Standard cell culture dishes are characterized by a hydrophilic surface that supports cell attachment and spreading; whereas, polyHEMA is a polymer that forms a hydrogel in water [61] and due to its uniformly nonionic nature it effectively prevents matrix deposition when used to coat cell culture dishes. However, it was only in 1994 that Frisch and Francis thoroughly described the process of cell apoptosis induced by loss of ECM anchorage using polyHEMA coated plates and for the first time used the term "*in vitro* anoikis" [30]. Furthermore, they found that epithelial cells transformed with H-RAS or V-SRC became resistant to *in vitro* anoikis. More recently, Douma and collegues used *in vitro* anoiks to perform a cDNA-based gain of function genetic screen in order to identify the genes able to overcome anoikis in rat intestinal epithelial cells (RIE). In this case, *in vitro* anoikis was performed with commercially available ultra-low attachment plates, and also with this approach RAS transformation induced resistance to *in vitro* anoikis. In particular with this functional screen the authors identified Trkb, a neurotrophic tyrosine kinase receptor, as a novel regulator of anoikis resistance with metastatic potential in vivo [62]. Consistently, human epithelial ovarian cancer displaying enhanced *in vitro* anoikis resistance were associated with greater aggressiveness and tumorigenicity *in vivo* [63].

1.2.4 Molecular bases of anoikis resistance

It has been widely assessed that, for tumor cells to became metastatic, they must develop resistance to the pro-apoptotic stimuli that ultimately lead to anoikis. Resistance to anoikis can be conferred by different mechanisms, mediated by RTKs (receptor tyrosine kinase), by MAPK and other kinases and by small GTPases (such as Ras) [53][64].

RTKs: RTKs are frequently aberrantly activated in cancer and may contribute to anoikis resistance. For example, IGFR1 (insulin growth factor receptor 1) is involved in anoikis resistance, as its inhibition reduces the ability of cancer cells to grow in an anchorage-independent state *in vitro* and impairs xenograft tumor formation *in vivo* [65][66]. A reciprocal regulation of EGFR (epidermal growth factor receptor) and ErbB2

(a member of EGFR family) has been found to regulate anoikis resistance. When non-malignant MCF-10A breast cells are forced to grow in suspension, the lack of ECM contacts causes a decrease of EGFR levels and an increase of pro-apoptotic Bim levels [67]. Interestingly, overexpression of ErbB2 or SRC can compensate deficiencies of EGFR, of α5 integrin and evade Bim-mediated anoikis via ERK signaling [68] [69].

MAPK and other kinases: Several members of the MAPK family can participate in the regulation of anoikis. In addition to ERK, that was already mentioned to regulate this process, another anoikis regulating MAPK is TGFβ-activated kinase 1 (TAK1; also known as MAP3K7). In particular, the activation of a non-canonical Wnt signaling pathway promotes evasion from anoikis, anchorage sphere formation and metastatic potential of circulating pancreatic cancer cells, and TAK1 inhibition represses these effects [70]. Integrin receptors can propagate downstream signaling through another kinase, ILK (integrin-linked kinase); interestingly, ILK can block anoikis in mammary epithelial cells via the activation of PI3K/Akt signaling pathway [71].

Small GTPases: As mentioned above, the transformation of non-tumorigenic epithelial cells with H-Ras confers resistance to anoikis [30]. In particular, the mechanism by which Ras confers protection during detachment from the ECM involves PI3K–AKT signaling pathway [72]. Moreover it was shown that Ras activation can block cytochrome *c* release and caspase 8 activation [73], important for the intrinsic apoptotic pathways leading to anoikis. Also the small GTPases RhoG, involved in cytoskeleton remodeling and migration, confers anoikis resistance via PI3K-AKT activation small GTPases RhoG, involved in cytoskeleton remodeling and migration, confers anoikis resistance via PI3K-AKT activation [74].

Additional mechanisms that confer resistance to anoikis are the overexpression of β catenin [75] and increased cytosolic FLIP, the primary endogenous inhibitor of extrinsic apoptosis [76].

1.2.5 Epithelial-mesenchymal transition

Epithelial-mesenchymal transition (EMT) is a de-differentiation process in which a polarized epithelial cell gradually loses its epithelial features (polygonal/columnar shape, apico-basolateral polarity, strong cell-cell adhesion, expression of: e-Cadherin, Claudin, Occludin, Zonula Occludens 1, Cytocheratins) and switches to a mesenchymal

phenotype (spindle shaped, anterior/posterior polarization, focal cell-cell contacts, expression of: N-cadherin, Vimentin, αSMA, Desmin, Fibronectin). Elevated resistance to apoptosis, enhanced migratory capacity, invasiveness, and increased production of ECM components are some of features that an epithelial cell acquires during EMT [31]. Diverse molecular processes are required for initiation and completion of the EMT program. In fact, epithelial cells undergoing EMT change their repertoire of specific cell-surface proteins, re-organize their cytoskeleton, produce proteases to degrade the surrounding ECM and finally convert themselves in mesenchymal cells that can migrate away from the epithelium of origin [31][77]. Moreover, changes in the expression of specific miRNAs can occur during EMT. To demonstrate the passage of a cell through EMT, some of the molecules that participate to this process are used as EMT biomarkers [31].

EMT can be classified into three different biological types, with very distinct functional consequences[31]: a) *Type 1*, comprising EMT associated with embryo formation and tissue organization; b) *Type 2,* comprising EMT associated with wound healing, tissue regeneration and organ fibrosis and, c) *Type 3*, occurring in cells that undergo genetic and epigenetic changes that favor tumorigenic transformation [78].

Type 1: EMT in embryogenesis and organ development

Type 1 EMT, unlike other types of EMT, is not associated with fibrosis, inflammation or induction of an invasive phenotype, and occurs at various stages of embryogenesis, including the earliest ones [31]. For example, Type 1 EMT occurs during gastrulation of the fertilized egg, when the three germ layers that will give rise to different adult body tissues are generated [79]. The canonical Wnt signaling is involved in gastrulation-associated EMT [80] and acts via proteins of the TGFβ superfamily, e.g. Nodal and Vg1 [81][82]. Transcription factors such as Snail, Eomes (Eomesdermin) and Mesps orchestrate gastrulation associated-EMT; for example, Snail transcription factor induces EMT by downregulating E-cadherin expression (see below) [83].

Type 2: EMT in inflammation and organ fibrosis

Type 2 EMT occurs in repair-associated programs. In fact, to intervene in case of inflammatory injury and trauma, the organism use this type of EMT to generate fibroblasts, which act in cooperation with other cell-types. In the case of organ fibrosis, in which the inflammatory stimulus is continuous, type 2 EMT eventually leads to organ

destruction [31]. Epithelial cells that undergo EMT associated with chronic inflammation are characterized by the expression of Desmin, Colloidin, Vimentin, FSP1 (fibroblast specific protein 1) and Collagen I [31] [84]. During inflammatory injury in mouse kidney, resident fibroblasts and macrophages release growth factors, such as TGFβ, EGF, FGF and PDGF (platelet derived growth factor), which induce epithelial cells to cooperate with inflammatory cells and to undermine basement membrane integrity via degradation of laminin and collagen IV; ultimately, under the influence of chemoattractants and growth factor gradients, delaminated epithelial cell migrate in the interstitial area [85].

1.2.6 Type 3: EMT in tumor invasion and metastasis

The EMT program has emerged as a critical mechanism in the induction of the early steps of metastatic cascade. In fact, both *in vitro* and *in vivo* experimental evidences prove that carcinoma cells can switch to mesenchymal phenotype by expressing relative mesenchymal markers, such as the above mentioned Vimentin, FSP1 and desmin [31][86]. Remarkably, these traits are frequently observed in the cells at the invasive front of primary tumors, that more likely will proceed to the subsequent steps of the metastatic cascade (see Figure 4) [87]. The molecular mechanisms that contribute the achievement of the EMT program can be grouped as follows:

1) **Signaling:** Tumor-associated stroma cells release a series of growth factors, which can act as EMT-inducing heterotypic signals for tumor cells [31][87]. In particular, EGF, HGF PDGF and TGFβ activate several intracellular signaling pathways inducing the expression of EMT-transcription factors, notably Snail, Slug (or Snail 2), Zeb 1 (zing finger E-box binding homeobox 1), Twist and others [87]. High expression of EMT-transcription factors correlated with poorer prognosis [88], as in the case of Snail in prostate and breast cancer [89], Slug in gastric cancer [90] and Zeb-1 in CRC [91]. TGFβ function is context dependent and has a dual role, acting as promoter of malignant transformation or as inducer of cell death depending on cell type, growth conditions, and the presence of other growth factors [92]. This ambiguity manifests also in the regulation of EMT [31]: TGFβ treatment induces EMT, invasion and metastasis in diverse murine and human systems, acting as promoter of malignancy [93]; in particular, TGFβ-mediated EMT can be triggered by different pathways, involving Smad proteins and ALK (anaplastic lymphoid kinase)-5 receptor [94], or

alternatively PGE2 (prostaglandin 2) and overexpression of Cox2 (cyclooxygenase 2)[31][95][96]. However, the pro-EMT and pro-oncogenic function of TGFβ is in contrast with its capability to promote apoptosis. A complex regulatory loop has been described in EMT induction by TGF-β: if on one hand TGF-β induces Snail, that in turn promotes EMT, at the same time it induces apoptosis signaling. Interestingly, however, cancer cells use Snail to counteract TGFβ-induced apoptosis, since Snail, in addition to inducing EMT, upregulates Akt and anti-apoptotic Bcl-xL [95][97].

2) **Cellular Junctions:** Cell-surface molecules, important for tissue homeostasis and cell polarity, dramatically decrease during EMT [95]. In epithelial cells, the apico-basal polarity is assured by the Apical-Junctional-Complex (AJC), that comprises tight junctions and adherent junctions [98]. Transmembrane proteins from both tight and adherent junctions are down-regulated in EMT. E-cadherin is a transmembrane glycoprotein that belongs to the family of Ca^{2+} dependent adhesion proteins, and is the main protein of the adherent junctions [98][99]. Through its extracellular domain, E-cadherin establishes homophilic interactions with adjacent E-cadherins expressed in neighboring cells; whereas, through its citoplasmic domain E-cadherin contacts intracellular proteins, like catenins (α, β, p120), which link E-cadherin to actin cytoscheleton [95][99]. The decrease of E-Cadherin, as well as the increase of N-Cadherin are considered the paradigm of the induction of the EMT program [31][86][95]. Indeed, Snail, Zeb-1, Zeb-2 and Twist EMT-transcription factors are all well-known repressors of E-cadherin transcription. Notably, Snail, Slug, Zeb-1 and Zeb-2 have been found to interact with specific E-boxes at the proximal promoter of E-cadherin during EMT [100].

3) **miRNAs:** Snail, Zeb and Twist transcription factors can epigenetically silence E-cadherin promoter by recruiting chromatin modifying enzymes, such as histone deacetylases HDAC1 and HDAC2 [95][101] [102]. MiRNAs extensively participate in the regulation of EMT circuitry, and regulate E-cadherin expression levels either directly or indirectly: for example, miR-9 directly targets E-cadherin mRNA, whereas members of the miR-200 and miR-205 families (miR-200a, miR-200b, miR-200c, miR-141, and miR-429) or members of the miR-34 family (miR-34a, miR-34b and

miR-34c) repress E-cadherin transcriptional repressors Zeb-1/Zeb-2 and Snail, respectively [103].

E-cadherin downregulation has been widely associated with tumor metastasis and invasiveness, both *in vitro* and *in vivo*. A paradigmatic evidence of this is a study conducted in various human carcinoma cell lines with epithelioid or fibroblastoid phenotype. In this study, the authors reported that carcinoma cell lines characterized by epithelioid phenotype and expressing E-cadherin, are noninvasive; conversely, carcinoma cell lines with a fibroblastoid phenotype, and which do not express E-cadherin, are invasive. Remarkably, the transfection of E-cadherin cDNA reversed invasiveness of these carcinoma cell lines [104]. Similar results have been observed *in vivo*; in fact, downregulation of E-cadherin levels *in vivo* is associated with the dedifferentiation, progression, and metastasis of CRC [105]. Finally, Onder and collegues found that E-cadherin silencing alone is able to induce EMT, invasiveness and the expression of a variety of EMT genes, such as transcription factors Twist and Zeb-1, suggesting that E-cadherin loss could be not only a consequence of the EMT program, but itself can act as an initiator of the EMT and metastasis processes [106].

1.2.7 EMT and stemness

Stem cells are undifferentiated cells which are characterized by two properties: 1) self-renewal capability, which consists in the ability of stem cells to perpetuate themselves and 2) potency, which is the ability of stem cells to generate differentiated cells [107]. Several lines of evidence indicate that tumor growth and proliferation are sustained by rare populations of stem cells, referred to as cancer stem cells (CSCs); CSCs, like normal stem cells, are capable of self-renewal and differentiation [107]. Moreover, in certain cases CSCs contribute to chemotherapy resistance and metastatic relapse [108][109][110][111].

EMT has been associated to stem-like phenotype and chemoresistance in different studies. For example, chemoresistance of ovarian cancer cells to cisplatin and paclitaxel induces EMT and stem-like phenotype via the up-regulation of FOXOM1 [112]. Moreover, the induction of EMT drives the acquisition of stem-like and tumorigenic properties in mammary epithelial cells [113]. EMT transcription factors have been found to induce CSCs properties: Zeb-1 not only induces EMT of pancreatic and CRC cells, but also stimulates sphere forming capacity and expression of stem cell transcription

factors, such as Sox2 and Bmi1. Zeb-1 mechanism of action involves repression of miR-200 family members, in particular stemness repressing miR-203 [114]. Moreover, TGF-β promotes EMT and stemness traits in liver cancer cells via the up-regulation of Snail-1, which induces stem cell transcription factor Nanog and tumor initating stem-like cells (TISCs) properties [115]. Yet, it must be mentioned that EMT not necessarily associates with stem-like phenotype. In fact, in mammary epithelial cells only transient activation of Twist initiates stem-like traits; whereas, permanent Twist-1 activation does not preserve stem-like traits [117].

1.2.8 Establishment of distant metastasis

A disseminating tumor cell (DTC), that has survived through all the metastatic steps and reached a distant organ, must proliferate to form a detectable metastatic lesion. It has been proposed that distant metastases in carcinomas can be divided in two groups based on their differentiation state: *plasticity type I* metastases are differentiated malignant lesions and retain the hierarchical organization of the corresponding primary tumor; these metastases are hypothesized to rely on cell plasticity, EMT and mesenchymal-epithelial transition (MET). Conversely, *genetic type II* metastases are undifferentiated lesions, and are hypothesized to rely on genetic alterations occurring at the primary site or along the metastatic cascade, rather than on phenotypic plasticity [118].

Plasticity Type I: It is important to remark that the term “transition” reflects in part the reversibility of the process; in fact, the existence of the inverse process of EMT, (MET), in which mesenchymal cells come back to the epithelial state, reveals the intrinsic plasticity of EMT. In plasticity type I metastases, EMT programs are turned on transiently during the initial steps of the metastatic cascade but then are switched back off during the establishment of the macroscopic distant lesion with induction of MET, with the recovery of epithelial traits in cells that previously underwent EMT. MET normally occurs during embryogenesis and organ development, as plasticity of epithelia is required to switch from epithelial to mesenchymal state and viceversa until completion of tissue development [119]. Colorectal adenocarcinoma is a good example of plasticity type I of metastasis: a de-differentiation similar to an EMT, with loss of E-cadherin expression has been observed in cells at the invasive front of primary tumor; whereas, corresponding liver metastases showed a reversal of this de-differentiated

phenotype with re-expression of E-cadherin. These clinical-pathological observations in primary and metastatic CRC samples enforce the hypothesis of a requirement for transient rounds of EMT-MET to enable metastasis formation [118]. Furthermore, analysis of gene expression in invasive cells revealed that EMT traits combine with a stem cell-like phenotype [118][120].

Several evidences suggest that cellular plasticity in type I metastases are governed by reciprocal feedback loops between EMT-transcription factors and miRNAs. One example is the regulatory loop existing between Zeb transcription family members (Zeb-1 and Zeb-2) and the miR-200 family [121][122][123]. In fact, Zeb transcription factors inhibit miR-200 family members, and miR-200 family members, in turn, repress the translation of Zeb transcription factor. Thus, Zeb-1 can use two different mechanisms to induce EMT: 1) by inhibiting the expression of epithelial molecules, such as E-cadherin and 2) by inhibiting its own repression by miR-200 family. In turn, miR-200 induces MET and differentiation in at least two different manners: 1) by inhibiting its own repression by Zeb-1 and 2) by inhibiting stem cell factors such as SUZ12 [124] and BMI1 [125][118]. Not only Zeb-1, but also Snail transcription factor is engaged in a similar feedback loop with a miRNA family, in this case miR-34 family. Notably, Snail transcription factor is a promoter of EMT, stemness and drug resistance; whereas, miR-34 family induce MET, differentiation and drug sensitivity and the two sides of the loop antagonize reciprocally [126].

The necessity for differentiation (and MET) in metastasis formation relies on the fact that in tumor cells the capability to proliferate (ideally represented by MET) and the capability to disseminate (ideally represented by EMT) are mutually exclusive [118]. In support of this, there is the observation that invasive cells that underwent EMT in CRC show low levels of Ki67 proliferative marker; conversely, Ki67 levels increase in the differentiated regions from both primary tumors and corresponding metastases [127], suggesting that MET is necessary to overcome the proliferative arrest induced by EMT and stemness.

What dictates the transition from non-proliferative/EMT/stem-like state to proliferative/MET/differentiated state? One hypothesis is that tumor microenvironment is involved in the regulation of this switch [118]. DTCs (disseminating tumor cells), i.e. tumor cells that infiltrate the target organ after extravasation but do not outgrow, are

reversibly in growth arrest at the G0/G1 phase of cell cycle [128]. In support of the hypothesis that tumor microenvironment regulates the proliferative state of DTCs, lung resident cells induce dormancy and growth arrest of infiltrated breast cancer through BMP (Bone Morphogenetic Protein) production [129]. Conversely, when dormant breast disseminated cancer cells get in touch with a fibrotic microenvironment rich in collagen I, they reactivate and colonize the lungs via β1 integrin activation of Src and FAK [130].

Genetic Type II: genetic type II metastases, different from plasticity type I, are undifferentiated lesions and do not seem to depend on cell plasticity to form. Two possible scenarios explain why these metastases do not need to re-differentiate [118]:

i) Intrinsic subtype metastases: In this scenario, primary tumors are intrinsically undifferentiated due to specific mutations that may have occurred very early during tumor development. This is the case of triple negative breast cancers, that are highly aggressive cancers, displaying a high proliferation rate and an early and high metastatic rate [131]. Importantly, primary tumors have intrinsic EMT and stem-like phenotype, with high levels of Slug, Zeb-1, CD133 and BMI1, low levels of miR-200, and expression of cancer stem cell markers $CD44^{high}CD24^{low}$ [118][132][133]. In this case stemness traits and EMT traits are associated with intrinsic limitless replicative potential.

ii) ***Induced subtype metastases****:* In this scenario, metastasis can be undifferentiated due to the selection for aggressive cells in the primary tumor, typically after repeated cycles of chemotherapy. Why do repeated cycles of chemotherapy lead to undifferentiated metastases? Uncontrolled proliferation and stem cell features, as self-renewal capability, are distinctive traits of undifferentiated metastases. These features likely derive from genetic alterations developed in the primary tumor; it has been hypothesized that repeated cycles of chemotherapy may select for these alterations by permitting the maintenance of EMT-stem-like traits [118]. In support of this, breast cancer patients with tumors that are initially differentiated, develop undifferentiated-EMT-stem-like tumors after three months treatment with hormone and chemotherapy [134].

1.3 INTERROGATING THE HUMAN GENOME

1.3.1 Molecular complexity of human genome

Human genome consists of 22 paired chromosomes, termed autosomes, and of two sex chromosomes: X chromosome (two in females, one in male), and in males only, one Y chromosome. Chromosomes are contained within the cell nucleus, but each mitochondrion contains a small circular DNA, termed mitochondrial DNA, which is part of the human genome. With a total length of 3 billion base pairs, haploid human genome is the largest genome to be extensively sequenced [135]. The final number of human protein-coding genes was estimated to fall between 40,000 and 100,000, but the completion of the Human Genome Project revised dramatically this view, indicating that protein-coding genes were between 20,000 and 25,000 [136]. Most recently, this number has been further revised to 20,500 protein-coding genes by using evolutionary comparisons [137], indicating that protein coding sequences account for only 1.5% of the genome.

Thus the genome sequencing project revealed an unexpected problem in our understanding of the molecular bases of developmental complexity in higher organisms: the fruitfly Drosophila Melanogaster and the nematode Caenorhabditis Elegans appear to have only half as many protein-coding genes (12,000-14,000) as humans [135]. The increased repertoire of protein isoforms expressed in higher organisms and humans is accounted by alternative splicing [138], but the other striking feature of the evolution of higher organisms seems to be the increase of the non-coding part of the genome, which in humans accounts for more than 98% of the whole genome [135] and includes introns, regulatory DNA sequences (promoters and enhancers), transposable elements such as LINEs (Long Interspersed Nuclear Elements), SINEs (Short Interspersed Nuclear Elements), and non-coding RNA genes. Moreover, 3' untranslated regions (3'UTR) of mRNAs are considered non-coding RNAs, and their length increases with the complexity of the genome [139].

1.3.2 Non-coding RNA

Non-coding RNA, i.e. RNA that is not translated into proteins, represents approximately 98% of the transcriptional output of the human genome. Two different regions of the genome can originate non-coding RNAs (ncRNAs): 1) the introns of protein-coding genes and 2) the exons and introns of non-protein coding genes [140], including many that are antisense to or overlapping protein coding genes.

There are two different types of ncRNA: infrastructural ncRNAs and regulatory ncRNAs. Infrastructural ncRNAs have a housekeeping role in RNA processing and translation, and include transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear (spliceosomal) RNAs and snoRNAs (small nucleolar RNAs) [141]. Regulatory ncRNAs regulate the expression of genes in different organisms and can be further divided in small ncRNAs that include miRNAs and piRNAs (piwi-interacting RNAs), and long RNAs, or lncRNAs [141].

MiRNAs: miRNAs (microRNAs) are a class of small ncRNAs which have a role in RNA silencing and post-transcriptional regulation of gene expression. The majority of worm and human miRNAs genes are isolated as independent genes, but other miRNAs are clustered in the genome and can be controlled by the same promoter and are often related to each other. Finally, a portion of miRNAs originate from introns of protein coding genes [142] [143].

MiRNA biogenesis is a strictly regulated process which is divided in different steps. First, RNA polymerase II transcribes a precursor, termed primary miRNA (pri-miRNA), that contains hundred to thousands of nucleotides, and is 5' capped and polyA [144]; interestingly, the pri-miRNA can contain a cluster of miRNA or a single miRNA in a folded hairpin stem structure. The nuclear "microprocessor complex", formed by the RNAse III enzyme Drosha and by the double stranded RNA binding domain protein DGCR8 (DiGeorge Critical Region 8)/Pasha, cleaves the pri-miRNA into a 70 nucleotides precursor known as pre-miRNA [145]. At this point, a Ran-GTP dependent nuclear export factor, Exportin 5, transports pre-miRNA from the nucleus to the cytoplasm, where it is processed through a series of cuts by the RNAse III Dicer in association with TRBP (HIV-transactivating response RNA-binding protein) and PACT (Protein kinase dsRNA dependent activator). The result of this processing, in humans, is the production of 18-25 nucleotides miRNA:miRNA duplex [146]. The RNA duplex is unwound to form

two single strands under the mediation of Ago-2 (Argonaut 2) protein, a key factor in the assembly and function of miR-RISC (miRNA-RNA induced silencing complexes). MiRNAs downregulate target gene expression by two mechanisms dictated by the level of complementarity between miRNA and mRNA. More in detail, it is suggested that specificity in choosing target transcript relies on complementarity between the 3'UTR (3' Untranslated Region) of target mRNA and the nucleotide sequence form position 2 to 8 at the 5' end of miRNA, referred to as "seed sequence" [143]; yet, it has been reported that many functional miRNA binding sites are located outside the 3'UTR of mRNAs [147]. MiRNA with close to perfect complementarity, bind to target mRNAs and induce cleavage and degradation of the transcript by decapping and deadenylation of the mRNA [148]. Most commonly, miRNAs mediate repression of translation without affecting mRNA expression levels, due to imperfect sequence complementarity between miRNA and mRNA [143]. Alternatively, miRNAs can direct targeted mRNAs to storage structures known as processing bodies (p-bodies); interestingly, this process may lead to the reversal of miR-induced mRNA repression through the release from the storage bodies and the re-entry into polysomes under stress conditions [149].

MiRNAs regulate an enormous number of cell functions, ranging from tissue and cell development to disease and cancer. Some mammalian miRNAs, for example, exhibit developmentally regulated expression profiles in a variety of tissues, such as brain [150] and many others [151] or appear to be specifically expressed in embryonic development [152]. As described above, miR-200 and miR34 families are engaged in a feedback regulation loop with Zeb and Snail transcription factors in the control of EMT and cell differentiation [121][122][123][126]. All known relationships between miRNAs and human diseases are listed in the publicly available database *miR2disease* [153]. For example, reduced expression of miR-1 and miR-133 have been found implicated in cardiac disease, such as cardiac hypertrophy [154].

Evidence is accumulating that miRNAs play an important role in cancer progression and metastasis [155]. In breast cancer miRNA signatures can distinguish primary tumors with metastases from metastasis-free tumors [156]. MiRNAs regulate CRC in different manners: as already mentioned, members of the miR-135 family affect Wnt signaling by downregulating APC protein in the earliest steps of transformation of CIN CRC [21]; whereas, the overexpression of miR-155 significantly downregulates MMR genes by inducing an MSI in CRC [26]. CRC invasion, intravasation and metastasis is

stimulated by miR-21 through the regulation of tumor suppressor Pdcd4 [157]; interestingly, high expression of miR-21 in 5-FU (5-fluorouracil) treated CRC patients correlate with poor therapeutic outcome [158].

piRNAs: piRNAs (piwi-interacting RNAs) are another class of small non-coding RNAs expressed in animal cells. Compared to miRNAs, piRNAs are longer (26-31 nucleotides) and show lack of conservation and increased complexity [159]. The majority of piRNAs map to multiple sites in the genome, and appear to be organized in clusters ranging from several to hundreds of kilobases; 95% of these clusters are located in pericentromeric and telomeric heterochromatin regions [160]. piRNAs function has been well described in animal gonads, where their role is to repress transposons for the maintenance of genome integrity. In fact, when piRNAs are lost, the consequent transposons de-repression has dramatic consequences, such as genome damage and impairment of fertility [161]. The biogenesis of piRNAs is not completely understood, but involves the formation of single stranded RNA precursors transcribed from piRNA clusters [160]. Different from miRNAs, piRNAs do not depend on RNAse III but are processed by the endonuclease Zucchini [162]. After maturation, piRNA interacts with a subclass of Argonaute proteins called Piwi proteins and form piRNAs silencing complexes (piRISCs) to target RNA transcripts by means of RNA-RNA base pairings [162]. In this section it is worth mentioning siRNAs (small interfering RNAs), that are double stranded RNAs which, similarly to miRNAs and piRNAs, act within the RNA interference pathway by using the same processing machineries [163]. Even if siRNAs have been first discovered in plants, the presence of endogenous siRNAs has been reported also in human genome [163][164][165]; typically, siRNAs are introduced artificially in cells to perform knock-down experiments. A comparison of miRNAs, piRNAs and siRNA biogenesis and mechanism of action is depicted in Figure 6.

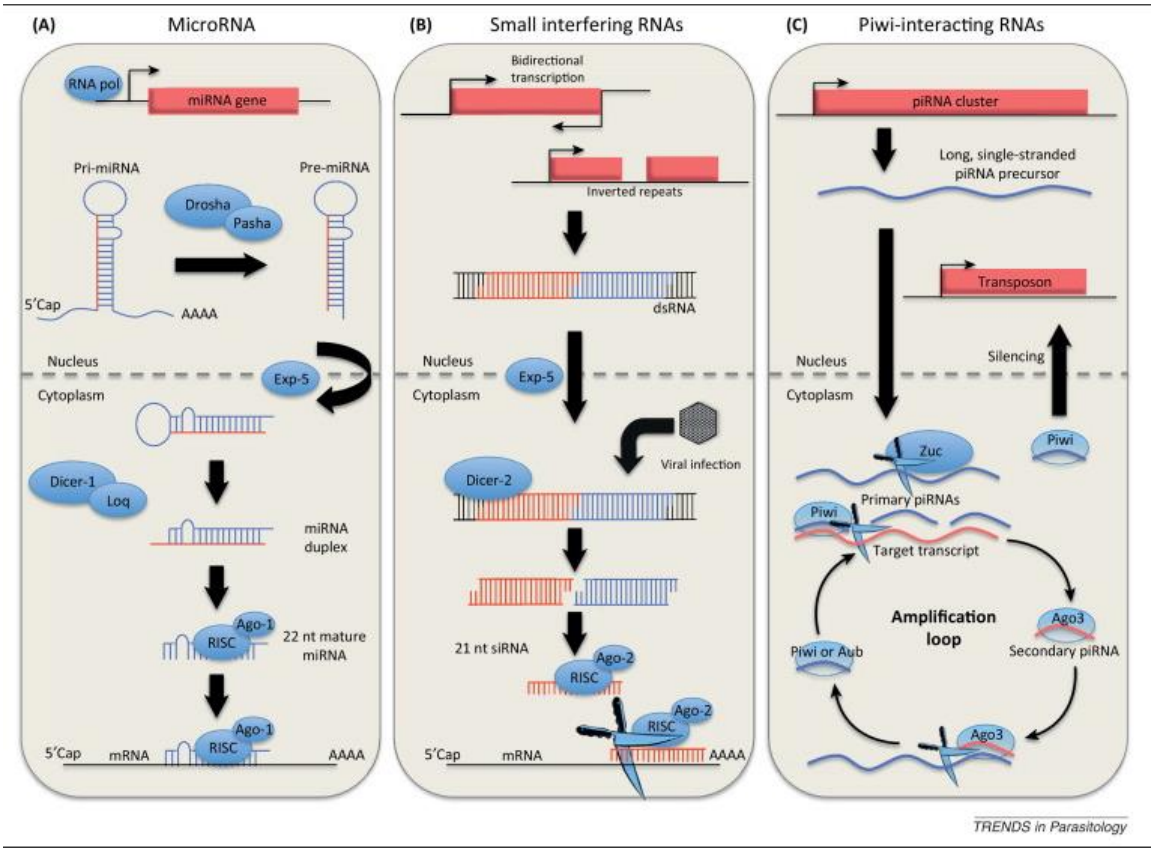


Figure 6 Comparison between miRNAs, siRNAs and piRNAs. Adapted from: Keira et al., Small RNAs: a new frontier in mosquito biology. 2013 Trends in Parasitology

LncRNAs: LncRNAs (long non-coding RNAs), differently from small ncRNAs (miRNAs, and piRNAs), are longer than 200 nucleotides; they are transcribed by RNA polymerase II and are subjected to 5' capping, splicing, polyadenylation and chemical base modification [166]. LncRNAs have been classified accordingly to their genomic position [167], including intergenic, enhancers and intronic lncRNAs.

Intergenic lncRNAs (or lincRNAs) are located between two protein-coding genes, and account for the majority of lncRNAs. Interestingly, the promoters of lincRNAs are enriched for active histone modifications such as H3K4me3, H3K9Ac and H3K27Ac, that are the same of mRNAs [168][169]. LincRNAs can function through different mechanisms: *cis* or trans transcriptional regulation, if lincRNA function is limited to neighboring genes or involves genes located at distant sites, respectively; translational control, splicing regulation and other functional mechanisms. For example, HOTAIR (Hot transcript antisense RNA), recruits PRC2 (Polycomb Repressive Complex 2) to silence developmental genes in distal chromosome, thus acting in *trans* [170]. Another example

of greatly studied intergenic lncRNA is Xist (X-inactive specific transcript) that coordinates mammalian X-chromosome inactivation by mediating epigenetic modifications [171][170].

Enhancer lncRNAs are transcribed from enhancer regions. This class of lncRNAs shows the ability to in *cis* activate neighboring genes, as in the case of LUNAR1 lncRNA that regulates the expression of the nearby IGF1-R (insulin growth factor 1 receptor) by mediating a chromosome looping [166].

Intronic lncRNAs are typically located within introns of protein-coding genes. According to the direction of transcription, intronic lncRNAs can be divided in: 1) sense lncRNAs, which are transcribed from the sense strand of protein-coding genes and can overlap introns and part of the exon; 2) antisense lncRNAs, which are transcribed from the antisense strand of protein-coding genes and can overlap an exon [167]. An example of antisense lncRNA is Bace1-AS (B secretase antisense); this lncRNA, in fact, is antisense to the gene Bace I (B secretase), displays base pair complementarity to Bace I and protects it from RNase H enzymatic degradation [172].

1.3.3 Transposons and transposon-like repetitive elements

Transposons are pieces of DNA able to jump around the genome through a mechanism known as transposition. Transposons and transposon-like repetitive elements occupy a large part of human genome (44%), but only a small proportion (less than 0.05%) has conserved its activity [173]. The impact of transposons on the evolution of primate genomes is enormous in term of structure and function [174].

Based on their mode of transposition, transposons are divided into two major classes. Class I transposons, also called retrotransposons, transpose through a "copy and paste mechanism" by reverse transcription of an RNA intermediate and class II, also called DNA transposons, which instead transpose through a "cut and paste" mechanism. Class I can be further divided in two subclasses: class I.1 includes retrotransposons, that are characterized by the presence of LTRs (Long Terminal Repeats) and encode products with structural homology to the retroviral gag- and pol-encoded proteins; class I.2 comprises the non-LTR retrotransposons that, similar to class I.1 retrotransposons, encode gag- and pol-like proteins but do not have LTRs. This class includes LINEs (Long

Interspersed Nuclear Elements), that are present in the genome as highly repeated sequences, 6-7 Kb in length, are transcribed by the RNA polymerase II and contain an open reading frame for the transcription of the reverse transcriptase of retroviruses. SINEs (Short Interspersed Nuclear Elements), or Alu-like sequences are highly repeated sequences 100-300 base long transcribed by the RNA polymerase III, that transpose via an RNA intermediate [175]. SINEs are not formally considered to be class I because they do not encode the reverse transcriptase. Interestingly, they are considered to be originated from the reverse transcription of cellular RNAs such as tRNAs [175][176]. Retrotransposons now account for approximately one third of the human genome, but had greatly proliferated during the past 80 million years of primate evolution. By generating insertion mutations, alterations in gene expression and genomic instability, and also by contributing to genetic innovations, retrotransposons influence human genome [174].

Class II transposons, also called DNA transposable elements, make up ~3% of the human genome and transpose via a "cut and paste" mechanism by excising themselves from the genome, moving as DNA and pasting themselves into new genomic sites. Differently from retrotransposons, the transposition mechanism requires a DNA rather than a RNA intermediate [176]. DNA transposons are characterized by the presence of two small inverted repeats (IRs) flanking one or more open reading frames that encode a transposase, which is the enzyme necessary for the transposition mechanism. It is believed that DNA transposons were active during early primate evolution, although they are currently not mobilizing in the human genome [174].

DNA transposons can be divided in three major groups that share this basic design: Ac-like transposons, P-like transposons, and the Tc1/mariner-like superfamily of transposons, that comprises Tc1, mariner-like (Sleeping Beauty) and Pogo transposons [175] [177].

Members of the Tc1/mariner-like superfamily are the most widespread transposons in nature, being present in ciliates, animals, plants and fungi [177]. Tc1/mariner transposons are typically of about 1300-2400 bp in length, and contain a single gene encoding a transposase flanked by terminal IRs. The members of this superfamily show sequence similarity as well as similar molecular mechanisms of transposition [176][177].

The length of the IRs and number of transposase binding sites can vary among the members of the superfamily (Figure 7). For example, the members with the simplest structure, such as Tc1 and Himar1 (a mariner transposon), possess IRs less than 100 bp in length with one single binding site per repeat [178]. Other members with a more complex structure, such as Tc3 transposons, present IRs 400 bp in length with two binding sites per repeat [177]. Instead, IR/DR transposons such as Sleeping Beauty, possess a 15-20 bp direct repeats (DRs) at the end of their IRs, that are 200-250 bp in length [179].

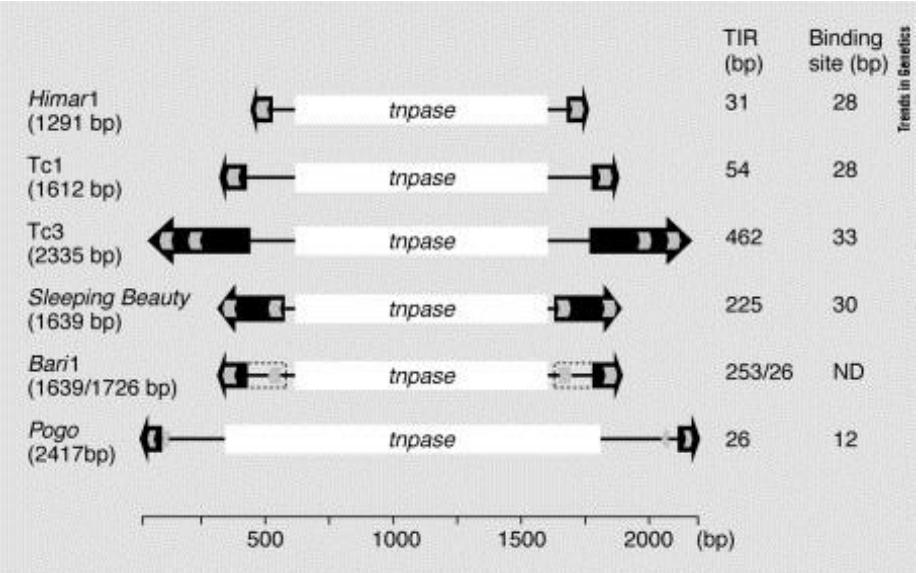


Figure 7: Tc1/mariner superfamily members differ in the length of IRs and in the number of transposase binding sites. Adapted from Plasterk et al., *Resident aliens: theTc1/mariner superfamily of transposable elements. 1999 Trends in Genetics*

Tc1/mariner transposons use similar mechanisms to transpose within the genome, generally involving a staggered double strand cut operated by the transposase enzyme at the ends of the transposon. This staggered cut causes the formation of characteristic "footprint" when cellular DNA repair processes intervene to repair the gap. Importantly, Tc1/mariner transposons have a preference for the integration into TA sequences; since the frequency of transposition at a certain TA site is determined by the sequence flanking the TA sites, it is believed that transposase recognizes some bases next to the TA sequence [180].

1.4 Functional screens used in mammalian cells: gene trap mutagenesis

The genome sequencing projects allowed the identification of almost all genes responsible for the biological complexity of various organisms. Yet, due to the enormous flux of information provided by genome sequencing, the assignment of a function to each of the identified genes results difficult [181]. Genetic screens represent a mean to identify gene functions, and have been successfully used in model organisms such as yeast, Drosophila, Caenorhabditis Elegans and mouse [182][183][184][185].

Also mammalian cell cultures arose as tools to elucidate gene function, as they permit to study *in vitro* biological processes such as apoptosis, senescence, cell proliferation and differentiation [186]. Genetic screens can be divided in forward and reverse genetic screens. In forward screens the starting point is a mutagenesis step that is followed by the detection of the genes responsible for a particular phenotype. Reverse screens, instead, analyze the phenotype of an organism following the disruption of a known gene [186] [187].

Similarly to the screens in model organisms, forward screens in cell cultures rely on mutagenesis, selection by functional assays of altered phenotype and identification of the causative mutation. Typically, mutations can be generated by chemicals or by introduction of genetic material.

Chemicals induced mutations are typically obtained with the use of mutagenizing agents such as N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) and ethyl metane sulphonate (EMS) [186]. Mutations generated by introduction of genetic material, instead, can be obtained with the use of retroviruses and DNA transposons. Alternatively, other types of forward screens based on the introduction of genetic material involve the use of cDNA libraries, SiRNAs, Sh-RNAs (short hairpin RNA) and miRNAs.

Retroviruses have been widely used in forward screens, in particular in gain-of-function screens. Oncogenic mutations can be can efficiently induced by slow transforming retroviruses via the insertion of the provirus into the genome; then, the determination of the site of insertion of the provirus will permit to identify these mutations [188]. Also DNA transposons have been extensively used in forward genetic screens. In fact, mutations can be generated by inserting DNA transposons into the

genome; subsequently, the mutated gene can be retrieved by using DNA transposons as tags [186].

An alternative approach in forward screens consists in the introduction of genetic material with the use of cDNA libraries, siRNA, Sh-RNA and miRNAs [186][189][190]. These screens have been used for both gain of function and loss of function approaches, based on the type of functional assay used for the selection of the phenotype. For example, in gain of function screens, a cDNA library of expression plasmids is introduced into the cells, followed by specific selective regimens [191]. In mammalian cells, these screens require iterative rounds of selection and usually only few functional gene products are identified by clone sequencing. Thus, gain-of-function screens are powerful tools to identify genes sufficient to confer a particular phenotype [186]. Oncogenes are typically identified with cDNA based gain of function screens: TYRO3 tyrosine kinase receptor for example, was identified as a pro-tumorigenic factor in melanoma cells through a genome-wide gain-of-function screen [192]. In recessive loss of function screens, instead of screening a collection of genes, a candidate cDNA suspected to be responsible for a phenotype is tested and cells are selected for the absence of specific properties [186]. For example, in a collection of cells that were unresponsive to NFKb after treatment with TNF (tumor necrosis factor), a recessive-loss-of-function screen identified RIP (transducer receptor-integrating protein) to transmit TNF-dependent apoptosis signals [193].

For the aim of this Thesis work we have decided to use transposon based genetic screen.

1.4.1 Transposon based screens

DNA transposons are precise and easily controllable DNA delivery vehicles capable of driving insertional mutagenesis, thus emerging as genetic tools for various applications, including genetic screens. The enormous potentiality of transposons resides in the fact that under laboratory conditions it is possible to virtually place any sequence of interest between the two IRs and trans-supplement transposase in the form of an expression plasmid [176]. Importantly, the desired insertion frequencies can be obtained by adjusting the amounts of the delivered plasmids [176]. Moreover, by transiently providing the transposase source, transposons can be remobilized from insertion sites, by allowing the monitoring of excision and re-integration events.

The most used transposons in mutagenesis screens are Sleeping Beauty (SB) and piggyBac (PB). SB belongs to the TC1/mariner-like superfamily and is an ancestral transposon found in salmonid fishes that was presumed to be active more than 10-15 million years ago, but has become inactive due to the accumulation of mutations during evolution. SB transposon was "kissed to life" *in vitro* by molecular reconstruction based on site-directed mutagenesis and now is widely used as genetic tool [177][194]. SB transposon presents a preference for integration in TA sites, in particular palindromic AT repeats [195].

PB transposon is a member of a group of short-inverted-repeat elements, characterized by their specificity for TTAA target sites and originally identified in Cabbage Looper [196]. PB transposon is 2,5 kB in length and has 13 bp identical TIRs (Terminal Inverted Repeats) with additional asymmetric internal 19 bp IRs. Similar to other DNA transposons, PB encodes a transposase that facilitates insertion and precise excision [196][197].

A great number of large scale transposon-based mutagenesis screens, for example, have been performed in embryonic cells with both SB and PB by transfecting transposon donor and transposase expression plasmids [198][199]. In recent years, however, PB transposons have become very popular as genetic tools due to some advantages compared to SB. First, PB can carry larger foreign DNA cargos (up to 9.1 kilobases) without significantly decreasing integration frequencies; SB transposons, conversely, show a 30% decrease of transposition frequency for every 1 kilobase increase in size

from their original 1.7 kilobases length [176][200]. Second, PB has a stronger transposition activity and usually leaves no footprint after excision [176] [201].

One of the main applications of transposon-based screens, as in the case of other screens, is gene trapping. Transposon based DNA delivery for insertional mutagenesis can be combined with various types of mutagenic cassettes. For example, 5'-gene trap cassettes include polyadenylation sequences and splicing acceptors to disrupt transcription of genes when the vector inserts in introns [176]. Typically, cassettes also encode a reporter gene, such as GFP (Green Fluorescent Protein), β galactosidase or antibiotic resistance, that is expressed only in case of correct splicing between exons of the trapped gene and splicing acceptor sites carried by the transposon vector [202]. Transposon based screens have been extensively used in oncogene trapping [176]. Transposon vectors can be fitted with oncogene trap cassettes, containing strong viral enhancers/promoters, splicing acceptor site and polyA sites, to induce gain-of-function mutations. The function of viral enhancers and promoters is to permit the overexpression of a full-length or a truncated protein product of the trapped gene; whereas, splice acceptor and polyA sites lead to gene truncation with dominant phenotypes [203].

An interesting application of transposon-based screens is the customization of tissue-specific screens for cancer development studies. To this aim, a possible strategy is to establish mouse lines that conditionally express the transposase from tissue specific promoters [204]. For example, a transposon based screen was used to help distinguish between CRC inducer and passenger mutations. In particular, mice harboring mutagenic SB transposon were crossed with mice expressing SB transposase in gastrointestinal tract epithelium. Retrieval of transposon insertions from the mice which developed intestinal lesions allowed the identification of 17 candidate genes that had never been implicated previously in CRC [205].

1.4.2 Toward the interrogation of non-coding genome

Our understanding over how much of the genome is functional was, until recently, restricted to the protein-coding sequences. This is underlined by the presence of a great number of genetic screens aimed to perturb protein-coding regions and their relative

protein products, mainly due to the use of protein-coding gene (PCG) traps [186] [187]. Non-protein-coding DNA has been regarded for long time as junk DNA; yet, its extent increases proportionally with developmental complexity [135]. Today, we know that over 90% of the genome is transcribed in ncRNAs that can be intergenic, intronic or overlapping with PCGs [206]. In the last few years, it has been shown that RNA *per se* plays regulatory functions; whereas, ncRNAs regulate expression of homeotic and metabolic genes [207]. Moreover ncRNA display precise tissue expression patterns and are differentially expressed in diseases, such as cancer [208]. These findings, together with genome wide association studies that have mapped disease associated variations to non-coding regulatory regions, are in favor of ncRNAs functional properties. All these notions underlie the importance of developing appropriate genetic tools to explore the non-coding part of the genome.

Chemical mutagens based screens, cDNA library based screens and retroviral based screens all present some limitations: 1) mutations induced by chemical mutagens (such as MNNG and EMS) are difficult to recover [186]; 2) the success in the expression of a cDNA library depends on the quality and source of cDNA libraries; moreover, cDNA expression libraries may bias toward shorter cDNAs [186]; 3) retroviruses carry strong enhancers in their LTRs, and can influence the expression of genes located hundreds of kb away [209]; moreover, retroviruses preferentially target the 5'end of expressed genes, thus limiting the number of candidate cancer genes accessible to retroviral mutagenesis [188][210].

For these reasons, transposon based screens are more appropriate tools for randomly interrogating the genome. Indeed, despite showing a preference for inserting into specific target site (AT for SB transposon and TTAA for PB transposon), transposons insert in the genome fairly randomly and show the potentiality to insert anywhere in the genome, comprising non-coding regions [176]. Indeed, transposon based screens have identified not only oncogenes but also non-coding regions promoting tumorigenesis. This is the case of a PB based screen in pancreatic cancer cells that identified a non-coding region (Ncruc) that is upstream to the *Cdkn2a* protein-coding gene. Interestingly, transposon insertions were associated with reduced levels of *Cdkn2a*, demonstrating the power of PB insertional mutagenesis in identifying a cancer relevant, non-coding genomic segment with *cis* regulatory function [211].

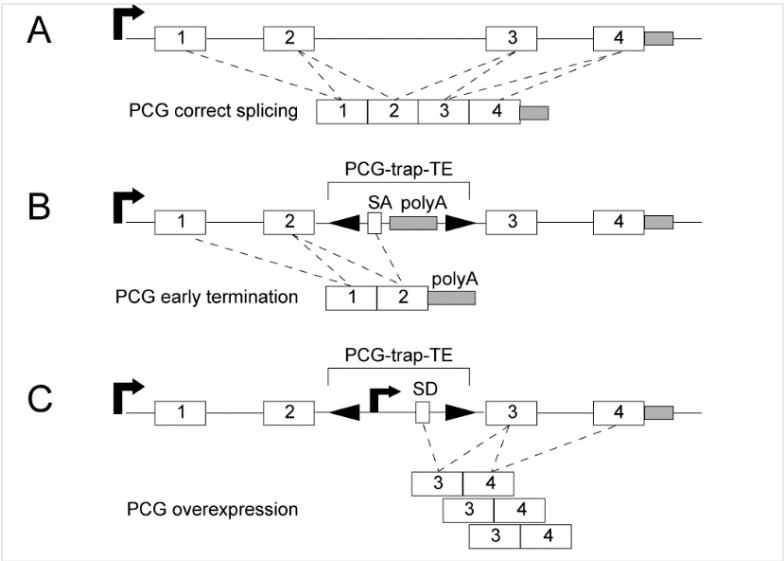


Figure 8: PCG splicing and PCG trap transposon insertion

White boxes=exons;
TE=transposable element;
Black arrows=TE IRs;
Grey boxes=termination signal;
Bent black arrows=promoter sequence;
SA=splicing acceptor site
SD=splicing donor site

Since the insertion of transposons occurs preferentially in intron or intergenic regions, in order to target protein-coding genes (PCGs) (fig. 8A), researchers created protein-coding genes traps within transposons (PCG-trap-TE, TE=transposable elements) by inserting splicing acceptor and donor sites (SA and SD), transcription stop signals (polyA), or promoters [176][212] (fig. 8B, C).

However, ncRNAs of unknown existence and function may reside within introns of PCGs or nearby PCGs (fig. 9 A). In the case scenario of a yet unknown ncRNAs that is within the intron of PCG (fig. 9 B), the PCG-trap-TE by landing within or near the unknown ncRNA sequence will modify both the PCG and the ncRNA by premature termination (fig. 9). This, in a functional screen, will not allow to discriminate if the observed effect is due to mutation of the PCG or of an unknown ncRNA, which remains undetected.

For these reasons, and for our ambition to interrogate the non-coding part of the genome, in this Thesis work we have decided to use a TN-based screen in which the TE will not contain neither SA, SD, polyA or promoters, but only the coding sequence of a selection gene (e.g. neomycin; ncRNA-TE) (fig. 9C); therefore, in the case of an intronic unknown ncRNA (fig. 9A), ncRNA-TE will still disrupt the sequence of the ncRNA where it inserts, but without exerting any effect on the surrounding PCG that will undergo regular splicing; consequently, the phenotype generated will be specific for the ncRNA and not due to modification of the surrounding PCG (fig. 9C).

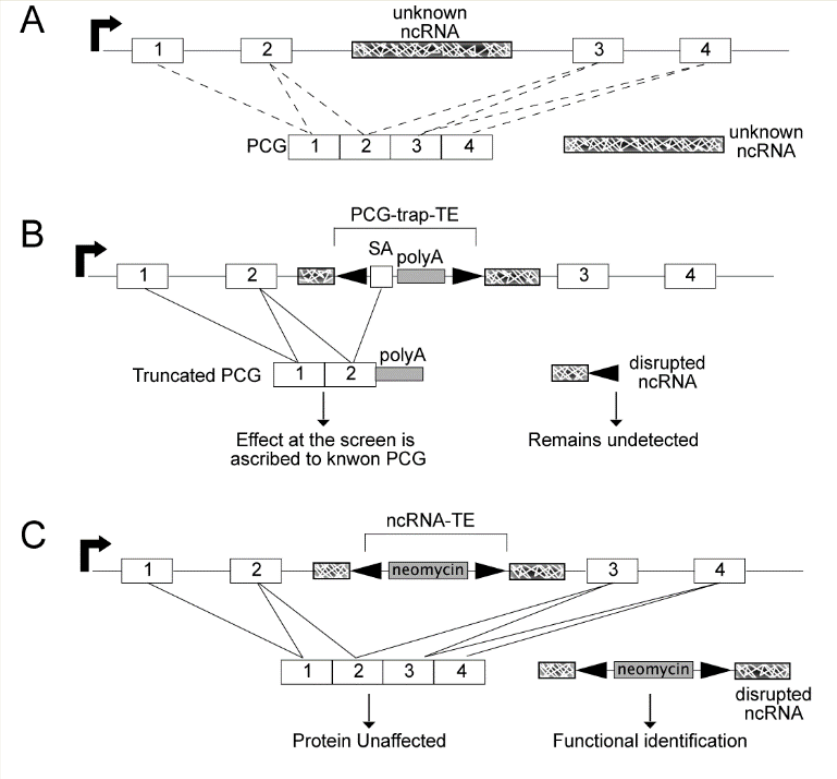


Figure 9: ncRNA can be found within PCG introns (9A); effect of classical PCG-trap-TE and of a modified ncRNA-TE for the discovery of unknown ncRNAs

2. AIM AND OBJECTIVES

Metastasis is the major cause of cancer-related deaths in many types of cancer, including colorectal cancer (CRC). Understanding the molecular mechanism driving metastasis is fundamental for the identification of alternative routes to contrast this dreadful disease.

The overall aim of this work is to identify genetic elements that orchestrate tumor plasticity and are necessary for CRC metastasis, by mainly focusing on transcribed non-coding regions (i.e. ncRNAs).

This research project stems on three interconnected considerations: 1) the hypothesis that metastasis requires plasticity of tumor cells that switch from an epithelial to a mesenchymal program (EMT) and back to an epithelial state (MET); 2) the part of the genome that does not encode for proteins (i.e. non-coding) represents more than 98% of the whole genome and has expanded during the evolution more than the protein coding part. According to such evidence, evolution appears to rely on protein regulation rather than on protein number and diversity; 3) functional genetic screens are great tools to interrogate the whole genome and recover genetic elements that control a specific phenotype. Yet, forward genetic screens have mainly been used to target the protein coding part of the human genome.

The above aim will be accomplished by combining forced Single Cell Suspension Assay (fSCS), an *in vitro* assay able to select cells with EMT traits, with an *in vitro* transposon (TN) based screen that randomly interrogates the genome of HCT116 CRC cell line. Retrieval of TN insertion from the genome of transposed fSCS selected cell and characterization of the genomic region mutated by the TN will help identifying crucial genomic regions regulating CRC metastasis.

3. MATERIALS AND METHODS

Cell lines and Cultures

HCT116, HT29, SW480 and SW420 colorectal cancer cell lines were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA) and were maintained in RPMI 1640 medium, containing 10% fetal bovine serum (FBS)(#FA30WS1810500, Carlo Erba). No antibiotics were used for cell cultures unless required for selection of transfected/transduced cells. Cell lines were maintained at 37°C, under 5% CO2 in humidified incubators and routinely tested for mycoplasma contamination using MycoAlert detection Kit (Lonza, Cologne, Germany). Only mycoplasma negative cells were used for experiments. 293FT cells (Invitrogen) were used for lentivirus production and were grown in DMEM supplemented with 10% FBS, 2 mM glutamine, Sodium Piruvate and non-essential amino acids (all from Life Technologies, Grand Island, NY, USA).

Lentivirus Production and Transduction

For lentivirus production, 293FT cells (Invitrogen) were plated 24h prior to transfection in order to be at 60-80% confluence. Standard calcium phosphate protocol was used to transfect pWZL Twist-ER (Plasmid # 18799, Addgene) or pLKO.1-sh vectors (Plasmid #10878), together with pLP1, pLP2, and pVSV-G at a 1.25: 1: 1 : 0.125 ratio. 16h post transfection medium was changed and 72h later lentivirus containing supernatant was collected, centrifuged and 0.4 um filtered to remove cell debris, and finally concentrated using centricon 10'000 MW cut off (Millipore). Transducing units per ml of supernatant were determined by limiting dilution titration in HCT116 cells. A MOI (multiplicity of infection) of about 5 was used for transducing cells using Polybrene (Sigma-Aldrich) at a final concentration of 8 μg/ml to increase transduction efficiency. 48h after transduction blasticydin selection was started.

Cell transfection

Lipofectamine 2000 reagent (#1668019 Life Technologies) was used to transfect plasmids or miRNAs precursors in HCT116, HT29, SW480 or SW620 cells. Cells were plated 24h prior to transfection in order to be at 60-80% confluence. Accordingly to Lipofectamine 2000 protocol, 24h later cells were washed and incubated with RPMI

1640 culture medium in the absence of antibiotics. The appropriate amount in μg of plasmid vectors or miRNAs precursors at the final concentration 100nM were incubated with Opti-MEM (Gibco, Life Technologies) and then mixed with Lipofectamine 2000 reagent plus Opti-MEM. This mixture was incubated for 20 minutes at room temperature and then incubated on cells for 24h.

Used plasmids or miRNAs precursors:

pT2 vector, carrying the Sleeping Beauty transposon and pCMV-SB100x, carrying the Sleeping Beauty transposase with high enzymatic activity (PMID: 19412179) were provided by Dr. Ivics (Max Delbrück Center for Molecular Medicine, Berlin, Germany) and Dr. Largaespada (University of Minnesota, USA). We modified the pT2 transposon vector in order to carry the coding sequence of eGFP under the control of CMV promoter, but in the absence of any polyA signal piggyBAC transposon system; empty eGFP vector (pEGFP-N1 Clontech); eGFP-Btbd7 (gently provided by Dr. Kenneth M. Yamada, National Institute of Health, Bethesda, Maryland, USA); Piggy Bac (PB) transposon system (purchased by System Bioscience). In particular, we used PBHR100A-1 as donor vector and PB210PA-1 as vector supplying the Super PB transposase; scramble (#AM17110, Life Technologies) and miR-23b precursors (#AM17100, Life Technologies).

Cell Treatments

4-Hydroxytamoxifen (4-OHT, Sigma) treatment was performed as follows: HCT116 Twist-ER cells and HT29 Twist-ER cells plated 24 hours earlier in 6 cm or 10 cm plates were exposed to 4-OHT at the final concentration of 0.1 μM in RPMI 1640 complete medium. Cells were continuously exposed to this treatment for 72 hours, during which time the medium was changed every day and cell cultures were eventually passaged by trypsinization after confluency was reached. At the end of the treatment period, cells were immediately used for in vitro *Anoikis* and *Forced Single Cell Suspension (fSCS)* assays and for protein and RNA extraction.

Cell response to chemotherapy was done by treating exponentially growing cells, plated in 24 multiwell plates, with 2 ug/ml of Fluorouracil (5-FU) for 24h in complete medium. Cell viability was assed 24 h later by MTT (M2003, Sigma) staining. MTT stock solution was added to cells at a final concentration of 0.28 mg/ml and incubated for 2 h

at 37° C. Supernatant was then discarded and cells were air-dried. Reduced MTT, measure of cellular metabolic activity and index of cell viability, was dissolved by adding 200 ul of DMSO and measured by reading the absorbance at 580 nm with Infinite M1000 microplate reader (Tecan, Männedorf, Switzerland).

EpCAM detection

Human epithelial adhesion molecule (EpCAM) was used as a marker of cellular differentiation and was measured by FACS using APC Mouse Anti-Human EpCAM (BD Biosciences). To this end cells were detached from culture dishes using 2mM EDTA (Ethylenediaminetetraacetic Acid, Sigma) in PBS, pelleted and suspended in 10% FBS at a final concentration 10x10^6 cells/ml. Following 30 min blocking at room temperature, 100 ul of cell suspension was stained at room temperature with 5 µl of antibody for 30 min. Cells were washed with ice cold PBS and analyzed using the 633 excitation laser of the BD FACSCanto™ flow cytometer system.

In vitro Anoikis and Forced Single Cell Suspension (fSCS) Assays

For the *in vitro* anoikis assay, 1x10^5 cells were suspended in serum free RPMI 1640 containing 1% BSA and cultured on ultra-low attachment six-well plates (ULA 3471, Corning) for 24 h. ULA plates have a covalently bound hydrogel layer that inhibits cellular attachment like polyhema coating with the advantage of better standardization of the assay. Next, cells were collected and washed with PBS without calcium phosphate, treated for 5 minutes with trypsin for the disruption of cell aggregates, plated on standard 10 cm dishes in complete medium and let grow for 8-10 days. For the fSCS assay, as for the anoikis assay, 1x10^5 cells were suspended in serum free RPMI 1640 containing 1% BSA and cultured on ultra-low attachment six-well plates (ULA 3471, Corning) but with the addition of 1mM EDTA which chelates calcium, a co-factor necessary for intercellular interactions, by preventing cell-cell contacts during cell culture on ULA plates. Time of culture on ULA plates was adjusted according to cell type to obtain the lowest amount of growing colonies: 24 hours for HCT116 and SW620, 36 hours for HT29 cells and 48 hours for SW480. Next, cells were collected, washed with PBS and plated on standard 10 cm dishes for 8-10 days. For both assays, survived cells, that grew and form colonies on 10 cm dishes were stained with Crystal Violet. Crystal violet stained dishes were acquired using GelDoc. Cell colonies were counted either

manually our automatically using Fiji [213]. Fiji was also used to determine colony circularity index, b

To obtain *fSCS* resistant pools from both parental and transposed HCT116 cells, two sequential rounds of *fSCS* were performed on ULA T75 flasks. By pooling together and expanding the cells surviving after the first round (T1) and by subjecting them again to *fSCS* assay T2 progeny was generated, with the same process used for T1. This was repeated three more times to obtain T3, T4 and T5 fSCS selected pools of cells. Early T3 cells represent T2 cells 7 days immediately after fSCS assay, before being further pooled and expanded.

Western Blotting

To obtain protein extracts, cells that had been previously scraped with ice cold PBS and pelleted were incubated for 20 min on ice in NP40 lysis buffer (0.5% NP40, 50 mM HEPES, 250 mM NaCl, 5 mM EDTA, 0.5 mM EGTA, 1 mM DTT) supplemented with Protease Inhibitors (Complete, Roche, Mannheim, Germany) and Phosphatase Inhibitor Cocktail 1 (P5726, Sigma Aldrich), while vortexing every 5 min. Cell lysates were clarified by centrifugation at maximum speed for 20 min. Proteins were quantified using the Protein Assay Dye Reagent Concentrate (Biorad, Hercules, CA, USA). For immunoblotting analysis, 40 μg of proteins were separated using 4-20% SDS-PAGE Criterion-TGX-stain free Precast Gels (Biorad) and transferred to Nitrocellulose membranes (Biorad). Membranes were blocked with 5% non-fat dried milk in TBS-0.1% Tween20, and incubated overnight with the following primary antibodies: BTBD7, e-CADHERIN, ZEB-1, VIMENTIN, TWIST, CAVEOLIN, TUBULIN, GAPDH, VINCULIN. Incubation with secondary antibodies, ECL IgG-HRP-linked anti-rabbit, anti-mouse or anti-goat (GE Healthcare) was performed at room temperature for one hour at a dilution of 1:3000-1:5000. According to signal intensity, either ECL Western Blotting Detection Reagents RPN 2106 (GE Healthcare Lifesciences, Little Chalfont, UK) or Luminata Forte Western HRP Substrate (Merck Millipore, Darmstadt, Germany), were used for secondary antibody detection using Chemidoc MP imager (Biorad). Bands were analyzed and quantified using Image Lab v5.2 (Biorad).

RNA extraction, cDNA synthesis, and Quantitative Real Time PCR (qRT-PCR)

RNA was extracted using the Isol-RNA Lysis Reagent (5 Prime, Hamburg, Germany) according to protocol instructions. Extraction was followed by PCA (Phenol: Chloroform: Isoamyl Alcohol 5:24:1, Sigma Aldrich), procedure that reduces salt contamination, and by DNase digestion (Turbo-Dnase, Ambion, Thermo Fisher Scientific). RNA quality was assessed using agarose gel electrophoresis after RNA exposure to 70°C for 5 min. cDNA was synthesized from 1 μg of RNA, using the AMV Reverse Transcriptase (Promega) with random primers (Promega). cDNA was diluted 10 times and used for qRT-PCR using iQ SYBR Green Supermix (Biorad) with the appropriate primers. For detection of miRNAs, we used reagents, primers and probes from Applied Biosystems. performing miRNA specific reverse transcriptase reactions and qReal-Time PCR according to the manufacturer instructions.

qRT-PCR reactions were carried out either in a 96-well optical reaction plates using Two-Color Real-Time PCR Detection System MyiQ2 (Biorad) or in 384-well optical reaction plates using Applied Biosystems 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific), according to manufacturer 's protocol. The $2-\Delta\Delta Ct$ method was used to calculate the relative abundance of RNA genes, measuring *GAPDH* or U6 RNA expression as housekeeping control. Primers used in this work were designed using Primer 3 Plus [214] were purchased from Sigma-Aldrich and are listed in Supplementary Table 3.

Genomic DNA extraction

Genomic DNA was extracted using Gentra Puregene Cell Kit (Qiagen Sciences, Maryland, USA), according to manufacturer's instructions. DNA quality was assessed using agarose gel electrophoresis. DNA concentration (absorbance at 260nm) and purity (ratio between absorbance at 260nm and absorbance at 280nm) were determined by using a spectrophotometer (Nanodrop, Thermo Scientific).

Linker-Mediated PCR

To identify the genomic position of TN insertion in HCT116 transposed cells, we performed LM-PCR according to [215] . In brief, genomic DNA from transposed cells was digested at 37° C digested with *Bfa*I (for cloning off the left IRDR) or *Nla*III (for cloning off the right IRDR). Linkers were ligated to NlaIII- (right-side) or BfaI- (left side)

digested genomic DNA using T4 DNA ligase. A secondary digestion (XhoI, right side; BamHI, left side) was performed to destroy concatamer-generated products. Primary and secondary PCR was performed using primers specific for linker and SB transposon sequences along with Fusion and barcode sequences. PCR amplicons were sequenced using the GS FLX (Roche). To purify digested DNA and PCR products we used QIAquick PCR purification kit from Qiagen following the directions provided by the kit. Final PCR products were ligated in pCR.2.1 Topo TA vector (Thermo Scientific) and transformed into DH5a competent cells. Two separate PCRs and nested PCRs were performed to obtain the left and the right boundaries of pT2-CMV-EGFP insertions in the genome.

DNA sequencing of minipreps (Promega) from topo TA clones containing LM-PCR products was done by direct Sanger sequencing, using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Fostercity, CA, USA). All primers and linkers used are listed in the following table:

hEcad-5	TGCCCAGAAAATGAAAAAGG
hEcad-3	GTGTATGTGGCAATGCGTTC
hNcad-5	ACAGTGGCCACCTACAAAGG
hNcad-3	CCGAGATGGGGTTGATAATG
hVim-5	GAGAACTTTGCCGTTGAAGC
hVim-3	GCTTCCTGTAGGTGGCAATC
HAS1-5	CGCTAACTACGTCCCTCTGC
HAS1-3	CCAGTACAGCGTCAACATGG
HAS2-5	GCCTCATCTGTGGAGATGGT
HAS2-3	ATGCACTGAACACACCCAAA
hSlug-5	GGGGAGAAGCCTTTTTCTTG
hSlug-3	TCCTCATGTTTGTGCAGGAG
hSnail-5	CCTCCCTGTCAGATGAGGAC
hSnail-3	CCAGGCTGAGGTATTCCTTG
hTwist-5	GGAGTCCGCAGTCTTACGAG
hTwist-3	TCTGGAGGACCTGGTAGAGG
ZEB1-5	CAGGCAGATGAAGCAGGATG
ZEB1-3	GACCACTGGCTTCTGGTGTG
ZEB2-5	GCGCTTGACATCACTGAAGG
ZEB2-3	ACCTGCTCCTTGGGTTAGCA
pT2-up	TCACTATAGGGCGAATTGGAG
pT2-within	CCAAGCTGTTTAAAGGCACA
pT2-down	TGTGGAATTGTGAGCGGATA
Bfa-Linker-S	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC
AS-Bfa-Linker-AS	[Phos]TAGTCCCTTAAGCGGAG[AmC3]
Nla-Linker-S	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACCATG
Nla-Linker-AS	[Phos]GTCCCTTAAGCGGAGCC[AmC3]
IRDR-R-primary	GCTTGTGGAAGGCTACTCGAAATGTTTGACCC
IRDR-L-primary	CTGGAATTTTCCAAGCTGTTTAAAGGCACAGTCAAC
Linker-primary	GTAATACGACTCACTATAGGGC
IRDR-R-nested	CCACTGGGAATGTGATGAAAGAAATAAAAGC
IRDR-L-nested	GACTTGTGTCATGCACAAAGTAGATGTCC
Linker-nested	AGGGCTCCGCTTAAGGGAC
BTBD7A_5(CDS)	AGTCAAATGCCTGGTTACGG
BTBD7A_3(CDS)	TGTCTGGCACATTGGACATT
BTBD7upA_5(3'UTR)	GTTTCCAATTTGCCTTCTGC
BTBD7up A_3(3'UTR)	GGCTTTGAGGCTTTTCAGTG

Statistical Analysis

Graphs presented in figures were obtained using GraphPad Prism v6.0d software and statistical analysis were done using JMP® v9.0.1 software. Data were examined using the two-tailed Student t test or un-paired two-tailed ManneWhitney U test. Differences were considered significant at $p < 0.05$.

Animal experiments

Animal experiments were reviewed and approved by the CRO (National Cancer Institute of Aviano) Institutional Animal Care and Use Committee and were conducted according to committee's and the Italian Ministry of Health guidelines. For all animal experiments, we used 4-week-old athymic nude mice (Charles River, UK Foxn1nu, females), except for orthotopic intracaecal injections in which were used 6-week-old to 8-week-old mice severe combined immunodeficiency (SCID) mice (Charles River, UK).

In vivo extravasation assay

I*n vivo extravasation assay* was performed as described in [230]. Cell suspensions were first stained with DiI (molecular Probes) 1:200 in serum free medium for 30' at 37°C. Then, 1,5x10^6 cells were suspended in 200 µl of PBS containing 0.1% Fluorescein and retro-orbitally inoculated in previously anesthetized nude mice to reach blood circulation. Fluorescein was used to detect the homogeneity of injection using the *in vivo* imaging system (Ivis Lumina, from Xenogen) at time 0, and it was not detectable any longer 72h later, when mice were sacrificed for analyses. Mice were humanly culled by cervical dislocation. During mice necroscopy lungs were inflated through the trachea with 1 ml of ice cold 4% PFA. Lungs were immediately observed at the stereomicroscope and pictures of whole lungs from both top and bottom were taken. Cryostat-cut sections were then obtained from OTC lung inclusions (see below).

Orthotopic intracaecal injections

Orthotopic mice injection were performed in collaboration with Dr. Alex Mirnezami (Cancer Research UK Centre, Southampton, UK). 6-week-old to 8-week-old severe combined immunodeficiency (SCID) mice (Charles River, UK) were anaesthetized prior to midline laparotomy and exteriorization of the caecum. A 1:1 suspension of cells and Matrigel was injected submucosally into the caecal wall under magnified vision, raising a bleb on the caecum. For each animal, 0.5×10^6 cells (TN4-sorted or TN4_20) were implanted orthotopically, with the entire experiment conducted in duplicate. Primary tumors grew in all animals. When showing signs of disease or >10% weight loss, mice were humanely culled, and colon, liver and lungs were harvested.

Histopathology and Immunohistochemistry

For orthotopic injection experiments, formalin-fixed organs were embedded in paraffin, and standard techniques were used by our institutional pathology facility to stain tissue sections with hematoxylin and eosin staining. Stained organ sections were evaluated also by one of CRO's pathologists.

Fluorescence Microscopy

To study the cellular localization of ectopic eGFP-Btbd7, three control clones expressing eGFP only (Empty clone #1, #2 and #3) versus three positive clones expressing eGFP-Btbd7 (eGFP-Btbd7 clone #1, #2, #3) were grown on glass coverslips, fixed with PFA 4% and observed using a confocal laser-scanning microscope (TSP2 Leica) interfaced with a Leica DMIRE2 fluorescent. To detect the presence of DiI stained HCT116 P0 AND HCT116 P2 infiltrated into the lungs of mice following *in vivo extravasation assay*, lung sections from OCT inclusions were fixed with PFA 4% and nuclei were counterstained with TOPRO (Molecular Probes). Stained tissue specimens were observed using a confocal laser-scanning microscope (TSP2 Leica) interfaced with a Leica DMIRE2 fluorescent.

4. RESULTS

4.1 In vitro Forced Single Cell Suspension Assay (fSCS) is a more stringent assay compared to in vitro anoikis resistance

To develop an *in vitro* assay that selects for metastatic features in epithelial CRC cell lines, we tested loss of cell matrix contact by growing cells on ultralow attachment plates as described in [216] and in materials and methods section of this Thesis (*in vitro* anoikis). We observed that both microsatellite stable (MSS; HT29, SW480, SW620) and microsatellite unstable (MSI; HCT116) CRC cells survived to low adherence condition with minimal cell death (untreated=anoikis, fig. 4.1B, upper part). In addition, by observing cells kept in low adhesion conditions, we noted that most cells formed cell aggregates, strengthening cell-cell contacts. Interestingly, this observation is consistent with [216], in which it was reported that transformed cells survived to *in vitro* anoikis by enforcing cell-cell contacts and by forming big cell aggregates in suspension; moreover, the same happened in cells transduced with the cDNA encoding Trkb, a neurotrophic tyrosine kinase receptor that was identified as regulator of anoikis resistance. Thus, we hypothesized that a physiological method to overcome loss of survival signal from ECM is obtained by increasing cell-cell contacts. To sum up, our results indicate that, at least in our hands and with the cells used, the *in vitro* anoikis assay, i.e. cell culture in absence of cell matrix-contact, is not sufficiently stringent to induce cell death and to select cells with more aggressive features. EMT and tumor metastasis rely on loss of epithelial cell-cell contacts [31][86][87]. Therefore, we hypothesized that by preventing cell-cell contacts in addition to loss of cell matrix contacts, we could impair cell survival and obtain a stronger selective pressure. In order to achieve this, we incubated HCT116 cells in low adherence in the presence of EDTA, which impairs Ca^{2+}-dependent cell-cell contacts. Thus, to determine the opportune EDTA concentration to use, we incubated HCT116 cells on ultralow attachment plates in the presence of several concentrations of EDTA (0.5mM; 1mM; 2mM; 5mM and 10mM); 24h later, we plated cells on culture dishes in complete growth medium and let them grow for additional 7-10 days; finally, we stained surviving cell colonies with crystal violet. We observed that the lowest EDTA concentration (0.5mM) was not stringent enough because, similar to untreated, almost all cells were still alive; conversely, higher

concentrations of EDTA (2mM, 5mM and 10mM) were too harsh, as all the cells died; only 1mM EDTA seemed to confer the perfect selective pressure, as it impaired cell survival but was not too stringent to prevent the recovery of survived cells (fig. 4.1A). Thus, the incubation in low adherence conditions with added 1mM EDTA seemed to confer the desired selective pressure without preventing the recovery of survived cells.

Figure 4.1

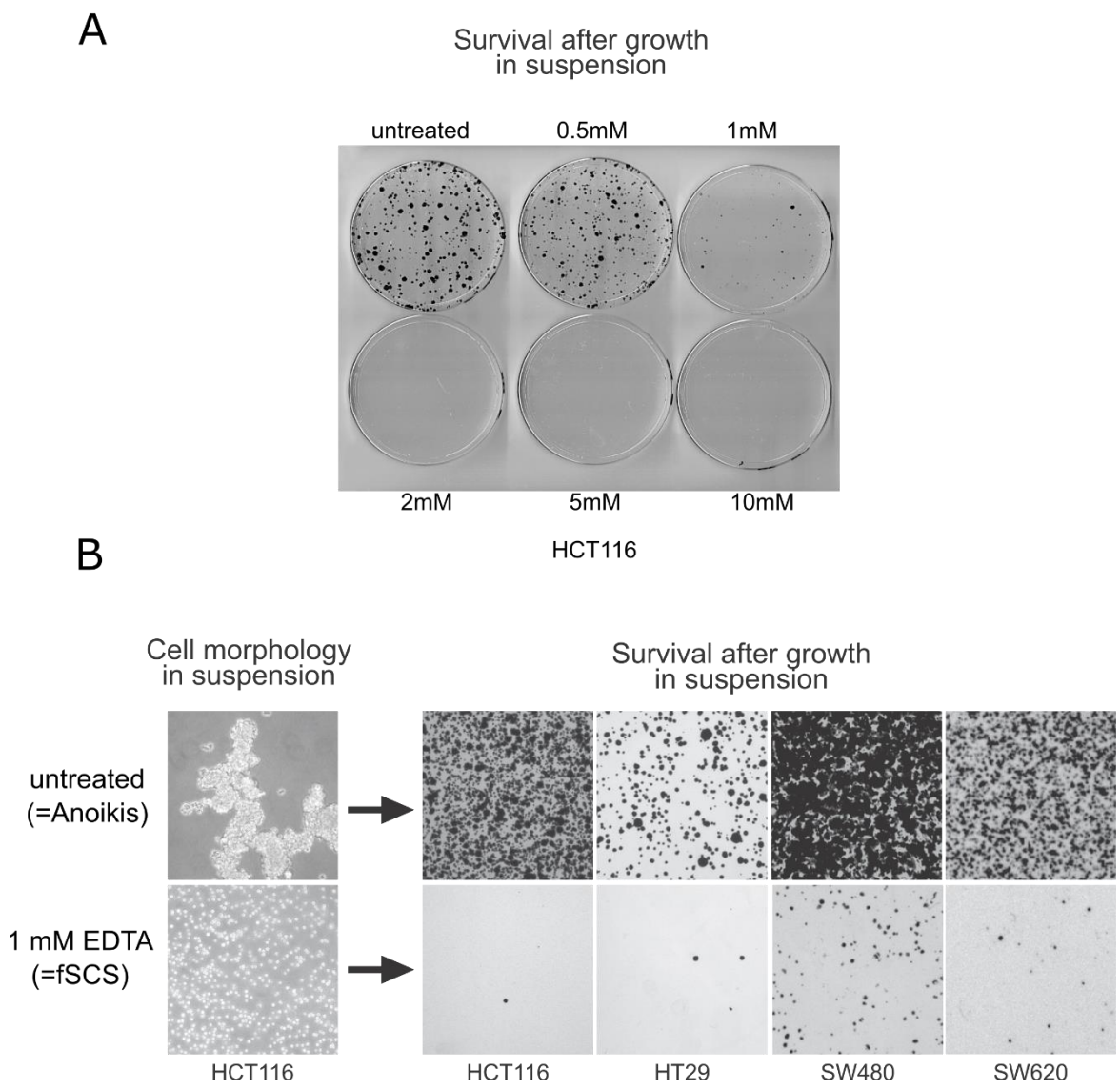


Figure 4.1: Comparison between *in vitro* anoikis and fSCS. A) HCT116 cells were incubated in low adherence conditions with different concentrations of EDTA (untreated, 0.5mM, 1mM, 2mM, 5mM and 10mM) in order to obtain a selective pressure; B) comparison between *in vitro* anoikis and fSCS with 1mM EDTA in several CRC cell lines.

Therefore, CRC cells were incubated on ultralow attachment plates in the presence of 1mM EDTA and, similarly to what previously described, 24-48 h later (depending on cell lines), we plated cells on culture dishes in complete growth medium, let them grow for additional 7-10 days and we stained surviving cell colonies with crystal violet. Interestingly, we noticed that: 1) compared to anoikis conditions, cells cultured on low adherence in the presence of 1 mM EDTA do not form any longer tight cell aggregates; 2) this assay, that we named *forced Single Cell Suspension* (fSCS), massively impaired cell survival without preventing the recovery of survived cells (fig.4.1B, lower part; untreated=anoikis; EDTA=fSCS). To further reinforce our hypothesis that cells survived to anoikis by strengthening cell-cell contacts, we also used other means (enzymatic, *i.e.* hyaluronidase treatment, or physical) to prevent cell contacts and keep cells separated Also in these conditions we observed a decrease in cell colony formation, similar to that obtained with the use of EDTA (data not shown).

To sum up, we set up a novel *in vitro* assay, namely fSCS, that stands on the combination of loss of cell-matrix and cell-cell contacts, and that is more stringent, compared to *in vitro* anoikis, in killing cells and selecting those with more aggressive features.

4.2 Forcing cell to a single cell suspension is a selective pressure for EMT features

To validate the pro-EMT selective pressure of fSCS, we verified whether EMT induction confers resistance to fSCS. EMT was obtained in HCT116 (MSI) and HT29 (MSS) by TWIST-ER overexpression and TWIST nuclear translocation by exposure to 4-hydroxy-tamoxifen (4OHT). Indeed, upon exposure to 4OHT, TWIST-ER overexpressing cells showed a more mesenchymal morphology and the expression of EMT markers, such as decreased levels of E-cadherin and increased levels of Slug and Vimentin (fig. 4.2A).

Thus, we performed fSCS in HCT116 and HT29 TWIST-ER cells exposed or not to 4OHT and we observed a greater survival in 4OHT treated cells compared to untreated cells in both cell lines (fig. 4.2B), indicating that TWIST, and therefore EMT, confers resistance to fSCS. This experiment proves that the EDTA addition is not toxic *per se* and that the fSCS negative selective pressure is not too strong to prevent the recovery of EMT cells.

Figure 4.2

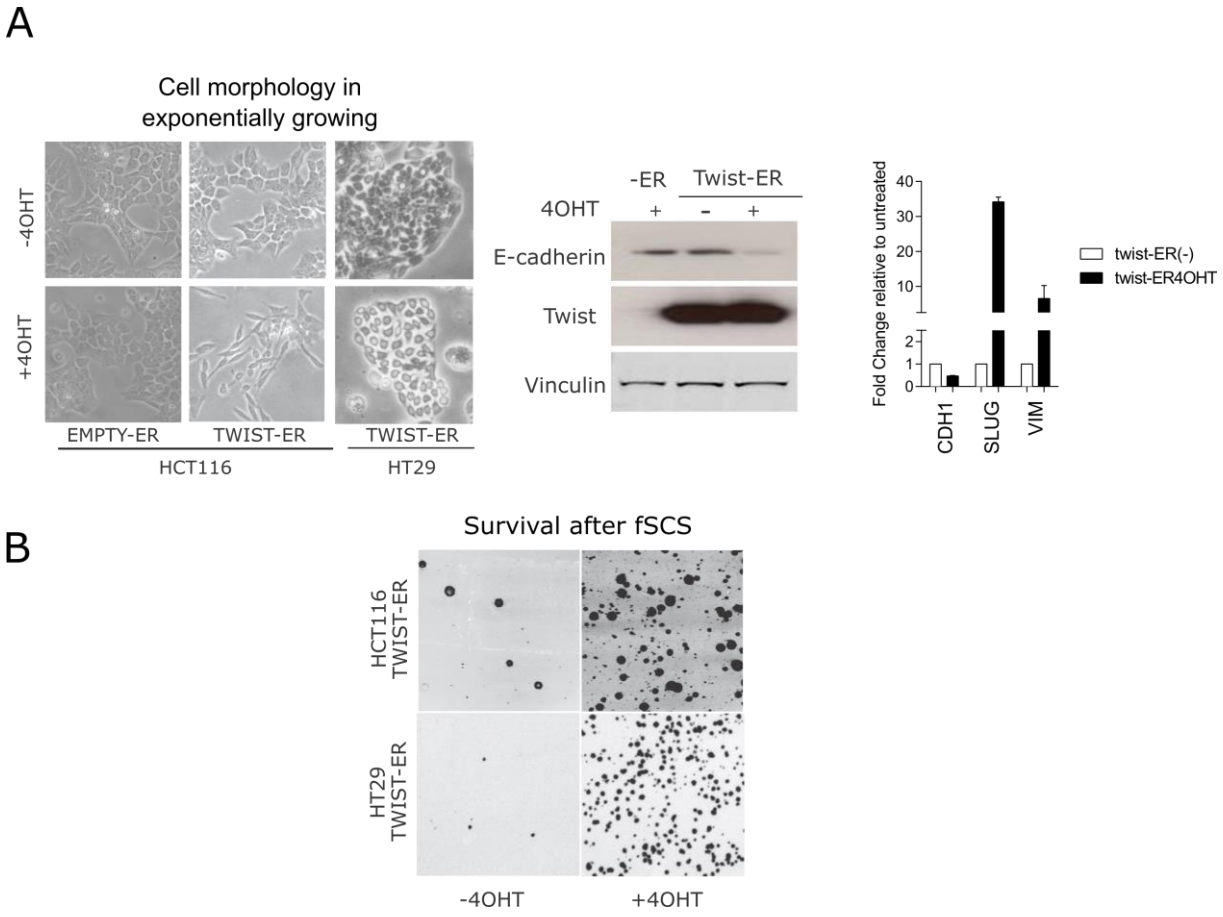


Figure 4.2: EMT cells gain resistance to fSCS. A) EMT induction in HCT116 and HT29 following overexpression of Twist-ER and treatment with 4OHT was demonstrated by mesenchymal morphology and expression of EMT markers; B) HCT116 and HT29 cells that underwent EMT gained resistance to fSCS.

4.3 A transposon-based forward genetic screen confers fSCS resistance by insertional mutagenesis

4.3a HCT116 mutagenesis by transposition

To investigate in an unbiased manner which genetic elements generate resistance to fSCS in CRC cell lines, we used the SB DNA transposon (TN) system (pT2BH vector carrying the TN and the pSB100 vector expressing the SB transposase)[194]. We adapted the pT2BH vector to carry the eGFP reporter coding sequence under the control of a CMV promoter (pT2-CMV-eGFP) (fig. 4.3): different from TNs traditionally used in genetic screens for gene trap, our modified SB TN does not include splicing acceptors,

splicing donors or polyA sites [176], thus it simply acts by insertional mutagenesis. In this manner, in the case of an intronic unknown ncRNA (see fig.9 in the introduction section of this Thesis), the TN will disrupt the sequence of the ncRNA where it inserts, but without exerting any effect on the surrounding PCG. After transfection of pT2-CMV-eGFP and pSB100 plasmids in HCT116 cell line, we expanded them, and after 14 days we sorted a pool of eGFP expressing cells by FACS (TN4_sorted). This was our initial pool of transposed cells demonstrated by the fact that eGFP expression was stable over time.

Figure 4.3

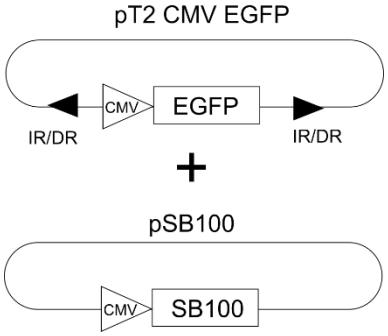


Figure 4.3: Transposition of HCT116 cells with a modified Sleeping Beauty transposon. HCT116 cells were co-transfected with a modified pT2BH (pT2-CMV-EGFP) transposon that contains two IR/DRs sequences, a CMV promoter and an EGFP expression cassette and with pSB100x transposase, that allows the transposition.

4.3b Functional selection by fSCS

Next, we used TN4_Sorted cells to perform fSCS assay and we collected single cell clones that survived to fSCS (see scheme depicted in figure 4.4A). Collected surviving clones (~100 per cell line) were challenged again for fSCS resistance. Only 4 displayed overt fSCS resistance, and among these, TN4_20 was the most strikingly resistant clone (Fig. 4.4B and 4.5A).

Figure 4.4

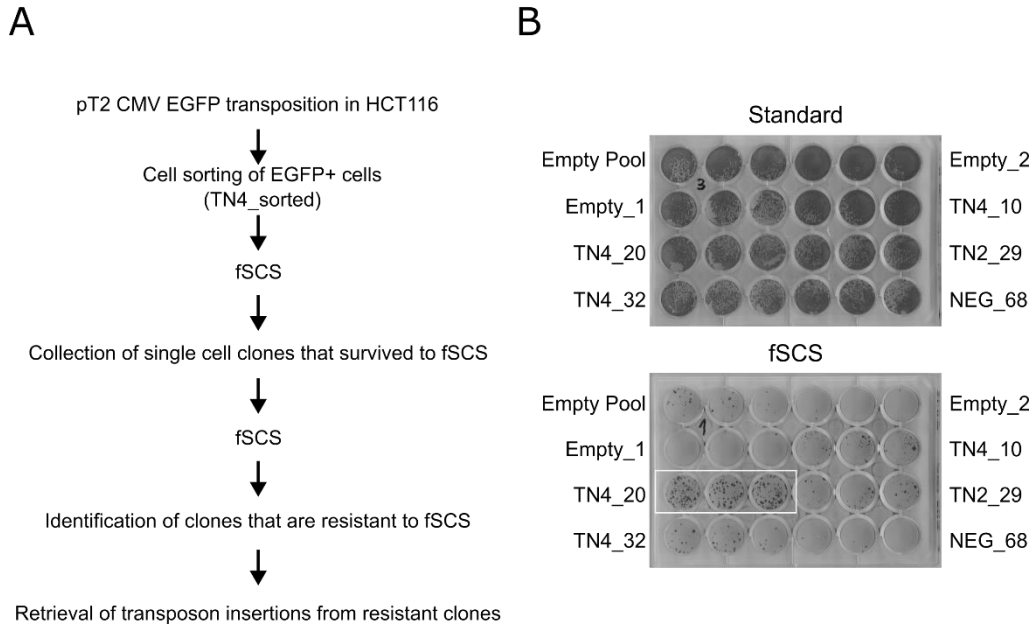


Figure 4.4: Functional selection by fSCS. A) overview scheme of the TN-f SCS based screening process; B) collected surviving clones after two rounds of fSCS

We further characterized TN4_20. From a morphological point of view, we observed that TN4_20 cells displayed a more mesenchymal shape (fig. 4.5A), with cells being more elongated and growing in a scattered fashion instead as compact and cuboidal. By qRT-PCR we detected that TN4_20 expressed some of the EMT markers (e.g. SLUG ↑, TWIST ↑, VIMENTIN ↑, HAS-2 ↑ (fig. 4.5B). We observed that E-cadherin reduction, another EMT trait, was not statistically significant by qRT-PCR (fig. 4.5B); yet, we observed a reduction of E-cadherin levels by Western Blotting, suggesting that E-cadherin regulation in TN4_20 occurs at a post-transcriptional level (fig. 4.5B).

In addition, by performing motility assay with 3D matrices cell inclusion, we observed that TN4_20 showed the ability to generate many more satellite colonies in matrigel evasion assay compared to TN4_Sorted cells (fig. 4.5C). Moreover, we have conducted a pilot *in vivo* experiment in mice to verify whether fSCS resistance was associated with increased metastatic potential. To this aim, we performed orthotopic intra-caecal injection of TN4_Sorted and TN4_20 cells. We observed that after 60 days, only TN4-20 injected mice developed distant metastasis to the lungs compared to TN4-Sorted injected mice that did not (2/2 vs 0/2) (fig 4.5D). All together, these results

indicate that the fSCS based selection process allowed us to identify a cell clone, TN4_20, that shows EMT and invasiveness traits and *in vivo* metastatic potential.

Figure 4.5

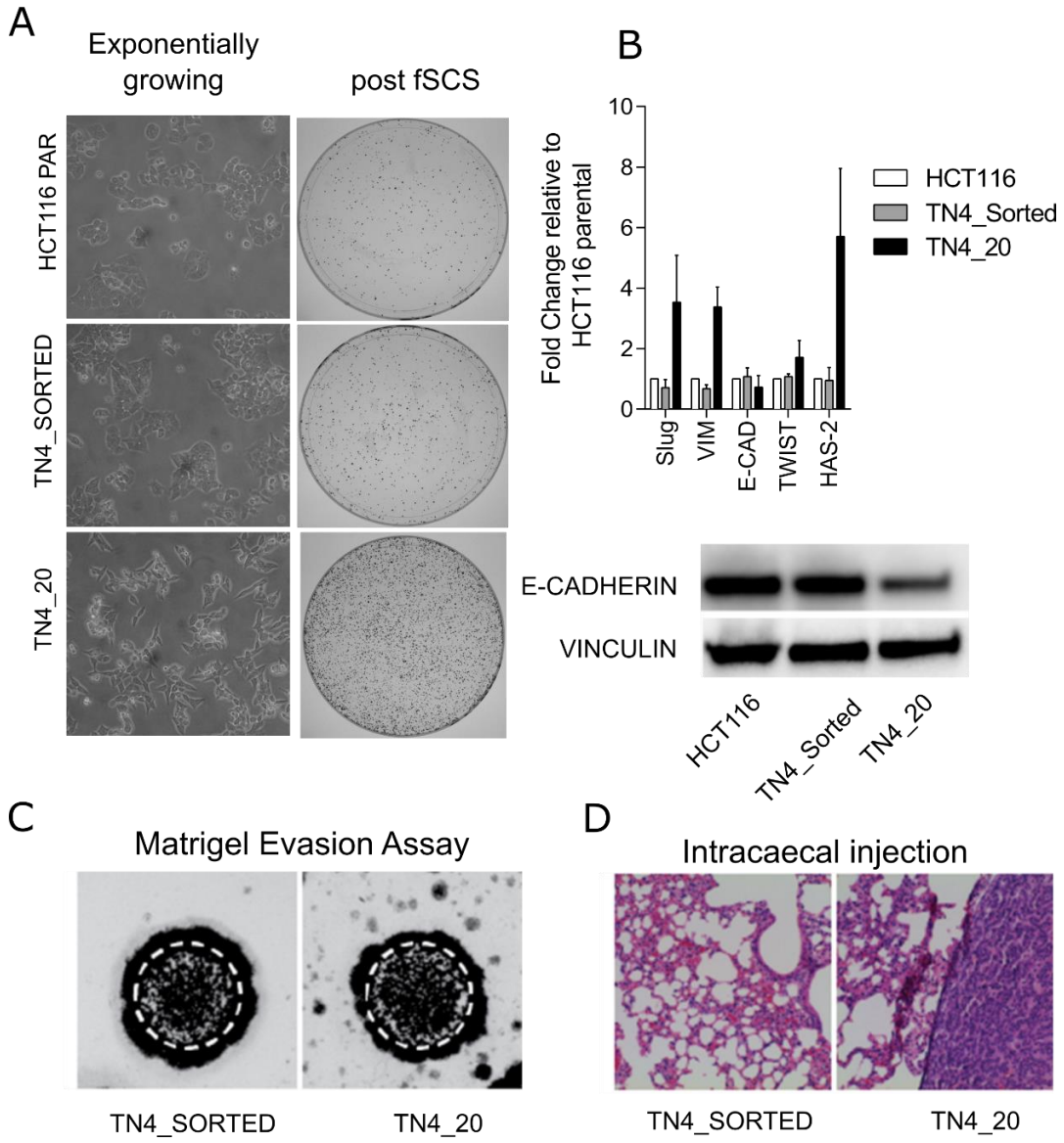


Figure 4.5: Characterization of TN4_20 clone. A) Left, TN4_20 clone displayed a more scattered and mesenchymal morphology compared to HCT116 parental and TN4_Sorted cells; A) Right, TN4_20 clone showed increased resistance to fSCS; B) TN4_20 clone expressed EMT markers (SLUG ↑, TWIST ↑, VIMENTIN ↑ E-CADHERIN ↓, HAS-2 ↑); C) and D) TN4_20 clone showed increased capability evade in Matrigel evasion assay and increased metastatic potential *in vivo*, respectively.

4.3c Fishing for the TN insertion responsible for fSCS resistance

To retrieve the genomic positions in which the pT2-CMV-eGFP TN had inserted in the TN4_20 clone, we used linker mediated PCR [215]. We identified seven distinct TN insertion sites, that were randomly distributed within the TN4_20 genome (Table 1). In fact, TNs were found in introns, in regulatory intergenic regions (i.e. promoter region) and in the 3'UTR region of a protein-coding gene, namely BTBD7 (BTB/POZ containing domain protein 7). Among the seven insertions found we decided to focus on the TN insertion located within the 3' UTR of BTBD7 for two reasons. First, BTBD7 is known to regulate salivary gland branching by EMT induction [217] as well as tumor invasiveness and metastasis in cancer [218][219]; second, the TN insertion inside the BTBD7 3'UTR located within the predicted target site of miR-23b, a known anti-metastatic miRNA [220][221].

Position (hg19)	Closest Transcripts and Biotype (Genecode V19)	Description
Chr13 :94,910,954-910,955	GPC6 (protein-coding)	Intronic
Chr1 :212,405,256-212,405,257	RP11 15I11 (lincRNA)	Intronic
Chr19 :3,501,100-3,501,101	DOHH (protein-coding)	Promoter
Chr1 :235,641,751-235,641,752	B3GALNT2 (protein-coding)	Intronic
Chr1 :45,885,179-45,885,180	TESK2 (protein-coding)	Intronic
Chr5 :12,473,909-12,473,910	Upstream: CNND2 (protein-coding); Downstream: CT49 (lincRNA)	Intergenic (with LINE L1PREC2)
Chr14 :93,705,840-93,705,841	BTBD7 (protein-coding)	3'UTR

Table 1: SB TN insertion sites retrieved in TN4_20 genome by linker mediated PCR

We confirmed by PCR that TN is located in the 3'UTR of BTBD7 gene (fig. 4.6).

Figure 4.6

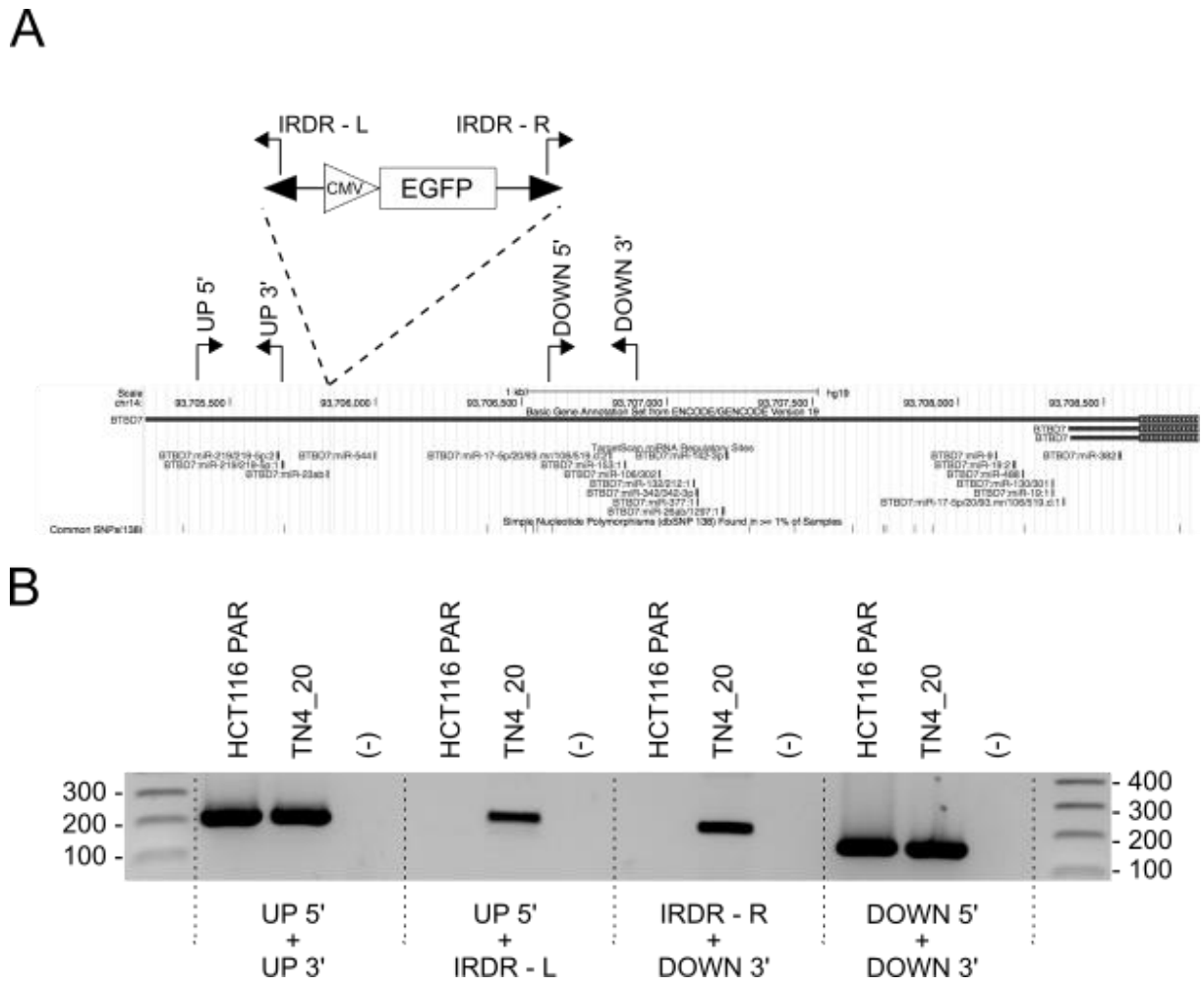


Figure 4.6: Retrieval of TN insertion in the 3'UTR of BTBD7 gene. A) Schematic representation of the exact position of pT2-CMV-EGFP TN insertion in the 3'UTR of BTBD7 gene; B) pT2-CMV-EGFP TN insertion in the 3'UTR of BTBD7 was confirmed by PCR: as expected, primers "Up 5'+Up 3' " and "Down 5'+Down 3' ", which were designed to recognize genomic DNA out of TN insertion (see A), worked in both HCT116 and TN4_20; whereas, primers "Up 5'+IRDR-L" and "IRDR-R+ Down 3' ", which were designed recognize TN IR/DRs left and right, respectively (see B), worked only in TN4_20.

Figure 4.7

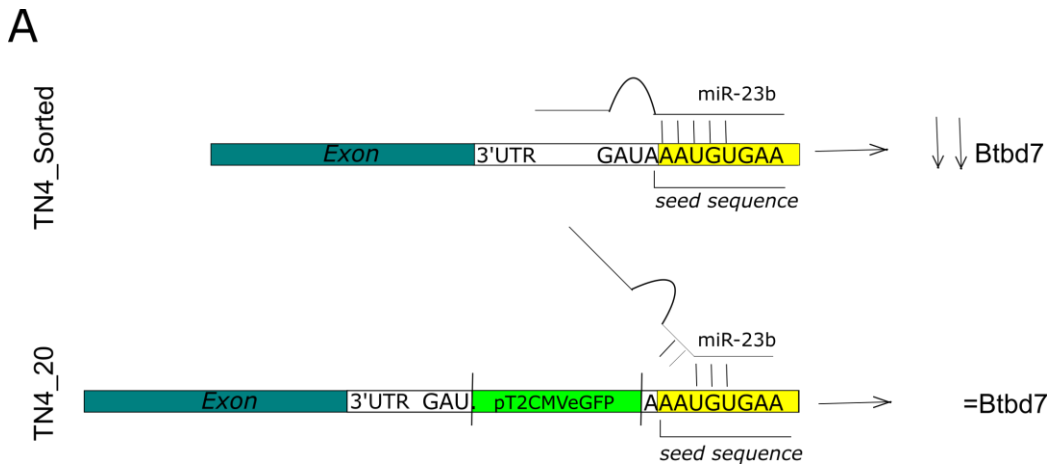


Figure 4.7 A: Hypothetic model. Upper part (=TN4_Sorted): miR-23b downregulates Btbd7 levels by interacting with its seed sequence in the 3' UTR of BTBD7 gene; Lower part (=TN4_20): TN insertion in the 3'UTR of BTBD7 impairs miR-23b capability to interact with its seed sequence, and Btbd7 levels remain unchanged.

Based on these notions, we hypothesized that the TN4_20 clone gained fSCS resistance because the TN insertion that occurred within BTBD7 3'UTR disrupted the interaction between miR-23b and BTBD7, leading to BTBD7 deregulation (fig. 4.7A).

Consistently, we checked Btbd7 protein expression levels in TN4_20. Supporting our hypothesis (see fig 4.7A), we found increased Btbd7 protein levels in TN4_20 cells compared to TN4_Sorted cells; yet, BTBD7 mRNA levels were not affected (fig. 4.7B).

Figure 4.7

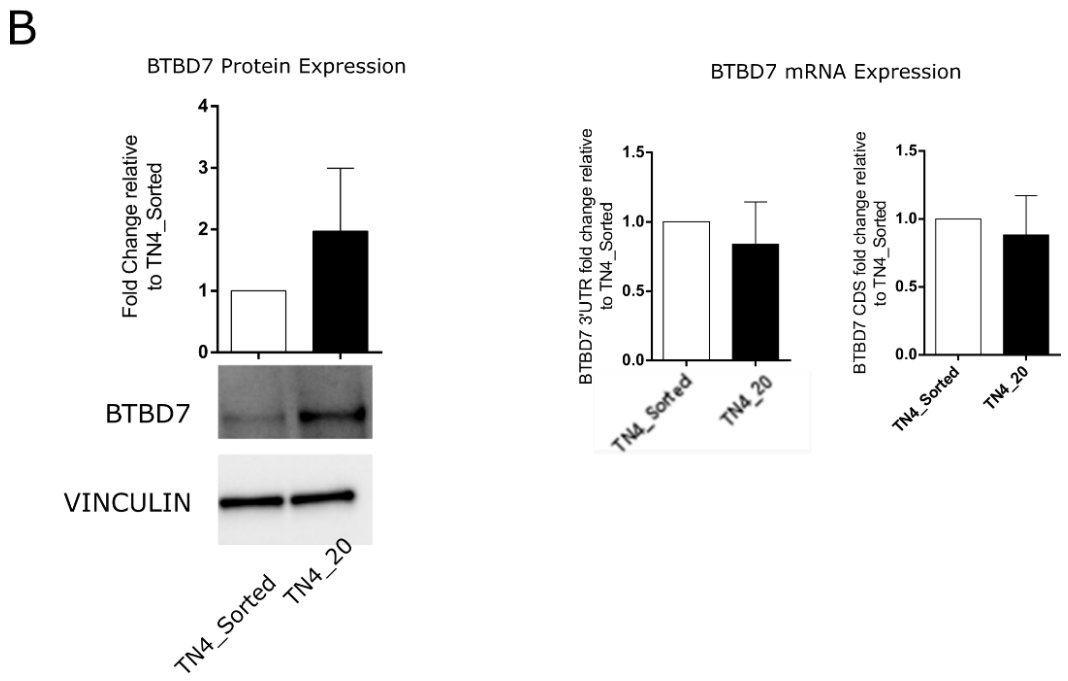


Figure 4.7B: Btbd7 protein levels increased in TN4_20. (Left), confirming the hypothesis depicted in fig. 4.7 A, Btbd7 protein levels were increased in TN4_20; (Right) BTBD7 mRNA levels were not affected.

4.4 miR-23b targets BTBD7 gene and this interaction is important in fSCS resistance

To further prove that miR-23b targets BTBD7 3'UTR, we transfected synthetic miR-23b precursor in HCT116 parental cells and observed downregulation of Btbd7 protein levels only in miR-23b transfected cells (fig. 4.8A). Then, to confirm the hypothesis that TN insertion within the 3'UTR of BTBD7 disrupted the interaction between miR-23b and BTBD7 (see fig. 4.7A), we transfected miR-23b also in TN4_20 cells. Consistently, we observed that miR-23b overexpression did not affect Btbd7 expression levels in TN4_20 (fig.4.8B).

Figure 4.8

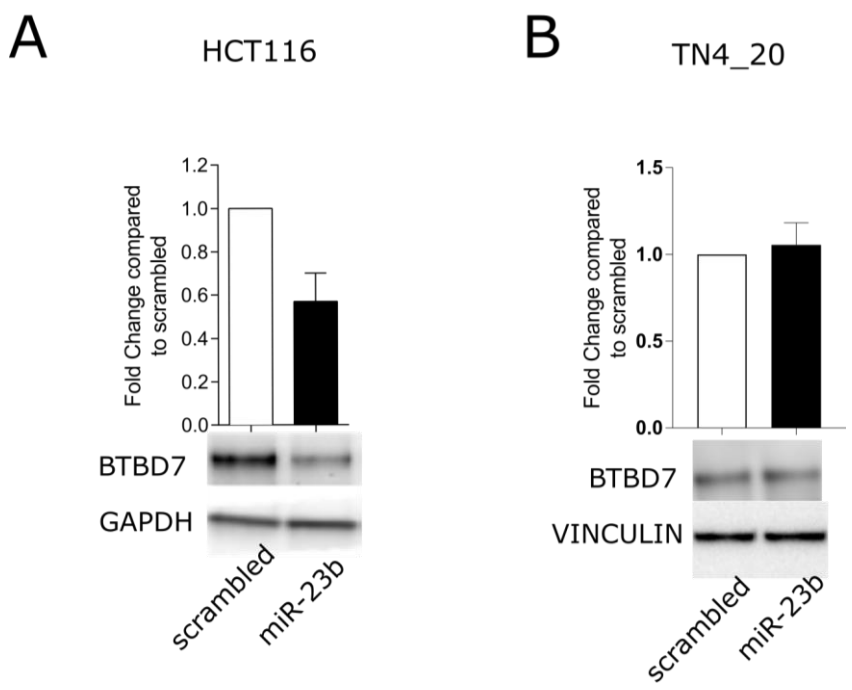


Figure 4.8 A and B: miR-23b targets BTBD7 gene and negatively regulates Btbd7 expression levels. A) Confirming the hypothesis in fig 4.7A, overexpression of miR-23b in HCT116 decreased Btbd7 protein levels; B) Overexpression of miR-23b in TN4_20 did not affect Btbd7 protein levels.

Next, to investigate the relevance of the interaction between BTBD7 and miR-23b in fSCS phenotype, we transfected miR-23b precursor in HCT116 parental, TN4_sorted and TN4_20 cells and performed fSCS assay. We observed that miR-23b prevented fSCS survival in HCT116 parental and TN_Sorted but not in TN4_20 cells (fig 4.8C). Additionally, we found that miR-23b prevented fSCS resistance also in other CRC cell lines (HT29, SW480) (4.8D).

Figure 4.8

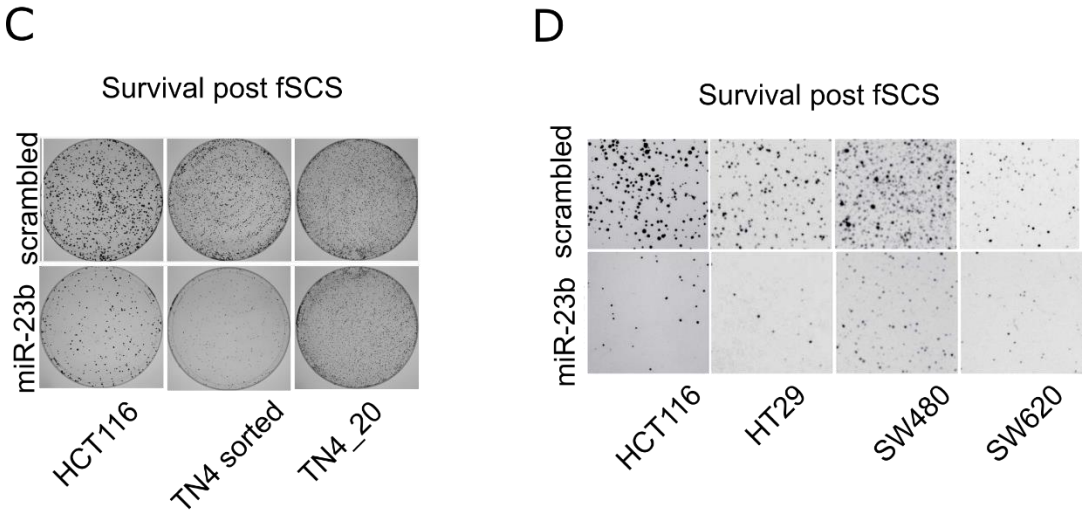


Figure 4.8 C and D: The interaction between Btbd7 and miR-23b is relevant for fSCS resistance. C) miR-23b overexpression affected fSCS resistance in HCT116 and TN4_Sorted but not in TN4_20; D) miR-23b overexpression affected fSCS resistance in several CRC cell lines.

Thus, these data suggest that miR-23b targets BTBD7 by negatively regulating Btbd7 protein levels and that the interaction between miR-23b and Btbd7 is important for fSCS resistance. In agreement, we showed that miR-23b overexpression did not impair either Btbd7 expression levels or fSCS resistance in TN4_20 cells, in which this interaction is putatively abolished due to TN insertion.

4.5 Btbd7 expression is necessary for fSCS resistance

To investigate whether Btbd7 expression is important for fSCS survival, we silenced BTBD7 by using Sh-RNA (short hairpin RNA). First, we tested five different Sh-RNAs directed against BTBD7 gene (Sh- BTBD7_1, Sh-BTBD7_2, Sh-BTBD7_3, Sh-BTBD7_4, Sh-BTBD7_5) in HCT116 parental cells. We observed that all the tested Sh-RNAs are effective in silencing Btbd7 expression both at the mRNA and at the protein level (fig. 4.9A), and we chose to use Sh-BTBD7_1 for further experiments.

Figure 4.9

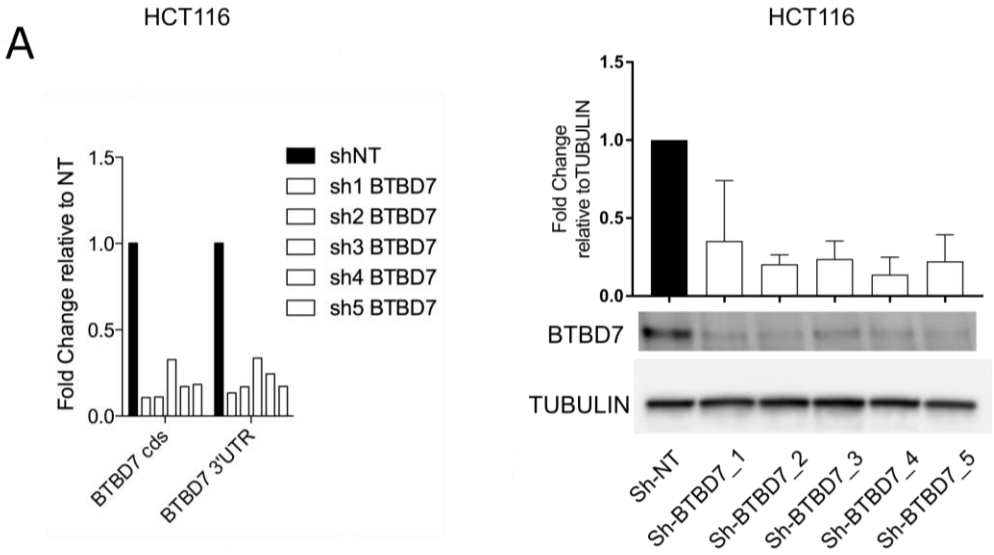


Figure 4.9 A: BTBD7 silencing in HCT116. Five different Sh-RNAs against BTBD7 were tested in HCT116 and validated by qRT-PCR (left) by using primers targeting BTBD7 3'UTR or BTBD7 CDS (coding sequence) and by Western Blotting (right).

Then, we transduced HCT116 parental and TN4_20 cells with Sh-BTBD7_1 and performed fSCS assay. We observed that the depletion of Btbd7 impairs the fSCS resistance in both HCT116 parental and TN4_20 cells (fig 4.9B), indicating that Btbd7 expression is a necessary for fSCS survival.

Figure 4.9

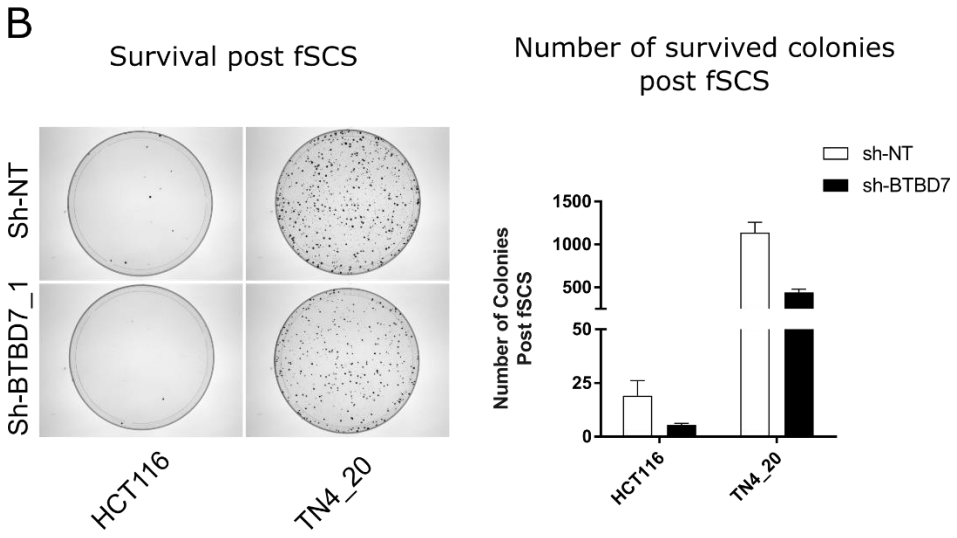


Figure 4.9 B: Btbd7 expression is necessary for fSCS resistance. (Left), silencing of Btbd7 with Sh-BTBD7_1 reduced fSCS resistance of HCT116 and TN4_20; (Right), quantification of fSCS survived colonies.

4.6 Btbd7 overexpression induces fSCS resistance, EMT and invasiveness

To further assess the importance of Btbd7 in fSCS resistance and in EMT phenotype, we generated HCT116 clones stably overexpressing ectopic eGFP-Btbd7 (fig. 4.10A, eGFP-Btbd7clone#1, eGFP-Btbd7 clone#2 and eGFP-Btbd7 clone#3 were compared with empty clone #1, empty clone #2, empty clone #3). Interestingly, we observed that ectopic eGFP-Btdb7 shows a diffuse cytoplasmic localization; moreover, consistent with [217], we noticed the absence of eGFP-Btbd7 from the nucleus (fig.4.10B).

Figure 4.10

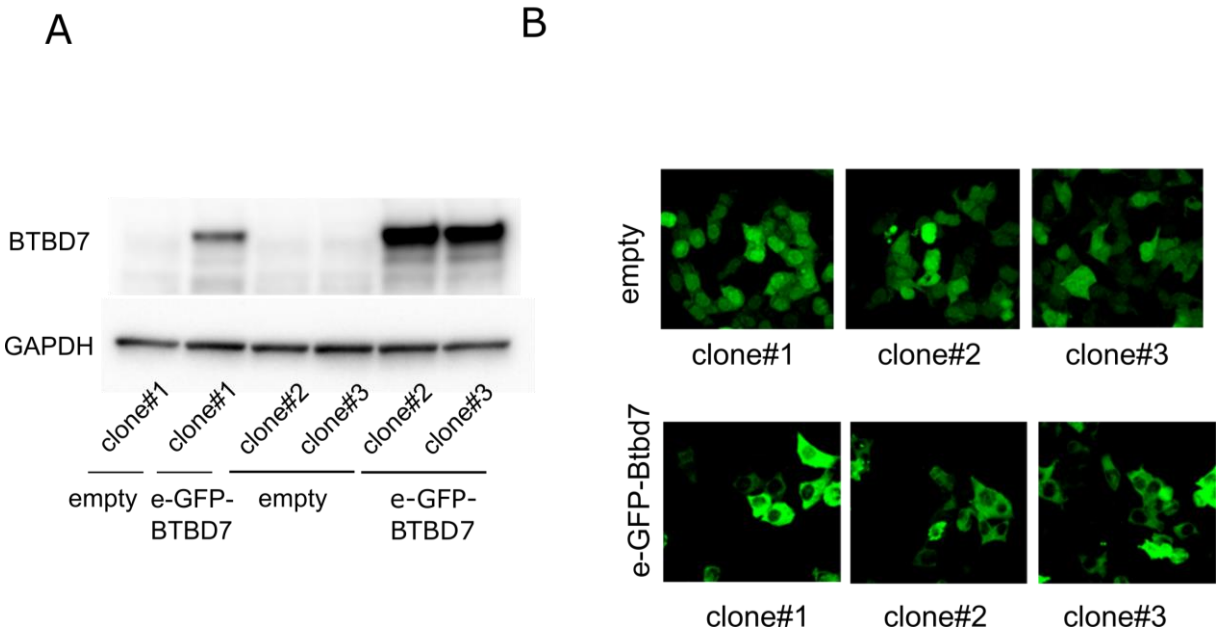


Figure 4.10 A and B Overexpression of ectopic Btbd7. A) Three clones stably overexpressing ectopic eGFP-Btbd7 protein were obtained in HCT116 (eGFP-Btbd7 clone#1, eGFP-Btbd7 clone#2, eGFP-Btbd7 clone #3); B) Localization of ectopic eGFP-Btbd7 protein in eGFP-Btbd7 clones compared to empty clones.

Next, we asked whether Btbd7 overexpression has an effect on fSCS survival and cell invasiveness. To this aim, we subjected eGFP-Btbd7 overexpressing cells to the fSCS assay and to matrigel evasion assay, and we observed increased fSCS resistance and increased matrigel evasion capability in these cells compared to control (fig. 4.10C).

Figure 4.10

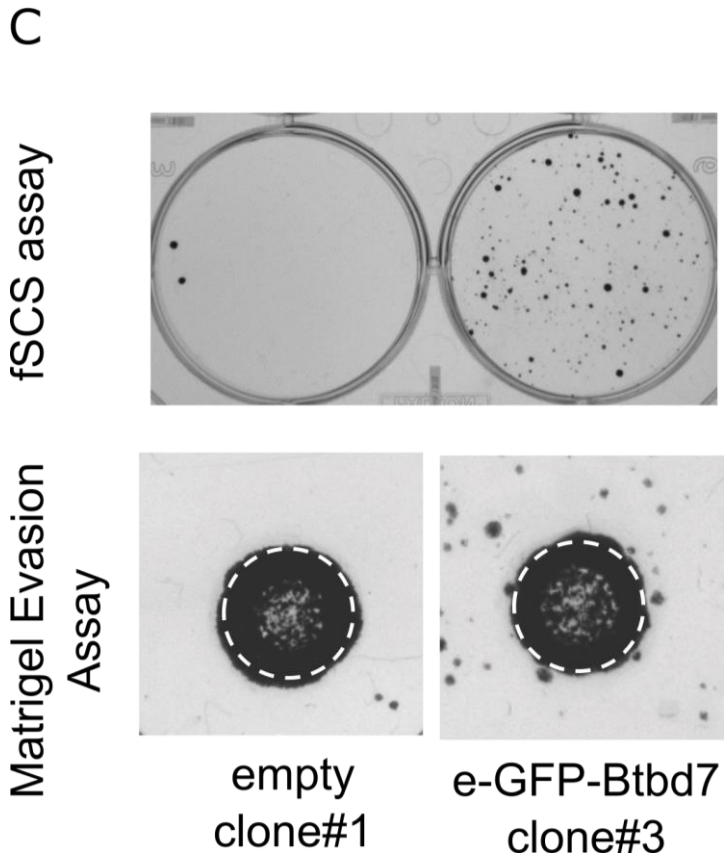


Figure 4.10C Overexpression of ectopic Btbd7 induced fSCS resistance C) eGFP-Btbd7 clone #3 showed increased fSCS resistance and Matrigel evasion capability compared to empty clone #1.

As already mentioned, Btbd7 has been found to regulate EMT and metastasis [217][218][219], and our data are consistent with these findings. However, the mechanism of action of Btbd7 has never been described. To shed light on Btbd7 molecular function, we investigated which are the genes related to Btbd7

overexpression. We analyzed by qRT-PCR a panel of EMT genes in eGFP-Btbd7 overexpressing cells and, consistent with literature observations, we observed E-cadherin downregulation and Vimentin upregulation (fig. 4.11A). Moreover, we noticed a significant upregulation of Zeb-1 transcription factor, that has never been described before, but not of Twist or Slug transcription factors (fig. 4.11A).

Figure 4.11

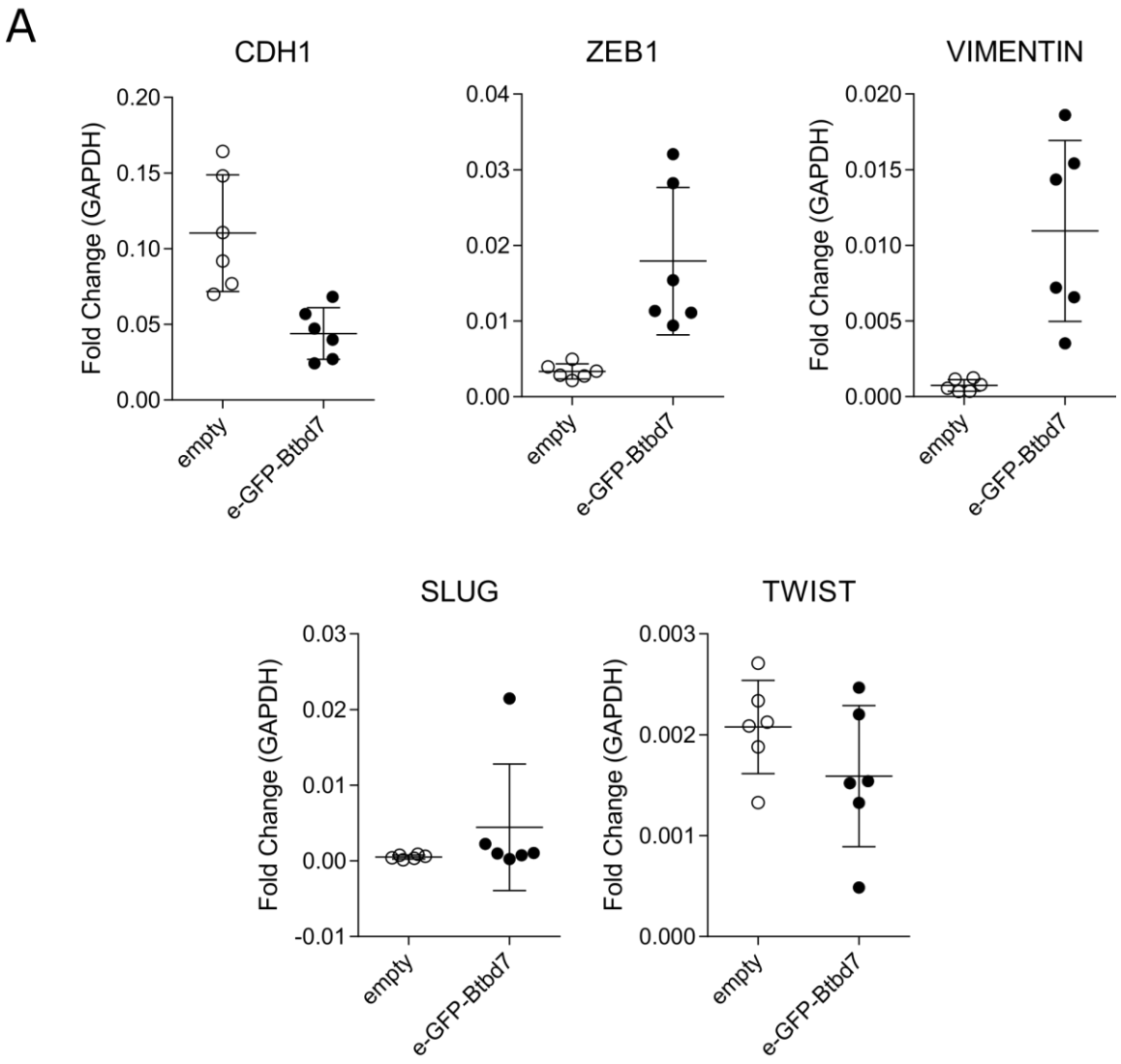


Figure 4.11 Btbd7 overexpression induced the expression of EMT proteins. A) eGFP-Btbd7 clones express EMT markers by q-RT-PCR.

Also at the protein level we found a decrease of E-cadherin and an increase of Zeb-1 levels (fig. 4.11B). These data indicate that Btbd7 may induce EMT and fSCS resistance by regulating E-cadherin and Zeb-1 expression levels. Moreover, the observation that the changes of E-cadherin and Zeb-1 protein levels are accompanied by changes of their mRNA levels suggests that Btbd7 could regulate the expression of E-cadherin and Zeb-1 at a transcription level. In contrast to this view, however, we observed that ectopic eGFP-Btbd7 does not localize to the nucleus (see fig.4.10B), suggesting its inability to act directly as a transcriptional gene regulator. This suggests the existence of intermediate interactors or of downstream effectors of Btbd7 contributing to the regulation of EMT genes expression. In alternative, it is possible that Btbd7 induces EMT by acting first on protein levels; then, given that EMT is capable of auto-maintaining, Zeb-1 and E-cadherin mRNAs levels may change due to the induction of EMT transcriptional program.

Figure 4.11

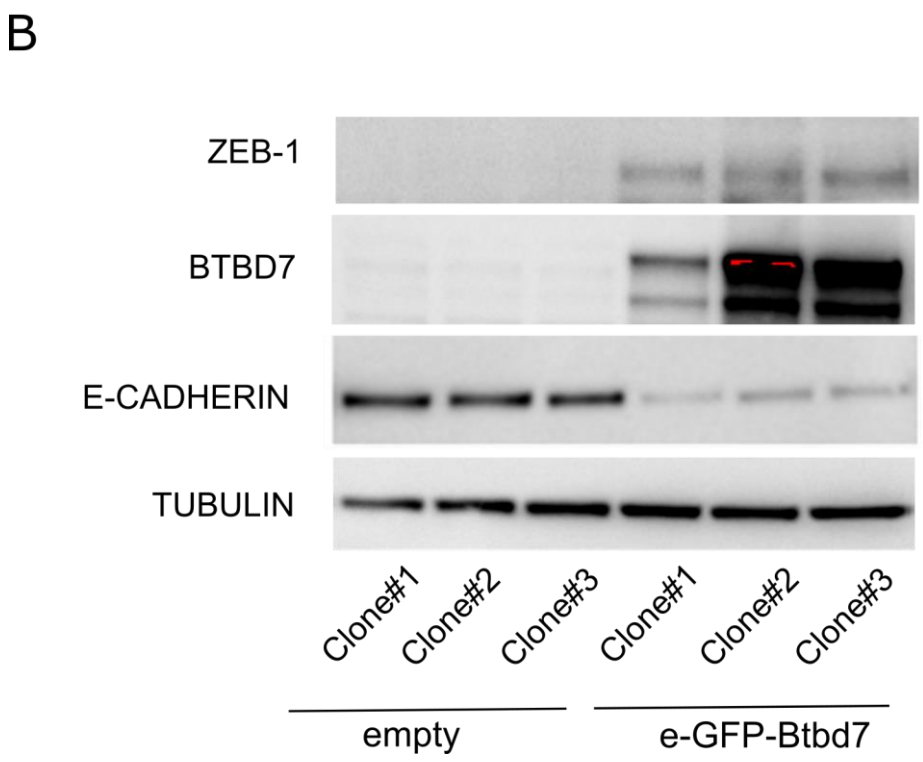


Figure 4.11 Btbd7 overexpression induced the expression of EMT proteins. B) Western Blotting analysis of Zeb-1 and E-caderin proteins in eGFP-Btbd7 clones.

4.7 Bioinformatic analysis of Btbd7 protein: a help to shed light on Btbd7 function?

To explore more in detail Btbd7 molecular function, we took advantage of bioinformatic tools and publicly available protein databases. First, to investigate the relevance of Btbd7 biological function, we analyzed its degree of conservation across species. To this aim, we performed a local alignment of Btbd7 in a *non-redundant protein sequences* database by using BLASTp. We found that Btbd7 protein shows more than 95% of alignment identity (E-value 0.0), not only in Primates, such as *Gorilla gorilla gorilla, Pan paniscus,* and *Pan troglodytes*, but also other Mammals, such as *Camelus bactrianus* and *Canis lupus familiaris*, indicating that the protein is conserved between species (fig 4.12A). Then, we repeated the local alignment in BLASTp by introducing the "*RefSeq protein-Homo sapiens"* restriction, in order to align Btbd7 exclusively with human proteins. With this approach, we were able to identify two isoforms of Btbd7, namely "Btbd7 isoform 2" and "Btbd7 isoform 3", that show 99% of alignment identity with Btbd7 (E-value 0.0). These alternative isoforms, that in UniProt publicly available database are referred to as "Btbd7 isoform 3" and "Btbd7 isoform 5", respectively, result to be shorter, since they lack some protein regions respect to the main isoform (fig. 4.12B, upper part). Interestingly, this alignment also identified several human proteins that, despite having less than 50% of alignment identity with Btbd7, share the same BTB/POZ and BTB/Kelch domains of Btbd7 (fig 4.12B, lower part), indicating that BTB/POZ and BTB/Kelch domains are conserved domains in humans. Consistently, the use of other bioinformatic tools for the study of protein structure, such as Pfam and PROSITE, confirmed that Btbd7 possesses 2 BTB/POZ domains and 1 BTB/Kelch domain (fig 4.12D). The BTB (BR-C, ttk and bab)/POZ (pox virus and zinc finger) domain, is present near the N-terminus of zinc finger proteins, in proteins that contain the Kelch domain, and in pox virus proteins. Interestingly, this domain mediates BTB/POZ homomeric and heteromeric dimerization, and in some zinc finger proteins, mediates transcriptional repression [222]. The BTB/Kelch domain, also known as BACK domain (BTB And C-terminal Kelch) is located C-terminal to a BTB domain and N-terminal to a Kelch domain in several Kelch and Kelch like proteins. Importantly, together with the BTB/POZ domain, it can participate in dimerization and interaction with other proteins, such as Cullin 3 (Cul 3) E3 ubiquitin ligase complex and ubiquitinate target proteins [223]. Actually, we noted that publicly available UniProt database (ID Q9P203) indicates

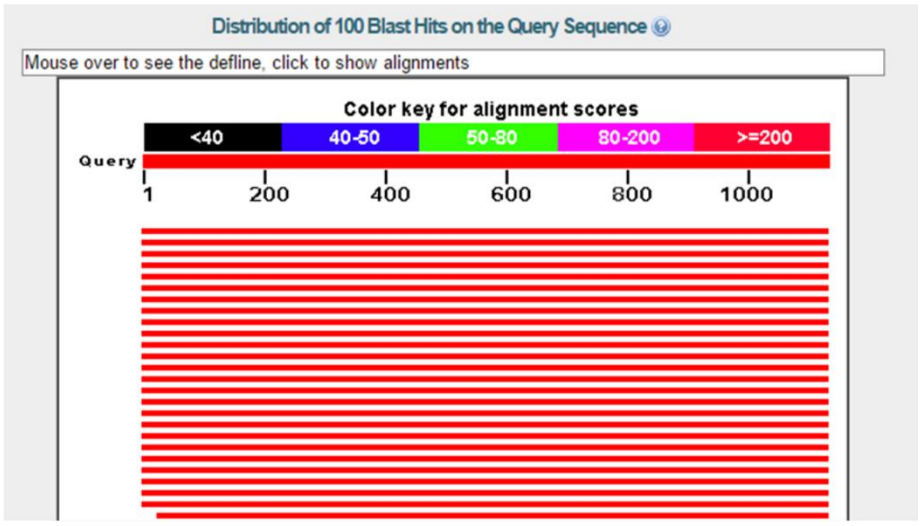
Cul 3 as a potential interactor of Btbd7; further analyses are required to investigate if Btbd7 interacts with Cul 3 (i.e. co-immunoprecipitation assays).

Thus, these results indicate that Btbd7 biological function is relevant, since it presents a remarkable degree of conservation not only in Primates, but also in other Mammals; moreover, a further analysis revealed that Btbd7 possesses two distinct domains, namely BTB/POZ and BTB/Kelch, that are conserved in humans and that, in association with each other, may be involved in the interaction with Cul 3, a E3 ubiquitin ligase, thus opening a new road to unravel the mechanism of action of Btbd7.

Figure 4.12

A

Aligment of Btbd7 protein in a non redundant protein database with BlasP



Description	Max score	Total score	Query cover	E value	Ident	Accession
BTB/POZ domain-containing protein 7 isoform 1 [Homo sapiens]	2353	2353	100%	0.0	100%	NP_001002860.2
KIAA1525 protein [Homo sapiens]	2352	2352	100%	0.0	100%	BAA96049.2
PREDICTED: BTB/POZ domain-containing protein 7 isoform 1 [Gorilla gorilla gorilla]	2344	2344	100%	0.0	99%	XP_004055649.1
PREDICTED: LOW QUALITY PROTEIN: BTB/POZ domain-containing protein 7 [Pan paniscus]	2340	2340	100%	0.0	99%	XP_008956470.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Pan troglodytes]	2339	2339	100%	0.0	99%	XP_001150552.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Rhinopithecus roxellana]	2332	2332	100%	0.0	99%	XP_010382815.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Mandrillus leucophaeus]	2330	2330	100%	0.0	99%	XP_011847128.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Cercocebus atys]	2330	2330	100%	0.0	99%	XP_011939143.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Nomascus leucogenys]	2329	2329	100%	0.0	99%	XP_003260957.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Pongo abelii]	2329	2329	100%	0.0	99%	XP_002825096.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Chlorocebus sabaeus]	2328	2328	100%	0.0	99%	XP_007985845.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Papio anubis]	2327	2327	100%	0.0	99%	XP_009210422.1
PREDICTED: BTB/POZ domain-containing protein 7 [Macaca fascicularis]	2327	2327	100%	0.0	99%	XP_005562123.1
hypothetical protein EGK_18484 [Macaca mulatta]	2319	2319	100%	0.0	99%	EHH28128.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Saimiri boliviensis boliviensis]	2314	2314	100%	0.0	99%	XP_010348849.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Callithrix jacchus]	2306	2306	100%	0.0	98%	XP_009004722.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Vicugna pacos]	2272	2272	100%	0.0	97%	XP_006208283.1
PREDICTED: BTB/POZ domain-containing protein 7 [Ceratotherium simum simum]	2265	2265	100%	0.0	96%	XP_004434279.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Camelus ferus]	2264	2264	100%	0.0	96%	XP_006184240.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Galeopterus variegatus]	2263	2263	100%	0.0	96%	XP_008593159.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Camelus bactrianus]	2263	2263	100%	0.0	96%	XP_010967020.1
PREDICTED: BTB/POZ domain-containing protein 7 isoform X1 [Canis lupus familiaris]	2262	2262	100%	0.0	96%	XP_005623806.1

Figure 4.12 A Alignement of Btbd7 protein in a non-reduntant protein database. Btbd7 shows a remarkable degree of conservation between species (Primates and other Mammals), suggesting the relevance of its function.

Figure 4.12

B

Alignment of Btbd7 protein in a non redundant protein database_
refseq restrinction

Description	Max score	Total score	Query cover	E value	Ident	Accession
BTB/POZ domain-containing protein 7 isoform 1 [Homo sapiens]	2353	2353	100%	0.0	100%	NP_001002860.2
BTB/POZ domain-containing protein 7 isoform 3 [Homo sapiens]	1544	1544	65%	0.0	99%	NP_001276062.1
BTB/POZ domain-containing protein 7 isoform 2 [Homo sapiens]	813	813	34%	0.0	100%	NP_060637.1
kelch repeat and BTB domain-containing protein 3 [Homo sapiens]	59.7	97.4	25%	3e-08	25%	NP_689646.2
PREDICTED: kelch repeat and BTB domain-containing protein 3 isoform X2 [Homo sapiens]	58.5	58.5	10%	6e-08	31%	XP_011540921.1
BTB/POZ domain-containing protein 9 isoform a [Homo sapiens]	55.5	89.7	14%	6e-07	33%	NP_443125.1
PREDICTED: BTB/POZ domain-containing protein 9 isoform X4 [Homo sapiens]	54.7	89.0	14%	8e-07	33%	XP_011512584.1
PREDICTED: BTB/POZ domain-containing protein 9 isoform X3 [Homo sapiens]	54.3	88.2	14%	9e-07	33%	XP_011512583.1
PREDICTED: BTB/POZ domain-containing protein 9 isoform X2 [Homo sapiens]	54.3	88.2	14%	1e-06	33%	XP_011512582.1
BTB/POZ domain-containing protein 8 [Homo sapiens]	50.4	50.4	9%	1e-05	29%	NP_899065.2
PREDICTED: speckle-type POZ protein-like isoform X2 [Homo sapiens]	48.1	48.1	7%	6e-05	30%	XP_005263714.1
speckle-type POZ protein-like [Homo sapiens]	48.5	48.5	7%	6e-05	30%	NP_001001664.1
PREDICTED: speckle-type POZ protein-like isoform X1 [Homo sapiens]	47.4	47.4	7%	1e-04	30%	XP_005263712.1
PREDICTED: kelch-like protein 40 isoform X1 [Homo sapiens]	47.4	47.4	11%	2e-04	25%	XP_005264923.1
kelch-like protein 40 [Homo sapiens]	47.4	47.4	11%	2e-04	25%	NP_689606.2
zinc finger and BTB domain-containing protein 24 isoform 2 [Homo sapiens]	44.7	44.7	6%	9e-04	29%	NP_001157785.1
kelch repeat and BTB domain-containing protein 2 [Homo sapiens]	44.3	82.8	27%	0.002	26%	NP_056298.2
kelch-like protein 24 [Homo sapiens]	43.9	43.9	7%	0.002	30%	NP_060114.2
zinc finger and BTB domain-containing protein 24 isoform 1 [Homo sapiens]	43.9	43.9	6%	0.002	29%	NP_055612.2
kelch-like protein 38 [Homo sapiens]	43.5	43.5	6%	0.003	30%	NP_001075144.2
PREDICTED: actin-binding protein IPP isoform X1 [Homo sapiens]	42.7	77.8	25%	0.005	30%	XP_006710685.1
kelch-like protein 41 [Homo sapiens]	42.7	42.7	9%	0.005	25%	NP_006054.2

C

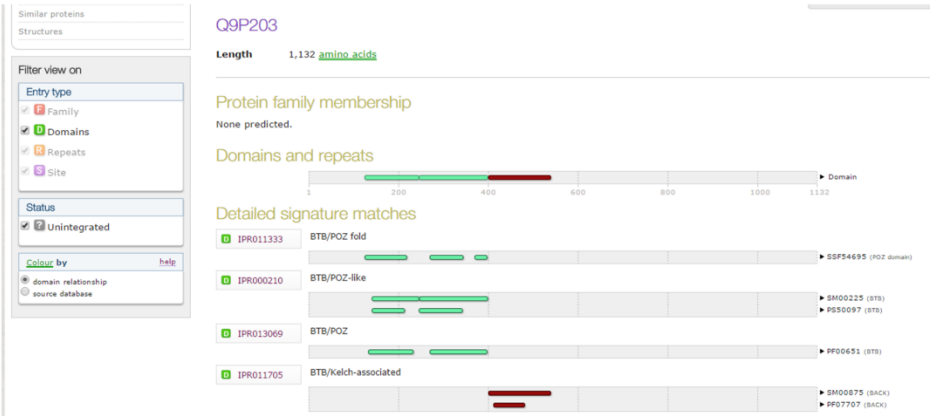


Figure 4.12 B and C: Btbd7 protein possesses BTB/POZ and BTB/Kelch conserved domains. B) Alignment of Btbd7 protein with Refseq restriction indicated that BTB/POZ and BTB/Kelch domains of Btbd7 are conserved domains in human proteins; C) Btbd7 possesses 2 BTB/POZ and 1 BTB/Kelch domains.

To obtain information on Btbd7 post-translational modifications, we consulted UniProt database (ID=Q9P203). We noted that Btbd7 protein presents a post-translational modification that consists in a lipidation, in particular N-myristilation, on the glycine at aa position 2 (data not shown). This post-translational modification has been described to target proteins to lipid rafts [224], that are membrane sub-domains rich in cholesterol and sphingolipids; interestingly, lipid rafts have been found to regulate EMT and tumor plasticity [225]. Moreover, Caveolin-1, a marker of caveolae, i.e. particular subtypes of lipid rafts, has an important role in the suppression of EMT [226]. Thus, we asked if Btbd7 function is related to lipid rafts, with particular attention to Caveolin-1 protein. First, to assess if Caveolin-1 negatively correlates with EMT, we checked its expression levels in TN4_20 clone, and we observed that it is decreased compared to control cells (fig. 4.13A). Then, in a preliminary experiment, we silenced Caveolin-1 in HCT116 parental cells; we observed that, especially in HCT116 Sh-Cav_2, cells acquired a scattered and mesenchymal morphology; moreover, we observed that Caveolin-1 silencing induces a decrease of E-cadherin protein levels, suggesting that the loss of Caveolin-1 may contribute to the induction of EMT phenotype (fig. 4.13B). Then, to investigate if there is a correlation between Caveolin-1 and Btbd7 expression, we first checked Caveolin-1 expression levels in HCT116 Sh-BTBD7_1 cells, and we found that upon Btbd7 silencing, Caveolin-1 increased (fig. 4.13C). Then, we checked Caveolin-1 expression levels in miR-23b overexpressing cells and we observed that, when Btbd7 is downregulated by miR-23b, Caveolin-1 is increased (fig. 4.13D), thus suggesting an inverse correlation between Btbd7 and Caveolin-1 expression. Given these results, we hypothesized that Caveolin-1 interacts with Btbd7 by negatively regulating its function. As for Cul3, a co-immunoprecipitation assay is required to assess if Btbd7 and Caveolin-1 actually interact.

To sum up, the use of the publicly available UniProt database indicated that Btbd7 is myristilated and may be addressed to lipid rafts. This prompted us to investigate the role of lipid rafts and Caveolin-1 in Btbd7 function. We found that Caveolin-1 silencing correlates with EMT induction; moreover, although we have not yet confirmed our hypothesis that Caveolin-1 and Btbd7 interact, we evidenced an inverse correlation between Caveolin -1 and Btbd7 expression.

Figure 4.13

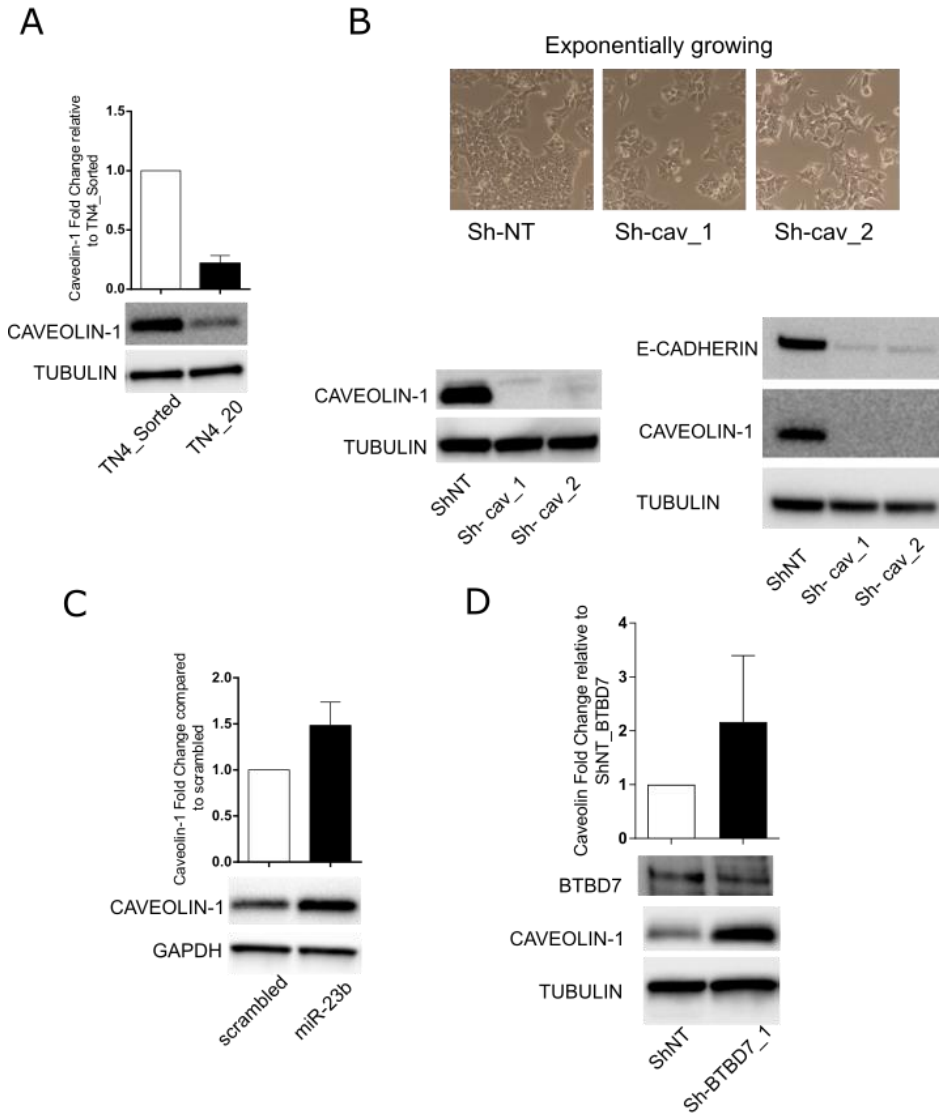


Figure 4.13 Btbd7 negatively correlates with Caveolin-1 and Caveolin-1 is a negative regulator of EMT. A) Caveolin-1 was decreased in TN4_20; B) silencing of Caveolin-1 induced mesenchymal morphology and decreased E-cadherin protein levels; C) Caveolin-1 was increased by miR-23b overexpression, a condition in which Btbd7 is decreased; D) Caveolin-1 was increased upon Btbd7 silencing.

4.8 Multiple rounds of fSCS assay enrich for cells with EMT/stem-cell traits and that are chemo-resistant

In order to scale-up our functional screen to a genome-wide level, we introduced a series of changes to our screening process that are depicted in Table 2. First, we used PB TN instead of SB TN. In fact, as mentioned in the Introduction section of this Thesis, PB TN presents some advantages compared to SB, such as the increased transposition efficiency and the absence of any footprint leaved after TN excision [176][201]. Second, we performed multiple, sequential rounds of fSCS in order to obtain pools of fSCS resistant cells instead of single clones.

	VERSION 1 of the screen	Version 2 of the screen	AIM
Transposition	Sleeping Beauty TN	PiggyBAC TN	-Higher Transposition efficiency
Assay	One Round fSCS	Multiple Rounds fSCS	-Progressive enrichment in fSCS resistant cells
Functional recovery	Single Cell Cloning	Pool of Resistant Cells	-Study the genetics of fSCS resistant cells at a genome wide level
TN insertion recovery	Linker Mediated PCR	Deep Sequencing	-Small RNA signature of fSCS resistant cells; -High throughput retrieval of TN insertions

Table 2: Comparison between version 1 and version 2 of the fSCS-TN based screen

To perform multiple rounds of fSCS, HCT116 parental and HCT116 PB transposed-cells were subjected to fSCS; after the first round, surviving cells (T1) were pooled together, expanded and subjected again to the fSCS assay to generate their progeny (T2) (being T0 the cells that never underwent fSCS) (fig. 4.14A). Interestingly, we observed that cells surviving to each additional round of fSCS were able to generate increasing numbers of surviving colonies, and these colonies had a greater scattered/mesenchymal morphology compared to the one generated by T0 cells (fig. 4.14A and 4.14B).

Figure 4.14

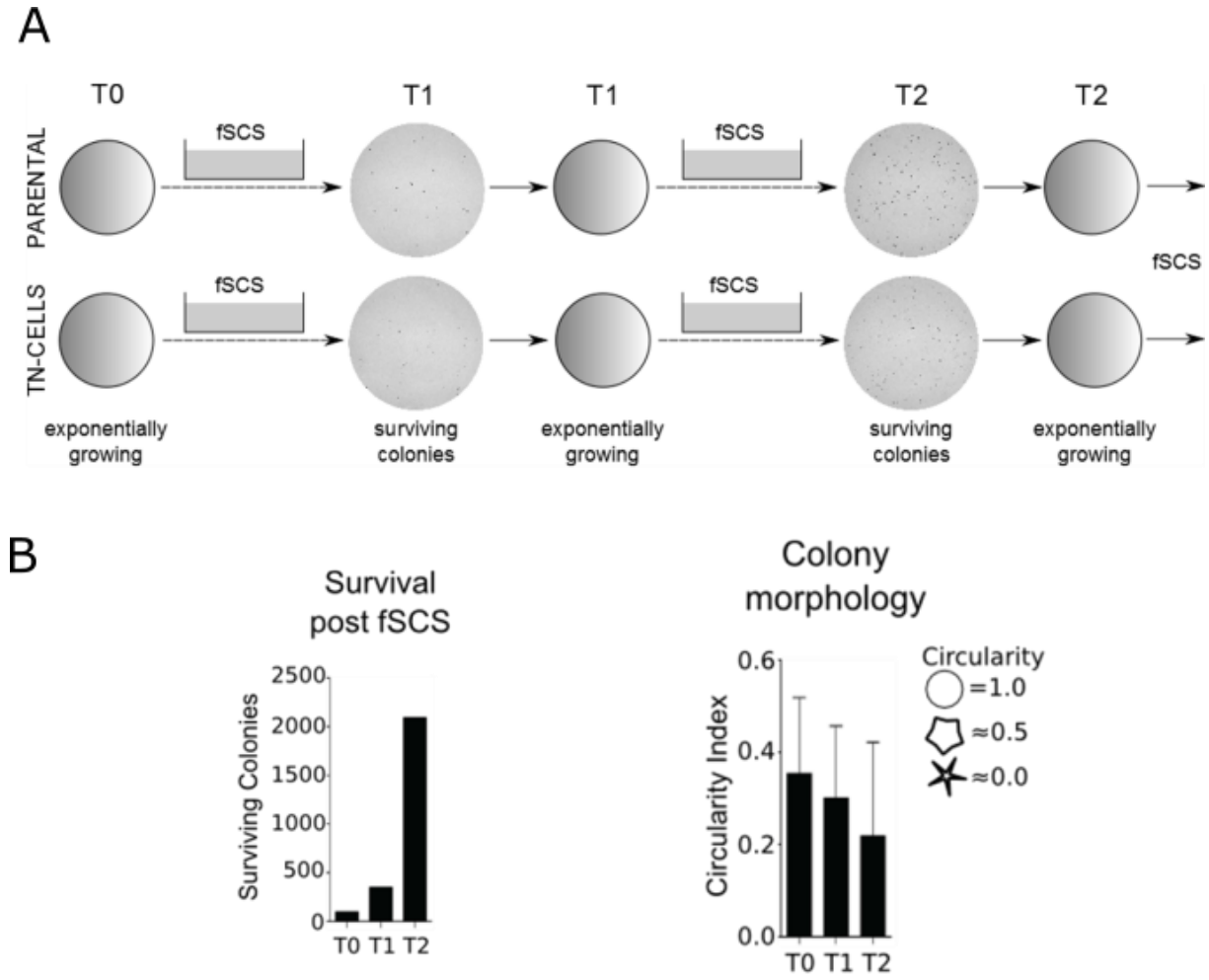


Figure 4.14 A and B: Updated version of fSCS based functional screen. A) Scheme of updated version of fSCS-TN based screen, with repeated and sequential rounds of fSCS: at each round the number of survived colonies increased; B) quantification of the number of survived colonies and of the colony morphology.

In particular, both in HCT116 parental and HCT116 PB-transposed cells we observed that the surviving colonies that were generated after fSCS assay of T2 cells (early T3) displayed EMT markers, such as lower E-Cadherin and higher Vimentin mRNA levels (fig. 4.14C). Given that EMT traits can be coupled to stemness traits [118][120], we investigated whether T3 cells display stemness traits. Since EpCAM expression is directly correlated with epithelial-cell differentiation [227], we used FACS analysis to evaluate EpCAM expression levels in T3 cells. Interestingly, we observed a greater number of cells with EpCAM low (dim) expression, i.e. with reduced EpCAM

expression (fig. 4.14D), suggesting that multiple rounds of fSCS enriches for EMT/stem-cell traits.

Finally, given that EMT has been found to regulate chemoresistance [228] [229], we asked if fSCS enrichment has an effect on chemotherapy response. We treated T2 and T0 cells with two different doses of 5-fluorouracil (5FU), and observed that T2 showed a greater survival to 5FU treatment compared to T0 cells (fig.4.14E). In conclusion, our results indicate that multiple rounds of fSCS can enrich for resistant cells with EMT/stem-cell features and with increased chemoresistance.

Figure 4.14

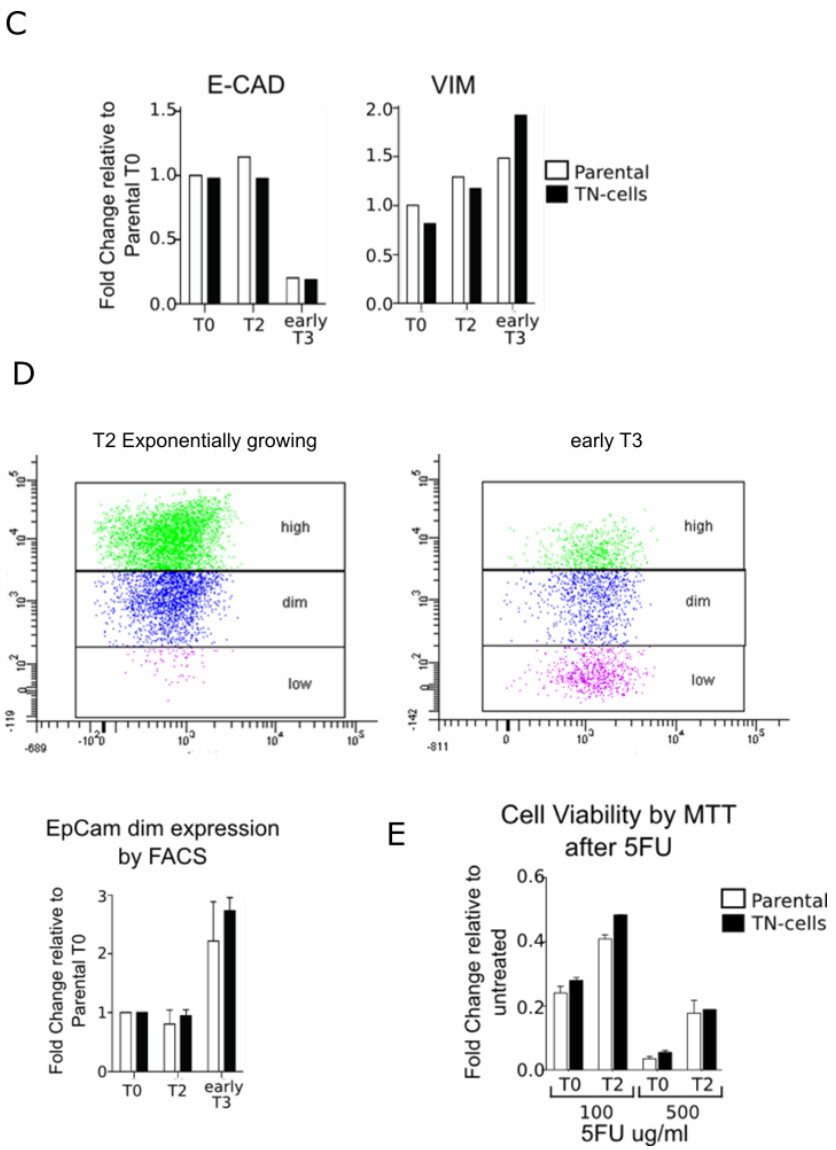


Figure 4.14 C, D, E: Repeated rounds of fSCS assay enrich for cells with EMT/stemness and chemoresistance traits. C) early T3 colonies, collected immediately after fSCS, expressed

EMT markers compared to T0 and T2 exponentially growing cells; D) early T3 colonies, collected immediately after fSCS, presented stemness traits (EpCAM low) compared to T0 and T2 exponentially growing cells E) Repeated rounds of fSCS conferred chemoresistance to T0 and T2 exponentially growing cells.

4.9 Multiple rounds of fSCS assay confer cells an increased *in vivo* metastatic potential

Next, to evaluate the *in vivo* metastatic potential of fSCS enriched cells, we subjected HCT116 parental T0 (that never underwent fSCS) and HCT116 parental T2 (that underwent 2 rounds of fSCS) cells to the *in vivo extravasation assay*. This *in vivo* assay aims to analyze the ability of tumor cells to achieve survival in the circulation and early colonization of the lungs following their injection in the blood circulation [230]. We quantitated cells that extravasated and dispersed in the lung parenchyma 72h after the inoculation (fig.4.15A and B) and found that the amount of T2 cells present in the lungs was higher compared to T0 cells. Additional immunostainings of mice lung for CD31 should be performed to actually confirm that cells found in the lungs have extravasated from the lumen vessel and have infiltrated the parenchyma. These data suggest that the "fSCS enrichment" confers to the cells the capacity to survive in the circulation, extravasate and start growing with overall greater metastatic potential (fig.4.15).

Figure 4.15

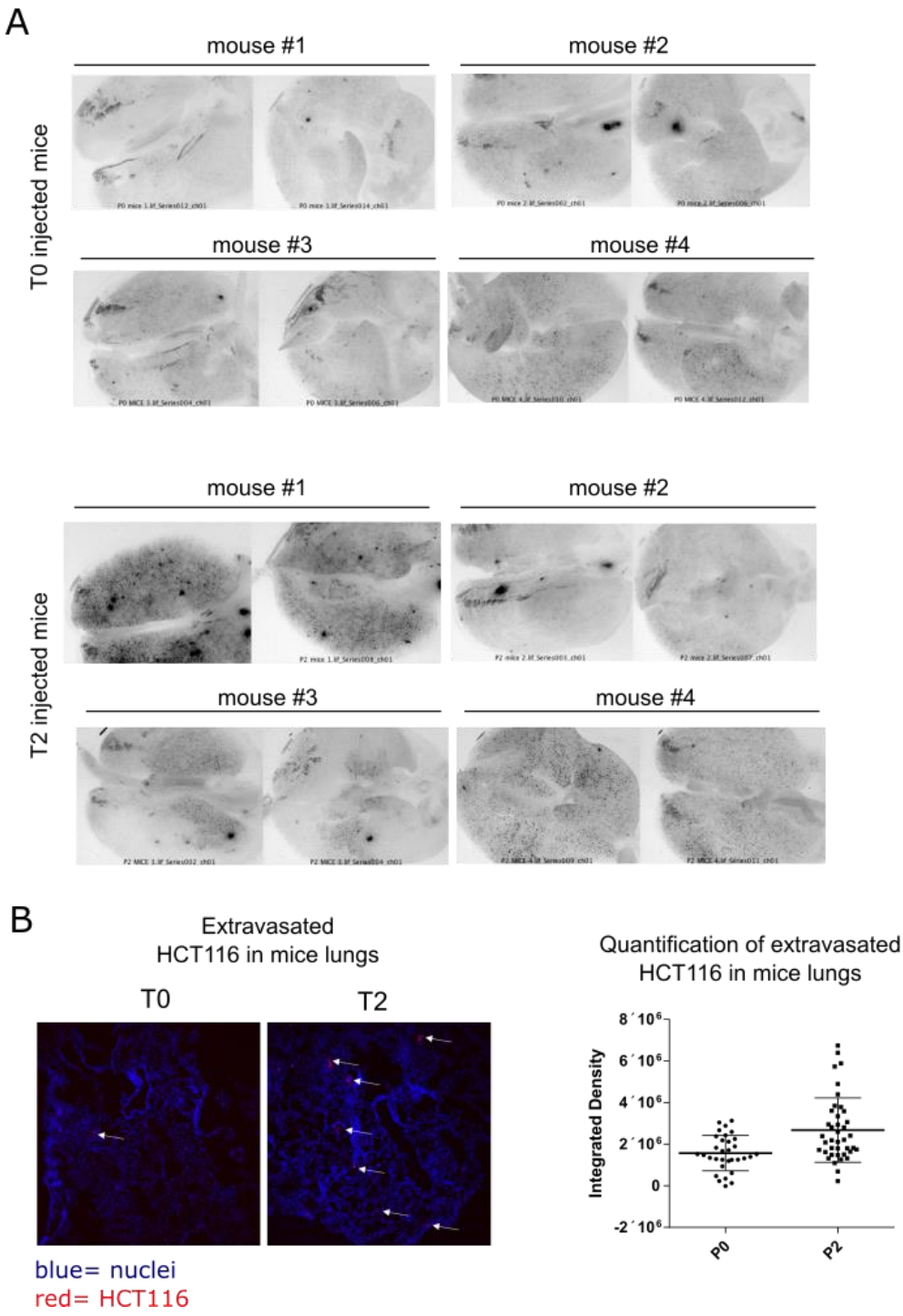


Figure 4.15 A and B: Multiple rounds of fSCS confer an increased *in vivo* metastatic potential. 72 hours after the injection in blood circulation, the capability of T2 cells to infiltrate the lungs is increased compared to T0 cells: A) Stereo-microscope acquired images (#4 mice T0 vs #4 mice T2); B) Lung section of T0 or T2 injected mice: white arrows indicate DII labeled T0 or T2 cells.

4.10 Two miRNA families involved in the regulation of EMT and metastasis are enriched by multiple rounds of fSCS

Last, we investigated whether fSCS-resistance, despite which TN-insertion or selection process triggered it, is regulated by few unique genetic elements (i.e. hubs). We reasoned that miRNAs are ideal hubs, since one miRNA can have multiple targets [143]. In order to identify miRNAs (and eventually other small RNAs) whose expression levels are significantly altered in cells that resist to fSCS compared to exponentially growing conditions we collected high quality RNA from HCT116 parental and HCT116 PB-TN T0, T2 (exponentially growing) and early T3 cells (colonies post fSCS) (fig. 4.16A) and used it for Next Generation Sequencing (NGS). We found that two families of miRNAs, miR-30 family (comprising miR-30a-3p, miR-30a-5p, miR-30c-2-3p) and miR-302 family (comprising miR-372-3p, miR-373-3p, miR-302-3p) are up-regulated upon fSCS enrichment (fig.416B). Interestingly, both miRNAs families are involved in the regulation of EMT and metastasis [231][232]. In summary, the analysis of miRNA signature in fSCS resistant cells allowed us to identify two miRNA families, namely miR-30 and miR-302 family, that have been previously described to regulate EMT and metastasis and that potentially regulate fSCS resistance.

Figure 4.16

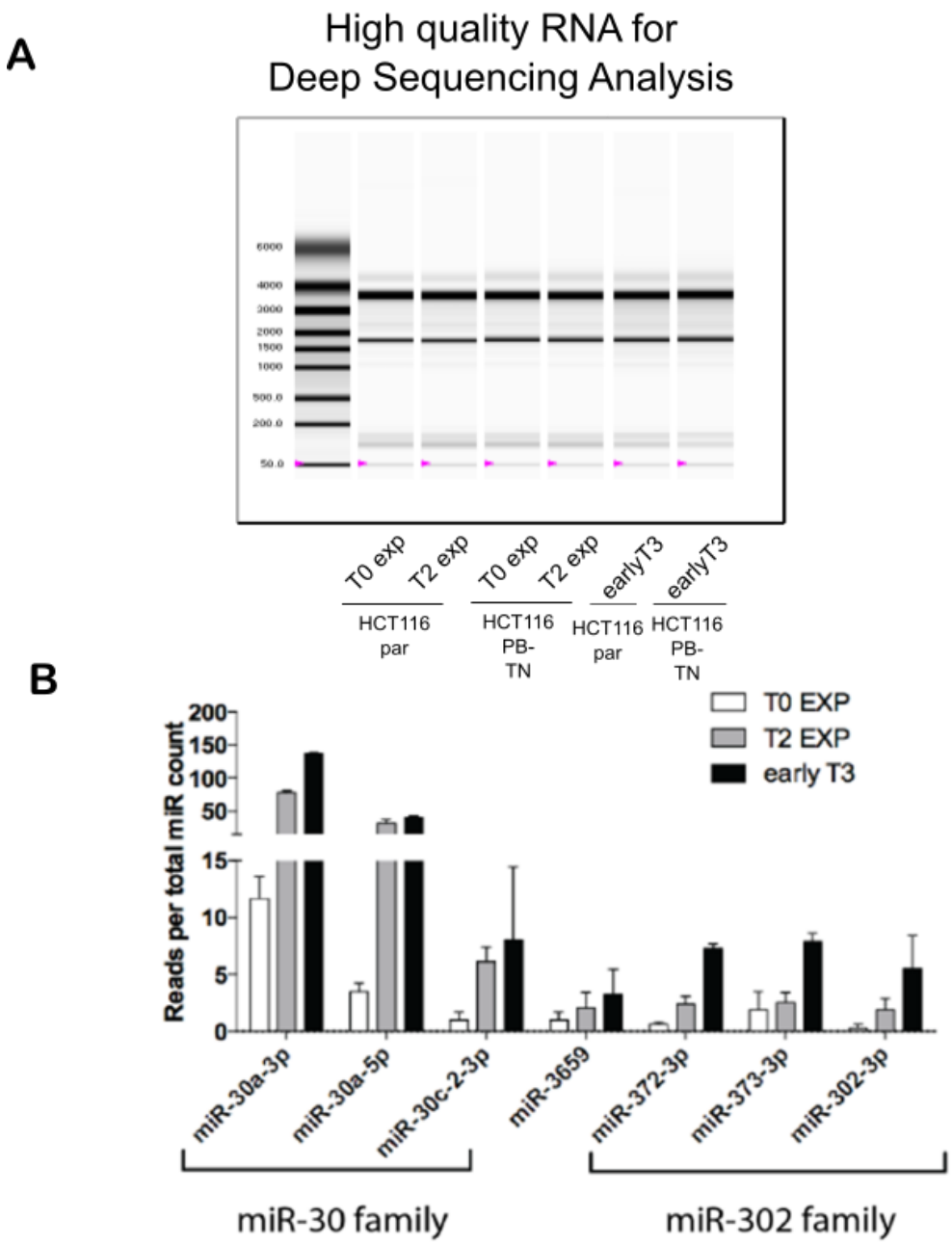


Figure 4.16: Multiple rounds of fSCS enrich for two families of miRNAs involved in the regulation of EMT and invasiveness. A) High quality RNA for Deep Sequencing from T0 and T2 cells (both Paretal and TN) and early T3 (both Parental and TN); B) two miRNAs families, miR-30 and miR-302 were found enriched with multiple rounds of fSCS.

5. DISCUSSION

Metastasis is a multistep process that ultimately leads to the formation of a new tumor lesion at a distant site [29]. During the first step of the metastatic cascade tumor cells have to detach from extracellular matrix (ECM) and from each other. Epithelial mesenchymal transition (EMT) is believed to occur at this step, since a loss of epithelial traits (*e.g.* adherens junctions) may represent a prerequisite for cells to detach from neighboring cells, to invade surrounding tissue and to proceed along the metastatic cascade. On the other hand, anoikis that is a type of cell death occurring upon cell detachment from the ECM, is an important barrier to metastasis, preventing survival of shed epithelial cells.

In vitro anoikis that consists in growing cells in the absence of cell-matrix contacts, has been used to select and study more aggressive tumor cells [62][63]. However, many tumor cells can easily overcome this form of cell death and protect themselves by strengthening cell-cell contacts and by forming big, viable suspension cell aggregates (i.e. HCT116 cells)[62]. In this Thesis work, we have set up a novel *in vitro* assay, named forced Sigle Cell Suspension (fSCS) assay, which consists in growing cells in low-adherence conditions with the concomitant inhibition of cell-cell contacts *e.g* with EDTA or enzymatically, a situation resembling an EMT. Our data indicate that: 1) fSCS is more stringent compared to classic *in vitro* anoikis, since a smaller fraction of cancer cells survive to it compared to *in vitro* anoikis; in addition, cells that survive to this constraint are more aggressive tumor cells (fig.4.1C) 2) fSCS effectively selects cells with a EMT phenotype: in fact, when EMT program is elicited by the overexpression of Twist transcription factor, EMT-switched cells gain resistance to fSCS (fig. 4.2B).

Moreover, the results of this Thesis give a contribution to the understanding of the molecular mechanisms regulating CRC EMT and metastasis. The unbiased interrogation of HCT116 genome through a modified TN based screen combined with fSCS assay, allowed us to identify: 1) a cell clone, TN4_20, that presents EMT and invasiveness traits as well as metastatic potential, confirming the effectiveness of fSCS to select EMT cells (fig. 4.5); 2) a novel interaction, revealed by TN insertion, between a non-coding RNA, miR-23b (with a known anti-metastatic function) and a protein-coding gene, BTBD7 (with a known pro-EMT and pro-metastatic function) (table 1 and figs 4.7 and 4.8). The fact that both miR-23b and BTBD7 gene were already described to regulate EMT and

metastasis [217][218][219][220][221] encouraged us about the effectiveness of our TN-fSCS based screen to detect regions of the genome that potentially regulate EMT and metastasis. Complementarily, even if we did not identify non-coding regions regulating EMT and metastasis in CRC, we unraveled the undescribed interaction between miR-23b and BTBD7 gene, that is important for fSCS phenotyoe. As a future perspective, we plan to study the relevance of BTBD7 and miR-23b also in CRC patients. To this aim, we will study by qRT-PCR whether miR-23b and BTBD7 are differentially expressed in a panel of samples comprising 10 normal colon mucosa, 10 primary CRC without metastasis (M0), 10 primary CRC with metastasis (M1) and 10 liver metastases. Samples have already been obtained from the tissue bank of CRO, Aviano. As validation, we will use a UK selected cohort of CRC samples.

We also investigated the molecular function of Btbd7 protein; although there is still much of its mechanism of action that needs to be determined, our results demonstrate that Btbd7 is necessary for fSCS resistance with an important role in the regulation of EMT and metastasis. Moreover, our data suggest that Btbd7 may regulate fSCS resistance and EMT by down-regulating the expression of E-cadherin (already shown by other groups) [217][218][219] and/or by up-regulating the expression of Zeb-1 EMT transcription factor (never described before)(fig. 4.11). Does Btbd7 directly act on E-cadherin or does it act on Zeb-1? On one hand, it is well described that E-cadherin is regulated at the transcriptional level by several EMT transcriptional repressors, such as Zeb-1; on the other hand, however, it has been shown that the loss of E-cadherin protein alone is able to induce an EMT phenotype [106], and the expression of transcription factors like Zeb-1. Thus, Btbd7 protein could induce EMT: 1) by up-regulating Zeb-1 transcription factor that in turn represses E-cadherin expression or 2) by down-regulating E-cadherin expression, that, accordingly to Onder et al., could up-regulate EMT transcription factors, such as Zeb-1. Further analysis are necessary to discriminate between these two possibilities. One strategy may be the use of Tetoff inducible regulation of Btbd7 expression, as described in [217], in order to study, at several time points, the reciprocal variations on Btbd7, E-cadherin and Zeb-1 mRNAs and protein levels. Another unclear point in Btbd7 function resides in the fact that the changes of E-cadherin and Zeb-1 protein levels are accompanied by changes of their mRNA levels, suggesting that Btbd7 affects directly or indirectly their transcription level. However, this is in contrast with the observation that ectopic Btbd7 does not localize in the

nucleus (fig. 4.10B), thus suggesting that other intermediate interactors or downstream effectors participate in the regulation of gene expression that follows Btbd7 activation. This contradiction may be alternatively explained with the hypothesis that Btbd7-mediated EMT induction relies first on changes in protein levels; then, given that EMT is capable of auto-maintaining itself, the changes of mRNA levels may occur secondly.

As alternative approach to decipher the Btbd7molecular function, we used bioinformatic tools and publicly available protein Databases (figs 4.12 and 4.13). When we used Blastp bioinformatic tool for the local alignment of Btbd7 protein in a *non-redundant protein database*, we found, as expected, that Btbd7 is conserved in Primates (*Gorilla gorilla gorilla, Pan paniscus,* and *Pan troglodytes*); more interestingly, Btbd7 protein resulted to be conserved also other Mammals (*Camelus bactrianus* and *Canis lupus familiaris*), thus suggesting the biological relevance of Btbd7 function. Blastp bioinformatic tool permits to restrict the local alignment to human proteins with "*RefSeq protein-Homo sapiens"* restriction option. By repeating the local alignment of Btbd7 with this restriction option, we were able to identify two Btbd7 shorter isoforms. Interestingly, this alignment also identified several proteins that show a low level of similarity to Btbd7 (less than 50%), except for the Btbd7 BTB/POZ and BTB/Kelch domains, thus suggesting that BTB/POZ and BTB/Kelch domains are conserved in human proteins. In fact, BTB/POZ is a protein-protein interaction domain that, alone or in association with BTB/Kelch domain, is present in several human proteins and regulates functions ranging from transcription repression, cytoskeletal regulation, tetramerization and gating of ion channels, ubiquitination and degradation of proteins [222][233]. Interestingly, when BTB/POZ domain is located N-terminal to the BTB/Kelch domain (as in Btbd7), it can participate in dimerization and interaction with Cul3, a E3 ubiquitin ligase [223]. Given these observations, we hypothesized that Btbd7 may interact with Cul3 and, as already described for proteins sharing its same arrangement of BTB/POZ and BTB/Kelch domains, may be implicated in the ubiquitination of target proteins for proteosomal degradation. In agreement, we noted that UniProt protein database indicates Cul3 as a putative interactor of Btbd7. This in an interesting point, since the expression levels of some EMT transcription factors, such as Snail-1, Slug and (more slightly) Zeb-1 have been demonstrated to be regulated by ubiquitination and proteasomal degradation [234]. A co-immunoprecipitation assay of

both endogenous Btbd7 and/or ectopic eGFP-Btbd7 with Cul3 may be a good approach to assess if these proteins actually interact.

Moreover, the use of Uniprot database revealed that Btbd7 protein is subjected to a post-translational modification, namely a N-myristylation at the aa position 2. We were immediately struck by this post-translational modification, since it has been described to address proteins to lipid rafts, which are particular membrane sub-domains rich in cholesterol and sphingolipids that may regulate EMT [224]. Moreover, Caveolin-1, a marker of caveolae, i.e. lipid rafts subtypes, is a negative regulator of EMT [226]. Given these notions, we asked if Btbd7 function is related to lipid rafts, and in particular to Caveolin-1. On one hand our data confirmed, at least in part, that Caveolin-1 has a role in EMT negative regulation. In fact, we found that: 1) Caveolin-1 is downregulated in TN4_20, that is fSCS resistant clone with EMT traits; 2) Caveolin-1 silencing induces a more mesenchymal and scattered cell morphology with reduction of E-cadherin protein levels. On the other hand, we found that Caveolin-1 expression seems to negatively correlate with Btbd7 expression, in fact: 1) Caveolin-1 expression is increased in cells in which Btbd7 is silenced and 2) Caveolin-1 is increased when Btbd7 is dowregulated by miR-23b overexpression. Given these results, we hypothesized that Caveolin-1 may negatively regulate Btbd7 function; in particular, we hypothesized that Caveolin-1 interacts with Btbd7. Co-immunoprecipitation and/or co-localization experiments are required to demonstrate the direct interaction between Caveolin-1 and Btbd7; thus, for now, we can only speculate on a role of Caveolin-1 in the regulation of Btbd7. To better address Btbd7 mechanism of action future area of research should be: 1) the use of biochemical approach to investigate if Btbd7 is actually myristylated; 2) to investigate if Btbd7 is actually addressed to lipid rafts; 3) to assess if Btbd7 and Caveolin-1 interact (e.g with a co-immunoprecipitation assay).

In conclusion, the use of bioinformatic tools provided us with some interesting inputs to further investigate Btbd7 function; even though we did not unravel its precise mechanism of action, we hypothesized two possible roads: 1) Btbd7, similar to others BTB/POZ and BTB/Kelch proteins, is engaged in the interaction with E3 ubiquitin ligase and is involved in the ubiquitination of target proteins, such as Zeb-1 transcription factor. This hypothesis may appear in contrast with the observation that Btbd7 seems to regulate Zeb-1 expression at a transcription level. However, as already mentioned, it is possible that the EMT program, once activated, is capable of auto-maintaining itself and

that the increase of Zeb-1 mRNA expression is secondary to Zeb-1 protein induction; 2) Btbd7 may be directed to lipid rafts by N-myristylation and regulated (directly or indirectly) Caveolin-1, a lipid rafts protein that acts as negative regulator of EMT.

A mass spectrometry based analysis of Btbd7 interactors could be a very useful approach to discriminate between these two hypotheses and may help to understand if they are interconnected.

Some tumors (including colon adenocarcinoma) form highly differentiated metastases; this observation can be explained by a plasticity model in which EMT/stemness state alternates with MET/differentiated state: an EMT is required for tumor cells to detach and migrate from the primary site, whereas they must come back to epithelial state (MET) in order to grow and colonize distant sites [118]. Thus, switching back and forth between repeated EMT/MET may be a prerequisite for metastasis formation. Our results show that repeated and sequential rounds of fSCS are able to generate pools of cells that are fSCS resistant. Regardless of the presence of TN insertion (in this case achieved by PB TN), the fSCS resistant colonies increase each round, acquiring a more scattered/mesenchymal morphology (T1= one round of fSCS; T2= two rounds; T0= never underwent to fSCS). Interestingly, early T3 cells, which are the early surviving cells obtained by performing an additional round of fSCS with T2, present increased EMT and stemness traits (decrease of E-cadherin, increase of Vimentin, increase of EpCAM dim/low population); whereas these traits are absent in expanded T1 and T2 (fig.4.14).

In some way, these observations may be linked to the EMT/MET model described in [118]. We hypothesize that T1 expanded and T2 cells, that are exponentially growing cells, may have lost their EMT traits to undergo a MET and can grow in the culture dish, similarly to transformed tumor cells that must undergo MET to grow at distant sites. Indeed, T1 and T2 cells do not do not show a change in the classic EMT markers (except a slight increase in Vimentin expression) nor show a decrease in EpCAM expression. Conversely, early T3 cells, that are fSCS surviving colonies that did not generate a full progeny (thus, they are not exponentially growing), may still be in a EMT and stemness state, resembling tumor cells that temporary undergo EMT during metastatic transformation. Our hypothesis needs to be further confirmed by the analysis of

exponentially growing T3 (late T3), which should display a decrease of EMT/stemness traits and an increase of MET/differentiated traits.

To sum up, we hypothesize that performing multiple rounds of fSCS may recapitulate two distinct phases of tumor cell plasticity: the early EMT/stemness phase, that is recapitulated by cells collected as surviving colonies that are still not in the exponentially growing phase, and the late MET/differentiated phase, recapitulated by cells collected in the exponentially growing phase. Thus, each pool of exponentially growing fSCS resistant cells (late, MET) should be generated by corresponding fSCS surviving colonies (early, EMT).

Extravasation is the step of the metastatic cascade following survival in the circulation is, in which CTCs (circulating tumor cells) leave lymphatic or blood circulation and penetrate the endothelium to invade the target organ. *In vivo* extravasation assay with T0 versus T2 cells, showed increased extravasation capability of T2 fSCS resistant cells compared to T0 cells. (fig.4.15). This preliminary result is an interesting indication that fSCS selection process enriches for more aggressive tumor cells, but needs to be further confirmed by challenging the capability of T2 cells to form distant macro-metastatic lesions. Given that T2 cells were generated by the fSCS enrichment process that selects cells with more EMT/stemness/chemoresistance traits, we strongly expect that T2 cells will be able to form distant macrometastases.

What are the genetic elements regulating fSCS resistance? In the first version of our functional screen, we answered to this question by identifying and studying the TN insertions occurred in one of the cell clones, TN4_20, that emerged from the selection process. In this manner, we dissected the molecular mechanism happening in a specific cell clone. In our extended version of the screen, however, we aimed to scale-up the identification of the genetic elements regulating fSCS to a genome-wide level. To this aim, on one hand we focused on small-RNAs, in particular miRNAs common signatures. Indeed, our expectation is that fSCS may be regulated by few unique genetic elements (i.e. the hubs), and that miRNAs represent the perfect hubs, because a single miRNA can target multiple mRNAs. Deep sequencing analysis revealed that fSCS selective process enriches for two families of miRNAs, namely miR-30 and miR-302 (fig.4.16). Similar to miR-23b and BTBD7 in the previous version of the screen, also miR-30 and miR-302 families were already found to be implicated in EMT and metastasis [231][232], thus

providing a further confirmation of the effectiveness of our fSCS selective process in the identification of genetic elements regulating EMT. However, a further analysis of the relevance of these miRNA families for fSCS is required. After confirming by q-RT-PCR that miR-30 and miR-302 family members are upregulated in fSCS resistant cells, a possible strategy to assess their role in fSCS resistance will be the overexpression of these miRNAs in parental cells, in order to understand whether they confer resistance to fSCS. In addition to differences in the expression of these two miRNA families, by allowing 2 mismatches when performing deep sequencing data alignment, we observed that fSCS induces also isomiRs. IsomiRs are isoforms of the same miRNA that differ from the canonical sequence for the 5' and/or 3' end. Most commonly, isomiRs present variations in the length of 3' end, which consist in shorter sequences or, conversely, in sequences with added nucleotides which potentially act as localization signals. Interestingly, TCGA (The Cancer Genome Atlas) data show that isomiRs distinguish tumor from normal tissue better than classic miRNAs expression [235]. Thus, we plan to: 1) re-analyze deep sequencing data we have generated from fSCS resistant cells using customized pipelines specific for the identification of differences in isomiRs; 2) analyze isomiRs expression profile in TCGA dataset to evaluate the differential expression of isomiRs between CRC (primary and metastatic) and normal samples.

Finally, even if fSCS resistant cells increase at each round and acquire EMT and stemness traits regardless of Piggy Back TN insertion, we plan to scale up retrieval of TN insertions from the pool of TN-T2-cells and by these means to identify the insertions (i.e. triggers) that most likely regulate fSCS-resistance and are responsible for EMT/stem cell and chemo-resistance traits.

BIBLIOGRAPHY

[1] F. Arvelo, F. Sojo, and C. Cotte, "Biology of colorectal cancer," *ecancermedicalscience*, vol. 9. 2015.

[2] L. A. Gloeckler Ries, M. E. Reichman, D. R. Lewis, B. F. Hankey, and B. K. Edwards, "Cancer survival and incidence from the Surveillance, Epidemiology, and End Results (SEER) program.," *Oncologist*, vol. 8, no. 6, pp. 541–552, 2003.

[3] A. Jemal, R. Siegel, J. Xu, and E. Ward, "Cancer statistics, 2010.," *CA. Cancer J. Clin.*, vol. 60, no. 5, pp. 277–300, 2010.

[4] OMS, "ORGANIZACIÓN MUNDIAL DE LA SALUD," *Centro de prensa*, 2014. .

[5] A. J. M. Watson and P. D. Collins, "Colon cancer: A civilization disorder," in *Digestive Diseases*, 2011, vol. 29, no. 2, pp. 222–228.

[6] O. S. Lin, "Colorectal cancer screening in patients at moderately increased risk due to family history.," *World J. Gastrointest. Oncol.*, vol. 4, no. 6, pp. 125–130, 2012.

[7] F. G. C. M. de Campos, I. Nicácio De Freitas, A. R. Imperiale, V. E. Seid, R. O. Perez, S. C. Nahas, and I. Cecconello, "Colorectal cancer in familial adenomatous polyposis: Are there clinical predictive factors?," *Cir. Esp.*, vol. 88, no. 6, pp. 390–397, 2010.

[8] S. Hassen, N. Ali, and P. Chowdhury, "Molecular signaling mechanisms of apoptosis in hereditary non-polyposis colorectal cancer.," *World J. Gastrointest. Pathophysiol.*, vol. 3, no. 3, pp. 71–9, 2012.

[9] A. Manuscript, "NIH Public Access," *Changes*, vol. 29, no. 6, pp. 997–1003, 2012.

[10] T. Muto, H. Bussey, and B. Morson, "The evolution of cancer of the colon and rectum," *Cancer*, vol. 36, no. March 1957, pp. 2251–2270, 1975.

[11] P. M. Colucci, S. H. Yale, and C. J. Rall, "Colorectal polyps.," *N. Y. State J. Med.*, vol. 1, no. 3, pp. 261–262, 2003.

[12] V. T. DeVita, S. Hellman, and S. A. Rosenberg, "Cancer : Principles and Practice of Oncology Collections :," in *Clinical Medicine*, 2005, p. 3120.

[13] J. R. Egner, "Abeloff's Clinical Oncology," *JAMA J. Am. Med. Assoc.*, vol. 303, no. 11, pp. 1097–1097, 2010.

[14] S. D. Markowitz, D. M. Dawson, J. Willis, and J. K. V Willson, "Focus on colon cancer," *Cancer Cell*, vol. 1, no. 3, pp. 233–236, 2002.

[15] T. André, C. Boni, L. Mounedji-Boudiaf, M. Navarro, J. Tabernero, T. Hickish, C. Topham, M. Zaninelli, P. Clingan, J. Bridgewater, I. Tabah-Fisch, and A. de Gramont, "Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer.," *N. Engl. J. Med.*, vol. 350, no. 23, pp. 2343–2351, 2004.

[16] T. Armaghany, J. D. Wilson, Q. Chu, and G. Mills, "Genetic alterations in colorectal cancer.," *Gastrointest. Cancer Res.*, vol. 5, no. 1, pp. 19–27, 2012.

[17] S. D. Markowitz and M. M. Bertagnolli, "Molecular Basis of Colorectal Cancer," *N. Engl. J. Med.*, vol. 361, no. 25, pp. 2449–2460, 2009.

[18] A. E. K. Ibrahim and M. J. Arends, "Molecular typing of colorectal cancer: Applications in diagnosis and treatment," *Diagnostic Histopathol.*, vol. 18, no. 2, pp. 70–80, 2012.

[19] A. Walther, E. Johnstone, C. Swanton, R. Midgley, I. Tomlinson, and D. Kerr, "Genetic prognostic and predictive markers in colorectal cancer.," *Nat. Rev. Cancer*, vol. 9, no. 7, pp. 489–99, 2009.

[20] B. Vogelstein, E. R. Fearon, S. R. Hamilton, S. E. Kern, A. C. Preisinger, M. Leppert, Y. Nakamura, R. White, A. M. Smits, and J. L. Bos, "Genetic alterations during colorectal-tumor development.," *N. Engl. J. Med.*, vol. 319, no. 9, pp. 525–32, 1988.

[21] R. Nagel, C. Le Sage, B. Diosdado, M. Van Der Waal, J. A. F. Oude Vrielink, A. Bolijn, G. A. Meijer, and R. Agami, "Regulation of the adenomatous polyposis coli gene by the miR-135 family in colorectal cancer," *Cancer Res.*, vol. 68, no. 14, pp. 5795–5802, 2008.

[22] H. Ling, K. Pickard, C. Ivan, C. Isella, M. Ikuo, R. Mitter, R. Spizzo, M. D. Bullock, C. Braicu, V. Pileczki, K. Vincent, M. Pichler, V. Stiegelbauer, G. Hoefler, M. I. Almeida, a. Hsiao, X. Zhang, J. N. Primrose, G. K. Packham, K. Liu, K. Bojja, R. Gafa, L. Xiao, S. Rossi, J. H. Song, I. Vannini, F. Fanini, S. Kopetz, P. Zweidler-McKay, X. Wang, C. Ionescu, a. Irimie, M. Fabbri, G. Lanza, S. R. Hamilton, I. Berindan-Neagoe, E. Medico, a. H. Mirnezami, G. a. Calin, and M. S. Nicoloso, "The clinical and biological significance of MIR-224 expression in colorectal cancer metastasis," *Gut*, pp. 1–13, 2015.

[23] M. Tuna and C. I. Amos, "Genomic sequencing in cancer," *Cancer Letters*, vol. 340, no. 2. pp. 161–170, 2013.

[24] W. M. Grady and J. M. Carethers, "Genomic and Epigenetic Instability in Colorectal Cancer Pathogenesis," *Gastroenterology*, vol. 135, no. 4. pp. 1079–1099, 2008.

[25] J. G. Herman, A. Umar, K. Polyak, J. R. Graff, N. Ahuja, J. P. Issa, S. Markowitz, J. K. Willson, S. R. Hamilton, K. W. Kinzler, M. F. Kane, R. D. Kolodner, B. Vogelstein, T. A. Kunkel, and S. B. Baylin, "Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma.," *Proc Natl Acad Sci U S A*, vol. 95, no. 12, pp. 6870–5, 1998.

[26] N. Valeri, P. Gasparini, M. Fabbri, C. Braconi, A. Veronese, F. Lovat, B. Adair, I. Vannini, F. Fanini, A. Bottoni, S. Costinean, S. K. Sandhu, G. J. Nuovo, H. Alder, R. Gafa, F. Calore, M. Ferracin, G. Lanza, S. Volinia, M. Negrini, M. A. McIlhatton, D. Amadori, R. Fishel, and C. M. Croce, "Modulation of mismatch repair and genomic stability by miR-155.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 15, pp. 6982–7, 2010.

[27] R. Parsons, L. L. Myeroff, B. Liu, J. K. Willson, S. D. Markowitz, K. W. Kinzler, and B. Vogelstein, "Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer," *Cancer Res.*, vol. 55, no. 23, pp. 5548–5550, 1995.

[28] M. F. Leber and T. Efferth, "Molecular principles of cancer invasion and metastasis (Review)," *International Journal of Oncology*, vol. 34, no. 4. pp. 881–895, 2009.

[29] G. P. Gupta and J. Massagu??, "Cancer Metastasis: Building a Framework," *Cell*, vol. 127, no. 4. pp. 679–695, 2006.

[30] S. M. Frisch and H. Francis, "Disruption of epithelial cell-matrix interactions induces apoptosis," *J. Cell Biol.*, vol. 124, no. 4, pp. 619–626, 1994.

[31] R. Kalluri and R. a Weinberg, "Review series The basics of epithelial-mesenchymal transition," *J. Clin. Invest.*, vol. 119, no. 6, pp. 1420–1428, 2009.

[32] D. Wirtz, K. Konstantopoulos, and P. C. Searson, "The physics of cancer: the role of physical interactions and mechanical forces in metastasis," *Nat Rev Cancer*, vol. 11, no. 7, pp. 512–522,

2011.

[33] P. Friedl, K. Wolf, and J. Lammerding, "Nuclear mechanics during cell migration," *Current Opinion in Cell Biology*, vol. 23, no. 1. pp. 55–64, 2011.

[34] K. N. Dahl, S. M. Kahn, K. L. Wilson, and D. E. Discher, "The nuclear envelope lamina network has elasticity and a compressibility limit suggestive of a molecular shock absorber.," *J. Cell Sci.*, vol. 117, no. Pt 20, pp. 4779–86, 2004.

[35] Y. Tseng, "Micro-organization and visco-elasticity of the interphase nucleus revealed by particle nanotracking," *J. Cell Sci.*, vol. 117, no. 10, pp. 2159–2167, 2004.

[36] A. K. Simi, A. S. Piotrowski, and C. M. Nelson, "Mechanotransduction, metastasis and genomic instability," *Cancer Metastasis - Biol. Treat.*, vol. 20, pp. 139–158, 2015.

[37] L. Borsig, R. Wong, J. Feramisco, D. R. Nadeau, N. M. Varki, and A. Varki, "Heparin and cancer revisited: mechanistic connections involving platelets, P-selectin, carcinoma mucins, and tumor metastasis.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, pp. 3352–3357, 2001.

[38] S. L. Napier, Z. R. Healy, R. L. Schnaar, and K. Konstantopoulos, "Selectin ligand expression regulates the initial vascular interactions of colon carcinoma cells: The roles of CD44V and alternative sialofucosylated selectin ligands," *J. Biol. Chem.*, vol. 282, no. 6, pp. 3433–3441, 2007.

[39] S. N. Thomas, R. L. Schnaar, and K. Konstantopoulos, "Podocalyxin-like protein is an E-/L-selectin ligand on colon carcinoma cells: comparative biochemical properties of selectin ligands in host and tumor cells.," *Am. J. Physiol. Cell Physiol.*, vol. 296, no. 3, pp. C505–C513, 2009.

[40] S. N. Thomas, F. Zhu, R. L. Schnaar, C. S. Alves, and K. Konstantopoulos, "Carcinoembryonic antigen and CD44 variant isoforms cooperate to mediate colon carcinoma cell adhesion to E- and L-selectin in shear flow," *J. Biol. Chem.*, vol. 283, no. 23, pp. 15647–15655, 2008.

[41] K. Konstantopoulos and S. N. Thomas, "Cancer cells in transit: the vascular interactions of tumor cells.," *Annu. Rev. Biomed. Eng.*, vol. 11, no. April, pp. 177–202, 2009.

[42] V. Patanaphan and O. M. Salazar, "Colorectal cancer: metastatic patterns and prognosis.," *Southern medical journal*, vol. 86, no. 1. pp. 38–41, 1993.

[43] M. Mendoza and C. Khanna, "Revisiting the seed and soil in cancer metastasis," *International Journal of Biochemistry and Cell Biology*, vol. 41, no. 7. pp. 1452–1462, 2009.

[44] B. Psaila and D. Lyden, "The metastatic niche: adapting the foreign soil.," *Nat. Rev. Cancer*, vol. 9, no. 4, pp. 285–293, 2009.

[45] R. N. Kaplan, R. D. Riba, S. Zacharoulis, A. H. Bramley, L. Vincent, C. Costa, D. D. MacDonald, D. K. Jin, K. Shido, S. a Kerns, Z. Zhu, D. Hicklin, Y. Wu, J. L. Port, N. Altorki, E. R. Port, D. Ruggero, S. V Shmelkov, K. K. Jensen, S. Rafii, and D. Lyden, "VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche.," *Nature*, vol. 438, no. 7069, pp. 820–7, 2005.

[46] A. P. Gilmore, "Anoikis," *Cell Death Differ.*, vol. 12, pp. 1473–1477, 2005.

[47] K. C. Zimmermann and D. R. Green, "How cells die: Apoptosis pathways," in *Journal of Allergy and Clinical Immunology*, 2001, vol. 108, no. 4 SUPPL.

[48] M. G. Valmiki and J. W. Ramos, "Death effector domain-containing proteins," *Cellular and Molecular Life Sciences*, vol. 66, no. 5. pp. 814–830, 2009.

[49] M. M. Hill, C. Adrain, and S. J. Martin, "Portrait of a killer: the mitochondrial apoptosome emerges

from the shadows.," *Mol. Interv.*, vol. 3, no. 1, pp. 19–26, 2003.

[50] R. O. Hynes, "Integrins: Bidirectional, allosteric signaling machines," *Cell*, vol. 110, no. 6. pp. 673–687, 2002.

[51] P. Mehlen and A. Puisieux, "Metastasis: a question of life or death.," *Nat. Rev. Cancer*, vol. 6, no. 6, pp. 449–58, 2006.

[52] H. Puthalakath, D. C. S. Huang, L. A. O'Reilly, S. M. King, and A. Strasser, "The proapoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex," *Mol. Cell*, vol. 3, no. 3, pp. 287–296, 1999.

[53] C. Horbinski, C. Mojesky, and N. Kyprianou, "Tales of Homeless (Cells) in Cancer," *Am J Pathol*, vol. 177, pp. 1044–1052, 2010.

[54] A. Apel, H. Zentgraf, M. W. Buchler, and I. Herr, "Autophagy-A double-edged sword in oncology," *Int. J. Cancer*, vol. 125, no. 5, pp. 991–995, 2009.

[55] S. Pattingre, L. Espert, M. Biard-Piechaczyk, and P. Codogno, "Regulation of macroautophagy by mTOR and Beclin 1 complexes," *Biochimie*, vol. 90, no. 2, pp. 313–323, 2008.

[56] J. Debnath, S. K. Muthuswamy, and J. S. Brugge, "Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures," *Methods*, vol. 30, no. 3. pp. 256–268, 2003.

[57] Y. F. Peng, Y. H. Shi, Z. Bin Ding, A. W. Ke, C. Y. Gu, B. Hui, J. Zhou, S. J. Qiu, Z. Dai, and J. Fan, "Autophagy inhibition suppresses pulmonary metastasis of HCC in mice via impairing anoikis resistance and colonization of HCC cells," *Autophagy*, vol. 9, no. 12, pp. 2056–2068, 2013.

[58] R. L. Macintosh, P. Timpson, J. Thorburn, K. I. Anderson, A. Thorburn, and K. M. Ryan, "Inhibition of autophagy impairs tumor cell invasion in an organotypic model," *Cell Cycle*, vol. 11, no. 10, pp. 2022–2029, 2012.

[59] S. Luo and D. C. Rubinsztein, "Apoptosis blocks Beclin 1-dependent autophagosome synthesis: an effect rescued by Bcl-xL.," *Cell Death Differ.*, vol. 17, no. 2, pp. 268–277, 2010.

[60] J. Folkman and a Moscona, "Role of cell shape in growth control.," *Nature*, vol. 273, pp. 345–349, 1978.

[61] F. Liu, X. Zhou, F. Cui, and D. Jia, "Synthesis and properties of poly(hydroxyethyl methacrylate) hydrogel for IOL materials," *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, vol. 24, no. 3, pp. 595–598, 2007.

[62] S. Information, F. Rt, and T. Pcr, "Letters To Nature," *Nature*, vol. 430, no. August, 2004.

[63] Q. Cai, L. Yan, and Y. Xu, "Anoikis resistance is a critical feature of highly aggressive ovarian cancer cells.," *Oncogene*, no. November 2013, pp. 1–10, 2014.

[64] C. L. Buchheit, K. J. Weigel, and Z. T. Schafer, "Cancer cell survival during detachment from the ECM : multiple barriers to tumour progression," *Nat. Publ. Gr.*, vol. 14, no. 9, pp. 632–641, 2014.

[65] S. E. Dunn, M. Ehrlich, N. J. Sharp, K. Reiss, G. Solomon, R. Hawkins, R. Baserga, and J. C. Barrett, "A dominant negative mutant of the insulin-like growth factor-I receptor inhibits the adhesion, invasion, and metastasis of breast cancer.," *Cancer research*, vol. 58, no. 15. pp. 3353–61, 1998.

[66] M. Resnicoff, D. Coppola, C. Sell, R. Rubin, S. Ferrone, and R. Baserga, "Growth inhibition of human melanoma cells in nude mice by antisense strategies to the type 1 insulin-like growth factor

receptor," *Cancer Res*, vol. 54, no. 18, pp. 4848–4850, 1994.

[67] M. J. Reginato, K. R. Mills, J. K. Paulus, D. K. Lynch, D. C. Sgroi, J. Debnath, S. K. Muthuswamy, and J. S. Brugge, "Integrins and EGFR coordinately regulate the pro-apoptotic protein Bim to prevent anoikis.," *Nat. Cell Biol.*, vol. 5, no. 8, pp. 733–740, 2003.

[68] M. J. Reginato, K. R. Mills, E. B. E. Becker, D. K. Lynch, A. Bonni, S. K. Muthuswamy, and J. S. Brugge, "Bim regulation of lumen formation in cultured mammary epithelial acini is targeted by oncogenes.," *Mol. Cell. Biol.*, vol. 25, no. 11, pp. 4591–601, 2005.

[69] K. K. Haenssen, S. A. Caldwell, K. S. Shahriari, S. R. Jackson, K. A. Whelan, A. J. Klein-Szanto, and M. J. Reginato, "ErbB2 Requires Integrin α5 for Anoikis Resistance via Src Regulation of Receptor Activity in Human Mammary Epithelial Cells," *J. Cell Sci.*, vol. 123, no. Pt 8, pp. 1373–1382, 2010.

[70] M. Yu, D. T. Ting, S. L. Stott, B. S. Wittner, F. Ozsolak, J. C. Ciciliano, M. E. Smas, D. Winokur, A. J. Gilman, M. J. Ulman, K. Xega, G. Contino, B. Alagesan, W. Brian, P. M. Milos, D. P. Ryan, L. V Sequist, N. Bardeesy, M. Toner, S. Maheswaran, D. a Haber, S. Paul, J. C. Ciciliano, M. E. Smas, D. Winokur, A. J. Gilman, M. J. Ulman, K. Xega, G. Contino, B. Alagesan, B. W. Brannigan, P. M. Milos, D. P. Ryan, L. V Sequist, N. Bardeesy, S. Ramaswamy, M. Toner, S. Maheswaran, and D. a Haber, "RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis.," *Nature*, vol. 487, no. 7408, pp. 510–3, 2013.

[71] S. Attwell, C. Roskelley, and S. Dedhar, "The integrin-linked kinase (ILK) suppresses anoikis.," *Oncogene*, vol. 19, no. 33, pp. 3811–5, 2000.

[72] A. Khwaja, P. Rodriguez-Viciana, S. Wennström, P. H. Warne, and J. Downward, "Matrix adhesion and Ras transformation both activate a phosphoinositide 3-OH kinase and protein kinase B/Akt cellular survival pathway," *EMBO J.*, vol. 16, no. 10, pp. 2783–2793, 1997.

[73] M. Rytömaa, K. Lehmann, and J. Downward, "Matrix detachment induces caspase-dependent cytochrome c release from mitochondria: inhibition by PKB/Akt but not Raf signalling.," *Oncogene*, vol. 19, no. 39, pp. 4461–8, 2000.

[74] N. Yamaki, M. Negishi, and H. Katoh, "RhoG regulates anoikis through a phosphatidylinositol 3-kinase-dependent mechanism," *Exp. Cell Res.*, vol. 313, no. 13, pp. 2821–2832, 2007.

[75] K. Orford, C. C. Orford, and S. W. Byers, "Exogenous expression of ??-catenin regulates contact inhibition, anchorage-independent growth, anoikis, and radiation-induced cell cycle arrest," *J. Cell Biol.*, vol. 146, no. 4, pp. 855–867, 1999.

[76] K. H. Shain, T. H. Landowski, and W. S. Dalton, "Adhesion-Mediated Intracellular Redistribution of c-Fas-Associated Death Domain-Like IL-1-Converting Enzyme-Like Inhibitory Protein-Long Confers Resistance to CD95-Induced Apoptosis in Hematopoietic Cancer Cell Lines," *J. Immunol.*, vol. 168, no. 5, pp. 2544–2553, 2002.

[77] R. Kalluri and E. G. Neilson, "Epithelial-mesenchymal transition and its implications for fibrosis," *Journal of Clinical Investigation*, vol. 112, no. 12. pp. 1776–1784, 2003.

[78] M. Zeisberg and E. G. Neilson, "Biomarkers for epithelial-mesenchymal transitions," *Journal of Clinical Investigation*, vol. 119, no. 6. pp. 1429–1437, 2009.

[79] E. D. Hay, "An overview of epithelio-mesenchymal transformation.," *Acta Anat. (Basel).*, vol. 154, no. 1, pp. 8–20, 1995.

[80] P. Liu, M. Wakamiya, M. J. Shea, U. Albrecht, R. R. Behringer, and A. Bradley, "Requirement for Wnt3 in vertebrate axis formation.," *Nat. Genet.*, vol. 22, no. 4, pp. 361–365, 1999.

[81] H. K. Chea, C. V. Wright, and B. J. Swalla, "Nodal signaling and the evolution of deuterostome gastrulation," *Developmental Dynamics*, vol. 234, no. 2. pp. 269–278, 2005.

[82] I. Skromne and C. D. Stern, "Interactions between Wnt and Vg1 signalling pathways initiate primitive streak formation in the chick embryo.," *Development*, vol. 128, no. 15, pp. 2915–2927, 2001.

[83] E. L. Whiteman, C.-J. Liu, E. R. Fearon, and B. Margolis, "The transcription factor snail represses Crumbs3 expression and disrupts apico-basal polarity complexes.," *Oncogene*, vol. 27, no. 27, pp. 3875–9, 2008.

[84] H. Okada, T. M. Danoff, R. Kalluri, and E. G. Neilson, "Early role of Fsp1 in epithelial-mesenchymal transformation.," *Am. J. Physiol.*, vol. 273, no. 4 Pt 2, pp. F563–F574, 1997.

[85] F. Strutz, M. Zeisberg, F. N. Ziyadeh, C.-Q. Yang, R. Kalluri, G. a Müller, and E. G. Neilson, "Role of basic fibroblast growth factor-2 in epithelial-mesenchymal transformation.," *Kidney Int.*, vol. 61, no. 5, pp. 1714–1728, 2002.

[86] J. Yang and R. A. Weinberg, "Epithelial-Mesenchymal Transition: At the Crossroads of Development and Tumor Metastasis," *Developmental Cell*, vol. 14, no. 6. pp. 818–829, 2008.

[87] J. P. Thiery, "Epithelial-mesenchymal transitions in tumour progression.," *Nat. Rev. Cancer*, vol. 2, no. 6, pp. 442–454, 2002.

[88] B. Smith and N. Bhowmick, "Role of EMT in Metastasis and Therapy Resistance," *J. Clin. Med.*, vol. 5, no. 2, p. 17, 2016.

[89] C. L. Neal, D. Mckeithen, and V. A. Odero-Marah, "Snail negatively regulates cell adhesion to extracellular matrix and integrin expression via the MAPK pathway in prostate cancer cells.," *Cell Adh. Migr.*, vol. 5, no. 3, pp. 249–57, 2011.

[90] Y. Uchikado, H. Okumura, S. Ishigami, T. Setoyama, M. Matsumoto, T. Owaki, Y. Kita, and S. Natsugoe, "Increased Slug and decreased E-cadherin expression is related to poor prognosis in patients with gastric cancer.," *Gastric Cancer*, vol. 14, no. 1, pp. 41–9, 2011.

[91] G.-J. Zhang, T. Zhou, H.-P. Tian, Z.-L. Liu, and S.-S. Xia, "High expression of ZEB1 correlates with liver metastasis and poor prognosis in colorectal cancer.," *Oncol. Lett.*, vol. 5, no. 2, pp. 564–568, 2013.

[92] S. B. Jakowlew, "Transforming growth factor-beta in cancer and metastasis.," *Cancer Metastasis Rev.*, vol. 25, no. 3, pp. 435–57, 2006.

[93] M. Oft, K.-H. Heider, and H. Beug, "TGFβ signaling is necessary for carcinoma cell invasiveness and metastasis," *Curr. Biol.*, vol. 8, no. 23, pp. 1243–1252, 1998.

[94] E. Piek, a Moustakas, a Kurisaki, C. H. Heldin, and P. ten Dijke, "TGF-(beta) type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells.," *J. Cell Sci.*, vol. 112 (Pt 2, pp. 4557–4568, 1999.

[95] S. Heerboth, G. Housman, M. Leary, M. Longacre, S. Byler, K. Lapinska, A. Willbanks, and S. Sarkar, "EMT and tumor metastasis," *Clin. Transl. Med.*, vol. 4, no. 1, p. 6, 2015.

[96] J. R. Neil, K. M. Johnson, R. A. Nemenoff, and W. P. Schiemann, "Cox-2 inactivates Smad signaling and enhances EMT stimulated by TGF-?? through a PGE2-dependent mechanisms," *Carcinogenesis*,

vol. 29, no. 11, pp. 2227–2235, 2008.

[97] S. J. Grille, A. Bellacosa, J. Upson, A. J. Klein-Szanto, F. Van Roy, W. Lee-Kwon, M. Donowitz, P. N. Tsichlis, and L. Larue, "The protein kinase Akt induces epithelial mesenchymal transition and promotes enhanced motility and invasiveness of squamous cell carcinoma lines," *Cancer Res.*, vol. 63, no. 9, pp. 2172–2178, 2003.

[98] R. Vogelmann and W. J. Nelson, "Fractionation of the epithelial apical junctional complex: reassessment of protein distributions in different substructures.," *Mol. Biol. Cell*, vol. 16, no. 2, pp. 701–16, 2005.

[99] A. Hartsock and W. J. Nelson, "Adherens and tight junctions: Structure, function and connections to the actin cytoskeleton," *Biochimica et Biophysica Acta - Biomembranes*, vol. 1778, no. 3. pp. 660–669, 2008.

[100] V. Bolós, H. Peinado, M. a Pérez-Moreno, M. F. Fraga, M. Esteller, and A. Cano, "The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors.," *J. Cell Sci.*, vol. 116, no. Pt 3, pp. 499–511, 2003.

[101] N. Herranz, D. Pasini, V. M. Díaz, C. Francí, A. Gutierrez, N. Dave, M. Escrivà, I. Hernandez-Muñoz, L. Di Croce, K. Helin, A. García de Herreros, and S. Peiró, "Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor.," *Mol. Cell. Biol.*, vol. 28, no. 15, pp. 4772–81, 2008.

[102] G. Maeda, T. Chiba, T. Aoba, and K. Imai, "Epigenetic inactivation of E-cadherin by promoter hypermethylation in oral carcinoma cells," *Odontology*, vol. 95, no. 1, pp. 24–29, 2007.

[103] J. Zhang and L. Ma, "MicroRNA control of epithelial-mesenchymal transition and metastasis," *Cancer Metastasis Rev.*, vol. 31, no. 3–4, pp. 653–662, 2012.

[104] U. H. Frixen, J. Behrens, M. Sachs, G. Eberle, B. Voss, A. Warda, D. Lochner, and W. Birchmeier, "E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells," *J. Cell Biol.*, vol. 113, no. 1, pp. 173–185, 1991.

[105] S. Dorudi, J. P. Sheffield, R. Poulsom, J. M. Northover, and I. R. Hart, "E-cadherin expression in colorectal cancer. An immunocytochemical and in situ hybridization study.," *Am. J. Pathol.*, vol. 142, no. 4, pp. 981–6, 1993.

[106] T. T. Onder, P. B. Gupta, S. A. Mani, J. Yang, E. S. Lander, and R. A. Weinberg, "Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways," *Cancer Res.*, vol. 68, no. 10, pp. 3645–3654, 2008.

[107] T. Reya, S. J. Morrison, M. F. Clarke, and I. L. Weissman, "Stem cells, cancer, and cancer stem cells.," *Nature*, vol. 414, no. 6859, pp. 105–111, 2001.

[108] L. N. Abdullah and E. K.-H. Chow, "Mechanisms of chemoresistance in cancer stem cells.," *Clin. Transl. Med.*, vol. 2, no. 1, p. 3, 2013.

[109] M. Bartucci, R. Dattilo, C. Moriconi, a Pagliuca, M. Mottolese, G. Federici, a Di Benedetto, M. Todaro, G. Stassi, F. Sperati, M. I. Amabile, E. Pilozzi, M. Patrizii, M. Biffoni, M. Maugeri-Saccà, S. Piccolo, and R. De Maria, "TAZ is required for metastatic activity and chemoresistance of breast cancer stem cells.," *Oncogene*, no. July 2013, pp. 1–10, 2014.

[110] K. M. Britton, J. A. Kirby, T. W. J. Lennard, and A. P. Meeson, "Cancer stem cells and side population

cells in breast cancer and metastasis," *Cancers*, vol. 3, no. 2. pp. 2106–2130, 2011.

[111] W.-T. Liao, Y.-P. Ye, Y.-J. Deng, X.-W. Bian, and Y.-Q. Ding, "Metastatic cancer stem cells: from the concept to therapeutics.," *Am. J. Stem Cells*, vol. 3, no. 2, pp. 46–62, 2014.

[112] W.-T. Chiu, Y.-F. Huang, H.-Y. Tsai, C.-C. Chen, C.-H. Chang, S.-C. Huang, K.-F. Hsu, and C.-Y. Chou, "FOXM1 confers to epithelial-mesenchymal transition, stemness and chemoresistance in epithelial ovarian carcinoma cells.," *Oncotarget*, vol. 6, no. 4, pp. 2349–65, 2015.

[113] A.-P. Morel, M. Lièvre, C. Thomas, G. Hinkal, S. Ansieau, and A. Puisieux, "Generation of breast cancer stem cells through epithelial-mesenchymal transition.," *PLoS One*, vol. 3, no. 8, p. e2888, 2008.

[114] U. Wellner, J. Schubert, U. C. Burk, O. Schmalhofer, F. Zhu, A. Sonntag, B. Waldvogel, C. Vannier, D. Darling, A. zur Hausen, V. G. Brunton, J. Morton, O. Sansom, J. Schüler, M. P. Stemmler, C. Herzberger, U. Hopt, T. Keck, S. Brabletz, and T. Brabletz, "The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs," *Nat. Cell Biol.*, vol. 11, no. 12, pp. 1487–1495, 2009.

[115] H. Dang, W. Ding, D. Emerson, and C. B. Rountree, "Snail1 induces epithelial-to-mesenchymal transition and tumor initiating stem cell characteristics," *BMC Cancer*, vol. 11, no. 1, p. 396, 2011.

[116] N. K. Kurrey, S. P. Jalgaonkar, A. V. Joglekar, A. D. Ghanate, P. D. Chaskar, R. Y. Doiphode, and S. A. Bapat, "Snail and slug mediate radioresistance and chemoresistance by antagonizing p53-mediated apoptosis and acquiring a stem-like phenotype in ovarian cancer cells," *Stem Cells*, vol. 27, no. 9, pp. 2059–2068, 2009.

[117] J. M. Schmidt, E. Panzilius, H. S. Bartsch, M. Irmler, J. Beckers, V. Kari, J. R. Linnemann, D. Dragoi, B. Hirschi, U. J. Kloos, S. Sass, F. Theis, S. Kahlert, S. A. Johnsen, K. Sotlar, and C. H. Scheel, "Stem-cell-like properties and epithelial plasticity arise as stable traits after transient twist1 activation," *Cell Rep.*, vol. 10, no. 2, pp. 131–139, 2015.

[118] T. Brabletz, "To differentiate or not — routes towards metastasis," *Nat. Rev. Cancer*, vol. 12, no. 6, pp. 425–436, 2012.

[119] J. M. Lee, S. Dedhar, R. Kalluri, and E. W. Thompson, "The epithelial-mesenchymal transition: new insights in signaling, development, and disease.," *J. Cell Biol.*, vol. 172, no. 7, pp. 973–81, 2006.

[120] T. Brabletz, F. Hlubek, S. Spaderna, O. Schmalhofer, E. Hiendlmeyer, A. Jung, and T. Kirchner, "Invasion and metastasis in colorectal cancer: epithelial-mesenchymal transition, mesenchymal-epithelial transition, stem cells and beta-catenin.," *Cells. Tissues. Organs*, vol. 179, no. 1–2, pp. 56–65, 2005.

[121] L. Hill, G. Browne, and E. Tulchinsky, "ZEB/miR-200 feedback loop: At the crossroads of signal transduction in cancer," *International Journal of Cancer*, vol. 132, no. 4. pp. 745–754, 2012.

[122] S. Brabletz and T. Brabletz, "The ZEB/miR-200 feedback loop--a motor of cellular plasticity in development and cancer?," *EMBO Rep.*, vol. 11, no. 9, pp. 670–677, 2010.

[123] P. A. Gregory, A. G. Bert, E. L. Paterson, S. C. Barry, A. Tsykin, G. Farshid, M. A. Vadas, Y. Khew-Goodall, and G. J. Goodall, "The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1.," *Nat. Cell Biol.*, vol. 10, no. 5, pp. 593–601, 2008.

[124] D. Iliopoulos, M. Lindahl-Allen, C. Polytarchou, H. A. Hirsch, P. N. Tsichlis, and K. Struhl, "Loss of

miR-200 Inhibition of Suz12 Leads to Polycomb-Mediated Repression Required for the Formation and Maintenance of Cancer Stem Cells," *Mol. Cell*, vol. 39, no. 5, pp. 761–772, 2010.

[125] Y. Shimono, M. Zabala, R. W. Cho, N. Lobo, P. Dalerba, D. Qian, M. Diehn, H. Liu, S. P. Panula, E. Chiao, F. M. Dirbas, G. Somlo, R. A. R. Pera, K. Lao, and M. F. Clarke, "Downregulation of miRNA-200c Links Breast Cancer Stem Cells with Normal Stem Cells," *Cell*, vol. 138, no. 3, pp. 592–603, 2009.

[126] H. Siemens, R. Jackstadt, S. Hünten, M. Kaller, A. Menssen, U. Götz, and H. Hermeking, "miR-34 and SNAIL form a double-negative feedback loop to regulate epithelial-mesenchymal transitions," *Cell Cycle*, vol. 10, no. 24, pp. 4256–4271, 2011.

[127] A. Jung, M. Schrauder, U. Oswald, C. Knoll, P. Sellberg, R. Palmqvist, G. Niedobitek, T. Brabletz, and T. Kirchner, "The invasion front of human colorectal adenocarcinomas shows co-localization of nuclear beta-catenin, cyclin D1, and p16INK4A and is a region of low proliferation.," *Am. J. Pathol.*, vol. 159, no. 5, pp. 1613–7, 2001.

[128] M. S. Sosa, P. Bragado, and J. A. Aguirre-Ghiso, "Mechanisms of disseminated cancer cell dormancy: an awakening field," *Nat. Rev. Cancer*, vol. 14, no. 9, pp. 611–622, 2014.

[129] H. Gao, G. Chakraborty, A. P. Lee-Lim, Q. Mo, M. Decker, A. Vonica, R. Shen, E. Brogi, A. H. Brivanlou, and F. G. Giancotti, "The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites," *Cell*, vol. 150, no. 4, pp. 764–779, 2012.

[130] D. Barkan, L. H. El Touny, A. M. Michalowski, J. A. Smith, I. Chu, A. S. Davis, J. D. Webster, S. Hoover, R. M. Simpson, J. Gauldie, and J. E. Green, "Metastatic growth from dormant cells induced by a Col-I-enriched fibrotic environment," *Cancer Res.*, vol. 70, no. 14, pp. 5706–5716, 2010.

[131] Z. Elsawaf and H.-P. Sinn, "Triple-Negative Breast Cancer: Clinical and Histological Correlations.," *Breast Care (Basel).*, vol. 6, no. 4, pp. 273–278, 2011.

[132] H. C. Lien, Y. H. Hsiao, Y. S. Lin, Y. T. Yao, H. F. Juan, W. H. Kuo, M.-C. Hung, K. J. Chang, and F. J. Hsieh, "Molecular signatures of metaplastic carcinoma of the breast by large-scale transcriptional profiling: identification of genes potentially related to epithelial-mesenchymal transition.," *Oncogene*, vol. 26, no. 57, pp. 7859–71, 2007.

[133] D. Sarrió, S. M. Rodriguez-Pinilla, D. Hardisson, A. Cano, G. Moreno-Bueno, and J. Palacios, "Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype," *Cancer Res.*, vol. 68, no. 4, pp. 989–997, 2008.

[134] C. J. Creighton, X. Li, M. Landis, J. M. Dixon, V. M. Neumeister, A. Sjolund, D. L. Rimm, H. Wong, A. Rodriguez, J. I. Herschkowitz, C. Fan, X. Zhang, X. He, A. Pavlick, M. C. Gutierrez, L. Renshaw, A. A. Larionov, D. Faratian, S. G. Hilsenbeck, C. M. Perou, M. T. Lewis, J. M. Rosen, and J. C. Chang, "Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 33, pp. 13820–13825, 2009.

[135] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D.

Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, and C. International Human Genome Sequencing, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[136] M. Hattori, "Finishing the euchromatic sequence of the human genome," *Tanpakushitsu Kakusan Koso.*, vol. 50, no. 2, pp. 162–168, 2005.

[137] M. Clamp, B. Fry, M. Kamal, X. Xie, J. Cuff, M. F. Lin, M. Kellis, K. Lindblad-Toh, and E. S. Lander, "Distinguishing protein-coding and noncoding genes in the human genome.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 49, pp. 19428–33, 2007.

[138] B. R. Graveley, "Alternative splicing: Increasing diversity in the proteomic world," *Trends in Genetics*, vol. 17, no. 2. pp. 100–107, 2001.

[139] C. Y. Chen, S. T. Chen, H. F. Juan, and H. C. Huang, "Lengthening of 3′UTR increases with morphological complexity in animal evolution," *Bioinformatics*, vol. 28, no. 24, pp. 3178–3181, 2012.

[140] J. S. Mattick, "Challenging the dogma: The hidden layer of non-protein-coding RNAs in complex organisms," *BioEssays*, vol. 25, no. 10. pp. 930–939, 2003.

[141] K. C. Pang, M. C. Frith, and J. S. Mattick, "Rapid evolution of noncoding RNAs: Lack of conservation

does not mean lack of function," *Trends in Genetics*, vol. 22, no. 1. pp. 1–5, 2006.

[142] Y. Lee, M. Kim, J. Han, K.-H. Yeom, S. Lee, S. H. Baek, and V. N. Kim, "MicroRNA genes are transcribed by RNA polymerase II.," *EMBO J.*, vol. 23, no. 20, pp. 4051–60, 2004.

[143] D. P. Bartel, "MicroRNAs: Genomics, Biogenesis, Mechanism, and Function," *Cell*, vol. 116, no. 2. pp. 281–297, 2004.

[144] X. Cai, C. H. Hagedorn, and B. R. Cullen, "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs.," *RNA*, vol. 10, no. 12, pp. 1957–66, 2004.

[145] R. I. Gregory, K.-P. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar, "The Microprocessor complex mediates the genesis of microRNAs.," *Nature*, vol. 432, no. 7014, pp. 235–240, 2004.

[146] R. F. Ketting, S. E. J. Fischer, E. Bernstein, T. Sijen, G. J. Hannon, and R. H. A. Plasterk, "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans," *Genes Dev.*, vol. 15, no. 20, pp. 2654–2659, 2001.

[147] N. Cloonan, "Re-thinking miRNA-mRNA interactions: Intertwining issues confound target discovery," *BioEssays*, vol. 37, no. 4, pp. 379–388, 2015.

[148] R. S. Pillai, S. N. Bhattacharyya, and W. Filipowicz, "Repression of protein synthesis by miRNAs: how many mechanisms?," *Trends in Cell Biology*, vol. 17, no. 3. pp. 118–126, 2007.

[149] S. N. Bhattacharyya, R. Habermacher, U. Martine, E. I. Closs, and W. Filipowicz, "Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress," *Cell*, vol. 125, no. 6, pp. 1111–1124, 2006.

[150] A. M. Krichevsky, K. S. King, C. P. Donahue, K. Khrapko, and K. S. Kosik, "A microRNA array reveals extensive regulation of microRNAs during brain development.," *RNA*, vol. 9, no. 10, pp. 1274–81, 2003.

[151] C.-G. Liu, G. A. Calin, B. Meloon, N. Gamliel, C. Sevignani, M. Ferracin, C. D. Dumitru, M. Shimizu, S. Zupo, M. Dono, H. Alder, F. Bullrich, M. Negrini, and C. M. Croce, "An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 26, pp. 9740–4, 2004.

[152] J. H. Mansfield, B. D. Harfe, R. Nissen, J. Obenauer, J. Srineel, A. Chaudhuri, R. Farzan-Kashani, M. Zuker, A. E. Pasquinelli, G. Ruvkun, P. a Sharp, C. J. Tabin, and M. T. McManus, "MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression.," *Nat. Genet.*, vol. 36, no. 19, pp. 1079–1083, 2004.

[153] Q. Jiang, Y. Wang, Y. Hao, L. Juan, M. Teng, X. Zhang, M. Li, G. Wang, and Y. Liu, "miR2Disease: A manually curated database for microRNA deregulation in human disease," *Nucleic Acids Res.*, vol. 37, no. SUPPL. 1, 2009.

[154] A. Carè, D. Catalucci, F. Felicetti, D. Bonci, A. Addario, P. Gallo, M.-L. Bang, P. Segnalini, Y. Gu, N. D. Dalton, L. Elia, M. V. G. Latronico, M. Høydal, C. Autore, M. a Russo, G. W. Dorn, O. Ellingsen, P. Ruiz-Lozano, K. L. Peterson, C. M. Croce, C. Peschle, and G. Condorelli, "MicroRNA-133 controls cardiac hypertrophy.," *Nat. Med.*, vol. 13, no. 5, pp. 613–8, 2007.

[155] M. S. Nicoloso, R. Spizzo, M. Shimizu, S. Rossi, and G. a Calin, "MicroRNAs--the micro steering wheel

of tumour metastases.," *Nat. Rev. Cancer*, vol. 9, no. 4, pp. 293–302, 2009.

[156] S. F. Tavazoie, C. Alarcón, T. Oskarsson, D. Padua, Q. Wang, P. D. Bos, W. L. Gerald, and J. Massagué, "Endogenous human microRNAs that suppress breast cancer metastasis.," *Nature*, vol. 451, no. 7175, pp. 147–52, 2008.

[157] I. a Asangani, S. a K. Rasheed, D. a Nikolova, J. H. Leupold, N. H. Colburn, S. Post, and H. Allgayer, "MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer.," *Oncogene*, vol. 27, no. 15, pp. 2128–36, 2008.

[158] A. J. Schetter, S. Y. Leung, J. J. Sohn, K. a Zanetti, E. D. Bowman, N. Yanaihara, S. T. Yuen, T. L. Chan, D. L. W. Kwong, G. K. H. Au, C.-G. Liu, G. a Calin, C. M. Croce, and C. C. Harris, "MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma.," *JAMA*, vol. 299, no. 4, pp. 425–36, 2008.

[159] A. G. Seto, R. E. Kingston, and N. C. Lau, "The Coming of Age for Piwi Proteins," *Molecular Cell*, vol. 26, no. 5. pp. 603–609, 2007.

[160] J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G. J. Hannon, "Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila," *Cell*, vol. 128, no. 6, pp. 1089–1103, 2007.

[161] C. Juliano, J. Wang, and H. Lin, "Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms," *Annu Rev Genet*, vol. 45, pp. 447–469, 2011.

[162] M. Ha and V. N. Kim, "Regulation of microRNA biogenesis," *Nat. Rev. Mol. Cell Biol.*, vol. 15, no. 8, pp. 509–524, 2014.

[163] N. Agrawal, P. V. N. Dasaradhi, A. Mohmmed, P. Malhotra, R. K. Bhatnagar, and S. K. Mukherjee, "RNA interference: biology, mechanism, and applications.," *Microbiol. Mol. Biol. Rev.*, vol. 67, no. 4, pp. 657–85, 2003.

[164] O. H. Tam, A. A. Aravin, P. Stein, A. Girard, E. P. Murchison, S. Cheloufi, E. Hodges, M. Anger, R. Sachidanandam, R. M. Schultz, and G. J. Hannon, "Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes.," *Nature*, vol. 453, no. 7194, pp. 534–8, 2008.

[165] P. Stein, N. V. Rozhkov, F. Li, F. L. C??rdenas, O. Davydenko, O. Davydenk, L. E. Vandivier, B. D. Gregory, G. J. Hannon, and R. M. Schultz, "Essential Role for endogenous siRNAs during meiosis in mouse oocytes," *PLoS Genet.*, vol. 11, no. 2, p. e1005013, 2015.

[166] J. J. Quinn and H. Y. Chang, "Unique features of long non-coding RNA biogenesis and function.," *Nat. Rev. Genet.*, vol. 17, no. 1, pp. 47–62, 2015.

[167] Y. Devaux, J. Zangrando, B. Schroen, E. E. Creemers, T. Pedrazzini, C.-P. Chang, G. W. Dorn, T. Thum, and S. Heymans, "Long noncoding RNAs in cardiac development and ageing.," *Nat. Rev. Cardiol.*, vol. 12, no. 7, pp. 415–25, 2015.

[168] M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn, and E. S. Lander, "Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.," *Nature*, vol. 458, no. 7235, pp. 223–7, 2009.

[169] J. T. Y. Kung, D. Colognori, and J. T. Lee, "Long noncoding RNAs: Past, present, and future," *Genetics*,

vol. 193, no. 3. pp. 651–669, 2013.

[170] J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, and H. Y. Chang, "Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs," *Cell*, vol. 129, no. 7, pp. 1311–1323, 2007.

[171] J. Grant, S. K. Mahadevaiah, P. Khil, M. N. Sangrithi, H. Royo, J. Duckworth, J. R. McCarrey, J. L. VandeBerg, M. B. Renfree, W. Taylor, G. Elgar, R. D. Camerini-Otero, M. J. Gilchrist, and J. M. A. Turner, "Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation," *Nature*, vol. 487, no. 7406, pp. 254–258, 2012.

[172] M. A. Faghihi, F. Modarresi, A. M. Khalil, D. E. Wood, B. G. Sahagan, T. E. Morgan, C. E. Finch, G. St Laurent, P. J. Kenny, and C. Wahlestedt, "Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase.," *Nat. Med.*, vol. 14, no. 7, pp. 723–730, 2008.

[173] R. E. Mills, E. A. Bennett, R. C. Iskow, and S. E. Devine, "Which transposable elements are active in the human genome?," *Trends in Genetics*, vol. 23, no. 4. pp. 183–191, 2007.

[174] R. Cordaux and M. A. Batzer, "The impact of retrotransposons on human genome evolution.," *Nat. Rev. Genet.*, vol. 10, no. 10, pp. 691–703, 2009.

[175] J. F. McDonald, "Evolution and consequences of transposable elements," *Curr. Opin. Genet. Dev.*, vol. 3, no. 6, pp. 855–864, 1993.

[176] Z. Ivics and Z. Izsvák, "The expanding universe of transposon technologies for gene and cell engineering.," *Mob. DNA*, vol. 1, no. 1, p. 25, 2010.

[177] R. H. Plasterk, "The Tc1/mariner transposon family.," *Curr. Top. Microbiol. Immunol.*, vol. 204, no. 1, pp. 125–43, 1996.

[178] D. J. Lampe, M. E. Churchill, and H. M. Robertson, "A purified mariner transposase is sufficient to mediate transposition in vitro.," *EMBO J.*, vol. 15, no. 19, pp. 5470–9, 1996.

[179] Z. Izsvák, Z. Ivics, and P. B. Hackett, "Characterization of a Tc1-like transposable element in zebrafish (Danio rerio)," *MGG Mol. Gen. Genet.*, vol. 247, no. 3, pp. 312–322, 1995.

[180] I. Mori, G. M. Benian, D. G. Moerman, and R. H. Waterston, "Transposable element Tc1 of Caenorhabditis elegans recognizes specific target sequences for integration," *Proc Natl Acad Sci U S A*, vol. 85, no. 3, pp. 861–864, 1988.

[181] A. E. Carpenter and D. M. Sabatini, "Systematic genome-wide screens of gene function.," *Nat. Rev. Genet.*, vol. 5, no. 1, pp. 11–22, 2004.

[182] S. L. Forsburg, "The art and design of genetic screens: yeast.," *Nat. Rev. Genet.*, vol. 2, no. 9, pp. 659–668, 2001.

[183] D. St Johnston, "The art and design of genetic screens: Drosophila melanogaster," *Nat. Rev. Genet.*, vol. 3, no. 3, pp. 176–188, 2002.

[184] E. M. Jorgensen and S. E. Mango, "the Art and Design of Genetic Screens: Caenorhabditis Elegans," *Nat. Rev. Genet.*, vol. 3, no. 5, pp. 356–369, 2002.

[185] B. T. Kile and D. J. Hilton, "The art and design of genetic screens: mouse.," *Nat. Rev. Genet.*, vol. 6, no. 7, pp. 557–67, 2005.

[186] S. Grimm, "The art and design of genetic screens: mammalian culture cells.," *Nat. Rev. Genet.*, vol. 5,

no. 3, pp. 179–89, 2004.

[187] M. L. Schmitz and P. A. Baeuerle, "The p65 subunit is responsible for the strong transcription activating potential of NF-kappa B.," *EMBO J.*, vol. 10, no. 12, pp. 3805–3817, 1991.

[188] a G. Uren, J. Kool, A. Berns, and M. van Lohuizen, "Retroviral insertional mutagenesis: past, present and future," *Oncogene*, vol. 24, no. 52, pp. 7656–7672, 2005.

[189] C. J. Echeverri and N. Perrimon, "High-throughput RNAi screening in cultured cells: a user's guide," *Nat. Rev. Genet.*, vol. 7, no. 5, pp. 373–384, 2006.

[190] J. Li, S. C. Ding, H. Cho, B. C. Chung, G. Michael, S. K. Chanda, and M. S. Diamond, "A short hairpin RNA screen of interferon-stimulated genes identifies a novel negative regulator of the cellular antiviral response," *MBio*, vol. 4, no. 3, 2013.

[191] B. Seed, "Developments in expression cloning," *Curr. Opin. Biotechnol.*, vol. 6, no. 5, pp. 567–573, 1995.

[192] S. Zhu, H. Wurdak, Y. Wang, A. Galkin, H. Tao, J. Li, C. A. Lyssiotis, F. Yan, B. P. Tu, L. Miraglia, J. Walker, F. Sun, A. Orth, P. G. Schultz, and X. Wu, "A genomic screen identifies TYRO3 as a MITF regulator in melanoma.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, pp. 17025–17030, 2009.

[193] A. T. Ting, F. X. Pimentel-Muinos, and B. Seed, "RIP mediates tumor necrosis factor receptor 1 activation of NF-kappaB but not Fas/APO-1-initiated apoptosis," *Embo J*, vol. 15, no. 22, pp. 6189–6196, 1996.

[194] Z. Ivics, P. B. Hackett, R. H. Plasterk, and Z. Izsvák, "Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells," *Cell*, vol. 91, no. 4, pp. 501–510, 1997.

[195] T. J. Vigdal, C. D. Kaufman, Z. Izsvák, D. F. Voytas, and Z. Ivics, "Common physical properties of DNA affecting target site selection of Sleeping Beauty and other Tc1/mariner transposable elements," *J. Mol. Biol.*, vol. 323, no. 3, pp. 441–452, 2002.

[196] M. J. Fraser, T. Ciszczon, T. Elick, and C. Bauser, "Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera.," *Insect Mol. Biol.*, vol. 5, no. 2, pp. 141–51, 1996.

[197] T. a Elick, C. a Bauser, and M. J. Fraser, "Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase.," *Genetica*, vol. 98, no. 1, pp. 33–41, 1996.

[198] C. Kokubu, K. Horie, K. Abe, R. Ikeda, S. Mizuno, Y. Uno, S. Ogiwara, M. Ohtsuka, A. Isotani, M. Okabe, K. Imai, and J. Takeda, "A transposon-based chromosomal engineering method to survey a large cis-regulatory landscape in mice.," *Nat. Genet.*, vol. 41, no. 8, pp. 946–952, 2009.

[199] W. Wang, A. Bradley, and Y. Huang, "A piggyBac transposon-based genome-wide library of insertionally mutated Blm-deficient murine ES cells," *Genome Res.*, vol. 19, no. 4, pp. 667–673, 2009.

[200] S. Ding, X. Wu, G. Li, M. Han, Y. Zhuang, and T. Xu, "Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice," *Cell*, vol. 122, no. 3, pp. 473–483, 2005.

[201] S. Zhao, E. Jiang, S. Chen, Y. Gu, A. J. Shangguan, T. Lv, L. Luo, and Z. Yu, "PiggyBac transposon vectors: the tools of the human gene encoding.," *Transl. lung cancer Res.*, vol. 5, no. 1, pp. 120–125,

2016.

[202] W. L. Stanford, J. B. Cohn, and S. P. Cordes, “Gene-trap mutagenesis: past, present and beyond.,” *Nat. Rev. Genet.*, vol. 2, no. 10, pp. 756–68, 2001.

[203] C. Miskey, Z. Izsvák, K. Kawakami, and Z. Ivics, “DNA transposons in vertebrate functional genomics,” *Cellular and Molecular Life Sciences*, vol. 62, no. 6. pp. 629–641, 2005.

[204] N. G. Copeland and N. A. Jenkins, “Harnessing transposons for cancer gene discovery.,” *Nat. Rev. Cancer*, vol. 10, no. 10, pp. 696–706, 2010.

[205] T. K. Starr, R. Allaei, K. A. Silverstein, R. A. Staggs, A. L. Sarver, T. L. Bergemann, M. Gupta, M. G. O’Sullivan, I. Matise, A. J. Dupuy, L. S. Collier, S. Powers, A. L. Oberg, Y. W. Asmann, S. N. Thibodeau, L. Tessarollo, N. G. Copeland, N. A. Jenkins, R. T. Cormier, and D. A. Largaespada, “A transposon-based genetic screen in mice identifies genes altered in colorectal cancer,” *Science (80-.).*, vol. 323, no. 5922, pp. 1747–1750, 2009.

[206] P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells, R. Kodzius, K. Shimokawa, V. B. Bajic, S. E. Brenner, S. Batalov, A. R. Forrest, M. Zavolan, M. J. Davis, L. G. Wilming, V. Aidinis, J. E. Allen, A. Ambesi-Impiombato, R. Apweiler, R. N. Aturaliya, T. L. Bailey, M. Bansal, L. Baxter, K. W. Beisel, T. Bersano, H. Bono, A. M. Chalk, K. P. Chiu, V. Choudhary, A. Christoffels, D. R. Clutterbuck, M. L. Crowe, E. Dalla, B. P. Dalrymple, B. de Bono, G. Della Gatta, D. di Bernardo, T. Down, P. Engstrom, M. Fagiolini, G. Faulkner, C. F. Fletcher, T. Fukushima, M. Furuno, S. Futaki, M. Gariboldi, P. Georgii-Hemming, T. R. Gingeras, T. Gojobori, R. E. Green, S. Gustincich, M. Harbers, Y. Hayashi, T. K. Hensch, N. Hirokawa, D. Hill, L. Huminiecki, M. Iacono, K. Ikeo, A. Iwama, T. Ishikawa, M. Jakt, A. Kanapin, M. Katoh, Y. Kawasawa, J. Kelso, H. Kitamura, H. Kitano, G. Kollias, S. P. Krishnan, A. Kruger, S. K. Kummerfeld, I. V Kurochkin, L. F. Lareau, D. Lazarevic, L. Lipovich, J. Liu, S. Liuni, S. McWilliam, M. Madan Babu, M. Madera, L. Marchionni, H. Matsuda, S. Matsuzawa, H. Miki, F. Mignone, S. Miyake, K. Morris, S. Mottagui-Tabar, N. Mulder, N. Nakano, H. Nakauchi, P. Ng, R. Nilsson, S. Nishiguchi, S. Nishikawa, F. Nori, O. Ohara, Y. Okazaki, V. Orlando, K. C. Pang, W. J. Pavan, G. Pavesi, G. Pesole, N. Petrovsky, S. Piazza, J. Reed, J. F. Reid, B. Z. Ring, M. Ringwald, B. Rost, Y. Ruan, S. L. Salzberg, A. Sandelin, C. Schneider, C. Schonbach, K. Sekiguchi, C. A. Semple, S. Seno, L. Sessa, Y. Sheng, Y. Shibata, H. Shimada, K. Shimada, D. Silva, B. Sinclair, S. Sperling, E. Stupka, K. Sugiura, R. Sultana, Y. Takenaka, K. Taki, K. Tammoja, S. L. Tan, S. Tang, M. S. Taylor, J. Tegner, S. A. Teichmann, H. R. Ueda, E. van Nimwegen, R. Verardo, C. L. Wei, K. Yagi, H. Yamanishi, E. Zabarovsky, S. Zhu, A. Zimmer, W. Hide, C. Bult, S. M. Grimmond, R. D. Teasdale, E. T. Liu, V. Brusic, J. Quackenbush, C. Wahlestedt, J. S. Mattick, D. A. Hume, C. Kai, D. Sasaki, Y. Tomaru, S. Fukuda, M. Kanamori-Katayama, M. Suzuki, J. Aoki, T. Arakawa, J. Iida, K. Imamura, M. Itoh, T. Kato, H. Kawaji, N. Kawagashira, T. Kawashima, M. Kojima, S. Kondo, H. Konno, K. Nakano, N. Ninomiya, T. Nishio, M. Okada, C. Plessy, K. Shibata, T. Shiraki, S. Suzuki, M. Tagami, K. Waki, A. Watahiki, Y. Okamura-Oho, H. Suzuki, J. Kawai, Y. Hayashizaki, Fantom Consortium, Riken Genome Exploration Research Group, and Genome Science Group, “The transcriptional landscape of the mammalian genome,” *Science (80-.).*, vol. 309, no. 5740, pp. 1559–1563, 2005.

[207] J. S. Mattick, “The genetic signatures of noncoding RNAs,” *PLoS Genetics*, vol. 5, no. 4. 2009.

[208] G. A. Calin, C. gong Liu, M. Ferracin, T. Hyslop, R. Spizzo, C. Sevignani, M. Fabbri, A. Cimmino, E. J.

Lee, S. E. Wojcik, M. Shimizu, E. Tili, S. Rossi, C. Taccioli, F. Pichiorri, X. Liu, S. Zupo, V. Herlea, L. Gramantieri, G. Lanza, H. Alder, L. Rassenti, S. Volinia, T. D. D. Schmittgen, T. J. Kipps, M. Negrini, and C. M. Croce, "Ultraconserved Regions Encoding ncRNAs Are Altered in Human Leukemias and Carcinomas," *Cancer Cell*, vol. 12, no. 3, pp. 215–229, 2007.

[209] J. C. Neil and E. R. Cameron, "Retroviral insertion sites and cancer: Fountain of all knowledge?," *Cancer Cell*, vol. 2, no. 4. pp. 253–255, 2002.

[210] X. Wu, Y. Li, B. Crise, and S. S. M. Burgess, "Transcription start regions in the human genome are favored targets for MLV integration," *Science (80-.).*, vol. 300, no. 5626, pp. 1749–51, 2003.

[211] R. Rad, L. Rad, W. Wang, A. Strong, H. Ponstingl, I. F. Bronner, M. Mayho, K. Steiger, J. Weber, M. Hieber, C. Veltkamp, S. Eser, U. Geumann, R. Öllinger, M. Zukowska, M. Barenboim, R. Maresch, J. Cadiñanos, M. Friedrich, I. Varela, F. Constantino-Casas, A. Sarver, J. Ten Hoeve, H. Prosser, B. Seidler, J. Bauer, M. Heikenwälder, E. Metzakopian, A. Krug, U. Ehmer, G. Schneider, T. Knösel, P. Rümmele, D. Aust, R. Grützmann, C. Pilarsky, Z. Ning, L. Wessels, R. M. Schmid, M. A. Quail, G. Vassiliou, I. Esposito, P. Liu, D. Saur, and A. Bradley, "A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer.," *Nat. Genet.*, vol. 47, no. 1, pp. 47–56, 2015.

[212] Z. Ivics, M. A. Li, L. Mates, J. D. Boeke, A. Nagy, A. Bradley, and Z. Izsvak, "Transposon-mediated genome manipulation in vertebrates," *Nat. Methods*, vol. 6, no. 6, pp. 415–422, 2009.

[213] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis," *Nat. Methods*, vol. 9, no. 7, pp. 676–82, 2012.

[214] A. Untergasser, H. Nijveen, X. Rao, T. Bisseling, R. Geurts, and J. A. M. Leunissen, "Primer3Plus, an enhanced web interface to Primer3," *Nucleic Acids Res.*, vol. 35, no. SUPPL.2, 2007.

[215] D. A. Largaespada and L. S. Collier, "Transposon-mediated mutagenesis in somatic cells: Identification of transposon-genomic DNA junctions," *Methods Mol. Biol.*, vol. 435, pp. 95–108, 2008.

[216] S. Douma, T. van Laar, J. Zevenhoven, R. Meuwissen, E. van Garderen, and D. S. Peeper, "Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB," *Nature*, vol. 430, no. August, pp. 1034–1040, 2004.

[217] T. Onodera, T. Sakai, J. C. Hsu, K. Matsumoto, J. A. Chiorini, and K. M. Yamada, "Btbd7 regulates epithelial cell dynamics and branching morphogenesis.," *Science (80-.).*, vol. 329, no. 5991, pp. 562–565, 2010.

[218] Y. M. Tao, J. L. Huang, S. Zeng, S. Zhang, X. G. Fan, Z. M. Wang, H. X. Yang, X. H. Yuan, P. Wang, F. Wu, J. Luo, D. Y. Zeng, and H. Shen, "BTB/POZ domain-containing protein 7: Epithelial-mesenchymal transition promoter and prognostic biomarker of hepatocellular carcinoma," *Hepatology*, vol. 57, no. 6, pp. 2326–2337, 2013.

[219] C. Fan, Y. Miao, X. Zhang, D. Liu, G. Jiang, X. Lin, Q. Han, L. Luan, Z. Xu, and E. Wang, "Btbd7 contributes to reduced E-cadherin expression and predicts poor prognosis in non-small cell lung cancer.," *BMC Cancer*, vol. 14, no. 1, p. 704, 2014.

[220] H. Zhang, Y. Hao, J. Yang, Y. Zhou, J. Li, S. Yin, C. Sun, M. Ma, Y. Huang, and J. J. Xi, "Genome-wide functional screening of miR-23b as a pleiotropic modulator suppressing cancer metastasis," *Nat Commun*, vol. 2, no. May, p. 554, 2011.

[221] L. Pellegrino, J. Stebbing, V. M. Braga, A. E. Frampton, J. Jacob, L. Buluwela, L. R. Jiao, M. Periyasamy, C. D. Madsen, M. P. Caley, S. Ottaviani, L. Roca-Alonso, M. El-Bahrawy, R. C. Coombes, J. Krell, and L. Castellano, "miR-23b regulates cytoskeletal remodeling, motility and metastasis by directly targeting multiple transcripts," *Nucleic Acids Res.*, vol. 41, no. 10, pp. 5400–5412, 2013.

[222] S. Zollman, D. Godt, G. G. Privé, J. L. Couderc, and F. A. Laski, "The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 22, pp. 10717–21, 1994.

[223] P. J. Stogios and G. G. Privé, "The BACK domain in BTB-kelch proteins," *Trends in Biochemical Sciences*, vol. 29, no. 12. pp. 634–637, 2004.

[224] M. D. Resh, "Fatty acylation of proteins: New insights into membrane targeting of myristoylated and palmitoylated proteins," *Biochimica et Biophysica Acta - Molecular Cell Research*, vol. 1451, no. 1. pp. 1–16, 1999.

[225] Z. W. and C. Y.-G., "Specific activation of mitogen-activated protein kinase by transforming growth factor-beta receptors in lipid rafts is required for epithelial cell plasticity," *Molecular Biology of the Cell*, vol. 20, no. 3. pp. 1020–1029, 2009.

[226] A. F. Salem, G. Bonuccelli, G. Bevilacqua, H. Arafat, R. G. Pestell, F. Sotgia, and M. P. Lisanti, "Caveolin-1 promotes pancreatic cancer cell differentiation and restores membranous E-cadherin via suppression of the epithelial-mesenchymal transition a n d e s i o s c i e n c e o n o t d i s t r i b u t e," vol. 10, no. 21, pp. 3692–3700, 2011.

[227] S. V Litvinov, W. van Driel, C. M. van Rhijn, H. A. Bakker, H. van Krieken, G. J. Fleuren, and S. O. Warnaar, "Expression of Ep-CAM in cervical squamous epithelia correlates with an increased proliferation and the disappearance of markers for terminal differentiation.," *Am. J. Pathol.*, vol. 148, no. 3, pp. 865–75, 1996.

[228] X. Zheng, J. L. Carstens, J. Kim, M. Scheible, J. Kaye, H. Sugimoto, C.-C. Wu, V. S. LeBleu, and R. Kalluri, "Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer," *Nature*, vol. 527, no. 7579, pp. 525–530, 2015.

[229] K. R. Fischer, A. Durrans, S. Lee, J. Sheng, F. Li, S. T. C. Wong, H. Choi, T. El Rayes, S. Ryu, J. Troeger, R. F. Schwabe, L. T. Vahdat, N. K. Altorki, V. Mittal, and D. Gao, "Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance," *Nature*, vol. 527, no. 7579, pp. 472–476, 2015.

[230] E. Penna and D. Taverna, "In vivo extravasation assay," vol. 4, pp. 1–6, 2014.

[231] Q. Huang, K. Gumireddy, M. Schrier, C. le Sage, R. Nagel, S. Nair, D. a Egan, A. Li, G. Huang, A. J. Klein-Szanto, P. a Gimotty, D. Katsaros, G. Coukos, L. Zhang, E. Puré, and R. Agami, "The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis.," *Nat. Cell Biol.*, vol. 10, no. 2, pp. 202–210, 2008.

[232] M. Ouzounova, T. Vuong, P.-B. Ancey, M. Ferrand, G. Durand, F. Le-Calvez Kelm, C. Croce, C. Matar,

Z. Herceg, and H. Hernandez-Vargas, “MicroRNA miR-30 family regulates non-attachment growth of breast cancer cells.,” *BMC Genomics*, vol. 14, p. 139, 2013.

[233] P. J. Stogios, G. S. Downs, J. J. S. Jauhal, S. K. Nandra, and G. G. Privé, “Sequence and structural analysis of BTB domain proteins.,” *Genome Biol.*, vol. 6, no. 10, p. R82, 2005.

[234] Y. Inoue, Y. Itoh, K. Sato, F. Kawasaki, C. Sumita, T. Tanaka, D. Morishita, and H. Hayashi, “Regulation of Epithelial-Mesenchymal Transition by E3 Ubiquitin Ligases and Deubiquitinase in Cancer.,” *Curr. Cancer Drug Targets*, vol. 16, no. 2, pp. 110–8, 2016.

[235] A. G. Telonis, P. Loher, Y. Jing, E. Londin, and I. Rigoutsos, “Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity,” *Nucleic Acids Res.*, vol. 43, no. 19, pp. 9158–9175, 2015.

Ringraziamenti

I miei più sinceri ringraziamenti vanno ai miei supervisori, Riccardo e Milena, che hanno ideato questo progetto di ricerca e ne hanno posto le basi, senza mai smettere di lavorarci in prima persona e soprattutto di crederci fino in fondo. Inoltre, li ringrazio per avermi supportata e guidata in questo percorso e per avermi aiutata a crescere e a credere in me stessa. Avervi come capi è una grande fortuna.

Desidero ringraziare il Prof. Paolo Bernardi per la sua disponibilità e per aver seguito il mio percorso di dottorato con grande interesse, dandomi consigli utili e spunti interessanti.

Allo stesso modo ringrazio il Dott. Gustavo Baldassarre, che segue i nostri progressi scientifici spingendoci a migliorarci sempre.

Un ringraziamento speciale devo farlo alle mie colleghe e soprattutto amiche del ncRNA lab, Michela e Lauretta: senza il vostro supporto scientifico e morale le giornate in lab non sarebbero mai state le stesse! Ringrazio anche Luigi, che ha fatto parte del nostro lab contribuendo attivamente a una parte dei risultati presentati in questa tesi.

Inoltre ringrazio le mie colleghe Giulia e Valentina, perché allo stesso modo mi hanno supportata e aiutata per qualsiasi difficoltà e mi hanno aiutata a credere in me stessa.

Un ringraziamento enorme va ai miei amici di sempre Chiara, Diana, Melania, Simona, Fabrizio e Francesca, che mi hanno dimostrato sempre una stima incredibile..! E Laura, con te il confronto scientifico e personale non è mai banale, senza di te non so cosa avrei fatto.

Ringrazio con tutto il cuore la mia famiglia: grazie mamma, papà e Alessio per avermi aiutato a diventare grande e per aver sopportato ancora le mie ansie da studio, sebbene questa volta un po' ve la siete scampata dato che ero lontana..

..e se ve la siete scampata è perché c'eri tu, Orlando, che in qualsiasi istante buio o felice che fosse, eri accanto a me e con infinita pazienza mi hai aiutato ad andare avanti quando volevo mollare tutto. Le parole non saranno mai abbastanza per ringraziarti.