scientific reports

Check for updates

OPEN

Transcription factor FOXD1 and miRNA-204-5p play a major role in B4GALNT2 downregulation in colon cancer

Martina Duca[1], Nadia Malagolini[1], Michela Pucci[1], Virginie Cogez[2], Anne Harduin-Lepers[2] & Fabio Dall'Olio[1]✉

The β1,4-N-acetylgalactosaminyltransferase 2 (B4GALNT2) which synthesizes the histo-blood group antigen Sd^a is highly expressed by normal colon, but it is dramatically down-regulated in colorectal cancer (CRC). High B4GALNT2 expression in CRC tissues is a marker of longer survival. The molecular bases of B4GALNT2 inhibition in CRC are largely obscure. A key role may be played by transcription factors and miRNA. Through an in silico analysis of The Cancer Genome Atlas and of the Cancer Cell Line Encyclopedia, we identified the transcription factors FOXD1, FOXF2 and PGR as well as mir-204-5p as potential inhibitory agents. Their transient transfection in the cell line GP2d, whose B4GALNT2 is closer to that of a normal mucosa, confirmed their inhibitory activity with a crucial role for FOXD1. The latter inhibited B4GALNT2 also in the middle B4GALNT2 expresser cell line Caco2. Deletion experiments of the putative FOXD1 binding sites in the ~2800 bp sequence upstream of the *B4GALNT2* transcriptional start site cloned in frame with the luciferase reporter gene, confirmed the regulatory role of FOXD1. Finally, FOXD1 knock down in the non-B4GALNT2 expresser cell line SW948 stimulated B4GALNT2. Thus, FOXD1 and miR-204-5p emerged as crucial new player of B4GALNT2 down-regulation in CRC.

Keywords Glycosylation, Sd^a antigen, Gene regulation, FOXD1, miR-204-5p, Colorectal cancer

The process of glycosylation, a major type of modification of proteins and lipids[1] is operated by glycosyltransferases, a class of enzymes which mediate the transfer of sugars from activated sugar donors to acceptors that can be proteins, lipids or other sugars[1]. However, the regulation of glycosylation involves also sugar-donor transporters, glycosidases, chaperons and other molecules. Neoplastic transformation results in profound glycosylation changes with profound effects on the biology and the clinic of the disease[2–4]. The Sd^a epitope is a terminal sugar structure formed by a α2,3-sialylated galactose residue to which a GalNAc unit is β4-linked (Fig. 1A). This epitope can be found at the end of sugar chains *N*- or *O*-linked to glycoproteins as well as some glycolipids[5]. The Sd^a antigen is expressed on erythrocytes, secretions and a few tissues, in particular the colon[6]. The addition of the β4-linked GalNAc, which is last and crucial step of Sd^a biosynthesis, is mediated by a single enzyme: the β1,4-N-acetylgalactosaminyltransferase 2[7], product of the *B4GALNT2* gene[8,9] (reviewed in[6,10,11]). Germline *B4GALNT2* mutations are responsible for some of the Sd^{a-} phenotypes[12–14]. The *B4GALNT2* gene maps on 17q21.33 and is comprised of at least 12 exons. There are two alternative first exon named exon 1S and exon 1L, both containing a translational start site (Fig. 1B,C).

Consequently, there are at least two different forms of the B4GALNT2 enzyme protein differing in the amino terminal portion: the one encoded by 1S (short form) contains a cytoplasmic domain of conventional length, while that encoded by 1L (long form) displays an exceptionally long cytoplasmic domain[8,9,15]. The major intracellular localization of both forms is the Golgi apparatus. However, the cytoplasmic domain of the long form dictates also post-Golgi vesicles and plasma-membrane localization[16]. The genomic region upstream of the exons 1 displays the features of a CpG island (Fig. 1C), indicating a potential regulatory role. In colon cancer (CRC), B4GALNT2 enzyme activity[17,18] and mRNA[19] are dramatically down-regulated (Fig. 1D). The mechanisms at the basis of this dramatic change have been previously investigated, with special focus on the methylation of the CpG island[20,21]. It has been reported that demethylating treatment of CRC cell lines resulted in partial activation

[1]Department of Medical and Surgical Sciences (DIMEC), General Pathology Building, University of Bologna, Bologna, Italy. [2]Unité de Glycobiologie Structurale et Fonctionnelle, CNRS, UMR 8576, UGSF, Univ. Lille, 59000 Lille, France. ✉email: fabio.dallolio@unibo.it

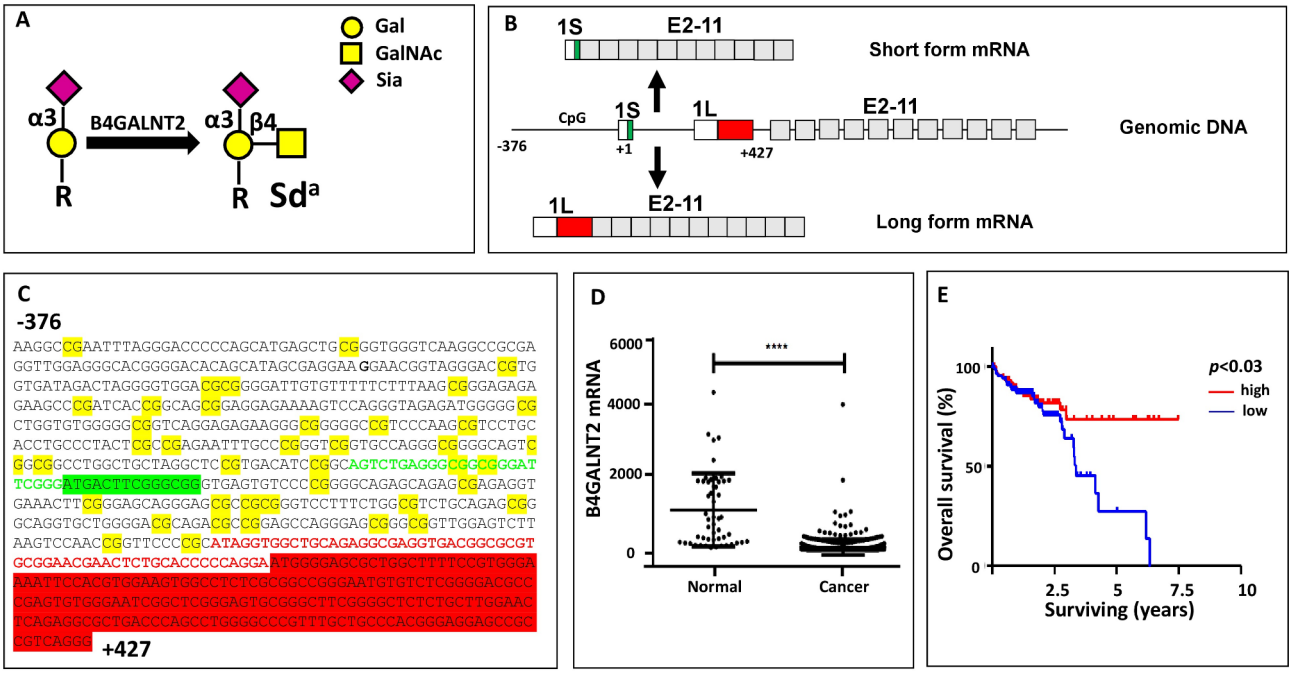

Fig. 1. Sd[a] epitope and *B4GALNT2*. (**A**) The Sd[a] epitope is comprised of a α3-sialylated galactose residue, substituted by a β4-linked GalNAc. R represents the underlying carbohydrate chains which can be *N*- or *O*-linked to proteins or to some glycolipids, as detailed in[6,10]. (**B**) Genomic and mRNA structure of B4GALNT2. The coding sequences of exons 1S and 1L are represented in green and red, respectively, while the 5′UTR are in white. Coding exons 2–11 are depicted in grey and not to scale. A more detailed representation is provided in[6,10]. (**C**) The coding sequences of exon 1S and 1L are highlighted in green and red respectively, while their untranslated sequences are in green or red characters. The CpG sequences are highlighted in yellow. Nucleotides are numbered taking the as + 1 the first adenine of the translation start codon of the short form. (**D**) B4GALNT2 mRNA expression in normal and cancer COAD cohort. (**E**) Kaplan Meier survival curves of COAD patients falling in upper or lower 15th percentile of B4GALNT2 mRNA expression.

of *B4GALNT2* transcription which, however, remained far below that of a normal mucosa. The recent availability of molecular and clinical data of large cohorts of CRC patients by "The Cancer Genome Atlas" (TCGA) has confirmed the general down-regulation of B4GALNT2 mRNA in CRC (Fig. 1D) and allowed to establish that patients with higher B4GALNT2 expression displayed longer overall survival[5,22,23] (Fig. 1E). In addition, TCGA data point to miR-204-5p as a potential inhibitory factor of B4GALNT2 expression in that its expression is mutually exclusive with that of B4GALNT2 in CRC samples[23]. The role of transcription factors in the regulation of colonic B4GALNT2 expression has been investigated only recently. The transcription factors ETS1 and to a lesser extent SP1 have been identified as necessary for *B4GALNT2* transcription[24] (Fig. 2A), although neither of the two was responsible for *B4GALNT2* down-regulation in CRC. In the vast majority of colon cancer cell lines, B4GALNT2 is barely or not detected[17]. The main purpose of this study was the identification of transcription factors potentially responsible for *B4GALNT2* down-regulation. To this aim, it was necessary to identify a colon cancer cell line whose B4GALNT2 expression was as close as possible to that of a normal colonic mucosa. Through database search, we have identified the cell line GP2d as the strongest B4GALNT2 expresser. This cell line, as well as the cell line Caco2 expressing moderate levels of B4GALNT2[25], was used for transfection studies to test the power of putative regulatory factors identified through database search.

Results

FOXD1, FOXF2 and PGR are potential inhibitors of B4GALNT2 expression

A ~ 2800 bp long genomic sequence upstream of the 1S exon of the *B4GALNT2* gene (Fig. 2A) was analyzed with the transcription binding sites prediction tools Consite and Promo. The 17 potential transcription factors identified were selected on the basis of their obvious association (positive or negative) with B4GALNT2 expression in TCGA colon adenocarcinoma (COAD) patients, TCGA normal samples and CRC cell lines, as reported in the Cancer Cell Line Encyclopedia (CCLE) (Fig. S1). Three transcription factors: FOXD1 (Forkhead Box D1), FOXF2 (Forkhead Box F2) and PGR (Progesterone Receptor) displayed the most consistent association with B4GALNT2 expression in the three cohorts. Their putative binding sites in the genomic region upstream the *B4GALNT2* transcription initiation site(s) is displayed in Fig. 2A. They were all negatively associated with B4GALNT2 (Fig. 2B). FOXF2 and PGR in normal tissues were the only showing a significant inverse non-linear correlation. The other plots failed to fit into a defined mathematical curve. Of the three transcription factors, FOXD1 was the only displaying a significant up-regulation in tumor tissues (Fig. 2C) and a negative association

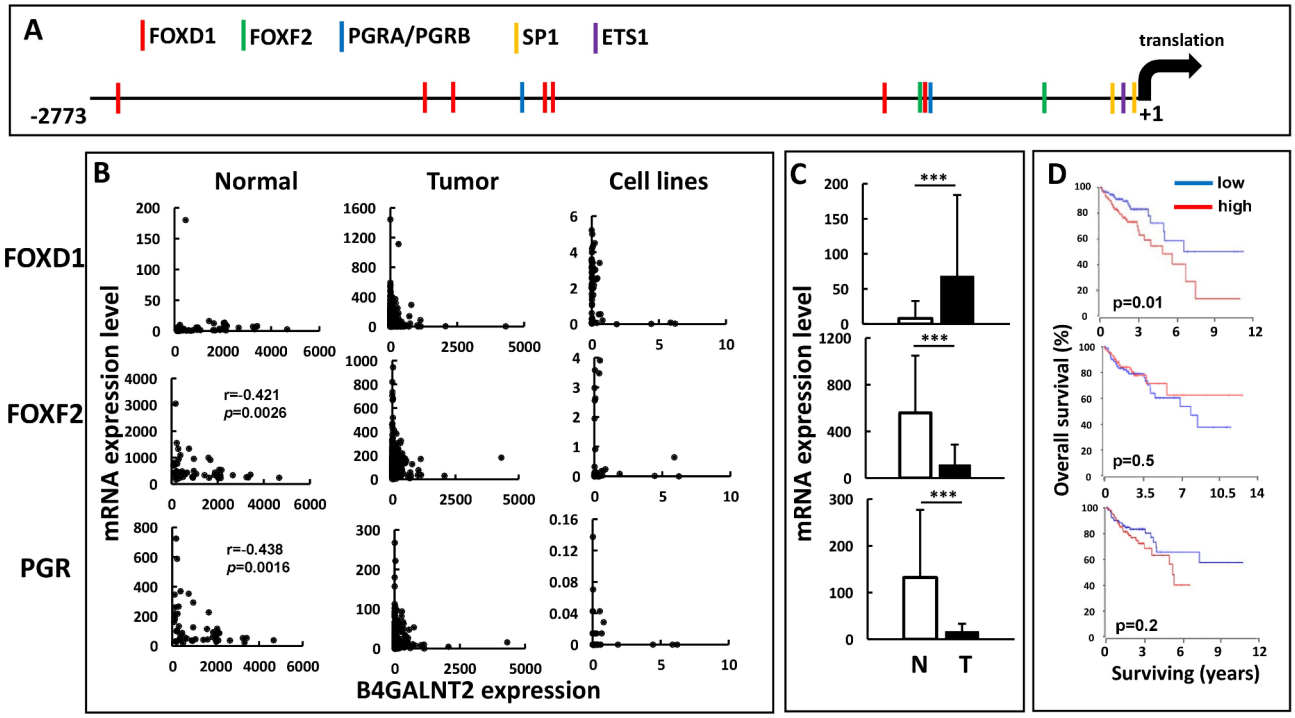

Fig. 2. Transcription factors and B4GALNT2. (**A**) Approximate position of the binding motifs of the three transcription factors in the B4GALNT2 genomic region preceding the translation initiation codon of the short form. The position of the two previously identified transcription factors ETS1 and SP1 in the promoter region[24] is also indicated. (**B**) Dot plot analysis of the three transcription factors and B4GALNT2 in the normal and cancer COAD cohort of TCGA and in colon cancer cell lines in CCLE. Analysis was performed by the Spearman correlation. (**C**) Expression of the three transcription factors in normal (N) and tumor (T) COAD TGCA samples**.** Data were analyzed by the Student's t test for paired samples (N = 49). (**D**) Kaplan Meier survival curves of COAD patients for the 25 upper (N = 110 red) and 25 lower (N = 110 blue) percentile of the three transcription factors.

with prognosis (Fig. 2D). Thus, FOXD1 appeared to be the most promising candidate as a B4GALNT2 inhibitor, although FOXF2 and PGR could also play an ancillary role.

The CRC cell lines GP2d and Caco2 as strong and medium B4GALNT2 expressers

To assess experimentally the three putative inhibitory transcription factors, it is necessary to utilize a colon cancer cell line whose B4GALNT2 expression level is comparable with that of a normal mucosa. This is not a trivial task, since the vast majority of the colon cancer cell lines display barely or not detectable B4GALNT2 levels. The CRC cell line Caco2 was so far considered the top B4GALNT2 expresser[25]. However, according to the CCLE the top expresser was the CRC cell line GP2d. In Fig. 3A, it is shown the comparison between the expression level of the Sd^a antigen and of the B4GALNT2 protein by immunoblot assay in Caco2 and GP2d. While the Sd^a antigen was barely expressed by Caco2 and strongly by GP2d, the B4GALNT2 enzyme protein was expressed by both cell lines albeit at a much higher level by GP2d (Fig. 3A). The B4GALNT2 enzyme activity (Fig. 3B) and the B4GALNT2 mRNA (Fig. 3C) were consistently much higher in GP2d than in Caco2. Since the *B4GALNT2* gene generates two major transcripts (encoding the short and the long polypeptide forms) (Fig. 1B), we established their expression in Caco2 and GP2d by RT-PCR, using primer pairs specific either for the short or long form. As shown in Fig. 3D, while Caco2 cells produced both transcripts, GP2d cells produced exclusively the short form. This was confirmed by Western blot analysis (Fig. 3E) in which a GP2d sample was run in parallel with samples of the cell lines LS174T-S11 and LS174T-L21. These are two previously established[18] variants of the cell line LS174T constitutively expressing either the short (-S11) or the long B4GALNT2 form (-L21). The MW of the B4GALNT2 form expressed by GP2d is fully consistent with the short form expressed by LS174T-S11 cells (Fig. 3E).

FOXD1, FOXF2 and PGR inhibit B4GALNT2 expression in GP2d at different levels

To determine experimentally the effect of FOXD1, FOXF2 and PGR on B4GALNT2 expression in GP2d, the corresponding expression vectors were transiently transfected and the level of expression of the transfected gene and of B4GALN2 was measured by qRT-PCR (Fig. 4A). For PGR, Addgene offers two different expression vectors encoding PGRA and PGRB, respectively. The two forms are identical except for the presence in PGRB of an additional 164 amino acid sequence at the NH_2-terminal domain. The mRNA of all 4 transcription factors was undetectable in mock-transfected cells but became strongly expressed upon transfection with the appropriate

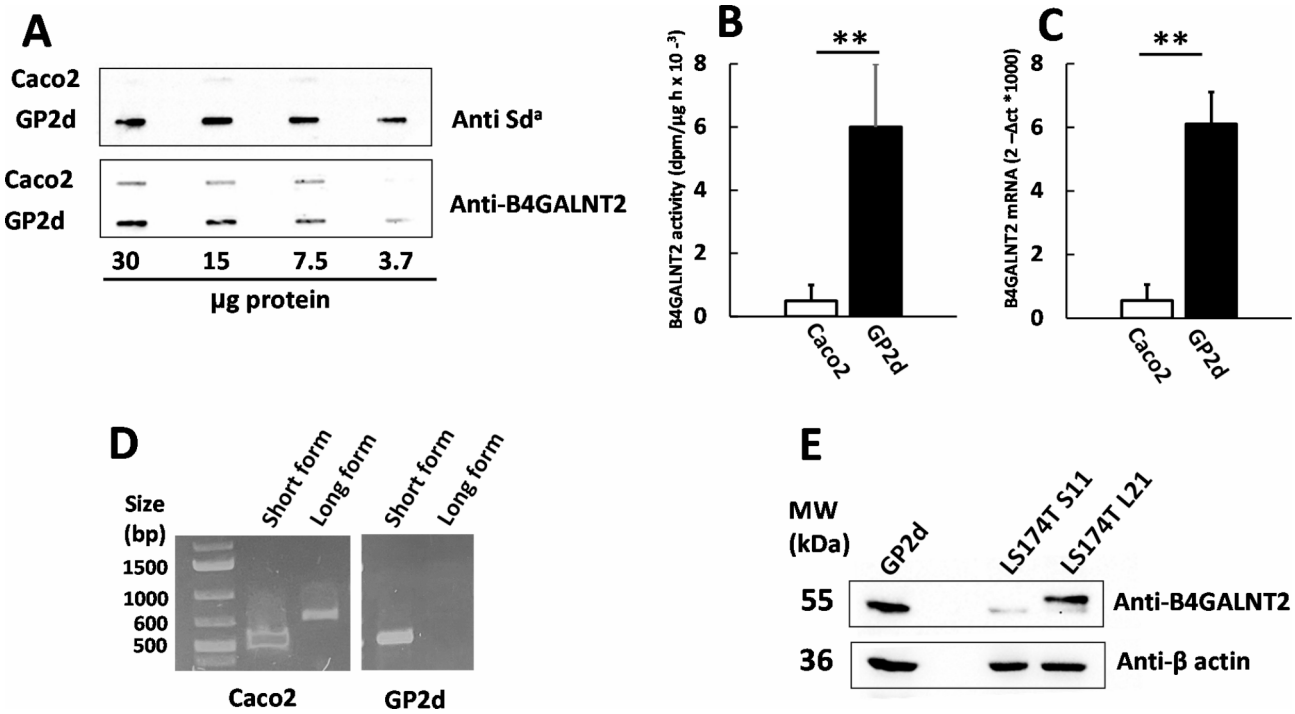

Fig. 3. Characterization of GP2d and Caco2 cells. (**A**) slot blot analysis of cell homogenates from GP2d and Caco2 with anti-Sd[a] or anti B4GALNT2 antibodies. (**B**) B4GALNT2 enzyme activity of the two cell lines. (**C**) qRT-PCR of the two cell lines. To the B4GALNT2 cycle threshold (CT), the CT of β actin was subtracted (ΔCT). Data are reported as $2^{-\Delta CT}$*1000. (**D**) End point RT-PCR with primers specific either for the short or the long form. Amplification was performed with primer pairs L.18/R.28 (short) and L5/R8 (long form) for 35 cycles. (**E**) Western blot analysis with anti B4GALNT2 antibodies of GP2d and the cell lines LS174T-S11 and LS174T-L21, expressing either the short or the long B4GALNT2 form[18]. In (**B**,**C**), statistical analysis was performed by the Student's t test for independent samples on at least 3 independent experiments. $^{**}p \leq 0.01$.

constructs. The B4GALNT2 mRNA level was significantly reduced by transfection with FOXD1, PGRA and PGRB but only partially by FOXF2 (Fig. 4B). However, the enzymatic activity was significantly inhibited by all 4 constructs, although the effect was particularly strong for FOXD1 (Fig. 4C).

FOXD1 inhibits B4GALNT2 expression also in Caco2 but not in a cell line constitutively expressing B4GALNT2

To confirm the effect of FOXD1 on *B4GALNT2* transcription in other cell models, we chose Caco2, in which B4GALNT2 is expressed at a much lower level than GP2d, and LS174T-S11 in which *B4GALNT2* is driven by the cytomegalovirus (CMV) promoter. In both, Caco2 and LS174T-S11, FOXD1 is barely or not expressed but becomes very strongly expressed upon transient transfection with FOXD1 expression vector (Fig. 5A). In Caco2 cells, FOXD1 induced an approximately 65% reduction of the B4GALNT2 mRNA level, which is even stronger than that induced in GP2d (Fig. 5B). However, owing to the much lower absolute level of B4GALNT2 enzyme activity in Caco2 and the poorer sensitivity of the enzymatic assay compared with qRT-PCR, the reduction of enzyme activity was not significant (Fig. 5C). In the cell line LS174T-S11, which results from the stable transfection with the short form B4GALNT2 cDNA under a strong CMV promoter, transfection with FOXD1 was ineffective both at the transcriptional as well as at the enzymatic level (Fig. 5B,C), indicating that only the physiological promoter of *B4GALNT2* is inhibited by FOXD1.

Progressive deletion of FOXD1 binding sites results in reduced FOXD1 response

To get insights into the relative importance of the different FOXD1 binding sites on transcriptional activity, the 2773 bp genomic region upstream the translational start up to the nucleotide (nt) position 72 of the coding region of the B4GALNT2 short form was cloned in frame with the firefly luciferase gene (Fig. 6). Regions containing putative FOXD1 binding sites (indicated by color bars) were progressively truncated, to obtain deletion mutants. Luciferase activity was measured in GP2d after transfection with the luciferase constructs with or without co-transfection with the FOXD1 expression vector. The expression of the full-length construct is quite low, but strongly inhibited by FOXD1 co-transfection, supporting the role of this site as inhibitory. Deletion of this site (Del.A) results in a roughly 6-fold elevation of luciferase activity. If it is considered that according to the CCLE and to our data (Fig. 4A), GP2d lack any endogenous FOXD1 expression, it is not clear the origin of this enhancement. However, also Del.A is strongly responsive to FOXD1 co-transfection. Further deletion of the cluster of 3 sites on both DNA strains (Del.B), results in little changes to luciferase activity but to

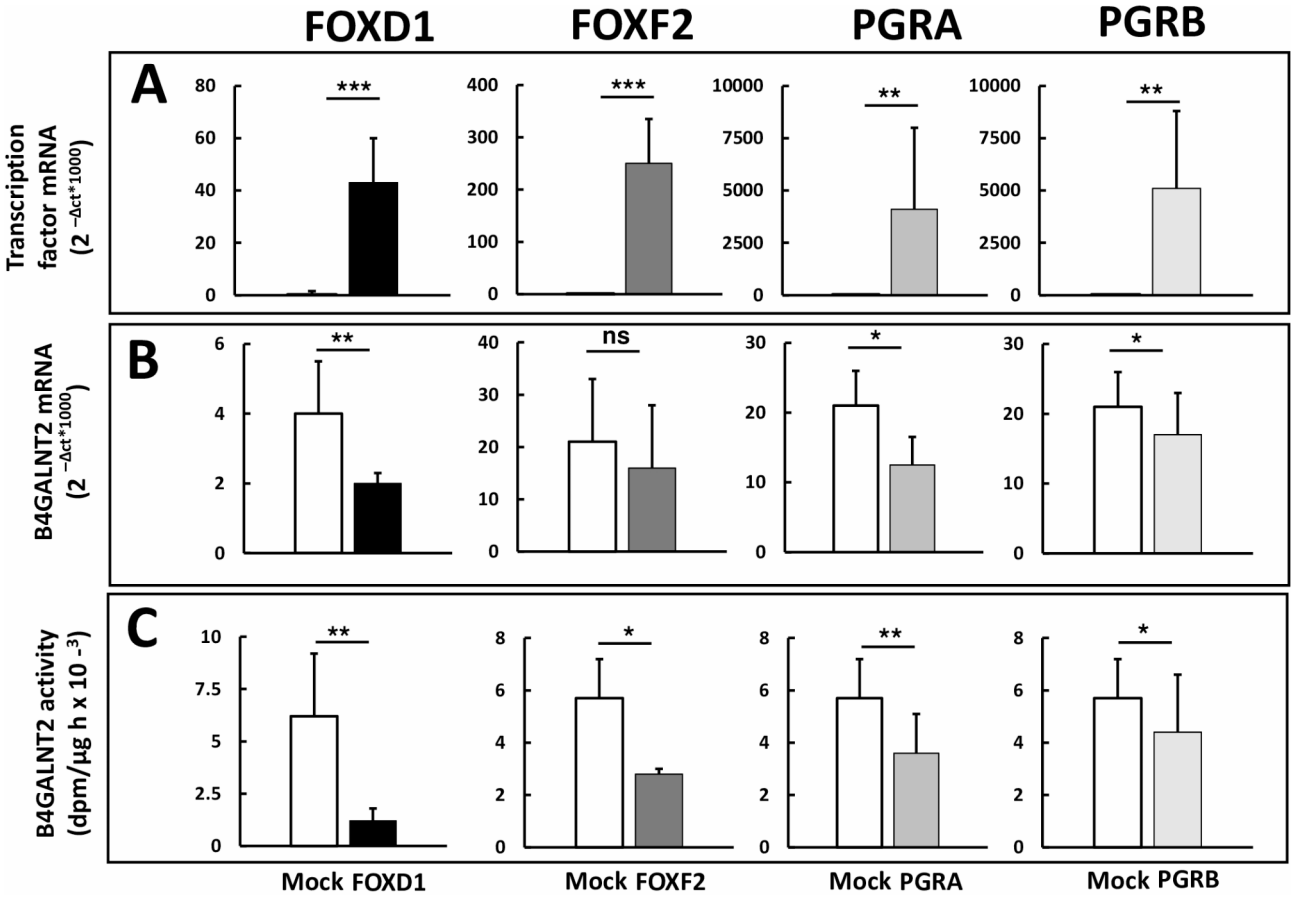

Fig. 4. Effect of transcription factor expression vectors in GP2d cells. GP2d cells were transiently transfected with expression vectors for FOXD1, FOXF2, PGRA and PGRB. Histograms represent the mean ± SD of at least 3 transfection experiments. The mRNA level of the transfected transcription factor (**A**) as well as that of B4GALNT2 (**B**) were measured by qRT-PCR and normalized as described in the legend of Fig. 3C. The B4GALNT2 enzyme activity was also measured (**C**). Statistical analysis was performed by the Student's t test for independent samples. $^{*}p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$.

reduced FOXD1 responsivity, compared with Del.A. Deletion of the additional 2 sites located on opposite DNA strains (Del.C) leads to a drastic reduction of luciferase activity and abolishes FOXD1 responsivity. Considering that constructs Del.B and Del.C differ for only 89 nt, such a dramatic decrease of expression was unexpected. It can be hypothesized that binding sequences for crucial factors necessary for B4GALNT2 transcription are present in this deleted region.

FOXD1 inhibition results in a partial B4GALNT2 activation in SW948 cells

To reinforce the notion that FOXD1 plays a major role in B4GALNT2 inhibition, we down-regulated FOXD1 in the cell line SW948. This cell line was chosen because of its negligible B4GALNT2 level and high FOXD1 level. Transient inhibition of FOXD1 was achieved by transfection with two different siRNA (siRNA1 and siRNA2). FOXD1 and B4GALNT2 mRNA were monitored by qRT-PCR. In Fig. 7A and B, it is shown the mean ± SD of the FOXD1 and B4GALNT2 mRNA expression levels in siRNA-treated or mock treated SW948 cells. As shown, treatment results in ~70% (siRNA1) and ~80% (siRNA2) inhibition of FOXD1. Concomitantly, B4GALNT2 level increased by ~40 and ~80%, respectively. Interestingly, dot plot analysis of the single experimental values of FOXD1 and B4GALNT2 (Fig. 7C), revealed a significant inverse relationship between the two parameters. These data reinforce the role of FOXD1 inhibition, although the B4GALNT2 level in siRNA-treated SW948 cells remained well below that of measured in GP2d.

Transfection with mir-204-5p strongly reduces B4GALNT2 expression

Data base searches reveal that high miR-204-5p is incompatible with high B4GALNT2 expression in both TCGA COAD cohort and CRC cell lines (Fig. 8A,B). Transient transfection in GP2d reveals that a miR-204-5p mimic resulted in a strong and significant inhibition of B4GALNT2 both at the mRNA (Fig. 8C) and at the enzyme activity level (Fig. 8D), compared with a mock transfection.

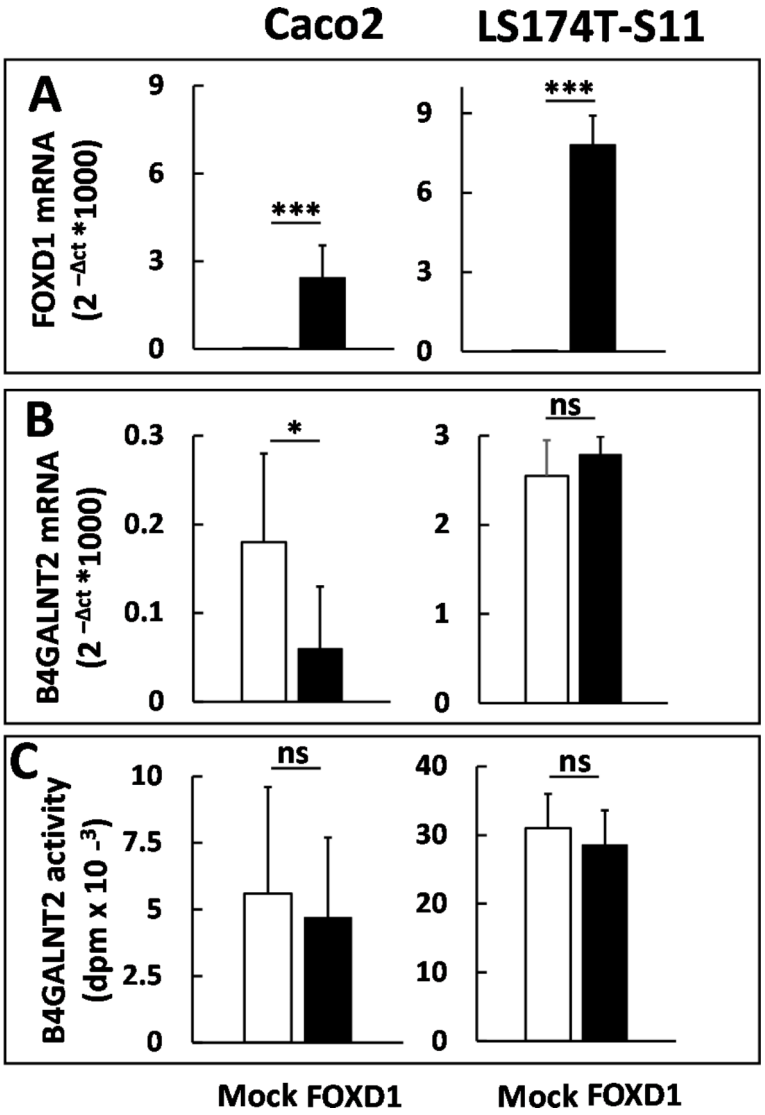

Fig. 5. Effect of FOXD1 expression on B4GALNT2 in Caco2 and LS174T-S11. Caco2 and LS174T-S11 were transiently transfected with FOXD1. mRNA expression was determined by qRT-PCR for FOXD1 (**A**) and B4GALNT2 (**B**) as described in the legend to Fig. 3. The B4GALNT2 enzyme activity was also determined (**C**). Data are the mean ± SD of at least three experiments. Statistical analysis was performed by the Student's t test for independent samples. $^{*}p \leq 0.05$; $^{***}p \leq 0.001$; ns = non significant.

Discussion

Since its first identification more than 40 years ago as the Sd^a synthase[7], B4GALNT2 has been found to be involved in the regulation of a variety of biological processes, including sheep prolificacy, fertilization and gestation, resistance to microbial and viral infections and attenuation of Duchenne muscular dystrophy (reviewed in[6,10]). However, the key issue of the molecular bases of its dramatic down-regulation in colon cancer is still open. Previous experimental investigations[20,21] have shown the importance of promoter methylation for the block of *B4GALNT2* expression. However, de-methylation treatments resulted only in a partial activation of *B4GALNT2* transcription, far below the level of a normal colonic mucosa. In addition, many unmethylated CRC cell lines fail to express B4GALNT2. Detailed analysis of the methylation status of *B4GALNT2* deduced from TCGA data[23] revealed that methylation of sites in the CpG island effectively blocks B4GALNT2 expression, although methylation of a "northern shore" in the promoter and of an "open sea" site in the intron between exons 6 and 7, results in a powerful activation[23] (Fig. 9). As shown, methylation is inversely correlated with B4GALNT2 in all sites, except in one of the "northern shore" and in one "open sea", in which it is positively correlated. In addition, TCGA data have indicated that the expression of the miR-204-5p is incompatible with B4GALNT2 expression in cancer samples. However, from a detailed analysis it appears that many cancer samples lacking blocking methylation of specific sites and/or miR-204-5p fail to express B4GALNT2. Thus, it appears that these two "brakes" so far identified are not the only responsible for *B4GALNT2* switch off in CRC. In this regard, a poorly investigated aspect is the role of transcription factors. A big step in this direction has been the recent

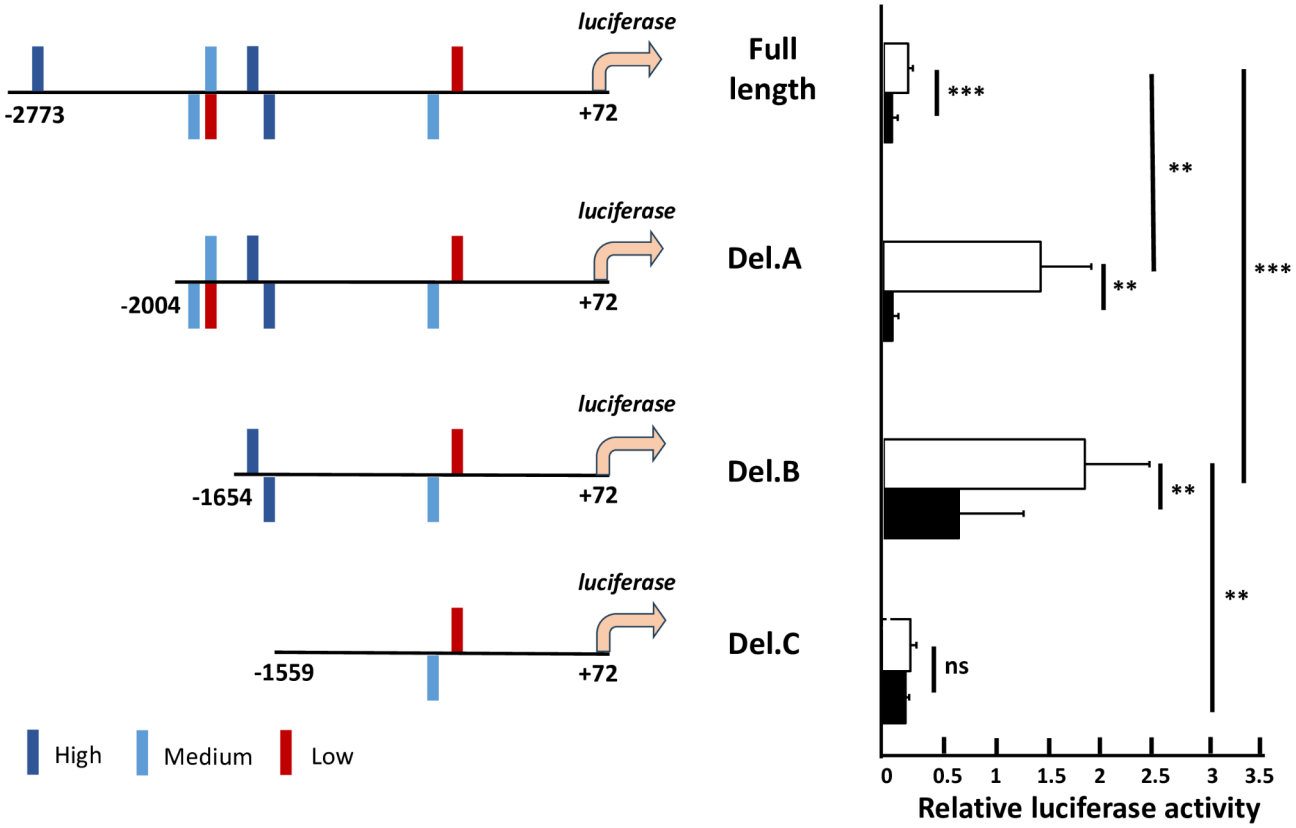

Fig. 6. Luciferase activity of deletion mutants. The genomic region encompassing the −2273/+72 nt (where nt + 1 is the A of the ATG translational start codon of the B4GALNT2 short form) was cloned in frame with the firefly luciferase gene (full length construct). Deletion mutants were obtained as described in Methods. Putative FOXD1 binding sites were indicated by vertical bars. The color of the bar represents of the binding score of the site. Bars above the DNA indicate that the site is on the + strain, while those below the DNA indicate that the site is on the − strain. Histograms represent the luciferase activity induced by transfection of the clone either in the absence (white) or in the presence (black) of FOXD1 co-transfection. Data are the mean ± SD of at least 3 experiments and were analyzed by the Student's t test for independent samples. $^{*}p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$; ns = non-significant.

identification of the transcription factors ETS1 and SP1 as responsible for *B4GALNT2* transcription, although neither of the two is differentially expressed in normal and cancer cases[24]. In this work, we have first identified several transcription factors with putative binding sites in a long genomic sequence upstream the *B4GALNT2* gene. We have then screened databases to identify putative transcription factors displaying a positive or negative relationship with B4GALNT2. FOXD1, FOXF2 and PGR appeared to be promising candidates. Of the three, FOXD1 appeared to be the most convincing as B4GALNT2 repressor. In fact, it was virtually not expressed by normal mucosa while it was inversely correlated with B4GALNT2 in CRC samples and cell lines. The number of FOXD1 binding sites in the B4GALNT2 regulatory region is much higher than that of the other two factors. Moreover, its high expression correlates with shorter survival[26] (Fig. 2D). Our transfection experiments have demonstrated the power of FOXD1 to suppress B4GALNT2 in cellular models of high (GP2d) and medium (Caco2) B4GALNT2 expression, while promoter deletion experiments have revealed the different importance of the binding sites. Attempts to activate B4GALNT2 in the cell line SW948 by FOXD1 knock down has achieved a level of B4GALNT2 expression which remained far lower than that of GP2d. This cell line was chosen because of its very low level of promoter methylation[21]. The only partial reactivation of B4GALNT2 may depend either from the residual FOXD1 expression after knock-down or from other still unidentified mechanisms. FOXD1 is an oncogene member of the forkhead box transcription factor family[27]. It promotes nuclear localization of β-catenin[28], increasing stemness and chemoresistance[29]. Considering that COAD patients expressing high B4GALNT2 do not express FOXD1, it may be thought that their better survival[22,23] was dependent by their lack of FOXD1 expression. However, this is not the case, since we have previously shown that B4GALNT2/Sd^{a} expression is per se responsible for reduced in vitro malignancy and reduced stemness[22,30].

Transfection with either FOXF2 or PGR also resulted in a significant inhibition of B4GALNT2 mRNA and/or enzyme activity, although at a much lower level than FOXD1. The role of FOXF2 and PGR in cancer and particularly in CRC is less clear. A meta-analysis study has shown that high FOXF2 levels are associated with good prognosis in various cancers prone to bone metastasis[31]. Although it appears to play opposite roles in different malignancies[32], FOXF2 is down-regulated in CRC (Fig. 2C), while its inhibition by miR-132a[33] and

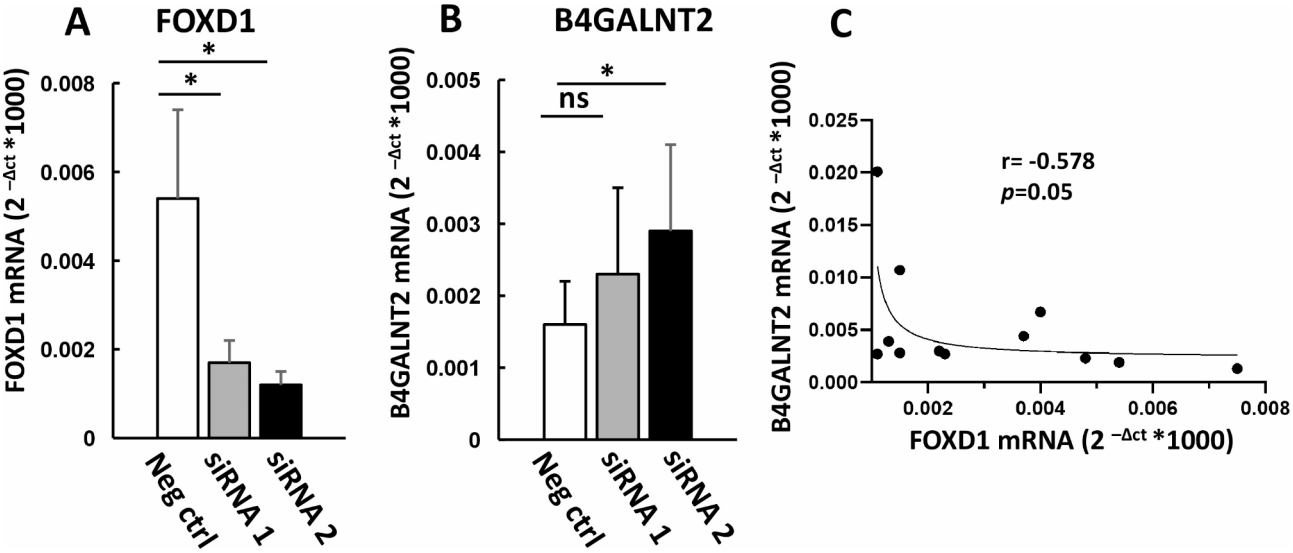

Fig. 7. Activation of B4GALNT2 expression by FOXD1 knock down in SW948 cells. Mean ± SD of qRT-PCR data of SW948 cells transfected with two different FOXD1 siRNA (siRNA1 and siRNA2) or with a negative control. (**A**) FOXD1 expression; (**B**) B4GALNT2 expression. (**C**) Relationship between FOXD1 and B4GALNT2 expression levels in each individual experiment. Analysis by Spearman correlation**.** In A and B, data are the mean ± SD of 4 experiments. Statistical analysis was performed by Student's t test for paired samples. $^{*}p \leq 0.05$; ns = non-significant.

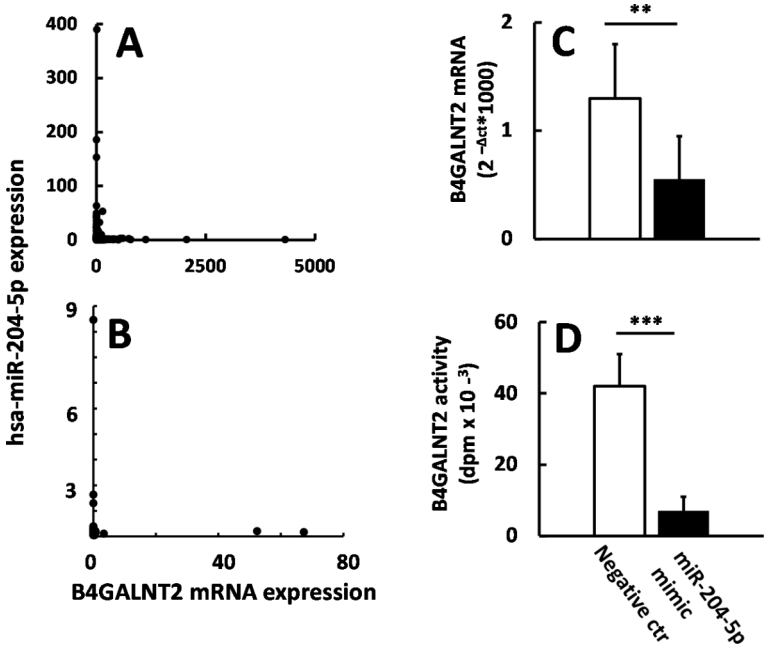

Fig. 8. B4GALNT2 inhibition by miR-204-5p. Correlation of B4GALNT2 and miR-204-5p expression in CRC TCGA samples (**A**) and CRC cell lines from CCLE (**B**). B4GALNT2 expression determined by qRT-PCR (**C**) and B4GALNT2 enzyme activity (**D**) in GP2d transfected with miR-204-5p mimic or mock transfected (negative ctr). Data are the mean ± SD of 4 experiments. Analysis was performed by Student's t test for paired samples. $^{**}p \leq 0.01$; $^{***}p \leq 0.001$.

miR-182[34] promotes malignancy in CRC. PGR is down-regulated in CRC, although high expression is associated with a tendency (not statistically significant) to worse prognosis (Fig. 2D). However, in vitro studies have shown an inhibitory and pro-apoptotic effect of progesterone and PGR in CRC cells[35]. Progesterone regulates gene transcription through the two forms (PGRA and PGRB) of the PGR, differing for the presence in PGRB of an

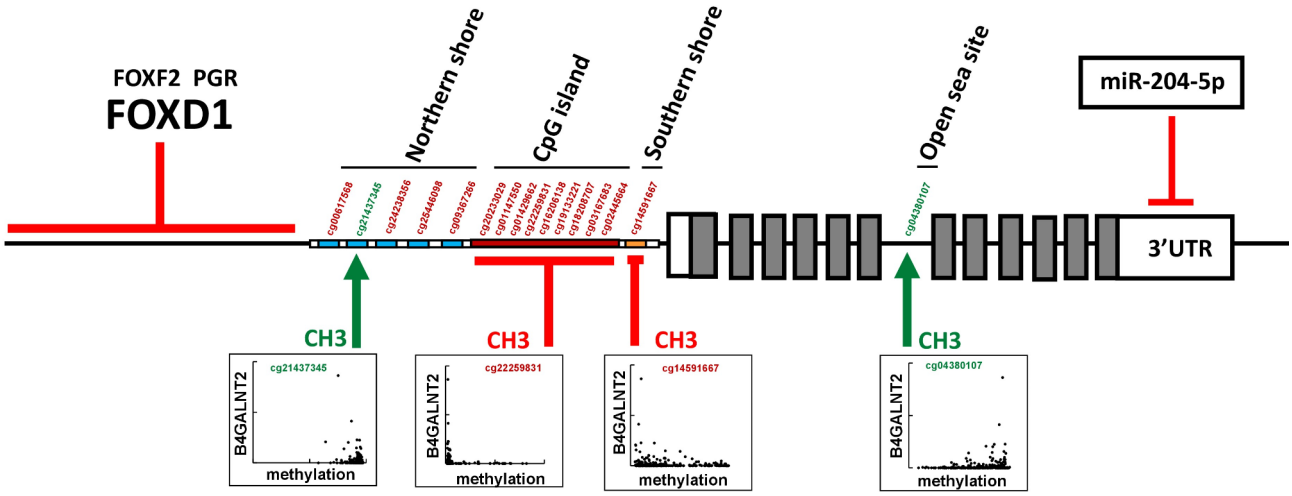

Fig. 9. Schematic representation of the *B4GALNT2* regulatory mechanisms. Coding exons are represented in gray, the untranslated regions (UTR) in white, while genomic DNA as a continuous black line. The length of the promoter region has been exaggerated, while exons are depicted of the same size for simplicity. Inhibitory signals are indicated in red, while activation signals are represented by green arrows. CH3 indicates methylation. Methylation sites in the CpG island, the "northern" and "southern shores" and the "open sea" site are indicated. In bottom plots is indicated the relationship between B4GALNT2 and methylation of the indicated sites. cg22259831 is representative of methylation of all sites in the CpG island. Methylation is negatively associated with B4GALNT2 in all sites except cg21437345 in the "northern shore" and cg04380107 in the "open sea" site.

extra 164 amino acid NH_2-terminal sequence. Even if PGRB is known to exhibit a greater transcriptional activity, compared with PGRA[36], our data indicate a comparable effect of the two forms in B4GALNT2 inhibition.

Among the many colon cancer cell lines available, GP2d is unique because of its very high B4GALNT2 level, close to that of a normal mucosa. The cell line Caco2, which was so far considered the top B4GALNT2 expresser[25] and served as a source for the cloning of the first human B4GALNT2 cDNAs[8,9], compared with GP2d appears to be only a medium expresser. Nevertheless, FOXD1 actively suppresses *B4GALNT2* transcription also in this cell line. On the other hand, the lack of FOXD1 sensitivity when *B4GALNT2* is driven by a non-physiological constitutive promoter, as occurs in LS174T-S2, runs out the possibility of post-translational inhibitory effects.

Although the long B4GALNT2 form appears to be involved in relevant biological phenomena[16], the short form was by far the predominant B4GALNT2 transcript in the vast majority of normal and CRC samples[11,15]. For this reason, this work was focused only on the regulation of the short form, which is the only expressed by GP2d. The expression of the long and the short forms is obviously regulated by different mechanisms, whose investigation will require additional investigations.

Beside the crucial role of FOXD1, in this work we have characterized miR-204-5p as a key player in B4GALNT2 inhibition in CRC. This miRNA exerts a tumor suppressor role in various malignancies[37], including CRC[38]. Thus, its inhibitory effect on the expression of *B4GALNT2*, a gene associated with reduced malignancy in CRC is surprising and requires further investigations. Together, available data from this and previous studies support a model in which the regulation of *B4GALNT2* expression in normal and cancer colon takes place at least at three different levels: presence/absence of FOXD1 (with FOXF2 and PGR minor contribution); methylation status of specific CpG sites in the promoter and in an open sea position; presence/absence of miR-204-5p. All the three factors have per se the power to block *B4GALNT2* expression, but the removal of each of the three is a necessary but not sufficient condition for *B4GALNT2* activation. It is likely that other, still unknown mechanisms, may play a role in the picture. In the light of the clear relevance as favorable prognostic marker of CRC, the elucidation of these mechanisms is highly relevant.

Methods

Bioinformatic analysis

Gene expression data for 623 colorectal adenocarcinoma (COAD) samples and 51 normal colonic tissues were downloaded from the TCGA database COAD cohort using the Firebrowse website (http://firebrowse.org) and normalized by RNA-Seq Expectation Maximization (RSEM). Gene expression data of CRC cell lines were obtained from the Cancer Cell Lines Encyclopedia (CCLE) (https://sites.broadinstitute.org/ccle/). The survival curves were estimated by the Kaplan Meier method through the oncolnc.org website (COAD cohort) and the Mantel–Cox log-rank test was performed to test differences between the survival curves. Search for transcription factor putative binding sites in the − 2800 bp genomic sequence upstream the translation start site of the B4GALNT2 short form was performed through Consite (https://bio.tools/consite) and Promo (https://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3) tools for binding sites prediction.

Cell lines and cell culture

The GP2d cell line (95090714)[39] was obtained from the European Collection of Authenticated Cell Cultures (ECACC, UK Health Security Agency, https://www.culturecollections.org.uk/products/cell-cultures/). Caco2 (HTB-37) and SW948 (CCL-237) cell lines were obtained from American Type Culture Collection (ATCC https://www.atcc.org). GP2d and Caco2 cell lines were cultured in DMEM, while SW948 was grown in Leibowitz's L-15 medium. Both media were supplemented with 10% foetal bovine serum, 1% glutamine and 1% penicillin-streptomycin. Cells were grown in incubator at 37 °C in a humidified atmosphere with (GP2d and Caco2) or without (SW948) 5% CO_2.

Slot and western blot analysis

Cell pellets were homogenized in distilled water and the protein concentration of the homogenates was quantified by the Lowry method. 80–100 µg of protein homogenates were electrophoresed in 8% polyacrylamide gel in reducing conditions, according to Laemmli[40]. Proteins were transferred to PVDF membranes (Bio-Rad Laboratories, Hercules, CA), using Trans Blot Turbo (Bio-Rad Laboratories), blocked with 0.1% BSA and probed 1 h with rabbit IgG polyclonal antibody anti-B4GALNT2 (HPA015721, Sigma-Aldrich, St. Louis, Missouri, US) or with anti-β actin rabbit IgG polyclonal (A2066, Sigma-Aldrich). Then membranes were incubated with secondary anti-rabbit IgG–peroxidase antibody (A6154, Sigma-Aldrich) at room temperature for one hour.

For slot blot analysis, different amounts of protein homogenates were spotted on a nitrocellulose blotting membrane (10600002, Amersham™, Amersham, UK) using SlotBlot™ (PR648, Amersham™). Membranes were subsequently blocked with 0.1% BSA, incubated 1 h at room temperature with anti-Sd^a monoclonal antibody KM694, kindly provided by Kyowa Hakko Kogyo Co. Ltd, Tokyo, Japan, followed by incubation with anti-mouse IgM (µ-chain specific)–peroxidase secondary antibody (A8786, Sigma-Aldrich). Chemiluminescence was detected using the WESTAR NOVA 2.0 ECL system (XLS071 Cyanagen, Bologna, Italy) and acquired with The ChemiDoc XRS + System (Biorad Laboratories).

Transfection of expression plasmids

Expression vectors harbouring a cDNA encoding for specific transcription factors were obtained from Addgene (https://www.addgene.org/). Flag-FOXD1 (Plasmid #153116) and Flag-FOXF2 (Plasmid #153123)[41], constitutively express Flag-tagged human FOXD1 and FOXF2, respectively, under CMV IE94 promoter in mammalian cells. p5M-PR-A (Plasmid #89124) and p5M-hPR-B (Plasmid #89114)[36] constitutively express full-length human progesterone receptor-A and B, respectively, under CMV promoter in mammalian cells. The control *Renilla* luciferase plasmid (phRL TK 5BoxB sp 36 # 115365) was also from Addgene. Control plasmid (pcDNA™3.1) was purchased from Invitrogen™ (V79020). Transient transfections were performed on 70% confluent cells with 1 µg of plasmid DNA using Lipofectamine™ 2000 Transfection Reagent (Invitrogen™, Waltham, Massachusetts, US) according to manufacturer's instructions. Cells were incubated with DNA-lipid complexes for 48–72 h, and harvested.

RNA extraction and reverse transcriptase polymerase chain reaction (RT-PCR)

Total RNA from transfected cells was extracted using TRIzol™ (Invitrogen) reagent. cDNA was obtained by using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems™, Waltham, Massachusetts, US) from a starting amount of 2 µg RNA. qRT-PCR was performed using SsoAdvanced™ Universal SYBR® Green Supermix (Biorad Laboratories), according to manufacturer instructions. Subsequently, the Ct values of the samples were determined on CFX96 Touch Real-Time PCR Detection System (Biorad Laboratories). *FOXD1*, *FOXF2*, *PGRA*, *PGRB* and *B4GALNT2* relative expressions were quantified with the $2^{-\Delta Ct}$*1000 method, using β-actin expression as the internal reference. *FOXD1* amplification was performed under denaturing conditions (DMSO 5%) to bypass the presence of repeated GC-rich regions in FOXD1 sequence.

Endpoint RT-PCR was performed using AmpliTaq Gold™ Fast PCR Master Mix (Applied Biosystems™). All primer sequences are listed in Table S1.

B4GALNT2 enzyme activity assay

B4GALNT2 enzyme activity was measured in cell homogenates by the difference of incorporation of [^{3}H] GalNAc on fetuin and asialofetuin, as previously described[8].

hsa-mir-204-5p transfection

Transient transfection with hsa-miR-204-5p mirVana™ miRNA mimics (Ambion, Catalog number 4464066, ID MC11116) was performed using Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen), according to manufacturer instructions. Simultaneously, cells were transfected with their counterpart mirVana™ miRNA Mimic, Negative Control #1 (Ambion, Catalog number: 4464058). 24–48 h after transfection, cells were harvested.

miRNA-specific RNA extraction and quantification

RNA was extracted using mirVana™ miRNA Isolation Kit, with phenol (Invitrogen). Reverse transcription was performed using TaqMan™ MicroRNA Reverse Transcription Kit. Five microliters of double-stranded template (up to 10 ng) and 3 µl of the specific 5× RT primer were mixed and incubated at 85 °C for 5 min, then at 60 °C for 5 min. Successive reverse transcription was performed according to manufacturer's instructions. Quantification was performed by a modified real-time approach with the TaqMan microRNA assay (Applied Biosystems) containing a hsa-miR-204-5p-specific stem-loop primer and pre-formulated assay respectively. 20× TaqMan MicroRNA assay contained the PCR primers and probes (5′-FAM) and TaqMan Universal Master mix no UNG (Applied Biosystems) according to manufacturer instructions. The reaction was first incubated at 95 °C

for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Normalization was achieved using the endogenous miRNA control RNU48. Data were expressed as $2^{-\Delta Ct}$*1000.

Construction of the reporter vector

Genomic DNA was extracted with the Universal kit for DNA isolation with GeDI reagent (EURX®, Molecular Biology Product, Gdańsk, Poland), following manufacturer's instructions. The approximate 3.9 kb genomic region upstream of the *B4GALNT2* translation start site was amplified by PCR using the Q5® high fidelity DNA polymerase (M0491, New England Biolab) and primer pairs Gib-For-PB4GALNT2 and Gib-Rev-PB4GALNT2 (Table S1). The PCR product was electrophoresed, extracted from agarose gel and purified using the Nucleospin® Gel and PCR clean-up purification kit (Macherey-Nagel). pGL3-Basic vector (Promega, Madison, USA) was linearized by digestion using *Sma*I restriction sites, and extracted as above. The purified fragment was cloned into pGL3-Basic vector (Promega, Madison, USA) upstream of the Firefly luciferase gene using the Gibson Assembly Master Mix kit (NEB #M5510) following the manufacturer's instructions. The resulting plasmid was designed pGL3 − 3850 /+ 72. After bacterial transformation using DH5-α competent *E. coli* (#C2987 NEB), plasmid was purified and sequenced (Eurofins Genomics, Germany). Specific primers for sequencing are listed in Table S1.

Construction of truncated promoter mutants

The reporter vector pGL3 − 3850/+ 72 was shortened of about 1kB in the upstream genomic regulatory region of *B4GALNT2* using *Nhe*I-HF restriction enzyme (R3131S, New England Biolabs). The resulting plasmid (pGL3 − 2773 /+ 72) was designed thereafter as "Full length". Starting from the full length, truncated promoter constructs with progressively reduced number of FOXD1 binding sites were generated by PCR (Table S1) followed by enzymatic digestions. PCR reaction was performed using Q5® high fidelity DNA polymerase, and primers listed in Table S1. PCR products obtained were extracted and purified from electrophoresis gel, as previously described, subjected to restriction enzyme digestion with *Nhe*I-HF (R3131S, New England Biolabs) and to ligation of sticky ends by T4 DNA ligase (M0202S, New England Biolabs). The PCR product was then digested with the *Dpn*I restriction enzyme (R0176S, New England Biolabs) for 1 h at 37 °C. The two resulting truncated constructs pGL3 − 2004/+ 72, and pGL3 − 1654/+ 72 were named Del.A and Del.B, respectively. After bacterial transformation as previously described, plasmid was purified and sequenced (Eurofins Genomics) together with pGL3 basic vector (negative control) and pGL3 control vector (positive control). Specific primers for sequencing are listed in Table S1.

Luciferase assay

To assay Luciferase activity, 70% confluent cells underwent transient transfection with 1 µg of pGL3 construct alone or in presence of 1 µg of FOXD1 plasmid, and with 20 ng of control *Renilla* plasmid. For each set of transfection experiments, in parallel we performed transfections with basic vector (negative control) and pGL3 control vector (positive control). After 4 h, the medium was replaced by fresh culture medium and after 24 h cells were lysed with Passive Lysis Buffer (PLB, Dual Luciferase Reporter Assay System, Promega) according to manufacturer instructions. For the measurement of luciferase activity, 20 µl of lysate were used for each condition, and luminescence was detected with the Centro luminometer (Berthold Technologies, Bad Wildbad, Germany). Experimental data were normalized according to the following formula: Normalized datum = (Experimental datum − negative control)/positive control.

Statistical analysis

Data were analyzed with the GraphPad Prism 8 software. The t test for paired or independent samples was usually used. Significance was indicated as follows: $*p \le 0.05$; $**p \le 0.01$; $*** p \le 0.001$.

Data availability

The dataset used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 19 September 2024; Accepted: 3 January 2025
Published online: 13 January 2025

References

1. Stanley, P. Genetics of glycosylation in mammalian development and disease. *Nat. Rev. Genet.* https://doi.org/10.1038/s41576-024-00725-x (2024).
2. Dall'Olio, F., Malagolini, N., Trinchera, M. & Chiricolo, M. Mechanisms of cancer-associated glycosylation changes. *Front. Biosci.* **17**, 670–699 (2012).
3. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: Mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555. https://doi.org/10.1038/nrc3982 (2015).
4. Pucci, M., Duca, M., Malagolini, N. & Dall'Olio, F. Glycosyltransferases in cancer: Prognostic biomarkers of survival in patient cohorts and impact on malignancy in experimental models. *Cancers (Basel)* **14**, 2128. https://doi.org/10.3390/cancers14092128 (2022).
5. Dall'Olio, F., Pucci, M. & Malagolini, N. The cancer-associated antigens Sialyl Lewis$^{a/x}$ and Sd^{a}: Two opposite faces of terminal glycosylation. *Cancers (Basel)* **13**, 5273. https://doi.org/10.3390/cancers13215273 (2021).
6. Dall'Olio, F., Malagolini, N., Chiricolo, M., Trinchera, M. & Harduin-Lepers, A. The expanding roles of the Sd^{a}/Cad carbohydrate antigen and its cognate glycosyltransferase B4GALNT2. *Biochim. Biophys. Acta* **1840**, 443–453 (2014).
7. Serafini-Cessi, F. & Dall'Olio, F. Guinea-pig kidney β-N-acetylgalactosaminyltransferase towards Tamm-Horsfall glycoprotein. Requirement of sialic acid in the acceptor for transferase activity. *Biochem. J.* **215**, 483–489 (1983).

8. Lo, P. L., Cabuy, E., Chiricolo, M. & Dall'Olio, F. Molecular cloning of the human β1,4 N-acetylgalactosaminyltransferase responsible for the biosynthesis of the Sd^a histo-blood group antigen: The sequence predicts a very long cytoplasmic domain. *J. Biochem. (Tokyo)* **134**, 675–682 (2003).
9. Montiel, M. D., Krzewinski-Recchi, M. A., Delannoy, P. & Harduin-Lepers, A. Molecular cloning, gene organization and expression of the human UDP-GalNAc:Neu5Acα2-3Galβ-R β1,4-N-acetylgalactosaminyltransferase responsible for the biosynthesis of the blood group Sda/Cad antigen: Evidence for an unusual extended cytoplasmic domain. *Biochem. J.* **373**, 369–379 (2003).
10. Duca, M., Malagolini, N. & Dall'Olio, F. The story of the Sd^a antigen and of its cognate enzyme B4GALNT2: What is new?. *Glycoconj. J.* **40**, 123–133. https://doi.org/10.1007/s10719-022-10089-1 (2023).
11. Groux-Degroote, S., Vicogne, D., Cogez, V., Schulz, C. & Harduin-Lepers, A. B4GALNT2 controls Sd^a and SLe^x antigen biosynthesis in healthy and cancer human colon. *Chembiochem* **22**, 3381–3390. https://doi.org/10.1002/cbic.202100363 (2021).
12. Stenfelt, L. et al. Glycoproteomic and phenotypic elucidation of B4GALNT2 expression variants in the SID histo-blood group system. *Int. J. Mol. Sci.* **23**, 3936. https://doi.org/10.3390/ijms23073936 (2022).
13. Stenfelt, L. et al. Missense mutations in the C-terminal portion of the *B4GALNT2*-encoded glycosyltransferase underlying the Sd^{a-} phenotype. *Biochem. Biophys. Rep.* **19**, 100659. https://doi.org/10.1016/j.bbrep.2019.100659 (2019).
14. Stenfelt, L., Hellberg, A. & Olsson, M. L. SID: A new carbohydrate blood group system based on a well-characterized but still mysterious antigen of great pathophysiologic interest. *Immunohematology* **39**, 1–10. https://doi.org/10.21307/immunohematology-2023-002 (2023).
15. Groux-Degroote, S. et al. *B4GALNT2* gene expression controls the biosynthesis of Sd^a and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract. *Int. J. Biochem. Cell Biol.* **53**, 442–449 (2014).
16. Groux-Degroote, S. et al. The extended cytoplasmic tail of the human B4GALNT2 is critical for its Golgi targeting and post-Golgi sorting. *FEBS J.* **285**, 3442–3463. https://doi.org/10.1111/febs.14621 (2018).
17. Malagolini, N. et al. Expression of UDP-GalNAc:NeuAc α2,3Gal β-R beta 1,4(GalNAc to Gal) N-acetylgalactosaminyltransferase involved in the synthesis of Sd^a antigen in human large intestine and colorectal carcinomas. *Cancer Res.* **49**, 6466–6470 (1989).
18. Malagolini, N., Santini, D., Chiricolo, M. & Dall'Olio, F. Biosynthesis and expression of the Sd^a and sialyl Lewis x antigens in normal and cancer colon. *Glycobiology* **17**, 688–697 (2007).
19. Dohi, T. et al. Detection of N-acetylgalactosaminyltransferase mRNA which determines expression of Sd^a blood group carbohydrate structure in human gastrointestinal mucosa and cancer. *Int. J. Cancer* **67**, 626–631 (1996).
20. Kawamura, Y. I. et al. DNA hypermethylation contributes to incomplete synthesis of carbohydrate determinants in gastrointestinal cancer. *Gastroenterology* **135**, 142–151 (2008).
21. Wang, H. R., Hsieh, C. Y., Twu, Y. C. & Yu, L. C. Expression of the human Sd^a β-1,4-N-acetylgalactosaminyltransferase II gene is dependent on the promoter methylation status. *Glycobiology* **18**, 104–113 (2008).
22. Pucci, M. et al. High expression of the Sd^a synthase B4GALNT2 associates with good prognosis and attenuates stemness in colon cancer. *Cells* **9**, 948. https://doi.org/10.3390/cells9040948 (2020).
23. Pucci, M., Malagolini, N. & Dall'Olio, F. Glycosyltransferase B4GALNT2 as a predictor of good prognosis in colon cancer: Lessons from databases. *Int. J. Mol. Sci.* **22**, 4331. https://doi.org/10.3390/ijms22094331 (2021).
24. Wavelet-Vermuse, C. et al. Analysis of the proximal promoter of the human colon-specific *B4GALNT2* (Sd^a synthase) gene: *B4GALNT2* is transcriptionally regulated by ETS1. *Biochim. Biophys. Acta Gene Regul. Mech.* **1864**, 194747. https://doi.org/10.1016/j.bbagrm.2021.194747[doi] (2021).
25. Malagolini, N., Dall'Olio, F. & Serafini-Cessi, F. UDP-GalNAc:NeuAc α 2,3Gal β-R (GalNAc to Gal) β 1,4-N-acetylgalactosaminyltransferase responsible for the Sda specificity in human colon carcinoma CaCo-2 cell line. *Biochem. Biophys. Res. Commun.* **180**, 681–686 (1991).
26. Zong, Y. et al. Combination of FOXD1 and Plk2: A novel biomarker for predicting unfavourable prognosis of colorectal cancer. *J. Cell Mol. Med.* **26**, 3471–3482. https://doi.org/10.1111/jcmm.17361 (2022).
27. Cheng, L., Yan, H., Liu, Y., Guan, G. & Cheng, P. Dissecting multifunctional roles of forkhead box transcription factor D1 in cancers. *Biochim. Biophys. Acta Rev. Cancer* **1878**, 188986. https://doi.org/10.1016/j.bbcan.2023.188986 (2023).
28. Feng, W. Q. et al. FOXD1 promotes chemotherapy resistance by enhancing cell stemness in colorectal cancer through β-catenin nuclear localization. *Oncol. Rep.* **50**, 134. https://doi.org/10.3892/or.2023.8571 (2023).
29. Koga, M. et al. Foxd1 is a mediator and indicator of the cell reprogramming process. *Nat. Commun.* **5**, 3197. https://doi.org/10.1038/ncomms4197 (2014).
30. Pucci, M., Gomes, F. I., Malagolini, N., Ferracin, M. & Dallolio, F. The Sd^a synthase B4GALNT2 reduces malignancy and stemness in colon cancer cell lines independently of Sialyl Lewis X inhibition. *Int. J. Mol. Sci.* **21**, 6558. https://doi.org/10.3390/ijms21186558 (2020).
31. Chen, Q. et al. Forkhead box F2 as a novel prognostic biomarker and potential therapeutic target in human cancers prone to bone metastasis: A meta-analysis. *J. Int. Med. Res.* **49**, 1–18. https://doi.org/10.1177/03000605211002372 (2021).
32. Wu, Q., Li, W. & You, C. The regulatory roles and mechanisms of the transcription factor FOXF2 in human diseases. *PeerJ* **9**, e10845. https://doi.org/10.7717/peerj.10845 (2021).
33. Chen, W., Tong, K. & Yu, J. MicroRNA-130a is upregulated in colorectal cancer and promotes cell growth and motility by directly targeting forkhead box F2. *Mol. Med. Rep.* **16**, 5241–5248. https://doi.org/10.3892/mmr.2017.7257 (2017).
34. Zhang, Y. et al. miR-182 promotes cell growth and invasion by targeting forkhead box F2 transcription factor in colorectal cancer. *Oncol. Rep.* **33**, 2592–2598. https://doi.org/10.3892/or.2015.3833 (2015).
35. Zhang, Y. L. et al. Progesterone suppresses the progression of colonic carcinoma by increasing the activity of the GADD45α/JNK/c-Jun signalling pathway. *Oncol. Rep.* **45**, 95. https://doi.org/10.3892/or.2021.8046 (2021).
36. Su, S. et al. Primate-specific melanoma antigen-A11 regulates isoform-specific human progesterone receptor-B transactivation. *J. Biol. Chem.* **287**, 34809–34824. https://doi.org/10.1074/jbc.M112.372797 (2012).
37. Yang, F., Bian, Z., Xu, P., Sun, S. & Huang, Z. MicroRNA-204-5p: A pivotal tumor suppressor. *Cancer Med.* **12**, 3185–3200. https://doi.org/10.1002/cam4.5077 (2023).
38. Yin, Y. et al. miR-204-5p inhibits proliferation and invasion and enhances chemotherapeutic sensitivity of colorectal cancer cells by downregulating RAB22A. *Clin. Cancer Res.* **20**, 6187–6199. https://doi.org/10.1158/1078-0432.CCR-14-1030 (2014).
39. Solic, N. et al. Two newly established cell lines derived from the same colonic adenocarcinoma exhibit differences in EGF-receptor ligand and adhesion molecule expression. *Int. J. Cancer* **62**, 48–57. https://doi.org/10.1002/ijc.2910620111 (1995).
40. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
41. Moparthi, L. & Koch, S. A uniform expression library for the exploration of FOX transcription factor biology. *Differentiation* **115**, 30–36. https://doi.org/10.1016/j.diff.2020.08.002 (2020).

Acknowledgements

Assistance of Dorothée Vicogne for cell culture and transfections, of Kawthar Agoudgil for the molecular cloning of genomic *B4GALNT2* sequence.This work has been funded in part by the Centre National de La Recherche Scientifique (CNRS), by the Ligue régionale contre le cancer (to A. H.-L.), in part by the University of Bologna and by the Pallotti Legacy for Cancer Research (to F.D.O).

Author contributions

M.D. performed experiments, analyzed data and contributed to write the manuscript; N.M. performed research; M.P. performed bioinformatic analysis; V.C. performed design and construction of deletion mutants; A.L-H designed deletion mutants, analyzed data, critically revised the manuscript; F.D.O. designed the research, analyzed data and wrote the manuscript. All authors have read and approved the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-85450-z.

Correspondence and requests for materials should be addressed to F.D.

Reprints and permissions information is available at www.nature.com/reprints.

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