Polycomb Group protein EZH2-mediated transcriptional repression of microRNA-338/-421 drives SPINK1-positive prostate cancer

Vipul Bhatia[1, 5], Anjali Yadav[1, 5], Ritika Tiwari[1], Shivansh Nigam[1], Sakshi Goel[1], Shannon Carskadon[2], Nilesh Gupta[3], Apul Goel[4], Nallasivam Palanisamy[2], Bushra Ateeq[1*]

[1]Molecular Oncology Lab, Department of Biological Sciences and Bioengineering, Indian Institute of Technology, Kanpur, 208016, U.P., INDIA
[2]Vattikuti Urology Institute, Department of Urology, Henry Ford Health System, Detroit, MI 48202, USA
[3]Department of Pathology, Henry Ford Health System, Detroit, MI 48202, USA
[4]Department of Urology, King George's Medical University, Lucknow, 226003, U.P., INDIA

[5]These authors contributed equally to this work.

Keywords: Prostate Cancer; SPINK1; EZH2; miRNA; *MALAT1*, EMT; Cancer Stem cells

Financial support: This work is supported by the Wellcome Trust/ DBT India Alliance grant (IA/I(S)/12/2/500635 to BA).

***Corresponding Author and Lead Contact**:
Phone: +91 512 2594083
Email: bushra@iitk.ac.in (B. Ateeq)

Conflict of interest: The authors declare no conflicts of interest or disclosures.

Abstract

Serine Peptidase Inhibitor, Kazal type-1 (SPINK1) overexpression defines the second largest subtype of prostate cancer (PCa), however, molecular mechanisms underlying its upregulation remains poorly understood. Here, we identified a critical role of miRNA-338-5p and miRNA-421 in post-transcriptional regulation of *SPINK1*. We show that SPINK1-positive PCa patients also exhibit overexpression of Polycomb group member EZH2, which confers repressive trimethylation marks on lysine 27 of histone 3 (H3K27me3) on the regulatory regions of these miRNAs. Further, we demonstrate that oncogenic lncRNA *MALAT1* interacts with EZH2, which in turn are targeted by miRNA-338-5p/miRNA-421, thus reinforcing a repressive molecular circuitry. Moreover, ectopic expression of miRNA-338-5p/-421 in SPINK1-positive PCa cells abrogate oncogenic properties including EMT, stemness and drug resistance, resulting in reduced tumor growth and distant metastases in mice. Collectively, we show that restoring miRNA-338-5p/miRNA-421 expression using epigenetic drugs or synthetic miRNA mimics could serve as a potential adjuvant therapy for treatment of *SPINK1*-positive malignancies.

Significance

SPINK1 overexpression is associated with aggressive prostate cancer subtype. We demonstrate EZH2-mediated epigenetic silencing of miR-338-5p/miR-421 leads to oncogenic overexpression of SPINK1. Ectopic expression of miRNA-338-5p/miRNA-421 in SPINK1+ cancer cells attenuate oncogenicity by targeting multiple pathways. Further, restoring miR-338-5p/miR-421 expression using synthetic mimics or epigenetic drugs could abrogate SPINK1-mediated oncogenicity.

Introduction

Prostate Cancer (PCa) is characterized by extensive molecular heterogeneity and varied clinical outcomes (1). Multiple molecular subtypes involving recurrent genetic rearrangements, DNA copy number alterations, and somatic mutations have been associated with this disease (2-6). Majority of the PCa patients harbor gene rearrangements between members of the *ETS* transcription factor family and the androgen-regulated transmembrane protease serine 2 (*TMPRSS2*), most recurrent (~50%) being *TMPRSS2-ERG*, a gene fusion involving the v-ets erythroblastosis virus E26 oncogene homolog (*ERG*) (5,7). The ERG transcription factor encoded by *TMPRSS2-ERG* fusion is known to drive cell invasion and metastases, induce DNA damage *in vitro,* and focal pre-cancerous prostatic intraepithelial neoplasia (PIN) lesions in transgenic mice (8,9).

While *TMPRSS2-ERG* fusion forms the most frequent molecular subtype, a significant subset of *ETS*-negative (–) PCa show overexpression of Serine Peptidase Inhibitor, Kazal type-1 (*SPINK1*) in ~10-15% of the total PCa patients, a distinct subtype defined by overall higher Gleason score, shorter progression-free survival and biochemical recurrence (10,11). SPINK1 promotes cell proliferation and invasion through autocrine/paracrine signaling and mediate its oncogenic effects in part through EGFR interaction by activating downstream signaling. Monoclonal antibody against EGFR showed only a marginal decrease in the growth of SPINK1-positive (+) xenografts in mice, supporting involvement of EGFR-independent oncogenic pathways (12).

Although, genomic events such as genetic rearrangements and somatic mutations constitute most recurrent oncogenic aberrations, many could also be attributed to epigenetic alterations. Earlier studies have shown that aberrant expression of Enhancer of Zeste Homolog 2 (EZH2) owing to genomic loss of miRNA-101 (13) or hypermethylation of miR-26a (14)

constitutes a common mechanism across several solid cancers including prostate. Moreover, EZH2, a key component of the Polycomb-Repressive Complex 2 (PRC2) mediates trimethylation on the histone 3 lysine 27 (H3K27me3), leading to gene silencing (15). However, phosphorylated form of EZH2 is known to switch its function from Polycomb repressor to transcriptional coactivator of androgen receptor in castration-resistant prostate cancers (CRPC) (16).

Nodal regulators such as the EZH2 are themselves under the control of other key players, such as lncRNAs and miRNAs, which regulate gene expression by mRNA degradation or translational inhibition (13,17). For instance, long non-coding RNA (lncRNA) such as Metastasis-associated Lung Adenocarcinoma Transcript 1 (*MALAT1*) specifically interacts with EZH2 and enhances EZH2-mediated gene repression in PRC2-dependent and -independent manner (18), imparting critical role in cancer progression and metastases. Moreover, PRC2 is known to epigenetically repress the expression of miR-181a/b, miR-200b/c, and miR-203, while these miRNAs in turn directly target PRC1 members, namely *BMI1* and *RING2*, in breast and prostate cancer (19).

Although SPINK1+ subtype forms a well-defined and second most prevalent subset of PCa, but the underlying mechanism involved in its upregulation is poorly understood and remains a matter of conjecture. Further, overexpression of *SPINK1* is not ascribed to chromosomal rearrangement, deletion, or amplification (10), and thus alludes to a possible transcriptional or post-transcriptional regulation. The present study uncovers the molecular mechanism involved in SPINK1 overexpression and shows how the *SPINK1* expression is regulated by miRs-338-5p and -421, which in turn are regulated by EZH2. The mechanism provides compelling evidence that EZH2 acts as an epigenetic switch, which promotes transcriptional silencing of miR-338-5p/miR-421 by establishing H3K27me3 repressive marks, thus leading to SPINK1 overexpression. Taken together, our findings suggest potential

benefits with epigenetic drugs such as EZH2 inhibitors or synthetic miR-338-5p/-421 mimics as an adjuvant therapy for the treatment of aggressive *SPINK1+* malignancies.

Results

Identification of differentially expressed miRNAs in *SPINK1+/ERG*-fusion-negative prostate cancer

We employed four miRNA prediction algorithms, namely PITA (omicstools.com), miRmap (mirmap.ezlab.org), miRanda (microRNA.org) and RNAHybrid (BiBiserv2-RNAhybrid) to examine putative binding of miRNAs to the 3' untranslated region (3'UTR) of *SPINK1* transcript. Notably, three miRs -338-5p, -421 and -876-5p were predicted as strong candidates by all four algorithms (Fig. 1A and Supplementary Table S1), and hence were taken forward for further investigation. To determine, whether these three miRNAs show any differential expression between *SPINK1+* and *ERG+* PCa patients' specimens, RNA-seq data available at public repository, The Cancer Genome Atlas Prostate adenocarcinoma (TCGA-PRAD) was analyzed. Interestingly, hierarchical clustering of TCGA-PRAD RNA-Seq dataset exhibit reduced expression of miR-338-5p and miR-421 (miR-338-5p/-421) in *SPINK1+/ERG*-negative patient specimens (Fig. 1B). To validate further, we examined the expression of miR-338-5p/-421 and miR-876-5p in our PCa patients' specimens. A significant lower expression of miR-338-5p and miR-421 was observed specifically in *SPINK1+* as compared to *ERG+* specimens (Fig. 1C), while no difference in miR-876-5p expression was noticed (Fig. 1C and Supplementary Fig. S1A). To understand the clinical significance of miR-338-5p/-421, we stratified TCGA-PRAD patients' data into high and low miRNAs expressing groups, intriguingly group with low miR-338-5p expression show significant (P=0.0024) association with decreased survival probability compared to high miRNAs group (Supplementary Fig. S1B), while no such association was found in case of miR-421. Moreover, lower expression of

miR-338-5p also associate with higher Gleason score, advanced clinical T score and lymph node status (Supplementary Fig. S1C). An association of higher Gleason score with lower expression of both miRNAs was further confirmed in another independent cohort (GSE45604) (Supplementary Fig. S1D). In summary, *SPINK1+* subtype show lower expression of miR-338-5p/-421, which strongly associate with over-all poor survival and aggressiveness of the disease.

MiR-338-5p and miR-421 directly target *SPINK1* and modulate its expression

Having established an association between miR-338-5p/ -421 and *SPINK1* expression in PCa specimens (Fig. 1, B-C), we next examined the ability of these miRNAs to bind to the 3'-untranslated region (3'UTR) of *SPINK1*. The wild-type (3'-UTR-WT) and mutant (3'-UTR-mut) *SPINK1* 3'-UTR cloned in Firefly/Renilla dual-luciferase reporter vectors were co-transfected with synthetic mimics for miR-338-5p or miR-421 in HEK293T cells, a significant reduction in the luciferase activity was noted with 3'-UTR-WT, while 3'-UTR-mut constructs failed to show any suppressive effect (Fig. 1D). We next evaluated the expression of these miRNAs in various PCa cell lines including 22RV1 (*SPINK1+*), *ETS*-fusion positive VCaP (*TMPRSS2-ERG+*) and LNCaP (*ETV1+*) cells. Supporting our observation in clinical specimens the cell line data also showed lower expression of miR-338-5p/-421 in the 22RV1 cells relative to fusion-positive cell lines (Supplementary Fig. S1E). To further ascertain that miR-338-5p/miR-421 specifically regulates *SPINK1*, we used antagomiRs to abrogate miR-338-5p and miR-421 expression (anti-338-5p and anti-421, respectively) in VCaP cells (Supplementary Fig. S1F). As expected, anti-338-5p or anti-421 significantly induced *SPINK1* expression in VCaP cells with concomitant increase in cell invasion and migration (Fig. 1E and 1F and Supplementary Fig. S1G and S1H), while there was no change in the endogenous ERG expression (Fig. 1E and Supplementary Fig. S1G). Conversely, we observed that 22RV1 cells stably overexpressing miR-338-5p or miR-421 (22RV1-miR-338-5p and 22RV1-miR-421,

respectively) show a significant reduction in *SPINK1* expression at both transcript (~80-90%) and protein (Fig. 1G) levels. Since, SPINK1 overexpression has also been implicated in colorectal, lung, pancreatic, and ovarian cancers (20), we sought to examine if SPINK1 is regulated by a similar mechanism in cancers of different cellular/tissue origins. Thus, we determined the status of SPINK1 expression in multiple cancer cell lines (Supplementary Fig. S2A and S2B). Furthermore, *SPINK1*+ cancer cell lines, namely, colorectal (WiDr), melanoma (SK-MEL-173), pancreatic (CAPAN-1) and prostate (22RV1) upon transfecting with mimics for miR-338-5p or miR-421 showed a significant decrease in SPINK1 expression both at transcript and protein levels (Supplementary Fig. S2C and S2D). This provides irrevocable evidence that these two miRNAs modulate the expression of *SPINK1* transcript irrespective of the tissue background. Furthermore, to ascertain whether decrease in oncogenic properties is indeed due to miR-338/-421 mediated reduction in SPINK1 expression, a rescue cell migration assay using human recombinant SPINK1 (rSPINK1) was performed. As expected, miR-338 and miR-421 overexpressing 22RV1 cells show decrease in cell migration, while adding rSPINK1 to these miRNAs overexpressing cells rescued the invasive phenotype, indicating that miR-338/-421 mediated effects are indeed due to decrease in SPINK1 expression (Supplementary Fig. S2E).

Ectopic expression of miR-338-5p and miR-421 attenuate SPINK1-mediated oncogenesis

SPINK1 overexpression is known to contribute to cell proliferation, invasion, motility and distant metastases (10,12,21). Hence, to understand the functional relevance of miR-338-5p/ -421, we examined 22RV1-miR-338-5p and 22RV1-miR-421 stable cells for any change in their oncogenic properties. Both 22RV1-miR-338-5p (C1 and C2) and 22RV1-miR-421 (pooled and C1) cells showed a significant decrease in cell proliferation compared to control (22RV1-CTL) cells (Fig. 2A). Similarly, reduced invasive properties of 22RV1-miR-338-5p and 22RV1-miR-421 cells were noted (~40% and 60% respectively) (Fig. 2B). While, only a

modest decrease in cell proliferation and invasion was observed in pooled 22RV1-miR-338-5p cells (Supplementary Fig. S3A and S3C). To assess neoplastic transformation, soft agar colony formation assay was performed, where both 22RV1-miR-338-5p and 22RV1-miR-421 cells exhibited marked reduction (~60% and ~80% respectively) in number and size of the colonies (Fig. 2C). Likewise, 22RV1-miR-338-5p and 22RV1-miR-421 cells demonstrate significantly lower numbers (~70% and ~60% respectively) of dense foci (Fig. 2D). Next, to examine whether overexpression of these miRNAs in benign immortalized prostate epithelial RWPE-1 cells show any phenotypic change, we performed cell-based functional assays. As expected, no significant change in cell proliferation or migration was observed in cells transfected with miR-338 mimic, while miR-421 mimic shows a marginal decrease in proliferation and migration (Supplementary Fig. S3D). Further, to examine the effect of these miRNAs in a SPINK1-independent context, we established stable miR-338-5p/-421 overexpressing prostate cancer PC3 cells and carried-out functional assays. Surprisingly, a significant decrease in cell proliferation, migration and foci formation was observed in miRNAs overexpressing PC3 cells (Supplementary Fig. S3E and S3G), suggesting that these miRNAs perhaps also target key regulators involved in cell division and motility. To demonstrate that miR-338-5p/-421 modulate *SPINK1* expression and attenuate SPINK1-mediated oncogenicity irrespective of the tissue background, we performed functional assays using colorectal carcinoma WiDr cells (SPINK1+) stably overexpressing these miRNAs. As anticipated, a significant decrease in the oncogenic potential of the miR-338-5p/-421 overexpressing WiDr cells was observed (Supplementary Fig. S3H and S3I).

To examine tumorigenic potential of 22RV1-miR-338-5p and 22RV1-miR-421 cells *in vivo*, chick chorioallantoic membrane (CAM) assay was performed, and relative number of intravasated cancer cells was analyzed. Consistent with *in vitro* results, 22RV1-miR-338-5p and 22RV1-miR-421 cells showed significant reduction in the number of intravasated cells

compared to control (Supplementary Fig. S4A and S4B). Likewise, a significant reduction in the tumor weight was recorded in the groups implanted with 22RV1-miR-338-5p and 22RV1-miR-421 cells (Fig. 2E). To evaluate distant metastases, lungs and liver excised from the chick-embryos were characterized for the metastasized cancer cells. The groups implanted with miRNAs overexpressing cells revealed ~80% reduction in cancer cell metastases to lungs (Fig. 2F), while no sign of liver metastases was observed in either group. Further, tumor xenograft experiment was recapitulated in immunodeficient NOD/SCID mice (n=8 per group) by subcutaneously implanting 22RV1-miR-338-5p, 22RV1-miR-421 and control 22RV1 cells into flank region, and trend of tumor growth was recorded. A significant reduction in the tumor burden was observed in the mice bearing miR-338-5p and miR-421 overexpressing xenografts as compared to control (~70% and 85% reduction respectively) (Fig. 2, G-H). To examine spontaneous metastases, lung, liver and bone marrow specimens were excised from the xenografted mice, and genomic DNA was quantified for the presence of human specific *Alu*-sequences. A significant decrease (~85% for miR-338-5p and ~90% for miR-421) in cancer cell metastases was observed in the group implanted with miRNAs overexpressing cells (Fig. 2I). Similar to CAM assay, cancer cells failed to metastasize to murine liver (data not shown). Furthermore, significant drop (~50%) in Ki-67-positive cells in the miRNAs overexpressing xenografts confirms that tumor regression was indeed due to decline in cell proliferation (Fig. 2J). Taken together, our findings indicate that miR-338-5p/-421 downregulate the expression of *SPINK1* and abrogate SPINK1-mediated oncogenic properties and tumorigenesis.

MiR-338-5p and miR-421 exhibit functional pleiotropy by regulating diverse biological processes

To explore critical biological pathways involved in the tumor-suppressive properties rendered by miR-338-5p/ -421 in SPINK1+ cancers, we determined global gene expression profiles of miRNAs overexpressing 22RV1 cells. Our analysis revealed 2,801 and 2,979 genes

significantly dysregulated in 22RV1-miR-338-5p and 22RV1-miR-421 cells respectively relative to control, when filtered by $\log_2$ fold change of 0.6, *FDR*<0.05 and *P*<0.05 (Supplementary Table S2 and S3). Remarkably, ~22% (704 genes) of the downregulated and ~15% (506 genes) of the upregulated transcripts show an overlap in miR-338-5p and miR-421 overexpressing cells (90% confidence interval) (Fig. 3A), indicating that these two miRNAs regulate a significant number of common gene sets and cellular processes. To examine biological processes commonly regulated by miR-338-5p/-421, we employed DAVID (Database for Annotation, Visualization and Integrated Discovery) and GSEA (Gene set enrichment analysis). Most of the downregulated genes were associated with DNA double-strand break repair by homologous recombination, cell cycle regulation including G2/M-phase transition, stem-cell maintenance, histone methylation and negative regulation of cell-cell adhesion. Whereas, genes involved in negative regulation of gene expression or epigenetics, intrinsic apoptotic signaling pathways, negative regulation of metabolic process and cell cycle were significantly upregulated (Fig. 3B, Supplementary Table S4 and S5). Moreover, GSEA also revealed enrichment of gene signatures associated with oncogenic pathways and cancer hallmarks. Conversely, 22RV1-CTL cells showed significant enrichment of genes involved in sustaining proliferative signaling (EGFR and MEK/ERK) and cell cycle regulators (E2F targets and G2/M transition). While, positive enrichment for tumor suppressive p53 signaling was found in miRNA overexpressing cells as compared to control (Fig. 3C), indicating its role in reduced oncogenicity. Additionally, an overlapping network of pathways using Enrichment map revealed regulation of cell-cycle phase transition and DNA repair pathways (overlap coefficient=0.8, *P*<0.001, *FDR*=0.01), as one of the significantly enriched pathways for both miRNAs (Supplementary Fig. S5).

Since MAPK signaling pathways involving a series of protein kinase cascades play a critical role in the regulation of cell proliferation, we examined the phosphorylation status of

MEK (pMEK) and ERK (pERK), as a read-out of this pathway. In agreement with our *in-silico* analysis, a significant decrease in pMEK and pERK was observed in 22RV1-miR-338-5p and 22RV1-miR-421 cells (Fig. 3D). E2F transcription factors are known to interact with phosphorylated retinoblastoma, and positively regulate genes involved in S-phase entry and DNA synthesis (22), thus we next examined the E2F1 level in miRNAs overexpressing cells, surprisingly a notable decrease in E2F1 was observed (Fig. 3D). Further, a significant decrease in the expression of genes involved in G1/S transition such as cyclin E2 (*CCNE2*), cyclin A2 (*CCNA2*) and cyclin-dependent kinase (*CDK1* and *CDK6*), including mini-chromosome maintenance (*MCM3* and *MCM10*), required for the initiation of eukaryotic replication machinery was recorded (Fig. 3E). Thus, these findings corroborate with previous literature that during DNA damage, CDKs being cell-cycle regulators crosstalk with the checkpoint activation network to temporarily halt the cell-cycle progression and promote DNA repair (23). Intriguingly, presence of putative miR-338-5p/miR-421 binding sites on the 3'UTRs of these cell cycle regulators (Supplementary Table S6) further support that these targets could be directly controlled by these miRNAs. Next, to validate that miR-338-5p/-421 overexpression leads to S-phase arrest, 22RV1 cells transfected with miR-338-5p or miR-421 mimics were subjected to cell cycle analysis, a significant increase in the S-phase arrested cells was noted (Supplementary Fig. S6A). To delineate that this increase in the S-phase cells is indeed due to cell-cycle arrest and not because of DNA replication, BrdU-7AAD-based cell cycle analysis was performed, which revealed a significant decrease in the percentage of BrdU incorporated cells in S-phase (Fig. 3F). Next, we performed Annexin V-PE staining to examine if overexpression of miR-338-5p/-421 lead to apoptosis, a marginal increase in the early apoptotic cells was evident in miR-338-5p/-421 mimics transfected cells (Supplementary Fig. S6B). Taken together, our findings strongly indicate that miR-338-5p/-421 overexpression led to S-

phase arrest, thus elucidating the mechanism for reduced cell proliferation and dramatic regression in tumor growth.

MiR-338-5p and miR-421 regulate oncogenic long non-coding RNA *MALAT1* post-transcriptionally

MiRNAs are known to regulate the expression of multiple coding as well as long non-coding RNAs (lncRNAs). Thus, we examined lncRNAs which are specifically dysregulated in 22RV1-miR-338-5p and 22RV1-miR-421 cells, gene expression profiling revealed several deregulated lncRNAs including *MALAT1* which is associated with metastatic cancers (24) (Fig. 4A). To ascertain the association between *MALAT1* and miRNAs, we analyzed *MALAT1* expression in 22RV1-miR-338-5p and 22RV1-miR-421 cells, a significant decrease in the *MALAT1* transcript levels were observed in miRNAs overexpressing cells as compared to control (Fig. 4B). Further, miRNAs target prediction indicates putative miR-338-5p/ -421 binding sites on the *MALAT1* transcript (Fig. 4C), thus *MALAT1* region harboring miR-338-5p (+7233-7421 bp) or miR-421 (+6501-6708 bp) binding sites was cloned in the luciferase reporter plasmid (MALAT1-Luc-338 and MALAT1-Luc-421 respectively). Subsequently, HEK-293T cells co-transfected with MALAT1-Luc-338 or MALAT1-Luc-421 and respective miRNA mimics, showed a significant reduction in the luciferase reporter activity (Fig. 4C). To ascertain that miR-338-5p and miR-421 target *MALAT1* post-transcriptionally, VCaP cells transfected with anti-338-5p or anti-421 were characterized for *MALAT1* expression, as expected a significant increase in *MALAT1* and *SPINK1* expression was observed (Fig. 4D).

Differential expression of *MALAT1* has been associated with G1/S and G2/M transitions of the cell cycle (25). Thus, we sought to understand the inference of miR338-5p/miR-421/*MALAT1* axis in cell cycle regulation. Putative binding sites of E2F1, an important regulator of G1/S phase transition, were found within critical region (+50bp to -700bp) of the

MALAT1 promoter (26). To explore the functional interplay between E2F1 and *MALAT1*, we silenced *E2F1* in VCaP cells, which resulted in significant decrease in *MALAT1* expression (Fig. 4E). Further, significant enrichment of E2F1 over input on the *MALAT1* promoter was confirmed by chromatin immunoprecipitation (ChIP) assay (Fig. 4F), indicating involvement of E2F1 in the regulation of *MALAT1*. Taken together, we demonstrated that miR-338-5p/-421 regulate the expression of two oncogenic drivers, *SPINK1* and *MALAT1*, and possibly mediate S-phase cell cycle arrest in *MALAT1*- and E2F1-dependent manner (Fig. 4G). Thus, our findings elucidate one of the plausible mechanisms involved in conferring aggressive phenotype of SPINK1+ subtype, typically associated with higher Gleason score and metastases.

Ectopic expression of miR-338-5p and miR-421 suppresses Epithelial-to-Mesenchymal Transition (EMT) and stemness

Association between EMT and cancer stem cells (CSCs) has been well-established, indicating that a subpopulation of neoplastic cells, which harbor self-renewal capacity and pluripotency, are associated with highly metastatic and drug-resistant cancers (27). Since miR-338-5p/-421 overexpression in 22RV1 cells show regression in tumor burden and metastases (Fig. 2, F-I), we evaluated our microarray data for the genes involved in EMT and stemness, and noted a marked decrease in their expression (Fig. 5A) including the EMT-inducing transcription factors (28) namely, *SNAI1* (*SNAIL*), *SNAI2* (*SLUG*), and *TWIST1* (Fig. 5, B-C). Since, SNAIL and SLUG are known to negatively regulate *CDH1 (E-Cadherin*) (29), an epithelial marker involved in cell-cell adhesion, we next examined E-Cadherin expression. Interestingly, miR-338-5p/ -421 overexpressing cells show a prominent increase in the membrane localization of E-Cadherin, while a significant decrease in the expression of vimentin, a mesenchymal marker was observed (Fig. 5D).

In addition, the expression of genes associated with cancer stem cell-like properties were examined in 22RV1-miR-338-5p and 22RV1-miR-421 cells. Strikingly, the expression of well-known pluripotency markers, such as *AURKA, SOX9* and *OCT-4,* and stem-cell surface markers *EPCAM, CD117* (*c-Kit*), and *ABCG2*, an ATP-binding cassette transporter, were markedly downregulated in miRNAs overexpressing cells (Fig. 5, E-F). Moreover, a sub-population (*CD117*$^{+}$*/ABCG2*$^{+}$) of 22RV1 cells, known as prostate carcinoma-initiating stem-like cells, exhibits stemness and multi-drug resistance (30). Having confirmed that miR-338-5p/-421 downregulate expression of *ABCG2* and *c-Kit*, we next examined the efflux of Hoechst dye *via* ABC-transporters in the absence or presence of verapamil, a competitive inhibitor for ABC transporters (31). As expected, 22RV1-miR-338-5p and 22RV1-miR-421 cells show a significant reduction (~91% and 89% respectively) in the side population (SP) cells involved in Hoechst dye efflux (Fig. 5G). Efflux assay performed in the presence of verapamil show substantial reduction in the SP cells due to inhibition of Hoechst efflux in both control and miRNAs overexpressing cells (Fig. 5G). Further, to confirm that overexpression of these microRNAs lead to decrease in CSC-like properties, prostatosphere assay, a surrogate model for testing enhanced stem cell-like properties was performed. As expected, 22RV1-miR-338-5p and 22RV1-miR-421 cells showed a significant decrease in the size and prostatosphere formation efficiency (Fig. 5, H-I). Moreover, prostatospheres formed by miRNAs overexpressing cells exhibit a significant reduction in the expression of genes implicated in cancer cells self-renewal and stemness (Fig. 5J). Intriguingly, miR-338-5p and miR-421 putative binding sites on the 3'UTR of *EPCAM, c-Kit, SOX9, SOX2* and *ABCG2* were also noticed (Supplementary Table S6), suggesting a possible mechanism involved in the downregulation of these genes.

Epigenetic regulators, such as ten-eleven-translocation (TET) family member, TET1, converts 5'-methylcytosine (5mC) to 5'-hydroxymethylcytosine (5hmC), are well-known to

induce pluripotency and maintain self-renewal capacity (32). Thus, we analyzed the expression of TET family members in miRNAs overexpressing cells; strikingly a significant decrease in TET1 was observed (Fig. 5K). Since, *ABCG2* and *c-Kit*, which are implicated in drug-resistance, were downregulated in miRNAs overexpressing cells, thus sensitivity of these cells to chemotherapeutic drug was evaluated. Interestingly, 22RV1-miR-338-5p and 22RV1-miR-421 cells show enhanced sensitivity to doxorubicin as compared to control (Supplementary Fig. S6C). Collectively, miR-338-5p/-421 downregulate the expression of genes implicated in multiple oncogenic pathways namely EMT, stemness and drug resistance, signifying that these two tumor suppressor miRNAs could represent a novel approach for integrative cancer therapy (Fig. 5L).

EZH2-mediated transcriptional repression of miR-338 and miR-421 drives SPINK1-positive prostate cancer

Aberrant transcriptional regulation, genomic loss or epigenetic silencing are well-known mechanisms involved in miRNAs deregulation (33,34). Since SPINK1+ PCa patients exhibit reduced expression of miR-338-5p/-421, we sought to decipher the mechanism involved in miRNAs silencing. EZH2, being a member of polycomb group protein play critical role in epigenetic gene silencing by promoting H3K27me3 marks. Thus, we interrogated Memorial Sloan Kettering Cancer Center (MSKCC) patients' cohort using cBioPortal (http://cbioportal.org) for any plausible association between *SPINK1* and *EZH2* expression. Interestingly, most of the SPINK1+ specimens comprising Gleason scores 3 and 4 show concordance with EZH2 expression (Fig. 6A). Further, TCGA-PRAD patients harboring higher expression of *EZH2* show increased levels of *SPINK1*, *MALAT1*, and decreased expression of miR-338-5p/-421 as compared to EZH2-low patients (Fig. 6B). To further confirm the concordance between these two oncogenes, we subsequently evaluated SPINK1 and EZH2 status by performing immunohistochemistry (IHC) and RNA in situ hybridization

(RNA-ISH) respectively on the paraffin embedded tissue microarrays (TMAs) comprising a total of 238 PCa specimens. Interestingly, of the 238 PCa specimens evaluated, 21% (50 cases) were positive for SPINK1 expression, and 88% (44 cases) of these SPINK1+ specimens show positive staining for EZH2 (Fig. 6C). While, 75% (141 cases) of the SPINK1-negative (SPINK1–) patients show EZH2 expression as well, notably 71% of these SPINK1–/EZH2+ cases exhibit lowest EZH2 intensity (score 1). Conversely, trend shown in Fig. 6C depicts about ~50% of the SPINK1+/EZH2+ patients fall into low EZH2 expression group (score 1), ~36% in medium EZH2 (score 2), and ~14% in high EZH2 range (score 3 and 4), indicating a significant association between SPINK1+ status and EZH2 expression (Fig. 6C; χ^2=13.66; *P*=0.008). Thus, in corroboration to previous reports (4,14), our data suggests a more pronounced role of epigenetic alterations in *ETS*-fusion negative cases. Although, 6 of the 50 SPINK1+ cases failed to show any expression of EZH2, pointing that an alternative mechanism may be involved in SPINK1 regulation or possibly miRNA-338/-421 genomic deletion could be a cause in such cases. Additionally, in another independent PCa cohort (GSE35988), increased expression of *SPINK1* and *EZH2* was observed in localized as well as metastatic specimens, while *MALAT1* was specifically upregulated in metastatic cases as compared to benign or localized (Supplementary Fig. S7A), indicating a plausible interplay of these oncogenic drivers in disease progression.

To investigate whether epigenetic silencing of these miRNAs is mediated by EZH2, we screened the promoters of miR-338, miR-421 and *FTX* (miR-421 host gene) for the putative transcription factor binding sites and identified MYC and MAX (Myc-Associated Factor X) elements within ~2 kb upstream of Transcription Start Site (TSS). MYC is known to form a repressive complex with EZH2 and HDACs, and downregulate multiple tumor suppressive miRNAs, which in turn target PRC2-interacting partners (35). In addition, EZH2-silenced DU145 cells miRNA expression data (GSE26996) indicates an increase in the expression of

numerous EZH2-regulated miRNAs including miR-338 and miR-421 (Supplementary Fig. S7B). We therefore examined the promoters of miR-338, miR-421 and *FTX* for the recruitment of EZH2, interestingly a significant enrichment of EZH2 over input was observed on the promoters of miR-338 and *FTX* (Fig. 6D). No enrichment on miR-421 promoter was observed (Supplementary Fig. S7C), indicating that the host gene *FTX* promoter regulates the expression of this intronic miRNA. Next, to confirm EZH2-mediated methyltransferase activity, we sought to identify H3K27me3 marks on these promoters, a remarkable enrichment of H3K27me3 marks on the miR-338 and *FTX* promoters were noted relative to IgG control (Fig. 6D and Supplementary Fig. S7C), confirming the role of EZH2 mediated epigenetic silencing of miRNA-338-5p/-421.

Comprehensive GSEA analysis revealed that miRNA-338/-421 overexpressing cells show an enrichment for EZH2 interacting partners, including PRC2 members (36) and EZH2 regulated genes (37,38) (Fig. 6E and Supplementary Fig. S7D), indicating that these two miRNAs in turn regulate EZH2 partners and their target genes. Thus, we next examined the putative binding of miRNA-338-5p/-421 on the 3'UTR of the PRC2 members, interestingly both miRNAs show negative mirSVR binding score (Supplementary Table S6). Moreover, a significant decrease in the transcript levels of *EZH2*, and its interacting partners *SUZ12*, *RBBP4*, *RBBP7* and *MTF2* were observed in miRNA-338-5p/-421 overexpressing cells (Supplementary Fig. S7E and Fig. 6F). Subsequently, we checked for EZH2 recruitment and H3K27me3 histone methylation marks on the promoters of miR-338 and *FTX* in stable 22RV1-miR-338-5p, 22RV1-miR-421 and control cells. As expected a significant decrease in the EZH2 occupancy and H3K27me3 repressive marks were observed on the promoter regions of miR-338 and *FTX* in miRNA overexpressing cells (Fig. 6G). Since, *MALAT1* is known to interact with EZH2 and facilitates its recruitment on its target genes (18), thus to confirm this interaction we performed RNA immunoprecipitation (RIP) assay in 22RV1 cells. Interestingly,

immune-complex pulled down by EZH2 antibody show ~22-folds enrichment of *MALAT1* as compared to IgG control (Fig. 6H), indicating that *MALAT1* directly binds to EZH2, might promote its occupancy, and H3K27me3 repressive marks at miR-338-5p/-421 promoters leading to epigenetic silencing. Collectively, our data also indicate that overexpression of miR-338-5p/-421 downregulates EZH2 expression and its interacting members, leading to impaired histone methyltransferase activity of PRC2 and reduced SPINK1-mediated oncogenicity, thereby establishing a double-negative feedback loop.

Since inhibitors for chromatin modifiers are known to erase epigenetic marks, we tested 3-Deazaneplanocin A (DZNep), an inhibitor of the histone methyltransferase; 2'-deoxy-5-azacytidine (5-Aza), a DNA methyltransferase (DNMT) inhibitor and Trichostatin A (TSA), a HDAC inhibitor in 22RV1 cells and examined the expression of miR-338-5p/-421. Treatment with TSA, DZNep, 5-Aza alone or a combination of DZNep and TSA in 22RV1 cells showed a modest increase in miR-338-5p/-421 expression, while 5-Aza and TSA together resulted in ~9-fold increase (Fig. 7A). Furthermore, 5-Aza and TSA combination results in significant increase in miRNAs expression accompanied with a notable decrease (~60-80%) in *SPINK1* levels (Fig. 7B). Since, 3'-arm of miR-338 (miR-338-3p) is known to negatively regulate *Apoptosis Associated Tyrosine Kinase (AATK*) expression (39), likewise a significant reduction in the *AATK* expression was noticed in our study (Fig. 7B). Furthermore, a deletion construct of *FTX* show decreased expression of miR-374/-421 cluster (40). In line with this, a significant increase in the *FTX* and miR-421 expression was reported upon 5-Aza and TSA combinatorial treatment, signifying the importance of host gene *FTX* in the regulation of miR-421 (Fig. 7B).

Furthermore, EZH2 is also known to interact with DNMTs, thus enabling chromatin remodeling and DNA methylation (41). Hence, we next examined the presence of methylated CpG marks on the promoters of miR-338-5p and *FTX*. Interestingly, methylated DNA immunoprecipitation (MeDIP) revealed locus-specific enrichment in the 5mC levels over

5hmC on these regulatory regions (Fig. 7C). To ascertain the presence of DNA methylation marks we performed bisulfite sequencing using PCa cell lines, a relative increase in the methylated CpG sites on miR-338 and *FTX* promoters was observed in 22RV1 cells (SPINK1-positive) as compared to VCaP (ERG-positive) cells (Fig. 7D). No significant difference in the methylated CpG sites on the *AATK* and miR-421 promoters was observed (Supplementary Fig. S7F and S7G). To understand clinical relevance, bisulfite sequencing was carried out on SPINK1-positive (n=5) and *ERG* fusion positive (n=5) PCa patients' specimens. Interestingly, all SPINK1-postive specimens exhibit increased methylation marks on the promoters of miR-338 and *FTX* as compared to *ERG* positive (Fig. 7D). Taken together, our results strongly indicate that epigenetic machinery comprising of EZH2 and its interacting partners play a critical role in the epigenetic silencing of miRNA-338-5p and miR-421 in *SPINK1*+ subtype, which in turn reaffirms its silencing by a positive feedback loop.

Discussion

In this study, we unraveled the underlying molecular mechanism involved in the overexpression of SPINK1 exclusively in *ETS*-fusion negative PCa. Our study provides a molecular basis for SPINK1 overexpression, brought about by the epigenetic repression of key post-transcriptional negative regulators of *SPINK1* namely, miR-338-5p and miR-421. We demonstrated the tumor suppressive roles of miR-338-5p/-421, which exhibits functional anti-cancer pleiotropy in *SPINK1*+ subtype, by attenuating oncogenic properties, tumor growth and metastases in murine model. Conversely, miR-421 has also been reported to be a potential oncogenic miRNA in multiple cancers (42,43). However, in corroboration with our findings, a recent report suggested tumor suppressive role of miR-421 in prostate cancer (44). We also established that miR-338-5p/-421 overexpressing cells display perturbed cell-cycle machinery triggered by dysregulated cyclins and CDKs, subsequently leading to S-phase arrest. It has been shown that miRNAs targeting multiple cyclins/CDKs are more effective than the FDA-

approved CDK4/6 inhibitor in triple-negative breast cancer (45), thus supporting our findings that replenishing these miRNAs may prove advantageous in SPINK1+ cancers. Moreover, besides SPINK1, miR-338-5p/-421 also targets *MALAT1*, which could possibly result in downregulation of E2F1, thus contributing to arrest in G1/S transition, although in-depth molecular mechanisms behind this remain to be elucidated.

Emerging evidences suggest a complex interaction between EMT and CSCs during cancer progression, and in developing resistance towards anti-cancer drugs. Previous studies have implicated the role of several miRNAs, such as miR-200 family, miR-205 and miR-34a (46,47) in regulating the expression of genes involved in metastases, stemness and drug resistance. Furthermore, miR-338 exhibits tumor suppressive role, and inhibits EMT by targeting *ZEB2* (48) and *PREX2a* (49) in gastric cancer. Here, we identified miR-338-5p/-421 as critical regulators of EMT-inducing transcription factors and -associated markers, which in turn led to decreased stem-cells like features. Moreover, CSCs are known to express ABC transporters, which efflux the chemotherapeutic drugs during resistance (50). Remarkably, miR-338-5p/-421 overexpression shows decreased expression of *ABCG2* and *c-KIT*, consequently a significant drop in the drug-resistant side population, indicating that these two miRNAs are highly effective in conferring drug-sensitivity and reducing the therapy-resistant CSCs. Collectively, our findings provide a solid foundation for qualifying these miRNAs as an adjuvant therapy for the *SPINK1+* as well as other drug resistant malignancies.

Numerous lncRNAs have been reported to be dysregulated in prostate cancer (51,52), for instance, *PCGEM1* and *PRNCR1* are highly expressed in aggressive PCa and enhance the AR-mediated gene activation program (53). Furthermore, *MALAT1* renders cell cycle arrest in G1/S transition and mitotic phase by modulating the expression of *E2F1* and *B-MYB* (25). Similarly, nuclear lncRNAs often interact with the components of chromatin-remodeling complexes such as EZH2, SUZ12, CBX7, CoREST and JARID1C, and mediate gene silencing

or activation by modulating their activity (52,54). For example, *MALAT1*, a known nuclear lncRNA interacts with EZH2, facilitates its occupancy and the H3K27me3 activity on the PRC2 target genes (18), in corroboration to this, we also found that *MALAT1* interacts with EZH2 and might facilitate its recruitment on miRNAs regulatory regions. Hence, we propose a molecular model for the functional interplay involving *SPINK1*, *MALAT1* and miR-338-5p/-421, wherein *MALAT1* facilitates recruitment of EZH2, which acts as an epigenetic switch and by its histone methyltransferase activity establishes H3K27me3 repressive marks on the promoters of miR-338 and *FTX*, a miR-421 host gene (Fig. 7E). This finding was further strengthened by a recent TCGA study (4), wherein a subset of PCa patients' harboring *SPOP*-mutation/*CHD1*-deletion exhibits elevated DNA methylation levels accompanied with frequent events of *SPINK1* overexpression. Recently, a new subtype of *ETS*-fusion-negative tumors has been defined by frequent mutations in the epigenetic regulators and chromatin remodelers (55). Yet another study, using genome-wide methylated DNA-immunoprecipitation sequencing revealed higher number of methylation events in *TMPRSS2-ERG* fusion-negative as compared to normal and *TMPRSS2–ERG* fusion-positive PCa specimens (14), thus collectively, these independent findings reaffirm the critical role of epigenetic pathways engaged in the pathogenesis of *SPINK1+* subtype.

Interestingly, increased methylated regions in the *ETS*-fusion negative patients have been attributed to hypermethylation of miR-26a, a post-transcriptional regulator of EZH2 (14). Thus, given the central role played by EZH2 and the epigenetic mechanism involved in *ETS*-fusion negative cases, our findings rationalize the role of EZH2-mediated epigenetic regulation of miR-338-5p and miR-421 in *SPINK1+*/*ETS*-negative subtype. In consonance with this, overexpression of miR-338-5p/-421 also results in decreased Tet1 expression. Converging lines of evidences suggest dual role of Tet1 in promoting transcription of pluripotency factors as well as recruitment of PRC2 on the CpG rich promoters (56). Taken together, miR-338-5p/-

421 mediated decrease in Tet1 expression might possibily contribute in reduced stemness and drug-resistance. We also conjecture that decrease in Tet1 expression may result in reduced PRC2 occupancy on the miRNA promoters, diminish epigenetic silencing marks, and consequently downregulate their targets including *SPINK1*.

Currently, there is no effective therapeutic intervention for *SPINK1*+/*ETS*-negative PCa as well as for other SPINK1+ malignancies, although use of monoclonal EGFR antibody has been suggested (57). Nevertheless, outcome of the phase I/II clinical trials using cetuximab (58) and small molecules inhibitors for EGFR has been largely unsuccessful (59,60). For instance, in a phase Ib/IIa clinical trial using cetuximab and doxorubicin combination therapy, only a fraction of CRPC patients (~8%) showed >50% PSA decline (58), revealing its limited effectiveness. Owing to the pleiotropic anti-cancer effects exhibited by miRNA-338-5p/-421, we propose microRNA-replacement therapy as one of the potential therapeutic approaches for SPINK1+ cancers; nonetheless *in-vivo* delivery methods and stability are some of the major challenges for successful translation into the clinic (61). While not restricted to this, the present study also suggests alternative avenues for the treatment of *SPINK1*+ malignancies, for instance inhibitors against DNMTs, HDACs or EZH2, several of which are already in clinical trials (62,63), or selective inhibition of *MALAT1* by using antisense oligonucleotides. Conclusively, we moved the field forward by addressing an important question that how SPINK1 is aberrantly overexpressed in *ETS*-fusion negative PCa, and stratification of patients based on *SPINK1*-positive and miRNA-338-5p/-421-low criteria could improve therapeutic modalities and overall management strategies.

METHODS

Animals

For mice xenograft studies, we used five to six weeks old NOD.CB17-Prkdc^{scid}/J (NOD/SCID) male mice (Jackson Laboratory) randomized into three groups (N=8 for each experimental condition) before implanting the cells. Mice were anesthetized using a cocktail of ketamine/xylazine (50 and 5 mg/kg respectively, via intraperitoneal route) and were subcutaneously implanted with 22RV1-CTL, 22RV1-miR-338-5p or 22RV1-miR-421 cells (2×10^6) suspended in 100µl of saline with 20% Matrigel into the dorsal both flank sides of the mice. A blinded assessment of tumor growth was conducted twice a week using digital Vernier's calipers, and tumor volumes were calculated using the formula ($\pi/6$) ($L \times W^2$), (L=length; W=width). Spontaneous metastasis to lungs and bone marrow of the xenografted mice was analyzed by performing qPCR using primers specific for human specific *Alu*-sequences as mentioned in the Supplementary Table S7. All procedures involving mice were approved by the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA) and conform to all regulatory standards of the Institutional Animal Ethics Committee of the Indian Institute of Technology, Kanpur.

Human Prostate Cancer Specimens

All prostate cancer (PCa) specimens used in this study were procured from King George's Medical University, Lucknow, India. Clinical specimens were collected after obtaining written informed consent from the patients and Institutional Review Board approvals from the King George's Medical University, Lucknow and Indian Institute of Technology, Kanpur, India. A total of 20 PCa specimens were selected for this study based on the *SPINK1* and *TMPRSS2-ERG* status, confirmed by qPCR, immunohistochemistry and Fluorescent in situ hybridization (FISH) for gene rearrangement (64). The PCa specimens used in this study were collected from men who underwent needle core biopsies and transurethral resection of the prostate (TURP) to relieve obstructive symptoms from locally advanced disease between year 2014 and 2016. None of the patients received preoperative radiation or androgen deprivation therapy. All

patients included in this study were of Indian descent residing in the northern part of India and were de-identified. Prostate cancer tissue microarrays (TMA) specimens (n=238) were obtained from Dept. of Pathology, Henry Ford Health System, Detroit, Michigan, USA, after getting written informed consent from the patients and approval from Institutional Review Board. TMAs were stained for SPINK1 and EZH2 by performing immunohistochemistry (IHC) and RNA *in situ* hybridization (RNA-ISH) respectively.

Cancer Cell Lines and authentication

Prostate cancer cell lines (22RV1, VCaP and PC3), Colorectal (WiDr), Pancreatic (CAPAN-1), Melanoma (SK-MEL-173), prostate epithelial cells (RWPE-1) and human embryonic kidney 293T cells (HEK293T) were procured from the American Type Culture Collection (ATCC) and were maintained using ATCC recommended medium supplemented with 10% fetal bovine serum and Gibco Penicillin-Streptomycin (Thermo-Fisher). Cell lines were cultured in CO_2 incubator (Thermo-Fisher) supplied with 5% CO_2 at 37°C temperature.

To ensure the identity, short tandem repeat (STR) profiling of all cell lines were performed at the Lifecode Technologies Private Limited, Bangalore and DNA Forensics Laboratory, New Delhi. The profiles were compared with reference STR genotypes available at ATCC, DSMZ-German Collection of Microorganisms and Cell Cultures, and Biosample databases to authenticate the identity and check for any cross contamination. All cell lines were routinely tested for *Mycoplasma* contamination using PlasmoTest mycoplasma detection kit (InvivoGen).

Transfection of microRNA mimetics, antagomiRs and Small interfering RNA

Synthetic mimics and antagomiRs for the miR-338-5p and miR-421, and negative controls (Exiqon) were transfected using Lipofectamine RNAiMAX (Invitrogen) with a final concentration of 30pmol. The cells were seeded at 40% confluency and were transfected with

respective miRNA mimics next day, followed by a second transfection after 24 hours. Subsequently, cells were processed for quantitative analysis and functional assays. Same transfection protocol was followed for On-Targetplus small interfering RNA (siRNA) for *SPINK1* (J-019724-07, GE Dharmacon) and *E2F1* (J-003259, GE Dharmacon).

Real-Time Quantitative PCR

Briefly, total RNA was extracted using miRNeasy Mini Kit (Qiagen) for miRNA related experiments or else TRIzol (Ambion), and 1μg of RNA with good integrity was reverse transcribed into cDNA using SuperScript III (Invitrogen) in the presence of random primers (Invitrogen). For Real Time Quantitative PCR (qPCR) all reactions were performed in triplicates using SYBR Green Master Mix (Applied Biosystems). The relative expression of the target gene was calculated for each sample by using the ΔΔCt method as described before (12,21). Sequences for all the primer sets used in this study are listed in the Supplementary Table S7.

TaqMan microRNA Assay

Total RNA including miRNA fraction was isolated using miReasy RNA extraction kit (Qiagen). QPCR for the miRNA stem-loop was performed using target-specific stem-loop reverse transcription primers using TaqMan microRNA reverse transcription kit (Thermo Fisher), followed by Taqman assays (Applied Biosystems) following manufacturer's instructions on the Step OnePlus Real Time PCR System (Applied Biosystems). Relative expression of the target miR-338-5p, miR-421, and miR-876-5p (Applied Biosystems Assays IDs: 4427975, 4427975, 4427975 respectively) was normalized to RNUB6 (Assay ID: 4427975).

MicroRNA 3'UTR *SPINK1* and *MALAT1* luciferase reporter assay

A Firefly/Renilla Dual-Luciferase reporter vector pEZX-MT01 (GeneCopoeia) was used for cloning full length *SPINK1* 3'UTR wild type, and mutant with altered residues in the binding sites of miR-338-5p and miR-421. Similarly, *MALAT1* wild type (250 bp) harboring miR-338-5p and miR-421 binding sites was also cloned into the same vector. Cells were seeded in a 24-well plate at 30-40% confluency, and co-transfected with 30pmol of miRNA mimics along with 25ng of pEZX-MT01 constructs using lipofectamine RNAiMax (Invitrogen). Luciferase assay was performed using Dual-Glo luciferase assay (Promega) 24 hours after the second transfection. Firefly luciferase activity was normalized to Renilla luciferase activity for each individual sample.

Cell proliferation, invasion and migration assays

For cell proliferation assay, cells were seeded in 12-wells culture plates (10,000 cells/well). At the indicated time points cells were trypsinized and counted on the Z-Series Coulter counter (Beckman Coulter). Cell invasion assays were performed using Transwell Boyden chambers of 8µm pore size (Corning) (21,65). Briefly, RPMI-1640 media supplemented with 20% FBS was added to the lower compartment, and 100,000 cells in serum-free media were added onto Transwell insert coated with Matrigel (BD Biosciences). After 24 hours incubation at 37°C with 5% CO_2, the non-invading cells and Matrigel were gently removed with a cotton swab from the Transwell inserts. Invasive cells located on the lower side of the inserts were fixed in formaldehyde (4% in PBS) and stained with crystal violet (0.5% w/v). Images of the representative field were taken on the Axio Observer Z1 microscope (Zeiss). The invaded cells were quantified by de-staining with 10% (v/v) acetic acid in distilled H_2O, and absorbance of de-staining solution was measured at 550nm. Same protocol was followed for cell migration assay, except no Matrigel was coated on the inserts.

Foci formation assay

For foci formation assay, cells (2×10^3) were plated in six-well culture dishes in cell line specific recommended culture media supplemented with 5% heat-inactivated fetal bovine serum (Invitrogen) and incubated at 37°C, media was changed every third day. The assay was terminated after 3 weeks and cells were fixed in formaldehyde (4% in PBS) and stained with crystal violet solution (0.05% w/v). Representative images were taken on the Axio Observer Z1 microscope (Carl Zeiss).

Soft agar colony assay

For anchorage-independent growth assay, soft agar plates were prepared by pouring 2ml of 0.6% low melting-point agarose (Sigma) dissolved in RPMI-1640 medium in 6-well dishes, after polymerization, second layer containing 2ml of 0.3% agar in RPMI-1640 medium supplemented with 10% FBS, and stable 22RV1-CTL, 22RV1-miR-338-5p and 22RV1-miR-421 cells (~1.5×10^4) resuspended were poured on the top of the first layer. Soft agar assay plates were incubated at 37°C for 20 days, and colonies greater than 40μm in size were counted.

Chick Chorioallantoic Membrane (CAM) assay

The chick embryo CAM assay was performed as explained previously (66). Briefly, fertilized eggs were incubated in a humidified incubator at 38°C for 10 days. The CAM was released by applying low pressure to the hole over the air sac and the shell was cut to make a square $1cm^2$ windows. Two million cells (22RV1-miR-338-5p, 22RV1-miR-421 or 22RV1-CTL) were implanted near the allantoic vein onto the CAM in 10 days post-fertilized eggs. The windows were subsequently sealed and the eggs were incubated at 38°C. For intravasation experiments, genomic DNA from lower CAM was isolated using Phenol/chloroform method and presence of the tumor cells was quantified by performing quantitative human *Alu*-specific PCR. The upper CAMs were isolated, fixed and immunostained for human-specific cytokeratin-18 as previously described (67). To assess tumor growth and metastasis, the assay was terminated on 18^{th} day post-implantation and extra-embryonic tumor mass were excised and weighed. For

metastases, embryonic lungs and liver was harvested, genomic DNA was isolated and subjected to quantitative human *Alu*-specific PCR as mentioned previously (66). Briefly, the standard curve for the *Alu*-specific PCR was prepared using different dilutions of genomic DNA (stock concentration 60ng/µl) isolated from 22RV1 cells along with 500ng of chicken DNA spiked in all the standards. About ~5ng of DNA template was used from each of these standards for the *Alu*-specific PCR. Further, the presence of *Alu* repeats in the human genomic DNA (cancer cells metastasized) isolated from embryonic lungs and liver was evaluated by performing *Alu* specific PCR using primers as listed in Supplementary Table S7.

H&E and immunostaining of tumor xenografts

Tumor tissues excised from the xenografted mice were fixed in 10% buffered formalin overnight, followed by dehydration using increasing concentration of ethanol. Subsequently, tumor specimens were embedded in paraffin and serially sectioned at 3µm thickness using microtome (Leica) as described earlier (21). Briefly, tissue sections were deparaffinized and dehydrated/rehydrated using standard protocol, followed by antigen-retrieval in the citrate buffer (pH 6.0) for 10 minutes at 100°C. Endogenous peroxidase activity was quenched using 3% hydrogen peroxide for 5 minutes. Sections were then blocked with 10% goat serum and probed with anti-mouse Ki-67 (1:400, CST, 9449S) at 4°C overnight, followed by secondary horseradish peroxidase (HRP)-conjugated antibody (DAKO), and HRP activity was detected using DAB (3, 3 -diaminobenzidine) peroxidase (HRP) substrate kit (DAKO). Quantification of IHC staining was performed in a blindfolded manner. The numbers of cells positive for Ki-67 staining were manually counted from ten random histological sections for each mouse.

Gene expression array analysis

For global gene expression profiling, total RNA was isolated from stable 22RV1-miR-338-5p, 22RV1-miR-421 and 22RV1-CTL cells as described earlier and subjected to Agilent Whole

Human Genome Oligo Microarray profiling (dual color) using Agilent platform (8x60K format) according to the manufacturer's protocol. A total of three microarray hybridizations were performed using each stable miRNA overexpressing cell line samples against control cells. Microarray data was normalized by following locally weighted linear regression (also known as Lowess) (68), and data was normalized using GeneSpringGX software for the raw data files. Differentially regulated genes were clustered using hierarchical clustering based on Pearson coefficient correlation algorithm to identify significant gene expression patterns. Further, for multiple hypotheses testing adjustments were applied using Benjamini and Hochberg procedure to calculate the FDR-corrected *P*-values (with FDR< 0.05) for the differentially expressed genes. Data was filtered to include only features with significant differential expression ($\log_2$ fold change greater than 0.6 or less than -0.6, $P< 0.05$) i.e. ~1.6-fold average over- or under-expressed genes, were then used for the enrichment of biological processes using DAVID bioinformatics platform. Further, enrichment of the biological pathways, and molecular signatures that were enriched upon miRNA overexpression with respect to control were analyzed using Gene set enrichment analysis (GSEA). A network based enrichment of critical miR-338-5p and miR-421 overlapping biological pathways was generated using Enrichment Map (69), a plug-in for Cytoscape network visualization software (http://baderlab.org/Software/EnrichmentMap/). The heatmap.2 function of R package 'gplots' was used to create the heat maps.

Western Blot analysis

Cell lysates were prepared in radioimmunoprecipitation assay (RIPA) lysis buffer, supplemented with complete protease (Roche) and phosphatase inhibitors mixture (Calbiochem). Protein samples were prepared in 1X SDS sample loading buffer; size fractionated on the SDS-PAGE and transferred onto a Polyvinylidene Difluoride membrane (PVDF) membrane (GE Healthcare). The PVDF membrane was then incubated for 1 hour at

room temperature in blocking buffer [Tris-buffered saline, 0.1% Tween (TBS-T), 5% non-fat dry milk], and were incubated overnight at 4^{o}C with the following primary antibodies: anti-phosphor -MEK or -ERK rabbit (1:1000, 9121S or 4377S) or total-MEK or –ERK (1:1000, 9126S or 4695S), anti-E2F1 rabbit (1:1000, 3742S), anti-TET1 rabbit (1:2000, ab121587) and anti-β-Actin rabbit (1:3000, 4970S). Subsequently, blots were incubated with horseradish peroxidase-conjugated secondary anti-mouse or anti-rabbit antibody (1:5000, Jackson ImmunoResearch Laboratories) for 2 hours at room temperature, and were washed with 1X TBS-T buffer, and the signals were visualized by enhanced chemiluminescence system (GE Healthcare) as described by the manufacturer.

Immunofluorescence analysis

Cells were grown on the glass coverslips and fixed in 4% para-formaldehyde, washed with 1X PBS, permeabilized with 0.3% Triton X-100 in PBS for 10 min and blocked with 5% goat serum in 0.1% Triton X-100 in 1X PBS for 2 hours at room temperature. Subsequently, cells were incubated with the following primary antibodies: SPINK1 (1:100, Abnova, H00006690-M01), E-cadherin (1:400, CST, 3195S), N-cadherin (1:400, Abcam ab98952), Slug (1:50, CST, 9585S), Snail (1:50, CST, 3895S), c-Kit or CD117 (1:400, CST, 3308S), SOX-9 (1:400, Merck millipore, AB5535, a kind gift from Dr. A. Bandyopadhyay, IITK), TET1 (1:500, Abcam, ab121587). Subsequently, cells were incubated with Alexa Fluor-488 conjugated secondary anti-mouse or anti-rabbit antibodies (1:600, CST, 4412 or 4408). The coverslips with the stained cells were mounted on the slides using Vectashield with DAPI (Vector laboratories). Images were captured on the Axio Observer Z1 microscope (Carl Zeiss) equipped with high-resolution CCD camera or at LSM780LNO Carl Zeiss Confocal microscope.

Prostatosphere Assay

Briefly, prostate cancer cells (10000 cells/ml) were cultured in suspension in low-adherence plate using serum-free DMEM-F12 (Invitrogen) supplemented with B27 (1:50, Invitrogen), 20

ng/ml EGF (Invitrogen), 20 ng/ml FGF (Invitrogen) and Penicillin-Streptomycin (Thermo-Fisher Scientific) as previously described (70). Small population of cells which formed prostatospheres were collected by gentle centrifugation and were mechanically dissociated into single cells suspension and then passaged for several generations (4-5 passages at an interval of 10-12 days) following similar culture conditions and were assessed for the sphere initiation/forming efficiency or self-renewal capacity. Spheres larger than 50µm in diameter were counted and plotted as percent sphere-forming efficiency. Representative images of the sphere were taken on phase contrast mode on the Axio Observer Z1 microscope (Carl Zeiss). Total RNA was isolated at day 6 and 12 as described before and was subjected to qPCR for detecting the expression of cancer stem cell like markers.

Flow cytometry

For BrdU-7AAD cell cycle analysis, 22RV1 cells were transfected with mimics for miR-338-5p, miR-421 or control for 2 consecutive days, followed by BrdU pulse labelling for 2 hours after second transfection. Subsequently, cells were stained with anti-BrdU antibody conjugated to fluorescein isothiocyanate (FITC) (BrdU-7AAD flow kit, BD Biosciences), and then with 7-Aminoactinomycin D (7-AAD), following manufacturer's instructions. Samples were subjected to FACS (BD FACS Calibur) excited at 488nm. Forward Scatter (FSC) and Side Scatter (SSC) parameters were adjusted to gate the population of interest. The 7-AAD signals were recorded on the linear scale while BrdU-FITC signals were recorded on the logarithmic scale.

For Hoechst side population assay, 22RV1-CTL, 22RV1-miR-338-5p or 22RV1-miR-421 cells were stained with Hoechst 33342 (5 µg/ml), in the presence or absence of an ABC efflux pump inhibitor, verapamil (used as a negative control) followed by incubating samples at 37^{o}C in water bath for 2 hours with periodic agitation. Later, cells were centrifuged at 2000 rpm for 5

min at 4°C and resuspended in cold 1X PBS and washed twice. Samples were kept at 4°C until analysis. Propidium iodide (PI) was added at a concentration of 5µg/ml to exclude dead cells. For detection of side population (SP), Hoechst blue and red signals were acquired using a 460/50 and 670/30 nm band-pass filters respectively. While 7-AAD was excited at 488 nm and its emission was measured in logarithmic scale through a band pass filter of 670/30. Since, Hoechst Red signals are comparatively lower than that of Blue, a relatively higher laser power was used, and an optimal resolution of the SP cells was found using 30-35mW of power with the UV laser. A dim tail of SP cells enriched faction was gated using a dot plot displaying Hoechst Blue and Hoechst Red scatter. A minimum of 100,000 live cell events were acquired to resolve the SP cell population in each sample.

For Annexin PI staining, 22RV1 cells were transfected with mimics for miR-338-5p, miR-421 or control miRNA. After 24 hours, cells were washed with cold 1X PBS and were resuspended in 1X Binding Buffer at a concentration of $1x10^6$ cells/ml. Subsequently, 5µl of FITC Annexin V and PI was added and incubated at room temperature for 15min. Subsequently, 400µl of 1X Binding Buffer was added to the samples and were analysed on BD FACSCalibur. Data acquisition was performed on BD FACSCalibur platform, and analysed using FlowJo version 10.7 (TreeStar).

Immunohistochemistry

TMA slides were incubated at 60°C for at least 2 hours. Slides were then placed in EnVision FLEX Target Retrieval Solution, High pH (Agilent DAKO, K800421-2) in a PT Link instrument (Agilent DAKO, PT200) at 75ºC, heated to 97ºC for 20 minutes, and then cooled to 75ºC. Slides were then washed in 1X EnVision FLEX Wash Buffer (Agilent DAKO, K800721-2) for 5 minutes. Slides were then treated with Peroxidazed 1 (Biocare Medical,

PX968M) for 5 minutes and Background Punisher (Biocare Medical, BP974L) for 10 minutes with a wash of 1X EnVision FLEX Wash Buffer for 5 minutes after each step. Mouse monoclonal SPINK1 (Novus Biologicals, H00006690-M01) diluted 1:100 in EnVision FLEX Antibody Diluent (Agilent DAKO, K800621-2) was added to each slide, which were then cover slipped with parafilm, placed in a humidifying chamber, and incubated overnight at 4ºC. The next day, slides were washed in 1X EnVision Wash Buffer for 5 minutes and then incubated in Mach2 Doublestain 1 (Biocare Medical, MRCT523L) for 30 minutes at room temperature in a humidifying chamber. Slides were then rinsed in 1X EnVision Wash Buffer 3 times for 5 minutes each. Slides were then treated with a Ferangi Blue solution (1 drop to 2.5ml buffer; Biocare Medical, FB813S) for 7 minutes. Slides were rinsed 2 times in distilled water, then treated with EnVision FLEX Hematoxylin (Agilent DAKO, K800821-2) for 5 minutes. Slides were rinsed several times in distilled water, immersed in a 0.01% ammonium hydroxide solution, and then rinsed twice in distilled water. Slides were then dried completely. Slides were dipped in xylene approximately 15 times. EcoMount (Biocare Medical, EM897L) was added to each slide, which was then cover slipped.

RNA *in situ* hybridization

TMA slides were incubated at 60°C for 1 hour. Tissues were then de-paraffinized by immersing in xylene twice for 5 minutes each with periodic agitation. The slides were then immersed in 100% ethanol twice for 3 minutes each with periodic agitation, then air-dried for 5 minutes. Tissues were circled using a pap pen (Vector, H-4000), allowed to dry, and treated with H_2O_2 for 10 minutes. Slides were rinsed twice in distilled water, and then boiled in 1X Target Retrieval for 15 minutes. Slides were rinsed twice in distilled water, and then treated with Protease Plus for 15 minutes at 40°C in a HybEZ Oven (Advanced Cell Diagnostics, 310010). H_2O_2, 1X Target Retrieval, and Protease Plus are included in the RNAscope pre-treatment kit (Advanced Cell Diagnostics, 310020). Slides were rinsed twice in distilled water, and then

treated with EZH2 probe (Advanced Cell Diagnostics, probe ID: 405491) for 2 hours at 40°C in the HybEZ Oven. Slides were then washed in 1X Wash Buffer (Advanced Cell Diagnostics, 310091) twice for 2 minutes each. Slides were then treated with Amp 1 for 30 minutes, Amp 2 for 15 minutes, Amp 3 for 30 minutes, and Amp 4 for 15 minutes, all at 40ºC in the HybEZ oven with 2 washes in 1X Wash Buffer for 2 minutes each after each step. Slides were then treated with Amp 5 for 30 minutes and Amp 6 for 15 minutes at room temperature in a humidity chamber with 2 washes in 1X Wash Buffer for 2 minutes each after each step. Red color was developed by adding a 1:60 solution of Fast Red B: Fast Red A to each slide and incubating for 10 minutes. Slides were washed twice in distilled water. Amps 1-6 and Fast Red are included in the RNAscope 2.5 HD Detection Reagents-RED (Advanced Cell Diagnostics, 322360). Slides were then treated with EnVision FLEX Hematoxylin (Agilent DAKO, K800821-2) for 5 minutes. Slides were rinsed several times in distilled water, immersed in a 0.01% ammonium hydroxide solution, and then rinsed twice in distilled water. Slides were then dried completely. Slides were dipped in xylene approximately 15 times. EcoMount (Biocare Medical, EM897L) was added to each slide, which was then cover slipped.

EZH2 and SPINK1 staining Evaluation Criteria

EZH2 expression intensity scoring by RNA-ISH for all the tumor foci was evaluated on the basis of the number of red dots/cell and were graded into five levels ranging from score of 0 to 4 as described previously (71). SPINK1 staining by IHC was used to evaluate SPINK1 positive and negative status of the PCa specimens. Further, an association between SPINK1 and EZH2 expression in patients' samples was calculated by applying Chi-Squared contingency test (72) on GraphPad Prism.

Chromatin immunoprecipitation

Briefly, cancer cells (~80-90% confluency) were crosslinked with a final concentration of 1% formaldehyde for 10 minutes, followed by quenching with Glycine (125mM) for 10 minutes at room temperature, followed by washing with 1X PBS twice. Next, cell lysis was performed using lysis buffer [1% SDS, 10mM EDTA, 50mM Tris-Cl and protease inhibitor (Roche)] followed by sonication using Bioruptor (Diagenode) to obtain an average length of ~500bp DNA fragments. Chromatin immunoprecipitation (ChIP) assays were carried out using antibodies against E2F1 (CST, 3742), EZH2/KMT6 (Abcam, ab191250), H3K27me3 (CST, 9733) and control rabbit IgG (Invitrogen). Supernatant containing sheared chromatin were incubated at 4°C overnight with 4µg of E2F1 or EZH2 or H3K27me3 and IgG antibodies. Concurrently, the Protein G coated Dynabeads (Invitrogen) were blocked with 100µg/ml BSA (HiMedia) and 500µg/ml sheared salmon sperm DNA (Sigma) and incubated at 4°C overnight. Blocked beads were washed twice with 9:1 dilution buffer: lysis buffer [1% Triton X-100; 150mM NaCl; 2mM EDTA (pH 8.0); 20mM Tris-HCl (pH 8.0) with protease inhibitors] and were incubated with respective antibodies to form antibody-bead conjugates. The antibody-bead conjugates were then washed three times in a low salt wash buffer 1 [1% Triton X-100; 0.1% SDS; 150mM NaCl; 2mM EDTA (pH 8.0); 20mM Tris-HCl (pH 8.0) with protease inhibitors] and once in high salt wash buffer 2 (same as wash buffer 1, except 500mM NaCl). The antibody/protein/DNA complexes were eluted using elution buffer [100mM $NaHCO_3$, 1% SDS, RNaseA and Proteinase K (500µg/ml each)]. DNA was isolated using phenol-chloroform-isoamyl alcohol extraction method, precipitated and washed with 70% ethanol, air-dried, and dissolved in nuclease free water (Ambion). QPCR was performed using appropriate primer sets as listed in Supplementary Table S7.

Methylated DNA Immunoprecipitation (MeDIP)

Genomic DNA was extracted from 22RV1 cells using QIAamp DNA Mini Kit and was sonicated to produce random fragments ranging from 300-1000 bp. About 4µg of fragmented

DNA was used for MeDIP assay. The DNA was denatured for 10 min at 95°C and immunoprecipitated with 4µg of monoclonal antibody against 5-mC (Abcam ab10805) or 5-hmC (Abcam ab106918) and IgG (Santa Cruz, sc-2027) in a final volume of 500µl IP buffer (10mM sodium phosphate (pH 7.0), 140 mM NaCl, 0.05% Triton X- 100) for 2 hours at 4°C. Dynabeads (40µl) were washed twice with 800 µl PBS-BSA (0.1%) for 5 minutes at room temperature, and then were resuspended in 40µl of 1X IP buffer. The resuspended Dynabeads were added to the samples and incubated for 4-5 hours at 4°C with end-over-end shaking using rotator stirrer. Beads were then collected and washed thrice with 700µl of 1X IP buffer. The beads were treated with proteinase K (500µg/ml) for 3 hours at 50°C, subsequently immunoprecipitated DNA was recovered by phenol-chloroform extraction followed by ethanol precipitation. Real-time PCR reactions were carried out with 40 ng of input DNA and 2µl of the immunoprecipitated DNA following manufacturer's instructions on the Step OnePlus Real Time PCR System (Applied Biosystems). All reactions were performed in triplicates and the relative fold enrichment of 5-mC over 5-hmC was plotted.

RNA Immunoprecipitation (RIP) Assay

About ~80% confluent 22RV1 cells were harvested for RIP assay. RIP was performed using EZH2 antibody (Abcam, ab191250). Briefly, the cells were washed twice with ice-cold 1X PBS and scraped in 1X PBS with protease inhibitor (Invitrogen). Next, cells were incubated in RIP buffer [50 mM Tris-HCl (pH 7.9), 0.25 M NaCl, 1% Nonidet P-40 (NP-40), 10 mM EDTA, protease inhibitor cocktail and RNase inhibitor] for 30 min at 4°C. Further, cell lysate was obtained by centrifugation at 12,000 rpm for 10 min at 4°C and was incubated overnight with 4µg of EZH2 antibody or IgG control (Invitrogen). Simultaneously, Protein G coated Dynabeads (Invitrogen) were pre-absorbed with 100µg/ml BSA at 4°C. The pre-absorbed beads were washed thrice with NT2 buffer [50 mM Tris-HCl (pH 7.4), 300 mM NaCl, 1 mM $MgCl_2$, 0.05% Nonidet P-40 (NP40), 1XPIC, RNase inhibitor] and then incubated with RNA-

antibody complex to form RNA-antibody-bead precipitates. Further, the beads were washed thrice with NT2 buffer, followed DNase I digestion for 15 min at 37°C and washed twice with NT2 buffer. Co-purified RNA was extracted by Trizol (Invitrogen), followed by cDNA synthesis using the SuperScript kit (Invitrogen). *MALAT1* and *GAPDH* expression was analysed by qPCR as described before.

Bisulfite sequencing

Bisulfite conversion of the genomic DNA was carried out using EpiTect Bisulfite Kit (Qiagen) following manufacturer's instructions. Briefly, bisulfite converted DNA was used as template for PCR amplification using primers (Macrogen Inc., South Korea) designed using the Methprimer software as listed in Supplementary Table S7. The amplified PCR product was purified using QIAquick PCR purification kit (Qiagen), cloned into pGEM-T Easy Vector (Promega) and transformed into One Shot TOP10 competent cells (Invitrogen). Plasmid DNA was isolated from eight independent colonies and was outsourced for conventional Sanger sequencing at Macrogen Inc., South Korea. The BiQ Analyzer online tool was used to calculate the methylation percentage and to generate the graphical plots.

Data Mining and Computational Analyses:

MicroRNA target Prediction by Multiple Programs

MicroRNA prediction programs, namely miRanda, miRMap, PITA and RNA Hybrid were used to predict miRNAs targeting 3'UTR of *SPINK1* (Fig. 1A and Supplementary Table S1). The correlation between the expression of the predicted miRNAs and SPINK1 was analysed by employing RNA Sequencing data for the TCGA-PRAD cohort. For Fig. 1A (lower panel), 4C and Supplementary Table S6 miRanda was used to predict the putative binding sites of the miR-338-5p and miR-421 on the 3'UTR of target genes.

Integrative analyses for TCGA-PRAD data

For gene association studies between miRs-338-5p, -421, miR-876-5p, *SPINK1*, *ERG*, *EZH2* and *MALAT1* Illumina HiSeq mRNA and miRNA-Seq data along with clinical information from TCGA-PRAD dataset was downloaded. Overexpression of SPINK1 in PCa exhibits outlier-expression in ~10-15% of the total PCa cases (Tomlins et al., 2008). Thus, to stratify patients with increased expression of *SPINK1*, we sorted TCGA patients' samples on the basis of increasing *SPINK1* expression (descending order), and divided the dataset into four equal parts by employing Quartile-based normalization method (73), the top 25% of the patients (N=119) corresponding to the upper quartile (QU, $\log_2$ (RPM+1)>5.468 or $\log_2$ (normalized count+1)>1.892), were assigned as *SPINK1* high or *SPINK1*-positive patient samples and the lower quartile (QL, $\log_2$ (RPM+1)<1.124 or $\log_2$ (normalized count+1)<-2.611), were considered as *SPINK1* low or *SPINK1*-negative samples. Also, we found about 18 patients with outlier expression of *SPINK1* with $\log_2$ (RPM+1) of greater than 11.984, which were included in the heat map representation of *SPINK1* positive TCGA patients in Figure 1B. No further cut-offs were applied for miR-338-5p, miR-421, miR-876-5p and *ERG* expression, corresponding expression values (based on *SPINK1* cut-off) were considered for these genes for further analysis. Hierarchical Clustering of miRs-338-5p, -421 or 876-5p, *SPINK1*, and *ERG* were employed using heatmap.2 of R's gplot package, which uses Euclidean distance to obtain distance matrix and complete agglomeration method for clustering between each genes and miRNA.

For Kaplan-Meier survival analysis, the survival data included sample type (primary tumors), days to first biochemical recurrence and days to last follow-up for TCGA-PRAD patients was considered. The samples were divided into two groups, higher and lower miRNA expression groups according to the expression level of a miR-338-5p and miR-421 using Cox proportional hazards regression model in R. Next, we performed a 13-year survival analysis of these miRNAs using Kaplan-Meier survival analysis (74) by employing survival package

(https://cran.r-project.org/web/packages/survival) in the R environment, and statistical significance was computed using the log-rank test. For clinical relevance of miR-338-5p and miR-421, TCGA-PRAD dataset was analyzed for the association of these miRNAs with clinical parameters such as primary Gleason score, Clinical T score and positive lymph node status. Data analysis was performed by one-way analysis of variance with Tukey's post hoc test for multiple comparisons, and student's *t*-test was applied for comparison between two groups.

The MSKCC cohort (Cancer Cell, 2010) data was retrieved from cBioPortal (http://www.cbioportal.org/) for *SPINK1* and *EZH2* expression in the prostate cancer patients (N=85), and oncoprints were generated using default parameters (mRNA expression z-score threshold ±2 vs normal). Further, to ascertain possible association between *EZH2*, *SPINK1*, *MALAT1* and miR-338-5p/-421, TCGA patients' samples were stratified on the basis of increasing *EZH2* expression, and divided the dataset into four equal quartiles, the top 25% of the patients (N=119) corresponding to the upper quartile (QU, $\log_2$ (normalized count+1) >7.313), were considered as *EZH2* high patient samples and the lower quartile $\log_2$ (normalized count+1) <6.36), were considered as *EZH2* low samples. The corresponding expression values for *SPINK1*, *MALAT1*, miR-338-5p and miR-421 in EZH2 high and EZH2 low groups (without further cut-offs) were considered to association with EZH2 expression (related to Fig. 6B).

Statistical analysis

Statistical significance was determined by either two-tailed Student's *t* test for independent samples or one-way Analysis of Variance (ANOVA), otherwise specified. The differences between the experimental groups were considered significant if the *P*-value of less than 0.05 was obtained. Error bars represent mean ± SEM. All experiments were repeated three times in triplicates.

Data availability

The gene expression microarray data from this study has been submitted to the NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE108558.

Acknowledgements

B.A. is an Intermediate Fellow of the Wellcome Trust/DBT India Alliance. This work is supported by the Wellcome Trust/DBT India Alliance Fellowship [grant number: IA/I(S)/12/2/500635] awarded to B.A. We thank Yuping Zhang, Brendan Veeneman, Mahendra Palecha, Ayush Praveen for their technical support and Anjali Bajpai for critically reading the manuscript. We also thank Jonaki Sen for extending the use of fertilized eggs facility. The IIT Kanpur has filed a patent (IN 201611016564) on the therapeutic applicability of miR-338-5p and miR-421 described in this study in which B.A., V.B. and A.Y. are named as inventors.

Authors' contributions

V.B., A.Y., and B.A. designed and directed the experimental studies. V.B. and A.Y. performed *in vitro* cell line-based studies. V.B., A.Y. and S.N. performed the bisulfite sequencing experiments and analysis. V.B., A.Y., R.T. and S.G. performed the gene expression studies, bioinformatics analysis and ChIP assays. A.Y. performed the immunofluorescence and RIP experiments. V.B., A.Y., and B.A. performed statistical analysis and interpreted the data. V.B., A.Y., and B.A. executed the *in vivo* experiments. A.G. provided PCa patient specimens. S.C.,

N.G., and N.P. performed immnuohistochemistry and RNA *in situ* staining on the PCa tissue microarrays. V.B., A.Y., and B.A. wrote the manuscript. B.A. directed the overall project.

References:

1. Rubin MA, Maher CA, Chinnaiyan AM. Common gene rearrangements in prostate cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2011**;29(27):3659-68 doi 10.1200/JCO.2011.35.1916.
2. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet **2012**;44(6):685-9 doi ng.2279 [pii]10.1038/ng.2279.
3. Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, *et al.* Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med **2010**;16(7):793-8 doi nm.2166 [pii]10.1038/nm.2166.
4. The Molecular Taxonomy of Primary Prostate Cancer. Cell **2015**;163(4):1011-25 doi S0092-8674(15)01339-2 [pii]10.1016/j.cell.2015.10.025.
5. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science **2005**;310(5748):644-8 doi 310/5748/644 [pii]10.1126/science.1117679.
6. Wu Y-M, Su F, Kalyana-Sundaram S, Khazanov N, Ateeq B, Cao X, *et al.* Identification of targetable FGFR gene fusions in diverse cancers. Cancer discovery **2013**.
7. Ateeq B, Kunju LP, Carskadon SL, Pandey SK, Singh G, Pradeep I, *et al.* Molecular profiling of ETS and non-ETS aberrations in prostate cancer patients from northern India. Prostate **2015**;75(10):1051-62 doi 10.1002/pros.22989.
8. Tomlins SA, Laxman B, Varambally S, Cao X, Yu J, Helgeson BE, *et al.* Role of the TMPRSS2-ERG gene fusion in prostate cancer. Neoplasia **2008**;10(2):177-88.
9. Brenner JC, Ateeq B, Li Y, Yocum AK, Cao Q, Asangani IA, *et al.* Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. Cancer cell **2011**;19(5):664-78 doi S1535-6108(11)00156-5 [pii]10.1016/j.ccr.2011.04.010.
10. Tomlins SA, Rhodes DR, Yu J, Varambally S, Mehra R, Perner S, *et al.* The role of SPINK1 in ETS rearrangement-negative prostate cancers. Cancer Cell **2008**;13(6):519-28 doi S1535-6108(08)00152-9 [pii]10.1016/j.ccr.2008.04.016.
11. Leinonen KA, Tolonen TT, Bracken H, Stenman UH, Tammela TL, Saramaki OR, *et al.* Association of SPINK1 expression and TMPRSS2:ERG fusion with prognosis in endocrine-treated prostate cancer. Clin Cancer Res **2010**;16(10):2845-51 doi 1078-0432.CCR-09-2505 [pii]10.1158/1078-0432.CCR-09-2505.
12. Ateeq B, Tomlins SA, Laxman B, Asangani IA, Cao Q, Cao X, *et al.* Therapeutic targeting of SPINK1-positive prostate cancer. Sci Transl Med **2011**;3(72):72ra17 doi 3/72/72ra17 [pii]10.1126/scitranslmed.3001498.
13. Varambally S, Cao Q, Mani RS, Shankar S, Wang X, Ateeq B, *et al.* Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science **2008**;322(5908):1695-9 doi 1165395 [pii]10.1126/science.1165395.
14. Borno ST, Fischer A, Kerick M, Falth M, Laible M, Brase JC, *et al.* Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. Cancer Discov **2012**;2(11):1024-35 doi 10.1158/2159-8290.CD-12-0041.

15. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science **2002**;298(5595):1039-43.
16. Xu K, Wu ZJ, Groner AC, He HH, Cai C, Lis RT, *et al.* EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. Science **2012**;338(6113):1465-9 doi 10.1126/science.1227604.
17. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer **2006**;6(4):259-69 doi 10.1038/nrc1840.
18. Wang D, Ding L, Wang L, Zhao Y, Sun Z, Karnes RJ, *et al.* LncRNA MALAT1 enhances oncogenic activities of EZH2 in castration-resistant prostate cancer. Oncotarget **2015**;6(38):41045-55 doi 10.18632/oncotarget.5728.
19. Cao Q, Mani RS, Ateeq B, Dhanasekaran SM, Asangani IA, Prensner JR, *et al.* Coordinated regulation of polycomb group complexes through microRNAs in cancer. Cancer Cell **2011**;20(2):187-99 doi 10.1016/j.ccr.2011.06.016.
20. Rasanen K, Itkonen O, Koistinen H, Stenman UH. Emerging Roles of SPINK1 in Cancer. Clinical chemistry **2016**;62(3):449-57 doi 10.1373/clinchem.2015.241513.
21. Tiwari R, Pandey SK, Goel S, Bhatia V, Shukla S, Jing X, *et al.* SPINK1 promotes colorectal cancer progression by downregulating Metallothioneins expression. Oncogenesis **2015**;4:e162 doi oncsis201523 [pii]10.1038/oncsis.2015.23.
22. Qin XQ, Livingston DM, Kaelin WG, Jr., Adams PD. Deregulated transcription factor E2F-1 expression leads to S-phase entry and p53-mediated apoptosis. Proc Natl Acad Sci U S A **1994**;91(23):10918-22.
23. Branzei D, Foiani M. Regulation of DNA repair throughout the cell cycle. Nature reviews Molecular cell biology **2008**;9(4):297-308 doi 10.1038/nrm2351.
24. Gutschner T, Hammerle M, Eissmann M, Hsu J, Kim Y, Hung G, *et al.* The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. Cancer Res **2013**;73(3):1180-9 doi 10.1158/0008-5472.CAN-12-2850.
25. Tripathi V, Shen Z, Chakraborty A, Giri S, Freier SM, Wu X, *et al.* Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. PLoS Genet **2013**;9(3):e1003368.
26. Li S, Wang Q, Qiang Q, Shan H, Shi M, Chen B, *et al.* Sp1-mediated transcriptional regulation of MALAT1 plays a critical role in tumor. Journal of cancer research and clinical oncology **2015**;141(11):1909-20.
27. Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell **2008**;133(4):704-15.
28. Puisieux A, Brabletz T, Caramel J. Oncogenic roles of EMT-inducing transcription factors. Nature cell biology **2014**;16(6):488.
29. Moreno-Bueno G, Cubillo E, Sarrió D, Peinado H, Rodríguez-Pinilla SM, Villa S, *et al.* Genetic profiling of epithelial cells expressing E-cadherin repressors reveals a distinct role for Snail, Slug, and E47 factors in epithelial-mesenchymal transition. Cancer research **2006**;66(19):9543-56.
30. Liu T, Xu F, Du X, Lai D, Liu T, Zhao Y, *et al.* Establishment and characterization of multi-drug resistant, prostate carcinoma-initiating stem-like cells from human prostate cancer cell lines 22RV1. Molecular and cellular biochemistry **2010**;340(1-2):265-73.
31. Zhou S, Schuetz JD, Bunting KD, Colapietro A-M, Sampath J, Morris JJ, *et al.* The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. Nature medicine **2001**;7(9):1028-34.

32. Gao Y, Chen J, Li K, Wu T, Huang B, Liu W, *et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell stem cell **2013**;12(4):453-69.
33. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, *et al.* Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proceedings of the National Academy of Sciences of the United States of America **2002**;99(24):15524-9 doi 10.1073/pnas.242606799.
34. Bueno MJ, Perez de Castro I, Gomez de Cedron M, Santos J, Calin GA, Cigudosa JC, *et al.* Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression. Cancer Cell **2008**;13(6):496-506 doi 10.1016/j.ccr.2008.04.018.
35. Zhang X, Zhao X, Fiskus W, Lin J, Lwin T, Rao R, *et al.* Coordinated silencing of MYC-mediated miR-29 by HDAC3 and EZH2 as a therapeutic target of histone modification in aggressive B-Cell lymphomas. Cancer cell **2012**;22(4):506-23.
36. Kamminga LM, Bystrykh LV, de Boer A, Houwer S, Douma J, Weersing E, *et al.* The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. Blood **2006**;107(5):2170-9.
37. Lu C, Han HD, Mangala LS, Ali-Fehmi R, Newton CS, Ozbun L, *et al.* Regulation of tumor angiogenesis by EZH2. Cancer cell **2010**;18(2):185-97.
38. Nuytten M, Beke L, Van Eynde A, Ceulemans H, Beullens M, Van Hummelen P, *et al.* The transcriptional repressor NIPP1 is an essential player in EZH2-mediated gene silencing. Oncogene **2008**;27(10):1449-60.
39. Kos A, Olde Loohuis NF, Wieczorek ML, Glennon JC, Martens GJ, Kolk SM, *et al.* A potential regulatory role for intronic microRNA-338-3p for its host gene encoding apoptosis-associated tyrosine kinase. PLoS One **2012**;7(2):e31022 doi 10.1371/journal.pone.0031022.
40. Chureau C, Chantalat S, Romito A, Galvani A, Duret L, Avner P, *et al.* Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. Hum Mol Genet **2011**;20(4):705-18 doi 10.1093/hmg/ddq516.
41. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. Nature **2006**;439(7078):871-4.
42. Hu H, Du L, Nagabayashi G, Seeger RC, Gatti RA. ATM is down-regulated by N-Myc–regulated microRNA-421. Proceedings of the National Academy of Sciences **2010**;107(4):1506-11.
43. Ge X, Liu X, Lin F, Li P, Liu K, Geng R, *et al.* MicroRNA-421 regulated by HIF-1α promotes metastasis, inhibits apoptosis, and induces cisplatin resistance by targeting E-cadherin and caspase-3 in gastric cancer. Oncotarget **2016**;7(17):24466.
44. Meng D, Yang S, Wan X, Zhang Y, Huang W, Zhao P, *et al.* A transcriptional target of androgen receptor, miR-421 regulates proliferation and metabolism of prostate cancer cells. The international journal of biochemistry & cell biology **2016**;73:30-40.
45. Hydbring P, Wang Y, Fassl A, Li X, Matia V, Otto T, *et al.* Cell-Cycle-Targeting MicroRNAs as Therapeutic Tools against Refractory Cancers. Cancer cell **2017**;31(4):576-90 e8 doi 10.1016/j.ccell.2017.03.004.
46. Liu C, Kelnar K, Liu B, Chen X, Calhoun-Davis T, Li H, *et al.* The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. Nature medicine **2011**;17(2):211-5.
47. Le MT, Hamar P, Guo C, Basar E, Perdigão-Henriques R, Balaj L, *et al.* miR-200–containing extracellular vesicles promote breast cancer cell metastasis. The Journal of clinical investigation **2014**;124(12):5109-28.

48. Huang N, Wu Z, Lin L, Zhou M, Wang L, Ma H, *et al.* MiR-338-3p inhibits epithelial-mesenchymal transition in gastric cancer cells by targeting ZEB2 and MACC1/Met/Akt signaling. Oncotarget **2015**;6(17):15222.
49. Guo B, Liu L, Yao J, Ma R, Chang D, Li Z, *et al.* miR-338-3p suppresses gastric cancer progression through a PTEN-AKT axis by targeting P-REX2a. Molecular Cancer Research **2014**;12(3):313-21.
50. Bunting KD. ABC transporters as phenotypic markers and functional regulators of stem cells. Stem cells **2002**;20(1):11-20 doi 10.1634/stemcells.20-3-274.
51. Bolton EM, Tuzova AV, Walsh AL, Lynch T, Perry AS. Noncoding RNAs in prostate cancer: the long and the short of it. Clin Cancer Res **2014**;20(1):35-43 doi 10.1158/1078-0432.CCR-13-1989.
52. Zhang A, Zhao JC, Kim J, Fong KW, Yang YA, Chakravarti D, *et al.* LncRNA HOTAIR Enhances the Androgen-Receptor-Mediated Transcriptional Program and Drives Castration-Resistant Prostate Cancer. Cell Rep **2015**;13(1):209-21 doi 10.1016/j.celrep.2015.08.069.
53. Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, *et al.* lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature **2013**;500(7464):598-602 doi 10.1038/nature12451.
54. Prensner JR, Iyer MK, Sahu A, Asangani IA, Cao Q, Patel L, *et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet **2013**;45(11):1392-8 doi 10.1038/ng.2771.
55. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, *et al.* The long tail of oncogenic drivers in prostate cancer. Nat Genet **2018** doi 10.1038/s41588-018-0078-z.
56. Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature **2011**;473(7347):389-93 doi 10.1038/nature09934.
57. Goldstein AS, Zong Y, Witte ON. A two-step toward personalized therapies for prostate cancer. Sci Transl Med **2011**;3(72):72ps7 doi 10.1126/scitranslmed.3002169.
58. Slovin SF, Kelly WK, Wilton A, Kattan M, Myskowski P, Mendelsohn J, *et al.* Anti-epidermal growth factor receptor monoclonal antibody cetuximab plus Doxorubicin in the treatment of metastatic castration-resistant prostate cancer. Clinical genitourinary cancer **2009**;7(3):E77-82 doi 10.3816/CGC.2009.n.028.
59. Nabhan C, Lestingi TM, Galvez A, Tolzien K, Kelby SK, Tsarwhas D, *et al.* Erlotinib has moderate single-agent activity in chemotherapy-naive castration-resistant prostate cancer: final results of a phase II trial. Urology **2009**;74(3):665-71 doi 10.1016/j.urology.2009.05.016.
60. Pezaro C, Rosenthal MA, Gurney H, Davis ID, Underhill C, Boyer MJ, *et al.* An open-label, single-arm phase two trial of gefitinib in patients with advanced or metastatic castration-resistant prostate cancer. American journal of clinical oncology **2009**;32(4):338-41 doi 10.1097/COC.0b013e31818b946b.
61. Adams BD, Parsons C, Walker L, Zhang WC, Slack FJ. Targeting noncoding RNAs in disease. The Journal of clinical investigation **2017**;127(3):761-71.
62. Knutson SK, Wigle TJ, Warholic NM, Sneeringer CJ, Allain CJ, Klaus CR, *et al.* A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. Nature chemical biology **2012**;8(11):890-6 doi 10.1038/nchembio.1084.
63. McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature **2012**;492(7427):108-12 doi 10.1038/nature11606.

64. Ateeq B, Kunju LP, Carskadon SL, Pandey SK, Singh G, Pradeep I, *et al.* Molecular profiling of ETS and non-ETS aberrations in prostate cancer patients from northern India. Prostate **2015**;Jul;75 (10)(10):1051-62. doi 10.1002/pros.22989.
65. Nunes JJ, Pandey SK, Yadav A, Goel S, Ateeq B. Targeting NF-kappa B Signaling by Artesunate Restores Sensitivity of Castrate-Resistant Prostate Cancer Cells to Antiandrogens. Neoplasia **2017**;19(4):333-45 doi 10.1016/j.neo.2017.02.002.
66. Mira E, Lacalle RA, Gomez-Mouton C, Leonardo E, Manes S. Quantitative determination of tumor cell intravasation in a real-time polymerase chain reaction-based assay. Clinical & experimental metastasis **2002**;19(4):313-8.
67. Brenner JC, Ateeq B, Li Y, Yocum AK, Cao Q, Asangani IA, *et al.* Mechanistic Rationale for Inhibition of Poly(ADP-Ribose) Polymerase in ETS Gene Fusion-Positive Prostate Cancer (vol 19, pg 664, 2011). Cancer Cell **2013**;23(4):557- doi 10.1016/j.ccr.2013.04.005.
68. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. Journal of the American statistical association **1979**;74(368):829-36.
69. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PloS one **2010**;5(11):e13984.
70. Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, *et al.* In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes & development **2003**;17(10):1253-70.
71. Warrick JI, Tomlins SA, Carskadon SL, Young AM, Siddiqui J, Wei JT, *et al.* Evaluation of tissue PCA3 expression in prostate cancer by RNA in situ hybridization—a correlative study with urine PCA3 and TMPRSS2-ERG. Modern Pathology **2014**;27(4):609.
72. Bewick V, Cheek L, Ball J. Statistics review 8: Qualitative data–tests of association. Critical care **2003**;8(1):46.
73. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, *et al.* A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Briefings in bioinformatics **2013**;14(6):671-83 doi 10.1093/bib/bbs046.
74. Efron B. Logistic regression, survival analysis, and the Kaplan-Meier curve. Journal of the American statistical Association **1988**;83(402):414-25.

Figure 1

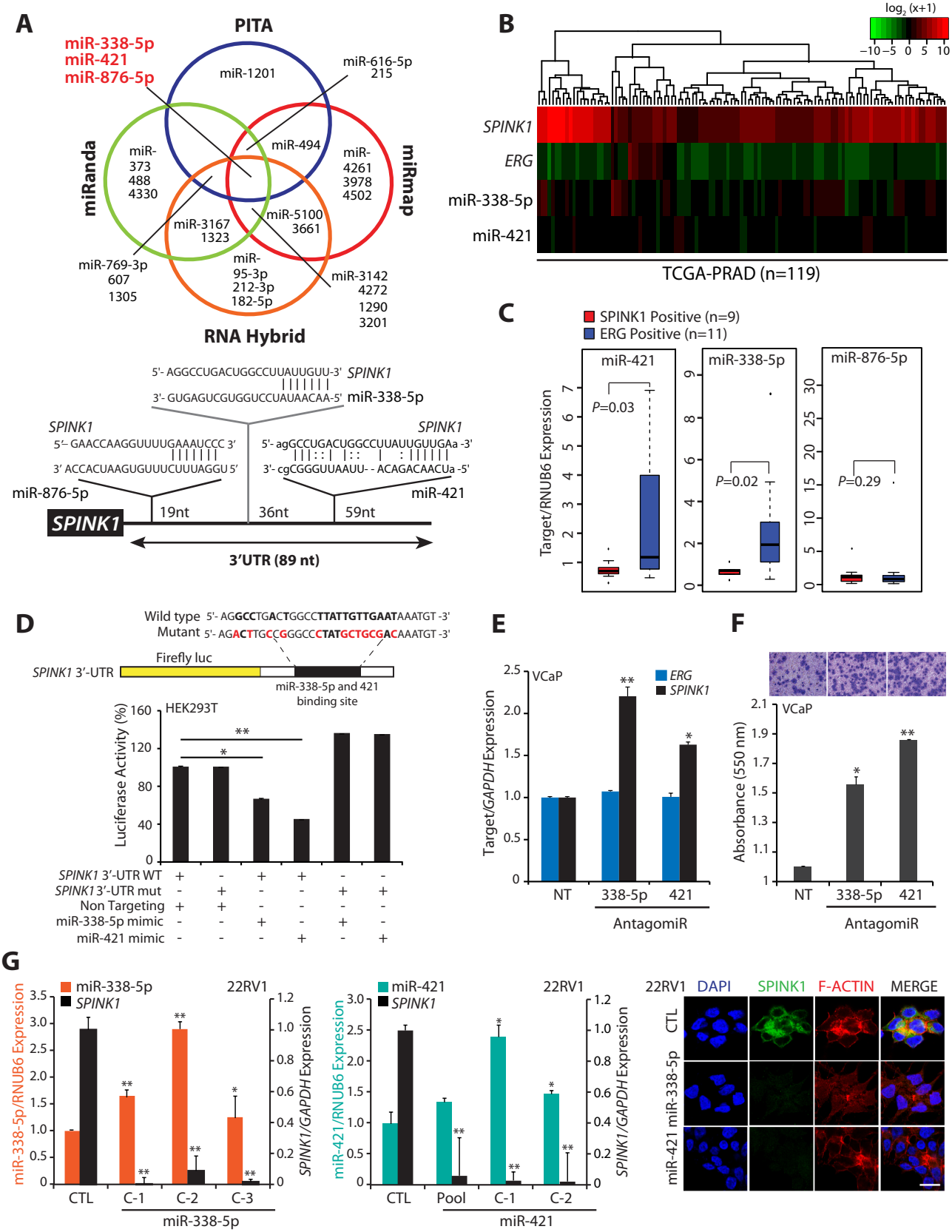

Figure 1. MiR-338-5p and miR-421 are differentially expressed in SPINK1+/ERG-fusion-negative prostate cancer.
(**A**) Venn diagram displaying miRNAs computationally predicted to target *SPINK1* by PITA, miRmap, miRanda and RNAHybrid (top panel). Schematic of predicted miR-338-5p, miR-421 and miR-876-5p binding sites on the 3'-UTR of *SPINK1* (bottom panel). (**B**) Heatmap depicting miR-338-5p and miR-421 expression in the SPINK1+/ERG-negative patients' (n=119) in TCGA-PRAD dataset. Shades of red and green represent log2 (x+1), where x represents the gene expression value. (**C**) Taqman assay showing relative expression for miR-338-5p, miR-421 and miR-876-5p in SPINK1+ and ERG+ PCa patients' specimens (n=20). Data represents normalized expression values with respect to RNUB6 control. (**D**) Schematic of luciferase reporter construct with the wild-type or mutated (altered residues in red) SPINK1 3' untranslated region (3'UTR) downstream of the Firefly luciferase reporter gene (top panel). Luciferase reporter activity in HEK293T cells co-transfected with wild-type or mutant 3'-UTR *SPINK1* constructs with mimics for miR-338-5p or miR-421. (**E**) QPCR data showing *SPINK1* and *ERG* expression in VCaP cells transfected with antagomiRs as indicated (n=3 biologically independent samples; data represent mean ± SEM). (**F**) Boyden chamber Matrigel invasion assay using same cells as in (C). (**G**) QPCR analysis demonstrating *SPINK1* and miRNAs expression in stable 22RV1-miR-338-5p (left panel) and 22RV1-miR-421 cells (middle panel) (n=3 biologically independent samples; data represent mean ± SEM). Immunostaining for SPINK1 (right panel). Scale bar represents 20µm.Representative fields of the invaded cells are shown in the inset. For all panels $*P \leq 0.05$ and $**P \leq 0.001$ using two-tailed unpaired Student's *t* test.

Figure 2

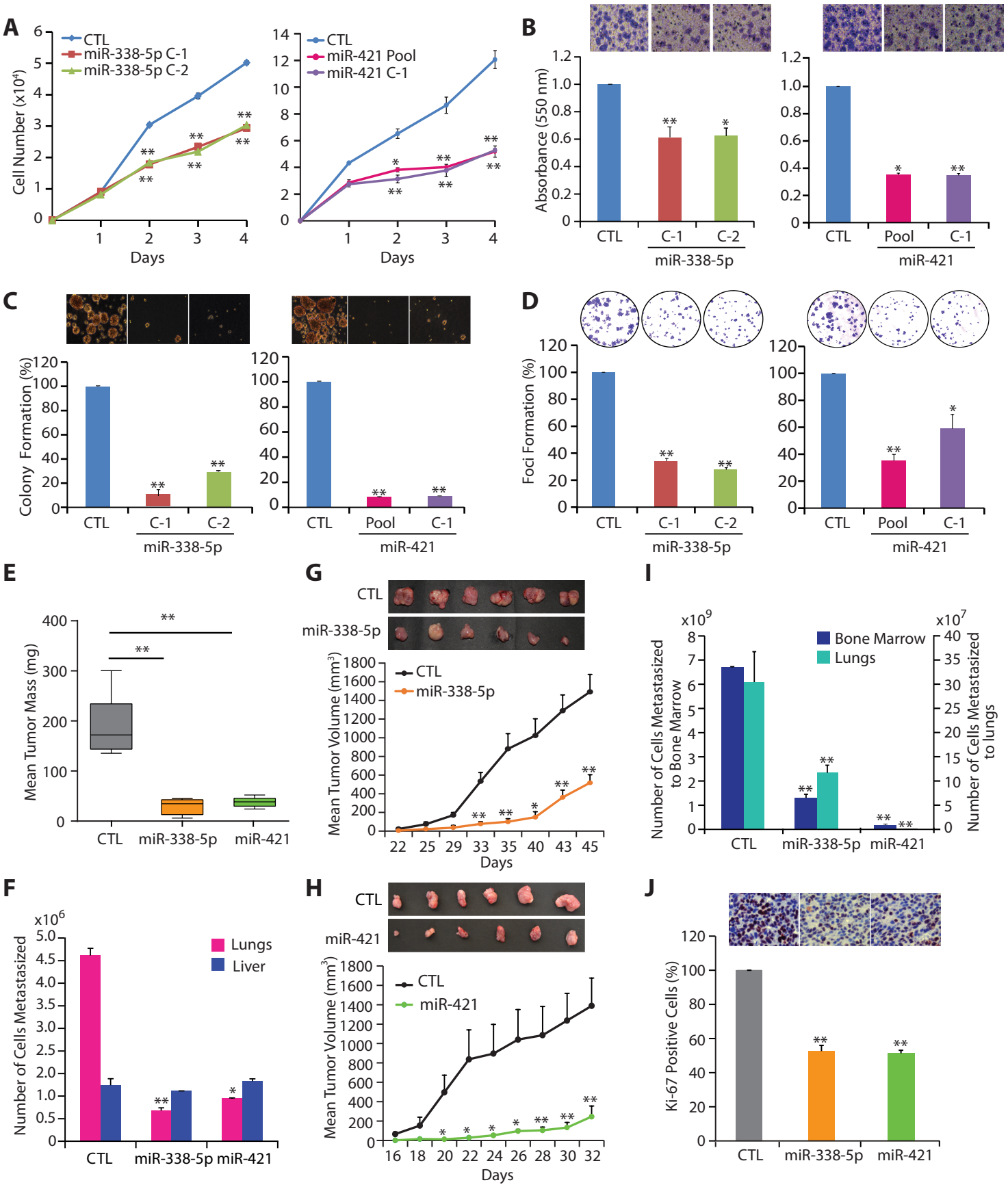

Figure 2. MiR-338-5p and miR-421 abrogates oncogenic properties of SPINK1-positive prostate cancer cells.
(**A**) Cell proliferation assay using 22RV1-miR-338-5p, 22RV1-miR-421 and 22RV1-CTL cells at the indicated time points. (**B**) Boyden chamber Matrigel invasion assay using same cells as in (A). Representative fields with invaded cells are shown in the inset (n=3 biologically independent samples; data represent mean ± SEM). (**C**) Soft agar assay for anchorage-independent growth using same cells as in (A). Representative soft agar colonies are shown in the inset (n=3 biologically independent samples; data represent mean ± SEM). (**D**) Foci formation assay using same cells as in (A). Representative images depicting foci are shown in the inset (n=3 biologically independent samples; data represent mean ± SEM). (**E**) Mean weight of the tumor mass (n=6) excised from the Chick CAM assay as indicated. (**F**) Human-specific *Alu* PCR quantification of metastasized human cells using genomic DNA extracted from the chick embryos' lungs and liver of CAM assay. (**G**) Mean tumor growth in NOD/SCID mice (n=8) subcutaneously implanted with stable 22RV1-miR-338-5p and 22RV1-CTL cells. (**H**) Same as (G), except stable 22RV1-miR-421 cells were implanted. (**I**) Same as (F), except genomic DNA extracted from the lung and bone marrow of the xenografted mice. (**J**) Ki-67 staining of tumor specimens excised from the xenografted mice. Ki-67 positive cells were counted from 10 independent fields for each tumor section. Data represent mean ± SEM. **P*≤ 0.05 and ***P*≤ 0.001 using two-tailed unpaired Student's *t* test.

Figure 3

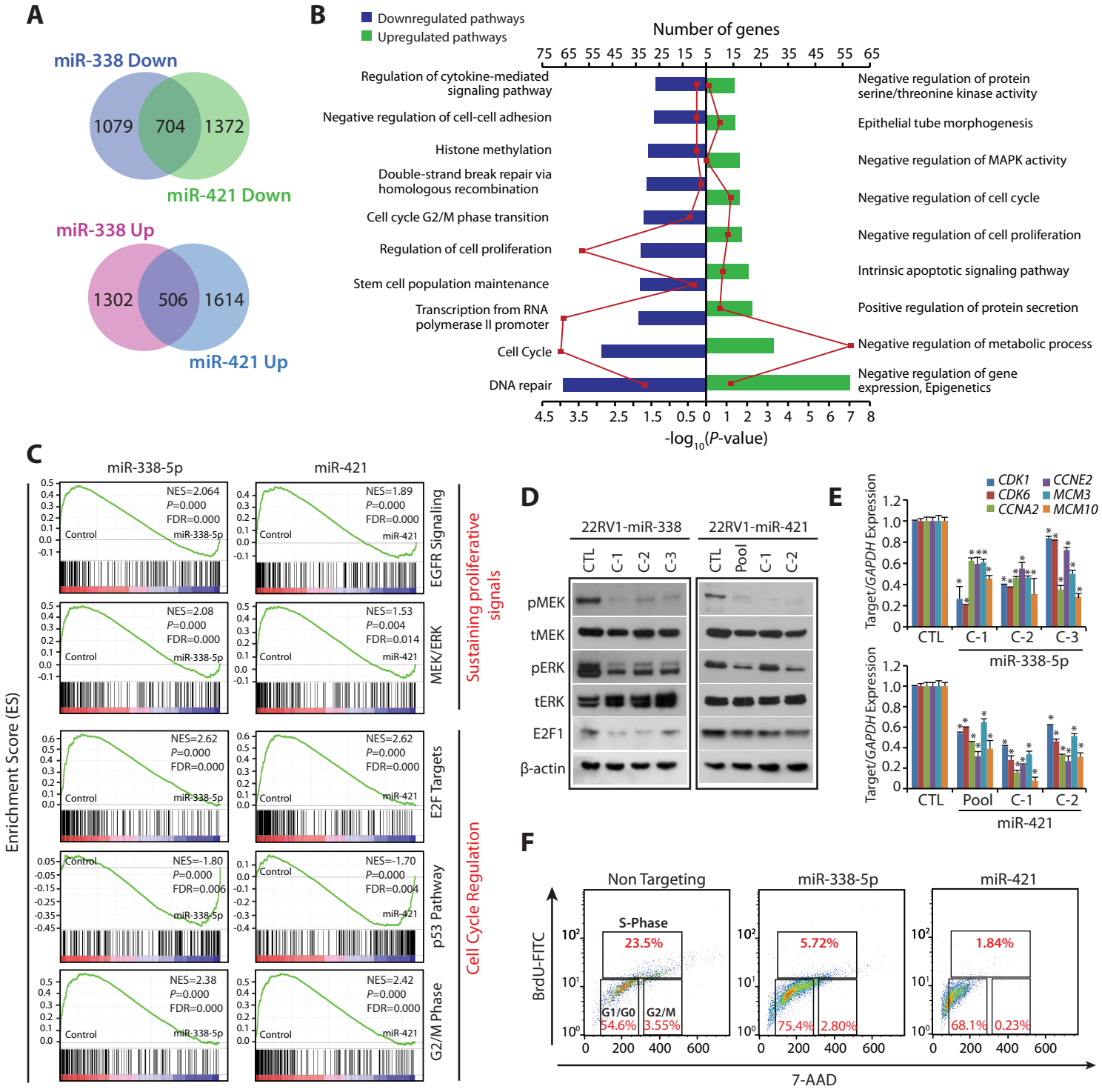

Figure 3. MiR-338-5p and miR-421 overexpression suppress oncogenic pathways and triggers G1/S arrest.
(**A**) Gene expression profiling data showing overlap of downregulated (upper panel) and upregulated genes (lower panel) in stable 22RV1-miR-338-5p and 22RV1-miR-421 cells relative to 22RV1-CTL cells (n=3 biologically independent samples). (**B**) Same as in (A), except DAVID analysis showing various downregulated (left) and upregulated (right) pathways. Bars represent $-\log_{10}$ of *P* values and frequency polygon (line in red) represents the number of genes. (**C**) Gene Set Enrichment Analysis (GSEA) plots showing various deregulated oncogenic gene signatures with the corresponding statistical metrics in the same cells as in (A). (**D**) Western blot analysis for phosphor (p) and total (t) MEK1/2, ERK1/2 and cell cycle regulator E2F1 levels. β-actin was used as a loading control. (**E**) QPCR analysis showing expression of cell cycle regulators for G1 and S phase as indicated. Expression level for each gene was normalized to *GAPDH* (n=3 biologically independent samples; data represent mean ± SEM). (**F**) BrdU/7-AAD cell cycle analysis for S-phase arrest in 22RV1 cells transfected with miR-338-5p or miR-421 mimics relative to control cells. Data represent mean ± SEM. *$P \leq 0.05$ and **$P \leq 0.001$ using two-tailed unpaired Student's *t* test.

Figure 4

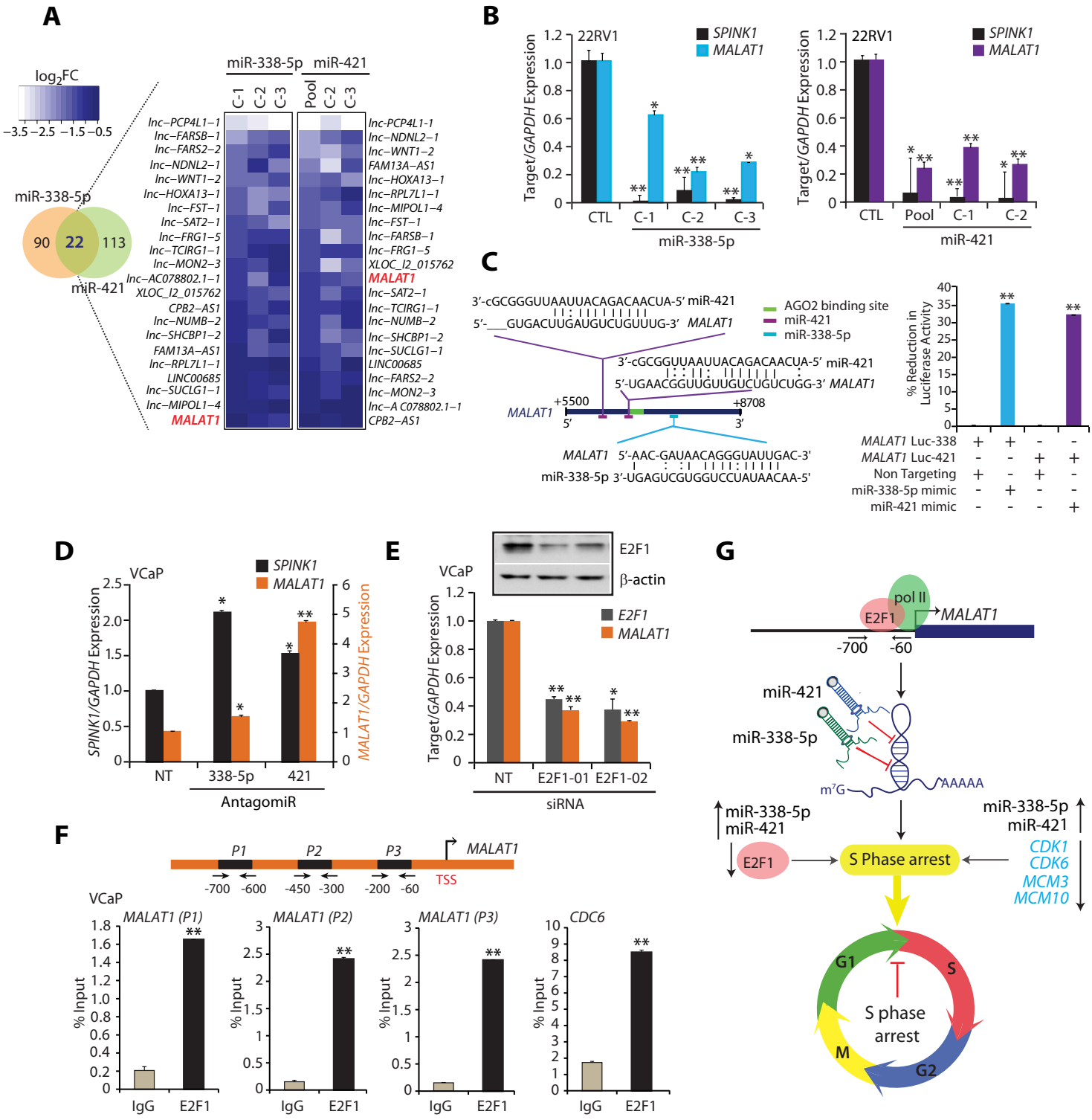

Figure 4. MiR-338-5p and miR-421 targets oncogenic long non-coding-RNA *MALAT1*.
(**A**) Heatmap showing lncRNAs downregulated in 22RV1-miR-338-5p and 22RV1-miR-421 cells relative to 22RV1-CTL cells. Shades of blue represent $\log_2$ fold-change in expression (n=3 biologically independent samples). (**B**) QPCR data showing expression of *MALAT1* and *SPINK1* using the same cells as in (A). (**C**) Schema showing miR-338-5p and miR-421 putative binding sites on the *MALAT1* transcript. Luciferase reporter activity in HEK293T cells co-transfected with *MALAT1* 3′ end constructs (*MALAT1*-Luc-338, *MALAT1*-Luc-421) and mimics for miR-338-5p or miR-421. (**D**) QPCR data show expression of *MALAT1* and *SPINK1* in VCaP cells transfected with anatogmiR as indicated. (**E**) Relative expression of *E2F1* and *MALAT1* in E2F1 silenced- and control VCaP cells. Western blot analysis for E2F1 using same cells. (**F**) Schema showing positioning of the ChIP primers on the *MALAT1* promoter. ChIP-qPCR analysis for E2F1 occupancy on the *MALAT1* promoter in VCaP cells. *CDC6* promoter was used as a positive control. (**G**) Schematic depicts plausible transcriptional regulation of *MALAT1* by E2F1, the master regulator of cell cycle, and post-transcriptional regulation by miR-338-5p and miR-421. In Fig panels (B-F) biologically independent samples were used (n=3); data represent mean ± SEM.*$P \leq 0.05$ and **$P \leq 0.001$ using two-tailed unpaired Student's *t* test.

Figure 5

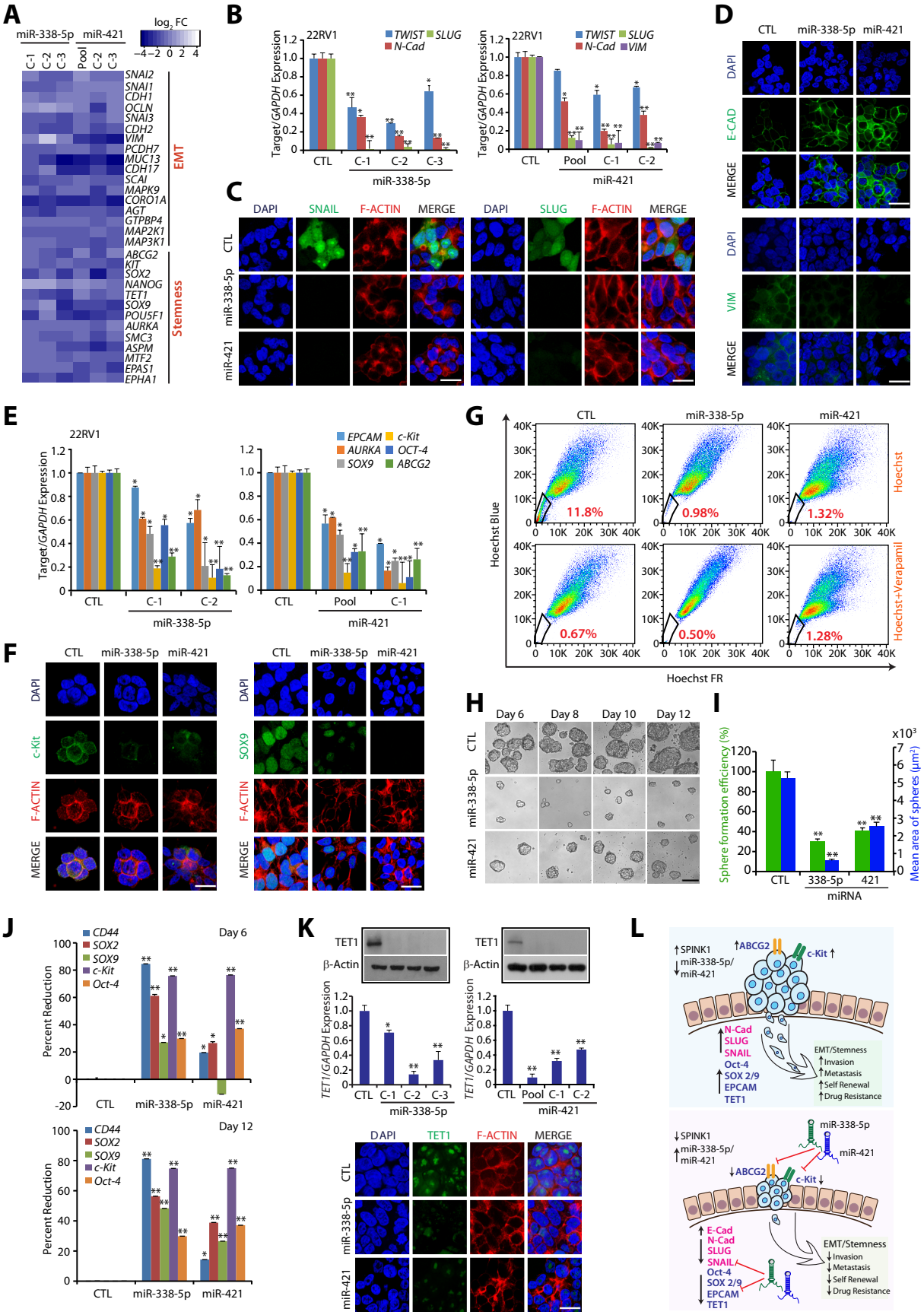

Figure 5. MiR-338-5p and miR-421 overexpression attenuates EMT and Stemness.

(**A**) Heatmap depicting change in the expression of EMT and pluripotency markers in 22RV1-miR-338-5p and 22RV1-miR-421 cells. Shades of blue represents $\log_2$ fold-change in gene expression (n=3 biologically independent samples). (**B**) QPCR analysis depicts expression of EMT markers in 22RV1-miR-338-5p, 22RV1-miR-421 and control cells. Expression for each gene was normalized to *GAPDH*. (**C**) Immunostaining showing SLUG and SNAIL expression in the same cells as in (B). (**D**) Same cells as in (B), except immunostained for E-cadherin and Vimentin. (**E**) Same cells as in (B), except qPCR analysis for stem cell markers. (**F**) Same cells as in (B), except immunostained for c-Kit and SOX-9. (**G**) Hoechst 33342 staining for side population (SP) analysis using same cells as in (B). Percentages of SP were analyzed using the blue and far red filters, gated regions as indicated (red) in each panel. (**H**) Phase contrast microscope images for the prostatospheres using same cells as in (B). (**I**) Bar plot depicts percent sphere formation efficiency and mean area of the prostatosphere. Scale bar 100μm. (**J**) QPCR analysis of stem cell markers using the prostatosphere as in (H). (**K**) Expression of TET1 by qPCR, Western blot and immunostaining using same cells as in (B). (**L**) Schematic describing the role of miR-338-5p and miR-421 in regulating EMT, cancer stemness and drug resistance in SPINK1+ cancer. Scale bar for panels (C), (D), (F) and (K) represents 20μm. In the panels (B), (E), (I-K) biologically independent samples were used (n=3); data represents mean ± SEM *$P \leq 0.05$ and **$P \leq 0.001$ using two-tailed unpaired Student's *t* test.

Figure 6

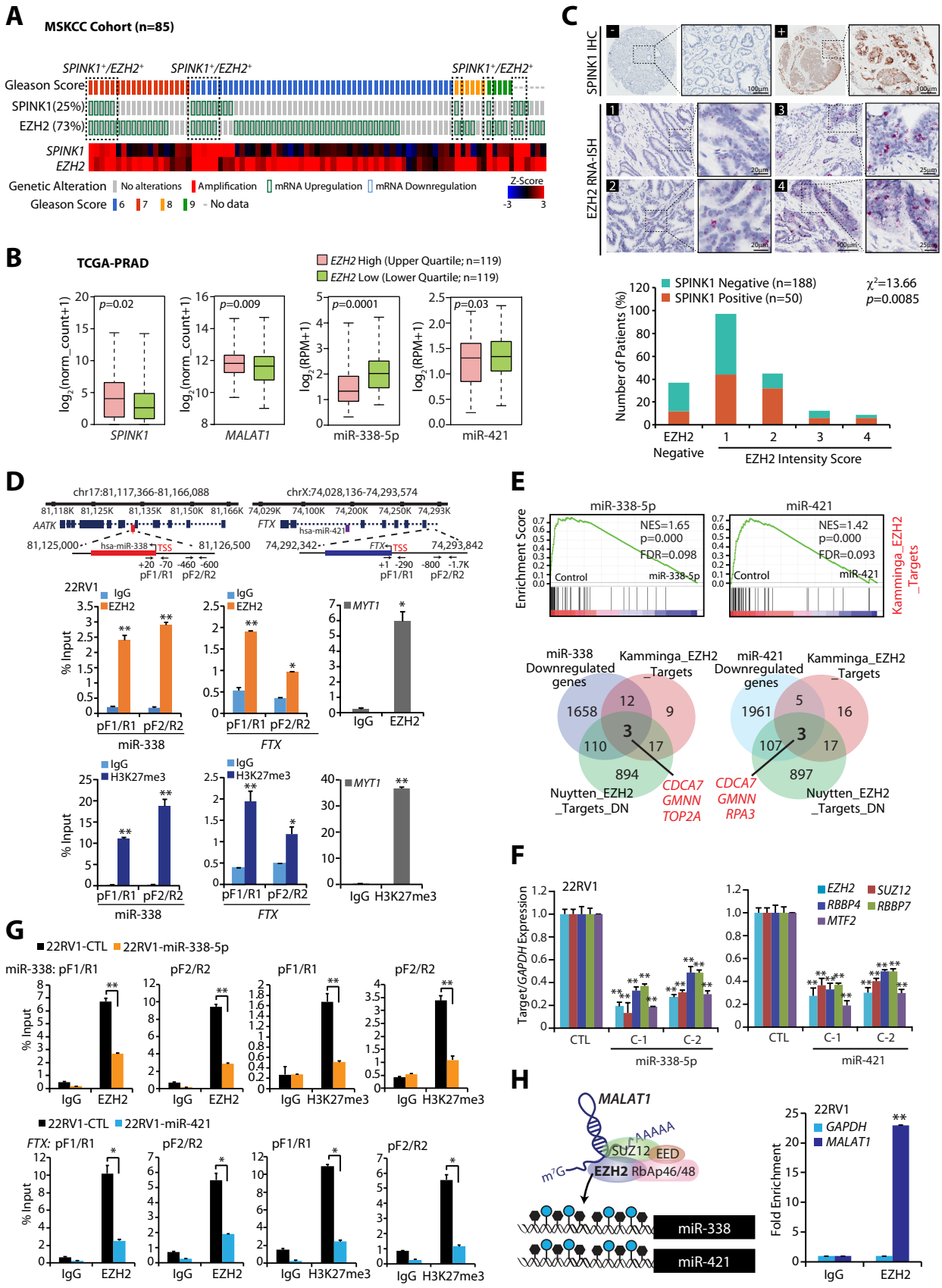

Figure 6. EZH2-mediated epigenetic silencing of miR-338-5p and miR-421 in SPINK1 positive prostate cancer.

(**A**) OncoPrint depicting mRNA upregulation of *EZH2* and *SPINK1* in MSKCC cohort using cBioportal. Each patient sample is represented by a bar and shades of color indicate specific genomic alteration and Gleason Score. At the lower panel shades of blue and red represents Z-score normalized expression for *EZH2* and *SPINK1*. (**B**) Box plot depicting *SPINK1*, *MALAT1*, miR-338-5p and miR-421 expression in *EZH2* high (n=119) and *EZH2* low (n=119) in PCa patients from TCGA-PRAD cohort. *P*-values were calculated using two-tailed unpaired Student's *t* test. (**C**) Representative micrographs depicting PCa tissue microarray (TMA) cores (n=238) stained for SPINK1 by immunohistochemistry (IHC) and EZH2 by RNA in-situ hybridization (RNA-ISH). Top panel represents SPINK1 IHC in SPINK1 negative (–) and SPINK1 positive (+) patients. RNA-ISH intensity score for EZH2 expression was assigned on a scale of 0 to 4 according to visual criteria for the presence of transcript at 40X magnification. Bar plot show EZH2 expression in the SPINK1-negative and SPINK1+ patient specimens. *P*-value for Chi-square test is indicated. (**D**) Genomic location for EZH2 binding sites on the miR-338 and *FTX* promoters and location of ChIP primers (top panel). ChIP data showing EZH2 occupancy and H3K27me3 marks on the miR-338, *FTX* promoters, and *MYT1* used as positive control. (**E**) GSEA plots showing the enrichment of EZH2 interacting partners (Kamminga) in 22RV1-miR-338-5p and 22RV1-miR-421 cells (top panel). Venn diagram showing overlap of genes downregulated upon miR-338-5p and miR-421 overexpression with EZH2 partners (Kamminga) and EZH2 downregulated genes (Nuytten). (**F**) QPCR data showing expression of *EZH2* and its interacting partners in the same cells as indicated. (**G**) ChIP-qPCR for the presence of *EZH2* and H3K27me3 marks on miR-338 and *FTX* promoters in stable 22RV1-miR-338-5p, 22RV1-miR-421 and 22RV1-CTL cells. (**H**) RNA-Immunoprecipitation (RIP) assay using EZH2 and IgG antibodies in 22RV1 cells. QPCR data show enrichment of lncRNA *MALAT1* in EZH2 pulldown over input. *GAPDH* was used as a negative control. Biologically independent samples (n=3) were used in panels (D), (F), (G), and (H); data represent mean ± SEM. **P*≤ 0.05 and ***P* ≤ 0.001 using two-tailed unpaired Student's *t* test.

Figure 7

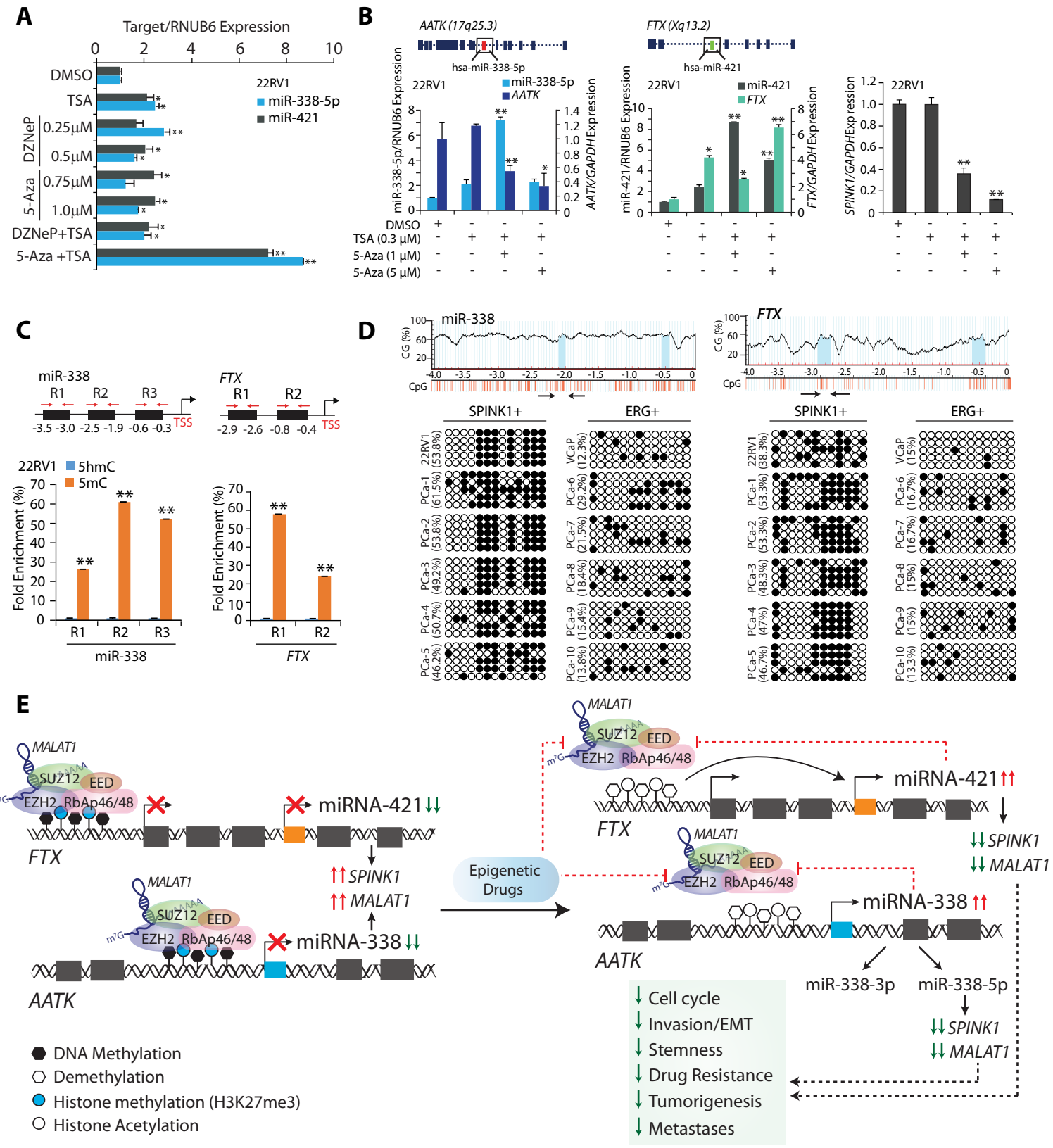

Figure 7: Epigenetic drugs ablate EZH2-mediated silencing of the miR-338-5p and miR-421.
(**A**) TaqMan assay for miR-338-5p and miR-421 expression in 22RV1 cells treated with different combination of epigenetic drugs. (**B**) QPCR showing relative expression of miR-338-5p, miR-421, *AATK*, *FTX* and *SPINK1* in 22RV1 cells treated with 5-Aza or TSA as indicated. (**C**) MeDIP-qPCR showing fold enrichment of 5-mC over 5-hmC in 22RV1 cells as indicated. (**D**) Bisulfite–sequencing showing CpG methylation marks on the region upstream of miR-338-5p (left) and *FTX* (right) in 22RV1, VCaP cells and patients' tumor specimens (PCa-1 to 5 are SPINK1 positive and PCa-6 to 10 are ERG fusion positive). PCR amplified regions are denoted by arrows. Data represents DNA sequence obtained from five independent clones. Hollow circles represent non-methylated CpG dinucleotides, whereas black solid circles show methylated-CpG sites. (**E**) Illustration depicting the molecular mechanism involved in EZH2-mediated epigenetic silencing of miR-338-5p and miR-421 in SPINK1-positive prostate cancer. In panels (A), (B) and (C) biologically independent samples (n=3) were used; data represent mean ± SEM. *$P \leq 0.05$ and **$P \leq 0.001$ using two-tailed unpaired Student's *t* test.