












ARTICLE

<https://doi.org/10.1038/s41467-019-13313-z>

OPEN

LncRNA REG1CP promotes tumorigenesis through an enhancer complex to recruit FANCI helicase for REG3A transcription

Hamed Yari ^{1,7}, Lei Jin^{1,7}, Liu Teng², Yufang Wang³, Yongyan Wu⁴, Guang Zhi Liu², Wei Gao ⁴, Jin Liang⁵, Yanfeng Xi⁵, Yu Chen Feng ¹, Chunming Zhang⁴, Yuan Yuan Zhang ¹, Hessam Tabatabaee ¹, Ting La ¹, Rui Hong Yang⁵, Fu Hua Wang⁵, Xu Guang Yan ¹, Margaret Farrelly¹, Rodney Scott ¹, Tao Liu ⁶, Rick F. Thorne ^{1,2}, Su Tang Guo^{5*} & Xu Dong Zhang ^{1,2*}

Protein products of the *regenerating islet-derived (REG)* gene family are important regulators of many cellular processes. Here we functionally characterise a non-protein coding product of the family, the long noncoding RNA (lncRNA) REG1CP that is transcribed from a DNA fragment at the family locus previously thought to be a pseudogene. REG1CP forms an RNA-DNA triplex with a homopurine stretch at the distal promoter of the *REG3A* gene, through which the DNA helicase FANCI is tethered to the core promoter of *REG3A* where it unwinds double stranded DNA and facilitates a permissive state for glucocorticoid receptor α (GR α)-mediated *REG3A* transcription. As such, REG1CP promotes cancer cell proliferation and tumorigenicity and its upregulation is associated with poor outcome of patients. REG1CP is also transcriptionally inducible by GR α , indicative of feedforward regulation. These results reveal the function and regulation of REG1CP and suggest that REG1CP may constitute a target for cancer treatment.

¹School of Biomedical Sciences and Pharmacy, School of Medicine and Public Health, School of Environmental and Life Sciences, The University of Newcastle, Newcastle, NSW 2308, Australia. ²Translational Research Institute, Henan Provincial People's Hospital, The Academy of Medical Sciences, Zhengzhou University, Zhengzhou, Henan 450053, China. ³Department of Pathophysiology, School of Preclinical and Forensic Medicine, Sichuan University, Chengdu, Sichuan 610041, China. ⁴Department of Otolaryngology, Shanxi Key Laboratory of Otorhinolaryngology Head and Neck Cancer, the first affiliated hospital, Shanxi Medical University, Taiyuan, Shanxi 030001, China. ⁵Department of Molecular Biology, Department of Pathology, Shanxi Cancer Hospital and Institute, Shanxi Medical University, Taiyuan, Shanxi 030013, China. ⁶Children's Cancer Institute Australia for Medical Research, University of New South Wales, Sydney, NSW 2750, Australia. ⁷These authors contributed equally: Hamed Yari, Lei Jin. *email: sutangguo@aliyun.com; Xu.Zhang@newcastle.edu.au

Regenerating islet-derived (REG) protein family members are involved in regulation of many cellular processes, including protection against cell death and promotion of cell proliferation¹. As such, their dysregulation is associated with various pathological conditions such as inflammation and cancer². There are five REG family protein-coding genes in humans, *REG protein 1 α* (*REG1A*), *REG protein 1 β* (*REG1B*), *REG protein 3 α* (*REG3A*) and *REG protein 1 γ* (*REG3G*) all clustered closely on chromosome 2p12 along with *REG protein 4* (*REG4*) situated on chromosome 1p12¹. Intriguingly, a segment of DNA located between the *REG1B* and *REG3A* genes on the opposite strand to *REG1A* that was previously thought to be a pseudogene has recently been annotated to represent a long noncoding RNA (lncRNA) termed REG family member 1 gamma pseudogene (*REG1CP*; HUGO Gene Nomenclature Committee: 9953)³. However, the biological function of *REG1CP* has not been defined.

lncRNAs commonly function to establish intermolecular interactions with other biomolecules, including proteins, DNAs, and other RNAs. This enables lncRNAs to act as tethers, guides, decoys or scaffolds to accomplish diverse arrays of biological functions⁴. For example, many lncRNAs interact with protein-coding genes and regulate their expression⁵. One emerging mode of the regulation involves lncRNAs forming RNA–DNA triplex structures, acting as “local address codes” to target chromatin-modifying enzymes to specific DNA sequences⁶. This in turn leads to alterations in chromatin structure and regulation of gene transcription as has been demonstrated for a number of lncRNAs such as *Khps1*, *MEG3* and *PARTICLE7–9*. Although functions of lncRNA–DNA triplexes are only beginning to be understood, bioinformatic prediction tools have revealed the existence of a large number of triplex-forming motifs across the genome⁸, suggesting that triplex-facilitated gene regulation could be far more common than currently appreciated.

Here, we present evidence that the lncRNA *REG1CP* binds distally to the transcription start site (TSS) of the *REG3A* gene through forming an RNA–DNA triplex. In concert with this, *REG1CP* binds to and recruits the RecQ helicase family member Fanconi anemia group J protein (*FANCI*) [also known as *BRCA1*-interacting helicase (*BRIP1*)] to the core promoter of *REG3A*, where it unwinds double-stranded DNA and derepresses transcriptional inhibition, resulting in a permissive state for *REG3A* transcription by glucocorticoid receptor α (*GR α*), the same transcription factor that drives *REG1CP* transcription. Thus, *REG1CP* is essential for the assembly of an enhancer complex that regulates *REG3A* transcription. Moreover, we demonstrate the functional importance of this mechanism, showing that *REG1CP* promotes cancer cell proliferation and tumorigenicity through activation of *REG3A*.

Results

REG1CP is frequently upregulated in colorectal cancer cells.

We compared lncRNA expression profiles using lncRNA arrays between laser capture micro-dissected (LCM) colon cancer cells of surgically resected primary colon cancer tissues and paired adjacent normal colon epithelial cells from five patients (Supplementary Table 1)¹⁰. The results revealed that *REG1CP* was the most prominently upregulated lncRNAs in colon cancer cells (Supplementary Table 2). This was subsequently verified by qPCR along with the lncRNAs *CCAT1* and *CCAL* as controls known to be upregulated in colon cancer (Fig. 1a)^{11,12}. Moreover, qPCR analysis of two independent cohorts (cohort 1 & 2) of formalin-fixed paraffin-embedded (FFPE) colon cancer tissues in comparison with paired normal mucosa also showed that *REG1CP* was upregulated in the majority of colon cancers (Fig. 1b, c and

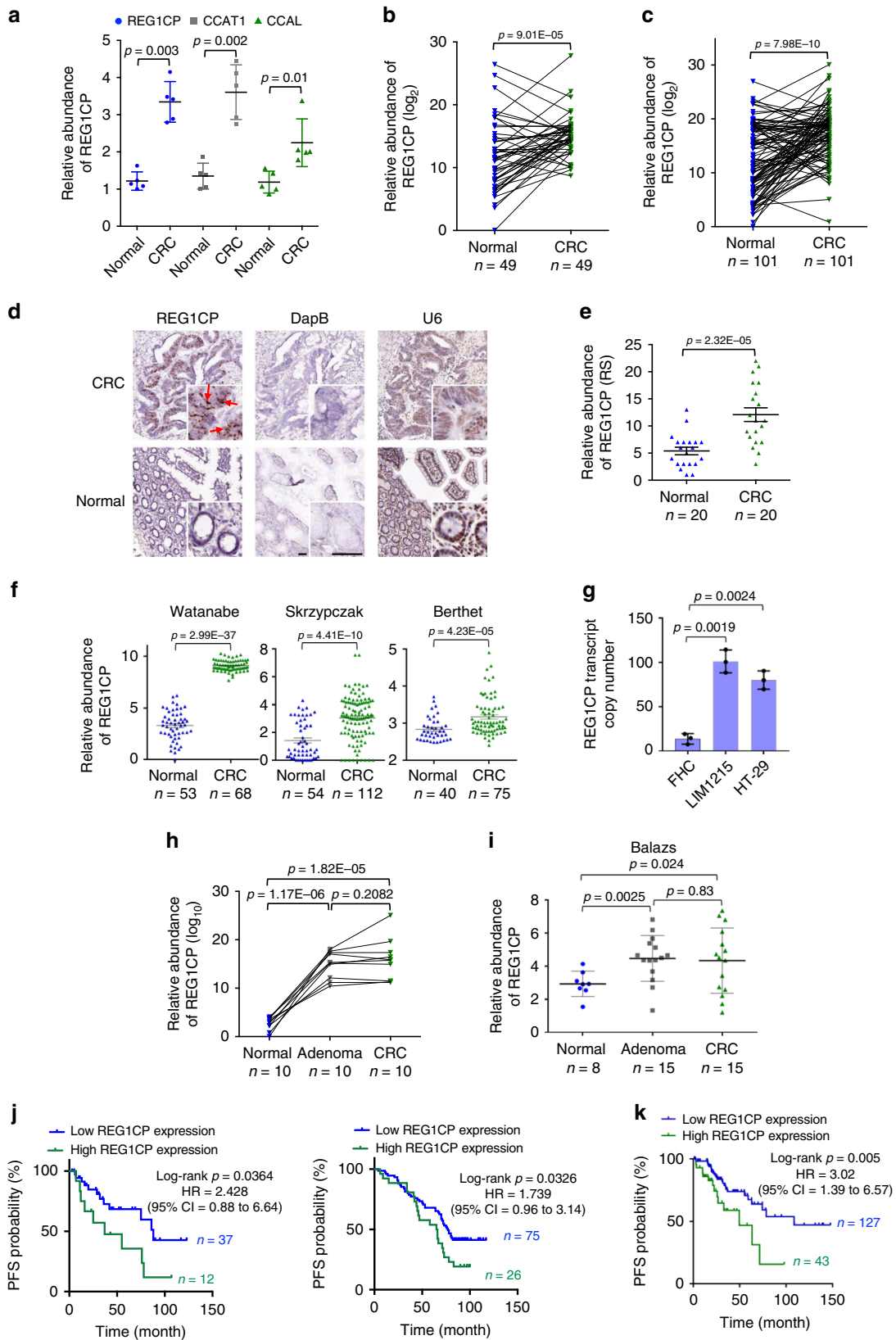
Supplementary Tables 3, 4). This was confirmed in randomly selected paired tissues from cohort 2 using in situ hybridization (ISH) (Fig. 1d, e). In support of these observations, analysis of datasets acquired from the R2 Genomics Analysis and Visualization Platform (R2: <http://r2.amc.nl>) revealed that *REG1CP* was similarly increased in colon cancers compared with normal colon tissues (Fig. 1f). Indeed, absolute quantitation using digital droplet PCR (ddPCR) demonstrated that there were ~80 and ~100 *REG1CP* molecules per HT-29 and LIM1215 colon cancer cell, respectively, compared with ~15 *REG1CP* molecules per normal colon epithelial (FHC) cell (Fig. 1g).

We next analysed the relationship between *REG1CP* expression, colon cancer development and progression. There were no significant differences in *REG1CP* levels among colon cancers of different clinicopathological groups defined by tumour stage and grade as well as gender (Supplementary Tables 3, 4). Similarly, there were no significant differences in *REG1CP* expression between tumours from patients stratified according to their median age at diagnosis (Supplementary Tables 3, 4). Moreover, no significant changes were found between *REG1CP* expression in colon cancer vs. colon adenoma (Fig. 1h, i). In contrast, *REG1CP* levels appeared higher in colon adenomas and dysplastic colon mucosa compared with paired normal mucosa (Fig. 1h, i), proposing that upregulation of *REG1CP* is an early event during colorectal tumorigenesis. Of note, high levels of *REG1CP* were associated with poorer progression-free survival (PFS) in colon cancer patients from both cohort 1 and 2 (Fig. 1j). Consistently, analysis of clinical colorectal cancer (CRC) data obtained from the TCGA showed that patients with colon cancers expressing higher levels of *REG1CP* had worse PFS (Fig. 1k).

REG1CP promotes CRC tumorigenesis through activating *REG3A*.

The gene encoding *REG1CP* is located between the *REG1B* and *REG3A* genes on the opposite strand to *REG1A* on chromosome 2.p12 (Supplementary Fig. 1A, B). This proximity suggests that *REG1CP* may have a role in regulating the expression of the REG family members¹³. Indeed, shRNA silencing of *REG1CP* reduced *REG3A* expression at both the transcript and protein levels in LIM1215 and HT-29 colon cancer cells that expressed relatively high levels of *REG1CP* among a panel of colon cancer cell lines (Fig. 2a–c and Supplementary Fig. 1C). However, it did not impinge on the expression of the *REG1A* nor *REG1B* (Fig. 2b). Co-introduction of a shRNA-resistant *REG1CP* mutant (*REG1CP-R*) reversed the inhibitory effect of *REG1CP* shRNA on *REG3A* expression (Fig. 2d), demonstrating the specificity of the *REG1CP* shRNAs and confirming the role of *REG1CP* in promoting *REG3A* expression. In accordance, *REG1CP* overexpression increased *REG3A* levels without affecting *REG1A* and *REG1B* expression in SW480 and COLO 205 cells that had relatively low levels of endogenous *REG1CP* (Fig. 2e, f and Supplementary Fig. 1C).

The turnover rates of *REG3A* mRNA remained similar in cells with or without *REG1CP* silenced (Supplementary Fig. 1D), suggesting that *REG1CP* promotes *REG3A* mRNA expression through a transcriptional increase^{14,15}. In support, analysis of histone transcriptional activation marks using chromatin immunoprecipitation (ChIP) assays showed that *REG1CP* silencing reduced the association of H3K4me3 with the promoter of *REG3A*, whereas *REG1CP* overexpression increased the association (Fig. 2g, h). Conversely, *REG1CP* knockdown increased binding of the transcriptional repression mark H3K27me3 to the *REG3A* promoter, whereas overexpression of *REG1CP* reduced binding (Fig. 2i, j). Knockdown of *REG1CP* did not alter the binding of H3K4me3 to either the *REG1A* or *REG1B* promoters (Fig. 2g). Therefore, the levels of *REG1CP* selectively dictate the



chromatin status of the *REG3A* promoter. Indeed, the levels of REG1CP were positively correlated with *REG3A* mRNA expression in our two cohorts (cohort 1 & 2) of CRC tissues and in a published CRC dataset acquired from the R2 platform (Fig. 2k, l). In contrast, there was no correlations between REG1CP

expression and *REG1A*, and *REG1B* mRNAs (Supplementary Fig. 1E).

As *REG3A* is mitogenic^{16,17}, we examined whether REG1CP has a similar effect on cell division. Silencing of REG1CP in LIM1215 and HT-29 cells reduced 5-bromo-2'-deoxyuridine

Fig. 1 REG1CP is upregulated in CRC cells and its high expression is associated with poor patient outcome. **a** Quantitation of REG1CP, CCAT1 and CCAL in LCM colon cancer cells and paired normal colon epithelial cells using qPCR. $n = 5$ paired biologically independent samples. **b, c** Quantitation of REG1CP expression in FFPE CRC tissues and paired normal mucosa. $n = 49$ (**b**) or $n = 101$ (**c**) paired biologically independent samples. **d** REG1CP was expressed at higher levels in FFPE colon cancers compared with normal colon tissues as shown in ISH. $n = 20$ paired biologically independent samples. Scale bar, 100 μm . Dihydrodipicolinate reductase (DapB) as a negative control and RNA, U6 Small Nuclear 1 (U6) as a positive control. **e** Quantitation of REG1CP expression according to results shown in (**d**). $n = 20$ paired biologically independent samples. RS, reactivation score. **f** Comparison of REG1CP expression between CRCs and normal colon tissues in the published datasets acquired from R2. **g** Absolute quantitation of REG1CP in the cell lines using ddPCR. $n = 3$ independent experiments. **h** Quantitation of REG1CP in FFPE colon adenoma, normal colon mucosa and colon cancer tissues. **i** Comparison of REG1CP expression between colon adenoma, normal colon mucosa and colon cancer tissues in a published dataset acquired from R2. **j** Kaplan–Meier analysis of the probability of PFS of CRC patients in cohort 1 (left) and cohort 2 (right) after surgical excision using the high quartile of REG1CP levels as the cut-off. **k** Kaplan–Meier analysis of the probability of PFS of CRC patients in a published dataset acquired from the TCGA using the high quartile of REG1CP levels as the cut-off. $n = 3$ independent experiments. Data are presented as the Mean \pm SEM (**a, e, f, g, i**), the Mean (**b, c, h**) or representatives (**d**). Statistical significance was calculated using a two-tailed *t*-test.

(BrdU) incorporation, arrested cell cycle progression at G0/G1 phase and retarded clonogenicity (Fig. 3a–c and Supplementary Fig. 2A, B). Nevertheless, inhibition of BrdU incorporation and cell cycle arrest caused by REG1CP silencing was diminished by overexpression of REG3A (Fig. 3a, b and Supplementary Fig. 2B). On the other hand, overexpression of REG1CP in SW480 and COLO 205 caused promotion of cell proliferation that was abolished by REG3A silencing (Fig. 3d, e). Therefore, REG1CP plays a role in promoting cell proliferation through REG3A.

To better facilitate investigation of REG1CP function *in vivo*, we established LIM1215 and HT-29 sublines (LIM1215-shREG1CP and HT-29-shREG1CP) with REG1CP conditionally knocked down in response to doxycycline (Dox). The addition of Dox readily reduced REG1CP expression and inhibited cell proliferation (Supplementary Fig. 2C, D). After Dox was withdrawn, REG1CP levels recovered and cell proliferation rates were restored as expected (Supplementary Fig. 2C, D). Treatment of nu/nu mice bearing tumours established by subcutaneous transplantation of LIM1215-shREG1CP and HT-29-shREG1CP cells with Dox retarded tumour growth (Fig. 3f, g and Supplementary Fig. 2E). This was associated with reduction in the proportion of Ki-67-expressing cells (Supplementary Fig. 2F). Nevertheless, cessation of Dox treatment led to recovery of the growth of LIM1215-shREG1CP and HT-29-shREG1CP xenografts (Fig. 3g). Moreover, overexpression of REG3A reversed the reduction in growth of LIM1215-shREG1CP xenografts in nu/nu mice treated with Dox (Fig. 3h), consistent with the notion that REG1CP functions through REG3A to sustain cancer cell proliferation *in vivo*.

REG1CP forms an RNA–DNA triplex with the REG3A gene.

REG1CP was highly enriched in the nucleus as shown by ISH analysis of HT29 cells grown on coverslips and by qPCR analysis of subcellular fractions (Fig. 4a and Supplementary Fig. 3A), pointing to the possibility that REG1CP interacts with the REG3A DNA to regulate its transcription. Bioinformatics analysis revealed that the REG3A gene comprised a homopurine region upstream of the TSS (–6176/–6195) that can potentially form RNA–DNA triplexes [triplex-forming region (TFR)] (Supplementary Fig. 3B). Intriguingly, this TFR complements to a fragment enriched of triplex-forming oligonucleotides (TFOs) contained in REG1CP (Supplementary Fig. 3B). To test whether REG1CP forms RNA–DNA complexes with the REG3A DNA, we employed a cell-free assay system. Our initial findings showed that *in vitro*-synthesized biotin-labelled REG1CP precipitated a DNA fragment containing the putative TFR of REG3A (Fig. 4b). This association was confirmed using electrophoretic mobility shift assays (EMSA) (Fig. 4c). However, the association was diminished when the TFOs within REG1CP or TFR in REG3A

were mutated (Fig. 4d). Furthermore, biotin-labelled REG1CP failed to precipitate a DNA fragment of REG3A (–4965/–5172) that did not contain the TFR (Fig. 4e). Notably the inclusion of 7-deaza-purine nucleotides, known to obviate Hoogsteen base-pairing required for RNA–DNA complexing, also diminished REG1CP–REG3A DNA association (Fig. 4e)⁹. In contrast, the addition of RNase H, which specifically hydrolyses the phosphodiester bonds of RNA when hybridized to DNA, did not affect the association (Fig. 4f). *In vitro*-synthesized REG1CP also precipitated the REG3A DNA from nuclear extracts (Fig. 4g). In addition, the TFR of endogenous REG3A could be co-pulled down with endogenous REG1CP using chromatin isolation by RNA purification (ChIRP) assays (Fig. 4h). Collectively, these results reveal an association between REG1CP and REG3A DNA with support for the formation of an RNA–DNA triplex.

To examine whether the association between REG1CP and REG3A is direct or alternatively is mediated via protein interactions, we treated nuclear extracts from LIM1215 and HT-29 cells with proteinase K and then monitored triplex formation. Treatment with proteinase K did not disrupt the RNA–DNA triplexes formed by REG1CP and REG3A (Fig. 4g), demonstrating the binding between REG1CP and REG3A is direct and not protein dependent. To confirm this, we employed biotinylated TFOs of REG1CP carrying a psoralen moiety at the 5' end to enable photoactivated fixation of nucleic acid interactions without crosslinking proteins bound to DNA or RNA⁹. Introduction of the psoralen-modified biotinylated TFOs into LIM1215 and HT-29 cells readily demonstrated an association between REG1CP and the TFR in REG3A after UV treatment whereas little association was found without photoactivation or when cells were introduced with a biotinylated control oligonucleotide not complementary to the TFR in REG3A (Fig. 4i). These results further consolidate that REG1CP and REG3A directly form triplexes without the involvement of proteins.

We then applied these findings to the context of colon cancer cell proliferation. Introduction of a REG1CP mutant lacking the TFOs (REG1CP- Δ TFO) into LIM1215 and HT-29 cells with endogenous REG1CP knocked down by shRNA had no effect on REG3A levels or proliferation (Fig. 4j, k), whereas introduction of the shRNA-resistant REG1CP (REG1CP-R) increased REG3A levels and reversed the inhibition of cell proliferation caused by REG1CP knockdown (Fig. 4j, k). Taken together with preceding data, our findings suggest that the formation of an RNA–DNA triplex is required for the REG1CP effects on REG3A expression and cell proliferation. In support, the entire REG1CP but not REG1CP without the TFOs caused the increase in cell proliferation (Supplementary Fig. 3C).

Using bioinformatics we also carried out a genome-wide search for potential TFRs complementary to the TFOs of REG1CP. This analysis identified numerous TFRs across the genome (Supplementary Data 1), although examining a random selection of 10

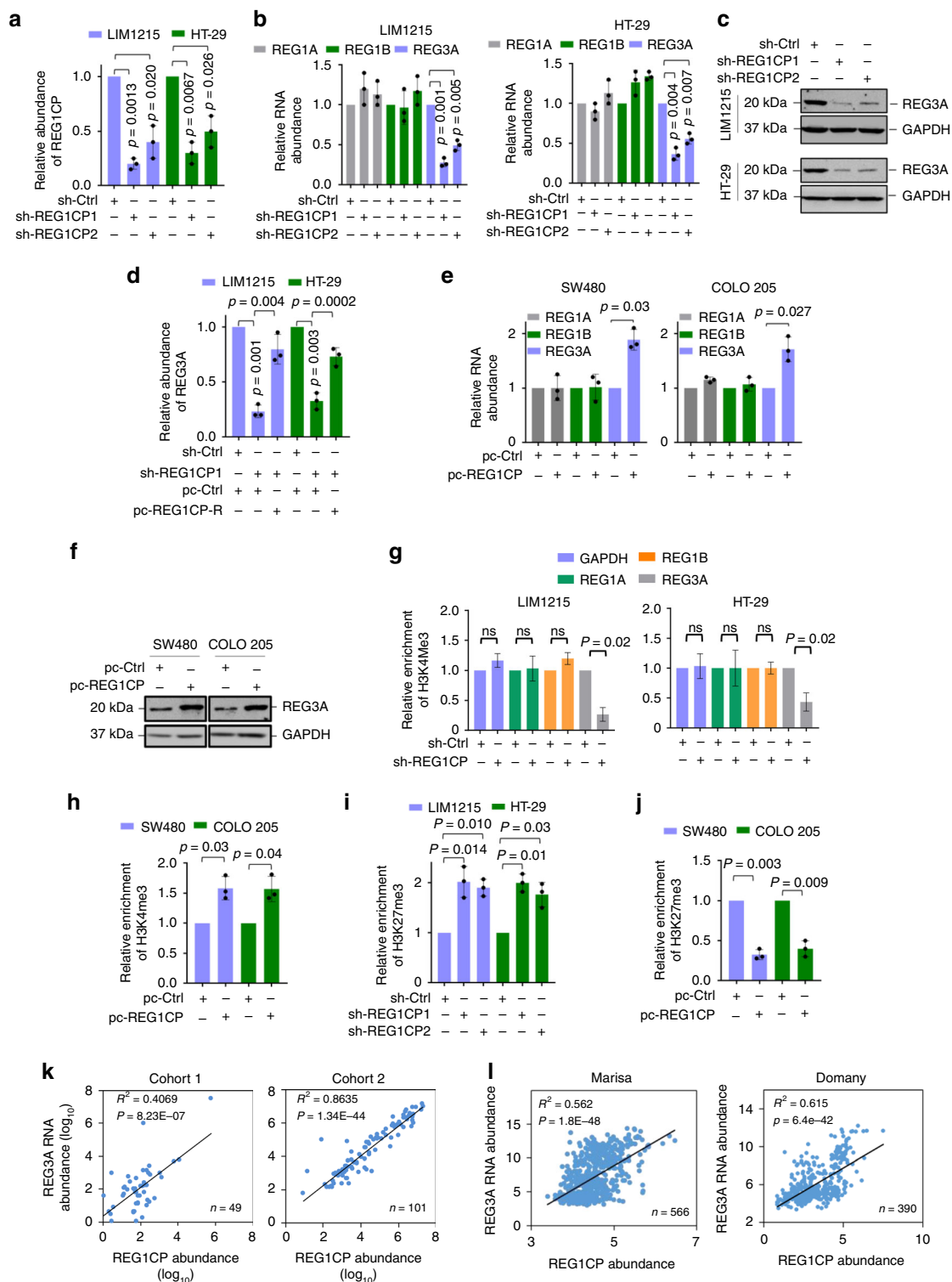


Fig. 2 REG1CP promotes transcription of *REG3A*. **a–c** Silencing of REG1CP (**a**) decreased *REG3A* mRNA (**b**) and protein levels (**c**). **d** Co-introduction of a shRNA-resistant REG1CP mutant (REG1CP-R) diminished the inhibitory effect of REG1CP silencing on *REG3A* mRNA expression. **e, f** Overexpression of REG1CP increased *REG3A* mRNA (**e**) and protein (**f**) levels. **g, h** Overexpression of REG1CP increased *REG3A* mRNA (**e**) and protein (**f**) levels. **g, h** Silencing of REG1CP decreased (**g**) whereas overexpression of REG1CP (**h**) increased enrichment of H3K4me3 to *REG3A* promoter, but not to *REG1A* or *REG1B* promoter (**g**). Relative enrichment of H3K4me3 to the -7/-109 fragment of *REG3A* promoter was measured using ChIP-qPCR assays in LIM1215 and HT-29 cells after silencing (**g**) or in SW480 and COLO205 cells after overexpression (**h**) of REG1CP. **i, j** Silencing of REG1CP increased (**i**) whereas overexpression of REG1CP decreased (**j**) enrichment of H3K27me3 to *REG3A* proximal promoter. Relative enrichment of H3K27me3 at the -7/-109 fragment of *REG3A* promoter was measured using ChIP-qPCR assays in LIM1215 and HT-29 cells after silencing (**i**) or in SW480 and COLO205 cells after overexpression (**j**) of REG1CP. **k, l** Regression analysis of the relationship between *REG3A* mRNA and REG1CP expression in two cohorts of colon cancer samples determined using qPCR (**k**) and in published datasets acquired from the R2 Genomics Analysis and Visualization Platform (**l**). $n = 3$ independent experiments unless specified differently. Data are presented as Mean \pm SEM (**a, b, d, e, g–j**) or representatives (**c, f**). Statistical significance was calculated using a two-tailed t-test unless specified.

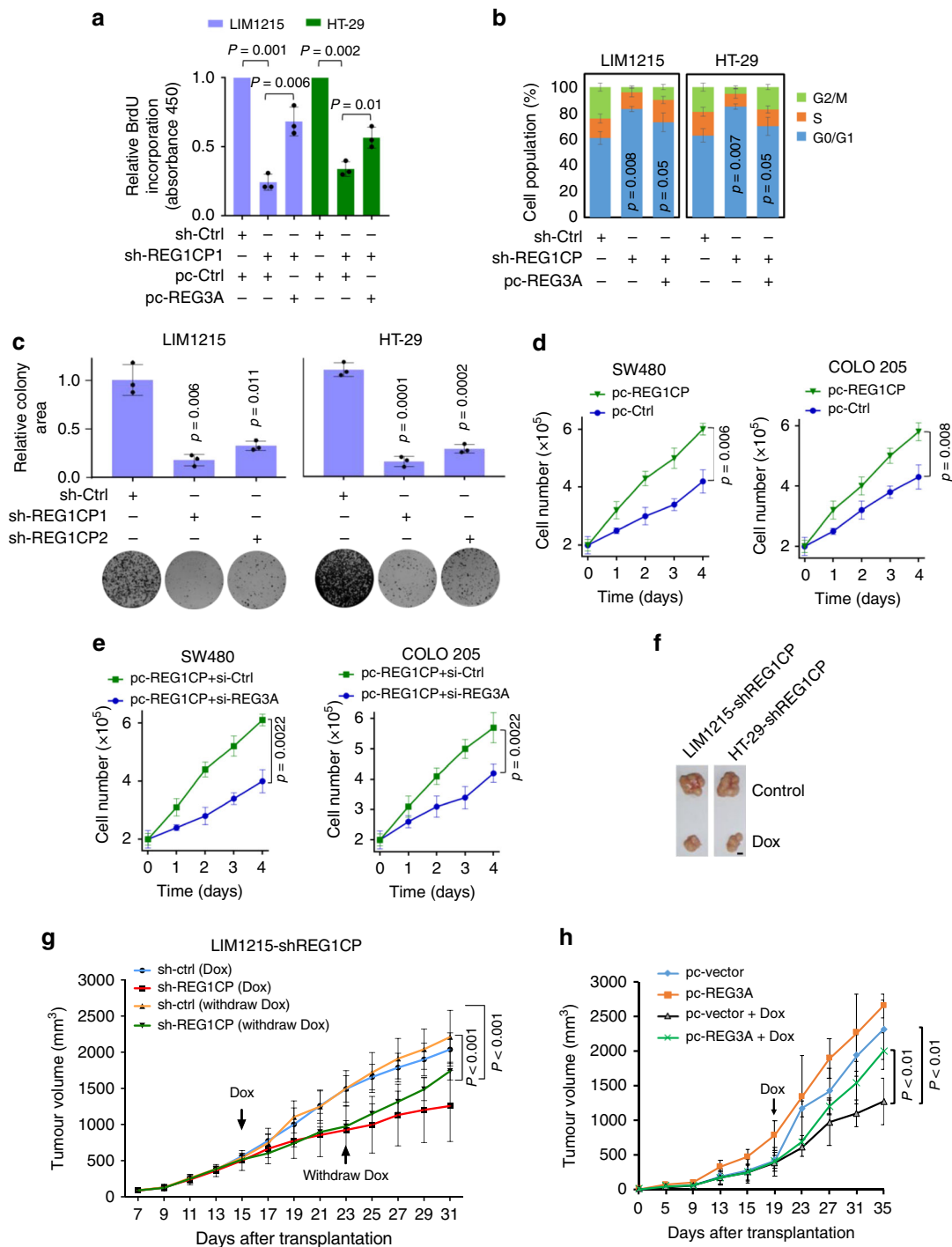


Fig. 3 REG1CP promotes cancer cell cycle progression and tumorigenicity through REG3A. **a** REG3A overexpression reversed the inhibitory effect of REG1CP silencing on BrdU incorporation. **b** REG3A overexpression reversed G0/G1 cell cycle arrest caused by REG1CP silencing. Relative cell cycle distribution was quantitated by propidium iodide staining using flow cytometry in cells transduced with the control or REG1CP shRNA1 with or without overexpression of REG3A. **c** Silencing of REG1CP inhibited clonogenicity. Relative clonogenicity of cells transduced with the control or REG1CP shRNAs was quantitated using ImageJ-plugin "ColonyArea". **d** Overexpression of REG1CP promoted cell proliferation. SW480 and COLO205 cells with or without overexpression of REG1CP were counted using an automated cell counter. **e** Silencing of REG3A reversed promotion of cell proliferation caused by REG1CP overexpression. SW480 and COLO205 cells overexpressing REG1CP transduced with the control or REG3A shRNA were counted using an automated cell counter. **f, g** Silencing of REG1CP inhibited colon cancer xenograft growth in nu/nu mice. Representative photographs (**f**) and growth curves (**g**) of LIM1215-shREG1CP xenografts in mice with or without treatment with doxycycline (Dox) and with or without Dox withdraw to induce silencing of REG1CP. $n = 6$ mice per group. **h** Overexpression of REG3A reversed the inhibitory effect of REG1CP silencing on colon cancer xenograft growth. Growth curves of LIM1215-shREG1CP xenografts with or without overexpression of REG3A in mice with or without treatment with doxycycline (Dox) to induce silencing of REG1CP. $n = 6$ mice per group. $n = 3$ independent experiments unless specified differently. Data are presented as Mean \pm SEM (**a, b, c** (upper), **d, e, g, h**) or representatives (**c** (lower) and **f**). Statistical significance was calculated using a two-tailed t-test.

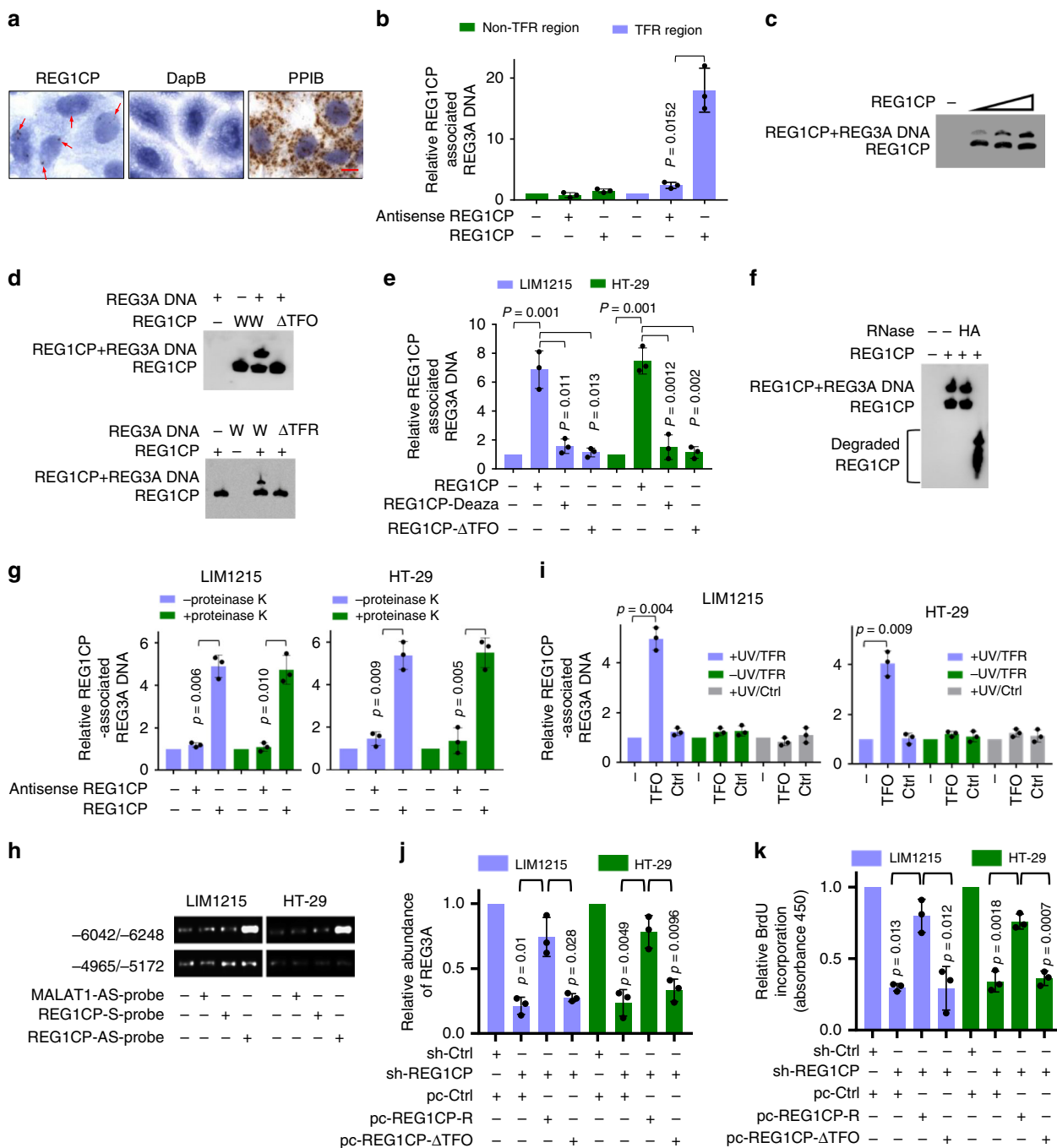


Fig. 4 REG1CP forms an RNA-DNA triplex with the REG3A gene. **a** Representative microphotographs of ISH staining using probes against REG1CP in HT29 cells. Dihydrodipicolinate reductase (DapB) as a negative control and peptidylprolyl isomerase B (PPIB) as a positive control. Scale bar: 20 μ m. **b** In vitro-synthesized biotin-labelled REG1CP pulled down a DNA fragment containing the TFR of the REG3A gene (-6072/-6248). **c** In vitro-synthesized biotin-labelled REG1CP at increasing concentrations was incubated with a DNA fragment containing the TFR of the REG3A gene followed by analysis using EMSA. **d** Deletion of the TFOs of REG1CP or the TFR of the REG3A gene abolished the association between REG1CP RNA and REG3A DNA as shown using EMSA. **e** In vitro-synthesized biotin-labelled wild-type REG1CP, a REG1CP mutant lacking the TFOs or wild-type REG1CP containing 7-deaza-purine nucleotides was incubated with a DNA fragment containing the TFR of the REG3A gene followed by pulldown with REG1CP sense RNA. **f** Treatment with RNase A abolished the binding of REG1CP to REG3A as shown using EMSA. **g** Nuclear extracts from LIM1215 and HT-29 cells were incubated with in vitro-synthesized biotin-labelled REG1CP or antisense RNA in the presence or absence of proteinase K. The baseline amount of REG3A DNA detected in the control assay (pulldown samples without RNA) was arbitrarily designated as one. **h** Endogenous REG3A DNA containing the TFR (-6072/-6248) was co-pulldown with endogenous REG1CP. LncRNA MALAT1 was included as a control. **i** UV treatment caused association between the TFR of REG3A and psoralen-modified biotinylated TFOs of REG1CP. Nuclear extracts from cells transfected with psoralen-modified biotinylated TFOs of REG1CP or a biotinylated RNA oligonucleotide without the TFOs with or without UV-crosslinking were subjected to pulldown with streptavidin beads. **(j, k)** Introduction of a shRNA-resistant REG1CP (REG1CP-R) but not a REG1CP mutant with the TFOs deleted (REG1CP- Δ TFO) into cells with endogenous REG1CP silenced increased REG3A expression **(j)** and promotes cell proliferation **(k)**. $n = 3$ independent experiments. Data are presented as Mean \pm SEM **(b, e, g, i-k)** or representatives **(a, c, d, f, h)**. Statistical significance was calculated using a two-tailed t -test.

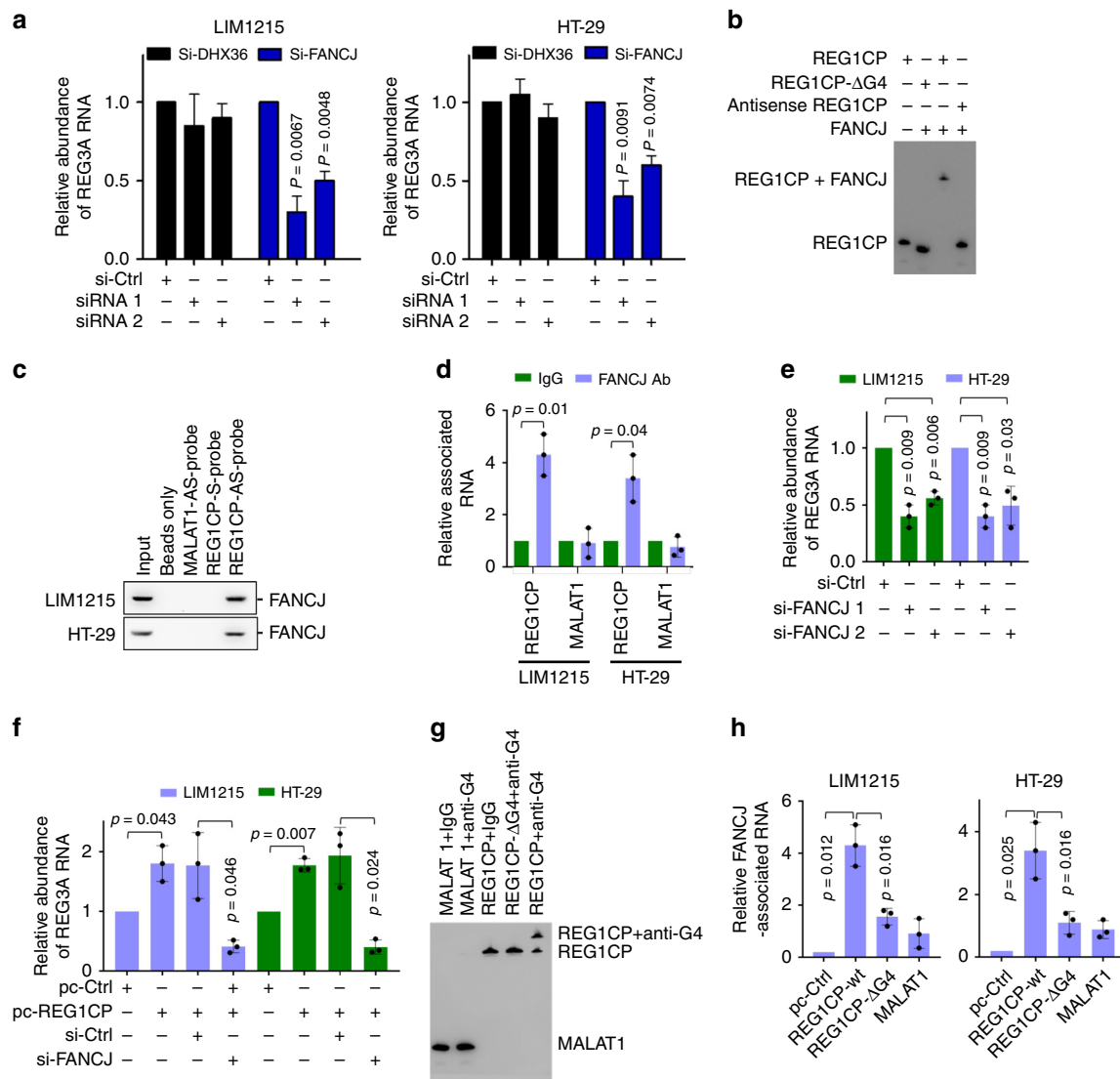


Fig. 5 REG1CP binds to FANCI. **a** Silencing of FANCI but not DHX36 inhibited REG3A mRNA expression. **b** FANCI bound to wild-type REG1CP but not a REG1CP mutant with the G-quadruplex deleted (REG1CP- Δ G4). In vitro-synthesized biotin-labelled REG1CP sense RNA (REG1CP) or antisense RNA (Antisense REG1CP) or REG1CP- Δ G4 was incubated with the immunoprecipitated FANCI from nuclear extracts of LIM1215 cells followed by EMSA. **c** Binding of FANCI to REG1CP was detected using RNA pulldown assays. Proteins co-pulled down with biotin-labelled REG1CP sense probe (REG1CP-S-probe) or REG1CP antisense probe (REG1CP-AS-probe) were analysed using western blotting. **d** Binding of FANCI to REG1CP was detected using RNA immunoprecipitation (RIP) assays. RNAs co-precipitated with an antibody against FANCI from nuclear extracts of LIM1215 and HT-29 cells were analysed using qPCR. **e** FANCI silencing reduced REG3A mRNA levels. Relative REG3A mRNA abundance was quantitated using qPCR in LIM1215 and HT-29 cells introduced with the control or FANCI siRNAs. **f** FANCI silencing diminished upregulation of REG3A by overexpression of REG1CP. Relative REG3A mRNA abundance was quantitated using qPCR in LIM1215 and HT-29 cells introduced with the control or FANCI siRNAs with or without overexpressing REG1CP. **g** An anti-G-quadruplex antibody bound to wild-type REG1CP but not a REG1CP mutant with the G-quadruplex deleted (REG1CP- Δ G4). In vitro-synthesized biotin-labelled REG1CP or REG1CP- Δ G4 were incubated with the isotype control (rabbit IgG) or an anti-FANCI antibody followed by analysis using EMSA. In vitro-synthesized biotin-labelled MALAT1 was included as a control. **h** FANCI bound to wild-type REG1CP but not REG1CP- Δ G4. LIM1215 and HT-29 cells with endogenous REG1CP stably silenced were introduced with exogenous REG1CP or REG1CP- Δ G4. FANCI-associated REG1CP was detected using RIP assays. $n = 3$ independent experiments. Data are presented as Mean \pm SEM (**a**, **d**, **e**, **f**, **h**) or representatives (**b**, **c**, **g**). Statistical significance was calculated using a two-tailed *t*-test.

TFR-containing genes from this list showed no evidence of changes in their expression in LIM1215 cells with and without REG1CP knockdown (Supplementary Fig. 3D). These observations implicate that regulation of REG3A expression by REG1CP through forming triplexes is highly specific.

REG1CP binds to FANCI. We analyzed the proteins that co-precipitated with biotin-labelled REG1CP by mass spectrometry. Among them were the DNA helicases, DEAH-Box Helicase 36

(DHX36) and FANCI that have been implicated in regulation of transcription (Supplementary Fig. 4A and Supplementary Table 5)^{18,19}. Depletion of FANCI but not DHX36 inhibited REG3A expression (Fig. 5a). We, therefore, focused on FANCI for its potential role in transcriptional activation of REG3A. The association between REG1CP and FANCI was readily detected using EMSAs (Fig. 5b) and was confirmed by RNA pulldown and RNA immunoprecipitation (RIP) assays (Fig. 5c, d). In contrast, there was no association between FANCI and the lncRNA MALAT1 included as a control (Fig. 5c, d). Instructively,

knockdown of FANCI inhibited REG3A expression and diminished upregulation of REG3A by REG1CP overexpression (Fig. 5e, f), suggesting that FANCI is involved in REG1CP-mediated transcriptional activation of REG3A.

Since FANCI can bind to G-quadruplexes of nucleic acids and an RNA G-quadruplex structure was identified in REG1CP (Supplementary Fig. 4B)²⁰, we investigated whether FANCI is physically associated with REG1CP via G-quadruplex interaction. Notably, the interaction between FANCI and REG1CP was abolished when the G-quadruplex was deleted (REG1CP- Δ G4) (Fig. 5g), suggesting that the G-quadruplex was necessary for binding of REG1CP with FANCI. In support, wild-type REG1CP but not REG1CP- Δ G4 was co-precipitated with FANCI when introduced into LIM1215 and HT-29 cells with endogenous REG1CP knocked down (Fig. 5h).

REG1CP tethers FANCI to REG3A promoter for its activation.

Analysis of an ENCODE dataset showed a DNase I hypersensitive site at the REG3A promoter (−2325/−2509 upstream the transcription start site) of HT-29 cells (Supplementary Fig. 5A) (<https://www.encodeproject.org>). The hypersensitivity site was validated in LIM1215 and HT-29 cells using PCR, thereby suggesting that the −2325/−2509 fragment may be the transcriptionally active region of the REG3A promoter (Fig. 6a)^{21,22}. Deletion of this fragment abolished the transcriptional activity of a REG3A reporter construct (Fig. 6b), confirming its importance in transcriptional activation of REG3A. Notably, FANCI bound to the −2325/−2509 fragment (Fig. 6c), and its binding required the association between REG1CP and the REG3A promoter since introduction of REG1CP-R into LIM1215 and HT-29 cells with endogenous REG1CP knocked down recovered FANCI binding to the REG3A promoter (Fig. 6d). In contrast, the expression of REG1CP- Δ TFO had no effect (Fig. 6d), collectively indicating that the REG1CP–REG3A RNA–DNA triplex is necessary for the binding of FANCI to the REG3A promoter.

We examined whether FANCI, as a helicase, is involved in unwinding double-stranded DNA of the REG3A promoter. Indeed, treatment of isolated DNA with a single-stranded DNA-specific DNase (ssDNase) caused degradation of the −2325/−2509 fragment of the REG3A promoter (Fig. 6e). This effect of the ssDNase was diminished in DNA isolated from LIM1215 and HT-29 cells with FANCI knocked down (Fig. 6e), suggesting that FANCI is necessary for unwinding double-stranded DNA of the fragment. This required REG1CP, as DNA from cells with REG1CP knocked down also displayed reduced sensitivity to the ssDNase (Fig. 6e). When REG1CP and FANCI were co-knocked down, the ability of the ssDNase to degrade the −2325/−2509 fragment was completely disabled (Fig. 6e). Collectively, these results suggest that the association between REG1CP and FANCI is important for tethering FANCI to the REG3A promoter where it unwinds double-stranded DNA of the active region of the REG3A promoter.

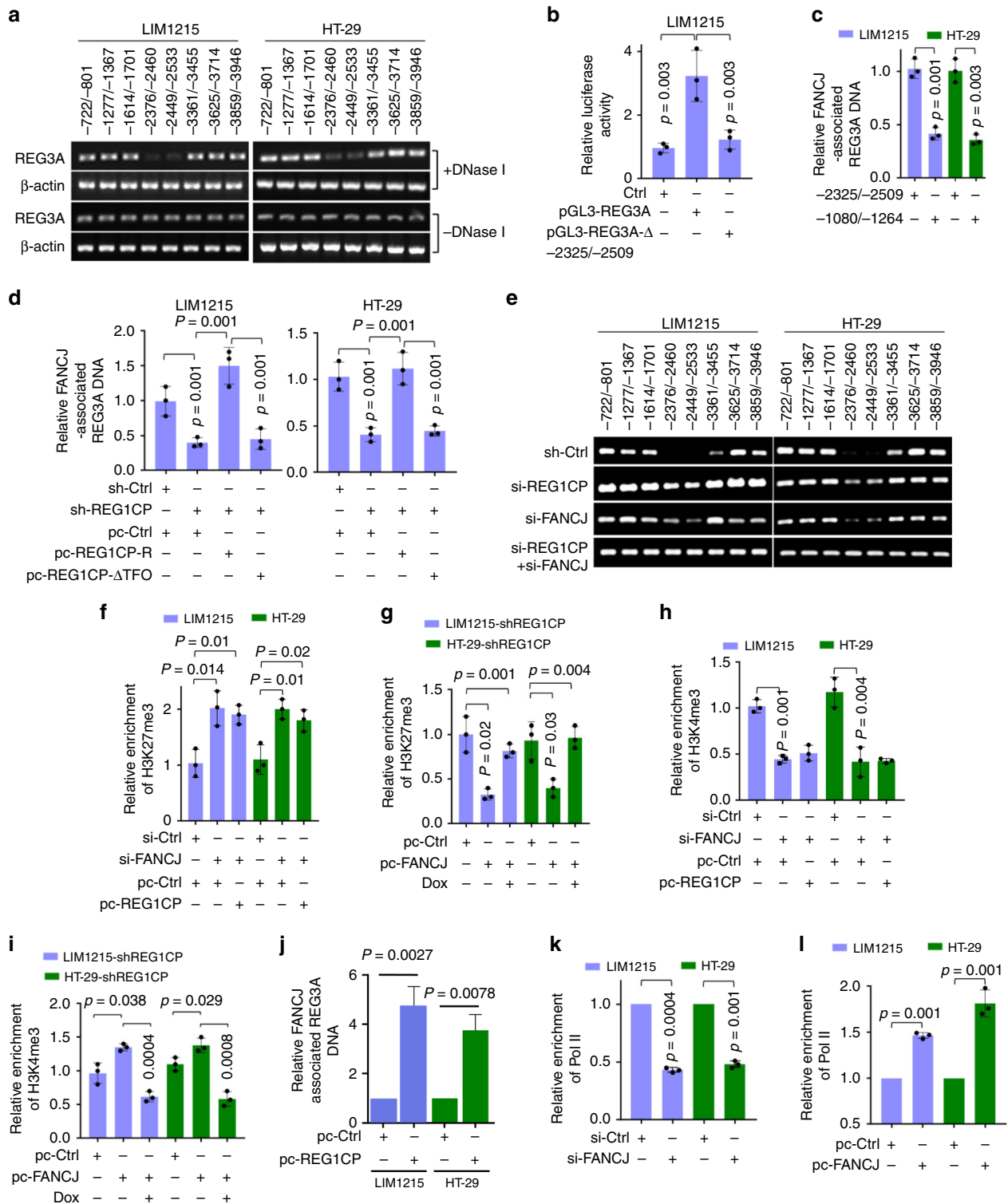
We investigated the functional significance of the REG1CP-mediated association between FANCI and the REG3A promoter. Downregulation of REG3A caused by FANCI knockdown was associated with an increase in H3K27me enrichment at the proximal promoter of REG3A, which was not affected by overexpression of REG1CP (Fig. 6f). On the other hand, overexpression of FANCI caused upregulation of REG3A and a decrease in H3K27me associated with the REG3A promoter (Fig. 6g and Supplementary Fig. 5B), which was however abolished by co-knockdown of REG1CP (Fig. 6g and Supplementary Fig. 5B). In contrast, knockdown of FANCI led to a decrease in the association between H3K4me3 and the REG3A promoter that was not affected by overexpression of REG1CP

(Fig. 6h). Conversely, overexpression of FANCI caused an increase in the H3K4me3 mark that was diminished by knockdown of REG1CP (Fig. 6i). These results suggest that tethering of FANCI to the REG3A promoter by REG1CP is necessary for forming a euchromatic region that promotes transcriptional activation of REG3A. Substantiating this, REG1CP knockdown reduced FANCI enrichment at the promoter of REG3A, whereas REG1CP overexpression increased enrichment (Fig. 6d, j). Furthermore, FANCI knockdown resulted in reduction in RNA polymerase II (Pol II) enrichment at the REG3A promoter, whereas overexpression of FANCI caused an increase in the association between Pol II and the promoter (Fig. 6k, l). Consistent with its effect on REG1CP-mediated transcriptional activation of REG3A, FANCI knockdown attenuated the increase in proliferation caused by REG1CP overexpression in LIM1215 and HT-29 cells (Supplementary Fig. 5C).

As DNA helicases are implicated in the unwinding of RNA–DNA triplex structures^{20,23}, we examined whether FANCI has any effect on the formation of REG1CP–REG3A RNA–DNA triplex. The addition of FANCI into a cell-free system only moderately reduced the triplex formation by in vitro-synthesized biotin-labelled REG1CP and a DNA fragment containing the putative TFR of REG3A (Supplementary Fig. 5D), suggesting that the REG1CP–REG3A RNA–DNA triplexes constantly exist regardless of the unwinding effect of FANCI that would ensure the transcriptional activation of REG3A through the complex. Consistently, neither knockdown nor overexpression of FANCI caused significant changes in the amount of REG1CP associated with REG3A TFR in LIM1215 and HT-29 cells (Supplementary Fig. 5E, F).

GRA transcriptionally activates REG3A and REG1CP. To gain further insights into how REG3A transcription is regulated, we interrogated its promoter for transcription factor binding sites using bioinformatics. This revealed multiple binding motifs for GR α localized to the −2325/−2509 fragment of REG3A promoter (Supplementary Fig. 6A). Indeed, GR α knockdown reduced REG3A levels (Fig. 7a and Supplementary Fig. 6B), suggesting that GR α is necessary for transcriptional activation of REG3A. Endorsing this, GR α knockdown inhibited the transcriptional activity of a REG3A reporter construct, whereas its overexpression enhanced reporter activity (Fig. 7b, c). Moreover, GR α was associated with the −2325/−2509 fragment (Fig. 7d). Nonetheless, this association was reduced by knockdown of FANCI or knockdown of REG1CP, and was further decreased by co-knockdown of FANCI and REG1CP (Fig. 7e), suggesting that FANCI and REG1CP cooperatively promote the binding of GR α to REG3A reporter. As further validation, knockdown of FANCI or REG1CP diminished the increase in the transcriptional activity of the REG3A reporter construct and reversed upregulation of endogenous REG3A caused by GR α overexpression (Fig. 7f). Thus the roles of FANCI and REG1CP are integral to the transcriptional activation of REG3A by GR α . Of note, despite the clear correlation between REG1CP and REG3A expression (Fig. 2k, l), there was no significant relationship between FANCI or GR α expression and the expression of REG3A as measured using immunohistochemistry (IHC) in CRC tissues (Supplementary Fig. 6C, D).

Interestingly, the active region of the REG1CP gene promoter also contains a consensus GR α -binding region (GR α -BR) (Supplementary Fig. 6E), which was co-precipitated with endogenous GR α (Fig. 7g). This region was required for REG1CP upregulation as the transcriptional activity of a REG1CP reporter construct was inhibited when the GR α -BR was deleted (Fig. 7h). Moreover, co-transfection of GR α selectively enhanced the transcriptional activity of the reporter with the GR α -BR, whereas knockdown of GR α diminished reporter activity (Fig. 7i and



Supplementary Fig. 6F). These data support the transactivation of *REG1CP* by GRα through the identified GRα-BR. In accordance, knockdown of GRα reduced the endogenous *REG1CP* levels, whereas overexpression of GRα caused *REG1CP* upregulation (Fig. 7j) and Supplementary Fig. 6G).

Discussion

The proteins encoded by *REG* gene family members are important regulators of many cellular processes and their roles in various diseases are being increasingly appreciated^{2,17,24}. In this

study, we demonstrate that *REG1CP*, the non-coding component of the *REG* gene cluster, fulfils an essential regulatory role in realising the actions of the *REG3A* gene. *REG1CP* was shown to promote cancer cell cycle progression and tumorigenicity through promoting transcription of *REG3A*. These findings are consistent with previous reports describing the function of *REG3A* and indeed *REG3A* has been proposed as an oncogene in various cancers including CRC^{16,17,25}. The increased *REG3A* expression is driven by feedforward regulation involving GRα that is responsible for transcriptional activation of *REG1CP* and *REG3A*.

Fig. 6 REG1CP tethers FANCI to the *REG3A* promoter. **a** the $-2325/-2509$ fragment of the *REG3A* gene was hypersensitive to DNase I. **b** Deletion (Δ) of the $-2325/-2509$ fragment abolished luciferase reporter activity of pGL3-basic based *REG3A* promoter constructs. **c** FANCI bound to the $-2325/-2509$ fragment of *REG3A* promoter as shown in ChIP assays. **d** The association between REG1CP and the *REG3A* gene was required for binding of FANCI to the *REG3A* promoter. Relative abundance of *REG3A* promoter associated with FANCI was measured using ChIP assays. **e** Silencing of REG1CP and FANCI reduced sensitivity of the $-2325/-2509$ fragment of *REG3A* promoter to a single-stranded DNA-specific DNase (ssDNase). **f** REG1CP overexpression did not affect the increase in H3K27me3 enrichment to *REG3A* promoter caused by FANCI silencing as measured using ChIP assays. **g** REG1CP silencing diminished reduction in H3K27me3 enrichment to *REG3A* promoter caused by overexpression of FANCI. Relative enrichment of H3K27me3 to *REG3A* promoter was measured using ChIP assays in LIM1215-shREG1CP and HT-29-shREG1CP cells with or without overexpression of FANCI in the presence or absence of doxycycline (Dox) to induce REG1CP silencing. **(h)** REG1CP overexpression did not affect the decrease in H3K4me3 enrichment to *REG3A* promoter caused by FANCI silencing measured using ChIP assays. **i** REG1CP silencing diminished the increase in H3K4me3 enrichment to *REG3A* promoter caused by FANCI overexpression. Relative enrichment of H3K4me3 to *REG3A* promoter was measured using ChIP assays in LIM1215-shREG1CP and HT-29-shREG1CP cells with or without overexpression of FANCI in the presence or absence of doxycycline (Dox) to induce REG1CP silencing. **j** REG1CP overexpression increased the enrichment of FANCI to the *REG3A* promoter. Relative abundance of *REG3A* promoter associated with FANCI was measured using ChIP assays. **k, l** FANCI silencing reduced (**k**), whereas its overexpression increased (**l**) RNA polymerase II (Pol II) enrichment to *REG3A* promoter measured using ChIP assays. $n = 3$ independent experiments. Data are presented as Mean \pm SEM (**b, c, d, f-l**) or representatives (**a, e**). Statistical significance was calculated using a two-tailed *t*-test.

The primary role of GR α involves anti-inflammation responses and although not previously linked to GR α , *REG3A* has similarly been demonstrated to play a role in inflammation¹⁶. However, while the upregulation of *REG3A* increased cell proliferation, GR α has been reported to suppress cell cycle progression, albeit utilizing transcriptional regulatory mechanisms distinct to those described here²⁶. These contrasting effects of GR α add a previously unrecognized perspective to its biological functions where outcomes appear contextually dependent.

Mechanistic investigations revealed that the formation of an RNA–DNA triplex was necessary for REG1CP to promote *REG3A* transcription, acting through an enhancer. TFOs located at 1637/1656 in REG1CP bound to a TFR homopurine region located at $-6176/-6195$ with respect to the *REG3A* TSS²⁷. The nature of the triplex interaction was substantiated through several independent approaches and importantly it was demonstrated that disruption of the triplex was sufficient to abolish upregulation of *REG3A* by REG1CP. The changes in *REG3A* expression were associated with modulation of the chromatin status of the *REG3A* promoter as shown by alterations in enrichment of active and repressive histone marks. Consistently, RNA Pol II enrichment to the promoter was regulated by the triplex formation. A number of lncRNAs have been shown to form RNA–DNA triplexes and thus regulate gene expression through modification of the epigenetic landscape of target promoters^{9,28}. This is commonly accomplished by the ability of lncRNAs to function as “local address codes” that target epigenetic regulators to specific gene sequences²⁹. For example, the lncRNA Khps1 functions through triplex-mediated binding to depress its target gene *sphingosine kinase 1 (SPHK1-B)* through recruiting the histone acetyltransferase p300/CBP, whereas the lncRNA PARTICLE forms a triplex with the *methionine adenosyltransferase 2A (MAT2A)* gene promoter to recruit the polycomb repressor complex 2 (PRC2) which in turn negatively regulates the expression of *MAT2A*⁷. Moreover, the lncRNA MEG3 binds by triplex to distal regulatory elements found in TGF- β pathway genes and recruits EZH2, a histone methyltransferase component of the PRC2 complex⁸. Of note, the mechanism of action of the RNA–DNA triplex formed by REG1CP appeared distinct, as we found that, instead of directly recruiting epigenetic regulators, the REG1CP triplex tethered the DNA helicase FANCI to *REG3A* promoter that in turn caused modulation of the chromatin status. Notably, REG1CP TFOs did not regulate the expression of other genes containing complementary TFRs, suggesting that the regulation of *REG3A* expression by REG1CP through forming RNA–DNA triplexes is specific. Although the mechanism responsible for this selectivity remains to be determined, it is

known that many nuclear-localised lncRNAs regulate gene expression in cis³⁰.

REG1CP was necessary for the association between FANCI and the active region of the *REG3A* promoter, an interaction mediated between FANCI and a G-quadruplex region in REG1CP and entirely different from the TFR. Thus these interactions serve as a two-stage address code to support the enrichment of transcriptional machinery to the *REG3A* promoter. Notably, depletion of FANCI was itself sufficient to cause downregulation of *REG3A*. Indeed, the actions of FANCI on the *REG3A* promoter phenocopied those of REG1CP, i.e. silencing of FANCI resulted in a more repressed state while its ectopic expression increased the activation status. Moreover, silencing or overexpression of FANCI decreased and increased RNA Pol II enrichment to the *REG3A* promoter, respectively. Therefore, FANCI is an essential component of REG1CP-mediated regulation of *REG3A*, although itself is not likely to be directly responsible for chromatin remodelling. Of note, the REG1CP–*REG3A* binding occurs nearly 4 kb upstream of the transcriptionally active region in the *REG3A* promoter that it influences. Distal elements and promoters interact through DNA looping are frequently associated with gene regulation events³¹. It is conceivable that the specific targeting of FANCI to the transcriptionally active region of *REG3A* is facilitated by structural folding of the *REG3A* DNA.

The involvement of DNA helicases in gene transcription has long been observed³². Helicase domain-containing proteins, such as the general transcriptional factors transcription factor II E (TFIIE) and TFIIH, constitute of an integral part of the RNA pol II holoenzyme and are important for derepression of transcriptional inhibition, acting to form an open chromatin structure necessary for the initiation of transcription³³. Even so, the mechanisms of functions and regulation of most helicases in gene transcription remain unclear. Our results showed that upon being tethered to the core promoter of the *REG3A* gene, FANCI helicase unwound double-stranded DNA and enabled a permissive chromatin structure for transcriptional activation of *REG3A*. This is likely to be independent of general transcription factors and occurs before the enrichment of the RNA pol II holoenzyme as FANCI silencing reduced enrichment of RNA Pol II to the *REG3A* promoter. Whether FANCI retains the same transcriptional activation role for other genes needs further investigations, either acting at enhancer sites or within promoters. More generally it also needs to be determined if accessibility of transcriptional machinery to other promoters is similarly unlocked by the two-stage address codes illustrated here (Fig. 8). In agreement with the notion that RecQ helicase family members unwind RNA–DNA triplexes^{32,34}, we found that FANCI reduced, albeit moderately,

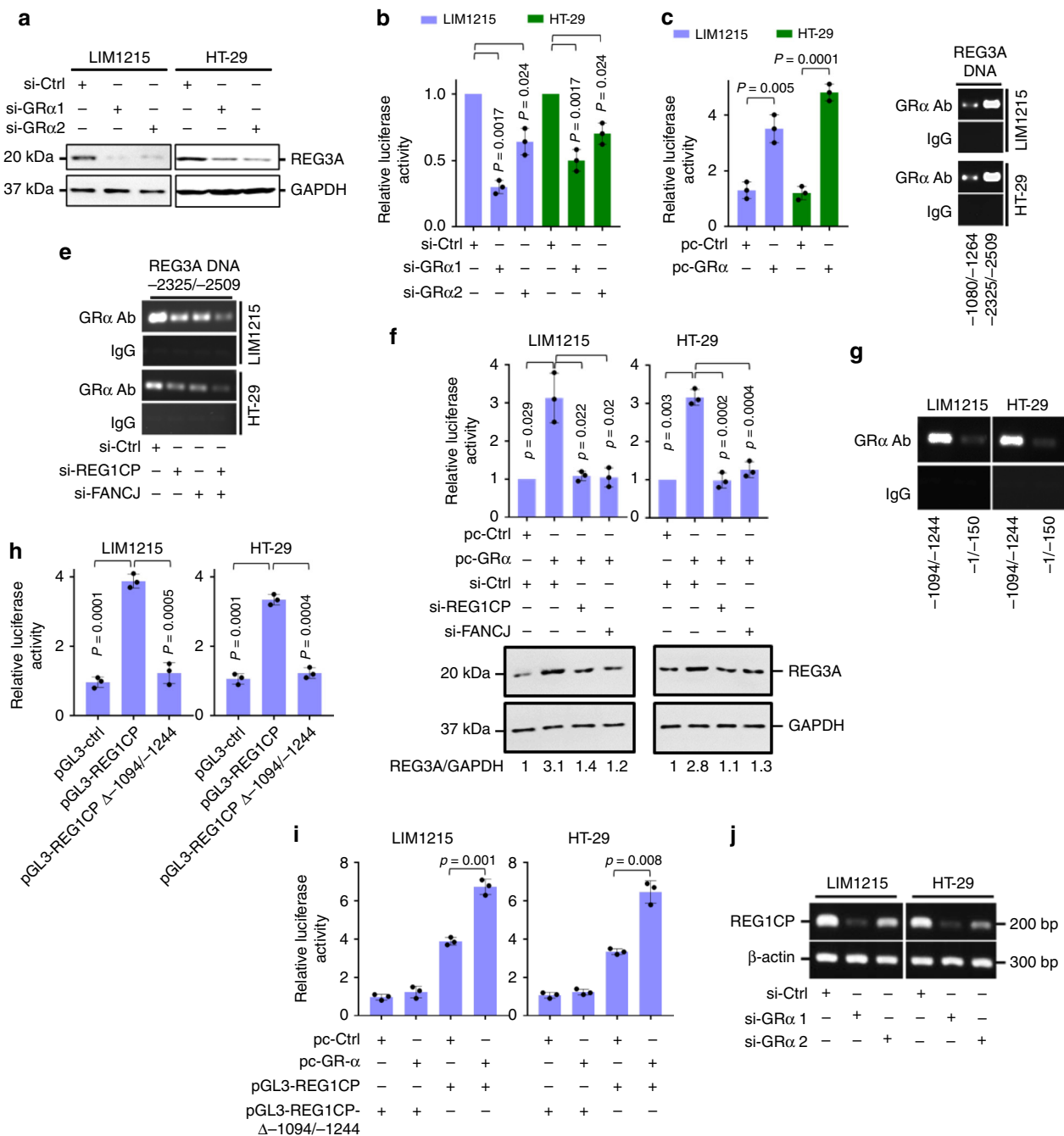


Fig. 7 GR α is responsible for transcriptional activation of REG3A and REG1CP. **a** GR α silencing reduced REG3A expression. **b, c** GR α silencing reduced (**b**) whereas its overexpression increased (**c**) luciferase reporter activity of pGL3-basic based REG3 promoter constructs. **d** GR α bound to the 2325/-2509 fragment of REG3A promoter. The precipitates of an antibody against GR α from formaldehyde-cross-linked chromatin of cells were subjected to qPCR using primers directed to indicated fragments of REG3A promoter. **e** Silencing of REG1CP or FANCI reduced the association between GR α and the 2325/-2509 fragment of REG3A promoter as shown using ChIP assays. **f** Silencing of REG1CP or FANCI diminished the increase in luciferase reporter activity of pGL3-basic based REG3 promoter constructs (upper) and the upregulation of endogenous REG3A protein levels (lower) caused by overexpression of GR α . The number below each western blot lane represents the level of REG3A protein relative the level of GAPDH expression. **g** GR α bound to the -1094/-1244 fragment of REG1CP promoter as measured using ChIP assays. **h** Deletion of the -1094/-1244 fragment abolished luciferase reporter activity of pGL3-basic based REG1CP promoter constructs. **i** Overexpression of GR α increased luciferase reporter activity of pGL3-basic based REG1CP promoter constructs with the -1094/-1244 fragment. **j** GR α silencing decreased endogenous REG1CP expression. $n = 3$ independent experiments. Data are presented as Mean \pm SEM (**b, c, f** (upper), **h, i**) or representatives (**a, d, e, f** (lower), **g, j**). Statistical significance was calculated using a two-tailed t-test.

the appearance of the REG1CP-REG3A RNA-DNA triplexes. Nevertheless, a significant amount of the triplexes constantly exist that is conceivably sufficient to ensure the transcriptional activation of REG3A through the complex. It seems that the formation and unwinding of the triplex structure is dynamic and the

unwinding effects of RecQ helicases on RNA-DNA triplexes have to be deliberated in a context-specific manner.

As a caveat to this study, we were unable to apply our findings to mouse models that would functionally consolidate the role of the identified REG1CP-mediated mechanism in the pathogenesis

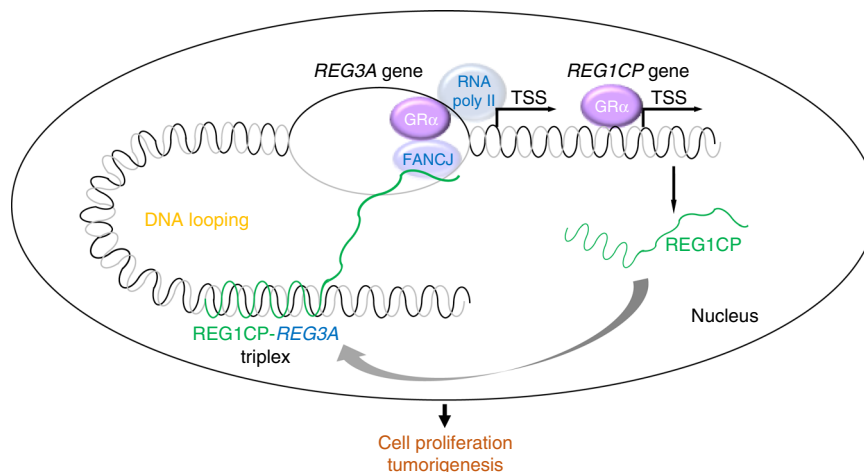


Fig. 8 Schematic illustration of the function of REG1CP. REG1CP acts as a two-stage address code to facilitate REG3A transcription through assembling a long-range enhancer complex via forming an RNA-DNA triplex with the REG3A gene. DNA looping through structural folding of the REG3A DNA is proposed to facilitate the specific targeting of FANCJ to the transcriptionally active region of REG3A.

of cancer *in vivo* since the lack of similarity between human REG1CP and transcripts of *Mus musculus* precludes this approach. Similarly, whether FANCJ helicase has a similar role in facilitating transcription of other genes remains an open question. Regardless, the results present here demonstrate a signaling pathway encompassing GR α , REG1CP, FANCJ and REG3A that promotes CRC cell proliferation and tumorigenesis and propose that REG1CP may constitute a molecular target for CRC treatment. Noticeably, although the expression of REG1CP and REG3A was positively correlated, there was no significant relationship between FANCJ or GR α expression and the expression of REG3A. These results support the notion that transcriptional activation of REG3A by FANCJ and GR α is dependent on the function of REG1CP, and further substantiate the importance of REG1CP in regulation of REG3A expression by the RNA-protein complex encompassing REG1CP, FANCJ and GR α . Nonetheless, our results did not exclude that independent mechanisms could regulate REG3A other than REG1CP/FANCJ/GR α . Since REG1CP is also present in the cytoplasm, albeit at low levels, it may function as competing endogenous RNA to repress microRNAs or as scaffolds to facilitate protein-protein interactions and modulator of mRNA stability as do many other cytoplasmic lncRNAs^{35,36}.

Methods

Cell culture and human tissues. The human normal colon epithelial cell line FHC (ATCC[®] CRL-1831[™]) and the human CRC cell lines HT-29 (ATCC[®] HTB-38[™]), SW480 (ATCC[®] CCL-228[™]) and COLO205 (ATCC[®] CCL-222[™]) were from American Type Culture Collection (ATCC). The human CRC cell line LIM1215 (ECACC 10092301) was from European Collection of Authenticated Cell Cultures (ECACC). Cells were cultured according to standard mammalian tissue culture protocols³⁷. Cells were cultured in a humidified incubator at 37 °C and 5% CO₂ and were tested using RT-PCR for mycoplasma contamination. Cell lines were authenticated using the AmpFISTR Identifier PCR Amplification Kit from Applied Biosystems and GeneMarker V1.91 software (SoftGenetics LLC). FFPE CRC tissues and paired normal colon tissues were retrieved from archives of the Anatomy Pathology Departments at Shanxi Cancer Hospital and Henan Provincial People's Hospital. Freshly removed colon cancer and paired adjacent normal colon tissues were obtained from patients undergoing surgical resection at the Department of Colorectal Surgery at Shanxi Cancer Hospital. Studies using human tissues were approved by the Human Research Ethics Committees of Henan Provincial People's Hospital and Shanxi Cancer Hospital in agreement with the guidelines set forth by the Declaration of Helsinki. All participants provided written informed consent.

Antibodies. H3K4me3 antibody (Abcam; ab8580, at 1:200 dilution), H3K27me3 antibody (Abcam; ab6002, at 1:200 dilution), FANCJ antibody (Novus Biologicals; NB100-416, at 1:200 dilution), GR α antibody (Novus Biologicals; NB300-633, at 1:200 dilution), REG3A antibody (Santa Cruz Biotechnology; sc-50967, at 1:1000

dilution) and G-quadruplex antibody (Absolute antibody; ab00174-24.1, at 1:500 dilution). Uncropped and unprocessed scans of all blots were supplied in the Source Data file.

Plasmid construction. PCR products obtained using full length primers were cloned into pGEM-T-easy as per manufacturer's instructions (Promega). Sanger sequencing was carried out by Australian Genome Research Facility (AGRF). REG1CP, REG3A, FANCJ and GR α DNAs were cloned into pcDNA 3.1⁽⁻⁾ respectively. DNA mutations were carried out using the QuikChange II site-directed mutagenesis kit (Agilent Technologies; Cat#200524-5) as per the manufacturer's instructions.

Polymerase chain reaction (PCR) assay. Total RNA was extracted using Bioline reagents (Isolate II RNA mini kit) according to the manufacturer's instructions. Two μ g of total RNA was used for cDNA synthesis using a reverse transcriptase kit from Bioline according to the manufacturer's instructions. Semi-quantitative PCR was conducted using BioRad reagents according to the manufacturer's instructions. Real-time PCR was carried out using the ABI PRISM 7500 Sequence Detection System (Applied Biosystems) and data were analysed using the 2^{- $\Delta\Delta$ CT} method³⁸. GAPDH and β -actin were used as internal controls.

siRNA and plasmid transfection. Cells were seeded 24 h before the transfection. siRNA and plasmid transfections were carried out using Lipofectamine 2000 reagent according to the manufacturer's instructions (Invitrogen).

Inducible shRNA knockdown. ShRNA sequences were constructed into FH1-tUTG inducible knockdown vector. The lentiviral particles were packaged via co-transfection with FH1-tUTG (44 μ g), pMDLg.pRRE (22 μ g), pMD2.g (13.2 μ g) and pRSU.pREV (11 μ g) plasmids into HEK293 cells³⁹. LIM1215 or HT-29 inducible knockdown cell sub-lines were established after the lentiviral transduction.

Cell cycle analysis. Cells were fixed by 70% ethanol on ice for 1 h and spun down at 1500 \times g. Cell pellets were resuspended in PBS containing 0.25% Triton X-100 and incubate on ice for 15 min. After discarding the supernatant, the cell pellet was resuspended in 0.5 ml PBS containing 10 μ g/ml RNase A and 20 μ g/ml PI stock solution and incubate at room temperature (RT) in the dark for 30 min. Cells were then subjected to analysis using a flow cytometer (FACSCanto, BD Biosciences).

BrdU cell proliferation assay. BrdU cell proliferation assays were carried out using the BrdU Cell Proliferation Assay kit (Cell Signaling). Briefly, cells were seeded at 5 \times 10³ cells per well in 96-well plates 24 h before treatment. BrdU (10 mM) was added and cells were incubated for 4 h before BrdU assays were carried out. Absorbance was read at 450 nm using a Synergy[™] 2 multidetection microplate reader (BioTek, VT).

Cell counting. Cells were stained with 0.4% Trypan blue solution and counted using Countess Automated Cell Counter (Invitrogen).

Clonogenic assay. Cells (2 \times 10³ cells/well) were seeded in a six-well plate. After 2–3 weeks of incubation, cells were fixed, stained with crystal violet, and

photographed. The percentage and intensity of area covered by crystal violet stained cell colonies were quantified using ImageJ-plugin “ColonyArea”⁴⁰.

Biotin RNA pull-down assay. 1×10^8 cells from were washed with PBS for 3 times and centrifuged at $500 \times g$ for 10 mins. Cells were lysed in RIP buffer (150 mM KCl, 25 mM Tris [pH 7.4], 0.5 mM DTT, 0.5% NP-40, protease inhibitors cocktail and RNase inhibitors) for 1 h at 4 °C and followed by sonication on ice for 10 cycles of 10 s with intervals 10 s. The cell lysates were centrifuged twice at $13,000 \times g$ for 15 min at 4 °C and supernatant was obtained. Simultaneously, probes were incubated with streptavidin magnetic beads at 4 °C for up to 4 h on an end-over-end rotator. The formed bead-probe complexes were mixed with the previous supernatant at 4 °C for overnight on a rotator. Beads were then washed five times with RIP buffer. RNA bound to the beads was isolated using Trizol reagent (Invitrogen) and quantified by qRT-PCR^{35,41}.

Mass spectrometry (MS) analysis. Co-precipitated proteins from REG1CP RNA pull-down assay were separated by 12.5% SDS-PAGE and visualized by silver staining. The specific peptides associated with REG1CP were sequenced by nanoflow reversed phased Liquid Chromatography (Dionex Ultimate 3000 RSLCnano, Dionex, Idstein, Germany) coupled directly to an ESI 3D Ion Trap Mass Spectrometer (AmaZon ETD, Bruker GmbH, Preston, VIC, Australia) operating in MS/MS (CID) mode ($n = 1$ technical replicate). Peptides were loaded at 5 μ l/min onto a 5 μ m C18 nanoViper trap column (100 μ m \times 2 cm, Acclaim PepMap100, Thermo) for desalting and pre-concentration. Peptide separation was then performed at 300 nl/min over an Acclaim nanoViper analytical column (2 μ m C18, 75 μ m \times 15 cm) utilising a gradient of 2–40% Buffer B (80% Acetonitrile, 0.1% Formic Acid) over 60 min. The peptides were eluted directly into the nanoflow ESI Ion source of the MS system for MS/MS analysis. The AmaZon Ion Trap system was tuned using Smart Parameter Settings tuned to 922 m/z and set to perform MS/MS on the top 6 ions present in each MS scan with an Ion exclusion time of 30 sec. Source settings were as follows: dry gas temperature, 180 °C; dry gas, 4.0 L min⁻¹; nebulizer gas, 0.4 bar; electrospray voltage, 4500 V; high-voltage end-plate offset, -200 V; capillary exit, 140 V; trap drive, 57.4; funnel 1 in 100 V, out 35 V, and funnel 2 in 12 V, out 3.3 V; MS/MS ICC target 500,000; maximum accumulation time, 50 ms. The sample was measured with the Ultrascan Scan Mode in polarity positive, scan range from m/z 100–3000, 3 MSn spectra averages. MS/MS spectra were triggered on ions higher than 50,000 in the Profile scan using a fragmentation amplitude of 100%.

Raw MS Files were converted into MASCOT Generic Format using DataAnalysis 4.1 and imported into ProteinScape 2.1 platform (both Bruker, Bremen, Germany) for database searching. Searches were performed against the UniProt Swiss-Prot Human database (retrieved January 2017) using an in-house licensed MASCOT server (version 2.3.02, Matrix Science). The number of allowed trypsin missed cleavages set to 2. Deamidation of Asparagine and Glutamine, Oxidation of Methionine and Phosphorylation of Serine, Threonine and Tyrosine were set as variable modifications. The parent ion tolerance was set to 1.2 Da with fragment ion tolerance set to 0.7 Da. Peptide thresholds were set requiring False Positive Rate less than 0.05% with a low stringency MASCOT score greater than 35. Those spectra meeting these criteria were validated by manual inspection to ensure accurate y- and b-ion detection with overlapping sequence coverage.

Chromatin isolation by rna purification (ChIRP) assay. 1×10^9 cells were washed with PBS for 3 times and centrifuged at $500 \times g$ for 10 min. Cells were crosslinked with 3% formaldehyde for 20 min with gentle end-over-end rotation at room temperature (RT). Crosslinking was terminated by adding 125 mM glycine to the cell suspension for 5 min. Cells were subsequently washed 3 times with cold PBS and centrifuged at $500 \times g$ for 10 min at 4 °C. Cells were lysed and sonicated to obtain DNA fragments around 800–1000 bp. Samples were then centrifuged and incubated with probes on an end-over-end rotator at 4 °C, overnight. DNA was purified and subjected to qPCR⁴².

RNA immunoprecipitation (RIP) assay. The RIP assay was performed using the Magna RIP kit (Merck Millipore) according to the manufacturer’s instructions. The FANCF antibody (Novus Biologicals, NB100-416) was used at 1:200 dilution in this study.

Chromatin immunoprecipitation (ChIP) assay. ChIP was performed using a commercially available Magnify chromatin immunoprecipitation system (Invitrogen). Briefly, LIM1215 or HT-29 cells were transfected with either siRNAs or overexpressing-plasmid vectors corresponding to REG1CP, FANCF or GRa. DNA-bound proteins were immunoprecipitated using antibodies against H3K4me3 (Abcam; ab8580, 1:200), H3K27me3 (Abcam; ab6002, 1:200), FANCF (Novus Biologicals; NB100-416, 1:200) or GRa (Novus Biologicals; NB300-633, 1:200). IgG was used as a negative control.

In vitro transcription assay. pcDNA3 vectors harbouring full length or mutant REG1CP respectively were used as templates for in vitro transcription of biotin-labelled RNAs. The forward primer contained the T7 RNA polymerase promoter

sequence to allow for subsequent in vitro transcription. PCR products were purified using DNA Gel Extraction kit (BioLigne; BIO-52058) and in vitro transcription was performed using T7-Flash Biotin-RNA Transcription Kit (Epicentre) according to the manufacturer’s instructions.

RNA electrophoretic mobility shift assay (EMSA). In vitro transcribed RNAs were heated at 80 °C for 30 s before snap cooling on ice. RNA folding was performed in a buffer containing 50 mM Hepes at pH 7.5, 150 mM KCl, 1.5 mM MgCl₂, 1 mM TCEP, and 20% glycerol⁴³. RNAs were incubated with G-quadruplex antibody (Absolute antibody; Ab00174-24.1, 1:500), FANCF protein, or PCR-amplified DNAs using a LightShift Chemiluminescent RNA EMSA Kit (cat# 20158) as per the manufacturer’s instructions. The samples were subjected to native protein electrophoresis.

DNase I hypersensitivity and ssDNase assay. LIM1215 or HT-29 cells (2×10^7) were washed 3 times with cold PBS and 2 times with RSB buffer [10 mM Tris (pH 7.4), 10 mM NaCl, 3 mM MgCl₂]. Cells were then gently lysed with RSB + 0.1% NP40 and spun down at $500 \times g$ for 10 mins at 4 °C to separate nuclei. The nuclei were resuspended in 720 μ l of cold RSB buffer, and 120 μ l of the nuclei suspension was then incubated with 4 U of DNaseI or ssDNase at 37 °C for 20 mins (except for one sample on ice as negative control). DNAs were extracted immediately by using Wizard SV Genomic DNA purification system (Promega) and subjected to PCR assay.

Luciferase assay. Cell lines were transfected with pGL3 plasmids containing various DNA fragments and 100 ng of phRL (Renilla luciferase) using Lipofectamine 2000 (Invitrogen). Twenty-four hours after transfection, cells were subjected to Dual-Luciferase reporter assay (Promega) according to the manufacturer’s instructions.

Quantification and statistical analysis. The analysis was performed using Microsoft Excel and GraphPad Prism to assess differences between varying groups. Statistical significance was defined by Student’s *t*-test and expressed as a *P*-value. *P*-values less than 0.05 were considered to be statistically significant.

Nucleotide sequences. Please refer to the Supplementary Table 6 for the list of primers, probes, siRNAs and shRNAs used in this study.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

A reporting summary for this Article is available as a Supplementary Information file. The source data underlying Fig. 1a–c, e–k, 2a–l, 3a–e, g, h, 4b–k, 5a–h, 6a–l, 7a–j, and Supplementary Figs. 1c–e, 2a, c–e, 3a, c, d, 4a, 5b–f, 6b, d, f, g are provided as a Source Data file. Raw data files for the RNA sequencing analysis have been deposited in the NCBI Gene Expression Omnibus under accession number GEO: GSE117830. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD015608. All data are available from the corresponding author upon reasonable request.

Received: 11 October 2018; Accepted: 1 November 2019;

Published online: 25 November 2019

References

- Parikh, A., Stephan, A. F. & Tzanakakis, E. S. Regenerating proteins and their expression, regulation and signaling. *Biomol. Concepts* **3**, 57–70 (2012).
- Zhang, Y. W., Ding, L. S. & Lai, M. D. Reg gene family and human diseases. *World J. Gastroenterol.* **9**, 2635–2641 (2003).
- Lennard, K. S., Goosen, R. W. & Blackburn, J. M. Bacterially-associated transcriptional remodelling in a distinct genomic subtype of colorectal cancer provides a plausible molecular basis for disease development. *PLoS One* **11**, e0166282 (2016).
- Lee, J. T. Epigenetic regulation by long noncoding RNAs. *Science* **338**, 1435–1439 (2012).
- Tsai, M. C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
- Batista, P. J. & Chang, H. Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298–1307 (2013).
- O’Leary, V. B. et al. PARTICLE, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. *Cell Rep.* **11**, 474–485 (2015).

8. Mondal, T. et al. MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. *Nat. Commun.* **6**, 1–17 (2015).
9. Postepska-Igielska, A. et al. LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via Triplex-mediated changes in chromatin structure. *Mol. Cell* **60**, 626–636 (2015).
10. Guo, S. T. et al. INPP4B is an oncogenic regulator in human colon cancer. *Oncogene* **35**, 3049–3061 (2016).
11. Ma, Y. et al. Long non-coding RNA CCAL regulates colorectal cancer progression by activating Wnt/beta-catenin signalling pathway via suppression of activator protein 2alpha. *Gut* **65**, 1494–1504 (2016).
12. Kim, T. et al. Long-range interaction and correlation between MYC enhancer and oncogenic long noncoding RNA CARLo-5. *Proc. Natl Acad. Sci. USA* **111**, 4173–4178 (2014).
13. Maamar, H., Cabili, M. N., Rinn, J. & Raj, A. linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. *Genes Dev.* **27**, 1260–1271 (2013).
14. Guo, Y. et al. Characterization of the mammalian miRNA turnover landscape. *Nucleic Acids Res.* **43**, 2326–2341 (2015).
15. Krol, J. et al. Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. *Cell* **141**, 618–631 (2010).
16. Liu, X. et al. REG3A accelerates pancreatic cancer cell growth under IL-6-associated inflammatory condition: Involvement of a REG3A-JAK2/STAT3 positive feedback loop. *Cancer Lett.* **362**, 45–60 (2015).
17. Chen, Z. F. et al. REG3A promotes the proliferation, migration, and invasion of gastric cancer cells. *OncoTargets Ther.* **10**, 2017–2023 (2017).
18. Booy, E. P. et al. The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1. *Nucleic Acids Res.* **42**, 3346–3361 (2014).
19. Brosh, R. M. Jr. & Cantor, S. B. Molecular and cellular functions of the FANCDJ DNA helicase defective in cancer and in Fanconi anemia. *Front. Genet.* **5**, 372 (2014).
20. Wu, C. G. & Spies, M. G-quadruplex recognition and remodeling by the FANCDJ helicase. *Nucleic Acids Res.* **44**, 8742–8753 (2016).
21. Wang, G. et al. Understanding transcription factor regulation by integrating gene expression and DNase I hypersensitive sites. *BioMed. Res. Int.* **2015**, 757530 (2015).
22. Chumsakul, O. et al. High-resolution mapping of in vivo genomic transcription factor binding sites using in situ DNase I footprinting and ChIP-seq. *DNA Res.* **20**, 325–338 (2013).
23. Creacy, S. D. et al. G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* **283**, 34626–34634 (2008).
24. Rafa, L. et al. REG4 acts as a mitogenic, motility and pro-invasive factor for colon cancer cells. *Int. J. Oncol.* **36**, 689–698 (2010).
25. Cao, G., Ma, J., Zhang, Y., Liu, B. & Li, F. Pancreatitis-associated protein is related closely to neoplastic proliferative activity in patients with colorectal carcinoma. *Anat. Rec.* **292**, 249–253 (2009).
26. Goya, L., Maiyar, A. C., Ge, Y. & Firestone, G. L. Glucocorticoids induce a G1/G0 cell cycle arrest of Con8 rat mammary tumor cells that is synchronously reversed by steroid withdrawal or addition of transforming growth factor-alpha. *Mol. Endocrinol.* **7**, 1121–1132 (1993).
27. Li, Y., Syed, J. & Sugiyama, H. RNA-DNA Triplex formation by long noncoding RNAs. *Cell Chem. Biol.* **23**, 1325–1333 (2016).
28. Mondal, T. et al. MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. *Nat. Commun.* **6**, 7743 (2015).
29. Morlando, M. & Fatica, A. Alteration of epigenetic regulation by long noncoding RNAs in Cancer. *Int. J. Mol. Sci.* **19**, pii: E570 (2018).
30. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu Rev. Biochem.* **81**, 145–166 (2012).
31. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
32. Maine, I. P. & Kodadek, T. Efficient unwinding of triplex DNA by a DNA helicase. *Biochem. Biophys. Res. Commun.* **204**, 1119–1124 (1994).
33. Graczyk, D., Ciesla, M. & Boguta, M. Regulation of tRNA synthesis by the general transcription factors of RNA polymerase III - TFIIB and TFIIC, and by the MAF1 protein. *Biochim Biophys. Acta Gene Regul. Mech.* **1861**, 320–329 (2018).
34. Guo, M. et al. A distinct triplex DNA unwinding activity of ChIR1 helicase. *J. Biol. Chem.* **290**, 5174–5189 (2015).
35. Hu, W. L. et al. GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. *Nat. Cell Biol.* **20**, 492–502 (2018).
36. Rashid, F., Shah, A. & Shan, G. Long Non-coding RNAs in the Cytoplasm. *Genomics, Proteom. Bioinforma.* **14**, 73–80 (2016).
37. Croft, A. et al. Functional identification of a novel transcript variant of INPP4B in human colon and breast cancer cells. *Biochem. Biophys. Res. Commun.* **485**, 47–53 (2017).
38. Rao, X., Huang, X., Zhou, Z. & Lin, X. An improvement of the 2'(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostat. Bioinforma. Biomath.* **3**, 71–85 (2013).
39. Herold, M. J., van den Brandt, J., Seibler, J. & Reichardt, H. M. Inducible and reversible gene silencing by stable integration of an shRNA-encoding lentivirus in transgenic rats. *Proc. Natl Acad. Sci. USA* **105**, 18507–18512 (2008).
40. Guzmán, C., Bagga, M., Kaur, A., Westermarck, J. & Abankwa, D. ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. *PLoS ONE* **9**, 1–9 (2014).
41. Marin-Bejar, O. & Huarte, M. RNA pulldown protocol for in vitro detection and identification of RNA-associated proteins. *Methods Mol. Biol.* **1206**, 87–95 (2015).
42. Chu, C., Quinn, J. & Chang, H. Y. Chromatin Isolation by RNA Purification (ChIRP). *J. Vis. Exp.* pii 3912 (2012).
43. Lister, N. et al. The molecular dynamics of long noncoding RNA control of transcription in PTEN and its pseudogene. *Proc. Natl Acad. Sci. USA* **114**, 9942–9947 (2017).

Acknowledgements

This work was supported by the National Health and Medical Research Council (NHMRC; APP1147271, APP1162753), Australia and National Natural Science Foundation of China (81471551). L.J. is a recipient of a Career Development Fellowship from Cancer Institute NSW, Australia. We thank Dr M.J. Herold (Walter and Eliza Hall Institute of Medical Research, Australia) for the plasmid FH1tUTG and Dr Nathan Smith (The University of Newcastle, Australia) for the Mass Spectrometry analysis.

Author contributions

H.Y., L.J., S.T.G. and X.D.Z. designed the experiments; S.T.G. and X.D.Z. supervised the work; Y.F. W., Y.Y.W., G.Z.L., R.H.Y., Y.X., J.L. and F.H.W., performed experiments using human tissues and ex vivo tissue cultures; Y.Y.W., W.G. and C.Z. conducted experiments in xenograft models; H.Y., Y.C.F., Y.Y.Z., H.T., L.T., X.G.Y. and M.F. conducted experiments using cell lines and related data collections; H.Y., R.S., T.A.L., T.L.L., R.F.T., S.T.G. and L.J. carried out analysis of publicly available data and bioinformatic analysis. X.D.Z. and R.F.T. wrote the manuscript. All authors commented on the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-13313-z>.

Correspondence and requests for materials should be addressed to S.T.G. or X.D.Z.

Peer review information *Nature Communications* thanks Valerie Bríd O'Leary, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019