Article

ColoSTEM Dx kit: innovative alliance of cancer stem cells and glycosylation for earlier prediction of colon cancer aggressiveness and prognosis

Sabrina Blondy[1]#, Stéphanie Durand[2]#, Aurélie Lacroix[2], Niki Christou[2], Charline Bouchaud[1], Maud Peyny[1], Serge Battu[2,3], Alain Chauvanel[4], Vincent Carré[1], Marie-Odile Jauberteau[2,5]#, Fabrice Lalloué[2]# and Muriel Mathonnet[2,6]#

[1] Carcidiag Biotechnologies company, Guéret, France; sabrina.blondy@carcidiag.com, maud.peyny@carcidiag.com, charline.bouchaud@carcidiag.com, vincent.carre@carcidiag.com

[2] EA3842 CAPTuR "Control of cell activation, Tumor progression and Therapeutic resistance", Faculty of Medicine, Limoges, France; stephanie.durand@unilim.fr , aurelie.lacroix@unilim.fr, niki.christou@unilim.fr, serge.battu@unilim.fr, m-o.jauberteau-marchan@unilim.fr, fabrice.lalloue@unilim.fr, muriel.mathonnet@unilim.fr

[3] Laboratory of Analytical Chemistry, Faculty of Pharmacy, Limoges, France; serge.battu@unilim.fr

[4] Department of Pathology, Dupuytren University Hospital, Limoges, France; Alain.Chauvanel@chu-limoges.fr

[5] Department of Immunology, Dupuytren University Hospital, Limoges, France; m-o.jauberteau-marchan@unilim.fr

[6] Department of Digestive Chirurgy, General and Endocrine Surgery, Dupuytren University Hospital, Limoges, France; mathonnet@unilim.fr

* Correspondence: fabrice.lalloue@unilim.fr; stephanie.durand@unilim.fr; muriel.mathonnet@unilim.fr

# These authors contributed equally to this work

Simple Summary:

Although colon cancer is diagnosed earlier and prognosis has improved, 5-years net survival after curative surgery still to widely vary by stage: 10% of stage I to 55% of stage III patients will present recurrences. In collaboration with Limoges University Hospital, Carcidiag Biotechnologies company has developed the ColoSTEM Dx kit. This prognosis tool consists in a MIX of biotinylated plant lectins that specifically detects colon cancer stem cells within heterogenous tumors, *via* their specific glycosylation profile. A retrospective clinical study was carried out on 208 colon tumor samples stained by immunohistochemistry with the ColoSTEM Dx kit. Diagnostic sensitivity and specificity as well as the association between MIX scoring with survival rates at 5- and 7 years, according to clinicopathological data, were evaluated. With 78.3% and 70.6% of diagnostic sensitivity and specificity respectively, our kit provides an innovative tool usefully in clinical routine for improving earlier prognosis and of the management of colon cancer patients curatively operated non-treated colon cancer patients.

Abstract: Nowadays, colon cancer prognosis still difficult to predict, especially in the early stages. Recurrences remain elevated, even in the early stages after curative surgery. Carcidiag Biotechnologies has developed an immunohistochemistry (IHC) kit called ColoSTEM Dx, based on a MIX of biotinylated plant lectins that specifically detects colon cancer stem cells (CSCs) through glycan patterns that they specifically (over)express. A retrospective clinical study was carried out on tumor tissues from 208 non-treated and 21 treated patients with colon cancer, that were stained by IHC with the MIX. Clinical performances of the kit were determined, and prognostic and predictive values were evaluated. With 78.3% and 70.6% of diagnostic sensitivity and specificity respectively, our kit shows great clinical performances. Moreover, patient prognosis is significantly

poorer when the MIX staining is "High" compared to "Low", especially at 5-years of overall survival and for early stages. The ColoSTEM Dx kit allows an earlier and a more precise determination of patients' outcome. Thus, it affords an innovating clinical tool for predicting tumor aggressiveness earlier and determining prognosis value regarding therapeutic response in colon cancer patients.

Keywords: Aggressiveness, biomarker, cancer stem cells, diagnosis, colon cancer, glycan, immunohistochemistry, lectin, prognosis

Graphical Abstract
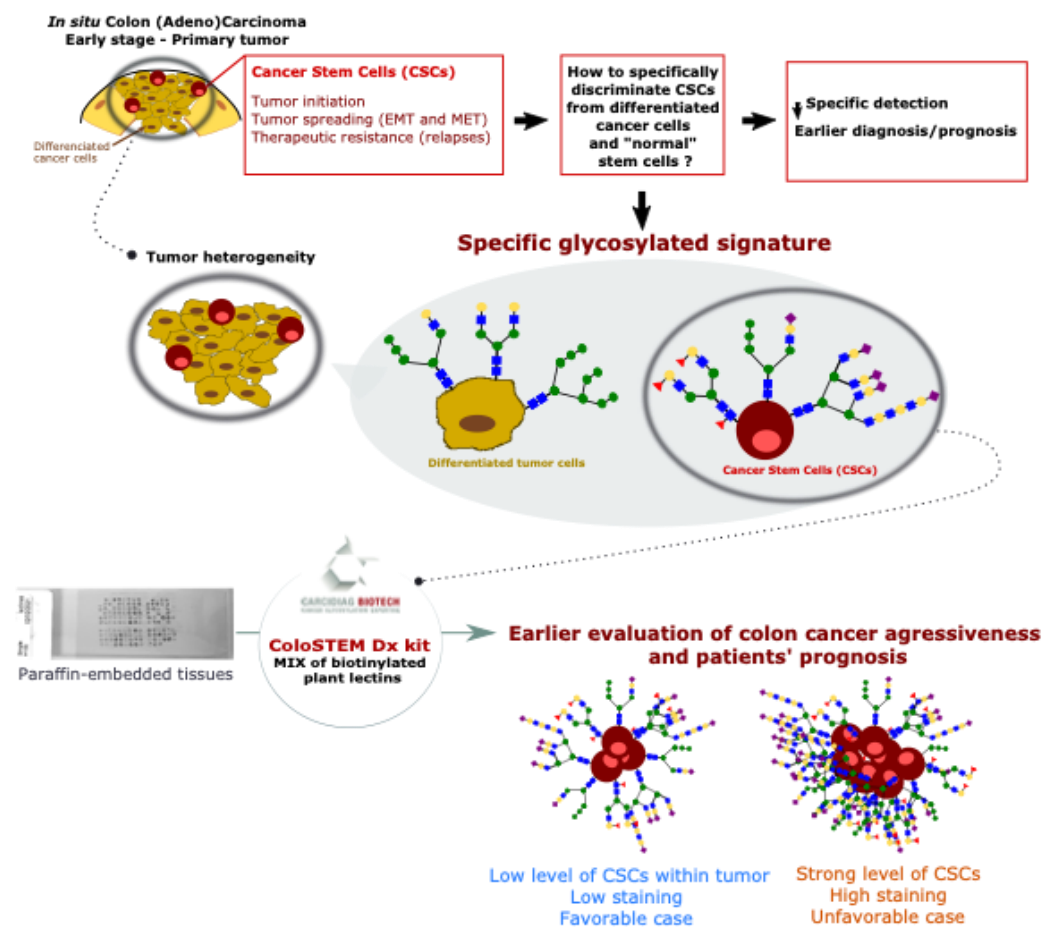
1. Introduction

Colon cancer represents the second leading cause of death from cancer [1–3]. Diagnosis is usually based on the pathological staging classification(pTNM) (stages I to IV) [4]. Surgical resection is the only curative method at present. Although the prognosis has improved in recent years, survival rates widely vary by stage, with 85% 5-years net survival for stage I and 50% for stage III [5]. Indeed, nearly 10% of stage I, 30% of stage II and 55% of stage III will present a metachronous cancer or a recurrence (locoregional or metastatic) within 5 years postoperatively [6]. This high risk of relapse requires to further improve earlier detection of colon cancer and to achieve personalized evaluation of patients' outcome and prognosis. This approach implies a systematic and precise determination of disease aggressiveness in order to strengthen patient follow-up and management [7].

Glycosylation is one of the most important posttranslational modifications of lipids (glycolipids) and proteins (glycoproteins), by the highly coordinated action of glycosyltransferases and glycosidases. Glycoproteins and glycolipids regulate a diverse range of key biological and cellular functions, including differentiation, proliferation, growth, pluripotency *etc...* Alterations in glycosylation processes (*i.e.* aberrant glycosylation) are linked to colon cancer development, progression, metastases and therapeutic failures [8–11]. Aberrant glycosylation constitutes a hallmark of Cancers and might even lead to the acquisition of a stemness phenotype [12].

CSCs are highly tumorigenic cells, *i.e.,* they are able to give rise to complete tumor mass. In addition to their ability to regenerate tumor mass, the main CSCs properties are self-renewal and multipotent differentiation capacity. It also have a unique property of resistance to treatments. Given all these properties, CSCs contribute to colon tumor initiation, colon cancer progression (metastasis formation), aggressiveness and relapses [13–17]. Various CSCs markers were defined to identify and isolate colon CSCs. Most of them are (i) membrane receptors and surface molecules (such as CD44 and its splice variants (CD44v), CD133/1 (AC133 epitope) or epithelial cell adhesion molecule (EpCAM)), (ii) cytosolic enzyme, such as aldehyde dehydrogenase 1 (ALDH1) enzymatic activity, and transcription factors including OCT-4 [8,13,15,18–26]. However, these biomarkers currently failed to be used in clinic because of their concomitant expression in adult (non-tumor) stem cells (SCs) [8,11]. These data underlie the importance to evidence more specific colon CSCs biomarkers.

Few data reported a correlation between the alteration of glycosylation processes with the induction and/or regulation of CSCs phenotype and properties. Overexpression of O-glycan truncated forms such as Tn antigen (Ag), is involved in the development and the induction of colon oncogenic features (tumorigenesis, cell growth, invasion, metastases and resistance to UV-induced apoptosis) [27]. The expression of ß-1,4-N-acetylgalactosaminyltransferase 3 is upregulated in colonospheres and its knockdown decreases sphere formation and CSCs marker expression (OCT-4 and NANOG) [28]. Overexpression of α-2, 6-Sialyltransferase and α-N-acetylgalactosaminide α-2,6-sialyltransferase 1, are both correlated with (i) colon CSCs enrichment (increase of CD133 and ALDH1 expressions, as well as sphere forming ability), and (ii) acquired resistance to chemotherapy (irinotecan and 5-Fluorouracil) and EGFR-targeted therapy (gefitinib) [29–32]. *FUT9* gene encoding the α-1,3 fucosyltransferase, plays a complex dual role in colon cancer development and malignancy. Alpha-1,3 fucosyltransferase knockdown strongly decreases sphere formation, growth of xenograft tumors and expression of OCT-4 and CD44, whereas it increases cell proliferation and migration. *FUT9* expression supports colon cancer aggressiveness. Indeed, its expression at early stages is required for CSCs expansion and colon cancer initiation. On the contrary, *FUT9* downregulation at later stages promotes colon cancer progression [33]. Most of colon CSCs surface markers are glycoproteins. They differ from their normal counterpart by the expression of tumor specific glycans [8,26]. CD44 splice variants carry oncofetal carbohydrate T and sialyl-Tn (sTn) Ag, correlating with increased metastatic potential of colon cancer cells [34]. Moreover, rather than the expression of total CD133 protein, it is the expression of a

specific glycan epitope (AC133) which could constitute a "bona fide" CSCs marker [25]. Altogether these data suggest that, a better characterization of colon CSCs glycosylation profiles could pave the way to identify more efficient new CSCs biomarkers in order to improve its specific detection within tumor and thus for targeting them.

Based on these knowledges and current clinical needs, Carcidiag Biotechnologies company has developed the ColoSTEM Dx kit, consisting in specific colon CSCs detection within heterogeneous tumor cell populations. There are currently no clinically standardized way (*i.e.*, efficient prognosis biomarkers) to provide reliable and earlier prognosis of colon cancer patients. In this context, our kit provide innovating and reliable biomarkers, specific to colon CSCs, for a better and an earlier stratification of low- or high-risk patients to develop an aggressive disease and relapse. The ColoSTEM Dx kit represents a tool perfectly adapted to the personalized management of patients. More precisely, it is an innovative tool that uses a MIX of biotinylated plant lectins (UEA-1, Jacalin and ACA, mixed in a particular ratio) recognizing glycan patterns specifically (over)expressed by CSCs only. This glycan patterns are not expressed by "normal" stem cells or differentiated cancer tumor cells. This colon CSCs specific MIX was evidenced by lectin-arrays and validated *in vitro* from research works carried out in collaboration with the University of Limoges, that have conducted to file two patents (national registration numbers WO2016FR53196 and WO2016FR53197).

Based on these results and in order to validate the ColoSTEM Dx kit for a routine clinical use by IHC, a retrospective clinical study was conducted in collaboration with Limoges University Hospital, on tumor tissues from non-treated (N=208) patients with colon cancer. Prognostic values of the MIX staining at 5- and 7-years of overall survival (OS) were validated according to clinicopathological data (*i.e.*, stages, age and gender). CSCs-specific glycosylated patterns evidenced by the MIX can be considered efficient biomarkers for a more accurate prediction of colon cancer patients' aggressiveness, outcomes and prognosis, compared to the non-specific SCs biomarker, OCT-4. With 78.3% and 70.6% of diagnostic sensitivity and specificity respectively, the ColoSTEM Dx kit constitutes an innovating tool usable in clinical practice in a standardized way, to provide a more accurate prognosis and to better stratify low- and high-risk patients for relapses. Finally, preliminary results obtained from treated patients' tissues (N=21) showed that the ColoSTEM Dx kit could also constitute a promising predictive tool for evaluating therapeutic response.

2. Materials and Methods

2.1. Cell culture

Colon adenocarcinoma cancer cell line HT-29 were obtained from ATCC (HTB-38™; ATCC®, France). Cells were incubated at 37°C with 5% CO_2 and 95% humidity and cultured in 1X McCoy's 5A modified medium (Gibco - ThermoFisher Scientific, France) supplemented with 10% fetal bovine serum (FBS; Gibco - ThermoFisher Scientific, France) and 1% penicillin/streptomycin (Gibco - ThermoFisher Scientific, France).

2.2. Indirect Magnetic Cell Sorting (MACS)

MACS was realized from 10^7 cells using the CELLectionTM Biotin Binder kit (Invitrogen - ThermoFisher Scientific, France) according to manufacturer's instructions, using 10µg of (i) the Mix (ColoSTEM Dx kit) or (ii) the AC133 biotinylated antibody (Ab) (Miltenyi Biotech, France), were used.

2.3. Evaluation of EpCAM^{High} immunostaining and ALDH1^{bright} activity by Flow Cytometry

EpCAM^{high} cell percentages within the AC133 – and + sorted-cells, were analyzed by flow cytometry (FCM) from 5.10^4 cells. After a saturation in 1%BSA in DPBS 1X calcium and magnesium free (10min, 4°C), cells were incubated for 45min at room temperature (RT) with an EpCAM mouse monoclonal Ab (clone VU1D9; Ozyme - Cell Signaling Technology, France) diluted at 1:150 in 1% BSA/DPBS. After a washing step in DPBS 1X

(g x 300, 10min, 4°C), cells were incubated for 30min at RT in the dark with an Alexa-Fluor 633-conjugated goat anti-mouse secondary Ab (ThermoFisher Scientific, France) diluted at 1:1000 in DPBS 1X.

Enzymatic activity of ALDH1 in MIX+ and MIX- sorted-cells, was analyzed from 10^5 cell/mL, using the ALDEFLUOR kit (Stem Cell Technologies, France) according to the manufacturer's recommendations.

Cells were extemporaneously stained with a DNA dye, *i.e.*, 0.5µL propidium iodide (PI; λex=475-581nm /λem=583-697nm; BD Biosciences, France). EpCAM^{high} immunostaining and ALDH1 enzymatic activity (ALDH1^{bright} cells) were analyzed among live cells (PI-), with the BD Accuri C6 Plus FCM (BD Biosciences, France). Mouse IgG1 isotype control (R&D systems, France) and N,N-diethylaminobenzaldehyde (DEAB, ALDH1 specific inhibitor) were used to control for background fluorescence.

2.4. Clonogenicity assay

MIX- and MIX+ sorted-cells were seeded in ultra-low attachment 96-wells plates (Falcon Corning brand, France) in increasing cell densities, i.e., 600, 1250, 2500, 5000, 10000 cells. For each condition, cells were seeded in 3 wells in 200µL of defined medium composed of 1X Dulbecco's Modified Eagle's Me (DMEM)/F12 (HAM 1:1) medium (Gibco - ThermoFisher Scientific, France) supplemented in 1X B27/Insulin (ThermoFisher Scientific, France), 10ng/mL Fibroblast Growth Factor (FGF; Peprotech, France) and 20ng/mL Epidermal Growth Factor (EGF; Peprotech, France). Twenty microliters of medium was added per well every week, for 4 weeks. The number of spheres formed per well and per condition was counted under optical microscope (Olympus CKX53; magnification, x100).

2.5. Patients and samples

Ninety colon tumor tissues came from TMA (HCol-Ade180Sur-08; AMSBIO, USA), which also includes the 90 corresponding tumor borders. Necrosed and absent tumor tissues (N=4) were excluded (N=86).

Forty-six colon tumor tissues were collected from non-treated patients with colon cancer, who underwent tumor resection at the Department of Digestive Surgery, General and Endocrine Surgery at Limoges University Hospital (France). Necrosed (N=4) were excluded (N=42).

Twenty-four tumor tissues were also collected from treated patients with colon cancer at Limoges University Hospital (France). Necrosed (N=3) were excluded (N=21).

Supplementary colon tumor tissues were collected from two TMA constituted in a cohort of non-treated colon cancer patients with early stages (I and II) from the "Centre de Ressources Biologiques – Institut Régional du Cancer Montpellier (CRB-ICM, Montpellier, France, ICM-CORT-2016-26). Necrosed and absent or non-interpretable tumor tissues (N=15) were excluded (N=80). Both TMAs also includes N=50 non-tumor samples.

Clinicopathological data including pTNM stages, gender, age and survival status at 5- and 7-years were provided after baseline examinations and diagnosis of the patients which was based on histological analyses of biopsies, according to the American Joint Committee on Cancer staging manual [37]. Survival rates analysis of non-treated and treated patients from all stages (refer to "statistical analysis" described below) were realized respectively from N=128 and N=21 samples (Tables S1 and S2). Survival rates analysis at 5 years of non-treated patients from early stages (refer to "statistical analysis" described below) were realized from N=70 samples (N=29 from CRB-ICM, N=27 from Limoges Hospital and N=14 from the TMA (AMSBIO) (Table S3).

2.6. MIX and OCT-4 IHC immunostaining

MIX staining was realized on N=208 and N=21 tumor tissues from respectively non-treated and treated patients with colon cancer (refer to "patients and samples" section; Tables S1, S2 and S3). MIX/OCT-4 co-staining was performed on some tumor tissues from

the non-treated patients' cohort, i.e., N=42 tumor tissues from Limoges University Hospital (refer to "patients and samples" section; Table S1). Each staining was realized by IHC on paraffin-embedded histological sections (4µm in thickness), in three main steps using the Leica Bond Max automatic staining platform (Leica Biosystems, France), according to the manufacturer's instructions.

1. Preparation and pretreatment of the tissues. Paraffin coating is removed using the Bond Dewax Solution (Leica Biosystems, France) and tissues are rehydrated under heat using the acidic buffer Bond Epitope Retrieval Solution 1, for 5min (pH 6; Leica Biosystems, France).

2. Immunostaining. Activity of endogenous peroxidases and biotins was blocked using the Bond Intense R Detection kit (Leica Biosystems, France). Tissues were incubated for 20 min with either the Mix alone pre-diluted at 1:2 ratio in the diluent supplied in the ColoSTEM Dx kit, or with both MIX (1:2) and OCT-4 (OCT-4 polyclonal rabbit IgG Ab; ThermoFisher Scientific, France). MIX and/or OCT-4 staining was revealed using respectively the Bond Intense R Detection kit and the Bond Polymer Refine Red Detection kit (Leica Biosystems, France). Nucleus were counter-stained by incubation with hematoxylin (Leica Biosystems, France) for 8 min.

3. Slides mounting. After dehydration by two successive baths of absolute ethanol (VWR, France) and toluene (ThermoFisher Scientific, France), for 5min each, tissue slides were mounted using the Leica CV Ultra (Leica Biosystems, France) and examined under the Leica photomicroscope DM4 B (Leica Biosystems, France; 200x magnification).

2.7. Scoring method

MIX staining appears in brown at apical membrane and/or in cytoplasm. OCT-4 staining appears in red/pink within nucleus (in blue) and/or cytoplasm. Scoring method of both staining was adapted from Ohara Y. *et al.* [38]. All tissues were stained either with the MIX alone, or with both the MIX and OCT-4. The total absence of staining (score 0) or the presence of stained cells constitutes the first element of analysis. The second element of analysis is related to the proportion of stained cells that is scored according to the followed gradation: 1 = 1-25%, 2 = 26-50%, 3 = 51-75% and 4 = 76-100% of stained cells. The third element of analysis is related to the staining intensity, graduated as followed: 1 = Low, 2 = Medium and 3 = High staining intensity. Scores obtained from both gradations are then added together and the total some obtained result into 6 intermediate scores, ranging from 2 to 7, that are finally grouped into 3 final scores (Figure 1). Final scores of 1 and 2 are considered as "Low staining" (MIX-Low and/or OCT-4-Low) and final score of 3 is considered as "High staining" (MIX-High and/or OCT-4-High).

Scoring of stained cells % \ Scoring of staining intensity	Low 1	Medium 2	High 3
1-25%: 1	2	3	4
26-50%: 2	3	4	5
51-75%: 3	4	5	6
76-100%: 4	5	6	7

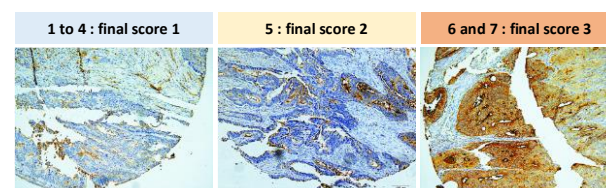| 1 to 4 : final score 1	5 : final score 2	6 and 7 : final score 3

Figure 1. Scoring method of Immunohistochemistry staining, according to percentages of stained cells and staining intensity. Percentages of stained cells are graduated into four scores (1 to 4) and staining intensity is graduated into three scores (1 to 3). The total some of both scoring results in intermediate scores ranging from 2 to 7. Intermediate scores ranging from 2 to 4 (in blue) result in a final score of 1. Intermediate score of 5 (in yellow) results in a final score of 2. Intermediate scores

of 6 and 7 (in orange) result in a final score of 3. Final scores of 1 and 2 are considered as "Low staining" and final score of 3 is considered as "High staining". Representative illustrations of MIX staining (in brown), as observed by IHC (magnification, 200x), are depicted below table.

2.9. Evaluation of clinical performances of the ColoSTEM Dx kit

Clinical performances of the ColoSTEM Dx kit, were determined from N=166 tumor tissues (N=86 of the commercial TMA (AMSBIO) and N=80 of the cohort of CRB-ICM Montpellier) and N=136 non-tumor tissues (N=86 tumor borders of the commercial TMA (AMSBIO) and N=50 non-tumor samples of the cohort of CRB-ICM Montpellier). Diagnostic sensitivity is related to the percentages of tumor tissues stained with the Mix (true positives) relative to the unlabeled ones (false negatives), as followed Sensitivity (%) = 100 x (True positives / (True positives + False negatives)). Diagnostic specificity is related to the percentages of tumor borders or non-tumor tissues unstained with the Mix (true negatives) relative to the labeled ones (false positives), as followed Specificity (%) = 100 x (True negatives / (True negatives + False positives)).

2.10. Statistical analysis

Statistical analysis and graphics were performed with StatView 5.0 (USA), Prism 7 (GraphPad, USA) and R environment (version 4.0.3). Statistical analysis of *in vitro* clonogenicity assay was made with an ANOVA/ANCOVA test. Survival rates according to MIX/OCT-4 co-staining were analyzed from the non-treated patients from the Limoges Hospital cohort, at 5 years, *i.e.*, only patients whose survival is ≤ 60 months at the last visit time, were retained (N=42; refer to "patients and samples" and Table S1). Survival rates of all non-treated patients were analyzed according to their clinicopathological data and Mix staining at (i) 5 years and (ii) 7 years, i.e., patients whose survival is ≤ 84 months at the last visit time (respectively N=79 and N=128; refer to "patients and samples" and Table S1). Survival rates analysis at 5 years of non-treated patients from early stages were achieved by combining three cohorts composed of patients from Limoges' hospital, CRB-ICM Montpellier as well as from a cohort provided by a AMSBIO (N=70; refer to "patients and samples" and Table S3). Survival rates of treated patients were analyzed according to MIX staining, at 5 years (N=21; refer to "patients and samples" and Table S2). The prognostic value of each parameter for outcome was assessed using the Kaplan-Meier method and log-rank test (Mantel-Cox). For each variable, hazard ratio (HR) was estimated using a univariate Cox model and expressed with their 95% confidence interval (95% CI). Multivariate analysis was carried out using a Cox regression model on single features identified by the univariate Cox modeling. Survival analysis were performed in R using *survival* and *survminer* packages. The proportional hazards assumption for Cox regression model fit was verified using *cox.zph* function of survival package. A p-value below 0.05 was considered as significant.

3. Results

3.1. The ColoSTEM Dx kit allows efficient isolation and enrichment of a cell subpopulation with stemness properties.

The ColoSTEM Dx kit originates from research works carried out by the EA3842 laboratory (Limoges' University; patent national registration number 1561763 – publication number 3044680). It aims at colon CSCs specific detection in both heterogeneous cell populations and tumors. Indeed, it is based on the use a MIX of biotinylated plant lectins that recognize glycan patterns specifically (over)expressed by these cells, *i.e.*, not by differentiated cancer cells, with heterogeneous tumor colon tissues. ColoSTEM Dx kit proofs of concept, *i.e.*, MIX evidence by lectin-arrays and its validation in specific colon CSCs detection and enrichment from several colon adenocarcinoma cell lines (including HT-29), are reported in detail in the patent mentioned above.

Main and most relevant results have been recalled in Figure S1. Briefly, HT-29 cells were sorted by MACS, with either the MIX (MIX+ and MIX- sorted-cells) or an AC133 Ab

(AC133+ and AC133- sorted-cells). Some of CSCs characteristics and properties were then evaluated: protein expression and enzymatic activity of stem cell markers (EpCAM and ALDH1), and sphere forming ability. While there are as many EpCAM^{High} cells in AC133+ sorted-cells as in AC133-, there are 7.5-times more EpCAM^{High} cells in MIX+ sorted-cells, compared to both MIX- (normalized to 1) and AC133+ cells (Figure S1A). Consistently, there are 4.7-times more ALDH1^{bright} cells in MIX+ sorted-cells (74.73%) than in MIX- (15.6%) (Figure S1B). Finally, MIX+ sorted-cells have a significant capacity to form spheres compared to MIX- cells, even when seeded at low densities (Figure S1C).

The use of the ColoSTEM Dx kit is more efficient for detecting and enriching in specific colon CSCs than cell sorting using AC133. These results suggest that ColoSTEM Dx kit improves CSC detection and cell sorting.

.

3.2. The ColoSTEM Dx kit improves colon CSCs detection and allows a more accurate prognosis than the standard stem cell marker OCT-4

MIX specificity in colon CSCs detection by IHC on tumor tissue, as well as its efficiency in patients' prognosis evaluation, were evaluated and compared to a standard stem cell marker, OCT-4. MIX and OCT-4 staining were realized on N=42 tumor tissues from non-treated patients (Table S1 and Figure S2). Among stained samples, half of samples are MIX-Low tissues or MIX-High (Figure S2A), while OCT-4-high staining is present in a broad panel of samples (almost 80%, Figure S2B) suggesting that OCT-4 is not able to discriminate CSC from the heterogeneous cell subpopulations. In addition, when cells are double-stained with MIX and OCT-4, samples are mainly divided in MIX-Low/OCT-4-High or MIX-High/OCT-4-High (Figure S2C). Intensity of MIX staining is not linked to OCT-4 staining and is independent of clinicopathological characteristics of patients except for gender (Table 1). However, this association is not found later on larger patient cohorts (Table 2). Altogether these results suggest that the ColoSTEM Dx kit is relevant for the discrimination of cancerous from healthy stem cells. It also evidences a better specificity to detect colon CSCs, than OCT-4 whose staining within tumor colon epithelium does not seem to be restricted to CSCs, but to all stem cells (healthy and cancer) and progenitors.

Table 1. Evaluation of the relationship between each clinicopathological characteristic (Sex, Age and pTNM staging) of non-treated patients (n=41, Table S1) and intensity of MIX staining (low/high) by Chi-squared statistic. Association between scoring according to MIX and OCT-4 staining was been also evaluated. *pTNM: pathology Tumor-Node-Metastasis; UICC: Union for International Cancer Control; yrs: years.*

		n	MIX staining		
			Low	High	P value
Sex	**Female**	18	12	6	**0.046**
	Male	23	7	16	
Age	**< 60yrs**	4	2	2	1
	≥ 60 yrs	37	17	20	
Stage (UICC)	**Early (I/II)**	27	14	13	0.514
	Late (III/IV)	14	5	9	
OCT4 staining	**Low**	8	6	2	0.109
	High	33	13	20	

Survival rates at 5 years were evaluated by Kaplan-Meier curves according to either OCT-4 staining (OCT-4-Low *versus* High; Figure 2A) or MIX staining (MIX-Low *versus*

High; Figure 2B). Univariate and multivariate Cox regression were performed to estimate prognosis value and risk scores associated to OCT-4 and MIX staining (Figure 2C). Representative pictures of MIX/OCT-4 co-staining, are depicted in Figure 2D.
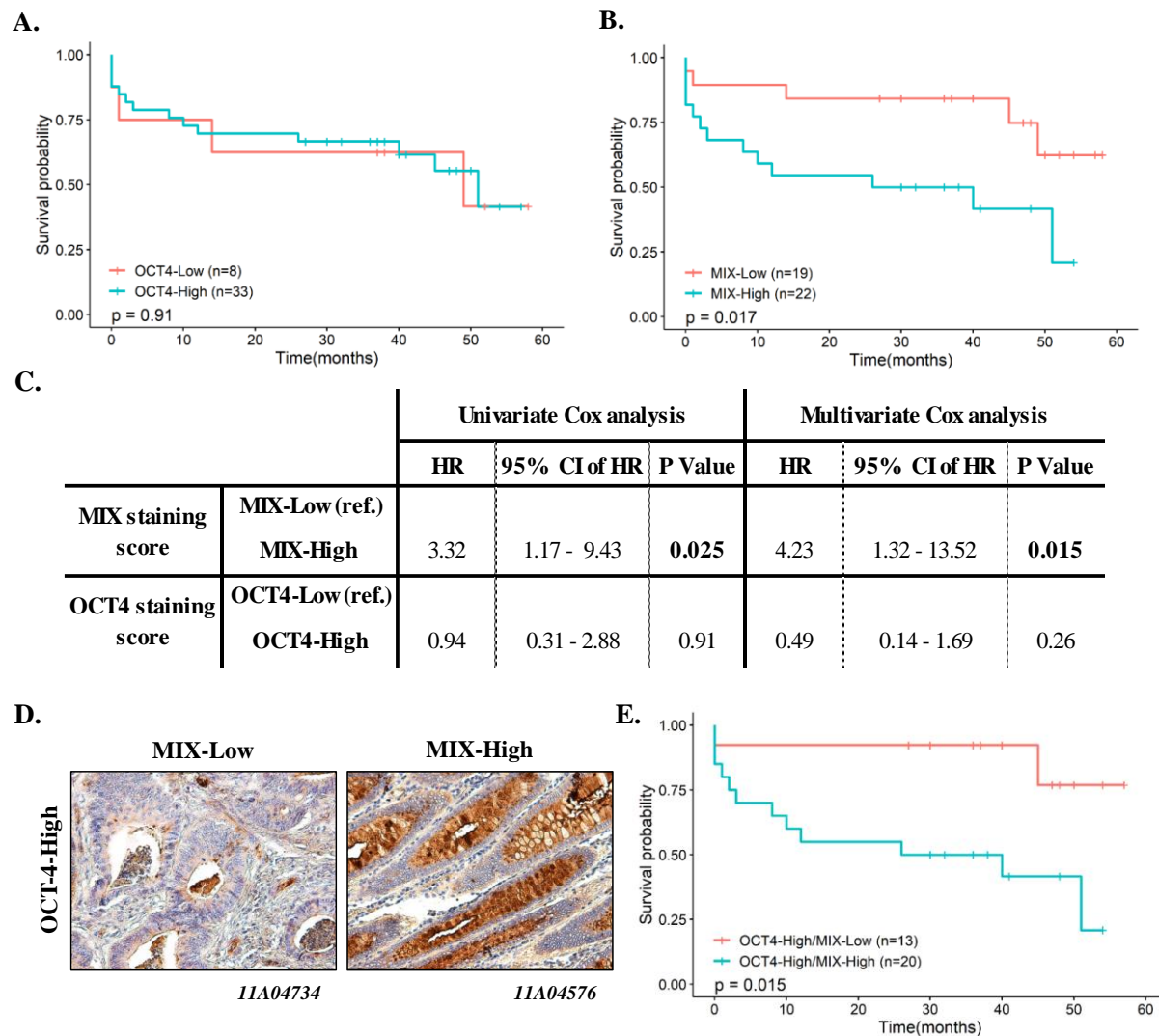

A.

B.

C.

		Univariate Cox analysis			Multivariate Cox analysis		
		HR	95% CI of HR	P Value	HR	95% CI of HR	P Value
MIX staining score	**MIX-Low (ref.)**						
	MIX-High	3.32	1.17 - 9.43	**0.025**	4.23	1.32 - 13.52	**0.015**
OCT4 staining score	**OCT4-Low (ref.)**						
	OCT4-High	0.94	0.31 - 2.88	0.91	0.49	0.14 - 1.69	0.26

D.

MIX-Low   MIX-High

OCT-4-High

11A04734   *11A04576*

E.

Figure 2. Association between MIX and OCT-4 scoring with survival rates at 5 years (60 months). Kaplan-Meier curves at 5-years are depicted according to OCT-4-Low *versus* -High staining (A), MIX-Low *versus* -High staining (B) and MIX-High/OCT-4-High *versus* MIX-Low/OCT-4-High co-staining (E). P value indicated in each panel correspond to log-rank test (Mantel-Cox) performed to survival curves comparison. Prognostic value of MIX and OCT-4 scoring was estimated for MIX and OCT-4 staining using univariate and multivariate Cox regression model and expressed with their hazard ratio (HR) and 95% confidence interval (CI). Representative illustrations of MIX (brown) and OCT-4 (red/pink) staining, as observed by IHC, are depicted (magnification, 200x) (D). *IHC: Immunohistochemistry; HR: Hazard ratio; CI: Confidence Interval.*

Survival rates at 5-years are not significantly different regardless of OCT-4- staining (Low or High; Figure 2A; p=0.91). On the contrary, 5-years survival rates of the MIX-High subgroup are significantly poorer than of the MIX-Low (Figure 2B; p=0.017). No relevant difference in survival median having been noted between a MIX-Low and an OCT-4-Low staining (data not shown; p=0.28). Interestingly, even if results are not significant, it has been observed a lower survival median (decrease of 18 months) with a MIX-High compared to an OCT-4-High staining (data not shown; p=0.2). Cox univariate analysis indicated that only MIX staining is a predictive factor for overall survival, with an significantly increasing risk associated to high MIX staining (HR: 3.3, 95% CI 1.17 to 9.43,

p=0.025) (Figure 2C, left panel). Multivariate model confirmed independence and relevance of MIX staining as prognostic factor (HR: 4.2, 95% CI 1.3 to 13.5, p=0.015) (Figure 2C, right panel).

Finally, survival rates at 5 years were also evaluated according to MIX/OCT-4 co-staining on same tissues (N=41), i.e., MIX-Low/OCT-4-Low, MIX-Low/OCT-4-High, MIX-High/OCT-4-Low, MIX-High/OCT-4-High. Due to not enough MIX-Low/OCT-4-Low and MIX-High/OCT-4-Low tumor tissues included (n=6 and 3, respectively; Figure S2C), Kaplan Meier curves were only depicted and analyzed for MIX-Low/OCT-4-High and MIX-High/OCT-4-High co-staining (Figure 2E). Interestingly, and consistently with previous observations, a MIX-High/OCT-4-High co-staining predicts significant poorer and worse prognosis than a MIX-Low/OCT-4-High co-staining, with strong survival median decrease (Figure 2E, p=0.015). High co-staining harbor a hazard ratio of 5.3 (95% CI 1.2 to 23.8, p=0.0298) in a univariate Cox regression model.

Contrary to OCT-4, MIX staining levels are closely associated with patient survival. Indeed, compared to OCT-4, MIX-High staining level improve significantly the detection and discrimination of colon CSCs. MIX-High staining might be a relevant cancer stem cell biomarker for monitoring disease aggressiveness and could be useful to establish the prognosis upon treatment.

3.3. The ColoSTEM Dx kit allows earlier evaluation of patients' disease agressiveness and prognosis, regardless of clinopathological data

In order to evaluate and confirm the prognosis value of the ColoSTEM Dx kit, all tumor tissues from non-treated patients (N=128, Table S1) were stained with the MIX. Survival rates were evaluated at 5- and 7-years of OS, according to clinicopathological data, *i.e.*, stages (early (I/II) and late (III/IV)), sex (men and women) and age (< and ≥ 60 years old) (Table 2).

Among the 79 tumor tissues included for survival rates analysis at 5-years, six were excluded due to a total absence of MIX staining. Fifty-six percent and 44% correspond respectively to early and late stages. Of the 41 early stages, 41% and 59% were respectively MIX-Low and MIX-High. Among the 32 late stages, 47% and 53% were respectively MIX-Low and MIX-High. Regarding OS at 5 years, we noted that MIX-staining is independent of tumor stage (Table 2A, p=0.809).

Overall survival rates at 7 years were then analyzed from 128 tumor samples. As previously described, we exclude tumor samples without MIX staining (n=13). Among the 115 retained tumors, 63% and 37% correspond respectively to early and late stages. Of the 73 early stages, 33% were MIX-Low whereas 67% were MIX-High. Regarding the 42 late stages, half of the population were MIX-Low or MIX-High. This result suggest that whatever evaluated time point of overall survival, MIX staining is independent of tumor stages (Table 2).

Table 2. Relationship of intensity of MIX staining and clinicopathological characteristics of non-treated patients included at 5 years and 7 years of OS. Numbers of tumor tissues from non-treated patients included at 5 years (A) and 7 years (B) of OS, for which there is a MIX-Low staining or a MIX-High staining, according to clinical and pathological data, *i.e.*, gender, age and stage, were indicated. The association between MIX staining and clinicopathological characteristic of patients was evaluated by Chi-squared statistic. (A) All tissues (n=79) were stained with the MIX but 6 samples show absence of MIX staining (not reported in the present table). They are distributed as follows: gender: 3 Female / 3 Male; age: 3 < 60 yrs / 3 ≥ 60 yrs ; stage: 3 early / 3 late ; vital status: 1 alive / 5 dead. (B) All tissues (n=128) were stained with the MIX but 13 samples show absence of MIX staining (not reported in the present table). They are distributed as follows: gender: 6 Female / 7 Male; age: 4 < 60 yrs / 9 ≥ 60 yrs ; stage: 7 early / 6 late ; vital status: 7 alive / 6 dead. *OS: Overall Survival; UICC: Union for International Cancer Control; yrs : years.*

A

		n	MIX Staining		
			Low	High	P value
Sex	**Female**	35	18	17	0.239
	Male	38	14	24	
Age	**< 60yrs**	8	6	2	0.139
	≥ 60 yrs	65	26	39	
Stage (UICC)	**Early (I/II)**	41	17	24	0.809
	Late (III/IV)	32	15	17	

B.

		n	MIX Staining		
			Low	High	P value
Sex	**Female**	54	22	32	0.855
	Male	61	23	38	
Age	**< 60yrs**	16	11	5	**0.014**
	≥ 60 yrs	99	34	65	
Stage (UICC)	**Early (I/II)**	73	24	49	0.087
	Late (III/IV)	42	21	21	

Kaplan-Meier curves and Cox regression models were displayed and patients' survival rates were analyzed at 5- and 7-years, according to MIX staining levels and stages (Figure 3 and Table 3). At 5-years, prognostic significance for OS of MIX scoring is clearly supported by survival curve (Figure 3A, p=0.011) and univariate Cox model (HR: 2.1 with 95% CI 1.17 to 3.75, p=0.013; Table 3A). On the contrary, no statistically relevant difference is shown between MIX-Low or -High at 7-years survival rates (Figure 3B and Table 3B).

Noted that sex and age have not a significant impact on survival rates, at 5- (Table 3A and Figure S3A) and 7-years (Table 3B and Figure S3B). Kaplan-Meier curves performed on separately groups of patients, stratifying according to sex (male and female) or age (inferior or superior to 60 years old), failed to show any difference in survival rates according to low or high MIX staining (data not shown).

In brief, no statistically relevant difference on survival rates between MIX-Low or -High staining was noted, regardless of age, even if a significant link was been previously identified by Chi-square test at 7 years patients' follow-up (p=0.014, Table 2). We conclude in the same way with regard to sex.
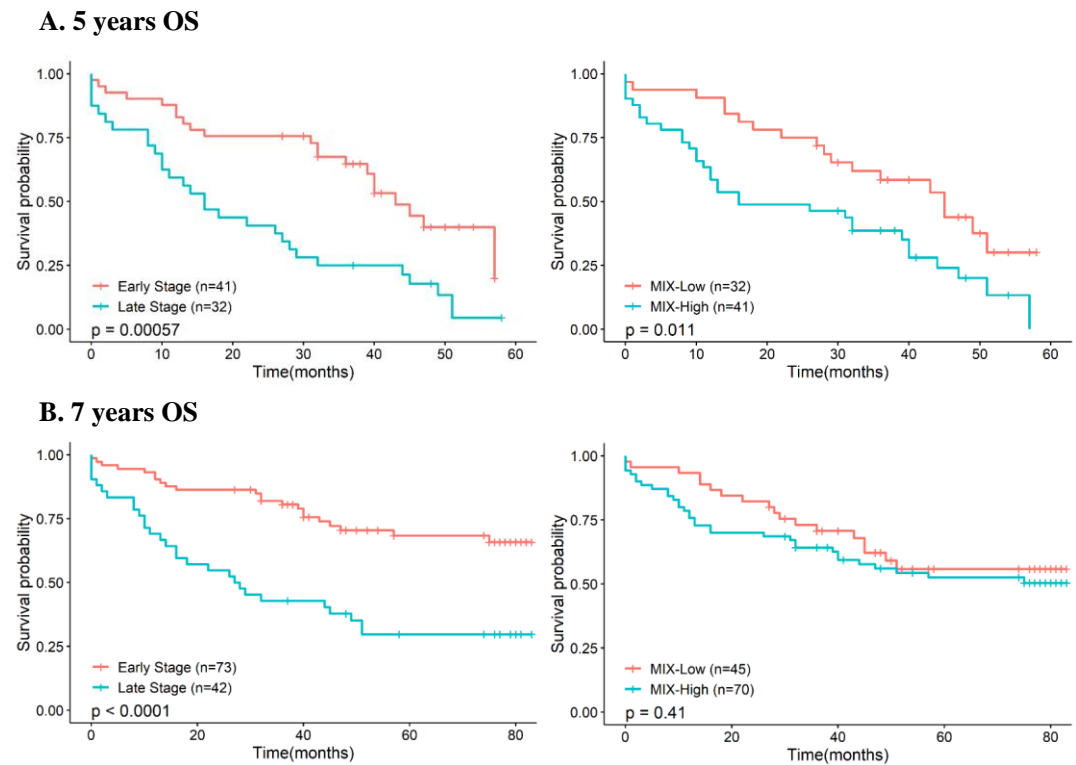
A. 5 years OS

B. 7 years OS

Figure 3. Association between pTNM staging and MIX scoring with survival rates at 5- and 7 years (60 and 84 months respectively). Prognostic value of each feature (pTNM staging and MIX staining) was assessed using the Kaplan-Meier method and log-rank test by stratification of patients according to early and late stages or MIX-Low and High scoring at 5-years of OS (A) and at 7-years of OS (B). *OS: Overall Survival; pTNM: pathology Tumor Node Metastasis.*

Table 3. Prognostic value of clinicopathological features (gender, age, stage and MIX staining) at 5- and 7 years of patients' follow-ups (60 and 84 months respectively). Resulting hazard ratio (with 95% confidence interval) for stratifying patients for 5-years of OS (A) and for 7-years of OS (B) using clinicopathological features were obtained by univariate Cox modeling (left panel). Multivariate analysis (right panel) was carried out using a Cox regression model using pTNM staging and MIX scoring. *OS: Overall Survival; pTNM: pathology Tumor Node Metastasis; HR: Hazard Ratio; CI: Confidence Interval.*

A.

		Univariate Cox analysis			Multivariate Cox analysis		
		HR	**95% CI of HR**	**P Value**	**HR**	**95% CI of HR**	**P Value**
Sex	**Female (ref.)**						
	Male	1.38	0.78 - 2.44	0.269	/	/	/
Age	**< 60 yrs (ref.)**						
	> 60 yrs	0.78	0.33 - 1.85	0.573	/	/	/
Stage (UICC)	**Early (I/II) (ref.)**						
	Late (III/IV)	2.62	1.49 - 4.62	**0.001**	2.997	1.69 - 5.32	**0.0002**
MIX staining score	**MIX-Low (ref.)**						
	MIX-High	2.09	1.17 - 3.75	**0.013**	2.461	1.37 - 4.44	**0.003**

B.

		Univariate Cox analysis			Multivariate Cox analysis		
		HR	95% CI of HR	P Value	HR	95% CI of HR	P Value
Sex	**Female (ref.)** **Male**	1.07	0.62 - 1.85	0.818	/	/	/
Age	**< 60 yrs (ref.)** **> 60 yrs**	1.30	0.56 - 3.06	0.542	/	/	/
Stage (UICC)	**Early (I/II) (ref.)** **Late (III/IV)**	3.21	1.84 - 5.6	**3.99e-05**	3.77	2.12 - 6.71	**6.51e-06**
MIX staining score	**MIX-Low (ref.)** **MIX-High**	1.27	0.72 - 2.26	0.411	1.82	1.00 - 3.30	**0.049**

Since multivariate analysis revealed that late stage and a high-MIX score were independent prognosis factors at 5-years and at 7-years of patients' follow-up (Table 3), we chose to combine these two parameters in order to assess their impact on survival rates. We confirmed high value of MIX score as risk factor for OS, regardless of pTNM staging (Figure 4). The survival rate of MIX-high staining associated to late stage patients is significantly poorer compared to MIX-low staining with a doubling hazard ratio, observed at 5-years and at 7-years of OS (Figure 4C, left panel). The same tendency was observed for the early stage patients, even if only results acquired at 5-years of OS were significant (Figure 4C, right panel).

Noteworthy that results are statistically more pronounced and relevant for 5-years follow-up: patients presented MIX-Low staining have a high survival rate, in comparison with MIX-High patients (Figure 4A and C). To accurately estimate the prognosis value of MIX according to given stage, we have performed survival analysis (Kaplan-Meier and univariate Cox regression) in function of low or high MIX subpopulations at early (I/II) or late (III/IV) stage (Figure S4). At 5 years of patients' follow-up, high MIX staining could be considered as a poor prognosis marker in early stage (HR: 3.3 with 95% CI 1.2 to 9.1, p=0.021; Figure S4A) and late stage (HR: 2.2 with 95% CI 1 to 4.6, p=0.039; Figure S4B). At 7 years of patients' follow-up, MIX prognosis benefit is lost for early stage (Figure S4C) but is slightly maintained for late stage (HR: 1.95 with 95% CI 0.93 to 4.1, p=0.076; Figure S4D). Altogether, these results suggest that the MIX could be considered as an efficient prognosis marker to predict disease aggressiveness from early phase post-surgery within the 5 years post resection. However, the significance of MIX prognosis value should be useful at early stage to adapt therapeutic strategy and improve patients' management. For this reason, we chose to increase the early stage subpopulation actually made up 16 stage 1 and 25 stages 2 (Table S1).

Thus, we completed a new cohort from early stage CRC tissue (Table S3) with a total of 41 patients of stage I and 29 patients of stage II at 5 years of follow-up. Survival analysis performed on this cohort confirms the High-MIX staining as poor prognosis factor. Although a slightly difference are observed between high and low MIX in early stage subpopulation (p=0.18), we show that MIX-high has a moderate bad prognosis value in early stage in univariate Cox model (p=0.18, HR=1.764 and 95% CI 0.6514-4.779; data not shown). Thus, in early stages, we tried to distinguish stage 1 and stage 2 by combining pTNM stages and MIX scoring. In this case, survival rate is collapsed in patients with stage II/high mixed compared to stage I/High MIX suggesting that the relative risk is markedly increased when high MIX staining is detected in stage 2 patients (Figure S5).
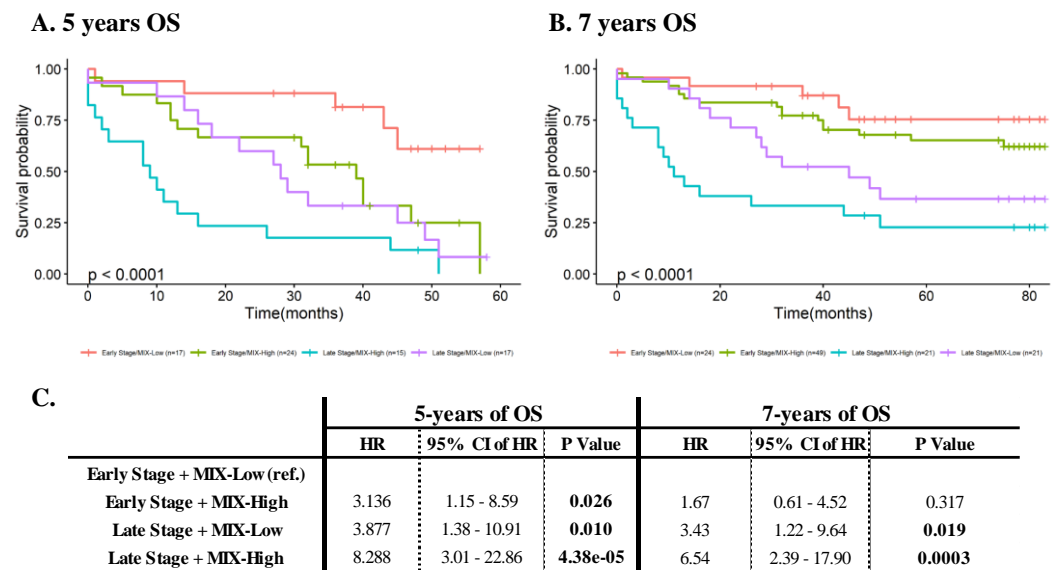
A. 5 years OS   **B. 7 years OS**

C.

	5-years of OS			7-years of OS		
	HR	95% CI of HR	P Value	HR	95% CI of HR	P Value
Early Stage + MIX-Low (ref.)						
Early Stage + MIX-High	3.136	1.15 - 8.59	**0.026**	1.67	0.61 - 4.52	0.317
Late Stage + MIX-Low	3.877	1.38 - 10.91	**0.010**	3.43	1.22 - 9.64	**0.019**
Late Stage + MIX-High	8.288	3.01 - 22.86	**4.38e-05**	6.54	2.39 - 17.90	**0.0003**

Figure 4. Combination of pTNM staging and MIX scoring for survival analysis at 5- and 7 years (60 and 84 months respectively). Comparison of survival curves was performed using Kaplan-Meier method (with log-rank test) by stratification of patients according early and late stages or Low- and High-MIX scoring, at 5- (A) and 7- (B) years. Prognostic values of pTNM staging and MIX scoring for survival analysis at 5- and 7 years of patients' follow-ups were analyzed (C) and resulting hazard ratio (with 95% confidence interval) for stratifying patients for 5- (left) and 7- (right) years of OS using stage and MIX scoring combination, were obtained by univariate Cox modeling. *OS: Overall Survival; pTNM: pathology Tumor Node Metastasis, HR: Hazard Ratio; CI: Confidence Interval.*

If we consider 7 years of overall survival, the ColoSTEM Dx kit does not allows prediction of disease evolution (i.e., patients' prognosis) regardless of their age or sex. However, concerning their stage, prognostic value of MIX staining appeared more reliable in the late stage of colorectal cancer patients (Figure S4C and D). On the contrary, if we consider 5 years of overall survival, the ColoSTEM Dx kit markedly predicts disease aggressiveness and allows stratifying patients of good or poor prognosis with a high or low risk of relapse after curative surgery, especially from early stages.

Altogether, these results evidence that specific glycan motif of colon CSCs detected by the ColoSTEM Dx kit, constitute independent prognosis factor from pTNM staging and other clinicopathological data, and allows to discriminate a better or worse prognosis. Importantly, our kit allows a great prediction of colon cancer aggressiveness and prognosis, from all stages, with more pronounced values for early stages, and within the first 5 years after curative surgery.

3.4. The ColoSTEM Dx kit displays great clinical performances

Clinical performances of the ColoSTEM Dx kit, *i.e.*, diagnostic specificity and sensitivity, have been determined from N=166 tumor tissues (N=86 from AMSBIO TMA and N=80 from the CRB-ICM Montpellier cohort) and N=136 tumor edges (N=86 from AMSBIO TMA and N=50 from the CRB-ICM Montpellier cohort). Among the N=166 tumor tissues, N=36 depicted an absence of MIX staining (false negatives) and N=130 were stained (true positives). Among the N=136 tumor tissues, N=96 depicted an absence of MIX staining (true negatives) and N=40 were stained (false negatives). According to formula described in "materials and methods" section, diagnostic sensitivity and specificity reach respectively 78.3% and 70.6%.

3.5. Specific glycan motifs of colon CSCs evidenced by the ColoSTEM Dx kit could also constitute promising predictive biomarkers

In order to evaluate predictive values of the ColoSTEM Dx kit, 21 tumor tissues from treated patients were stained with the MIX: 42.8% and 57.1% tissues were respectively Mix-Low and Mix-High (Table S2 and Figure 5A). Kaplan Meier curves were achieved with treated patients' overall survival rates at 5-years according to MIX staining levels (Low *versus* High; Figure 5B). It appeared that 5-years OS of the MIX-High subgroup is significantly poorer than the MIX-Low, with a strong decrease in survival median (p=0.016). Univariate Cox analysis revealed MIX score (p=0.03) and, to a lesser extent, age (p=0.066) and pTNM staging (p=012), as predictive factor for overall survival in treated patients (Table 4, left panel). Multivariate analysis confirm that MIX score is an independent prognostic factor (HR: 6.98 with 95%CI 1.1 to 44.03, p=0.0387, Table 4 right panel).
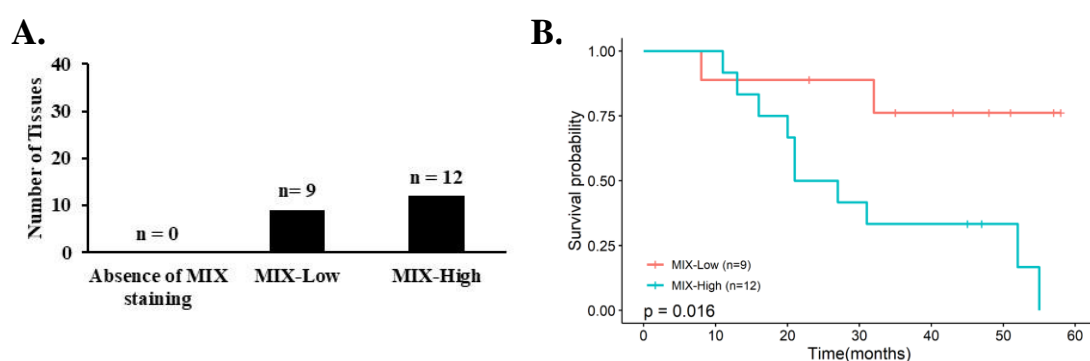
Figure 5. Association between MIX scoring with treated patients' survival rates at 5-years. (A) Graph represent numbers (N) of tumor samples from treated patients for which there is an absence of MIX, a MIX-Low or a MIX-High staining. All samples included (N=21) were stained with the MIX. (B) Corresponding Kaplan-Meier curves are displayed according to MIX staining (i.e., Mix-Low and -High). P value correspond to log-rank test. *OS: Overall Survival.*

Table 4. Prognostic value of clinicopathological features (gender, age, stage and MIX staining) at 5 years of patients' follow-ups (60 months) of treated patients' cohort. Univariate and multivariate Cox regression were displayed for analysis of clinicopathological parameters and MIX scoring impact on prognostic value. *OS: Overall Survival ; HR: Hazard Ratio; CI: Confidence Interval.*

		Univariate Cox analysis			Multivariate Cox analysis		
		HR	95% CI of HR	P Value	HR	95% CI of HR	P Value
Sex	**Female (ref.)** **Male**	2.01	0.65 - 6.26	0.229	/	/	/
Age	**< 60 yrs (ref.)** **> 60 yrs**	7.01	0.88 - 56.06	**0.067**	15.48	1.52 - 158.16	**0.021**
Stage (UICC)	**Early (I/II) (ref.)** **Late (III/IV)**	2.91	0.76 - 11.19	0.12	4.23	0.94 - 18.98	0.0596

MIX staining score	MIX-Low (ref.)						
	MIX-High	5.42	1.17 - 25.07	**0.031**	6.98	1.11 - 44.03	**0.039**

These preliminary results evidence that the ColoSTEM Dx kit might also constitute a promising predictive tool, i.e., companion test, in order to (i) allow better prediction of therapeutic responses and relapses' risk and (ii) improve therapeutic management for each patient.

4. Discussion

Cancer stem cells play a key role in colon cancer evolution and has major implications to cancer therapy. Currently, CSC are not used as biomarkers in clinical routine although these cells could reflect tumor aggressiveness and might be of prime importance for diagnosis or prognosis. Indeed, CSC enable cancer therapeutic resistance to conventional treatments thereby conduce to therapeutic failure. Thus, the reliable detection of CSCs from patient samples might improve future patient management and survival. Nevertheless, no kits or devices developed for clinical or translational research are currently likely to detect cancer stem cells within tissues. Furthermore, the clinical use and significance of CSC biomarkers are still restricted due to the risk of confusing detection with biomarkers expressed by adult non-cancerous stem cells [8,11]. In this context, Carcidiag Biotechnologies decided to develop a new device for detecting CSC in patient solid biopsies. ColoSTEM Dx kit, developed by Carcidag Biotechnologies, is based on the detection of glycoproteins, known to be specifically expressed at the surface of colon cancer stem cells (colon CSCs). This diagnosis tool uses a MIX of biotinylated plant lectins that recognize glycan patterns specifically expressed or overexpressed by CSCs only. Normal stem cells or differentiated tumor cells are not detected by the lectin MIX. HT-29 cell sorting of MIX-positive cells show an enrichment of EpCAM-high and ALDH1-high cell subpopulation, characteristics consistent with a stem cell phenotype. EPCAM and ALDH1 activity is currently been used to define CSC populations in digestive cancers [37]. Their expressions were associated with poor prognosis in both disease-free and overall survival for colorectal cancer. In addition, MIX positive cells are highly able to form colonospheres, up to 8 times more than their negative counterpart, again reflecting the stem cell status of these cells. Based on this *in vitro* evaluation, the potential stem cell detection capacity of MIX was tested by immunohistochemistry on 42 CRC samples, in comparison with OCT-4 staining, a classical stem cell marker. While MIX-high staining seems restricted to a subset of tissues, OCT4-high staining is present in a broad panel of tissue. No significant association was been put in evidence between intensity of staining with both markers. These results are consistent with previous work demonstrating that although OCT-4 is considered as a pluripotent stem cell marker required to enhance the self-renewal ability, its expression was reported to be restricted in normal colon, polyp and CRC. Thus, OCT-4 analysis by using immunohistochemistry is not useful to identify cancer and has restricted interest to characterize cancer stem cells for diagnosis [38]. Thus, survival curves according low or high OCT4 staining were homogeneous, while MIX-high staining reveals a significant decrease of overall survival suggesting this biomarker is clearly more relevant for monitoring patients and might be of prime interest for patient management. High MIX staining is considered as an independent bad prognosis factor as validated by univariate and multivariate Cox regression (HR: 4.2). Poor prognosis value was confirmed by the decreased median survival of CRC patients characterizing by an OCT4-high/MIX-high staining combination. As previously mentioned, we confirmed that OCT-4 staining alone (High vs. Low) is unable to discriminate good or poor prognosis patients suggesting that OCT-4 is not likely to detect CSC from tissue samples. These

results suggest that MIX staining is specific of CSCs and could be useful for its detection in tumor samples and thus could predict the presence of CSCs as well as the associated risk of recurrence. In this context, the significance of the ColoSTEM Dx kit in the detection of CSC has been assessed on several cohorts of patients in order to demonstrate that it might provide a novel prognosis tool. The MIX staining has revealed a prognosis value at early stage in the five years patient cohort. The prognosis value of MIX staining on overall survival at 5 years was confirmed but not demonstrated at 7 years suggesting that the MIX could be of prime importance at early stages of CRC and might be used to predict treatment outcome at this stage. Since the MIX staining constitute an independent poor prognosis factor, from pTNM staging and from other clinical parameters (age and gender), it could be crucial in the future management of patients. In a similar way, patients stratification using combination of MIX and stage allows improve the patient's classification. Indeed, this combined analysis highlights that high MIX staining is a marker of poor prognosis for the patient. Indeed, whatever the stage, MIX-high is associated with a poor prognosis with a HR of 3 to 8 in early or late stage, respectively, at 5 years of follow-up. The same trend was observed at 7 years, but it is no significant for early stage. Thus, patients from 7 years cohort are probably very good prognosis CRC. The bad prognosis value of MIX staining has been confirmed in treated patients supporting that the MIX could also be likely to predict patients treatment outcome at early stages and maybe in future to prevent tumor burden by early detection of recurrence. Even if new tools such as Immunoscore and circulating tumor DNA aid to accurately characterize patients with minimal residual disease, they don't allow to identify the specific presence of CSCs within tumor. Thus, although these approaches permit to adapt patients' management with personalized adjuvant treatment approaches, they fail to eradicate CSCs thereby increase the risk of patient recurrence. On the contrary, the ColoSTEM Dx kit is efficient to detect CSCs even at early stage tumor. It can therefore be complementary to current approaches. Nevertheless, further developments are required and will include validation in prospective multicenter interventional outcome studies in order to confirm on a wide cohort of treated patients that MIX staining has a bad prognosis value in early stages.

5. Conclusions

In conclusion, it appears that ColoSTEM Dx kit demonstrated its significance to detect CSCs, more efficiently than OCT-4 and could be a new tool usefully in clinical management of colon cancer, due to their potential to predict tumor aggressiveness, even on early stage tumors.

6. Patents

The patent under the national registration number 1561763 (N° WO2016FR53196 and WO2016FR53197 publication number 3044680 and 3044681) results from a part of the work reported in this manuscript, *i.e.*, experimental data depicted in Figure S1.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: Characteristics of non-treated patients with colon cancer (N=128) and corresponding MIX and OCT-4 scoring , Table S2: Characteristics of treated patients with colon cancer (N=21) and corresponding MIX scoring, Table S3: Characteristics of non-treated early stages (I and II) patients with colon cancer (N=70) and corresponding MIX scoring, Figure S1: *In vitro* characterization of MIX (ColoSTEM Dx kit) efficiency in colon CSCs detection and enrichment, Figure S2: Distribution of tumor tissues from non-treated patients according MIX or Oct-4 or both staining, Figure S3: Association between gender and age with survival rates at 5- and 7 years, Figure S4: Association between MIX-staining with survival rates at 5- and 7 years according to early stage (I/II) or late stage (III/IV), Figure S5: Combination of pTNM staging (early stages, *i.e.,* I and II) and MIX scoring for survival analysis at 5-years.

References

1. Dekker, E.; Tanis, P.J.; Vleugels, J.L.A.; Kasi, P.M.; Wallace, M.B. Colorectal Cancer. The Lancet 2019, 394, 1467–1480, doi:10.1016/S0140-6736(19)32319-0.

2. Keum, N.; Giovannucci, E. Global Burden of Colorectal Cancer: Emerging Trends, Risk Factors and Prevention Strategies. Nat Rev Gastroenterol Hepatol 2019, 16, 713–732, doi:10.1038/s41575-019-0189-8.

3. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians 2018, 68, 394–424, doi:10.3322/caac.21492.

4. Brierley, J.D.; Gospodarowicz, Ma.K.; Wittekind, C. TNM Classification of Malignant Tumours, 8th Edition; Wiley-Blackwell.; 2017; ISBN 978-1-119-26357-9.

5. Drouillard, A.; Bouvier, A.-M.; Rollot, F.; Faivre, J.; Jooste, V.; Lepage, C. Conditional Net Survival: Relevant Prognostic Information for Colorectal Cancer Survivors. A French Population-Based Study. Digestive and Liver Disease 2015, 47, 597–601, doi:10.1016/j.dld.2015.03.013.

6. Manfredi, S.; Bouvier, A.M.; Lepage, C.; Hatem, C.; Dancourt, V.; Faivre, J. Incidence and Patterns of Recurrence after Resection for Cure of Colonic Cancer in a Well Defined Population. Br J Surg 2006, 93, 1115–1122, doi:10.1002/bjs.5349.

7. Dienstmann, R.; Salazar, R.; Tabernero, J. Personalizing Colon Cancer Adjuvant Therapy: Selecting Optimal Treatments for Individual Patients. JCO 2015, 33, 1787–1796, doi:10.1200/JCO.2014.60.0213.

8. Barkeer, S.; Chugh, S.; Batra, S.K.; Ponnusamy, M.P. Glycosylation of Cancer Stem Cells: Function in Stemness, Tumorigenesis, and Metastasis. Neoplasia 2018, 20, 813–825, doi:10.1016/j.neo.2018.06.001.

9. Jang, H.; Kim, T.W.; Yoon, S.; Choi, S.-Y.; Kang, T.-W.; Kim, S.-Y.; Kwon, Y.-W.; Cho, E.-J.; Youn, H.-D. O-GlcNAc Regulates Pluripotency and Reprogramming by Directly Acting on Core Components of the Pluripotency Network. Cell Stem Cell 2012, 11, 62–74, doi:10.1016/j.stem.2012.03.001.

10. Wang, Y.-C.; Peterson, S.E.; Loring, J.F. Protein Post-Translational Modifications and Regulation of Pluripotency in Human Stem Cells. Cell Res 2014, 24, 143–160, doi:10.1038/cr.2013.151.

11.   Very, N.; Lefebvre, T.; El Yazidi-Belkoura, I. Drug Resistance Related to Aberrant Glycosylation in Colorectal Cancer. Oncotarget 2017, 9, 1380–1402, doi:10.18632/oncotarget.22377.

12.   Peixoto, A.; Relvas-Santos, M.; Azevedo, R.; Santos, L.L.; Ferreira, J.A. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. Front Oncol 2019, 9, doi:10.3389/fonc.2019.00380.

13.   Batlle, E.; Clevers, H. Cancer Stem Cells Revisited. Nat Med 2017, 23, 1124–1134, doi:10.1038/nm.4409.

14.   O'Brien, C.A.; Pollett, A.; Gallinger, S.; Dick, J.E. A Human Colon Cancer Cell Capable of Initiating Tumour Growth in Immunodeficient Mice. Nature 2007, 445, 106–110, doi:10.1038/nature05372.

15.   Zhou, Y.; Xia, L.; Wang, H.; Oyang, L.; Su, M.; Liu, Q.; Lin, J.; Tan, S.; Tian, Y.; Liao, Q.; et al. Cancer Stem Cells in Progression of Colorectal Cancer. Oncotarget 2017, 9, 33403–33415, doi:10.18632/oncotarget.23607.

16.   Iyer, D.N.; Sin, W.-Y.; Ng, L. Linking Stemness with Colorectal Cancer Initiation, Progression, and Therapy. World J Stem Cells 2019, 11, 519–534, doi:10.4252/wjsc.v11.i8.519.

17.   Effendi-YS, R. Cancer Stem Cells and Molecular Biology Test in Colorectal Cancer: Therapeutic Implications. Acta Medica Indonesiana 2018, 49, 351.

18.   Zhu, L.; Gibson, P.; Currle, D.S.; Tong, Y.; Richardson, R.J.; Bayazitov, I.T.; Poppleton, H.; Zakharenko, S.; Ellison, D.W.; Gilbertson, R.J. Prominin 1 Marks Intestinal Stem Cells That Are Susceptible to Neoplastic Transformation. Nature 2009, 457, 603–607, doi:10.1038/nature07589.

19.   Hatano, Y.; Fukuda, S.; Hisamatsu, K.; Hirata, A.; Hara, A.; Tomita, H. Multifaceted Interpretation of Colon Cancer Stem Cells. Int J Mol Sci 2017, 18, doi:10.3390/ijms18071446.

20.   CHEN, X.; WEI, B.; HAN, X.; ZHENG, Z.; HUANG, J.; LIU, J.; HUANG, Y.; WEI, H. LGR5 Is Required for the Maintenance of Spheroid-Derived Colon Cancer Stem Cells. Int J Mol Med 2014, 34, 35–42, doi:10.3892/ijmm.2014.1752.

21.   Pang, R.; Law, W.L.; Chu, A.C.Y.; Poon, J.T.; Lam, C.S.C.; Chow, A.K.M.; Ng, L.; Cheung, L.W.H.; Lan, X.R.; Lan, H.Y.; et al. A Subpopulation of CD26+ Cancer Stem Cells with Metastatic Capacity in Human Colorectal Cancer. Cell Stem Cell 2010, 6, 603–615, doi:10.1016/j.stem.2010.04.001.

22.   Todaro, M.; Gaggianesi, M.; Catalano, V.; Benfante, A.; Iovino, F.; Biffoni, M.; Apuzzo, T.; Sperduti, I.; Volpe, S.; Cocorullo, G.; et al. CD44v6 Is a Marker of Constitutive and Reprogrammed Cancer Stem Cells Driving Colon Cancer Metastasis. Cell Stem Cell 2014, 14, 342–356, doi:10.1016/j.stem.2014.01.009.

23.   Fan, X.; Ouyang, N.; Teng, H.; Yao, H. Isolation and Characterization of Spheroid Cells from the HT29 Colon Cancer Cell Line. Int J Colorectal Dis 2011, 26, 1279–1285, doi:10.1007/s00384-011-1248-y.

24.   Slater, C.; De La Mare, J.-A.; Edkins, A.L. In Vitro Analysis of Putative Cancer Stem Cell Populations and Chemosensitivity in the SW480 and SW620 Colon Cancer Metastasis Model. Oncol Lett 2018, 15, 8516–8526, doi:10.3892/ol.2018.8431.

25.   Kemper, K.; Sprick, M.R.; Bree, M. de; Scopelliti, A.; Vermeulen, L.; Hoek, M.; Zeilstra, J.; Pals, S.T.; Mehmet, H.; Stassi, G.; et al. The AC133 Epitope, but Not the CD133 Protein, Is Lost upon Cancer Stem Cell Differentiation. Cancer Res 2010, 70, 719–729, doi:10.1158/0008-5472.CAN-09-1820.

26.   Karsten, U.; Goletz, S. What Makes Cancer Stem Cell Markers Different? SpringerPlus 2013, 2, 301, doi:10.1186/2193-1801-2-301.

27.   Jiang, Y.; Liu, Z.; Xu, F.; Dong, X.; Cheng, Y.; Hu, Y.; Gao, T.; Liu, J.; Yang, L.; Jia, X.; et al. Aberrant O-glycosylation Contributes to Tumorigenesis in Human Colorectal Cancer. J Cell Mol Med 2018, 22, 4875–4885, doi:10.1111/jcmm.13752.

28.   Zhang, L.; Ten Hagen, K.G. Pleiotropic Effects of O-Glycosylation in Colon Cancer. J Biol Chem 2018, 293, 1315–1316, doi:10.1074/jbc.H117.812826.

29.   Lavrsen, K.; Dabelsteen, S.; Vakhrushev, S.Y.; Levann, A.M.R.; Haue, A.D.; Dylander, A.; Mandel, U.; Hansen, L.; Frödin, M.; Bennett, E.P.; et al. De Novo Expression of Human Polypeptide N-Acetylgalactosaminyltransferase 6 (GalNAc-T6) in Colon Adenocarcinoma Inhibits the Differentiation of Colonic Epithelium. J Biol Chem 2018, 293, 1298–1314, doi:10.1074/jbc.M117.812826.

30.   Che, M.-I.; Huang, J.; Hung, J.-S.; Lin, Y.-C.; Huang, M.-J.; Lai, H.-S.; Hsu, W.-M.; Liang, J.-T.; Huang, M.-C. B1, 4-N-Acetylgalactosaminyltransferase III Modulates Cancer Stemness through EGFR Signaling Pathway in Colon Cancer Cells. Oncotarget 2014, 5, 3673–3684.

31.   Park, J.-J.; Lee, M. Increasing the α 2, 6 Sialylation of Glycoproteins May Contribute to Metastatic Spread and Therapeutic Resistance in Colorectal Cancer. Gut Liver 2013, 7, 629–641, doi:10.5009/gnl.2013.7.6.629.

32.   Park, J.-J.; Yi, J.Y.; Jin, Y.B.; Lee, Y.-J.; Lee, J.-S.; Lee, Y.-S.; Ko, Y.-G.; Lee, M. Sialylation of Epidermal Growth Factor Receptor Regulates Receptor Activity and Chemosensitivity to Gefitinib in Colon Cancer Cells. Biochemical Pharmacology 2012, 83, 849–857, doi:10.1016/j.bcp.2012.01.007.

33.   Swindall, A.F.; Londoño-Joshi, A.I.; Schultz, M.J.; Fineberg, N.; Buchsbaum, D.J.; Bellis, S.L. ST6Gal-I Protein Expression Is Upregulated in Human Epithelial Tumors and Correlates with Stem Cell Markers in Normal Tissues and Colon Cancer Cell Lines. Cancer Res 2013, 73, 2368–2378, doi:10.1158/0008-5472.CAN-12-3424.

34.   Ogawa, T.; Hirohashi, Y.; Murai, A.; Nishidate, T.; Okita, K.; Wang, L.; Ikehara, Y.; Satoyoshi, T.; Usui, A.; Kubo, T.; et al. ST6GALNAC1 Plays Important Roles in Enhancing Cancer Stem Phenotypes of Colorectal Cancer via the Akt Pathway. Oncotarget 2017, 8, 112550–112564, doi:10.18632/oncotarget.22545.

35.   Auslander, N.; Cunningham, C.E.; Toosi, B.M.; McEwen, E.J.; Yizhak, K.; Vizeacoumar, F.S.; Parameswaran, S.; Gonen, N.; Freywald, T.; Bhanumathy, K.K.; et al. An Integrated Computational and Experimental Study Uncovers FUT9 as a Metabolic Driver of Colorectal Cancer. Mol Syst Biol 2017, 13, doi:10.15252/msb.20177739.

36.   Singh, R.; Campbell, B.; Yu, L.; Fernig, D.; Milton, J.; Goodlad, R.; FitzGerald, A.; Rhodes, J. Cell Surface-Expressed Thomsen-Friedenreich Antigen in Colon Cancer Is Predominantly Carried on High Molecular Weight Splice Variants of CD44. Glycobiology 2001, 11, 587–92.

37.   AJCC Cancer Staging Manual; Amin, M.B., Edge, S.B., Greene, F.L., Byrd, D.R., Brookland, R.K., Washington, M.K., Gershenwald, J.E., Compton, C.C., Hess, K.R., Sullivan, D.C., Jessup, J.M., Brierley, J.D., Gaspar, L.E., Schilsky, R.L., Balch, C.M., Winchester, D.P., Asare, E.A., Madera, M., Gress, D.M., Meyer, L.R., Eds.; Springer International Publishing: Cham, 2017; ISBN 978-3-319-40617-6.

38.   Ohara, Y.; Oda, T.; Sugano, M.; Hashimoto, S.; Enomoto, T.; Yamada, K.; Akashi, Y.; Miyamoto, R.; Kobayashi, A.; Fukunaga, K.; et al. Histological and Prognostic Importance of CD44+/CD24+/EpCAM+ Expression in Clinical Pancreatic Cancer. Cancer Science 2013, 104, 1127–1134, doi:10.1111/cas.12198.