HAL
open science

CDX2 inducible microRNAs sustain colon cancer by targeting multiple DNA damage response pathway factors

Swati Priya, Ekjot Kaur, Swati Kulshrestha, Awadhesh Pandit, Isabelle Gross, Nitin Kumar, Himanshi Agarwal, Aamir Khan, Radhey Shyam, Prakash Bhagat, et al.

To cite this version:

Swati Priya, Ekjot Kaur, Swati Kulshrestha, Awadhesh Pandit, Isabelle Gross, et al.. CDX2 inducible microRNAs sustain colon cancer by targeting multiple DNA damage response pathway factors. Journal of Cell Science, 2021, 134 (15), 10.1242/jcs.258601 . hal-03788695

HAL Id: hal-03788695

https://hal.science/hal-03788695v1

Submitted on 26 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

© 2021. Published by The Company of Biologists Ltd.

CDX2 inducible microRNAs sustain colon cancer by targeting multiple DNA damage response pathway factors

Swati Priya[1], Ekjot Kaur[1], Swati Kulshrestha[1], Awadhesh Pandit[1], Isabelle Gross[2], Nitin Kumar[1], Himanshi Agarwal[1], Aamir Khan[1], Radhey Shyam[1], Prakash Bhagat[3], Jyothi S. Prabhu[4], Perumal Nagarajan[1], S. V. S. Deo[3], Avinash Bajaj[5], Jean-Noël Freund[2], Arnab Mukhopadhyay[1], Sagar Sengupta[1,6]

[1] National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi, India.
[2] Université de Strasbourg, Inserm, UMR_S1113, FMTS, Strasbourg, France.
[3] Department of Surgical Oncology, All India Institute of Medical Sciences, New Delhi, India.
[4] Division of Molecular Medicine, St. John's Research Institute, Bengaluru, Karnataka, India.
[5] Regional Centre for Biotechnology, NCR Biotech Science Cluster, Faridabad, Haryana, India.

[6] Corresponding author: Sagar Sengupta, National Institute of Immunology
Aruna Asaf Ali Marg
New Delhi 110067
India
Phone: 91-11-26703786
Fax: 91-11-2616 2125
Email: sagar@nii.ac.in

Keywords: microRNA, BRCA1, ATM, RNF8, Chk1, BLM, DNA damage response, DNA repair, colon cancer

Abstract

Meta-analysis of transcripts in colon adenocarcinoma patient tissues led to the identification of a DNA damage responsive miR signature called DNA damage sensitive miRs (DDSMs). DDSMs were experimentally validated in the cancerous colon tissues obtained from an independent cohort of colon cancer patients and in multiple cellular systems with high levels of endogenous DNA damage. All the tested DDSMs were transcriptionally upregulated by a common intestine-specific transcription factor, CDX2. Reciprocally, DDSMs were repressed via the recruitment of HDAC1/2 containing complexes onto the CDX2 promoter. These miRs downregulated multiple key targets in the DNA damage response (DDR) pathway, namely BRCA1, ATM, Chk1 and RNF8. CDX2 directly regulated the DDSMs which led to increased tumor volume and metastasis in multiple preclinical models. In colon cancer patient

Journal of Cell Science • Accepted manuscript

tissues the DDSMs negatively correlated with BRCA1 levels, were associated with decreased probability of survival, and thereby could be used as a prognostic biomarker.

Introduction

MicroRNAs (miRs) are among the many important factors that control gene expression. The role of miRs on genome instability and ultimate progression to different types of cancers have been well studied. It is interesting to note that changes in the miR expression levels occur during the above processes, by generating a threshold in target gene expression (Mukherji et al., 2011). Microarray expression or small RNA sequencing data from a large number of different cancers show both increased or decreased miR levels (Chung et al., 2017; Croce, 2009), which results in transcriptional rewiring of the cognate genes. A significant number of miRs have been reported to be dysregulated in multiple types of cancers including colorectal carcinoma (CRC) (Chen et al., 2019; Ding et al., 2018).

To prevent the development of cancers, maintenance of genomic integrity is of utmost importance. Normal cells initiate DNA damage response (DDR) pathway, when exposed to multiple types of DNA damages like stalled replication, ionizing irradiation (IR), ultraviolet lights (UV) and genotoxic drugs (Ciccia and Elledge, 2010; Tikoo and Sengupta, 2010). One of the key factors which regulate genome integrity is BLM. BLM mutations lead to the Bloom Syndrome (BS) and are associated with DNA damage sensing, DNA repair, recombination (Sengupta et al., 2004; Tikoo et al., 2013; Tripathi et al., 2018) and cancer predisposition (Cunniff et al., 2017; Kaur et al., 2021). MRN and ATM are upstream DDR factors which recognize and accumulate at the double strand breaks within seconds after the IR exposure (Smith et al., 2010; Syed and Tainer, 2018). Multiple other factors take part in this choreographed process including E3 ligases like RNF8 (which facilitate recruitment of DNA damage response proteins) (Zhou et al., 2019), damage specific kinases like Chk1 and Chk2 (Smith et al., 2010) and BRCA1. Tumor suppressor BRCA1 maintains genome stability by being a part of multiple protein complex and is involved in DNA damage repair, DNA damage-induced cell cycle checkpoint activation, protein ubiquitination, transcriptional regulation and apoptosis (Savage and Harkin, 2015). Loss of any of the genome stabilizers (like BLM, ATM, BRCA1) leads to accumulation of DNA damage which progresses to neoplastic transformation, cancer predisposition and finally the development of cancer

Journal of Cell Science • Accepted manuscript

(Hanahan and Weinberg, 2011; Negrini et al., 2010). Links between miRs and BRCA1 have been documented in literature (Chang and Sharan, 2012; Petrovic et al., 2017). For example, miR-182 has been shown to downregulate BRCA1, impacting DNA repair and sensitivity to PARP inhibitors (Moskwa et al., 2011).

In this context, we identified via meta-analysis of the TCGA database the DNA damage sensitive miRs (DDSMs) which are upregulated across all stages of colon adenocarcinoma tissue samples. We validated DDSMs in an independent cohort of colon adenocarcinoma patient tissues and in multiple cellular models with high levels of endogenous DNA damage. All the DDSMs were regulated by a common transcription factor elevated in the colon, CDX2. In normal cells transcription of CDX2 was repressed due to BLM-dependent recruitment of HDAC1/2 containing Sin3b and NuRD complexes on its promoter. DDSMs target multiple key proteins involved in DNA damage response, DDR (BRCA1, ATM, Chk1 and RNF8). Specifically, DDSMs downregulate BRCA1 expression in cancer cell lines and colon adenocarcinoma patient tissue samples, thereby allowing neoplastic transformation. Hence this report serves as an integrated study where we have demonstrated how a patient-derived colon cancer specific miR signature is epigenetically silenced in normal tissues, becomes deregulated during DNA damage, thereby repressing DNA damage response factors and subsequently worsening the pathological consequence.

Results

Identification of a DNA damage dependent miR signature

To identify the miRs which were upregulated in all stages of colon adenocarcinoma, we carried out an unbiased meta-analysis of the TCGA database. We found that of the total 1323 miRs for whom the transcript expression values were available, only 647 miRs were found to be expressed in at least 50% of samples. Differential expression analysis indicated that 382 miRs to be down-regulated while 138 miRs were found to be up-regulated as compared to the normal samples. Of the 138 miRs which were upregulated, the transcript levels of 62 miRs were increased all across the four different stages of colon cancer, which was depicted in the form of a heat-map (Figure S1). All the predicted targets for these 62 miRs were then analyzed by miRanda, miRTarBase and Target Scan. The common targets were then analyzed by four pathway analysis softwares (Reactome, KEGG, Wiki Pathways and GSEA).

Journal of Cell Science • Accepted manuscript

Ten upregulated miRs were found to be enriched in pathways having association with colorectal cancer and having direct implications in DNA damage response, namely DNA repair, apoptosis, DNA replication, chromatin remodelling. Recognizing the fact that intrinsic DNA damage is enhanced within the cancerous lesions, we named these miRs as DNA damage sensitive miRs (DDSMs) (Figure 1A). We randomly chose six of the DDSMs (miR-29a-5p, miR-29b-3p, miR-96-5p, miR-182-5p, miR-183-5p, miR-335-3p) for further studies. Among the 62 miRs which were upregulated across four stages of cancers, we also chose three miRs (miR-130a-5p, miR-148a-5p, miR-429), which did not fit the above criteria for DDSMs. These three miRs were called non-DDSMs.

The six chosen DDSMs were significantly upregulated in the tissues of the colon cancer patients in the TCGA database across all the four stages of cancer progression (Figure 1B). Next, we expanded our studies to 54 paired tissue samples obtained from an Indian cohort. These six DDSMs were also found to be upregulated in the patient cancerous tissue samples compared to their matched adjacent normal tissues across stages of cancer progression (Figure 1C). Kaplan-Meier analyses indicated lesser overall survival for patients with higher risk score (i.e. patients which show higher levels of the combined expression of the six miRs) in the tissue samples (Figure 1D).

To further investigate the functions of DDSMs, we chose cells from BS patients which show high levels of endogenous DNA damage (Kharat et al., 2016), hyper-recombination, genome instability and increased sister chromatid exchanges (Cunniff et al., 2017). BS patients are predisposed to multiple types of cancers. The most common type of solid tumour in BS patients is colorectal cancer (Cunniff et al., 2017). Further when BLM is heterozygous, enhanced risk of colon cancer had been shown to occur in both mice and human (Goss et al., 2002; Gruber et al., 2002). Hence, we carried out small RNA sequencing with RNA isolated from two pairs of isogenic cell lines created from two different BS patients (see materials and method section for details about isogenic lines). Small RNA sequencing indicated that the DDSMs were upregulated in cells lacking BLM (Figure 2A, S2A). The higher expression of the six identified DDSMs were validated by RT-qPCR in BLM-deficient cells compared to BLM-restored cells in both the above isogenic pairs (Figure 2B, S2B). We further demonstrated the increased expression of the DDSMs in BLM-knockout HCT116 cells compared to parental cells (Figure 2C), as well as in siRNA-mediated BLM knockdown SW480 cells compared to control cells (Figure S2C).

Journal of Cell Science • Accepted manuscript

Next, we wanted to determine if the matured DDSMs were also enriched in the RNA-induced silencing complex (RISC) under the same conditions. HeLa S3 cells stably integrated with either vector pREV or Flag Ago2 were used for the experiment. BLM was ablated in both the cell lines using BLM siRNA (Figure 2D). In either absence or presence of BLM, immunoprecipitations were carried out with either anti-GFP (which acted as an antibody control) or with anti-Flag antibody. RNA bound to the immunoprecipitated Ago2 complex was isolated followed by RT-qPCR which showed that the five tested DDSMs were associated with Ago2 protein (i.e. the RISC complex) (Figure 2E).

Absence of BLM leads to persistence of damaged DNA, which culminates in hyper-recombination (Croteau et al., 2014; Tikoo and Sengupta, 2010). Hence, we hypothesized that in a self-perpetuating loop, the DDSMs maybe responding to the increased levels of cellular DNA damage. To test this hypothesis, BLM isogenic cells were exposed to a range of IR doses (leading to increased DSB generation). While the levels of the DDSMs increased over the range of IR dosage (Figure 2F), the three non-DDSMs did not respond to the IR gradient (Figure S2D).

Overexpression of DDSMs cause neoplastic transformation

Since the levels of DDSMs depend on the extent of DNA damage, we next wanted to determine whether the DDSMs can potentiate neoplastic transformation in both *in vitro* and *in vivo* models. We found that the ablation of DDSMs by miR inhibitors in the BLM-deficient GM03509 GFP100 cells caused a decrease in the extent of residual DNA damage as determined by neutral Comet assays (Figure 3A) and spontaneous level of SCEs (Figure 3B). Reciprocally, over-expression of the same DDSMs in the BLM-restored GM03509 GFP-BLM Clone 4.3.4 cells led to an enhancement in SCEs (Figure 3C). Subsequently, inhibition of two DDSMs (miR-29a-5p and miR-96-5p) in HCT116 BLM KO cells culminated in decrease in the invasive potential of these cells, as determined using both matrigel (Figure 3D) and soft agar (Figure 3E) assays.

Journal of Cell Science • Accepted manuscript

To determine whether modulation of the levels of DDSMs affected the ability to initiate and propagate tumors, we generated three stable lines in HCT116 BLM -/- cells which expressed either a control inhibitor or specific inhibitors of miR-29a-5p or miR-96-5p. Each of the stable lines, which also constitutively expressed EGFP (Figure S3A), were validated for the lowering of the levels of miR-29a-5p or miR-96-5p (Figure S3B). Xenograft assays were carried out in NOD SCID mice by subcutaneously injecting these cells in the animals and monitoring for tumor development. Results indicated that inhibition of the two DDSMs decreased the rate of tumor growth (Figure 3F). As a second *in vivo* assay, 100 cubic mm tumors were subcutaneously generated using HCT116 BLM -/- cells. A nanoparticle-mediated delivery system was used to deliver miR inhibitor control, miR inhibitor 29a-5p or miR inhibitor 96-5p directly to the base of the tumors every third day. Tumor formation was decreased when miR inhibitor 29a-5p or miR inhibitor 96-5p were injected (Figure 3I). Analysis of the levels of miR-29a-5p and miR-96-5p in these tumors (excised at the end point in both the experiments) confirmed the decrease in the levels of the two tested DDSMs (Figure 3G, 3J). Thus, the DDSMs are miRs which sustain tumor growth.

CDX2 regulates DNA damage dependent miRs

Next, we wanted to determine how the DDSMs were regulated in the cellular milieu. Using an *in-silico* approach we first analyzed the upstream 5 kb promoter of each of these miRs. We found that only one transcription factor, CDX2, was common among the DDSMs (Table S1-S5). We hypothesized that CDX2 (known to be highly expressed in colon) may control the expression of the DDSMs. CDX2 expression was found to slightly increase 1 hour after 3Gy IR exposure in HCT116 cells. This effect was significantly enhanced by the loss of BLM in HCT116 BLM KO cells. However, by 6 hours after IR exposure (when certain amount of DNA repair has occurred) the levels of CDX2 began to decrease (Figure 4A). EMSA demonstrated that CDX2 bound to the promoter sequence of the tested DDSM, miR-96-182-183. The binding was specific as a super-shift was observed when an anti-CDX2 antibody was used in the EMSA reactions (Figure 4B). The binding of CDX2 to the miR promoter was lost in the presence of an excess of unlabeled competitor (Figure 4C) or when a mutant oligo in which CDX2 binding site was destroyed was used as the substrate (Figure S3D). Using immunoprecipitated (IPed) CDX2, it was shown that the binding of CDX2 to the miR promoter was enhanced 1-hour post-IR (Figure 4D). The ability of CDX2 to transactivate the DDSMs in a dose-dependent manner was demonstrated for the full-length protein but not for

Journal of Cell Science • Accepted manuscript

the mini CDX2 splicing variant lacking the transactivation domain (Figure 4E). Mutating the CDX2 binding site on the miR-96-182-183 promoter prevented its transactivation by CDX2 (Figure 4F). Finally, ablation of CDX2 by siRNA (Figure 4G), decreased the levels of primary, precursor and mature miRs within the cells (Figure 4H, S3E).

We next hypothesized that modulating the levels of CDX2 should also lead to changes in the DDSM levels and thereby control in the tumor formation process. First, we used HT29-derived colon cancer cells with doxycycline (Dox) inducible expression of CDX2 + GFP (TW6 cell line) or GFP (TG8 cell line) (Figure 5A). Induction of CDX2 by Dox treatment in TW6 cells (Figure 5A) led to an increase in the levels of all the six DDSMs (Figure 5B), which was concomitant with enhanced wound healing (Figure 5C) and colony formation (Figure 5D) efficiencies. The stimulatory effect of CDX2 was also tested in another pair of cell lines based on HCT116 cells – namely HW2 (constitutively expressing Flag CDX2) and HC1 (the corresponding vector control) (Figure S4A). Overexpression of Flag CDX2 in HW2 cells led to the upregulation of the DDSMs (Figure S4B). Next, *in vivo* experiments were carried out using xenograft models where TG8 and TW6 cells were subcutaneously injected into NOD SCID mice, a subset of which were fed orally with Dox. Tumors obtained from TW6 cells expressing inducible CDX2 had the maximum tumor volume (Figure 5E). Tumors were excised at the end point of the above experiment. Tumors obtained from TW6 cells showed increased levels of all the DDSMs (Figure 5F), increased levels of the proliferation marker PCNA (Figure 5G) and angiogenesis marker CD31 (Figure 5H). We also carried out *in vivo* subcutaneous xenograft studies using the HC1/HW2 cells. The sizes of tumors obtained from HW2 cells were consistently larger (Figure S4C), they expressed higher levels of the DDSMs (Figure S4D), PCNA (Figure S4E) and CD31 (Figure S4F). Finally, we wanted to determine whether expression of the DDSMs caused increased dissemination of the cancer cells. Using the TG8/TW6 lines, we carried out an orthotopic model where the cells were implanted into the cecal wall of the mice (Figure 5I). All the mice were orally fed with Dox. At 21 days post-initiation of the experiments the mice were subjected to whole body imaging and GFP fluorescence tracked. TW6 cells expressing CDX2 showed enhanced *in vivo* dissemination to distal organs (Figure 5I).

Journal of Cell Science • Accepted manuscript

DDSMs target DDR proteins

Having established how the DDSMs are regulated in the colon, we wanted to decipher how these miRs function and sought to determine their putative targets. Using MiRanda, we found 4840 common targets for the six DDSMs (Figure S5A, Table S6, S7). Using Reactome, KEGG, Wiki Pathways and GSEA, we determined that the DDSMs regulated many vital cellular processes including DNA repair, DNA replication and gene expression (Figure S5B). From the list of common targets of the DDSMs (Table S7) we chose four known genes involved in the DDR pathways, namely BRCA1, ATM, Chk1, RNF8, for further in-depth studies. Decrease of the high levels of endogenous DDSMs by treatment with specific miR inhibitors led to increase in the transcript levels of BRCA1 (Figure 6A, S5C), ATM (Figure 6B), Chk1 (Figure 6C), RNF8 (Figure 6D) in GM03509 GFP Clone 100 (Figure 6A-6D) and HCT116 BLM -/- (Figure S5C) cells. Reciprocally, enhancement of the levels of DDSMs by treatment with specific mimics led to the decrease in the transcript levels of BRCA1 (Figure 6E, S5D), ATM (Figure 6F), Chk1 (Figure 6G), RNF8 (Figure 6H) in both GM03509 GFP-BLM Clone 4.3.4 (Figure 6E-6H) and HCT116 WT (Figure S5D) cells. The effect of the miR inhibitors and mimics at the RNA level was phenocopied at the protein levels in both the cell types tested (Figure 6I, 6J, S5E, S5F).

We chose BRCA1 for mechanistic and clinical studies as meta-analyses have shown that mutations of this gene conferred increased risk to colon cancer (Oh et al., 2018). The RNA and the protein levels of BRCA1 were both enhanced when the miRs were downregulated in diverse experimental systems - including transient transfection of miR inhibitors in multiple cell types (Figure 6A, 6I, S5C, S5E), stable lines expressing DDSM inhibitors (Figure S3B, S3C) and in the tumors obtained in two xenograft models where expression of DDSMs are inhibited (Figure 3H, 3K). Reciprocally, we also observed that the levels of BRCA1 was downregulated when the DDSM levels were enhanced in multiple experimental systems – cells exposed to DDSM mimics (Figure 6E, 6J, S5D, S5F), in tumors where the DDSM levels were increased by inducible or constitutive enhancement of CDX2 levels (Figure 5G, S4E) and when cells are exposed to increasing dosage of IR (Figure S5G). Analysis of the 3' UTR of BRCA1 revealed binding sites of four of the DDSMs (Table S8). We used a luciferase reporter assay using either wildtype (WT) or mutant (MT) sequence of the BRCA1 3'UTR which binds to miR-183. Overexpression of the DDSMs led to the reduction of luciferase activity when WT BRCA1 3'UTR but not MT BRCA1 3'UTR was used (Figure 6K). We

Journal of Cell Science • Accepted manuscript

hypothesized that if DDSMs act through BRCA1, increasing or decreasing BRCA1 levels should phenocopy the effects seen by modulating the levels of the miRs in cellular invasion assays (shown in Figure 3D, 3E). Indeed, ablation of BRCA1 by siRNA in three different cell types [verified at both transcript and protein levels (Figure S6A, S6B)], led to an increase in the levels of endogenous DNA damage (Figure S6C), SCEs (Figure S6D) and invasion (Figure S6E). Conversely, overexpression of BRCA1 (Figure S6F), decreased the DNA damage levels (Figure S6G), SCEs (Figure S6H) and invasion (Figure S6I).

BLM represses CDX2 expression

We next wanted to determine the regulatory circuit which regulates DDSM expression. We hypothesized that BLM itself may negatively regulate CDX2 expression as the levels of CDX2 were elevated in HCT116 BLM -/- cells (Figure 7A, S7A). Further, overexpression of BLM reduced the transcript level of CDX2 (Figure 7B). We next wanted to determine whether BLM is recruited to the promoter of CDX2. On analysis of the 5kb upstream to the CDX2 promoter TSS, we identified potential binding sites for several transcription factors - MAD, AP3β, SMAD3, AP2β, E2F1 (Figure S7B). Using ChIP, we found that BLM was specifically recruited to all these putative binding sites except to the E2F1 site (Figure 7C). This recruitment of BLM was independent of its helicase activity as it occurred to similar extent even in presence of ML216, a specific BLM helicase inhibitor (Nguyen et al., 2013) (Figure S7C). Two transcriptional factors, SMAD3 and AP2β, were themselves recruited to their cognate binding sites in a BLM dependent manner (Figure 7D, 7E).

Transcriptional repression is controlled by two major protein complexes – NuRD and Sin3. These two repressor complexes have specific subunits and also share common subunits (Baymaz et al., 2015). LC MS/MS analysis of immunoprecipitated BLM from the extracts of GM03509 GFP BLM Clone 4.3.4 cells indicated its interaction with both Sin3b and CHD4, the core ATPase subunits of the two complexes. The peptide sequences found associated with BLM immunoprecipitates were SQSIDTPGVIR (for Sin3b) and APEPTPQQVAQQQ (for CHD4) (Figure S7D, S7E). BLM or Sin3b immunoprecipitations further revealed that BLM interacted with Sin3b, CHD4, HDAC1 and SMAD3 (Figure S8A-S8E). Direct interaction was also observed between GST-tagged BLM and *in vitro* transcribed and translated SMAD3 as well as GST-tagged BLM and *in vitro* transcribed and translated HDAC1 (Figure S8F).

Journal of Cell Science • Accepted manuscript

Next, we wanted to determine whether BLM is co-recruited with members of the NuRD and Sin3b complexes onto the CDX2 promoter. Using ChIP, we found that the core ATPase subunits of the two co-repressor complexes, Sin3b and CHD4, are recruited to different binding sites on the CDX2 promoter. While Sin3b was recruited to the two SMAD3 binding sites (Figure 7F), CHD4 is recruited to one of the two AP2β sites (Figure 7G). The extent of recruitment of both Sin3b and CHD4 was always enhanced in cells which express BLM. HDAC1 and HDAC2 are the two common factors present in both NuRD and Sin3b complexes. We found that while HDAC1 was recruited exclusively to SMAD3 binding sites (Figure 7H), HDAC2 was recruited to the regions where MAD, AP3, SMAD3 and AP2β binding sites were present (Figure 7I) in a BLM dependent manner. Sequential re-ChIP experiments were carried out which validated that BLM-Sin3b and BLM-CHD4 were binding to specific DNA recognition sequences on the CDX2 promoter (Figure S8G, S8H). Based on these results we hypothesized that BLM repressed CDX2 expression using both Sin3b and NuRD repressor complexes. To obtain direct validation, we carried out ablation experiments of the two co-repressor complexes. Hence depletion of Sin3b (Figure 7J, 7K), CHD4 (Figure 7L, 7M), HDAC1 (Figure 7N, 7O) and HDAC2 (Figure 7P) enhanced the expression of CDX2 at both transcript and protein levels.

Further, we wanted to address the physiological significance of BLM recruitment to the CDX2 promoter. We focused on ten paired samples from the Indian colon cancer patients. ChIP experiments revealed that the recruitment of BLM to the CDX2 promoter was lower in the cancerous regions compared to the adjacent normal regions (Figure 8A). This indicates that in colon cancer patients increasing levels of CDX2 expression correlated with the absence of BLM from the promoter of this homeobox gene.

DDSMs target BRCA1 in colon cancer patients

Next, we wanted to determine whether the increased levels of the DDSMs in the colon cancer patients corelated with the changes in the expression levels of their common target, BRCA1. We found that the transcript levels of BRCA1 decreased in the colon cancer patient tissue samples (for both stage I + II and stage III + IV) compared to their respective matched normal tissue controls (Figure 8B). Analyzing only the cancerous tissues, we found that the levels of BRCA1 was significantly decreased in Stage III + compared to their Stage I + II counterparts. Using Western analysis (Figure 8C, 8D) and immunohistochemistry (Figure 8E,

Journal of Cell Science • Accepted manuscript

8F), decreased BRCA1 protein levels was observed in cancerous sections (across all stages) compared to their adjacent normal control. Finally using the TCGA dataset we observed that in the cancerous colon tissues, a negative correlation exists between the overexpression of the six DDSMs and the reduction of BRCA1 transcripts (Figure 8G), thereby revealing the pathophysiological significance of the existence of these miRs.

Discussion

Several miRs signatures have been proposed to be predictive of the evolution of CRC. These include (a) a five miR signature identified through bioinformatics and subsequently validated in the tissues from two cohorts of patients (Ozawa et al., 2018); (b) an eight miR signature identified by three independent miR profile analyses which predicts recurrence of tumors in stages II and III patients (Kandimalla et al., 2018); (c) a four miR signature related to the relapse after curative surgery (Grassi et al., 2018); (d) a three miR signature which predicts both distant metastasis and hepatic recurrence (Coebergh van den Braak et al., 2018) and (e) a 16-miR signature which serves as a prognostic biomarker for Stage II and III patients (Jacob et al., 2017). Distilling the results from these studies indicated that there was hardly any overlap in these lists of upregulated miRs in the different cohorts. Apart from miR-182, none of the DDSMs was present in any of the above studies. Such diverse results are possibly due to the fact that multiple miRs can regulate the same pathway.

This study identified the CDX2 transcription factor as a common regulator of the DDSMs in colon cancer cells. A lot of studies have documented an altered pattern of CDX2 in CRC resulting in a decreased or heterogenous expression, which correlated with compromised cell differentiation and adverse prognosis (Baba et al., 2009; Balbinot et al., 2018; Brabletz et al., 2004; Dalerba et al., 2016; Pilati et al., 2017). Several mouse models have supported the idea of a tumor suppressor activity played by this homeobox gene since its reduction or loss of function facilitates the development of genetically- and chemically-induced intestinal cancers (Aoki et al., 2003; Balbinot et al., 2018; Bonhomme et al., 2003; Hryniuk et al., 2014; Sakamoto et al., 2017). This notion was further reinforced in cellular models by the inverse correlation between CDX2 levels and cell aggressiveness (Gross et al., 2008; Mallo et al., 1998; Yu et al., 2019; Zheng et al., 2017). However, increased CDX2 has also been reported in CRC (Witek et al., 2005), especially associated with short amplifications overlapping the

Journal of Cell Science • Accepted manuscript

gene locus and conferring a context-dependent oncogenic activity to this homeobox gene (Salari et al., 2012). The context-dependent effect is further illustrated in cases of abnormal ectopic expression outside the gut where CDX2 drives the formation of precancerous intestinal-type metaplasia mainly in foregut-derived organs [reviewed in (Chawengsaksophak, 2019)] and also promotes leukemogenesis in the hematopoietic lineage (Galland et al., 2021; Vu et al., 2020). Interestingly, we showed here that increasing the CDX2 levels in colon cancer cells embedded in a growth factor- and matrix-rich environment stimulates tumor growth when grafted in NOD/SCID mice impaired for both T and B cells, unlike the less immunodeficient nude mice having only a reduced number of T cells (Gross et al., 2008). However, the possibility exists that the effect of CDX2 expression may also regulate other factors which can also affect the process of tumour growth sustenance.

Although we show here that CDX2 stimulates DDSMs in colon cancer cells, the question remains of why these miRs are not expressed in the normal intestinal epithelium? This could reflect the context-dependent molecular functions of CDX2, since besides its inductive transcriptional activity, it also has permissive effects by maintaining chromatin regions open and accessible to other sets of transcription factors that could be differentially expressed in the nucleus of malignant vs healthy cells (Saxena et al., 2017; Verzi et al., 2013). CDX2 being a bivalent gene, it is not surprising to find BLM being recruited with its promoter together with HDAC1 and HDAC2 containing Sin3b and NuRD repressor complexes in the normal colonic mucosa and also in colon cancer cells like HCT116. The present study conducted on the CDX2 promoter in HCT116 colon cancer cells suggests that BLM can also play the role of an adaptor protein in unstressed cells, and that DNA damage leads to BLM redistribution to the site of damage and to its release from the CDX2 locus. Interestingly an adaptor function has already been attributed to BLM whereby it helps in the interaction of c-Jun and c-Myc oncoproteins to their common E3 ligase, Fbw7α, and facilitates the degradation of the substrates (Chandra et al., 2013; Priyadarshini et al., 2018).

In an effort to understand how DDSMs carry out their biological functions in the cells, we have identified and validated common targets with the aim of narrowing down our search for targets with known activities in DNA damage recognition, signaling and repair (i.e. the classical DDR response). While four targets, BRCA1, ATM, Chk1 and RNF8 were initially

Journal of Cell Science • Accepted manuscript

validated, further functional and clinical studies focused on BRCA1 showed consistent effects on DNA repair, DNA damage and invasion using gain and loss of function studies and cell-based assays. It is interesting to note that BRCA1 and ATM are targeted by multiple miRs. For example, BRCA1 is downregulated by miR-182 (which is also one of the DDSMs) (Moskwa et al., 2011), miR-1255b, miR-148b*, miR-193b* (Choi et al., 2014), miR-146a, miR-146b-5p (Garcia et al., 2011) and miR-498 (Matamala et al., 2016). Similarly ATM is targeted by miR-421 (Hu et al., 2010), miR-101 (Yan et al., 2010) and miR-203 (Zhou et al., 2014). As demonstrated previously (Moskwa et al., 2011), we also found that the miR-182 mediated suppression of BRCA1 impeded DNA repair which led to greater amount of residual DNA damage in miR-182 expressing cells compared to the control cells. Further analogous to the results in breast tumor lines (Moskwa et al., 2011), we also found that BRCA1 is targeted by miR-182 in colon tumor cells and patient tissues. The different miRs reported in the literature for ATM and BRCA1 could reflect specific types of DNA damage and/or different regulatory factors involved in their inductive mechanism(s). We also believe that the need for several miRs within the DDSMs signature illustrates the relevance for appropriately controlling such an important process as DDR. The DDSMs bind to neighboring yet discrete sites on targets like BRCA1 and possibly carry out cooperative repression *in vivo* (Broderick et al., 2011; Grimson et al., 2007; Saetrom et al., 2007). This also might be the reason why the DDSMs probably have neighboring non-canonical binding sites in the 3'UTR of the target genes like BRCA1.

miRNA–miRNA interaction has been implicated for potential mutual regulatory pattern (Guo et al., 2012). miRNA-miRNA synergistic networks via co-regulatory functional modules have also been implicated in pathological conditions (Xu et al., 2011) and specifically in breast cancer (Cilek et al., 2017). We believe that similar inter-regulatory network also possibly exists for the DDSMs. Hence when one of the DDSMs was inhibited, the phenotypes associated with neoplastic transformation and cancer progression are cancelled out to a large extent.

Journal of Cell Science • Accepted manuscript

This study provides evidence that the DDSMs are biologically active moieties present in colon cancer patient tissue samples. Their overexpression in colon cancer patients corelated with decreased BRCA1 expression. Compared to matched normal tissues, BRCA1 is decreased in all stages of the cancerous tissues and also negatively corelates with the cancer progression. The fact that high DDSM expression is associated with decreased probability of survival, adds to the mechanistic reasons why the decrease in BRCA1 and ATM levels in colon cancer tissues correlates with reduced overall survival of the patients (Bai et al., 2004; Grabsch et al., 2006; Wang et al., 2018). If DDSMs respond to the extent of DNA damage and have a role in tumor growth sustenance, it may be possible to hinder tumor progression by inhibiting the miRs. Indeed, we showed here using two different experimental strategies that DDSMs inhibition led to the regression of tumors. Similar approaches to target tumors using either RNA aptamers or chemical ligands have also been reported (Shu et al., 2014). Other complementary approaches, like the usage of miRNA sponges, miRNA masking, antisense oligonucleotides or small molecule inhibitors (Hernandez et al., 2018), could also be alternative methods to reach the same objective. The low levels of BRCA1 and ATM in cells having overexpressed DDSMs indicates defective DNA damage response and homologous recombination pathways. Hence, we suggest that colon cancer patients with increased DDSMs in tissues could be treated with PARP inhibitors like olaparib and veliparib, as already proposed in certain studies (Clark et al., 2012; Davidson et al., 2013; Wang et al., 2017).

In conclusion, the identification of DDSMs has importance as it presents an entire workflow for a colon cancer specific miR signature, whereby the identity of a common upstream regulator (CDX2) and the downstream effectors (DDR proteins like BRCA1, ATM, Chk1 and RNF8) have been elucidated and validated in both preclinical mice models and patient samples (Figure 8H). Further, this study opens up the possibility of the usage of the DDSMs as potential prognostic biomarkers of colon cancer and as targets for therapeutic miR inhibition, all attractive avenues of future research.

Journal of Cell Science • Accepted manuscript

Materials and Methods

Antibodies, Plasmids, siRNAs

The antibodies used in the study have been listed in Table S9, the recombinants listed in Table S10 and the reagents used are in Table S11. Information about the mammalian expression vectors of CDX2 used in this study is in Table S10 and has been described in (Balbinot et al., 2017). pGEX4T-1 CDX2 was obtained by cloning full-length CDX2 into BamH1 and EcoR1 sites of pGEX4T-1. The DDSM promoter sequences with CDX2 binding sites are depicted in Table S2-S5. The sequence containing the CDX2 binding site in the miR-96/182/183-5p promoter was cloned in pGL3 vector using the KpnI and HindIII to generate the pGL3-miR-96/182/183-5p WT promoter luc recombinant. BRCA1 coding and 3'UTR sequence with four identified DDSM binding sites is depicted in Table S8. The BRCA1 3'UTR sequences containing the miR-183 binding sequences were inserted into the KpnI and HindIII sites of the pGL3 vector to generate pGL3-BRCA1 3'UTR miR-183 WT. All site-directed mutagenesis was carried out using QuikChange II XL Site-Directed Mutagenesis Kit. The siRNA and shRNA sequences used were: CDX2 (AAC CAG GAC GAA AGA CAA AUA), BLM (AGC AGC GAU GUG AUU UGC A), BRCA1 (UCA CAG UGU CCU UUA UGU A), CHD4 (CCC AGA AGA GGA UUU GUC A), HDAC1 (GGC UCC UAA AGU AAC AUC AUU), HDAC2 (CCA CCA UGC UUU AUG UGA UUU) and Sin3b (AGG CUG UAG ACA UCG UCC A). Information for miR mimics and inhibitors (used for DDSM overexpression and shutdown) are in Table S11.

TCGA analysis

miRNA expression levels of 332 colon tumor samples and 8 normal samples were downloaded from the Genomic Data Commons (GDC) data portal (TCGA COAD) using the miR quantitation file and files with extension FPKM.txt.gz, respectively on 27th April, 2017. The clinical information for all these samples were also downloaded from the GDC data portal. For the meta-analysis miRNAs that were lowly expressed were first filtered out, followed by removing the entries that were not expressed in at least half of the samples. DESeq2 (v.1.29.8) from R package was used to normalize the count values and then differentially expressed miRNAs compared to the normal samples were identified after performing Wald statistical test. The miRNAs with cut-off of log2fold-change (FC) > 1.5 and

Journal of Cell Science • Accepted manuscript

adjusted $p < 0.05$ were considered significant. The heatmap of the 62 up-regulated miRNAs were plotted using R package gplots clustered based on group average clustering method. Further, we identified miRs upregulated across all stages of cancer as compared to normal samples. One-way Anova was used to identify the significantly upregulated miRNAs consistently across the tumour samples.

For individual miR analysis, the $\log_2(x+1)$ transformation was carried out on the miR levels and gene expression values. The expression of these miRs in different stages of the colon cancer as compared to the normal samples was then examined by comparing the means of the expression values across different stages. Kruskal-Wallis test was carried out in SPSS v.24 to compare the means of the miR expression across stages with that in the normal samples. The number of patients (n) in each classification is as follows: Normal: 8, Stage I: 67, Stage II: 166, Stage III: 119, Stage IV: 61.

For Kaplan–Meier analysis, the expression value in each tissue sample was subtracted from the average value of normal samples. These values were then used to calculate the risk score of the 6 miRs significantly found upregulated in TCGA samples as described earlier (Ji et al., 2018). Briefly, regression coefficients (β) of the individual miRNA were determined by Cox regression analysis. The risk score was calculated for each patient using the formula: ($\beta_{miR29a-5p*}$ expression value of miR-29a-5p) + ($\beta_{miR29a-3p*}$ expression value of miR-29b-3p) + ($\beta_{miR96-5p*}$ expression value of miR-96-5p) + ($\beta_{miR182-5p*}$ expression value of miR-182-5p) + ($\beta_{miR183-5p*}$ expression value of miR-183-5p) + ($\beta_{miR335-5p*}$ expression value of miR-335-3p). The tissue samples with risk score more than or equal to 75% quartile or 0.9618 were grouped as high-risk samples (n=78) while the patients with risk score less than or equal to 25% quartile or -0.3654 were grouped as low-risk samples (n=76). The overall survival (OS) curve was plotted using Kaplan–Meier analysis in SPSS followed by log–rank test to detect the significant difference between the high and low risk group of patients. Spearman correlation was carried out with 207 patient samples for which both the transcriptome and miR expression datasets were available.

Journal of Cell Science • Accepted manuscript

Indian Cohort

All samples (detailed in Table S13) pertaining to the Indian cohort were obtained from All India Institute of Medical Science (according to Institute Human Ethics approval number RP-23/2017) after obtaining informed consent from all subjects. All experimental work on these samples was carried out in National Institute of Immunology (according to Institute Human Ethics approval number IHEC#92/17 and Institutional Bio-Safety Committee approval number IBSC/VKN/2014/63). All samples obtained were paired, the adjacent normal tissues been excised 7-10 cm from the periphery of the tumor. Each tissue sample was graded, subjected to routine histology and H and E staining, based on which the core of the tumour was used for RNA/protein/IHC analysis. Patients in polyp, stages I and II were combined together while stages III and IV were combined together. The samples were acquired in two phases – Phase 1 (n=40) and Phase 2 (n=14). In Phase 1 out of the 40 patients, 26 patients were in Stage I + II, while 14 patients were in Stage III + IV. In Phase 2 out of the 14 patients, 5 patients were in Stage I + II, while 9 patients were in Stage III + IV. Significantly altered expression of miRs, BRCA1 mRNA and BRCA1 protein (estimated by both western and IHC) were identified in the tissues of the Indian cohort.

Animal studies

All animal studies were carried out in National Institute of Immunology according to an approved animal ethics protocol (IAEC approval reference number: IAEC#398/15). Following animal studies were carried out: (A) Tumorigenic potential of HCT116 BLM -/- cells expressing miR inhibitors using a subcutaneous model; (B) Tumorigenic potential of HCT116 BLM -/- cells which were challenged with a nanoparticle coated miR inhibitors in a subcutaneous model; (C) Tumorigenic potential of HC1/HW2 cells using a subcutaneous model; (D) Tumorigenic potential of TG8/TW6 cells using a subcutaneous model; (E) Tumorigenic potential of TG8/TW6 cells using an orthotopic model. In all subcutaneous models (A-D) approximately 2 million cells were resuspended in Fetal Bovine Serum. These cells were injected subcutaneously in healthy 8-week-old male or female NOD SCID mice along with matrigel. In the TG8/TW6 models (D, E) the mice were additionally treated with doxycycline (10mg/ml/kg body weight) by oral gavage every day throughout the experiments. In all the subcutaneous models (A-D), tumor formation started after 7-10 days. In case of (B) a nanoparticle mediated delivery system was used to deliver miR inhibitor

Journal of Cell Science • Accepted manuscript

Control, miR inhibitor 29a-5p or miR inhibitor 96-5p directly to the base of the tumors every third day for 4 times (further details of the delivery system are below in a separate section). In case of orthotopic model (E), 50,000 TG8/TW6 cells were used and were implanted into the cecal wall. In (E), at the end point whole body imaging was done for the mice using an *in vivo* imaging system (Perkin Elmer) to check the expression of GFP and thereby determine the invasive potential of TG8/TW6 cells. For all the above models, at the end point the mice were sacrificed by cervical dislocation. The excised tumors (in case of subcutaneous models, A-D) were imaged, measured, used for lysate preparation (using RIPA), RNA extraction (using Trizol LS) and immunohistochemistry (IHC). The staining in IHC is presented as H-score (Rajarajan et al., 2020).

Cells

All pre-existing cell lines (Table S11) were maintained as described in the original publications or as recommended by the suppliers. Immortalized cell lines from BS patient, GM03509, were complemented with either GFP-BLM (Clone 4.3.4) or with GFP (Clone 100) (Kharat et al., 2016). Immortalized cell lines from another BS patient GM08505 were also used (Hu et al., 2001). Cells from the BS patients were derived from the respective patients' skin fibroblasts. HC1, HW2 cells were generated by stably transfecting HCT116 cells with pCB6 (for HC1) or pCB6-Flag$_2$-mCDX2 WT (for HW2) and selecting with G418. The cells were grown in DMEM + 10% FBS + G418 (1 mg/ml) + antibiotics. To generate the HCT116 BLM -/- Inhi Control cells, pLenti-III-mir-Off Control Vector (Abm Inc.) was used. The lentivirus was generated by using Lenti-X HT packaging mix in Lenti-X 293T. To obtain HCT116 BLM -/- Inhi-29a-5p and HCT116 BLM -/- Inhi-96-5p lines, commercial lentivirus particles were used (Abm Inc.). BLM -/- cells were plated in six well cluster and transduced with the three different lentiviral particles. Transduction was carried out with 2ug/ml polybrene. Medium was changed 24 hours post-transduction. For selection, 1μg/ml puromycin was added to the cells. The transduced cells were grown in presence of 1μg/ml puromycin for stable line generation for 7 days after which the clones were analysed for the expression of the two miRs. ML216 (12.5μM) treatment was carried out for 24 hrs prior to making the chromatin for the ChIP assays. For DNA damage-dependent experiments cells were exposed to a particular IR (3Gy) or a range of IR indicated in the specific experiments.

Journal of Cell Science • Accepted manuscript

Lysates were made or RNA extracted after 1 hour or 6 hours post-IR exposure. All cells used tested negative for mycoplasma contamination.

RNA immunoprecipitation

RNA immunoprecipitation was carried out in HeLa pREV and Ago2 cells according to published protocols (Keene et al., 2006). Cells were scraped in PBS and resuspended in polysome lysis buffer [100mM KCl, 5mM $MgCl_2$, 10mM HEPES (pH 7.0), 0.5% NP-40, 1mM DTT, 100units/ml RNase Out, 400µM VRC, Protease inhibitor cocktail supplemented with RNase inhibitor and protease inhibitors]. The lysates (2mg) were used to set up RNA immunoprecipitations with either Protein G bound GFP antibody (2µg/IP) or with anti-Flag beads (4µl/IP). The immunoprecipitations were for 4 hours on an end to end rotor at 4^0 C, after which beads were pelleted at 1200rpm for 5 minutes and washed with ice cold NT2 buffer 3-4 times. Beads were resuspended in 100µl of NT2 buffer [50mm Tris-HCl (pH 7.4), 150mM NaCl, 1mM $MgCl_2$, 0.05% NP-40] supplemented with 30µg of Proteinase K to release the RNP components. This mixture was incubated at 55^0 C for 30 minutes. Thereafter, 200µl of Trizol was added directly to the beads and RNA was isolated. It was followed by cDNA synthesis and qPCR to determine the enrichment of miRNA bound with Argonaute 2 protein. The main steps of traditional immunoprecipitations were similar to the above, except 1XPBS + 0.1% NP-40 was used as the buffer to make up the volume and carry out washes. Post-IP, the bound proteins were run in SDS-PAGE gels to determine their co-immunoprecipitating partners.

Overexpression and ablation studies

Cells were transfected with 20nM of either miRNA inhibitors or miRNA mimics specific to the respective miRs. Both miR inhibitors and miRNA mimics were commercially purchased and details are described in Table S11. Lipofectamine 2000 was used for the transfections in 1:1 ratio with the amount of inhibitor or mimic being used. Control miRNA inhibitor or control miRNA mimic at the same concentrations was always used in parallel. Transfections were for 6 hours. 36 hours post-transfection either RNA was isolated or lysate was prepared using RIPA buffer [1 mM Tris HCl pH 7.8, 150 mM NaCl, 2 % Triton X-100, 1% (w/v) Sodium deoxycholate, 0.1% (w/v) SDS) supplemented with 1X PIC and 1mM PMSF].

Journal of Cell Science • Accepted manuscript

Transfections involving plasmids were carried out using the respective plasmids in 6-well cluster plates for 48 hours. All siRNA transfections were carried out using 200pmole of the respective siRNAs for 60 hours. For shSin3b induction, cells transfected with pTRIPZ shSin3b were treated with doxycycline (1μg/ml) for 48 hours. Corresponding siRNA or shRNA controls were always used in parallel for all the ablation experiments.

Electrophoretic Mobility Shift Assay (EMSA)

For each EMSA reaction the radiolabelled substrate (10^4 cpm) was incubated with 500ng of recombinant CDX2 protein in a binding buffer (10mM Tris pH-7.5,50mM KCl, 2.5mM $MgCl_2$., 0.5mM DTT and 4% glycerol) for 15 minutes at 4°C. 1μg/μl of poly dI-dC was added in the reaction mixture to prevent non-specific binding. When "supershift" was desired, anti-CDX2 antibody (1μg/reaction) was added and incubation continued at 37°C for another 15 minutes. In certain reactions, 1000X fold-excess cold competitor was also added to confirm the specificity of the assay. Post-completion of the reactions, loading dye (5XTBE, 10% glycerol, 10% bromophenol blue, 1% xylene cyanol, autoclaved H_2O) was added in all the tubes. Samples were then loaded on 6% native PAGE gel which had already been pre-run in TBE buffer at 20mA and 300V for 30 minutes. Sample without CDX2 was used as control for the EMSA reactions. After the run was completed, the gel was dried at 65°C for 1 hour and exposed overnight, followed by autoradiography.

Small RNA sequencing

Total RNA was isolated from asynchronously growing BLM isogenic pair of cells – (a) GM03509 GFP-BLM 4.3.4/ GM03509 GFP (b) Clone 100 and GM08505 GFP-BLM/ GM08505 GFP. RNA was extracted using Trizol and the isolated RNA was used for library preparation using Illumina Small RNA sample preparation kit v1.5 according to the manufacturer's instructions. The total RNA (700-800ng) were ligated to 3′ and 5′ RNA adapters. The ligation products were reverse transcribed using Superscript II Reverse Transcriptase and amplified with 12 cycles of PCR. The PCR products constituting the small RNA cDNA libraries were resolved on 6% Novex TBE PAGE Gel and ~150 bp fragments excised. The library was eluted from the PAGE gel and analyzed on Agilent 2100 Bioanalyzer using DNA high sensitivity kit (Agilent Technologies, USA). Sequencing of

Journal of Cell Science • Accepted manuscript

miRNA libraries (~150bp fragments) were performed using Illumina GAIIX sequencing platform for 36 cycles. CLC genomic software was used to determine quantitatively the levels of the differentially miRNAa in the two isogenic pairs. With the help of this software, adaptor trimming was also done. The remaining sequence was mapped with the known miRNAs in miR Base database. Subsequently, the mapped miRNAs were normalized to get TPM (Transcript per million). Further analysis was carried out to determine the miRNAs whose expression was either increased or decreased by the presence of BLM. Only those miRNAs which showed a change above or below two-fold and a p-value either equal to or below 0.05 were chosen for further analysis. The raw sequence reads have been deposited in GEO (see Table S11 for details).

In silico predictions

In order to determine the transcription factors involved in the regulation of the miRNAs, initial *in silico* analysis was carried out using the database called ChIP base (*rna.sysu.edu.cn/chipbase/)*. The region upto 5kb upstream of the TSS was considered as the promoter region of the miRs. Using the ChIP Base database, a list of transcription factors binding to the promoter regions of the DDSMs was determined. Only the transcription factors binding to atleast four out of the six DDSM promoters are depicted in Table S1.

From putative targets for each of the miRs (Table S6), the common targets of all DDSMs were predicted (Table S7). Predicting the binding sites of the DDSMs in the 3'UTR of BRCA1 (Table S8) were determined from MiRanda tool (www.microrna.org). The potential gene targets for each of the miRNAs were predicted and then plotted using jvenn software (Bardou et al., 2014). Pathway analysis was carried out using Reactome, KEGG, Wiki Pathways and GSEA online databases and the significant pathways with p-value <0.05 were selected.

Nanoparticle mediated miRNA delivery

The cationic polymer (TAC6) was used for the *in vivo* delivery of miRNA inhibitors (Yavvari et al., 2019). miRNA inhibitors (10 µL of 20 µM) were mixed with TAC6 polymer (final volume 100 µL at 1 mg/mL) and incubated for 20 min at room temperature. Complexes were then mixed with sodium aspartate (final volume 10µL at 1mg/mL) for 10 min. The

Journal of Cell Science • Accepted manuscript

nanogels thus created were diluted with PBS. Each mouse was given a dose of 200ng miRNA (50µL of nanogels). A total of four doses were used.

Mass spectrometry

The BLM immunoprecipitate was electrophoresed on SDS-PAGE and stained by Coomassie. Each lane of the gel containing BLM-interacting proteins was subjected to mass spectrometry. Briefly, each lane was cut out from the gel, separately into smaller pieces. Coomassie stained gel pieces were de-stained using 25mM ammonium bicarbonate and 50% (v/v) Acetonitrile solution. Subsequently, they were treated with 0.1M TCEP for 45 min at 37°C, followed by 0.5M Iodoacetamide for 1 hr at 37°C. Overnight tryptic in-gel digestion was then carried out, trypsin: protein ratio of 1:100. The next day, peptides were recovered. pH of the supernatant was set to acidic pH (pH~3) using trifluoroacetic acid. The supernatant was dried in a Speed Vac. Resuspension was carried out in 5% Acetonitrile, 0.1% Formic acid. Desalting was carried out using ZIP TIP (C18 P-10, Millipore). The eluted peptides were dried using speed vac. The peptides were finally resuspended in 5% Acetonitrile, 0.1% Formic acid).

All mass spectrometry experiments were performed using EASY-nLC system (Thermo Fisher Scientific) coupled to LTQ Orbitrap-Velos mass spectrometer (Thermo Fisher Scientific) equipped with nano-electrospray ion source. A 10 cm PicoFrit Self-Pack microcapillary column (New Objective) was used to resolve the peptide mixture and the peptides were eluted. The LTQ Orbitrap-velos was operated using the Top20 CID (High/High) data-dependent acquisition mode with a full scan in the Orbitrap and a MS/MS scan in the CID. The target values for the full scan MS spectra were set at 0.5 x 10^6 charges with maximum injection time of 300 ms and a resolution of 60,000 at m/z 400. Spectra obtained were queried against Human Uniprot database. The precursor and fragment mass tolerances were set at 10 ppm and 0.8 Da, respectively. Proteome Discoverer 1.3 was used as the search algorithm with oxidation of methionine and carbamido-methylation of cysteine as static modification. All PSMs were identified at 1% false discovery rate (FDR). For peptide identification, a peptide posterior error probability (PEP) threshold of 0.01 was specified. A comprehensive and nonredundant list of all Human proteins along with their Uniprot ID and functional classification was generated using Uniprot database. The same Sin3b and CHD4 peptides were detected in the BLM immunoprecipitates three times.

Journal of Cell Science • Accepted manuscript

Statistical analysis and primary data availability

All quantitations are presented as mean ± S.D. except Figure 1B where it is median ± range. The number of replicates are mentioned in figure legends. The p values or calculated probability is as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ while n.s. indicates that the result is not significant. The statistical analyses employed for every experiment are shown in Table S14. All primary data has been deposited in Mandeley Data (see Table S11 for details).

Author Contributions

SP carried out cell biology, biochemical, clinical sample experiments. EK carried out all *in silico* analysis. SK carried out cell biology and *in vitro* assays. AP carried out small RNA sequencing and analysis. HA carried out proteomic analysis. AK carried out *in vitro* assays. SP, NK, RS carried out animal experiments. PB, SVSD provided clinical samples. JSP provided pathology support. IG, J-NF generated reagents and provided intellectual input for the project. PN, AB, SS supervised animal experiments. AM supervised small RNA sequencing and its analysis. SP, SK, EK, AP, AM, SS analyzed the data. J-NF provided critical input for manuscript preparation. SS wrote the manuscript.

Acknowledgement

SS acknowledges National Institute of Immunology (NII) core funds, Department of Biotechnology (DBT), India (BT/MED/30/SP11263/2015, BT/PR23545/BRB/10/1593/2017, BT/PR27681/GET/119/269/2018), Council of Scientific and Industrial Research (CSIR), India (37(1699)/17/EMR-11), Science & Engineering Research Board (SERB), India (EMR/2017/000541) and J C Bose Fellowship (JCB/2018/000013) for financial assistance. AM was supported by the Ramalingaswami Reentry Fellowship (BT/HRD/35/02/2008), the National Bioscience Award for Career Development (BT/HRD/NBA/38/04/2016) and SERB-STAR award (STR/2019/000064). SP acknowledges Indian Council of Medical Research for salary (2020-6197/CMB-BMS). EK acknowledges DST Inspire Faculty Fellowship (DST/INSPIRE/04/2017/000088) for salary and funding. The authors acknowledges Dipanjan Chowdhury (Dana-Farber Cancer Institute, USA), Nathan Ellis (The University of Arizona Cancer Center, USA), Gregory David (New York University School of

Journal of Cell Science • Accepted manuscript

Medicine, USA), Jeff Wrana (University of Toronto, Canada), Joan Massague (Slaon Kettering Institute, USA) for recombinants and Nathan Ellis (The University of Arizona Cancer Center, USA), Annick Harel-Bellan (University of Paris-Saclay, France), Bert Vogelstein (Johns Hopkins Medicine, USA) for cells, Aasheesh Srivastava (IISER Bhopal) for polymer, Vaibhav Jain for help with bioinformatic analysis, Kunal Dhall for clinical data compilation. SS, AM are grateful to the Department of Biotechnology (DBT), Government of India, for a generous infrastructure grant for establishment of the Next Generation Sequencing core facility in NII.

References

Aoki, K., Tamai, Y., Horiike, S., Oshima, M. and Taketo, M. M. (2003). Colonic polyposis caused by mTOR-mediated chromosomal instability in Apc+/Delta716 Cdx2+/- compound mutant mice. *Nat Genet* **35**, 323-30.

Baba, Y., Nosho, K., Shima, K., Freed, E., Irahara, N., Philips, J., Meyerhardt, J. A., Hornick, J. L., Shivdasani, R. A., Fuchs, C. S. et al. (2009). Relationship of CDX2 loss with molecular features and prognosis in colorectal cancer. *Clin Cancer Res* **15**, 4665-73.

Bai, A. H., Tong, J. H., To, K. F., Chan, M. W., Man, E. P., Lo, K. W., Lee, J. F., Sung, J. J. and Leung, W. K. (2004). Promoter hypermethylation of tumor-related genes in the progression of colorectal neoplasia. *Int J Cancer* **112**, 846-53.

Balbinot, C., Armant, O., Elarouci, N., Marisa, L., Martin, E., De Clara, E., Onea, A., Deschamps, J., Beck, F., Freund, J. N. et al. (2018). The Cdx2 homeobox gene suppresses intestinal tumorigenesis through non-cell-autonomous mechanisms. *J Exp Med* **215**, 911-926.

Balbinot, C., Vanier, M., Armant, O., Nair, A., Penichon, J., Soret, C., Martin, E., Saandi, T., Reimund, J. M., Deschamps, J. et al. (2017). Fine-tuning and autoregulation of the intestinal determinant and tumor suppressor homeobox gene CDX2 by alternative splicing. *Cell Death Differ* **24**, 2173-2186.

Bardou, P., Mariette, J., Escudie, F., Djemiel, C. and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**, 293.

Baymaz, H. I., Karemaker, I. D. and Vermeulen, M. (2015). Perspective on unraveling the versatility of 'co-repressor' complexes. *Biochim Biophys Acta* **1849**, 1051-6.

Bonhomme, C., Duluc, I., Martin, E., Chawengsaksophak, K., Chenard, M. P., Kedinger, M., Beck, F., Freund, J. N. and Domon-Dell, C. (2003). The Cdx2 homeobox gene has a tumour suppressor function in the distal colon in addition to a homeotic role during gut development. *Gut* **52**, 1465-71.

Brabletz, T., Spaderna, S., Kolb, J., Hlubek, F., Faller, G., Bruns, C. J., Jung, A., Nentwich, J., Duluc, I., Domon-Dell, C. et al. (2004). Down-regulation of the homeodomain factor Cdx2 in colorectal cancer by collagen type I: an active role for the tumor environment in malignant tumor progression. *Cancer Res* **64**, 6973-7.

Broderick, J. A., Salomon, W. E., Ryder, S. P., Aronin, N. and Zamore, P. D. (2011). Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. *RNA* **17**, 1858-69.

Journal of Cell Science • Accepted manuscript

Chandra, S., Priyadarshini, R., Madhavan, V., Tikoo, S., Hussain, M., Mudgal, R., Modi, P., Srivastava, V. and Sengupta, S. (2013). Enhancement of c-Myc degradation by BLM helicase leads to delayed tumor initiation. *J Cell Sci* **126**, 3782-95.

Chang, S. and Sharan, S. K. (2012). BRCA1 and microRNAs: emerging networks and potential therapeutic targets. *Mol Cells* **34**, 425-32.

Chawengsaksophak, K. (2019). Cdx2 Animal Models Reveal Developmental Origins of Cancers. *Genes (Basel)* **10**.

Chen, B., Xia, Z., Deng, Y. N., Yang, Y., Zhang, P., Zhu, H., Xu, N. and Liang, S. (2019). Emerging microRNA biomarkers for colorectal cancer diagnosis and prognosis. *Open Biol* **9**, 180212.

Choi, Y. E., Pan, Y., Park, E., Konstantinopoulos, P., De, S., D'Andrea, A. and Chowdhury, D. (2014). MicroRNAs down-regulate homologous recombination in the G1 phase of cycling cells to maintain genomic stability. *Elife* **3**, e02445.

Chung, I. F., Chang, S. J., Chen, C. Y., Liu, S. H., Li, C. Y., Chan, C. H., Shih, C. C. and Cheng, W. C. (2017). YM500v3: a database for small RNA sequencing in human cancer research. *Nucleic Acids Res* **45**, D925-D931.

Ciccia, A. and Elledge, S. J. (2010). The DNA damage response: making it safe to play with knives. *Mol Cell* **40**, 179-204.

Cilek, E. E., Ozturk, H. and Gur Dedeoglu, B. (2017). Construction of miRNA-miRNA networks revealing the complexity of miRNA-mediated mechanisms in trastuzumab treated breast cancer cell lines. *PLoS One* **12**, e0185558.

Clark, C. C., Weitzel, J. N. and O'Connor, T. R. (2012). Enhancement of synthetic lethality via combinations of ABT-888, a PARP inhibitor, and carboplatin in vitro and in vivo using BRCA1 and BRCA2 isogenic models. *Mol Cancer Ther* **11**, 1948-58.

Coebergh van den Braak, R. R. J., Sieuwerts, A. M., Lalmahomed, Z. S., Smid, M., Wilting, S. M., Bril, S. I., Xiang, S., van der Vlugt-Daane, M., de Weerd, V., van Galen, A. et al. (2018). Confirmation of a metastasis-specific microRNA signature in primary colon cancer. *Sci Rep* **8**, 5242.

Croce, C. M. (2009). Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* **10**, 704-14.

Croteau, D. L., Popuri, V., Opresko, P. L. and Bohr, V. A. (2014). Human RecQ helicases in DNA repair, recombination, and replication. *Annu Rev Biochem* **83**, 519-52.

Cunniff, C., Bassetti, J. A. and Ellis, N. A. (2017). Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. *Mol Syndromol* **8**, 4-23.

Dalerba, P., Sahoo, D., Paik, S., Guo, X., Yothers, G., Song, N., Wilcox-Fogel, N., Forgo, E., Rajendran, P. S., Miranda, S. P. et al. (2016). CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. *N Engl J Med* **374**, 211-22.

Davidson, D., Wang, Y., Aloyz, R. and Panasci, L. (2013). The PARP inhibitor ABT-888 synergizes irinotecan treatment of colon cancer cell lines. *Invest New Drugs* **31**, 461-8.

Ding, L., Lan, Z., Xiong, X., Ao, H., Feng, Y., Gu, H., Yu, M. and Cui, Q. (2018). The Dual Role of MicroRNAs in Colorectal Cancer Progression. *Int J Mol Sci* **19**.

Galland, A., Gourain, V., Habbas, K., Guler, Y., Martin, E., Ebel, C., Tavian, M., Vallat, L., Chenard, M. P., Mauvieux, L. et al. (2021). CDX2 expression in the hematopoietic lineage promotes leukemogenesis via TGFbeta inhibition. *Mol Oncol*.

Garcia, A. I., Buisson, M., Bertrand, P., Rimokh, R., Rouleau, E., Lopez, B. S., Lidereau, R., Mikaelian, I. and Mazoyer, S. (2011). Down-regulation of BRCA1 expression by miR-146a and miR-146b-5p in triple negative sporadic breast cancers. *EMBO Mol Med* **3**, 279-90.

Journal of Cell Science • Accepted manuscript

Goss, K. H., Risinger, M. A., Kordich, J. J., Sanz, M. M., Straughen, J. E., Slovek, L. E., Capobianco, A. J., German, J., Boivin, G. P. and Groden, J. (2002). Enhanced tumor formation in mice heterozygous for Blm mutation. *Science* **297**, 2051-3.

Grabsch, H., Dattani, M., Barker, L., Maughan, N., Maude, K., Hansen, O., Gabbert, H. E., Quirke, P. and Mueller, W. (2006). Expression of DNA double-strand break repair proteins ATM and BRCA1 predicts survival in colorectal cancer. *Clin Cancer Res* **12**, 1494-500.

Grassi, A., Perilli, L., Albertoni, L., Tessarollo, S., Mescoli, C., Urso, E. D. L., Fassan, M., Rugge, M. and Zanovello, P. (2018). A coordinate deregulation of microRNAs expressed in mucosa adjacent to tumor predicts relapse after resection in localized colon cancer. *Mol Cancer* **17**, 17.

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P. and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91-105.

Gross, I., Duluc, I., Benameur, T., Calon, A., Martin, E., Brabletz, T., Kedinger, M., Domon-Dell, C. and Freund, J. N. (2008). The intestine-specific homeobox gene Cdx2 decreases mobility and antagonizes dissemination of colon cancer cells. *Oncogene* **27**, 107-15.

Gruber, S. B., Ellis, N. A., Scott, K. K., Almog, R., Kolachana, P., Bonner, J. D., Kirchhoff, T., Tomsho, L. P., Nafa, K., Pierce, H. et al. (2002). BLM heterozygosity and the risk of colorectal cancer. *Science* **297**, 2013.

Guo, L., Sun, B., Wu, Q., Yang, S. and Chen, F. (2012). miRNA-miRNA interaction implicates for potential mutual regulatory pattern. *Gene* **511**, 187-94.

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* **144**, 646-74.

Hernandez, R., Sanchez-Jimenez, E., Melguizo, C., Prados, J. and Rama, A. R. (2018). Downregulated microRNAs in the colorectal cancer: diagnostic and therapeutic perspectives. *BMB Rep* **51**, 563-571.

Hryniuk, A., Grainger, S., Savory, J. G. and Lohnes, D. (2014). Cdx1 and Cdx2 function as tumor suppressors. *J Biol Chem* **289**, 33343-54.

Hu, H., Du, L., Nagabayashi, G., Seeger, R. C. and Gatti, R. A. (2010). ATM is down-regulated by N-Myc-regulated microRNA-421. *Proc Natl Acad Sci U S A* **107**, 1506-11.

Hu, P., Beresten, S. F., van Brabant, A. J., Ye, T. Z., Pandolfi, P. P., Johnson, F. B., Guarente, L. and Ellis, N. A. (2001). Evidence for BLM and Topoisomerase IIIalpha interaction in genomic stability. *Hum Mol Genet* **10**, 1287-98.

Jacob, H., Stanisavljevic, L., Storli, K. E., Hestetun, K. E., Dahl, O. and Myklebust, M. P. (2017). Identification of a sixteen-microRNA signature as prognostic biomarker for stage II and III colon cancer. *Oncotarget* **8**, 87837-87847.

Ji, D., Qiao, M., Yao, Y., Li, M., Chen, H., Dong, Q., Jia, J., Cui, X., Li, Z., Xia, J. et al. (2018). Serum-based microRNA signature predicts relapse and therapeutic outcome of adjuvant chemotherapy in colorectal cancer patients. *EBioMedicine* **35**, 189-197.

Kandimalla, R., Gao, F., Matsuyama, T., Ishikawa, T., Uetake, H., Takahashi, N., Yamada, Y., Becerra, C., Kopetz, S., Wang, X. et al. (2018). Genome-wide Discovery and Identification of a Novel miRNA Signature for Recurrence Prediction in Stage II and III Colorectal Cancer. *Clin Cancer Res* **24**, 3867-3877.

Kaur, E., Agrawal, R. and Sengupta, S. (2021). Functions of BLM Helicase in Cells: Is It Acting Like a Double-Edged Sword? *Front Genet* **12**, 634789.

Journal of Cell Science • Accepted manuscript

Keene, J. D., Komisarow, J. M. and Friedersdorf, M. B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* **1**, 302-7.

Kharat, S. S., Tripathi, V., Damodaran, A. P., Priyadarshini, R., Chandra, S., Tikoo, S., Nandhakumar, R., Srivastava, V., Priya, S., Hussain, M. et al. (2016). Mitotic phosphorylation of Bloom helicase at Thr182 is required for its proteasomal degradation and maintenance of chromosomal stability. *Oncogene* **35**, 1025-38.

Mallo, G. V., Soubeyran, P., Lissitzky, J. C., Andre, F., Farnarier, C., Marvaldi, J., Dagorn, J. C. and Iovanna, J. L. (1998). Expression of the Cdx1 and Cdx2 homeotic genes leads to reduced malignancy in colon cancer-derived cells. *J Biol Chem* **273**, 14030-6.

Matamala, N., Vargas, M. T., Gonzalez-Campora, R., Arias, J. I., Menendez, P., Andres-Leon, E., Yanowsky, K., Llaneza-Folgueras, A., Minambres, R., Martinez-Delgado, B. et al. (2016). MicroRNA deregulation in triple negative breast cancer reveals a role of miR-498 in regulating BRCA1 expression. *Oncotarget* **7**, 20068-79.

Moskwa, P., Buffa, F. M., Pan, Y., Panchakshari, R., Gottipati, P., Muschel, R. J., Beech, J., Kulshrestha, R., Abdelmohsen, K., Weinstock, D. M. et al. (2011). miR-182-mediated downregulation of BRCA1 impacts DNA repair and sensitivity to PARP inhibitors. *Mol Cell* **41**, 210-20.

Mukherji, S., Ebert, M. S., Zheng, G. X., Tsang, J. S., Sharp, P. A. and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. *Nat Genet* **43**, 854-9.

Negrini, S., Gorgoulis, V. G. and Halazonetis, T. D. (2010). Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* **11**, 220-8.

Nguyen, G. H., Dexheimer, T. S., Rosenthal, A. S., Chu, W. K., Singh, D. K., Mosedale, G., Bachrati, C. Z., Schultz, L., Sakurai, M., Savitsky, P. et al. (2013). A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. *Chem Biol* **20**, 55-62.

Oh, M., McBride, A., Yun, S., Bhattacharjee, S., Slack, M., Martin, J. R., Jeter, J. and Abraham, I. (2018). BRCA1 and BRCA2 Gene Mutations and Colorectal Cancer Risk: Systematic Review and Meta-analysis. *J Natl Cancer Inst* **110**, 1178-1189.

Ozawa, T., Kandimalla, R., Gao, F., Nozawa, H., Hata, K., Nagata, H., Okada, S., Izumi, D., Baba, H., Fleshman, J. et al. (2018). A MicroRNA Signature Associated With Metastasis of T1 Colorectal Cancers to Lymph Nodes. *Gastroenterology* **154**, 844-848 e7.

Petrovic, N., Davidovic, R., Bajic, V., Obradovic, M. and Isenovic, R. E. (2017). MicroRNA in breast cancer: The association with BRCA1/2. *Cancer Biomark* **19**, 119-128.

Pilati, C., Taieb, J., Balogoun, R., Marisa, L., de Reynies, A. and Laurent-Puig, P. (2017). CDX2 prognostic value in stage II/III resected colon cancer is related to CMS classification. *Ann Oncol* **28**, 1032-1035.

Priyadarshini, R., Hussain, M., Attri, P., Kaur, E., Tripathi, V., Priya, S., Dhapola, P., Saha, D., Madhavan, V., Chowdhury, S. et al. (2018). BLM Potentiates c-Jun Degradation and Alters Its Function as an Oncogenic Transcription Factor. *Cell Rep* **24**, 947-961.e7.

Rajarajan, S., C, E. A., Jose, B., Correa, M., Sengupta, S. and Prabhu, J. S. (2020). Identification of colorectal cancers with defective DNA damage repair by immunohistochemical profiling of mismatch repair proteins, CDX2 and BRCA1. *Mol Clin Oncol* **13**, 57.

Saetrom, P., Heale, B. S., Snove, O., Jr., Aagaard, L., Alluin, J. and Rossi, J. J. (2007). Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res* **35**, 2333-42.

Journal of Cell Science • Accepted manuscript

Sakamoto, N., Feng, Y., Stolfi, C., Kurosu, Y., Green, M., Lin, J., Green, M. E., Sentani, K., Yasui, W., McMahon, M. et al. (2017). BRAF(V600E) cooperates with CDX2 inactivation to promote serrated colorectal tumorigenesis. *Elife* **6**.

Salari, K., Spulak, M. E., Cuff, J., Forster, A. D., Giacomini, C. P., Huang, S., Ko, M. E., Lin, A. Y., van de Rijn, M. and Pollack, J. R. (2012). CDX2 is an amplified lineage-survival oncogene in colorectal cancer. *Proc Natl Acad Sci U S A* **109**, E3196-205.

Savage, K. I. and Harkin, D. P. (2015). BRCA1, a 'complex' protein involved in the maintenance of genomic stability. *FEBS J* **282**, 630-46.

Saxena, M., Roman, A. K. S., O'Neill, N. K., Sulahian, R., Jadhav, U. and Shivdasani, R. A. (2017). Transcription factor-dependent 'anti-repressive' mammalian enhancers exclude H3K27me3 from extended genomic domains. *Genes Dev* **31**, 2391-2404.

Sengupta, S., Robles, A. I., Linke, S. P., Sinogeeva, N. I., Zhang, R., Pedeux, R., Ward, I. M., Celeste, A., Nussenzweig, A., Chen, J. et al. (2004). Functional interaction between BLM helicase and 53BP1 in a Chk1-mediated pathway during S-phase arrest. *J Cell Biol* **166**, 801-13.

Shu, Y., Pi, F., Sharma, A., Rajabi, M., Haque, F., Shu, D., Leggas, M., Evers, B. M. and Guo, P. (2014). Stable RNA nanoparticles as potential new generation drugs for cancer therapy. *Adv Drug Deliv Rev* **66**, 74-89.

Smith, J., Tho, L. M., Xu, N. and Gillespie, D. A. (2010). The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. *Adv Cancer Res* **108**, 73-112.

Syed, A. and Tainer, J. A. (2018). The MRE11-RAD50-NBS1 Complex Conducts the Orchestration of Damage Signaling and Outcomes to Stress in DNA Replication and Repair. *Annu Rev Biochem* **87**, 263-294.

Tikoo, S., Madhavan, V., Hussain, M., Miller, E. S., Arora, P., Zlatanou, A., Modi, P., Townsend, K., Stewart, G. S. and Sengupta, S. (2013). Ubiquitin-dependent recruitment of the Bloom Syndrome helicase upon replication stress is required to suppress homologous recombination. *EMBO J* **32**, 1778-92.

Tikoo, S. and Sengupta, S. (2010). Time to bloom. *Genome Integr* **1**, 14.

Tripathi, V., Agarwal, H., Priya, S., Batra, H., Modi, P., Pandey, M., Saha, D., Raghavan, S. C. and Sengupta, S. (2018). MRN complex-dependent recruitment of ubiquitylated BLM helicase to DSBs negatively regulates DNA repair pathways. *Nat Commun* **9**, 1016.

Verzi, M. P., Shin, H., San Roman, A. K., Liu, X. S. and Shivdasani, R. A. (2013). Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. *Mol Cell Biol* **33**, 281-92.

Vu, T., Straube, J., Porter, A. H., Bywater, M., Song, A., Ling, V., Cooper, L., Pali, G., Bruedigam, C., Jacquelin, S. et al. (2020). Hematopoietic stem and progenitor cell-restricted Cdx2 expression induces transformation to myelodysplasia and acute leukemia. *Nat Commun* **11**, 3021.

Wang, C., Jette, N., Moussienko, D., Bebb, D. G. and Lees-Miller, S. P. (2017). ATM-Deficient Colorectal Cancer Cells Are Sensitive to the PARP Inhibitor Olaparib. *Transl Oncol* **10**, 190-196.

Wang, G. H., Zhao, C. M., Huang, Y., Wang, W., Zhang, S. and Wang, X. (2018). BRCA1 and BRCA2 expression patterns and prognostic significance in digestive system cancers. *Hum Pathol* **71**, 135-144.

Witek, M. E., Nielsen, K., Walters, R., Hyslop, T., Palazzo, J., Schulz, S. and Waldman, S. A. (2005). The putative tumor suppressor Cdx2 is overexpressed by human colorectal adenocarcinomas. *Clin Cancer Res* **11**, 8549-56.

Xu, J., Li, C. X., Li, Y. S., Lv, J. Y., Ma, Y., Shao, T. T., Xu, L. D., Wang, Y. Y., Du, L., Zhang, Y. P. et al. (2011). MiRNA-miRNA synergistic network: construction via

Journal of Cell Science • Accepted manuscript

co-regulating functional modules and disease miRNA topological features. *Nucleic Acids Res* **39**, 825-36.

Yan, D., Ng, W. L., Zhang, X., Wang, P., Zhang, Z., Mo, Y. Y., Mao, H., Hao, C., Olson, J. J., Curran, W. J. et al. (2010). Targeting DNA-PKcs and ATM with miR-101 sensitizes tumors to radiation. *PLoS One* **5**, e11397.

Yavvari, P. S., Verma, P., Mustfa, S. A., Pal, S., Kumar, S., Awasthi, A. K., Ahuja, V., Srikanth, C. V., Srivastava, A. and Bajaj, A. (2019). A nanogel based oral gene delivery system targeting SUMOylation machinery to combat gut inflammation. *Nanoscale* **11**, 4970-4986.

Yu, J., Liu, D., Sun, X., Yang, K., Yao, J., Cheng, C., Wang, C. and Zheng, J. (2019). CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/beta-catenin signaling via transactivation of GSK-3beta and Axin2 expression. *Cell Death Dis* **10**, 26.

Zheng, J., He, S., Qi, J., Wang, X., Yu, J., Wu, Y., Gao, Q., Wang, K. and Sun, X. (2017). Targeted CDX2 expression inhibits aggressive phenotypes of colon cancer cells in vitro and in vivo. *Int J Oncol* **51**, 478-488.

Zhou, T., Yi, F., Wang, Z., Guo, Q., Liu, J., Bai, N., Li, X., Dong, X., Ren, L., Cao, L. et al. (2019). The Functions of DNA Damage Factor RNF8 in the Pathogenesis and Progression of Cancer. *Int J Biol Sci* **15**, 909-918.

Zhou, Y., Wan, G., Spizzo, R., Ivan, C., Mathur, R., Hu, X., Ye, X., Lu, J., Fan, F., Xia, L. et al. (2014). miR-203 induces oxaliplatin resistance in colorectal cancer cells by negatively regulating ATM kinase. *Mol Oncol* **8**, 83-92.

Journal of Cell Science • Accepted manuscript

Figures

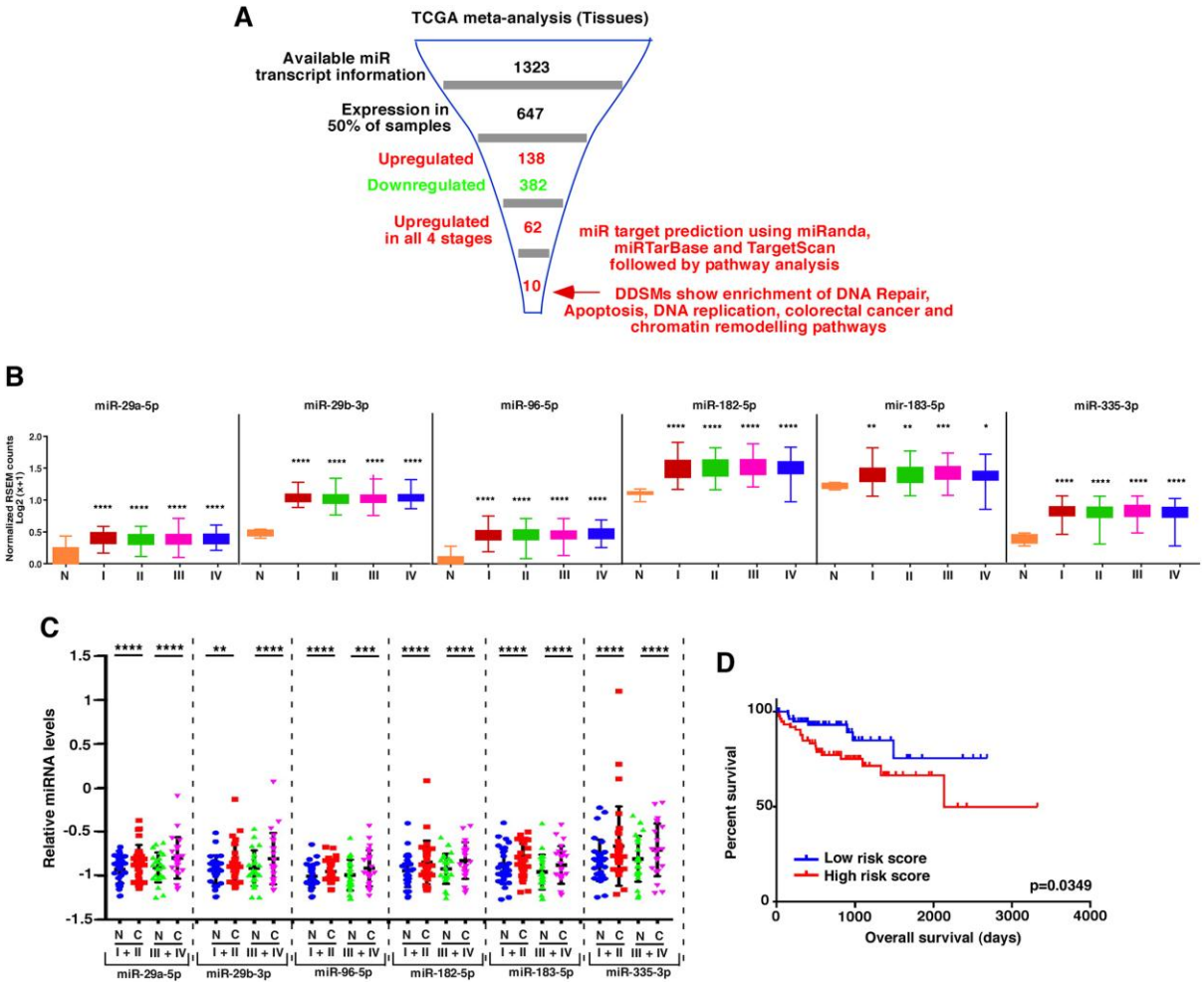


Figure 1: Levels of DDSMs are increased in colon adenocarcinoma patients.

A. Schematic diagram of the meta-analysis carried out for miR expression in colon adenocarcinoma tissue samples present in TCGA database. Out of the 1323 miRs for which the miR expression information was available, 62 miRs were identified to be upregulated in all four stages of colon cancer. The targets of these 62 miRs were analyzed by multiple pathway prediction software. Ten out these 62 miRs were characterized as DDSMs.

B. Increased levels of DDSMs were observed in the tissue samples of colon adenocarcinoma patients in TCGA database. The expression levels of the DDSMs were quantitated in the tissues of the four stages of colon cancer patients and normal individuals available in TCGA database.

Journal of Cell Science • Accepted manuscript

C. Increased levels of DDSMs were observed in the tissue samples of colon cancer patients from India. The levels of the DDSMs were analyzed by RT-qPCR in colon cancer tissues and their adjacent normal tissues in the Indian cohort and healthy normal individuals (n=54).

D. Survival function of colon cancer patients increased with decreased expression of the six DDSM signature. Kaplan-Meier curve was generated from TCGA dataset to determine the survival function of the colon cancer patients showing the combined expression of six DDSM signature in the colonic tissues from colon cancer patients.

Journal of Cell Science • Accepted manuscript

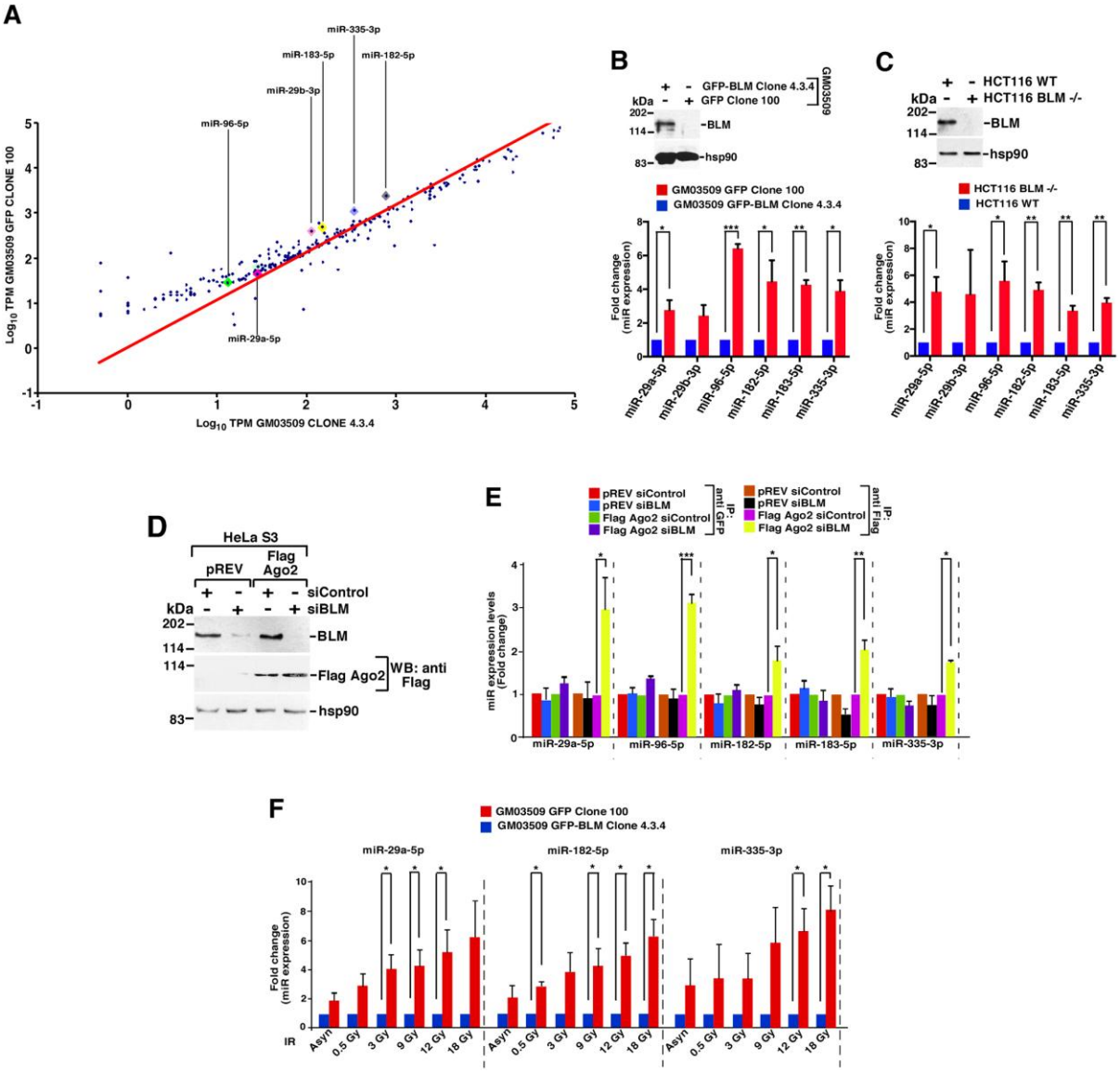


Figure 2: Validation of DDSMs.

A. Alteration in microRNA levels in absence or presence of BLM in GM03509 cells. Small RNA sequencing was done in isogenic cell lines derived from BS patient, GM03509 GFP-BLM Clone 4.3.4 and GM03509 GFP Clone 100. Six DDSMs have been indicated.

B, C. Validation of DDSMs whose expressions were increased in absence of BLM. Two isogenic pairs of cell lines (B) GM03509 GFP-BLM Clone 4.3.4 and GM03509 GFP Clone 100 and (C) HCT116 WT and HCT116 BLM -/- were validated by western analysis using antibodies against BLM, hsp90. DDSMs were validated by RT-qPCR analysis. Three biological replicates were used for protein analysis and RT-qPCR validation.

Journal of Cell Science • Accepted manuscript

D, E. Upregulated DDSMs are bound to the RISC complex. (D) Ablation of BLM in HeLa S3 pREV and HeLa S3 Flag Ago2 cells. HeLa S3 pREV and HeLa S3 Flag Ago2 cells were transfected with either siControl or siBLM. The lysates made were probed with antibodies against BLM, Flag, hsp90. Three biological replicates were carried out and representative blots presented. (E) Lysates from (D) were used for immunoprecipitations with either anti-GFP or anti-Flag antibody. RNA was isolated from the immunoprecipitated material and RT-qPCR was carried out to estimate the levels of the indicated DDSMs. Quantitation from three biological replicates.

F. Levels of DDSMs increased in absence of BLM after IR exposure. Isogenic cell lines, GM03509 GFP-BLM Clone 4.3.4 and GM03509 GFP Clone 100 were exposed to a gradient of IR. RNA was prepared 1 hr post-exposure. Expression of indicated DDSMs were validated by RT-qPCR analysis. Quantitation from three biological replicates.

Journal of Cell Science • Accepted manuscript

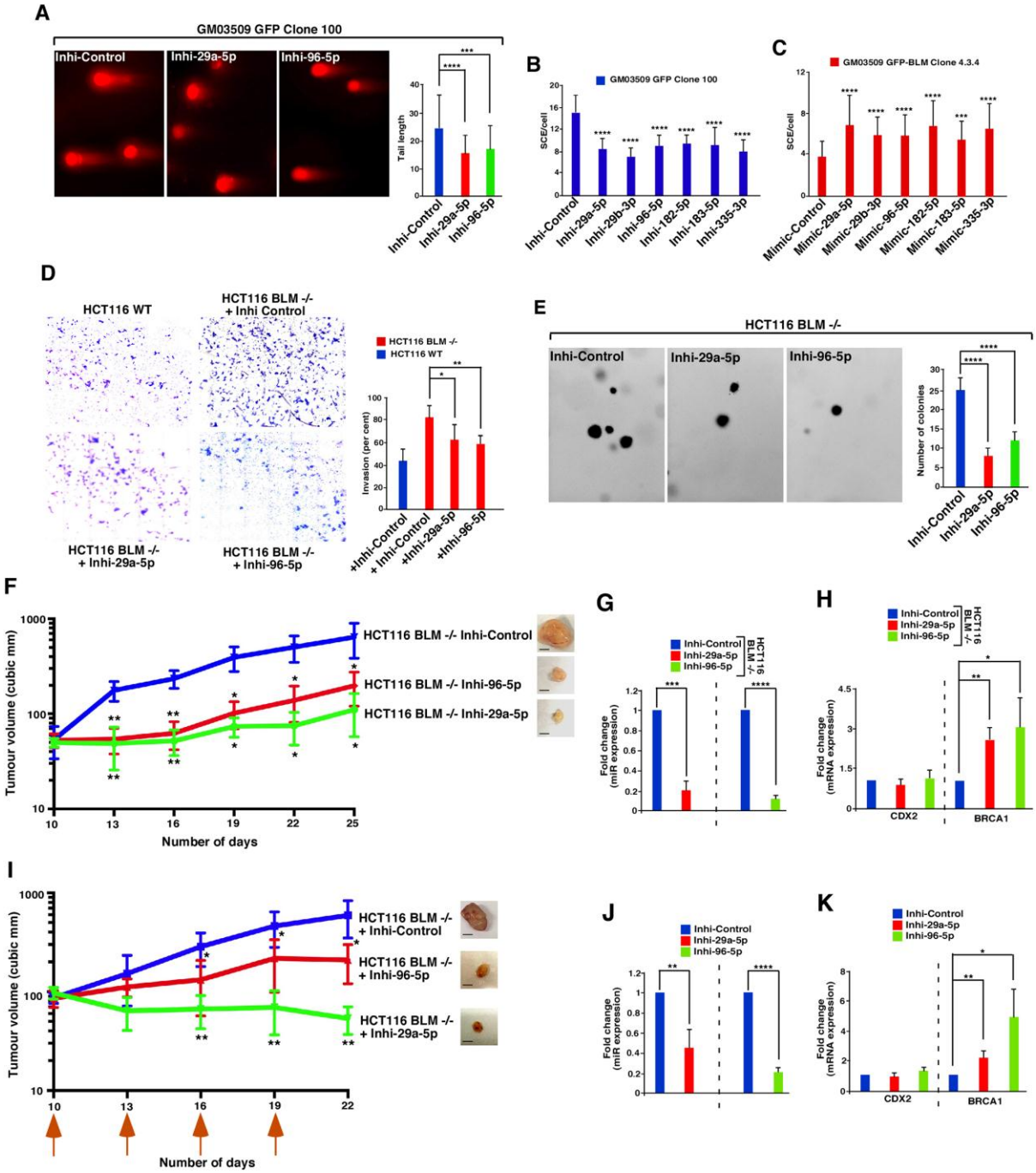


Figure 3: DDSMs sustain tumor growth.

A. Inhibition of DDSMs decreased DNA damage. Neutral comet assays were carried out in GM03509 GFP Clone 100 cells transfected with indicated miR inhibitors. (Left) Representative images of cells with comets. (Right) Quantitation (n=50).

B, C. DDSMs modulated SCEs. SCEs were carried out in (B) GM03509 GFP Clone 100 cells transfected with the indicated miR inhibitors and (C) GM03509 GFP-BLM Clone 4.3.4 cells transfected with the indicated miR mimics. Quantitation (B) 29 spreads, (C) 31 spreads.

Journal of Cell Science • Accepted manuscript

D, E. Inhibition of DDSMs decreased invasion and colony formation in soft agar assay. HCT116 BLM -/- cells transfected with the indicated miR inhibitors. HCT116 WT was used as a control in (D). (D) Matrigel invasion assay (n=6) and (E) soft agar assay colony formation assay (n=9) were carried out. (D, E - left) Representative images of the two assays. (D, E - right) Quantitation.

F, I. Inhibition of DDSMs decreased the rate of tumor formation in mice xenograft models. (F) HCT116 BLM -/- derived stable lines expressing GFP and the indicated miR inhibitors were injected subcutaneously into NOD SCID mice (n=4). (I) Nanoparticle encoded miRs were injected into the base of 100 cubic cm tumors generated by subcutaneously injecting HCT116 BLM -/- cells in NOD SCID mice (n=5). Days on which injections were carried out have been indicated by arrows. Tumor formation was monitored over the period indicated. One representative excised tumor for each condition has been shown. All statistical significance was calculated relative to their respective controls.

G, J. Presence of inhibitors decreased DDSM levels in tumors excised at the end of xenograft experiments. RNA was isolated from the tumors at the end of both the xenograft experiments (F and I). Levels of the indicated miRs were determined by RT-qPCR analysis. The quantitation performed using RNA isolated from four mice in each condition.

H, K. Presence of DDSM inhibitors increased BRCA1 levels in tumors excised at the end of the xenograft experiments. RNA was isolated from the tumors at the end of both xenograft models (F and I). Levels of CDX2 and BRCA1 transcript were determined by RT-qPCR analysis. Quantitation from four mice in each condition.

Journal of Cell Science • Accepted manuscript

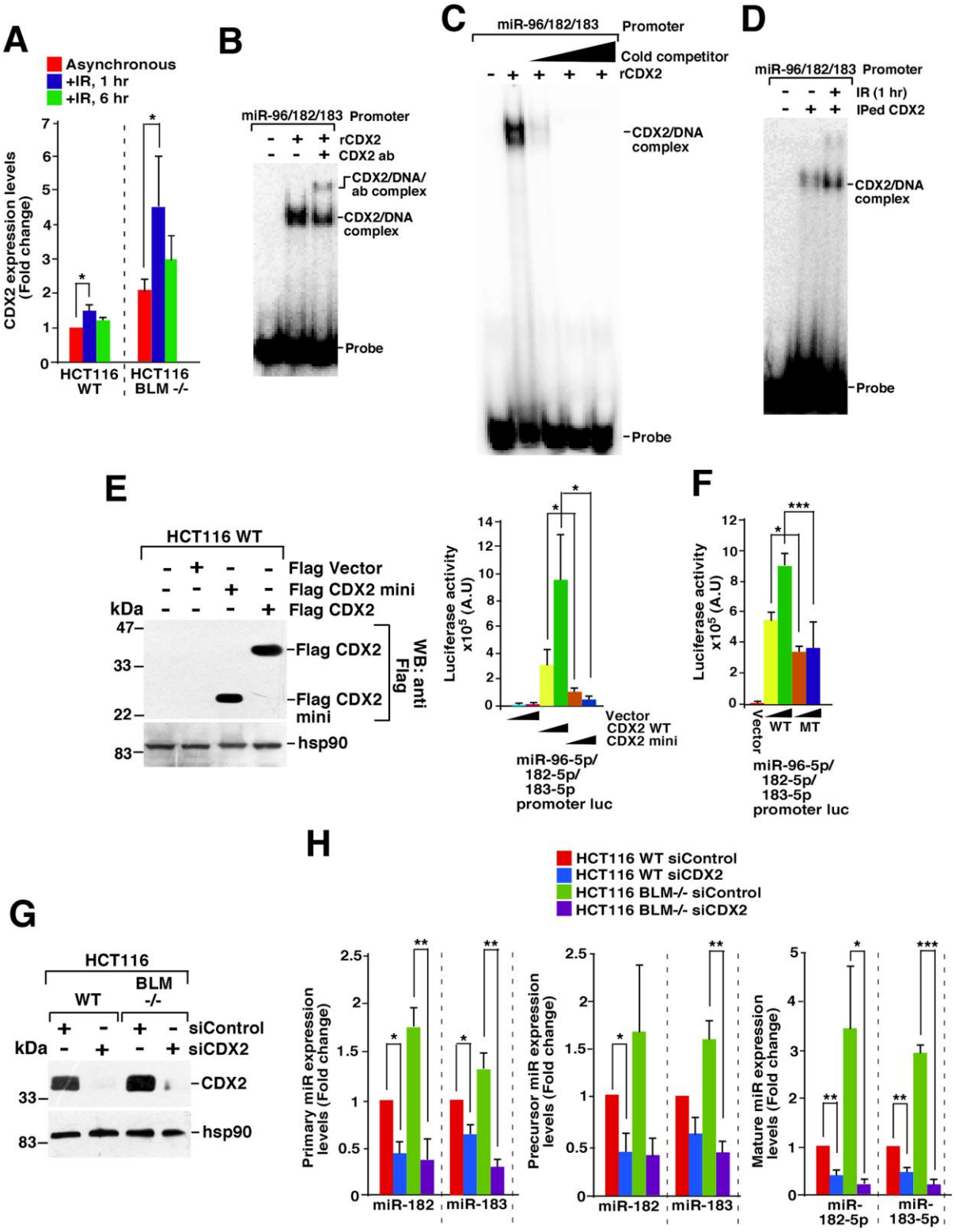


Figure 4: CDX2 regulates the DDSMs.

A. CDX2 expression corelated with the extent of DNA damage in cells. RNA was isolated from HCT116 WT and HCT116 BLM -/-. RNA grown in asynchronous conditions or 1 hour

Journal of Cell Science • Accepted manuscript

or 6 hours post-IR exposure. Levels of CDX2 transcript were determined by RT-qPCR analysis. Quantitation from four biological replicates.

B-D. CDX2 bound to miR promoters. Radiolabeled double stranded annealed oligos containing CDX2 binding site present in the miR-96/182/183 promoter were generated. EMSAs were carried out in presence of (B) recombinant CDX2 alone in absence or presence of anti-CDX2 antibody, (C) recombinant CDX2 alone without or with increasing amounts of the cold competitor, (D) immunoprecipitated CDX2 from cells which were either left unirradiated or were exposed to IR (3 Gy, 1 hr). CDX2/DNA complexes were visualized by autoradiography. All experiments were done three times and representative EMSAs have been presented.

E. Lack of transactivation domain of CDX2 led to decreased promoter activity of the DDSMs. (Left) Expression of Flag tagged CDX2 WT and mini CDX2 proteins in HCT116 WT cells were determined by western analysis using antibodies against Flag and hsp90. Around 200ng of the respective plasmids were used for the transfection. (Right) Luciferase-based miR-96/182/183 promoter activity were carried out with lysates expressing either CDX2 WT or mini CDX2. Quantitation from four biological replicates.

F. Mutation of the CDX2 binding site in DDSM promoter abrogated the promoter activity. Same as (E) except luciferase assays were carried out in cells expressing CDX2 WT and either the wildtype or mutant miR-96/182/183 promoter where CDX2 binding site had been mutated. Quantitation from four biological replicates.

G, H. Ablation of CDX2 led to decreased DDSM levels. HCT116 WT or HCT116 BLM -/- cells were transfected with either siControl or siCDX2. (G) Levels of CDX2 were determined by western analysis using antibodies against CDX2 and hsp90. Representative western blots from one of the three biological replicates is presented. (H) RNA isolated from HCT116 cells transfected with siControl or siCDX2. Levels of the indicated primary, precursor or mature DDSMs were determined by RT-qPCR analysis. Quantitation from three biological replicates.

Journal of Cell Science • Accepted manuscript

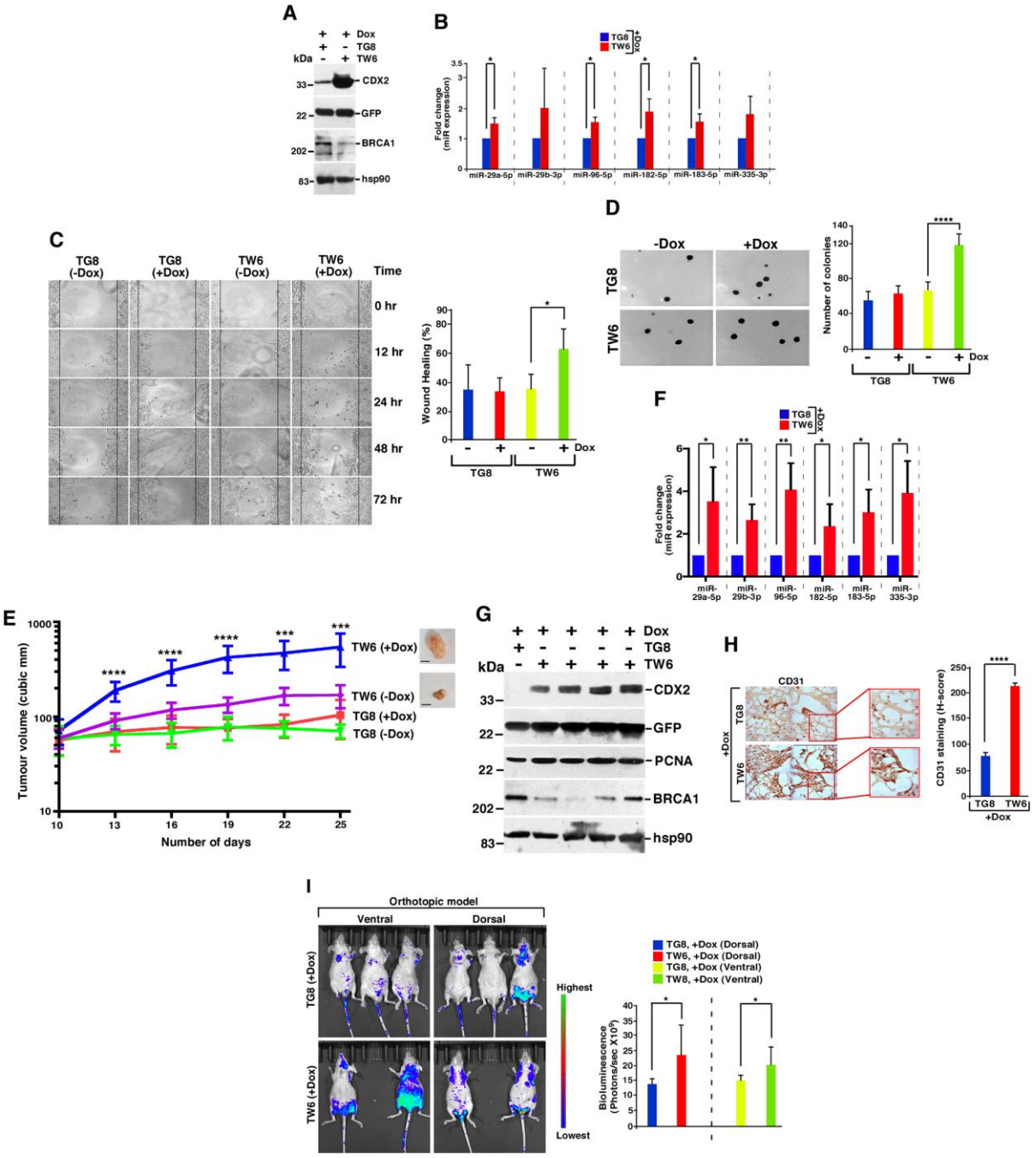


Figure 5: CDX2 regulated DDSMs disseminates colon cancer cells *in vivo*.

A. CDX2 was induced in TW6 cells. Lysates were made from TG8 and TW6 cells in presence of Doxycycline (Dox). Western blots were carried out using antibodies against CDX2, GFP, BRCA1 and hsp90. Representative western blots from one of the three biological replicates is presented.

Journal of Cell Science • Accepted manuscript

B. DDSMs were induced by CDX2 expression. RNA was isolated from TG8, TW6 cells grown asynchronously but in presence of Dox. Levels of the DDSMs were determined by RT-qPCR analysis. Quantitation from three biological replicates.

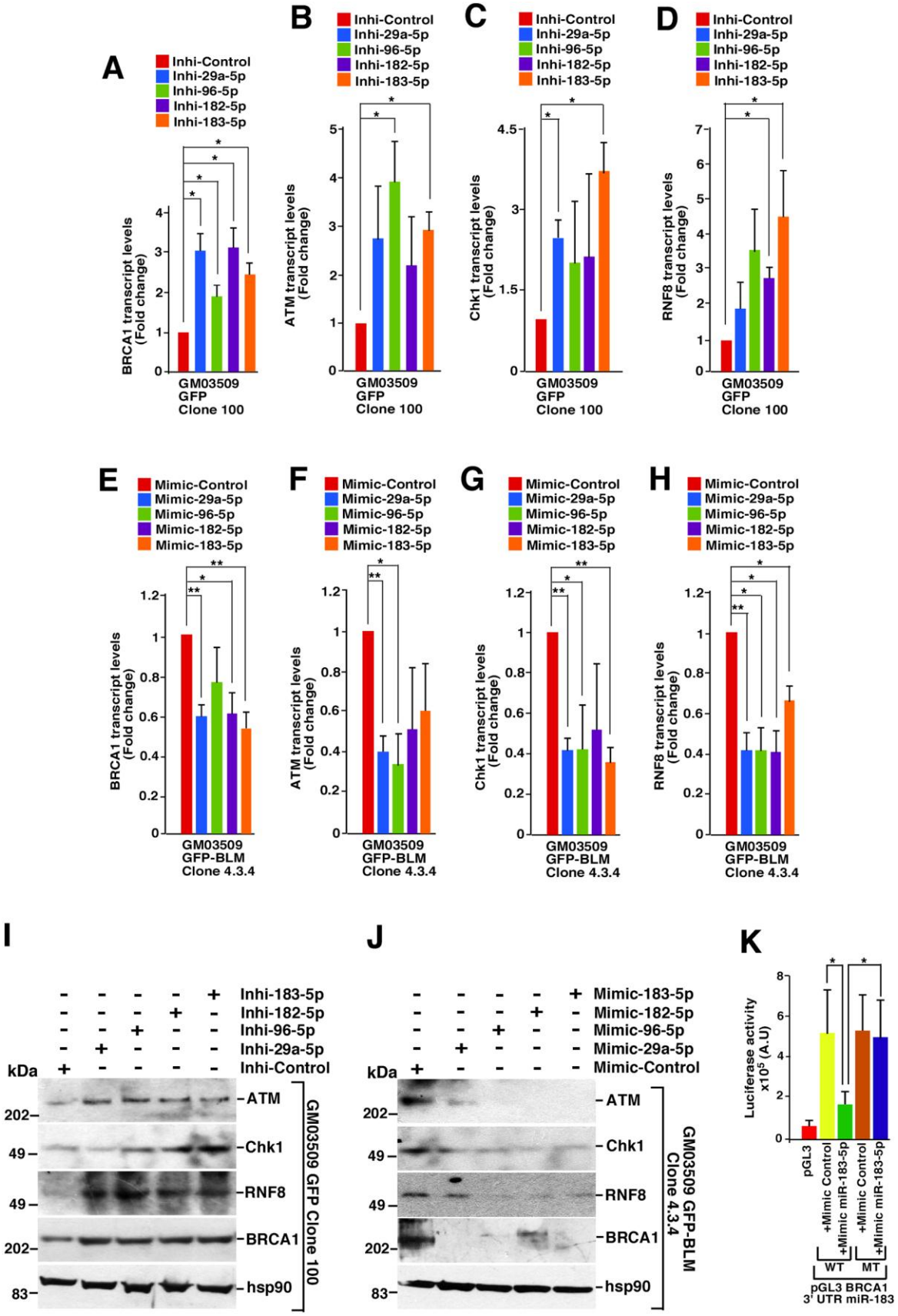
C, D. Induction of CDX2 led to enhanced wound healing and colony formation. (C) Scratch assay (n=3) and (D) Colony formation assay (n=4) were carried out with TG8, TW6 in ±Dox conditions. The time allowed (in hours) for wound healing has been indicated. (Left) Representative images of the assays. (Right) Quantitation.

E. Induction of CDX2 led to enhanced tumor formation in mice xenograft model. Tumor formation in a mice xenograft model was carried out by subcutaneously injecting TG8 and TW6 cells into NOD SCID mice (n=6). Mice were fed with Dox every day. Tumor formation was monitored over the indicated period. One representative tumor for each condition has been also represented. The statistical significance was determined by comparing +Dox condition with -Dox condition for both cell types. Tumors formed by TW6 in ±Dox conditions was found to be statistically significant.

F-H. Induction of DDSMs, proliferation and angiogenesis markers in CDX2 induced tumors derived in a xenograft model. (F) RNA and (G) protein were extracted from the tumors at the end point of the xenograft experiment (E). (F) Levels of the indicated DDSMs were determined by RT-qPCR analysis (RNA was from five different mice injected with TW6 cells) while (G) protein levels of CDX2, GFP, PCNA, BRCA1, hsp90 in the tumors was determined by carrying out western analysis with the indicated antibodies (extracts were from four different mice injected with TW6 cells were used). (H) IHC was carried out with tumor sections (from six different mice in each condition) with anti-CD31 antibodies. (Left) Representative images of CD31 staining. (Right) Quantitation.

I. Induction of CDX2-dependent DDSMs led to increase *in vivo* dissemination of cancer cells. *In vivo* dissemination of GFP expressing TG8 and TW6 cells were determined in an orthotopic model. Cells were implanted into the cecal wall of the mice (n=8) and fed with Dox every day. At the end of 21 days *in vivo* imaging of the mice (both ventral and dorsal) were carried out. (Left) Representative images of TG8, TW6 cell migration. (Right) Quantitation.

Journal of Cell Science • Accepted manuscript



Journal of Cell Science • Accepted manuscript

Figure 6: DDSMs target BRCA1.

A-D. Ablation of DDSMs enhanced the transcript levels of its targets. GM03509 GFP Clone 100 cells were transfected with the indicated DDSM inhibitors. The transcript levels of (A) BRCA1, (B) ATM, (C) Chk1, (D) RNF8 were determined by RT-qPCR. Quantitation from three biological replicates.

E-H. Overexpression of DDSMs decreased the transcript levels of its targets. GM03509 GFP-BLM Clone 4.3.4 cells were transfected with DDSM mimics. The levels of (E) BRCA1, (F) ATM, (G) Chk1, (H) RNF8 were determined by RT-qPCR. The statistical significance was calculated relative to Mimic Control. Quantitation from three biological replicates.

I, J. Levels of DDSMs determined the protein levels of its targets. (I) GM03509 GFP Clone 100 cells or (J) GM03509 GFP-BLM Clone 4.3.4 cells were transfected with either (I) DDSM inhibitors or (J) DDSM mimics. Levels of ATM, Chk1, RNF8, BRCA1, hsp90 were determined by carrying out western analysis with their corresponding antibodies. Three biological replicates were used for both experiments and representative blots presented.

K. miR-183 mimic did not bind to its mutated binding sequence in 3'UTR of BRCA1. Luciferase assays were carried out with extracts from HCT116 cells transfected with either mimic Control or the mimics for miR-183 in presence of either miR-183 BRCA1 3'UTR WT or miR-183 BRCA1 3'UTR MT. Quantitation from four biological replicates.

Journal of Cell Science • Accepted manuscript

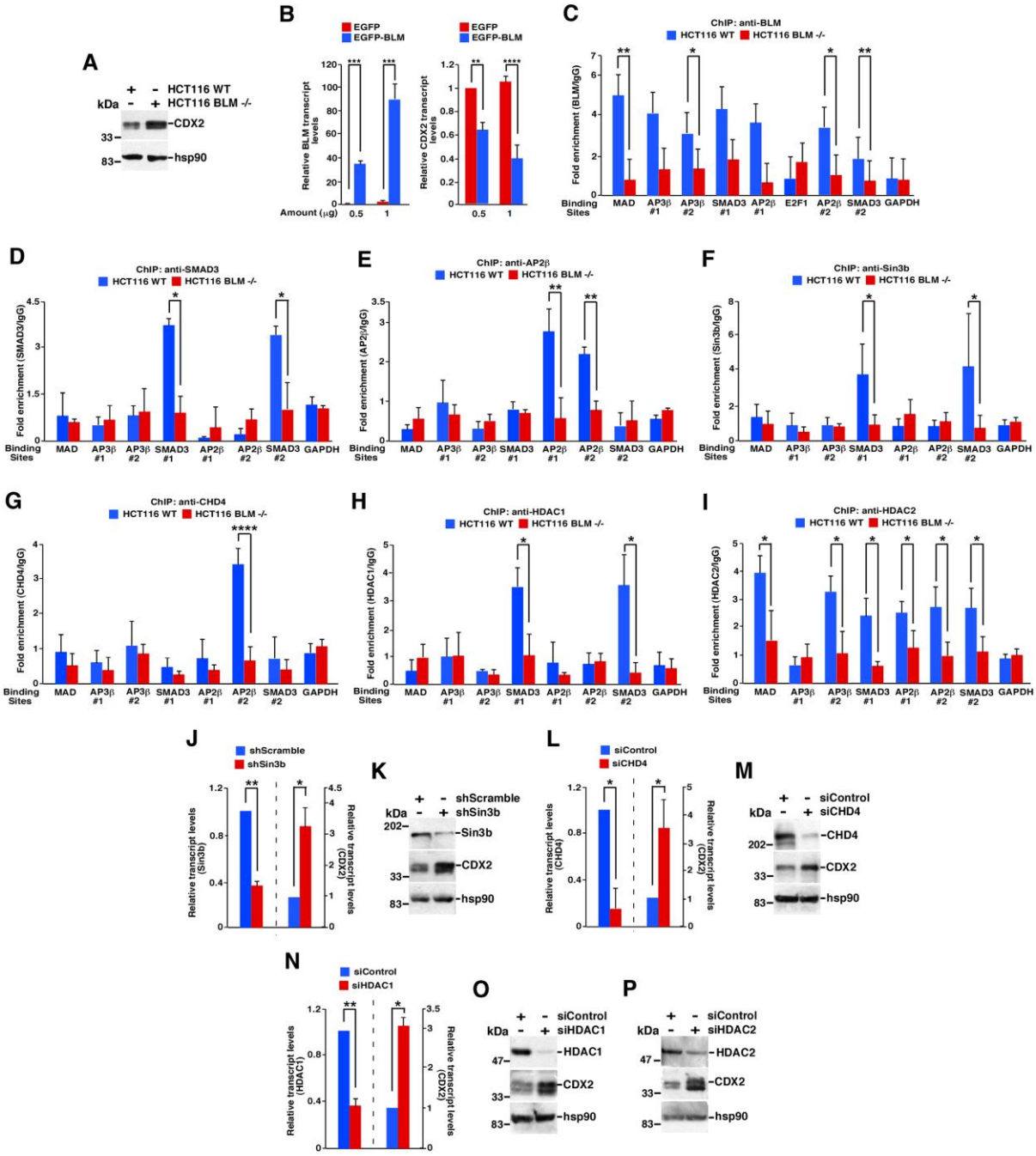


Figure 7: BLM repressed CDX2 transcription.

A. CDX2 protein levels increased in absence of BLM. Lysates made from HCT116 WT and HCT116 BLM -/- cells were probed with antibodies against CDX2 and hsp90. The experiment was done three times and representative blots presented.

Journal of Cell Science • Accepted manuscript

B. Expression of BLM decreased CDX2 transcript levels. HCT116 BLM -/- cells were transfected with either EGFP or EGFP-BLM. BLM (left) and CDX2 (right) transcript levels were determined by RT-qPCR. Quantitation from four biological replicates.

C. BLM was recruited to the CDX2 promoter. BLM ChIP were carried out with chromatin isolated from HCT116 WT and HCT116 BLM-/- cells. The recruitment of BLM to the putative binding sites of the transcriptional repressors and activator on CDX2 promoter has been shown. The corresponding IgG was used as the antibody control. Quantitation from four biological replicates.

D, E. Transcriptional repressors were recruited to the CDX2 promoter in BLM dependent manner. Same as (C) except ChIP was carried out with antibodies against (D) SMAD3, (E) AP2β. Quantitation from three biological replicates.

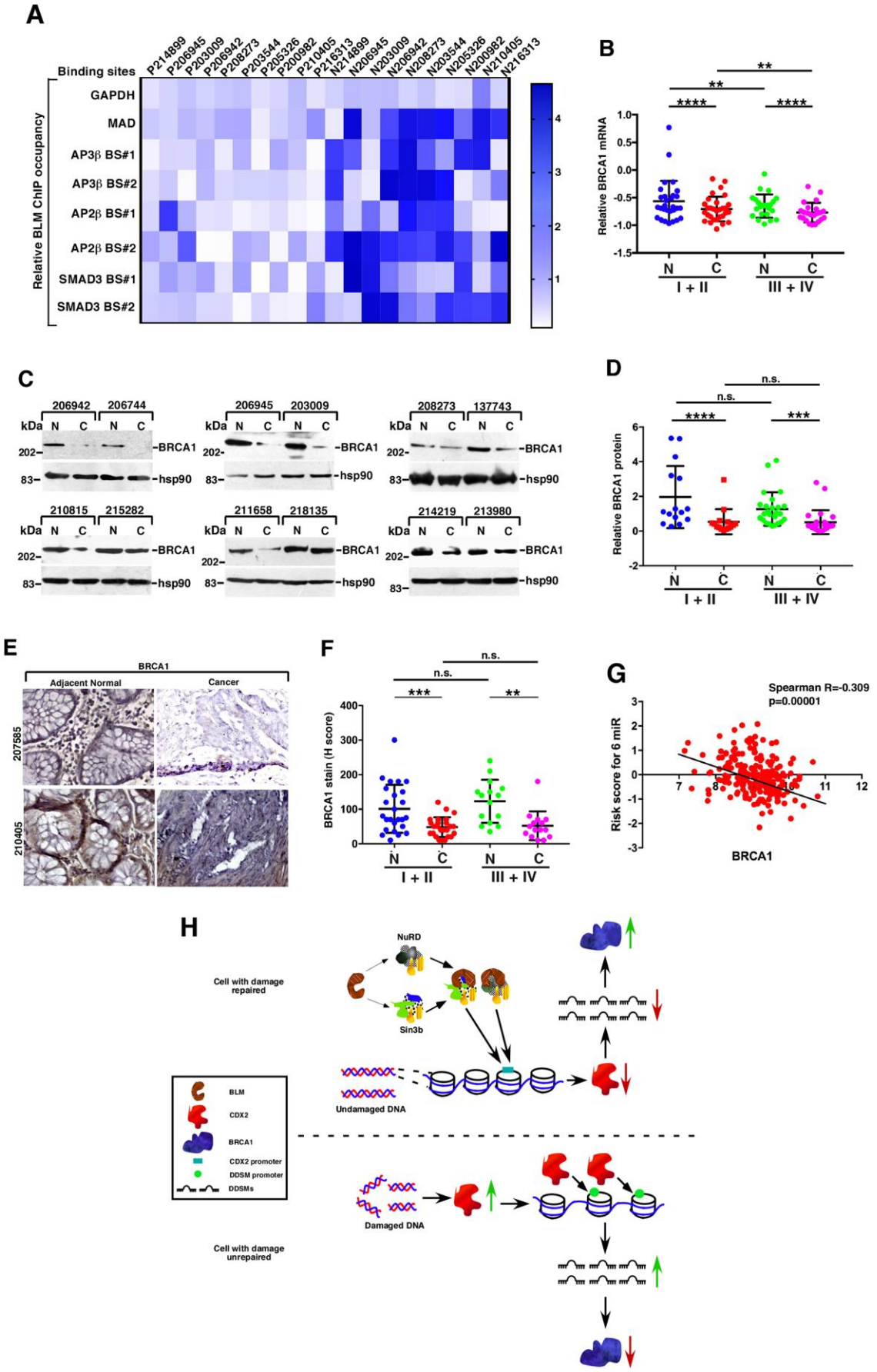
F, G. Co-repressor complexes were recruited to the CDX2 promoter in BLM dependent manner. Same as (C) except ChIP was carried out with antibodies against (F) Sin3b (G) CHD4. Quantitation from four biological replicates.

H, I. Histone deacetylases were recruited to the CDX2 promoter in BLM dependent manner. Same as (C) except ChIP was carried out with antibodies against (H) HDAC1 (I) HDAC2. Quantitation from three biological replicates.

J-M. Ablation of Sin3b and CHD4 enhanced CDX2 transcript and protein levels. HCT116 cells were transfected with either (J, K) shScramble and shSin3b or (L, M) siControl and siCHD4. CDX2 transcript levels were determined by RT-qPCR. Quantitation was from three biological replicates. Protein levels of Sin3b, CHD4, hsp90 were determined by carrying out western analysis with the corresponding antibodies. Three biological replicates were done for both experiments and representative blots presented.

N-P. Ablation of HDAC1 and HDAC2 enhanced CDX2 levels. Same as (J-M) except HCT116 cells were transfected with either (N, O) siControl and siHDAC1; (P) siControl and siHDAC2. Quantitation in (N) was from three biological replicates. For westerns (O, P) three biological replicates were done for both experiments and representative blots presented.

Journal of Cell Science • Accepted manuscript



Journal of Cell Science • Accepted manuscript

Figure 8: DDSMs target BRCA1 in colon adenocarcinoma tissues.

A. BLM was recruited to the CDX2 promoter in the adjacent normal regions. ChIP with anti-BLM antibody was carried out on colon cancer samples and their adjacent normal control regions (n = 10). The amount of BLM recruitment to the different transcriptional repressor sites were determined, quantitated and represented in form of a heat-map.

B. Levels of BRCA1 mRNA decreased in the cancerous tissues of colon cancer patients from India. RT-qPCR quantitation for BRCA1 from RNA isolated from colon cancer tissues (C) and their adjacent normal tissues (N). Quantitation from fifty-four paired samples.

C, D. Levels of BRCA1 protein decreased in the colon cancer tissues in the Indian cohort as detected by Western blot analysis. (C) Representative western blot analysis obtained for BRCA1 and hsp90 from the tissue extracts of colon cancer tissues (C) and their adjacent normal tissues (N). (D) Quantitation from forty paired samples.

E, F. Levels of BRCA1 protein decreased in the colon cancer tissues in the Indian cohort as detected by immunohistochemistry. (E) Representative immunohistochemistry staining obtained for BRCA1 in the colon cancer tissues (C) their adjacent normal tissues (N). (F) Quantitation from forty paired samples.

G. Expression of the six DDSM signature inversely correlated with BRCA1 expression. Spearman correlation analysis was carried out for the transcript levels of six DDSM signature and BRCA1 from the TCGA dataset. The correlation coefficient R and p value have been indicated.

H. Schematic diagram showing the upregulation of DDSMs in colon cancer cells. In cancerous cells or tissues higher levels of damaged DNA led to the upregulation of CDX2, which allowed CDX2 to bind to the promoters of DDSMs. The levels of DDSMs increased which caused a decrease in the levels of its targets involved in DDR response (like BRCA1). In cells or tissues where DNA damage has been repaired or not present (like in adjacent normal tissues), CDX2 expression was transcriptionally repressed as BLM recruited co-repressor complexes (Sin3b and NuRD) to the CDX2 promoter. Lack of CDX2 induction prevented the upregulation of the DDSMs, due to which the levels of the DDR proteins (like BRCA1) remained elevated.

Journal of Cell Science • Accepted manuscript

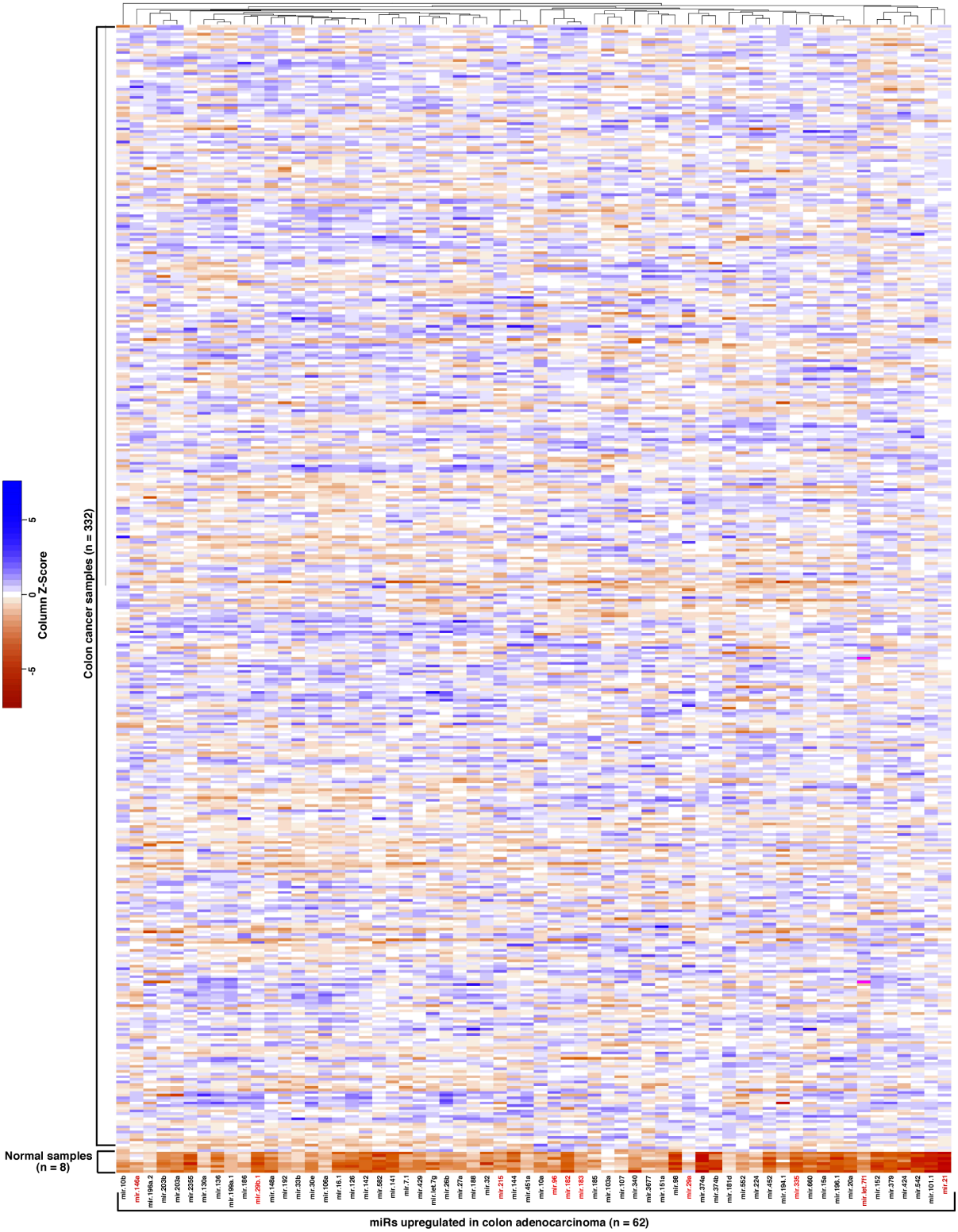
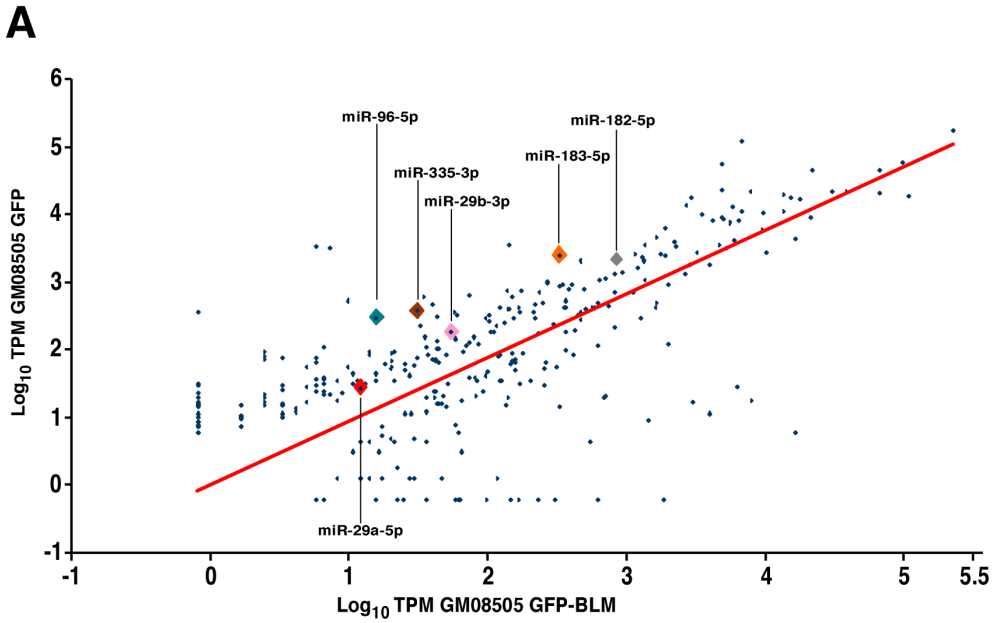
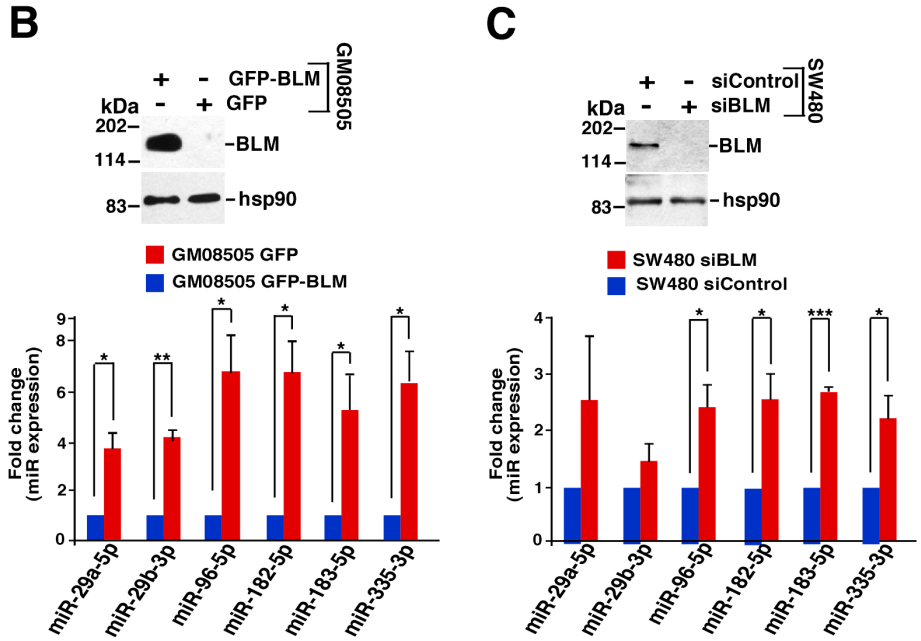

Column Z-Score
5
0
-5
Colon cancer samples (n = 332)
Normal samples
(n = 8)
miR-10b
miR-146a
miR-196a-2
miR-203b
miR-203a
miR-2355
miR-130a
miR-136
miR-199a-1
miR-186
miR-29a-1
miR-148a
miR-192
miR-33b
miR-30e
miR-106a
miR-16-1
miR-126
miR-142
miR-582
miR-141
miR-7-1
miR-429
miR-let-7g
miR-26b
miR-27a
miR-188
miR-32
miR-215
miR-144
miR-451a
miR-10a
miR-96
miR-182
miR-183
miR-185
miR-103a
miR-107
miR-340
miR-3677
miR-151a
miR-98
miR-29a
miR-374a
miR-374b
miR-181d
miR-552
miR-224
miR-452
miR-194-1
miR-335
miR-660
miR-15a
miR-196-1
miR-20a
miR-let-7f1
miR-152
miR-379
miR-424
miR-542
miR-101-1
miR-21
miRs upregulated in colon adenocarcinoma (n = 62)

Fig. S1: Heat-map representation of the miRs upregulated in colon cancer samples.

Representation of miRs upregulated all across in different stages of colon cancer tissues (n=332) as compared to normal samples (n=8). DeSeq2 analysis using the R package was carried out on the 647 miRs to identify the miRNAs (n=62) which were upregulated in all stages of colon adenocarcinoma tumor samples as compared to the normal samples. The 10 miRs highlighted in red have been characterized as DDSMs based on the targets being responsive to DNA damage response (DDR), DNA repair, DNA replication, apoptosis, chromatin remodeling and colorectal cancer. Six of the DDSMs were chosen for further analysis.

Journal of Cell Science • Supplementary information


A
miR-96-5p
miR-182-5p
miR-183-5p
miR-335-3p
miR-29b-3p
miR-29a-5p
Log10 TPM GM08505 GFP
Log10 TPM GM08505 GFP-BLM


B
GFP-BLM
GFP
GM08505
kDa
202
114
83
BLM
hsp90
GM08505 GFP
GM08505 GFP-BLM
Fold change (miR expression)
miR-29a-5p
miR-29b-3p
miR-96-5p
miR-182-5p
miR-183-5p
miR-335-3p
C
siControl
siBLM
SW480
SW480 siBLM
SW480 siControl

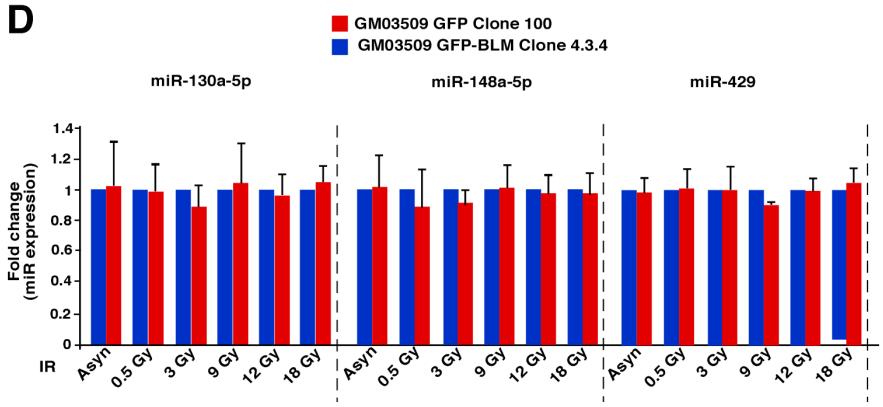

D
GM03509 GFP Clone 100
GM03509 GFP-BLM Clone 4.3.4
miR-130a-5p
miR-148a-5p
miR-429
Fold change (miR expression)
IR
Asyn
0.5 Gy
3 Gy
9 Gy
12 Gy
18 Gy

Fig. S2: Validation of the DDSMs.

A. Alteration in DDSM levels in absence or presence of BLM in GM08505 cells. Small RNA sequencing was done in isogenic cell lines derived from BS patient, GM08505 GFP-BLM and GM08505 GFP. Six DDSMs of interest have been indicated.

B, C. Validation of DDSMs whose expressions were increased in absence of BLM. Two isogenic pairs of cell lines (B) GM08505 GFP-BLM and GM08505 GFP and (C) SW480 siControl and SW480 siBLM were validated by western analysis with antibodies against BLM and hsp90. DDSMs which were upregulated in absence of BLM in (A) were validated by RT-qPCR analysis. Three biological replicates were done for protein analysis and RT-qPCR validation.

D. Levels of non-DDSMs did not increase after IR exposure. Isogenic cell lines, GM03509 GFP-BLM Clone 4.3.4 and GM03509 GFP Clone 100 were exposed to a gradient of IR. RNA was prepared 1 hr post-exposure. Expression of indicated the non-DDSMs were validated in both the isogenic pair of cell lines by RT-qPCR analysis. The quantitation was performed using three biological replicates.

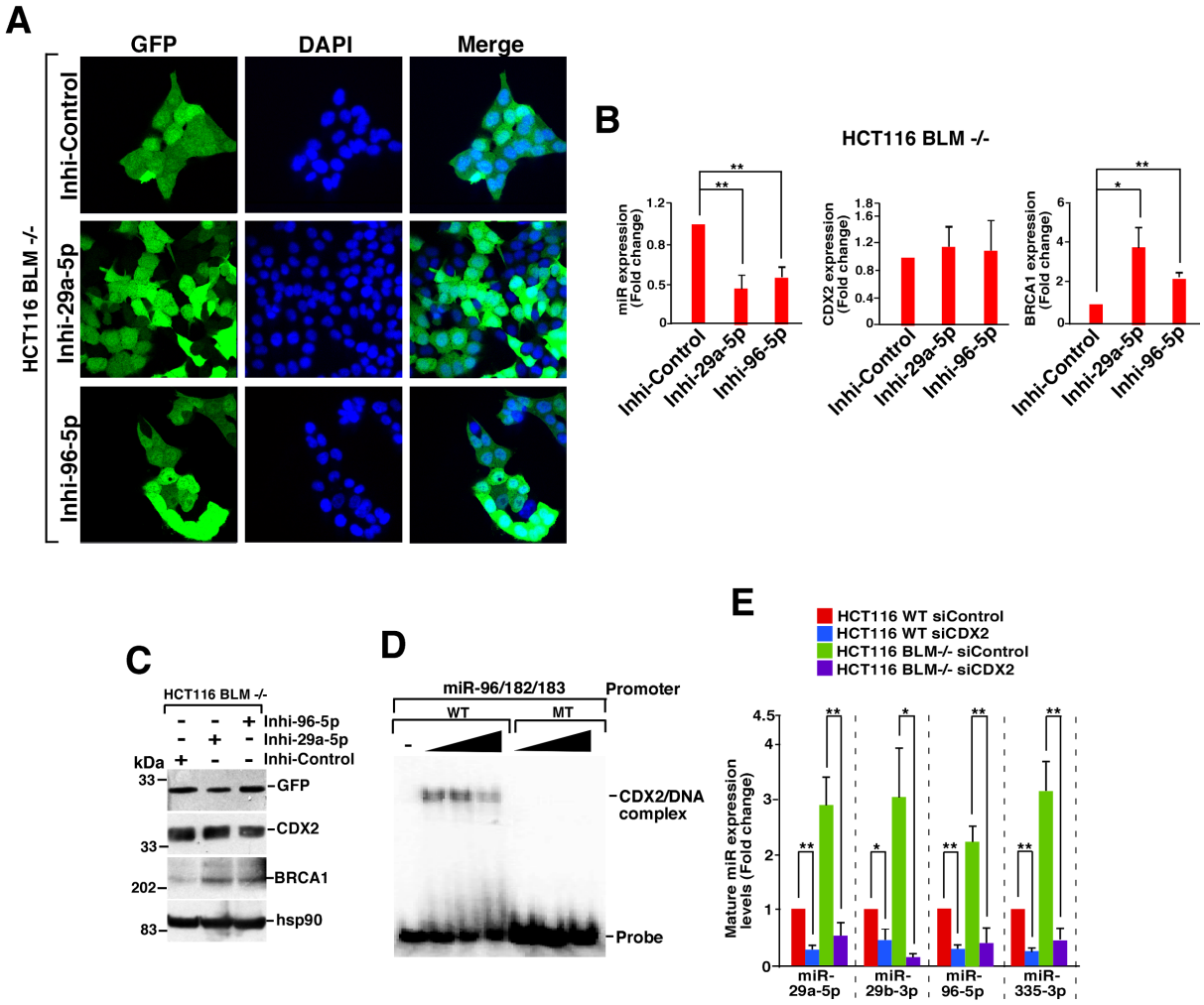


Fig. S3: CDX2 regulates DDSMs.

A-C. Characterization of HCT116 BLM -/- cells expressing inhibitors to the DDSMs. (A) HCT116 BLM -/- cells stably expressing Inhibitor Control, Inhibitor miR-29a-5p or Inhibitor miR-96-5p were stained with antibodies against GFP. Nuclei were stained with DAPI. (B) Levels of the three DDSMs, CDX2, BRCA1 were determined in HCT116 BLM -/- cells stably expressing Inhibitor Control, Inhibitor miR-29a-5p or Inhibitor miR-96-5p by RT-qPCR. Quantitation was performed using four biological replicates. (C) Lysates were made from HCT116 BLM -/- cells stably expressing Inhibitor Control, Inhibitor miR 29a-5p or Inhibitor miR 96-5p. Western analysis was done with antibodies against GFP, CDX2, BRCA1, hsp90. The experiment was done three times and representative western blot from one of three biological replicates is presented.

D. Recombinant CDX2 bound specifically to its recognition sequence on miR-96/182/183 promoter. Radiolabelled double stranded annealed oligos containing either CDX2 binding site (WT) or where CDX2 binding site was destroyed (MT) were generated. EMSAs were carried out in presence of recombinant CDX2. CDX2/DNA complexes were visualized by autoradiography. The experiment was done three times. Representative EMSA from one of three biological replicates is presented.

E. Ablation of CDX2 led to decreased levels of mature DDSMs. HCT116 WT or HCT116 BLM -/- cells were transfected with either siControl or siCDX2. RNA isolated from each of the cell types was quantitated by RT-qPCR analysis for the levels of the indicated mature DDSMs. Quantitation was performed using three biological replicates.

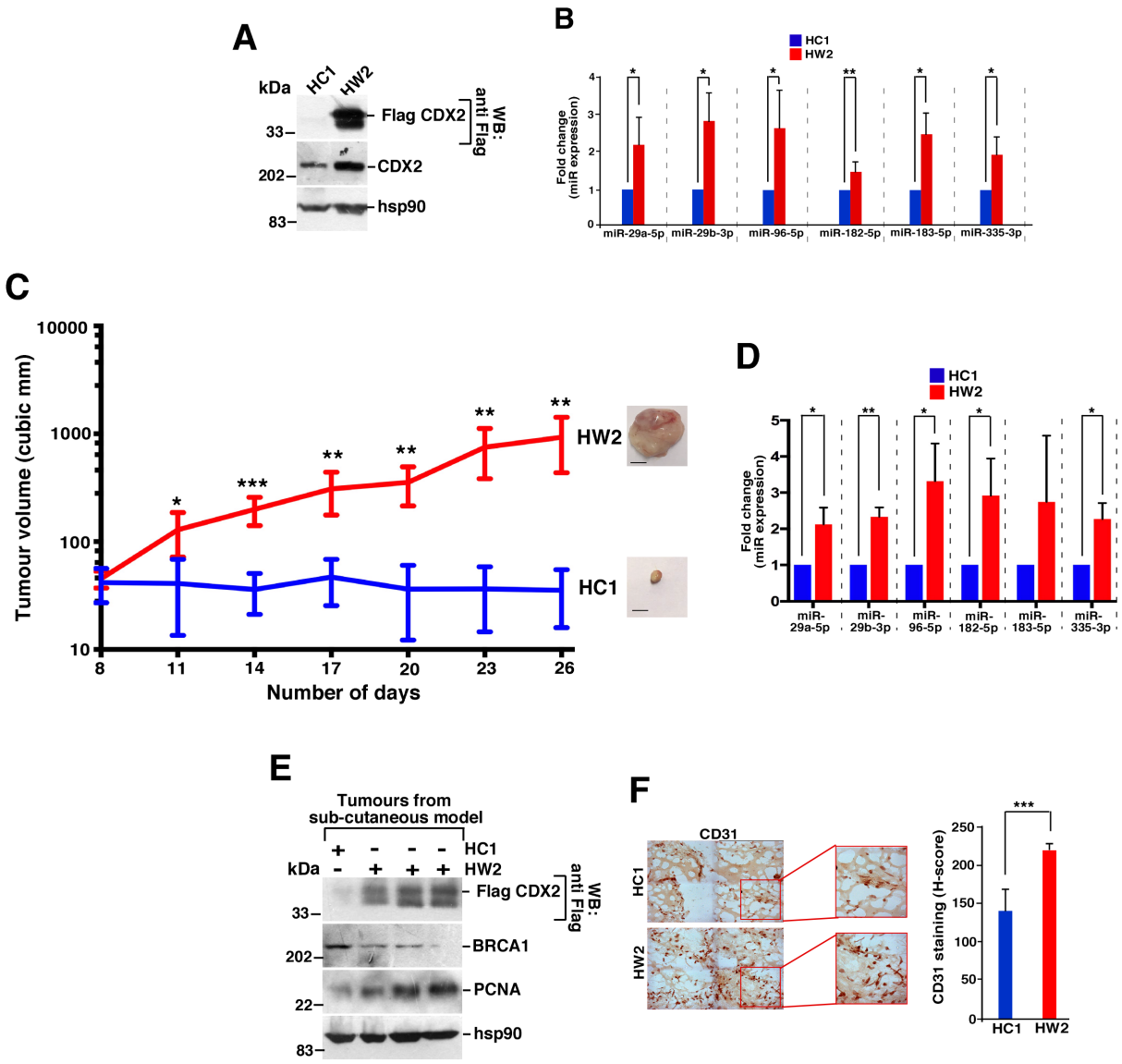


Fig. S4: CDX2 inducible DDSMs cause tumour sustenance.

A. Flag tagged CDX2 was overexpressed in HW2 cells. Lysates were made from HC1 and HW2 cells. Western blots were carried out with antibodies against Flag, CDX2 and hsp90. The experiment was performed using three times and representative western blot from one of three biological replicates is presented.

B. DDSMs were induced by CDX2 expression in HW2 cells. RNA was isolated from HC1, HW2 cells. Levels of the indicated miRs were determined by RT-qPCR analysis. Quantitation was performed using four biological replicates.

C. Overexpression of CDX2 led to enhanced tumor formation in mice xenograft model. Tumor formation in a mice xenograft model was carried out by injecting HC1 and HW2 cells subcutaneously into NOD SCID mice. Tumor formation was monitored over the period indicated. Six mice were used for each condition. One representative tumor for each condition has been represented. The statistical significance was determined by comparing tumor formation in HW2 cells compared to HC1 cells.

D-F. Induction of DDSMs, proliferation and angiogenesis markers in CDX2 overexpressing tumors derived in a xenograft model. (D) RNA and (E) protein were extracted from the tumors at the end point of the xenograft experiment (C). (D) Levels of the indicated miRs were determined by RT-qPCR analysis (RNA was from four different mice injected with HW2 cells). (E) Levels of CDX2, GFP, PCNA, BRCA1, hsp90 protein levels in the tumors were determined by carrying out western analysis with the corresponding antibodies (extracts from 3 different mice injected with the HW2 cells were used). (F) IHC was carried out with tumor sections with anti-CD31 antibodies (from 6 different mice in each conditions). (Left) Representative images of CD31 staining. (Right) Quantitation is presented as H-score.

Journal of Cell Science • Supplementary information

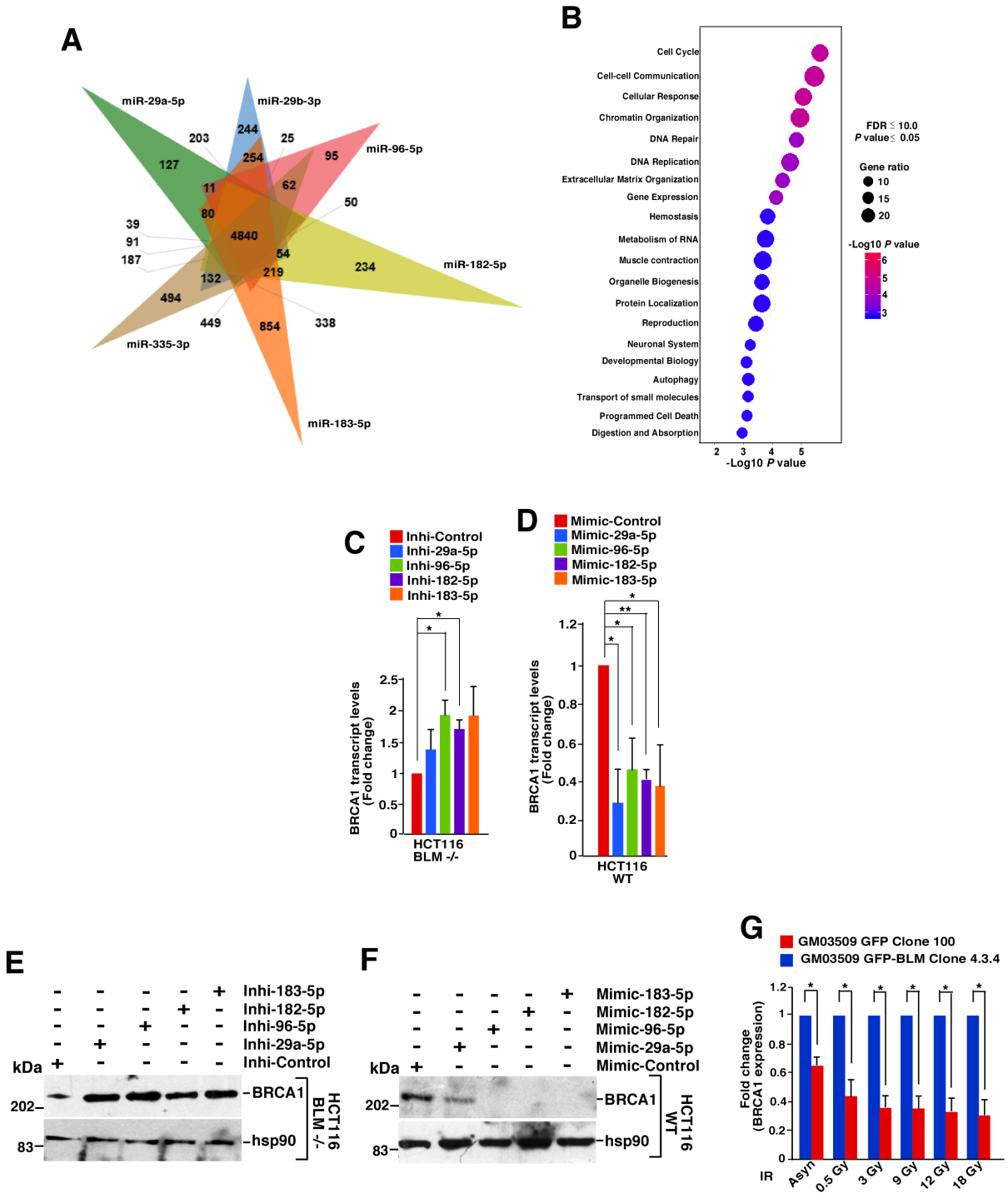


Fig. S5: DDSMs target DDR proteins.

A. Common targets of DDSMs. Using miRanda the targets of each of the DDSMs were determined. The Venn Diagram showed the common targets of the DDSMs.

B. DDSMs target multiple cellular pathways utilizing their targets. All the DDSM targets (as determined in A) were subjected to analysis using Reactome, KEGG, Wiki Pathways and GSEA online databases. The number of enriched genes in each pathway is indicated by the size of the circle. The color of each circle represents the -Log10 p-value for each of the pathway as determined by Fisher Exact Test.

C. Transcript levels of BRCA1 were enhanced upon inhibition of the DDSMs. HCT116 BLM -/- cells were transfected with either inhibitor Control or specific DDSM inhibitors. Transcript levels of BRCA1 were determined by RT-qPCR. The statistical significance was calculated relative to Inhibitor Control. Quantitation was performed using three biological replicates.

D. Transcript levels of BRCA1 were decreased upon overexpression of the DDSMs. HCT116 WT cells were transfected with either mimic Control or specific DDSM mimics. Transcript levels of BRCA1 were determined by RT-qPCR. The statistical significance was calculated relative to Mimic Control. Quantitation was performed using three biological replicates.

E. Protein levels of BRCA1 were enhanced upon inhibition of the DDSMs. Same as (C) except protein levels of BRCA1 were determined. Western blots were carried out with antibodies against BRCA1 and hsp90. The experiment was done three times and representative blots presented.

F. Protein levels of BRCA1 were decreased upon overexpression of the DDSMs. Same as (D) except mimic Control or specific miR mimics were used to transfect the cells. The experiment was done three times and representative blots presented.

G. BRCA1 transcript levels were decreased in cells exposed to IR. BRCA1 transcript levels in GM03509 GFP-BLM Clone 4.3.4 and GM03509 GFP Clone 100 cells, exposed to different doses of IR. RNA was extracted 1 hr post-IR and BRCA1 levels were determined by RT-qPCR. Quantitation was performed using three biological replicates.

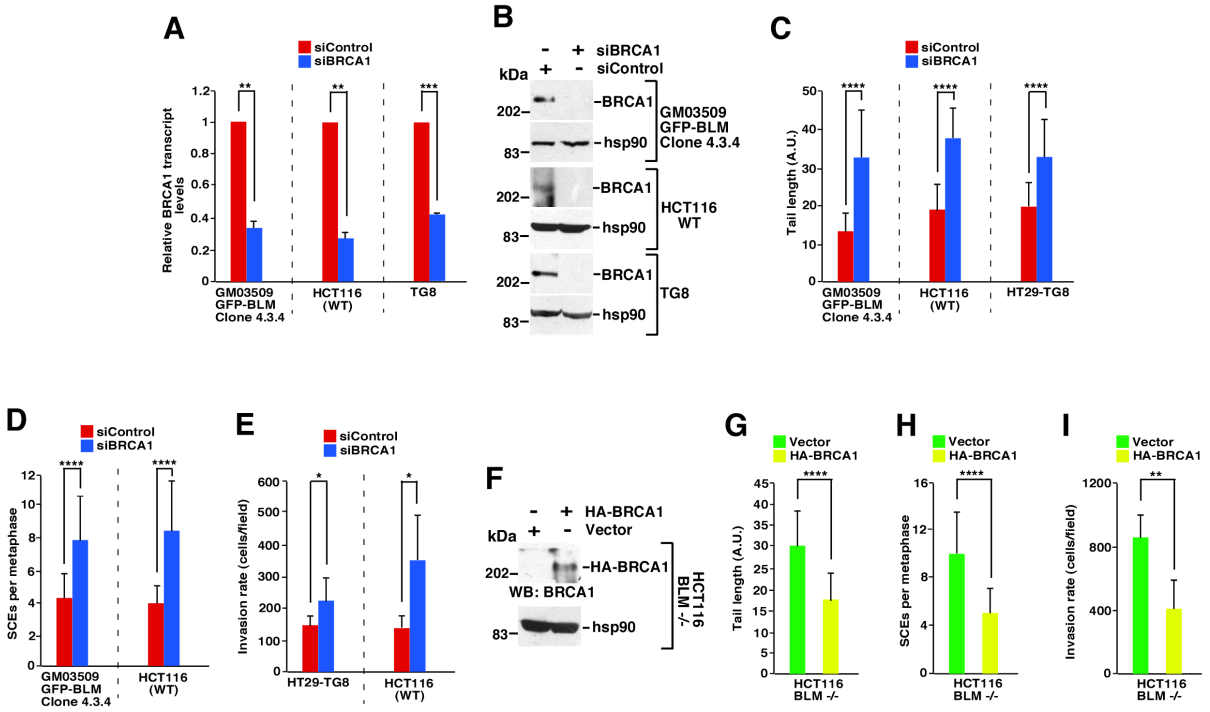


Fig. S6: Levels of BRCA1 regulate DDR and invasive capability of cells.

A, B. Depletion of BRCA1 levels. GM03509 GFP BLM Clone 4.3.4, HCT116 or TG8 cells were either transfected with siControl or siBRCA1. The levels of BRCA1 were determined at the (A) RNA level by RT-qPCR or (B) protein levels by western analysis with antibodies against BRCA1 and hsp90. Quantitation was performed using three biological replicates.

C-E. Ablation of BRCA1 increased the amount DNA damage, SCEs and invasion. Cells obtained from (A, B) were subjected to (C) Comet assays (n=21 comets quantitated for each pair of cell types) (D) SCEs (n=21 spreads quantitated for each pair of cell types) and (E) invasion assays (n=6). Quantitation of each assay is presented.

F. Overexpression of BRCA1 levels. HCT116 BLM -/- cells were transfected with HA-BRCA1 or Vector. The lysates were probed with antibodies against BRCA1 and hsp90. The experiment was done three times and representative blots presented.

G-1. Overexpression of BRCA1 decreases the amount DNA damage, SCEs and invasion. Cells obtained from (F) were subjected to (G) Comet assays (n=36 comet tails quantitated) (H) SCEs (n=21 metaphase spreads quantitated) and (I) invasion assays (n=6). Quantitation of each assay is presented.

Journal of Cell Science • Supplementary information

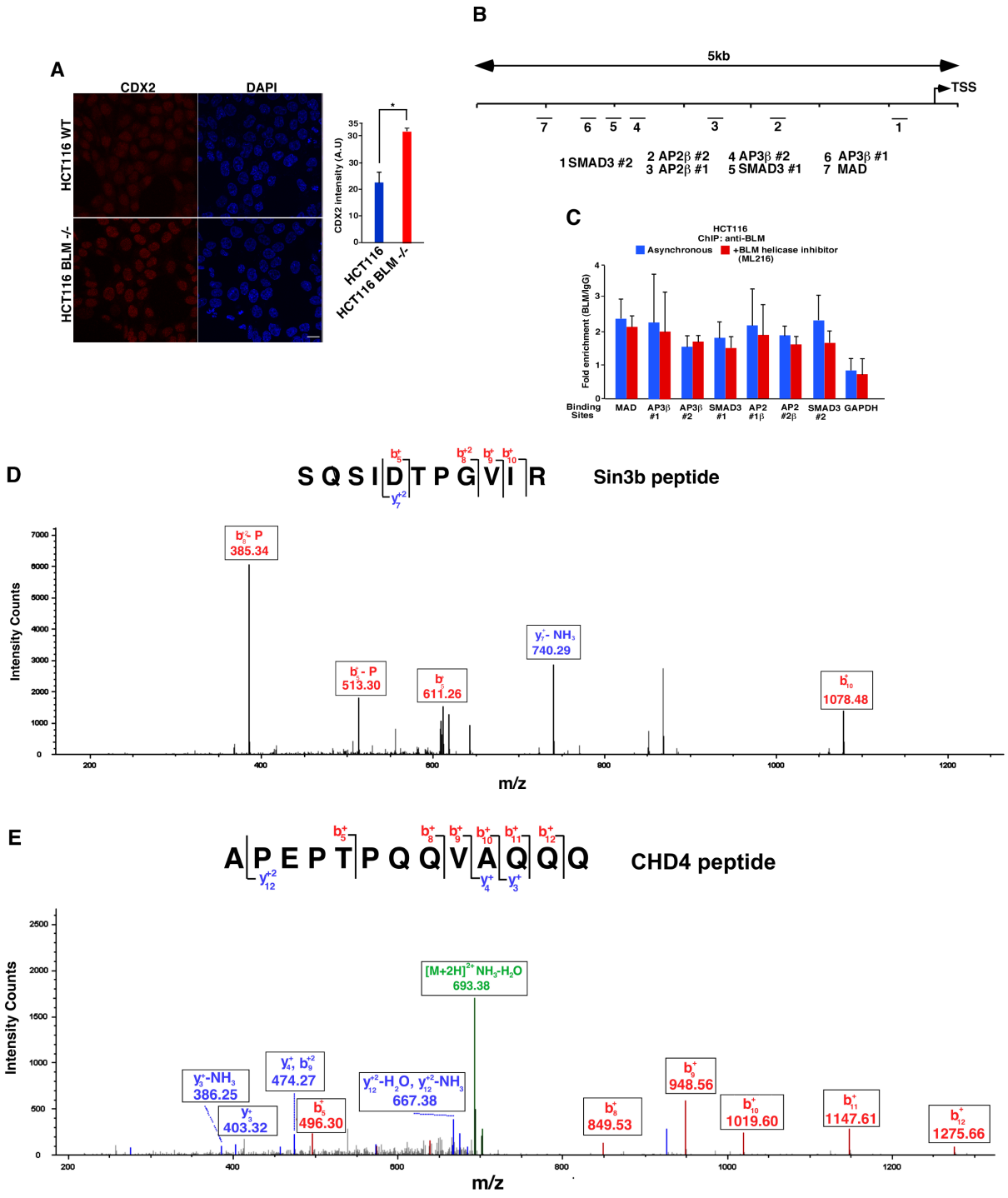


Fig. S7: Absence of BLM increases CDX2 levels.

A. CDX2 levels increased in absence of BLM. Immunofluorescence was carried out with HCT116 WT and HCT116 BLM -/- cells (n=49 cells quantitated for staining). Staining was done with antibodies against CDX2. DNA is stained with DAPI.

B. Schematic diagram of CDX2 promoter. Approximately 5kb upstream to the TSS in CDX2 promoter was analysed. Binding sites for the different transcription factors are indicated.

C. BLM recruitment to the CDX2 promoter was independent of its helicase activity. BLM ChIP were carried out with chromatin isolated from HCT116 WT cells grown in either asynchronous condition or in presence of ML216. The recruitment of BLM to the putative binding sites of the transcriptional repressors has been shown. The corresponding IgG was used as the antibody control. As specificity control, the recruitment of BLM to the GAPDH promoter was also determined. Quantitation was performed using four biological replicates.

D, E. Identification of BLM interacting proteins by mass spectroscopy. The identified peptides for (D) Sin3b (E) CHD4 in the BLM immunoprecipitate are depicted. Fragment series (b- and y-type ions are labelled) confirmed the sequence of the indicated peptides.

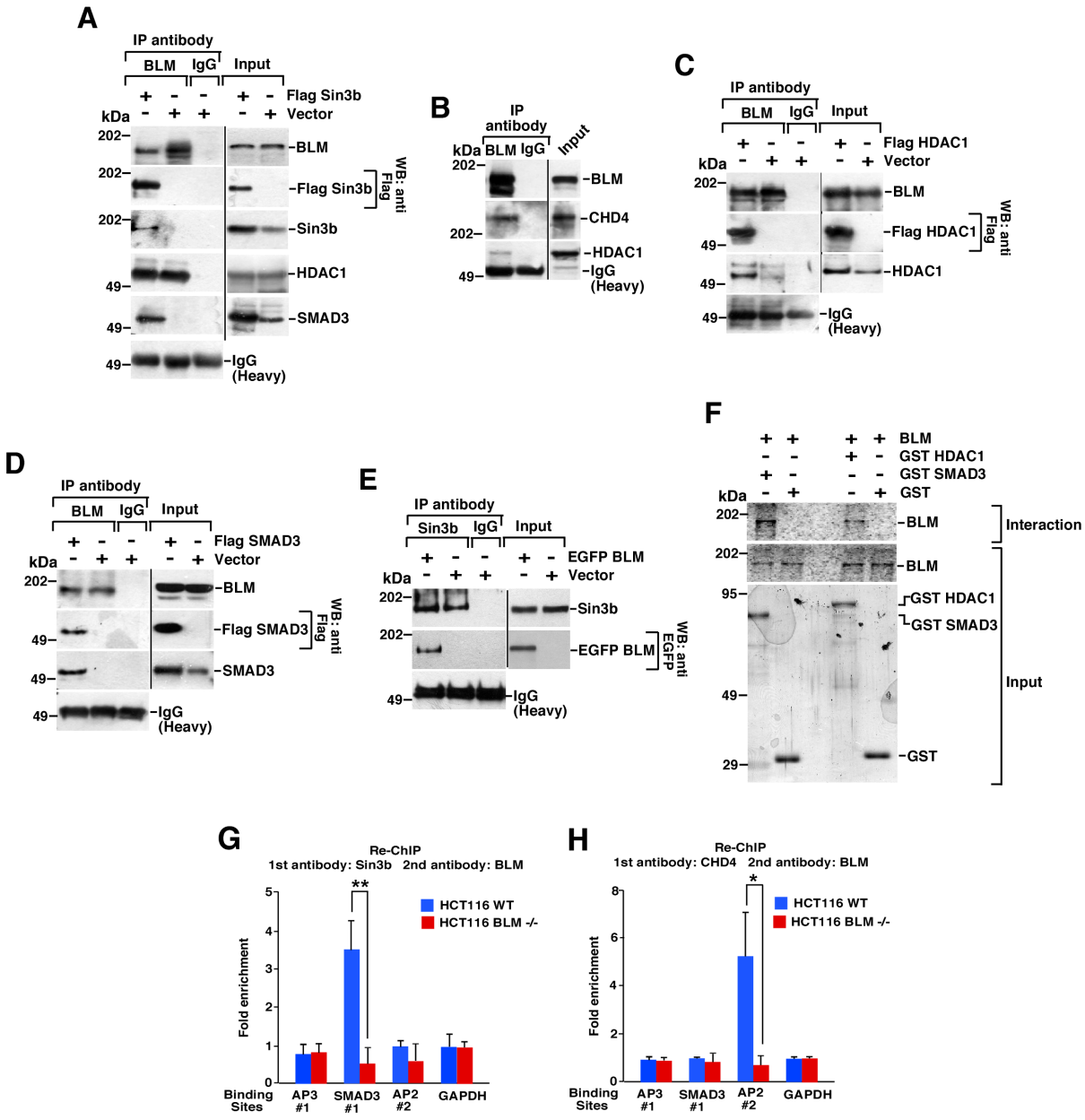


Fig. S8: BLM interacts with members of Sin3b, NuRD complexes and SMAD3.

A-E. BLM interacted *in vivo* with components of the Sin3b and NuRD co-repressor complexes and SMAD3. HCT116 cells were transfected with the plasmids indicated. Reciprocal immunoprecipitations were carried out with antibodies against (A-D) BLM or the corresponding IgG or (E) Sin3b or the corresponding IgG. The immunoprecipitates were probed with antibodies against (A) BLM, Flag, Sin3b, HDAC1, SMAD3, (B) BLM, CHD4,

Journal of Cell Science • Supplementary information

HDAC1, (C) BLM, Flag, HDAC1, (D) BLM, Flag, SMAD3, (E) Sin3b, GFP. Each experiment was done three times and representative blots presented.

F. BLM directly interacted with SMAD3 and HDAC1. (Top) *In vitro* interactions were carried out with S^{35} methionine radiolabelled BLM and bound (left) GST and GST SMAD3, (right) GST and GST HDAC1. Post-interaction the bound radioactive BLM was detected by autoradiography. (Bottom) Inputs used for *in vitro* translated BLM (detected by autoradiography) and GST, GST SMAD3, GST HDAC1 (detected by Coomassie). The experiment was done three times and representative blots presented.

G, H. BLM was bound with Sin3b and CHD4 on the CDX2 promoter. In HCT116 WT cells, sequential re-ChIP assays were carried out with (G) Sin3b, BLM and (H) CHD4, BLM combinations. Three indicated binding sites of the transcriptional repressors were chosen to check for the recruitment. The corresponding IgGs were used as the antibody controls. As specificity control, the recruitment of BLM to the GAPDH promoter was also determined. Quantitation was performed using three biological replicates.

Journal of Cell Science • Supplementary information

Table S1: Distribution of transcription factor binding sites on the DDSM promoters

miRNA	CDX2	ZNF263	TCF12	Ini1
miR-29a-5p	1	0	0	0
miR-29b-3p	1	0	1	0
miR-96-5p	1	1	1	1
miR-182-5p	1	1	0	1
miR-183-5p	1	1	1	1
miR-335-3p	1	1	1	1

"1" indicates the presence of one or more transcription factor binding site while "0" indicates the lack of any binding site on the promoter of the indicated DDSM. Only factors which bind to the promoters of atleast four of the six DDSMs are represented.

Table S2: Information regarding CDX2 binding sites on miR-96/182/183 promoter

CDX2 binding sequences: TTAATA, TTAACA, TTTACA, TTTATA

miR-96/182/183 promoter

(hg19_dna range=chr7:129418617-129421925)

CGAGAATCCGCCACCTGGCTCGCAGGATTTTCGCCCTCAGGAGCTTTGCG
AAGACGGGCGCAAACCTGGGAGGCGCCGCCGCACCTTGCCTGTGGGGGTG
GGGAGGTCGATTACGGCACCGCCCGCCCTCGGCCCTAATCGCTCAGAGCC
GGCCTCAGTGCCCCTGCCCCCGGCGGGTATTACTGTACACGGTGTCCGGG
AGCCCGGGTTGGAAGAGCAAACGAATGGCTCAGCCTGGAGGACACCGTGG
AATCGGGCGACCCCTGACCCCGGGTGACACCCTCACTGCCCATAGTCCTG
GGATGCGCTGACCGGTGAGGCGCCAGAGGCAAGCCGTCCAGCCGCCGATC
TTGGGGAAGGGCTGGGAACCTGGGTGCCACTAGCGCTA**TTAACA**GGAATG
GGGCAGTTGAGGGCGGGGAACGGTGTCTCATTCGCGTGCGCCGGGGTTCC
CGGGATGAGTGGCCGCGAAGTGGAGGCACACGGGAGGTCTGAGAGCCGCT
TTCCCCCAGAGACTGGGCGGGGCGGCTCCGGGCCACGGAAGCAGCCTCGA
ACCAGCCGCGCAGCCCTGACAAGACGCTGGAGGGGCCAGCTCCGGGGCCC
CGGCGCTCTCCGCAGCCCCCTCCCAGCCTGGCGCTTCCGCCTGAGCCCAC
CAGGGCGGCGAAACGCAGCCCGACCCCAGCCAAACTCCTCGCGGGTAGGC
ACCTCGCAAGCACCGGCCGCACTCGGACCGCAGCCGCTCCCGCCCCCGCC
CCCGCCCGGGAACAGCGCTGGGCGCTGGGCAGACGCCACTGGCGGACCCG
GGACCCAGGCGCCCCGTGGGCGCCTCCGCCTGAAGGCGCGGAAGAGCGAG
TAGGCACTGCCGCGTCATCCCCGAAAGCCAGCTGCGCCGGGACTTCTGCC
CTGCACCCGGGTTCTAGTGCGGCCGGTCCCGGGCGCCCGCCTCCGCCTCC
TCTACTGAGCCCCGCTCTGGCTCCGCCCGGCCCGCCCCCCCGCTCCCGGC
CCCAGACACTCACCTGCGCCGGCTCCAGGCTGGGCGCACAGCCGCCTCTC
AGCTCCGCGCGGAGTCCACACCAGAGAAGCCAGGCAAAAACGGCCCCGGT
TTGGAAAGAGCCGCATGGACGGAGCTTGAAGCGCGCCCGGCCCGGGCCCA
AACCTCCGCTGCCCCGCGAGAGGGGGCGGCGCCGGAGCGCCGAGCCCAGC
GGGGGAACAAAGGGGAGCCGGGACCCGGCCGCGCGCTTCCTGGGTGTCGG
GTGACAGGTCTGCCGGGGAAGGCCCCGGCCCGGCCACCTGAAGAGGAGCC
GCCTCCGCGCTGCAGGCGGGGGAGGTCCAGAGGGGTACCTGGGCCCGAGG
ACCGGAAGAACAGGGGGCGCGGGCTGCCCGGACCGCGCGGCTGGACGGAG
GCCCTGCGGCTCCGGCACGGCTGCGCGGGGGCTCGACGGAGTGCACAGAG
AATGACTCCAGGAGGGTTTTACCGCCCCCCGCCCCGGGCCCCGCCCGCCC
CTCGCGCAGCGCCATTGGCTCAGGTGCCGCGAGGCGGGGAGGGGCAGGGG
GCGGGCCGGGGATGATTCACCTTGAATGTGGCTCCGGTTGGTGCGCGGCG
CAGGGCACGCCCGTCCGGGACTGCGGAGACGCCGGGAGCCCCGCCTGCTT
GCCCAACCCGGCCGGGTCTGGCCCCAAGGGGTCCCGTCGCCCGCCTTCTA
TCCCCGCCTCCTGCCCTCGGGCTAGGGTGCCGCTAGCCGAGGTGCGCCCG
GACGCCCTGCCACCCCAGCGCCCTCTGAGCCCGGAGCTAGCCGAGGTGTG
TGTGGCGTGGCGTGCTTGGAAATCCGACCCTGGGAGGGGGGAGGGGAAGT
GCGTTGGGAGCTGGGGCGCACCGTAGGGCCACTGGACGATCTCTGCAGTC
TCCCCGCAGAAAGGTGCTCAAGGCTGACCTGTCTCAGGGGGCGCACCCGG
ACCCAGGGGCGCAGGGCTGGAATGGGGACAAGGGGCCTCCATGTCCTCAC
ACGCCCACCCTCGGCCAGTTCTGCGGCCTCACAGGTGTCTGGCCACCGCT
GCCGCCGCCGCCGGGAACAAAGGCTGTCTGTCCCAAGAGAGGGAGGGGGA
GGGGAGCCGCCCCCCAGACCCGTTCCCATCTCCCCCTCCCCCCACTGCGA

CAGTGGCTCCCTCGGGAGGGGGCGGGGCCAGCCGGGGCTGGGGCGCGGAC
GCGGATTAGATCCTTCTCAACCGTCCTTTGGTCTGACACCACCCGGAGCA
TTCAGAAGGGACCGGGGGGCAGGGGCTAGGTCCTCCAAGCTGTCAGCGTT
GGGGTGGCCTCTTCCTTCTTCCCCAGCATCCTGAGCACCGGAGCGCCCCC
TCCCTCCCTGCTGCTTTGGCTGCCCTTGGCAGTCCCAGCTCCGCCCCGCC
CGGCCGCCCCAGACGGCAACGGCAGCTCAAAACCAGATTAAAATAAAACG
TTAATCCGCCCTTCTCCTTTTTAGCTTCTAGTTCTTCCCAATCGACCCCC
GCAGATCTGGCACCCTAGAAGCCCCTGTAAAATGAGGGTGGTGCTTAGCC
TATGCTCCCACCAAACCTAACCCTTGCACTTGAAGACCG**TTAACA**CCTTG
TCTTGGGCTCGAGCAAAAGAAGGGGGTGAGGCAGTGGAAGCGGGGCCCGG
GAAATCTATAGTTTCCCCACCCCCAGCTCCTGCCGAGCCGGGAGGGGCCA
AGGGTGACGGCAAGGATACGAACCCTTCAATATGCCCGACTCCCGGGCAG
GGCCTGCTGAAGCCGTAGATGTCGCAGCCAGATGGATCGGGAAGCAAGAT
TCCCACCCCAGGTGGATAACTGGACGAGTCTCAGGGCCTCGGGGCCCCTC
TCCCCATGCAGCCAGGGATGGGCACGGTGAGCGAGAACCCGGACAGCGGA
GATAATGAGCACGTCACAGAGCCCAACGGGCCAAAAGGGGGTGATAGCTC
CACATTAACCCCTCCCCGTCCTACTAGCATTCCCCCGCCCATCTAGGTCG
CCTCGAGCCTCCCTGGCTGTCTCTGCACCCGCGCCCTACCTGGTCCTCAG
ACCCATGCGATGCTGGGCAACCCGGGCCGGGTGTGCCCAGGACCGCAAAG
CGGGAAGCGCAGGCCTTGCGGACTGCGAGTGGCCCCCGCGGCCTGCTGCT
CCCC**TTTATA**ACGCGCGCCCCGCCCCCCGCGCTCCGGGCGCCCCGCCCTG
GCCCATTCACCTGCTGTCGAGAACCAGACACCGCGCAGCAGAAAGCTAGC
CGCTGGGCCGCGAGGTCCGGCCGTCGGAGCGGGCGGGGCTGGGGCGCCGCCGAG
AC

Highlighted region (556bp) – cloned in KpnI + HindIII sites in pGL3 basic vector

Table S3: Information regarding CDX2 binding sites on miR-335 promoter

CDX2 binding sequences: TTAATA, TTAACA, TTTACA, TTTATA

miR-335 promoter

(hg19_dna range=chr7:130128665-130131907)

CTAACCAGCATAATAGAGTAGGATTAGCTTATTTTTGGTACTTCCCCTCA
GTCTGCAGTCACAGTAGCCTTTTTAAGGGTCCTGTCTTACTGATTCAGTA
AGTTTCTCAACCAAAACCCCTTCG**TTAACA**TGTGCTCCTGTTCAACCAAG
TTTCTCCCATTATGAAGTGTTTTAGTTGATTTGAATCTACAATGATGGCT
CTTGAAAGTGGCAGTCCCGGTTGTGTCTTGTTGTCTCCATAAATATCCTG
CACTTGTCATATTATGGGTCTCAAACTGGGGTTAGAAAAATATGGTTCTT
GAATCTGTGGTACCATCTCTGCTGATCCCATACCACTGTTGTGAATCAGC
GAGATAATAAATGCAGAGCTTTCTACATGAATTTTTGCACAGATCTGAGG
TCTTATCATGAGGCGGAACTGTGTTAGACTTGGAATGTGACAAGAATGTG
TGATAAGCCACTGGCTGAGGGAGTACTACTGCATGAGATATCCTGCACAG
TCCAGGATAGCATCTGTCCAGAGGAACTGAAAGGGTGCAACTTCGAGCAA
GCCAGAGCCGACTTTGAATTTATGTGTGGCTGTGTGTTTTCTCAGGGCTC
CAAGTGT**TTTATA**CCTTGGGTCAAATTCTCCCCCTGCGGAGGAAGCTCCT
TTCTGGAAGTAAATCAGGCACATTTAGAAGATTCGTTAACTCTTTGGCTC
CTGGGGGATTATGGCCAGCATTTTCTCTCTTAGACTTCCCACAGCGAATT
CCTGGTTGGGCCAAGCTGATATCTTTGAGCTGACTTCTGTACTCAGGGTT
CTTCCAAACAGGAATATGCTTTCTCTGCCTAAGATCTGGGTTCCATGCCC
ATCTCAACCTCAATCCCTCCTTCAAGGAAGTTTTTCCTAACTCCTCGGCC
AGACATCAATAATGTCTGAAGGGAGGATTTCTCCCGAGTGCCAGACTTGG
AATTCCTGTTGTCTTCAGGAGCTCTCCCTGGCGCTCCCTCCTTCCCCCAG
GGCCTTTAAACTGTTTCTCCTCCAGTCTGCCCACTTGATCCATGTGGGTA
GACATGTTCCATGGCCTTCAGAATAAGATCTTTTGTGGGGTACTCAGTCC
ATAAACAGAAAATTGCCTTCATCTTGGATGCTGGTGTTTCTGGCCTGGTG
AGAATGAAAGTGTGGAATTTGTTTGTAGGACTCAAAGTTTCCA**TTAACA**A
TAAATATGAAAAATATATAATGCAAAGGCTTTTGTTATTGTTATTATTAA
AATAGGAAAGCCTTTACCTTTTTGGAAGAACCGATTTAGATTTTGCAGAG
TAATTCAGTCTTTAAAAGGTAGTGGTAGTTAAAATAGGTAATCTTTGACT
TGTAAAGAGACTAATTTAAATTTTGTTAAGTAACCAATTAAGGCCATAAA
GAGTTTAAATCCATCTGTGGCTTCTCATTTTCTCTAGAACAAATTCCTAA
TTCCCTACGATGAAATTCTCTTGCTATCCTAATTTTTTGCCGGCTCAAAG
GACTAACCGTTTTCCCCAAACACGCGGCCCTGAGCCTCCGGTGCTTTGCG
AAAGCCGTCCCCTCCGGCTGGAATGTTCTTTGTCTCTCTGGTAAATGTCT
ATTCATCATTCAAGACCACCCTTCGCCCCTCCCCTTAATTACTTCCTCCC
CTGTGCGCCCTCAAATTCTTGTACTCATTCAAGCAGTATTTATTAGGGGC
CAGCTTTGTGGCCGGCACCGTGGCGGGCTCTGGGGCTACAAAAGGTGAAT
AATACTTGGGCTCTGCCTCTGAGGGCCTTACACGTTAGGGAGGAGTGGGT
CAACTGCCACAAACGTCGCTAGGAAATTAAAAGGAAAAT**TTTACA**AAGTG
GCAGTTCTTGTCTGTCTTCCCCTCCAGATGGCCCGTGTGTTGTTTTCGGG
CCGGGGCTATTTCTCATTTATTTCGCACCCCCGGTTCTTAGTGCCCTGTA
GGTGCTAAATCAGCATTTGTTTCATGAGTGCTTTTTCTGGGGGCAACCAG
ACCCCTGCAGAAGTGTACCTGTGTTGTGCCAGAGGTTCTGATGATAGGCT
TATAGGCGGTAGTTTCCTCAGTGTCCGTGGGTCGCCCCCGGTCCCGGGTT
GGATGCCCCGCGGTCCAGCACCGAACCTTTCGGGGTGCAGAGTTGCAGAG

CCGCGGAGGGCCCGGGCCGTGCGCAGCCGAAGGGAGGCCTGCAGCGCCCC
CTCTGGATGCAGCGGGCACCGGCCGGCCGCCCCGCTCACCCGCTCGCACC
CCACGTTTGTTCACCAGTATTTCAGTTTACGGTCAGAAAATGAACACAGA
CACTTCGTGATACTCTACACTTTTCAAAGGCGTAAGGGATGCCTTTTAAA
GGATTATGGATTAGAAAAATTCCTCCCTCTTTCTTGTGCCTCTGGGCCCT
TGCATTGTGATTCTATCTTACGTAAATAAAGGGGGCTTTGCTCTCCTAAT
TGTGCCCACTGTTCTGTGCAGCGCGGACCGGCGCATGCAGCGAGCGGGGC
TGCGAGGGCGCTGCTGTGGCCAGGCGTCTGGCATGCTGACCACGTCGCGC
TGCTGTAAAGGAAACCTGCCCCGCGCAGCGGCGGTGGCTGGAGCGGGAGA
AACCGGACTTTGTGCAACTTTGGCCATAGTGGCCATCCCATGAATCTGTT
TACTAGCTTGGTGGTGGGTCCAACAGAGCTTGTTGCTCCCTAGCCGCTTG
CTCGTGCCCTTGGTGGTTACCGGTAGTTAAGCTTAGGGCGCATAGGGCCC
TCGTGGCTCGCCACCTCTCACGGTTCAGTACCCACGCTTCGAACGAGGGA
TGGGAGCAGGCGCCACGGCCGGCACCCCAGAGCCCTGCTGCCCCTTAGTT
CGAGCGGCCATCCTCCTGTGGGGCTTGTGGGCAGCCTGTGGGGTTTGTGG
GCGGCCTGTGGGGTTTGTGGGTGGTCTAAGGAAAGAGTTGGGGCACTCAG
GGGTCTGCTGTTTTTGCCCGTGGCCTTAACTCATCAGGGGAGGGTTTCTG
CAGCAGAATCTCGGGCTCAGGGTTGGCGGTTAACGAGGGAGCAGCGGGGT
CTTGGGGAGGGGGCTCGACACCCCTGAAGGTGCCCCCTAAAGGAGCCACT
GTTAGAGGGGCACCCCATCTTTGTGGCCATGGCGGTGGTAGAGCGGCTGG
GAGGGGCTCTGCGGCGAGCAAGGGAGCAGGCGGTAGGGGTTCTGCGGCGA
TGGGCGGGCTAGGGGCGGGGCGCGGGTGGGCTCTAAAAGTCGG

Table S4: Information regarding CDX2 binding sites on miR-29a promoter

CDX2 binding sequences: TTAATA, TTAACA, TTTACA, TTTATA

miR-29a
(hg38_dna range=chr7:130876613-130881813)

AGAAACTGGAAGAAGAAAAATCACCTTTGCATTATTGCTTTGCATTTGTT
TTCTTAGTTCTTGTCCATGGCCCCAACGGTCACCAATACATTTCCTCTCA
GCAGTCAGCATCATGGTGCTCTTCCCCAATCATTATAACCGATTTCAGAT
GGTGCTAGAAAATTATATTGACTCTGAACACCAAAAGAAATCAGTCATCC
TCTAAGGGGTCACAGAAGGTCGTGAGGGTTGGGAGGGCCCAAATCTTACT
AAATCCAGCATACATTGGGCTCTTTAACCCAGGCACAATGGAATGAGTCT
GTACTATCTATTAAGTGCTATAGGTTTATTCCACAAAGAAGTATAAAATT
CCAACAATGCGATATCCTGTACAATTACATTCATCCCTTGGTATCCTTGA
GGGATTGGTTCCAGAAAGCCCCTTGGATACCAAAATCCACAGATGCTCAA
GTCTCTGATAGAAAATGGTGTAGTATTTGCATACAACCTACACACATCCT
CCCGTATACTTTAAATCGTCTCTAGATTACTTAGAATACTTAATGCAATG
CAAATGTATGCAAATAATTGTTATGCAGTATTGTTAGAGAATAATGACAA
GAAGAAAGTCTGTTCATGTTCAATACAGACACAACTTATTTTTGGAAAAC
TTTCGATCACAGTTGGATCCGCCAGTGCAGAGACCTGACTGCCATTTGTG
ATATATGCCACCACTGTTGGCAGGGGTGAGGAAGGGGTCAGCTACATGTG
AGGCAGGTTCTCACAGCCATCCTGCTGTTGCTTTCAAATACTTCAGAGCT
GTCCCATTCACTTCTTTTGCTGTTGGTAGTGCAGCGCAGCTGGTCTCCCC
CAAGAACACTGATTTCAAATGGTGCTAGACAATCACTATTTAAATCTAAA
CCACCATATGAAACCAGCTTCCTGAAGAAGCTTTATGATCTATTTTTTTC
TTCATAATGCTCTCTTACAATAAACCCAAGACAACCTAGAAAGGAAAAGT
TGAAACCTAGCTGTGAGTTACTCAGTGGTCTAAACAGAGCTTTGGCAAAT
TTCATATATTCATAAACTGATGTCACTGAATTTGTTATTGATGGGAGAGC
ATGCCTGGGACTCCCTTATTCTTAAGTTAAGTTGCCCAGACAAAGGTTCA
GCTTCATCATTTAGAGGCAAGGGCACATGTCAGCCCCGTCACTCTTGGGT
GGGAAGACAAGGAAAATCCTTTGCTTGGGTAAAGGCAGGAGAAGTCAGAG
GCATCCTGTGATCAAGTACCCAGGCAGGCGGAGGACGACTAGCAGTGGTG
CTGGAGAACCGTCCGTCGGACATGCAAATAGGACAAAGCAAGCAAGTGCT
CCATAGAAGTTGAAGAAAAACGCTTTCCTAAACCTATTATTTTCACTTGG
AGATAACTTGGATCCAAAGGATTACTGAGATTTTGGGGTCTAGGGAAGAA
AATAGACTAGCTCAGAGGCCAAGTAAGTCACATAATGGCACTAATGAGAT
GCTAATGTCGCTTCCGGAATGAATTCCTTAGACTGCATCTATGAGTTAGG
GGTTTTTGTGAGCTTTACTTAAGACTCTCCTTCCCTCTACTCCAAAGGTA
AGGCTTTGTACTCTGAGTTCTCCTTGGCAAATTATTTTTGCAGCAGTTC
TTTTTCGCCGTCATGGTATAGTTCTTGGTATCATCTTATTAGGTGGCCTG
TTGTCGTCCCTTTTTGTGTTCTAGGCATCCAAAAAGATACTTCATGTACC
CAGGCCACCTAATTAAAGGGGTGTGCACTGTT**TTTATA**AACTGTTATAAA
AAAT**TTTATA**CATTAAATTGGAATGACAATACAATGAATTCCTAAATCCA
TACTATCTAAATTCAATAACTATTCCTTTTTATCATGTTTTGCTTTAATT
ATATGTAATATAGATACATATTTTTTGCTGCATTGTTTGAAAGTAAATTG
ATAACCACCCATACATACACTGCTCCTAACTACTTTGAGGTCGTTCTATA
AAATGAAGACACTCTTCTACATAACCACAAAATCTTACCATACCTAAGAA
AACAAACAATTCCTTGGTACCCACAAATATCCAGCCTACATCCAACTTCT

CCCACTTTCTCCAAACAAAAGGTTT**TTTATA**GGTTTTTCAAACCCAGATC
GCATCAAGGTTGATTCACTAAGACTCTTTGGTCTCTTTTAGCCTCAAAAC
ACTACCCAACCTTTTCTTTGGCTTGTTTTATCATGACATACCACCTTGTG
AAAACCAGGCTCCAATCCTGGCTGTACCATTCACTTAGTCTGTTCCCTTG
GCTTGTATTTAAGAACTGAACTTTAGTTTCTCCGTCTGACTTTCCTCATA
TATCATCCTCAT**TTTACA**ATCTCACTATGTGTATTAGAAGAACATCCAAT
TACCACGTGCCTGTGGTTGATGATTCCAAGTGGGTTCAGTGCTAATCAGT
CAGGAGGTGCACTGCAGGCTGTTTACCTTGGGTCTGATTAGCATAACATT
CTTGTTCCTCATGGGCTTGGCCAGTTATCTTTTCATAAGGAGATCCCAGC
ACAAACTCATCCTTTATCATCATGGAGAAGCATCCACAGCACATACCTAT
TGCTGCTAGGCTGGTTTTTAAAAAAAATCTCTGTGAATAAGGGAATAGAA
CACCAGGGCTATGGAGATGCGTGGCAATCATTTCATCAAAGATGTAAACA
GCTTGTTGACTGCAGTGAAAGGCACATACTTTGTGAAGTGACAGGCAGTG
ACACATTTCAAGTGGATAGTAAATTTAAGCACCAAAACTGGGAAAAGTGC
CAGGACAGAAGAATCACTGCAACTCTCAAATAATGGTTTCACTAATTCTG
TGATTCCTTTAGGGACCCCCTGCATCCTTCTGTATATGCTACTCTAATTC
GCAGTATCAGTGACTGAAGTACAAAAGAAGCAGCCAGATTATTCTCATCA
GTTACAAAACTCTAAAGTCTGCTTATTTTTGATAAATAACTAGTCACAGT
GTTTGCCTACTGCCTAGAAGTTTTTTCAAATGATTGACACTTCAGTTATG
GGGAAAATTTTTCCTGAAGAATGAATTTCTTTTACTTATGTCTGATAAGA
TCATATATATCTCATACACCAAATACATCACACACACTATATATTCAACC
ATGCCCTTAGCAGGGCCAAACAAATATGCTAAGTTT**TTTACA**TTGTCTTT
ATGTATGCTTCTCAAAAATTCAGCAGAAGGCTAAGCGTGGTGGCTCATAC
CTGTAATCCCAGCATTTTGGGAGGCTGAGGCAGGCAGATCACTTGAGGTC
AGAAGTTTGAGACCAACCTGGCCAACATGGTGAAACCCCATATCTACTAA
AAAAATACAAAAAAAAAAAAACCAAACTAGCCAGGCGTGGTGGCACATGCC
TGTAGTGAGGCTGAGGCCGAGGCAGGAGAGTTGCTTGAGCCCAGTAGGCG
GAGGTTGCACTGAGCTGAGATCGTGCCATTGCACTCCAGCCTGGGTGACA
GAGCAAGCCTCTGTTTCAAAAAAATTAGCCGGGTATGGGATGCCTGCAAT
CCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGC
AGAGGTTGCAGTGAGCCAAGATCATGCCACTGCACTCCAGCCTGGGCAAT
ACAGCCAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGCC
AGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCA
GGTGTATCACGAGGTCAGGAGTTCAAGACTAGCCTGGCCAAGAAGGTGAA
ACCCTCTCTCTACTAA**TTAATA**ATACAAAAATTAGCTGGGCGGGAGGCTG
AGGCAGGAGAATTGCCTGAACCCGGGTGGCAGAGGTTGCAGTGAGCCAAG
ATCGCGCCACTGGATTCCAGCCTGGGCAAAAGAGTGAGACTGTCTCCGAA
AAAAAAAAAAAAAAAAAAAATCAGAAGAATAAAGGTATGATACATTATTT
TGCATCTGTGGGCTCTTCATATTAGCCCAGTGTGACAAGGTCATTTGGTA
TATCTTCCCACATTGCTTTTATGTCCTTGGGGAGAAAGTGGGGTACTGAG
TTGGCAAGTAGTGACCACGCCAGGTACTCCTGTGTGGCAACAGCTGAGTA
GTGGCAGTCGATTGTTTTCTGTGACATGAAAAGCCAGCAAACTGGCTTAA
GGTCACTTCCCCTGGGGTTTGAGGATTTCACAAGGAGAGCTCTATCTGAG
CAGGACATTTCTTTAAGAACGTAGCTTTGAAATGTTCGAGACATCAGGAC
AAAACAGTCCCAACCTGAGTATTGAGGCTGGACTACAGATGCAACTGTTA
CGCAACAGAACACACTGTGATAAGACATCTTGAGTTGGGTCCTTGTTTTC
ACATAAAGCAAGAGGCTAGTTTCCTGAAAATAAGTCTCCATGAAGAGACT
GGCTACCCAGTTAACGGCAAAAAACCCCCATGGCTCACTAGTAAAGAGAT
GGAAGAAAACCAGTTGATAAATTAAAAAGGGCTCATTCATTCTCTTTCCA
GGGCATGCATAGGAGAACAGAGTGAGACAAGTGGGATCTGCTTATTTTGT

TGAGTCTTTCAGTGACTTTCAGCTGCAAGGGCAACTCTACTGAGAAGGAA
AGATTTCACTCAGTAAACATTAGCATGTTTCCAACCCTGCCTAGCCTCTC
ATTTCCAACAATCAACTTGATTGTTCTGTATTTTTAGGCCAGGATACACT
AGGGCTGAATTTCTGTAACTCAGGATGTTCTATTTTAATCATCAGCAATT
TTAAGGCATGGCAACATCTGAAAGAAAAAAAAAACGTAATGACTCTTTCT
TACACACAGAACATGCACTTAAAATTGTCAAAACTCTCCATCTCCAGCAT
TAAAACAATCATCTTTCAAGAACAAAATCCTTTCTCAGTGTCTCTCAGCC
CACAAGCCATGAGAATGTGTCTCCATTTCAAAACCCAGTCTGTACCGTCA
TGCTTCCGCTGTTTCTCAGCAAGCTGCCTTTCTTTGTAACTTTTTTCTTC
CAAATCTGATTGATAGTGAAGCTGGAGCCTGAGATGTAAA**TTAATA**AACC
ACACCAGGAACTCCTTGCATGGCAAAAGCTGAACAGTACAAATCCATTGT
TTTCCTGGATGTGAAATGCCAGTCTGTTCTTGCATGAGCAAACTGGATTA
A

Table S5: Information regarding CDX2 binding sites on miR-29b promoter

CDX2 binding sequences: TTAATA, TTAACA, TTTACA, TTTATA

miR-29b
(hg38_dna range=chr7:130913180-130918380)

ATCTTGAGAAATTCTGATAAAACCACCAACTGAAAACCTGCCATTATAAA
GGAGGGATTTCACAACGTTCATTCCAAAAAACTGTGGCTATTCATTTCTG
GAAGTGATTCACTGCAGAAGAATCGCTGTTTGGGGGTGGAGGGAGCGTTG
ATTTTTAGAAGCCAGCCTGCCCTCTGAGAAGTGAGCAGCAACCTCTGAGA
TGGCAGCCACGTGCCTTCTGACCCCCAGCCCTGCTCCCTGAATCACCACC
CCTCGAGATGACACGCAGTGACCAATTTCCTCTTAAAACTCCCTTCTGTT
CAAAAACCCTATGATCCATGCACCCTCTGTCTAATCCTCTGCGATGGGGC
CAGTCTGCAGATCCGTCTGGGCGGGCTGCGGTAGCATTCTCCACCTGCCA
ACACTCAGTTGCTGGTGAGTAACCAGCATACGGACTCTCACTTCCTGTGA
GTATTTCTGAAAGAGCTTATGATATTTCCTTCAAAAAAAAAAAATTGTGG
GGAGGAAAGCCACTCGGGGAGGAGAAGCAACTTCAACAAGCAAAGCTAAA
AGCATGCAACGTGATCACTGTTGACCAGACACGCCTAATTTTAATTTGGT
TTCTCAACCCCTGTGTAATTAAGACAGCCCAAAGAGAACAGCATCAAGAG
GAAAGAAAAAAAAAAAAAGACATTTCACAAATGCTTAGACCTTCGGTGCGC
GTGACCAGAAAAGTAATTTTAAGATTCAAGCGAACCAATCTGAGGCTGGT
TGGAGCATTAACCCTTCAAGGAAAGAGAAAGAATTATTTTTGTTTGTCCC
CAAATTTCAAATACGGAATTCGTCTTCCCTACCAGGAACTGAGATACCAA
AAATCAGCCTCTAAATACCATCCTCTGGGCTTGTAAGATGAAAAAACTTT
TTCAGATTTGTCCATCACAGGGGTAGAAAAATACCTATTTATGAGACCCA
GTCACA**TTAACA**GAGCAGCCGACATGGTGTTCACACTCAGGCACCCTTTG
AAAACCGGCCACTAAGGCTAAAATGTGACATTTTGTATTTCTCTAGGCTG
TAGCCAAACCTGTTAATTTATGTTAGAGAGAAGGGCAGCATTTCTTGACA
ACTGGGCAAAATGAGTAAGAAAACTTGCTTTCATCCCACCTTACAGAGTA
TCAATGGTGACAAAGCTCACTTTCTGAGATTAGTAGACGTTTTGCTGAAA
GAAAGATTTTTCAGAGCCTGTCCTTCAGCTTCAGGCTGTGAATGTGCTTC
CAACTTGGGGCCTCCATGTGCCTGCATGGTGGGGAGGGGGGCTCTTGAAG
GGGGAAGGAGGAAAATACAAGCTTCTACTTAATGGGCACTGAAGAATAAA
ACAAATCACCCACATTCACACCTGTATTTATCACACAATCCATGGAAGCC
CTTAACTTGAAACACGGCAGCACTTTGCCACTTAGACCCCAACTCTTTTA
TTTCACATTTTTGGATGTGAAATTCTGAAAGGGAATTTTCACTTTTTTTT
TTTGTTTGTTTGAGACGGGGTCTCACTCTTGTCGCCCAGGCTGGAGTGCA
GTGGTGCGATCCTGGCTCACGGAAGCCTCGACTTCCTGGGATGTGATAAG
CTTTTACCATCTTTACCAAACTGATATT**TTAACA**AGAAGAGTATGGAAAT
CATTTGCCTGATATACATAATTGATTTATTTATTTATTTTAGAGGCAGGG
TCTCACTTCATCTCAGCCTCCAGAGTAGCTGAAACTACAGGTAAGTACCA
CCATGCCCAGATAATTTTTTTTTTTTTGGTAGAGATGACGCCTTGCTGTG
TTGCCCAGGGTGGTCTCTAATTCCTGGCCTTTAGCAATTCTCCCGCCTTG
GCCTCCTAAAGCACTGGGGTTACAGGCATGAACCACTGTGCCCAGCCTGC
ATGATGTAATCTCCAAATTGAGGGTAAGTAAAAAGTAGCACCTATAACTA
GAGGTAAATTTTTAATTTAAAACAACTACACTACGCAATAGTCATCTCCT
CAGCCCTTGCCCACGAGATCTTATGAACCTCCTCAATAGAAAAAAAGGAA
GTCTGTGATAGGATCAAAGGAACAGGAACCC**TTTACA**TGGAGAGTAATTT

Journal of Cell Science • Supplementary information

TAAATAGCTACAATTTTTGAGCTTGTCCGTTGATACCGAGCTCAGCAAAA
GAAAAAGCAGTGATTACAGGCCTGACCACTACTCAATGATTAGAATATTC
ACTCTTCTCTATTACTTAGAAGGAAGCCCCCCTCTCCCACTATGTCAGGT
ATACACACAAGTTCACAGCTCTCAGTGTTTCTGTGTAATTTGTTCCTGTT
TCAGACAGCTTTCCAGAAAAGCTCCTCTTAAATCCTGGGTGGCCAATGCA
GAGGTTTTAAAAAGAGGAAAAGTAGAGCAAAACAACAAAAAGATCATGAA
ATAGAAGCTTGGGGAGAAGGGAGCAAAGAGAACATGAATCTAACAGATGC
AGCAGTTGCAAAAAAGCTGCTATCCTTTAGAAACTCCTGTTAATTTGGTG
ACCTAAAGGGAGCTATTTCTAAGGCTGGATGTTGAGAATGTGATGAAACC
ATGGAGGCTGTCCAGGTCCTTGCTATGCTGTGGACACCATTCAAGAGTTA
TAGAGTCCACCCTCTCCAAACAAAGTAAAACCCCGTGACAAATGACCTTC
ATGT**TTAACA**TTTACTCAGTTCTCTTTTAAACAATGAAATGGAGCGCACT
CTTTCGTGGCACT**TTTATA**GGCAGGAGAGAGAGTGGTCTCACAACGGAAG
GGCACCAGCCACTGAGAAGTCCTAAGGCCGTGTTTGAGTTCTGGCTCATC
TTTTTTGCTAATTTGCTCTTATGCATATCACCCAACCTCGCTAAGCCAGT
TTCATCATCAGCGAGAACAGCACACTGACATGCAGGGCCAGCTCAGGCGC
TACCACAAGCTACACATCCGTCCATTCGTTTAATCTTCACAGCAGCCCTG
GAGGTAAGTTCTATTTTTAATCCTCATCCTACAGATAAGGAAACTGAGGC
CCTGGGTGCTTAAGGACATGCCCAGGGTTACACAATCAGTAGGTGGTTGA
GCCAGGATTCAGAGCCAGGGAGCCTGACCTTGACTCCACATCCAGAGCCT
GCCCTCAGCATGTCTGCCAGTAACACCCAGGTTGGCTGCACAGCAAGGAA
AACAGCATATAAGAAACTACCTTAAAACCTACAAAGACCATAAAAACACA
AGGTAGATTTTTAATCATGTCACATGCAAATATATTGTCTGTAAAAATGT
AATGAAGATTCAGGACAACGAAGCAGTGGTAACTCTGCTGATCCTGTGAA
AACAGTTTCACATAACTTGCAAGTTACTTAGTTGCACAATGTGTGGTACC
GAGTTGTAGAGGTAGAAGGGATTCAGAGATCCTGTTTTCCGTCAGGGTGG
TTGGGTGAGGCTCCCAGACACAGGGCTGAAGTCTGAACCTTCATCCTTTC
ACTCCCAGCCCAATGATGTGCCACTGTGTCATAGCCACATGGTGGTTATG
AATGCATCTGGAACTCCACATTTTAATCTGCCAGGCATCCACCCATTCGT
TCACATAATTCACGTGTGTGTAATAGTAACCTTCAGAAAGGAATGTCACA
ATATTTTTCTCATTACAACATCTTTAGGCTTTAAGATGAAAAGGGTGAAG
AAAGAAATCTTCTCTAAATCTTGTATTCCCGCTATTCATTTTTTTATTCA
AGGAAAAACAAAGCCTCTGTTAAAGCTCTGCTTACAATGAAATGTAACTG
AAGTTCTATCCAATTCTTTCCTCTTCTCTGTAAACTTGTTCACATGCAAC
AGCCTCTCTGATG**TTAACA**GGCTGTTCCTCCCAAGTGCCTGCAGATCATG
GCTATAGGAGTCACTGGAAGGAACTTGACTCGGTGGCAGTTTGGAGGCTT
CCCGGGCTTTAATGATGAAAAATGATTTGGTCATTTAAGACAGAGCCCCC
CGCTCTGTTGCCTAGGCTGGCGTACAGTGGCGTGATCATAGCTCACTGCA
GCCTCAAACTCCTGGGCTCAAGTGATCCTCCTGCCCCAGTTTCTTGAGTA
GTTGGGACTACAGGCGTGTGCCACATCTGGCTATTTTTTTTTT**TTAATA**G
AGAAGGGGTCTTGCTATGTTGCCCAGGCTCGTCTTGAACTTCTGGGCTCA
AGCAATCCTTTCACCTCAGCCCCCCAAAGCAATGGTTCACAGGTTCAGAG
GTATGAACTACCATGCCTAGCCTAGCTATTTACTCTTAAAGGAAAAAATA
ATTTTTTATTTAGTTTTCAAGCCTTTCATGCTA**TTTACA**CTAAAGTTTGG
CCTTATTATTTTTGCTTAGAATCTTTTAATGGTTTTAGGTCACGTTGGGT
CACATCAAAACACTTGAAAATGCCTCTTTCTAAAACAAAGACATTCGTTG
ACAAAAAAAAGTTAAAAGAAAAAAGACCAGCCAGGTATTTAGTTCCATTT
AAGAGAGCACAGGCGACTGTATGCAAGATTCTGGAAAATTTATGTACGTT
AAA**TTTATA**AGCCAGCACTTTTTTCCCAAAGATTTATTTGGTGTGAAAAG
TTACTTCCTGTACTGTAATCATAAAATCAACTGTTATTTACTCCTGTCAT

AACCATTCACATACTTCCTGCCATAAAAACTTAAACCTGGCTAGAATATT
AAATAAAACATGCAAAAAACTACAAGTTCAGGAAGTGACTGTATGTATTA
TATACATTAAGTTCATATTCCAATCTAAGAGAAAGCTGAATCACCCCTAC
TTAGGAAAACCAGTATTATGTTTTATCATCTGGGGTGACTAATCCCACAA
TGGCCCGTGCAGTGAACTTTAAGGGTAAGAAGGGCTTTCTCTTTTGCACT
AAGCTCAGTCCATAGTTTTGTGGGGGCTGTCAGGACGTCTCGCAGGTCTG
CCGTGTGTAGTGGGCTGAAATACACTGTACATGGCCCAGAGTTAAAACTG
TAAAGATCCAAAGACTTACTTAGGGTAGATGTGCACACTTTAAATGCATA
TGCAGGTTTTTCTTTGGTGCCTGCCTACTTCTAAAAGAAAACAAGCTTCT
AAATGCTTAGGCTTTGGCGTGTCGGGTGGGAAAATCAGAGCCAGTGTACA
CATCCAATGATAGAAGAAAGCGTGCGTGAACAGATTTG**TTTACA**AAACAG
GGCCCAGCAAAGTCTGTCTCTCAGTGCCTCGGCAGGATCCATCCAACAAA

Table S6: Targets for individual DDSMs
(presented in MS Excel format)

Table S6
Click here to download Table S6

Journal of Cell Science • Supplementary information

Table S7: List of common targets for all DDSMs
(presented in MS Excel format)

Table S7
Click here to download Table S7

Journal of Cell Science • Supplementary information

Table S8: BRCA1 coding and 3'UTR sequence with the DDSM binding sites

GTACCTTGATTTCGTATTCTGAGAGGCTGCTGCTTAGCGGTAGCCCCTTGGTTTCCGTGGCAACGGAAAAGCGCGGGAAT
TACAGATAAATTAAAACTGCGACTGCGCGGCGTGAGCTCGCTGAGACTTCCTGGACGGGGGACAGGCTGTGGGGTTTCTC
AGATAACTGGGCCCCTGCGCTCAGGAGGCCTTCACCCTCTGCTCTGGGTAAAGTTCATTGGAACAGAAAGAAATGGATTT
ATCTGCTCTTCGCGTTGAAGAAGTACAAAATGTCATTAATGCTATGCAGAAAATCTTAGAGTGTCCCATCTGTCTGGAGT
TGATCAAGGAACCTGTCTCCACAAAGTGTGACCACATATTTTGCAAATTTTGCATGCTGAAACTTCTCAACCAGAAGAAA
GGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAAAGGAGCCTACAAGAAAGTACGAGATTTAGTCAACTTGT
TGAAGAGCTATTGAAAATCATTTGTGCTTTTCAGCTTGACACAGGTTTGGAGTATGCAAACAGCTATAATTTTGCAAAAA
AGGAAAATAACTCTCCTGAACATCTAAAAGATGAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCCAAAAGA
CTTCTACAGAGTGAACCCGAAAATCCTTCCTTGCAGGAAACCAGTCTCAGTGTCCAACTCTCTAACCTTGGAACTGTGAG
AACTCTGAGGACAAAGCAGCGGATACAACCTCAAAAGACGTCTGTCTACATTGAATTGGGATCTGATTCTTCTGAAGATA
CCGTTAATAAGGCAACTTATTGCAGTGTGGGAGATCAAGAATTGTTACAAATCACCCCTCAAGGAACCAGGGATGAAATC
AGTTTGGATTCTGCAAAAAAGGCTGCTTGTGAATTTTCTGAGACGGATGTAACAAATACTGAACATCATCAACCCAGTAA
TAATGATTTGAACACCACTGAGAAGCGTGCAGCTGAGAGGCATCCAGAAAAGTATCAGGGTAGTTCTGTTTCAAACTTGC
ATGTGGAGCCATGTGGCACAAATACTCATGCCAGCTCATTACAGCATGAGAACAGCAGTTTATTACTCACTAAAGACAGA
ATGAATGTAGAAAAGGCTGAATTCTGTAATAAAAGCAAACAGCCTGGCTTAGCAAGGAGCCAACATAACAGATGGGCTGG
AAGTAAGGAAACATGTAATGATAGGCGGACTCCCAGCACAGAAAAAAAGGTAGATCTGAATGCTGATCCCCTGTGTGAGA
GAAAAGAATGGAATAAGCAGAAACTGCCATGCTCAGAGAATCCTAGAGATACTGAAGATGTTCCTTGGATAACACTAAAT
AGCAGCATTCAGAAAGTTAATGAGTGGTTTTCCAGAAGTGATGAACTGTTAGGTTCTGATGACTCACATGATGGGGAGTC
TGAATCAAATGCCAAAGTAGCTGATGTATTGGACGTTCTAAATGAGGTAGATGAATATTCTGGTTCTTCAGAGAAAATAG
ACTTACTGGCCAGTGATCCTCATGAGGCTTTAATATGTAAAAGTGAAAGAGTTCACTCCAAATCAGTAGAGAGTAATATT
GAAGACAAAATATTTGGGAAAACCTATCGGAAGAAGGCAAGCCTCCCCAACTTAAGCCATGTAACTGAAAATCTAATTAT
AGGAGCATTTGTTACTGAGCCACAGATAATACAAGAGCGTCCCCTCACAAATAAATTAAAGCGTAAAAGGAGACCTACAT
CAGGCCTTCATCCTGAGGATTTTATCAAGAAAGCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATAAATCAGGGAACT
AACCAAACGGAGCAGAATGGTCAAGTGATGAATATTACTAATAGTGGTCATGAGAATAAAACAAAAGGTGATTCTATTCA
GAATGAGAAAAATCCTAACCCAATAGAATCACTCGAAAAAGAATCTGCTTTCAAAACGAAAGCTGAACCTATAAGCAGCA
GTATAAGCAATATGGAACTCGAATTAAATATCCACAATTCAAAAGCACCTAAAAAGAATAGGCTGAGGAGGAAGTCTTCT
ACCAGGCATATTCATGCGCTTGAACTAGTAGTCAGTAGAAATCTAAGCCCACCTAATTGTACTGAATTGCAAATTGATAG
TTGTTCTAGCAGTGAAGAGATAAAGAAAAAAAAGTACAACCAAATGCCAGTCAGGCACAGCAGAAACCTACAACTCATGG
AAGGTAAAGAACCTGCAACTGGAGCCAAGAAGAGTAACAAGCCAAATGAACAGACAAGTAAAAGACATGACAGCGATACT
TTCCCAGAGCTGAAGTTAACAAATGCACCTGGTTCTTTTACTAAGTGTTCAAATACCAGTGAACTTAAAGAATTTGTCAA
TCCTAGCCTTCCAAGAGAAGAAAAAGAAGAGAAACTAGAAACAGTTAAAGTGTCTAATAATGCTGAAGACCCCAAAGATC
TCATGTTAAGTGGAGAAAGGGTTTTGCAAACTGAAAGATCTGTAGAGAGTAGCAGTATTTCATTGGTACCTGGTACTGAT
TATGGCACTCAGGAAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAAGGCAAAAACAGAACCAAATAAATGTGTGAG
TCAGTGTGCAGCATTTGAAAACCCCAAGGGACTAATTCATGGTTGTTCCAAAGATAATAGAAATGACACAGAAGGCTTTA
AGTATCCATTGGGACATGAAGTTAACCACAGTCGGGAAACAAGCATAGAAATGGAAGAAAGTGAACTTGATGCTCAGTAT
TTGCAGAATACATTCAAGGTTCAAAGCGCCAGTCATTTGCTCCGTTTTCAAATCCAGGAAATGCAGAAGAGGAATGTGC
AACATTCTCTGCCCACTCTGGGTCCTTAAAGAAACAAAGTCCAAAAGTCACTTTTGAATGTGAACAAAAGGAAGAAAATC
AAGGAAAGAATGAGTCTAATATCAAGCCTGTACAGACAGTTAATATCACTGCAGGCTTTCCTGTGGTTGGTCAGAAAGAT
AAGCCAGTTGATAATGCCAAATGTAGTATCAAAGGAGGCTCTAGGTTTTGTCTATCATCTCAGTTCAGAGGCAACGAAAC
TGGACTCATTACTCCAAATAAACATGGACTTTTACAAAACCCATATCGTATACCACCACTTTTTCCCATCAAGTCATTTG
TTAAAACTAAATGTAAGAAAAATCTGCTAGAGGAAAACTTTGAGGAACATTCAATGTCACCTGAAAGAGAAATGGGAAAT
GAGAACATTCCAAGTACAGTGAGCACAATTAGCCGTAATAACATTAGAGAAAATGTTTTTAAAGAAGCCAGCTCAAGCAA
TATTAATGAAGTAGGTTCCAGTACTAATGAAGTGGGCTCCAGTATTAATGAAATAGGTTCCAGTGATGAAAACATTCAAG
CAGAACTAGGTAGAAACAGAGGGCCAAAATTGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGTCTATAAACAA
AGTCTTCCTGGAAGTAATTGTAAGCATCCTGAAATAAAAAAGCAAGAATATGAAGAAGTAGTTCAGACTGTTAATACAGA
TTTCTCTCCATATCTGATTTCAGATAACTTAGAACAGCCTATGGGAAGTAGTCATGCATCTCAGGTTTGTTCTGAGACAC
CTGATGACCTGTTAGATGATGGTGAAATAAAGGAAGATACTAGTTTTGCTGAAAATGACATTAAGGAAAGTTCTGCTGTT
TTTAGCAAAAGCGTCCAGAAAGGAGAGCTTAGCAGGAGTCCTAGCCCTTTCACCCATACACATTTGGCTCAGGGTTACCG
AAGAGGGGCCAAGAAATTAGAGTCCTCAGAAGAGAACTTATCTAGTGAGGATGAAGAGCTTCCCTGCTTCCAACACTTGT
TATTTGGTAAAGTAAACAATATACCTTCTCAGTCTACTAGGCATAGCACCGTTGCTACCGAGTGTCTGTCTAAGAACACA
GAGGAGAATTTATTATCATTGAAGAATAGCTTAAATGACTGCAGTAACCAGGTAATATTGGCAAAGGCATCTCAGGAACA
TCACCTTAGTGAGGAAACAAAATGTTCTGCTAGCTTGTTTTCTTCACAGTGCAGTGAATTGGAAGACTTGACTGCAAATA
CAAACACCCAGGATCCTTTCTTGATTGGTTCTTCCAAACAAATGAGGCATCAGTCTGAAAGCCAGGGAGTTGGTCTGAGT
GACAAGGAATTGGTTTCAGATGATGAAGAAAGAGGAACGGGCTTGGAAGAAAATAATCAAGAAGAGCAAAGCATGGATTC
AAACTTAGGTGAAGCAGCATCTGGGTGTGAGAGTGAAACAAGCGTCTCTGAAGACTGCTCAGGGCTATCCTCTCAGAGTG
ACATTTTAACCACTCAGCAGAGGGATACCATGCAACATAACCTGATAAAGCTCCAGCAGGAAATGGCTGAACTAGAAGCT
GTGTTAGAACAGCATGGGAGCCAGCCTTCTAACAGCTACCCTTCCATCATAAGTGACTCTTCTGCCCTTGAGGACCTGCG
AAATCCAGAACAAAGCACATCAGAAAAAGCAGTATTAACTTCACAGAAAAGTAGTGAATACCCTATAAGCCAGAATCCAG
AAGGCCTTTCTGCTGACAAGTTTGAGGTGTCTGCAGATAGTTCTACCAGTAAAAATAAAGAACCAGGAGTGGAAAGGTCA
TCCCCTTCTAAATGCCCATCATTAGATGATAGGTGGTACATGCACAGTTGCTCTGGGAGTCTTCAGAATAGAAACTACCC
ATCTCAAGAGGAGCTCATTAAGGTTGTTGATGTGGAGGAGCAACAGCTGGAAGAGTCTGGGCCACACGATTTGACGGAAA
CATCTTACTTGCCAAGGCAAGATCTAGAGGGAACCCCTTACCTGGAATCTGGAATCAGCCTCTTCTCTGATGACCCTGAA
TCTGATCCTTCTGAAGACAGAGCCCCAGAGTCAGCTCGTGTTGGCAACATACCATCTTCAACCTCTGCATTGAAAGTTCC
CCAATTGAAAGTTGCAGAATCTGCCCAGAGTCCAGCTGCTGCTCATACTACTGATACTGCTGGGTATAATGCAATGGAAG
AAAGTGTGAGCAGGGAGAAGCCAGAATTGACAGCTTCAACAGAAAGGGTCAACAAAAGAATGTCCATGGTGGTGTCTGGC
CTGACCCCAGAAGAATTTATGCTCGTGTACAAGTTTGCCAGAAAACACCACATCACTTTAACTAATCTAATTACTGAAGA
GACTACTCATGTTGTTATGAAAACAGATGCTGAGTTTGTGTGTGAACGGACACTGAAATATTTTCTAGGAATTGCGGGAG
GAAAATGGGTAGTTAGCTATTTCTGGGTGACCCAGTCTATTAAAGAAAGAAAAATGCTGAATGAGCATGATTTTGAAGTC
AGAGGAGATGTGGTCAATGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCCCAGGACAGAAAGATCTTCAGGGG
GCTAGAAATCTGTTGCTATGGGCCCTTCACCAACATGCCCACAGATCAACTGGAATGGATGGTACAGCTGTGGTGCTT
CTGTGGTGAAGGAGCTTTCATCATTCACCCTTGGCACAGGTGTCCACCCAATTGTGGTTGTGCAGCCAGATGCCTGGACA
GAGGACAATGGCTTCCATGCAATTGGGCAGATGTGTGAGGCACCTGTGGTGACCCGAGAGTGGGTGTTGGACAGTGTAGC
ACTCTACCAGTGCCAGGAGCTGGACACCTACCTGATACCCCAGATCCCCCACAGCCACTACTGACTGCAGCCAGCCACAG
GTACAGAGCCACAGGACCCCAAGAATGAGCTTACAAAGTGGCCTTTCCAGGCCCTGGGAGCTCCTCTCACTCTTCAGTCC
miR-335 BS
TTCTACTGTCCTGGCTACTAAATATTTTATGTACATCAGCCTGAAAAGGACTTCTGGCTATGCAAGGGTCCCTTAAAGAT
miR-183 BS
TTTCTGCTTGAAGTCTCCCTTGGAAATCTGCCATGAGCACAAAATTATGGTAATTTTTCACCTGAGAAGATTTTAAAACC
ATTTAAACGCCACCAATTGAGCAAGATGCTGATTCATTATTTATCAGCCCTATTCTTTCTATTCAGGCTGTTGTTGGCTT
AGGGCTGGAAGCACAGAGTGGCTTGGCCTCAAGAGAATAGCTGGTTTCCCTAAGTTTACTTCTCTAAAACCCTGTGTTCA
CAAAGGCAGAGAGTCAGACCCTTCAATGGAAGGAGAGTGCTTGGGATCGATTATGTGACTTAAAGTCAGAATAGTCCTTG
GGCAGTTCTCAAATGTTGGAGTGGAACATTGGGGAGGAAATTCTGAGGCAGGTATTAGAAATGAAAAGGAAACTTGAAAC
miR-29b BS
CTGGGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAGATCACTGGAGGTCAGGAGTTCGA

Journal of Cell Science • Supplementary information

AACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAGAAATTAGCCGGTCATGGTGGTGGACACCTGTA
ATCCCAGCTACTCAGGTGGCTAAGGCAGGAGAATCACTTCAGCCCGGGAGGTGGAGGTTGCAGTGAGCCAAGATCATACC
ACGGCACTCCAGCCTGGGTGACAGTGAGACTGTGGCTCAAAAAAAAAAAAAAAAAAAAGGAAAATGAAACTAGAAGAGATT

miR-182 BS

TCTAAAAGTCTGAGATATATTTGCTAGATTTCTAAAGAATGTGTTCTAAAACAGCAGAAGATTTTCAAGAACCGGTTTCC
AAAGACAGTCTTCTAATTCCTCATTAGTAATAAGTAAAATGTTTATTGTTGTAGCTCTGGTATATAATCCATTCCTCTTA
AAATATAAGACCTCTGGCATGAATATTTCATATCTATAAAATGACAGATCCCACCAGGAAGGAAGCTGTTGCTTTCTTTG
AGGTGATTTTTTTCCTTTGCTCCCTGTTGCTGAAACCATACAGCTTCATAAATAATTTTGCTTGCTGAAGGAAGAAAAAG
TGTTTTTCATAAACCCATTATCCAGGACTGTTTATAGCTGTTGGAAGGACTAGGTCTTCCCTAGCCCCCCCAGTGTGCAA
GGGCAGTGAAGACTTGATTGTACAAAATACGTTTTGTAAATGTTGTGCTGTTAACACTGCAAATAAACTTGGTAGCAAAC
ACTTCCAAAAAAAAAAAAAAAAAA

ATG - Start codon of BRCA1
TGA – Stop codon of BRCA1
3'UTR sequences – highlighted in Red
DDSM Binding sites in 3'UTR of BRCA1, BS indicates "binding site"
Primers used for cloning pGL3-BRCA1 3'UTR miR-183 (KpnI + HindIII sites in pGL3 basic vector)

Table S9: List of antibodies used in the study

Name of antibody	Source	Identifier (Dilution used)
Anti-Flag (used for WB)	Merck	Cat# F1804; RRID:AB_262044 (WB: 1:1000)
Anti-Flag M2 affinity gel (used for IP)	Merck	Cat# F2220 (IP: 1μg/IP)
Anti-BLM (used for WB, ChIP)	Bethyl Laboratories	Cat# A300-110A; RRID: AB_2064794 (WB: 1:2500, ChIP: 2 μg/IP)
Anti-CDX2 (used for WB, IF, EMSA, IHC)	Abcam	Cat# ab76541; RRID: AB_1523334 (WB: 1:1000, IF: 1:250, EMSA: 1μg/reaction, IHC: 1:250)
Anti-CDX2 (used for WB)	Santa Cruz Biotechnology	Cat# sc-134468; RRID: AB_2260275 (WB: 1:1000)
Anti-CDX2 (used for WB)	Santa Cruz Biotechnology	Cat# sc-166830; RRID: AB_2260278 (WB: 1:1000)
Anti-BRCA1 (used for WB, IHC)	Abcam	Cat# ab16780; RRID: AB_2259338 (WB: 1:2000)
Anti-GFP (used for WB, IF)	Santa Cruz Biotechnology	Cat# sc-9996; RRID:AB_627695 (WB: 1:5000, IF: 1:500)
Anti-hsp90 (used for WB)	Santa Cruz Biotechnology	Cat# sc-7947; RRID:AB_2121235 (WB: 1:2000)
Anti-PCNA (used for WB)	Santa Cruz Biotechnology	Cat# sc-56; RRID:AB_628110 (WB: 1:1000)
Anti-CD31 (used for IHC)	Abcam	Cat# ab28364; RRID:AB_726362 (IHC: 1:250)
Anti-ATM (used for WB)	Santa Cruz Biotechnology	Cat# sc-377293 (WB: 1:1000)
Anti-Chk1 (used for WB)	Cell Signaling Technology	Cat# 2360; RRID:AB_2080320 (WB: 1:750)
Anti-RNF8 (used for WB)	Santa Cruz Biotechnology	Cat# sc-271462; RRID:AB_ 10648902 (WB: 1:1000)
Anti-SMAD3 (used for ChIP, WB)	Abcam	Cat# ab28379; RRID:AB_ 2192903 (WB: 1:2500, ChIP: 2 μg/IP)
Anti-Sin3b (used for ChIP, IP, WB)	Santa Cruz Biotechnology	Cat# sc-13145; RRID:AB_ 628254 (WB: 1:1000, ChIP: 2 μg/IP, I: 1 μg/IP)
Anti-AP2β (used for ChIP)	Santa Cruz Biotechnology	Cat# sc-390119 (ChIP: 2 μg/IP)
Anti-HDAC1 (used for ChIP, WB)	Abcam	Cat# ab7028; RRID:AB_ 305705 (WB: 1:1000, ChIP: 2 μg/IP)
Anti-HDAC2 (used for ChIP, WB)	Abcam	Cat# ab7029; RRID:AB_ 305706 (WB: 1:1000, ChIP: 2 μg/IP)
Anti-CHD4 (used for ChIP, WB)	Abcam	Cat# ab70469; RRID:AB_2229454 (WB: 1:750, ChIP: 2 μg/IP)

WB: Western blotting
IP: Immunoprecipitation
IF: Immunofluorescence
IHC: Immunohistochemistry
EMSA: Electrophoretic Mobility Shift Assay
ChIP: Chromatin Immunoprecipitation Assay

Journal of Cell Science • Supplementary information

Table S10: List of recombinant DNAs used in the study

Name of the recombinant DNA	Source	Identifier
CMV-β gal	Present in the lab of corresponding author	N/A
pcDNA3 Flag BLM (1-1417)	Present in the lab of corresponding author	(Kharat et al., 2016)
pFlag CDX2	Jean-Noel Freund (Université de Strasbourg, France)	(Balbinot et al., 2017)
pFlag CDX2 mini	Jean-Noel Freund (Université de Strasbourg, France)	(Balbinot et al., 2017)
pGEX4T-1 CDX2	This study	N/A
pGL3-miR-96/182/183 WT promoter luc	This study	N/A
pGL3-miR-96/182/183 MT promoter luc	This study	N/A
pGL3-BRCA1 3'UTR miR-183 WT	This study	N/A
pGL3-BRCA1 3'UTR miR-183 MT (mutated)	This study	N/A
pcDNA3 HA-BRCA1	Dipanjan Chowdhury (Dana-Farber Cancer Institute, USA)	N/A
EGFP-C1-BLM	Nathan Ellis (University of Arizona, USA)	(Hu et al., 2001)
pcDNA3 Sin3b Flag	Gregory David (New York University School of Medicine, USA)	(Bainor et al., 2018)
pTRIPZ shSin3b	Gregory David (New York University School of Medicine, USA)	(Bainor et al., 2018)
pTRIPZ sh Scrambled	Gregory David (New York University School of Medicine, USA)	(Bainor et al., 2018)
Flag-HDAC1	Present in the lab of corresponding author	N/A
GST-HDAC1	Gordon Hager (National Cancer Institute, National Institutes of Health, USA)	(Qiu et al., 2011)
pCMV5B-Flag-SMAD3	Jeff Wrana (University of Toronto, Canada) via Addgene (Cat# 11742)	(Labbe et al., 1998)
pGEX6-SMAD3 WT	Joan Massague (Slaon Kettering Institute, USA) via Addgene (Cat# 27010)	(Gao et al., 2009)
pLenti-III-mir-Off Control Vector	Abm Inc.	Cat# m007

Table S11: List of reagents used in the study
(presented in MS Excel format)

Table S11
Click here to download Table S11

Journal of Cell Science • Supplementary information

Table S12: List of primers used in the study

RT-qPCR primers for genes		
Identifier	**Forward primer (5' to 3')**	**Reverse primer (5' to 3')**
BLM	AGA CAG GAT TCT CTG CCA CCA GGA	TGG TGT TTC AGC CCA GTT GCT
CDX2	GCG CCC GGA GGC GCC GCG GCC	CGC CGG CTT CCG CAT CCA CTC
BRCA1	GTG AGT CAG TGT GCA GCA TTT GAA	TCT ATG CTT GTT TCC CGA CTG TGG
Chk1	TCT ATG GTC ACA GGA GAG AAG G	CAT GCC TAT GTC TGG CTC TAT TC
RNF8	TAG GAC GAG GAT TTG GTG TCA	TAA TTG TCC ATT GGC CCT CAG
ATM	CAG GGT AGT TTA GTT GAG GTT GAC AG	C TAT ACT GGT GGT CAG TGC CAA AGT
ChIP primers for CDX2 promoter		
Identifier	**Forward primer (5' to 3')**	**Reverse primer (5' to 3')**
MAD	CAA AGG TCA TTC CAG GCA CG	TCC ACC CCA AAG GAA ACC TC
AP3β #1	CCT GGT CCC TGA GTG GTA AC	TCC CAA GCA AAC TGG GTG TC
AP3β #2	GGC CCT ACA GAA AGG ACG AT	GCC ATT GCA CTC AAT CAG CC
SMAD3 #1	GGA CCT TTG AAC AGG GAG GGA	CTC CAA GCA GAT GAA AAA CTC CT
AP2β #1	GGA GCA CTC CAA GTG AGC CT	CCT GCT CAT CCA ACT GCT TTG
E2F1	GTC TCC CAC CTC TTG AGT GC	TGC AGA CAC CTG AAA GCG AT
AP2β #2	CGA GCG AAA CGC TTG ACA AT	TCG GGA ATG AGG ATC ACT GC
SMAD3 #2	TTA TGT TTC GAG GGG TTG TGC	GAC ACA GAC ACC AAT GGT TGG A
ChIP primers for GAPDH promoter		
Identifier	**Forward primer (5' to 3')**	**Reverse primer (5' to 3')**
GAPDH	ATG GTT GCC ACT GGG GAT CT	TGC CAA AGC CTA GGG GAA GA
Primers for EMSA and SDM		
Identifier	**Top strand**	**Complementary strand**
miR-182,miR-183, miR-96-5p (WT)	CTG CTC CCC TTT ATA ACG CGC GCC	GAC GAG GGG AAA TAT TGC GCG CGG
miR-182,miR-183, miR-96-5p (MT)	CTG CTC CCC TTG CTA ACG CGC GCC	GGC GCG CGT TAG CAA GGG GAG CAG
Primers for measuring levels of mature miRs (DDSMs and non-DDSMs)		
Identifier	**Forward primer (5' to 3')**	**Reverse primer (5' to 3')**
miR-29a-5p	GCA GAC TGA TTT CTT TTG GTG TT	GGT CCA GTT TTT TTT TTT TTT TCT GAA
miR-29b-3p	CAG TAG CAC CAT TTG AAA TCA GT	GGC TCA GTT TTT TTT TTT TTT TAA CAC
miR-96-5p	CAG TTT GGC ACT AGC ACA TTT TT	GGT CCA GTT TTT TTT TTT TTT TAG CAA
miR-182-5p	GCA GTT TGG CAA TGG TAG AAC	GTC CAG TTT TTT TTT TTT TTT GTG TGA
miR-183-5p	GCA GTA TGG CAC TGG TAG AAT	GGT CCA GTT TTT TTT TTT TTT TAG TGA
miR-335-3p	CGC AGT TTT TCA TTA TTG CTC CT	GTC CAG TTT TTT TTT TTT TTT GGT CA
miR-130a-5p	GCT CTT TTC ACA TTG TGC	GCG CAG TTT TTT TTT TTT TTT AGT A
miR-148a-5p	G CAG AAA GTT CTG AGA CA	CAG TTT TTT TTT TTT TTT AGT CGG AG
miR-429	GCA GTA ATA CTG TCT GGT AAA A	G CAG TTT TTT TTT TTT TTT ACG GT
Primers for measuring levels of precursor miRs (only DDSMs)		
Identifier	**Forward primer (5' to 3')**	**Reverse primer (5' to 3')**
miR-182	GAG CTG CTT GCC TCC	CCC ATA GTT GGC AAG TCT AG
miR-183	CCG CAG AGT GTG ACT CCT	TCG TGG ATC TGT CTC TGC TC
Primers for measuring levels of primary miRs (only DDSMs)		

Identifier	Forward primer (5' to 3')	Reverse primer (5' to 3')
miR-182	GTC TCT GTC TCT TCC TCA G	CCC ATA GTT GGC AAG TCT AG
miR-183	GGT AGA GAC CGT AGC AG	CCA GTG CCA TAC ACA GAA C
Primers for measuring levels of control RNA		
Identifier	**Forward primer (5' to 3')**	**Reverse primer (5' to 3')**
U6snRNA	CTC GCT TCG GCA GCA CA	AAC GCT TCA CGA ATT TGC GT

Table S13: Patient information pertaining to the Indian cohort
(presented in MS Excel format)

Table S13
Click here to download Table S13

Journal of Cell Science • Supplementary information

Table S14: Statistical analysis performed in this study

Figure number	Statistical Analysis performed	Software used
Figure 1B	Kruskal-Wallis test	GraphPad Prism
Figure 1C	Wilcoxon test	GraphPad Prism
Figure 1D	Log-rank (Mantel-Cox) test	GraphPad Prism
Figure 2B	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 2C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 2E	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 2F	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3A	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3B	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3D	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3E	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3F	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3G	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3H	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3I	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3J	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 3K	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 4A	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 4E	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 4F	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 4H	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 5B	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 5C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 5D	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 5E	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 5F	Student's t-test, two tailed, paired	GraphPad Prism
Figure 5H	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 5I	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 6A	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 6B	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 6C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 6D	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 6E	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 6F	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 6G	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 6H	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 6K	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7B	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7D	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7E	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7F	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7G	Student's t-test, two tailed, unequal variance	Microsoft Excel

Figure 7H	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7I	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7J	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7L	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 7N	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure 8B	Wilcoxon test	GraphPad Prism
Figure 8D	Wilcoxon test	GraphPad Prism
Figure 8F	Wilcoxon test	GraphPad Prism
Figure 8G	Spearman Correlation	GraphPad Prism
Figure S2B	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S2C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S3B	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S3E	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S4B	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S4C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S4D	Student's t-test, two tailed, paired	GraphPad Prism
Figure S4F	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S5C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S5D	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S5G	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S6A	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S6C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S6D	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S6E	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S6G	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S6H	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S6I	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S7C	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S8G	Student's t-test, two tailed, unequal variance	Microsoft Excel
Figure S8H	Student's t-test, two tailed, unequal variance	Microsoft Excel

References

Bainor, A. J., Saini, S., Calderon, A., Casado-Polanco, R., Giner-Ramirez, B., Moncada, C., Cantor, D. J., Ernlund, A., Litovchick, L. and David, G. (2018). The HDAC-Associated Sin3B Protein Represses DREAM Complex Targets and Cooperates with APC/C to Promote Quiescence. *Cell Rep* **25**, 2797-2807 e8.

Balbinot, C., Vanier, M., Armant, O., Nair, A., Penichon, J., Soret, C., Martin, E., Saandi, T., Reimund, J. M., Deschamps, J. et al. (2017). Fine-tuning and autoregulation of the intestinal determinant and tumor suppressor homeobox gene CDX2 by alternative splicing. *Cell Death Differ* **24**, 2173-2186.

Gao, S., Alarcon, C., Sapkota, G., Rahman, S., Chen, P. Y., Goerner, N., Macias, M. J., Erdjument-Bromage, H., Tempst, P. and Massague, J. (2009). Ubiquitin ligase Nedd4L targets activated Smad2/3 to limit TGF-beta signaling. *Mol Cell* **36**, 457-68.

Hu, P., Beresten, S. F., van Brabant, A. J., Ye, T. Z., Pandolfi, P. P., Johnson, F. B., Guarente, L. and Ellis, N. A. (2001). Evidence for BLM and Topoisomerase IIIalpha interaction in genomic stability. *Hum Mol Genet* **10**, 1287-98.

Kharat, S. S., Tripathi, V., Damodaran, A. P., Priyadarshini, R., Chandra, S., Tikoo, S., Nandhakumar, R., Srivastava, V., Priya, S., Hussain, M. et al. (2016). Mitotic phosphorylation of Bloom helicase at Thr182 is required for its proteasomal degradation and maintenance of chromosomal stability. *Oncogene* **35**, 1025-38.

Labbe, E., Silvestri, C., Hoodless, P. A., Wrana, J. L. and Attisano, L. (1998). Smad2 and Smad3 positively and negatively regulate TGF beta-dependent transcription through the forkhead DNA-binding protein FAST2. *Mol Cell* **2**, 109-20.

Qiu, Y., Stavreva, D. A., Luo, Y., Indrawan, A., Chang, M. and Hager, G. L. (2011). Dynamic interaction of HDAC1 with a glucocorticoid receptor-regulated gene is modulated by the activity state of the promoter. *J Biol Chem* **286**, 7641-7.