HMGA2 functions as a competing endogenous RNA to promote lung cancer progression

Madhu S. Kumar1, Elena Armenteros-Monterroso1, Philip East2, Probir Chakravorty2, Nik Matthews3, Monte M. Winslow4 & Julian Downward1,5

Non-small-cell lung cancer (NSCLC) is the most prevalent histological cancer subtype worldwide1. As the majority of patients present with invasive, metastatic disease2, it is vital to understand the basis for lung cancer progression. Hmga2 is highly expressed in metastatic lung adenocarcinoma, in which it contributes to cancer progression and metastasis3–6. Here we show that Hmga2 promotes lung cancer progression in mouse and human cells by operating as a competing endogenous RNA (ceRNA)7–11 for the let-7 microRNA (miRNA) family. Hmga2 can promote the transformation of lung cancer cells independent of protein-coding function but dependent upon the presence of let-7 sites; this occurs without changes in the levels of let-7 isoforms, suggesting that Hmga2 affects let-7 activity by altering miRNA targeting. These effects are also observed in vivo, where Hmga2 ceRNA activity drives lung cancer growth, invasion and dissemination. Integrated analysis of miRNA target prediction algorithms and metastatic lung cancer gene expression data reveals the TGFB-competing co-receptor Tgfb3 (ref. 12) as a putative target of Hmga2 ceRNA function. Tgfb3 expression is regulated by the Hmga2 ceRNA through differential recruitment to Argonaute 2 (Ago2), and TGFB-signalling driven by Tgfb3 is important for Hmga2 to promote lung cancer progression. Finally, analysis of NSCLC-patient gene-expression data reveals that Hmga2 and TGFBR3 are coordinately regulated and that TGFBR3 is important for Hmga2 to promote lung cancer progression. Notably, expression of the wild-type (WT) and ATG mutant (ATG m7) constructs efficiently expresses full-length Hmga2 protein (M. Narita, personal communication), we found that the Hmga2 WT and m7 constructs constructs efficiently express full-length Hmga2 protein (m7 overexpresses Hmga2 owing to loss of let-7-mediated suppression), whereas the Hmga2 ATG WT and ATG m7 constructs do not (Fig. 1b). Importantly, we observed similar levels of Hmga2 transcript expression in the allelic series (in the case of the 482N1 cell line, the allelic series was mutated to abrogate binding to a short hairpin RNA (shRNA) against Hmga2) (Fig. 1c). Moreover, expression of the allelic series has no effect on the expression of various let-7 family members (Extended Data Fig. 1a). Taken together, this Hmga2 allelic series allows us to compare specifically the roles of Hmga2 protein and transcript function on lung cell transformation.

We therefore compared the ability of the Hmga2 allelic series to promote anchorage-independent growth of the lung cancer cell lines. We observed a striking promotion of soft-agar growth by both Hmga2 WT and ATG WT in the 368T1 and 482N1 cells (Figs. 1d, e); more modest growth was observed with Hmga2 m7, despite elevated protein expression relative to Hmga2 WT, and no growth was provided by Hmga2 ATG m7. This effect can be observed further in two additional human lung cancer cells (H1299 and SK-MES-1), as suppression of soft-agar growth by Hmga2 depletion can be rescued robustly by Hmga2 WT and ATG WT but more modestly by Hmga2 m7 (Extended Data Fig. 1b–e). Importantly, exonous expression of let-7 reversed the ability of the Hmga2 ceRNA to promote anchorage-independent growth, suggesting that let-7 regulates this effect (Extended Data Fig. 2b). To demonstrate that the effect of the Hmga2 ceRNA is driven by let-7 sites in the 3′ UTR, we expressed only the wild-type or let-7-mutant 3′ UTRs in 368T1 cells and examined the consequences on anchorage-independent growth. Notably, expression of the wild-type but not let-7-mutant 3′ UTR was sufficient to promote soft-agar growth in 368T1 cells (Extended Data Fig. 2c).

Beyond direct Hmga2 depletion, the Hmga2 WT and Hmga2 ATG WT constructs substantially rescued anchorage-independent growth in 482N1 cells stably overexpressing Nkx2.1, which we have previously shown to suppress lung cancer progression12 (Extended Data Fig. 2d). Notably, this effect is not due to a general proliferative benefit of Hmga2 WT and ATG WT cells, as BrdU (bromodeoxyuridine; 5-bromo-2′-deoxyuridine) incorporation in adherent conditions was comparable across the allelic series (Extended Data Fig. 3a). However, when lung cancer cells are placed in suspension, Hmga2 depletion suppressed proliferation and this proliferation was rescued substantially by Hmga2 WT and ATG WT, and more marginally by Hmga2 m7 (Extended Data Fig. 3b); in contrast, the rate of apoptosis in the allelic series was...
Magnification is representative mean technical triplicates, have been performed independently three times, and anchorage-independent growth in both 368T1 and 482N1 cells. Values are with propagated error.

To elucidate the mechanism of Hmga2 ceRNA function on lung carcinogenesis, we analysed the set of genes differentially expressed between metastatic and non-metastatic KrasG12D;Trp53−/− lung cancer cells and compared them to the list of predicted let-7 target genes based upon the miRNA target prediction algorithm TargetScan (Supplementary Table 5). Kras was not a candidate in this analysis, in spite of previous description of Kras as an important let-7 target. Moreover, the Hmga2 allelic series had no impact on either expression of K-Ras protein or activity of downstream K-Ras signalling pathways (Extended Data Fig. 3c). In contrast, we observed several known Hmga2 transcriptional targets, including components of the Igf2bp family, validating this approach.

To elucidate more broadly which transcripts are Hmga2 ceRNA targets, we initially examined whether let-7 sites are enriched among transcripts induced by the Hmga2 ceRNA through RNA-seq of the 482N1 allelic series combined with Sylamer analysis, which detects miRNA seed sites as nucleotide strings enriched within the 3′ UTRs of transcripts. We first compared control to Hmga2-knockdown cells and observed a specific enrichment of let-7 sites lost with Hmga2 depletion (Extended Data Fig. 4a). We then determined whether this was specific to ceRNA activity by determining let-7-site enrichment upon re-expression of either the Hmga2 WT or ATG WT constructs in Hmga2-knockdown cells. In both conditions, let-7 sites were enriched among the upregulated transcripts (Extended Data Fig. 4b, c). Importantly, let-7 sites were not enriched with re-expression of either Hmga2 m7 or ATG m7 in the Hmga2-depleted background (Extended Data Fig. 4d, e). Moreover, analysis of fragments per kilobase of exon per million fragments mapped (FPKM) in the RNA sequencing (RNA-seq) results from control 482N1 cells showed Hmga2 was among the most highly expressed predicted let-7 target transcripts, suggesting that Hmga2 constitutes a physiologically germane fraction of the let-7 target milieu (Supplementary Table 6). Taken together, these results indicate that the Hmga2 ceRNA broadly regulates let-7 targets.

To assess Hmga2 ceRNA targets more specifically, we examined which transcripts were suppressed in response to Hmga2 depletion; 13 out of 34 predicted targets were suppressed by Hmga2 knockdown (Extended Data Fig. 5a). To delineate which of these were Hmga2 transcriptional targets versus ceRNA targets, we re-expressed either Hmga2 WT or ATG WT in knockdown cells. As seen in Extended Data Fig. 5b, 6 out of 13 transcripts were rescued by both Hmga2 WT and ATG WT, suggesting they are putative ceRNA targets; the remaining targets were rescued only by Hmga2 WT, suggesting they are targets of Hmga2 transcription factor function. These Hmga2 ceRNA targets were markedly enriched in 3′-regulated transcripts, as their repression by Hmga2 loss could be reversed with the use of a ‘tough decoy’ let-7 sponge transcript, designed to be an efficient and long-term
suppressor of miRNA function (Extended Data Fig. 5c)\textsuperscript{18}. Conversely, overexpression of let-7 suppressed these transcripts, although Hmga2 transcriptional targets were also affected owing to depletion of Hmga2 (Extended Data Fig. 5d). Taken together, these studies outline a collection of putative target transcripts regulated by Hmga2 ceRNA function.

Among these six Hmga2 ceRNA target transcripts, we found the TGF-β co-receptor Tgfbr3 (ref. 12) to be both upregulated in metastatic lung cancer cells and a putative let-7 target. Furthermore, several Hmga2 ceRNA targets have been described as targets of TGF-β signalling\textsuperscript{9}. Thus, we examined whether Hmga2 exerts ceRNA function through enhanced TGF-β signalling via Tgfbr3. Consistent with this, we found that in both 368T1 and 482N1 cells, Hmga2 WT and ATG WT promote the expression of Tgfbr3 protein (Fig. 3a). This Tgfbr3 upregulation also occurs to a lesser degree at the messenger RNA level, as has been described previously for miRNA targets\textsuperscript{20} (Fig. 3b). Moreover, exogenous expression of let-7 reversed the ability of the Hmga2 ceRNA to upregulate Tgfbr3, suggesting that this effect is controlled by let-7 (Extended Data Fig. 2a). An important consideration in ceRNA–target analysis is the absolute levels of Hmga2, Tgfbr3 and let-7 transcripts in cells, so we determined the copies per cell of these factors (Extended Data Fig. 4f). We observed that Hmga2 and Tgfbr3 had similar levels of transcript, as might be expected for two factors that can titrate expression of one another; similar results were observed in FPKM analysis of these transcripts in control 482N1 cells by RNA-seq (Supplementary Table 6). Furthermore, total let-7 family expression was within an order of magnitude of Hmga2 and Tgfbr3. As this pool of let-7 regulates the entire target set, it is possible that miRNA occupancy to be a limiting factor, allowing for the contribution of a ceRNA-like Hmga2. Taken together, these results suggest that Hmga2 may regulate Tgfbr3 expression as a let-7 ceRNA.

In line with these observations of Hmga2 promoting Tgfbr3 expression, Hmga2 WT and ATG WT activated TGF-β signalling through phosphorylation of Smad2 (Fig. 3a). This effect was let-7-dependent, as exogenous let-7 suppressed Smad2 phosphorylation (Extended Data Fig. 2a). It is likely that the TGF-β pathway is active in the absence of exogenous ligand owing to low but detectable levels of TGF-β in serum during cell culture\textsuperscript{18}. We further examined whether Hmga2 ceRNA function affects TGF-β pathway activation by two methods. First, we found that a TGF-β reporter (CAGA\textsubscript{12})\textsuperscript{22} was potently induced by Hmga2 WT and ATG WT (Extended Data Fig. 5e). Second, analysis of TGF-β target transcript levels revealed specific expression of these genes with the Hmga2 WT and ATG WT constructs (Extended Data Fig. 5f). Notably, we observed little activation of the TGF-β pathway by Hmga2 m7, despite previous reports of Hmga2 functioning as a co-activator for Smad2, Smad3 and Smad4 in the epithelial–mesenchymal transition (EMT)\textsuperscript{23}; this is likely to be due to the lack of upstream activation of the pathway. Consistent with this, the Hmga2 ceRNA does not induce an EMT in either 368T1 or 482N1 cells (Extended Data Fig. 6a). Overall, these results indicate that the Hmga2 ceRNA induces expression of Tgfbr3 and potentiates TGF-β signalling.

To determine whether the effects of Hmga2 on Tgfbr3 occur through let-7-mediated derepression, we first examined the effect of Hmga2 ceRNA function on a reporter containing the Tgfbr3 3’ UTR. In both 368T1 and 482N1 cells, Hmga2 WT and ATG WT expression induced expression of luciferase under the control of the wild-type Tgfbr3 3’ UTR, but not if the let-7 site was mutated (Fig. 3c). Furthermore, we found that the effect of the Hmga2 ceRNA was broadly miRNA-dependent, as Hmga2 WT and ATG WT induced the reporter expression in Dicer1-intact sarcoma cells, but not in a Dicer1-null derivative cell line\textsuperscript{24} (Extended Data Fig. 6b). To assess directly whether Hmga2 induces Tgfbr3 through competition away from Ago2, we performed RNA immunoprecipitation (RIP) on Ago2 in lung cancer cells expressing the Hmga2 allelic series. As shown in Fig. 3d, we found that Hmga2 WT and ATG WT were recruited to Ago2 at levels comparable to the

**Figure 2 | Hmga2 ceRNA activity enhances lung cancer progression in vivo.** a, Hmga2 WT and ATG WT restore lung tumour growth in response to endogenous Hmga2 knockdown. B6129SF1/Tac males were intravenously injected with 482N1 cells expressing either a control shRNA and empty vector (shluc empty) or shHmga2 with the Hmga2 allelic series. Three weeks afterwards, animals were scanned by micro-CT and representative transverse images are shown. The heart is demarcated (H’) and white arrows identify lung tumours. b, Representative histological images of lungs transplanted with 482N1 cells from the series described in a. Magnification is ×1. c, Lung surface tumour counts were taken from animals transplanted with 482N1 cells from the series described in a (n = 3 animals per group). Values are technical triplicates and represent mean ± s.e.m. d, Hmga2 WT and ATG WT substantially reduce survival of animals transplanted with 482N1 cells expressing the shRNA targeting Hmga2. Animals were intravenously transplanted with cells as in a. Animals were subsequently aged for survival and a Kaplan–Meier analysis was performed (n = 9 animals per group). Median survival was 34 days for shluc empty and shHmga2 WT transplants; 37 days for shHmga2 ATG WT transplants; 43 days for shHmga2 m7 transplants; and 50 days for shHmga2 empty and ATG m7 transplants. Statistical significance was assessed by log-rank tests compared to shHmga2 empty. ***P < 0.00005; **P < 0.0005; *P < 0.005; *P < 0.05; NS, not significant.
constructs of the control siCXCR4 multimer and either the expression of a luciferase (IgG) or immunoprecipitation for Ago2. RNA was purified and qRT–PCR was performed for Hmga2 and Tgfbr3 on both the immunoprecipitated and input RNAs. Values are depicted as the percentage of input RNA, are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. c, Hmga2 WT and ATG WT significantly induce expression of a luciferase Tgfbr3 3′ UTR reporter in a let-7-site-dependent manner in both 368T1 and 482N1 cells. Cells were transfected with Renilla constructs of the control siCXCR4 multimer and either the Tgfbr3 wild-type or let-7-mutant 3′ UTR reporter. Values are normalized to co-transfected pGL3 plasmid. 368T1 values are normalized to empty and 482N1 values are normalized to shLuc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. d, Hmga2 WT and ATG WT displace Tgfbr3 from Ago2-based RNA-induced silencing complexes. Lyssates from 368T1 and 482N1 cells of the Hmga2 allelic series underwent either control immunoprecipitation (IgG) or immunoprecipitation for Ago2. RNA was purified and qRT–PCR was performed for Hmga2 and Tgfbr3 on both the immunoprecipitated and input RNAs. Values are depicted as the percentage of input RNA, are technical triplicates, have been performed independently twice, and represent mean ± s.d. Multiple shRNAs elicit substantial knockdown of Tgfbr3 mRNA in both 368T1 and 482N1 cells. 482N1 cells were infected with control shRNA (shLuc) or a set of shRNAs targeting Tgfbr3 (shTgfbr3.1–3.5), and 368T1 WT and ATG WT cells were infected with shLuc or shTgfbr3.1, 3.2, 3.4 and 3.5. RNA was purified and qRT–PCR was performed. Tgfbr3 expression is normalized to Gapdh and 368T1 WT and ATG WT and 482N1 values are normalized to shLuc empty. Values are technical triplicates, have been performed independently twice, and represent mean ± s.d. with propagated error. f, Multiple shRNAs induce knockdown of Tgfbr3 and suppress TGF–β pathway activity in 368T1 and 482N1 cells. Cells were infected with shRNAs as in e and Western blot analysis was performed for Tgfbr3, pSmad2 and total Smad2 (Smad2). g, Tgfbr3 depletion reduces anchorage-independent growth of 368T1 WT and ATG WT and 482N1 cells. Cells were infected with the listed shRNAs and plated for anchorage-independent growth and colonies were counted as above. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. **p < 0.0005; ***p < 0.005; *p < 0.05.

Figure 3 | Hmga2 cRNA activity enhances TGF–β signalling through overexpression of Tgfbr3. a, Hmga2 WT and ATG WT substantially induce both Tgfbr3 protein expression and phosphorylation of Smad2 (pSmad2) in both 368T1 and 482N1 cells. Tgfbr3 expression is normalized to Gapdh. 368T1 values are normalized to empty and 482N1 values are normalized to empty and 482N1 values are normalized to shLuc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. c, Hmga2 WT and ATG WT specifically induce expression of a luciferase Tgfbr3 3′ UTR reporter in a let-7-site-dependent manner in both 368T1 and 482N1 cells. Cells were transfected with Renilla constructs of the control siCXCR4 multimer and either the Tgfbr3 wild-type or let-7-mutant 3′ UTR reporter. Values are normalized to co-transfected pGL3 plasmid. 368T1 values are normalized to empty and 482N1 values are normalized to shLuc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. d, Hmga2 WT and ATG WT displace Tgfbr3 from Ago2-based RNA-induced silencing complexes. Lyssates from 368T1 and 482N1 cells of the Hmga2 allelic series underwent either control immunoprecipitation (IgG) or immunoprecipitation for Ago2. RNA was purified and qRT–PCR was performed for Hmga2 and Tgfbr3 on both the immunoprecipitated and input RNAs. Values are depicted as the percentage of input RNA, are technical triplicates, have been performed independently three times, and represent mean ± s.d. Multiple shRNAs elicit substantial knockdown of Tgfbr3 mRNA in both 368T1 and 482N1 cells. 482N1 cells were infected with control shRNA (shLuc) or a set of shRNAs targeting Tgfbr3 (shTgfbr3.1–3.5), and 368T1 WT and ATG WT cells were infected with shLuc or shTgfbr3.1, 3.2, 3.4 and 3.5. RNA was purified and qRT–PCR was performed. Tgfbr3 expression is normalized to Gapdh and 368T1 WT and ATG WT and 482N1 values are normalized to shLuc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. f, Multiple shRNAs induce knockdown of Tgfbr3 and suppress TGF–β pathway activity in 368T1 and 482N1 cells. Cells were infected with shRNAs as in e and Western blot analysis was performed for Tgfbr3, pSmad2 and total Smad2 (Smad2). g, Tgfbr3 depletion reduces anchorage-independent growth of 368T1 WT and ATG WT and 482N1 cells. Cells were infected with the listed shRNAs and plated for anchorage-independent growth and colonies were counted as above. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. **p < 0.0005; ***p < 0.005; *p < 0.05.
phosphorylation, CAGA12 reporter activity and expression of TGF-β target genes (Fig. 3f and Extended Data Fig. 8a, b). We then assessed the functional effect of Tgfbr3 loss on Hmga2 ceRNA-driven soft-agar colony formation. In all cells, Tgfbr3 knockdown notably suppressed anchorage-independent growth, but not to the same extent as Hmga2 depletion in 482N1 cells (Fig. 3g). This occurred without generally affecting proliferation, as measured by BrdU incorporation (Extended Data Fig. 8c). We further functionally analysed the broader set of six Hmga2 ceRNA targets by individually depleting them by short interfering RNA (siRNA) (Extended Data Fig. 8d). When we compared their effects on anchorage-independent growth, both Hmga2 and Tgfbr3 loss strongly suppressed growth, Hmgal depletion modestly reduced colony formation, whereas the remaining targets had little effect (Extended Data Fig. 8e). It should be noted that these other targets include extracellular factors like Angptl2 and Col1a2 that might promote lung cancer progression in vivo in a non-cell-autonomous manner, and could thus still be relevant to Hmga2 ceRNA activity in lung cancer progression. This considered, these results still indicate that Tgfbr3, although certainly not the only relevant Hmga2 ceRNA target, is an important effector of Hmga2 ceRNA function in lung cancer cell transformation.

To determine whether this effect of Tgfbr3 is driven through potentiation of TGF-β signalling, we inhibited the TGF-β pathway with the TGF-β-receptor-kinase inhibitor SB-431542 (SB)25. In 368T1 and 482N1 cells, SB treatment led to a substantial inhibition of Smad2 phosphorylation (Extended Data Fig. 9a). In addition, SB treatment of 482N1 and 482T1 Hmga2 WT and ATG WT cells markedly suppressed CAGA12 reporter activity and expression of TGF-β target genes (Extended Data Fig. 9b, c). We then examined whether SB could inhibit Hmga2 ceRNA-driven soft-agar colony formation and observed a striking reduction in anchorage-independent growth (Extended Data Fig. 9d). This impaired colony formation was not due to general proliferative arrest, as SB treated cells had a similar rate of BrdU incorporation (Extended Data Fig. 9e). Notably, many Hmga2 ceRNA targets are in fact TGF-β target genes, as their expression is suppressed by SB treatment and induced by exogenous addition of TGF-β (Extended Data Fig. 9f, g). Thus, it is possible that Hmga2 could function in a feed-forward loop in which it regulates TGF-β target genes directly through ceRNA function and indirectly through TGF-β signalling via Tgfbr3. In summary, these results indicate that TGF-β signalling through Tgfbr3 is an important pathway downstream of Hmga2 ceRNA function.

Based on the above findings, we wanted to examine whether Hmga2 functions as a ceRNA for TGFBR3 (the human orthologues of Hmga2 and Tgfbr3, respectively) in NSCLC patients. An important corollary of the ceRNA hypothesis is the coordinate regulation of a competing RNA and its targets, such that upregulation of the ceRNA should lead to higher expression of the target RNA and vice versa16. To assess this, we used NSCLC gene-expression data generated by the Cancer Genome Atlas (TCGA) and sorted the patient cohort into the top and bottom quartiles of Hmga2 expression (Extended Data Fig. 9a). As seen in Fig. 4a, we observed significantly higher levels of TGFBR3 transcript in Hmga2 high versus low patient samples. To address the converse relationship, we sorted the TCGA data set into top and bottom quartiles of TGFBR3 expression (TGFBR3 high and low, respectively) (Fig. 4c). When we compared Hmga2 transcript levels between the groups, we found Hmga2 to be significantly overexpressed in TGFBR3 high versus low patient samples (Fig. 4d). To extend and validate these findings, we carried out similar gene-expression analyses of Hmga2 and TGFBR3 in an independent lung-adenocarcinoma-patient gene-expression cohort, the Director’s Challenge data set26. Similar to the findings with the TCGA cohort, we observed Hmga2 and TGFBR3 to be coordinately expressed in the Director’s Challenge data set (Extended Data Fig. 10a–d). Although we focused specifically on high and low expressors of Hmga2 and TGFBR3, for which ceRNA activity is more likely to occur16, Hmga2 and TGFBR3 expression was broadly correlated across both data sets (Extended Data Fig. 10e, f). As these two data sets constitute two of the largest collections of NSCLC gene-expression data available, we believe that these findings are consistent with Hmga2 functioning as a ceRNA for TGFBR3 in NSCLC patients. It is possible that the coexpression of Hmga2 and TGFBR3 could correspond to additional tumour characteristics that these data sets do not include; future studies in independent data sets would be needed to assess this issue. In total, our results suggest a model in which Hmga2 promotes lung cancer progression by competing for let-7 occupancy with other targets, including Tgfbr3, leading to the upregulation of those targets (Fig. 4e). Importantly, this occurs without changes in the levels of let-7 family microRNAs, reflecting specific competition for microRNA binding among targets.

Here we have outlined a novel gene-expression pathway in which a protein-coding gene, Hmga2, operates largely independently of its protein-coding function to promote lung cancer progression as a competing endogenous RNA. Although much of this ceRNA activity is
driven by overexpression of TGF-β signalling through Tgfb3, there are likely to be additional Hmg2a ceRNA targets to be found in future studies. Moreover, HMG2A is overexpressed in many other cancer types, so it is possible that HMG2A functions as a ceRNA in cancer sites beyond lung. Taken more broadly, these findings raise the possibility that many protein-coding genes differentially expressed in cancer might contribute to tumorigenesis through this distinct mode of regulatory gene expression. Moreover, these results raise issues with the validation of candidates in RNA interference screens. The ‘gold standard’ assay for validating an siRNA target is expression of an siRNA-resistant form of the coding sequence; however, such an approach overlooks the possibility that depletion of both the full-length RNA and protein might contribute to a given phenotype, requiring complementation by the full-length transcript. Such dual-function ceRNA and protein activities necessitate a deeper exploration of the coding genome in biological systems.

**METHODS SUMMARY**

Soft-agar assays. Soft-agar assays were performed essentially as described previously. Assays were carried out in triplicate and quantified by microscope.

Intravenous transplantation. Intravenous injection was performed on 12-week-old B6129SF1/Tac male mice (Taconic), essentially as described previously. In short, 10^5 cells in 50 µl PBS were injected into 12 animals per group. Three weeks post injection, animals were scanned using the SkyScan 1176 micro-CT scanner as described previously. Three mice per group were then euthanized at random for surface tumour and histopathological analysis. Surface tumours were quantified by counting all the visible tumours on the lung pleura; quantification was carried out blind to the expression construct. The remaining nine mice were aged for survival analysis. All procedures were performed under an approved project license as per UK Home Office regulations.

RNA immunoprecipitation. RNA immunoprecipitation was carried out on 482N1 and 368T1 cells with control antibodies (immunoglobulin G (IgG)) or antibody targeting Ago2 as per manufacturers’ instructions (Millipore). Total RNA was used for either qRT–PCR (quantitative polymerase chain reaction with reverse transcription) of mRNAs or miRNA-specific qRT-PCR as above.

Public gene-expression array analysis. NSCLC gene-expression data sets (from both The Cancer Genome Atlas (TCGA) and the Director’s Challenge) were downloaded and processed using standard methods. Patient expression profiles were sorted by HMG2A- or TGFBR3-expression status, and the top and bottom quartiles of both groups were selected. Target gene expression was then analysed and represented as box-and-whisker plots. Statistical significance was assessed using Mann–Whitney tests with correction for multiple hypothesis testing.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 February; accepted 16 October 2013.

Published online 4 December 2013.

1. Ferlay, J. et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer 127, 2893–2917 (2010).
2. Herbst, R. S., Heymach, J. V. & Lippman, S. M. Lung cancer. Nature Rev. Mol. Cell Biol. 13, 610–620 (2012).
3. Di Cello, F. et al. HMGA2 participates in transformation in human lung cancer. Mol. Cancer Res. 6, 743–750 (2008).
4. Meyer, B. et al. HMGA2 overexpression in non-small cell lung cancer. Mol. Carcinog. 46, 503–511 (2007).
5. Sarhadi, V. K. et al. Increased expression of high mobility group A proteins in lung cancer. J. Pathol. 209, 206–212 (2006).
6. Winslow, M. M. et al. Suppression of lung adenocarcinoma progression by Nkx2-1. Nature 473, 101–104 (2011).
7. Ebert, M. S., Neison, J. R. & Sharp, P. A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nature Methods 4, 721–726 (2007).
8. Karreth, F. A. et al. In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. Cell 147, 382–395 (2011).
9. Polisen, L. et al. A coding-independent function of gene and pseudogene miRNAs regulates tumour biology. Nature 465, 1035–1039 (2010).
10. Salmena, L., Polisen, L., Tay, Y., Kats, L. & Pandolfi, P. P. A. ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 146, 353–358 (2011).
11. Tay, Y. et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous miRNAs. Cell 147, 346–357 (2011).
12. Wang, X. F. et al. Expression cloning and characterization of the TGF-β type III receptor. Cell 67, 797–805 (1991).
13. Mayr, C., Hemann, M. T. & Bartel, D. P. Disrupting the pairing between let-7 and Hmg2a enhances oncogenic transformation. Science 315, 1576–1579 (2007).
14. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian miRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105 (2009).
15. Johnson, S. M. et al. RS is regulated by the let-7 microRNA family. Cell 120, 635–647 (2005).
16. Brants, J. R. et al. Differential regulation of the insulin-like growth factor II mRNA-binding protein genes by architectural transcription factor HMGA2. FEBS Lett. 569, 277–283 (2004).
17. van Dongen, S., Abreu-Goedcker, C. & Enright, A. J. Detecting microRNA binding and siRNA off-target effects from expression data. Nature Methods 5, 1023–1025 (2008).
18. Xie, J. et al. Long-term, efficient inhibition of microRNA function in mice using rAAV vectors. Nature Methods 9, 403–409 (2012).
19. Massagué, J. TGFβ signalling in context. Nature Rev. Mol. Cell Biol. 13, 610–630 (2012).
20. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466, 835–840 (2010).
21. Oida, T. & Weiner, H. Deletion of TGF-β from fetal bovine serum. J. Immunol. Methods 362, 195–198 (2010).
22. Denninger, S. et al. Direct binding of Smad3 and Smad4 to critical TGF-β-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. EMBO J. 17, 3091–3100 (1998).
23. Thouault, S. et al. HMGA2 and Smads co-regulate SNAIL1 expression during induction of epithelial-to-mesenchymal transition. J. Biol. Chem. 283, 33437–33446 (2008).
24. Ravi, A. et al. Proliferation and tumorigenesis of a murine sarcoma cell line in the absence of DICER1. Cancer Cell 21, 848–855 (2012).
25. Imman, G. J. et al. SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. Mol. Pharmacol. 62, 65–74 (2002).
26. Shedden, K. et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nature Med. 14, 822–827 (2008).
27. Park, S. M. et al. Let-7 prevents early cancer progression by suppressing expression of the embryonic gene BMI1. Cell Cycle 6, 2585–2590 (2007).
28. Kaelin, W. G., Jr. Molecular biology. Use and abuse of RNAi to study mammalian gene function. Science 337, 421–422 (2012).
29. Kumar, M. S. et al. The Deltex-A lineal network is requisite for RAS oncogene-driven non-small cell lung cancer. Cell 149, 642–655 (2012).
30. Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. Genes Dev. 17, 436–442 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the laboratory of T. Jacks for providing the 482N1 and 368T1 lung cancer cell lines and the laboratories of P.A. Sharp and P. Chambon for providing the KPD sarcoma cell lines. We thank the LRI FACS Facility for the BrdU analyses and the LRI Biological Resources Unit for assistance with the animal studies. We thank M.S. Ebert and E. De Bruin for critical review of the manuscript. M.S.K. is a Long-Term Fellow of the Human Frontier Science Program and the European Molecular Biology Organization. This work was funded by Cancer Research UK and by the European Commission’s Seventh Framework Programme (FP7/2007–2013) under the grant agreement Lungtarget (project no. 259770).

Author Contributions M.S.K. and J.D. designed the study. M.S.K. and E.A.-M. performed the experiments described. M.S.K., P.E. and P.C. conducted bioinformatics analyses. N.M. performed the next-generation sequencing studies. M.M.W. provided necessary reagents and conceptual advice. M.S.K. and J.D. wrote the manuscript.

Author Information Sequence data have been uploaded to the Gene Expression Omnibus database under accession number GSE50932. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D. (julian.downward@cancer.org.uk).

©2014 Macmillan Publishers Limited. All rights reserved
METHODS

Cell lines. 368T1 and 482N1 cells were generated previously38. KD P- and /-sarcoma cells were a gift of P. A. Sharp41. Cancer Research UK Cell Services provided HEK293, Phoenix amphi, NCI-H1299 and SK-MES-1 cells. All cell lines were cultured in DMEM medium with 10% FBS, 10 mM glutamine and 1% penicillin and streptomycin at 37 °C in a 5% CO2-95% air incubator.

Drug and reagent treatment. For Western blot analysis, luciferase, qRT–PCR and BrdU analysis, 368T1 cells were co-transfected with 2 ng ml−1 TGF-β and either DMSO or 2.5 μM SB-431452 (SB) and 482N1 cells were treated with either DMSO or 2.5 μM SB-431452 for 24 h. For soft-agar analysis, cells were fed with media containing either DMSO or SB. For Western blot analysis, 368T1 and 482N1 cells were plated (2.5 × 105 cells) into 6-well plates overnight. Cells were then treated with DMSO, 1 μM AZD6244 or 1 μM PFK-75 for 2 h and lysates were prepared.

Lentivirus and retrovirus production. Short hairpin RNA (shRNA) lentiviruses (which are listed in Supplementary Table 1) were generated by co-transfection of HEK293 cells with shRNA vector and packaging vectors pCMV-MSGV and pCMV-8.2. MSCV-based retroviruses were generated by transfection of Phoenix amphi cells with retroviral vectors. Forty-two and seventy-two hours after transfection, virus particles in the medium were harvested, filtered and transduced onto cells.

Generation of stable cell lines. Expression vectors were initially linearized overnight using BglII. Linearised DNA was then transfected into cells in 24-well plates, with media replaced 6 h later. Cells were then re-plated into 6-well plates in the presence of 500 μg ml−1 G418 for selection over 2 weeks. For lentiviral and retroviral infection, cells were infected in 6-well plates and subsequently split into 10-cm dishes in the presence of 2.5 μg ml−1 puromycin for selection over 72 h.

Western blotting. Western blot analysis was performed using standard methods. Antibody details are provided in Supplementary Table 2.

qRT–PCR. RNA was isolated using RNA Bee according to the manufacturer’s instructions (Promega). Total RNA samples were initially quality controlled using the 6000 TOPO clone inserts and pcDNA3.1-Hmga2 WT

Luciferase assays. For the CAGA12 reporter, cells were seeded into 96-well plates in triplicate and co-transfected with pRL-TK (Promega) and either control pGL3 or the TGF-β reporter pCAGA12-GL3 (ref. 22). Twenty-four hours later, luciferase activities were determined using the Dual-Luciferase Assay System as per the manufacturer’s instructions (Promega). For the Tgfbr3 3′ UTR, cells were seeded into 96-well plates in triplicate and co-transfected with pGL3 (Promega) and either the control pRL-CXCR4 (ref. 30), the wild-type or mutant fragment of the Tgfbr3 3′ UTR. Luciferase activities were determined as described above.

BrdU analysis. Cells were plated in triplicate into 6-well plates overnight (2.5 × 105 cells per well). They were then treated with 10 μM BrdU for 1 h, trypsinized, washed in PBS and stained for BrdU and DAPI. Analysis was performed using standard methods described previously42. Antibody details are provided in Supplementary Table 2.

Statistical analysis. Statistical analysis was performed using GraphPad Prism software. When possible, t-tests were used. Where indicated, the Student’s t-test was used.

Hmga2-TOPO. For Hmga2: pCDNA3.1-Hmga2 WT and Hmga2 m7 (’WT’ and ’m7’) were generated from all 8mer miRNA strings whose target sites were enriched significantly (P < 0.001 by hypergeometric testing). Statistical testing. Unless otherwise specified, statistical significance was assessed by the Student’s t-test.

Vector cloning. For Hmga2: pCDNA3.1-Hmga2 WT and Hmga2 m7 (’WT’ and ’m7’) were acquired from Addgene (Plasmid nos 14789 and 14792). To generate pCDNA3.1-Hmga2 ATG WT and pCDNA3.1-Hmga2 ATG m7 (’ATG WT’ and ’ATG m7’), the WT and m7 constructs were mutated by site-directed mutagenesis as per the manufacturer’s instructions (Stratagene) using primers 5′-GGTAG GCCGCCGGGAGACGACGTGACGAGCTGGG-3′ and 5′-CCCTC ACCCGCTGTCGCTCCTCCTGCCGGCTGCTACC-3′. To generate 482N1 shHmga2 stable cell lines, WT, m7, ATG WT and ATG m7, the shRNA-binding site was mutated using primers 5′-CCATTTCGTCAAGCTATGTATGTGTCGGGAGACCGTCGGAGATGCTATGGCAGACCTGAGTCGCA-3′. The differential analysis was carried out using edgeR33, applying TMM (trimmed Mean of M-values) library normalization and a 0.0005 false discovery rate (FDR) to select expressed transcripts. In addition, we calculated gene-level fragments per kilobase of exon per million fragments mapped (FPKM) values for the same gene set using Cufflinks44.

Sylamer analysis. A list of Ensembl gene identifiers was generated from the set of transcripts expressed in the samples described above. This set of transcripts was then placed in rank order by their differential expression between two groups and analysed using the Sylamer algorithm via the SylArray platform46. A Sylamer plot was then generated for all 8mer miRNA strings whose target sites were enriched significantly (P < 0.001 by hypergeometric testing).

Histopathological analysis. After mice were euthanized by CO2 asphyxiation, BrdU analysis was performed as described above. The total RNA samples were initially quality controlled using the 6000 Nano RNA Chip on the BioAnalyzer 2100 (Agilent) to ensure RNA integrity and estimate concentration before starting the procedure. If the samples passed the initial quality control, the total RNA samples were subjected to poly-A selection using Sera-Mag oligo deoxy-thymine (dT) beads (Thermo Fisher Scientific), and the bound poly-A RNA species utilized for downstream library preparation using the Illumina mRNA kit RS-122-2101 (TrueSeq Stranded mRNA Sample Prep, Illumina). The standard PCR cycles suggested in the protocol were also altered to match the concentration of the total RNA from the initial quality control. After production of the mRNA libraries a final quality control was performed on a DNA 100 Chip using the BioAnalyzer 2100 (Agilent). If the mRNA libraries passed the quality control they were ready for flow-cell cluster formation on a cBot and then 100-bp paired-end sequencing by synthesis on the HiSeq 2500 was performed.

RNA-seq data processing and analyses. We aligned 100-bp paired-end sequencing reads to the mouse genome (mm9, UCSC) using TopHat2/Bowtie2 (ref. 32) allowing for five mismatches, a mate-inner-distribution of −40 and a mate-standard-deviation of 50. These library fragment metrics were derived from an alignment of five mRNA samples to the mouse transcriptome. We identified read-pair mappings to gene structures derived from RefSeq (RefGene table, UCSC) using the summarizeOverlaps function with mode IntersectStrict (GenomicRanges, Bioconductor). This generated a mean fragments per sample value of 33.5e6. Using these raw counts we identified genes expressed across the sample groups. We removed genes with a count ≤10 across all samples prior to statistical analysis. The differential analysis was carried out using edger33, applying TMM (trimmed Mean of M-values) library normalization and a 0.0005 false discovery rate (FDR) to select expressed transcripts. In addition, we calculated gene-level fragments per kilobase of exon per million fragments mapped (FPKM) values for the same gene set using Cufflinks44.
31. Liau, S. S., Jazag, A., Ito, K. & Whang, E. E. Overexpression of HMGA1 promotes anoikis resistance and constitutive Akt activation in pancreatic adenocarcinoma cells. Br. J. Cancer 96, 993–1000 (2007).
32. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).
33. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
34. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnol. 31, 46–53 (2013).
Extended Data Figure 1 | Hmga2 functions as a ceRNA in human cancer cells without changes in let-7 expression. a, RNA was purified from 368T1 and 482N1 cells expressing the Hmga2 allelic series as in Fig. 1c and qRT–PCR was performed for a control miRNA (miR-10a) and various members of the let-7 family (let-7a, let-7c, let-7e, let-7f and let-7i). Expression of let-7 members is normalized to miR-10a. 368T1 values are normalized to empty and 482N1 values are normalized to shluc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± standard deviation (s.d.) with propagated error.

b, H1299 cells infected with control shRNA (shluc) or two different shRNAs targeting human HMGA2 (shHMGA2.1 and shHMGA2.2). Cells were then stably transfected with the Hmga2 allelic series as in Fig. 1a, b. Lysates were then prepared and Western analysis was performed for HMGA2, Rho, Tgfbr3, phosphorylated Smad2 (pSmad2) and Smad2 as in Figs 1b and 3a. c, H1299 cells with the Hmga2 allelic series described in b were plated in soft agar as in Fig. 1d. Values are technical triplicates, have been performed independently twice, and represent mean ± s.d.

d, SK-MES-1 cells were infected with shRNAs and stably transfected with the Hmga2 allelic series and Western blot analysis was performed as in b. e, SK-MES-1 cells with the Hmga2 allelic series described in d were plated in soft agar as above. Values are technical triplicates, have been performed independently twice, and represent mean ± s.d. ***P < 0.0005; **P < 0.005; *P < 0.05.
Extended Data Figure 2 | Let-7 regulates Hmga2 ceRNA activity, which functions through the 3' UTR and controls Nkx2.1-mediated anchorage-independent growth. a, Lysates were prepared from the 368T1 Hmga2 WT and Hmga2 ATG WT; 482N1 shluc empty, shHmga2 WT and shHmga2 ATG WT cells reverse transfected with either control miRNA (miR-10a) or let-7e. Western blot analysis was subsequently performed for pSmad2, Smad2, Tgfbr3 and Rho. b, Cells were transfected as in a and soft-agar analysis was performed as in Fig. 1d. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. c, 368T1 cells were transfected with empty vector, Hmga2 WT (as in Fig. 1b), the wild-type 3' UTR of Hmga2 (WT 3' UTR) or the let-7-site-mutated 3' UTR (m7 3' UTR). Soft-agar analysis was performed as in Fig. 1d. Values are technical triplicates, have been performed independently twice, and represent mean ± s.d. d, 482N1 cells were infected with empty retrovirus or a retrovirus overexpressing Nkx2.1. The Hmga2 allelic series was then introduced into Nkx2.1 cells (with just empty vector in the empty retrovirus cells) and soft-agar analysis was performed as in Fig. 1d. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. ***P < 0.0005; **P < 0.005; *P < 0.05.
Extended Data Figure 3 | Hmga2 ceRNA activity affects proliferation in suspension without altering K-Ras signalling. a, 368T1 and 482N1 cells with the Hmga2 allelic series from Fig. 1 were treated with 10 μM BrdU for 1 h. Cells were subsequently analysed for percentage BrdU positive cells. Values are mean ± s.d (n = 3).

b, 482N1 cells with the Hmga2 allelic series from Fig. 1 were plated under normal adherent conditions (adherent) or on plates coated with polyHEMA (suspension) overnight. Cells were then treated with BrdU as in a and subsequently analysed for percentage BrdU-positive cells. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d (n = 3). ***P < 0.0005; **P < 0.005; *P < 0.05.

c, Lysates were prepared from the 368T1 and 482N1 Hmga2 allelic series and from 482N1 cells treated with either DMSO, 1 μM AZD6244 or 1 μM PIK-75 for 2 h. Western blot analysis was subsequently performed for phosphorylated Akt (pAkt), Akt, phosphorylated ERK1 and ERK2 (pERK1/2), ERK1/2, phosphorylated ribosomal protein S6 (pS6), ribosomal protein S6 (S6), K-Ras and Rab11.
Extended Data Figure 4 | Hmga2 ceRNA function specifically upregulates let-7 target transcripts. RNA-seq was performed for cells from the 482N1 Hmga2 allelic series. a, The set of transcripts expressed in shluc empty and shHmga2 empty cells were ranked in order of differential expression and Sylamer analysis was performed. Sylamer plots were then generated across the transcript set for all octomer miRNA strings statistically enriched (P < 0.001 by hypergeometric testing). The Sylamer plot for mmu.let-7a.8(A1) is delineated in red. b, The set of transcripts expressed in shHmga2 WT and shHmga2 empty cells were ranked in order and Sylamer analysis and plots were generated as in a. c, The set of transcripts expressed in shHmga2 ATG WT and shHmga2 empty cells were ranked in order and Sylamer analysis and plots were generated as in a. d, The set of transcripts expressed in shHmga2 m7 and shHmga2 empty cells were ranked in order and Sylamer analysis and plots were generated as in a. e, The set of transcripts expressed in shHmga2 ATG m7 and shHmga2 empty cells were ranked in order and Sylamer analysis and plots were generated as in a. f, Copies-per-cell analysis of Hmga2, Tgfbr3, and let-7 family members in 482N1 cells. Expression levels of transcripts in 482N1 cells were compared to a dilution series of in vitro transcribed mRNA (Hmga2 and Tgfbr3) or chemically synthesized microRNA (let-7a, let-7c, let-7e, let-7f and let-7i). Comparison of expression in 482N1 cells to the respective standard curves allowed calculation of copies per cell. Values are technical triplicates, have been performed independently three times, and represent mean with the 99% confidence interval. Confidence intervals were determined using a t-statistic.
Extended Data Figure 5 | Hmga2 regulates target transcripts through both protein and ceRNA function to promote TGF-β signalling. 

a, Effect of Hmga2 depletion on potential Hmga2 target genes from Supplementary Table 5. RNA was purified from 482N1 shluc and shHmga2 cells and qRT–PCR was performed for Gapdh and the overlapping set of transcripts upregulated in metastasis-generating lung cancer cells (Tmet) and predicted let-7 target transcripts described in Supplementary Table 5. Values are normalized to shluc. Values are technical triplicates, have been performed independently three times, and represent mean ± standard deviation (s.d.) with propagated error.
b, Effect of re-expression of either Hmga2 WT or ATG WT on target transcripts in knockdown cells. RNA was purified from 482N1 shluc empty, shHmga2 empty, shHmga2 WT and shHmga2 ATG WT cells and qRT–PCR was performed for Gapdh and the Hmga2 regulated transcripts from a. Values are normalized to shluc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± standard deviation (s.d.) with propagated error.
c, Induction of target transcripts by let-7 loss-of-function. RNA was purified from 482N1 shHmga2 cells expressing tough decoys targeting miR-122 or let-7 and qRT–PCR was performed on the transcripts from a. Values are normalized to the miR-122 tough decoy. Values are technical triplicates, have been performed independently three times, and represent mean ± standard deviation (s.d.) with propagated error.
d, Suppression of target transcripts by let-7 overexpression. RNA was purified from 482N1 cells transfected with either miR-10a or with let-7e and qRT–PCR was performed on the transcripts from a. Values are normalized to the miR-10a expressing cells. Values are technical triplicates, have been performed independently three times, and represent mean ± standard deviation (s.d.) with propagated error.
e, Hmga2 WT and ATG WT specifically induce expression of a TGF-β luciferase reporter (CAGA12) but not a control luciferase reporter (GL3) in both 368T1 and 482N1 cells. Values are normalized to co-transfected pRL-CXCR4 plasmid™. 368T1 values are normalized to empty and 482N1 values are normalized to shluc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error.
f, RNA was purified from 368T1 and 482N1 cells expressing the Hmga2 allelic series as in Fig. 1c and qRT–PCR was performed for various TGF-β target genes. Values are normalized to Gapdh. 368T1 values are normalized to empty and 482N1 values are normalized to shluc empty. Values are technical triplicates, have been performed independently three times, and represent mean ± standard deviation (s.d.) with propagated error. ***P < 0.0005; **P < 0.005.
Extended Data Figure 6 | Hmga2 ceRNA activity does not induce an epithelial–mesenchymal transition and regulates the Tgfbr3 3’ UTR in a miRNA-dependent manner.  

a, Lysates were prepared from the 368T1 and 482N1 Hmga2 allelic series and from Dicer1 expressing and Dicer1-null sarcoma cells (KPD F/- and -/-). Western blot analysis was subsequently performed for mesenchymal markers N-cadherin and vimentin, epithelial marker E-cadherin, and loading control Rho. 

b, Hmga2 WT and ATG WT specifically induce expression of a luciferase Tgfbr3 3’ UTR reporter in a let-7-site-dependent manner in KPD F/- cells but not KPD -/- cells. Cells were transfected with Renilla constructs of the control siCXCR4 multimer_ENREF_30 and either the Tgfbr3 wild-type or let-7-mutant 3’ UTR reporter. Values are normalized to co-transfected pGL3 plasmid. KPD F/- and -/- values are each normalized to empty. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. ***P < 0.0005.
Extended Data Figure 7 | Hmga2 ceRNA function does not alter Ago2 occupancy of let-7 family members. Lysates from 368T1 and 482N1 cells of the Hmga2 allelic series underwent either control immunoprecipitation (IgG) or immunoprecipitation for Ago2 as in Fig. 3d. RNA was purified and qRT–PCR was performed for miR-10a and let-7 family members on both the immunoprecipitated and input RNAs. Values are depicted as the percentage of input RNA, are technical triplicates, have been performed independently three times, and represent mean ± s.d.
Extended Data Figure 8 | Depletion of Tgfbr3 specifically represses TGF-β signalling and anchorage-independent growth without affecting general proliferation. a, Tgfbr3 depletion reduces TGF-β reporter activity in 368T1 WT and ATG WT and 482N1 cells. Cells were infected with the listed shRNAs and transfected with either GL3 or CAGA12 luciferase reporters as in Extended Data Fig. 5e. Values are normalized to co-transfected pRL-CXCR4 plasmid. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. b, Tgfbr3 knockdown suppresses TGF-β target gene expression. RNA was purified from 368T1 WT and ATG WT and 482N1 cells expressing the listed shRNAs and qRT–PCR was performed for various TGF-β target genes as in Extended Data Fig. 5f. Values are normalized to Gapdh and shluc. Values are technical triplicates, have been performed independently three times, and represent mean ± standard deviation (s.d.) with propagated error. c, Tgfbr3 knockdown does not generally affect proliferation. Cells were infected with the listed shRNAs and BrdU analysis was performed as in Extended Data Fig. 3a. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. (n = 3). d, 482N1 cells were transfected with a scrambled siRNA (siscr) or siRNAs targeting Hmga2, Angptl2, Colla1, Fndc3, Hmga1, Skil or Tgfbr3. RNA was purified and qRT–PCR was performed for Gapdh and the relevant target transcripts. Values are normalized to both Gapdh and specific transcript expression in siscr cells. Values are technical triplicates, have been performed independently twice, and represent mean ± standard deviation (s.d.) with propagated error. e, 482N1 cells transfected with the siRNAs described in d and were plated in soft agar, and the number of colonies was analysed as in Fig. 1d. Values are technical triplicates, have been performed independently twice, and represent mean ± s.d. ***P < 0.0005; **P < 0.005; *P < 0.05.
**RETRACTED**

(a) 482N1  368T1

- SB  + SB  + TGF-β, - SB  + SB  - TGF-β, - SB  - TGF-β, + SB  + SB

pSmad2  Smad2

(b) Fold Induction

|        | 368T1 WT | 368T1 ATG WT | 482N1 |
|--------|----------|--------------|-------|
| - SB   |          |              |       |
| + SB   |          |              |       |
| + TGF-β, - SB |      |              |       |
| + SB   |          |              |       |
| - TGF-β, - SB |      |              |       |
| + TGF-β, + SB |      |              |       |

(c) Normalised Expression

|        | 368T1 WT | 368T1 ATG WT | 482N1 |
|--------|----------|--------------|-------|
| Cdkn2b | + SB     |              |       |
| Tgfβ1  | + SB     |              |       |
| Tβra1  | + SB     |              |       |
| + SB   |          |              |       |
| - SB   | + SB     |              |       |
| - TGF-β, + SB |      |              |       |
| - TGF-β, - SB |      |              |       |
| + TGF-β, + SB |      |              |       |

(d) Colony Number

|        | 368T1 WT | 368T1 ATG WT | 482N1 |
|--------|----------|--------------|-------|
| - SB   | 150      |              |       |
| + SB   | 200      |              |       |
| + TGF-β, - SB |      |              |       |
| + SB   | 160      |              |       |
| - TGF-β, - SB |      |              |       |
| + TGF-β, + SB |      |              |       |

(e) Percentage BrdU+

|        | 368T1 WT | 368T1 ATG WT | 482N1 |
|--------|----------|--------------|-------|
| - SB   | 30       |              |       |
| + SB   | 50       |              |       |
| + TGF-β, - SB |      |              |       |
| + SB   | 40       |              |       |
| - TGF-β, - SB |      |              |       |
| + TGF-β, + SB |      |              |       |

(f) Normalised Expression

|        | Gapdsh  | Hmgs2  | Tgfb3  | Hmgat1 | Fndc3 | Skil  | C01a2 | Angpt2 |
|--------|---------|--------|--------|--------|-------|-------|-------|--------|
| - SB   | 0       |        |        |        |       |       |       |        |
| + SB   | 10      |        |        |        |       |       |       |        |
| + TGF-β, - SB |      |        |        |        |       |       |       |        |
| + SB   | 20      |        |        |        |       |       |       |        |
| - TGF-β, - SB |      |        |        |        |       |       |       |        |
| + TGF-β, + SB |      |        |        |        |       |       |       |        |

(g) Normalised Expression

|        | DMSO    | TGF-β | DMSO    | TGF-β | DMSO   | TGF-β | DMSO   | TGF-β | DMSO   | TGF-β | DMSO   | TGF-β |
|--------|---------|-------|---------|-------|--------|-------|--------|-------|--------|-------|--------|-------|
| - SB   | 0       |       |        |       |        |       |        |       |        |       |        |       |
| + SB   | 10      |       |        |       |        |       |        |       |        |       |        |       |
| + TGF-β, - SB |      |       |        |       |        |       |        |       |        |       |        |       |
| + SB   | 20      |       |        |       |        |       |        |       |        |       |        |       |
| - TGF-β, - SB |      |       |        |       |        |       |        |       |        |       |        |       |
| + TGF-β, + SB |      |       |        |       |        |       |        |       |        |       |        |       |

©2014 Macmillan Publishers Limited. All rights reserved
Extended Data Figure 9 | Inhibition of TGF-β signalling suppresses Hmga2 ceRNA-driven lung cancer cell transformation without altering general proliferation. a, SB-431542 (‘SB’) suppresses TGF-β pathway activation. 368T1 cells were treated with or without TGF-β ± SB, and 482N1 cells were treated with or without SB. Twenty-four hours later, lysates were prepared and Western blot analysis was performed for pSmad2 and total Smad2. b, SB suppresses TGF-β reporter activity. 368T1 WT and ATG WT and 482N1 cells were treated with or without SB for 24 hours and transfected with either GL3 or CAGA12 luciferase reporters as in Extended Data Fig. 5e. Values are normalized to co-transfected pRL-CXCR4 plasmid. Values are normalized to vehicle treatment (−SB). Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. with propagated error. c, SB suppresses TGF-β target gene expression. 368T1 WT and ATG WT and 482N1 cells were treated with or without SB for 24 hours. RNA was purified and qRT–PCR was performed for various TGF-β target genes as in Extended Data Fig. 5f. Values are normalized to Gapdh and to vehicle treatment (−SB). Values are technical triplicates, have been performed independently three times, and represent mean ± standard deviation (s.d.) with propagated error. d, SB suppresses anchorage-independent growth driven by Hmga2 ceRNA activity. 368T1 WT and ATG WT and 482N1 cells were plated for soft-agar colony growth in the continual absence or presence of SB. Soft-agar colony analysis was performed as in Fig. 1d. Values are technical triplicates, have been performed independently three times, and represent mean ± s.d. e, SB does not generally affect proliferation. 368T1 WT and ATG WT and 482N1 cells were treated with or without SB for 24 hours and BrdU analysis was performed as in Extended Data Fig. 3a. Values are technical triplicates, have been performed independently three times, mean ± s.d (n = 3). f, SB treatment suppresses multiple Hmga2 ceRNA target genes. 482N1 cells were treated with or without SB for 24 hours. RNA was purified and qRT–PCR was performed for Hmga2 and Hmga2 ceRNA targets. Values are normalized to vehicle treatment (−SB). Values are technical triplicates, have been performed independently twice, and represent mean ± standard deviation (s.d.) with propagated error. g, TGF-β treatment induces multiple Hmga2 ceRNA target genes. 368T1 cells were treated with vehicle (DMSO) or TGF-β for 24 hours. RNA was purified and qRT–PCR was performed for Hmga2 and Hmga2 ceRNA targets. Values are normalized to Gapdh and vehicle treatment (DMSO). Values are technical triplicates, have been performed independently twice, and represent mean ± standard deviation (s.d.) with propagated error. ***P < 0.0005; **P < 0.005; *P < 0.05.
Extended Data Figure 10 | HMGA2 and TGFBR3 are reciprocally and coordinately upregulated in an independent NSCLC-patient gene-expression data set. **a**, The Director’s Challenge lung adenocarcinoma gene expression data set was sorted on HMGA2 expression. The top and bottom quartiles (HMGA2 low and high, respectively) were selected (117 patients per group) and HMGA2 expression was compared using box-and-whisker plots. **b**, The Director’s Challenge data set was sorted into top and bottom quartiles of HMGA2 expression as in **a**, and TGFBR3 expression was compared using box-and-whisker plots. **c**, The Director’s Challenge data set was sorted into top and bottom quartiles of TGFBR3 expression (TGFBR3 low and high, respectively) as in **a**, and TGFBR3 expression was compared using box-and-whisker plots. **d**, The Director’s Challenge data set was sorted into top and bottom quartiles of TGFBR3 expression as in **c**, and HMGA2 expression was compared by box-and-whisker plots. In all box-and-whisker plots, values are presented on a log₂ scale. Significance was assessed by the Mann–Whitney test with a Bonferroni correction for multiple-hypothesis testing. **e**, Correlation analysis of HMGA2 and TGFBR3 expression was performed on all samples in the TCGA data set. All values are presented on a log₂ scale. Significance was assessed by Spearman correlation analysis. **f**, Correlation analysis of HMGA2 and TGFBR3 and Spearman analysis were performed on all samples in the Director’s Challenge data set as in **e**. ***P < 0.0005; **P < 0.005.***
Retraction: HMGA2 functions as a competing endogenous RNA to promote lung cancer progression

Madhu S. Kumar, Elena Armenteros-Monterroso, Philip East, Probir Chakravorty, Nik Matthews, Monte M. Winslow & Julian Downward

In this Letter, we reported that Hmga2 promotes lung cancer progression in mouse and human cells by operating as a competing endogenous RNA for the let-7 microRNA family. It has been brought to our attention that the cell lines used in the RNA sequencing (RNA-seq) experiment presented in Extended Data Fig. 4 of the Letter cannot be those specified in the figure legend. Moreover, these data cannot generate the Sylamer plots presented in this figure. The cell line substitution also casts doubt on which cells were used for experiments presented elsewhere in the paper. Although replication of other experiments in the Letter have not uncovered any further inconsistencies, given this uncertainty and the clear issues with Extended Data Fig. 4, we think it prudent to retract the paper pending more thorough investigation. We apologize for any adverse consequences that this may have caused.